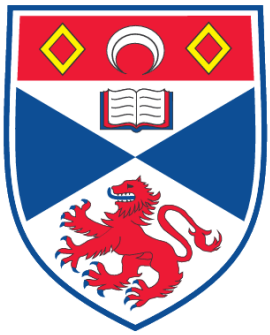
TOOLS FOR PROBING 2A SEQUENCE SPACE

Helena Escuin Ordinas

A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews



2008

Tools for Probing 2A Sequence Space

By
Helena Escuin Ordinas
School of Biology
University of St. Andrews

A thesis submitted for the Degree of Doctor of Philosophy
at the University of St. Andrews
June 2008

ABSTRACT

Foot-and-mouth disease virus (FMDV) 2A is an oligopeptide composed of only 18 amino acids that can mediate a co-translational cleavage at its own C-terminus. It has been observed that 2A sequences do not show cleavage activity within bacterial organisms. Why 2A lacks activity in a prokaryotic organism such as *E.coli* is unclear. A series of plasmids designed to provide a phenotypic screen for 2A-mediated cleavage (in prokaryotes) were developed. Even though no active 2A sequences were found in bacteria, this system can easily be adapted to eukaryotic cells and will also be very useful in mutagenic studies on 2A sequences. Furthermore, $2A_{FMDV}$ has been used in the construction of a reporter of stress in the cell. This may allow us to open a new approach in the use of 2A oligopeptide, which had already been widely used to co-express genes of interest with reporter proteins, in biotechnology and gene therapy.

Theiler's murine encephalomyelitis cardiovirus (TMEV) 2A has the same role as in FMDV but is 150 aa in length instead of the 18 aa in FMDV. It also presents the same C-terminal motif but what is the function of the remaining -85% of the cardiovirus 2A sequence remains a mystery. To this end we have produced antibodies against TMEV-2A, to study the role of $2A_{TMEV}$ within the cell.

Database searches probing for 2A's C-terminal conserved motif (-DxExNPGP-) has identified many 2A-like sequences, not only within picornaviruses but also in trypanosomes, insect and cellular genes. These remarkable findings indicate that the control of protein synthesis by 2A is not solely confined to the *Picornaviridae*. Bioinformatics analyses of all the known 2A-like sequences, comparing all the different upstream sequences, show a clear pattern on the organization of residues in the upstream region.

The discovery of this 2A oligopeptide has led to a breakthrough in protein co-expression technology. It has been used as a highly effective new tool for the co-expression of multiple proteins from a single ORF in plant biotechnology and also gene therapy applications. Although we have gained substantial insights into the general features and biological significance of this process, a great deal still needs to be uncovered about the structural and mechanistic details of this unique mechanism of action.

Declarations

i) I, Helena Escuin Ordinas, hereby certify that this thesis, which is approximately 40,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

ii) I was admitted as a research student in September 2004 as a candidate for the degree of Doctor of Philosophy in Molecular Virology; the higher study for which this is a record was carried out in the University of St Andrews between 2004 and 2008.

Date ……. Signature of candidate ……

iii) 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.

Date …….. Signature of supervisor ……

Unrestricted copyright declaration

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 that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration.

Date ……. Signature of candidate ……

Acknowledgments

I would like to thank all the people who have given me their support and the necessary strength during the process of completing this project.

Special thanks to my supervisor Martin Ryan for his guidance, valuable advice and encouragement. The completion of this thesis would not have been possible without him. Thanks also to the members of Ryan's lab, always ready to help and answer all my infinite questions.

Special thanks also to Garry and Jane, my Scottish family—thanks for all your support and for making me feel at home! I have to mention Jane's carrot cake, the best in the world!

My friends deserve a special mention too for giving me all the reassurance needed and for always believing in me. Thanks to my three little grasshoppers, Kacie, Claire and Mandy, for always being there and for all the wonderful moments spent together. Xenilin 'danke shoen' for always listening to me and Sandrita, my amazing office neighbour—always so supportive! Sylvie for all the 'shakira' moments and Neil, thanks for your encouragement at any time of the day; thanks to Andy and Elisa for those relaxing lunch, coffee and yoga times together; and thanks to Pitu, Joe and Carlitos (the three other musketeers) for always making me smile no matter what.

Sara, Rakel (ewes), Almusan y Noe (minime) por vuestros buenos consejos y todo el cariño que me dais. Gregorio por todos eso momentos de risas juntos, tan necesarios durante la tesis. A la Topita, Markius i Trini, per fer-me saber que sempre puc contar amb vosaltres, encara que la distancia ens separi.

Finally, I would like to give immense thanks to my family who have been an invaluable source of strength. Moritz y Suspi, gracias sisters por estar ahi en los buenos y, tambien, malos momentos. Mami y papi (super cluecos) gracias por no dejar de creer en mi, por vuestro apoyo y amor incondicional. Me siento muy afortunada de teneros.

Contents

1. INTRODUCTION

1.1. Viruses

Viruses are obligate intracellular parasites which consist of a genetic material (RNA or DNA) encapsidated within a protein coat which may also be surrounded by a lipid membrane.

The concept of infectious particles smaller than a bacterium, such as viruses, was developed in 1892 by Dimitri Ivanosfsky (1864-1920), who found such particles in the sap of mosaic tobacco plants (reviewed by Horzinek, 1997 and Lustig & Levine, 1992). These studies were followed by Martinus Beijerinck (1851-1931) and lead him to propound a new concept: a filterable agent too small to observe in the light microscope but able to cause disease by multiplying in living cells.

In 1898 Friedrich Loeffler and Paul Frosh isolated the first infectious filterable particle from animals, foot and mouth disease virus (FMDV). In 1901 Walter Reed described the first human virus, the causative agent of yellow fever.

Further studies not only helped in the description of new viruses and their properties but also in the successful production of vaccines to prevent specific diseases. From the 1960's virologists began to use viruses as tools to gain an in-depth knowledge and understanding of life processes, from the replication of nucleic acid to protein synthesis and transport.

Viruses have been classified in different ways. These classifications are based on phenotypic characteristics such as morphology, mode of replication, cell host, disease that the virus causes and also the type of nucleic acid that it carries. The Baltimore Classification separates viruses into several groups, based on their mode of replication and type of genome. Other classifications use chemical and physical characteristics, such as type of nucleic acid, symmetry and presence/absence of envelope, while others are based on the host organism that the virus attacks.

1.2. Positive stranded RNA viruses

Positive stranded RNA viruses comprise over one-third of all virus genera and include pathogens such as poliovirus, hepatitis C virus, severe acute coronavirus syndrome SARS, winter vomiting calicivirus, among others. These viruses have a relatively small genome that can, directly, be translated in the first step of infection without having to be transcribed first. Their RNA acts like cellular mRNA and can be translated by the host's cells machinery. They need to encode their own RNA-dependent RNA-polymerase to replicate their RNA and they use different strategies to express their proteins. These expression strategies will be discussed in section 1.4.

1.3. Picornaviruses

Picornaviruses are non-enveloped positive-stranded RNA viruses, which encode a single, long, open reading frame (ORF) comprising a polyprotein of ~225 kDa. The *Picornaviridae* is one of the largest families of human and animal pathogens and contains many important human and animal viruses, including: poliovirus, hepatitis A virus and foot-and-mouth disease virus.

The *Picornaviridae* consists of 9 genera: ***Enterovirus*** (*Poliovirus, Human enterovirus A, Human enterovirus B, Human enterovirus C, Human enterovirus D, Simian enterovirus A, Bovine enterovirus and Porcine enterovirus B.*)***, Rhinovirus*** *(human rhinovirus1-2),* ***Cardiovirus*** *(Encephalomyocarditis virus and Theilovirus),* ***Aphthovirus*** *(Foot-and-mouth disease virus, Equine rhinitis A virus, Bovine rhinovirus 2),* ***Hepatovirus*** *(Hepatitis A virus and Avian encephalomyelitis-like virus),* ***Parechovirus*** *(Human parechovirus and Ljungan virus),* ***Erbovirus*** *(Equine rhinitis B virus),* ***Kobuvirus*** *(Aichi virus and Bovine Kobuvirus)* and ***Teschovirus*** *(Porcine Teschovirus).* Three new genera have been proposed and provisionally named: ***Sapelovirus*** *(Porcine enterovirus A, SV2-like virus, Duck Picornavirus TW90A),* ***Senecavirus*** *(Seneca valley virus)* and ***Tremovirus*** *(Avian encephalomyelitis virus, which now belongs to Hepatovirus)*. In the near future, the genera Rhinovirus will be removed, its two members placed in the genus Enterovirus. In addition, two new species have recently been identified: *duck hepatitis virus 1* and *seal picornavius 1*. These will form two novel genera. The family *Picornaviridae* will thus consist of 13 genera and 28 species (Table 1).

Family Picornaviridae	
Genus	***Species***
Enterovirus	*Poliovirus, Human enterovirus A, Human enterovirus B, Human enterovirus C, Human enterovirus D, Simian enterovirus A, Bovine enterovirus and Porcine enterovirus B*
Rhinovirus	*Human rhinovirus 1-2*
Cardiovirus	*Encephalomyocarditis virus and Theilovirus*
Aphthovirus	*Foot-and-mouth disease virus and Equine rhinitis A virus Bovine rhinovirus*
Hepatovirus	*Hepatitis A virus and Avian encephalomyelitis-like virus*
Parechovirus	*Human parechovirus and Ljungan virus*
Erbovirus	*Equine rhinitis B virus*
Kobuvirus	*Aichi virus and Bovine Kobuvirus*
Teschovirus	*Porcine Teschovirus*
Sapelovirus	*Porcine enterovirus A, SV2-like virus,Duck Picornavirus TW90A*
Senecavirus	*Seneca valley virus*
Tremovirus	*Avian encephalomyelitis virus, which now belongs to Hepatovirus*
New unnamed genera:	*duck hepatitis virus 1 and seal picornavirus 1*

Table 1. The *Picornaviridae*. Table showing the classification of all the genera and species within the family.

1.3.1 Picornavirus genera

1.3.1.1 Enteroviruses

Enteroviruses have been implicated in chronic as well as acute diseases. These chronic diseases include dermatomyositis, polymyositis, dilated cardiomyopathy and diabetes mellitus.

Poliovirus is a well-known virus within this family, which causes poliomyelitis, an acute viral infectious disease that spreads from person to person *via* the faecal-oral route.

Most of these virus infections are asymptomatic, although, in a few cases, the virus can enter the central nervous system leading to acute flaccid paralysis (figure 1). Poliovirus is a widely studied virus, whose genome (~7400 nucleotides) functions as single genome sized RNA and is representative of most positive-sense RNA viruses (shown in figure 2). Extensive studies of this virus allied to successful vaccine production, and vaccination program have led to the almost complete eradication of polio. This will be the second virus eradicated from the world, smallpox being the first.

Figure 1. Poliomyelitis, flaccid muscular paralysis disease cause by poliovirus.

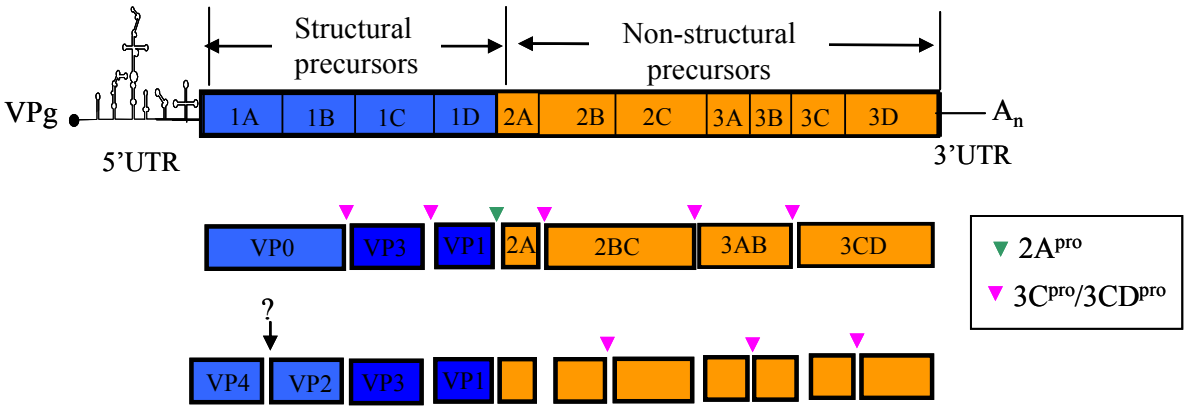


Figure 2. Schematic diagram of the poliovirus genome and proteolytic processing of its polyprotein. Poliovirus has a single-stranded, positive-sense RNA genome, which encodes a single polyprotein. Its genome is covalently liked at the 5′end to a viral protein encoded by 3B (VPg). The 5' untranslated region (UTR) contains a 5'terminal stem-loop, cloverleaf-like structure, and a type I or II internal ribosome entry site (IRES). The 3'UTR is poly-adenylated. Proteolytic processing is mediated by the virally-encoded proteinases $2A^{pro}$ and $3C^{pro}/3CD^{pro}$ to ultimately generate eleven mature viral proteins. Three intermediate products are produced (2BC, 3AB, and 3CD).

1.3.1.2 Rhinovirus

Rhinoviruses, aetiological agents of the common cold, are the most commonly isolated viruses from individuals experiencing mild upper respiratory illnesses (Figure 3). In contrast to enteroviruses, rhinoviruses do not replicate in the intestinal tract. Constant efforts towards producing effective vaccines have been unsuccessful and remain as an important goal. This genus contains 99 human rhinovirus serotypes classified into 2 species, *Human rhinovirus A and B*. This genus will disappear soon, its 2 species moved into the enterovirus genera (http://www.picornaviridae.com/).

Figure 3. One of the classic symptoms caused by rhinovirus in the common cold.

1.3.1.3 Cardiovirus

This genus consists of two species: encephalomyocarditis virus (EMCV), which is represented by a single serotype of the same name, and theilovirus, which comprises 4 different serotypes: Theiler´s murine encephalomyelitis virus (TMEV), Vilyuisk human encephalomyelitis virus (VHEV), Theiler's-like virus (TLV) isolated from rats and human pathogen Saffold virus (SAF-V). These are distinguished from other genera by special features of their genome organization (presence of a poly(C) tract - except TMEV - and the dissociability of their capsid at pH 5-7, among others).

Several strains of EMCV have been determined: encephalomyocarditis virus (strain EMC-B nondiabetogenic), Encephalomyocarditis virus (strain EMC-D diabetogenic), Maus-Elberfeldvirus, Mengo virus and Porcine encephalomyocarditis virus (http://www.picornaviridae.com/).

TMEV strains were first isolated by Max Theiler at the Rockefeller Foundation during the 1930's, (Lipton, 1975) from the central nervous system of paralysed mice and later from the intestine of apparently uninfected mice.

These investigations demonstrated that the virus caused widespread, asymptomatic, enteric infection and chronic progressive demyelination in mice. Such symptoms were later identified as being similar to those observed in humans with multiple sclerosis (Lipton, 1975; Roos, 2002).

There are several strains classified on the basis of differences in their biological activities: i) *GDVII sub-group* (GDVII and FA), which are extremely virulent and produce an acute disease that is similar to poliomyelitis; and ii) *TO sub-group* (DA, BeAn, TO, WW, Yale, etc), which produce a biphasic disease (reviewed by Racaniello, 2001; Roos, 2002).

1.3.1.4 Aphthovirus

Foot-and-mouth disease virus (FMDV) is a member of this genus and is responsible for Foot-and-mouth disease (FMD). This disorder principally affects domesticated and wild cloven-hoofed animals and is known to spread by direct contact between infected and susceptible animals by the airborne route, by animal products such as meat and milk, and mechanical transfer *via* people, wild animals, birds, and by vehicles. It is accepted as a significant epidemic disease that infects primarily cattle, goats, pigs, and sheep and rarely humans (reviewed by: Rossmann, 2002; Agol, 2002).

FMD is one of the most contagious animal diseases, with important economic losses. It is endemic in Asia, Africa, the Middle East and South America. On the domestic front, the 2001 outbreak cost the UK something in the region of £10 billion with the loss of eight million cattle, sheep, pigs and goats (DEFRA, March 2002)

Unfortunately, presently available inactivated vaccines are not entirely effective. Vaccination blocks disease symptoms, making detection of infection difficult, but does not always block transmission of the virus to other animals. Sheep can harbour the virus for several months; cows for up to a year or even longer. Occasional vaccine-linked disease outbreaks occur as a result.

Its genome is unique in that it encodes three VPg genes whereas the other picornaviruses encode only one, and it also has the shortest poly(A) tract compared to other genera. Equine rhinitis A virus (ERAV) also belongs to the aphthoviruses

(Pringle *et al*., 1999) and is the only non-FMDV member of this genus. The genome organisation of this virus possesses two features which distinguish it from FMDV; (i) there is no poly C tract and (ii) ERAV has only one copy of 3B (VPg) (Wutz *et al.,* 1996).

1.3.1.5 Hepatovirus

Hepatitis A is an acute infectious disease caused by Hepatitis A virus (HAV). It is transmitted *via* the faecal-oral route and may be mistaken for flu. Symptoms typically appear 2 to 6 weeks after the start of infection and may return over the following 6-9 months. The most common symptoms are: fatigue, fever, nausea, abdominal pain and diarrhoea.

In contrast with other picornaviruses, Hepatovirus 2A protein does not show any proteolytic function with regards to polyprotein processing. Cleavage between 2A/2B, leading to separation of 5' terminal structural proteins from the non-structural proteins, is achieved by the 3C protease (Jia *et al*., 1993; Martin *et al.,* 1995).

1.3.1.6 Parechovirus

Two species form this genus, Human parechovirus (HePV), which comprises five serotypes (Ito *et al*., 2004; Al-Sunaidi *et al*., 2007), and the recently described Ljungan virus (LV), which may be comprised of two or more serotypes.

Human parechovirus causes mostly mild gastrointestinal or respiratory illness, with a few cases of myocarditis and encephalitis. It commonly infects children between 2 to 5 years old. In contrast, Ljungan virus causes diabetes, neurological disease, myocarditis and intrauterine foetal death in humans (Niklasson *et al.*, 2007). It is a zoonotic virus, which was isolated from bank voles (Niklasson *et al.*, 1999). Its genome contains two 2A proteins in tandem. The N-terminal $2A_1$ has the -DxExNPGP- motif and is related to the type of 2As encoded by aphthoviruses, cardioviruses, erboviruses and teschoviruses. In contrast, the $2A_2$ belongs to the type of 2As encoded by kobuviruses and hepatoviruses (Johansson *et al*., 2002).

1.3.1.7 Erbovirus

Equine rhinitis B virus (ERBV) is the only member of this genus and causes acute upper febrile respiratory disease in horses (Dynon *et al*., 2007). There are two serotypes, ERV-1 and ERV-2, which are closely related to FMDV, although the

symptoms they cause more closely resemble those produced by rhinovirus. ERV-1 proteins are very similar to the FMDV proteins, whereas, most ERV-2 proteins are more closely related to EMCV proteins (Wutz *et al*., 1996).

1.3.1.8 Kobuvirus

This genus consists of 2 species, Aichi virus, which infects humans (Yamashita *et al*., 1998) and Bovine Kobuvirus, which infects cattle (Yamashita *et al*., 2003).

Aichi virus, first recognized in 1989 as the cause of oyster-associated non-bacterial gastroenteritis in humans, has been recently classified into the new Kobuvirus genus (Pham *et al*., 2007)

1.3.1.9 Teschovirus

Teschovirues constitute a recently defined genus within the *Picornaviridae*. It is represented by a single species, Porcine Teschovirus (PTV), which causes porcine enteroviral encephalomyelitis disease. PTV-1 Talfan is the reference strain for this genus, and, as mentioned earlier, its genome contains a significantly shorter IRES compared to the other genera, and requires part of the 5´UTR sequence for its activity (Kaku *et al*., 2002).

1.3.1.10 Proposed novel genera

Sapelovirus consists of three species: Porcine enterovirus A, SV2-like virus, Duck Picornavirus TW90A. The name of this proposed genus derives from **S**imian, **A**vian and **P**orcine **E**ntero-**L**ike viruses.

Senecavirus contains Seneca valley virus (SVV), whose proteins closely resemble some cardiovirus proteins (P1, 2C, $3C^{pro}$ and $3D^{pol}$). However, the 5'UTR, leader, 2A and 3A proteins are very different to all known picornaviruses. SVV-2A protein presents the same conserved motif found in aphthoviruses (Luke *et al*., 2008)

Tremovirus consists of a single species, Avian encephalomyelitis virus (AEV), which currently belongs to hepatovirus.

1.3.1.11 Novel species

Two novel picornaviruses, whose entire genomes have recently been sequenced, have been allocated their own genera.

Duck hepatitis virus 1 (DHV-1) is one of the three viruses causing Duck virus hepatitis, an acute highly contagious disease affecting young ducklings up to the age of 28 days. Its genome features three in-tandem 2A genes. The $2A_1$ and $2A_3$ proteins are aphthovirus-like and human parechovirus-like 2A proteins respectively, whereas $2A_2$ is not related to any known picornavirus protein. These 2A proteins are only 12 amino acid long (Tseng *et al*., 2007b) Another exclusive feature of this new species is the length of 3' UTR, which is composed of 314 nucleotides, the largest among the picornaviruses (Tseng *et al*., 2007a). On the bases of these findings it is proposed that DHV-1 should be assigned to a new genus in the *Picornaviridae*.

Seal picornavius 1 (SePV-1) is a new picornavirus species that was isolated from Arctic ringed seals (*Phoca hispida*) in Canada. The SePV-1 genome contains two 2A genes, whose sequence correspond to the canonical co-translational cleavage site DxExNPG↓P, found in cardioviruses, apthoviruses, teschoviruses and erboviruses (explained in sections 1.4.5 and 1.4.6). Furthermore, the absence of a predicted maturational cleavage site between 1A and 1B (VP0) was also observed in other species within the family, such as LV, HPeV and DHV.

1.3.2 Genome structure and organization

The name *Picornaviridae* conveys two important features of the family, the small size (*Pico*-) and the type of nucleic acid that the viral genome carries (-***rna****viridae*). This family has played a major role in the development of modern virology. Foot-and-mouth disease virus was the first animal virus discovered, followed by poliovirus, which was isolated ten years later. Picornavirus virions are spherical particles with a diameter of 30 mm and are simply composed of a protein shell surrounding the naked RNA genome. Their capsid is composed of four structural proteins (VP1, VP2, VP3, and VP4), excepting parechoviruses, DHV-1 and SePV with only three (VP1, VP3 and VP0, which is equivalent to the uncleaved VP2 plus VP4 precursor of other picornaviruses).

Picornavirus RNAs have certain features that differ from most mammalian mRNAs. The lack of the 5' terminal cap structure (m^7GpppN) and the presence of 5' untranslated regions are among their distinctive properties. The naked RNA genome of picornaviruses is infectious because it can be translated on entry into the cell using the host´s cell machinery. Its genome is covalently liked at the 5´end to a protein

called VPg (3B; virion protein, genome linked). FMDV is the only one in the family that has three Vpgs instead of one (Forss & Schaller, 1982). The picornavirus genome can be divided in three regions: a *5´untranslated region* (5´UTR), which is 600 to 1,200 nucleotides long, the *coding region*, which comprises 6,500 to 7000 nucleotides and a short *3´untranslated region* (3´UTR) that contains a heteropolymeric segment and a poly(A) tail (Rueckert, 1996; Palmenberg, 1990). The 5'UTR contains a 5'terminal stem-loop, cloverleaf-like structure, and a type I or II internal ribosome entry site (IRES), which allows a cap-independent mode of translation by binding to eIF4G (Kolupaeva *et al*., 1998, 2003; Clark *et al.*, 2003). In aphthoviruses, cardioviruses (except TMEV) and erboviruses the 5'UTR also includes a poly(C) tract. The poly(C) tail is longer in cardioviruses and is associated with higher virulence in animals (Duke *et al*., 1990; Hahn & Palmenberg, 1995; Wutz *et al.*, 1996). IRES's enable the eukaryotic ribosome to bind directly to the internal site without first having to scan from the 5' terminus, thus allowing 5'cap-independent translation (reviewed in Jang, 2005). There are four different IRES types, one of them belatedly discovered in porcine teschovirus-1 (Kaku *et al.,* 2002). These viral elements are widely used in vectors for gene therapy and their ability to enhance expression of an upstream gene in dicistronic vectors, has been recently discovered (Niepman, 2007). The 3' poly (A) tail is essential for infectivity. Its removal in poliovirus leaves the viral genome non-infectious (Spector & Baltimore, 1974).

The picornavirus polyprotein is divided into three regions: P1, which encodes the structural proteins, and P2-P3, which encodes the non-structural proteins that are involved in polyprotein processing (2A or $2A^{pro}$, $3C^{pro}$ and $3CD^{pro}$) and genome replication (2B, 2C, 3AB, $3B^{VPg}$, $3CD^{pro}$, $3D^{pol}$).

Unlike other genera, cardio-, aphtho-, erbo-, kobu-, tescho- and sapelovirus have an L protein region at the N-terminus of the polyprotein (figure 4). In aphtho- and erboviruses this acts as a protease (L^{pro}), cleaving at its own C-terminus, to produce the first cleavage of the polyprotein (van Pesch *et al., 2001;* Hinton, 2002; van Eyll *et al*., 2002).

The next stage in processing is the primary even that separates the encapsidation functions (P1; capsid proteins) from the replicative functions of the polyprotein (referred to as the P2-P3 region).

The 3C protease ($3C^{pro}$) accomplishes the final primary cleavage, between the 2C and 3A regions and is also responsible for secondary proteolytic processing. Both

the 3C and L-proteases are involved in the degradation of certain host-cell proteins, to enhance virus replication (for reviews see: Leong *et al.*, 2002; Grubman *et al.*, 1995; Ryan & Flint, 1997; Piccone *et al.*, 1995; Robert & Belsham, 1995; Gradi *et al.*, 2004). Each of these proteins are described in the following sections and important intra-genera variations discussed.

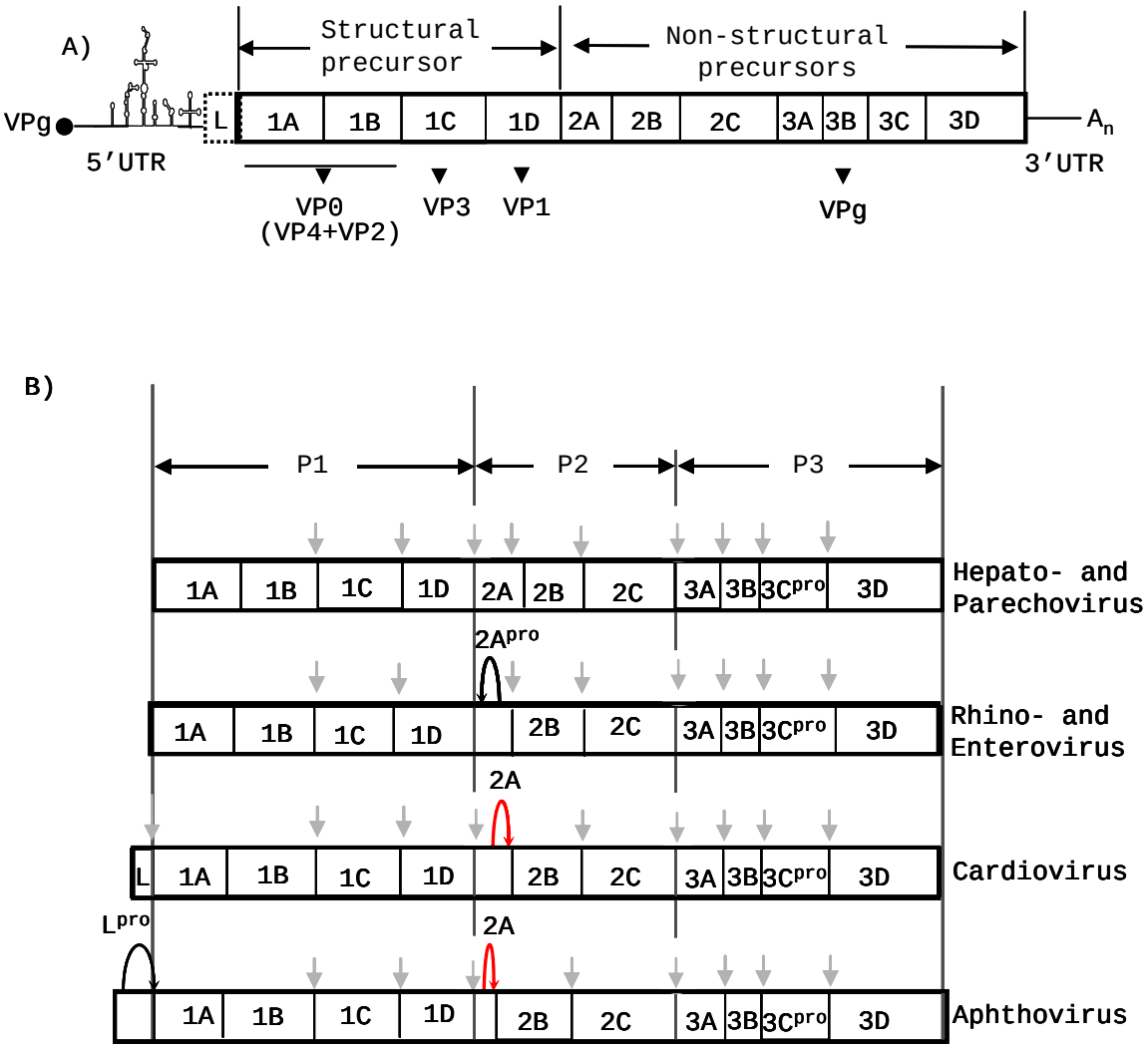


Figure 4. Picornavirus genome and proteolytic processing. A) General representation of picornavirus genome. Leader protein (L) is only present in aphthoviruses and cardioviruses and is a protease in the former genus. B) Proteolytic processing in picornaviruses. Grey vertical arrow: $3C^{pro}$ cleavages, also including those for which $3CD^{pro}$ are required. Black curved arrow: $2A^{pro}$, a trypsin family member found in rhinoviruses and enteroviruses. Red curved arrow: 2A co-translational cleavage in aphthoviruses and cardioviruses. 1AB cleavage mechanism is unknown and it does not occur in parechoviruses.

1.3.3 Replication cycle

The replication cycle starts with attachment of the viral particle to specific cell receptors on the membrane, followed by uncoating of the capsid to release the RNA genome inside the cell. The entire replication process takes place in the cytoplasm, where the RNA is translated to yield viral proteins necessary for replication, polyprotein processing and encapsidation of new virus particles (figure 5).

The single positive RNA strand replicates itself by producing a negative stranded RNA intermediate that is used as a template to create new single positive strands. This process occurs in small membranous vesicles that are stimulated by certain virus proteins. The synthesis rates of the positive RNA strand is ~100 fold greater than the negative strand since the daughter positive strands may fulfil one of three functions (i) as templates for the negative strand synthesis, (ii) they act as mRNAs to direct the synthesis of viral proteins or (iii) they are encapsidated into virions.

Once the pool of capsid proteins is sufficient for encapsidation, coat protein precursor P1 is cleaved and assembled into pentamers. These associate with newly synthesized positive-stranded RNAs to form fresh infectious particles that are released from the cell in a number of ways depending on the virus type. The majority of picornaviruses release their new particles by cell lysis, in others, such as Hepatitis A virus, release occurs in the absence of as cytophatic effect.

This process takes from 5 to 10 hours, depending on the virus species, temperature, pH, host cell, and multiplicity of infection.

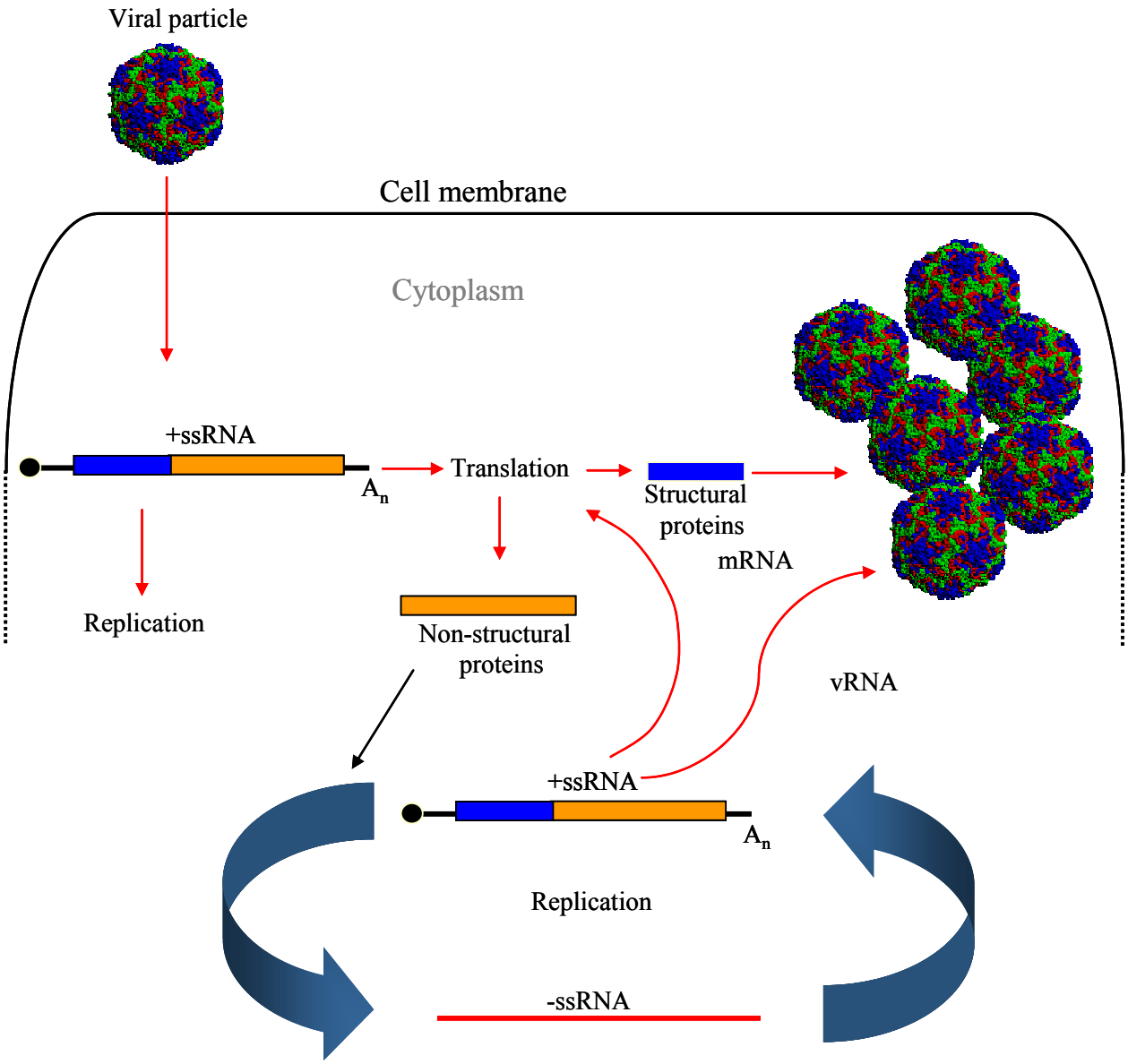


Figure 5. Picornavirus replication cycle. The entire replication takes place in the cytoplasm. The viral capsid attaches to specific cell membrane receptors and releases the positive single stranded RNA (+ssRNA) genome into the cytoplasm. The genome is translated into a large polyprotein, which can process itself during and after translation leading to the production of non-structural and structural proteins, required for replication and capsid formation, respectively. The replication machinery produces a negative-stranded RNA from the input positive-stranded template. This new strand will be used as a template to produce new positive-stranded RNAs that will be encapsidated once the pool of capsid proteins is large enough. Thus, new infectious viruses will be produced.

1.3.4. Picornavirus polyprotein

1.3.5.1 L protein

In contrast to other picornaviruses, the first protein encoded in the genome of cardioviruses, aphthoviruses, erboviruses, kobuviruses, teschoviruses as well as two members of the proposed genera Sapelovirus (Human enterovirus-8 and simian virus 2), is the Leader protein (L) (Zell *et al*., 2005). In aphthoviruses, it is a protease (L^{pro}) (Strebel & Beck, 1986), which appears in two different forms (Lb and Lab) in FMDV (Clarke *et al.*, 1985). L^{pro} cleaves at its own C-terminus releasing itself from the N-terminus of VP4. Furthermore, it also cleaves the translation initiation factors eIF4GI and eIF4GII, together with $3C^{pro}$, leading to inhibition of cap dependent translation in infected cells (Gradi *et al.*, 2004; Belsham *et al.*, 2000). Like aphthovirus, erbovirus L protease cleaves at the L/VP4 junction but does not have a role in eIF4G cleavage (Hinton *et al.,* 2002).

In cardioviruses, the L protein does not have proteolytic activity but plays other important roles during infection. The TMEV L protein interferes with trafficking of the cytoplasmic interferon regulatory factor 3 (IRF-3), a factor critical for transcriptional activation of alpha/beta interferon genes (Delhaye *et al.*, 2004) and interacts with Ran-GTPase, disrupting nucleocytoplasmic transport (Porter *et al*., 2006). L protein also plays an essential role in persistence by inhibiting the production of alpha/beta interferon (van Pesch *et al.*, 2001). Moreover, Cardiovirus L protein also has a role in alteration of nucleocytoplasmic traffic, although this function is not essential for viral reproduction (Lidsky *et al*., 2006).

1.3.4.2 L protein*

L* protein is a unique feature of translation of the TO subgroup strains of TMEV (DA, BeAn). It encodes an 18 kDa protein, which is translated from an alternative open reading frame starting 13 nucleotides downstream from the AUG codon of the main protein (Kong & Roos, 1991; Yamasaki *et al*., 1999). *In vivo* studies of the L* protein have shown a direct effect in viral persistence of infections and also enhancement in infection of macrophage cell lines (reviewed by Brahic *et al.,* 2005). This complementary AUG codon appears as an ACG in other strains, such as GDVII. L* can not only be translated from the second downstream AUG codon but it can also be initiated from this uncommon ACG codon in this non-persistent strain

(van Eyll & Michiels, 2000). Saffold virus (SAF-V) is a novel human cardiovirus, which has been recently discovered. In contrast to TMEV strain DA, which has 156 amino acids, the L* open reading frame of SAF-V encodes only 57 amino acids. It is still unknown if this open reading frame encodes for a protein (Jones *et al.*, 2007).

1.3.4.3 Capsid proteins

The structural proteins are encoded towards the 5'end of the open reading frame, which appears to fold into a structure closely related antigenically to the viral proteins VP0, VP1 and VP3 (Jackson *et al.*, 2003). The structural capsid proteins, VP1, VP2, VP3, and VP4 (VP1, VP3 and VP0 in parechoviruses) form an assembled icosahedral structure (shown in figure 6).

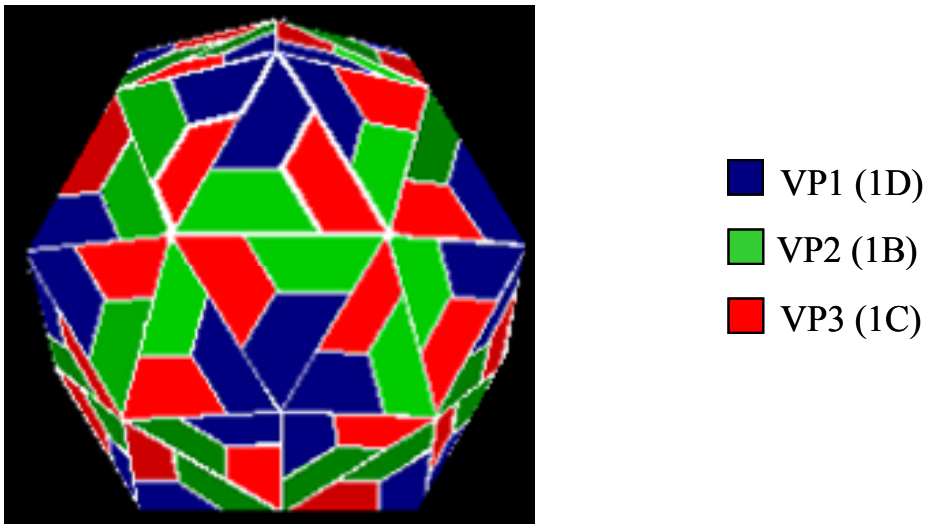


Figure 6. Icosahedral symmetry of the capsid if picornaviruses. The structural capsid proteins, VP1, VP2 and VP3 are shown in different colours. VP4 is not shown because it is internal to the capsid. Figure by courtesy of Professor Martin Ryan.

1.3.4.4 Protein 2A

The 2A protein is variable between genera, in contrast to the other non-structural proteins, which have the same structure. Kobuviruses, HePV, DHV-1 ($2A_3$) and hepatovirus 2A proteins are not involved in polyprotein processing (Jia *et al*., 1993; Schultheiss *et al.,* 1995; Yamashita *et al.,* 1998). In other genera, 2A is involved in the separation of the P1 structural proteins from the P2 and P3 non-structural proteins. In enteroviruses and rhinoviruses 2A is a protease ($2A^{pro}$), while in cardio-, aphtho-, erbo- and tescho-, LV ($2A_1$), DHV-1 ($2A_1$) and SePV ($2A_{1-2}$) it is an oligopeptide also responsible for the primary cleavage between structural and non-structural proteins (Donnelly *et al*., 2001a). This 'cleavage' is not proteolysis but a translational effect -'ribosome skipping' (Ryan *et al.,* 1991; Ryan & Drew, 1994; Ryan *et al*., 1999; Donnelly *et al*., 2001b; Ryan *et al.*, 2002). This novel mechanism is explained in more detail in section 1.2.5. 2A in the aphtho-, erbo- and teschovirus is an oligopeptide of ~18 aa whereas in the cardioviruses 2A is a longer track (~150 aa). EMCV-2A protein activates the translation initiator factor 4E binding protein 1, which binds eIF4E, the cap binding subunit of the initiation factors complex. Thus, it suppresses cap-dependent translation (Gingras *et al.,* 1996; Aminev *et al.,* 2003 a, b; Svitkin *et al.,* 2005). In entero- and polioviruses 2A protease also plays a role in translation repression by cleaving eIF4GI, both directly and indirectly, through the activation of cellular proteases (Zamora *et al*., 2002). In addition, 2A protease, together with 3C protease, cleaves the poly(A)-binding protein (PABP) during infection, and thus, inhibits host cell translation (Joachims *et al.,* 1999; Kuyumcu-Martinez *et al.,* 2002). Poliovirus 2A protease also increases RNA stability (Jurgens *et al.,* 2006).

New 2A sequences with the same conserved motif 'DxExNPGP' (where x= any amino acid) found in aphtho- and cardioviruses have also been found within the *Picornaviridae*. In addition, this motif has been identified in species outwith this family (section 1.3.6).

1.3.4.5 Protein 2B

Enterovirus 2B protein is a viroporin, a transmembrane pore-forming protein, which participates in different viral functions (Gonzalez & Carrasco, 2003). This kind of protein alters membrane permeability through the formation of pores, and thus, induces disassembly of the Golgi complex (Sandoval & Carrasco, 1997). Coxsakievirus 2B protein inhibits vesicular protein transport by reducing the endoplasmic reticulum (ER) and Golgi´s Ca^{2+} content. This disturbance on intracellular Ca^{2+} homeostasis leads to two different events: the enhancement of viral RNA genome replication and the suppression of apoptotic host-cell responses (Kuppeveld *et al*., 2006; Doedens & Kirkegaard, 1995). Little is known about the function of the other picornavirus 2B proteins, although it is been recently shown that rhinovirus 2B is also localized in the ER and Golgi apparatus and functions similarly to enterovirus 2B. In contrast, HAV, FMDV and EMCV 2B protein is not localized in the ER and Golgi apparatus and does not cause relevant effects on Ca^{2+} homeostasis and intracellular protein trafficking (de Jong *et al.,* 2008).

1.3.4.6 Protein 2C

Protein 2C is a highly conserved non-structural protein that binds to membranes and RNA, and is crucial in poliovirus replication (reviewed in Goodfellow *et al*., 2003). It also plays a role in encapsidation (Vance *et al*., 1997) and also has ATPase/GTPase activity (Rodríguez & Carrasco, 1993). 2C and its precursor 2BC are responsible for poliovirus RNA binding to the cytoplasmic vesicles, whose formation is also induced by this precursor. Its mechanism of action has yet to be determined. FMDV 2C protein is localized in juxtanuclear structures, vesicles that could be derived from Golgi compartements. However, the origin of these vesicles is not clear yet; recent studies argue against the relation with the Golgi (Knox *et al.,* 2005; Moffat *et al.,* 2005). For instance, in FMDV and EMCV infected cells treated with Brefeldin A (BFA), which inhibits membrane transport between the ER and the Golgi by preventing the formation of COPI-dependent secretory transport vesicles (Duden *et al.,* 1994), 2C juxtanuclear localization and replication is not inhibited (Gazina *et al*., 2002). In contrast, replication is inhibited in PV and Rhinoviruses when infected cells are treated with BFA (Cuconati *et al.,* 1998).

1.3.4.7 Protein 3A

3A protein has a role in disrupting ER-to-Golgi protein trafficking, which inhibits the secretion of cytokines such as interleukins (IL-6, IL-8) and interferon-β (IFN-β) (Dodd *et al*., 2001). It has an anti-apoptotic effect; it provokes the release of intracellular calcium *via* permeabilization of cellular membranes (Liu *et al*., 2003). Poliovirus 3A exists as a dimer and is a critical component of the viral replication complex (Strauss *et al*., 2003). Additionally, it also inhibits TNF-induced apoptosis by elimination of TNF receptor from the cell surface due to the inhibition of protein trafficking (Neznanov *et al*., 2001). FMDV 3A plays an essential role in the determination of host-range. For instance, it has been shown that a single mutation within 3A mediated adaptation of FMDV to the guinea pig (Nuñez *et al.,* 2001). Furthermore, mutations within poliovirus 3A also affected host-range (Lama *et al.,* 1998).

1.3.4.8 Protein 3B (VPg)

Protein 3B is covalently bound to the 5'end of viral RNA and functions as a primer for the initiation of viral genome replication. FMDV 3B encodes three different types of VPg, which are uridylylated by $3D^{pol}$, making possible the initiation of the viral RNA replication (Nayak *et al*., 2005). The poliovirus VPg NMR structure has been solved and will hopefully improve our understanding of the mechanism of action of VPg, which is essential for virus replication, and also the interactions between this protein and the viral polymerase (Shein *et al.,* 2006).

1.3.4.9 Protein 3C

Protein 3C is a chymotrypsin-like protease responsible for the primary 2C/3A cleavage of the polyprotein (Palmenberg *et al.*, 1992; reviewed by Ryan & Flint, 1997; Ryan *et al*., 2004). In the case of HePV, kobuvirues, DHV-1 ($2A_3$) and hepatoviruses the primary 2A/2B polyprotein cleavage is mediated by the 3C protease (Stanway & Hyypiä, 1999). $3C^{pro}$ also induces the cleavage of the translation initiator factors eIF4A and eIF4GI-II (Belsham *et al.*, 2000) and is involved in cell apoptosis in poliovirus and enterovirus (Barco *et al.*, 2000). Poliovirus $3C^{pro}$, combined with unknown host-cell activity, degrades p53 (Weidman *et al*., 2001). Furthermore, this protease cleaves poly(A)-binding protein (PABP) and removes the C-terminal domain

(CTD) that interacts with several translation factors. This mechanism of translation inhibition complements the effect of eIF4G cleavage by $2A^{pro}$ (Kuyumcu-Martinez *et al.,* 2004). In addition, $3C^{pro}$ contains RNA-binding domains (Blair *et al.,* 1998).

1.3.4.10 Protein $3D^{pol}$

$3D^{pol}$ is a RNA-dependent RNA polymerase (RdRp) that does not possess proof-reading activity. Its mechanism of action is still unknown. Studies have shown specificity of $3D^{pol}$ for each virus type and specificity also between $3D^{pol}$ and the other viral replicative proteins within the same virus species. A chimeric poliovirus, its $3D^{pol}$ replaced by coxsackievirus $3D^{pol}$, showed a lack of replication due to the inefficient recognition of the P1 protein substrate by the chimeric 3CD protease ($3CD^{pro}$) (Bell *et al*., 1999). Atomic structures of $3D^{pol}$ are available for poliovirus (Hansen *et al*., 1997) and FMDV (Ferrer-Orta *et al*., 2004), both showing the classical architecture: 'fingers', 'palm' and 'thumb' domains.

1.3.4.11 Cleavage intermediates

There are three different intermediates, 2BC, 3AB and $3CD^{pro}$, which have different roles from their cleavage products. Accumulation of small ER- and Golgi-derived membrane vesicles in the cytosol has been observed in enterovirus-infected cells, and is where viral replication takes place. 2BC is responsible for this accumulation (Bienz *et al*., 1994). 3AB is thought to be an integral membrane protein (Ciervo *et al.*, 1998) and it also induces $3D^{pol}$ activity, most likely by recruiting $3D^{pol}$ to the 3'termini of chain elongation sites (Richards & Ehrenfeld, 1998). On the other hand, $3CD^{pro}$ binds the 5' RNA 'cloverleaf' structure, an essential step in replication (Blair *et al*., 1998).

Poliovirus 3CD forms a ribonucleoprotein complex (RNP), together with a 3kDa ribosome-associated cellular protein at the 5' UTR region. This complex is essential for positive RNA synthesis but not for the negative strands (Andino *et al.,* 1993).

1.3.5. Polyprotein processing

In picornaviruses the first step of infection, following cell entry, is translation of the RNA genome into a polyprotein (reviewed in Palmenberg, 1990). The viral RNA, which has a poly (A) sequence at its 3' end, like cellular mRNAs, sequesters the cell's own translational machinery for its protein synthesis. One important peculiarity of most viral mRNAs is their uncapped 5' end. Picornaviruses have a ~22 amino acid long protein, VPg, which is covalently attached to its 5´end that has an important role in initiation of RNA synthesis and may also have a role to play in the virulence of the virus (reviewed in Reuckert, 1996). This polyprotein precursor does not appear in infected cells because it undergoes co-translational cleavage while it is being translated**.** The different products appear in equimolar quantities due to their common origin from a single precursor (reviewed in Palmenberg, 1990). $3C^{pro}$ and $3CD^{pro}$ are responsible for the main polyprotein processing events cleaving all the structural and non-structural protein precursors, except i) L^{pro}, which cleaves itself at its C-terminus, ii) 1AB precursor, whose cleavage mechanism is still unknown, and iii) 2AB, except in hepato- and parechoviruses. $2A^{pro}$, is the other enzyme involved in polyprotein processing and is responsible for the 1D/2A cleavage in rhino- and enteroviruses. $2A^{pro}$ separates the structural protein domain from the non-structural domain and it was first thought to be present in a wide range of picornaviruses. However, we now know that the majority of picornaviruses use another method (see section 1.5.1.4: ribosome skipping). This novel 2A protein, first discovered in FMDV (Ryan *et al*., 1991) expands the repertoire of translation strategies used by RNA viruses.

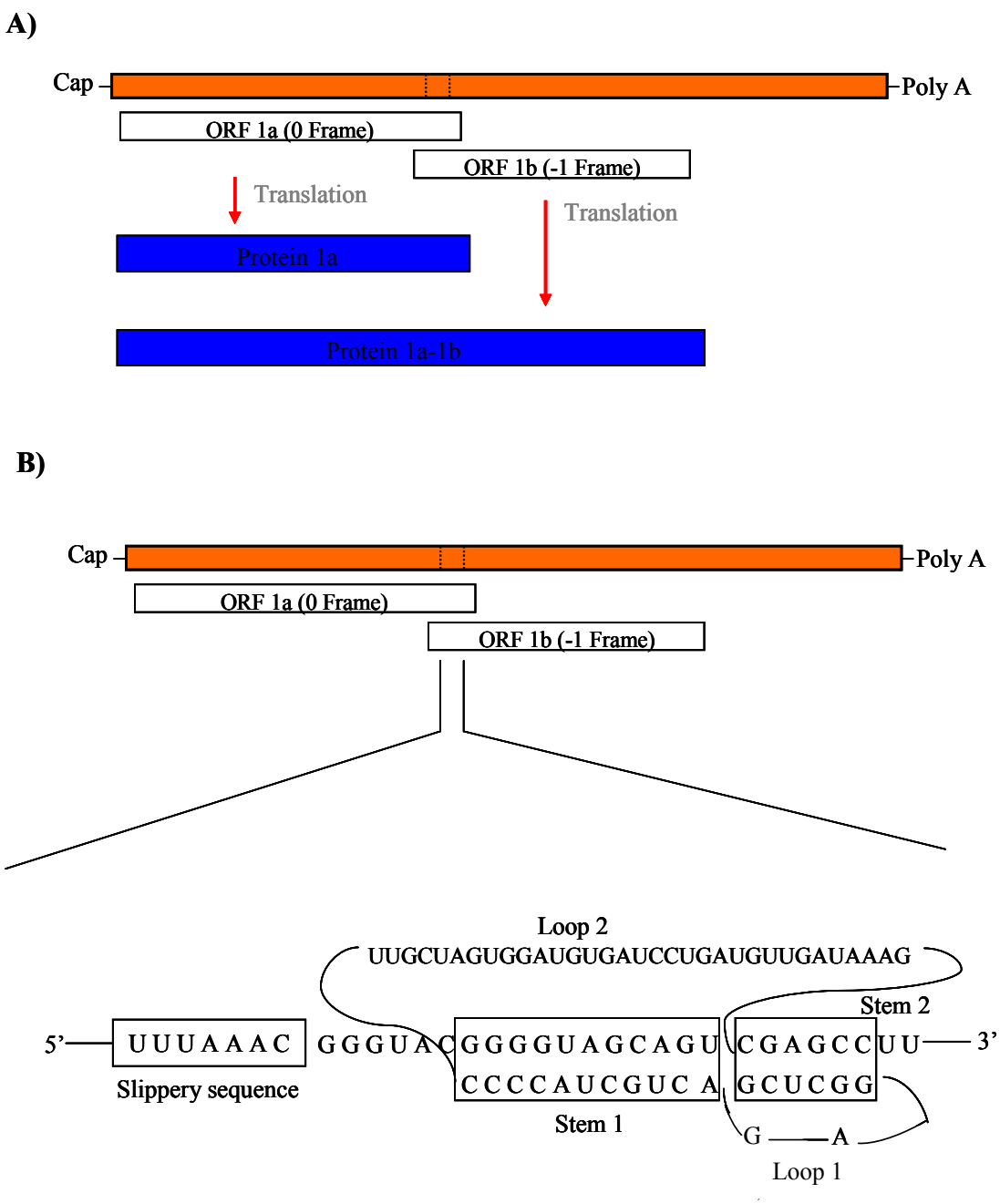
RNA viruses have evolved a remarkable variety of ways to translate their RNA genomes into structural and nonstructural proteins. Production of subgenomic mRNAs, synthesis of polyprotein precursors from single mRNAs, ribosomal readthrough, and frameshifting are all ways to produce more than one protein product from a single RNA genome. Picornaviruses translation strategy is the synthesis of polyprotein precursors that will be co-translationally or immediately post-translationally cleaved.

1.4. Control of protein biogenesis within positive stranded RNA viruses

1.4.1. Ribosomal Frameshifting

The majority of RNA viruses are single-stranded. The majority of these are positive-sense, single-stranded RNA viruses, such as the family *Picornaviridae, Coronaviridae, Caliciviridae and Flaviviridae.* Their genome is transcribed into a polycistronic mRNA, which is translated into a polyprotein that is subsequently cleaved into individual mature proteins.

One of the replication strategies that positive-sense, single-stranded RNA viruses RNA viruses use is programmed ribosomal frameshifting, in which ribosomes change reading frame within the mRNA, leading to the synthesis of alternative proteins. There are two essential signals needed for this process to occur; (i) a 'slippery' hepta-nucleotide sequence (eg. UUUAAAC), where the ribosome changes frame and (ii) an RNA pseudoknot structure situated downstream of the slippery sequence and consisting of two helical segments connected by single-stranded regions or loops. The current model proposes that the ribosome finds the pseudoknot while the slippery sequence is being translated causing the ribosome to pause on the slippery sequence, where it slips back one nucleotide and subsequently continues translation in the -1 reading frame (Somogyi *et al*., 1993). This recoding mechanism is found in several positive-stranded viruses such as infectious bronchitis virus (IBV), which is a member of the *Coronaviridae* (Brierley *et al.*, 1987; reviewed by Spaan *et al.*, 1988; reviewed by Brierley *et al*., 2007; figure 7).



.

Figure 7. Ribosomal frameshift representation in IBV. A) Frameshifting allows translation to continue at the end of protein 1a, thus, producing protein 1a-1b. B) A 'slippery' sequence (UUUAAAC), where the ribosome changes frame and an RNA hepta-nucleotide sequence pseudoknot structure situated downstream of the slippery sequence are two essential signals needed for this frameshift process to occur.

1.4.3. Leaky scanning

The majority of eukaryotic mRNAs are monocistronic and initiate translation when the 40S ribosomal subunit and initiation factors bind to the 5' end of the mRNA (m^7G cap). This preinitiation complex moves in a 3´direction on the mRNA in a process called scanning, until it reaches (generally) the first AUG codon. Then initiation factors are released allowing the 60S subunit to associate with the small subunit. The efficiency of initiation is affected by the nucleotide sequence surrounding a certain codon. The consensus sequence accepted as the most efficient in eukaryotic cells is 5'GCC(**A/G**)CCAUG**G** 3', where the presence of a purine at the position -3 followed by a G at position +4 is essential for high levels of translation. Nevertheless, only 5% of eukaryotic sequences contain this ideal consensus sequence, most of them have suboptimal ones (reviewed in Flint *et al.*, 2000). Some viral mRNAs encode two or more proteins in overlapping reading frames, initiation of translation occurring not only at the first 5' AUG but also at downstream AUGs. This occurs when ribosome preinitiation complexes bypass the 5´AUG due to the fact that this codon is surrounded by suboptimal nucleotide sequences. An example of this can be found in Sendai virus P/C gene mRNA, a member of the family *Paramyxoviridae* that contains a non-segmented negative sense single-stranded RNA genome, from which 6 mRNAs are transcribed. One of them, the P/C mRNA, is polycistronic and has two open reading frames that start near the 5´end (Girogi *et al*., 1983; figure 8). The P protein starts at the 5' proximal ATG in the first open reading frame, the last 95 amino acids being expressed as a different protein (X). The second open reading frame produces a nested set of non-structural C proteins (C', C, Y1, and Y2). Translation of the first 3 initiation sites on P/C mRNA is arranged in such a manner that leaky scanning is enhanced. The first non-structural protein, C', starts at an unusual non-ATG unusual codon, ACG, whereas the rest of proteins initiate on AUGs. Although ACG is in a good context, because it is an unusual start codon, this leads to an inefficient initiation, in which some ribosomes bypass ACG and initiate translation at the next initiator codon (AUG). This second initiator, which translates P protein, is an AUG but is in a poor initiation context (a pyrimidine at position -3), while the third AUG is in good context (an A at -3) and translates for C protein (Curran & Kolakofsky, 1989).

Another case of leaky scanning is observed in influenza B virus RNA segment 6. Influenza viruses belong to the family of *Orthomyxoviridae*, which are negative-

sense single-stranded RNA viruses. Influenza B virus RNA segment 6 is bicistronic and encodes NB and NA proteins in overlapping reading frames. NB translation starts at the 5' proximal AUG codon, while the NA AUG codon is four nucleotides downstream. Even if the NA initiation codon is in a good context for translation initiation compared to the NB one, both proteins accumulate in almost equal amounts in infected cells. This suggests a unique model in which ribosomes randomly choose which AUG to use. Certain cellular factors and surrounding mRNA sequences might have an effect on the selection of these codons (Williams & Lamb, 1988).

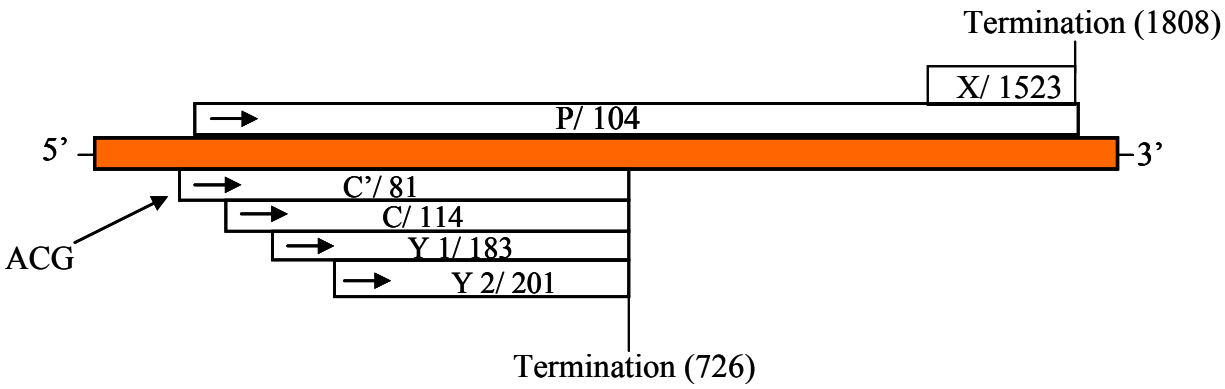


Figure 8. Schematic representation of the P/C mRNA of Sendai virus. The mRNA is shown in orange. The different ORFs are shown in boxes. The initiation site of C' is a non-ATG codon (ACG), whereas C, Y1 and Y2 start on ATGs. Numbers refer to the position of the first base of the initiation codons. The position of the termination codons is also indicated.

1.4.4. Reinitiation

Reinitiation is a very common strategy among prokaryotic cellular and viral RNAs. It is another mechanism by which two proteins are produced from a single mRNA. An example for this strategy is influenza virus RNA segment 7, which encodes two proteins, M1 and BM2, from a bicistronic mRNA. The ATG initiation codon for BM2 protein overlaps the termination codon for the M1 protein. Therefore, BM2 protein synthesis is dependent upon the initiation and termination of the upstream protein M1 (figure 9). A coupled translational termination-initiation mechanism is essential for the downstream protein to be produced (Horvath *et al*., 1990).

In caliciviruses, which are non-segmented positive single-stranded RNA viruses, translation initiation of the 3'terminal open reading frame is also achieved *via* a termination-reinitiation process. Its genome contains two open reading frames for member of the genera *Lagovirus* and *Sapovirus*, and three open reading frames in *Vesivirus* and *Norovirus* (Green *et al.,* 2000). The minor capsid protein VP2 is expressed *via* reinitiation of translation after termination of the upstream VP1 protein synthesis. A sequence of about 80 nucleotides, called termination upstream ribosomal binding site (TURBS), is essential for this termination-reinitiation mechanism to occur. TURBS has a specific conserved motif, which is complementary to 18S rRNA, and may be important to prevent release of post-termination ribosomes, hence, increasing the chance of reinitation (Meyers, 2007).

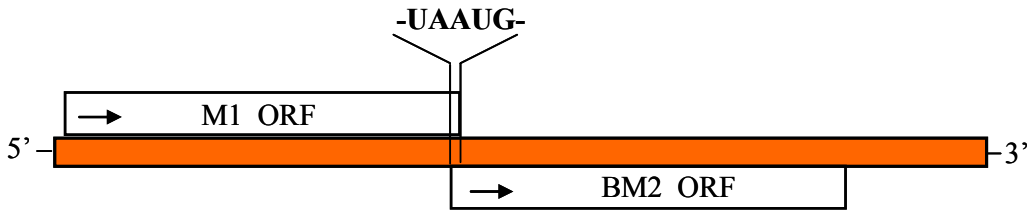


Figure 9. Influenza virus RNA segment 7. Bicistronic mRNA is represented in orange. The two ORFs are represented in boxes. The BM2 initiation codon (AUG) overlaps the M1 termination codon (UAA).

1.4.5. Suppression of termination

Suppression of termination can lead to the generation of a second protein with an extended carboxy-terminus. This occurs when one of the three termination codons UAG (amber), UGA (opal) and UAA (ochre) are suppressed as a result of leaky termination.

In murine leukaemia virus (MuLV), which is member of the family *Retroviridae,* the *gag* and *pol* coding regions are separated by an in-frame UAG termination codon. Read-through suppression of the UAG codon results in the production of a gag-pol fusion protein, which is cleaved later to produce the Pol proteins (protease, reverse transcriptase and integrase). This event happens because the tRNA is misreading the UAG termination codon for a Gln codon. A purine-rich sequence 3' of the termination codon, as well as a pseudoknot structure further downstream enhances suppression of termination (Yoshinaka *et al*., 1985; Feng *et al*., 1992) (shown in figure 10).

Another example of translational suppression is observed in Sindbis virus, which belongs to the genus alphaviruses from the *Togaviridae* family. This phenomenon is required for the synthesis of nsP4 protein, the viral RNA-dependent RNA polymerase. A single cytidine residue immediately 3' to the termination codon is essential for translational read-through (Li & Rice, 1993).

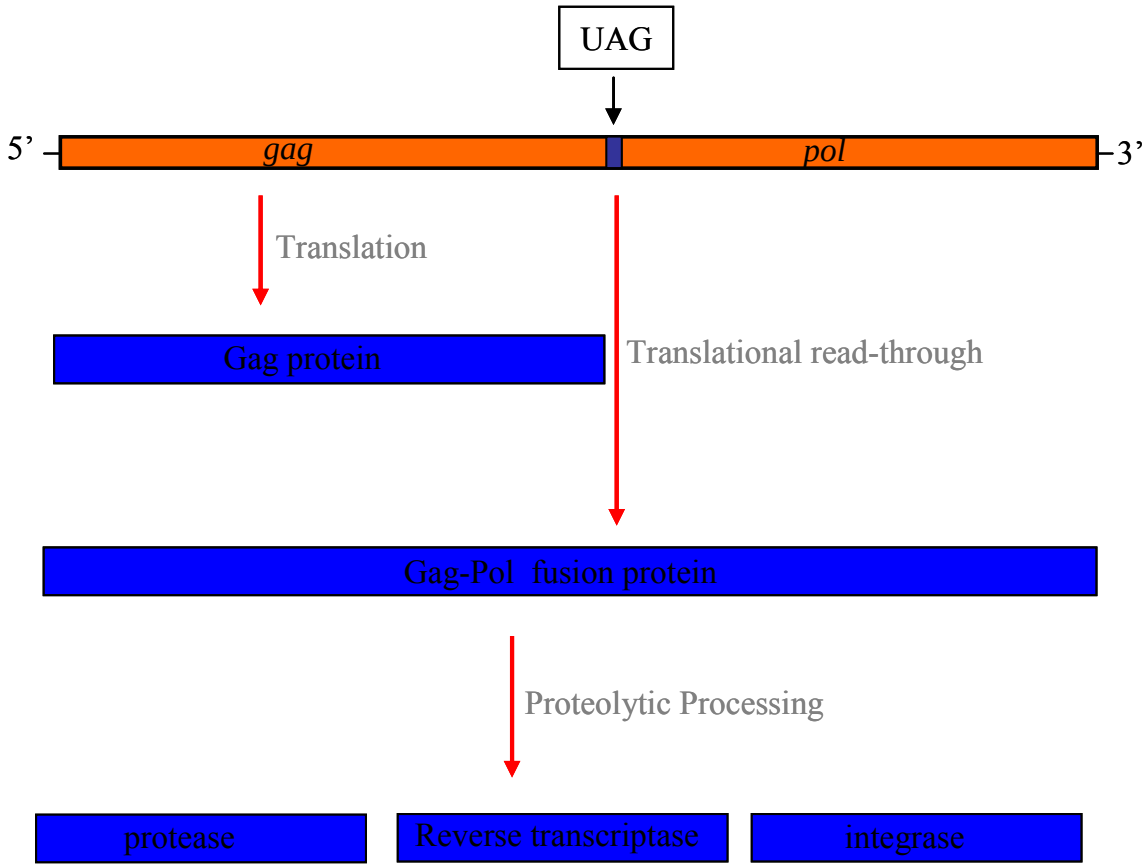


Figure 10. Supression of termination in MuLV. Read-through suppression of the termination codon (UAG) produces the gag-pol fusion protein, which is, later, processed to produce the Pol proteins. This event happens because the tRNA is misreading the UAG termination codon for a Gln codon.

1.4.6. Subgenomic mRNA

Another gene expression strategy involves subgenomic mRNAs, which allows structural and replicative proteins to be synthesized separately and in different amounts. Togaviruses, which are the smallest enveloped animal viruses, use this translation strategy. Their genome is dicistronic, the gene closest to 5´end encodes the replication proteins and the one at the 3´end encodes the structural proteins. The replication proteins are produced in the early stage of infection, using the viral RNA as a template. Structural proteins are not synthesized until the later stages since its translation initiation codon is masked from the ribosome. A subgenomic mRNA, which is produced from an internal initiation site on the negative strand RNA, is used as a template for capsid protein production (reviewed in Schlesinger & Schlesinger 1996). These structural and non-structural products will be proteolytically processed once they have been translated (shown in figure 11). The replicative product is cleaved by a viral proteinase, which is located in the nsP2 region (Ding & Schlesinger, 1989 and Hardy & Strauss, 1989), whereas the capsid proteins are obtained in different ways depending on the genera. In rubivirus the polyprotein is cleaved by host cell proteinases while a mixture of viral and host cell enzymes process the alphavirus subgenomic mRNA (ten Dam *et al*., 1999).

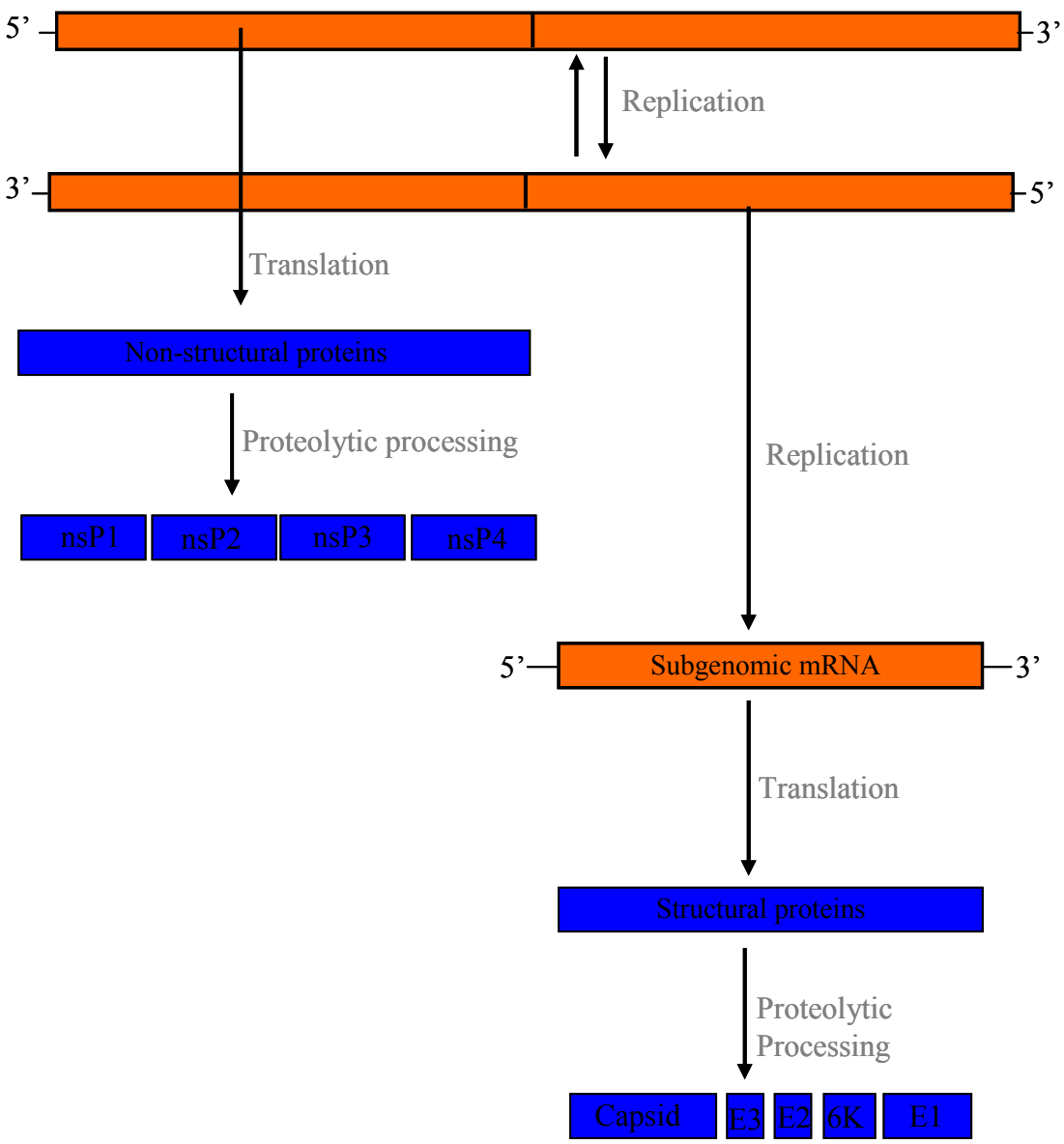


Figure 11. Togavirus subgenomic mRNA. mRNA is represented in orange, and the protein in blue. Structural proteins are encoded by a subgenomic mRNA copied from the negative strand RNA.

1.4.7. Nested subgenomic mRNAs

The order *Nidovirales* includes the *Coronaviridae* (torovirus and coronavirus) *Roniviridae* and *Arteriviridae*. They have single-stranded, polycistronic RNA genomes of positive polarity. The non-structural proteins are encoded at the 5'end, and structural proteins at the 3'end. After uncoating inside the host cell cytoplasm, the genome is translated into two replicase open reading frames (ORF1a and ORF1b) by the host ribosome, ORF1b is produced by ribosomal frameshifting. The large polyprotein precursor is autoproteolytically cleaved to produce a membrane-bound replicase/transcriptase complex that mediates the synthesis of the genome RNA and a nested set of subgenomic RNAs. Structural proteins are synthesized from these 3´coterminal set of subgenomic RNAs, which are composed of a leader and a body (illustrated in figure 12). The leader and body are transcribed from sequences in the 3' end and 5'-terminal one third of the genomic negative-strand, respectively, by discontinuous RNA synthesis. These two segments are connected by a conserved junction site sequence, which is found both at the 3' end of the common leader sequence and at the 5' end of the mRNA body (Snijder & Meulenberg, 1998). During synthesis of the negative-stranded RNA, the nascent RNA strand is transferred from one site in the genomic template to another to yield subgenomic RNA molecules. This process is guided by conserved transcription-regulating sequences (TRSs) found at the genomic positive-stranded RNA (Pasternak *et al*., 2004).

The number of subgenomic RNAs produced varies between families, being 4 in Torovirus, 7 in *Arteriviridae* and 9 in Coronavirus (Gorbalenya *et al.*, 2006; Snijder & Meulenberg, 1998).

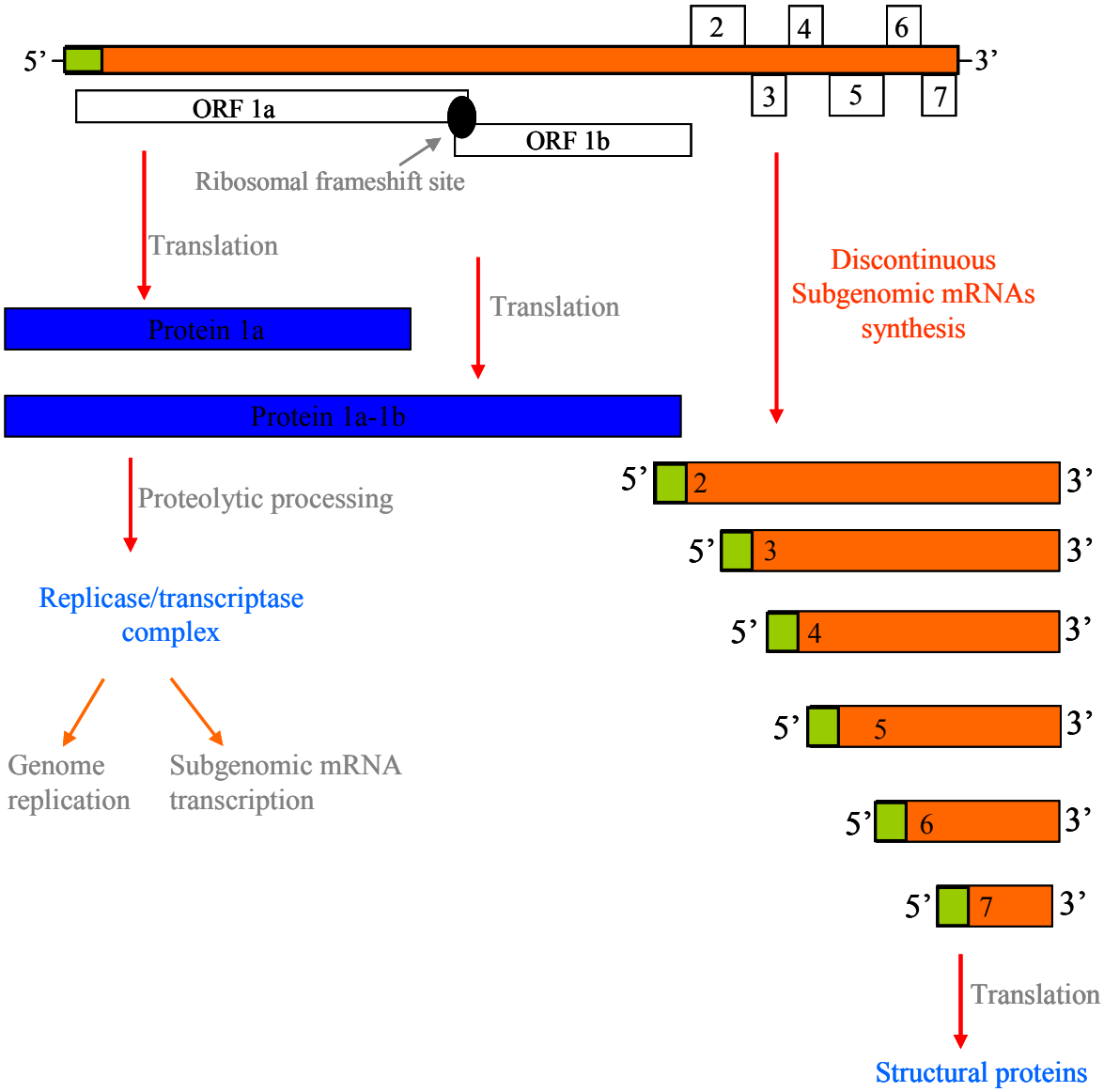


Figure 12. Schematic diagram of the genome organization and expression of equine arteritis virus, the artevirus prototype. The mRNA is represented in orange, the proteins in blue and the leader sequence in green. The nested set of subgenomic mRNAs are produced by a discontinuous transcription mechanism, where the body translocates to the 5' end of the genomic RNA to complete synthesis of the leader sequence.

1.5 2A oligopeptide

1.5.1 Aphthovirus 2A

1.5.1.1 Characteristics

2A is a short peptide located between the capsid protein domain and the downstream replicative domain. It is a proteolytic enzyme in some viruses, such as enteroviruses and rhinoviruses whereas in cardioviruses, aphthoviruses, erboviruses, DHV-1 ($2A_1$), LV ($2A_1$), SePV ($2A_{1-2}$) and teschoviruses, it appears to be an oligopeptide with a self-cleaving activity (reviewed by Glaser *et al*., 2003).

FMDV-2A is an oligopeptide of only 18 amino-acids in length, whereas in Cardioviruses it is ~150 aa in length. This protein is responsible for the separation of the structural protein domain from the non-structural domain during primary polyprotein processing. The 'cleavage' of the 2A protein occurs at its C-terminus, during translation. Thus, 2A remains attached to the upstream capsid protein precursor (P1) after processing. Subsequent secondary polyprotein processing, mediated by $3C^{pro}$ and $3CD^{pro}$, separates 2A from P1, "delineating" 2A as just 18 amino acids.

Both FMDV 2A and TMEV 2A have a completely conserved C-terminal sequence that consists of an Asparagine, Proline and Glycine (-NPG-) (shown in figure 13). Furthermore, the first amino acid of the next protein (2B), a proline, is essential for the cleavage. Without this highly conserved amino acid cleavage does not occur and translation of the uncleaved polyprotein precursor continues (Hahn & Palmenberg, 1996; Hahn & Palmenberg, 2001; Donnelly *et al*., 1997).

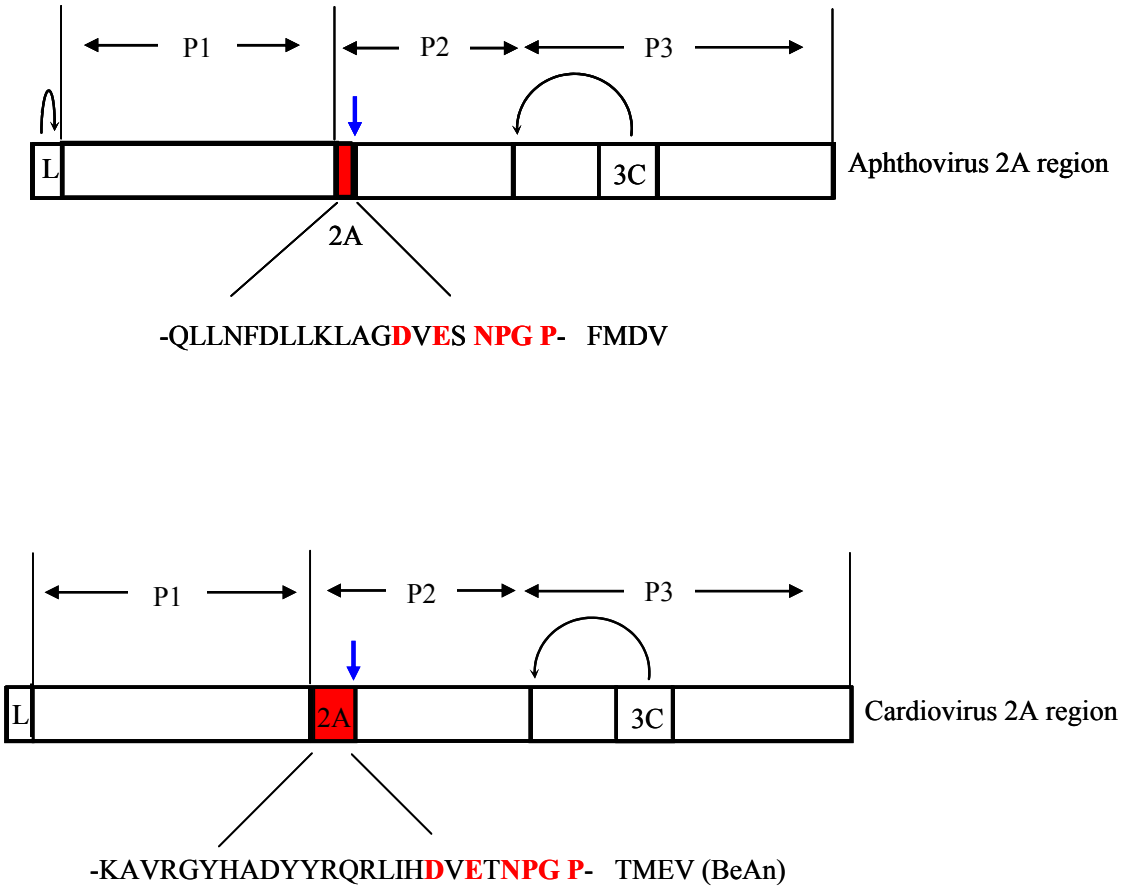


Figure 13. Illustration of aphthovirus and cardiovirus 2A protein. 2A is responsible for the separation between the capsid protein (P1) and the replicative proteins (P2-P3). 3C is also represented, as responsible for the primary cleavage between P2 and P3. L is a protease in aphthovirus but not in cardioviruses. 2A protein sequence is shown; the most conserved part in red.

1.5.1.2 Studies on FMDV 2A

The mechanism by which 2A is separated from 2B protein is now becoming understood and remains the focus of a number of studies. Several models of 2A cleavage have been proposed in the past 20 years. It was first suggested that 2A sequence may be recognized by host cell proteinases. However, these proteinases would have to be conserved among different cell types and also be present in *in vitro* systems. Other explanations suggest that 2A protein possesses a novel type of proteolytic activity or is able to disrupt normal peptide bond formation. Some of them are explained below.

Initial suggestions focused on the possibility that FMDV-2A had proteolytic activity like the other picornavirus proteins involved in polyprotein processing (Ryan *et al.*, 1989). Deletions in the N-terminus coding regions around 2A showed that the active site was only 19 amino acids long and contained the first amino acid of protein 2B, a proline (Ryan *et al*., 1991). A plasmid with the 18 amino acid 2A sequence together with the N-terminal proline of 2B, inserted between chloramphenicol acetyltransferase (CAT) and β-glucuronidase (GUS) in a single long open reading frame [CAT-2A-GUS], demonstrated that 2A cleavage was still active in a completely foreign context (shown in figure 14). This experiment clearly eliminated the involvement of additional sequence elements elsewhere in the FMDV polyprotein, such as L^{pro} and $3C^{pro}$. Uncleaved protein was observed *in vitro*, using rabbit reticulocyte lysate systems but not *in vivo*. Further, deletions at the N-terminus of FMDV 2A showed that cleavage required only 13 amino acids, thus, the N-terminal region is not essential for cleavage activity (Ryan *et al.*, 1994). However, it has an influence on efficiency of cleavage and it also comprises the cleavage site required by $3C^{pro}$ and $3CD^{pro}$ to cleave 2A away from P1 (Ryan *et al.*, 1989).

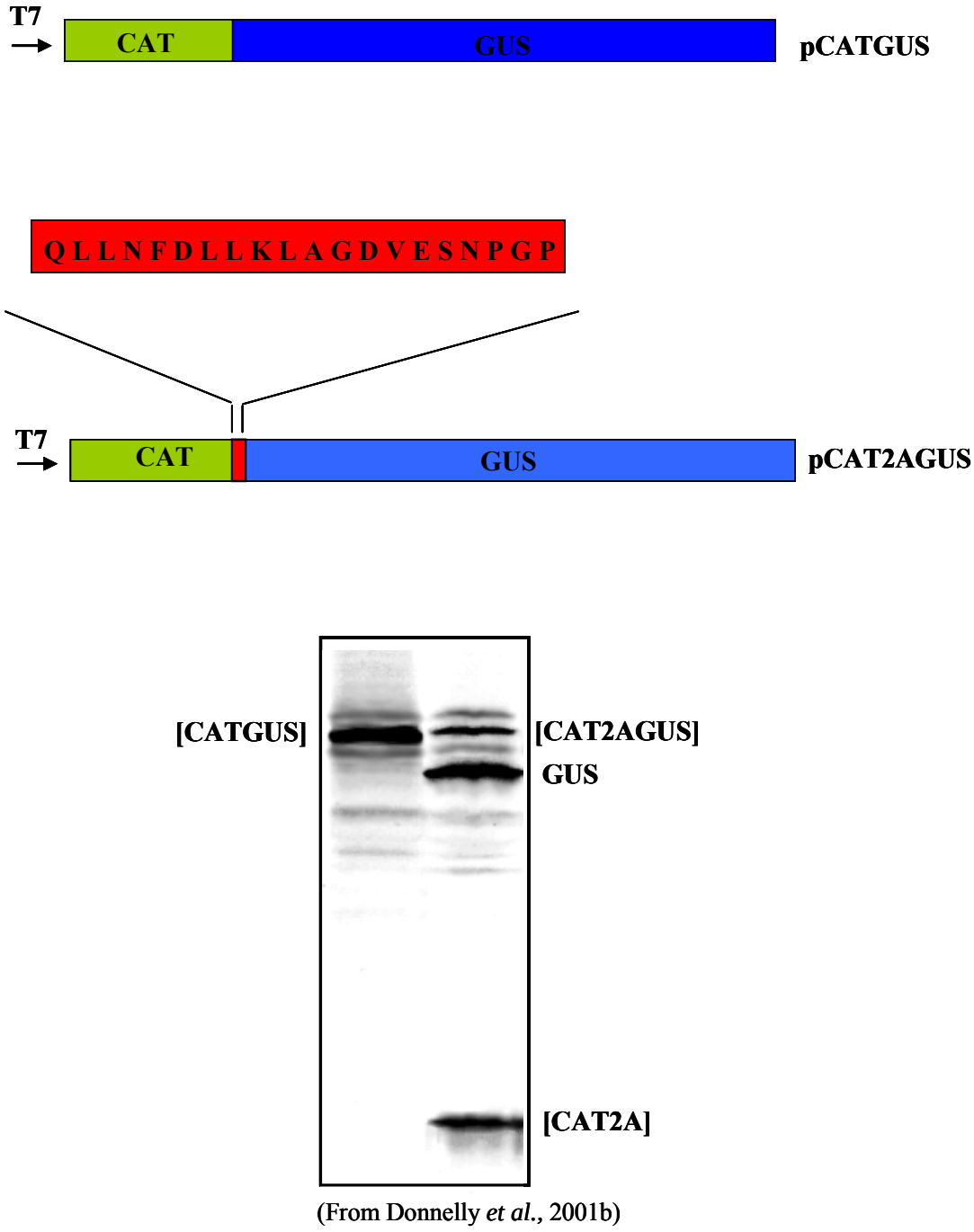


Figure 14. Artificial reporter polyprotein [CAT-2A-GUS]. FMDV-2A sequence was inserted between CAT and GUS, in a single long ORF. The translation profiles showed three products, the uncleaved protein [CAT-2A-GUS] and the two processed products [CAT-2A] and [GUS]. CAT-GUS polyprotein was used as a control.

Further experiments using a [CAT-2A-GUS] artificial polyprotein and *in vitro* translation systems were performed. As a result of translation *in vitro*, it was observed that not only the uncleaved product [CAT-2A-GUS] appeared, but also GUS and [CAT-2A] cleavage products were produced. It was also shown that the products are separated co- and not post-translationally since prolonged incubation of the "uncleaved" translation products did not result in increased cleavage. Additionally, in these experiments it was observed that an imbalance in the translation products occurred. Analysis of the translation profiles obtained from [CAT-2A-GUS] constructs showed higher accumulations of [CAT-2A] compared to GUS (Donnelly *et al.*, 1997). These results provided the first indication that the mechanism of 2A-mediated 'cleavage' was not proteolytic. If this was the case, equimolar quantities of the cleavage products would be obtained.

Although degradation rates of neither protein were significant, studies of these translation products showed a slower degradation of GUS with respect to [CAT-2A]. It was concluded from this work that more [CAT-2A] than [GUS] was being translated. This was further verification than 2A cleavage was not caused by a proteolytic event but more likely by a translational phenomenon (Donnelly *et al.*, 2001b).

Another artificial reporter polyprotein system comprising green fluorescent protein (GFP) linked *via* FMDV-2A to β-glucuronidase (GUS), [GFP-2A-GUS] was analysed. The results of translation reaffirmed earlier findings and showed a molar excess of the upstream product [GFP2A] over [GUS]. When GUS was the N-terminal protein, [GUS-2A-GFP], then it too was the one in molar excess. This demonstrated that the N-terminal product was always in excess. This imbalance was observed in *in vitro* systems but not *in vivo*. A possible explanation for this is the suboptimal functioning of 2A because, in these experiments, the whole native protein was not being used but only a fragment. Increasing the length of FMDV 2A sequences, by adding parts of the 1D region to the N-terminus of 2A reduces the amount of unprocessed material detected in *in vitro* systems. Therefore, three possible outcomes were observed during translation of [GFP-2A-GUS]: (i) the cessation of the process at the C-terminus of [GFP-2A] consequently with the dissociation of the ribosome, (ii) the continuation of translation after cleavage with the synthesis of the downstream

GUS product and (iii) the translation of the uncleaved full-length protein [GFP-2A-GUS] (Ryan & Drew, 1994).

1.5.1.4 Translational Model and Mechanism of 2A Action

The translational model for 2A activity suggested that 2A protein may have esterase activity rather than of a proteinase (Ryan *et al.*, 1999). In this model 2A mediates an attack on the ester linkage between the nascent peptide and the tRNA moiety preventing peptide bond formation between the Glycine residue of 2A and the N-terminal proline of 2B.

During elongation, translocation of the peptidyl-tRNA from the acceptor site (A) into the peptidyl site (P) site occurs, mediated by eukaryotic translation elongation factor 2 (eEF2). This allows the ingress of prolyl-tRNA into the A site. The next step would be formation of the peptide bond between the glycine and proline but nucleophilic attack by the prolyl-tRNA amide nitrogen upon the peptidyl-$tRNA^{gly}$ carbonyl carbon is inhibited by 2A. Thus, prolyl-tRNA is unable to attack the peptidyl(2A)-$tRNA^{gly}$ ester linkage, and the nascent peptide is released by the hydrolysis of the glycyl-tRNA ester bond. The ribosome re-initiates translation of the downstream protein, not with the normal initiator amino acid methionine but with proline.

This model is represented in figure 15 (from Donnelly *et al.*, 2001b).

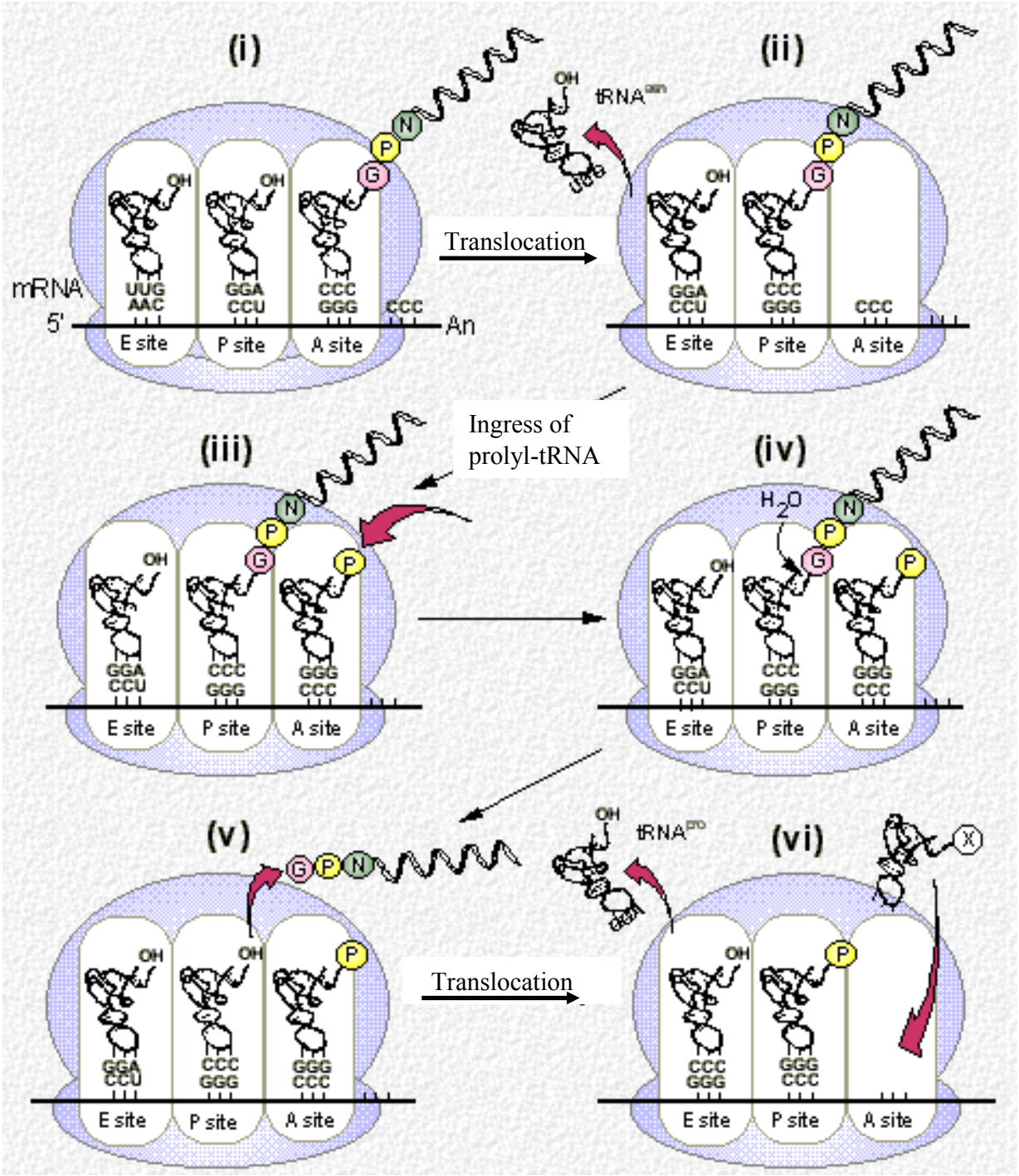


Figure 15. Proposed model of the mechanism 2A's action. The nascent 2A peptidyl-tRNA present in the A site (i), is translocated to the P site (ii), leaving the A site free for the ingress of the prolyl-tRNA (iii). The nascent 2A peptide interacts with the ribosome exit tunnel and re-orients the tight-turn at the base of the peptide. This re-orientation of the peptide inside the exit tunnel inhibits the peptide formation between glycyl-tRNA and prolyl-tRNA, leading to the release of the nascent peptide by hydrolysis of the glycyl-tRNA ester bond (iv-v). The ribosome re-initiates translation of the downstream protein with a proline.

Recent studies have shown translation ('release' or 'termination') factors that are the key for 2A-mediated 'cleavage' (Doronina *et al*., 2008).

Firstly, another line of evidence proved what was already suspected: the 2A reaction occurs in the ribosomal peptidyl transferase centre (PTC) of the ribosome. PTC is formed at the interface between the ribosomal small and large subunits, at the entrance of the ribosome exit tunnel and comprises the activity responsible for peptide bond formation. A series of mRNAs ending at positions spanning the region from glutamic acid at position 14 to the final proline 19 of a 2A were generated and used to program translation reactions assembled with wheat germ extract. Ribosomes stalled at the 3' end of truncated transcripts, with nascent chains remaining covalently attached to ribosome-associated tRNA. Further, the ribosomal pausing that occurs during translation of 2A is consistent with the hypothesis that the 2A peptide interacts with the ribosomal exit tunnel. Toe-print signals showed that the ribosome pause when the glycine 18 codon (-DxExNP**G**P-) is at the P site and the proline final codon is at the A site of the ribosome. Replacement of the proline final codon resulted in loss of 2A activity.

It was also shown that translation terminating release factors (RFs) play an essential role in this reaction. Termination of protein synthesis requires two classes of RFs. Class-I release factors (RF1 and RF2 in prokaryotes, and eRF1 in eukaryotes) recognize the three different stop codons (UAG, UGA, UAA) and trigger hydrolysis of peptidyl-tRNA at the ribosomal peptidyl transferase center. Although the function of class-I RFs is similar in prokaryotes and eukaryotes, they exhibit different structural and functional features. eRF1, recognizes all three termination codons, whereas each prokaryotic factor recognizes two out of three stop codons (Scolnick *et al*., 1968). Class-II release factors (RF3 and eRF3 in prokaryotes and eukaryotes, respectively) are guanine-nucleotide-binding proteins possessing GTPase activity. eRF3's activity depends entirely on the ribosome and the eRF1 (Frolova *et a*l., 1996) and it binds to eRF1. eRF3 can recycle both class-1 RFs (mediated by RF3 in bacteria) and ribosomes (mediated by the ribosome recycling factor (RRF) in bacteria) (Kisselev *et al.*, 2003; Mitkevich *et al*., 2006). It has also been observed that RF3 can abort protein synthesis by inducing premature dissociation of peptidyl-tRNA from the ribosome (`drop-off') (Heurgue-Hamard *et al*., 1998), but the mechanism for this side reaction remains unexplained.

Recent experiments where RF activity was altered revealed an influence in the outcome of the 2A reaction both *in vitro* and *in vivo*. Reduced eRF1 levels were accompanied by reduced synthesis of separated upstream and downstream products, consistent with RF catalysing the hydrolytic termination event. Impaired GTP hydrolysis on eEF3 led to increased production of upstream product and reduction in both extension and downstream products (Doronina *et al*. 2008). The way this factors release the protein without recognizing any stop codon is still unknown, but several hypotheses have been suggested.

To date release factors are only known to function at stop codons, and the A site must be empty for them to gain entry into the ribosome. Prolyl-tRNA^{Pro} could be unstably bound to 2A paused ribosomes, possibly due to inability to form a peptide bond or, as a second possibility, the ribosomal conformation imposed by 2A could disfavour entry of the tRNA. Prolyl-tRNA^{Pro} and RF might compete on ribosomes paused by 2A. However, over-expression of tRNA^{Pro} did not reverse the toxic effects of over-expressing 2A in RF-limited cells. Thus the peptidyl(2A)-ribosome interaction and the conformation of the complex disfavours further extension unless RF acts to release the nascent chain or the 2A-ribosome interaction is lost. 2A may also influence and by-pass the necessity for stop codon decoding by RF through directing the ribosome into a conformation similar to that which it takes once RF has bound productively to the A site.

It is interesting to contrast eRF's action with the well-studied tmRNA system in prokaryotes. tmRNA is a specialized RNA involved in trans-translation, a ubiquitous pathway for removing stalled translational complexes from bacterial cells. Alanine-charged tmRNA, in association with the SmpB protein, recognizes stalled ribosomes, binds like a tRNA to the A-site and donates its alanine to the nascent polypeptide chain in a standard transpeptidation reaction. TmRNA then acts as a surrogate mRNA, replacing the defective mRNA with the self-encoded peptide reading frame, to direct translation (*trans*-translation) of the degradation tag. Translation terminates normally at a stop codon provided by the mRNA-like domain of tmRNA. The final translation product of this process carries an 11-residue degradation tag at its C-terminus and thus becomes a substrate for C-terminal specific cellular proteases (Keiler *et al*., 1996; Yakamoto *et al*., 2003). This translation quality control system is only present in prokaryotic organisms, leaving a gap in the

knowledge of the control of eukaryotic stalled ribosomes, making possible the involvement of RFs in this case.

All these new findings need to be more extensively studied to find out how the process occurs, however, it is clear that interaction between the ribosome and 2A must exist, as seen in other systems (see section 1.5.1.5), and the presence of RFs add a new piece to the yet to complete puzzle of the 2A mechanism of action.

1.5.1.5 Ribosome-Nascent peptide interactions

Inside the ribosome there is an exit channel, which is made of RNA and where the nascent peptides are located. During synthesis this original protein, once it reaches a certain length, enters the tunnel.

It has been shown that this tube is not totally neutral for nascent peptides. There can be some interactions between them that modify the ribosomes action, for example the bacteriophage T4 gene 60 (Farabaugh, 1996; Gesteland & Atkins, 1996) and, also, chloramphenicol (Cm)-resistance genes cat and cmlA (Rogers & Lovett, 1994; Harrod & Lovett, 1994; 1995). T4 gene 60 interacts with the ribosome, leading to a ribosomal "hop" from codon 47 to codon 50 downstream (Weiss *et al.*, 1990; Herr *et al.*, 2000). It is also called "translational bypassing" or "subversion of contiguity" because the determinate gene sequence can control the translation in this way. As has been demonstrated it is essential for the peptide to be a certain length (16 amino acids in T4). Thus the peptide reaches the interior of the tunnel and can interact with the ribosome producing the translational bypass (Gesteland & Atkins; 1996). Another striking example of this nascent peptide-ribosome exit tunnel interaction is the one reported by Nakatogawa and Ito (2002). SecA is a protein involved in peptide export, whose expression is regulated by another protein called SecM, which is encoded in an ORF upstream of SecA ORF. SecM contains an export signal at its N-terminus and an effector motif at its C-terminus, which can block protein elongation stalling the ribosome complex. This occurs only when translocation of SecM through the membrane is impaired by low SecA activity. Thus, the translation initiation region for SecA, which is normally 'hidden' by a strong RNA secondary structure preventing initiation of SecA synthesis, is accessible by rearranging of this mRNA secondary structure once the ribosome is stalled. This mechanism creates an efficient intracellular feedback loop for adjusting the supply of SecA proteins to the intracellular demand for protein export. The way this interaction (and others) works is

still unclear but experiments, where mutations where produced in different parts of the ribosome exit tunnel, revealed the interaction to be in the segment of ribosomal protein L22, which is located at the entrance of the exit tunnel (illustrated in figure 16).

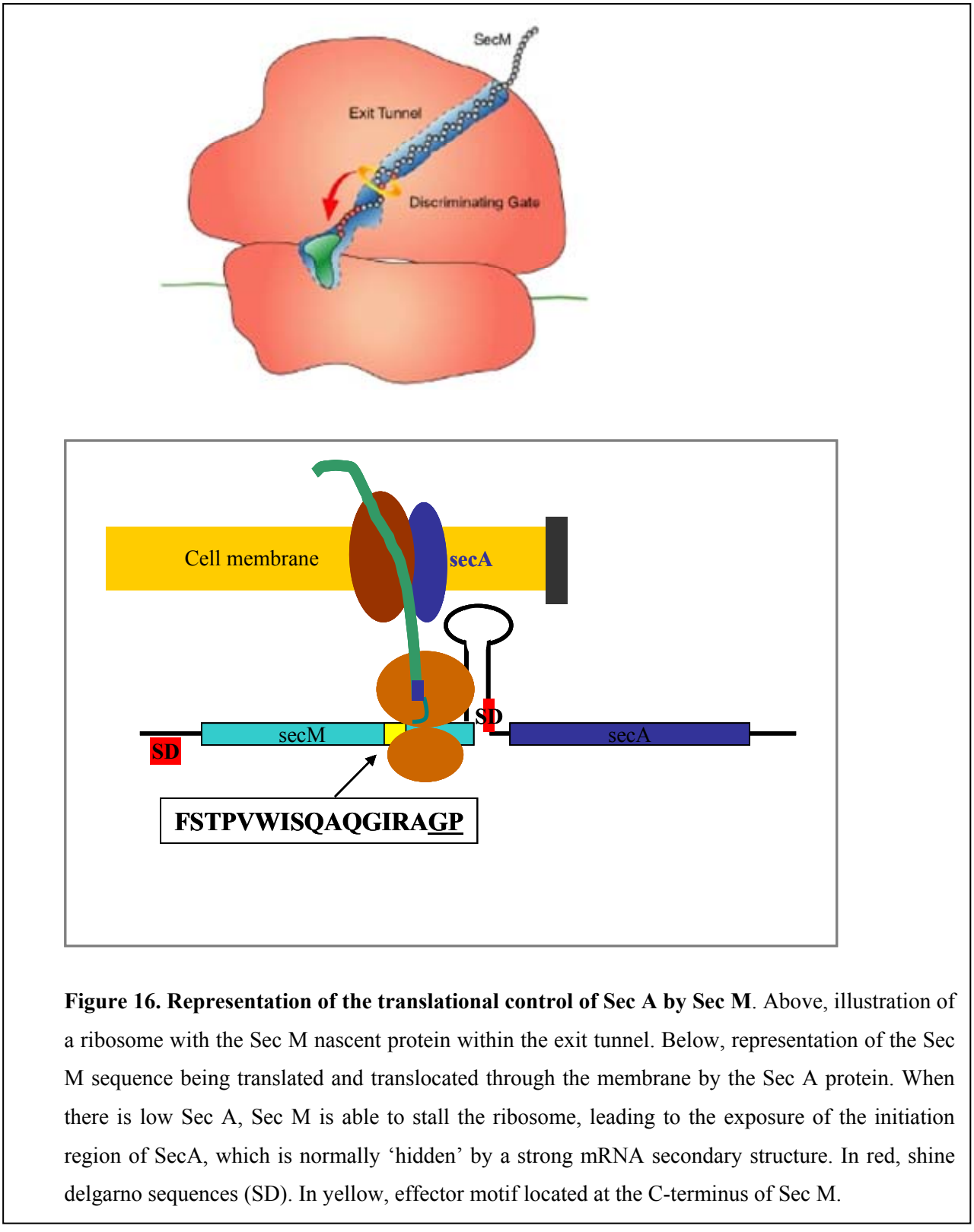


Figure 16. Representation of the translational control of Sec A by Sec M. Above, illustration of a ribosome with the Sec M nascent protein within the exit tunnel. Below, representation of the Sec M sequence being translated and translocated through the membrane by the Sec A protein. When there is low Sec A, Sec M is able to stall the ribosome, leading to the exposure of the initiation region of SecA, which is normally 'hidden' by a strong mRNA secondary structure. In red, shine delgarno sequences (SD). In yellow, effector motif located at the C-terminus of Sec M.

The peptidyl transferase is responsible for catalysis of the transpeptidation reaction in the elongation cycle of translation. In prokaryotes the exit tunnel measures ~80-100Å in length and has a bend some 20-35Å from the peptidyl transferase centre. 2A interacts with the exit channel of eukaryotic ribosomes and this interaction leads to the inhibition of the peptide bond formation between the last amino acid of 2A and the first amino acid of 2B.

The C-terminal –NPG- residues of 2A could play a role in the reorientation of the peptidyl (2A)-tRNA^{gly} substrate to inhibit peptide bond formation and then, stimulate hydrolysis when the A site is occupied by prolyl-tRNA (Donnelly *et al.*, 2001). The length of the suggested helical part of the 2A sequence is 27Å, which could fit inside the exit tunnel of the ribosome.

This translational model presents a plausible explanation for 2A cleavage activity. First of all a structural model of 2A has been developed that consists of an N-terminal helical portion followed by a tight turn fragment at its own C-terminus (represented in figure 17).

The upstream sequence stabilises and extends the helix and increases interaction with the ribosome exit pore. This α-helix, which interacts with the ribosome exit pore, fixes stereochemistry of tight-turn in peptidyl transferase centre and could have a dipole moment. Inside this fragment there is a glutamate possibly involved in interaction with tRNA and inhibition of peptidyl transferase.

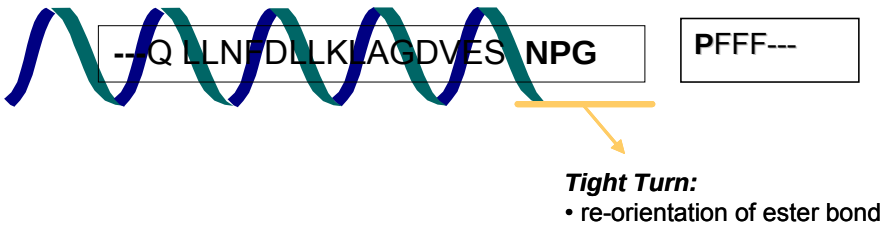


Figure 17: FMDV 2A structural model. 2A sequence is formed by an N-terminal helical portion followed by a tight turn fragment at its own C-terminus.

The highly conserved sequence NPG is proposed to form a tight-turn that reorients the ester bond. Once the peptide bond between proline and glycine (2A fragment NPG) is synthesized, translocation occurs and the 2A peptidyl-tRNA, which was in the A site, is located into the P site. Then, the helical part of 2A interacts with the exit pore and this promotes a specific orientation of the base of the helix within the peptidyl-tranferase centre of the ribosome. Thus, the NPG portion with its tight turn structure reorients the peptide tRNA ester linkage promoting an unusual conformation, different to the usual one that leads to peptide bond formation.

In the model proposed earlier the peptidyl-tRNA ester group would be attacked and thus cut because of the coordination of a water molecule and magnesium ions (Mg^{2+}), which are necessary for the peptidyl-transferase activity. Magnesium ions attach at the base of the 2A helix axis and promote the cleavage in conjunction with a water molecule (Ryan *et al.*, 2004). Thus, when the prolyl-tRNA enters the A site the formation of the peptide bond is not possible because the usual nucleophilic attack performed by prolyl-tRNA upon the electrophilic centre is hampered. The 2A-peptidyl-$tRNA^{gly}$ ester bond is, then, hydrolysed and the nascent polypeptide released (reviewed by Ryan *et al.*, 2002).

Prolyl-tRNA is the poorest nucleophile among all the aminoacyls-tRNAs since its secondary amino group is sterically constrained due to its specific structure in a five membered pyrrolidine ring.

The suggested model represents yet another trick used by viruses to modify the host cell's translation apparatus to their own ends. However, as other modifications such as suppression of termination or leaky scanning, this one has not only been observed in picornaviruses but in other viruses and cellular sequences (see section 1.6).

1.5.1.6 Length of 2A oligopeptide

Although FMDV 2A region was first described as an oligopeptide of only 16 amino acids in length (Roberston *et al*., 1985) its activity was mapped at 19 amino acids (Ryan *et al*., 1991). Further experiments have shown an increase in cleavage efficiency when 2A length is up to 25 amino acids (adding the upstream 1D sequence). 2A activity maps, therefore, not only to the 18 amino acids but 25-30 amino acids. The seven C-terminal amino acids of 2A plus the N-terminal proline of 2A, form the highly conserved motif, the upstream sequence being variable. However,

even if the C-terminal region of 1D is quite variable, it also presents some conserved residues. 2A is cleaved from 1D by $3C^{pro}$, or more efficiently $3CD^{pro}$ (Ryan *et al.*, 1989). This provides an additional constraint on the 2A region.

1.6. 2A-like sequences

Although originally identified in foot-and-mouth disease virus, new active 2A sequences have been found in other picornaviruses, plus quite different RNA viruses. The identification of these functional 2A-like sequences has led to speculation about the possible roles of the 2A in regulation of translation in the different organisms in which the sequence has been found. It also demonstrates that this type of protein biogenesis is not only confined to the picornaviruses but is a more widely adopted strategy.

1.6.1 Insect 2A-like sequences

2A-like sequences were identified in single-stranded RNA insect viruses; *Dicistroviridae* and *Tetraviridae*, and also the double-stranded RNA (dsRNA) viruses: *Reoviridae*. These families are represented in table 2.

The activity of representatives of these 2A-like sequences was tested in artificial polyproteins encoding a single open reading frame (ORF) comprising [GFP-2A-GUS]. *In vitro* transcription-translation reactions were performed to check the efficiency of the cleavage within different species. Some 2A sequences not completely matching the -DxExNPGP- motif were also tested. In contrast to previous experiments (Donnelly *et al.*, 2001a), longer versions of 2A (30aa) were tested to more closely represent the situation *in vivo* (Luke et *al.,* 2008).

Single-stranded RNA insect viruses contain the *Dicistroviridae* (cripaviruses), the unassigned iflaviruses and tetraviruses. Three species belonging to iflavirus (*infectious flacherie virus* (IFV), *Perina nuda picorna-like virus* (PnPV) and *ectropis obliqua picorna-like virus* (EoPv)) were tested and showed ~99% cleavage activity, in contrast with the previously tested shorter versions with only ~63% active cleavage (Donnelly *et al.,* 2001a). Interestingly, PnPV and EoPV have a second 2A between the structural VP2 and VP4 proteins (shown in figure 18) that is also highly efficient. In Cripaviruses, whose 2A is in the N-terminal region of the replicative proteins open reading frame (figure 18), 2A is highly efficient in agreement with earlier reports in

drosophila C virus (DCV) and *acute bee paralysis virus* (ABPV), except in *cricket paralysis virus* (CrPV), where cleavage was not very efficient and is hardly improved in the longer 2A.

The *Tetraviridae* has 2 genera, betatetravirus (*euprosterna elaeasa virus* (EeV) and *providence virus* (PrV), which contains three 2As (figure 18) and omegatetravirus (*thosea asigna virus* (TaV)). The tested members of this family also showed high cleavage activity. PrV $2A_3$, which is situated at the N-terminus of the structural open reading frame, and $2A_1$, within a non-structural open reading frame, cleave very efficiently; whereas $2A_2$ showed lower activity.

The *Reoviridae* is a family of segmented dsRNA viruses, which contains 13 genera that affect a wide host range from humans to bacteria. 2A has been found in two genera within this family, rotavirus, which are mammalian 2A-like sequences (see section 1.6.2) and cypovirus (*Bombyx mori cypovirus 1* (BmCPV-1) and *Operophtera brumata cypovirus 18* (OpbuCPV-18). These 2A-like sequences are located on the segments encoding non-structural proteins and are highly active (Luke *et al.*, 2008).

Phylogenies were constructed from polymerase domains and 2A-like sequences. Examination of the distribution of those viruses possessing 2A within the phylogeny lead the authors to conclude that 2As in different viruses can either be homologous (common ancestral origin), or, homoplasic (arising from multiple, independent, evolutionary origins). The latter is not surprising due to their short length and the location of these sequences in known recombinational hot-spots (Luke *et al.,* 2008).

Insect viruses with 2A-like sequences		
Family	Genus	*Species*
Unassigned	Iflavirus	*Ectropis obliqua picorna-like virus (EoPV)* *Infectious flacherie virus* (IFV) *Perina nuda picorna-like virus* (PnPV) *Varroa destructor virus-1* (VDV-1)
Dicistroviridae	Cripavirus	*Cricket paralysis virus* (CrPv) *Drosophila C virus* (DCV) *Acute bee paralysis virus* (ABPV) *Kashmir bee virus* (KBV) *Israel acute paralysis virus of bees* (IAPV)
Tetraviridae	Betatetravirus Omegatetravirus	*Providence virus* (PrV) *Euprosterna elaeasa virus* (EeV) *Thosea asigna virus* (TaV)
Reoviridae	Cypovirus	*Bombyx mori cypovirus type1* (BmCPV-1) *Operophtera brumata cypovirus-18* (OpbuCPV-18)

Table 2. Classification of insect viruses containing 2A-like sequences.

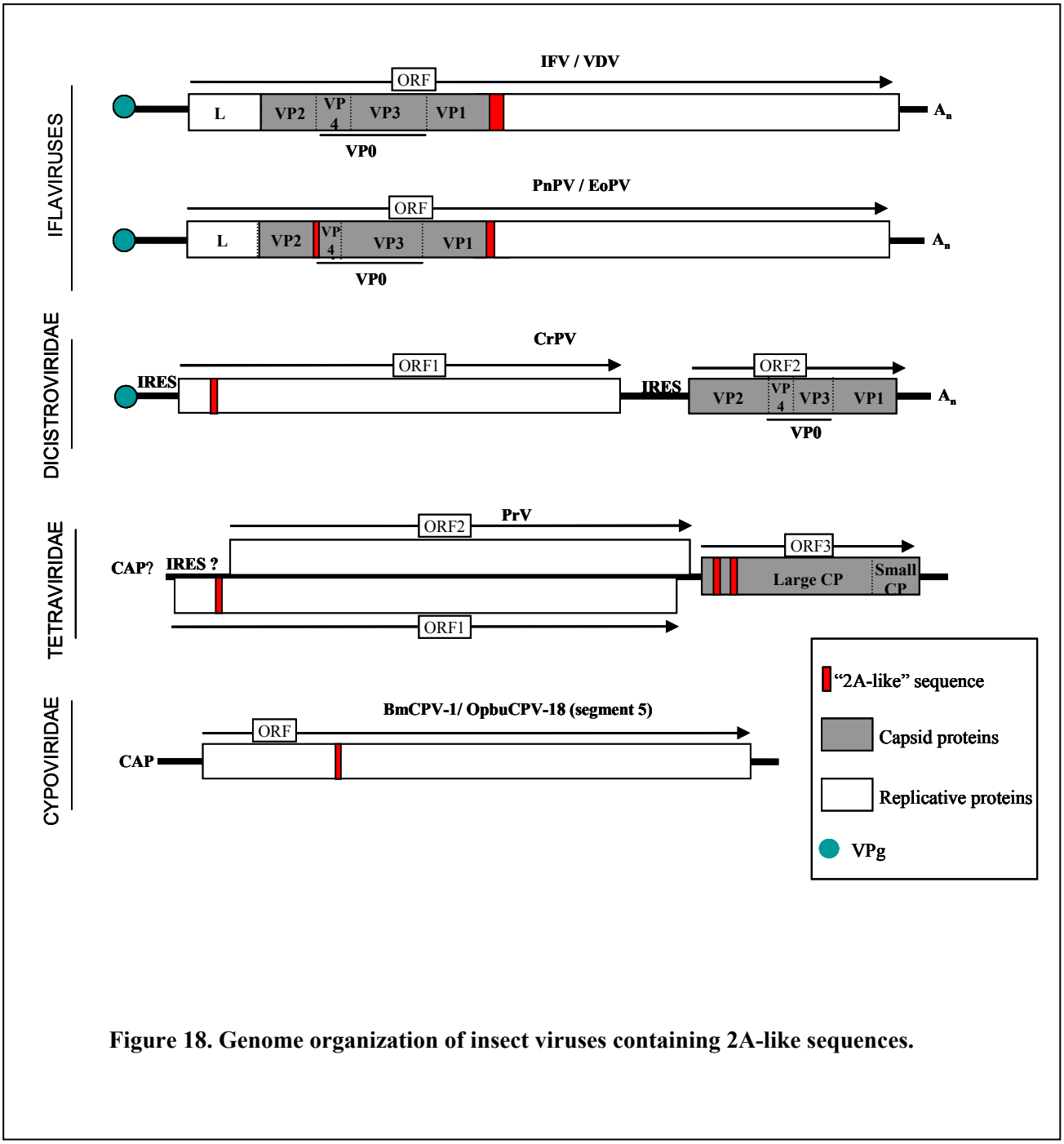


Figure 18. Genome organization of insect viruses containing 2A-like sequences.

1.6.2. Other 2A-like sequences

2A-like sequences are also present in the genome of four species of the eukaryotic unicellular parasite Trypanosoma, *T.cruzi*, *T.brucei, T.congolense and T.vivax.* These 2A-like sequences are located at the N-terminus of the non-LTR retro-element, being L1Tc in *T.cruzi* and ingi in *T.brucei*. 20% of L1Tc-2A sequences contain a histidine residue instead of asparagine at position 17 (**N**^{17}→**H**; **H**PGP) that is highly active in common with the normal motif *in vivo* (Heras *et al.*, 2006).

A 2A-like sequence was also identified in the eubacterial α-glucuronidase augA gene of *Thermatoga maritima*, a hyperthermophilic bacterium, but does not show any activity (Donnelly *et al.,* 2001a). No active 2A-like sequences have been identified in prokaryotic organisms (Donnelly *et al.,* 1997). It should be noted that the exit tunnels of prokaryotic and eukaryotic ribosomes bear substantial differences (see discussion section 4.1). This could explain why there are no active 2As in bacteria and supports the hypothesis that 2A needs to interact with eukaryotic ribosomes to be active.

Rotaviruses, which belong to the family Reoviridae, consist of *bovine rotavirus C* (BoRV) *porcine rotavirus C* (PoRV), *human rotavirus C* (HuRV) and the *new adult diarrhoea virus* (ADRV-N), which has recently been added to the family and is closer to group B than group A or C rotaviruses (Yang *et al.*, 2004). The genome of rotaviruses is composed of 11 segments of dsRNA that encode six structural and five non-structural proteins (James *et al.,* 1999). The type C rotavirus gene 6 encodes the non-structural protein 3 (NSP3; NS34), which is an ssRNA-binding protein equivalent to the group A rotaviruses (Qian *et al.,* 1991). 2A-like sequences are present on the segment 6 in bovine, porcine and human rotavirus C, whereas in ADRV-N is encoded in segment 5 with the non-structural protein 1 (NSP1). In type C rotaviruses, 2A links the NSP3 with the dsRNA-binding protein, which sequesters viral dsRNA from the cellular sensors of dsRNA, neutralizing the activation of the cellular antiviral interferon system (Lagland *et al.,* 1994). NSP3 also binds the non-polyadenylated 3' end of the rotavirus mRNAs, enhancing its translation and subverting the host translation machinery (Piron *et al.,* 1999; Jayaram *et al.,* 2004). *In vivo* and *in vitro* translation analysis on [NSP3-2A-dsRBP] showed three products, a small amount of the uncleaved protein [NSP3-2A-dsRBP] and the two processed products [NSP3-2A] and [dsRBP] in equimolar amounts (Lagland *et al.,* 1994). It is remarkable that NSP3 forms dimers, which may add a further

complexity since NSP3 could form heterodimers with [NSP3-2A-dsRBP]. Analyses of 2A- mediated cleavages showed that the 2A-like sequence in segment 5 ADRV-N was highly active, whereas in segment 6 of PoRV and HuRV it was lower. This lower cleavage efficiency is involved in translational control of products and allows the production of a complex array of products at high levels (Luke *et al.*, 2008).

Penaeid shrimp infectious myonecrosis virus (IMNV) was recently isolated from farmed *Penaeus vannamei* in north-eastern Brazil and classified as a member of *Totiviridae,* which is a family of non-segmented dsRNA viruses. Their genome contains two overlapping ORFs, ORF1, which encodes a 179 kDa protein that includes the N-terminal sequence of the major capsid protein (MCP) and ORF2, which encodes a 85 kDa protein that contains a series of motifs characteristic of an RNA-dependent RNA polymerase (RdRp) (Poulos *et al*., 2006). New examinations of the IMNV genome revealed the presence of two 2A-like motifs located at the N-terminal third of ORF1 preceding MCP and probably involved in co-translational processing of the IMNV polyprotein into consecutive 93, 284 and 1228 amino acids fragments. The fragment 1228 would be further cleaved co- or post-translationally by a different mechanism to yield a 327 fragment and the 901 amino acids MCP. The functions of the 93 and 284 and 327 amino acids fragments of ORF1 are still unknown although the 93 amino acids fragment encompasses the N-terminal region previously noted, by Poulos *et al*. (2006), to share sequence similarities with double-stranded RNA-binding proteins (Nibert, 2007). IMNV 2A-like proteins have been tested and are highly active (Luke *et al.*, 2008).

The recent determination of the genome sequence of a marine invertebrate, the purple sea urchin *Strongylocentrotus purpuratus* (Sea Urchin Genome Sequencing Consortium, 2006), revealed ~150 2A-like sequences present within a range of genes. These sequences are found in two types of cellular sequences. The first type are genes involved in innate immunity, such as, nucleotide-binding oligomerization domain/leucine-rich repeat family (NOD-LRR), NACHT domain/leucine-rich repeat (NACHT-LRR family) or 'CATERPILLER' (CARD, transcription enhancer, R (purine)-binding, pyrin, lots of LRRs) genes. The second group of genes are associated with non-LTR retrotransposons, although is still unclear if 2A is part of these elements or forms a genomic site favourable for their transposition. All of the 2A sequences tested mediate ribosome 'skipping' except one that has a very short N-terminal tract. This is consistent with previous analysis of truncated FMDV 2A forms

(Ryan *et al.*, 1994). Although many of them still need to be tested, results obtained recently show 2A as an essential mechanism for the control of protein biogenesis, not only in viruses but also in cellular organisms. (Illustrated in figure 19).

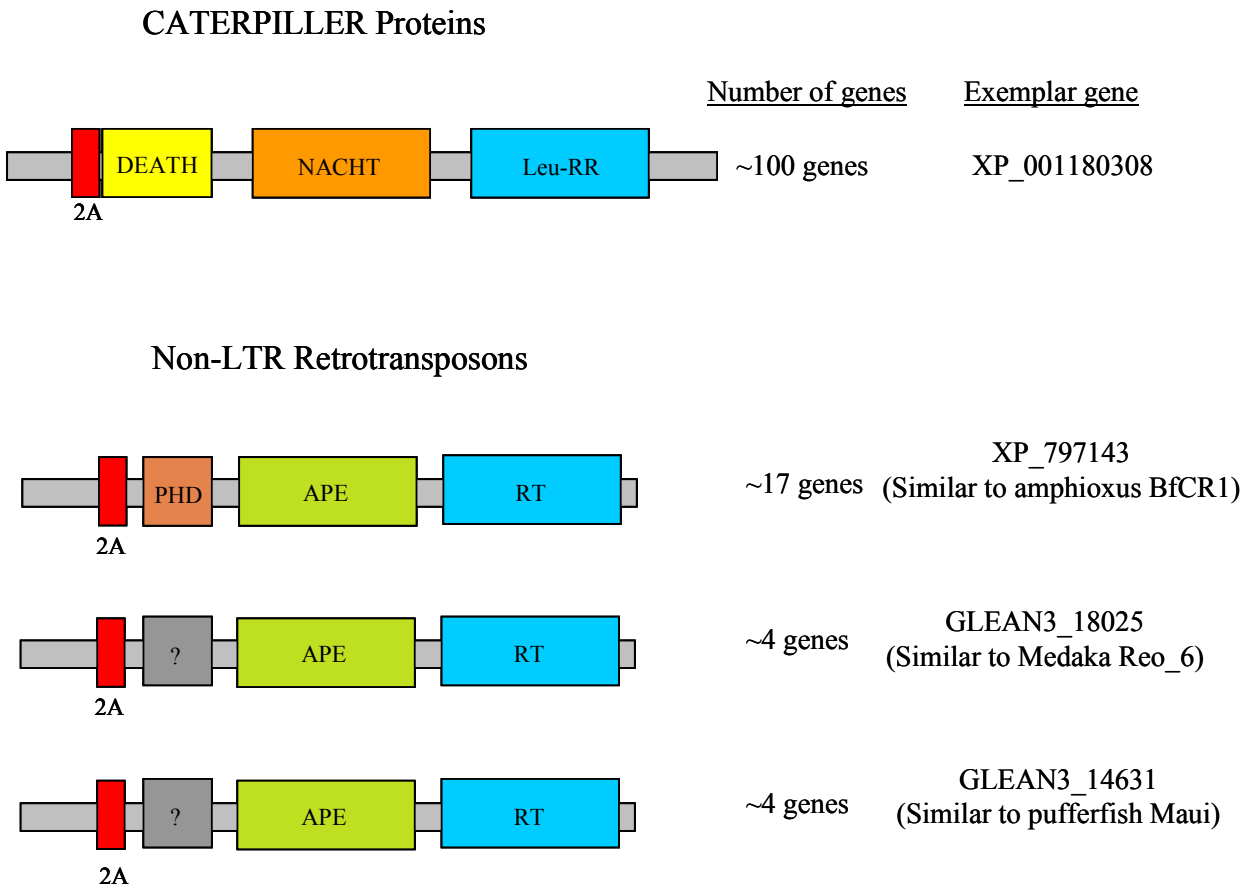


Figure 19. Schematic representation of the two major groups of genes (CATERPILLER proteins and Non-LTR retrotransposons), where 2A is found within the *S. pupuratus* genome. Protein domains are indicated in boxed areas. Grey areas indicate other part of the genome that do not correspond to these two group of genes (? = unidentified domains).

1.6.3 Picornavirus 2A-like sequences

New 2A sequences within picornaviruses include: bovine rhinovirus 2 (BRV-2), Theiler-like virus of rats (T-LV), human Saffold virus (SAF-V), porcine teschovirus (PTV), Ljungan virus, Seneca valley virus (SVV) and duck hepatitis virus (DHV1).

The activity of these 2A-like sequences was tested in artificial polyproteins encoding a single open reading frame (ORF) comprising [GFP-2A-GUS]. In common with the insect virus experiment (see 1.6.1) longer versions of 2A (30 aa) were also tested. The longer 2A versions within the *Picornaviridae* showed highly efficient cleavage. Furthermore, a change in the conserved motif (Ser→Pro; -DVEPNPGP-) showed efficient cleavage (Luke *et al.,* 2008).

By studying more functional 2A-like sequences and comparing them with FMDV 2A oligopeptide, we will be able to achieve a better understanding about this protein and its mechanism of action.

1.6.3.1 TMEV 2A

Like FMDV 2A, the TMEV 2A is a polypeptide responsible for the autocatalytic, co- translational primary cleavage of the polyprotein (Ryan *et al*., 2002). It cleaves between LP12A and 2BCP3 and although it is approximately 144 amino acids longer than FMDV 2A, both act in the same way. It also shares the highly conserved sequence at its C-terminus (-DxExNPGP-), which is responsible for the cleavage (Donnelly *et al.,* 2001b). The upstream sequence of TMEV 2A diverges remarkably from that of EMCV and its function is not known (Donnelly *et al.,* 1997).

1.6.3.1.1 Studies on TMEV 2A

A plasmid with the C-terminal part of Theiler´s murine encephalomyelitis virus (TMEV) 2A together with the N-terminal proline of 2B, inserted between chloramphenicol acetyltransferase (CAT) and β-glucuronidase (GUS) was constructed. This particular recombinant polyprotein was able to mediate cleavage with a very high efficiency (85%). As before, this means that no other viral protein domain is contributing to the cleavage event. In addition, it is not necessary to have all the upstream 2A sequence to make the cleavage possible. However, the efficiency can be significantly reduced when the length of the 2A sequence is decreased. Thus, this N-terminal

sequence influences cleavage but it is not essential. In some cases, the efficiency is hardly reduced (Donnelly *et al*., 1997).

The variability of the 2A protein among picornaviruses is remarkable: this small region of the genome is highly plastic compared to the other non-structural proteins, which have homologous structures and functions. The 2A self-cleaving peptide, first discovered in FMDV, has also been found in the majority of picornavirus genera. Something that seemed to be very particular has turned out to be widely spread between different organisms, except bacteria. The mechanism of 2A action is becoming much clearer but there are still lots of questions to answer, such as; why is it not active in bacteria? How this interaction between the nascent peptide and the ribosome occurs and how the peptide bond formation is avoided without the presence of a stop codon? By studying all these new 2A sequences, which present the same C-terminal motif but variable upstream sequences, and comparing them, we should be able to increase our knowledge of how 2A functions.

We now know that the presence of the –DxExNPGP- motif alone does not confer cleavage activity, for example, it is not active in *Thermatoga maritima* aguA gene. This conserved motif, found in all 2As, must be accompanied by an appropriate upstream context to be active. The combination of this conserved motif with a determinate upstream sequence may make the interaction between the sequence and the ribosomal exit tunnel suitable and, thus the process to occur. Recent work has proved the involvement of release factors in this reaction. Further mutational studies, will expand our understanding of this 'ribosome skipping' process and will fill some of the gaps in our knowledge of this unique mechanism of action.

2. MATERIALS AND METHODS

2.1 Cloning:

2.1.1 Polymerase Chain Reaction (PCR)

PCR was used to amplify segments of DNA, introduce mutations and insert 2A-like sequences. Typically, reactions were catalysed by Taq DNA polymerase (Promega Ltd., Southampton, UK) containing 10 x buffer (330mM Tris-acetate, pH 7.9, 660mM potassium acetate, 100mM magnesium acetate, 5mM DTT). This reaction was carried out in a final volume of 50µl using 80pmol of the forward and reverse primer, 5ng of template DNA and dNTPs at a final concentration of 250 µM. Each round of PCR synthesis involves three steps: denaturation (94ºC for 1 minute), annealing (56ºC for 30 seconds) and elongation (72ºC, 2 minutes per kb of template). This three-step 'PCR-cycle' was repeated 35 times. The termination is the final step, where elongation is carried on at 72ºC for 10 minutes.

All PCR products were purified using the Wizard PCR Prep purification system (Promega Ltd., Southampton, UK) as per the manufacturers instructions.

2.1.2 Preparative Restriction Enzyme Digests

Restriction enzymes (purchased from Promega and New England Biolabs) were used following the conditions specified by the suppliers.

Typically, 1µg of DNA was digested in 1 unit of enzyme, to a final volume of 20 µl including 2 µl of 10 x restriction buffer. The reaction was incubated at the optimum temperature for the specific enzyme (normally 37ºC) for 2 hours.

2.1.3 Analytical Restriction Enzyme Digests

Typically, 0.4 µg of DNA were digested with 0.2 units of enzyme, in a total volume of 10 µl, including 1 µl of x10 restriction buffer. The reactions were incubated at the optimum temperature indicated for the specific enzyme for 1 hour.

2.1.4 Agarose-gel Preparation

0.8 % (w/v) agarose gels were prepared with 50 x TAE buffer (242 g Tris base, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA, pH) with a final addition of 4 µl of ethidium bromide (5mg/ml) to a final volume of 100 ml. The 50 x TAE running buffer was diluted to 1 x with distilled water when preparing the gel.

2.1.5 Gel Electrophoresis

DNA samples were prepared by adding 6 x agarose gel loading buffer (50% [v/v] glycerol, 0.005% [w/v] bromophenol blue) and loaded onto the gel. The electrophoresis was performed at 100 V in 1 x TAE buffer (40mM Tris-acetate, 1mM EDTA). DNA bands were subsequently visualized by illumination using a UV transilluminator.

2.1.6 Purification of DNA fragments from Agarose gel

The bands of interest were localized and the area in front of them excised and filled with low-melting-point agarose (LMP). The band was run into the LMP agarose and cleaned using the Wizard Prep DNA Purification system (Promega Ltd., Southampton, UK) as per the manufacturers' instructions.

2.1.7 Ligations

Typically, ligation reactions were set up in a final volume of 10 µl using DNA ligase (Promega Ltd., Southampton, UK) and 10 x buffer (300mM Tris-HCl, pH 7.8, 100mM $MgCl_2$, 100mM DTT, 10mM ATP). The concentrations of vector and insert were estimated during electrophoresis along with molecular weight standards of a known concentration. Normally, the concentrations used were 50 ng of vector DNA and an equal to 3-fold concentration of insert. The reactions were incubated at 4ºC overnight.

2.1.8 Preparation of E. coli (JM109) by Calcium Chloride Method

1 ml of a JM109 overnight culture was added to 50 ml in a LB broth flask and incubated shaking at 37ºC until the OD_{600} of 0.3 was reached. The cells were then cooled to 0ºC for 10 minutes in ice cold polypropylene tubes and centrifuged at 4000rpm for 10 minutes at 4ºC. The pellets were re-suspended in 10 ml of ice-cold 0.1 M calcium chloride after the supernatants were discarded. Then, the samples were

kept on ice for 15 minutes and centrifuged 3,000 rpm for 10 minutes to recover the pellets. Pellets were re-suspended in 1 ml of ice cold 0.1M calcium chloride and stored on ice for 30 minutes.

20 µl of ligation was added into 200 µl of the prepared cells in a 0.5 ml eppendorf and kept on ice for 30 minutes. The cells were heat shocked in a water bath at 42ºC for 50 seconds and then, placed on ice for 2 minutes. After this, 950 µl of LB medium (10g Bacto-tryptone, 5g Bacto-yeast extract, 5g NaCl, deionized water, pH 7.5) was added to the mixture and incubated with shaking at 37ºC for 1 hour. Cells were then spread on agar plates supplemented with a suitable antibiotic.

2.1.9 Transformation of Competent E. coli (JM109) cells

DNA constructs were transformed in competent *E. coli* cells JM109 (purchased from Promega or prepared by the Calcium Chloride technique), by adding 50 µl of cells to the 10 µl ligation reaction in a 1.5 ml microcentrifuge tube. The tubes were flicked gently to mix the contents and placed on ice for 15 minutes. Heat-shock was carried out at 42ºC in a water bath for 50 seconds. The tubes were immediately placed on ice for 2 minutes. After this, 950 µl of LB medium (10g Bacto-tryptone, 5g Bacto-yeast extract, 5g NaCl, deionized water, pH 7.5) were added to the mixture and placed in a shaking incubator at 37ºC for 1 hour. Cells were spread on agar plates supplemented with a suitable antibiotic and incubated at 37ºC overnight.

Individual colonies from plates were picked, the following day, placed in universals with 5 ml of LB medium containing antibiotic, and incubated overnight at 37 ºC with moderate shaking.

2.1.10 Mini-preparation of plasmid DNA

Plasmid DNA was extracted from cells using Wizard Plus SV Minipreps DNA purification System (Promega Ltd., Southampton,UK).

2.1.11 Midi-preparation of plasmid DNA

Midi-preps were prepared using the Midi-prep Pure Yield^{TM} Plasmid Midiprep System (Promega Ltd., Southampton, UK) as per the manufacturers instructions. A colony was grown in a universal containing 25 ml of LB and the appropriate amount of antibiotic and incubated overnight with shaking at 37ºC.

2.1.12 DNA sequencing

600 ng of DNA, in a final volume of 15 µl, was sent to the Sequencing Service MSI/WTB Complex (School of Life Sciences, University of Dundee). T7 and SP6 promoter primers were normally used for sequencing and provided by the sequencing service. All clones were verified by the Sequencing Service MSI/WTB Complex (School of Life Sciences, University of Dundee).

2.1.13 Alkaline Phosphatase, Calf Intestinal treatment (CIAP)

Removal of 5' phosphate groups from DNA was performed using CIAP (purchased from Promega Ltd., Southampton, UK) to avoid self-ligation and re-circularization of linearized plasmids. This enzyme is active on 5' overhangs, 5' recessed and blunt ends.

A reaction containing linearized DNA (up to 10 pmol of 5' ends), 10 x CIAP buffer and CIAP enzyme (0.01 u/µl) was incubated at 37ºC for 30 minutes in a final volume of 250 µl. The reaction was stopped by adding 0.8 µl of EDTA (0.5M) and, subsequently purified using the Wizard SV gel and PCR clean up kit (Promega Ltd., Southampton, UK) as per manufacturers instructions.

2.1.14 TOPO® cloning

PCR products were inserted into pcDNA3.1/V5-His-TOPO® vector (Invitrogen Ltd, Paisley, UK) (shown in figure 20). This vector was linearized with single 3' thymidine (T) overhangs for TA cloning. The BIO-X-ACT™ DNA polymerase (Bioline Ltd, London, UK) used for the PCR assays has a nontemplate-dependent terminal transferase activity that adds a single deoxyadenosine (A) to the 3' ends of PCR products. This allows PCR inserts to ligate efficiently with the TOPO vector.

The TOPO cloning reaction was carried out in a final volume of 6 µl containing 3 µl PCR product, 1 µl salt solution, 1 µl sterile water and 1 µl TOPO vector. After 30 to 60 minutes of incubation at room temperature (22-23ºC), transformation was carried out using *E. coli* competent cells (JM109).

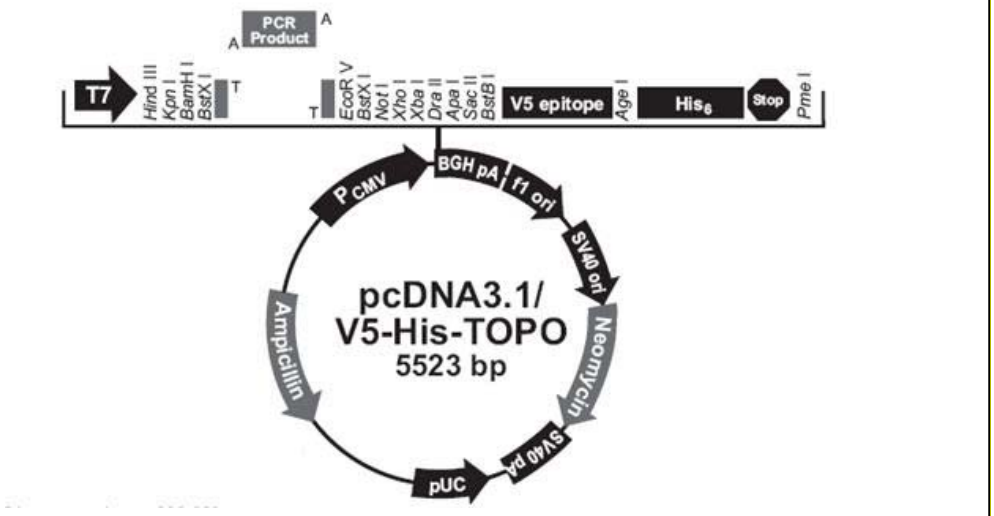


Figure 20. Schematic representation of the pcDNA 3.1/V5-His-TOPO vector. Vector diagram showing positions of the multiple cloning site. (pcDNA 3.1/V5-His-TOPO vector map; www.invitrogen.com)

2.1.15 *pGEM®-T Easy cloning*

PCR products were inserted into pGEM®-T Easy Vector (Promega Ltd., Southampton, UK) (represented in figure 21). This vector was prepared by cutting with EcoRV and adding a 3´terminal thymidine to both ends, creating single 3' T overhangs, which improves the efficiency of ligation and prevents recircularization. The multiple cloning site is flanked by restriction enzymes sites BstZI, NotI and EcoRI, providing three options for removal of the insert with a single digest.

The ligation reaction can be preformed in 1 hour at room temperature. Furthermore, this vector contains T7 and SP6 RNA polymerase promoters flanking the multiple cloning region inside the α-peptide coding region of β-galactosidase. Thus, there will be inactivation of the α-peptide if the PCR product is inserted correctly. This allows the possibility of selection of recombinant clones by colour screening through the use of indicator plates that contain IPTG (Isopropyl β-D-1-thiogalactopyranoside; 0.1M) and X-Gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside; 50 μg/μl).

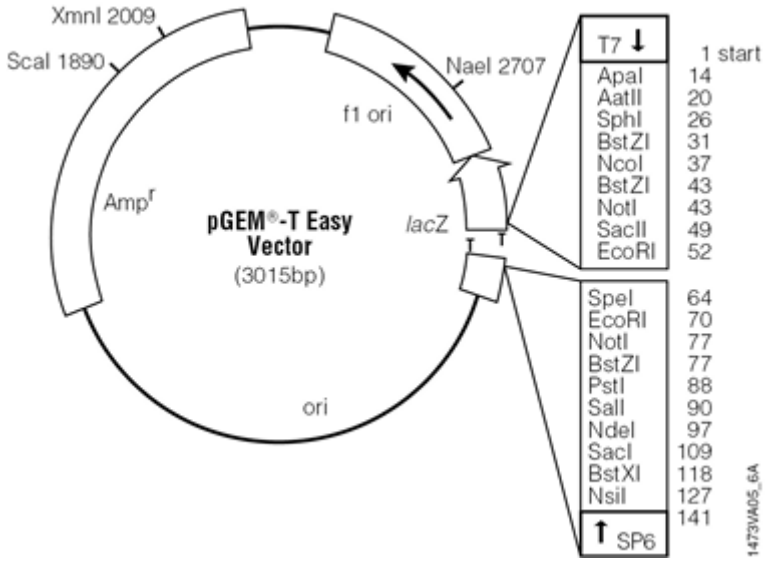


Figure 21. Schematic representation of the pGEM-T Easy expression vector. Vector diagram of pGEM-T Easy expression vector showing positions of the multiple cloning site. **(**pGEM-T Easy expression vector map; www.promega.com)

2.1.16 Oligonucleotide annealment

100mM stock solutions of the oligonucleotide adapters were prepared. 5 µl of each oligonucleotide were mixed together and boiled at 95ºC for 5 minutes to destroy any secondary structures and then cooled down at room temperature to allow the primers to anneal. A series of dilutions of the annealed primers were prepared (1:10, 1:100 and 1:1000) and ligated into a suitably restricted vector.

2.1.17 Primer sequences

Oligonucleotides used for cloning are presented in the following table.

Primer name	Sequence (5'→ 3')
HERf	5' GGATCCGGAGCTTGGTCTAGACAGCTGTTGAATTTTGACCTTCTTAAGCTTGCGG GAGACGTCGAGTCCAACCCCGGGCCCCTGTTGCCCGTCTCACTGGTGAAA 3′
HERr	5' ATAGAGATTCGGGATTTCGGCGCTCC 3′
pGEMV5f	5' GGTTTCTTAGACGTCAGGTG GCACTTTTCGGG 3'
pGEMV5r	5' TGATGAGATCTGTAGAATCGAGACCGAGGAGAGGGTTAGGGATAGGCTTACCTGA GTATTCTATAGTGTCACCTAAATAG 3'
ApepV5f	5' CATCATGGGCCCATGACCATGATTACGCCAAGCT 3'
ApepV5r	5' ATGATGCCATGGTTACAATTTCCATTCGCCATTCAGGC 3'
HEPf	5' CCATGGAA T CCCCTTCTCTCTACCGCATTGATCTT 3'
HEPr	5' GGATCCTCAGCCTGG GTTCATTTCTACATCATGTAT 3'
Lh135f	5' CATCATGCTAGCACCATGGGAAAGCCGATCCCAAACCCT 3'
Lh135r	5' ATG ATGGGATCCGGGCCCAGGGTTGGACTCGACGTCTCCCGC 3'
ScFvf	5' CATCATGGATCCGCCGATGTGCAGGTGGTGGAGTCAGGGGGAGGC 3'
ScFvr	5'ATGAT GAAGCTTTTACCGTCTTATTTCCAACTTTGTCC 3'
CFPf	5' CATCATGGATCCGTGAGCAAGGGCGAGGAGCTGTTCACC 3'
CFPr	5' ATGATGTCTAGACAGGTCTTCTTCTGAGATGAGTTTTTGTTCCTTGTACAGCTCGT CCATGC 3'
Str-14f	5' TTCGCCGCTAGCTCTAGAACCACTGATGATCCTGCCATGCAAGAGAGCACCTGCC TCCCTGAGATGATCATGGTAAAGGCTGGAGATGTAGAGCTAAATSCAGGGCCCTTT GGCCACCGTCGGCG 3'
Str-34f	5' TTCGCCGCTAGCTCTAGAAACTCAACCCCTGCAGCCATGTTTGTGTGCGTGTTCAT ACTGATATCAGTATTGCTACTGAGTGGTGATGTGGAAATAAGTSCCGGGCCCTTCC TGGCCACCGTCGGCG 3'
ΔPACrev	5' GGGCCTAAGCTTGAATTCTTACGTAGAATCGAGACCGAGGAGAGGGTTAGGGATA GGCTTACCCATCATCATCATGGCACCGGGCTTGCGGGTCATGCAC 3'

Table 3. Primers used for cloning. Names and sequences of all the primers used for cloning. **S** = C or G. C being the wild type version and G the mutated version.

2.2 Analysis of translation profiles

2.2.1 Antibodies

Antibodies were used in Western Blotting, Immuno-precipitation and Immuno-fluorescence analysis described in the following sections. The primary antibodies used in this study are shown in Table 4.

The following antibodies were used as secondary antibodies:

- ELC anti-mouse IgG-peroxidase from sheep (Amersham Biosciences, Buckinghamshire, UK)
- Anti-rabbit IgG peroxidase goat (Sigma-Aldrich company LDT. Dorset, UK)
- Texas red goat anti-mouse (Invitrogen, Ltd, Paisley, UK)
- Texas red goat anti-rabbit (Invitrogen Ltd, Paisley, UK)

Antibody	Target Protein	Source
Anti-β-Galactosidase, Rabbit IgG fraction	β-Galactosidase	Invitrogen Ltd, Paisley, UK
Anti-GFP (mAb)	GFP, CFP	Roche Diagnostics Ltd., Burgess Hill, UK
Anti V5-tag (mAb)	V5 protein	Kindly provided by Prof. Rick Randall (University of St. Andrews)
Anti-2A rabbit polyclonal	$2A_{TAV}$	Kindly provided by Dr. Dario Vignali (St. Jude Faculty, Memphis)
Anti-2A rabbit polyclonal	$2A_{FMDV}$	Kindly provided by Dr. Garry Luke (University of St. Andrews)
Anti-Ub rabbit monoclonal	Ubiquitin	Kindly provided by Dr. Dimitris Xirodimas (University of Dundee)

Table 4. Primary antibodies and target proteins.

2.2.2 Denaturing Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Analysis of protein mixtures was carried out under denaturing (SDS) conditions using discontinuous buffer systems similar to that developed by Laemmli (1970). SDS-PAGE gels consisted of an upper layer of 4% polyacrylamide stacking gel and a lower layer of 12% or 15% polyacrylamide resolving gel.

A Hoefer mini-gel apparatus were used to run the gel using a constant current of 40mA. Then, the gel was incubated in Coomassie brilliant staining solution (0.2% [w/v] Coomassie brilliant blue (R-250), 20% [v/v] methanol, 10% [v/v] acetic acid) to stain the proteins.

2.2.3 Preparation of samples to run on SDS-PAGE

Proteins were obtained by preparing a starter culture consisting of the inoculation of a single colony of *E.coli*, transformed with the plasmid of interest, into 2 ml LB medium containing the appropriate antibiotic and grown overnight at 37ºC with moderate shaking. This starter culture was diluted 1:10 and grown at 37ºC with moderate shaking to an A_{600} of 0.6. Protein expression was induced by adding IPTG to a final concentration of 0.5mM. The incubation was continued for 4-6 hours, and the culture harvested by centrifugation at 13000rpm for 10 min at 4ºC. The supernatant was transferred to a fresh tube and treated with Lysis Buffer (Cell lytic B bacterial lysis reagent, protease inhibitors cocktail, Benzonase (240u/µl) (Cell Lytic B Kit, Sigma-Aldrich company LDT. Dorset, UK). 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, 100mM Tris, pH 6.8) was added to the reactions.

Samples were analyzed in SDS-PAGE gels, loading 5 µl of the reaction with loading buffer (see 2.2.1).

2.2.4 Coupled in vitro Transcription/Translation reactions (TNT)

Proteins were expressed *in vitro* using TNT® Coupled Reticulocyte Lysate Systems or Wheat Germ Extract systems (Promega Ltd., Southampton, UK) for eukaryotic expression. Radiolabelling of the proteins was performed using [^{35}S]-methionine (10µCi/µl). Reactions were incubated at 30ºC for 90 minutes.

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, 100mM Tris, pH 6.8) was added to the reactions.

5 µl aliquots of the translation reactions were analyzed by denaturing SDS-PAGE, loading 5 µl of the reaction with loading buffer (see 2.2.1).

2.2.5 In vitro Immune precipitation (IP)

Proteins expressed *in vitro* (see 2.2.3) were incubated with an adequate antibody for 2 hours at 4ºC. This and subsequent incubations were carried out on an orbital shaker. Protein A/G Sepharose beads (Amersham Biosciences, Buckinghamshire, UK) were washed three times with PBS to remove the ethanol in which they are supplied. The beads were recovered from washes by centrifugation at 12,000 x g for 10 seconds. A 50% slurry was prepared by adding an equal volume of PBS and stored at 4ºC. 20 µl of this slurry was added to the antibody/protein mixture and incubated at 4 ºC for 2 hours. The beads were recovered by centrifugation at 12,000 x g for 10 seconds and re-suspended in 2x SDS-PAGE sample loading buffer (2%[w/v] SDS, 20% [v/v]glycerol, 2% [v/v] β-mecaptoethanol, 0.2% [w/v] bromophenol blue, 100mM Tris, pH 6.8). Samples were boiled at 95ºC for 2 minutes and centrifuged at 12,000 x g for 10 seconds. The supernatant was analyzed on SDS-PAGE (see 2.2.1 and 2.2.5).

2.2.6 Visualization of Radiolabelled Translation products

The distribution of radiolabel between the translation products was visualized by autoradiography or phosphor-imaging using FUJIX Bio-imaging analyzer.

2.2.7 Western blot analysis

Polypeptides were transferred to PDVF membranes (ImmobilonTM-P Transfer membranes, Millipore Corporation, UK), directly after SDS-PAGE electrophoresis. The membrane was methanol-activated for 5 minutes, rinsed with dH_2O and incubated in transfer buffer (2.9g glycine, 5.8g Tris base, 0.37g SDS, 200 ml methanol in a final volume of 1 litre).

A blotting sandwich was prepared as follows: 1 fibre pad, 1 Whatmann 3 MM paper, membrane, protein gel, 1 Whatmann 3 MM paper, 1 fibre pad. Both fibre pad

and Whatmann 3 MM paper were previously soaked in transfer buffer. The blotting apparatus was set at a constant current of 30 mA overnight or 400 mA for 1 hour.

Following transfer, the membrane was incubated in blocking buffer (5% (w/v) skimmed milk powder, 0.1% (v/v) Tween 20 in PBS) for 1 hour. The membrane was rinsed twice with blocking buffer for 10 min.

The antibody to be used (previously diluted in blocking buffer depending on its efficiency) was incubated with the membrane for 1 hour at room temperature, with agitation, followed by three 10 minute washes of the membrane using blocking buffer. In order to detect bound primary antibodies, the membrane was incubated with the appropriate secondary HRP-conjugated antibody (in blocking buffer) for 45 minutes. Then 2 brief washes were carried out using wash buffer (0.1% [v/v] Tween 20 in PBS). The membrane was incubated for 15 minutes in wash buffer, and subsequently washed twice for 10 min with the same buffer.

The bound HRP-conjugated secondary antibodies were visualized by enhanced chemiluminescence (ELC plus Western Blotting Detection System, Amersham Biosciences, Buckinghamshire, UK).

2.3 Bacterial protein expression and purification

Protein purification is essential for the characterisation of the function, structure and interactions of the protein of interest. Affinity purification tags can be fused to any recombinant protein allowing fast and easy purification. These tags are biochemical indicators that are normally small proteins, polypeptides or enzymes added to the N- or C- terminus of the protein of interest. They can have a role as an aid to solubilization, stabilization and expression of the protein to which they are attached.

2.3.1 Glutathione S-transferase (GST) gene fusion system

The GST Gene Fusion System (Amersham Biosciences, Buckinghamshire, UK) was used to express, purify and detect fusion protein produced in *E.coli* BL21 (DE3) pLys cells (Promega Ltd., Southampton, UK). GST fusion proteins were purified from bacterial lysates by affinity chromatography using immobilized glutathione. GST fusion proteins were captured by the affinity medium and the fusion proteins were eluted under mild, non-denaturing conditions using reduced glutathione

after several washes to remove contaminating bacterial proteins. These GST-tagged plasmids also include a site-specific protease domain, such as thrombin, which is situated upstream of the multiple cloning site. Thus, separation of GST from the protein of interest can be achieved.

2.3.1.1 GST-tagged protein expression

For protein expression, transformation of the GST-tagged plasmids of interest was performed, using *E.coli* BL21(DE3) pLys cells (Promega Ltd., Southampton, UK), which are cells that allow high-efficiency protein expression of any gene that is under the control of a T7 promoter and has a ribosome binding site.

The vector of interest was transformed in BL21 (DE3) pLysS *E.coli* competent cells (Promega Ltd., Southampton, UK), and, the next morning, a single colony was inoculated into 2ml LB medium containing Ampicillin (100 µg/ml) and grown overnight at 37ºC with moderate shaking. This starter culture was diluted 1:10 and grown at 37ºC with moderate shaking to an A_{600} of 0.6. Protein expression was induced by adding IPTG to a final concentration of 0.5mM and incubating the culture for 4-6 hours more. The culture was harvested by centrifugation at 4,000 rpm for 20 min at 4ºC. The pellet was re-suspended in Lysis Buffer (PBS containing 0.1% [v/v] NP40, 1mM DTT and protease inhibitor cocktail tablet [Complete Mini EDTA-free, Roche Diagnostics Ltd., Burgess Hill, Uk]), followed by sonication. The culture was then harvested by centrifugation at 18,000rpm for 30 min at 4ºC and the supernatant transferred to a fresh tube. 20µl of 50 % slurry of Protein G Sepharose 4 Fast Flow (GE Healthcare UK Ltd, Buckinghamshire, England) were added and the tube was left rotating for 30 min at 4ºC. The beads/fusion protein complex was washed 5 times with PBS containing 0.1% [v/v] NP40. The complex was recovered between washes at 12000 rpm for 1 min. The pellet was re-suspended in 2x SDS-PAGE sample loading buffer and the solution was dissociated at 100ºC for 2 minutes and analysed on SDS-PAGE gel (10 % polyacrylamide, described in section 2.2.1.)

2.3.2 Polyhistidine gene fusion system (His-tag)

A polyhistidine-tag is an amino acid motif in proteins that consists of at least six histidine (*His*) residues, often at the N- or C-terminus of the protein. This purification method relies on the affinity of histidine residues for immobilized metal such as nickel, which allows specific purification of the proteins. This mechanism is

only dependent on the primary structure of proteins, thus it is very efficient for purifying recombinant proteins in denaturing conditions.

2.3.2.1 His-tagged protein expression

2A protein from TMEV virus was inserted into Pehistev vector (a kind gift from Dr. Huanting Liu, figure 38) and transformed in BL21 (DE3)pLysS E.coli competent cells (Promega Ltd., Southampton, UK). A single colony of *E.coli* transformed with this plasmid was inoculated into 2ml LB medium containing Kanamycin (30µg/ml) and grown overnight at 37ºC with moderate shaking. This starter culture was diluted 1:10 and grown at 37ºC with moderate shaking to an A_{600} of 0.6. Protein expression was induced by adding IPTG to a final concentration of 0.5mM. Incubation continued for 4-6 hours, and the culture was harvested by centrifugation at 5000 rpm for 10 min. Bacterial cells were lysed by re-suspending the pellet in 0.5 ml of sample buffer (PBS containing 0.3M NaCl, 10mM imidazol, protease inhibitor cocktail tablet [Complete Mini EDTA-free, Roche Diagnostics Ltd., Burgess Hill, UK]), followed by sonication. The lysate was centrifuged for 10 min at 13000 rmp at 4 ºC, and the supernatant transferred to a fresh tube with a 50% slurry of Nickel beads inside. The tube was left rotating for an hour at 4ºC and then, washed 3 times with 0.5ml of washing buffer (PBS containing 0.3 NaCl, 30mM imidazole and 1mM PMSF). The beads/fusion protein complex was recovered between washes by centrifugation at 13000 rpm for 1 min. After washing, 50µl of elution buffer (PBS containing 0.3M NaCl, 500nM imidazole, 1mM PMSF) was added to the complex and incubated for 2 min. Re-suspension of the pellet in 2x SDS-PAGE sample loading buffer was performed after centrifugation of the cells for 1 min at 13000 rpm. The sample was dissociated at 100ºC for 2 min and analysed by SDS-PAGE gel (10% polyacrylamide, described in section 2.2.1).

2.3.4 Purification of the His-tagged protein using Immobilised metal affinity columns

The bacterial pellet (obtained as explained above) was re-suspended in lysis buffer (20mM Tris-HCl pH 8.0, 0.5M NaCl, 0.1% Triton-X-100, 1mM $MgCl_2$, 1 Complete EDTA-free Protease Inhibitor tablet) and sonicated (3 x 1 minute) on ice. The cell lysate was centrifuged at 20,000 rpm for 30 minutes at 4°C and the supernatant filtered (Acrodisc® 0.45mm syringe filter), and diluted 2 fold with buffer A (20mM Tris-HCl pH 8.0, 0.5M NaCl, 30mM Imidazole, 30mM NaH_2PO_4). The clear lysate was loaded onto an equilibrated 5 ml HisTrap Chelating Nickel column

(GE Healthcare UK Ltd, Buckinghamshire, England) and the His-tagged protein was eluted off the column over a 115 ml 30-500mM Imidazole linear gradient. The fractions were analysed by SDS-PAGE (see 2.2.1)

2.3.5 Purification of the GST-tagged protein using Glutathione agarose columns

2.3.5.1 Glutathione-Agarose Column

Affinity chromatography using Glutathione Agarose allows rapid, mild, non-denaturing and highly selective purification of glutathione binding enzymes such as glutathione-S-transferase.

Purification of the protein of interest was achieved by using a 5 ml column filled with Glutathione Agarose (Sigma-Aldrich Company LDT. Dorset, UK). The entire process was performed at 4˚C to avoid protein degradation. The column was regenerated by loading 5 resin volumes of Cleansing Buffer (0.1M borate buffer, pH 8.5 containing 0.5M NaCl), followed by washes with 5 column volumes of deionized water. The column was washed with 5 column volumes of Cleansing buffer 2 (0.1M acetate buffer, pH 4.5, containing 0.5M NaCl) and washed again with deionized water. After the regeneration process, the column was equilibrated with Equilibration buffer (PBS [Phosphate buffered saline: 10mM phosphate buffer pH7.4, 150mM NaCl]).

The supernatant (with the soluble protein of interest, prepared as explained previously) was loaded onto the column under gravity flow. The flowthrough was kept for subsequent SDS-analysis, to check the efficiency of binding).

Once all the supernatant was loaded, the resin was washed with PBS-T (PBS containing 1% [v/v] Triton X-100) until the OD_{280} was 0. Then the GST-tagged protein was eluted from the resin with Elution Buffer (5mM to 10mM reduced glutathione in 50mM Tris-HCl, pH 9.5).

The fractions containing GST-tagged protein were identified by SDS-PAGE analysis and Mass Spectrometry.

2.3.5.2 Pre-packed Glutathione-Agarose Column

Purification of the protein of interest using a 1ml GSTrap™ FF Column (GE Healthcare UK Ltd, Buckinghamshire, England). These columns are simple to use and bind approx. 10 mg of recombinant GST/ml medium.

The pre-packed column was equilibrated with 5-10 column volumes of Equilibration Buffer (PBS with 0.5M NaCl). The supernatant containing the protein

of interest was loaded onto the column and left re-circulating overnight. The column was then washed with Equilibration Buffer until OD_{280}= 0. Elution buffer (5mM to 10mM reduced glutathione in 50mM Tris-HCl, pH 9.5) was added at this stage and 3ml fractions collected. The fractions were analysed by SDS-PAGE.

2.3.6 Ultrafiltration

Millipore's Amicon® Ultra-15 centrifugal filter devices (Millipore UK Ltd, Watford, England) contain a Millipore Ultracell® regenerated cellulose low binding membrane, which provides high concentration factors and excellent sample recoveries from dilute protein fractions. The sample was loaded onto the filter device and this was spun in a swinging bucket at 4000 x g 15-45 minutes at 4ºC.

2.3.7 Thrombin protease digestion

PGEX vectors, used to purify the protein of interest *via* Glutathione-Agarose column, also have a Thrombin protease recognition site for cleaving the desired protein from the fusion product.

100µg of GST-fusion protein was digested with 1 unit of enzyme in 1x PBS [Phosphate buffered saline: 10mM phosphate buffer pH7.4, 150mM NaCl]) and incubated at 4˚C overnight.

2.3.8 Silver staining

Silver staining is a sensitive method for detecting proteins in SDS-PAGE gels. The gel was incubated in fresh Fixer (10ml glycerol, 100ml methanol, 20ml acetic acid, 70ml dH_2O) for 20 minutes, then, washed twice with 400ml dH_2O.The Staining solution (Silver complex (5 ml) [0.1g $AgNO_3$ and 0.1g NH_4NO_3 in a final volume of 5 ml dH_2O], Moderator (5ml) [0.5g tungstosilicic acid monohydrate in 5ml of dH_2O] and Developer (5ml) [0.43ml in 5ml dH_2O] in a final volume of 30ml dH_2O) was prepared freshly and mixed with 50ml Accelerator (6.75g $Na_2CO_3 \cdot 10\ H_2O$ in 50ml dH_2O). This mixture was immediately added to the gel and incubated in a clean dish. Development was stopped (5% acetic acid) when the bands reached the desired intensity in relation to background.

2.4 Cell culture

2.4.1 Cell lines

2.4.1.1 Mammalian cell lines

The mammalian cell lines were kindly provided by Dr. Rick Randall (University of St. Andrews).

- HeLa: cell line derived from cervical cancer cells
- BHK: baby hamster kidney cells.
- 293T: human embryo kidney cells (HEK).

2.4.2 Splitting cells

The old medium was poured off and cells washed with PBS to remove the remaining residual medium. Trypsin/EDTA (Becton Dickinson, Plymouth, UK), which is a buffered salt solution containing 0.5% (w/v) trypsin and 0.2% (w/v) EDTA, was added to dissociate adherent cells. The flask was incubated at 37ºC for 5 minutes and then, 10ml of Dulbecco's modified Eagle's medium (DMEM; Sigma-Aldrich company LDT. Dorset, UK) supplemented with 10% [v/v] fetal calf serum (FCS; Invitrogen Ltd, Paisley, UK) was added to inactivate the trypsin. The cells were subsequently dispersed by agitation, transferred to a 15ml tube and centrifuged at 750 x g for 3 min. The supernatant was removed and the cell pellet re-suspended in 10 ml of DMEM supplemented with 10% FCS. 2 ml of cell solution was added.

Cells were maintained in 75cm^2 flasks at 37ºC in a 5% CO_2 humidified incubator.

2.4.3 Mammalian protein expression

Transfection is a method that facilitates the introduction of negatively charged molecules (phosphate backbone of DNA) into cells with negatively charged membranes. It is a non-viral method used to introduce nucleic acids into eukaryotic cells using different physical, lipid or chemical methods. Transfection is a very useful technique for studying the function of a certain gene in the context of a cell.

2.4.3.1 Transfection reagents

2.4.3.1.1 Chemical reagents

A cationic polymer associates with the negatively charged nucleic acid leading to the formation of a positively charged complex. This DNA/polymer complex will associate with the negatively charged cell membrane, which uptakes it by endocytosis.

2.4.3.1.1.1 GeneJuice transfection reagent (Novagen, Nottingham, UK)

GeneJuice is a proprietary formulation optimized for maximal transfection efficiency and minimal cytotoxicity. It consists of a non-toxic cellular protein and a small amount of a novel polyamine. Furthermore, this chemical reagent is compatible with both serum-containing and serum-free media.

On the day prior to transfection, cells were seeded to achieve a cell density of 50-80% needed at the time of transfection. A GeneJuice®:DNA ratio of 3:1 was used (although different ratios can be used for optimization (see manufacturer´s protocol guidelines)).

GeneJuice transfection reagent was diluted with serum-free medium (Opti-MEM® I Reduced Serum Medium (1X), liquid - with GlutaMAX™ I; Invitrogen Ltd, Paisley, UK). The reagent was added drop-wise to the medium and mixed thoroughly by vortexing. The GeneJuice/serum-free medium mixture was incubated at RT for 5 minutes. DNA was added to the GeneJuice/serum-free medium mixture and mixed gently by pipetting. The mixture was incubated at RT for 5-15 min and then added in a drop-wise manner to the cells in complete growth Medium. The dish was gently rocked to ensure even distribution. The cells were incubated for 24-48 hours (5% CO_2).

2.4.3.1.2 Cationic lipids

Liposome-mediated transfection is considered to be relatively more efficient in gene transfer than chemical reagents, thus, it is recommendable for certain resistant cell types. The DNA is delivered into cells by synthetic cationic lipids. The cationic head group of the lipid compound associates with the negatively charged nucleic acids, leading to the formation of a positively charged liposome/nucleic acid complex.

The complex effectively associates with the negatively charged cell membrane inducing very high transfer efficiency *via* endocytosis or fusion with the plasma

membrane (Gao & Huang 1995). Following cellular internalization, the complex will be released from the endosomes by an unclear mechanism, and will later appear in the nucleus.

2.4.3.1.2.1 FuGENE 6 transfection reagent (Roche Diagnostics Ltd, Burgess Hill, UK)

FuGENE 6 is a proprietary blend of lipids that forms a complex with DNA, and transports it into animal cells. Transfection can be performed in serum-containing media. Cells were seeded one day prior to transfection, so they achieve an appropriate density of 50-80% confluency during overnight incubation. Initially a FuGENE 6 : DNA ratio of 3:1 was used. Different ratios can be used for optimization (see manufacturer´s protocol).

First FuGENE 6 reagent was diluted with serum-free medium (Opti-MEM® I Reduced Serum Medium (1X), liquid - with GlutaMAX™ I, Invitrogen Ltd, Paisley, UK). Serum-free medium was added first, followed by the addition of FuGENE 6 directly into the medium without allowing contact with the walls of the plastic tubes. The solution was mixed, flicking the tube gently, and the mixture was subsequently incubated for 5 min at RT. Addition of DNA to the mixture was performed followed by incubation of the transfection reagent/DNA complex for 15-45 min at RT. The complex was then added to the cells in a drop-wise manner. The wells or flasks were swirled to ensure distribution over the entire plate surface, and finally returned to the incubator 24-48 hours to allow gene-expression.

2.4.4 Analyses of protein expressed in mammalian cells

2.4.4.1 Lysis of cells and Western Blotting

Cells were harvested 48 hours after transfection. The medium was removed and the cells washed with PBS 3 times for 3 minutes each. The PBS was removed and the cells re-suspended in 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, 100mM Tris, pH 6.8). The re-suspension was transferred to a 1.5 ml tube and sonicated for 5 seconds. Samples were boiled at 95ºC for 2 minutes, and the supernatant analyzed on SDS-PAGE (see 2.2.1).

2.4.4.2 Fixing cells

Cells grown in DMEM medium supplemented with 10% FCS on coverslips (10mm in diameter, General Scientific Co. Ltd., Redhill, UK) were washed with PBS, and subsequently fixed with 4% paraformaldehyde (PFA) for 30 min. The fixed cells were washed twice with PBS before mounted using Mowiol mounting solution (2.4g Mowiol 4-88 [Calbiochem, San Diego, USA], 6g glycerol, 6ml H_2O, 0.2M Tris-HCl pH 8.5, 12 ml 1.4-diazabicyclo[2.2.2]octane [DABCO; Sigma-Aldrich company LDT. Dorset, UK) supplemented with diamino phenylindole (DAPI, 0.5µg/ml; Sigma-Aldrich company Ltd. Dorset, UK) for nuclear staining.

2.4.4.3 Immunofluorescence

Cells, grown in DMEM medium supplemented with 10% FCS, were seeded onto glass coverslips in a 6-well plate and incubated overnight before being transfected (see 2.4.4). After 36-48 hours, cells were fixed with 400µl PFA for 30 min and washed with PBS before incubation with permeabilization buffer (500ml PBS, 10% [w/v] sucrose, 0.5% [v/v] NP40) for 20 min. The cells were blocked for 15 minutes with blocking buffer (permeabilization buffer containing 30% [v/v] goat serum; Sigma-Aldrich company Ltd. Dorset, UK). The proteins of interest were detected by incubating the cells with an appropriately diluted primary antibody (in blocking buffer 1/500) for at least 45 min. The cells were washed 3 times with wash buffer (0.1% [v/v] Tween 20 in PBS) for 5 min. The cells were then incubated with the fluorochrome-conjugated secondary Antibody (in blocking buffer 1/500) in order to detect bound primary antibody. Coverslips were kept in the dark to reduce bleaching of fluorescence probe and washed 3 times with wash buffer after 30 minutes of incubation. The coverslips were subsequently mounted in Mowiol mounting solution (2.4g Mowiol 4-88 [Calbiochem, San Diego, USA], 6g glycerol, 6ml H_2O, 0.2M Tris pH 8.5, 12ml 1.4-diazabicyclo [2.2.2]octane [DABCO; Sigma-Aldrich company Ltd. Dorset, UK) supplemented with diamino phenylindole (DAPI, 0.5 micrograms/ml; Sigma-Aldrich company Ltd. Dorset, UK).

2.4.4.4 Imaging

DeltaVision® microscope system (applied Precision, Marlborough, UK) consists of an inverted microscope (Olympus 1x70, Olympus, Tokyo, Japan) with a 1.40 NA 100x or 60 x oil immersion objectives and Photometric CH300 CCD camera. Images were processed with the softWoRx® Imaging software package (applied Precision, Marlborough, UK).

2.4.4.5 In vivo Immune precipitation (IP)

A radioactive mix containing: 1 ml DMEM without met/glu, 30 µCi of [^{35}S]-methionine (10µCi/µl) and 10 µl 100X glutamine; was added to 25 cm^2 flasks with transfected cells incubated for 48 hours at 37 ºC. The cells were incubated for 45-60 minutes at 37 ºC and, then, harvested and lysed in 500 µl chilled IP buffer (20mM Tris pH7.8, 650mM NaCl, 5mM EDTA, 0.5% [v/v] NP40, protease inhibitor cocktail tablet [Complete Mini EDTA-free, Roche Diagnostics Ltd., Burgess Hill, UK]). The cells were kept at -70 ºC for 10 minutes and transferred to 1.5 ml tubes. Centrifugation of cells was performed at 13,000 rpm for 30 minutes at 4 ºC, and the supernatant transferred to a new 1.5 ml tube. A specific antibody was added to the cleared cell lysate and the antibody/lysate mixture incubated at 4 ºC for 2 hours, rotating constantly. Protein A/G Sepharose beads (Amersham Biosciences, Buckinghamshire, UK) were washed three times with PBS to remove the ethanol in which they are supplied. The beads were recovered from washes by centrifugation at 12,000 x g for 10 seconds. A 50% slurry was prepared by adding an equal volume of PBS and stored at 4 ºC. 20 µl of this slurry was added to the antibody/lysate mixture and incubated at 4 ºC for 2 hours rotating. The beads were recovered by spinning at 12,000 x g for 10 seconds and re-suspended in 2x SDS-PAGE sample loading buffer (2%[w/v] SDS, 20% [v/v] glycerol, 2% [v/v] β-mecaptoethanol, 0.2% [w/v] bromophenol blue, 100mM Tris, pH 6.8). Samples were boiled at 95 ºC for 2 minutes and centrifugated at 12,000 x g for 10 seconds. The supernatant was analyzed by SDS-PAGE (see 2.2.1 and 2.2.5).

3. RESULTS

3.1 Development of a bacterial screen for 2A activity

This work is based on the hypothesis that 2A interacts with the ribosome exit tunnel while it is being translated, and is thus, able to stall the ribosome and inhibit peptide bond formation. There is consistent data to support this nascent peptide-ribosome interaction (reviewed by Tenson & Ehrenberg, 2002).

After several unsuccessful attempts, a reliable and efficient bacterial screen has been successfully developed. It consists of a reporter system that allows the detection of 2A activity directly in live cells without the need for Western blots. It is based on the construction of a destabilized LacZ reporter (Tobias *et al*., 1991a; Tobias *et al*., 1991b). This is achieved by the N-terminal fusion of the ubiquitin (Ub) protein linked to LacZ by an arginine (Arg; R) residue. Upon co-transformation with the ubiquitin proteinase (UBP), the Ub is proteolytically removed from the fusion and the Arg becomes the N-terminal residue of the LacZ. This destabilising residue directs LacZ to a degradation pathway. The introduction of the FMDV 2A between the destabilising Arg and LacZ allows the stabilisation of LacZ if 2A is active and removes the N-terminal Arg.

3.1.1 Construction of the destabilized reporter pHE12

The plasmid pUb23R (illustrated in figure 22), was digested with HindIII and ApaI, and the large DNA restriction fragment purified by agarose gel electrophoresis, and then, religated to create pHE9. A PCR fragment was amplified from pHE9 using the oligonucleotide primer forward HERf, which has an N-terminal BamHI restriction site and contains also the $2A_{FMDV}$ sequence and the reverse primer HERr (see table 3 in section 2.1.17). The PCR product was gel purified and, then, ligated into pcDNA 3.1/V5-His TOPO®. This vector was doubly restricted with BamHI and ClaI, gel purified and ligated into pHE9 vector similarly restricted to produce pHE12.

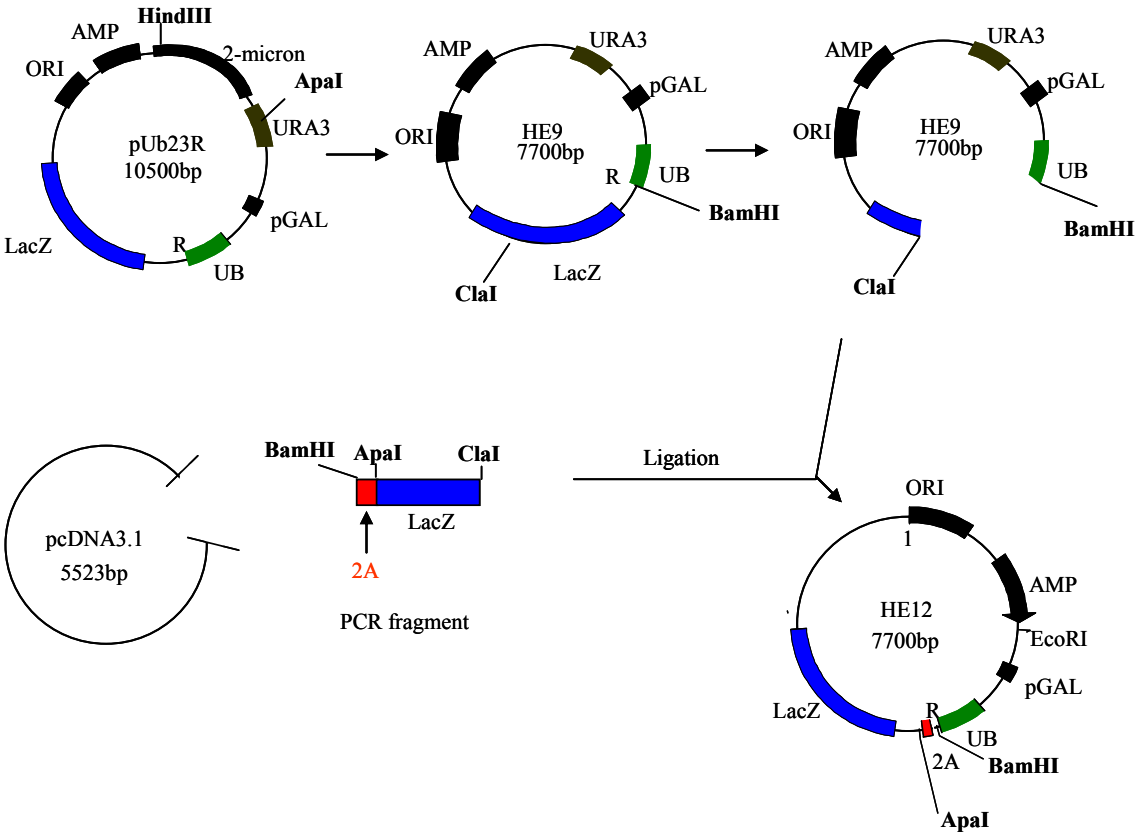


Figure 22: Overview of the cloning strategy utilized for the formation of pHE12. The pUb23R fragment was cut with HindIII and ApaI and religated in order to reduce its size. The PCR fragment was cloned into pcDNA 3.1/V5-His TOPO® and, then restricted with BamHI and ClaI, ligated into pHE9 vector similarly restricted, to produce pHE12.

3.1.2 pJT184

A plasmid essential for this novel bacterial screening system is pJT184 (figure 23), encoding the ubiquitin protease UBP1 (Tobias *et al.*, 1991b).

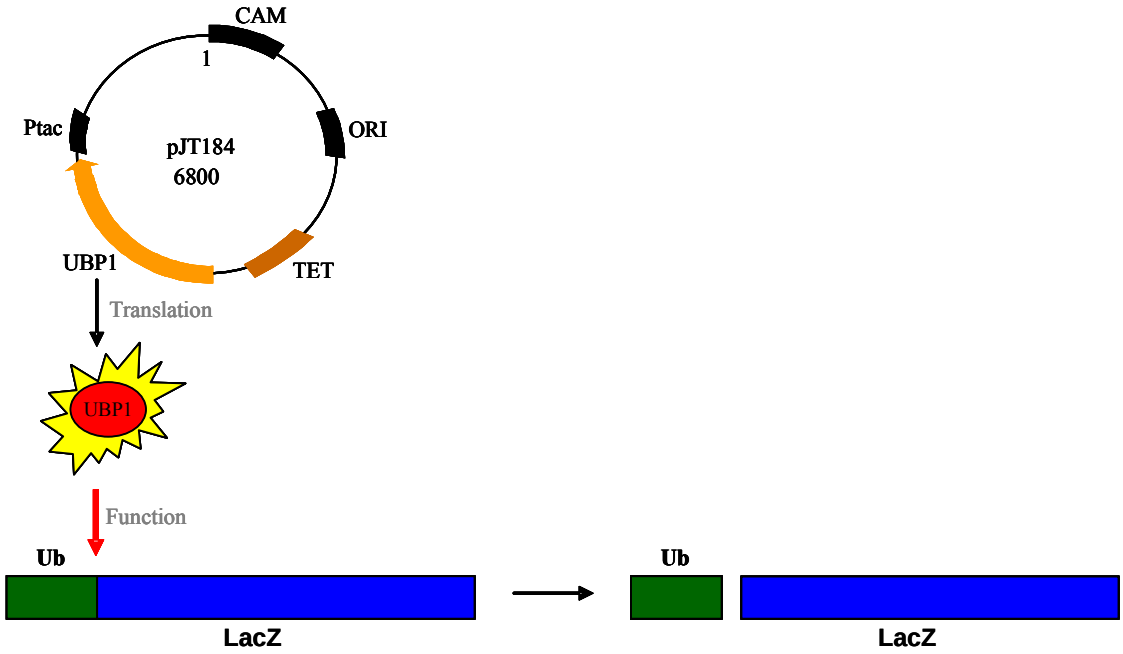


Figure 23: Illustration of pJT184. This plasmid contains the ubiquitin protease UBP1, which is able to cleave at the C-terminus of ubiquitin, releasing LacI/LacZ protein.

3.1.3 Testing the system

There are two versions of the plasmid pUb23, one with an arginine at the N-terminus of LacI/LacZ (pUb23R) and one with methione (pUb23M). *E.coli* was co-transformed with plasmids pUb23M and pJT184 or plasmids pUb23M and pJT184. Ampicillin indicator plates, which contain IPTG, and X-Gal were used to grow the colonies. In the first co-transformation, the colonies obtained were blue, whereas in the second one they were white (Figure 24a). This is because once the LacZ protein is cleaved away from the ubiquitin by UBP1, its N-terminus is exposed. When this N-terminus is methionine, LacZ is stable and reacts with X-gal to give blue coloured colonies. By contrast, an N-terminal arginine is very unstable, thus LacZ is degraded, giving white colonies (Tobias *et al.*, 1991a).

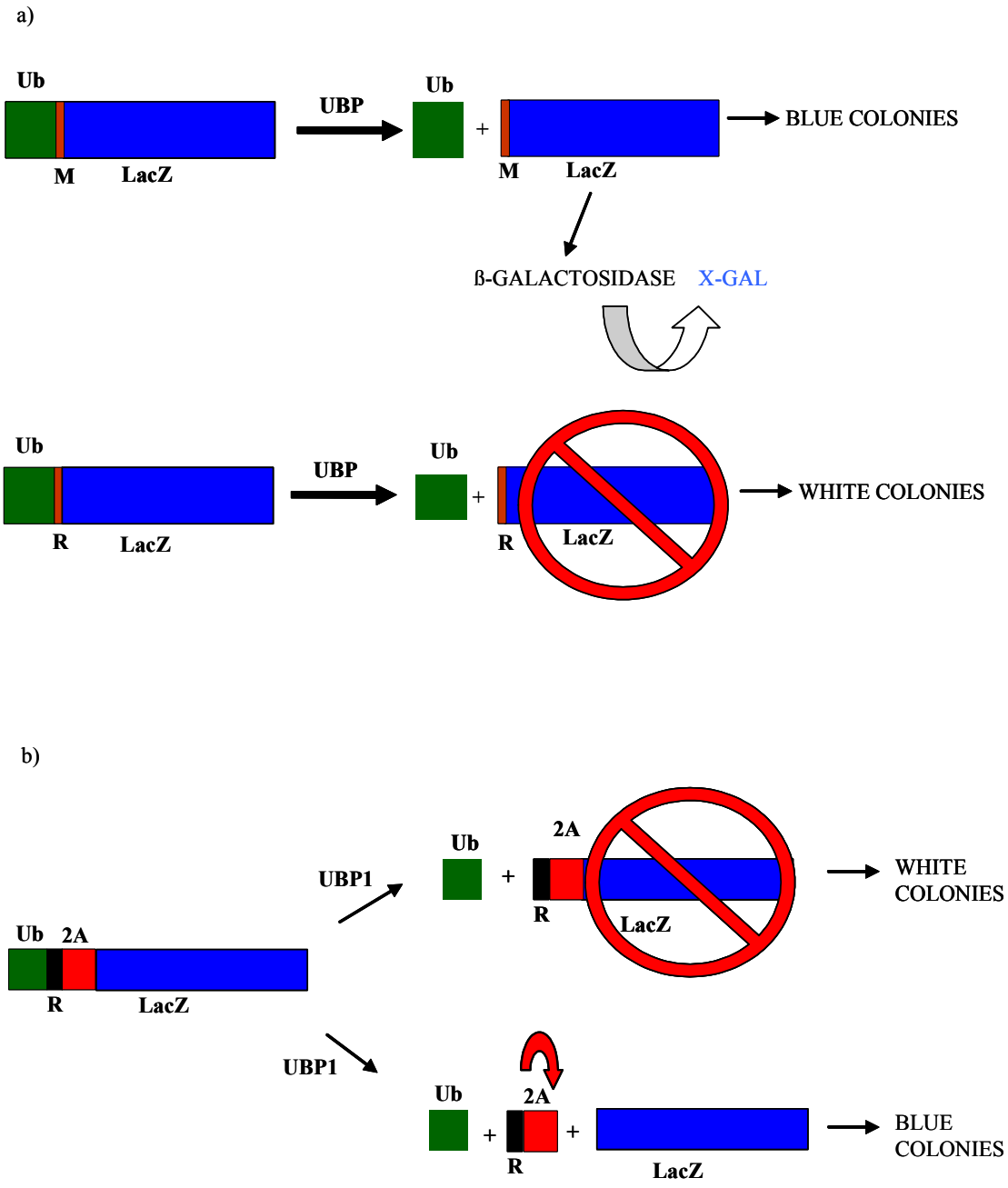


Figure 24: Scheme of the bacterial screen system. a) Two control plasmids: pUb23M co-transformed with UBP1 gives blue colonies, whereas pUb23R co-transformation with UBP1 gives white ones. b) Model of 2A screen system. Different 2As can be cloned into plasmid pHE12 and its activity tested through Blue/White screen.

In pHE12, the 2A sequence is situated in between the arginine and the N-terminus of LacZ. It is flanked by BamHI and ApaI, which are unique restriction sites in the plasmid. Thus, mutated 2As can be cloned there and their activity tested using our Blue/White screening system (Figure 24b).

To test the system MC1061 cells (genotype: hsdR2 hsdM+ hsdS+ araD139 Δ(ara-leu)7697 Δ(lac)X74 galE15 galK16 rpsL (StrR) mcrA mcrB1) were used for the co-transformations. Additionally aatΔ* cells, which derive from MC1061, were used as a control. These cells are deficient in the L/F transferase, which is a component of the N-end rule pathway necessary for the degradation of proteins carrying N-terminal Arg (R) or Lys (L) (Shrader *et al.,* 1993). Thus, pUb23R will be stable in these cells and will produce blue colonies (shown in figure 25).

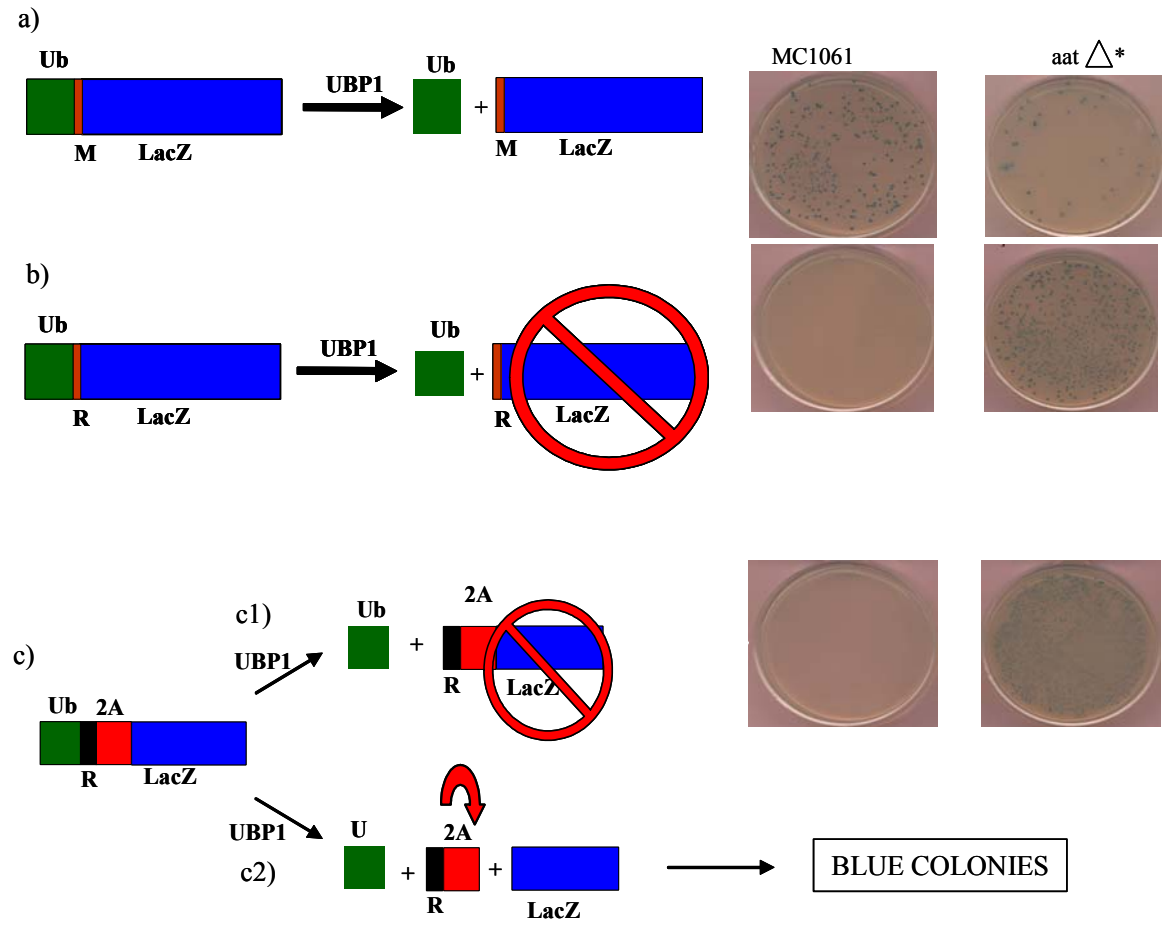


Figure 25. Testing the bacterial screen system a) pUb23M co-transformed with pJT184 produces blue colonies in MC1061 cells and the mutated strain aatΔ* b) pUb23R co-transformed with pJT184 (UBP1) produces white colonies in MC1061 and blue colonies in aatΔ* c1) pHE12 gives the same result as pUb23R because wild type FMDV-2A is not active in bacteria. c2) representation of a mutated 2A that 'cleaves' in bacteria.

Western blot analysis using anti-ß-galactosidase antibodies was performed to check the presence of the LacZ product in the cell. pUb23M (Ub-M-LacI/Z) and pHE12 (Ub-R-2A-LacI/Z) were transformed separately and also, co-transformed with pJT184 (UBP1). ß-galactosidase, whose size is 150 kDa, appeared in all the cases except when pHE12 was co-transformed with pJT184 (R-2A-LacI/Z), where LacZ is

degraded. Because 2A is not active in bacteria, the N-terminal arginine will not be cleaved away, and LacZ will therefore be degraded.

The results shown in figure 26 are consistent with prior observations seen in other types of artificial polyprotein systems, where wild type $2A_{FMDV}$ is not active in prokaryotic systems (Donnelly *et al.*, 1997)

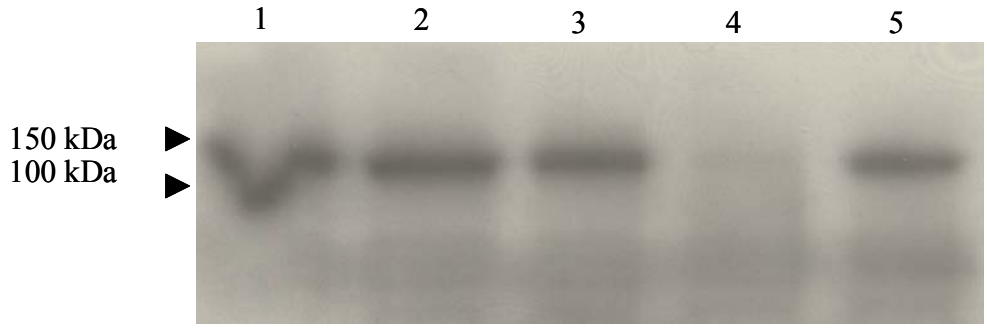


Figure 26: Western analysis of the destabilized reporter using anti-β-galactosidase. Lane 1: pUb23M (Ub-M-Lacz). Lane 2: Co-transformation of pUb23M with pJT184. Lane 3: HE12. Lane 4: Co-transformation of HE12 with pJT184 (MC1061 cells). Lane 5: Co-transformation of HE12 with pJT184 (aatΔ* cells), as positive control. pUb23M and HE12 were transformed in MC1061 cells and show a 150 kDA band, which is β-galactosidase. The co-transformation of HE12 with pJT184 in MC1061 cells does not show the LacZ band, because is degraded. In contrast, there is LacZ when using aatΔ* cells, as a positive control. The difference between Ub-M-LacZ and M-Lac is only 8.6 kDa, therefore is difficult to see this small difference in size on a protein gel (lanes 1 and 2).

To test that the transformations worked, DNA was purified from MC1061 cells transformed with pHE12 alone, pHE12 together with pJT184 and pJT184 by itself (figure 27).

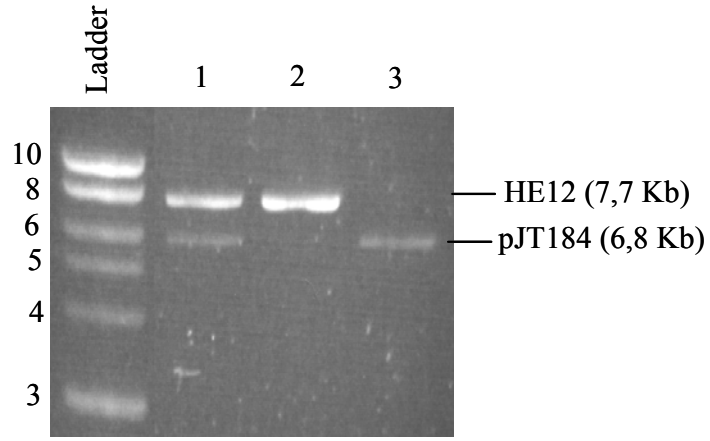


Figure 27: Agarose gel showing the two plasmids involved in the co-transformation. HE12 and pJT184 were linearized with the restriction enzyme ClaI. Lane 1: Linearized DNA obtained from a co-transformation with both plasmids was loaded in the gel. Both DNA bands are observed. Lane 2: HE12. Lane 3: pJT184. The positions of the size markers (Kb) are shown on the left of the figure.

pUb23M/R and pHE12 expression products were probed by Western blotting, using anti-ubiquitin antibodies (figure 28). Recombinant ubiquitin was used as a positive control (20ng/µl) and was the only recognized protein. The conjugated ubiquitin from pUb23M/R and pHE12 constructs was not recognized, possibly due to the Ub-epitope being masked at the C-terminal fusion.

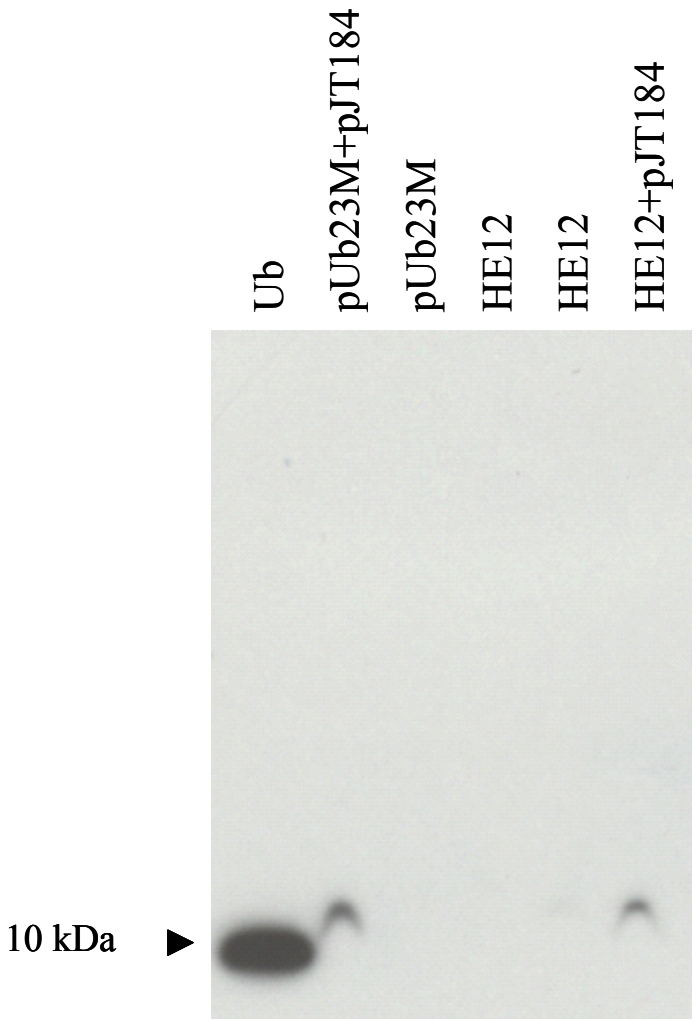


Figure 28. Western Blot using anti-ubiquitin antibody. Recombinant ubiquitin (20ng/µl), kindly provided by Dr. Xirodimas, was used as a positive control. Unfortunately the antibody only recognizes non-conjugated or non-fused ubiquitin.

3.1.4 Semi-random Mutagenesis

A double-stranded oligonucleotide adapter comprising the 2A sequence with 'degeneracy' built-in to the nucleotides highlighted in red (figure 29) was designed (Eurogentec Ltd, Hampshire, UK). The sequence has sticky ends with a 5' BamHI site and a 3' ApaI site, and the consensus motif is not mutated. The level of spiking is 0.7 probability of the wild-type nucleotide, 0.3 of incorporating a mutant base. That means 4 mutations in total on a 42 nucleotide sequence (figure 31).

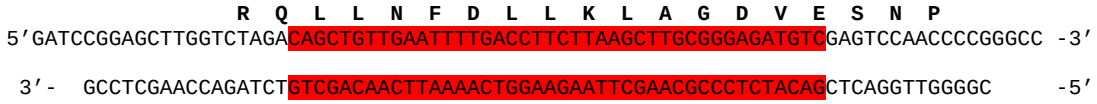


Figure 29. **Representation of the double-stranded oligonucleotide adapter comprised of the 2A sequence.** The 2A sequence has sticky ends with 5' BamHI site and a 3' ApaI site, and the consensus motif is not mutated. The mutated nucleotides are highlighted in red.

The binomial distribution (figure 30) was used to calculate the best level of spiking. This distribution is based on two possible outcomes (eg. Mutated or non-mutated amino acid in each position of the sequence) and the outcome of one trial does not affect the outcome on other trials.

$$P(X=k) = \binom{n}{k} p^k q^{n-k}$$

Figure 30. Binomial formula. X= k, indicates that the outcome X occurs k times; n is the number of repeated events; q is the probability of the second possible outcome and p the probability of success in one trial. (http://www.stat.yale. edu/ Courses/ 199798/ 101/binom.htm)

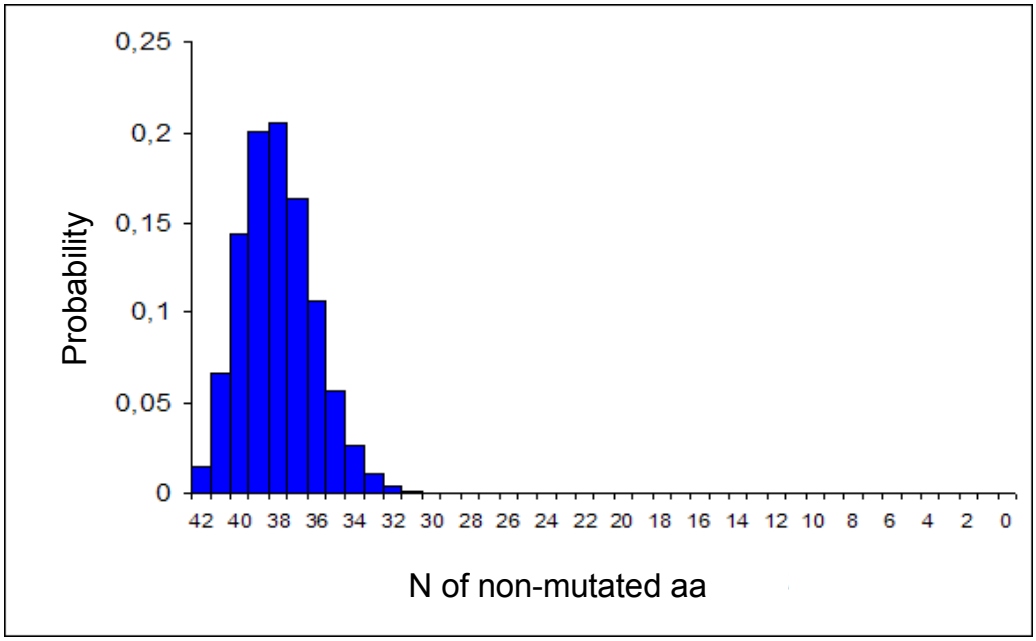


Figure 31: Graph representing the probability of introducing mutations depending on the length of the sequence. The optimum level of degeneracy was obtained by calculating the probability of incorporating a mutated nucleotide based on the length of the sequence. The calculations are based on binomial distribution, where there are two possible outcomes and the outcome of one trial does not affect the outcome on other trials.

The oligonucleotide adapters were annealed (as described in section 2.1.16) and then ligated into the pHE12 vector, restricted with BamHI and ApaI, and gel purified.

The mutated clones were transformed into pJM109 cells and grown on Amp-IPTG-X-Gal plates. Potentially active 2As would be selected from blue colonies and tested with anti-β-galactosidase and anti-ubiquitin.

The advantage of this system is that it would offer the possibility to check a very large number of mutated 2As through white/blue colour screening. Furthermore, it is a very efficient system because it avoids possible clones with frameshift mutations or premature stop codons.

3.1.5 Searching for active 2As in prokaryotic systems

Thirty different ligations of semi-random mutated 2As within the pHE12 plasmid were performed, and transformed on Amp-IPTG-XGAL plates. Unfortunately, no blue colonies were observed. A positive control was based on the substitution of the FMDV-2A with a ribosome binding site (RBS) to create a similar situation to that which happens in eukaryotic cells with the FMDV wild type 2A (figure 31).

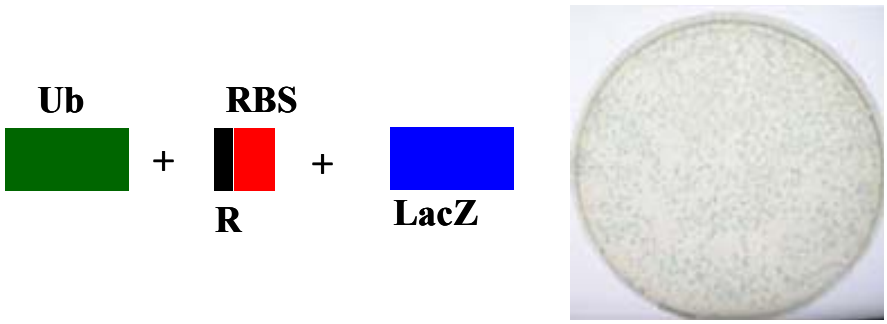


Figure 31. Positive control for the bacterial screen system based on the substitution of FMDV-2A for a ribosome binding site. This intergenic region mimics what an active 2A would do in eukaryotic organisms, thus, LacZ is released from the N-terminal arginine (R) and, therefore, blue colonies are obtained from the transformation of bacterial cells.

A number of plasmids prepared from white colonies were analyzed to assess the strategy of mutagenesis; check the ratio of mutation: wild-type sequences and, thus, the quality of the oligonucleotides (figure 32).

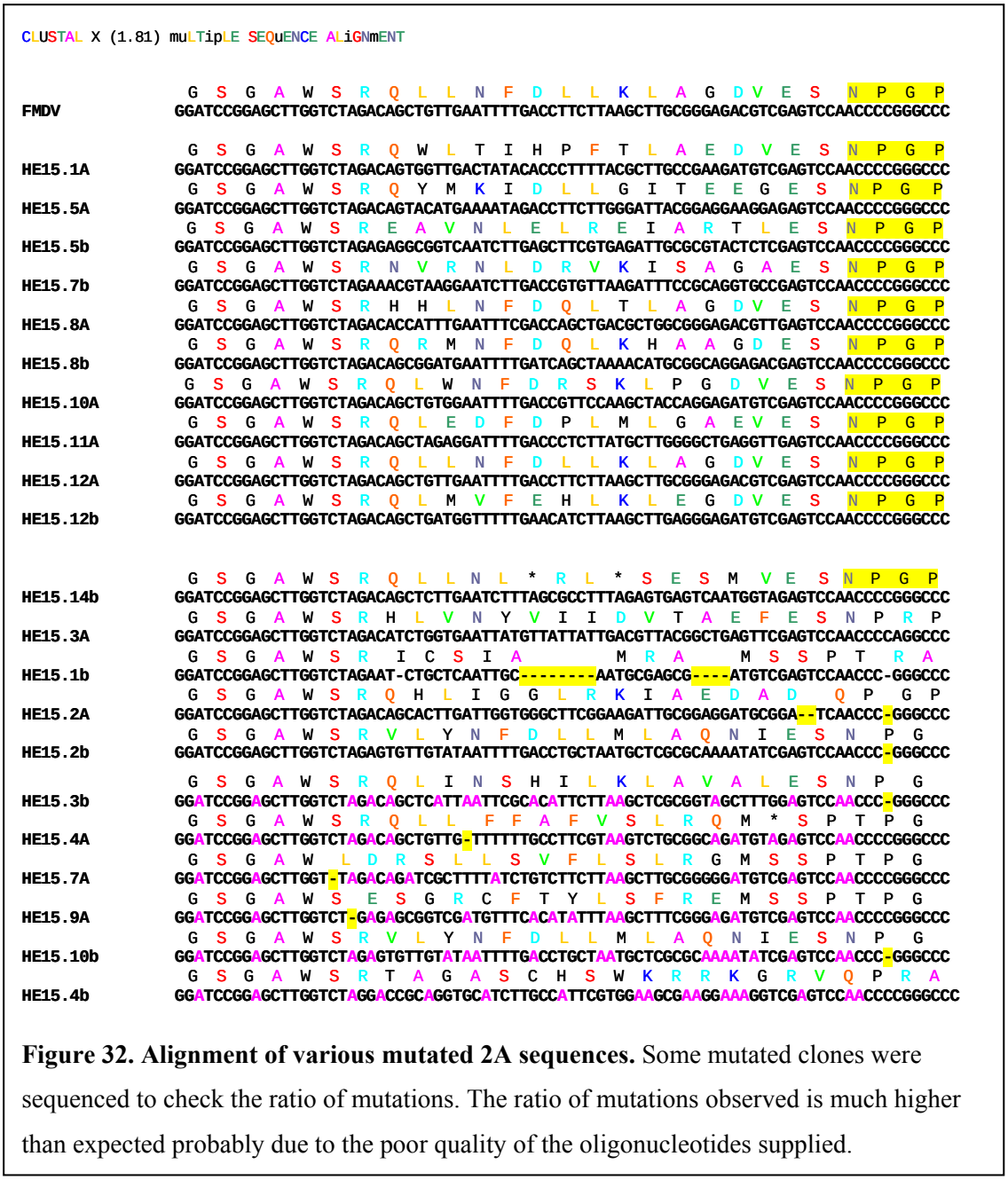


Figure 32. Alignment of various mutated 2A sequences. Some mutated clones were sequenced to check the ratio of mutations. The ratio of mutations observed is much higher than expected probably due to the poor quality of the oligonucleotides supplied.

Unfortunately, the quality of the oligonucleotides supplied was not as good as expected, the majority of mutated 2As showing between 7-10 mutations, much higher than the expected average of 4 mutations in a 42 nucleotide sequence. Furthermore, some versions with deletions and additions were also obtained showing the poor quality of these annealed nucleotides. It should be noted that many other oligonucleotides ordered (Eurogentec Ltd, Hampshire, UK) during this period were supplied with incorrect sequences resulting from faulty syntheses.

3.1.6 Improvement of the bacterial screen system

3.1.6.1 pHE29 vector

As shown previously, it is difficult to see a difference in size between Ub-M-LacI/Z and M-LacI/Z on an SDS-PAGE because ubiquitin is only 77 amino acids long, equivalent to a protein of 8.6 kDa, and ß-gal is 150 kDa. Therefore, using the α-peptide, which is only 10.4 kDa, instead of the entire Lac Z protein, the difference in size between the reporter protein (with and without Ub) is much better visualized on an SDS-PAGE using this system.

pGEM3Zf(-) vector (Promega Ltd., Southampton, UK) was used as a template to amplify the α-peptide fragment, which is a part of the LacZ gene, used for the white/blue screen. The oligonucleotides used were pGEMV5f as a forward primer, which has an N-terminal AatII, and the reverse primer pGEMV5r, which contains the V5 epitope sequence and a C-terminal BglII (see table 3 in section 2.1.17).

The PCR fragment was digested with BglII and AatII and ligated into pGEM3ZF(-), which was restricted with BamHI and AatII (figure 33). Thus BamHI site was deleted in this new pGEM3ZF(-) vector. A PCR was performed using the modified pGEM3ZF(-) vector using the forward primer ApepV5f, which has an N-terminal ApaI restriction site, and the primer reverse ApepV5r with a C-terminal NcoI restriction site (see table 3 in section 2.1.17).

The PCR fragment was restricted with ApaI and NcoI and ligated into the similarly restricted pHE12 vector to give HE29 (figure 33).

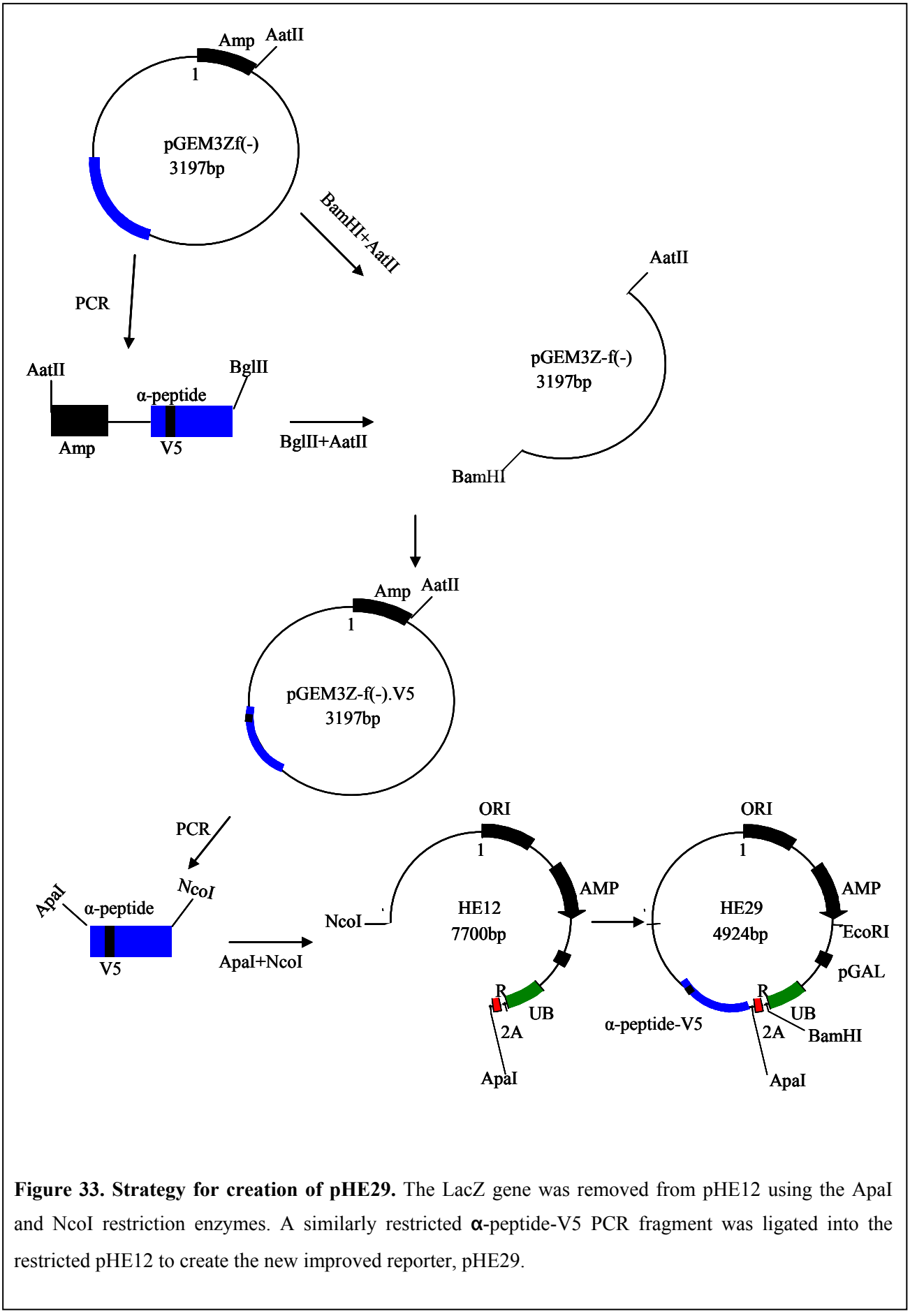


Figure 33. Strategy for creation of pHE29. The LacZ gene was removed from pHE12 using the ApaI and NcoI restriction enzymes. A similarly restricted α-peptide-V5 PCR fragment was ligated into the restricted pHE12 to create the new improved reporter, pHE29.

3.1.6.2 Western Blot anti-V5

First pGEM3Z-f(-).V5 was tested against anti-V5 in a Western Blot. A plasmid consisting of V5-CFP-2A-V5-PAC was used as a positive control. The size of α-peptide-V5 fusion protein is ~12 kDa (figure 34).

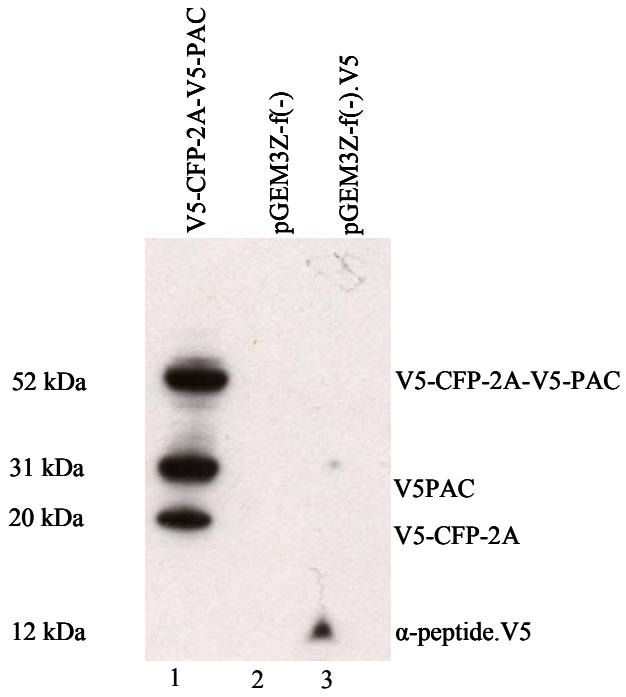


Figure 34. Western blot anti-V5. Lane 1: positive control containing 2 V5. Lane 2: Negative control, pGEM3Z-f(-). Lane 3: pGEMZ-f(-).V5. The size of the proteins recognized by V5 antibodies are shown on the left (kDa).

To test the "improved" system, transformations of pHE29 with and without pJT184 were performed in the same cell strains as before (see 3.1.3). Unfortunately, anti-V5 antibodies did not recognize for pHE29 (figure 35).

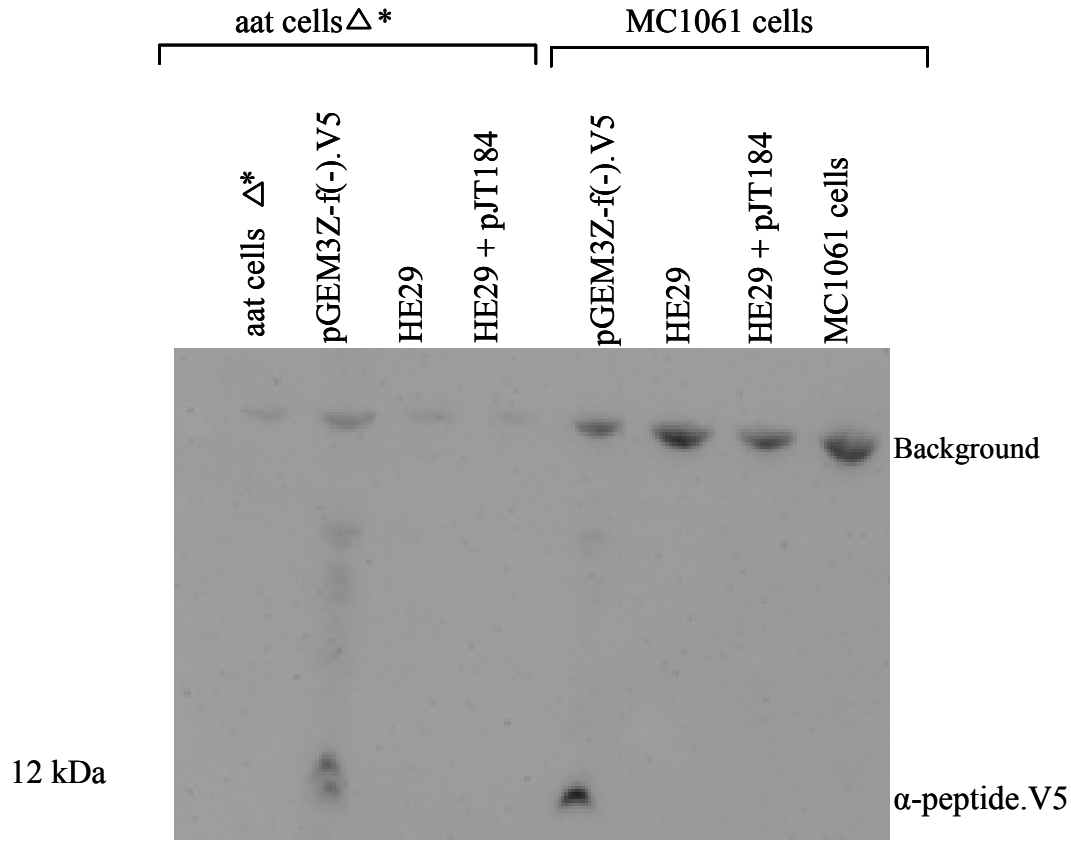


Figure 35. Prokaryotic expression. Artificial polyproteins were expressed in *E.coli* (MC1061 and its mutated version aatΔ*) and the cleavage activities assessed by probing a Western Blot of total cellular proteins with anti-V5 antibodies.

One must conclude, therefore, that the V5 epitope on the α-peptide encoded by the pHE29 is masked. Mutated 2A were cloned inside pHE29 and co-transformations with pJT184 were preformed. Only white colonies were observed. Due to time constrictions this work was not pursued further.

3.2 Purification of TMEV-2A

The aim of this part of the project was to purify $2A_{TMEV}$ protein in order to learn more about its structure and its behaviour within the cell. Furthermore, it was hoped to produce antibodies against $2A_{TMEV.}$ Due to time constraints and difficulties in the purification process (see below) insufficient material was obtained to proceed with 2A crystallization.

3.2.1 Optimization of bacterial expression systems

A construct containing GST-tagged (Glutathione Sepharose transferase) $2A_{TMEV}$ protein (pGEX-2T.$2A_{TMEV}$, kindly provided by Dr. Garry Luke) was transformed in BL21 (DE3) cells and the protein induced with IPTG (figure 36).

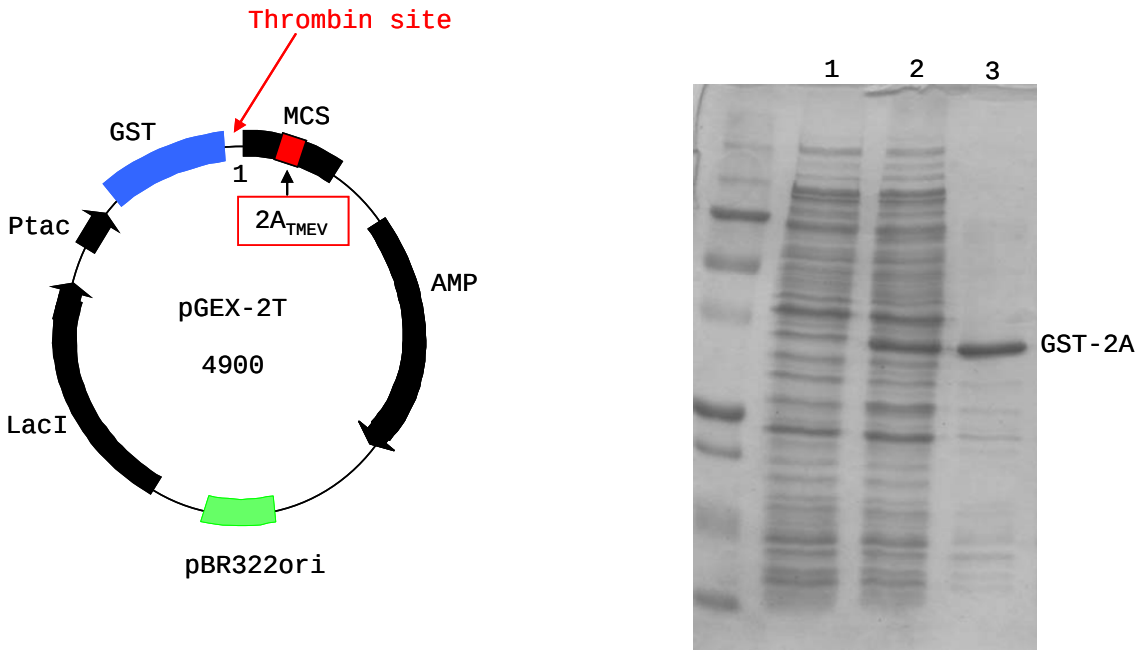


Figure 36. Schematic illustration of pGEX-2T on the left and SDS-PAGE showing the induction and purification of pGEX-2T.2A$_{TMEV}$ on the right. pGEX-2T multiple cloning site (MCS) is where 2A$_{TMEV}$ was cloned. As a result pGEX-2T.2A$_{TMEV}$ was obtained and, thereafter, transformed in BL21(DE3) *E.coli* cells. IPTG induction of the cells and purification of GST-2A protein were performed. Lane 1, uninduced; lane 2, induced; lane 3, GST-2A$_{TMEV}$ purified with Glutathione Sepharose Beads.

pGEX-2T.2A$_{TMEV}$ was transformed in BL21(DE3) *E.coli* cells and the colonies were grown overnight. The next morning 9 universals were set up with 10 ml of the overnight culture. The cultures were grown at different temperatures and the protein induced using various IPTG concentrations to test the optimum conditions for GST-tagged 2A$_{TMEV}$ protein induction (figure 37).

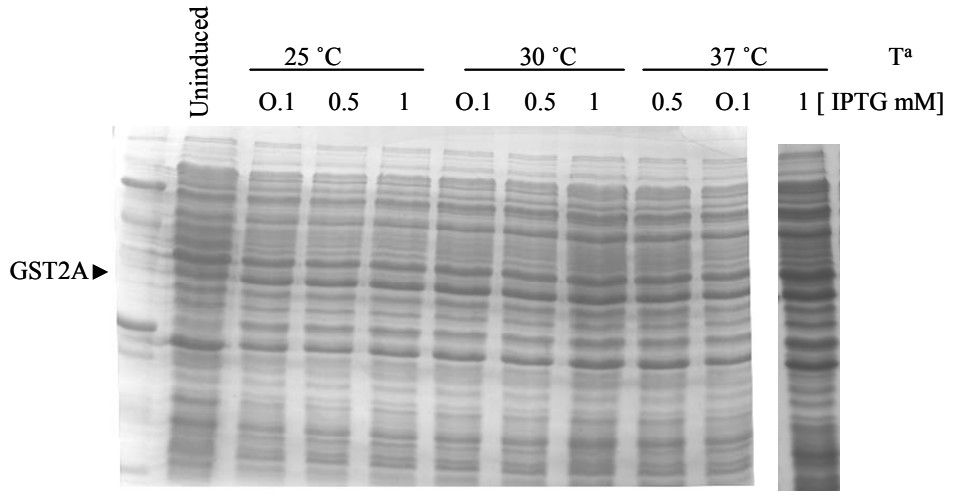


Figure 37. GST-2A induction with IPTG. Different temperatures of induction and different concentrations of IPTG were used during GST-2A protein induction to establish the optimum conditions. No differences in protein concentration were observed between different conditions.

In a parallel experiment, $2A_{TMEV}$ was cloned into a His_{6x}-tagged vector and the expression of the new plasmid compared to the GST-tagged construct. This was influenced by the fact that the Scottish Structural Proteomic Facility (SSPF) in the University of St. Andrews works only with His_{6x}-tagged constructs and, therefore, has all the high-throughput technologies for protein expression, purification, crystallization, structure determination specific for these constructs instead of the GST-tagged ones.

The $2A_{TMEV}$ sequence was amplified from pcDNA containing the sequence of TMEV (GDVII) (provided by Dr. Garry Luke) using the oligonucleotide forward primer HEPf, which has an N-terminal NcoI restriction site, and the reverse primer HEPr, which has a C-terminal BamHI restriction site (see table 3 in section 2.1.17). The PCR fragment was ligated into pcDNA3.1/ V5-His-TOPO® vector and, then, doubly restricted with NcoI and BamHI. The NcoI/BamHI product was gel purified and ligated into into pEHISTEV (a kind gift from Dr. Huanting, figure 38) similarly restricted.

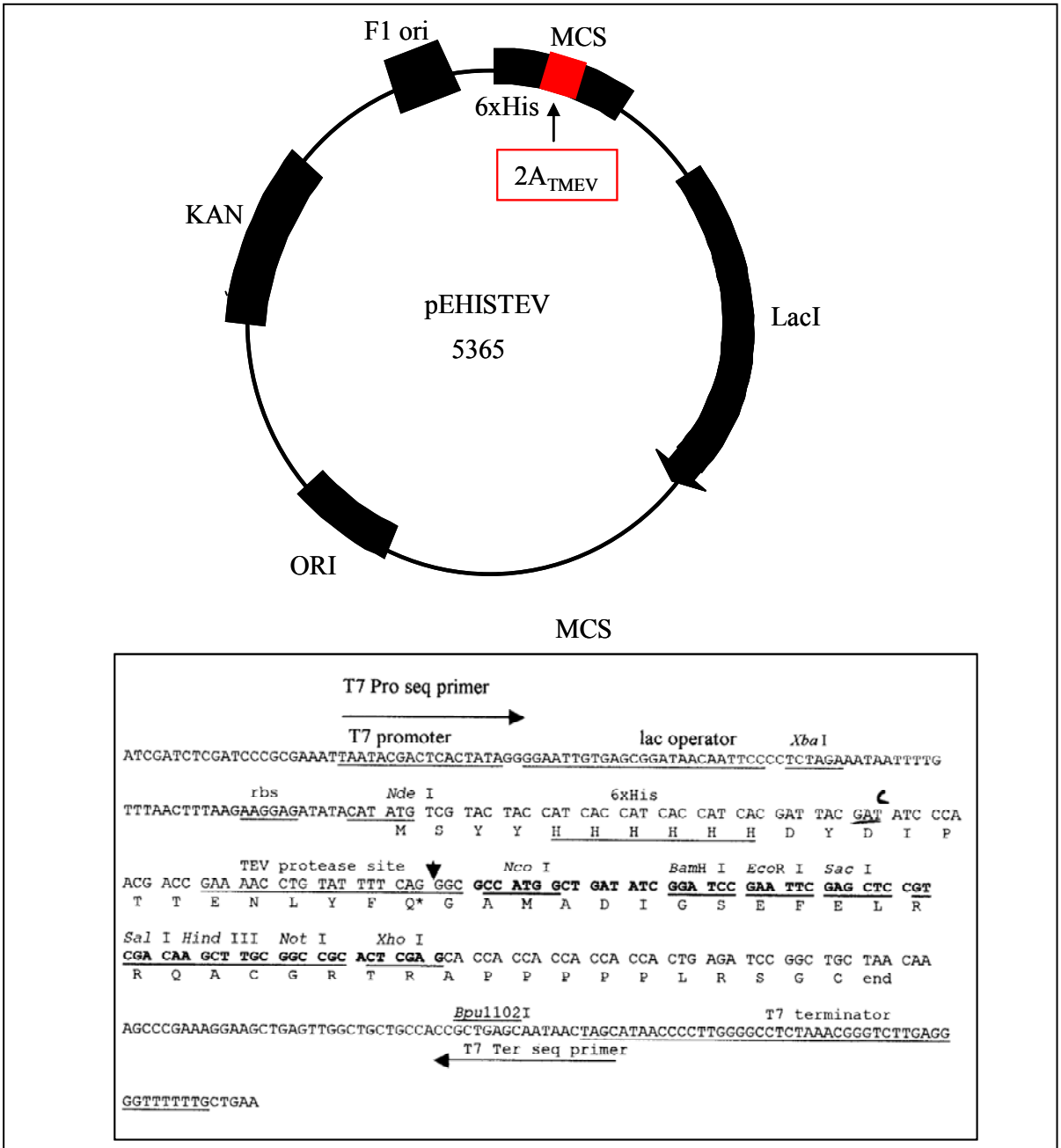


Figure 38. His-tagged prokaryotic expression vector. Top: Schematic pEHISTEV vector. Bottom: sequence detail of the multicloning site sequence (MCS).

Unfortunately, the resulting His_{6x}-tagged construct (pHE11) did not show any significant His_{6x}-2A protein expression (16 kDa) when compared to the GST-tagged construct (41 kDa) (figure 39). Therefore, it was discarded and pGEX-2T.2A$_{TMEV}$ was used for GST-2A protein induction and purification.

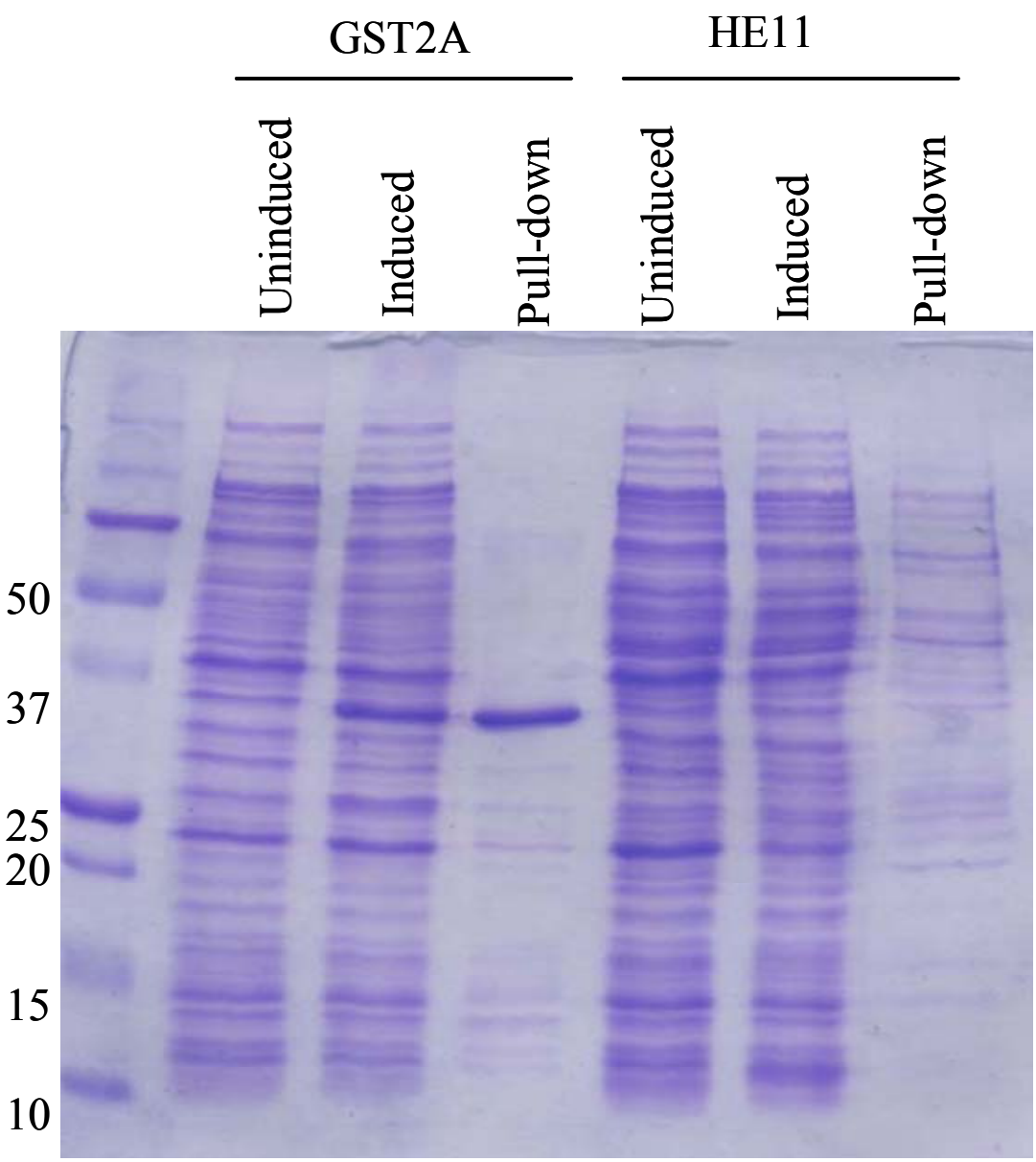


Figure 39. Comparison between the GST-tagged $2A_{TMEV}$ and the His-tagged $2A_{TMEV}$ IPTG induction and purification. Lanes 1-3: GST-tagged $2A_{TMEV}$. Lane 1: Uninduced. Lane 2: Induced. Lane 3: GST-$2A_{TMEV}$ pull- down with Glutathione Sepharose beads. Lanes 4-6: His-tagged $2A_{TMEV}$. Lane 4: Uninduced. Lane 5: Induced. Lane 6: GST-$2A_{TMEV}$ pull-down with Nickel beads.

3.2.2 Small scale induction, purification and thrombin digestion trials

The construct pGEX-2T.2A$_{TMEV}$ was transformed in BL21 (DE3) *E.coli* cells and, thereafter, a 50 ml culture protein induction with IPTG was performed. The cells were, then, lysed and the cleared lysate was loaded into a 5 ml Glutathione-Agarose column (Sigma-Aldrich Company Ltd. Dorset, UK). The column was washed with PBS-T (PBS containing 1% [v/v] Triton X-100) and the GST-tagged 2A$_{TMEV}$ protein eluted with Elution Buffer (5mM to 10mM reduced glutathione in 50mM Tris-HCl, pH 9.5). GST-tagged 2A$_{TMEV}$ protein fractions were collected (1.5 ml/fraction) and identified by SDS-PAGE analysis (figure 40).

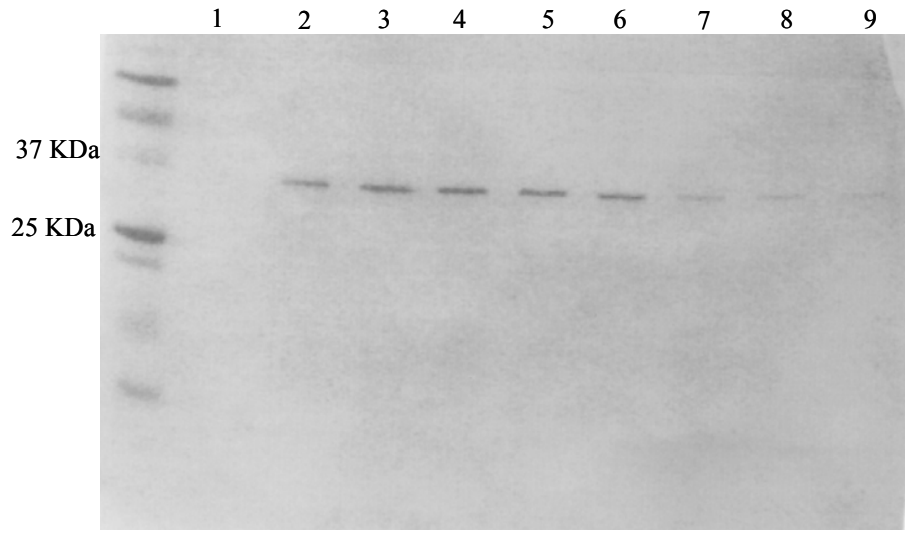


Figure 40. Small scale GST-tagged 2A$_{TMEV}$ protein purification. The lanes are labeled 1-9 corresponding to the different fractions collected.

All the protein containing fractions (1-9) were pooled and ultra-filtrated through a 20 ml Amicon Ultra centrifugal device (Millipore UK Ltd, Watford, England) to concentrate the sample (figure 41).

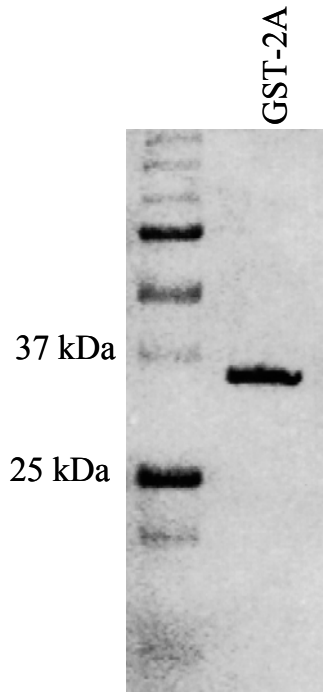


Figure 41. GST-tagged $2A_{TMEV}$ concentration. The 9 pooled GST-2A fractions were concentrated using a 20 ml Amicon Ultra centrifugal device.

The concentrated GST-tagged $2A_{TMEV}$ protein was digested with 1 unit of thrombin enzyme (Amersham Biosciences, Buckinghamshire, UK) to cleave the $2A_{TMEV}$ away from the GST. To optimize digestion conditions, three experiments were set up using either 5 µl, 10 µl or 20 µl GST-tagged $2A_{TMEV}$ (figure 42). After digestion samples were centrifuged at 15,000 rpm for 3 minutes. The supernatants were transferred to fresh tubes and, although a pellet was not visible, it was re-suspended in 5 µl of PBS to make sure there was not insoluble protein present. The fact that 2A was found in the precipitate fraction (figure 42, lane 3c) suggests the protein is insoluble however; it may be because of the fact that there is more concentration of purified protein at the bottom of the tube, rather than being the pellet.

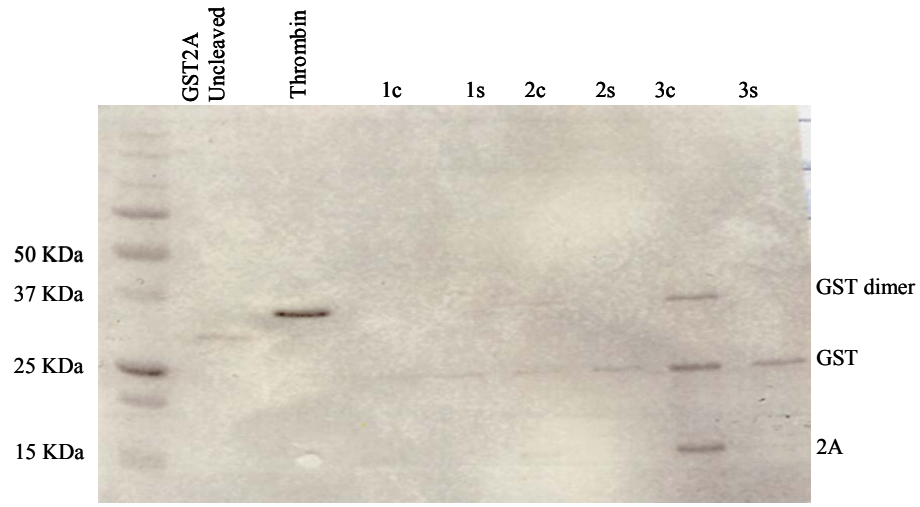


Figure 42. Small scale thrombin digestion trial. First lane: GST-2A uncleaved. Second lane: Thrombin (37 KDa). The following lanes represent GST-2A digested with thrombin using three different starting volumes of GST-2A protein: 1) 5 µl, 2) 10 µl and 3) 20 µl. s: supernatant, c: centrifuged. The digestion was performed at room temperature for 16 hours.

GST-2A purification was also performed using Glutathione Sepharose Beads (Amersham Biosciences, Buckinghamshire, UK) tested and compared to the glutathione agarose column.

Another 50 ml small scale protein induction was performed (as explained) and the protein purification preformed using Glutathione Sepharose Beads. Thrombin was added to the GST-2A/Beads mixture and incubated by rotating at 16ºC for 16 hrs. The mixture was spun down and the supernatant transferred to a fresh tube.

The purification was not as clean as the column procedure and most 2A protein was observed in the pellet with the Glutathione Sepharose Beads. No 2A band was observed in the supernatant. However, this was not due to the possibility of insolubility but due to a much diluted sample. Insolubility was discarded when Trichloroacetic acid (TCA) precipitation was performed in the supernatant samples and 2A protein was observed (figure 43).

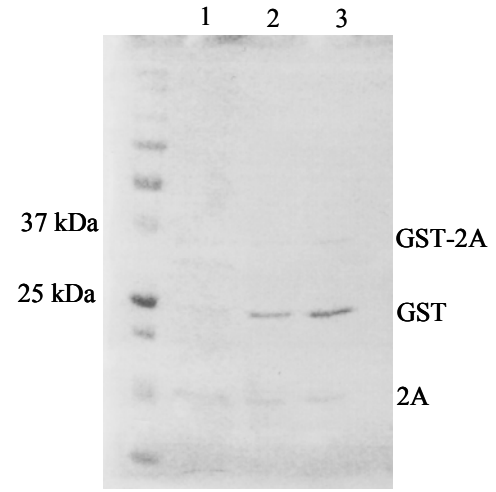


Figure 43. Thrombin digestion on-beads. GST-2A was digested on Beads, the supernatant recovered and concentrated with Trichloroacetic acid (TCA). Lane 1: Supernatant after thrombin digestion. 2A protein is present. Lane 2 and 3: Glutathione sepharose beads centrifuged from digested sample. Majority of GST is observed on the beads.

The purified samples from the 50 ml glutathione-agarose column (see above) were used to test the optimum conditions of thrombin digestion. Two different temperatures of incubation were used (4ºC and 16ºC) and samples were taken at 15, 30, 90, 120 and 240 minutes and overnight. As shown in figure 44 the optimum conditions were at 16 ºC overnight.

Purification using beads was less efficient than the column procedure. Although some 2A protein did precipitate with the glutathione sepharose beads, the bulk of 2A was recovered from the soluble fraction by TCA precipitation.

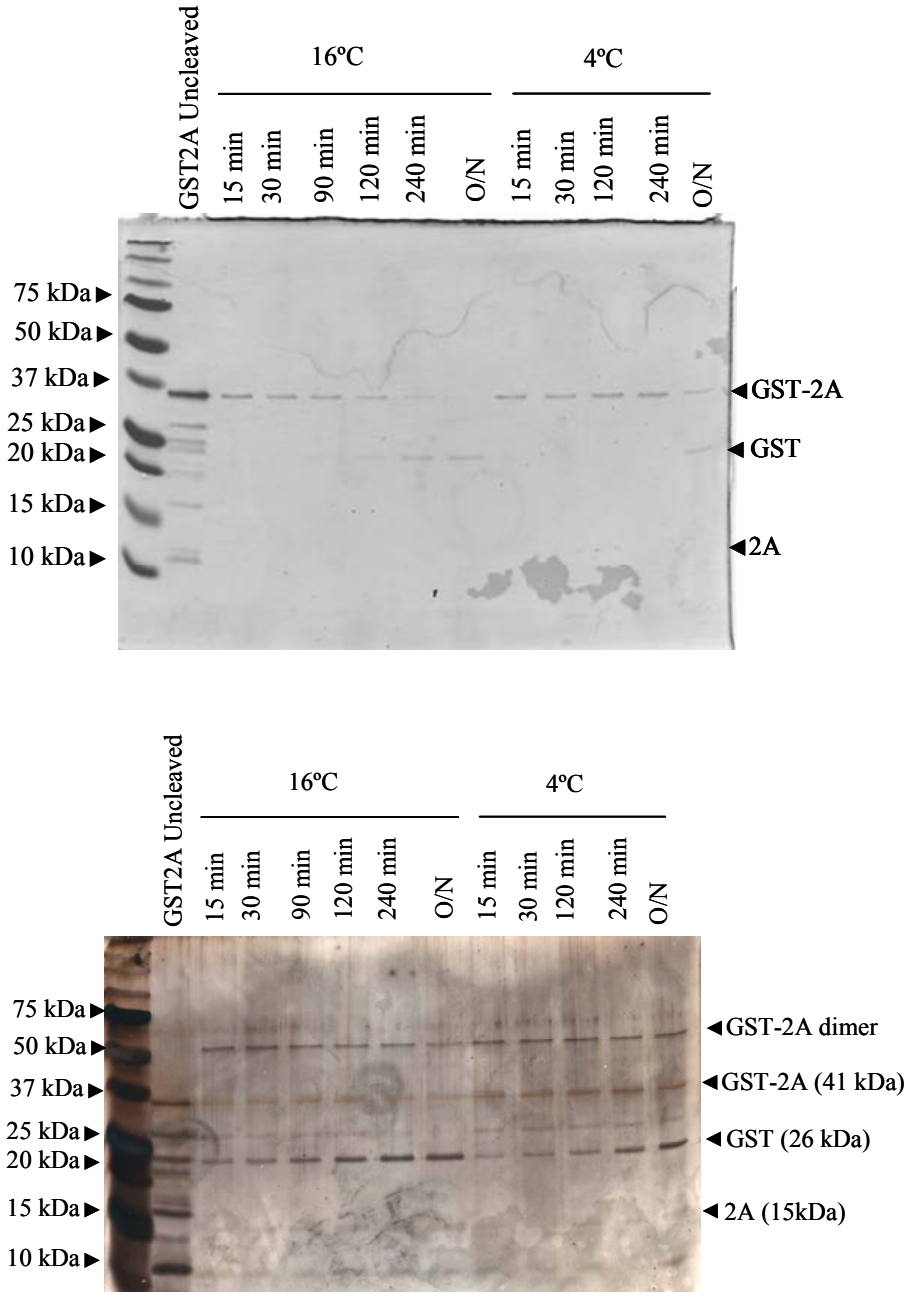


Figure 44. Small scale GST-2A thrombin digestion trial using different conditions. SDS-PAGE above: Coomassie stained; SDS-PAGE below: Silver stained. The optimum digestion conditions were 16˚C and overnight. 2A was difficult to see due to its small size and small amounts digested. The positions of the proteins size markers are indicated. Minor band staining at 25 kDa could be 2A dimer.

3.2.3 Large scale induction, purification and thrombin digestion

Based on the findings of small scale experiments, a large scale experiment was set up. GST-2A$_{TMEV}$ protein was induced (4 litres culture) and the tagged protein purified (2mg/ml) through 1ml Glutathione-Sepharose column (Invitrogen Ltd, Paisley, UK). GST-2A eluted samples were dialyzed (Dialysis tubing size 7; 23.8 mm diameter; MW: 12-14 kDa; Medicell International Ltd, London, UK) and digested with thrombin (1unit/100µg of protein). The cleavage was not as efficient as expected, with a high proportion of the fusion protein not being cleaved by thrombin (figure 45).

Despite our previous findings cleavage of GST was not as efficient as we expected. Failure to separate GST from 2A, allied to the need for larger amounts of pure 2A for antibody production, led us to abandon this approach in favour of procedure 3.2.4 (see below), whereby GST-2A was purified for anti-GST-2A antibody production.

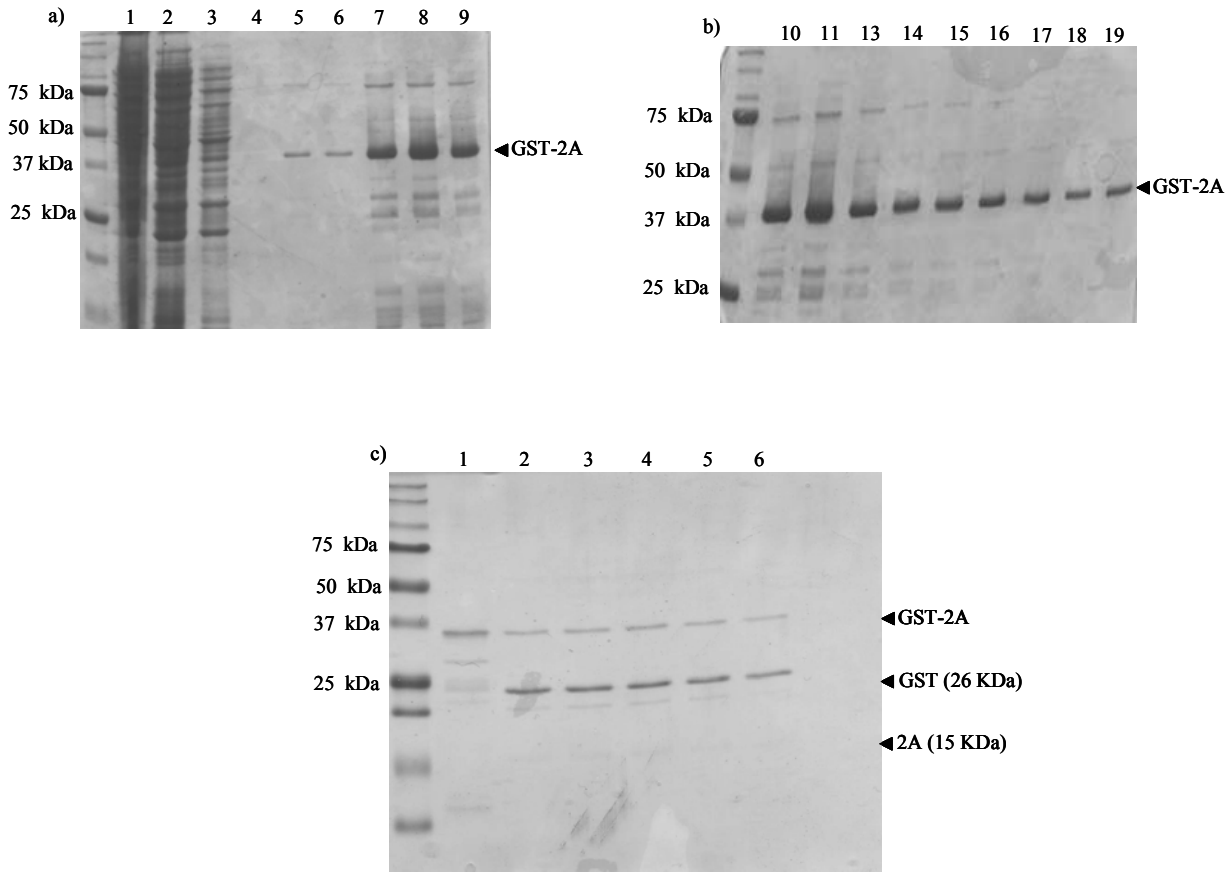


Figure 45: Large scale GST-2A purification and digestion with thrombin. a) and b) GST-2A purification from 4 litres IPTG induction of *E.coli* BL21 (DE3) cells. The lanes are labeled 1-19 and represent all the fractions collected. Lane 1: GST-2A supernatant before loading onto the column. Lane 2: Flow-through from column after sample added. Lane 3: Flow-through during washes with PBS after loading sample. Lanes 4 to 19: Fractions collected during elution with Elution Buffer (5mM to 10mM reduced glutathione in 50mM Tris-HCl, pH 9.5). c) GST-2A samples digested with thrombin (16 ºC, overnight). Lane 1: GST-2A uncleaved sample after dialysis.

3.2.4 Induction of large amounts of protein for TMEV-2A antibody production

` 8 litres of *E.coli* BL21 (DE3) transformed with pGEX-2T, containing $2A_{TMEV}$ (figure 36), were induced with IPTG (0.5mM) and the GST-2A complex purified. 2 mg/ml of protein complex were digested with thrombin (1u/100µg of protein) overnight.

Eluted fractions are shown in figure 47. The bands were excised and identified by MALDI, MALDI-TOF MS and MS/MS methods (Mass spectrometry and proteomics facility, Biomolecular Sciences Building, University of St. Andrews) (data not shown).

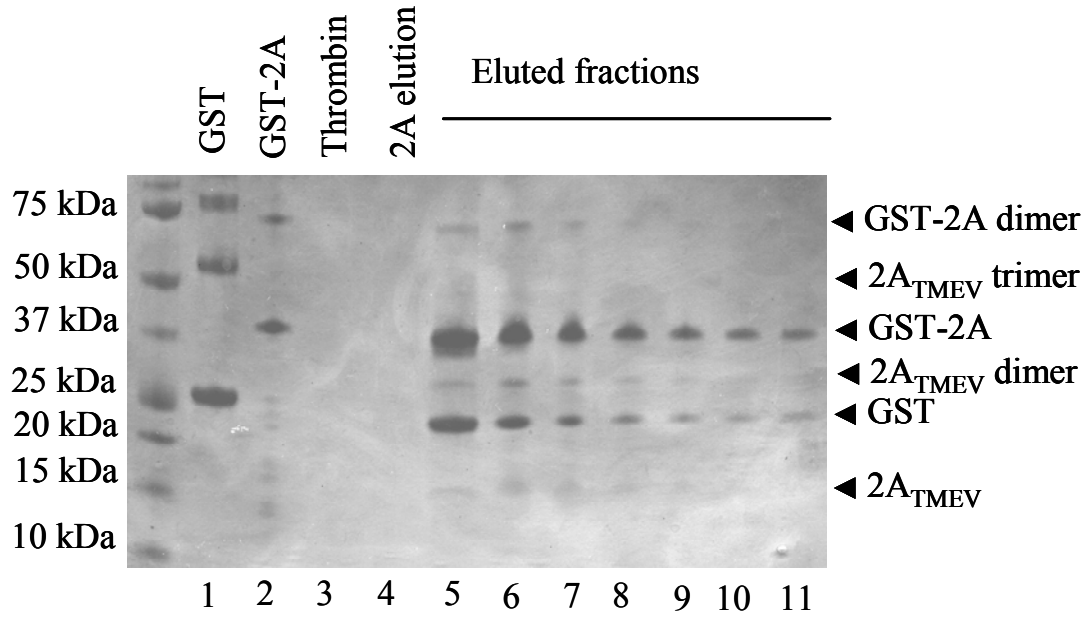


Figure 47. SDS-PAGE of GST-2A purified fractions, digested with thrombin eluted with glutathione. Lane 1: GST induced form pGEX-5X Lane 2: uncleaved GST-2A. Lane 3: Thrombin in PBS (too diluted to be observed (37 kDa). Lane 4: 2A eluted with PBS washes, after thrombin cleavage. Lanes 5 to 11: Eluted fractions with Glutathione (to check the induction worked).

These results indicate 2A forms multimers (dimers, trimers). This may explain why 2A was not being eluted through the column and staying attached to the bound uncleaved GST-2A on column.

Due to the difficulties purifying 2A protein from GST-2A, the strategy was shifted to obtaining large amounts of induced GST-2A, which was, then, used for antibody production (Pentlands Science Park, Penicuik Midlothian). 2mg/ml GST-2A was used as immunogen to raise antibodies in sheep. Antibodies were produced not only against $2A_{TMEV}$ but also GST. GST antibodies were extracted by using Glutathione Sepharose Beads, bound to GST. Furthermore, $2A_{TMEV}$ antibodies were tested *via* ELISA.

2 litres of *E.coli* BL21 (DE3) transformed with GST-2A plasmid were induced as before and purified (figure 48).

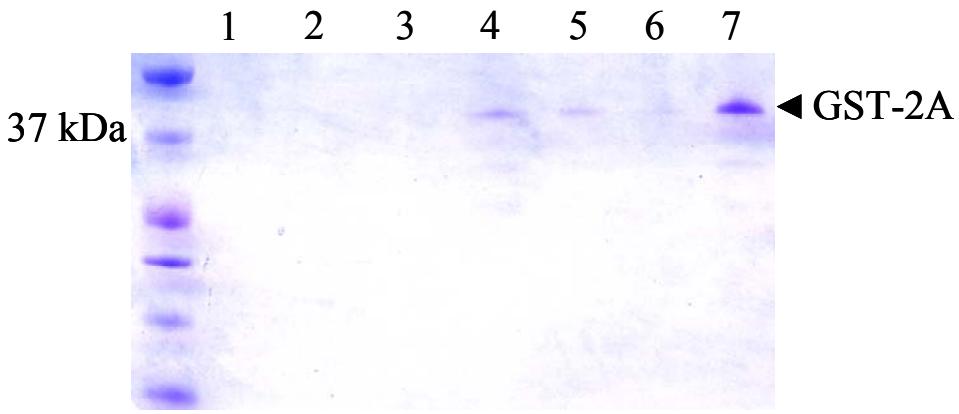


Figure 48. SDS-PAGE with GST-2A eluted fractions.

The eluted fraction number 7, shown in figure 48, was determined to be 2.39 mg/ml (figure 49).

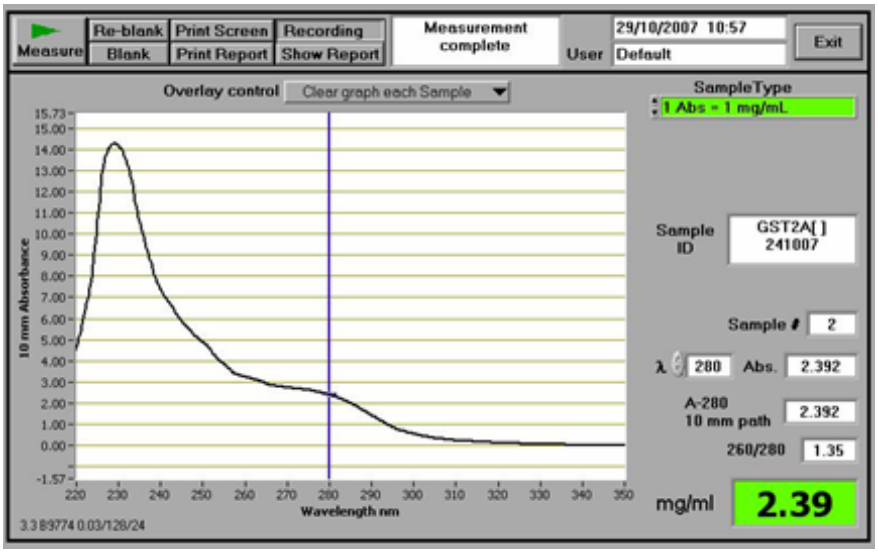


Figure 49. Measurement of protein concentration and purity. Absorbance at 230nm is due to peptide bonds and the absorbance at 280nm is due to light absorption by aromatic amino acids, such as, tyrosine, tryptophan and phenylalanine. The concentration of the eluted GST-2A fraction number 7 shown in figure 48 was 2.39 mg/ml.

Future work will involve testing these $2A_{TMEV}$ antibodies against artificial polyprotein constructs containing this protein, and probing them within the cell.

Using this new resource we should be able to study $2A_{TMEV}$ behaviour within the cell and resolve a number of interesting questions eg. What happens when 2A is cleaved away from the capsid proteins?

3.3. Creation of a reporter of stress in the cell using $2A_{FMDV}$ protein

The aim of this part of the project was to create an artificial construct as a reporter of stress in the cell. This is based on the ability to produce two products from a single ORF. The hypothesis tested here is that when cells are not stressed, normal levels of eEF2 activity are present resulting in efficient 're-initiation' of the translation of sequences downstream of 2A - obtaining a 1:1 ratio of the products up- and down-stream of 2A. When cells become stressed, eEF2 becomes phosphorylated, elongation rates are reduced and the efficiency of re-initiation of sequences downstream of 2A is reduced. This would produce a molar excess of translation products up- over those downstream of 2A. The system being tested here is a reporter protein (CFP) tagged with a V5 epitope upstream of 2A and a single-chain antibody (ScFv binds the V5 epitope) downstream of 2A (an "intrabody"). The rationale being that whilst (free) [V5-CFP-2A] is small enough to diffuse through nuclear pores into the nucleus, a [V5-CFP-2A]: ScFv complex cannot.

Analysis of this construct in cells would provide evidence that this type of reporter system could be used in cells. The final reporter system would be a V5 epitope tagged transcriptional transactivator (TTA) linked via 2A to the ScFv ([V5-TTA-2A-ScFv]). The uncleaved protein (V5-TTA-2A-ScFv) or the cleaved protein [V5-TTA-2A] bound to ScFv is too large to diffuse into the nucleus and, thus, is localized in the cytoplasm. In contrast, [V5-TTA-2A] not bound to ScFv is small enough to diffuse through the nuclear pores, and would be localized both in the cytoplasm and inside the nucleus.

In normal conditions for the cell, [V5-TTA-2A] and scFv are produced in a 1:1 ratio. However, in stressful conditions the [V5-TTA-2A] product will be produced in a higher proportion than the scFv, and will enter the nucleus, since not enough scFv will be produced to bind [V5-TTA-2A]. Using stable cell lines expressing luciferase under the control of the tetracycline operator system, [V5-TTA-2A] will activate luciferase expression and the amount of this luciferase expression will represent the level of stress in the cell (see section 4.3).

The first step of the project was to test the ability of the singe chain variable fragment (scFv) to bind V5 inside the cell. A 'test' construct, which has CFP instead of TTA, was created (pHE27; figure 50) and used to transfect 293 T cells. The results were positive; showing a good binding between V5 and ScFv.

V5-CFP-2A fragment was amplified from Lh135 plasmid (V5-CFP-2A-V5-ΔPAC; kindly provided by Dr. Hughes) using the oligonucleotide primer forward Lh135f, which has an N-terminal NheI restriction site, and the reverse primer Lh135r, which has a C-terminal BamHI site (see table 3 in section 2.1.17). The PCR product was gel purified, restricted with both Nhe I and BamHI enzymes and gel purified again.

ScFv was amplified from pPDF83 (kindly provided by Dr. de Felipe) using the oligonucleotide primer forward ScFvf, which has an N-terminal BamHI restriction site, and the reverse primer ScFvr, with a C-terminal HindIII restriction site (see table 3 in section 2.1.17). The PCR fragment was gel purified, restricted with BamHI and HindIII and gel purified again. Lh135 vector was digested with NheI and HindIII and gel purified. This restricted plasmid was used as a vector for a triple ligation together with V5-CPF-2A and ScFv digested PCR products. This triple ligation led to the formation of a new vector, pHE27, which will be used to test the future stress cell selector system (figure 50).

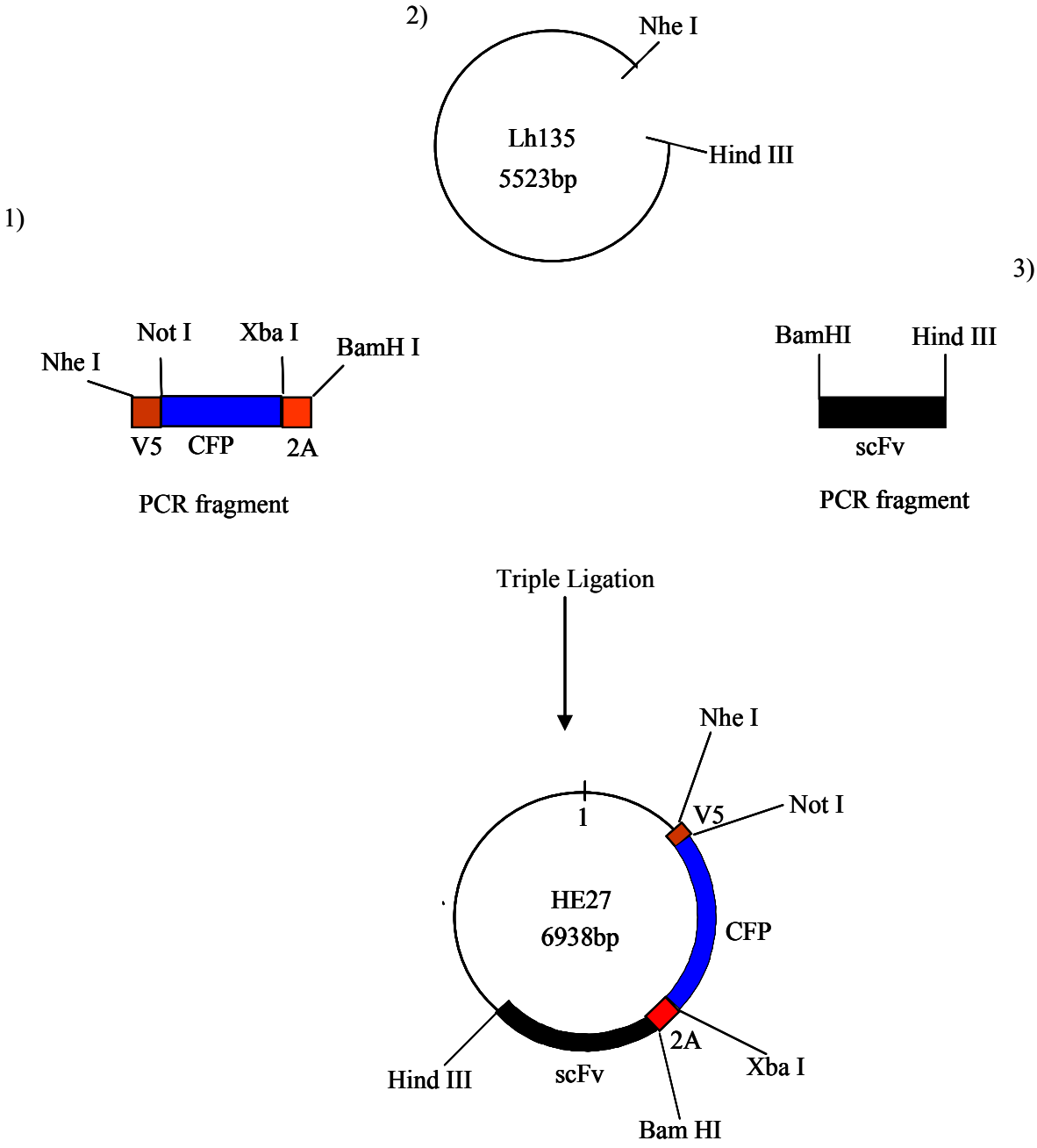


Figure 50. Overview of the cloning strategy utilized for the formation of HE27 construct.1) and 3) are the inserts, while 2) is the vector for the performed triple ligation that led to the formation of HE27 vector.

3.3.1 Testing pHE27 in vitro

Wheat germ *in vitro* transcription/translation with the pHE27 plasmid construct was performed and the obtained products analyzed on SDS-PAGE. As expected, three products were observed: full-length protein [V5-CFP-2A-scFv] and processed products, [V5-CFP-2A] and [scFv] (figure 51).

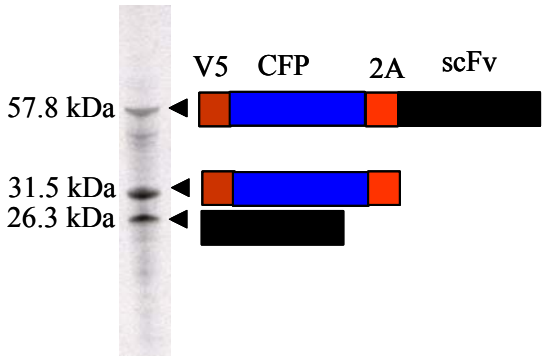


Figure 51. Translation profile of pHE27. Coupled *in vitro* transcription/translation wheat germ extracts were programmed with the pHE27 plasmid construct and radiolabelled with [^{35}S]-methionine (10µCi/µl). The full-length and processed translation products are shown.

3.3.2 Testing pHE27 in cells: Western Blot, Immunofluorescence and Immunoprecipitation

pHE27 vector was transfected into 293T cells, the protein expression assessed by probing a Western Blot of total cellular proteins with anti-V5, anti-2A$_{FMDV}$ and anti-CFP antibodies. Immunofluorescence and immunoprecipitation was also performed using anti-V5, anti-2A$_{FMDV}$ and anti-CFP antibodies.

In these experiments Lh135 plasmid was used as a positive control (figure 52) and untransfected cells were used as a negative control.

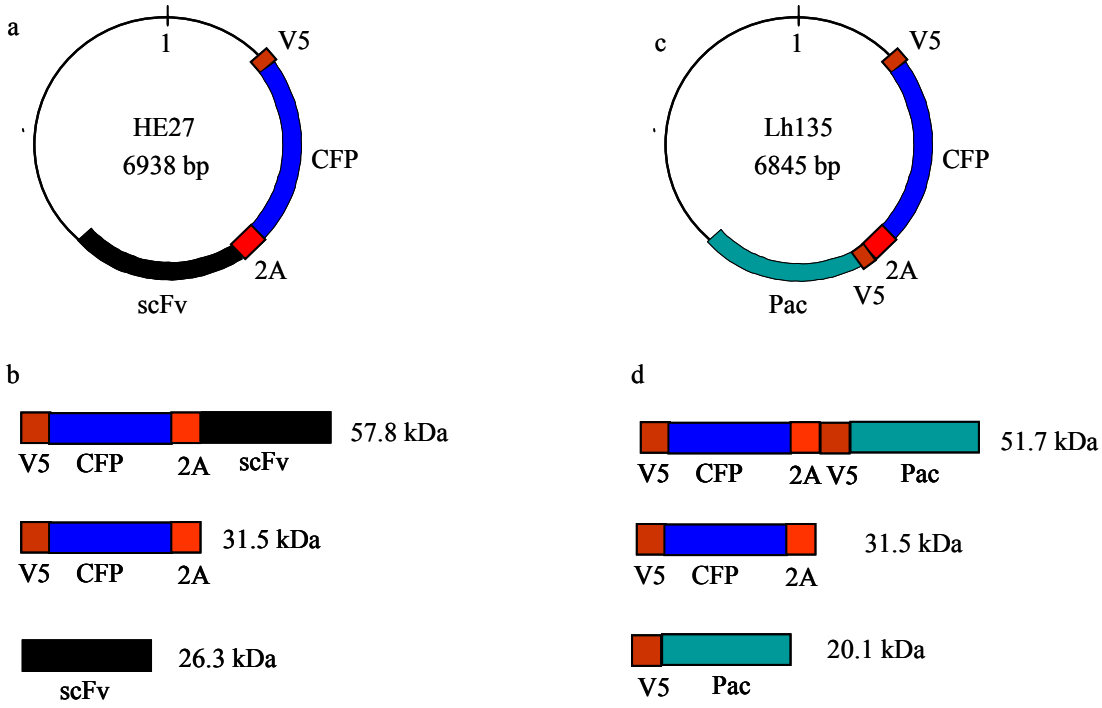


Figure 52. Schematic illustration of pHE27 and pLh135 vectors and their translation products. Both vectors, a and c, contain an artificial polyprotein, which is encoded by a single open reading frame. b and d show the full-length polyproteins and the processed products of pHE27 and pLh135, respectively. V5: small epitope, 14 amino acids long, in the C-terminal of P and V Simian virus 5. CFP: cyan fluorescent protein. scFv: single chain variable fragment from anti-V5. Pac: puromycin-N-acetyltransferase.

Several Western Blots were performed after transfection, and revealed the expected translation products (figure 53, 54 and 55).

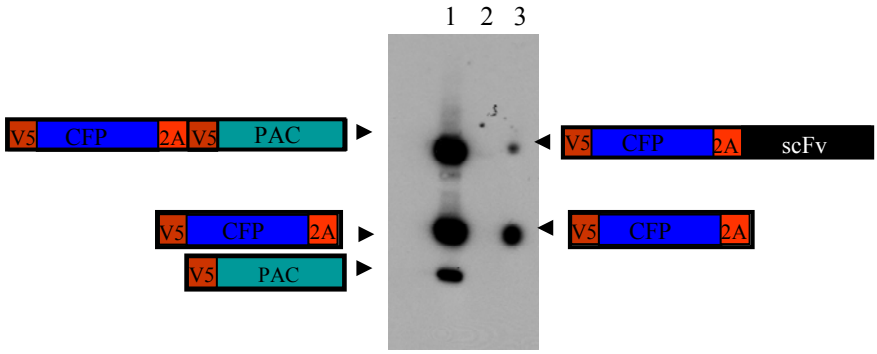


Figure 53. Western Blot anti-V5. Artificial polyproteins were expressed in eukaryotic cells and probed against anti-V5 antibodies. Lane 1: positive control, pLh135, showing 3 translation products, the full-length protein, and the two processed ones. Lane 2: Negative control. Lane 3: pHE27 vector shows the uncleaved product and the cleaved one, recognized by anti-V5 antibody.

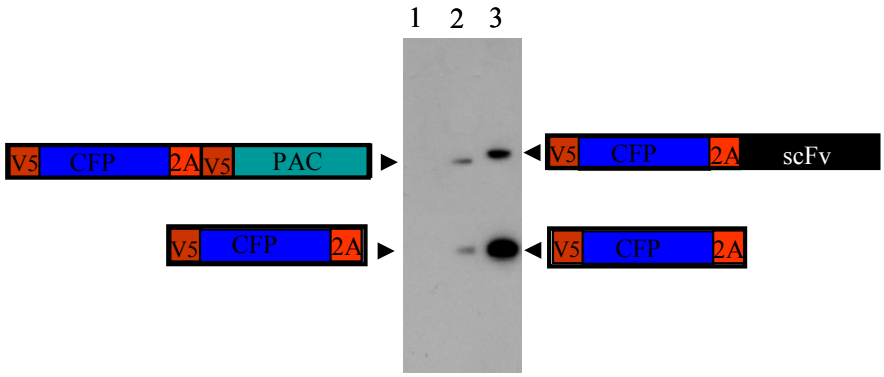


Figure 54. Western Blot anti-CFP. Artificial polyproteins were expressed in eukaryotic cells and probed against anti-CFP antibodies. The full length and processed translation products are indicated in each case. Lane 1: negative control. Lane 2: positive control; 2 translation products were recognized by the antibody. Lane 3: pHE27; full-length protein and upstream processed product were recognized by anti-CFP.

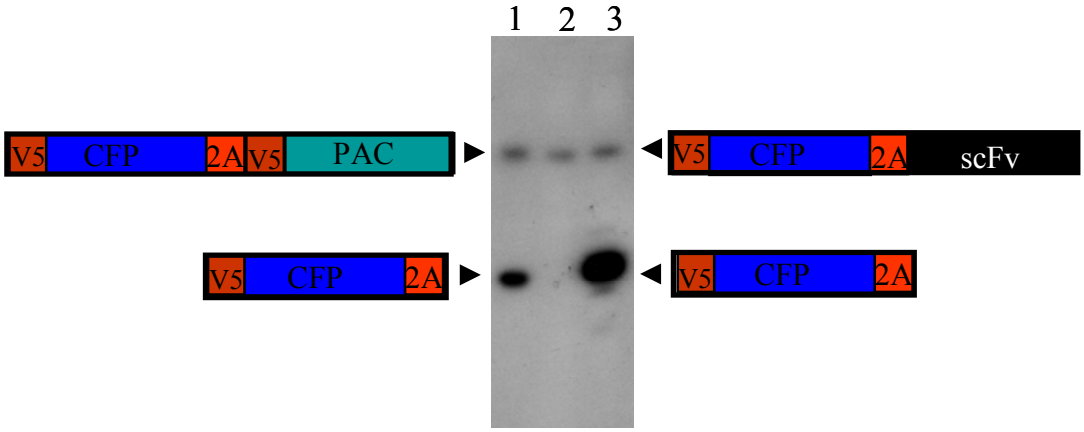


Figure 55. Western Blot anti-2A. The full length and processed translation products are indicated in each case. Lane 1: positive control, showing two translation products. Lane 2: negative control. Lane 3: pHE27; full-length protein and upstream processed product were recognized by anti-2A.

Furthermore, immunofluorescence experiments showed clear cytosolic localization of the pHE27 translation products (figure 56). These images suggest ScFv binds V5 within the cytoplasm, thus acting as an 'intrabody'.

Immunofluorescence using antibodies against V5 and 2A were performed. Unfortunately, anti-V5 antibodies did not show any signal, probably due to V5 epitope being masked in pHE27, and the fluorescence signal obtained from CFP was too weak (data not shown).

Anti-2A antibodies used in cells transfected with pHE30 (positive control consisting of [V5-CFP-2A] without scFv) showed [V5-CFP-2A] to be uniformly distributed throughout the cell (nucleus + cytoplasm). In contrast, cells infected with pHE27, indicated [V5-CFP-2A] was localized in the cytoplasm, and probably excluded from the nucleus. NOTE: the weak fluorescence appearing to arise from the nucleus can be accounted for by the small amount of cytoplasm above and below the nucleus.

These promising data indicate that further development of this system may provide a 'read-out' of cellular stress.

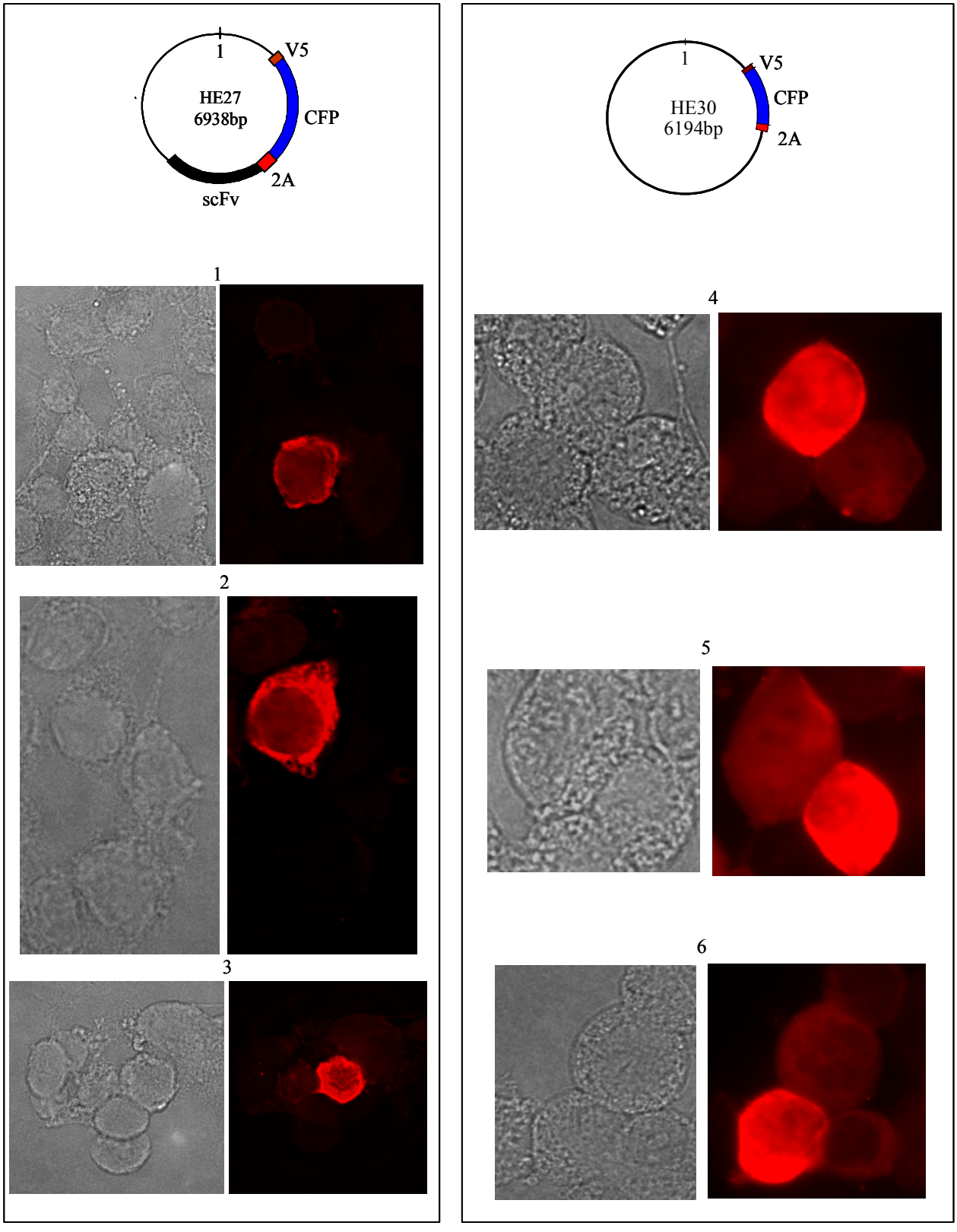


Figure 56. scFv binds to V5 within the cell. Imunofluorescent images of 293T cells, transfected with pHE27 and pHE30, using Texas red anti-2A antibody. Images 1-3: pHE27 shows cytosolic localization. In contrast, pHE30 (images 4-6) is localized everywhere in the cell. Black and white pictures are the phase/reference images.

Immunoprecipitation results from *in vivo* expression in 293T cells were consistent with the cell imaging results shown in figure 56 (figure 57).

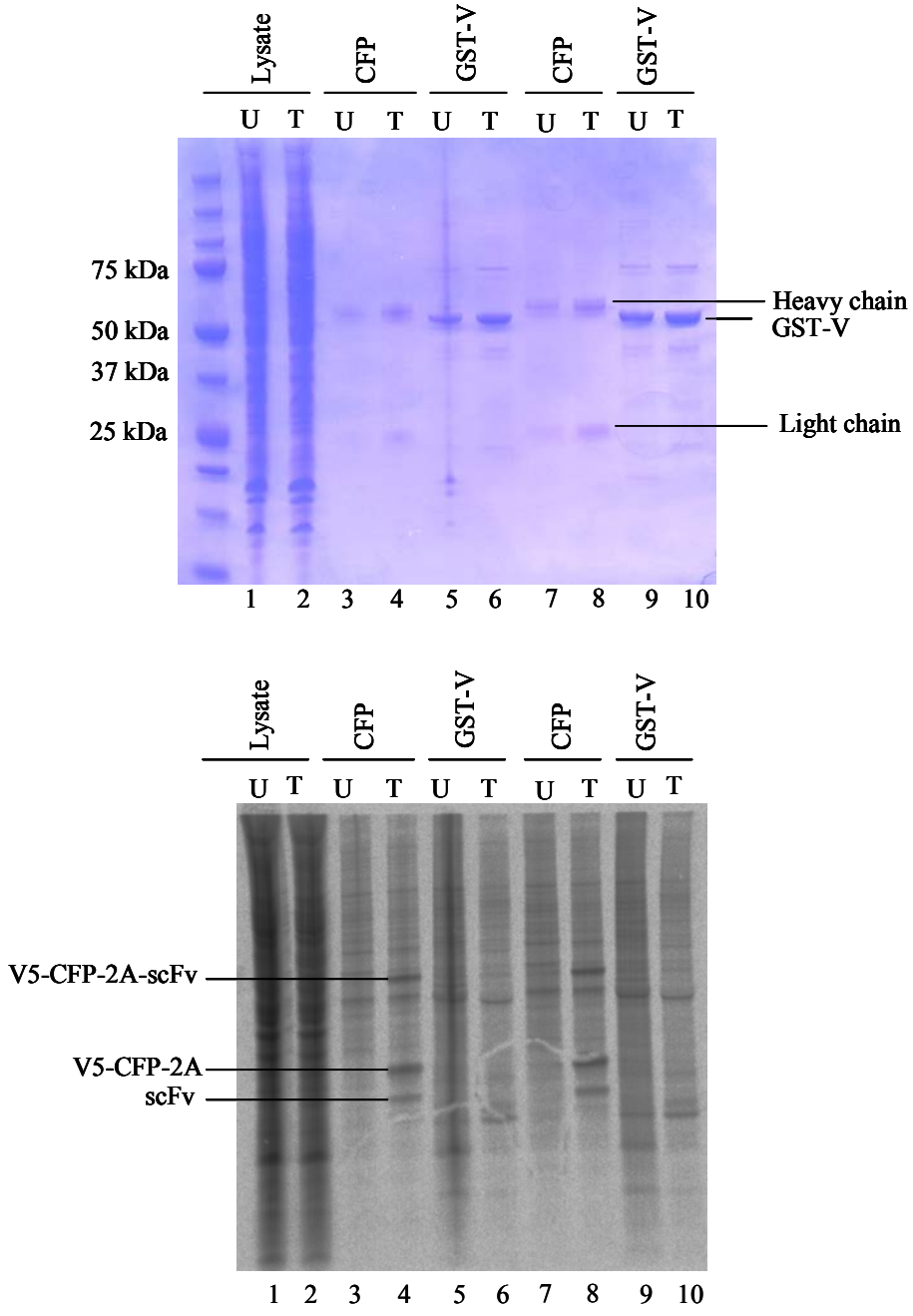


Figure 57. *In vivo* expression and immunoprecipitation using anti-CFP and GST-V protein. Above, SDS-PAGE; U: untransfected; T: transfected. Lanes 7-10 are the same as lanes 3-6 but 10 µl and 20 µl were added to the gel, respectively. Below, Phosphorimaging image of SDS-PAGE gel.

In vitro immunoprecipitation (IP) was also performed and confirmed the results shown in the immunofluorescence images; scFv was clearly binding V5 within the cell (figure 58).

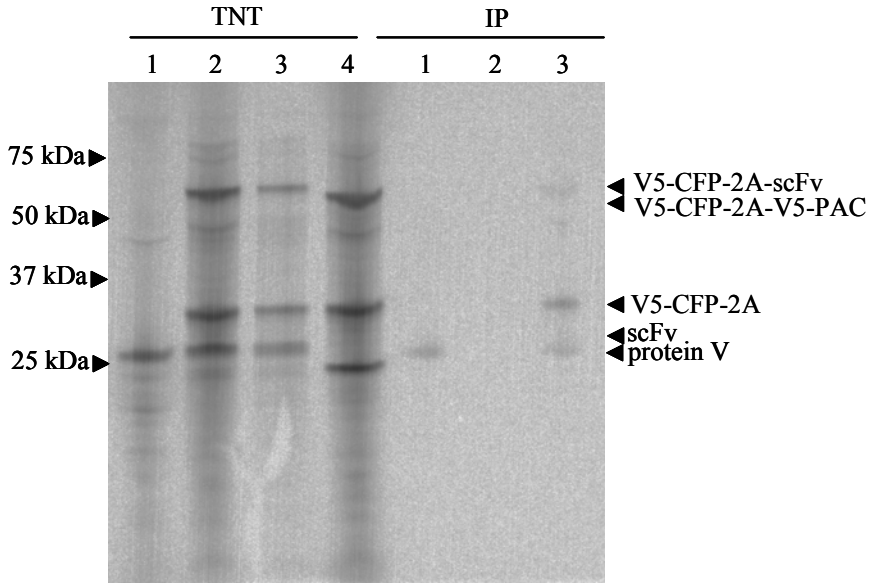


Figure 58. Translation *in vitro* and immunoprecipitation (IP) from these *in vitro* translations. Coupled transcription/translation (TNT) products were immunoprecipitated using anti-V5 antibody. The *in vitro* translation and immunoprecipitations were analyzed by SDS-PAGE (4-12%). The first four lanes show the *in vitro* translated products: 1) pGEM3.V as a control, which expressed Simian parainfluenza virus 5 V protein (23.8 kDa); 2) pHE27, as shown in figure 51; 3) plasmids pGEM3.V and pHE27 together; 4) pLh135 as a control. The 3 lanes on the right showed the results for the immunoprecipitation for 1-3 TNTs.

These results confirmed that the scFv binds the V5 protein *in vivo* (figure 57) and also *in vitro* (figure 58). Immunoprecipitation experiments showed anti-CFP antibodies immunopreciptated not only [V5-CFP-2A] but also the scFv protein (bound to the V5 epitope; figure 57). In contrast, in the immunoprecipitation with GST-V, no scFv was pulled down: presumably due to the sequestration of the scFv by the V5 epitope of [V5-CFP-2A] (figure 57). Immunoprecipitation from the *in vitro* translation reactions was consistent with the *in vivo* results (figure 58). IP of Lh135 was not performed because pGEM3.V was used as a control for the IPs and Lh135 as

a control for the *in vitro* translation reactions. No proteins were detected in the IP anti-V5 against pHE27 translation products (lane 2) since all [V5-CFP-2A] was bound to scFv. On the other hand, in the IP of [V5-CFP-2A]:scFv complex and pGEM3.V together (figure 58, lane 3) some [V5-CFP-2A-scFv], [V5-CFP-2A] and V protein were recognized by the antibody anti-V5, probably because scFv was not only binding the full-length and processed product derived from pHE27 but also pGEM3.V. Thus, there is some free, unbound, [V5-CFP-2A] from pHE27 and pGEM3.V that can be recognized by the antibody.

The immunoprecipitated proteins (figure 58) were determined by probing a Western Blot with anti-2A antibodies. [V5-CFP-2A-scFv] full-length (uncleaved) protein and processed translation products were identified (figure 59).

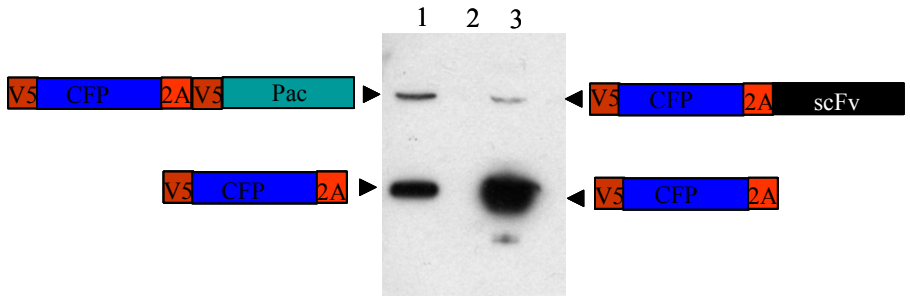


Figure 59: Western Blot of immuno-precipitations. Immuno-precipitated samples (fig) were identified by Western Blot using anti-2A antibodies. Lane 1: Lh135 as positive control. Lane 2: Untransfected 293T cells as negative control. Lane 3: HE27. The full-length and processed translation products are indicated.

Evidence from co-immunoprecipitation and the cell imaging suggest that this strategy may well provide a method to study the effects of stress on the translational apparatus of the cell. Future work will involve the creation of the definitive, more sensitive, reporter (V5-TTA-2A-scFv), which may allow us to detect stress using transcriptional transactivation.

3.4 2A-like sequences

3.4.1 Insect virus 2A-like sequences

Previously identified insect 2A-like sequences (Donnelly *et al*., 2001a) together with newly identified sequences (Luke *et al*., 2008), were re-cloned in a new vector pBiEx-3 (figure 60). This vector contains the *Autographa californica* nuclear polyhedrosis virus (AcNPV) immediate early promoter (IE1), to direct expression in insect cells, and the T7 lac promoter for expression in *E.coli*.

The purpose of this study was to test these newly identified insect 2A-like sequences (30aa) and longer (30 aa) versions of those previously identified (Luke *et al*., 2008), in a new insect *in vitro* system (Qiagen Ltd., West Sussex, UK).

The 2A-like 30 amino acids long sequences are represented in figure 61.

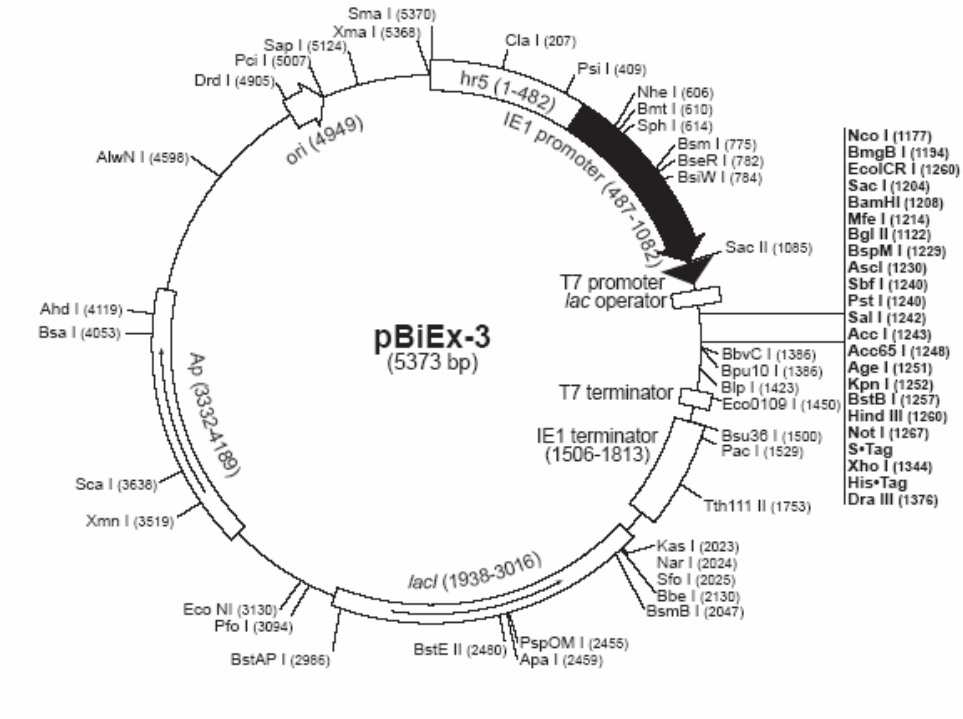


Figure 60: pBiEx-3 vector (Novagen, Nottingham, UK) contains both insect (IE1) and bacterial (T7) promoters.

IFV	P	S	I	G	N	V	A	R	T	L	T	R	A	E	I	E	D	E	L	I	R	A	G	I	E	S	N	P	G	P
PnPV-2A1	G	Q	R	T	T	E	Q	I	V	T	A	Q	G	W	V	P	D	L	T	V	D	G	D	V	E	S	N	P	G	P
PnPV-2A2	T	R	G	G	L	R	R	Q	N	I	I	G	G	G	Q	K	D	L	T	Q	D	G	D	I	E	S	N	P	G	P
VDV-1	E	Y	E	L	E	C	V	T	S	L	L	Q	L	S	N	P	V	S	A	K	P	E	M	D	N	P	N	P	G	P
CrPV	L	V	S	S	N	D	E	C	R	A	F	L	R	K	R	T	Q	L	L	M	S	G	D	V	E	S	N	P	G	P
PrV-2A1	L	E	M	K	E	S	N	S	G	Y	V	V	G	G	R	G	S	L	L	T	C	G	D	V	E	S	N	P	G	P
PrV-2A2	N	S	D	D	E	E	P	E	Y	P	R	G	D	P	I	E	D	L	T	D	D	G	D	I	E	K	N	P	G	P
PrV-2A3	T	L	M	G	N	I	M	T	L	A	G	S	G	G	R	G	S	L	L	T	A	G	D	V	E	K	N	P	G	P
B. mori CPV1-I	R	T	A	F	D	F	Q	Q	D	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	I	E	S	N	P	G	P
O. brumata CPV18	I	H	A	N	D	Y	Q	M	A	V	F	K	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P

Figure 61. 2A-like amino acid sequences (30 aa long). *Infectious flacherie virus (IFV), perina nuda picorna-like virus (PnPV), ectropis obliqua picorna-like virus (EoPv), varroa destructor virus-1 (VDV-1) cricket paralysis virus (CrPV), providence virus (PrV), bombyx mori cypovirus- 1* (BmCPV-1) and *operophtera brumata cypovirus 18* (OpbuCPV-18) 2A-like sequences are shown. The conserved C-terminal motif is highlighted in yellow.

Reporter plasmid pSTA1, containing a single ORF encoding green fluorescent protein (GFP), the 30aa version of the insect 2A-like sequences and β-glucoronidase (GUS), were restricted with the restriction enzymes BamHI and XbaI and the large fragment purified by agarose gel electrophoresis.

Cyan fluorescent protein (CFP) was PCR amplified from pMK3 (CFP-2A-Δpac; kindly provided by Dr. de Felipe) using the forward oligonucleotide primer CFPf, which has an N-terminal BamHI restriction site, and the reverse primer CFPr, which contains the human c-Myc epitope tag sequence and a C-terminal XbaI site (see table 3 in section 2.1.17). The PCR product was doubly restricted with BamHI and XbaI, gel purified, then ligated into pSTA1 similarly restricted. These 10 new plasmids ($HE24_{1\text{-}10}$) were restricted with BamHI and XhoI, and the small 2.6 Kb fragments (CFP-2A-GUS) gel purified. The purified inserts were ligated into similarly restricted pBiEx-3 vector to give $HE25_{1\text{-}10}$ (figure 62).

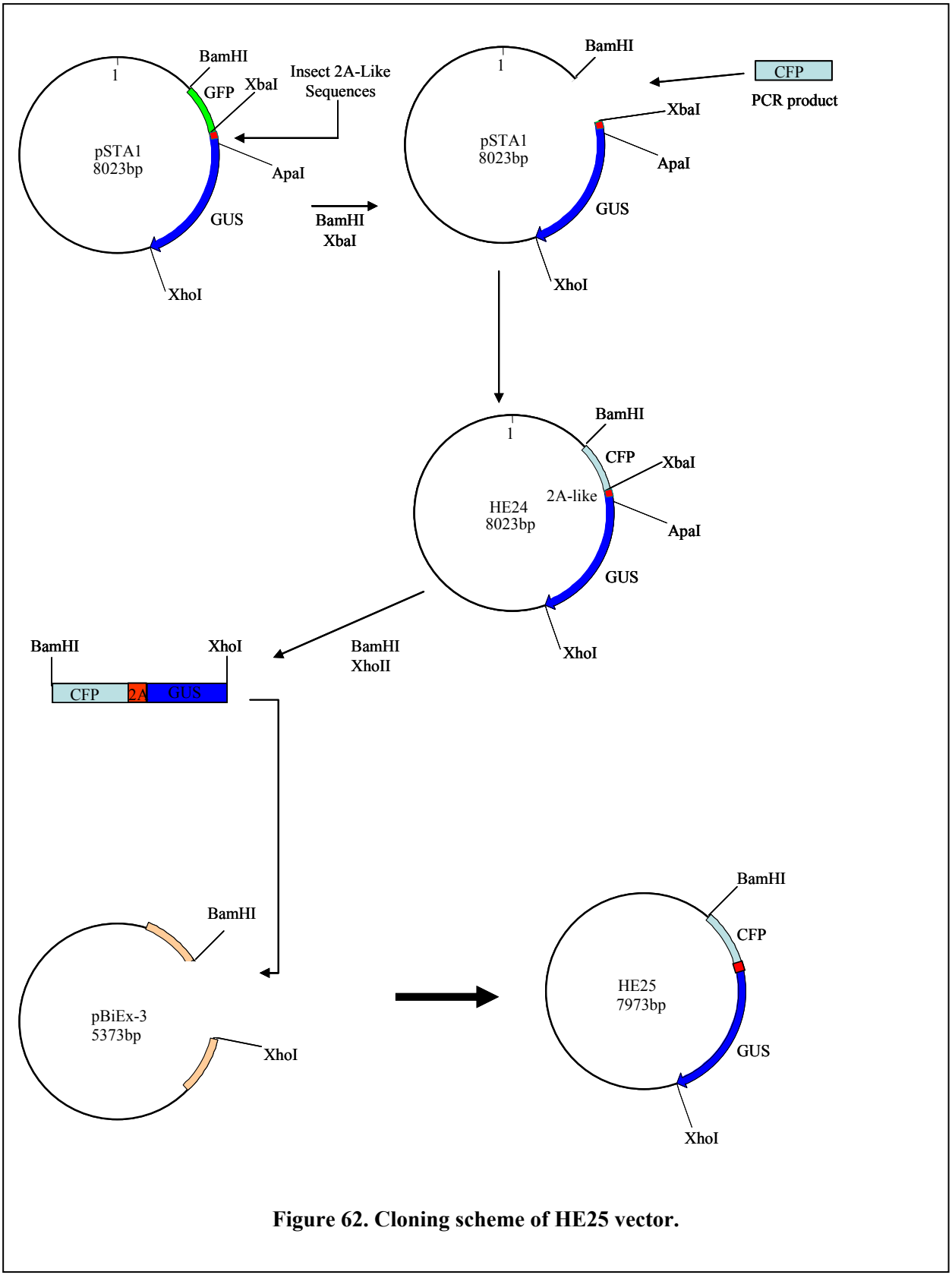


Figure 62. Cloning scheme of HE25 vector.

$HE25_{1\text{-}10}$ constructs were used to programme insect extract transcription/ translation reactions (Insect Direct™ System, Novagen**,** Nottingham, UK). Surprisingly, no translation profiles were obtained (figure 63 a).

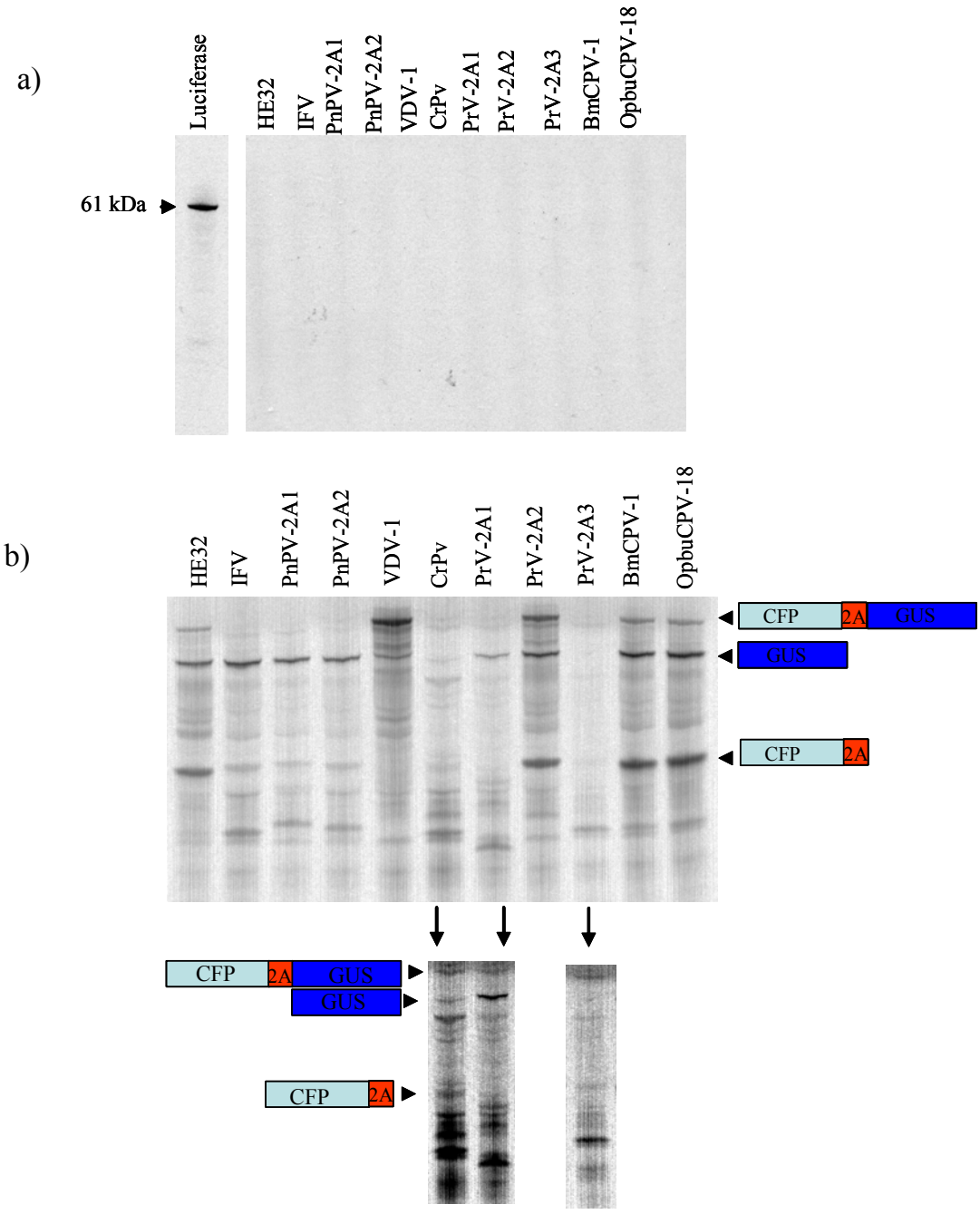


Figure 63. a) Insect extract transcription/translation reactions. Luciferase as a positive control, monomeric protein, 61 kDa. DNA is pIX4 vector. **b)Translation profiles of the constructs containing 2A-like sequences of insect origin.** Coupled transcription/translation wheat germ extract reactions were programmed with constructs encoding the 2A-like sequences in the CFP-2A-GUS artificial polyprotein system. HE32 vector, containing CFP-2A$_{FMDV}$-GUS profile was included for reference. The lanes are labelled with the names of the insect species. The full length translation product (CFP-2A-GUS), and processing products ([CFP-2A] and [GUS]) are indicated.

To check the integrity of the constructs, the plasmids were used to programme wheat germ extract coupled transcription/translation reactions. The radiolabelled translation products were separated by 4-12% Bis-Tris SDS-PAGE and visualized by autoradiography or phosphorimaging. HE32 (CFP-$2A_{FMDV}$-GUS) was used as a positive control, and showed an expected gel profile with three major translation products; full-length polyprotein ([CFP-$2A_{FMDV}$-GUS]) and the processing products ([CFP-$2A_{FMDV}$] and [GUS]) (figure 63 b).

These data are consistent with Luke *et al.* (2008). Iflaviruses, *Infectious flacherie virus* (IFV) and *Perina nuda picorna-like virus* (PnPV), both $2A_1$ and $2A_2$, showed high cleavage activity. *Varroa destructor virus-1* (VDV-1) did not show cleavage activity. Its 2A-like sequence does not completely match the conserved motif -**D**x**E**xNPGP- but contain the motif –**M**D**N**PNPGP.

Cricket paralysis virus (CrPV) showed low cleavage activity together with *Providence virus* $2A_2$ (PV). In contrast, PV-$2A_1$ and PV-$2A_3$ showed higher cleavage activity. The 2A-like sequences belonging to *Cypoviridae*, *Bombyx mori cypovirus type1* (BmCPV-1) and *Operophtera brumata cypovirus-18* (OpbuCPV-18) cleavage is higher than the observed in CrPV and PV-$2A_2$ but lower than PnPV $2A_1$ and $2A_2$.

3.4.2 Strongylocentrotus purpuratus 2A-like sequences

Many new 2A-like sequences were found in the genome of the purple sea urchin (*Strongylocentrotus purpuratus*), when probing databases with the conserved motif at the C-terminal of 2A (-DxExNPGP-).

More than 150 2A-like sequences have been identified within the purple sea urchin genome, and analyses show these sequences are distributed between two types of cellular sequences: intracellular genes involved in innate immunity and non-LTR retrotransposons, similar to Trypanosome 2A-like sequences. 2A is within or close to the N-terminus of the single ORF in both groups.

Str-14 and Str-34, which are two different *Strongylocentrotus purpuratus* 2A-like sequences belonging to the first group of genes (i) and the second one (ii), respectively, were cloned into a reporter plasmid (Lh136, kindly provided by Dr. de Felipe) and its activity analysed. Reporter plasmid Lh136, containing a single ORF encoding cyan fluorescent protein (CFP), $2A_{FMDV}$ sequence and puromycin-N-acetyltransferase fragment (Δpac), (kindly provided by Dr. de Felipe) was used as a

template to insert the different *Strongylocentrotus purpuratus* 2A-like (Str) sequences using PCR. Primer forward Str-14, which contains the Str 2A sequence and an N-terminal XbaI restriction site, and reverse primer ΔPACrev, with a C-terminal HindIII, were used to amplify Str-14 (figure 64). Furthermore, primer forward Str-34, with the same characteristics as Str-14 but with a different Str 2A sequence, and ΔPACrev, were used to amplify Str-34 (figure 64). There are two versions of forward primers, the wild type version and the mutated version (see table 3 in section 2.1.17). These two PCR products were doubly restricted with the restriction enzymes XbaI and HindIII, gel purified, then ligated into Lh136 similarly restricted (figure 64).

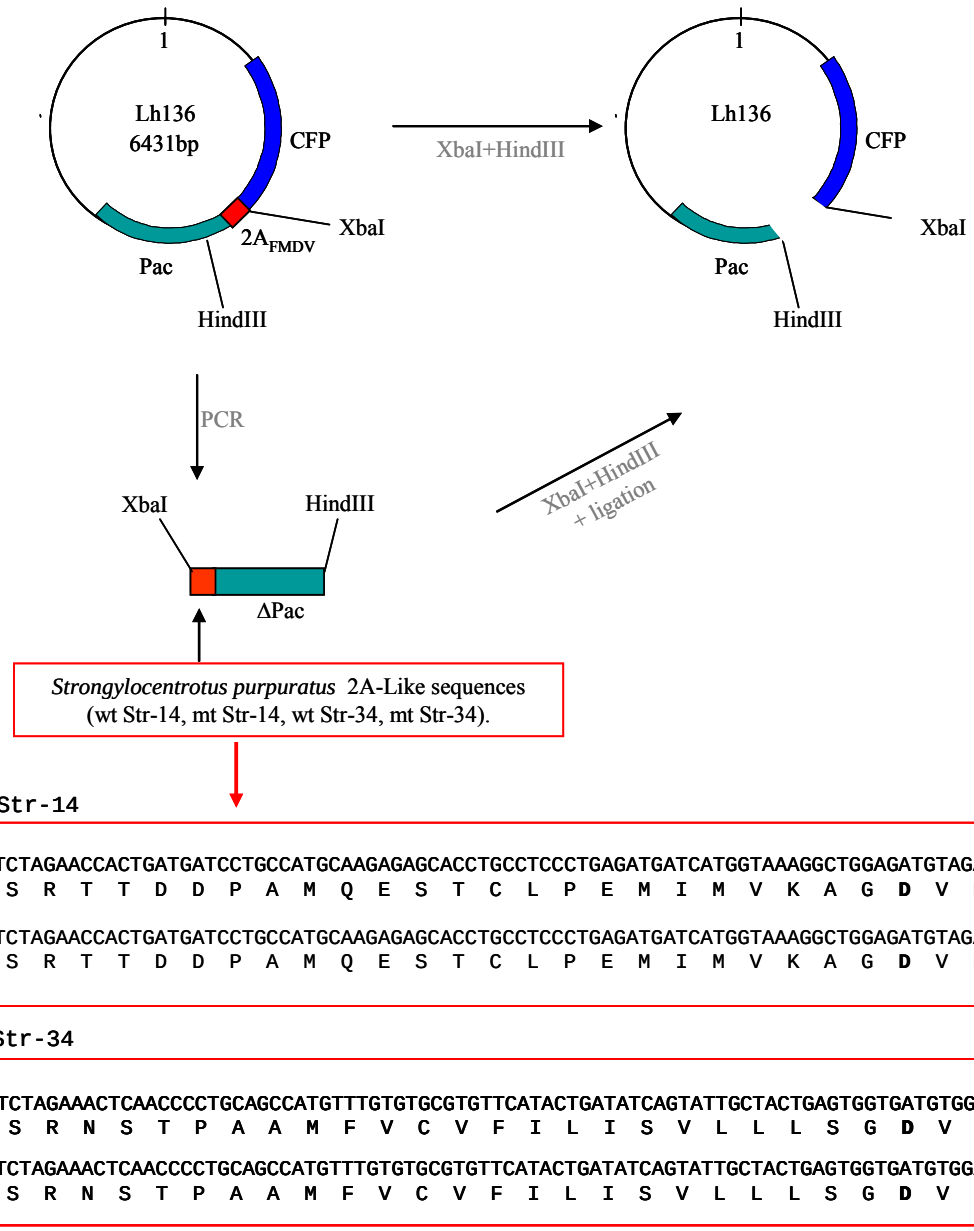


Figure 64. *Strongylocentrotus purpuratus* 2A-like (Str) sequences cloning. Lh136 was cut with XbaI and HindIII, and the big fragment gel purified and used as a vector for the four different PCR inserts, which contain wild type (wt) and mutated (mt) Str-14 and Str-34 2A-like sequences. Wt: C; Mt: G.

In vitro coupled transcription/translation wheat germ extract reactions were performed and showed Str-14 2A-like sequence to be active whereas Str-34 did not show any activity (figure 65). Lack of activity was expected, in the case of Str-34, due to the presence of Serine (S) instead of an asparagine (N) in the C-terminal conserved motif (–DxEx**S**PGP-).

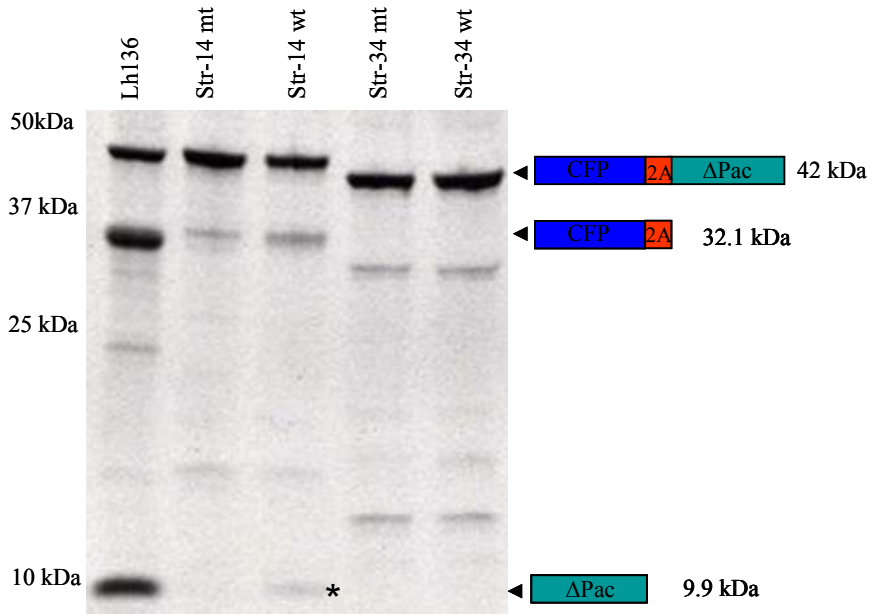


Figure 65. Translation profiles of constructs containing *Strongylocentrotus purpuratus* 2A-like sequences. Coupled transcription/ translation wheat germ extract reactions were programmed with constructs encoding 2A-like sequences in the CFP-2A-ΔPac artificial polyprotein system. Lh136 was used as a positive control, contains the wild type 2AFMDV. The full-length translation products (CFP-2A-ΔPac) and the processing products (CFP-2A and ΔPac) are indicated. The processed products are absent in the Str-34 lanes, indicating inactive 2A-like sequences, although extra bands of ~30 kDa and ~15 kDa were observed.

3.6 Bioinformatic Analyses

All the 2A-like sequences found in the database were classified and analyzed using Excel. This study focused on the upstream sequence rather than the conserved C-terminus motif (-DxExNPGP-).

The aim of this part of the project was to determine whether the 2A upstream sequences follow a pattern among all the different families with 2A-like sequences.

To be sure we include all the upstream amino acids that might be influencing 2A's mechanism of action, the sequences analyzed were 30 amino acids long.

Each position of the sequence was analyzed by counting the number of each amino acid of each position. The amino acids were grouped in different categories depending on the properties of their side-chains. The amino acid classification used was the one shown in figure 66, where the hydrophobicity, charge, and size of the amino acids were taken into account.

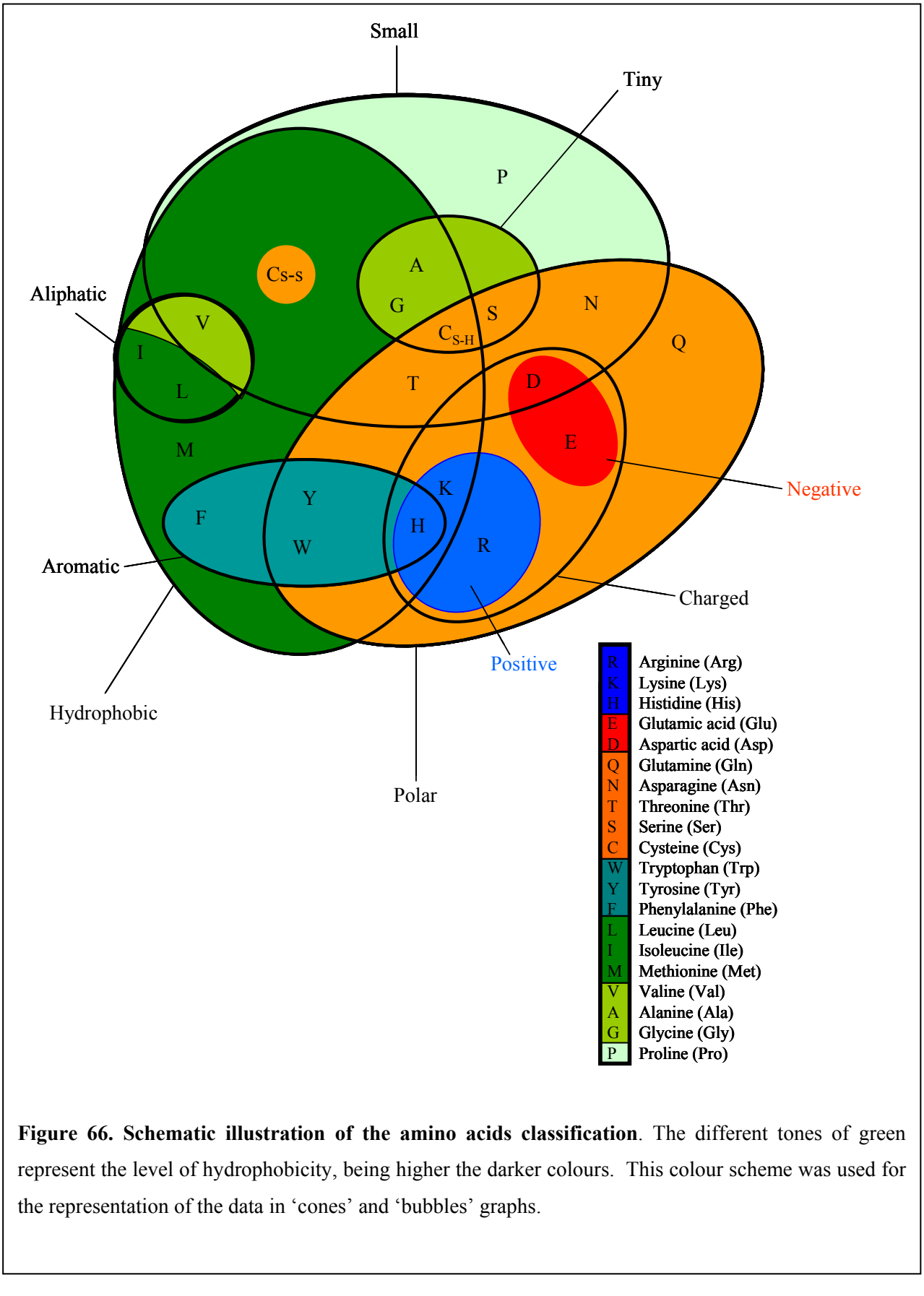


Figure 66. Schematic illustration of the amino acids classification. The different tones of green represent the level of hydrophobicity, being higher the darker colours. This colour scheme was used for the representation of the data in 'cones' and 'bubbles' graphs.

3.6.1 Picornaviruses

This analysis includes the 2A sequences within the *Picornaviridae*. The genera and species analyzed are shown in table 5.

Genera	*Species*	*Number of strains*
Aphthovirus	*Foot-and-mouth disease virus (FMDV)*	282
	Equine rhinitis A virus (ERAV)	7
	Bovine rhinovirus (BRV-2)	1
Erbovirus	*Equine rhinitis B virus (ERBV)*	2
Teschovirus	*Porcine teschovirus (PTV)*	61
Cardiovirus	*Encephalomyocarditis virus (EMCV)*	14
	Theiler's murine encephalomyelitis virus (TMEV)	4
	Theiler's like virus (TLV)	1
	Saffold virus (SAF-V)	2
Parechovirus	*Ljungan virus (LV) 2A*$_1$	5
Unclassified	*Duck Hepatitis Virus-1 (DHV-1)*	24
Seneca virus	*Seneca valley virus (SVV)*	1
Unclassified	*Seal picornavirus-1 (SePV-1)*	8

Table 5. Picornaviruses analyzed. The table shows the genera, species and number of strains of each species analyzed.

The sequences analyzed appear to be active. SAF-V, SVV and DHV-1 2A-like sequence's activity have not yet been analyzed.

The analysis focuses more in the variety of sequences within species and between different species, more than the quantity of them. The number of FMDV strains is bigger than the rest of species; this is not taken into account in this analysis.

The amino acids sequences of the 2A protein analyzed in this chapter are shown in table 6. Not all of them are shown since there are 281 strains of FMDV, thus a maximum of 10 strains per species were selected randomly for representation.

First, the 2A amino acid sequences within picornaviruses were analyzed altogether and then they were analyzed separately in different genera. FMDV was analyzed also in a separate group since the number of strains is quite large. All the group analyses were compared to check the variability of amino acids in each position, and also the possibility of the existence of a pattern in amino acid distribution in all or some of them. Two different graphs were performed to represent the data, a cone graph (see figures 67, 69, 71, 73, 75) and a bubble graph (see figures 68, 70, 72, 74, 76).

Species	Accession number	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
FMDV-63	AY593836	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV-64	AY593831	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV-65	AY593832	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV-75	AY593823	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV-76	AY593824	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV-78	AY593811	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV-79	AY593812	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV-83	AY593828	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV-110	AF506822	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV-128	AJ539138	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
ERAV-393/76	L43052	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	S	L	L	K	L	A	G	D	V	E	S	N	P	G	P
ERAV-Plowright	DQ272127	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	S	L	L	K	L	A	G	D	V	E	S	N	P	G	P
ERAV-U188	DQ272128	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	S	L	L	K	L	A	G	D	V	E	S	N	P	G	P
ERAV-T3	DQ268580	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	S	L	L	K	L	A	G	D	V	E	S	N	P	G	P
ERAV-T10	DQ272577	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	S	L	L	K	L	A	G	D	V	E	S	N	P	G	P
ERAV-PERV	X96870	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	A	L	L	K	L	A	G	D	V	E	S	N	P	G	P
ERAV-PERV-1	DQ272578	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	A	L	L	K	L	A	G	D	V	E	S	N	P	G	P
BRV2	EU236594	L	R	L	T	G	E	I	V	K	Q	G	A	T	N	F	E	L	L	Q	Q	A	G	D	V	E	T	N	P	G	P
ERBV-1-P1436/71	X96871	E	A	T	L	S	T	I	L	S	E	G	A	T	N	F	S	L	L	K	L	A	G	D	V	E	L	N	P	G	P
ERBV-2-P313/75	AF361253	V	A	D	W	E	N	L	L	S	Q	G	A	T	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
PTV-1-5-D-VIII	AF296106	A	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PTV-1-D 61/96	AY392535	A	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PTV-1-PS 34	AF296105	A	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PTV-1-Sek 549/98	AF296101	A	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
EMCV (EMC-PV21)	X74312	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
EMCV (BEL-2887A/91)	AF356822	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
EMCV (pEC9)	DQ288856	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
EMCV (HB1)	DQ464063	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
EMCV (BJC3)	DQ464062	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
EMCV (EMCV-CBNU)	DQ517424	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
EMCV (EMC-B)	M22457	I	F	G	L	Y	R	I	F	S	T	H	Y	A	G	Y	F	S	D	L	L	I	H	D	I	E	T	N	P	G	P
MENGO (Rz-pMwt)	DQ294633	V	F	G	L	Y	H	V	F	E	T	H	Y	A	G	Y	F	S	D	L	L	I	H	D	V	E	T	N	P	G	P
TMEV (GDVII)	X56019	F	R	E	F	F	K	A	V	R	G	Y	H	A	D	Y	Y	K	Q	R	L	I	H	D	V	E	M	N	P	G	P
TMEV	M20562	F	R	E	F	F	K	A	V	R	G	Y	H	A	D	Y	Y	K	Q	R	L	I	H	D	V	E	M	N	P	G	P
TMEV (DA TO)	M20301	F	G	E	F	F	R	A	V	R	A	Y	H	A	D	Y	Y	K	Q	R	L	I	H	D	V	E	M	N	P	G	P
TMEV (BeAn 8386)	M16020	F	G	E	F	F	K	A	V	R	G	Y	H	A	D	Y	Y	R	Q	R	L	I	H	D	V	E	T	N	P	G	P
TLV(NGS910)	AB090161	F	S	D	F	F	K	H	V	R	E	Y	H	A	A	Y	Y	K	Q	R	L	M	H	D	V	E	T	N	P	G	P
SAF-V	EF165067	F	T	D	F	F	K	A	V	R	D	Y	H	A	S	Y	Y	K	Q	R	L	Q	H	D	V	E	T	N	P	G	P
SAF-V	AM922293	F	T	D	F	F	K	A	V	R	D	Y	H	A	S	Y	Y	K	Q	R	L	Q	H	D	I	E	A	N	P	G	P
LV (174F)	AF327921	Y	F	N	I	M	H	S	D	E	M	D	F	A	G	G	K	F	L	N	Q	C	G	D	V	E	T	N	P	G	P
LV (87-012)	AF327920	Y	F	N	I	M	H	S	D	E	M	D	F	A	G	G	K	F	L	N	Q	C	G	D	V	E	T	N	P	G	P
LV (87-012G)	EF202833	Y	F	N	I	M	H	S	D	E	M	D	F	A	G	G	K	F	L	N	Q	C	G	D	V	E	T	N	P	G	P
DHV-1 (DRL-62)	DQ219396	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
DHV-1 (R85952)	DQ226541	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
DHV-1 (F)	EU264072	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
SVV	DQ641257	R	A	W	C	P	S	M	L	P	F	R	S	Y	K	Q	K	M	L	M	Q	S	G	D	I	E	T	N	P	G	P
SePV-1	EU152976	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
SePV-1	EU142040	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
SePV-1	EU152979	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P

Table 6. 2A amino acid sequences within picornaviruses. Representation of the species and the different strains analyzed. The conserved motif is highlighted in yellow, and the differences in this region are shown in red.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	72	3	0	17	0	10	2	0	7	5	1	0	0	24	0	0	8	0	10	1	0	0	0	0	0	0	0	0	3	0
K	1	**269**	9	**185**	0	6	9	1	3	**284**	0	0	0	1	1	23	23	0	**344**	0	0	0	0	0	0	0	0	0	0	0
H	**188**	10	1	0	7	9	1	0	0	0	14	7	0	0	0	8	0	0	0	0	0	21	0	0	0	0	0	0	0	0
E	2	0	28	39	1	1	24	0	83	3	0	0	0	0	0	3	0	0	1	0	1	24	0	0	**412**	61	0	0	0	0
D	6	0	9	0	9	0	1	4	3	2	29	1	0	4	0	**276**	0	38	0	0	7	0	**388**	0	0	0	0	0	0	0
Q	0	5	**257**	5	0	0	0	61	0	2	**289**	24	0	0	1	0	0	7	4	67	2	0	0	0	0	0	0	1	0	0
N	1	0	6	0	14	1	8	0	28	0	0	0	0	**349**	23	3	8	0	5	0	0	0	0	0	0	0	**412**	0	0	0
T	8	2	75	41	0	26	1	4	2	2	0	23	67	0	7	0	0	0	0	24	24	0	0	0	0	23	0	0	0	0
S	6	1	0	4	1	35	3	0	6	24	0	3	7	6	0	67	2	0	0	0	2	0	0	0	0	**295**	0	0	0	0
C	2	0	8	1	8	3	3	0	0	0	0	7	13	0	1	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0
W	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	32	0	0	0	15	0	7	0	0	0	15	16	1	0	28	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	7	48	0	22	7	0	33	14	0	1	0	3	6	0	**343**	14	5	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	2	1	39	51	25	30	3	1	0	0	**217**	**256**	0	2	0	**353**	**359**	46	**320**	0	0	0	0	0	3	0	10	0	0
I	1	0	1	5	**235**	131	15	31	0	0	0	0	19	0	1	0	0	0	0	0	18	0	0	22	0	0	0	0	0	0
M	0	59	0	0	48	7	1	0	0	5	0	10	8	0	0	0	1	0	2	0	1	0	0	1	0	3	0	0	0	0
V	16	0	1	13	2	**142**	1	13	98	8	0	32	0	0	0	7	0	8	0	0	0	0	0	**389**	0	0	0	0	0	0
A	69	3	1	13	0	6	**268**	0	**114**	13	0	69	27	1	0	3	12	0	0	0	**352**	0	0	0	0	1	0	0	0	11
G	0	10	14	1	1	2	5	0	66	3	64	0	0	19	5	1	0	0	0	0	0	**367**	24	0	0	0	0	0	**409**	0
P	0	0	0	26	13	8	0	**281**	1	60	0	0	8	8	0	0	0	0	0	0	0	0	0	0	0	26	0	**401**	0	**401**
	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412

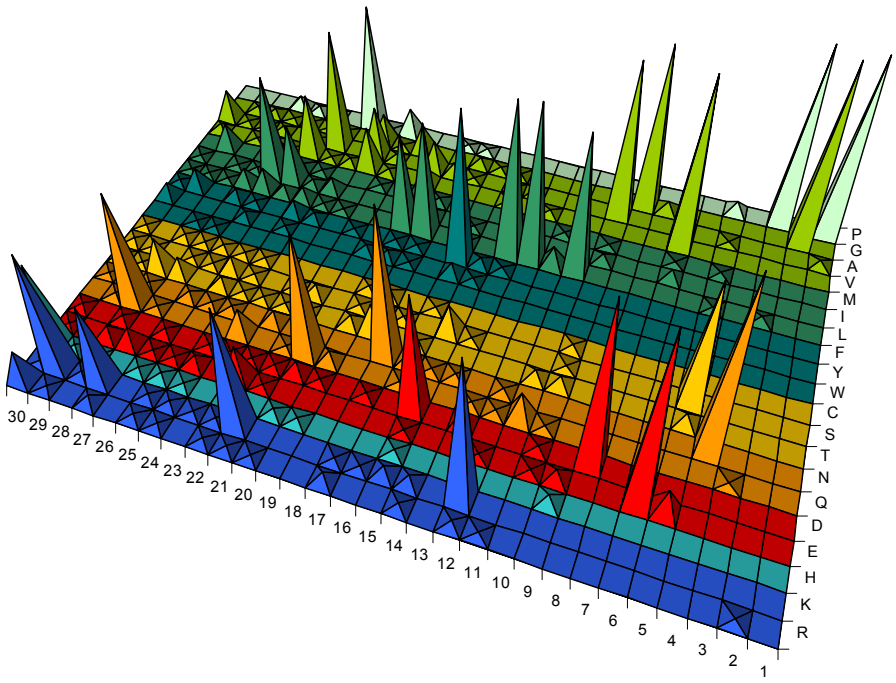


Figure 67. **Analysis of all the 2A-like sequences within the *Picornaviridae* family**. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Cone graph, below, representing the results from the table. The data set analyzed is shown in Appendix 1, sequences 1 to 412.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	72	3	0	17	0	10	2	0	7	5	1	0	0	24	0	0	8	0	10	1	0	0	0	0	0	0	0	0	3	0
K	1	**269**	9	**185**	0	6	9	1	3	**284**	0	0	0	1	1	23	23	0	**344**	0	0	0	0	0	0	0	0	0	0	0
H	**188**	10	1	0	7	9	1	0	0	0	14	7	0	0	0	8	0	0	0	0	0	21	0	0	0	0	0	0	0	0
E	2	0	28	39	1	1	24	0	83	3	0	0	0	0	0	3	0	0	1	0	1	24	0	0	**412**	61	0	0	0	0
D	6	0	9	0	9	0	1	4	3	2	29	1	0	4	0	**276**	0	38	0	0	7	0	**388**	0	0	0	0	0	0	0
Q	0	5	**257**	5	0	0	0	61	0	2	**289**	24	0	0	1	0	0	7	4	67	2	0	0	0	0	0	0	1	0	0
N	1	0	6	0	14	1	8	0	28	0	0	0	0	**349**	23	3	8	0	5	0	0	0	0	0	0	0	**412**	0	0	0
T	8	2	75	41	0	26	1	4	2	2	0	23	67	0	7	0	0	0	0	24	24	0	0	0	0	23	0	0	0	0
S	6	1	0	4	1	35	3	0	6	24	0	3	7	6	0	67	2	0	0	0	2	0	0	0	0	**295**	0	0	0	0
C	2	0	8	1	8	3	3	0	0	0	0	7	13	0	1	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0
W	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	32	0	0	0	15	0	7	0	0	0	15	16	1	0	28	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	7	48	0	22	7	0	33	14	0	1	0	3	6	0	**343**	14	5	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	2	1	39	51	25	30	3	1	0	0	**217**	**256**	0	2	0	**353**	**359**	46	**320**	0	0	0	0	0	3	0	10	0	0
I	1	0	1	5	**235**	**131**	15	31	0	0	0	0	19	0	1	0	0	0	0	0	18	0	0	22	0	0	0	0	0	0
M	0	59	0	0	48	7	1	0	0	5	0	10	8	0	0	0	1	0	2	0	1	0	0	1	0	3	0	0	0	0
V	16	0	1	13	2	**142**	1	13	98	8	0	32	0	0	0	7	0	8	0	0	0	0	0	**389**	0	0	0	0	0	0
A	69	3	1	13	0	6	**268**	0	**114**	13	0	69	27	1	0	3	12	0	0	0	**352**	0	0	0	0	1	0	0	0	11
G	0	10	14	1	1	2	5	0	66	3	64	0	0	19	5	1	0	0	0	0	0	**367**	24	0	0	0	0	0	**409**	0
P	0	0	0	26	13	8	0	**281**	1	60	0	0	8	8	0	0	0	0	0	0	0	0	0	0	0	26	0	**401**	0	**401**
	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412	412

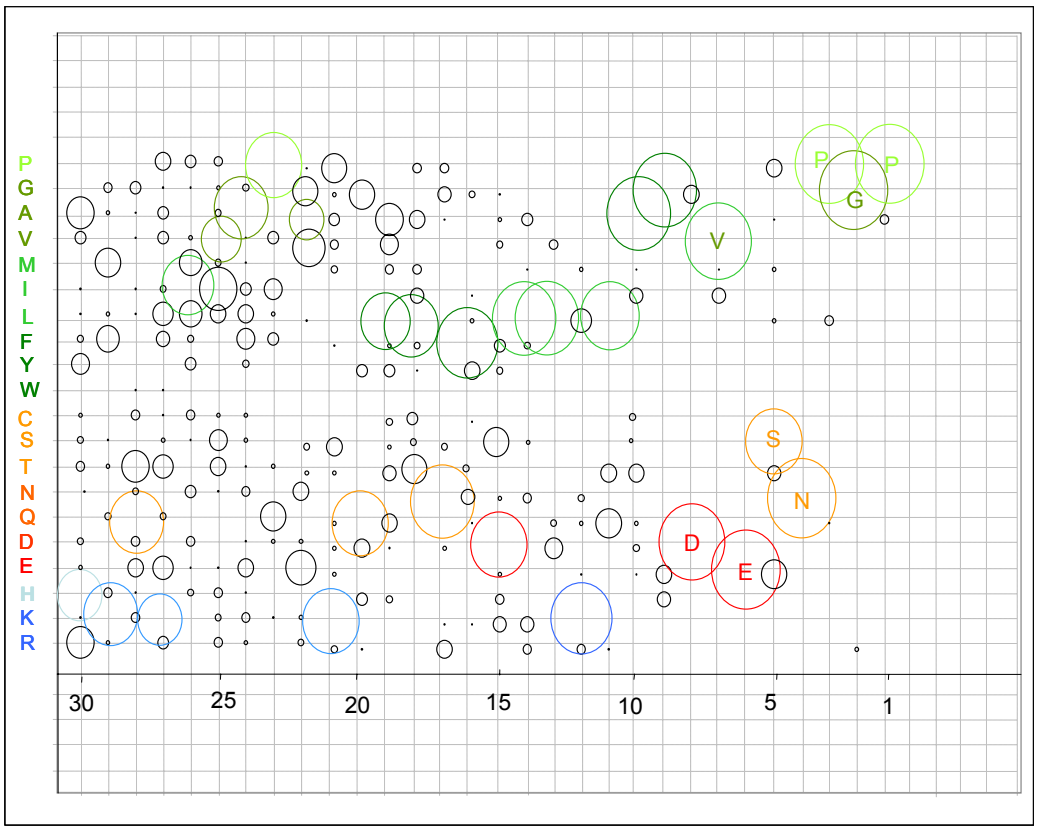


Figure 68. Bubble graph representing the data for all picornavirus 2A-like sequences. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Below, bubble graph; the bubbles represent the amino acids in each position. The size of the bubbles is related with the amount of amino acids in each position. The predominant amino acids are shown in colour. The data set analyzed is shown in Appendix 1, sequences 1 to 412

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	71	1	0	11	0	0	2	0	0	4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0
K	1	**269**	8	**185**	0	0	9	0	3	**284**	0	0	0	0	0	0	0	0	**283**	0	0	0	0	0	0	0	0	0	0	0
H	**188**	10	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	39	0	1	0	0	54	1	0	0	0	0	0	3	0	0	1	0	0	0	0	0	**290**	0	0	0	0	0
D	0	0	5	0	6	0	0	0	2	0	0	1	0	0	0	**275**	0	0	0	0	0	0	**290**	0	0	0	0	0	0	0
Q	0	5	**257**	5	0	0	0	0	0	1	**289**	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	1	0	0
N	0	0	2	0	0	0	7	0	8	0	0	0	0	**287**	0	3	0	0	0	0	0	0	0	0	0	0	**290**	0	0	0
T	2	0	14	2	0	10	1	4	2	0	0	23	8	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
S	0	0	0	4	0	0	0	0	3	0	0	2	3	3	0	5	0	0	0	0	1	0	0	0	0	**284**	0	0	0	0
C	0	0	0	0	0	3	3	0	0	0	0	7	13	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	27	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	5	0	7	0	0	0	0	0	0	0	0	1	0	**280**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	0	1	0	37	0	0	0	1	0	0	**217**	**256**	0	2	0	**290**	**290**	0	**289**	0	0	0	0	0	2	0	10	0	0
I	0	0	1	0	**234**	131	1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	10	8	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
V	0	0	1	0	0	**142**	0	6	98	0	0	24	0	0	0	0	0	0	0	0	0	0	0	**289**	0	0	0	0	0	0
A	0	0	0	10	0	0	**262**	0	**114**	0	0	6	1	0	0	3	0	0	0	0	**289**	0	0	0	0	0	0	0	0	11
G	0	0	0	1	1	0	5	0	5	0	1	0	0	0	0	1	0	0	0	0	0	**290**	0	0	0	0	0	0	**288**	0
P	0	0	0	26	12	0	0	**273**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	**279**	0	**279**
	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	**290**

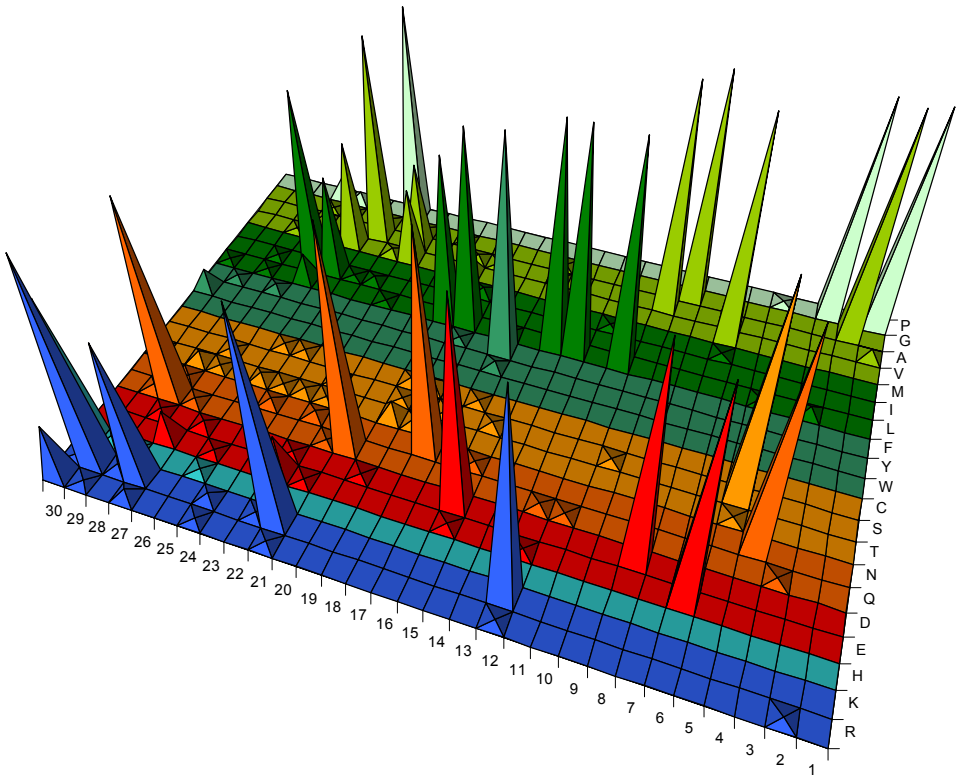


Figure 69. **Analysis of all the 2A-like sequences within aphthoviruses**. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Cone graph, below, representing the results from the table. The data set analyzed is shown in Appendix 1, sequences 1 to 290.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	71	1	0	11	0	0	2	0	0	4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0
K	1	**269**	8	**185**	0	0	9	0	3	**284**	0	0	0	0	0	0	0	0	**283**	0	0	0	0	0	0	0	0	0	0	0
H	**188**	10	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	39	0	1	0	0	54	1	0	0	0	0	0	3	0	0	1	0	0	0	0	0	**290**	0	0	0	0	0
D	0	0	5	0	6	0	0	0	2	0	0	1	0	0	0	**275**	0	0	0	0	0	0	**290**	0	0	0	0	0	0	0
Q	0	5	**257**	5	0	0	0	0	0	1	**289**	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	1	0	0
N	0	0	2	0	0	0	7	0	8	0	0	0	0	**287**	0	3	0	0	0	0	0	0	0	0	0	0	**290**	0	0	0
T	2	0	14	2	0	10	1	4	2	0	0	23	8	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
S	0	0	0	4	0	0	0	3	0	0	2	3	3	0	5	0	0	0	0	1	0	0	0	0	**284**	0	0	0	0	0
C	0	0	0	0	0	3	3	0	0	0	0	7	13	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	27	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	5	0	7	0	0	0	0	0	0	0	0	1	0	**280**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	0	1	0	37	0	0	0	1	0	0	**217**	**256**	0	2	0	**290**	**290**	0	**289**	0	0	0	0	0	2	0	10	0	0
I	0	0	1	0	**234**	131	1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	10	8	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
V	0	0	1	0	0	**142**	0	6	98	0	0	24	0	0	0	0	0	0	0	0	0	0	0	**289**	0	0	0	0	0	0
A	0	0	0	10	0	0	**262**	0	**114**	0	0	6	1	0	0	3	0	0	0	0	**289**	0	0	0	0	0	0	0	0	11
G	0	0	0	1	1	0	5	0	5	0	1	0	0	0	0	1	0	0	0	0	0	**290**	0	0	0	0	0	0	**288**	0
P	0	0	0	26	12	0	0	**273**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	**279**	0	**279**
	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290	290

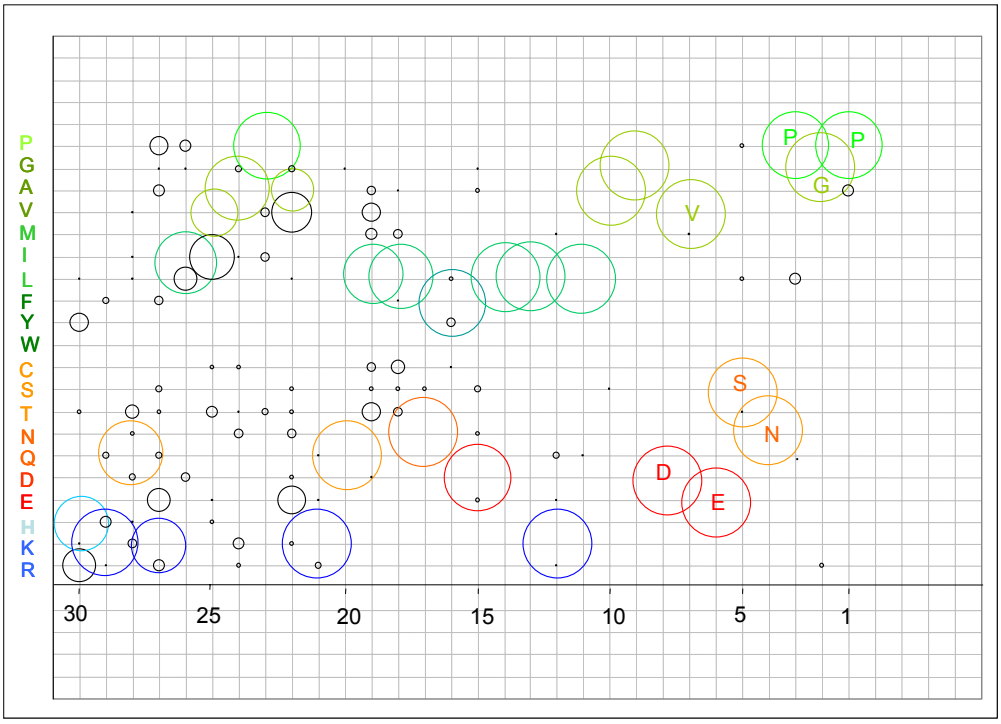


Figure 70. Bubble graph representing the data for aphthovirus 2A-like sequences. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Below, bubble graph; the bubbles represent the amino acids in each position. The size of the bubbles is related with the amount of amino acids in each position. The predominant amino acids are shown in colour. The data set analyzed is shown in Appendix 1, sequences 1 to 290.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	64	0	0	11	0	0	2	0	0	4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0
K	1	**269**	1	**185**	0	0	9	0	2	**277**	0	0	0	0	0	0	0	0	**276**	0	0	0	0	0	0	0	0	0	0	0
H	**188**	3	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	39	0	0	0	0	54	1	0	0	0	0	0	2	0	0	1	0	0	0	0	0	**282**	0	0	0	0	0
D	0	0	5	0	6	0	0	0	2	0	0	1	0	0	0	**275**	0	0	0	0	0	0	**282**	0	0	0	0	0	0	0
Q	0	5	**257**	5	0	0	0	0	0	0	**282**	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	1	0	0
N	0	0	2	0	0	0	0	0	1	0	0	0	0	**279**	0	3	0	0	0	0	0	0	0	0	0	0	**282**	0	0	0
T	2	0	14	1	0	3	1	4	2	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	4	0	0	0	0	3	0	0	2	3	3	0	0	0	0	0	0	1	0	0	0	0	**277**	0	0	0	0
C	0	0	0	0	0	3	3	0	0	0	0	0	13	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	5	0	0	0	0	0	0	0	0	0	0	1	0	**279**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	37	0	0	0	1	0	0	**217**	**256**	0	2	0	**282**	**282**	0	**282**	0	0	0	0	0	2	0	10	0	0
I	0	0	1	0	**234**	131	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	10	8	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
V	0	0	1	0	0	**142**	0	5	98	0	0	24	0	0	0	0	0	0	0	0	0	0	0	**281**	0	0	0	0	0	0
A	0	0	0	10	0	0	**262**	0	**114**	0	0	5	1	0	0	1	0	0	0	0	**281**	0	0	0	0	0	0	0	0	11
G	0	0	0	1	0	0	5	0	5	0	0	0	0	0	0	1	0	0	0	0	0	**282**	0	0	0	0	0	0	**280**	0
P	0	0	0	26	5	0	0	**273**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	**271**	0	**271**
	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282

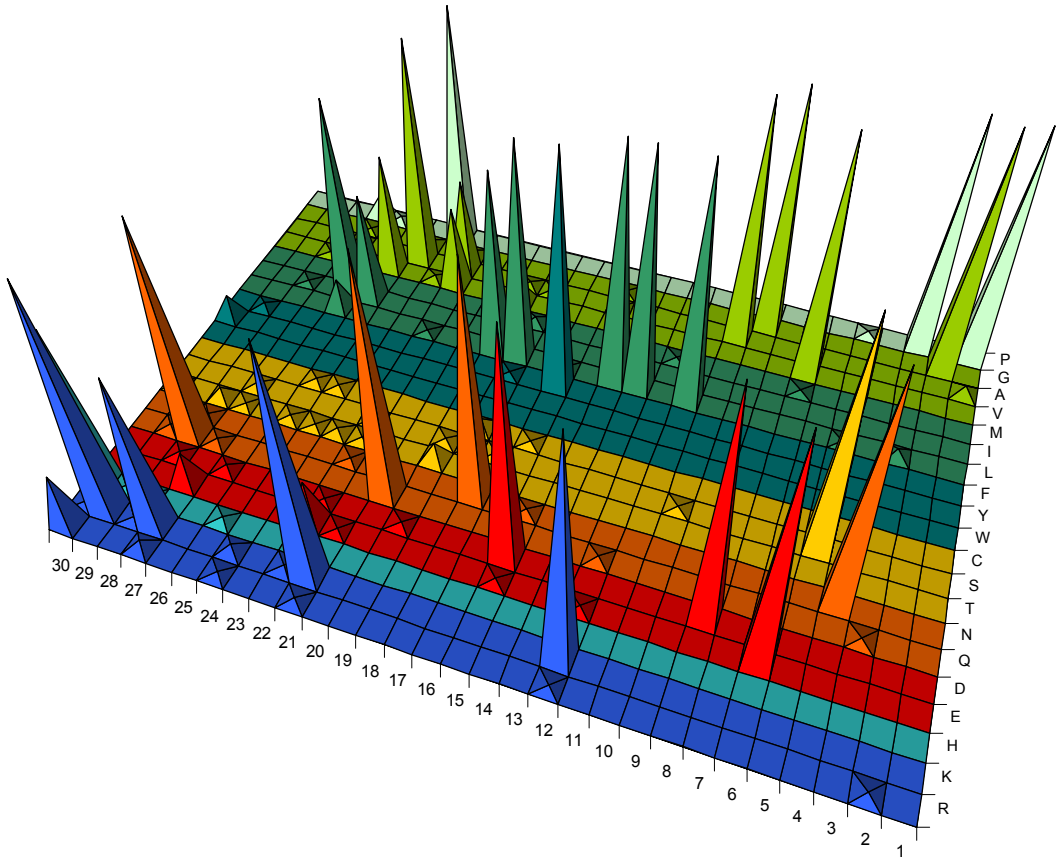


Figure 71. **Analysis of the 2A-sequences within FMDV strains**. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Cone graph, below, representing the results from the table. The data set analyzed is shown in Appendix 1, sequences 1 to 282.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	64	0	0	11	0	0	2	0	0	4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0
K	1	**269**	1	**185**	0	0	9	0	2	**277**	0	0	0	0	0	0	0	0	**276**	0	0	0	0	0	0	0	0	0	0	0
H	**188**	3	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	39	0	0	0	0	54	1	0	0	0	0	0	2	0	0	1	0	0	0	0	0	**282**	0	0	0	0	0
D	0	0	5	0	6	0	0	0	2	0	0	1	0	0	0	**275**	0	0	0	0	0	0	**282**	0	0	0	0	0	0	0
Q	0	5	**257**	5	0	0	0	0	0	0	**282**	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	1	0	0
N	0	0	2	0	0	0	0	0	1	0	0	0	0	**279**	0	3	0	0	0	0	0	0	0	0	0	0	**282**	0	0	0
T	2	0	14	1	0	3	1	4	2	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	4	0	0	0	0	3	0	0	2	3	3	0	0	0	0	0	0	1	0	0	0	0	**277**	0	0	0	0
C	0	0	0	0	0	3	3	0	0	0	0	0	13	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	5	0	0	0	0	0	0	0	0	0	0	1	0	**279**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	37	0	0	0	1	0	0	**217**	**256**	0	2	0	**282**	**282**	0	**282**	0	0	0	0	0	2	0	10	0	0
I	0	0	1	0	**234**	131	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	10	8	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
V	0	0	1	0	0	**142**	0	5	98	0	0	24	0	0	0	0	0	0	0	0	0	0	0	**281**	0	0	0	0	0	0
A	0	0	0	10	0	0	**262**	0	**114**	0	0	5	1	0	0	1	0	0	0	0	**281**	0	0	0	0	0	0	0	0	11
G	0	0	0	1	0	0	5	0	5	0	0	0	0	0	0	1	0	0	0	0	0	**282**	0	0	0	0	0	0	**280**	0
P	0	0	0	26	5	0	0	**273**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	**271**	0	**271**
	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282	282

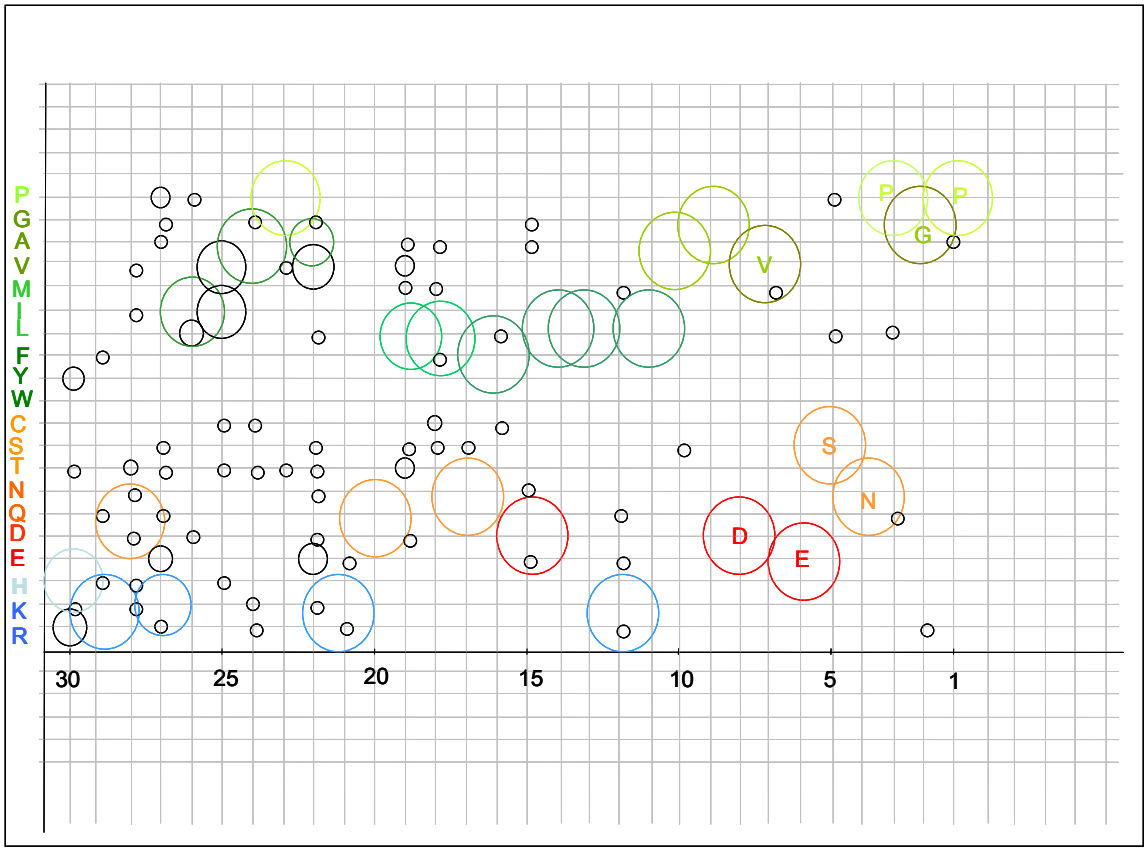


Figure 72. Bubble graph representing the data for FMDV 2A sequences. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Below, bubble graph; the bubbles represent the amino acids in each position. The size of the bubbles is related with the amount of amino acids in each position. The predominant amino acids are shown in colour. The data set analyzed is shown in Appendix 1, sequences 1 to 282.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	0	2	0	0	0	**10**	0	0	7	0	0	0	0	0	0	0	1	0	7	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	1	1	0	0	0	**14**	7	0	0	0	0	0	0	0	0	0	**21**	0	0	0	0	0	0	0	0
E	0	0	4	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0	0	0	0	0	0
D	0	0	3	0	0	0	0	0	0	2	0	0	0	4	0	0	0	**14**	0	0	0	0	**21**	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	2	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0	0	0
T	0	2	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**17**	0	0	0	0
S	0	1	0	0	0	2	0	0	1	0	0	0	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	**14**	0	0	0	0	0	7	**14**	0	0	**21**	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	7	**14**	0	7	7	0	0	**14**	0	0	0	0	0	0	0	**14**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	**14**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**14**	**21**	0	0	0	0	0	0	0	0	0	0
I	1	0	0	0	0	0	**13**	0	0	0	0	0	0	0	0	0	0	0	0	0	**18**	0	0	**14**	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	3	0	0	0	0
V	**13**	0	0	0	0	0	1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0
A	0	0	0	0	0	0	6	0	0	**13**	0	0	**21**	1	0	0	**12**	0	0	0	0	0	0	0	1	0	0	0	0	0
G	0	2	**14**	0	0	2	0	0	0	3	0	0	0	**14**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0	**21**
	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21

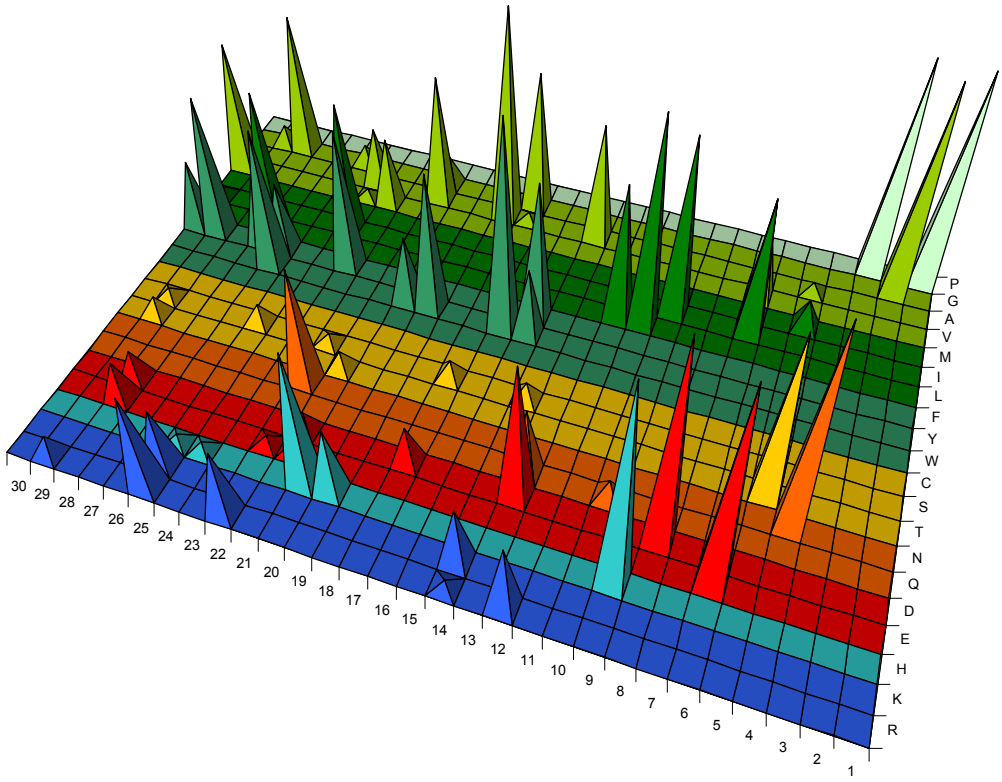


Figure 73. **Analysis of the 2A-sequences within cardioviruses**. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Cone graph, below, representing the results from the table. The data set analyzed is shown in Appendix 1, sequences 354 to 374.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	0	2	0	0	0	**10**	0	0	7	0	0	0	0	0	0	0	1	0	7	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	1	1	0	0	0	**14**	7	0	0	0	0	0	0	0	0	0	**21**	0	0	0	0	0	0	0	0
E	0	0	4	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0	0	0	0	0
D	0	0	3	0	0	0	0	0	0	2	0	0	0	4	0	0	0	**14**	0	0	0	0	**21**	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	2	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0	0	0
T	0	2	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**17**	0	0	0	0
S	0	1	0	0	0	2	0	0	1	0	0	0	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	**14**	0	0	0	0	0	7	**14**	0	0	**21**	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	7	**14**	0	7	7	0	0	**14**	0	0	0	0	0	0	0	**14**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	**14**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**14**	**21**	0	0	0	0	0	0	0	0	0	0
I	1	0	0	0	0	0	**13**	0	0	0	0	0	0	0	0	0	0	0	0	0	**18**	0	0	**14**	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	3	0	0	0	0
V	**13**	0	0	0	0	0	1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0
A	0	0	0	0	0	0	6	0	0	**13**	0	0	**21**	1	0	0	**12**	0	0	0	0	0	0	0	0	1	0	0	0	0
G	0	2	**14**	0	0	2	0	0	0	3	0	0	0	**14**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0	**21**
	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21

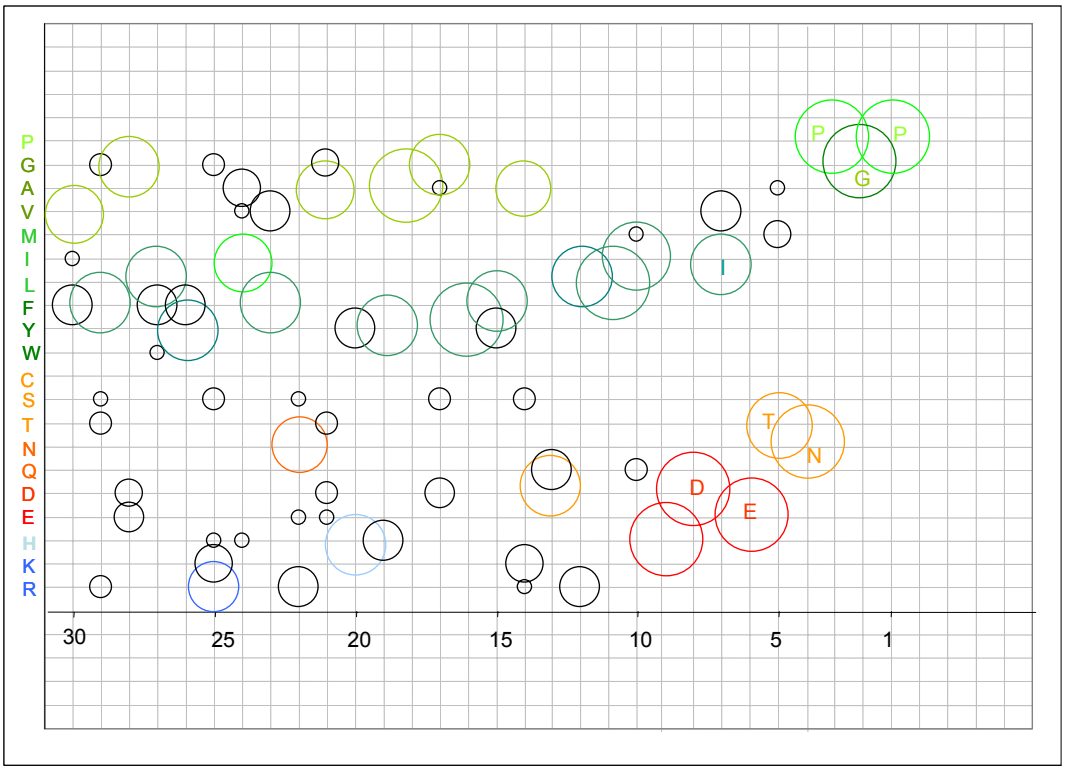


Figure 74. Bubble graph representing the data for cardiovirus 2A-like sequences. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Below, bubble graph; the bubbles represent the amino acids in each position. The size of the bubbles is related with the amount of amino acids in each position. The predominant amino acids are shown in colour. The data set analyzed is shown in Appendix 1, sequences 354 to 374.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	0	0	0	6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**59**	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**61**	**61**	0	0	0	0
D	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**61**	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	**61**	0	0	0	0	0	0	0	0	0	0	0	**60**	0	0	0	0	0	0	0	0	0	0
N	1	0	0	0	0	0	0	0	0	0	0	0	0	**60**	0	0	0	0	0	0	0	0	0	0	0	0	**61**	0	0	0
T	6	0	**60**	**39**	0	15	0	0	0	0	0	0	**57**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	**32**	0	0	0	0	0	0	3	1	0	**61**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	**33**	0	0	0	0	0	0	0	**61**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	2	0	0	14	1	21	0	0	0	0	0	0	0	0	0	**61**	**61**	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0
M	0	59	0	0	**44**	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V	2	0	0	13	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**54**	0	0	0	0	0	0
A	**45**	0	1	3	0	6	0	0	0	0	0	**61**	1	0	0	0	0	0	0	0	**61**	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	**61**	0	**61**	0	0	0	0	0	0	0	0	0	0	**61**	0	0	0	0	0	0	**61**	0
P	0	0	0	0	0	0	0	0	0	**60**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**61**	0	**61**
	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61

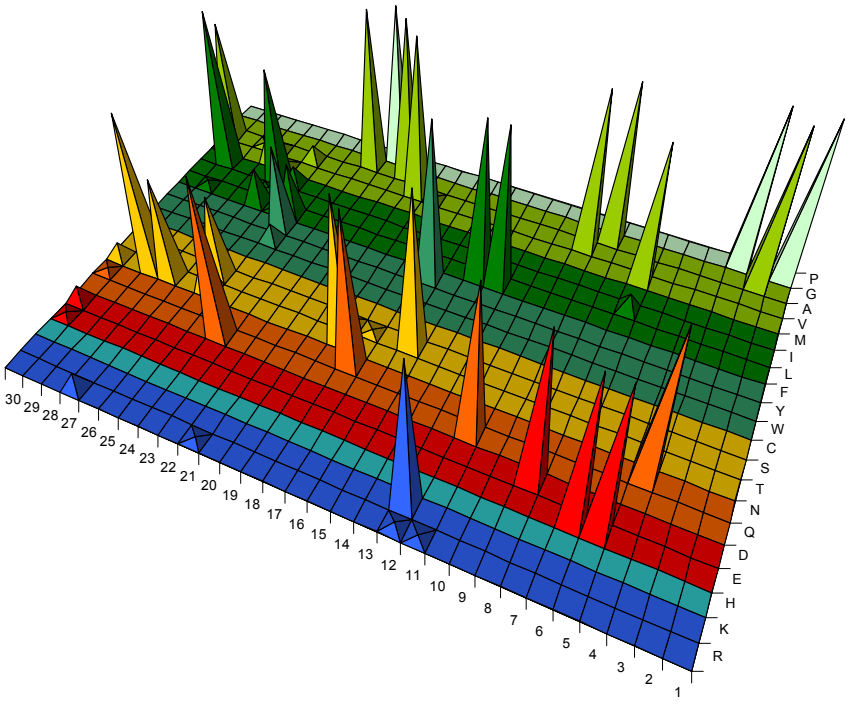


Figure 75. **Analysis of all the 2A-like sequences within teschoviruses**. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Cone graph, below, representing the results from the table. The data set analyzed is shown in Appendix 1, sequences 293 to 353.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	0	0	0	6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**59**	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**61**	**61**	0	0	0	0
D	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**61**	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	**61**	0	0	0	0	0	0	0	0	0	0	0	**60**	0	0	0	0	0	0	0	0	0	0
N	1	0	0	0	0	0	0	0	0	0	0	0	0	**60**	0	0	0	0	0	0	0	0	0	0	0	0	**61**	0	0	0
T	6	0	**60**	**39**	0	15	0	0	0	0	0	0	**57**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	**32**	0	0	0	0	0	0	3	1	0	**61**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	**33**	0	0	0	0	0	0	0	**61**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	2	0	0	14	1	21	0	0	0	0	0	0	0	0	0	**61**	**61**	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0
M	0	59	0	0	**44**	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V	2	0	0	13	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**54**	0	0	0	0	0	0
A	**45**	0	1	3	0	6	0	0	0	0	0	**61**	1	0	0	0	0	0	0	0	**61**	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	**61**	0	**61**	0	0	0	0	0	0	0	0	0	0	**61**	0	0	0	0	0	0	**61**	0
P	0	0	0	0	0	0	0	0	0	**60**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**61**	0	**61**
	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61	61

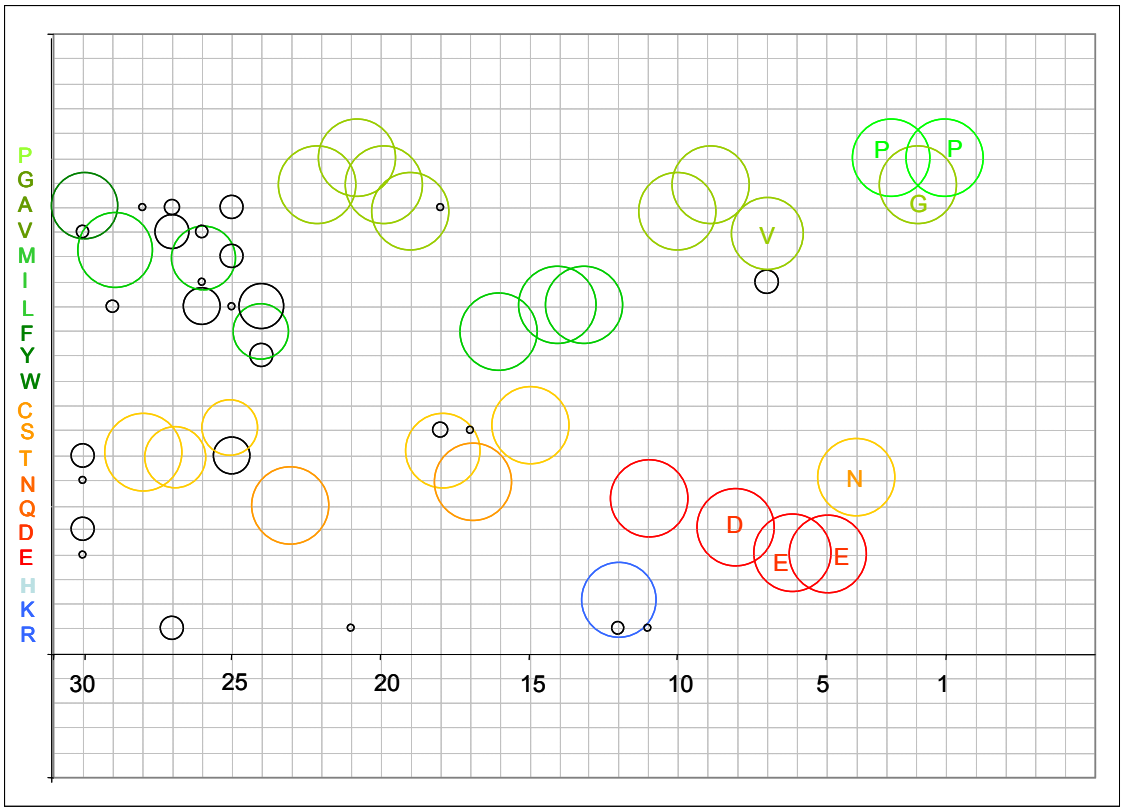


Figure 76. Bubble graph representing the data for teschovirus 2A-like sequences. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Below, bubble graph; the bubbles represent the amino acids in each position. The size of the bubbles is related with the amount of amino acids in each position. The predominant amino acids are shown in colour. The data set analyzed is shown in Appendix 1, sequences 293 to 353.

There is a clear pattern in all picornaviruses 2A-like sequences (figures 67, 68). There is conservation from position 1 to 15 and it is notable that tryptophan is never observed in this region. This could be due to its bulky aromatic side chain, although phenylalanine is observed in the majority of sequences at position 16, although. tyrosine and asparagine are also observed in this position.

Greater heterogeneity is observed from position 15 to 30: charged amino acids are predominant in the middle of the sequence whilst hydrophobic residues being more common at the N-terminus. Positions 27, 29 and 30 mostly are occupied by the hydrophilic residues, histidine and lysine. There is a large predominance of leucine residues in positions 11, 13, 14, 15, 18 and 19, whereas position 12 is a lysine in most sequences.

Aphthovirus 2A sequences are the most conserved (figures 69, 70), together with teschovirus 2A-like sequences (figures 75, 76). This is probably due to the presence of $3C^{pro}$ cleavage site. Position 19 and 20 in FMDV comprise a consensus sequence, which is the cleavage site for $3C^{pro}$ or more efficiently $3CD^{pro}$, to separate 2A from 1D (Ryan *et al*., 1989). This conserved amino acid pair is mostly glutamine and leucine at position 19 and 20, respectively (figures 71, 72). Conservation at these positions is also observed in teschoviruses (figures 75, 76), where the cleavage site is completely conserved at a glutamine/glycine (Q/G) pair (see section 4.5 for discussion).

In contrast, cardiovirus 2A-like sequences are more heterogeneous. These sequences do not present any consensus motif at position 19, 20 since they are longer (~150aa) than tescho- and aphthoviruses (~19aa) and the C-terminal portion of cardiovirus 2A is not proteolytically cleaved from the remainder of 2A.

Tryptophan seems to be avoided in this part of the sequence too. It is only present in two cases: SVV (position 28) and ERBV-2 (position 27).

It is also remarkable the periodicity of aliphatic residues along 2A's upstream sequence, mostly leucine and valine, which is consistent with the suggestions about 2A's structural model being in a helical conformation (Donnelly *et al.*, 1997; Hahn & Palmenberg, 2001).

3.6.2 Mammalian 2A-like sequences

Analysis on the mammalian 2A-like sequences shows a similar pattern to that observed in picornaviruses although with a higher range of heterogeneity. Neutral non-polar residues are predominant along the sequence, some of them being bulky aromatic residues, which are not found in picornaviruses (except phenylalanine at position 16). Phenylalanine is found in positions 18 and 27, whereas position 20 has, mostly, tyrosine. Moreover, tryptophan is also observed in a minority of sequences.

The C-terminal motif (-DxExNPGP-) is highly conserved except in ADRV-N where there is a point mutation at the $D^{8}\rightarrow C$ (table 8).

The mammalian 2A-like amino acid sequences analyzed are shown in table 8, and the analyses, in figures 77 and 78.

Family	*Genera*	*Species*	*Number of strains*
Picornaviridae	*Seneca virus*	*Seneca valley virus (SVV)*	1
	Unclassified	*Seal picornavirus-1 (SePV-1)*	8
Reoviridae	*Cypovirus*	*New adult diarrhoea virus* (ADRV-N)	3
		Bovine rotavirus C (BoRV)	1
		Porcine rotavirus C (PoRV)	1
		Human rotavirus C (HuRV).	4

Table 7. Classification of mammalian 2A-like sequences. The table shows the families, genera, species and number of strains analysed.

Species	Accession number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
SVV	DQ641257	R	A	W	C	P	S	M	L	P	F	R	S	Y	K	Q	K	M	L	M	Q	S	G	D	I	E	T	N	P	G
SePV-1	EU152976	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G
SePV-1	EU142040	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G
SePV-1	EU152979	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G
SePV-1	EU152978	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G
SePV-1	EU152975	C	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G
SePV-1	EU152974	C	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	E	G	D	V	E	S	N	P	G
SePV-1	EU152980	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	I	H	N	V	L	L	D	G	D	V	E	S	N	P	G
SePV-1	EU152977	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	R
ADRV-N	AY632079	F	F	D	S	V	W	V	Y	H	L	A	N	S	S	W	V	R	D	L	T	R	E	C	I	E	S	N	P	G
ADRV-N (J19)	DQ113901	F	F	D	S	V	W	V	Y	H	L	A	N	S	S	W	V	R	D	L	T	R	E	C	I	E	S	N	P	G
ADRV-N (B219)	DQ168032	F	F	D	S	I	W	V	Y	H	L	A	N	S	S	W	V	R	D	L	T	R	E	C	I	E	S	N	P	G
Human-C (V508)	AY941781	G	V	G	Y	P	L	I	V	A	N	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G
Human-C (V966)	AY941782	G	V	G	Y	P	L	I	V	A	N	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G
Human-C (Bristol)	AJ132203	G	A	G	Y	P	L	I	V	A	N	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G
Human-C (V460)	AY941780	G	T	G	Y	P	L	I	V	A	N	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G
Bovine-C (Shintoku)	L12390	G	I	G	N	P	L	I	V	A	N	S	K	F	Q	I	D	R	I	L	I	S	G	D	I	E	L	N	P	G
Porcine-C (Cowden)	M69115	G	N	G	N	P	L	I	V	A	N	A	K	F	Q	I	D	K	I	L	I	S	G	D	V	E	L	N	P	G

Table 8. Table showing the mammalian 2A-like sequences analyzed. The conserved amino acids at the C-terminal motif are highlighted in yellow, and the differences in the region are shown in red.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	4	0	0	0	3	0	0	0	0	0	0	0	1	0
K	0	0	0	0	0	0	0	0	0	0	0	6	0	1	0	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	**8**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	18	0	0	0	0	0
D	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	6	0	3	0	0	**7**	0	15	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
N	0	1	0	2	0	0	0	0	**8**	6	0	3	0	0	0	0	**8**	0	0	0	0	0	0	0	0	0	18	0	0	0
T	0	1	0	0	0	0	0	0	0	0	0	0	0	0	**7**	0	0	0	0	3	0	0	0	0	0	1	0	0	0	0
S	**6**	0	0	3	0	1	0	0	0	0	5	1	3	3	0	0	0	0	0	0	**7**	0	0	0	0	11	0	0	0	0
C	2	0	**8**	1	**8**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
W	0	0	1	0	0	3	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	4	0	0	0	3	0	0	**8**	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	3	3	0	**8**	0	0	0	0	0	1	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	6	**8**	**1**	0	3	0	0	0	0	0	0	0	1	**17**	**8**	0	0	0	0	0	6	0	0	0	0
I	0	1	0	0	1	0	6	0	0	0	0	0	0	0	**7**	0	0	6	0	6	0	0	0	9	0	0	0	0	0	0
M	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
V	0	2	0	0	2	0	3	6	0	**8**	0	**8**	0	0	0	3	0	**8**	0	0	0	0	0	9	0	0	0	0	0	0
A	0	2	0	0	0	0	0	0	6	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	**6**	**8**	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**15**	0	0	0	0	0	0	17	0
P	0	0	0	0	7	**8**	0	8	1	0	0	0	**8**	**8**	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0	18
	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	**18**	**18**

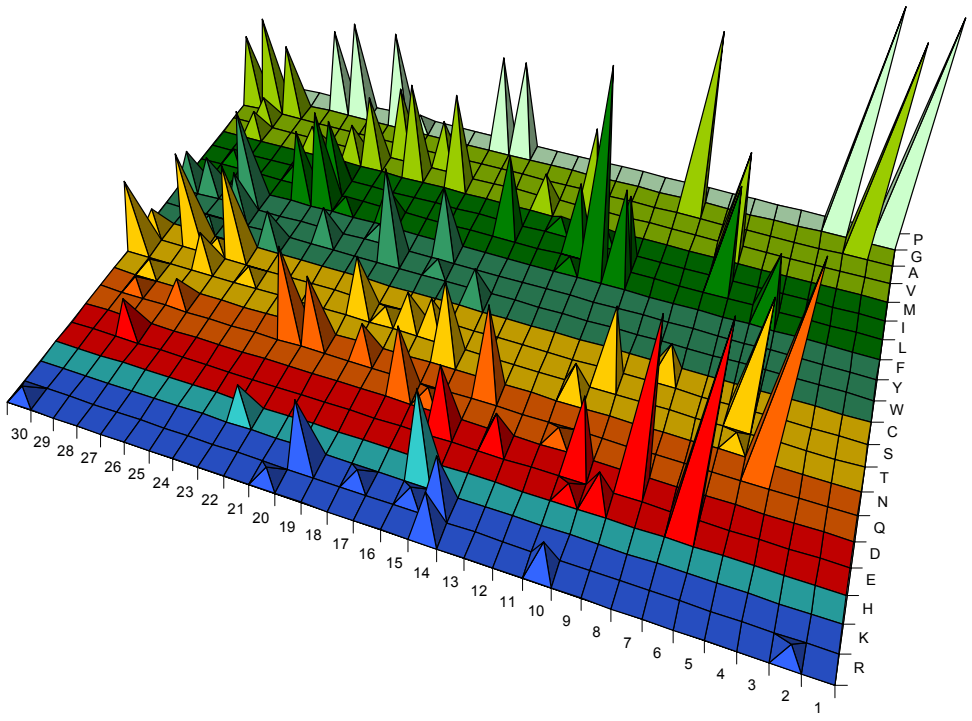


Figure 77. **Analisys of all the 2A-like sequences within Mammalian viruses**. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Cone graph, below, representing the results from the table. The data set analyzed is shown in Appendix 1, sequences 404 to 421

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	4	0	0	0	3	0	0	0	0	0	0	0	1	0
K	0	0	0	0	0	0	0	0	0	0	0	6	0	1	0	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	**8**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	18	0	0	0	0	0
D	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	6	0	3	0	0	**7**	0	15	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
N	0	1	0	2	0	0	0	0	**8**	6	0	3	0	0	0	0	**8**	0	0	0	0	0	0	0	0	0	18	0	0	0
T	0	1	0	0	0	0	0	0	0	0	0	0	0	0	**7**	0	0	0	0	3	0	0	0	0	0	1	0	0	0	0
S	**6**	0	0	3	0	1	0	0	0	0	5	1	3	3	0	0	0	0	0	0	**7**	0	0	0	0	11	0	0	0	0
C	2	0	**8**	1	**8**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
W	0	0	1	0	0	3	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	4	0	0	0	3	0	0	**8**	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	3	3	0	**8**	0	0	0	0	0	1	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	6	**8**	**1**	0	3	0	0	0	0	0	0	0	1	**17**	**8**	0	0	0	0	0	6	0	0	0	0
I	0	1	0	0	1	0	6	0	0	0	0	0	0	0	**7**	0	0	6	0	6	0	0	0	9	0	0	0	0	0	0
M	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
V	0	2	0	0	2	0	3	6	0	**8**	0	**8**	0	0	0	3	0	**8**	0	0	0	0	0	9	0	0	0	0	0	0
A	0	2	0	0	0	0	0	0	6	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	**6**	**8**	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**15**	0	0	0	0	0	17	0	
P	0	0	0	0	7	**8**	0	8	1	0	0	0	**8**	**8**	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0	18
	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18

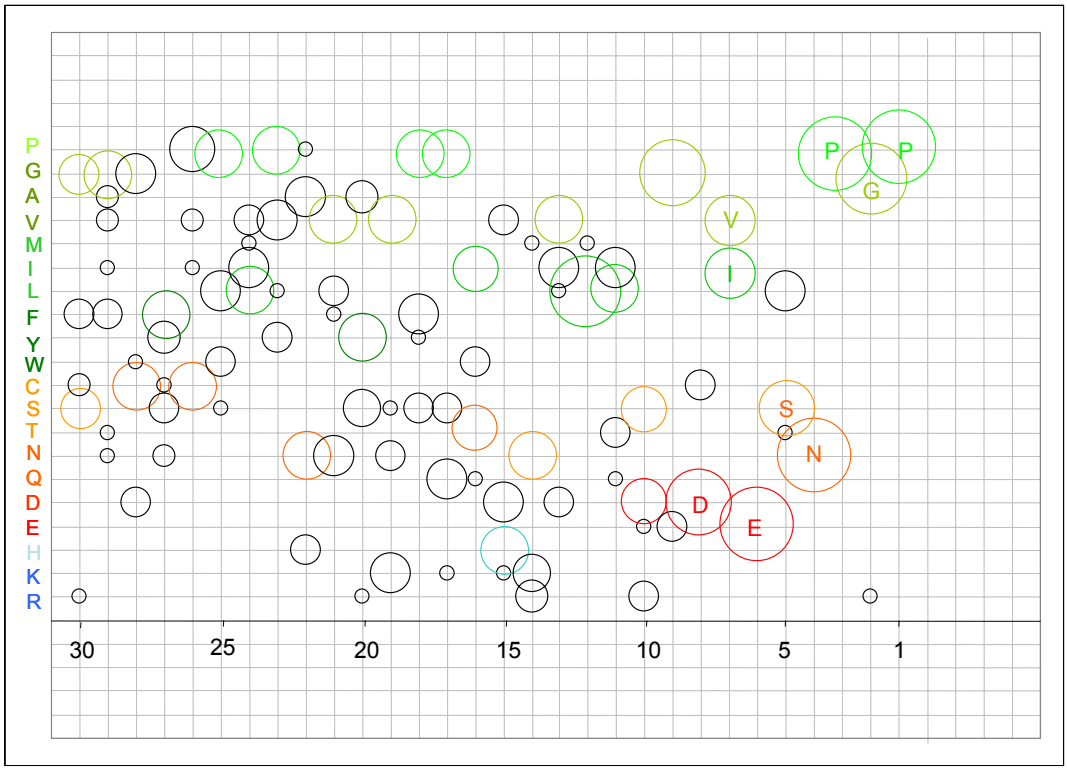


Figure 78. Bubble graph representing the data for 2A-like sequences within mammalian viruses. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Below, bubble graph; the bubbles represent the amino acids in each position. The size of the bubbles is related with the amount of amino acids in each position. The predominant amino acids are shown in colour. The data set analyzed is shown in Appendix 1, sequences 404 to 421

3.6.3 Insect 2A-like sequences

The analysis on the insect 2A-like sequences showed a similar pattern to that observed in picornaviruses, but with a much higher heterogeneity of residues in the upstream region, from position 10 to 30. Notable is the tract of amino acids with a preponderance of longer, aliphatic, sidechains (leucine, aspartate, glutamine, glutamate, lysine) observed in positions 11 to 14.

The sequences analyzed are represented in table 8, and the analysis in figures 79 and 80.

Species	Accession number	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
EoPV-2A1	AY365064	G	Q	R	T	T	E	Q	I	V	T	A	Q	G	W	A	P	D	L	T	Q	D	G	D	V	E	S	N	P	G	P
EoPV-2A2	AY365064	T	R	G	G	L	Q	R	Q	N	I	I	G	G	G	Q	R	D	L	T	Q	D	G	D	I	E	S	N	P	G	P
EoPV-2A1	AY341824	G	Q	R	T	T	E	Q	I	V	T	A	Q	G	W	A	P	D	L	T	Q	D	G	D	V	E	S	N	P	G	P
EoPV-2A2	AY341824	T	R	G	G	L	Q	R	Q	N	I	I	G	G	G	Q	R	D	L	T	Q	D	G	D	I	E	S	N	P	G	P
PnPV-2A1	AF323747	G	Q	R	T	T	E	Q	I	V	T	A	Q	G	W	V	P	D	L	T	V	D	G	D	V	E	S	N	P	G	P
PnPV-2A2	AF323747	T	R	G	G	L	R	R	Q	N	I	I	G	G	G	Q	K	D	L	T	Q	D	G	D	I	E	S	N	P	G	P
IFV	AB000906	P	S	I	G	N	V	A	R	T	L	T	R	A	E	I	E	D	E	L	I	R	A	G	I	E	S	N	P	G	P
ABPV (U.K.)	AF150629	T	G	F	L	N	K	L	Y	H	C	G	S	W	T	D	I	L	L	L	L	S	G	D	V	E	T	N	P	G	P
ABPV (Poland-1)	AF486073	T	G	F	L	N	K	L	Y	H	C	G	S	W	T	D	I	L	L	L	L	S	G	D	V	E	T	N	P	G	P
ABPV (Hungary-1)	AF486072	T	G	F	L	N	K	L	Y	H	C	G	S	W	T	D	I	L	L	L	W	S	G	D	V	E	T	N	P	G	P
KBV	AY275710	I	G	F	L	N	K	L	Y	K	C	G	T	W	E	S	V	L	N	L	L	A	G	D	I	E	L	N	P	G	P
IAPV	EF219380	I	G	F	L	N	K	L	Y	R	C	G	D	W	D	S	I	L	L	L	L	S	G	D	I	E	E	N	P	G	P
CrPV	AF218039	L	V	S	S	N	D	E	C	R	A	F	L	R	K	R	T	Q	L	L	M	S	G	D	V	E	S	N	P	G	P
DCV (EB)	AF014388	Q	G	I	G	K	K	N	P	K	Q	E	A	A	R	Q	M	L	L	L	L	S	G	D	V	E	T	N	P	G	P
TaV	AF062037	R	G	P	R	P	Q	N	L	G	V	R	A	E	G	R	G	S	L	L	T	C	G	D	V	E	E	N	P	G	P
TaV	AF282930	R	G	P	R	P	Q	N	L	G	V	R	A	E	G	R	G	S	L	L	T	C	G	D	V	E	E	N	P	G	P
EeV	AF461742	R	R	L	P	E	S	A	Q	L	P	Q	G	A	G	R	G	S	L	V	T	C	G	D	V	E	E	N	P	G	P
PrV-2A1	AF548354	L	E	M	K	E	S	N	S	G	Y	V	V	G	G	R	G	S	L	L	T	C	G	D	V	E	S	N	P	G	P
PrV-2A2	AF548354	N	S	D	D	E	E	P	E	Y	P	R	G	D	P	I	E	D	L	T	D	D	G	D	I	E	K	N	P	G	P
PrV-2A3	AF548354	T	L	M	G	N	I	M	T	L	A	G	S	G	G	R	G	S	L	L	T	A	G	D	V	E	K	N	P	G	P
D. punctatus CPV1	AY163248	M	T	A	F	D	F	Q	Q	A	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P
D. punctatus CPV1	AY185594	M	T	A	F	D	F	Q	Q	A	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P
L. dispar CPV1	AF389466	M	T	A	F	D	F	Q	Q	A	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P
B. mori	AF433660	R	T	A	F	D	F	Q	Q	D	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	I	E	S	N	P	G	P
B. mori CPV1-H	AB035733	R	T	A	F	D	F	Q	Q	D	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	I	E	S	N	P	G	P
B. mori CPV1-I	AB035732	R	T	A	F	D	F	Q	Q	D	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	I	E	S	N	P	G	P
O. brumata CPV18	DQ192245	I	H	A	N	D	Y	Q	M	A	V	F	K	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P

Table 8. Representation of all the insect 2A-like sequences analysed. The conserved C-terminal motif is highlighted in yellow.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	**6**	4	3	2	0	1	3	1	2	0	3	**7**	1	1	6	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0
K	0	0	0	1	1	**6**	0	0	2	0	0	1	0	1	0	1	0	0	7	0	0	0	0	0	0	2	0	0	0	0
H	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	1	0	0	3	4	1	1	0	0	1	0	2	2	0	2	0	1	0	0	0	0	0	0	**27**	4	0	0	0	0
D	0	0	1	1	7	1	0	0	3	0	0	1	1	1	3	**7**	8	0	0	1	7	0	**26**	0	0	0	0	0	0	0
Q	1	3	0	0	0	4	**10**	**10**	0	1	1	3	0	0	4	0	1	0	0	5	0	0	0	0	0	0	0	0	0	0
N	1	0	0	1	**8**	0	4	0	3	0	0	0	0	**7**	0	0	0	1	0	0	0	0	0	0	0	0	**27**	0	0	0
T	7	**6**	0	3	3	0	0	1	1	3	1	1	0	3	0	1	0	0	7	5	0	0	0	0	0	4	0	0	0	0
S	0	2	1	1	0	2	0	1	0	0	0	4	7	0	2	0	5	0	0	0	6	0	0	0	0	**16**	0	0	0	0
C	0	0	0	0	0	0	0	1	0	5	0	0	0	0	0	0	0	0	0	0	**11**	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	5	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	1	0	5	1	1	0	0	0	0	**7**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	**5**	**6**	0	**6**	0	0	0	0	**8**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	2	1	1	5	3	0	5	2	2	1	0	1	0	0	0	0	**13**	**25**	**12**	**12**	0	0	0	0	0	1	0	0	0	0
I	3	0	2	0	0	1	0	3	0	3	3	0	0	0	2	4	0	0	0	1	0	0	0	10	0	0	0	0	0	0
M	3	0	2	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
V	0	1	0	0	0	1	0	0	3	**9**	1	1	0	0	1	1	0	0	1	1	0	0	0	**17**	0	0	0	0	0	0
A	0	0	7	0	0	0	2	0	**4**	2	3	3	3	0	2	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0
G	3	8	3	6	0	0	0	0	3	0	6	5	**8**	8	0	5	0	0	0	0	0	**26**	1	0	0	0	0	0	**27**	0
P	1	0	2	1	2	0	1	1	0	2	0	0	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0	**27**	0	**27**
	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27

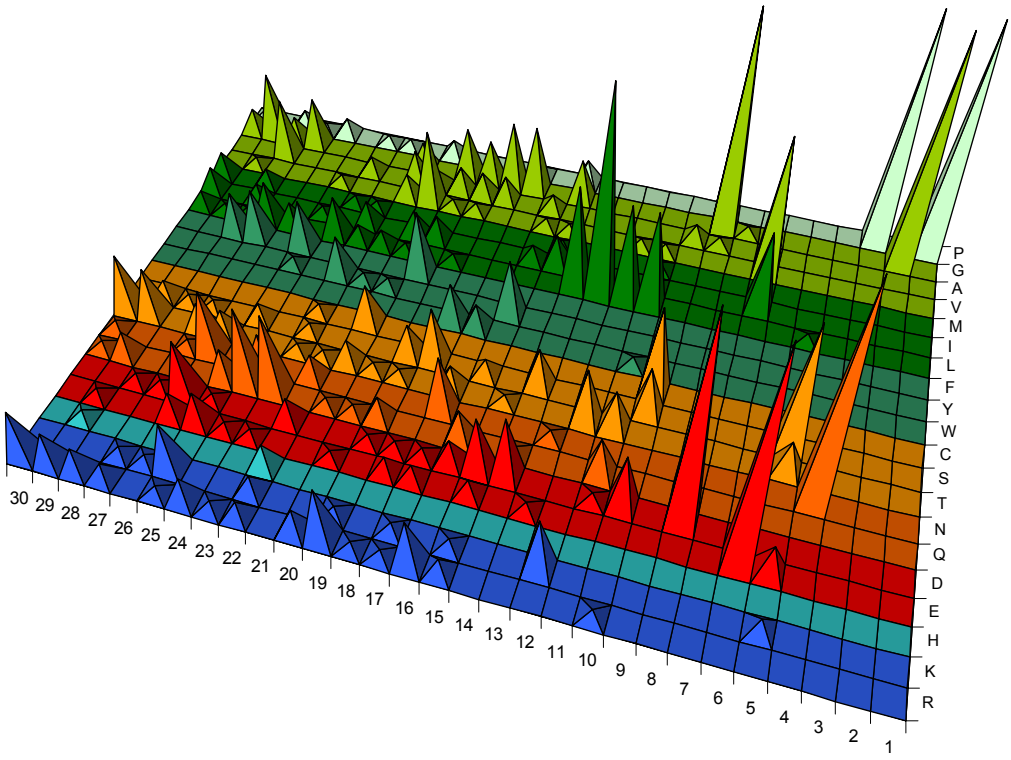


Figure 79. **Analisys of all the 2A-like sequences within insect viruses**. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Cone graph, below, representing the results from the table. The data set analyzed is shown in Appendix 1, sequences 422 to 448.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	**6**	4	3	2	0	1	3	1	2	0	3	**7**	1	1	6	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0
K	0	0	0	1	1	**6**	0	0	2	0	0	1	0	1	0	1	0	0	7	0	0	0	0	0	0	2	0	0	0	0
H	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	1	0	0	3	4	1	1	0	0	1	0	2	2	0	2	0	1	0	0	0	0	0	0	**27**	4	0	0	0	0
D	0	0	1	1	7	1	0	0	3	0	0	1	1	1	3	**7**	8	0	0	1	7	0	**26**	0	0	0	0	0	0	0
Q	1	3	0	0	0	4	**10**	**10**	0	1	1	3	0	0	4	0	1	0	0	5	0	0	0	0	0	0	0	0	0	0
N	1	0	0	1	**8**	0	4	0	3	0	0	0	0	**7**	0	0	0	1	0	0	0	0	0	0	0	0	**27**	0	0	0
T	7	**6**	0	3	3	0	0	1	1	3	1	1	0	3	0	1	0	0	7	5	0	0	0	0	0	4	0	0	0	0
S	0	2	1	1	0	2	0	1	0	0	0	4	7	0	2	0	5	0	0	0	6	0	0	0	0	**16**	0	0	0	0
C	0	0	0	0	0	0	0	1	0	5	0	0	0	0	0	0	0	0	0	0	**11**	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	5	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	1	0	5	1	1	0	0	0	0	**7**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	**5**	**6**	0	**6**	0	0	0	0	**8**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	2	1	1	5	3	0	5	2	2	1	0	1	0	0	0	0	**13**	**25**	**12**	**12**	0	0	0	0	0	1	0	0	0	0
I	3	0	2	0	0	1	0	3	0	3	3	0	0	0	2	4	0	0	0	1	0	0	0	10	0	0	0	0	0	0
M	3	0	2	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
V	0	1	0	0	0	1	0	0	3	**9**	1	1	0	0	1	1	0	0	1	1	0	0	0	**17**	0	0	0	0	0	0
A	0	0	7	0	0	0	2	0	**4**	2	3	3	3	0	2	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0
G	3	8	3	6	0	0	0	0	3	0	6	5	**8**	8	0	5	0	0	0	0	0	**26**	1	0	0	0	0	0	**27**	0
P	1	0	2	1	2	0	1	1	0	2	0	0	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0	**27**	0	**27**
	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27

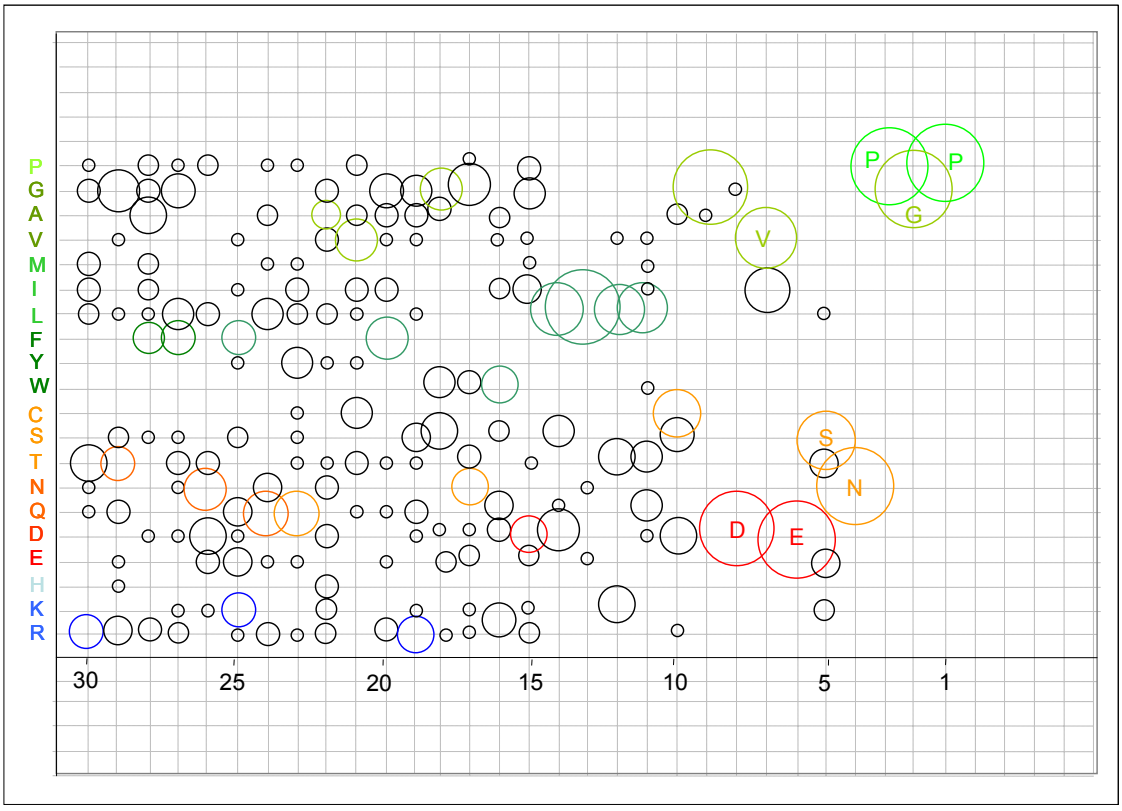


Figure 80. Bubble graph representing the data for insect 2A-like sequences. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. The bubbles represent the amino acids in each position. The size of the bubbles is related with the amount of amino acids in each position. The predominant amino acids are shown in colour. The data set analyzed is shown in Appendix 1, sequences 422 to 448.

3.6.4 Trypanosomal 2A-like sequences

The sequences analyzed belong to *T.cruzi* (14 strains), *T.brucei* (3 strains), *T.congolense* (37 strains) *and T.vivax* (83 strains). The 2A-like sequences are represented in table 9. Not all are shown since the number of strains in some cases is very high: a maximum of 17 strains are represented in the table.

Trypanosomal 2A-like sequences show a similar pattern observed in picornaviruses although their C-terminal motif is not totally conserved and point mutations are present in a wide range of these 2A-like sequences (see table 9). Charged amino acids predominate at positions 16, 17, 19 and 23. A majority of neutral residues are found in the other positions, some of them being aromatic (tyrosine, tryptophan and phenylalanine) similar to that previously observed in insect 2A-like sequences.

The tract of leucine residues (positions 12 to 15) is commonly flanked by a pair of basic residues at the N-terminal side, and a neutral residue (S) at the C-terminal side, like in insect 2A-like sequences (C). This neutral residue is mostly followed by a glycine in both cases.

It should be borne in mind that these non-LTR elements were active when transposed, but in many cases have accumulated mutations throughout evolution and become inactive. Indeed, the 'age' of such elements may be assessed by the accumulations of such mutations.

The analyses are shown in figures 81 and 82.

Species	Accession number	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
T.brucei	CAA29181	R	S	L	G	T	C	K	R	A	I	S	S	I	I	R	T	K	M	L	V	S	G	D	V	E	E	N	P	G	P
T.brucei	CAD21861	R	S	L	G	T	C	Q	R	A	I	S	S	I	I	R	T	K	M	L	L	S	G	D	V	E	E	N	P	G	P
T.brucei	CAD21860	R	S	L	G	T	C	Q	R	A	I	S	S	I	I	R	T	K	M	L	L	S	G	D	V	E	E	N	P	G	P
T.congo	354h04.q1k_6	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
T.congo	335b10.q1k_3	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
T.congo	432g10.q1k_7	I	L	P	C	T	C	G	C	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	V	E	R	N	P	G	P
T.congo	400g12.q1k_4	I	L	P	C	T	C	G	R	T	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
T.congo	876g11.p1k_3	I	L	P	C	T	C	G	R	T	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
T.congo	1381h11.q1k_4	I	V	P	C	T	C	G	R	T	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
T.congo	1071g10.p1k_14	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	F	L	L	P	V	R	G	D	V	G	R	N	P	G	P
T.congo	1294e07.p1k_3	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	F	L	L	P	V	R	G	D	V	G	R	N	P	G	P
T.congo	1473f10.p1k_5	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	F	L	L	P	V	R	G	D	V	G	R	N	P	G	P
T.congo	1305b04.p1k_2	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	F	L	L	P	V	R	G	D	V	G	R	N	P	G	P
T.congo	530f06.q1kbw_10	I	L	P	C	T	C	I	C	P	T	L	E	A	R	R	L	L	V	L	V	S	G	G	I	E	R	N	P	R	P
T.congo	1463e05.p1k_0	A	L	S	C	V	C	G	H	G	N	S	L	L	C	R	L	L	L	F	L	S	G	D	V	E	Y	N	P	G	S
T.congo	800b12.p1k_3	A	L	S	C	V	C	G	H	G	N	S	L	L	C	R	L	L	L	F	L	S	G	N	V	E	Y	N	P	G	S
T.congo	47d01.q1k_6	A	L	S	C	V	C	G	H	G	N	S	L	L	C	R	L	L	L	F	L	S	G	N	V	E	Y	N	P	G	S
T.congo	987a11.q1k_0	A	L	S	C	V	C	G	H	G	N	S	L	L	C	R	L	L	L	F	L	S	G	N	V	E	Y	N	P	G	S
T.congo	1423d04.p1k_0	T	L	S	C	T	C	G	S	A	L	P	K	A	L	G	P	L	L	L	L	S	R	V	E	D	H	N	P	G	P
T.congo	1182h09.q1k_0	F	T	C	T	C	W	R	G	R	A	L	L	C	R	P	F	L	M	P	L	S	G	D	V	G	Q	N	P	E	P
T.cruzi	AAA67559	Q	P	Y	T	Y	C	L	R	A	L	C	D	A	Q	R	Q	K	L	L	L	I	G	D	I	E	Q	N	P	G	P
T.cruzi	CAB41692	Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	R	R	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	Q	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	R	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	P	Q	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	G	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	N	Q	A	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	N	P	S	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	H	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	R	R	Q	K	L	L	L	S	G	D	I	E	Q	H	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	N	G	D	I	E	Q	H	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	Y	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	H	P	G	P
T.vivax	638g08.p1k_1	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
T.vivax	262a12.p1k_4	M	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
T.vivax	1734a06.p1k_4	M	L	P	C	A	C	G	R	A	T	L	D	A	R	R	L	T	L	L	V	S	G	D	V	E	R	D	P	G	P
T.vivax	720g04.q1k_0	I	L	P	C	T	C	E	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
T.vivax	814g01.p1k_9	T	L	P	F	A	R	W	H	I	A	L	D	M	R	R	P	L	L	L	I	S	G	D	V	D	S	K	P	G	P
T.vivax	1814e03.p1k_1	L	L	P	C	T	C	G	R	A	T	L	D	A	W	R	L	L	L	L	I	C	G	G	V	G	R	N	P	G	P
T.vivax	346a10.p1k_3	M	L	L	C	T	R	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	P	G	P
T.vivax	1198e11.p1k_1	M	L	L	C	T	R	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	P	G	P
T.vivax	104g02.p1k_0	M	L	L	C	T	R	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	P	G	P
T.vivax	961a05.q1k_2	I	L	P	C	T	C	G	R	A	A	L	D	A	Q	W	R	L	L	L	I	F	V	D	A	E	R	N	P	G	P
T.vivax	1278c04.p1k_6	I	L	P	C	T	C	G	R	A	A	L	D	A	Q	W	R	L	L	L	I	F	V	D	A	E	R	N	P	G	P
T.vivax	856h07.q1k_12	I	L	P	C	T	R	G	R	A	M	L	S	A	R	W	L	L	L	L	I	S	G	G	V	E	R	K	P	G	P
T.vivax	1858e01.q1k_5	I	L	P	F	T	C	G	R	A	A	L	D	A	W	R	L	L	L	L	I	G	G	G	V	G	R	N	P	G	P
T.vivax	158a04.q1k_12	I	L	P	C	L	C	V	H	A	A	S	D	A	R	W	L	L	L	L	I	S	G	D	V	E	R	R	P	C	P
T.vivax	1890c02.p1k_12	M	L	L	C	T	S	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	S	G	P
T.vivax	1013b08.p1k_5	S	Q	V	R	W	S	N	G	A	E	K	K	V	Q	R	L	L	L	L	S	G	G	D	V	E	R	N	P	G	P

Table 9. Trypanosome 2A-like sequences. The C-terminal motif is highlighted in yellow and the differences in this region in red. Sequences without an accession number were taken from Heras *et al.*, 2006.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	3	17	0	7	1	27	3	**110**	1	13	1	6	1	**84**	**110**	11	0	0	2	0	5	7	0	0	0	**91**	1	0	1	0
K	1	0	1	0	0	0	1	0	0	0	2	4	0	0	0	1	22	0	0	0	0	0	0	0	6	0	4	0	0	0
H	0	1	4	0	0	0	0	8	0	0	0	0	13	0	0	1	0	0	0	0	0	0	0	0	0	3	4	0	0	0
E	0	1	0	0	0	1	3	0	0	1	0	6	1	0	0	0	0	0	0	0	0	1	0	3	**115**	3	0	0	1	0
D	0	0	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**97**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	2	**0**	**114**	**0**	**4**	**0**	**8**	**0**	**0**	**0**
Q	14	1	**0**	**0**	**0**	**0**	7	**0**	**0**	**0**	5	**0**	**0**	16	2	18	**0**	**0**	**0**	**0**	1	**0**	**0**	**0**	**0**	15	**0**	1	**0**	**0**
N	1	**0**	**0**	2	5	**0**	1	**0**	**0**	4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	3	**0**	4	**0**	**0**	**0**	**116**	**0**	**0**	1
T	5	2	**0**	16	**100**	**0**	**0**	**0**	4	**60**	**0**	**0**	2	**0**	**0**	5	1	**0**	**0**	1	1	**0**	4	**0**	**0**	**0**	1	**0**	**0**	1
S	13	14	8	10	**0**	14	1	5	**0**	**0**	11	4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	1	**106**	**0**	**0**	**0**	**0**	3	**0**	2	1	4
C	**0**	**0**	1	**87**	2	**91**	**0**	9	1	1	12	12	1	4	**0**	13	**0**	**0**	**0**	**0**	7	**0**	**0**	**0**	**0**	2	**0**	**0**	1	**0**
W	**0**	**0**	**0**	**0**	1	1	1	**0**	**0**	**0**	**0**	**0**	**0**	2	8	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	1	**0**
Y	1	**0**	14	1	14	**0**	**0**	1	**0**	**0**	1	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	1	**0**	**0**	**0**	**0**	**0**	4	**0**	**0**	**0**	**0**
F	1	2	1	4	**0**	1	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	5	**0**	1	17	**0**	2	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
L	13	**96**	20	2	1	**0**	27	**0**	1	9	**89**	5	5	11	**0**	**59**	**95**	**128**	**109**	29	**0**	**0**	**0**	**0**	**0**	13	**0**	**0**	**0**	**0**
I	**69**	**0**	**0**	**0**	**0**	**0**	1	**0**	1	3	**0**	**0**	3	3	**0**	17	**0**	**0**	**0**	**70**	4	**0**	**0**	24	**0**	**0**	2	**0**	**0**	**0**
M	9	**0**	**0**	**0**	1	**0**	**0**	**0**	**0**	7	**0**	**0**	1	**0**	**0**	**0**	**0**	4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	1	**0**	**0**	**0**
V	2	2	1	**0**	5	1	1	**0**	12	16	13	**0**	7	13	**0**	**0**	**0**	3	1	30	**0**	17	3	**107**	**0**	**0**	**0**	**0**	1	**0**
A	5	0	5	0	7	0	0	2	**112**	20	0	1	**103**	1	1	1	13	1	0	4	0	0	1	3	0	3	0	1	2	2
G	0	0	0	3	0	1	**91**	2	4	2	0	2	0	0	14	0	0	0	0	0	6	**112**	11	0	12	0	0	0	**129**	0
P	0	1	**82**	5	0	0	0	0	1	1	3	0	0	3	2	6	6	0	8	1	0	0	0	0	0	0	0	**133**	0	**129**
	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137

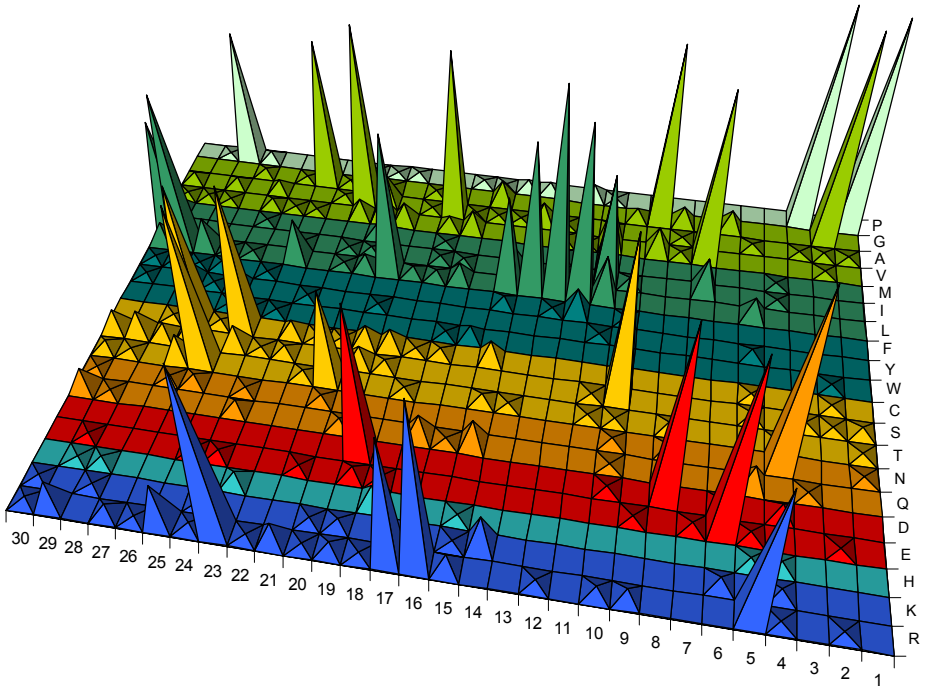


Figure 81. **Analisys of all the 2A-like sequences within Trypanosomes**. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Cone graph, below, representing the results from the table. . The data set analyzed is shown in Appendix 1, sequences 449 to 585.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	3	17	0	7	1	27	3	**110**	1	13	1	6	1	**84**	**110**	11	0	0	2	0	5	7	0	0	0	**91**	1	0	1	0
K	1	0	1	0	0	0	1	0	0	0	2	4	0	0	0	1	22	0	0	0	0	0	0	0	6	0	4	0	0	0
H	0	1	4	0	0	0	0	8	0	0	0	0	13	0	0	1	0	0	0	0	0	0	0	0	0	3	4	0	0	0
E	0	1	0	0	0	1	3	0	0	1	0	6	1	0	0	0	0	0	0	0	0	1	0	3	**115**	3	0	0	1	0
D	0	0	0	0	0	0	0	0	0	0	0	**97**	0	0	0	0	0	0	0	0	2	0	**114**	0	4	0	8	0	0	0
Q	14	1	0	0	0	0	7	0	0	0	5	0	0	16	2	18	0	0	0	0	1	0	0	0	0	15	0	1	0	0
N	1	0	0	2	5	0	1	0	0	4	0	0	0	0	0	0	0	0	0	0	3	0	4	0	0	0	**116**	0	0	1
T	5	2	0	16	**100**	0	0	0	4	**60**	0	0	2	0	0	5	1	0	0	1	1	0	4	0	0	0	1	0	0	1
S	13	14	8	10	0	14	1	5	0	0	11	4	0	0	0	0	0	0	0	1	**106**	0	0	0	0	3	0	2	1	4
C	0	0	1	**87**	2	**91**	0	9	1	1	12	12	1	4	0	13	0	0	0	0	7	0	0	0	0	2	0	0	1	0
W	0	0	0	0	1	1	1	0	0	0	0	0	0	2	8	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Y	1	0	14	1	14	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	4	0	0	0	0
F	1	2	1	4	0	1	0	0	0	0	0	0	0	0	0	5	0	1	17	0	2	0	0	0	0	0	0	0	0	0
L	13	**96**	20	2	1	0	27	0	1	9	**89**	5	5	11	0	**59**	**95**	**128**	**109**	29	0	0	0	0	0	13	0	0	0	0
I	**69**	0	0	0	0	0	1	0	1	3	0	0	3	3	0	17	0	0	0	70	4	0	0	24	0	0	2	0	0	0
M	9	0	0	0	1	0	0	0	0	7	0	0	1	0	0	0	0	4	0	0	0	0	0	0	0	0	1	0	0	0
V	2	2	1	0	5	1	1	0	12	16	13	0	7	13	0	0	0	3	1	**30**	0	17	3	**107**	0	0	0	0	1	0
A	5	0	5	0	7	0	0	2	**112**	20	0	1	**103**	1	1	1	13	1	0	4	0	0	1	3	0	3	0	1	2	2
G	0	0	0	3	0	1	**91**	2	4	2	0	2	0	0	14	0	0	0	0	0	6	**112**	11	0	12	0	0	0	**129**	0
P	0	1	**82**	5	0	0	0	0	1	1	3	0	0	3	2	6	6	0	8	1	0	0	0	0	0	0	0	**133**	0	**129**
	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137

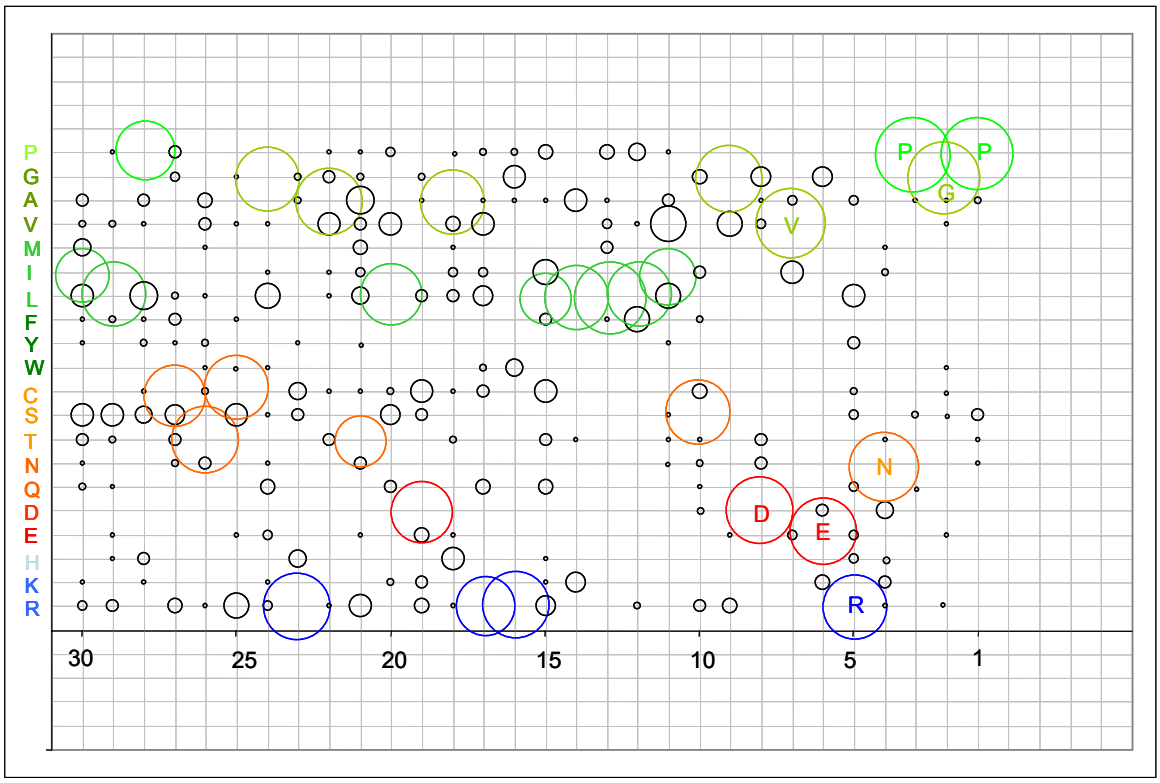


Figure 82. Bubble graph representing the data for trypanosomal 2A-like sequences. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Below, bubble graph; the bubbles represent the amino acids in each position. The size of the bubbles is related with the amount of amino acids in each position. The predominant amino acids are shown in colour. The data set analyzed is shown in Appendix 1, sequences 449 to 585.

3.6.5 Strongylocentrotus purpuratus 2A-like sequences

The analyses were divided in two groups of 2A-like sequences: CATERPILLER sequences and non-LTR retrotransposons. Both groups show a pattern similar than picornavirus 2A sequences being the C-terminal motif more conserved in CATERPILLER 2As than non-LTR retrotransposons. A remarkable difference between the two types of 2A-like sequences should be noted; this is the distribution of amino acids surrounding the patch of leucine residues, which is observed in both cases (also observed in insect 2A-like sequences). Non-LTR 2A-like sequences comprise a patch of leucine residues at positions 11 to 14, flanked by neutral residues. In contrast, CATERPILLER 2A-like sequences, which have the tract of leucine residues at positions 12 to 15, possess a basic residue (K) and an acidic residue (E) at positions 11 and 16, respectively. This is consistent with their recently observed function as signal sequences (see discussion, section 4.4.2)

The results of the analyses are represented in Tables 10 and 11 and figures 83, 84, 85, 86.

3.6.5.1 CATERPILLER sequences

Species	Accession number	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
STR-20	XP_001196456	S	K	T	D	L	I	S	G	Q	F	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
STR-50	GLEAN3_03186	S	K	T	D	L	I	S	G	Q	I	P	H	L	S	E	L	L	L	M	K	S	G	D	V	E	L	N	P	G	P
STR-65	GLEAN3_09160	S	K	T	E	L	M	S	G	Q	I	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
STR-70	GLEAN3_22394	S	K	T	D	L	I	S	G	Q	I	P	S	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
STR-76	GLEAN3_03448	S	K	T	D	L	I	S	G	Q	I	P	P	L	S	K	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
STR-81	GLEAN3_21478	S	K	T	D	L	I	S	G	Q	I	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
STR-83	GLEAN3_22780	S	K	T	D	L	I	S	G	Q	I	P	P	L	S	E	L	L	L	M	K	S	G	D	V	E	L	N	P	G	P
STR-100	GLEAN3_15340	S	K	T	D	L	I	S	G	Q	I	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
STR-111	GLEAN3_20436	S	K	T	D	L	I	S	G	Q	F	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
STR-127	GLEAN3_23550	S	K	T	D	L	I	S	G	Q	I	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
STR-147	GLEAN3_11433	S	K	T	D	L	I	S	G	Q	I	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
STR-23	XP_001198729	L	H	P	A	I	L	C	S	A	S	L	C	F	R	P	Y	L	L	L	M	A	G	D	V	E	P	N	P	G	P
STR-35	XP_001200466	N	S	S	C	V	L	N	I	R	S	T	S	H	L	A	I	L	L	L	L	S	G	Q	V	E	P	N	P	G	P
STR-54	GLEAN3_25204	S	Q	N	I	D	V	L	S	Q	Q	P	Y	L	T	E	L	L	L	V	K	A	G	D	V	E	L	N	P	G	P
STR-60	GLEAN3_09111	Q	N	L	D	F	N	L	Y	L	L	M	I	L	L	M	I	L	L	M	R	S	G	D	V	E	T	N	P	G	P
STR-67	GLEAN3_08283	P	Q	Q	D	L	Q	G	F	C	L	L	Y	L	L	M	I	L	L	M	R	S	G	D	V	E	T	N	P	G	P
STR-82	GLEAN3_25914	T	T	D	D	P	V	V	Q	E	S	T	C	L	P	E	M	I	L	V	K	A	G	D	V	E	Q	N	P	G	P
STR-106	GLEAN3_23532	L	H	P	A	I	L	C	S	A	S	L	C	F	R	P	Y	L	L	L	M	A	G	D	V	E	P	N	P	G	P
STR-110	GLEAN3_06203	Q	D	L	D	V	K	E	A	D	K	P	H	I	T	Q	S	L	I	L	K	A	G	D	V	E	S	N	P	G	P
STR-136	GLEAN3_20380	G	A	V	D	V	V	L	S	Q	Q	P	Y	L	T	E	L	L	L	V	K	A	G	D	V	E	L	N	P	G	P
STR-143	GLEAN3_22449	S	R	P	I	L	Y	Y	S	N	T	T	A	S	F	Q	L	S	T	L	L	S	G	D	I	E	P	N	P	G	P

Table 10. CATERPILLER 2A-like sequences analyzed within *Strongylocentrotus purpuratus*. C-terminal conserved motif is shown in yellow.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	0	1	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
K	0	**11**	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	**15**	0	0	0	0	0	0	0	0	0	0
H	0	2	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	1	0	0	1	0	1	0	0	0	0	0	**13**	0	0	0	0	0	0	0	0	0	**21**	0	0	0	0	0
D	0	1	1	**15**	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**20**	0	0	0	0	0	0	0
Q	2	2	1	0	0	1	0	1	**13**	2	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
N	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0	0	0
T	1	1	**11**	0	0	0	0	0	0	1	3	0	0	3	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0
S	**13**	1	1	0	0	0	**11**	5	0	4	0	2	1	**11**	0	1	1	0	0	0	**15**	0	0	0	0	1	0	0	0	0
C	0	0	0	1	0	0	2	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	1	1	1	0	0	0	3	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	1	0	0	1	0	2	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	2	0	2	0	**13**	3	3	0	1	2	3	0	**16**	3	0	**14**	**19**	**19**	**14**	2	0	0	0	0	0	**13**	0	0	0	0
I	0	0	0	2	2	**10**	0	1	0	**9**	0	1	1	0	0	3	1	1	0	0	0	0	0	1	0	0	0	0	0	0
M	0	0	0	0	0	1	0	0	0	0	1	0	0	0	2	1	0	0	4	2	0	0	0	0	0	0	0	0	0	0
V	0	0	1	0	3	3	1	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	**20**	0	0	0	0	0	0
A	0	1	0	2	0	0	0	1	2	0	0	1	0	0	1	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
G	1	0	0	0	0	0	1	**11**	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0	0	0	0	0	0	**21**	0
P	1	0	3	0	1	0	0	0	0	0	**14**	**9**	0	1	2	0	0	0	0	0	0	0	0	0	0	4	0	**21**	0	**21**
	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21

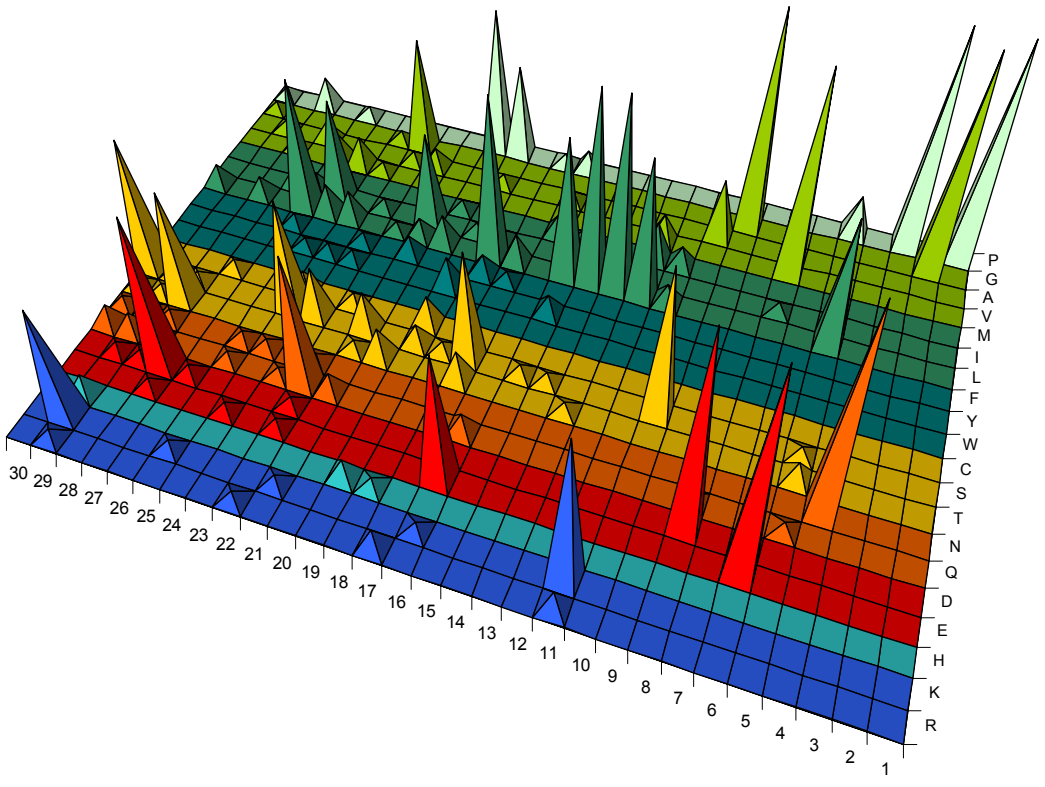


Figure 83. **Analisys of the CATERPILLER 2A-like sequences within *Strongylocentrotus purpuratus***. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Cone graph, below, representing the results from the table. . The data set analyzed is shown in Appendix 1, sequences 586 to 606.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	0	1	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
K	0	**11**	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	**15**	0	0	0	0	0	0	0	0	0	0
H	0	2	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	1	0	0	1	0	1	0	0	0	0	0	**13**	0	0	0	0	0	0	0	0	0	**21**	0	0	0	0	0
D	0	1	1	**15**	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**20**	0	0	0	0	0	0	0
Q	2	2	1	0	0	1	0	1	**13**	2	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
N	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0	0	0
T	1	1	**11**	0	0	0	0	0	0	1	3	0	0	3	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0
S	**13**	1	1	0	0	0	**11**	5	0	4	0	2	1	**11**	0	1	1	0	0	0	**15**	0	0	0	0	1	0	0	0	0
C	0	0	0	1	0	0	2	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	1	1	1	0	0	0	3	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	1	0	0	1	0	2	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	2	0	2	0	**13**	3	3	0	1	2	3	0	**16**	3	0	**14**	**19**	**19**	**14**	2	0	0	0	0	0	**13**	0	0	0	0
I	0	0	0	2	2	**10**	0	1	0	**9**	0	1	1	0	0	3	1	1	0	0	0	0	0	1	0	0	0	0	0	0
M	0	0	0	0	0	1	0	0	0	0	1	0	0	0	2	1	0	0	4	2	0	0	0	0	0	0	0	0	0	0
V	0	0	1	0	3	3	1	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	**20**	0	0	0	0	0	0
A	0	1	0	2	0	0	0	1	2	0	0	1	0	0	1	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
G	1	0	0	0	0	0	1	**11**	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0	0	0	0	0	0	**21**	0
P	1	0	3	0	1	0	0	0	0	0	**14**	**9**	0	1	2	0	0	0	0	0	0	0	0	0	0	4	0	**21**	0	**21**
	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21

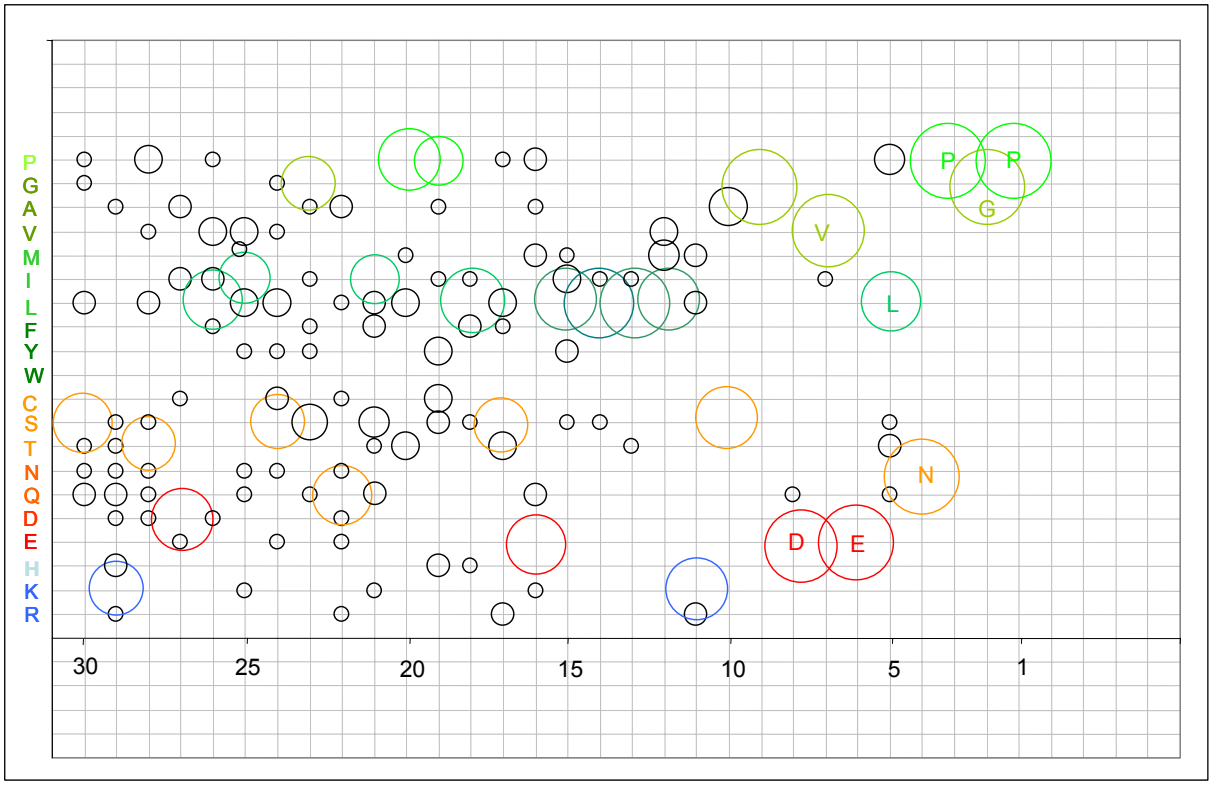


Figure 84. Bubble graph representing the data for CATERPILLER 2A-like sequences within *Strongylocentrotus purpuratus*. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Bubble graph, below; the bubbles represent the amino acids in each position. The size of the bubbles is related with the amount of amino acids in each position. The predominant amino acids are shown in colour. . The data set analyzed is shown in Appendix 1, sequences 586 to 606.

3.6.5.2 Non-LTR sequences

Species	Accession number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
STR-1	XP_797143	N	S	T	P	A	A	M	F	V	C	A	F	I	L	I	S	V	L	L	L	S	G	D	V	E	I	N	P	G	P
STR-34	XP_001196844	N	S	T	P	A	A	M	F	V	C	V	F	I	L	I	S	V	L	L	L	S	G	D	V	E	I	S	P	G	P
STR-24	XP_001196407	S	Q	R	D	L	S	C	S	Q	P	R	T	I	I	L	G	L	I	M	C	A	G	D	V	Q	P	N	P	G	P
STR-25	XP_001186348	S	Q	R	D	L	S	C	S	Q	P	R	T	I	I	L	G	L	I	M	C	A	G	D	V	Q	P	N	P	G	P
STR-32	XP_001185404	N	S	S	C	V	L	N	I	R	S	T	S	H	L	A	I	L	L	L	L	S	G	Q	V	E	P	N	P	G	P
STR-35	XP_001200466	N	S	S	C	V	L	N	I	R	S	T	S	H	L	A	I	L	L	L	L	S	G	Q	V	E	P	N	P	G	P
STR-164	XR_025775	N	S	S	C	V	L	N	I	R	S	T	S	H	L	A	I	L	L	L	L	S	G	Q	V	E	P	N	P	G	P
STR-27	XP_001185149	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
STR-28	XP_001179204	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
STR-29	XP_791376	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
STR-30	XP_001199602	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
STR-31	XP_001200060	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
STR-33	XP_001184905	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
STR-36	XP_001180489	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
STR-163	XP_001192137	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	D	P
STR-116/160	XR_026225	T	T	C	Q	C	K	A	L	S	V	M	Y	L	T	L	L	L	L	T	N	A	S	D	I	E	L	N	P	G	P
STR-40/141	GLEAN3_18025	K	S	C	I	S	Y	Y	S	N	S	T	A	C	F	N	I	E	I	M	C	C	G	D	V	K	S	N	P	G	P
STR-55	GLEAN3_24854	G	A	R	I	S	Y	H	P	N	T	T	A	T	F	Q	L	R	L	L	V	S	G	D	V	N	P	N	P	G	P
STR-61	GLEAN3_22393	G	A	R	I	R	Y	Y	N	N	S	S	A	T	F	Q	T	I	L	M	T	C	G	D	V	D	P	N	P	G	P
STR-89	GLEAN3_19055	G	R	R	I	Q	Y	Y	N	N	S	I	S	T	F	R	S	E	L	L	R	C	G	D	V	E	S	N	P	G	P
STR-38	XP_793501	K	T	R	I	P	Y	S	V	N	S	N	A	S	F	Q	L	E	L	L	H	A	G	D	V	H	P	N	P	G	P
STR-51	GLEAN3_22449	S	R	P	I	L	Y	Y	S	N	T	T	A	S	F	Q	L	S	T	L	L	S	G	D	I	E	P	N	P	G	P
STR-69	GLEAN3_27016	C	R	R	I	A	Y	Y	S	N	S	D	C	T	F	R	L	E	L	L	K	S	G	D	I	Q	S	N	P	G	P
STR-133	GLEAN3_00868	K	R	R	I	P	Y	N	P	N	S	T	A	S	F	Q	L	E	L	L	H	A	G	D	V	H	P	N	P	G	P

Table 11. Non-LTR 2A-like sequences analyzed within *Strongylocentrotus purpuratus.* The C-terminal conserved amino acids are highlighted in yellow, and the differences in this region are shown in red.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	0	4	8	0	1	0	**8**	0	3	0	2	0	0	0	2	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
K	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
H	0	0	0	0	0	0	1	0	0	0	0	0	**11**	0	0	0	0	0	0	2	0	0	0	0	2	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	**16**	0	0	0	0	0
D	0	0	0	2	**8**	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	**13**	0	1	0	0	0	1	0
Q	0	2	0	1	1	0	0	0	2	0	0	0	0	0	5	0	0	0	0	0	0	0	11	0	3	0	0	0	0	0
N	5	0	0	0	0	0	4	2	**8**	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	**23**	0	0	0
T	1	2	2	0	0	0	0	0	**8**	2	7	2	4	1	**8**	1	0	1	1	1	0	0	0	0	0	8	0	0	0	0
S	3	6	3	0	2	2	1	**13**	1	**17**	1	**12**	3	0	0	3	1	0	0	0	**16**	1	0	0	0	3	1	0	0	0
C	1	**8**	2	3	1	0	2	0	0	2	0	1	1	0	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	**8**	5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	**8**	0	2	0	0	0	2	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	**8**	0	0	**8**	3	3	0	1	0	0	**8**	0	1	**13**	3	6	**14**	**20**	**19**	**14**	0	0	0	0	0	1	0	0	0	0
I	0	0	0	**8**	0	0	0	3	0	0	1	0	4	2	2	**12**	1	3	0	0	0	0	0	3	0	2	0	0	0	0
M	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
V	0	0	0	0	3	0	0	1	2	1	1	0	0	0	0	0	2	0	0	1	0	0	0	**21**	0	0	0	0	0	0
A	0	2	0	0	3	2	1	0	0	0	1	6	0	0	3	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0
G	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	**23**	0	0	0	0	0	0	**23**	0
P	0	0	**9**	2	2	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	**24**	0	**24**
	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24

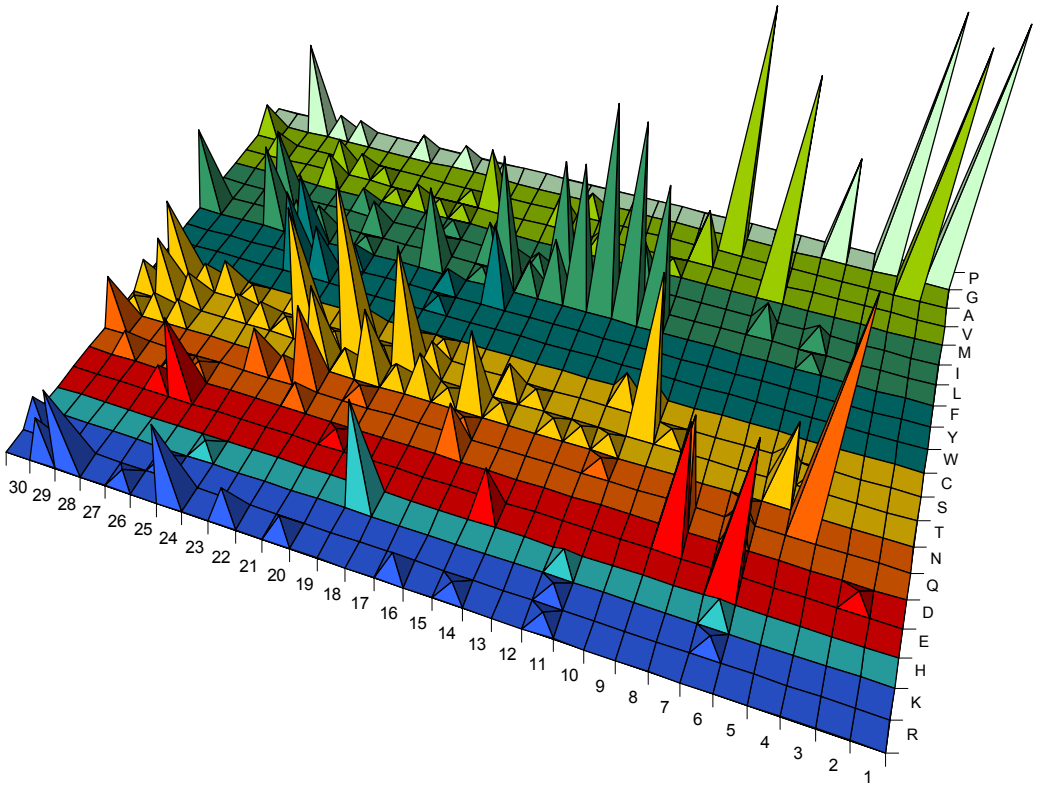


Figure 85. **Analisys of the non-LTR 2A-like sequences within *Strongylocentrotus purpuratus***. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Cone graph, below, representing the results from the table. The data set analyzed is shown in Appendix 1, sequences 607 to 630.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	0	4	8	0	1	0	**8**	0	3	0	2	0	0	0	2	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
K	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
H	0	0	0	0	0	0	1	0	0	0	0	0	**11**	0	0	0	0	0	0	2	0	0	0	0	2	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	**16**	0	0	0	0	0
D	0	0	0	2	**8**	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	**13**	0	1	0	0	0	1	0
Q	0	2	0	1	1	0	0	0	2	0	0	0	0	0	5	0	0	0	0	0	0	0	11	0	3	0	0	0	0	0
N	5	0	0	0	0	0	4	2	**8**	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	**23**	0	0	0
T	1	2	2	0	0	0	0	0	**8**	2	7	2	4	1	**8**	1	0	1	1	1	0	0	0	0	0	8	0	0	0	0
S	3	6	3	0	2	2	1	**13**	1	**17**	1	**12**	3	0	0	3	1	0	0	0	**16**	1	0	0	0	3	1	0	0	0
C	1	**8**	2	3	1	0	2	0	0	2	0	1	1	0	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	**8**	5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	**8**	0	2	0	0	0	2	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	**8**	0	0	**8**	3	3	0	1	0	0	**8**	0	1	**13**	3	6	**14**	**20**	**19**	**14**	0	0	0	0	0	1	0	0	0	0
I	0	0	0	**8**	0	0	0	3	0	0	1	0	4	2	2	**12**	1	3	0	0	0	0	0	3	0	2	0	0	0	0
M	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
V	0	0	0	0	3	0	0	1	2	1	1	0	0	0	0	0	2	0	0	1	0	0	0	**21**	0	0	0	0	0	0
A	0	2	0	0	3	2	1	0	0	0	1	6	0	0	3	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0
G	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	**23**	0	0	0	0	0	0	**23**	0
P	0	0	**9**	2	2	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	0	**24**	0	**24**
	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24

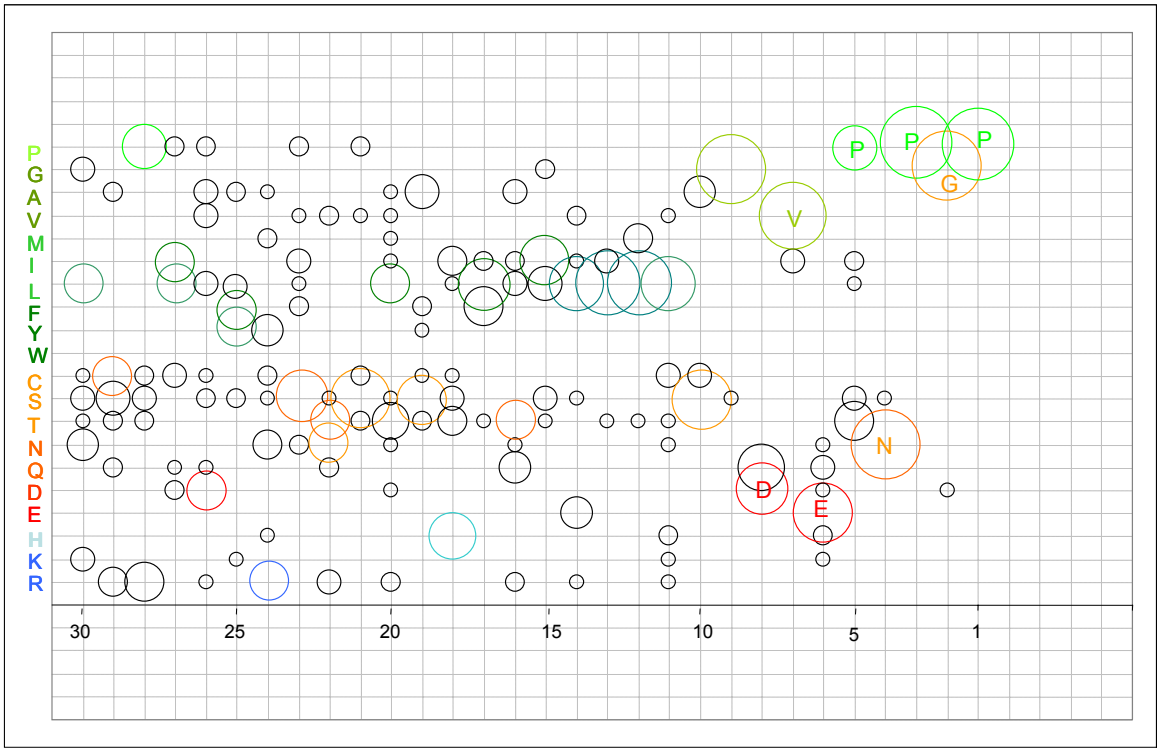


Figure 86. Bubble graph representing the data for non-LTR 2A-like sequences within *Strongylocentrotus purpuratus*. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Below, bubble graph; the bubbles represent the amino acids in each position. The size of the bubbles is related with the amount of amino acids in each position. The predominant amino acids are shown in colour. The data set analyzed is shown in Appendix 1, sequences 607 to 630.

3.6.6 Non-LTR 2A-like sequences

All the non-LTR 2A-like sequences within Trypanosomes and *Strongylocentrotus purpuratus* were analyzed together. The distribution of amino acids is very similar in both cases, with a tract of leucine residues at positions 12 to 15. This hydrophobic tract is very commonly flanked at its N-terminus by a pair of basic residues and at the C-terminus by either serine or cystine (hydroxyl or sulphydryl groups). The majority of the residues present are neutral, polar or non-polar, except in positions 12, 17 and 23, where arginine predominates.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	3	21	8	7	2	27	11	**110**	4	13	3	6	1	**84**	**112**	11	1	0	2	1	5	7	0	0	0	**91**	1	0	1	0
K	4	0	1	0	0	1	1	0	0	0	2	4	0	0	0	1	22	0	0	1	0	0	0	0	7	0	4	0	0	0
H	0	1	4	0	0	0	1	8	0	0	0	0	24	0	0	1	0	0	0	2	0	0	0	0	2	3	4	0	0	0
E	0	1	0	0	0	1	3	0	0	1	0	6	1	0	0	0	5	0	0	0	0	1	0	3	**131**	3	0	0	1	0
D	0	0	0	2	8	0	0	0	0	0	1	**97**	0	0	0	0	0	0	0	0	2	0	**127**	0	5	0	8	0	1	0
Q	14	3	0	1	1	0	7	0	2	0	5	0	0	16	7	18	0	0	0	0	1	0	11	0	3	15	0	1	0	0
N	6	0	0	2	5	0	5	2	8	4	1	0	0	0	1	0	0	0	0	1	3	0	4	0	1	0	**139**	0	0	1
T	6	4	2	16	**100**	0	0	0	12	**62**	7	2	6	1	8	6	1	1	1	2	1	0	4	0	0	8	1	0	0	1
S	16	20	11	10	2	16	2	18	1	17	12	16	3	0	0	3	1	0	0	1	**122**	1	0	0	0	6	1	2	1	4
C	1	8	3	**90**	3	**91**	2	9	1	3	12	13	2	4	0	13	0	0	0	3	10	0	0	0	0	2	0	0	1	0
W	0	0	0	0	1	1	1	0	0	0	0	0	0	2	8	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Y	1	0	14	1	14	8	5	1	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	4	0	0	0	0
F	1	2	1	4	0	9	0	2	0	0	0	2	0	8	0	5	0	1	17	0	2	0	0	0	0	0	0	0	0	0
L	21	**96**	20	10	4	3	27	1	1	9	**97**	5	6	24	3	**65**	**109**	**148**	**128**	43	0	0	0	0	0	14	0	0	0	0
I	**69**	0	0	8	0	0	1	3	1	3	1	0	7	5	2	29	1	3	0	**70**	4	0	0	27	0	2	2	0	0	0
M	9	0	0	0	1	0	2	0	0	7	1	0	1	0	0	0	0	4	4	0	0	0	0	0	0	1	0	0	0	0
V	2	2	1	0	8	1	1	1	14	17	14	0	7	13	0	0	2	3	1	31	0	17	3	**128**	0	0	0	0	1	0
A	5	2	5	0	10	2	1	2	**112**	20	1	7	**103**	1	4	1	13	1	0	4	5	0	1	3	0	3	0	1	2	2
G	3	0	0	3	0	1	**91**	2	4	2	0	2	0	0	14	2	0	0	0	0	6	**135**	11	0	12	0	0	0	**152**	0
P	0	1	**91**	7	2	0	0	2	1	3	3	0	0	3	2	6	6	0	8	1	0	0	0	0	0	10	0	**157**	0	**153**
	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161

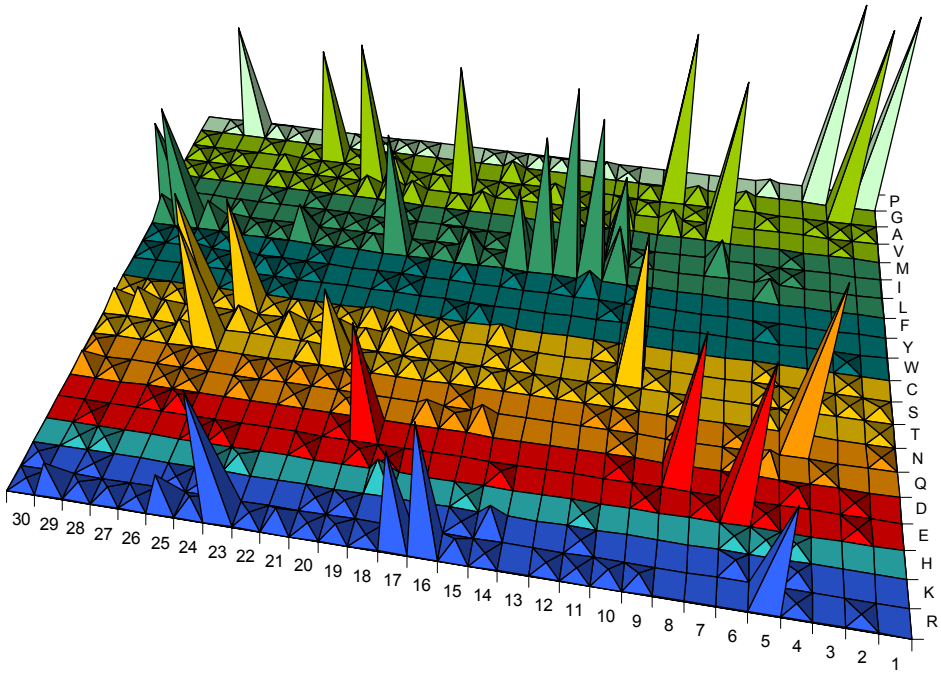


Figure 87. **Analisys of all the non-LTR 2A-like sequences within Trypanosomes and *Strongylocentrotus purpuratus***. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Cone graph, below, representing the results from the table. The data set analyzed is shown in Appendix 1, sequences from 449 to 585 and from 607 to 630.

	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
R	3	21	8	7	2	27	11	**110**	4	13	3	6	1	**84**	**112**	11	1	0	2	1	5	7	0	0	0	**91**	1	0	1	0
K	4	0	1	0	0	1	1	0	0	0	2	4	0	0	0	1	22	0	0	1	0	0	0	0	7	0	4	0	0	0
H	0	1	4	0	0	0	1	8	0	0	0	0	24	0	0	1	0	0	0	2	0	0	0	0	2	3	4	0	0	0
E	0	1	0	0	0	1	3	0	0	1	0	6	1	0	0	0	5	0	0	0	0	1	0	3	**131**	3	0	0	1	0
D	0	0	0	2	8	0	0	0	0	0	1	**97**	0	0	0	0	0	0	0	0	2	0	**127**	0	5	0	8	0	1	0
Q	14	3	0	1	1	0	7	0	2	0	5	0	0	16	7	18	0	0	0	0	1	0	11	0	3	15	0	1	0	0
N	6	0	0	2	5	0	5	2	8	4	1	0	0	0	1	0	0	0	0	1	3	0	4	0	1	0	**139**	0	0	1
T	6	4	2	16	**100**	0	0	0	12	**62**	7	2	6	1	8	6	1	1	1	2	1	0	4	0	0	8	1	0	0	1
S	16	20	11	10	2	16	2	18	1	17	12	16	3	0	0	3	1	0	0	1	**122**	1	0	0	0	6	1	2	1	4
C	1	8	3	**90**	3	**91**	2	9	1	3	12	13	2	4	0	13	0	0	0	3	10	0	0	0	0	2	0	0	1	0
W	0	0	0	0	1	1	1	0	0	0	0	0	0	2	8	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Y	1	0	14	1	14	8	5	1	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	4	0	0	0	0
F	1	2	1	4	0	9	0	2	0	0	0	2	0	8	0	5	0	1	17	0	2	0	0	0	0	0	0	0	0	0
L	21	**96**	20	10	4	3	27	1	1	9	**97**	5	6	24	3	**65**	**109**	**148**	**128**	43	0	0	0	0	0	14	0	0	0	0
I	**69**	0	0	8	0	0	1	3	1	3	1	0	7	5	2	29	1	3	0	**70**	4	0	0	27	0	2	2	0	0	0
M	9	0	0	0	1	0	2	0	0	7	1	0	1	0	0	0	0	4	4	0	0	0	0	0	0	1	0	0	0	0
V	2	2	1	0	8	1	1	1	14	17	14	0	7	13	0	0	2	3	1	31	0	17	3	**128**	0	0	0	0	1	0
A	5	2	5	0	10	2	1	2	**112**	20	1	7	**103**	1	4	1	13	1	0	4	5	0	1	3	0	3	0	1	2	2
G	3	0	0	3	0	1	**91**	2	4	2	0	2	0	0	14	2	0	0	0	0	6	**135**	11	0	12	0	0	0	**152**	0
P	0	1	**91**	7	2	0	0	2	1	3	3	0	0	3	2	6	6	0	8	1	0	0	0	0	0	10	0	**157**	0	**153**
	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161	161

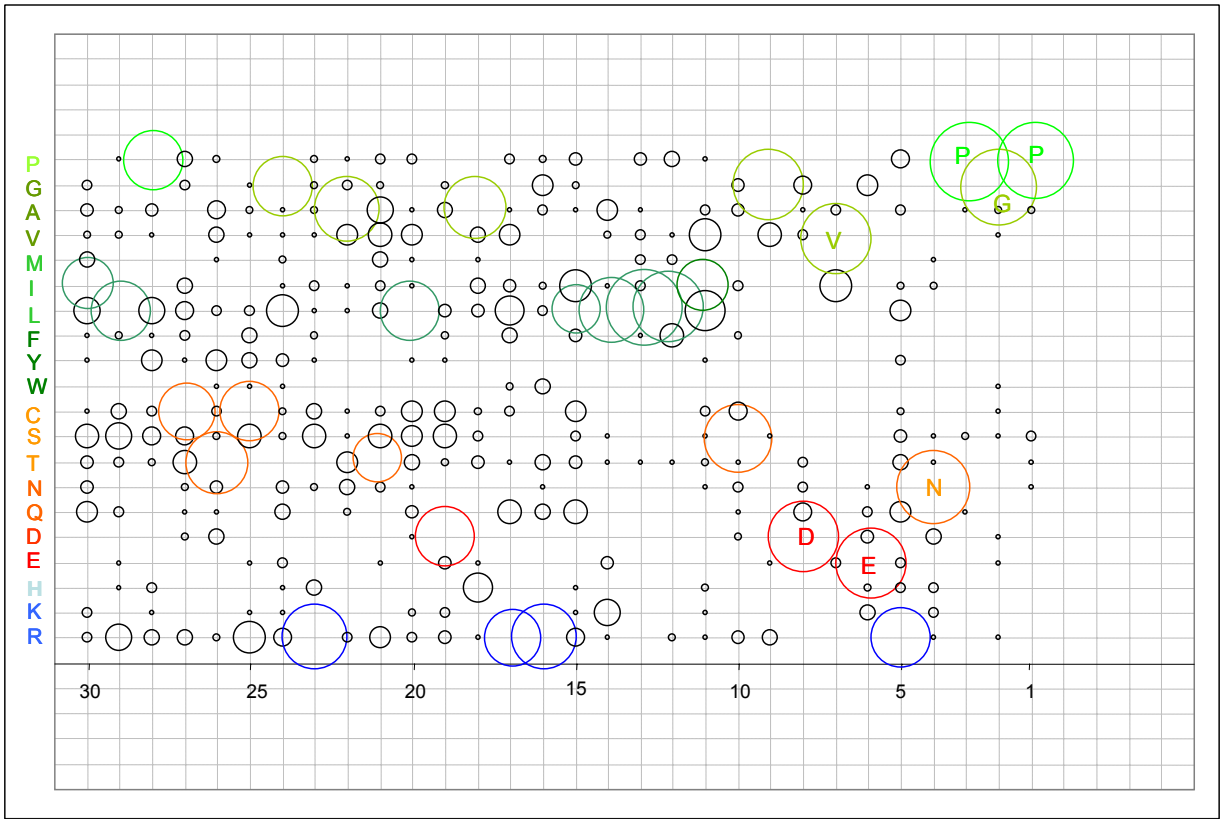


Figure 88. Bubble graph representing the data for all the non-LTR 2A-like sequences within Trypanosomes and *Strongylocentrotus purpuratus*. Above, table showing the amount of the different amino acids in each position of the 30 amino acids sequence analyzed. Below, bubble graph; the bubbles represent the amino acids in each position. The size of the bubbles is related with the amount of amino acids in each position. The predominant amino acids are shown in colour. The data set analyzed is shown in Appendix 1, sequences from 449 to 585 and from 607 to 630.

4. DISCUSSION

4.1 Development of a bacterial screen for 2A activity

This work is based on the observation that some nascent peptides may regulate translation by interacting with the ribosome exit tunnel during translation. This has not only been observed in eukaryotic systems (Delbecq, *et al*., 2000; Bachursky *et al*., 1994) but also in prokaryotic organisms (Harrod & Lovett, 1994; Nakatogawa and Ito, 2002). Some examples of nascent peptides that interact with the ribosome exit tunnel are presented in table 12 (adapted from Tenson and Ehrenberg, 2002).

These findings repudiate earlier assumptions of the ribosome exit tunnel being a 'neutral' path that does not interact with nascent peptides during translation. It is now clear that the ribosome plays an active role in sequence-specific gating of nascent peptides and in responding to cellular signals. These interactions not only affect protein elongation but also peptide termination (Lovett & Rogers, 1996). The molecular mechanisms of such interactions and of the ribosome response are currently unknown. Recent studies have, however, revealed elements of the tunnel which might be involved in 'sensing' the nascent-peptide sequence. In the case of prokaryotes, at approximately one-third of the tunnel length from the peptidyltransferase centre (PTC), the nascent peptide reaches a constriction formed by the tunnel walls formed by the extensions of two ribosomal proteins, L22 and L4, which are exposed in the lumen from opposite walls of the tunnel. In prokaryotes, this region seems to be the most crucial, whereas, in eukaryotes, other segments of the tunnel might also be involved in the ribosomal response (Rospert, 2004; Johnson, 2005).

Nascent peptides causing ribosome stalling			
Gene	*Organism*	*Active sequence*	*Co-effector*
Cat / cmlA	*eubacteria*	*VKTD/KNAD*	*chloramphenicol*
ermC	*eubacteria*	*SFVI*	*erythromycin*
tnaC	*eubacteria*	*KWFNID*	*tryptophan*
secM	*eubacteria*	*FXXXXWIXXXXGIRAGP*	*membrane trasnlocation (SecA)*
CPA1	*yeast*	*NSQYTCQDYISDHIWKTS*	*arginine*
arg	*fungi*	*PSXFTSQDYXSDHLWXAX*	*arginine*
AdoMet	*mammals*	*MAGDIS*	*spermidne*
Other active nascent peptides			
Gene	*Organism*	*Active sequence*	*Co-effector*
Gene 60	*bacteriophage T4*	*KYKLQNNVRRSIKSSSM*	*No co-effector*
ß-Tubulin	*eukaryotes*	*MREI*	*unknown*
Export signal	*universal*		*Signal recognition particle (SRP)*

Adapted from Tenson & Ehrenberg, 2002

Table 12. Examples of Active Nascent peptides

The motif FxxxxWIxxxxGIRAGP (x being any amino acid) was shown to induce elongation arrest in *E. coli* while SecM protein was being synthesized (Nakatogawa and Ito, 2002). This sequence also hindered translation elongation in *E. coli* when present in the unrelated sequence of LacZα protein (Nakatogawa and Ito, 2002), demonstrating that its action is independent of the sequence context around this motif. The proline (P), tryptophan (W) and isoleucine (I) were identified to be the main residues involved in ribosome stalling. P incorporation in the nascent chain allows W and I to reach the L22 ß-hairpin protein, triggering its rearrangement and thus inducing elongation arrest of the nascent peptide (Berisio *et al*., 2003).

Consistent with these studies are the findings made by Gong and Yanofsky (2002) on the *E. coli* tryptophanase (tnaC) operon, which stalls the ribosome at the last sense codon of tnaC in the presence of a high concentration of tryptophan. The model put forward by Gong and Yanofsky postulates that specific interactions of the nascent peptide with the exit tunnel generate a tryptophan-binding site at, or near, the A-site in the large ribosomal subunit. Binding of tryptophan at this site might hinder functions of the PTC, thus, preventing termination and release of the nascent tnaC peptide from the tRNA. Mutational analyses have revealed P24, the C-terminal proline, K11 and W12 to be involved in the ribosome stalling. K11 could be cross-linked to A750 in the loop of the helix 35 of 23S rRNA, which is located at the PTC and contains the peptidyl bond synthesis activity at the central loop in domain V (Nissen *et al.*, 2000). In contrast,W12 may interact with L22 once the P24 is being translated, suggesting the importance of the position of W and also the spacing between P and W, which seems to be crucial for elongation arrest of the nascent peptide. Mutations near the tip of the ß-hairpin of protein L22 alter the stalling effects (Berisio *et al*., 2003). All these findings suggest that the exact position of the nascent peptide in the tunnel might depend on the peptide sequence. Therefore, having the right residues at the right positions on 2A sequence, we may be able to create the right interactions between the 2A oligopeptide and the prokaryotic ribosomal exit tunnel, which would lead to the 'ribosome skipping' mechanism observed in eukaryotes.

A remarkable point is that in both tnaC and SecM, stalling occurs at the proline, underscoring the importance of the C-terminal P residue, which, in the stalled complex resides in the peptidyl-transferase active site and is still linked to the tRNA moiety. Furthermore, in both tnaC and SecM nascent peptides, a tryptophan residue, located 11 and 10 amino acids from the C-terminal proline, in tnaC and SecM,

respectively, is crucial for stalling. Mutations in the loop of helix 35 of 23S rRNA or in the tip of the L22 ß-hairpin, abolish ribosome stalling at both nascent peptides. However, a mutation at A2058, which is also located in the tunnel, has a profound effect on the ribosome response to SecM, but it does not greatly affect the tryptophan-induced stalling (Mankin, 2006). Potentially, different placement of SecM and tnaC in the exit tunnel might explain the involvement of different rRNA and protein positions in the ribosome stalling. It must also be mentioned that these two nascent peptides stalling action leads to different outcomes. Sec M stalling interferes in elongation of translation, whereas tnaC stalls at the last sense codon and fails to terminate translation (Mankin, 2006).

It is an interesting fact that the W residue is not present in any region of the FMDV 2A oligopeptide or the majority of Eukaryotic 2A-like sequences, excepting some insect and trypanosome 2A-like sequences, where the W is in a minority of these sequences (see section 4.5, Bionformatic analysis). This suggests that W is disfavoured residue for these nascent peptide-ribosome interactions: it may, however, play a role in these interactions in some insect ribosomes or at least does not hamper 2A mechanism of action. It should be remarked that eukaryotic ribosomes are not only different from prokaryotic ribosomes but also present distinctive features among different origins (insect, trypanosomes, mammals, yeast, etc).

Could the fact that FMDV 2A does not comprise W residues in the sequence disfavour the interaction of 2A nascent peptide with the prokaryotic ribosome exit tunnel? It could that be one of the reasons for lack of 2A activity in prokaryotic systems, having seen these two striking examples of two sequences (SecM, tnaC) containing W in conserved positions, and being able to stall prokaryotic ribosomes?

A good approach to test the effect of W residues in the interaction between 2A and the ribosome would be to change several amino acids from the upstream 2A region, for W residues to see the effect in 2A's activity and compare it between eukaryotic and prokaryotic systems. However, having the right interactions may not be the only important factor involved in 2A's activity. Other factors that seem to be involved are also discussed here.

4.1.1 Predicted model of 2A oligopeptide within the ribosome

2A oligopeptide predicted model structure shows 2A forming an N-terminal α-helix with a tight-turn at its C-terminus (Ryan *et al*., 1999). This is consistent with previous studies, where analysis searching for sterically allowed conformations within the ribosome exit tunnel predicted a unique solution: the α-helical conformation. The α-helix possesses a required rigidity for pushing the nascent peptide through the ribosome during translocation and it is also geometrically the most suitable conformation, being the most self-saturated structure with hydrogen bonds and, thermodynamically, the most stable single-stranded structure, universal for any amino acid sequence (Lim & Spirin, 1986).

The mechanism by which ribosomes function as molecular machines is one of the paramount unsolved problems of molecular biology. Advancements toward ribosome structure at high levels of resolution hold great promise of providing a structural basis of understanding the details of the reactions that are involved. A tunnel, in the large ribosomal subunit, was first detected in *Bacillus stearothermophilus* ribosomes (Yonath *et al.*, 1987). The tunnel was estimated to have a diameter of about 24Å and a length of 100-120Å, starting from a point at the region in which the PTC has been located (Yonath *et al*., 1987), and predicted to be able to accommodate 40 amino acids (Hardesty & Krammer, 2001). The PTC is located at the interface side of the 30 and 50 S ribosomal subunits close to the base of the central protuberance (5S RNA) (Oakes *et al*., 1990).

These findings are consistent with the 2A structural model, where the length of the proposed helical region of 2A sequence (~15aa) is some 27Å, which could be entirely accommodated within the exit tunnel of the ribosome.

The upstream region of 2A and 2A-like sequences form an essential interaction within the ribosome exit tunnel creating an environment that promotes 2A 'ribosome skipping' activity. The C-terminal –NPG- residues of 2A could play a role in the reorientation of the peptidyl tRNA-glycine substrate to disfavour peptide formation and then stimulate hydrolysis of the linkage between the nascent peptide and its tRNA when the A site is occupied by prolyl-tRNA.

The upstream helical part of 2A interacts with the exit pore and this promotes specific orientation of the base of the helix (the tight-turn) within the PTC of the

ribosome. Thus the NPG portion with its tight turn structure reorients the peptide-tRNA ester linkage promoting an unusual conformation: one different to that required for peptide bond formation. The ester bond between the nascent peptide and its tRNA is hydrolysed and the protein released. Therefore, not only the –DxExNPGP- motif is required for a 2A-like sequence to be active, but also the upstream region is critical for the functionality of the 2A protein. This model represents another *modus operandi* of viruses in their way to modify the host cell´s translation process for their own ends.

Suboptimal functioning of 2A could be due to the loss of interactions between the nascent peptide and the ribosome since the length of some tested 2As is not optimum. These interactions are at specific positions within the ribosome and re-orient the tight-turn (NPG) at the base of the helix, preventing the peptide bond formation. If the upstream sequence does not have the appropriate composition to allow these nascent peptide-ribosome interactions, then 2A's action will not be optimal, and that is why uncleaved protein has been obtained in past experiments using sub-optimal lengths of 2As (Donnelly *et al*., 2001a). Increasing the length of the FMDV-2A and 2A-like sequences, by extending it N-terminally including regions of the 1D protein, improves the efficiency of 2A 'cleavage' (Donnelly *et al*., 2001b; Luke *et al*., 2008)

Having the correct conformation of 2A is an essential factor for 'ribosome skipping' to occur but other factors must also be taken into account, such as the release factors, which have been recently found to be an integral part in this 'cleavage' event (Doronina *et al*. 2008). It should be pointed out that prokaryotic and eukaryotic ribosomes present different features - discussed below.

4.1.2 Involvement of the Release factors in 2A's mechanism of action.

Our model predicts that peptidyl(2A)-tRNA occupies the P site of the ribosome and prolyl-tRNA in the A site – but that the peptide bond cannot be formed. The orientation of the ester linkage (electrophilic centre) of the peptidyl-tRNA in the P-site preventing it from attack by the nucleophile. Normally, once the aminoacyl-tRNA binds into the A site the ribosome undergoes specific conformation changes,

changing the orientation of the α-amino group from the A site aa-tRNA and the carbonyl carbon from the P site peptidyl-tRNA thus allowing peptide bond formation and hydrolysis of the peptydil-tRNA (Nissen *et al*., 2000). This suggests that this change of conformation of the ribosome, once the A site is filled, makes the peptidyl-tRNA susceptible to hydrolysis. The first suggestions about 2A's mechanism of action mentioned the possibility of the C-terminal –NPG- residues, which form a tight turn, promoting a specific 2A peptidyl-tRNA substrate orientation that disfavours peptide bond formation and enhances hydrolysis of the peptidyl-tRNA ester bond (Ryan *et al*., 1999). In this case the A site is not empty but occupied by a prolyl-tRNA, which will form the N-terminal of the protein downstream 2A. The first hypothesis suggested that hydrolysis of the ester bond between the nascent peptide and the tRNA was mediated by the ribosome or that 2A may act itself as a hydrolytic element, activating a water molecule (Ryan *et al*., 1999). However, the recent findings of the involvement of the release factors in 2A's action add a new insight in the model of 2A's mechanism of action, where eRF1 and eRF3 promote hydrolysis of the 2A peptidyl-tRNA ester linkage.

The new model predicts that (i) due to there is no peptide bond formation between 2A peptidyl-tRNA and prolyl-tRNA, eventually the latter dissociates, allowing ingress of eRF1/eRF3 into the A site, which promotes hydrolysis of 2A peptidyl-tRNA ester linkage. Thus, the upstream (in the case of FMDV – the capsid proteins) domain plus the C-terminal 2A are released. Subsequently the release factors egress the A site and there is ingress of prolyl-tRNA. Two outcomes are possible at this stage, the dissociation of the unstable translation complex and, thus, termination of translation or translocation of prolyl-tRNA into the P site followed by ingress of the next aminoacyl-tRNA into the A site, thus, synthesis of the downstream protein domains. In the case of FMDV, the replication proteins.

The mechanism of translation termination has been widely studied, and it still remains uncertain how the three stop codons (UAA, UGA, and UAG) are recognized and how the process of termination and ribosome dissociation occurs. The release factors (RFs) are responsible for the recognition of these three stop codons and the hydrolysis of the nascent peptide. There are two types of RFs, codon specific class I and codon non-specific class II. Prokaryotes have two class I RFs, RF1, specific for

UAG and/or UAA, and RF2, specific for UGA and/or UAA) whereas eukaryotes have only one class I factor (eRF1) that recognizes all three stop codons. Class II factors RF3 and eRF3, which are GTPases, are present in prokaryotes and eukaryotes, respectively. RF3 is not, however, present in small-genome bacteria or Archaea. eRF3 function is not yet understood but is essential for cell growth, in contrast with RF3, which is not necessary in cell growth but is known to recycle class-I RFs.

In response to a stop codon in the ribosomal A-site, formation of a quaternary complex comprising the ribosome, eRF1, GTP and eRF3 triggers GTP hydrolysis and enhances the rate of peptidyl release (Frolova *et al*., 1994). Unlike eRF3, RF3 is non-essential and does not form stable complexes with the RF1 and RF2, implying these latter RFs can function in the absence of RF3 (Grentzmann *et al.,* 1994). RF3 may stimulate termination efficiency by ejecting the class I release factors (RF1/2) from the A site following peptidyl-tRNA hydrolysis (Freistroffer *et al*., 1997). The way eRF3 acts is still unclear but some observations suggest that it might act as a complex with eRF1 (Zhouravleva *et al*., 1995). However, some contradictory studies have shown eRF1 alone having peptidyl-release activity *in vitro* (Bertram *et al*., 2001). It may, therefore, be unnecessary to form an eRF1-eRF3 complex in order to terminate translation, although it remains possible that eRF3 may still have affinity for ribosome-eRF1-stop codon complex.

Termination of translation is induced by recognition of the stop codon in the mRNA by class I RFs, which promotes hydrolysis of the ester bond between the nascent peptide and the tRNA on the P site of the ribosome. The release of the nascent peptide is followed by the disassembly of the post-termination complex. This is achieved by the ribosome recycling factor (RRF) and the elongation factor EF-G in prokaryotes (reviewed by Nakamura and Ito, 2003). Eukaryotes do not encode a RRF homologue, and their mechanism of ribosomal recycling is unknown. Recent studies in eukaryotic recycling have shown that the initiation factors eIF3, eIF1, eIF1A, and eIF3j, a loosely associated subunit of eIF3, can promote recycling of eukaryotic post-translation complexes. eIF3 is the principal factor that promotes splitting of posttermination ribosomes into 60S subunits and tRNA- and mRNA-bound 40S subunits. Its activity is enhanced by eIFs 3j, 1, and 1A. eIF1 also mediates release of P site tRNA, whereas eIF3j ensures subsequent dissociation of mRNA (Pisarev *et al*., 2007).

4.1.3 Eukaryotic and prokaryotic release factors

Eukaryotic and bacterial RFs exhibit little conservation when their sequences are aligned. However, RF1 and RF2 share highly conserved regions in their primary sequences. Further, they all contain Gly-Gly-Gln (GGQ motif), which is essential for peptidyl-tRNA hydrolysis (Frolova *et al*., 1999). It was shown that the tripeptides Pro-Ala-Thr (-PAT-) in RF1 and Ser-Pro-Phe (-SPF-) in RF2 are responsible for RF specificity. Hence, the first and third amino acids of these tripeptides 'anticodons' discriminate the second and third purine bases, respectively (Ito *et al*., 2000).

The human eRF1 and *Escherichia coli* RF2 crystal structures were solved at high resolution (Song *et al*., 2000; Vestergaard *et al*., 2001) and revealed substantial structural differences. This is remarkable since they carry out similar tasks, although in different organisms. eRF3 and RF3 show limited sequence similarity to one another at the amino acid level, and while the eRF3 C-terminal domain shows most sequence similarity to eukaryotic elongation factor eEF-1 (which brings aminoacyl tRNA to the A-site), RF3 is most similar to prokaryotic elongation factor EF-G (the ribosome translocase), implying their precise functions may differ (Bertram *et al*., 2001).

Much more detailed biochemical evidence is required before the mechanism of eukaryote termination is elucidated. It may be possible that eRF3 enhances peptidyl-release. Interestingly, the recent determination of the genome sequence from the ciliated protozoan *Tetrahymena thermophila* showed it encodes an eRF1 protein highly homologous at the amino acid level to other more complex eukaryotic and archaeal release factors (Karamyshev *et al*., 1999).

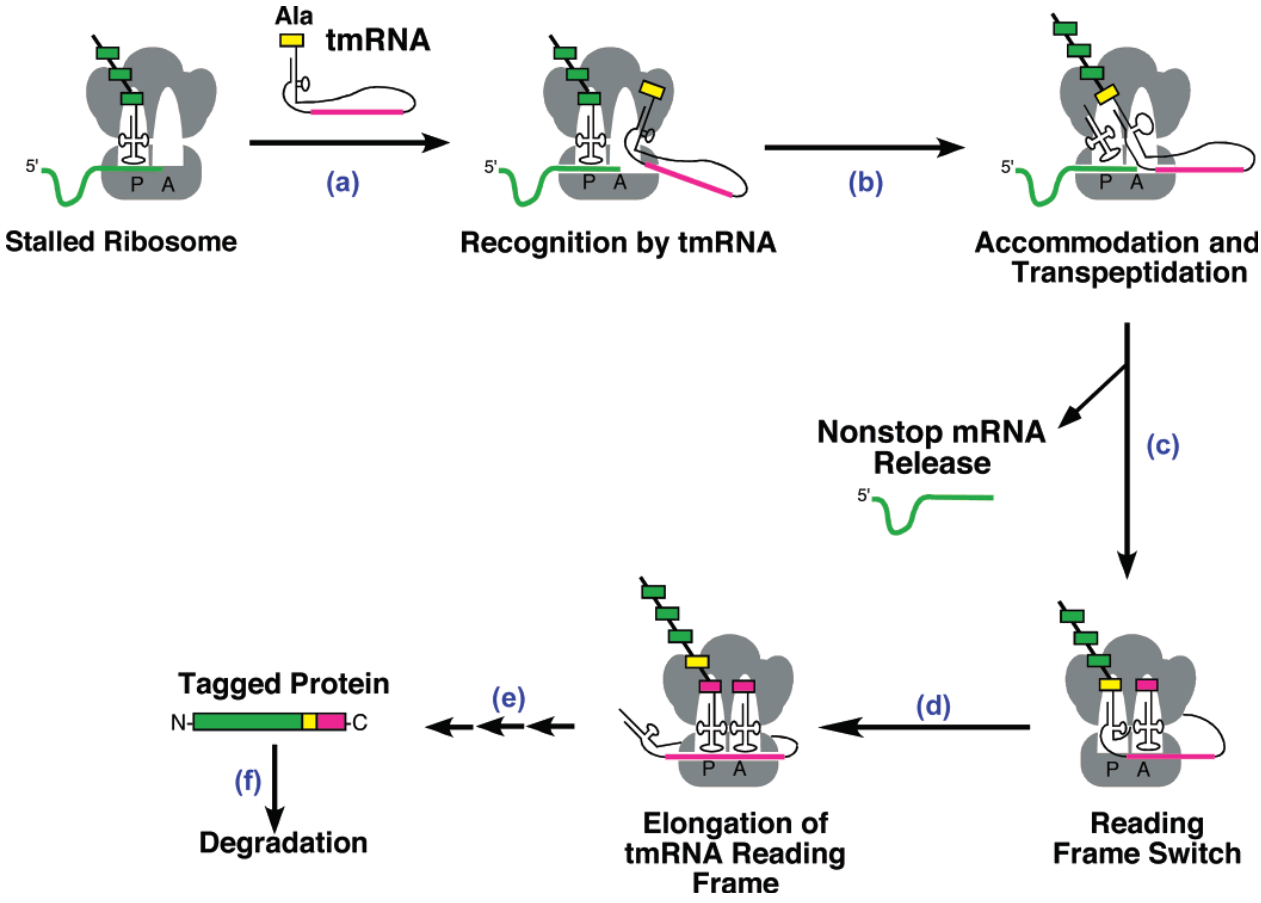
4.1.4 TmRNA rescue system versus Release factors function in 2A's mechanism of action?

Transfer-messenger RNA (tmRNA), also known as ssrA and 10Sa RNA, is a unique hybrid between tRNA and mRNA that is present in prokaryotes, which recycles 70S ribosomes stalled on problematic mRNA and also contributes to the degradation of incompletely synthesized peptides (reviewed by Gillet & Felden, 2001). A complex of alanyl-tmRNA, SmpB protein, and EF-Tu·GTP binds stalled ribosomes; the nascent peptide is transferred to the alanine on the tmRNA, and translation switches from the original message to a short tmRNA ORF that encodes a degradation tag. Translation of the ORF and normal termination releases the tagged polypeptide for degradation and permits disassembly and recycling of ribosomal subunits for new rounds of protein synthesis. Archaea and eukaryotic cells lack tmRNA (illustrated in figure 89, adapted from Dulebohn *et al*., 2007).

In eukaryotes deadenylated mRNA is degraded by the exosome, although the fate of the ribosome on a truncated mRNA (and the associated polypeptide) remains to be elucidated (reviewed by Moore & Sauer, 2007).

Recently, a novel mRNA decay pathway called No-Go decay (NGD) has been identified in *Saccharomyces cerevisiae*. This pathway releases ribosomes stalled in translation due to the presence of a stable stem-loop within the mRNA and this complex triggers direct endonucleolytic cleavage of aberrant mRNA. It involves a protein called Dom34, which is related to eRF1, and physically interacts with another protein Hbs1 that is itself related to eRF3 (Clement & Lykke-Andersen, 2006). The crystal structure of Dom34 has been resolved revealing two domains structurally homologous to eRF1 (Graille *et al.,* 2008). This could mean that they both interact with the ribosome in similar ways as they both trigger ribosome release by catalysing chemical bond hydrolysis.

The similarities of these two factors reinforce the new 2A's model mechanism of action, where a stalled ribosome is recognized and released by eRF1 (eRF3). Therefore, suggesting that these eukaryotic release factors may also be part of a ribosome rescue system similar to that observed in prokaryotes (tmRNA) and, recently, in yeast.



From Dulebohn et al., 2007

Figure 89. TmRNA-mediated protein tagging and ribosome rescue. A ribosome stalls on an incomplete or untranslatable message, leading to (a) the recruitment of aminoacylated tmRNA to the ribosomal A site and (b) transfer of the nascent chain to the alanine-charged tRNA-like domain of tmRNA. A message-switching event (c) then replaces the faulty mRNA with an open reading frame within tmRNA (d), which is translated until a stop codon is reached (e) and the tagged protein is released and degraded by C-terminal specific proteases (f).

It could well be, therefore, that viruses employing 2A as part of their replication strategy have utilised aspects of general mechanism that has evolved in eukaryotes to release ribosomes stalled on truncated mRNAs. In the case of 2A, in the middle of an ORF, this event occurs not at a stop codon but in the midst of the ORF such that translation can either (i) terminate, or, (ii) essentially pseudo re-initiate to synthesise the downstream sequences

4.2 Purification of TMEV-2A

2A is only 18aa long in aphthoviruses whereas in cardioviruses 2A is ~150aa. In spite of this difference, their C-terminal regions are similar and posses the same self-cleaving activity (Donnelly *et al.*, 1997).

The high level of cleavage activity of the C-terminal region in cardioviruses, without the need of the upstream sequence raises interesting questions, Why do cardioviruses posses a 2A region of 150aa whilst those of aphthoviruses are a mere 18aa? What is the function of the remaining -85% of the cardiovirus 2A sequence?

The role of the upstream sequences has been previously highlighted (Ryan *et al.,* 1991) and several experiments have been performed to determine the influence of these upstream sequences on aphthovirus and cardiovirus 2A activity (Donnelly *et al*., 1997). It is most improbable that the only function of cardiovirus 2A is to enhance the C-terminal region's cleavage activity to 100% since FMDV 2A can achieve that with a total length of 30-55 aa. Furthermore, previous studies demonstrated that deletions of two-third of the N-terminal region of cardiovirus 2A had no effect on the cleavage of 2A/2B junction (Palmenberg *et al*., 1992).

Another role, other than the separation between the capsid and replicative domains, could be invoked, similar to that shown for the $2A^{pro}$. The $2A^{pro}$ in enteroviruses and rhinoviruses not only cleaves the polyprotein at its N-terminus, separating the capsid precursor from the replicative domain, but also mediates host cell shut off by cleaving eIF4GI (Zamora *et al*., 2002). In FMDV this activity is found not in the 2A region but in the N-terminal protease unique to aphthoviruses, L^{pro} (Lloyd *et al*., 1988). Therefore, FMDV 2A's sole activity appears to be separating the capsid protein precursor form the replicative domain whilst the longer 2A found in cardioviruses may have an additional role other than the primary cleavage of 2A/2B. Such a function could be cis-acting in capsid protein processing, encapsidation or, trans-acting, in replication once it is cleaved from 1D.

Not much is known about the biological functions of cardiovirus 2A. EMCV 2A has been found to have non-specific RNA-binding properties and to be associated with ribosomes isolated from infected cells (Medvedkina *et al.,* 1974).

Recent studies reported EMCV-dependent changes in 4EBP1 phosphorylation patterns and a link between these changes and the presence of cardiovirus 2A (Svitkin *et al*., 1998). 4EBP1 is a regulator of eIF4E availability and direct competitor of eIF4G-eIF4E interactions. In BHK cells infected with EMCV 2A the mammalian target of rapamycin (mTOR) protein kinase pathway is activated, leading to dephosphorylation of 4EBP1, inactivating 4E. Deletions within EMCV 2A abrogated host protein shutoff and, additionally were partially defective in viral protein processing (Svitkin *et al*., 1998).

More recent studies in HeLa cells infected with EMCV show EMCV 2A tightly associated with some of the free 40S ribosome subunits, but not present in the 80S pool which accumulated after infection. Expression of 2A protein in cells in the absence of infection was able to modulate the cellular translational environment to increase the ratio of internal ribosome entry site-dependent translation to cap-dependent translation of a reporter construct. These results provide further evidence for a role of 2A protein in the mechanism of cardiovirus induced host translational shutoff although the mechanism by which 2A affects this response remains elusive (Groppo & Palmenberg, 2007).

Important goals for future work must include studies on the localization of TMEV 2A within the cell. The production of antibodies against TMEV 2A will, therefore, be very useful to resolve the yet unknown localization of 2A once it is separated from 1D, and to provide an insight into its possible role in the inhibition of host´s cells protein synthesis (reported for EMCV 2A). Bacterial expression will allow another important goal to be achieved: the resolution the atomic structure of 2A by nuclear magnetic resonance (NMR) spectroscopy.

4.3 Creation of a reporter of cellular stress using $2A_{FMDV}$ protein

Protein synthesis consumes a large amount of energy in the cell, with the highest proportion used in peptide-chain elongation. Therefore, under conditions of temporary increased energy demand or reduced energy supply, the cell has regulatory mechanisms to reduce the rate of protein synthesis allowing energy to be diverted to other processes, such as maintaining the plasma membrane potential and ion gradients.

The elongation cycle of protein synthesis starts with the ribosome bound to the mRNA with peptidyl-tRNA in the P site, and aminoacyl-tRNA binding to the A site in a codon-dependent manner. This reaction requires an elongation factor (EF-Tu and eEF-1 in prokaryotes and eukaryotes, respectively) and GTP, which together form a high-affinity ternary complex with the A-site aminoacyl-tRNA. Codon recognition triggers GTP hydrolysis followed by release of the aminoacyl-tRNA from the elongation factor-GDP complex. The aminoacyl-tRNA is accommodated in the PTC where transpeptidation occurs. A second elongation factor (EF-G and eEF-2 in prokaryotes and eukaryotes, respectively) catalyses the translocation of the peptidyl-tRNA from the A site to the P site, with hydrolysis of GTP. The deacylated tRNA is then transferred from the P site to the E site, from where it dissociates. Thus, translocation restores the initial state of the ribosome, which then enters the next elongation cycle unless a stop codon in the A site cause termination (reviewed by Rodnina *et al*., 2000).

Under environmental stress, such as temperature changes or viral infections, the cell can regulate these factors by phosphorylation. EF-2 is a monomeric protein with a mass of ~93 kDa, which binds guanine nucleotides and is active when bound to GTP. Phosphorylation of eEF-2 inhibits its activity by preventing it binding to the ribosome (Carlberg *et al.,* 1990). eEF-2 is phosphorylated in a Ca^{+2}/calmodulin-dependent manner (Ryazanov, 1987) by the Ca^{+2}/calmodulin-dependent kinase III, which is known as eEF-2 kinase (Nairn & Palfrey, 1987). The activity of this kinase is also regulated by cAMP-dependent protein kinase (PKA), which is activated under cellular conditions of increased energy demand (Mitsui *et al*., 1993). Different types of cellular stress (osmotic, oxidative or heat stress) has been shown to cause inhibition of peptide synthesis at different levels (initiation, elongation and termination). For

instance, phosphorylation of eEF-2 occurs with oxidative stress of the cell (Patel *et al.,* 2002). As said before, translation elongation is expensive in terms of metabolic energy, thus, may be inhibited when the cellular energy supply is restricted, for instance under stressing conditions, such as, viral infection leading to the activation of stress signaling pathways within the cell (reviewed by Browne & Proud, 2002).

eEF-2 levels have an effect on the outcome of translation products produced upstream and downstream of the 2A oligopeptide. Following release of the nascent peptide by eRF1/3, deacylated tRNA is present in the P site. The model proposes that these release factors leave the A site allowing prolyl-tRNA to bind. For continued synthesis of downstream sequences to occur, the deacylated tRNA needs to be translocated from the P to E site concomitant with proly-tRNA translocated from the A to P site. eEF-2 is responsible for this translocation reaction, thus if its activity is inhibited, further elongation cannot occur: ribosomes dissociate and termination occurs (Ryan *et al*., 1999). This is consistent with the previous observations of Svitkin and Agol (1983) where a 'translational barrier' prevented the synthesis of translation products downstream of EMCV 2A. These translational studies of cardiovirus RNA were made in Krebs-2 cell-free extracts, which were found to have low levels of active eEF-2. When eEF-2- was added to the cellular extracts, enhancement of the translation of the downstream of 2A was achieved. This may be because, as with FMDV 2A, eEF-2 promotes translocation of the prolyl-tRNA from the A to the P site thus allowing synthesis of the downstream products.

As previously mentioned, one of the innate responses of cellular stress is the phosporylation of eEF-2, to inhibit the level of translation. Viruses may have adapted to this situation, taking advantage in a very subtle manner. Therefore, the effects of eEF-2 levels on 2A-mechanism may be favouring the virus production, when in the latter stages of infection (increased stress) only the capsid proteins are produced. This would allow the generation of the large excess of capsid proteins (60 copies per viral genome) required for encapsidation, avoiding the further production of the replicative proteins. Indeed, preliminary data indicate that whilst the level of eEF2 remains constant through TMEV infection, the proportion of *phosphorylated* eEF2 increases throughout infection (M. Ryan, pers. comm.)

The creation of the construct [V5-TTA-2A-scFv] (TTA: transcriptional transactivator) from pHE27 may open up a new approach in the use of 2A oligopeptide, which had already been widely used to co-express genes of interest with

reporter proteins in biotechnology and gene therapy (Percy *et al.,* 1994; Chaplin *et al.,* 1999; de Felipe et al., 1999; reviewed by Szymczak & Vignali, *et al.*, 2005).

This artificial reporter may allow us to detect the level of stress in the cell. As explained above one of the cellular outcomes in stressful conditions is the phosphorylation of eEF-2, thus, inhibiting translocation. In normal conditions, the translation of [V5-TTA-2A-scFv] leads to two products: [V5-TTA-2A] and [scFv] in equimolar quantities. The downstream product [scFv] will bind the V5 from the upstream protein, and, therefore the complex is too big to pass through the nuclear pore and stays in the cytoplasm. In contrast, when the cell is stressed and, thus, translocation is repressed, there will be an excess of the upstream product compared to the downstream one. This will, consequently, allow the free [V5-TTA-2A] protein to pass through the nuclear pore since it is not bound to [scFv]. TTA activates gene expression, thus, this will be used in cell lines expressing luciferase under the control of tetracyline, where the amount of expressed luciferase will represent the levels of stress in the cell.

This system will allow us to measure stress in the cell in a range of conditions of stress: temperature, pH, osmolarity, nutrient limitation, infection etc. This artificial reporter will also be a very useful tool for cell biology and biotechnology and will shed important new light on the cellular stress studies not only in animals but in plants.

4.4 2A-like sequences

First identified and characterized in picornaviruses, 2A-like sequences are found in other mammalian viruses and a wide range of insect viruses. These naturally occurring 2A-like sequences have been recognized by database searches and, then analyzed with the intention to provide insights into the role of this oligopeptide and to determine the requisites of the N-terminal region.

4.4.1 Insect virus 2A-like sequences

Many insect 2A-like sequences have been identified and analyzed in the past 12 years. First analyses of 2A were performed using the "short" 2A version (18 aa) (Donnelly *et al*., 2001a) – the length of the first 2A sequence characterised. Later studies on FMDV 2A demonstrated that residues located upstream of 2A, located within the 1D capsid protein domain influenced 2A by enhancing the cleavage activity (Donnelly *et al*., 1997). Thus, recent experiments on insect 2A-like sequences have been performed using longer 2A versions (30aa), including upstream sequences (Luke *et al*., 2008). The results obtained with the longer versions show a clear improvement on 2A cleavage activity in the majority of species compared to the previous results using the shorter 2A version (18aa).

A first approach here was to compare 2A activity between two different *in vitro* systems: the wheat germ extract system and a new insect *in vitro* system (Qiagen Ltd., West Sussex, UK). Although translation profiles were routinely obtained using other *in vitro* translation systems, unfortunately (other than the positive control supplied by the manufacturer) no translation products were obtained using the insect translation system. The results obtained using the wheat germ extract systems were consistent with those seen in previous experiments (Luke *et al*., 2008). *Infectious flacherie virus* (IFV), *Varroa destructor virus 1*(VDV-1) and *Perina nuda picorna-like virus* (PnPV) belong to an unassigned genera of Iflaviruses, which are positive single stranded RNA viruses. IFV, which causes flacherie disease in the silkworm *Bombyx mori*, has the same genome arrangement as picornaviruses (see figure 18), 2A is situated in between the N-terminal structural protein domain and the replicative domain at the C-terminus of the polyprotein (Isawa, *et al.*, 1998). In IFV 2A-like sequence the –DxExNPGP- motif is not conserved, but differs from the consensus by

a single change, a glycine residue instead of an aspartate, -**G**xExNPGP-. It is, however, still active. It is likely that IFV 2A-like sequence has the same role as in picornaviruses, to separate the capsid protein from the replicative proteins during translation. Interestingly, when this glycine residue was mutated to be the canonical –**D**xExNPGP- motif, the sequence showed no cleavage activity (Donnelly *et al*., 2001a). This suggests that the 2A sequence requirements to be active are more complex than having the –DxExNPGP- motif alone. The upstream sequence of 2A may also have a major influence in the efficiency of cleavage. VDV has also the same genome structure than picornaviruses, although 2A consensus motif is not conserved, nor is this 2A-like sequence active. The –NPGP- motif is present but the upstream sequence is different (–**M**D**N**PNPGP-).

The rest of the insect viruses analyzed have a genome organization that is very different from picornaviruses. *Perina nuda picorna-like virus* (PnPV) has two 2A-like sequences: the first 2A is in the same position as picornaviruses, flanked by the capsid and replicative protein domains, whilst the second 2A lies between the structural VP2 and VP4 proteins (figure 18). Both 2As are highly active. Phylogenetic studies of 2A show that IFV is much more distantly related to the other members of the family with active 2As. There are two possible explanations; one where 2A would be acquired at early stage in evolution accompanied by divergence of 2A between IFV and PnPV/EoPV (acquisition of a second 2A in PnPV/EoPV and loss of 2A from the other lineages), or 2A would be acquired independently, one in the PnPv lineage and another one in the IFV lineage (Luke *et al*., 2008).

Cricket paralysis virus (CrPV), which belongs to the *Dicistroviridae* family, has a monopartite and bicistronic genome. This genome encodes two large non-overlapping ORFs with viral non-structural and structural polyprotein precursors encoded by the upstream and downstream ORFs, respectively (figure 18). Each ORF is preceded by an internal ribosome entry site (IRES), which control translation of the two different ORFs independently. It was observed that the downstream IRES was more active than the upstream IRES, providing a mechanistic explanation for the previously observed increased expression of viral structural proteins relative to the non-structural protein in infected cells. This genome organization and the utility of IRES elements of different strengths could be an economical strategy for gene expression (Wilson *et al.*, 2000). The 2A-like sequence is located in the N-terminal region of the replicative ORF1 and is able to mediate cleavage. However, 2A-like

sequence is less active in CrPV than the other members of the family, *drosophila C virus (DCV)* and *acute bee paralysis virus* (ABPV), where 2A is highly active (Donnelly *et al*., 2001a). Nevertheless, all were capable of mediating cleavage and this translational model of 2A activity would predict that the translation of the replicative proteins (ORF1) would result in a 'primary' N-terminal cleavage product of 96aa (DCV) and 166aa (CrPV and ABPV) (Donnelly *et al.*, 2001a).

Studies where 2As were aligned using ClustalX 1.81, showed clusters that do not necessarily corresponded to the same clusters formed when the RNA-dependent RNA-polymerase (RdRp) sequences were aligned. Whilst 2A appears to have been acquired at a relatively early stage in picornavirus evolution, the reverse seems to be the case in the dicistroviruses (Luke *et al*., 2008).

Providence virus (PrV), a member of the family *Tetraviridae*, infects midgut cell line derived from the corn earworm (*Helicoverpa zea*). The virus possesses a bicistronic genome with two N-terminal overlapping ORFs (ORF1 and ORF2) encoding the non-structural proteins and a third C-terminal ORF (ORF3), which encodes the structural proteins (Pringle *et al*., 2003). Three 2As can be found in its genome; $2A_{1,}$ within ORF1 and $2A_2$-$2A_3$ in ORF3 (figure 18). PrV $2A_2$ shows lower cleavage activity than $2A_1$ and $2A_3$. Appreciable levels of uncleaved polyprotein can be observed in *in vitro* experiment using PrV $2A_2$, similar to CrPV 2A (figure 63). Currently no data is available from PrV and CrPV infected cells. The tetraviruses could not be included in the polymerase based phylogenetic analysis since the domains in the polymerase of these viruses are 'shuffled' compared to all other positive stranded RNA virus polymerases.

The 2A-like sequences found in *Bombyx mori cypovirus type1* (BmCPV-1) and *Operopthera brumata cypovirus-18* (OpbuCPV-18), members of *Reoviridae*, are highly active and located in the segment 5 of its genome, encoding non-structural proteins. This family contains a segmented genome of 10 to 12 segments of linear double stranded RNA. Similar to dicistroviruses, 2A acquisition appears to have occurred at a relatively late stage in the evolution of this family (Luke *et al*., 2008).

4.4.2 Strongylocentrotus purpuratus 2A-like sequences

Active 2A-like sequences have been found not only in a wide range of RNA viruses but also in trypanosomal sequences, such as in *Trypanosoma cruzi (L1Tc)* and *T. brucei (igni)*, where 2A is associated with non-LTR retrotransposons (Donnelly *et al*., 2001a; Heras *et al*., 2006). 2A-like sequences have been also found in cellular genes such as the carbamyl-phosphate synthetase I (CPS1) protein of North American bullfrog *Rana catesbeiana* (with a histidine instead of an aspartic acid in the C-terminal conserved sequence (-**H**PEANPGP-), 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 hyperthermophilic bacterium *Thermatoga maritima.* All of them were analyzed and found to be inactive (Hughes, 2003). It seemed, therefore, that this method of controlling protein biogenesis was limited to viruses or retroelements.

The recent discovery of new cellular 2A-like sequences in the purple sea urchin genome, *Strongylocentrotus purpuratus* (Sea Urchin Genome Sequencing Consortium, 2006) gives us insight into the role of these 2A-like sequences. Some of these sequences have been analyzed and found to be active. More than 100 sequences have been found in the genome and have been subsequently classified in two major groups. The first group of 2A-like sequences is found in the genes encoding intracellular proteins involved in innate immunity: nucleotide-binding oligomerization domain/leucine-rich repeat family (NOD-LRR), NACHT domain/leucine-rich repeat (NACHT-LRR family) or 'CATERPILLER' (CARD, transcription enhancer, R (purine)-binding, pyrin, lots of LRRs) genes (Lich & Ting, 2007). CATERPILLER proteins have been recently discovered and form the second important family involved in inflammatory responses to pathogens, together with the mammalians toll-like receptor (TLR) family. They have not only been found in animals but also in the plant kingdom, in great numbers (~100 genes). They are greatly expanded in *Strongylocentrotus purpuratus* (~200 genes) and, surprisingly, 2A is found in ~50% of them, where it is located at the N-terminus of the ORF. The 2A sequence is mainly followed by a DEATH domain (DD; mediating protein:protein interactions), followed by a NATCH domain (predicted ATP/GTPase activity with similarity to apoptopic proteinase activating factor-1), then LRR (thought to bind variously pathogen proteins, lipids or carbohydrates) (figure 19).

The second major group of 2A-like sequences is associated with non-LTR retrotransposons and are also found forming or close the N-terminus of the single ORF. The non-LTR sequences downstream of 2A correspond to ORF2 of these elements (figure 19). Bioinformatic analyses of the sequences show three groups of non-LTR retrotransposons, all clustering within the chicken repeat (CR) 1 lineage of these elements, in contrast with L1Tc and *ingi* of trypanosomes. The first group of retroelements (~17 genes) are more similar to the Bf CR1 retroelement, the second group (~4 genes) most similar to ReO_6 and the third (~4 genes) most similar to Maui. The role of 2A within these elements remains unclear; it may form a genomic site favourable for their transposition.

Most of the 2A sequences analyzed so far appear to be active, and, as previously shown in a wide range of RNA viruses, 2A may also be playing a role in the regulation of the synthesis of these novel cellular polyproteins. Furthermore, it has been recently observed that CATERPILLAR 2A-like sequences act as signal sequences (personal communication, Dr. Ryan). Two artificial constructs, consisting of [CATERPILLAR 2A-Cherry fluorescent protein] and a point mutated form of the CATERPILLAR 2A were assembled. *In vitro* translation analyses showed the expected products, with ~90% 'cleavage'. When the wild-type 2A plasmid construct was transfected into HeLa cells (2A cleaving at its own C-terminus), Cherry fluorescent protein was localized in the cytosol. In contrast, the mutated, inactive, 2A resulted in the cherry fluorescent protein being localized to the ER and Golgi. Thus, 2A when is at the N-terminus of the Cherry protein ('uncleaved'), it acts as a signal sequence: is recognized by SRP and translocated into the ER co-translationally. If 2A mediates 'cleavage', then the downstream protein (lacking a signal) is located in the cytoplasm. Recently, the intriguing possibility of dual localization of membrane-targeted single translation products has emerged. Sorting of a protein to a membrane-sealed organelle involves targeting to the organelle, interaction with its surface receptors and translocation through the membrane by specific import machineries (post-translational import). Therefore dual distribution of a single translation product must reflect competition and/or promiscuity in one or more of these sorting steps. The cell uses different 'tricks' to achieve dual localization, such as competition between two signals on the same polypeptide, one ambiguous signal that is recognized by two organelles, inaccessibility of a signal by folding or protein binding, among others. If

the signal is hidden, this may lead to the retention of the protein in the cytosol (reviewed by Karniely and Pines, 2005)

The *Strongylocentrotus* CATERPILLAR 2A-like sequences represent another case of dual protein targeting, where 2A protein acts as an N-terminal 'self-cleaving' signal for the CATERPILLER protein. Presumably the proportions of the secreted / cytosolic forms of the protein depend upon the 'cleavage' activity of each particular 2A-like sequence. This range of 2A-like sequences is currently being investigated.

4.4.2 2A-like sequences within Dicistroviridae

The N-terminal peptide upstream of the 2A-like sequences within DCV, ABPV and CrPV of the *Dicistroviridae* is reported to be involved in the shut off of the innate immune system during infection (personal communication, Professor Martin Ryan). These proteins inhibit the formation of the RNA-induced silencing complex (RISC). Small interfering double-stranded RNAs (siRNAs) are part of the RISC formed as part of the innate immune response. This complex cleaves one of the strands of siRNA, leaving the other strand as a guide to bind complementary (target) mRNA sequences, which are then cleaved by Argonaute endonucleases and then, degraded (reviewed by Sioud, 2007).

Interestingly, other 2A-like sequences have also been observed in the genome of type C rotaviruses and *Penaeid shrimp infectious myonecrosis virus* (IMNV) as a 'linker' between these dsRNA binding motifs and various proteins (Luke *et a*l, 2008). In Rotaviruses the dsRNA binding motif is at the C-terminus of 2A (figure 90), and shares homology with the interferon-induced, double stranded RNA-dependent protein kinase (PKR) (Langland, *et al*., 1994). PKR, like many components of the innate immune system, depend on dsRNA for its activation. One of the virus strategies to repress the immune response consists in the ability to sequester dsRNAs, and thus, inhibit the activation of these effector proteins.

IMNV presents the dsRNA binding motif at the N-terminus of $2A_1$ (figure 90). This protein arrangement is also the case for many other 2A-like sequences, such as non-LTR sequences in Trypanosomes. The function of these short proteins upstream of the 2A-like sequences of trypanosome non-LTRS is not known, but it may be

considered that, by analogy with Dicistroviruses, suppression of the formation of the RISC may promote transposition.

2A may, therefore, represent a method of adding – 'bolting-on' - an extra-function to these viral genomes, which seems to have a common theme in the inhibition of the cellular innate immune response.

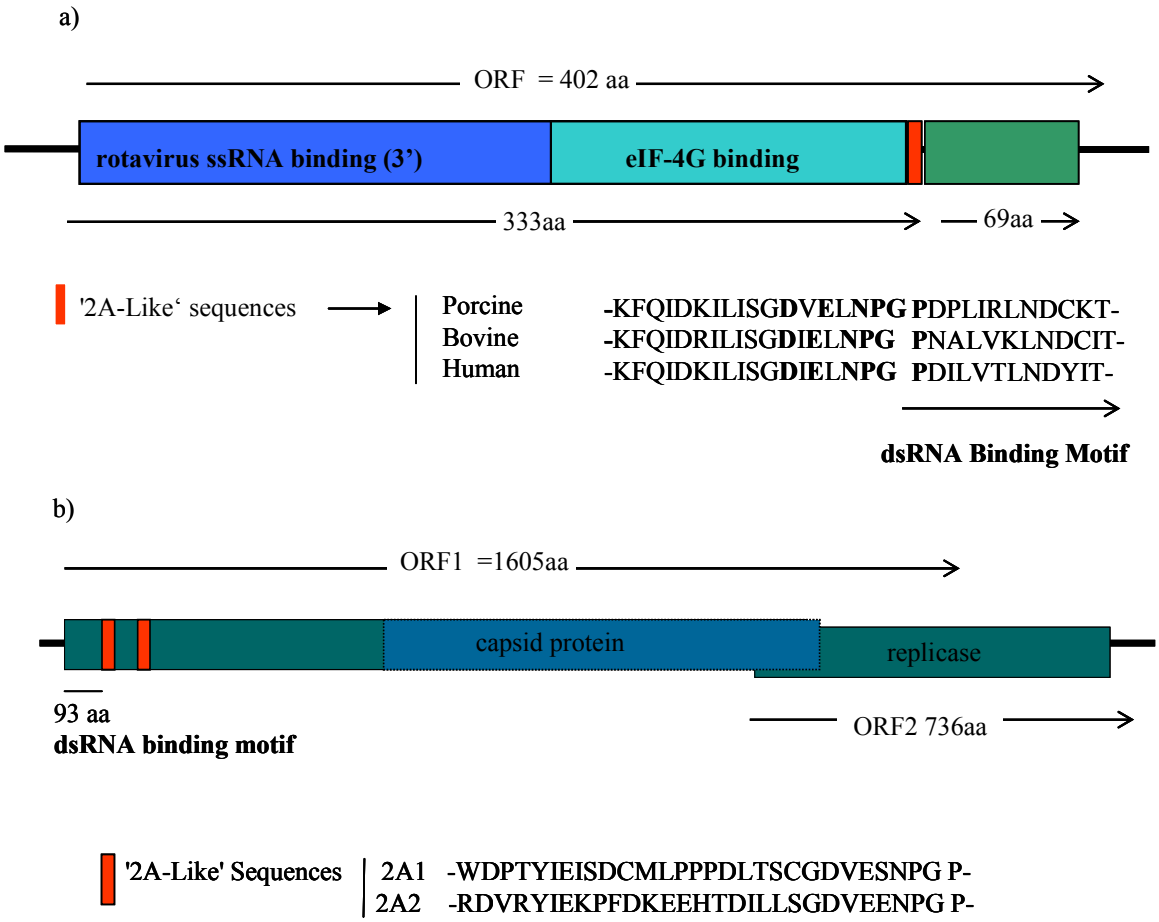


Figure 90. 2A-like sequences within dsRNA binding motifs. a) Type C Rotavirus Protein NSP3, segment 6. The segment is divided in three different parts, the ssRNA binding region, the eIF-4G region and the dsRNA binding fragment at the C-terminus of the 2A-like sequence represented in red. b) *Penaeid shrimp infectious myonecrosis virus* genome. The replicative domains are shown in green, and the structural domain in blue. The 2A-like sequences are represented in red and the dsRNA binding motifs are indicated

4.5 Bioinformatic Analyses

The 2A-like sequences found in the database searches containing the 2A motif are variable in the content upstream of the conserved motif. The analyses made, comparing this region between different species and different families has provided more knowledge about this region of 2A: those residues which are variable or conserved amongst different families of viruses.

As discussed in section 4.1 it is known that the ribosome exit tunnel is not neutral and interactions exist between the tunnel and nascent peptides. This tunnel is known to be able to accommodate approximately 40 amino acids in a helical conformation. This is consistent with the proposed and mapped interaction of 2A with the exit tunnel being ~30aa long (see section 4.1).

4.5.1 Picornaviruses

2A upstream sequences present a very similar pattern within picornaviruses. In spite of the fact that the range of amino acids is not conserved, there is a higher presence of hydrophobic amino acids at the N-terminus of the sequence except from the more upstream positions 27, 29 and 30, where there is a prominence of hydrophilic, positively charged, amino acids (table 13).

Tryptophan is avoided in all 2A sequences except SVV and ERBV-2. ERBV has been tested and is active (Donnelly *et al*., 2001a), whereas SVV has yet to be tested.

The similarities and differences between sequences and reasons for the presence of certain conserved residues on the N-terminal sequence will be discussed below.

Species	Accession number	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
FMDV-63	AY593836	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
ERAV-393/76	L43052	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	S	L	L	K	L	A	G	D	V	E	S	N	P	G	P
BRV2	EU236594	L	R	L	T	G	E	I	V	K	Q	G	A	T	N	F	E	L	L	Q	Q	A	G	D	V	E	T	N	P	G	P
ERBV-1-P1436/71	X96871	E	A	T	L	S	T	I	L	S	E	G	A	T	N	F	S	L	L	K	L	A	G	D	V	E	L	N	P	G	P
ERBV-2-P313/75	AF361253	V	A	D	W	E	N	L	L	S	Q	G	A	T	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
PTV-1-Sek 549/98	AF296101	A	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
EMCV (EMC-PV21)	X74312	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
TMEV (GDVII)	X56019	F	R	E	F	F	K	A	V	R	G	Y	H	A	D	Y	Y	K	Q	R	L	I	H	D	V	E	M	N	P	G	P
SAF-V	EF165067	F	T	D	F	F	K	A	V	R	D	Y	H	A	S	Y	Y	K	Q	R	L	Q	H	D	V	E	T	N	P	G	P
LV (174F)	AF327921	Y	F	N	I	M	H	S	D	E	M	D	F	A	G	G	K	F	L	N	Q	C	G	D	V	E	T	N	P	G	P
DHV-1 (DRL-62)	DQ219396	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
SVV	DQ641257	R	A	W	C	P	S	M	L	P	F	R	S	Y	K	Q	K	M	L	M	Q	S	G	D	I	E	T	N	P	G	P
SePV-1	EU152976	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P

Table 13. Picornaviral 2A sequences. Only one strain per species is shown. The conserved motif is in yellow and the tryptophan residues in orange. Differences at the conserved motif are shown in red.

During processing of the P1-2A precursor, 2A is cleaved from 1D by the $3C^{pro}$ or more efficiently by $3CD^{pro}$ (Ryan *et al*., 1989). This provides an additional constraint on the 2A sequence: specifically the need to form a suitable substrate for $3C^{pro}$.

In comparison with other picornavirus 3C proteinases, FMDV $3C^{pro}$ cleaves a wide range of amino acid pairs (Ryan *et al*., 2004). The primary cleavage of P1-2A occurs mostly at the Q/L site, and in some cases, at the Q/M, Q/T, Q/V, Q/S and Q/C, whereas in PTV the cleavage site between P1-2A is completely conserved at a Q/G pair (Zell *et al*., 2001). The presence of this cleavage site presents a constraint in the position 19-20 and the 2-3 residues surrounding the 1D/2A cleavage site in FMDV. This is not the case in cardioviruses, since 2A is not 19 amino acids in length but ~150 amino acids. This constraint is also observed in other viruses in which 2A is proteolytically trimmed from 1D; BRV, ERBV (position 20-21) and PTV (position 22-23). It is known that $3C^{pro}$ recognizes a short sequence of four or more residues (reviewed by Ryan & Flint, 1997). The constrains are, therefore, on a longer tract (5-6 aa) than just the amino acid pair where the cleavage occurs (see table 14).

Species	Accession number	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
FMDV	AY593836	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593831	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593832	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593823	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593824	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593811	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593812	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593828	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AF506822	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AJ539138	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AJ539139	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AJ539140	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AJ539141	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	DQ296518	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	DQ296517	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	DQ296514	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	DQ296511	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	DQ296507	H	K	Q	K	I	V	A	P	V	K	**Q**	**L**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	DQ296506	H	K	Q	K	I	V	A	P	V	K	**Q**	**T**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593814	H	K	Q	K	I	V	A	P	V	K	**Q**	**T**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593815	H	K	Q	K	I	V	A	P	V	K	**Q**	**T**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593816	H	K	Q	K	I	V	A	P	V	K	**Q**	**T**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593817	H	K	Q	K	I	V	A	P	V	K	**Q**	**T**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593799	R	K	Q	E	I	I	A	P	E	K	**Q**	**V**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY304994	R	K	Q	E	I	I	A	P	E	K	**Q**	**V**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593800	R	K	Q	E	I	I	A	P	E	K	**Q**	**V**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	AY593798	R	K	Q	E	I	I	A	P	E	K	**Q**	**V**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	X88856	R	K	Q	E	I	I	A	P	E	K	**Q**	**V**	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	DQ989309	R	K	Q	E	I	I	A	P	E	K	**Q**	**M**	M	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	DQ989308	R	K	Q	E	I	I	A	P	E	K	**Q**	**M**	M	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	DQ989304	R	K	Q	E	I	I	A	P	E	K	**Q**	**D**	L	N	L	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	DQ989303	R	K	Q	E	I	I	A	P	E	K	**Q**	**V**	L	N	L	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
FMDV	DQ119643	H	K	Q	K	I	V	A	P	A	K	**Q**	**S**	L	N	F	D	L	L	R	L	A	G	D	V	E	S	N	P	G	P
ERAV	L43052	R	H	K	F	P	T	N	I	N	K	**Q**	**C**	T	N	Y	S	L	L	K	L	A	G	D	V	E	S	N	P	G	P
BRV2	EU236594	L	R	L	T	G	E	I	V	K	**Q**	**G**	A	T	N	F	E	L	L	Q	Q	A	G	D	V	E	T	N	P	G	P
ERBV-1	X96871	E	A	T	L	S	T	I	L	S	**E**	**G**	A	T	N	F	S	L	L	K	L	A	G	D	V	E	L	N	P	G	P
ERBV-2	AF361253	V	A	D	W	E	N	L	L	S	**Q**	**G**	A	T	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
PTV-1	AF296106	A	M	T	V	M	T	F	**Q**	**G**	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PTV-1	AY392535	A	M	T	V	M	T	F	**Q**	**G**	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PTV-1	AF296105	A	M	T	V	M	T	F	**Q**	**G**	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PTV-1	AF296101	A	M	T	V	M	T	F	**Q**	**G**	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PTV-1	AY392553	A	M	T	V	M	T	F	**Q**	**G**	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PTV-1-	AF231769	A	M	T	T	L	S	Y	**Q**	**G**	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PTV-1-	AB038528	A	M	T	T	L	S	Y	**Q**	**G**	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PTV-1	Q355222	A	M	T	T	L	S	Y	**Q**	**G**	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PTV-1	AJ011380	A	M	T	V	M	A	F	**Q**	**G**	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PTV-1	AY392552	A	M	T	V	M	A	F	**Q**	**G**	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
EMCV	AY392555	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
TMEV	AF231768	F	R	E	F	F	K	A	V	R	G	Y	H	A	D	Y	Y	K	Q	R	L	I	H	D	V	E	M	N	P	G	P
TLV	AF296097	F	S	D	F	F	K	H	V	R	E	Y	H	A	A	Y	Y	K	Q	R	L	M	H	D	V	E	T	N	P	G	P
SAF-V	AF296103	F	T	D	F	F	K	A	V	R	D	Y	H	A	S	Y	Y	K	Q	R	L	Q	H	D	V	E	T	N	P	G	P

Table 14. Aphthoviruses, teschoviruses and cardioviruses different strains (not all shown). The conserved motif is highlighted in yellow. The different $3C^{pro}$ cleavage sites that separate P1 from 2A are shown in purple. Not shown in cardioviruses, since 2A is longer, thus $3C^{pro}$ cleavage site is upstream the 30 amino acids showed in the figure.

Molecular modelling studies of FMDV 2A (cited in Ryan *et al.*, 1999) revealed a predicted structure wherein an amphipathic α-helix is stabilized by a number of side-chain/side-chain interactions (figure 91). Therefore, there are interactions not only between side chains and the ribosomal exit tunnel but also between the side chains within 2A protein. One of these interactions might occur across the 1D/2A $3C^{pro}$ cleavage site (-Q/Xaa) between the capsid protein 1D C-terminal K and an N residue (figure 91).

This α-helix is characterized by consecutive main chain i→i+4 hydrogen bonds between each amide hydrogen and a carbonyl oxygen from the adjacent helical turn (Pauling and Corey, 1951).

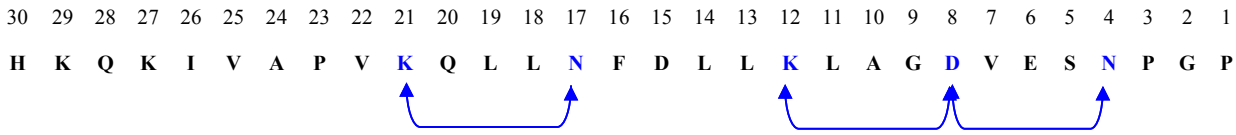


Figure 91. **FMDV 2A sequence, including the C-terminus o 1D.** Side-chain interactions are shown in blue.

4.5.2 Trypanosomes, insects and mammalian 2A-like sequences

2A-like sequences, in *T. cruzi* and *T. brucei,* are located at the N-terminus of apurinic/apyrimidinic-endonuclease (APE) domain and repeated sequence TRS-1, respectively. *T. congolense* and *T. vivax* 2A-like sequences are also situated at the N-terminus of the APE domain.

It should be noted that these repeated elements may no longer be active and, thus, may have accumulated mutations. The high rate of mutations present in the C-terminal motif is shown in table 15 and 16. Using translation systems *in vitro*, both types of trypanosome 2A-like sequences were shown to mediate cleavage (Heras *et*

al., 2006). Alignments of the last 19 residues of *T.cruzi* 2A-like sequences, showed a high degree of amino acid conservation maintaining identity of 93.7 %. 72.5 % of them maintained the consensus motif (-DxExNPGP-), whereas 27.5 % contained point mutations at the consensus motif, the 20% being N^4→H (-DxEx**H**PGP) (figure 92). The presence of this H residue at this position in a high percentage of sequences suggests that it may provide a selective advantage and a functional role.

In the bioinformatics analysis only one type of each sequence with point mutation is represented, taking into account the variability but not the predominance of each type of sequence.

2A sequences with this N^4→H point mutation are active in trypanosomes but, like FMDV, they are less active reducing the efficiency of cleavage to 30% in FMDV (Donnelly *et al*., 2001a) and to 23.9% in trypanosomes (Heras *et al*., 2006). Further experiments, where a point mutation N^4→Q was introduced into the sequence, showed only 10.91% activity, confirming the significance of this N^4 residue in the cleavage activity. These studies suggested that this 2A-like sequence may play a role in regulation of L1Tc translation. Similar results were obtained for FMDV 2A, where this N^4→Q mutations was introduced and showed 10% of cleavage activity (Donnelly *et al*., 2001a). This suggests that L1Tc 2A has similar mechanism of action than FMDV 2A. However, it must be taken into account that the experiments were made in heterologous ribosomes and translation systems. From these earlier studies of L1Tc 2A the conclusion made was that this sequence plays a role in determining the ratio of the cleavage products, where the sequence with the conserved motive mediates molar excess of the upstream over the downstream product. In contrast, the N^4→H point mutated sequence showed equimolar ratios of translation products, which was also obtained when the length of the sequence was increased up to 58 residues (Heras *et al*., 2006). These results are consistent with the results obtained for FMDV 2A, where the addition of residues located upstream of the consensus motif favours hydrolysis of the tRNA-peptide ester linkage together with 'pseudo-reinitiation' instead of 'pseudo-termination' (Donnelly *et al*., 2001b).

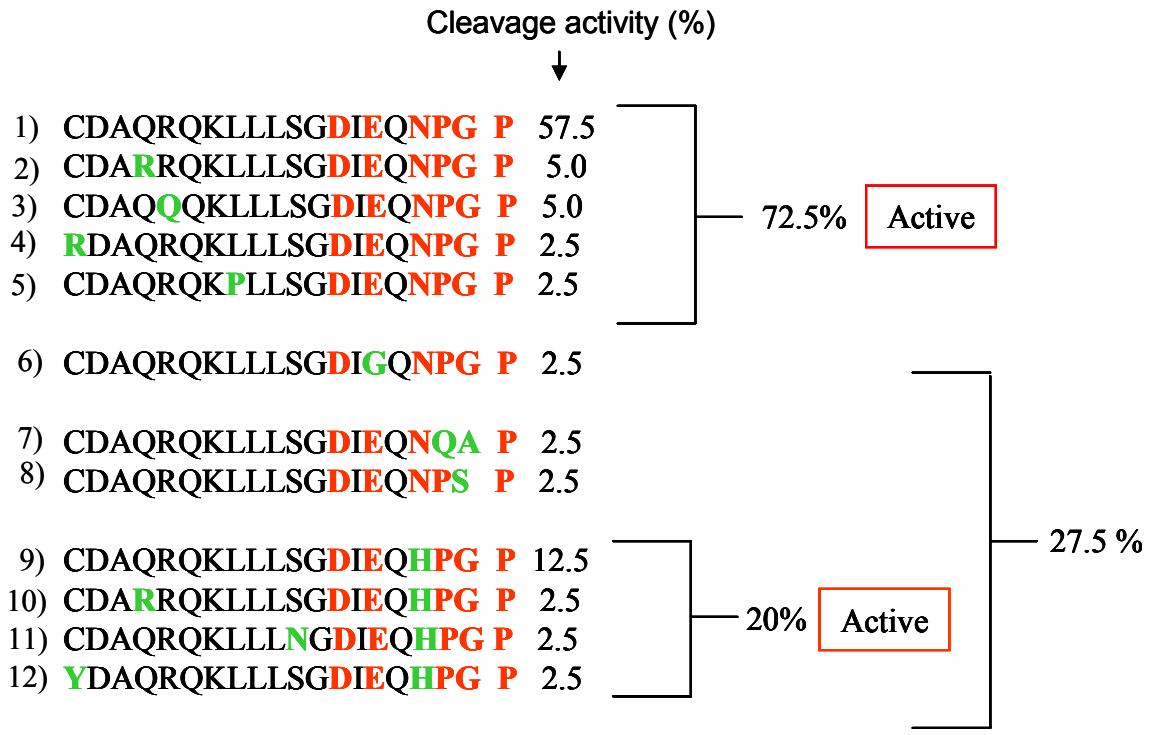


Figure 92. 2A-like *T.cruzi* sequences contained in L1Tc elements. Sequences were grouped considering the point mutations they contain (shown in green). The percentage (%) of elements that contain each type of 2A-like sequence are shown, being 75.5% sequences with conserved motif, 27.5% sequences with mutations at the conserved motif, and 20% sequences with the mutation N^{17}→H and tested to be active. Sequences 7, 8 and 9 have not been analyzed (adapted from Heras *et al.*, 2006).

Species	Accession number	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
FMDV-63	AY593836	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
T.cruzi	AAA67559	Q	P	Y	T	Y	C	L	R	A	L	C	D	A	Q	R	Q	K	L	L	L	I	G	D	I	E	Q	N	P	G	P
T.cruzi	CAB41692	Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	R	R	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	Q	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	R	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	P	Q	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	G	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	N	Q	A	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	N	P	S	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	H	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	R	R	Q	K	L	L	L	S	G	D	I	E	Q	H	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	N	G	D	I	E	Q	H	P	G	P
T.cruzi		Q	R	Y	T	Y	R	L	R	A	V	Y	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	H	P	G	P
T.vivax	936e06.q1k_21	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
T.vivax	1768f01.q1k_0	M	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
T.vivax	1768f01.q1k_7	M	L	P	C	A	C	G	R	A	T	L	D	A	R	R	L	T	L	L	V	S	G	D	V	E	R	D	P	G	P
T.vivax	302f07.p1k_20	I	L	P	C	T	C	E	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
T.vivax	1681d10.q1k_7	T	L	P	F	A	R	W	H	I	A	L	D	M	R	R	P	L	L	L	I	S	G	D	V	D	S	K	P	G	P
T.vivax	1664g12.q1k_4	L	L	P	C	T	C	G	R	A	T	L	D	A	W	R	L	L	L	L	I	C	G	G	V	G	R	N	P	G	P
T.vivax	1768f01.q1k_0	M	L	L	C	T	R	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	P	G	P
T.vivax	1768f01.q1k_7	M	L	L	C	T	R	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	P	G	P
T.vivax	563c09.q1k_9	M	L	L	C	T	R	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	P	G	P
T.vivax	696d02.p1k_1	I	L	P	C	T	C	G	R	A	A	L	D	A	Q	W	R	L	L	L	I	F	V	D	A	E	R	N	P	G	P
T.vivax	395e02.q1k_1	I	L	P	C	T	C	G	R	A	A	L	D	A	Q	W	R	L	L	L	I	F	V	D	A	E	R	N	P	G	P
T.vivax	938f02.q1k_5	I	L	P	C	T	R	G	R	A	M	L	S	A	R	W	L	L	L	L	I	S	G	G	V	E	R	K	P	G	P
T.vivax	733e05.p1k_4	I	L	P	F	T	C	G	R	A	A	L	D	A	W	R	L	L	L	L	I	G	G	G	V	G	R	N	P	G	P
T.vivax	73h08.q1k_2	I	L	P	C	L	C	V	H	A	A	S	D	A	R	W	L	L	L	L	I	S	G	D	V	E	R	R	P	C	P
T.vivax	390g10.p1k_23	M	L	L	C	T	S	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	S	G	P
T.vivax	892h02.p1k_5	S	Q	V	R	W	S	N	G	A	E	K	K	V	Q	R	L	L	L	L	S	G	G	D	V	E	R	N	P	G	P

Table 15. Comparison between FMDV 2A and Trypanosome (*T. vivax and T. cruzi)* 2A-like sequences. *T. cruzi* 2A-like sequences without accession numbers are taken from Heras *et al.*, 2006. The conserved motif is highlighted in yellow and the differences in red.

The *T. congolense* and *T. vivax* 2A-like sequences analyzed showed a high level of heterogeneity compared to FMDV 2A. Interestingly, some of the sequences contain a tryptophan (W) residue, which is never present in any picornavirus 2A sequences, except ERBV-2 and SVV (table 16). However, most of the sequences with W, do not have the consensus motif conserved and they may be inactive, which would be consistent with the fact that a bulky residue like tryptophan apparently does not favour the correct 2A conformation inside the ribosome. This is still unclear yet, since these sequences have not been tested.

It is interesting to note that where tryptophan do occur, the *i*+4 residue is very frequently leucine, or a residue with an aliphatic, hydrophobic, side chain (table 16). This is also observed in insect and mammalian 2A-like sequences (tables 17 and 18).

The fact that eukaryotic ribosomes from different origins are slightly different must also be taken into account. It has been reported using cryo-electron microscopy, that the overall structure of the *T. cruzi* 80S ribosome shows the phylogenetically conserved eukaryotic rRNA core structure, having well-defined small (40S) and large (60S) subunits, but also together with some distinctive structural features. It was also observed that the 60S ribosomal subunit from *T.cruzi* presents a shape that is similar to those from bacteria (Gao *et al*., 2005).

Species	Accession number	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
FMDV-63	AY593836	H	K	Q	K	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
T.congo	354h04.q1k_6	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
T.congo	335b10.q1k_3	I	L	P	C	T	C	G	C	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	V	E	R	N	P	G	P
T.congo	432g10.q1k_7	I	L	P	C	T	C	G	R	T	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
T.congo	400g12.q1k_4	I	L	P	C	T	C	G	R	T	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
T.congo	876g11.p1k_3	I	V	P	C	T	C	G	R	T	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
T.congo	1381h11.q1k_4	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	F	L	L	P	V	R	G	D	V	G	R	N	P	G	P
T.congo	1071g10.p1k_14	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	F	L	L	P	V	R	G	D	V	G	R	N	P	G	P
T.congo	1294e07.p1k_3	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	F	L	L	P	V	R	G	D	V	G	R	N	P	G	P
T.congo	1473f10.p1k_5	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	F	L	L	P	V	R	G	D	V	G	R	N	P	G	P
T.congo	1305b04.p1k_2	I	L	P	C	T	C	I	C	P	T	L	E	A	R	R	L	L	V	L	V	S	G	G	I	E	R	N	P	R	P
T.congo	530f06.q1kbw_10	A	L	S	C	V	C	G	H	G	N	S	L	L	C	R	L	L	L	F	L	S	G	D	V	E	Y	N	P	G	S
T.congo	1463e05.p1k_0	A	L	S	C	V	C	G	H	G	N	S	L	L	C	R	L	L	L	F	L	S	G	N	V	E	Y	N	P	G	S
T.congo	800b12.p1k_3	A	L	S	C	V	C	G	H	G	N	S	L	L	C	R	L	L	L	F	L	S	G	N	V	E	Y	N	P	G	S
T.congo	47d01.q1k_6	A	L	S	C	V	C	G	H	G	N	S	L	L	C	R	L	L	L	F	L	S	G	N	V	E	Y	N	P	G	S
T.congo	987a11.q1k_0	T	L	S	C	T	C	G	S	A	L	P	K	A	L	G	P	L	L	L	L	S	R	V	E	D	H	N	P	G	P
T.congo	1423d04.p1k_0	F	T	C	T	C	W	R	G	R	A	L	L	C	R	P	F	L	M	P	L	S	G	D	V	G	Q	N	P	E	P
T.congo	245e02.p1k_7	T	V	P	P	N	R	Q	C	A	L	Q	E	A	L	R	K	K	L	L	L	C	G	D	V	E	S	N	P	W	N
T.vivax	1198e11.p1k_1	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
T.vivax	961a05.q1k_2	M	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
T.vivax	1875a05.p1k_16	M	L	P	C	A	C	G	R	A	T	L	D	A	R	R	L	T	L	L	V	S	G	D	V	E	R	D	P	G	P
T.vivax	302f07.p1k_20	I	L	P	C	T	C	E	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
T.vivax	1681d10.q1k_7	T	L	P	F	A	R	W	H	I	A	L	D	M	R	R	P	L	L	L	I	S	G	D	V	D	S	K	P	G	P
T.vivax	1664g12.q1k_4	L	L	P	C	T	C	G	R	A	T	L	D	A	W	R	L	L	L	L	I	C	G	G	V	G	R	N	P	G	P
T.vivax	1768f01.q1k_0	M	L	L	C	T	R	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	P	G	P
T.vivax	1768f01.q1k_7	M	L	L	C	T	R	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	P	G	P
T.vivax	563c09.q1k_9	M	L	L	C	T	R	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	P	G	P
T.vivax	696d02.p1k_1	I	L	P	C	T	C	G	R	A	A	L	D	A	Q	W	R	L	L	L	I	F	V	D	A	E	R	N	P	G	P
T.vivax	395e02.q1k_1	I	L	P	C	T	C	G	R	A	A	L	D	A	Q	W	R	L	L	L	I	F	V	D	A	E	R	N	P	G	P
T.vivax	938f02.q1k_5	I	L	P	C	T	R	G	R	A	M	L	S	A	R	W	L	L	L	L	I	S	G	G	V	E	R	K	P	G	P
T.vivax	733e05.p1k_4	I	L	P	F	T	C	G	R	A	A	L	D	A	W	R	L	L	L	L	I	G	G	G	V	G	R	N	P	G	P
T.vivax	73h08.q1k_2	I	L	P	C	L	C	V	H	A	A	S	D	A	R	W	L	L	L	L	I	S	G	D	V	E	R	R	P	C	P
T.vivax	390g10.p1k_23	M	L	L	C	T	S	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	S	G	P
T.vivax	892h02.p1k_5	S	Q	V	R	W	S	N	G	A	E	K	K	V	Q	R	L	L	L	L	S	G	G	D	V	E	R	N	P	G	P

Table 16. Classification of a selection of different 2A-like sequences in *T. congolense* and *T.vivax* compared to FMDV 2A. Differences in the conserved motif are highlighted in red, and tryptophan (W) in orange. Residues in blue represent i, i+4 interactions.

Tryptophan residues are also found in the upstream region of some insect 2A-like sequences (table 17). Interestingly all the tested sequences containing this bulky residue are active. However, some caution must be taken in making some conclusions since the analysis were performed in heterologous *in vitro* ribosomes/translation systems (wheat germ extracts). These sequences have not been tested yet in insect *in vivo* systems, and the *in vitro* insect extracts that were used in this project did not show any expression (see section 3.4.1). They all keep the conserved motif, and a very heterogenic upstream region. Eukaryotic ribosomes from different origins, such as insect and mammals, show distinctive features, which, in the insect case, may allow the presence of tryptophan.

Species	Accession number	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
EoPV-2A1	AY365064	G	Q	R	T	T	E	Q	I	V	T	A	Q	G	W	A	P	D	L	T	Q	D	G	D	V	E	S	N	P	G	P
EoPV-2A2	AY365064	T	R	G	G	L	Q	R	Q	N	I	I	G	G	G	Q	R	D	L	T	Q	D	G	D	I	E	S	N	P	G	P
EoPV-2A1	AY341824	G	Q	R	T	T	E	Q	I	V	T	A	Q	G	W	A	P	D	L	T	Q	D	G	D	V	E	S	N	P	G	P
EoPV-2A2	AY341824	T	R	G	G	L	Q	R	Q	N	I	I	G	G	G	Q	R	D	L	T	Q	D	G	D	I	E	S	N	P	G	P
PnPV-2A1	AF323747	G	Q	R	T	T	E	Q	I	V	T	A	Q	G	W	V	P	D	L	T	V	D	G	D	V	E	S	N	P	G	P
PnPV-2A2	AF323747	T	R	G	G	L	R	R	Q	N	I	I	G	G	G	Q	K	D	L	T	Q	D	G	D	I	E	S	N	P	G	P
IFV	AB000906	P	S	I	G	N	V	A	R	T	L	T	R	A	E	I	E	D	E	L	I	R	A	G	I	E	S	N	P	G	P
ABPV	AF150629	T	G	F	L	N	K	L	Y	H	C	G	S	W	T	D	I	L	L	L	L	S	G	D	V	E	T	N	P	G	P
ABPV	AF486073	T	G	F	L	N	K	L	Y	H	C	G	S	W	T	D	I	L	L	L	L	S	G	D	V	E	T	N	P	G	P
ABPV	AF486072	T	G	F	L	N	K	L	Y	H	C	G	S	W	T	D	I	L	L	L	W	S	G	D	V	E	T	N	P	G	P
KBV	AY275710	I	G	F	L	N	K	L	Y	K	C	G	T	W	E	S	V	L	N	L	L	A	G	D	I	E	L	N	P	G	P
IAPV	EF219380	I	G	F	L	N	K	L	Y	R	C	G	D	W	D	S	I	L	L	L	L	S	G	D	I	E	E	N	P	G	P
CrPV	AF218039	L	V	S	S	N	D	E	C	R	A	F	L	R	K	R	T	Q	L	L	M	S	G	D	V	E	S	N	P	G	P
DCV	AF014388	Q	G	I	G	K	K	N	P	K	Q	E	A	A	R	Q	M	L	L	L	L	S	G	D	V	E	T	N	P	G	P
TaV	AF062037	R	G	P	R	P	Q	N	L	G	V	R	A	E	G	R	G	S	L	L	T	C	G	D	V	E	E	N	P	G	P
TaV	AF282930	R	G	P	R	P	Q	N	L	G	V	R	A	E	G	R	G	S	L	L	T	C	G	D	V	E	E	N	P	G	P
EeV	AF461742	R	R	L	P	E	S	A	Q	L	P	Q	G	A	G	R	G	S	L	V	T	C	G	D	V	E	E	N	P	G	P
PrV-2A1	AF548354	L	E	M	K	E	S	N	S	G	Y	V	V	G	G	R	G	S	L	L	T	C	G	D	V	E	S	N	P	G	P
PrV-2A2	AF548354	N	S	D	D	E	E	P	E	Y	P	R	G	D	P	I	E	D	L	T	D	D	G	D	I	E	K	N	P	G	P
PrV-2A3	AF548354	T	L	M	G	N	I	M	T	L	A	G	S	G	G	R	G	S	L	L	T	A	G	D	V	E	K	N	P	G	P
D. punctatus CPV1	AY163248	M	T	A	F	D	F	Q	Q	A	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P
D. punctatus CPV1	AY185594	M	T	A	F	D	F	Q	Q	A	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P
L. dispar CPV1	AF389466	M	T	A	F	D	F	Q	Q	A	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P
B. mori	AF433660	R	T	A	F	D	F	Q	Q	D	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	I	E	S	N	P	G	P
B. mori CPV1-H	AB035733	R	T	A	F	D	F	Q	Q	D	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	I	E	S	N	P	G	P
B. mori CPV1-I	AB035732	R	T	A	F	D	F	Q	Q	D	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	I	E	S	N	P	G	P
O. brumata CPV18	DQ192245	I	H	A	N	D	Y	Q	M	A	V	F	K	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P

Table 17. Insect 2A-like amino acid sequences. The C-terminal conserved motif is highlighted in yellow and tryptophan (W) residues are highlighted in orange**.** Blue residues represent i+4 interactions.

Tryptophan is also observed in a minority of mammalian 2A-like sequences. Like observed in trypanosomal and insect 2A-like sequences (tables 16 and 17), where tryptophan occurs, a leucine residue is observed at the i+4 position, suggestive of an i/i+4 hydrophobic sidechain interaction that could occur in an alpha helical conformation of the nascent peptide (table 18).

Species	Accession number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
SVV	DQ641257	R	A	W	C	P	S	M	L	P	F	R	S	Y	K	Q	K	M	L	M	Q	S	G	D	I	E	T	N	P	G	P
SePV-1	EU152976	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
SePV-1	EU142040	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
SePV-1	EU152979	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
SePV-1	EU152978	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
SePV-1	EU152975	C	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
SePV-1	EU152974	C	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	E	G	D	V	E	S	N	P	G	P
SePV-1	EU152980	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	I	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
SePV-1	EU152977	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	R	P
ADRV-N	AY632079	F	F	D	S	V	W	V	Y	H	L	A	N	S	S	W	V	R	D	L	T	R	E	C	I	E	S	N	P	G	P
ADRV-N (J19)	DQ113901	F	F	D	S	V	W	V	Y	H	L	A	N	S	S	W	V	R	D	L	T	R	E	C	I	E	S	N	P	G	P
ADRV-N (B219)	DQ168032	F	F	D	S	I	W	V	Y	H	L	A	N	S	S	W	V	R	D	L	T	R	E	C	I	E	S	N	P	G	P
Human-C (V508)	AY941781	G	V	G	Y	P	L	I	V	A	N	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G	P
Human-C (V966)	AY941782	G	V	G	Y	P	L	I	V	A	N	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G	P
Human-C (Bristol)	AJ132203	G	A	G	Y	P	L	I	V	A	N	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G	P
Human-C (V460)	AY941780	G	T	G	Y	P	L	I	V	A	N	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G	P
Bovine-C (Shintoku)	L12390	G	I	G	N	P	L	I	V	A	N	S	K	F	Q	I	D	R	I	L	I	S	G	D	I	E	L	N	P	G	P
Porcine-C (Cowden)	M69115	G	N	G	N	P	L	I	V	A	N	A	K	F	Q	I	D	K	I	L	I	S	G	D	V	E	L	N	P	G	P

Table 18. Mammalian 2A-like amino acid sequences. The C-terminal conserved motif is highlighted in yellow, and the point mutations in the region, in red. Tryptophan (W) residues are highlighted in orange**.** Blue residues represent i+4 interactions.

In summary, all 2A-like sequences present a similar pattern: a region with low conservation but containing a high proportion of aliphatic hydrophobic residues followed by the conserved motif. We have proposed 2A adopts an α-helical conformation within the ribosome exit tunnel and this preliminary analysis reveals a series of helix stabilizing hydrogen bond and hydrophobic i,i+4 interactions. Not only is the length of the sequence important but also the nature of each amino acid in each position to achieve the correct interaction with the exit tunnel to mediate ‘cleavage’.

4.6 SUMMARY

The concluding points that summarize this work are as follows:

- A reliable and efficient artificial reporter system that allows the detection of 2A activity directly in live cells without the need for Western blots has been created and used for the screening of an active 2A in bacterial systems. In spite of the fact that no active 2A within prokaryotic organisms was found, this reporter system can easily be adapted to eukaryotic organisms for mutagenic studies of 2A's activity.

- TMEV 2A antibodies have been produced and will be used for the localization of TMEV 2A within the cell and, also to provide an insight into its possible role in the inhibition of host´s cells protein synthesis.

- The creation of a reporter of stress in the cell (from the tested pHE27 vector) will form the basis of future work and represents a new tool in the biotechnology field.

- The study of new 2A-like sequences, found in the genome of the sea-urchin *Strongylocentrotus purpuratus,* has shown these cellular sequences to be active. These analyses provide useful information on 2A's role within *Strongylocentrotus* cells and the function of the innate immune system of this organism. Recent unpublished experiments show these new 2A-like sequences are acting as signal sequences within the cell.

-We know that all 2A/2A-like sequences present a conserved motif at the C-terminus, but what about the upstream sequence? Does it follow any pattern, any conservation, etc.? The bioinformatics analysis has provided insight on these questions.

We believe that this upstream sequence is interacting with the ribosomal exit tunnel; therefore, some grade of conservation in the amino acid organization within the sequence must be present. The bioinformatics analyses reveal a clear pattern.

In picornaviruses, there is a large predominance of leucine residues in positions 11, 13, 14, 15, 18 and 19, whereas position 12 is a lysine in most sequences. Conservation is observed in aphtho- and teschoviruses, at positions 19-20 (Q/L, M, T) and 22-23 (Q/G), respectively, due to presence of $3C^{pro}$ cleavage site. In contrast, insect 2A-likes sequences show little conservation upstream of the conserved motif, and represent the most heterogenic sequences.

Bulky residues with aromatic side chains (W, F and Y) seem to be widely avoided in the sequence except in trypanosome and insect 2A-like sequences, and also at position 16 in picornaviruses, where phenylalanine is present in the majority of cases.

Charged hydrophobic residues are more common at the N-terminal region whereas neutral, polar or non-polar, residues are found in the middle of 2A-like sequences. A conserved patch of leucine residues is found in all non-LTR 2A-like sequences within trypanosomes, the sea urchin *Strongylocentrotus purpuratus*, and also insect viruses (positions 11 to 14). This patch of leucines is also found in CATERPILLER 2A-like sequences within *Strongylocentrotus purpuratus.* However, in that case, the patch of leucine amino acids is flanked by an acidic and a basic residue instead of neutral residues. This is consistent with the recently determined role of CATERPILLER 2A-like sequences as signal sequences.

5. REFERENCES

Agol V.A. (2002) Picornavirus genome: an overview. In *Molecular Biology of Picornaviruses*, pp. 127-149. Edited by Semler, B.L., Wimmer, E. ASM Press Washington.

Al-Sunaidi M., Williams C.H., Hughes P.J., Schnurr D.P. & Stanway G. (2007) Analysis of a new human parechovirus allows the definition of parechovirus types and the identification of RNA structural domains. *Journal of Virology*. 81 (2): 1013-21.

Aminev A.G., Amineva S.P. & Palmenberg A.C. (2003a) Encephalomyocarditis viral protein 2A localizes to nucleoli and inhibits cap-dependent mRNA translation. *Virus Research*. 95: 45-57.

Aminev A.G., Amineva S.P. & Palmenberg A.C. (2003b) Encephalomyocarditis virus (EMCV) proteins 2A and 3BCD localize to nuclei and inhibit cellular mRNA transcription but not rRNA transcription. *Virus Research*. 95: 59-73.

Andino R., Rieckhof G.E., Achacoso P.L. & Baltimore D. (1993) Poliovirus RNA synthesis utilizes an RNP complex formed around the 5'end of the viral RNA. *The EMBO J.* 12 (9): 3587-3598.

Bachurski C.J., Theodorakis N.G., Coulson R.M.R. & Cleveland D.W. (1994) An amino-terminal tetrapeptide specifies cotranslational degradation of ß-tubulin but not α-tubulin mRNAs. *Molecular and Cellular Biology*. 14 (6): 4076-4086.

Barco A., Feduchi E. & Carrasco L. (2000) Poliovirus protease $3C^{pro}$ kills cells by apoptosis. *Virology*. 266 (2): 352-360.

Bell Y.C., Semler B.L. & Eherenfeld E. (1999) Requirements for RNA Replication of a Poliovirus Replicon by Coxsackievirus B3 RNA Polymerase. *Journal of Virology*. 9413-9421.

Belsham G.J., McInerney G.M. & Ross-Smith N. (2000) Foot-and-mouth disease virus 3C protease induces cleavage of translation initiation factors eIF4A and eIF4G within infected cells. *Journal of Virology*. 272-280.

Berisio R., Schluenzen F., Harms J., Bashan A., Auerbach T., Baram D. & Yonath A. (2003) Structural insight into the role of the ribosomal tunnel in cellular regulation. *Nature Structural Biology*.10: 367-370.

Bertram G., Innes S., Minella O., Richardson J.P. & Stansfield I. (2001) Endless possibilities: translation termination and stop codon recognition. *Microbiology*. 147: 255-269.

Bienz K., Egger D. & Pfister T. (1994) Characteristics of the picornavirus replication complex. *Arch Virol Suppl*. 9: 147-157.

Blair W.S., Parley T.B., Bogerd H.P., Towner J.S., Semler B.L. & Cullen B.R. (1998) Utilization of a mammalian cell-base RNA binding assay to characterize the RNA binding properties of picornavirus 3C proteinases. *RNA*. 4: 215-225.

Brahic M., Bureau J.F. & Michiels T. (2005) The genetics of the persistent infection and demyelinating disease caused by Theiler´s virus. *Annual Review Microbiology*. 59: 279-298.

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 frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J*. 12: 3779-3785.

Browne G.J & Proud C.G. (2002) Regulation of peptide-chain elongation in mammalian cells. *FEBS*. 269: 5360-5368.

Brierley I., Pennell S. & Gilbert R.J. (2007) Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nature Reviews*. 5 (8): 598-610.

Carlberg U., Nilsson A. & Nygard O. (1990) Functional properties of phosphorylated elongation factor 2. *European Journal of Biochemistry*. 191: 639-645.

Chaplin P.J., Camon E.B., Villareal-Ramos B., Flint M., Ryan M.D. & Collins R.A. (1999) Production of interleukin-12 as a self-processing 2A polypeptide. *Journal of .Interferon and Cytokine Research.* 19: 235-241.

Ciervo A., Beneduce F., Morace G. (1998) Polypeptide 3AB of Hepatitis A Virus Is a Transmembrane Protein. *Biochemical and Biophysical Research Communications*. 249 (1): 266-274.

Clark A.T., Robertson M.E.M., Conn G.L. & Belsham G.J. (2003). Conserved nucleotides within the J domain of the encephalomyocarditis virus internal ribosome entry site are required for activity and for interaction with eIF4G. *Journal of Virology.* 77: 12441-12449.

Clarke B.E., Sangar D.V., Burroughs J.N., Newton S.E., Carroll A.R. & Rowlands D.J. (1985) Two initiation sites for Foot-and-Mouth disease virus polyprotein *in vivo*. *Journal of General Virology*. 66: 2615-2626.

Clement S.L. & Likke-Andersen J. (2006) No mercy for messages that mess with the ribosome. *Nature Structural and Molecular biology.* 13 (4): 299-301.

Cuconati A., Molla A. & Wimmer E. (1998) Brefeldin A inhibits cell-free, de novo synthesis of poliovirus. *Journal of Virology*. 72: 6456-6464.

Curran J. & Kolakofsky D. (1989) Scanning independent ribosomal initiation of the Sendai virus Y proteins in vitro and in vivo. *EMBO J.*. 8 (2): 521-526.

de Felipe P., Luke G.A., Hughes L.E., Gani D., Halpin C. & Ryan M.(2006) *Eunum pluribus*: multiple proteins from a self-processing polyprotein. *Trends in Biotechnology*. 24 (2): 68-75.

De Jong A. S., de Mattia F., van Dommelen M.M., Lanke K., Melchers W.J.G., Willems P.H.G.M. & van Kuppeveld F.J.M. (2008) Functional analysis of picornavirus 2B proteins: effects on calcium homeostasis and intracellular protein trafficking. *Journal of Virology*. 82 (7): 3782-3790.

Delbecq P., Calvo O., Filipkowski R.K., Pierard A. & Messenguy F. (2000) Functional analysis of the leader peptide of the yeast gene CPA1 and heterologous regulation by other fungal peptides. *Current Genetics*. 38: 105-112.

Delhaye S., van Pesch V. & Michiels T. (2004).The leader protein of Theiler's virus interferes with nucleocytoplasmic trafficking of cellular proteins. *Journal of Virology*. 78: 4357–4362.

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

Dodd D.A., Giddings T.H. & Kirkegaard K. (2001) Poliovirus 3A protein limits Interleukin-6, IL-8 and beta interferonsecretion during viral infection. *Journal of Virology*. 8158-8168.

Doedens J.R. & Kirkegaard K. (1995) Inhibition of protein secretion by poliovirus proteins 2B and 3A. *EMBO J*. 14: 894–907.

Donnelly M.L.L., Gani D., Flint M., Monaghan S. & Ryan M.D. (1997) The cleavage activities of aphthovirus and cardiovirus 2A proteins. *Journal of General Virology.* 78: 13-21.

Donnelly M.L.L., Hughes L.E., Luke G., Mendoza H., Gani D. & Ryan M.D. (2001a) The cleavage activities of foot-and-mouth disease virus 2A site directed

mutants and naturally occurring ´2A-like´sequences. *Journal of General Virology.* 82: 1027-1041.

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

Doronina V.A., de Felipe P., Wu C., Sharma P., Sachs M.S., Ryan M.D., Brown J.D. (2008) Dissection of a co-translational nascent chain separation event. *Biochemical Society Transactions*. 36 (Pt 4):712-6.

Duden R., Hosobuchi M., Hamamoto S., Winey M., Byers B. & Schekman R. (1994). Yeast beta-coat and beta9-coat proteins (COP). Two coatomer subunits essential for endoplasmic reticulum-to-Golgi protein traffic. *Journal of Biological Chemistry.* 269: 24486–24495.

Duke G.M., Osorio J.E. & Palmenberg A.C. (1990) Attenuation of Mengo virus through genetic engineering of the 5' noncoding poly(C) tract. *Nature*. 343: 474-476.

Dulebohn D., Choy J., Sundermeier T., Okan N. & Karzai A.W. (2007) Trans-Translation: The tmRNA-Mediated Surveillance Mechanism for Ribosome Rescue, Directed Protein Degradation, and Nonstop mRNA Decay. *Current Topics*. 46 (16): 4681-4693.

Dynon K., Black W., Ficorilli N., Hartley C. & Studdert M. (2007). Detection of viruses in nasal swab samples from horses with acute, febrile, respiratory disease using virus isolation, polymerase chain reaction and serology. *Australian Veterinary Journal.* 85 (1-2): 46-50.

Feng Y., Yuan H., Rein A. & Levin J. (1992) Bipartite signal for read-through suppression in murine leukemia virus mRNA: an eight-nucleotide purine-rich

sequence immediately downstream f the gag termination codon followed by the RNA pseudoknot. *Journal of General Virology*. 5127-532.

Ferrer-Orta C., Arias A., Agudo R., Pérez-Luque R., Escarmís C.,Domingo E. & Verdaguer N. (2004) Structure of Foot-and-Mouth Disease Virus RNA-dependent RNA Polymerase and Its Complex with a Template-Primer RNA. *Journal of Biological Chemistry.*279 (45): 47212-47221.

Ferrer-Orta C., Arias A., Agudo R., Pérez-Luque R., Escarmís C.,Domingo E. & Verdaguer N. (2006) The structure of a protein primer-polymerase complex in the initiation of genome replication. *EMBO J.* 25(4): 880–888.

Flint S.J., Enquist L.W., Krug R.M., Racaniello V.R. & Skalka A.M. (2000) Principles of Virology: molecular biology, pathogenesis, and control. ASM Press, Washington.

Forss S. & Schaller H.A. (1982) A tandem repeat gene in picornavirus. *Nucleic Acids Research.*10: 6441-6450.

Freistroffer D.V., Pavlov M.Y., MacDougall J., Buckingham R.H. & Ehrenberg M. (1997) Release factor RF3 in *E.coli* accelerates the dissociation of release factors. *EMBO J.* 16: 4126-4133.

Frolova L., Le Goff X., Rasmussen H.H., Cheperegin S., Drugeon G., Kress M., Arman I., Haenni A., Celis J.E., Philippe M., Justesen J. & Kisselev L. (1994) A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factors. *Nature*. 372: 701-703.

Frolova L., Le Goff X., Zhouravleva G., Davydova E., Philippe M. & Kisselev L. (1996) Eukaryotic polypeptide chain release factor eRF3 is an eRF1- and ribosome-dependent guanosine triphosphatase. *RNA*. 2: 334–341.

Frolova L. Y., Tsivkovskii R. Y., Sivolobova G. F., Oparina N. Y., Serpinsky O. I., Blinov V. M. & Tatkov S. I. (2003) Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA*. 5: 1014-1020.

Gao X. & Huang L. (1995) Cationic liposome-mediated gene transfer. *Gene Therapy*. 2 (10): 710-722.

Gao H., Ayub M.J., Levin M.J. & Frank J. (2005) The structure of the 80S ribosome from *Trypanosoma cruzi* reveals unique rRNA components. *PNAS*. 102 (29): 10206-10211.

Gazina E. V., Mackenzie J. M., Gorrell R. J. & Anderson D. A. (2002) Differential requirements for COPI coats in formation of replication complexes among three genera of Picornaviridae. *Journal of Virology*. 76: 11113–11122.

Gesteland R.F. & Atkins J.F. (1996) Recoding: dynamic reprogramming of translation. *Annual Review of Biochemistry*. 65: 741-768.

Gillet R. & Felden B. (2001) Emerging views of tmRNA-mediated protein tagging and ribosome rescue. *Molecular Microbiology*. 42 (4): 879-885.

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

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

Glaser W., Triendl A. & Skern T. (2003) The processing of eIF4GI by human rhinovirus type 2 2A (pro): Relationship to self-cleavage and role of zinc. *Journal of Virology*. 77 (8): 5021-5025.

Gong F. & Yanofsky C. (2002) Instruction of translating ribosome by nascent peptide. *Science*. 297: 1864-1867.

Goodfellow I.G., Polacek C., Andino R. & Evans D.J. (2003). The poliovirus 2C cis-acting replication element-mediated uridylylation of VPg is not required for synthesis of negative-sense genomes. *Journal of General Virology*. 83: 2359-2363.

Gonzalez M.E. & Carrasco L. (2003) Viroporins. *FEBS letter*. 552 (1): 28-34.

Gorbalenya A.E., Enjuanes L., Ziebuhr J. & Snijder E.J. (2006) Nidovirales: Evolving the largest ENA virus genome. *Virus Research*. 117: 17-37.

Gradi A., Foeger N., Strong R., Svitkin Y.V., Sonenberg N., Skern T. & Belsham G.J.(2004) Cleavage of eukaryotic translation initiation factor 4GII within foot-and-mouth disease virus-infected cells: Identification of the L-protease cleavage site in vitro. *Journal of Virology*. 78 (7): 3271-3278.

Graille M., Chaillet M. & van Tilbeurgh G. (2008) Structure of yeast Dom34: a protein related with translation terminator factor eRF1 and involved in Non-Go decay. *The Journal of Biological Chemistry.* 1-17.

Grentzmann G., Brechemier-Baey D., Heurgue V., Mora L. & Buckingham R.H. (1994) Localization and Characterization of the Gene Encoding Release Factor RF3 in *Escherichia coli*. *PNAS*. 91: 5848-5852.

Green K. Y., Ando T., Balayan M. S., Berke T., Clarke I. N., Estes M. K., Matson D. O., Nakata S., Neill J. D., Studdert M. J. & Thiel H. J. (2000) Taxonomy of the caliciviruses. *J. Infect. Dis*. 181: S322-S330

Groppo R. & Palmenberg A.C. (2007) Cardiovirus 2A protein associates with 40S but not 80S ribosome subunits during infection. *Journal of Virology*. 81 (23): 13067-13074.

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. *Journal of Virology.* 213 (2): 581-589

Hahn, H. & Palmenberg, A. C. (1995) Encephalomyocarditis viruses with short poly(C) tracts are more virulent than their mengovirus counterparts. *Journal of Virology*. 69: 2697-2699.

Hahn, H. & Palmenberg, A. C. (1996) Mutational analysis of the encephalo-myocarditis virus primary cleavage. *Journal of Virology.* 70: 6870-6875.

Hahn H. & Palmenberg A.C. (2001) Deletion mapping of the encephalomyocarditis virus primary cleavage site. *Journal of Virology*. 75 (15): 7215-7218.

Hansen J. L., Long A. M. & Schultz S. C. (1997) Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure.* 5: 1109-1122.

Hardesty B. & Kramer G. (2001) Folding of a nascent peptide on the ribosome. *Progress in Nucleic Acid Research and Molecular Biology*. 66: 41-67.

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

Harrod R. & Lovett P.S. (1994) Anti-peptidyltransferase leader peptides of attenuation-regulated chloramphenicol-resistance genes. *PNAS*. 91: 5612-5616.

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 peptide control of translation. *PNAS*. 92: 8650-8654.

Heras S.R., Thomas M.C., García-Canadas M., de Felipe P., García- Pérez J.L., Ryan M.D. & López M.C. (2006) L1Tc non-LTR retrotransposons from *Tripanosoma cruzi* contain a functional viral-like self-cleaving 2A sequence in frame

with the active proteins they encode. *Cellular Molecular Life Sciences*. 63: 1449-1460.

HeurgueÂ-Hamard V., Karimi R., Mora L., MacDougall J., Leboeuf C., Grentzmann G., Ehrenberg M. & Buckingham R. H. (1998) Ribosome release factor RF4 and termination factor RF3 are involved in dissociation of peptidyl-tRNA from the ribosome. *EMBO J*. 17: 808-816.

Horzinek, M.C. (1997) The birth of virology. *Antonie van Leeuwenhoek.* 71: 15–20.

Hughes L. (2003) Analysis of the foot-and-mouth disease virus 2A-mediated polyprotein processing event. Ph.D. Thesis, Department of Biomolecular Sciences, University of St.Andrews, St. Andrews.

Igarashi A. (1978). Isolation of a Singh's Aedes albopictuscell clone sensitive to dengue and chikungunya virues. *Journal of General Virology.* 40: 531-544.

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 insects, mammalian and plant picorna(-like) viruses. *Archives of Virology*. 143: 127-143.

Ito M., Uno M. & Nakamura Y. (2000) A tripeptide anticodon deciphers stop codons in mesesenger RNA. *Nature*. 403 (6770): 680-684.

Ito M., Yamashita T., Tsuzuki H., Takeda N. & Sakae K. (2004) Isolation and identification of a novel human parechovirus. *Journal of General Virology.* 85 (2): 391-398.

Jackson T., King A.M., Stuart D. & Fry E. (2003) Structure and receptor binding. *Virus research*. 91: 33-46.

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

Jang S.K. (2006) Internal initiation: IRES elements of picornaviruses and hepatitis c virus. *Virus Research*. 119 (1): 2-15.

Jayaram H., Estes M.K. & Venkataram Prasad B.V. (2004) Emerging themes in rotavirus cell entry, genome organization, transcription and replication. *Virus research.* 101: 67-81.

Jia X.Y., Summers D.F. & Ehrenfeld E. Primary cleavage of the HAV capsid protein precursor in the middle of the proposed 2A coding region. *Virology*. 193: 515-519.

Joachims M., van Breugel P.C. & Lloyd R.E. (1999) Cleavage of poly(A)-binding protein by enterovirus proteases concurrent with inhibition of translation in vitro. *Journal of Virology.* 73: 718-727.

Johansson S., Niklasson B., Maizel J., Gorbalenya A. & Lindberg A.M. (2002) Molecular Analysis of Three Ljungan Virus Isolates Reveals a New, Close-to-Root Lineage of the *Picornaviridae* with a Cluster of Two Unrelated 2A Proteins. *Journal of Virology.* 76 (17): 8920-8930.

Johnson A.E. (2005) The co-translational folding and interactions of nascent protein chains: a new approach using fluorescence resonance energy transfer. *FEBS Letters*. 579 (4): 916-920.

Jones M.S., Lukashov V.V., Ganac R.D. & Schnurr D.P. (2007). Discovery of a Novel Human Picornavirus in a Stool Sample from a Pediatric Patient Presenting with Fever of Unknown Origin. *Journal of Clinical Microbiology*. 45: 2144-2150.

Jurgens C.K., Barton D.J., Sharma N., Morasco B.J., Ogram S. A. & Flanegan J.B (2005) $2A^{pro}$ is a multifunctional protein that regulates the stability, translation and replication of poliovirus RNA. *Virology*. 345:346-357.

Kaku Y., Chard L., Inoue T. & Belsham G.J. (2002) Unique characteristics of a Picornavirus internal ribosomal entry site from the porcine teschovirus-1 talfan. *Journal of Virology*. 11721-11728.

Karniely S. & Pines O. (2005) Single translation-dual destination: mechanism of dual protein targeting in eukaryotes. *EMBO Reports*. 6 (5): 420-425.

Karamyshev A.L., Ito K. & Nakamura Y. (1999) Polypeptide release factor eRF1 from Tetrahymena thermophila: cDNA cloning, purification and complex formation with yeast eRF3. *FEBS Letters*. 457: 483-488.

Karpf A. R., Blake J. M. & Brown D. T., (1997). Characterisation of the infection of *Aedes albopictus* cell clones by Sindbis virus. *Virus Research.* 50: 1-13.

Keiler K. C., Waller P. R. & Sauer R. T. (1996) Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science 271*: 990-993.

Kisselev L., Ehrenbergh M. & Frolova L. (2003) Termination of translation: interplay of mRNA, rRNAs and release factors. *EMBO J*. 22 (2): 175-182.

Knox C., Moffat K., Shireen A., Ryan M. & Wileman T. (2005) Foot-and-mouth disease virus replication sites form next to the nucleus and close to the Golgi apparatus, but exclude marker proteins associated with host membrane compartments. *Journal of General Virology*. 86: 687-696.

Kolupaeva V.G., Pestova T.V., Hellen C.U.T. & Shatsky I.N. (1998). Translation eukaryotic initiation factor 4G recognizes a specific structural element within the

internal ribosome entry site of encephalomyocarditis virus RNA. *Journal of Biological Chemistry.* 273: 18599–18604.

Kolupaeva V.G., Lomakin I.B., Pestova T.V. & Hellen C.U.T. (2003). Eukaryotic initiation factors 4G and 4A mediate conformational changes downstream of the initiation codon of the encephalomyocarditis virus internal ribosomal entry site. *Molecular Cellular Biology*. 23: 687–698.

Kong W.P. & Roos R.P. (1991). Alternative translation initiation site in the DA strain of Theiler's murine encephalomyelitis virus. *Journal of Virology.* 65: 3395–3399.

Kuyumcu-Martinez N.M., Joachims M. & Lloyd R.E. (2002) Efficient cleavage of ribosome-associated poly(A)-binding protein by enterovirus 3C protease. *Journal of Virology*. 76: 2062-2074.

Kuyumcu-Martinez N.M., van Eden M.E., Younan P. & Lloyd R.E. (2004) Cleavage of poly(A)-binding protein by poliovirus 3C protease inhibits host cell translation: a novel mechanism for host translation shutoff. *Molecular and Cellular Biology.* 24 (4): 1779-1790.

Laemmli U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 227 (5259): 680-685.

Lama J., Sanz M.A. & Carrasco L. (1998) Genetic analysis of poliovirus protein 3A: characterization of a non-cytopathic mutant virus defective in killing Vero cells. *Journal of Virology*.79: 1911-1921.

Langland J.O., Pettiford S., Jiang B. & Jacobs B.L. (1994) Products of the porcine group C rotavirus NSP3 gene bind specifically to double-stranded RNA and inhibit

activation of the interferon-induced protein kinase PKR. *Journal of Virology*. 68 (6): 3821-3829.

Leong L.E.C., Cornell C.T. & Semler B.L. (2002) Processing determinants and functions of cleavage products of Picornavirus polyproteins. In *Molecular biology of picornaviruses*, pp. 187-198. Edited by Semler B.L., Wimmer E. ASM Press, Washington.

Li G. & Rice C.M. (1993) The signal for translational read-through of a UGA codon in Sinbis virus RNA involves a single cytidine residue immediately downstream of the termination codon. *Journal of Virology*. 67: 5062-5067.

Lidsky P.V., Hato S., Bardina M.V., Aminev A.G., Palmenberg A.C., Sheval E.V., Polyakv V.Y., van Kuppeveld F.J.M. & Agol V.I. (2006) Nucleocytoplasmic traffic disorder induced by Cardioviruses. *Journal of Virology*. 80 (6): 2705-2717.

Lodish H., Berk A., Zipursky L.S., Matsudaira P., Baltimore D. & Darnell J. (2000) Molecular cell biology. Fourth Edition. Edited by FREEMAN W.H and Company Molecular Cell Biology.

Lovett P. & Rogers E.J. (1996) Ribosome regulation by nascent peptide. *Microbiological Reviews*. 60: 336-385.

Lloyd R.E., Grubman I.J. & Ehrenfeld E. (1988). Relationship of p220 cleavage during picornavirus infection to 2A proteinase. *Journal of Virology* 62: 4216-4223.

Lich J.D. & Ting J.P-Y. (2007) CATERPILLER (NCR) Family members as positive and negative regulators of inflammatory responses. *Proceedings of the American Thoracic Society*. 4: 263-266.

Lim V.I. & Spirin A.S. (1986) Stereochemical analysis of ribosomal transpeptidation conformation of nascent peptide. *Journal of Molecular Biology*. 188: 565-577.

Lipton H. (1975) Theiler's virus infection in mice: an unusual biphasic disease process leading to demyelination. *Infection and Immunity*. 11: 1147-1155.

Liu J., Wei T. & Kwang J. (2003) Membrane-association properties of avian encephalomyelitis virus protein 3A. *Virology*. 321 (2): 297-306.

Luke G.A., de Felipe P., Lukashev A., Kallioinen S., Bruno E.A. & Ryan M.D. (2008) The occurrence, function and evolutionary origins of 2A-like sequences in virus genomes. *Journal of General Virology.* 89: 1-7

Lustig A. & Levine A.J. (1992) One Hundred Years of Virology. *Journal of Virology*: 4629-4631.

Mankin A.S. (2006) Nascent peptide in the 'birth canal' of the ribosome. *Trend in the Biochemical Sciences*. 31 (1): 11-13.

Martin A., Escriou N., Chao S-F., Girard M., Lemon S.M. & Wychowsky C. (1995) Identification and site-directed mutagenesis of the primary (2A/2B) cleavage site of the hepatitis A virus polyprotein: functional impact on the infectivity of HAV RNA transcripts. *Virology*. 213 (1): 213-222

Medvedkina O.A., Scarlat I.V., Kalinina N.O. & Agol V.I. (1974) Virus-specific proteins associated with ribosomes of Kreb-II cells infected with encephalomyocarditis virus. *FEBS Letters*. 39 (1): 4-8.

Merz A.J. & Wickner W.T. (2004) Trans-SNARE interactions elicit Ca2+ efflux from the yeast vacuole lumen. *Journal of Cellular Biology*. 164: 195-206.

Meyers G. (2007) Characterization of the Sequence Element Directing Translation Reinitiation in RNA of the Calicivirus Rabbit Hemorrhagic Disease Virus. *Journal of Virology*. 9623-9632.

Mitsui K., Brady M., Palfrey H.C. & Nairn A.C. (1993) Purification and characterization of calmodulin-dependent protein kinase III from rabbit reticulocytes and rat pancreas. *The Journal of Biological Chemistry.* 268: 13422-13433.

Moffat K., Howell G., Knox C., Belsham G.J., Monaghan P., Ryan M.D. & Wileman T. (2005) Effects of Foot-and-mouth disease virus non-structural proteins on the structure and function of the early secretory pathway: 2BC but not 3A blocks endoplamic reticulum-to-Golgi transport. *Journal of Virology*. 4382-4395.

Moore S.D. & Sauer R.T. (2007) The tmRNA system for translational surveillance and ribosome rescue. *Annual Review of Biochemistry.* 76: 101-124.

Nairn A.C. & Palfrey H.C. (1987) Identification of the major Mr 100,000 substrate for calmodulin-dependent protein kinase III in mammalian cells as elongation factor-2. *The Journal of Biological Chemistry*. 262: 17299-17303.

Nakatogawa H. & Ito K. (2002) The ribosomal exit tunnel functions as a discriminating gate. *Cell.* 108: 629-636.

Nakamura Y. & Ito K. (2003) Making sense of mimic in translation termination. *Trends in Biochemical Sciences.* 28: 99-105.

Nayak A., Goodfellow I.G. & Belsham G.J. (2005) Factors Required for the Uridylylation of the Foot-and-Mouth Disease Virus 3B1, 3B2, and 3B3 Peptides by the RNA-Dependent RNA Polymerase ($3D^{pol}$) *In Vitro*. *Journal of Virology*. 7698-7706.

Neznanov N., Kondratova A., Konstantin C.M., Angres B., Zhumabayeva B., Agol V.I. & Gudkov A.V. (2001) Poliovirus Protein 3A Inhibits Tumor Necrosis Factor (TNF)-Induced Apoptosis by Eliminating the TNF Receptor from the Cell Surface. *Journal of Virology*. 10409-10420.

Nibert M.L. (2007) 2A-like' and 'shifty heptamer' motifs in penaeid shrimp infectious myonecrosis virus, a monosegmented double-stranded RNA virus. *Journal of General Virology*. 88: 1315-1318.

Niklasson B., Kinnunen L., Hörnfeldt B., Hörling J., Benemar C., Hedlund K.O., Matskova L., Hyypiä T. & Winberg G.(1999).A new picornavirus isolated from bank voles (*Clethrionomys glareolus*). *Virology*. 255 (1): 86-93.

Niklasson B., Samsioe A., Papadogiannakis N., Kawecki A., Hörnfeldt B., Saade G.R. & Klitz W. (2007) Association of zoonotic Ljungan virus with intrauterine fetal deaths. *Birth Defects Research Part A: Clinical and Molecular Teratology.* 79 (6): 488-93.

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 (5481): 920-930.

Nuñez J.I., Baranowvsky E., Molina N., Ruiz-Jarabo C.M., Sanchez C., Domingo E. & Sobrino F. (2001) A Single Amino Acid Substitution in Nonstructural Protein 3A can Mediate Adaptation of Foot-and-Mouth Disease Virus to the Guinea Pig. *Journal of Virology*. 75 (8): 3977-3983.

Oakes M. I., Scheinman A., Atha T., Shankweiler G. & Lake J. A. (1990) In *The Ribosome: Structure, Function and Evolution,* pp. 180-193. Edited by Hill W. E., Dahlberg A., Garrett R. A., Moore P. B., Schlessinger D. and Warner J. S. American Society for Microbiology, Washington, D.C.

Palmenberg A.C. (1990) Proteolytic processing of picornaviral polyprotein. *Annual Review of Microbiology*. 44: 603-623.

Palmenberg A.C., Parks G.D., Hall D.J., Ingraham R.H., Seng T.W. & Pallai P.V. (1992) Proteolytic processing of the cardioviral P2 region: Primary 2A/2B cleavage in clone-derived precursors. *Virology*.190: 754-762.

Pham N.T.K., Khamrin P., Nguyen T.A., Kanti D.S., Phan T.G., Okitsu S. & Ushijima H. (2007) Isolation and Molecular Characterization of Aichi Viruses from Fecal Specimens Collected in Japan, Bangladesh, Thailand, and Vietnam. *Journal of Clinical Microbiology*. 2287-2288.

Pasternak A. O., Spaan W.J.M. & Snijder E.J. (2004) Regulation of Relative Abundance of Arterivirus Subgenomic mRNAs. *Journal of Virology.* 78 (15): 8102–8113.

Pauling L. & Corey R.B. (1951) The structure of synthetic polypeptides. *PNAS*. 37: 241-250.

Patel J., McLeod L.E., Vries R.G., Flynn A., Wang X. & Proud C.G. (2002) Cellular stresses profoundly inhibit protein synthesis and modulate the states of phosphorylation of multiple translation factors. *European Journal of biochemistry.* 269: 3076-3085.

Percy N., Barclay W.S., Garcia-Sastre A. & Palese P. (1994) Expression of a foreign protein in influenza A virus. *Journal of Virology*. 68: 4486-4492

Pevear D.C., Calenoff M., Rozhon E. & Lipton H. L (1987) Analysis of the complete nucleotide sequence of the Picornavirus Theiler's Murine Encephalomyelitis virus indicates that it is closely related to Cardioviruses. *Journal of Virology*. 61: 1507-1516.

Piccone, M.E., Zellner, M., Kumosinski, T.F., Manson, P.W. & Grubman, M.J. (1995) Identification of the active-site residues of the L-proteinase of Foot-and-Mouth-disease virus. *Journal of Virology.* 69 (8): 4950-4956.

Piron M., Delaunay T., Grosclaude J. & Poncet D. (1999) Identification of the RNA-binding, dimerization and eIF4GI-binding domains of rotavirus non-structural protein NSP3. *Journal of Virology*. 73 (7): 5411-5421.

Pisarev A.V., Hellen C.U.T. & Pestova T.V. (2007) Recycling of Eukaryotic Posttermination Ribosomal Complexes. *Cell*. 131 (2): 286-299.

Porter F.W., Bochkov Y.A., Albee A.J., Wiese C. & Palmenberg A.C. (2006) A Picornavirus protein interacts with Ran-GTPase and disrupts nucleocytoplasmic transport. *PNAS*. 15: 12417-12422.

Poul B.T., Tang K.F.J., Pantoja C.R., Bonami J.R. & Lightner D.V. (2006) Purification and characterization of infectious myonecrosis virus of penaeid shrimp. *Journal of General Virology*. 87: 987-996.

Pringle C.**R**. (1999) Virus Taxonomy at the XIth International Congress of Virology, Sydney, Australia. *Archives of Virology*. 144 (10): 1432-8798.

Pringle F.M., Johnson K.N., Goodman C.L., McIntosh A.H. & Ball L.A. (2003) Providence virus: a new member of the *Tetraviridae* that infects cultured insect cells. *Virology*. 306: 359-370.

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 encoded the NS34 equivalent of group A rotaviruses. *Virology*. 184: 752-757.

Richards O. C. & Ehrenfeld E. (1998). Effects of poliovirus 3AB protein on 3D polymerase-catalyzed reaction. *Journal of Biological Chemistry*. 273**:** 12832–12840.

Roberts P.J. & Belshman G.J. (1995)Identification of critical amino-acids within the Foot-and-Mouth disease virus Leader protein A cysteine protease. *Journal of Virology*. 213 (1): 140-146.

Robertson B.H., Grubman M.J., Weddell G.N., Moore D.M., Welsh J.D., Fischer T., Dowbenko D.J., Yansura D.G., Small B. & Kleid D.G. (1985) Nucleotide and amino acid sequence coding for polypeptides of foot-and-mouth disease virus type A12. *Journal of Virology*. 54: 651-660.

Rodnina M.V., Pape T., Savelsbergh A., Mohr D., Matassova N.B. & Wintermeyer W. (2000) Mechanism of partial reactions of the elongation cycle catalyzed by elongation factors Tu and G. In *The Ribosome: structure, function, antibiotics and cellular interactions,* pp.301-317. Edited by Garrett R.A., Douthwaite S.R., Liljas A., Matheson A.T., Moore P.B., Noller H.F. ASM Press, Washington, D.C.

Rodríguez P.L. & Carrasco L. (1993) Poliovirus protein 2C has ATPase/GTPase activities. *The Journal of Biological Chemistry*. 268 (11): 8105-8110.

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

Roos R.P (2002) Pathogenesis of Theiler's Murine Encephalomyelitis Virus-induced disease. In *Molecular biology of picornaviruses,* pp. 427-435. Edited by Semler B.L., Wimmer E. ASM Press, Washington.

Rossmann M.G. (2002) Picornaviruses structure overview. In *Molecular biology of picornaviruses,* pp. 27-39. Edited by Semler B.L., Wimmer E. ASM Press, Washington.

Rospert S. (2004) Ribosome function: governing the fate of a nascent polypeptide. *Current biology*. 14 (10): 386-388.

Ryan M.D., Belshman G.J. & King A.M.Q. (1989) Specificity of enzyme-substrate interactions in foot-and-mouth disease virus polyprotein processing. *Virology*. 173 (1): 35-45.

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

Ryan M.D. & Drew J. (1994). Foot-and-mouth disease virus 2A oligopeptide mediated
cleavage of an artificial polyprotein. *EMBO J*. 134: 928-933.

Ryan M.D. & Flint M. (1997) Virus-encoded proteinases of the picornavirus super-group. *Journal of General Virology.* 78: 699-723.

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

Ryan M.D., Luke G., Hughes L.E., Cowton V.M., ten Dam E., *Li* X., Donnelly M.L.L., Mehrotra A. & Gani D. (2002) The aphtho- and cardiovirus "primary" 2A/2B polyprotein "cleavage". In *Molecular Biology of Picornaviruses,* pp. 213-223. Edited by: Semler BL, Wimmer E . Washington, ASM Press.

Ryan M.D., Donnelly M.L.L., Flint M., Cowton V.M., Luke G., Hughes L.E., Knox C. & de Felipe P. (2004) Foot and mouth disease virus proteinases .In *Foot and mouth disease. Current perspectives*, pp. 53-76. Edited by: Sobrino, F., Domingo, E. Madrid, Spain.

Rueckert R.R. (1996) Picornaviridae: the viruses and their replication. In *Fields Virology*, 3rd edition, pp. 609-645. Edited by Fields B., Knipe D. and Howley P.M. Raven publishers, Philadelhia, USA.

Ryazanov A.G. (1987) Ca^{2+}/calmodulin-dependent phosphorylation of elongation factor 2. *FEBS Letters*. 214: 331-334.

Sandoval I.V. & Carrasco L. (1997) Poliovirus infection and expression of the poliovirus protein 2B provoke the disassembly of the Golgi complex, the organelle target for the antipoliovirus drug Ro-090179. *Journal of Virology*. 71: 4679-4693.

Sarver N. & Stollar V. (1977) Sindbis virus-induced cytopathic effect in clones of *Aedes albopictus* (Singh) cells. *Virology*. 80: 390-400.

Scolnick E., Tompkins R., Caskey T. & Nirenberg M. (1968) Release factors differing in specificity for termination codons. *PNAS*. 61: 768–774.

Sea Urchin Genome Sequencing Consortium (2006) The genome of sea urchin *Strongylocentrotus purpuratus*. *Science*. 314: 941-952.

Shein C.H., Oezguen N.,Volk D.E. & Garimella R. (2006) NMR structure of the viral peptide linked to the genome (VPg) of poliovirus. *Peptides*. 7: 1676-84.

Shrader T., Tobias J.W. & Varshavsky A. (1993) The N-End rule in *Escherichia coli*: Cloning and analysis of the Leucyl, Phenylalanyl-tRNA-protein transferase gene aat. *Journal of Bacteriology*. 175 (14): 4364-4374.

Shlesinger S., Shlesinger M.J. (1996) Togaviridae: the viruses and their replication. . In *Fields Virology*, 3rd edition, pp. 15-57. Edited by Fields B., Knipe D. and Howley P.M. Raven publishers, Philadelhia, USA.

Schultheiss T., Emerson S.U., Purcell R.H. & Gaussmuller V. (1995) Polyprotein processing in echovirus 22 (EV22): A first assessment. *Biochemical and Biophysical Research Communications.* 217 (3): 1120-1127.

Singh K. R. P. (1967). Cell cultures derived from larvae of *Aedes albopictus* (Skuse) and Aedes aegypti (L.) *Current Science in India*. 36: 506-508.

Sioud M. (2007) RNA interference and innate immunity. *Advance drug delivery reviews.* 59 (2-3): 153-163.

Snijder E. J. & Meulenberg J.J.M. (1998) The molecular biology of arteriviruses. *Journal of General Virology*. 79: 961-979.

Song H., Mugnier P, Das A.K., Webb H.M., Evans D.R., Tuite M.F., Hemmings B.A. & Barford D. (2000) The crystal structure of human eukaryotic release factor

eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell*. 100: 311-321.

Somogyi P., Jenner A., Brierley I. & Inglis S.C. (1993) Ribosomal Pausing during Translation of an RNA PseudoKnot. *Molecular and Cellular Biology*. 13 (11): 6931-6940.

Spaan W., Cavanagh D. & Horzinek M.C. (1988) Coronaviruses: Structure and genome expression. *Journal of General Virology.* 69: 2939-2952.

Spector D.H. & Baltimore D. (1974) Requirement of 3' terminal polyadenylic acid for infectivity of poliovirus RNA. *PNAS.* 71: 2983-2987.

Stanway G. & Hyypiä T. (1999) Parechoviruses. *Journal of Virology.* 73 (7): 5249-5254.

Strebel K. & Beck E. (1986) A second protease of foot-and-mouth disease virus. *Journal of Virology*. 58: 893-899.

Strauss D.M., Glustrom L.W. & WuttkeD.S. (2003) Towards an Understanding of the Poliovirus Replication Complex: The Solution Structure of the Soluble Domain of the Poliovirus 3A Protein. *Journal of Molecular Biology*. 330 (2): 225-234.

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

Svitkin Y., Belsham G.J., Pause A. & Sonenberg. N. (1996). Activation of the translational suppressor 4E-BP1 following infection with encephalomyocarditis virus and poliovirus. *PNAS.* 93: 5578–5583.

Svitkin Y.V., Herdy B., Costa-Mattioli M., Gingras A.C., Raught B. & Sonenberg N. (2005).Eukaryotic translation initiation factor 4e availability controls The switch between cap-dependent and internal ribosomal entry site-mediated

translation. *Molecular and Cellular Biology.* 25: 10556-10565.

Szymczak A. & Vignali D. (2005) Development of 2A peptide-based strategies in the design of multicistronic vectors. *Expert Opinion on Biological Therapy*. 5: 627-638.

Ten Dam E., Flint M. & Ryan M.D. (1999) Virus-encoded proteinases of the Togaviridae. *Journal of General Virology.* 80: 1879-1888.

Tenson T. & Ehrenberg M. (2002) Regulatory nascent peptides in the ribosomal tunnel. *Cell*. 108: 591-594.

Tobias J.W. & Varshavsky A. (1991a) The N-end rule in bacteria. *Science*. 254 (5036): 1374-1377.

Tobias J.W. & Varshavsky A. (1991b) Cloning and functional analysis of the Ubiquitin-Specific Protease Gene UBP1 of *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry.* 266: 12021-12028.

Tseng C.H., Knowles N.J. & Tsai H.J. (2007) Molecular analysis of duck hepatitis virus type 1 indicates that it should be assigned to a new genus. *Virus Research*. 123 (2): 190-203.

Tseng C.H. & Tsai H.J. (2007) Sequence analysis of a duck picornavirus isolate indicates that it together with porcine enterovirus type 8 and simian picornavirus type 2 should be assigned to a new picornavirus genus. *Virus Research*. 129 (2): 104-114.

van Eyll O. & T. Michiels T. (2000) Influence of the Theiler's virus L*protein on macrophage infection, viral persistence, and neurovirulence. *Journal of General Virology.* 74: 9071–9077.

van Eyll, O. & Michiels T. (2002) Non-AUG-initiated internal translation of the L* protein of Theiler's virus and importance of this protein for viral persistence. *Journal of Virology.* 76 (21): 10665-10673.

van Kuppeveld F.J.M., de Jong A.S., Melchers W.G.J. & Willems P.H.G.M. (2005) Enterovirus protein 2B po(u)res out the calcium: a viral strategy to survive? *Trends in Microbiology*. 13:41-44.

van Pesch, V., van Eyll, O. & Michiels T. (2001) The leader protein of Theiler's virus inhibits immediate-early alpha/beta interferon production. *Journal of Virology.* 75 (17): 7811-7817.

Vance L.M., Moscufo N., Chow M. & Heinz B.A. (1997) Poliovirus 2C region functions during encapsidation of viral RNA. *Journal of Virology*. 71 (11): 8759-8765.

Vestergaard B., Van L.B., Andersen G.R., Nyborg J., Buckingham R.H. & Kjeldgaard M. (2001) Bacterial polypeptide release factor RF2 is structurally distinct from eukaryotic eRF1. *Molecular Cell*. 8: 1375-1382.

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. *Journal of Virology*. 63:28-35.

Weidman M.K., Yalamanchili P., Ng B., Tsai W. & Dasgupta A. (2001) Poliovirus 3C protease-mediated degradation of transcriptional activator p53 requires a cellular activity. *Virology*. 291 (2): 260-271.

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

Wilson J.E., Powell M.J., Hoover S.E. & Sarnow P. (2000) Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Molecular and Cellular Biology*. 20: 4990-4999.

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. *Journal of General Virolgy.* 77: 1719-30

Yamasaki K., Weihl C.C. & Roos R.P. (1999) Alternative translation initiation of Theiler's murine encephalomyelitis virus. *Journal of Virology*. 73:8519–8526.

Yamashita T., Sakae K., Tsuzuki H., Suzuki Y., Ishikawa N., Takeda N., Miyamura T. & Yamazaki S. (1998) Complete nucleotide sequence and genetic organization of Aichi virus, a distinct member of the *Picornaviridae* associated with acute gastroenteritis in humans. *Journal of Virology*. 72: 8408-8412.

Yamamoto Y., Sunohara T., Jojima K., Inada T. & Aiba H. **(2003)** Ssr-SsrA mediated trans-translation plays a role in mRNA quality control by facilitating degradation of truncated mRNAs. *RNA*. 9: 408-418.

Yang H., Makeyev E.V., Kang Z., Ji S., Bamford D.H. & van Dijk A.A. (2004) Cloning and sequence analysis of dsRNA segments 5, 6 and 7 of a novel non-group A, B, C adult rotavirus that caused an outbreak of gastroenteritis in China. *Virus Research*. 106: 15-26.

Yonath A., Leonard K.R. & Wittman H.G. (1987) A tunnel in the large ribosomal subunit revealed by three-dimensional image reconstruction. *Science*. 236 (4803): 813-816.

Yoshinaka Y., Katoh I., Copeland T.D. & Oroszlan S. (1985) Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. *PNAS*. 82: 1618-1622.

Zamora M., Marissen W.F. & Lloyd R.E. (2002) Multiple eIF4GI-specific protease activities present in uninfected and poliovirus-infected cells. *Journal of Virology*. 76 (1): 165-177.

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. *Journal of Virology*. 75 (4): 1620-1631.

Zell R., Seitz S., Henke A., Munder T. & Wutzler P. (2005) Linkage map of protein–protein interactions of porcine teschovirus. *Journal of General Virology*. 86: 2763-2768.

Zhouravleva G., Frolova L., Le Goff X., Le Guellec R., Inge-Vechtomov S. Kisselev L. & Philippe M. (1995) Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO J*. 14 (16): 4065-4072.

Appendix 1

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	FMDV-63	AY593836	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
2	FMDV-64	AY593831	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
3	FMDV-65	AY593832	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
4	FMDV-75	AY593823	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
5	FMDV-76	AY593824	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
6	FMDV-78	AY593811	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
7	FMDV-79	AY593812	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
8	FMDV-83	AY593828	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
9	FMDV-110	AF506822	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
10	FMDV-128	AJ539138	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
11	FMDV-129	AJ539139	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
12	FMDV-130	AJ539140	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
13	FMDV-131	AJ539141	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
14	FMDV-141	DQ296518	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
15	FMDV-143	DQ296517	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
16	FMDV-149	DQ296514	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
17	FMDV-153	DQ296511	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
18	FMDV-157	DQ296507	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
19	FMDV-159	DQ296506	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
20	FMDV-160	DQ296531	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
21	FMDV-163	DQ296504	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
22	FMDV-164	DQ296528	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
23	FMDV-167	DQ296502	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
24	FMDV-169	DQ296501	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
25	FMDV-170	DQ296524	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P

26	FMDV-178	AY960759	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
27	FMDV-179	AJ633821	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
28	FMDV-180	AF506822	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
29	FMDV-184	AY312587	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
30	FMDV-186	AY333431	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
31	FMDV-205	EU140964	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
32	FMDV-232	EF611987	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
33	FMDV-248	DQ296525	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
34	FMDV-256	DQ404180	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
35	FMDV-257	DQ404179	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
36	FMDV-258	DQ404178	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
37	FMDV-259	DQ404177	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
38	FMDV-260	DQ404176	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
39	FMDV-261	DQ404175	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
40	FMDV-262	DQ404174	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
41	FMDV-263	DQ404173	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
42	FMDV-264	DQ404172	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
43	FMDV-265	DQ404171	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
44	FMDV-266	DQ404169	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
45	FMDV-267	DQ404170	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
46	FMDV-269	DQ404167	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
47	FMDV-270	DQ404166	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
48	FMDV-271	DQ404165	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
49	FMDV-272	DQ404164	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
50	FMDV-273	DQ404163	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
51	FMDV-275	DQ404161	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
52	FMDV-276	DQ404160	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
53	FMDV-277	DQ404159	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
54	FMDV-278	DQ404158	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
55	FMDV-281	DQ767863	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P

56	FMDV-185	AY312589	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	P	N	P	G	P
57	FMDV-138	DQ296522	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	A
58	FMDV-139	DQ296519	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	A
59	FMDV-151	DQ296512	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	A
60	FMDV-161	DQ296505	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	A
61	FMDV-165	DQ296503	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	A
62	FMDV-168	DQ296526	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	A
63	FMDV-155	DQ296508	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	Q	G	A
64	FMDV-137	DQ296523	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	L	G	P
65	FMDV-145	DQ296516	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	L	G	P
66	FMDV-147	DQ296515	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	L	G	P
67	FMDV-166	DQ296527	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	L	G	P
68	FMDV-1	AY593769	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
69	FMDV-2	AY593789	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
70	FMDV-3	AY593767	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
71	FMDV-13	AY593787	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
72	FMDV-14	AY593788	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
73	FMDV-19	AY593768	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
74	FMDV-21	AY593771	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
75	FMDV-22	AY593773	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
76	FMDV-24	AY593775	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
77	FMDV-28	AY593776	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
78	FMDV-33	AY593792	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
79	FMDV-39	AY593765	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
80	FMDV-41	AY593760	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
81	FMDV-43	AY593762	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
82	FMDV-44	AY593763	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
83	FMDV-45	AY593764	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
84	FMDV-46	AY593761	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
85	FMDV-47	AY593766	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P

86	FMDV-115	X74812	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
87	FMDV-144	DQ296543	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
88	FMDV-148	DQ296537	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
89	FMDV-150	DQ296536	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
90	FMDV-152	DQ296535	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
91	FMDV-154	DQ296534	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
92	FMDV-156	DQ296533	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
93	FMDV-158	DQ296532	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
94	FMDV-135	DQ296548	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
95	FMDV-174	DQ063707	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
96	FMDV-176	AJ251476	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
97	FMDV-177	AY960760	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
98	FMDV-210	EF117837	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
99	FMDV-237	EF405981	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
100	FMDV-238	EF405980	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
101	FMDV-239	EF159977	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
102	FMDV-282	DQ767857	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
103	FMDV-27	AY593759	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	P	N	P	G	P
104	FMDV-204	M38362	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
105	FMDV-243	DQ296548	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
106	FMDV-146	DQ296542	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	A
107	FMDV-134	X00871	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
108	FMDV-66	AY593814	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
109	FMDV-67	AY593815	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
110	FMDV-68	AY593816	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
111	FMDV-69	AY593817	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
112	FMDV-70	AY593818	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
113	FMDV-71	AY593819	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
114	FMDV-72	AY593820	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
115	FMDV-73	AY593821	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P

116	FMDV-77	AY593825	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
117	FMDV-80	AY593813	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
118	FMDV-82	AY593827	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
119	FMDV-84	AY593829	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
120	FMDV-85	AY593830	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
121	FMDV-123	AJ320488	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
122	FMDV-133	V01131	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
123	FMDV-192	M95781	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
124	FMDV-190	A00276	H	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
125	FMDV-48	AY593799	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
126	FMDV-114	AY304994	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
127	FMDV-49	AY593800	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
128	FMDV-50	AY593798	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
129	FMDV-200	X88856	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
130	FMDV-211	DQ989323	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
131	FMDV-212	DQ989322	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
132	FMDV-213	DQ989321	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
133	FMDV-214	DQ989320	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
134	FMDV-215	DQ989319	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
135	FMDV-216	DQ989318	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
136	FMDV-217	DQ989317	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
137	FMDV-218	DQ989316	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
138	FMDV-219	DQ989315	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
139	FMDV-220	DQ989314	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
140	FMDV-229	DQ989305	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
141	FMDV-228	DQ989306	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	L	G	P
142	FMDV-55	AY593804	H	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
143	FMDV-56	AY593805	H	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
144	FMDV-119	AJ133358	H	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
145	FMDV-120	AJ133359	H	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P

146	FMDV-121	AJ133357	H	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
147	FMDV-175	X00130	H	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
148	FMDV-208	AM503966	H	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
149	FMDV-209	AM503965	H	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
150	FMDV-38	AY593755	H	K	Q	R	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
151	FMDV-140	DQ296545	H	K	Q	R	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
152	FMDV-142	DQ296544	H	K	Q	R	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	L	G	P
153	FMDV-136	DQ296547	H	K	Q	R	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	A
154	FMDV-244	DQ296547	H	K	Q	R	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	A
155	FMDV-116	AY317098	H	K	Q	K	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
156	FMDV-117	AY359854	H	K	Q	K	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
157	FMDV-111	AF511039	H	K	Q	K	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
158	FMDV-173	DQ248888	H	K	Q	K	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
159	FMDV-183	AY686687	H	K	Q	K	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
160	FMDV-240	EF175732	H	K	Q	K	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
161	FMDV-274	DQ404162	H	K	Q	K	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
162	FMDV-268	DQ404168	H	K	Q	K	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
163	FMDV-17	AY593756	Y	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
164	FMDV-25	AY593751	Y	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
165	FMDV-18	AY593757	Y	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
166	FMDV-132	X00429	Y	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	L	G	P
167	FMDV-255	V01130	Y	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	L	G	P
168	FMDV-298	E00082	Y	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	L	G	P
169	FMDV-299	BD437421	Y	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	L	G	P
170	FMDV-235	AM409190	R	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
171	FMDV-283	DQ409191	R	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
172	FMDV-284	DQ409190	R	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
173	FMDV-287	DQ409187	R	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
174	FMDV-289	DQ409185	R	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
175	FMDV-290	DQ409184	R	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P

176	FMDV-291	DQ409183	R	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
177	FMDV-9	AY593786	H	K	Q	K	I	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
178	FMDV-37	AY593793	H	K	Q	K	I	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
179	FMDV-10	AY593790	H	K	Q	K	I	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
180	FMDV-11	AY593801	H	K	Q	K	I	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
181	FMDV-12	AY593802	H	K	Q	K	I	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
182	FMDV-86	AY593835	H	K	Q	R	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
183	FMDV-87	AY593833	H	K	Q	R	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
184	FMDV-105	AF026168	H	K	Q	R	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
185	FMDV-106	AF154271	H	K	Q	R	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
186	FMDV-108	AF308157	H	K	Q	R	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
187	FMDV-29	AY593781	H	K	Q	K	I	I	A	P	A	R	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
188	FMDV-36	AY593754	H	K	Q	K	I	I	A	P	A	R	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
189	FMDV-34	AY593778	H	K	Q	K	I	I	A	P	A	R	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
190	FMDV-32	AY593780	H	K	Q	K	I	I	A	P	A	R	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
191	FMDV-54	AY593810	H	K	Q	P	L	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
192	FMDV-57	AY593806	H	K	Q	P	L	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
193	FMDV-188	AF274010	R	K	Q	Q	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
194	FMDV-285	DQ409189	R	K	Q	Q	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
195	FMDV-286	DQ409188	R	K	Q	Q	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
196	FMDV-288	DQ409186	R	K	Q	Q	L	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
197	FMDV-224	DQ989310	R	K	Q	E	I	I	A	P	E	K	Q	M	M	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
198	FMDV-206	EF149010	R	K	Q	E	I	I	A	P	E	K	Q	M	M	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
199	FMDV-223	DQ989311	R	K	Q	E	I	I	A	P	E	K	Q	M	M	N	F	D	L	L	K	L	A	G	D	V	E	P	N	P	G	P
200	FMDV-162	DQ296529	R	K	Q	E	I	I	A	P	E	K	Q	M	M	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	A
201	FMDV-6	AY593783	R	K	Q	K	I	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
202	FMDV-7	AY593784	R	K	Q	K	I	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
203	FMDV-8	AY593785	R	K	Q	K	I	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
204	FMDV-26	AY593752	H	K	Q	K	I	I	A	P	G	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
205	FMDV-113	M10975	H	K	Q	K	I	I	A	P	G	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	R	P

206	FMDV-202	J02187	H	K	Q	K	I	I	A	P	G	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	R	P
207	FMDV-53	AY593797	R	K	Q	E	I	I	A	P	E	K	Q	A	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
208	FMDV-118	AY390432	R	K	Q	E	I	I	A	P	E	K	Q	A	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
209	FMDV-221	DQ989313	R	K	Q	E	I	I	A	P	E	K	Q	A	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
210	FMDV-126	AJ539136	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	N	L	L	K	L	A	G	D	V	E	S	N	P	G	P
211	FMDV-127	AJ539137	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	N	L	L	K	L	A	G	D	V	E	S	N	P	G	P
212	FMDV-189	AF167307	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	N	L	L	K	L	A	G	D	V	E	S	N	P	G	P
213	FMDV-227	DQ989307	R	K	Q	K	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
214	FMDV-222	DQ989312	R	K	Q	K	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
215	FMDV-201	X88855	R	K	Q	K	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
216	FMDV-16	AY593753	Y	K	Q	K	I	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
217	FMDV-23	AY593758	Y	K	Q	K	I	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
218	FMDV-30	AY593777	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	Q	L	A	G	D	V	E	S	N	P	G	P
219	FMDV-31	AY593779	H	K	Q	K	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	Q	L	A	G	D	V	E	S	N	P	G	P
220	FMDV-51	AY593795	R	K	Q	E	I	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
221	FMDV-191	AF024509	R	K	Q	E	I	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
222	FMDV-58	AY593807	H	K	Q	P	L	I	A	P	A	K	Q	L	S	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
223	FMDV-60	AY593809	H	K	Q	P	L	I	A	P	A	K	Q	L	S	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
224	FMDV-74	AY593822	H	K	Q	K	I	V	A	P	A	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
225	FMDV-81	AY593826	H	K	Q	K	I	V	A	P	A	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
226	FMDV-101	AY593851	Y	K	T	P	L	V	K	P	E	K	Q	L	C	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
227	FMDV-102	AY593852	Y	K	T	P	L	V	K	P	E	K	Q	L	C	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
228	FMDV-103	AY593853	Y	K	T	P	L	V	K	P	D	K	Q	M	C	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
229	FMDV-104	AB079061	H	K	Q	K	I	V	A	P	V	K	Q	L	L	S	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
230	FMDV-109	AF377945	H	K	Q	K	I	V	A	P	V	K	Q	L	L	S	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
231	FMDV-124	AJ007347	Y	K	Q	P	L	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
232	FMDV-125	AJ007572	Y	K	Q	P	L	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
233	FMDV-293	DQ478937	Y	K	Q	K	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
234	FMDV-294	DQ478936	Y	K	Q	K	I	V	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
235	FMDV-4	AY593770	H	K	Q	K	I	I	A	P	T	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P

236	FMDV-5	AY593782	H	K	Q	K	I	I	A	P	A	K	Q	S	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
237	FMDV-15	AY593803	H	K	Q	S	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
238	FMDV-20	AY593794	Y	K	Q	Q	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
239	FMDV-35	AY593774	H	K	Q	R	I	I	A	P	A	K	Q	L	L	N	F	D	L	L	Q	L	A	G	D	V	E	S	N	P	G	P
240	FMDV-40	AY593772	H	K	Q	K	I	I	A	P	S	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
241	FMDV-42	AY593791	H	K	Q	K	I	I	T	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
242	FMDV-52	AY593796	R	K	Q	E	I	I	A	P	E	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
243	FMDV-59	AY593808	H	K	Q	P	L	I	A	P	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
244	FMDV-61	AY593834	H	K	Q	K	I	V	A	P	T	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
245	FMDV-62	AY593837	Y	K	Q	K	I	V	A	P	V	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
246	FMDV-88	AY593845	Y	K	T	T	L	V	K	P	A	K	Q	L	S	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
247	FMDV-89	AY593846	Y	K	T	A	I	T	K	P	V	K	Q	L	C	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
248	FMDV-90	AY593838	Y	K	T	S	I	V	R	P	A	K	Q	L	C	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
249	FMDV-91	AY593839	Y	K	T	A	I	T	K	P	A	K	Q	M	C	S	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
250	FMDV-92	AY593840	Y	Q	T	A	L	T	K	P	A	K	Q	L	C	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
251	FMDV-93	AY593841	Y	Q	T	A	L	V	R	P	A	K	Q	L	C	N	F	D	L	L	M	L	A	G	D	V	E	S	N	P	G	P
252	FMDV-94	AY593842	H	K	T	A	L	V	K	P	A	K	Q	L	C	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
253	FMDV-95	AY593843	Y	K	T	A	L	V	K	P	A	K	Q	L	C	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
254	FMDV-96	AY593844	Y	K	V	S	L	V	A	P	E	K	Q	M	A	N	F	A	L	L	K	L	A	G	D	V	E	S	N	P	G	P
255	FMDV-97	AY593847	R	F	D	A	P	I	G	V	E	K	Q	L	F	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
256	FMDV-98	AY593848	R	F	D	A	P	I	G	V	E	K	Q	L	C	N	C	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
257	FMDV-99	AY593849	R	F	D	A	P	I	G	V	E	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
258	FMDV-100	AY593850	Y	K	I	K	L	V	A	P	D	K	Q	L	C	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
259	FMDV-107	AF189157	H	K	H	K	I	V	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	M	E	S	N	P	G	P
260	FMDV-112	AF540910	R	F	D	S	P	I	G	V	K	K	Q	L	C	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
261	FMDV-122	AJ251473	R	F	D	A	P	I	G	V	A	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
262	FMDV-171	AY687334	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	F	D	L	L	K	L	S	G	D	V	E	S	N	P	G	P
263	FMDV-172	AY687333	R	K	Q	E	I	I	A	P	A	K	Q	M	M	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
264	FMDV-182	AY881014	R	K	Q	K	I	I	A	P	G	K	Q	V	M	N	F	D	L	L	K	L	A	G	D	V	E	L	N	P	G	P
265	FMDV-187	AY145897	H	K	Q	K	I	V	A	P	V	K	Q	L	L	N	F	E	L	L	K	L	A	G	D	V	E	S	N	P	G	P

266	FMDV-193	X88862	R	H	N	E	D	C	A	T	L	E	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
267	FMDV-194	X88861	R	H	K	E	D	C	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
268	FMDV-195	X88863	R	H	N	E	D	C	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
269	FMDV-196	X88860	T	Q	T	G	D	H	C	T	S	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
270	FMDV-197	X88859	K	Q	T	E	D	H	C	T	N	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
271	FMDV-198	X88858	T	Q	T	E	D	H	C	T	S	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
272	FMDV-199	X88857	R	K	Q	E	I	I	A	P	K	K	Q	V	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
273	FMDV-207	EF149009	R	K	Q	K	I	I	A	P	E	K	Q	T	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
274	FMDV-225	DQ989309	R	K	Q	K	I	I	A	P	E	K	Q	M	M	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
275	FMDV-226	DQ989308	R	K	Q	E	I	I	A	P	E	K	Q	M	M	N	F	E	L	L	K	L	A	G	D	V	E	S	N	P	G	P
276	FMDV-230	DQ989304	R	K	Q	E	I	I	A	P	E	K	Q	D	L	N	L	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
277	FMDV-231	DQ989303	R	K	Q	E	I	I	A	P	E	K	Q	V	L	N	L	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
278	FMDV-233	EF600684	R	K	Q	K	I	I	A	P	G	K	Q	A	L	N	F	D	L	L	K	L	A	G	D	V	E	L	N	P	G	P
279	FMDV-234	AM409325	R	K	Q	P	L	V	A	P	A	K	Q	L	L	N	F	G	L	L	K	L	A	G	D	V	E	S	N	P	G	P
280	FMDV-236	EF405982	H	K	Q	K	I	I	A	P	V	K	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
281	FMDV-279	DQ119643	H	K	Q	K	I	V	A	P	A	K	Q	S	L	N	F	D	L	L	R	L	A	G	D	V	E	S	N	P	G	P
282	FMDV-292	DQ533483	R	K	Q	E	I	I	A	P	E	K	Q	A	L	N	F	D	L	L	E	L	A	G	D	V	E	S	N	P	G	P
283	ERAV-393/76	L43052	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	S	L	L	K	L	A	G	D	V	E	S	N	P	G	P
284	ERAV-Plowright	DQ272127	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	S	L	L	K	L	A	G	D	V	E	S	N	P	G	P
285	ERAV-U188	DQ272128	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	S	L	L	K	L	A	G	D	V	E	S	N	P	G	P
286	ERAV-T3	DQ268580	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	S	L	L	K	L	A	G	D	V	E	S	N	P	G	P
287	ERAV-T10	DQ272577	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	S	L	L	K	L	A	G	D	V	E	S	N	P	G	P
288	ERAV-PERV	X96870	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	A	L	L	K	L	A	G	D	V	E	S	N	P	G	P
289	ERAV-PERV-1	DQ272578	R	H	K	F	P	T	N	I	N	K	Q	C	T	N	Y	A	L	L	K	L	A	G	D	V	E	S	N	P	G	P
290	BRV2		L	R	L	T	G	E	I	V	K	Q	G	A	T	N	F	E	L	L	Q	Q	A	G	D	V	E	T	N	P	G	P
291	ERBV-1-P1436/71	X96871	E	A	T	L	S	T	I	L	S	E	G	A	T	N	F	S	L	L	K	L	A	G	D	V	E	L	N	P	G	P
292	ERBV-2-P313/75	AF361253	V	A	D	W	E	N	L	L	S	Q	G	A	T	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
293	PTV-1-5-D-VIII	AF296106	A	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
294	PTV-1-D 61/96	AY392535	A	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
295	PTV-1-PS 34	AF296105	A	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P

296	PTV-1-Sek 549/98	AF296101	A	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
297	PTV-1-Sek 655/97	AY392553	A	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
298	PTV-1-Talfan-A	AF231769	A	M	T	T	L	S	Y	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
299	PTV-1-Talfan-B	AB038528	A	M	T	T	L	S	Y	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
300	PTV-1-swine/CH/IMH/03	Q355222	A	M	T	T	L	S	Y	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
301	PTV-1-F65	AJ011380	A	M	T	V	M	A	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
302	PTV-1-Sek 65/97	AY392552	A	M	T	V	M	A	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
303	PTV-1-Sek 1042/97	AY392555	A	M	T	V	M	A	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
304	PTV-1-Teschen-Konratic	AF231768	A	M	T	T	M	S	Y	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
305	PTV-1-Teschen-Tirol	AF296097	A	M	T	T	M	S	Y	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
306	PTV-1-Vir 1626/89	AF296103	D	M	T	R	M	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
307	PTV-1-Vir 1627/89	AF296104	D	M	T	R	M	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
308	PTV-1-Sek 2498/96	AY392551	A	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	I	E	E	N	P	G	P
309	PTV-1-Sek 736/97	AY392554	A	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	I	E	E	N	P	G	P
310	PTV-1-DS 562/91	AF296100	A	M	T	T	M	T	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
311	PTV-1-Teschen-Bozen 65	AF231767	A	M	T	T	I	S	Y	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
312	PTV-1-Teschen-199	AF296098	A	M	T	V	V	T	Y	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	I	E	E	N	P	G	P
313	PTV-1-IBRSV-VII	AF296099	D	M	T	R	L	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
314	PTV-1-Vir 2236/99	AF296102	A	M	T	A	M	A	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
315	PTV-1-DS 1520/93	AY392532	D	M	T	V	M	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
316	PTV-1-RD 181/01	AY392536	N	M	A	R	M	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
317	PTV-2-Vir 480/87	AF296109	A	M	T	T	M	S	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
318	PTV-2-Vir 6711-12/83	AF296107	A	M	T	T	M	S	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
319	PTV-4-Vir 918-19/85	AF296111	A	M	T	T	M	S	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
320	PTV-2-DS 183/93	AY392533	A	M	T	T	M	T	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	I	E	E	N	P	G	P
321	PTV-2-DS 756/93	AY392534	A	M	T	T	M	T	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	I	E	E	N	P	G	P
322	PTV-2-12-PL	AY392541	T	M	T	T	M	S	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
323	PTV-2-2-AK-III	AY392542	T	M	T	T	M	S	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
324	PTV-2-T80	AF296087	V	M	T	T	M	M	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
325	PTV-2-Sek 49/99	AF296110	E	M	T	T	M	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P

326	PTV-2-Vir 6793/83	AF296108	A	M	T	T	L	S	L	Q	G	P	G	A	T	N	F	S	L	L	R	Q	A	G	D	V	E	E	N	P	G	P
327	PTV-2-Stendal 2532	AY392537	A	M	T	T	M	M	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
328	PTV-3-O 2b	AF296088	T	M	T	T	M	S	F	Q	G	P	G	A	S	S	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
329	PTV-3-1-AA-VI	AY392540	A	M	T	T	M	T	F	Q	G	R	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
330	PTV-4-PS 36	AF296089	V	M	T	T	M	M	L	Q	G	P	G	A	S	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
331	PTV-4-Vir 3764/86	AF296112	A	M	T	A	L	T	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
332	PTV-4-Vir 2500/99	AF296113	A	M	T	T	L	T	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
333	PTV-5-Vir 1806/89	AF296114	A	M	T	T	M	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
334	PTV-5-F 26	AF296090	A	M	T	T	M	L	F	Q	G	P	G	A	A	N	F	S	L	L	R	Q	A	G	D	V	E	E	N	P	G	P
335	PTV-6-Vir 289/89	AF296116	A	M	T	T	M	M	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
336	PTV-6-21-SZ	AF296117	A	M	T	T	M	M	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
337	PTV-6-Vir 3634/85	AF296115	A	M	T	T	M	M	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
338	PTV-6-121-E-IX	AY392546	A	M	T	T	M	M	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
339	PTV-6-PS 37	AF296091	A	M	T	T	M	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
340	PTV-7-F 43	AF296092	T	M	T	V	V	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
341	PTV-8-UKG 173/74	AF296093	A	L	T	T	M	S	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	I	E	E	N	P	G	P
342	PTV-8-25-T-VII	AF296118	A	L	T	T	M	S	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	I	E	E	N	P	G	P
343	PTV-9-Vir 2899/84	AF296094	A	M	T	T	M	A	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
344	PTV-10-UKG/170/80	AY392539	A	M	T	T	L	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
345	PTV-10-12/15 Ge (1980)	AY392547	A	M	T	T	L	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
346	PTV-10-S 776/83	AY392548	A	M	T	T	L	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
347	PTV-10-VS7/92	AY392549	A	M	T	T	L	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
348	PTV-11-UKG 53/81	AF296120	A	M	T	T	L	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
349	PTV-10-Vir 460/88	AF296095	T	M	T	T	L	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
350	PTV-10-Vir 461/88	AF296119	T	M	T	T	L	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
351	PTV-11-Dresden	AF296096	D	M	T	R	M	S	F	Q	G	P	G	A	T	N	F	S	L	L	K	R	A	G	D	V	E	E	N	P	G	P
352	PTV-11-1008/88	AY392550	D	M	T	R	M	S	L	Q	G	P	G	A	S	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
353	PTV-11-DS 1696/91	AF296121	A	M	T	A	M	A	L	Q	G	P	G	A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
354	EMCV(Rueckert)	M81861	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
355	EMCV	X00463	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P

356	EMCV (EMC-PV21)	X74312	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
357	EMCV (BEL-2887A/91)	AF356822	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
358	EMCV (pEC9)	DQ288856	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
359	EMCV (HB1)	DQ464063	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
360	EMCV (BJC3)	DQ464062	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
361	EMCV (EMCV-CBNU)	DQ517424	V	F	G	L	Y	R	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
362	EMCV (EMC-B)	M22457	V	F	G	L	Y	G	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
363	EMCV (EMC-D)	M22458	V	F	G	L	Y	G	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
364	EMCV (PV2)	X87335	V	F	G	L	Y	S	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
365	EMCV (M)	M37588	V	F	G	L	Y	S	I	F	N	A	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
366	EMCV (EMCV-30)	AY296731	I	F	G	L	Y	R	I	F	S	T	H	Y	A	G	Y	F	S	D	L	L	I	H	D	I	E	T	N	P	G	P
367	MENGO (Rz-pMwt)	DQ294633	V	F	G	L	Y	H	V	F	E	T	H	Y	A	G	Y	F	S	D	L	L	I	H	D	V	E	T	N	P	G	P
368	TMEV (GDVII)	X56019	F	R	E	F	F	K	A	V	R	G	Y	H	A	D	Y	Y	K	Q	R	L	I	H	D	V	E	M	N	P	G	P
369	TMEV	M20562	F	R	E	F	F	K	A	V	R	G	Y	H	A	D	Y	Y	K	Q	R	L	I	H	D	V	E	M	N	P	G	P
370	TMEV (DA TO)	M20301	F	G	E	F	F	R	A	V	R	A	Y	H	A	D	Y	Y	K	Q	R	L	I	H	D	V	E	M	N	P	G	P
371	TMEV (BeAn 8386)	M16020	F	G	E	F	F	K	A	V	R	G	Y	H	A	D	Y	Y	R	Q	R	L	I	H	D	V	E	T	N	P	G	P
372	TLV(NGS910)	AB090161	F	S	D	F	F	K	H	V	R	E	Y	H	A	A	Y	Y	K	Q	R	L	M	H	D	V	E	T	N	P	G	P
373	SAF-V (Saffold virus)	EF165067	F	T	D	F	F	K	A	V	R	D	Y	H	A	S	Y	Y	K	Q	R	L	Q	H	D	V	E	T	N	P	G	P
374	SAF-V (Saffold virus)	AM922293	F	T	D	F	F	K	A	V	R	D	Y	H	A	S	Y	Y	K	Q	R	L	Q	H	D	I	E	A	N	P	G	P
375	LV (174F)	AF327921	Y	F	N	I	M	H	S	D	E	M	D	F	A	G	G	K	F	L	N	Q	C	G	D	V	E	T	N	P	G	P
376	LV (87-012)	AF327920	Y	F	N	I	M	H	S	D	E	M	D	F	A	G	G	K	F	L	N	Q	C	G	D	V	E	T	N	P	G	P
377	LV (87-012G)	EF202833	Y	F	N	I	M	H	S	D	E	M	D	F	A	G	G	K	F	L	N	Q	C	G	D	V	E	T	N	P	G	P
378	LV (145SL)	AF327922	Y	F	N	I	M	H	N	D	E	M	D	Y	S	G	G	K	F	L	N	Q	C	G	D	V	E	S	N	P	G	P
379	LV (M1146)	AF538689	Y	F	K	I	Y	H	D	K	D	M	D	Y	A	G	G	K	F	L	N	Q	C	G	D	V	E	T	N	P	G	P
380	DHV-1 (DRL-62)	DQ219396	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
381	DHV-1 (R85952)	DQ226541	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
382	DHV-1 (F)	EU264072	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
383	DHV-HS	DQ812094	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
384	DHV-HSS	DQ812092	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
385	DHV-1 (03D)	DQ249299	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P

386	DHV-1 (R)	EF585200	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
387	DHV-1 (E53)	EF151313	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
388	DHV-1 (HP-1)	EF151312	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
389	DHV-1 (S)	EF417871	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
390	DHV-1 (ZJ)	EF382778	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
391	DHV-1 (H)	DQ249300	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
392	DHV AP-03337	DQ256132	A	F	E	L	H	L	E	I	E	S	D	Q	F	R	N	V	R	D	L	T	T	E	G	V	E	P	N	P	G	P
393	DHV AP-04114	DQ812093	A	F	E	L	H	L	E	I	E	S	D	Q	F	R	N	V	R	D	L	T	T	E	G	V	E	P	N	P	G	P
394	DHV AP-04203	DQ256134	A	F	E	L	H	L	E	I	E	S	D	Q	F	R	N	V	R	D	L	T	T	E	G	V	E	P	N	P	G	P
395	DHV AP-04009	DQ256133	A	F	E	L	H	L	E	I	E	S	D	Q	F	R	N	V	R	D	L	T	T	E	G	V	E	P	N	P	G	P
396	DHV AP-03337	DQ256132	A	F	E	L	H	L	E	I	E	S	D	Q	F	R	N	V	R	D	L	T	T	E	G	V	E	P	N	P	G	P
397	DHV-1 (JX)	EF093502	A	F	E	L	D	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
398	DHV-1 (A66)	DQ886445	A	F	E	L	D	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
399	DHV-1 (C80)	DQ864514	A	F	E	L	D	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
400	DHV-1 (90D) N-DHV	EF067924	A	F	E	L	H	L	E	I	E	S	D	Q	I	R	N	V	R	D	L	T	T	E	G	V	E	P	N	P	G	P
401	DHV-1 (04G)	EF067923	A	F	E	L	H	L	E	I	E	S	D	Q	I	R	N	V	R	D	L	T	T	E	G	V	E	P	N	P	G	P
402	DHV-1 (JX)	EU371557	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	K	K	K	D	L	T	T	E	G	V	E	P	N	P	G	P
403	DHV-1 (5886)	DQ249301	A	F	E	L	N	L	E	I	E	S	D	Q	I	R	N	K	K	D	L	T	T	E	G	V	E	S	N	P	G	P
404	SVV	DQ641257	R	A	W	C	P	S	M	L	P	F	R	S	Y	K	Q	K	M	L	M	Q	S	G	D	I	E	T	N	P	G	P
405	SePV-1	EU152976	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
406	SePV-1	EU142040	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
407	SePV-1	EU152979	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
408	SePV-1	EU152978	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
409	SePV-1	EU152975	C	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
410	SePV-1	EU152974	C	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	E	G	D	V	E	S	N	P	G	P
411	SePV-1	EU152980	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	I	H	N	V	L	L	D	G	D	V	E	S	N	P	G	P
412	SePV-1	EU152977	S	G	C	F	C	P	L	P	N	V	Y	V	P	P	T	H	N	V	L	L	D	G	D	V	E	S	N	P	R	P
413	ADRV-N	AY632079	F	F	D	S	V	W	V	Y	H	L	A	N	S	S	W	V	R	D	L	T	R	E	C	I	E	S	N	P	G	P
414	ADRV-N (J19)	DQ113901	F	F	D	S	V	W	V	Y	H	L	A	N	S	S	W	V	R	D	L	T	R	E	C	I	E	S	N	P	G	P
415	ADRV-N (B219)	DQ168032	F	F	D	S	I	W	V	Y	H	L	A	N	S	S	W	V	R	D	L	T	R	E	C	I	E	S	N	P	G	P

416	Human-C (V508)	AY941781	G	V	G	Y	P	L	I	V	A	N	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G	P
417	Human-C (V966)	AY941782	G	V	G	Y	P	L	I	V	A	N	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G	P
418	Human-C (Bristol)	AJ132203	G	A	G	Y	P	L	I	V	A	N	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G	P
419	Human-C (V460)	AY941780	G	T	G	Y	P	L	I	V	A	N	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G	P
420	Bovine-C (Shintoku)	L12390	G	I	G	N	P	L	I	V	A	N	S	K	F	Q	I	D	R	I	L	I	S	G	D	I	E	L	N	P	G	P
421	Porcine-C (Cowden)	M69115	G	N	G	N	P	L	I	V	A	N	A	K	F	Q	I	D	K	I	L	I	S	G	D	V	E	L	N	P	G	P
422	EoPV-2A1	AY365064	G	Q	R	T	T	E	Q	I	V	T	A	Q	G	W	A	P	D	L	T	Q	D	G	D	V	E	S	N	P	G	P
423	EoPV-2A2	AY365064	T	R	G	G	L	Q	R	Q	N	I	I	G	G	G	Q	R	D	L	T	Q	D	G	D	I	E	S	N	P	G	P
424	EoPV-2A1	AY341824	G	Q	R	T	T	E	Q	I	V	T	A	Q	G	W	A	P	D	L	T	Q	D	G	D	V	E	S	N	P	G	P
425	EoPV-2A2	AY341824	T	R	G	G	L	Q	R	Q	N	I	I	G	G	G	Q	R	D	L	T	Q	D	G	D	I	E	S	N	P	G	P
426	PnPV-2A1	AF323747	G	Q	R	T	T	E	Q	I	V	T	A	Q	G	W	V	P	D	L	T	V	D	G	D	V	E	S	N	P	G	P
427	PnPV-2A2	AF323747	T	R	G	G	L	R	R	Q	N	I	I	G	G	G	Q	K	D	L	T	Q	D	G	D	I	E	S	N	P	G	P
428	IFV	AB000906	P	S	I	G	N	V	A	R	T	L	T	R	A	E	I	E	D	E	L	I	R	A	G	I	E	S	N	P	G	P
429	ABPV (U.K.)	AF150629	T	G	F	L	N	K	L	Y	H	C	G	S	W	T	D	I	L	L	L	L	S	G	D	V	E	T	N	P	G	P
430	ABPV (Poland-1)	AF486073	T	G	F	L	N	K	L	Y	H	C	G	S	W	T	D	I	L	L	L	L	S	G	D	V	E	T	N	P	G	P
431	ABPV (Hungary-1)	AF486072	T	G	F	L	N	K	L	Y	H	C	G	S	W	T	D	I	L	L	L	W	S	G	D	V	E	T	N	P	G	P
432	KBV	AY275710	I	G	F	L	N	K	L	Y	K	C	G	T	W	E	S	V	L	N	L	L	A	G	D	I	E	L	N	P	G	P
433	IAPV	EF219380	I	G	F	L	N	K	L	Y	R	C	G	D	W	D	S	I	L	L	L	L	S	G	D	I	E	E	N	P	G	P
434	CrPV	AF218039	L	V	S	S	N	D	E	C	R	A	F	L	R	K	R	T	Q	L	L	M	S	G	D	V	E	S	N	P	G	P
435	DCV (EB)	AF014388	Q	G	I	G	K	K	N	P	K	Q	E	A	A	R	Q	M	L	L	L	L	S	G	D	V	E	T	N	P	G	P
436	TaV	AF062037	R	G	P	R	P	Q	N	L	G	V	R	A	E	G	R	G	S	L	L	T	C	G	D	V	E	E	N	P	G	P
437	TaV	AF282930	R	G	P	R	P	Q	N	L	G	V	R	A	E	G	R	G	S	L	L	T	C	G	D	V	E	E	N	P	G	P
438	EeV	AF461742	R	R	L	P	E	S	A	Q	L	P	Q	G	A	G	R	G	S	L	V	T	C	G	D	V	E	E	N	P	G	P
439	PrV-2A1	AF548354	L	E	M	K	E	S	N	S	G	Y	V	V	G	G	R	G	S	L	L	T	C	G	D	V	E	S	N	P	G	P
440	PrV-2A2	AF548354	N	S	D	D	E	E	P	E	Y	P	R	G	D	P	I	E	D	L	T	D	D	G	D	I	E	K	N	P	G	P
441	PrV-2A3	AF548354	T	L	M	G	N	I	M	T	L	A	G	S	G	G	R	G	S	L	L	T	A	G	D	V	E	K	N	P	G	P
442	D. punctatus CPV1	AY163248	M	T	A	F	D	F	Q	Q	A	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P
443	D. punctatus CPV1	AY185594	M	T	A	F	D	F	Q	Q	A	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P
444	L. dispar CPV1	AF389466	M	T	A	F	D	F	Q	Q	A	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P
445	B. mori	AF433660	R	T	A	F	D	F	Q	Q	D	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	I	E	S	N	P	G	P

446	B. mori CPV1-H	AB035733	R	T	A	F	D	F	Q	Q	D	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	I	E	S	N	P	G	P
447	B. mori CPV1-I	AB035732	R	T	A	F	D	F	Q	Q	D	V	F	R	S	N	Y	D	L	L	K	L	C	G	D	I	E	S	N	P	G	P
448	O. brumata CPV18	DQ192245	I	H	A	N	D	Y	Q	M	A	V	F	K	S	N	Y	D	L	L	K	L	C	G	D	V	E	S	N	P	G	P
449	*T.brucei*	CAA29181	R	S	L	G	T	C	K	R	A	I	S	S	I	I	R	T	K	M	L	V	S	G	D	V	E	E	N	P	G	P
450	*T.brucei*	CAD21861	R	S	L	G	T	C	Q	R	A	I	S	S	I	I	R	T	K	M	L	L	S	G	D	V	E	E	N	P	G	P
451	*T.brucei*	CAD21860	R	S	L	G	T	C	Q	R	A	I	S	S	I	I	R	T	K	M	L	L	S	G	D	V	E	E	N	P	G	P
452	*T.congolense*	354h04.q1k_6	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
453	*T.congolense*	335b10.q1k_3	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
454	*T.congolense*	432g10.q1k_7	I	L	P	C	T	C	G	C	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	V	E	R	N	P	G	P
455	*T.congolense*	400g12.q1k_4	I	L	P	C	T	C	G	R	T	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
456	*T.congolense*	876g11.p1k_3	I	L	P	C	T	C	G	R	T	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
457	*T.congolense*	1381h11.q1k_4	I	V	P	C	T	C	G	R	T	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
458	*T.congolense*	1071g10.p1k_14	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	F	L	L	P	V	R	G	D	V	G	R	N	P	G	P
459	*T.congolense*	1294e07.p1k_3	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	F	L	L	P	V	R	G	D	V	G	R	N	P	G	P
460	*T.congolense*	1473f10.p1k_5	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	F	L	L	P	V	R	G	D	V	G	R	N	P	G	P
461	*T.congolense*	1305b04.p1k_2	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	F	L	L	P	V	R	G	D	V	G	R	N	P	G	P
462	*T.congolense*	530f06.q1kbw_10	I	L	P	C	T	C	I	C	P	T	L	E	A	R	R	L	L	V	L	V	S	G	G	I	E	R	N	P	R	P
463	*T.congolense*	1463e05.p1k_0	A	L	S	C	V	C	G	H	G	N	S	L	L	C	R	L	L	L	F	L	S	G	D	V	E	Y	N	P	G	S
464	*T.congolense*	800b12.p1k_3	A	L	S	C	V	C	G	H	G	N	S	L	L	C	R	L	L	L	F	L	S	G	N	V	E	Y	N	P	G	S
465	*T.congolense*	47d01.q1k_6	A	L	S	C	V	C	G	H	G	N	S	L	L	C	R	L	L	L	F	L	S	G	N	V	E	Y	N	P	G	S
466	*T.congolense*	987a11.q1k_0	A	L	S	C	V	C	G	H	G	N	S	L	L	C	R	L	L	L	F	L	S	G	N	V	E	Y	N	P	G	S
467	*T.congolense*	1423d04.p1k_0	T	L	S	C	T	C	G	S	A	L	P	K	A	L	G	P	L	L	L	L	S	R	V	E	D	H	N	P	G	P
468	*T.congolense*	1182h09.q1k_0	F	T	C	T	C	W	R	G	R	A	L	L	C	R	P	F	L	M	P	L	S	G	D	V	G	Q	N	P	E	P
469	*T.congolense*	1264h04.p1k_15	L	L	S	T	C	G	S	A	L	P	K	A	L	R	P	P	L	L	L	L	S	R	D	E	D	H	N	P	G	P
470	*T.congolense*	245e02.p1k_7	T	V	P	P	N	R	Q	C	A	L	Q	E	A	L	R	K	K	L	L	L	C	G	D	V	E	S	N	P	W	N
471	*T.congolense*	372c07.q1k_4	L	R	H	P	N	R	Q	Y	A	L	Q	E	A	L	R	Q	K	F	L	L	C	G	D	V	E	S	N	P	G	P
472	*T.congolense*	endsN14h02.p1k_0	L	R	H	P	N	R	Q	C	A	L	Q	E	A	L	R	Q	K	L	L	L	C	G	D	V	E	A	N	P	G	P
473	*T.congolense*	1158a09.p1k_3	L	R	H	P	N	R	Q	C	A	L	Q	E	A	L	R	Q	K	L	P	L	C	G	D	V	E	A	N	P	G	P
474	*T.congolense*	791f06.q1k_6	L	R	H	P	N	R	Q	C	A	L	Q	E	A	L	R	Q	K	L	L	L	C	G	D	V	E	A	N	P	G	P
475	*T.congolense*	664f07.p1k_5	T	L	S	C	T	C	G	S	A	L	P	K	A	L	R	P	L	L	L	P	S	R	D	V	E	R	N	P	G	P

476	*T.congolense*	1242b10.p1k_5	I	L	P	C	M	C	G	R	A	T	L	D	A	R	R	L	L	L	L	V	S	E	D	I	E	R	N	P	G	P
477	*T.congolense*	668g12.q1k_4	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	V	E	R	N	P	G	P
478	*T.congolense*	276e06.p1k_0	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
479	*T.congolense*	464h12.q1k_4	I	L	P	C	T	C	G	R	A	T	L	D	A	Q	R	I	L	L	L	V	S	G	D	V	E	R	N	P	G	P
480	*T.congolense*	335e01.p1k_0	I	L	P	R	T	C	G	R	A	T	L	D	A	Q	R	I	L	L	L	V	S	G	D	V	K	R	N	P	G	P
481	*T.congolense*	_endsn12b02.p1k_1	I	L	P	C	T	C	G	R	A	T	L	D	A	P	R	I	L	L	L	V	S	G	D	V	E	R	N	P	G	P
482	*T.congolense*	242b03.p1k_7	I	L	P	C	T	C	G	C	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	V	E	R	N	P	G	P
483	*T.congolense*	612c08.p1k_1	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
484	*T.congolense*	1476a12.p1k_1	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	I	E	R	N	P	G	P
485	*T.congolense*	559g05.q1k_4	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	V	E	R	N	P	G	P
486	*T.congolense*	419f08.q1k_1	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	I	L	L	L	V	S	G	D	V	E	R	N	P	G	P
487	*T.congolense*	55d01.p1k_3	T	L	F	C	T	C	G	S	A	L	P	K	A	L	R	P	L	L	L	L	S	R	V	E	D	H	N	P	G	P
488	*T.congolense*	519b02.p1k_8	V	L	P	C	T	C	G	R	A	T	L	D	A	R	R	I	L	L	L	I	S	G	D	V	E	R	N	P	A	P
489	*T.cruzi*	AAA67559	Q	P	Y	T	Y	C	L	R	A	L	C	D	A	Q	R	Q	K	L	L	L	I	G	D	I	E	Q	N	P	G	P
490	*T.cruzi*	CAB41692	Q	R	Y	T	Y	R	L	R	A	V	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
491	*T.vivax*	638g08.p1k_1	I	L	P	Y	T	C	E	C	A	T	L	D	A	L	R	L	L	L	L	T	C	G	D	V	E	R	N	P	G	P
492	*T.vivax*	262a12.p1k_4	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
493	*T.vivax*	1734a06.p1k_4	I	L	P	C	T	C	G	R	A	A	L	D	V	R	R	H	L	L	L	I	I	G	D	V	E	R	N	P	G	P
494	*T.vivax*	720g04.q1k_0	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	T	L	L	L	I	S	G	D	V	E	R	N	P	G	P
495	*T.vivax*	814g01.p1k_9	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	T	L	L	L	I	S	G	D	V	E	R	N	P	G	P
496	*T.vivax*	1814e03.p1k_1	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
497	*T.vivax*	346a10.p1k_3	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
498	*T.vivax*	1198e11.p1k_1	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
499	*T.vivax*	104g02.p1k_0	M	H	P	C	T	R	G	R	A	V	L	D	A	R	R	L	P	L	L	I	S	G	D	V	E	R	N	P	G	P
500	*T.vivax*	961a05.q1k_2	M	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
501	*T.vivax*	1278c04.p1k_6	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
502	*T.vivax*	856h07.q1k_12	I	L	P	C	T	R	G	R	A	T	L	D	A	R	R	P	L	L	L	I	S	G	V	V	E	R	N	P	G	P
503	*T.vivax*	1858e01.q1k_5	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
504	*T.vivax*	158a04.q1k_12	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
505	*T.vivax*	1890c02.p1k_12	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P

506	*T.vivax*	1013b08.p1k_5	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
507	*T.vivax*	786e10.q1k_5	I	L	P	C	T	C	G	R	A	V	S	D	A	R	R	L	L	L	L	I	S	G	D	V	G	R	N	P	G	P
508	*T.vivax*	1875a05.p1k_16	M	L	P	C	A	C	G	R	A	T	L	D	A	R	R	L	T	L	L	V	S	G	D	V	E	R	D	P	G	P
509	*T.vivax*	302f07.p1k_20	I	L	P	C	T	C	E	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
510	*T.vivax*	1681d10.q1k_7	T	L	P	F	A	R	W	H	I	A	L	D	M	R	R	P	L	L	L	I	S	G	D	V	D	S	K	P	G	P
511	*T.vivax*	174f04.p1k_4	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
512	*T.vivax*	1797d11.q1k_3	I	L	P	C	T	C	G	H	A	A	L	D	A	R	R	R	P	L	L	V	G	R	D	V	K	R	N	P	G	P
513	*T.vivax*	1152b09.q1k_1	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	A	S	G	D	V	E	R	N	P	G	P
514	*T.vivax*	992d03.q1k_1	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	A	S	G	D	V	E	R	N	P	G	P
515	*T.vivax*	333f01.p1k_22	I	L	P	R	T	C	G	S	A	T	L	D	A	R	R	R	L	L	L	I	S	G	D	V	E	R	T	P	G	P
516	*T.vivax*	1108e04.q1k_0	I	L	P	C	T	C	G	R	A	T	L	D	V	L	R	L	L	L	L	V	S	G	D	V	E	R	N	S	G	P
517	*T.vivax*	676h09.p1k_0	I	L	P	C	T	C	G	R	A	A	S	D	V	R	R	L	L	L	L	I	G	G	D	A	E	R	N	P	G	P
518	*T.vivax*	580d11.p1k_4	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	A	V	E	R	N	P	G	P
519	*T.vivax*	13b07.q1k_4	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
520	*T.vivax*	697h03.p1k_0	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
521	*T.vivax*	208f09.q1k_0	I	L	P	C	T	C	G	R	A	A	L	D	A	R	R	L	L	L	L	I	S	G	N	V	E	C	N	P	G	P
522	*T.vivax*	1664g12.q1k_4	L	L	P	C	T	C	G	R	A	T	L	D	A	W	R	L	L	L	L	I	C	G	G	V	G	R	N	P	G	P
523	*T.vivax*	615g10.p1k_36	I	L	P	R	T	C	G	S	A	T	L	D	A	R	R	R	L	L	L	I	S	G	D	V	E	R	M	P	G	P
524	*T.vivax*	1296c04.p1k_1	I	L	P	C	T	R	G	R	A	T	L	D	A	R	R	L	L	L	L	V	S	G	G	V	E	R	N	P	G	P
525	*T.vivax*	389g02.q1k_2	I	L	P	C	T	C	G	R	A	M	L	D	A	R	R	L	L	L	L	I	S	V	D	V	E	R	N	P	G	P
526	*T.vivax*	915h11.q1k_1	I	L	P	C	T	C	G	R	A	T	L	G	A	R	R	L	L	L	L	I	S	V	D	V	E	R	N	P	G	P
527	*T.vivax*	395d11.q1k_1	I	L	P	C	A	C	G	R	A	T	L	D	A	R	R	L	L	V	L	I	S	G	D	V	E	R	N	P	G	A
528	*T.vivax*	1090d05.q1k_9	L	L	P	C	T	C	G	R	A	A	L	D	A	R	R	L	L	L	L	I	I	G	G	V	E	R	K	P	G	P
529	*T.vivax*	749d12.p1k_0	I	L	P	R	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	D	G	D	V	E	R	I	P	G	P
530	*T.vivax*	407c12.p1k_11	I	L	P	R	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	D	G	D	V	E	R	I	P	G	P
531	*T.vivax*	1357e07.q1k_4	L	L	A	C	T	C	G	R	A	A	L	D	V	R	R	R	L	L	L	I	S	G	T	V	K	R	N	P	G	P
532	*T.vivax*	396d05.q1k_10	I	L	P	C	T	C	G	H	A	A	L	D	A	R	R	R	L	L	L	I	S	G	D	V	E	R	N	P	G	A
533	*T.vivax*	936e06.q1k_21	I	L	P	C	A	C	G	R	A	A	L	D	A	R	R	L	L	L	L	A	S	G	D	V	G	R	N	P	G	P
534	*T.vivax*	1768f01.q1k_0	M	L	L	C	T	R	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	P	G	P
535	*T.vivax*	1768f01.q1k_7	M	L	L	C	T	R	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	P	G	P

536	*T.vivax*	563c09.q1k_9	M	L	L	C	T	R	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	P	G	P
537	*T.vivax*	169g11.q1k_0	I	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	V	P
538	*T.vivax*	1062d12.p1k_10	K	L	P	C	T	C	R	R	A	A	L	D	A	R	R	L	L	L	L	I	N	G	G	V	E	R	N	P	G	P
539	*T.vivax*	604d12.p1k_1	M	L	L	C	T	R	G	C	A	M	L	D	A	R	R	L	L	L	P	V	R	G	D	V	E	R	N	P	G	T
540	*T.vivax*	1362h08.p1k_4	I	L	P	R	T	C	G	R	A	T	L	D	A	R	R	R	P	L	L	V	G	R	G	V	E	R	N	P	G	P
541	*T.vivax*	915e09.p1k_12	L	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	N	G	D	V	E	R	N	P	G	P
542	*T.vivax*	242g08.q1k_7	I	L	P	C	A	C	G	R	A	T	L	G	A	R	R	L	L	L	L	I	S	G	D	V	E	R	N	P	G	P
543	*T.vivax*	1878e03.p1k_0	A	L	P	C	T	C	G	R	A	A	L	D	A	R	R	L	L	L	L	A	S	G	D	V	E	R	N	P	G	P
544	*T.vivax*	1143a11.p1k_25	M	L	P	C	T	C	G	R	A	T	L	D	A	R	R	L	L	L	L	I	I	G	D	V	E	R	D	P	G	P
545	*T.vivax*	127f05.p1k_10	I	L	P	C	A	C	G	R	A	V	S	D	A	L	R	L	L	L	L	I	S	G	D	V	E	C	N	P	G	P
546	*T.vivax*	696d02.p1k_1	I	L	P	C	T	C	G	R	A	A	L	D	A	Q	W	R	L	L	L	I	F	V	D	A	E	R	N	P	G	P
547	*T.vivax*	395e02.q1k_1	I	L	P	C	T	C	G	R	A	A	L	D	A	Q	W	R	L	L	L	I	F	V	D	A	E	R	N	P	G	P
548	*T.vivax*	1204f09.p1k_16	S	F	L	N	T	S	L	R	V	R	V	R	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
549	*T.vivax*	1669f09.p1k_0	S	S	L	N	T	S	L	R	V	R	V	C	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
550	*T.vivax*	379d02.q1k_1	L	E	K	L	V	E	R	R	T	R	V	C	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
551	*T.vivax*	1208b05.q1k_3	S	S	L	S	T	S	L	R	V	R	V	C	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
552	*T.vivax*	610d02.p1k_6	S	S	L	S	T	S	L	R	V	R	V	C	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
553	*T.vivax*	1368b05.p1k_6	S	S	L	S	T	S	L	R	V	R	L	C	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
554	*T.vivax*	532f08.q1k_24	S	S	L	S	T	S	L	R	V	R	V	C	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
555	*T.vivax*	1322e01.p1k_12	S	S	L	S	T	S	L	R	V	R	V	C	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
556	*T.vivax*	252g03.p1k_8	S	S	L	S	T	S	L	R	V	R	V	C	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
557	*T.vivax*	357e12.q1k_5	S	S	L	S	T	S	L	R	V	R	V	C	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
558	*T.vivax*	1265f12.p1k_12	S	S	L	S	T	S	L	R	V	R	V	C	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
559	*T.vivax*	1660h01.q1k_4	S	S	L	S	T	S	L	R	V	R	V	C	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
560	*T.vivax*	960h06.q1k_1	Y	F	A	C	T	C	E	R	A	A	L	D	A	P	R	L	P	V	L	I	S	G	D	V	E	R	N	P	G	P
561	*T.vivax*	631h12.q1k_13	S	S	L	S	T	S	L	R	V	R	V	C	H	V	G	C	A	L	F	I	S	V	D	V	E	L	N	P	G	P
562	*T.vivax*	1351h04.q1k_31	I	L	P	F	T	C	G	R	A	G	L	D	T	R	R	L	L	L	L	I	S	G	G	V	G	R	N	P	G	P
563	*T.vivax*	87c09.p1k_2	L	L	A	C	T	C	G	R	A	A	L	D	V	R	R	R	L	L	R	I	T	G	T	V	K	R	N	P	G	P
564	*T.vivax*	615c05.p1k_1	I	L	P	F	T	C	G	R	A	G	L	D	T	R	R	L	P	L	L	I	S	G	G	V	G	R	N	P	G	P
565	*T.vivax*	533b05.q1k_2	V	L	P	C	A	C	G	R	A	T	L	D	A	R	R	L	L	L	P	V	G	G	G	V	E	R	N	A	G	P

566	*T.vivax*	938f02.q1k_5	I	L	P	C	T	R	G	R	A	M	L	S	A	R	W	L	L	L	L	I	S	G	G	V	E	R	K	P	G	P
567	*T.vivax*	314e11.q1k_9	L	L	A	C	T	C	G	R	A	A	L	D	V	R	R	R	L	L	L	I	S	G	T	V	K	R	D	P	G	P
568	*T.vivax*	733e05.p1k_4	I	L	P	F	T	C	G	R	A	A	L	D	A	W	R	L	L	L	L	I	G	G	G	V	G	R	N	P	G	P
569	*T.vivax*	73h08.q1k_2	I	L	P	C	L	C	V	H	A	A	S	D	A	R	W	L	L	L	L	I	S	G	D	V	E	R	R	P	C	P
570	*T.vivax*	772h12.q1k_1	N	T	S	L	R	V	L	A	C	C	V	R	R	A	A	A	P	A	V	Y	Q	R	D	V	E	R	K	P	G	P
571	*T.vivax*	390g10.p1k_23	M	L	L	C	T	S	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	S	G	P
572	*T.vivax*	1122f12.p1k_3	L	L	A	C	T	F	G	R	A	A	L	D	E	R	R	R	L	L	R	I	S	G	T	V	K	R	D	P	G	P
573	*T.vivax*	892h02.p1k_5	S	Q	V	R	W	S	N	G	A	E	K	K	V	Q	R	L	L	L	L	S	G	G	D	V	E	R	N	P	G	P
574	*T.vivax*	1351h04.q1k_31	I	L	P	F	T	C	G	R	A	G	L	D	T	R	R	L	L	L	L	I	S	G	G	V	G	R	N	P	G	P
575	*T.vivax*	87c09.p1k_2	L	L	A	C	T	C	G	R	A	A	L	D	V	R	R	R	L	L	R	I	T	G	T	V	K	R	N	P	G	P
576	*T.vivax*	615c05.p1k_1	I	L	P	F	T	C	G	R	A	G	L	D	T	R	R	L	P	L	L	I	S	G	G	V	G	R	N	P	G	P
577	*T.vivax*	533b05.q1k_2	V	L	P	C	A	C	G	R	A	T	L	D	A	R	R	L	L	L	P	V	G	G	G	V	E	R	N	A	G	P
578	*T.vivax*	938f02.q1k_5	I	L	P	C	T	R	G	R	A	M	L	S	A	R	W	L	L	L	L	I	S	G	G	V	E	R	K	P	G	P
579	*T.vivax*	314e11.q1k_9	L	L	A	C	T	C	G	R	A	A	L	D	V	R	R	R	L	L	L	I	S	G	T	V	K	R	D	P	G	P
580	*T.vivax*	733e05.p1k_4	I	L	P	F	T	C	G	R	A	A	L	D	A	W	R	L	L	L	L	I	G	G	G	V	G	R	N	P	G	P
581	*T.vivax*	73h08.q1k_2	I	L	P	C	L	C	V	H	A	A	S	D	A	R	W	L	L	L	L	I	S	G	D	V	E	R	R	P	C	P
582	*T.vivax*	772h12.q1k_1	N	T	S	L	R	V	L	A	C	C	V	R	R	A	A	A	P	A	V	Y	Q	R	D	V	E	R	K	P	G	P
583	*T.vivax*	390g10.p1k_23	M	L	L	C	T	S	G	R	A	M	L	R	A	R	W	L	L	L	L	I	S	G	D	V	E	R	D	S	G	P
584	*T.vivax*	1122f12.p1k_3	L	L	A	C	T	F	G	R	A	A	L	D	E	R	R	R	L	L	R	I	S	G	T	V	K	R	D	P	G	P
585	*T.vivax*	892h02.p1k_5	S	Q	V	R	W	S	N	G	A	E	K	K	V	Q	R	L	L	L	L	S	G	G	D	V	E	R	N	P	G	P
586	STR-20	XP_001196456	S	K	T	D	L	I	S	G	Q	F	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
587	STR-50	GLEAN3_03186	S	K	T	D	L	I	S	G	Q	I	P	H	L	S	E	L	L	L	M	K	S	G	D	V	E	L	N	P	G	P
588	STR-65	GLEAN3_09160	S	K	T	E	L	M	S	G	Q	I	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
589	STR-70	GLEAN3_22394	S	K	T	D	L	I	S	G	Q	I	P	S	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
590	STR-76	GLEAN3_03448	S	K	T	D	L	I	S	G	Q	I	P	P	L	S	K	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
591	STR-81	GLEAN3_21478	S	K	T	D	L	I	S	G	Q	I	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
592	STR-83	GLEAN3_22780	S	K	T	D	L	I	S	G	Q	I	P	P	L	S	E	L	L	L	M	K	S	G	D	V	E	L	N	P	G	P
593	STR-100	GLEAN3_15340	S	K	T	D	L	I	S	G	Q	I	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
594	STR-111	GLEAN3_20436	S	K	T	D	L	I	S	G	Q	F	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
595	STR-127	GLEAN3_23550	S	K	T	D	L	I	S	G	Q	I	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P

596	STR-147	GLEAN3_11433	S	K	T	D	L	I	S	G	Q	I	P	P	L	S	E	L	L	L	L	K	S	G	D	V	E	L	N	P	G	P
597	STR-23	XP_001198729	L	H	P	A	I	L	C	S	A	S	L	C	F	R	P	Y	L	L	L	M	A	G	D	V	E	P	N	P	G	P
598	STR-35	XP_001200466	N	S	S	C	V	L	N	I	R	S	T	S	H	L	A	I	L	L	L	L	S	G	Q	V	E	P	N	P	G	P
599	STR-54	GLEAN3_25204	S	Q	N	I	D	V	L	S	Q	Q	P	Y	L	T	E	L	L	L	V	K	A	G	D	V	E	L	N	P	G	P
600	STR-60	GLEAN3_09111	Q	N	L	D	F	N	L	Y	L	L	M	I	L	L	M	I	L	L	M	R	S	G	D	V	E	T	N	P	G	P
601	STR-67	GLEAN3_08283	P	Q	Q	D	L	Q	G	F	C	L	L	Y	L	L	M	I	L	L	M	R	S	G	D	V	E	T	N	P	G	P
602	STR-82	GLEAN3_25914	T	T	D	D	P	V	V	Q	E	S	T	C	L	P	E	M	I	L	V	K	A	G	D	V	E	Q	N	P	G	P
603	STR-106	GLEAN3_23532	L	H	P	A	I	L	C	S	A	S	L	C	F	R	P	Y	L	L	L	M	A	G	D	V	E	P	N	P	G	P
604	STR-110	GLEAN3_06203	Q	D	L	D	V	K	E	A	D	K	P	H	I	T	Q	S	L	I	L	K	A	G	D	V	E	S	N	P	G	P
605	STR-136	GLEAN3_20380	G	A	V	D	V	V	L	S	Q	Q	P	Y	L	T	E	L	L	L	V	K	A	G	D	V	E	L	N	P	G	P
606	STR-143	GLEAN3_22449	S	R	P	I	L	Y	Y	S	N	T	T	A	S	F	Q	L	S	T	L	L	S	G	D	I	E	P	N	P	G	P
607	STR-1	XP_797143	N	S	T	P	A	A	M	F	V	C	A	F	I	L	I	S	V	L	L	L	S	G	D	V	E	I	N	P	G	P
608	STR-34	XP_001196844	N	S	T	P	A	A	M	F	V	C	V	F	I	L	I	S	V	L	L	L	S	G	D	V	E	I	S	P	G	P
609	STR-24	XP_001196407	S	Q	R	D	L	S	C	S	Q	P	R	T	I	I	L	G	L	I	M	C	A	G	D	V	Q	P	N	P	G	P
610	STR-25	XP_001186348	S	Q	R	D	L	S	C	S	Q	P	R	T	I	I	L	G	L	I	M	C	A	G	D	V	Q	P	N	P	G	P
611	STR-32	XP_001185404	N	S	S	C	V	L	N	I	R	S	T	S	H	L	A	I	L	L	L	L	S	G	Q	V	E	P	N	P	G	P
612	STR-35	XP_001200466	N	S	S	C	V	L	N	I	R	S	T	S	H	L	A	I	L	L	L	L	S	G	Q	V	E	P	N	P	G	P
613	STR-164	XR_025775	N	S	S	C	V	L	N	I	R	S	T	S	H	L	A	I	L	L	L	L	S	G	Q	V	E	P	N	P	G	P
614	STR-27	XP_001185149	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
615	STR-28	XP_001179204	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
616	STR-29	XP_791376	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
617	STR-30	XP_001199602	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
618	STR-31	XP_001200060	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
619	STR-33	XP_001184905	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
620	STR-36	XP_001180489	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	G	P
621	STR-163	XP_001192137	L	C	P	L	D	F	R	S	T	S	L	S	H	L	T	I	L	L	L	L	S	G	Q	V	E	T	N	P	D	P
622	STR-116/160	XR_026225	T	T	C	Q	C	K	A	L	S	V	M	Y	L	T	L	L	L	L	T	N	A	S	D	I	E	L	N	P	G	P
623	STR-40/141	GLEAN3_18025	K	S	C	I	S	Y	Y	S	N	S	T	A	C	F	N	I	E	I	M	C	C	G	D	V	K	S	N	P	G	P
624	STR-55	GLEAN3_24854	G	A	R	I	S	Y	H	P	N	T	T	A	T	F	Q	L	R	L	L	V	S	G	D	V	N	P	N	P	G	P
625	STR-61	GLEAN3_22393	G	A	R	I	R	Y	Y	N	N	S	S	A	T	F	Q	T	I	L	M	T	C	G	D	V	D	P	N	P	G	P

626	STR-89	GLEAN3_19055	G	R	R	I	Q	Y	Y	N	N	S	I	S	T	F	R	S	E	L	L	R	C	G	D	V	E	S	N	P	G	P
627	STR-38	XP_793501	K	T	R	I	P	Y	S	V	N	S	N	A	S	F	Q	L	E	L	L	H	A	G	D	V	H	P	N	P	G	P
628	STR-51	GLEAN3_22449	S	R	P	I	L	Y	Y	S	N	T	T	A	S	F	Q	L	S	T	L	L	S	G	D	I	E	P	N	P	G	P
629	STR-69	GLEAN3_27016	C	R	R	I	A	Y	Y	S	N	S	D	C	T	F	R	L	E	L	L	K	S	G	D	I	Q	S	N	P	G	P
630	STR-133	GLEAN3_00868	K	R	R	I	P	Y	N	P	N	S	T	A	S	F	Q	L	E	L	L	H	A	G	D	V	H	P	N	P	G	P