2A-INDUCED RIBOSOME STALLING

Valèrie Odon

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

2014

2A-induced Ribosome Stalling

A thesis submitted for the Degree of Doctor of Philosopy

By
Valèrie Odon
School of Biology
University of Saint-Andrews

February 2014

1. Candidate's declarations:

I, Valerie Odon, hereby certify that this thesis, which is approximately 51 000 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.

I was admitted as a research student in May 2009 and 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 2009 and 2012.

Date 10th Feb. 2014 signature of candidate

2. Supervisor's declaration:

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 10th Feb. 2014 signature of supervisor

3. Permission for electronic publication:

In submitting this thesis to the University of St Andrews we understand that we are 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. We also understand that the title and the 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 unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. We have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:
(i) Access to printed copy and electronic publication of thesis through the University of St Andrews.

Date 10th Feb. 2014 signature of candidate signature of supervisor

Abstract

Originally 2A was characterised in foot- and -mouth disease virus. Site directed mutagenesis identified a C-terminus consensus motif [D(V/I)ExNPGP] and it is proposed that 2A interacts with the exit tunnel of the ribosome in a way that a specific peptide bond is skipped between the last glycine of 2A and the proline of 2B, thus providing a discontinuity in translation, resulting in release of discrete proteins from one single ORF. 2A was also identified in other picornaviruses, positive, single and double-stranded RNA insect viruses and mammalian rotaviruses. A motif present at the C-terminus of the 2A oligopeptide [D(V/I)ExNPGP] is very highly, though not completely conserved . The sequence upstream of this motif shows, however, no apparent conservation between 2As of different viruses.

In this study, extensive site-directed mutagenesis were performed on several 2A sequences and a series of 'hybrid' 2As comprising different consensus motifs juxtaposed with different upstream contexts were created as part of a detailed analysis of the mechanism of 2A-mediated ribosome stalling. The results demonstrated that a minimal region of twenty to twenty-three amino acids interacts with the exit tunnel of the ribosome to bring about a pause in processivity, alter the peptidyl transferase centre geometry and restrict the ribosome A site *via* two distinctive stalling mechanisms. Other molecular analyses tested here will require further optimisations or alternative methods: a visual method to explore the dynamics of re-initiation of translation from proline codon, purification of the translation-regulating factors and structural resolution of 2A sequences.

Previously, cellular 2As were identified in non-LTR retrotransposons of trypanosomes. It is reported here as part of two other cellular organisms *Saccoglossus kowalevskii* (acorn worm) and *Branchiostoma floridae* (amphioxus). In the acorn worm, the nucleotides sequences corresponding to 2A motifs were part of the untranslated genome. In amphioxus, three 2A elements were identified in hypothetical proteins, and at the N-terminus of twenty non-LTR retrotransposons.

Table of contents

Chapter 2 Materials and generic methods

Chapter 3 Analysis of re-initiation following ribosome stalling

List of abbreviations

aa	amino acid
BCA	bicinchoninic acid assay
BHK21	baby hamster kidney cells
bp	base pair
CherryFP	cherry fluorescent protein
DHV	duck hepatitis virus
DMEM	dulbecco's modified eagle medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
eEF2	elongation factor 2
eIF4	initiation factor 4
eRF1 and 3	release factor 1 and 3
FCS	fetal calf serum
GFP	green fluorescent protein
GUS	beta-glucuronidase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Luria Bertani broth
MES	2-(N-morpholino) ethanesulfonic acid
[^{35}S]-Met	radiolabelled [^{35}S]- methionine
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MW	molecular weight
NC	negative control
NMR	nuclear magnetic resonance
nt(s)	nucleotide(s)
OD	optic density
o/n	overnight
ORF	open reading frame
PBS	phosphate buffered saline
PC	positive control
PCR	polymerase chain reaction
PEI	polyethileneimine
PTC	peptidyl transferase centre
RNA	ribonucleic acid
rpm	revolutions per minute
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TAE	tris base, acetic acid and EDTA buffer
TeV	tobacco etch virus protease
Tm	melting temperature
TnT	transcription and translation
U	units
uORF	upstream open reading frame
UV	ultraviolet
v/v	volume for volume
w/v	weight for volume
XGal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

Virus species and abbreviations

ABPV	*Acute bee paralysis virus*
ADRV	*Non A,B, C novel adult diarrhea virus*
BmCPV-1	*Bombyx mori cypovirus 1*
BoRV-C	*Bovine rotavirus C*
CrPV	*Cricket paralysis virus*
DcpCPV-1	*Dendrolimus punctatus cypovirus 1*
DCV	*Drosophila C virus*
DHV	*Duck hepatitis virus 1*
EeV	*Euprosterna eleasa virus*
EMCV	*Encephalomyocarditis virus*
EoPV	*Ectropis oblique picorna-like virus*
ERAV	*Equine rhinitis A virus*
ERBV	*Equine rhinitis B virus*
FMDV	*Foot-and-mouth disease virus*
HuRV-C	*Human rotavirus C*
IAPV	*Israeli acute paralysis virus*
IFV	*Infectious flacherie virus*
IMNV	*Penaeid shrimp infectious myonecrosis virus*
KBV	*Kashmir bee virus*
LdCPV-1	*Lymantria dispar cypovirus 1*
LV	*Ljungan virus*
OpbuCPV-18	*Operophtera brumata cypovirus 18*
PnPV	*Perina nuda picorna-like virus*
PoRV-C	*Porcine rotavirus C*
PrV	*Providence virus*
PTV	*Porcine teschovirus*
SAF-V	*Saffold virus*
TaV	*Thosea asigna virus*
TMEV	*Theiler's murine encephalomyelitis virus*

List of tables

List of figures

Chapter 1 Introduction

The 2A element is a peptide able to stall the eukaryotic ribosome during the course of its translation and prevent the formation of a peptide bond between a specific glycine and proline pair at the C-terminus. Surprisingly, the ribosome having skipped the formation of this peptide bond is able to resume translation of the remaining RNA sequence past the C-terminal glycine of 2A. This gives rise to two discrete proteins from a single open reading frame (ORF).
A model of action has previously been proposed to explain 2A-induced ribosome stalling and the purpose of this body of work was to further define and elaborate this model.

2A was termed after the genomic region of the viruses where it was first identified; the 2A region of the picornaviruses foot-and-mouth disease virus (FMDV), encephalomyocarditis virus (EMCV) and Theiler's murine encephalomyelitis virus (TMEV). The C-terminal consensus motif of 2A [D(V/I)ExNPGP] was later identified encoded by the genomes of several other single-stranded and double-stranded RNA viruses.

The introduction begins with a section on the occurrence of the 2A motif in viruses. The second section of the introduction summarises the experimental data that lead to the elaboration of the model for the mechanism of action of 2A. This section is followed with a review of the elongation step on the ribosome and a description of four others well-studied nascent ribosome-stalling peptides.

1.1 Terminology and numbering of 2A

In this thesis, for simplicity all 2A and 2A-like elements will be referred to as 2A, preceded by the initial of the virus in which it occurs, and if appropriate the length of the 2A sequence tested.
e.g.: DHV20 2A the twenty-residues-long-2A sequence from Duck hepatitis virus (DHV).
The numbering of residues follows the N- to C- numbering adopted in the literature for other ribosome-arresting peptides : SecM, arginine attenuator peptide (AAP), *TnaC*, (Ito and Nakatogawa, 2002, Gong et al, 2007 and Vasquez-Laslop et al., 2008). In the context of 2A the numbering begins at the first residue of the 2A sequence.
e.g.:

```
Residue number          1        10        20        30
                        ....|....|....|....|....|....|
FMDV30  2A              RHKEDCAPVKQLLNFDLLKLAGDVESNPGP
```

Based on sequence alignments (Luke et al., 2008) and for the purpose of site-directed mutagenesis (chapter 4 and 5) it has been decided to distinguish two regions in the 2A sequences: the variable region also called the upstream context (underlined in the example above) and the consensus motif, which is more conserved (in blue in the above example).

Throughout this text the ribosome numbering follows *E. coli* numbering consistent with the ribosome-arresting peptide literature.

1.2 Occurrence of the ribosome-arrest 2A element in viruses

Viruses consist of RNA or DNA genome encapsulated within a virion. They lack the translational machinery necessary to replicate and are obligate parasites of cellular systems. The Baltimore classification divides all viruses into seven categories based upon genome type and mode of replication (Baltimore, 1971). Positive-stranded RNA viruses, which are the classes of virus relevant to this thesis, have an RNA genome acting like cellular mRNAs. They are translated directly in the cytoplasm of the host cell. Each family of RNA viruses showcase distinctive strategies to express and process their proteins.

The 2A-induced ribosome stalling is one of the strategies employed by several virus species from the families *Picornaviridae, Dicistroviridae, Tetraviridae, Iflaviridae, Reoviridae* and *Totiviridae*. The mechanism permits the segregation of two virally encoded proteins without the need for proteolytic activity.

1.2.1 Occurrence in *Picornaviridae*

1.2.1.1 Picornaviruses

'Pico' means small and 'rna' refers to the type of genome. Picornaviruses are infectious agents to human and animals. The classification of the *Picornaviridae* family has been amended extensively in the past four years as newly-identified viruses have been included. The picornavirus study group publishes online (http//www.picornaviridae.com, 2013) the most up-to-date classification. In 2013, to reflect the changes approved by the International Committee on Taxonomy of Viruses (Adams et al., 2013), the *Picornaviridae* family is organised into the seventeen genera presented in table 1.1.

Their genome is 7500 to 8500 nucleotides (nts) long and organised in a single ORF encoding a polyprotein (figure 1.1). This single precursor is sequentially processed into final-mature proteins, twelve in the case of FMDV. The 5' end of the picornavirus RNA is the Leader region and comprises 600 to 1100 nts (Palmenberg, 1990). The ORF has three distinct regions called P1, P2 and P3. The underlying rationale for this subdivision lies in the structural or enzymatic functions of the viral proteins encoded. The ORF constitutes 85 to 90 % of the coding capacity, and the remaining 10 to 15 % are shared between the 3' and the 5' region (Palmenberg *et al*., 2009). The 3' region is between 40 to 120 nts long (Palmenberg, 1990).

The unprocessed polyprotein would be a large protein of about 250 kDa, but is not observed in cultures. It is processed sequentially into precursors and then individual viral proteins by a proteolytic cascade involving a primary and secondary series of events (Palmenberg, 1992). The primary processing releases the structural from the replication precursors. The purpose of the secondary processing, performed by the virally encoded 3C protease ($3C^{pro}$) is to cleave each individual protein unit from the precursors.

Table 1.1: The picornaviruses genera, species and serotypes

(From Fernandez- Miragall *et al*., 2009, Adams et al., 2013 and http//www.picornaviridae.com, 2013)

Picornaviruses using the 2A-induced ribosome stalling mechanism during the course of their polyprotein processing are highlighted in red showing the serotypes relevant to this thesis.

genus	species	relevant serotypes
Aphthovirus	*Foot-and-mouth disease virus*	Foot-and-mouth disease virus O, A, C, Asia 1, SAT 1, SAT 2 and SAT 3
	Bovine rhinitis A virus	Bovine rhinitis A virus 1and 2
	Bovine rhinitis B virus	Bovine rhinitis B virus 1
	Equine rhinitis A virus	Equine rhinitis A virus 1
Aquamavirus	*Aquamavirus A*	Aquamavirus A
Avihepatovirus	*Duck hepatitis A virus*	Duck hepatitis A virus 1 to 3
Cardiovirus	*Encephalomyocarditis virus*	Encephalomyocarditis virus 1
		Encephalomyocarditis virus 2
	Theilovirus	Theiler's murine encephalomyelitis virus
		Vilyuisk human encephalomyelitis virus
		Thera virus
		Saffold virus 1 to 9
Cosavirus	*Cosavirus A, B, C and D*	
Dicipivirus	*Cadicivirus A*	
Enterovirus	*Enterovirus A, B, C, D, E, F, G, H, J*	
	Rhinovirus A, B, C	
Erbovirus	*Equine rhinitis B virus*	Equine rhinitis B virus 1 to 3
Hepatovirus	*Hepatitis A virus*	
Kobuvirus	*Aichi virus A, B, C*	
Megrivirus	*Melegrivirus A*	
Parechovirus	*Human parechovirus*	Human parechovirus 1 to 14
	Ljungan virus	Ljungan virus 1 to 4
Salivirus	*Salivirus A*	
Sapelovirus	*Porcine sapelovirus*	
	Simian sapelovirus	
	Avian sapelovirus	
Senecavirus	*Seneca Valley virus*	Seneca Valley virus 1
Teschovirus	*Porcine teschovirus*	Porcine teschovirus 1 to 11
Tremovirus	*Avian encephalomyelitis virus*	

In the 5' region, the RNA has a VPg protein capping the 5'end (figure 1.1) attached to the 5' uridylyl nucleotide of the RNA through a tyrosine residue to form a phosphodiester bond (Forss and Schaller, 1982 and Steil et al., 2010). Other key features of the 5' region are the presence of a cloverleaf structure and an internal ribosome entry site (IRES). This structural element is the key factor controlling the viral-RNA translation by its ability to recruit cellular ribosomes (Kolupaeva et al., 1998).

From the coding region, region P1 yields all the proteins (VP1, VP2, VP3 and VP4) required for the virion formation. The middle part of the picornaviral RNA yields peptides 2A, 2B and 2C. The P3 region yields four final proteins *via* a series of active intermediates 3AB, 3B ^{VPg}, 3CD^{pro}, 3C^{pro} and the polymerase 3D^{pol} (Palmenberg, 1990). Picornavirus non-structural proteins are involved in viral genome replication, cellular shut-down and cellular membranes shuffle. 2A protein is highly variable amongst *Picornaviridae* and is involved in primary processing of the viral polyprotein. 2B protein is a transmembrane protein able to form pores, alter membranes permeability and therefore disturb calcium homeostasis (Sandoval and Carrasco, 1997). In *Enterovirus* 2B disrupts membrane permeability in the endoplasmic reticulum and Golgi apparatus (Sanchez-Martinez et al., 2012). It is not established at present if all the picornaviruses 2B functions similarly. 2C has ATPase/GTPase activity and is an essential element of the replication process (Rodriguez and Carrasco, 1993). The 2C and the precursor 2BC bind viral RNA to cellular vesicles. 2C has been shown to participate in the encapsidation step of the viral genome (Vance et al., 1997). 3A disrupts the traffic between endoplasmic reticulum and Golgi apparatus thereby inhibiting host cellular responses such as production of interferon, interleukins and cytokins mainly by releasing intracellular calcium (Dodd et al., 2001). The host range is defined essentially by 3A (Lama et al., 1998). 3B or VPg is covalently bound to the next viral RNA and is a primer for initiation of replication (Schein et al., 2006). 3C^{pro} is a chymotrypsin like protease responsible for most of the polyprotein cleavage (Palmenberg, 1992). 3C also takes part outside of the viral protein processing to cellular disruption by cleaving eIF4A, eIF4GI and II, PABP, and degrading p53, thereby controlling cell apoptosis (Blair et al., 1998, Belsham et al., 2000, Barco et al., 2000, Weidman et al., 2001). 3D^{pol} is a RNA-dependent RNA polymerase, which shows much specificity to each virus type and is the crucial element in viral replication, the enzyme lacks proof-reading and frequent errors induced during replication allow the virus flexible adaptation (Kerkvliet et al., 2010). The cleavage intermediates 2BC, 3AB and 3CD^{pro} have additional functions. 2BC creates small cellular vesicles, 3AB induce 3D^{pol} activity and 3CD^{pro} participate in polyprotein processing and enhances replication by binding the cloverleaf structure at the 5'UTR (Chase and Semler, 2012).

The last region of the viral RNA, the 3' region is similar to eukaryotic mRNA as it hosts a poly(A) tail, thought to be involved on the stability of the RNA translation through forming a loop back, binding to the ribosomal initiation complex (Steil et al., 2010).

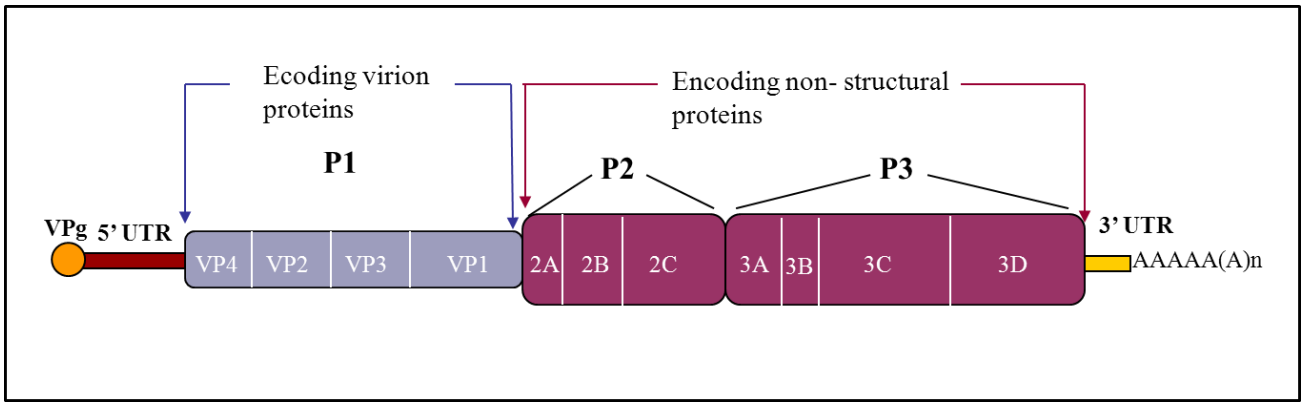


Figure 1.1: The generic genomic organisation of *Picornaviridae*

The positive-sense RNA genome has a small protein VPg at the 5' end (encoded by 3B). The 5' UTR contains an internal ribosome entry site (IRES). The 3' end has a poly (A) tail. $3C^{pro}$ carries out most of the proteolytic processing. The genome is organised in three regions P1, 2 and 3. The schema indicates the peptides encoded by each region (written in white).

1.2.1.2 The 2A product in *Picornaviridae*

Picornaviral RNA is translated once the IRES has successfully recruited a ribosome. Eukaryotic initiation factor 4 (eIF4) is a principal target for most picornaviruses. The genus *Enterovirus* uses the 2A protease ($2A^{pro}$) to cleave eIF4 (Novoa and Carrasco, 1999). Members of the *Aphthovirus* and *Cardiovirus* genera shut cellular-mRNA translation down by the activity of their leader protein. For apthoviruses, the leader protein is a protease L^{pro} which is able to cleave eIF4G (Gradi et al., 2004). For cardioviruses, there is no proteolysis but dephosphorylation of eIF4E, in which case mRNA cap recognition is inhibited (Jen and Thach, 1982). The cleavage of eIF4 family serves the purpose of releasing the central component eIF4G most useful for virus translation as it assists in recruitment of 43S ribosome subunit by IRESes type I and II (Daijogo and Semler, 2011). *Hepatitis A virus* however has a type III IRES and can recruit 43S subunit without the need for eIF4G. This virus does not inhibit host-cellular translation (Cohen et al., 2002).

Regions of the picornavirus ORF encodes for proteins with similar functions across genera. The 2A locus however is an exception. For the genus *Hepatovirus*, the 2A protein is cleaved from VP4 late in the replication. The primary processing is performed by $3C^{pro}$ (the only protease this genus encodes) at the 2A/2B junction. Cohen and colleagues (2002) performed serial N- and C-terminal deletions of the 2A sequence. It was found that 40 % of the N-terminal sequence was necessary for infectivity whereas the C-terminal (the remaining 60 %) of the sequence was required for adequate cleavage of VP1/2A by an elusive cellular protease. For this genus, the 2A product is involved in virion assembly and maturation.

Human parechovirus bioinformatics analyses showed that the 2A product is related to the cellular protein H-rev107 involved in cell-growth regulation. The conserved features of this protein are a long hydrophobic domain and an Asn-Cys-Glu motif (Hughes and Stanway, 2000). The 2A from *Human parechovirus* was then called 'H-box 2A' and is also identified in *Kobuvirus*, *Megrivirus* and *Tremovirus* (http//www.picornaviridae.com, 2013). For this category of 2A, $3C^{pro}$ acts between the VP1 and 2A.

In*Enterovirus*, the 2A product is a protease. The primary cleavage has two events, the self-processing cleavage by $2A^{pro}$ at its C-terminus and the release of the 3ABC precursor by $3C^{pro}$. Apart from performing the primary processing of the picornavirus polyprotein, $2A^{pro}$ also cleaves eIF4 (inhibition of cap-dependent translation) and nucleoporins (inhibition of mRNA transport to cytoplasm) (Redondo et al., 2011). *Sapelovirus* also has this type of 2A protein.

The 2A product for *Dicipivirus* and *Salivirus* has not been characterised to date.

Finally, 2A for *Aphthovirus* and *Cardiovirus*, as well as other genera (*Aquamavirus*, *Avihepatovirus*, *Erbovirus*, *Senecavirus*, *Teschovirus* and the species Ljungan virus from genus *Parechovirus*) highlighted in red in the table 1.1, is characterised by the C-terminally conserved motif [D(V/I)ExNPGP] and is the focus of this thesis. The primary processing involves a scission at the 2A/2B junction, by ribosome- stalling and re-initiation of translation (Luke et al., 2008).

Table 1.2: Viruses employing 2A-induced ribosome stalling.

Family	Genus	Species	reference virus for 2A studies
Picornaviridae	*Aphthovirus*	*Foot-and-mouth disease virus*	FMDV-O1K (X00871)
		Bovine rhinitis A virus	
		Bovine rhinitis B virus	
		Equine rhinitis A virus	
	Aquamavirus	*Aquamavirus A*	
	Avihepatovirus	*Duck hepatitis A virus*	Duck hepatitis virus 1 (DQ219396)
	Cardiovirus	*Encephalomyocarditis virus*	Encephalomyocarditis virus (Ruckert) (M81861)
			Mengovirus (DQ294633)
		Theilovirus	Theiler's murine encephalomyelitis virus (NGS910_AB)
	Erbovirus	*Equine rhinitis B virus*	
	Parechovirus	*Ljungan virus*	
	Senecavirus	*Seneca Valley virus*	
	Teschovirus	*Porcine teschovirus*	
Dicistroviridae	*Aparavirus*	*Acute bee paralysis virus*	Acute bee paralysis virus (AF150629)
		Israeli acute paralysis virus	
		Kashmir bee virus	
	Cripavirus	*Cricket paralysis virus*	Cricket paralysis virus (AF218039)
		Drosophila C virus	
Iflaviridae	*Iflavirus*	*Ectropis oblique picorna-like virus*	
		Infectious flacherie virus	Infectious flacherie virus (AB000906)
		Perina nuda picorna-like virus	
Tetraviridae	*Permutotetravirus*	*Thosea asigna virus*	Thosea asigna virus (AF062037)
		Euprosterna eleasa virus	
	Cormotetravirus	*Providence virus*	Providence virus (AF548354)
Reoviridae	*Rotavirus*	*Human rotavirus C*	
		Porcine rotavirus C	Porcine rotavirus C (Cowden) (M69115)
		Bovine rotavirus C	
		Non A,B,C novel adult diarrhea virus	novel adult diarrhea virus (AY632079)
	Cypovirus	*Bombyx mori cypovirus 1*	*Bombyx mori* cypovirus 1(AF433660)
		Lymantria dispar cypovirus 1	
		Dendrolimus punctatus cypovirus 1	
		Operophtera brumata cypovirus 18	
Totiviridae	*Giardavirus*	*Penaeid shrimp infectious myonecrosis virus*	Penaeid shrimp infectious myonecrosis virus (AF323747)

1.2.2 Occurrence of the ribosome-stalling 2A sequence in other RNA viruses

Worldwide efforts in genome sequencing provided an unprecedented opportunity to compare and align viral genomes. Using local alignment algorithms programmed with the [D(V/I)ExNPGP] consensus motif previously identified in *Picornaviridae*, revealed that outside aphthoviruses and cardioviruses other mammalian and insect RNA viruses contained the 2A motif (Luke et al., 2008). The classification for these viruses is provided in table 1.2. The 2A elements were cloned and tested in the generic expression-system (figure 1.6). They were able to separate a single ORF into two discrete products with varying efficiencies (summarised in table 1.3) (Luke et al., 2008). A nucleotide sequence coding for 2A was discovered in the genomes of positive single stranded RNA virus families: *Dicistroviridae*, *Iflaviridae* and *Tetraviridae*. Figure 1.2 provides the position of 2A in relation to these virus genomes.
The virus classifications reported throughout this section follows the latest taxonomy released by the International Committee on Taxonomy of Viruses in 2012 (http//www.ictv.org).

The *Dicistroviridae* family consists of two genera: *Aparavirus* and *Cripavirus* infecting invertebrates such as bees and cricket worldwide (De Miranda et al., 2010).
The generic dicistrovirus genome is organised in two non-overlapping ORFs encoding for two distinct polyproteins. The structural proteins (products of the ORF2) of insect picorna-like viruses have many common features with picornaviruses. However, dicistroviruses are distinguished by the position of their structural protein-coding sequences at the 3' rather than the 5' end as for picornaviruses and iflaviruses. Dicistroviruses also have an IRES preceding each ORF (De Miranda et al., 2010). Three out of five species of *Aparavirus* have a NPGP 2A motif: *Acute bee paralysis virus* (ABPV), *Israeli acute paralysis virus* and *Kashmir bee virus* (KBV). Two species out of nine of *Cripavirus* have a 2A, *Drosophila C virus* (DCV), and *Cricket paralysis virus* (CrPV). The 2A for these viruses is located at the N-terminus of the replicative ORF1. The glycine- proline of the [NPGP] motif of 2A is found at residue 96 in DCV and residue 166 in ABPV and CrPV polyproteins (Luke et al., 2008). The N-terminal region of the replicative polyprotein encodes an RNA helicase (Govan et al., 2000). RNA helicases are a family of proteins involved in a wide range of RNA-processing events such as mRNA degradation and RNA editing. The RNA helicases have a core region of 290 to 360 aa homologous to initiaton factor 4A, however the amino and carboxy termini are of variable length and sequence (Soto-Rifo and Ohlman, 2013). 2A is located within the RNA helicase. The sequence upstream of the 2A motif is thought to be involved in the modulation of the host immune response by interfering with the formation of double-stranded RNA complexes targeting the viral RNA for degradation (Luke et al., 2008). CrPV 2A was tested to be only 88 % efficient and ABPV 94 % (Luke et al, 2008).

Iflavirus is the only genus in the *Iflaviridae* family. There are seven species currently classified and 2A was found in the genome of three of these: *Infectious flacherie virus* (IFV), *Peruna nuda picorna-like virus* (PnPV), *Ectropis oblique picorna-like virus* (EoPV). The iflaviruses are the causative agents of flacherie disease for invertebrates such as honey bees and silkworm (Ribiere et al., 2010).
The iflaviruses have a single ORF encoding the capsid protein from the 5' genomic region and non-structural proteins from the 3' region (Wu et al., 2002).
Purified EoPV particles were analysed by SDS-PAGE and only two protein masses were detected: 31.5 kDa and 28.8 kDa. Further bioinformatic analyses of the particle proteins revealed homologies to several other ssRNA viruses-virion proteins (Wang et al., 2004). The position of 2A coincides with two junctions. In PnPV and EoPV, 2A is found between VP4 and VP1 (at residues 572- 575) and between the last capsid protein and the helicase of the non-replicative precursor (at residues 1189-1192). Both 2A sequences were highly efficient *in vitro* (Luke et al., 2008). It is highly likely that the two virion particles detected correspond to processing events induced by the 2A sequences. The 28.8 kDa protein was detected in twice greater amounts than the 31.5 kDa (Wang et al., 2004). The phylogenetic study published by Wang and colleagues indicates that EoPV, PnPV are closely related to IFV. These three virus species share a common organisation and some sequence similarities. There are several strains of IFV reported and potential new members for the *Iflavirus* genus, however only one sequence is available to date (GenBank accession number AB000906 (Isawa et al., 1998). IFV has one 2A element at position 1081 to 1084 aa segregating the capsid precursors from non-structural proteins precursor.

The *Tetraviridae* family taxonomy has been reassessed extensively. The current classification (http://www.ictvonline.org) recognises three genera: *Cormotetravirus*, *Permutotetravirus* and *Alphatetravirus*. *Permutotetravirus* consists of two species *Europrosterna elaeasa virus* (EeV) and *Thosea asigna virus* (TaV), both containing a 2A sequence. *Cormotetravirus* has a single species, *Providence virus* (PrV) known to have three 2A sequences.
Tetraviruses have been isolated from lepidopterans such as moths and butterflies (Gordon and Waterhouse, 2010). The permutotetraviruses have two ORFs in a linear genome. The first ORF encodes the polymerase while the second ORF encodes the capsid proteins (Zeddam et al., 2010). The second ORF was derived by bioinformaric analyses from the aa and corresponding nucleotides sequences of the viral particle. ORF2 overlaps by 520 nts the first ORF. Translation from the second ORF AUG codon requires a +1 frameshift (Pringle et al., 2001). The 2A motif in EeV and TaV is found on the second ORF. TaV particles were isolated from *Setothosea asigna* larvae from Sumatra (Pringle et al., 2001). Analysis on SDS-PAGE of the particle shows that the virion is assembled from proteins of 58.3 kDa (L protein) and 6.8 kDa (S protein). The expression of these two proteins on separate vectors failed to replicate any virus. Pringle and colleagues (2001) concluded that the L and S proteins are joined and cleaved after translation during virion synthesis; the L and S aa sequences

correspond to the C-terminal segment of the capsid precursor encoded by the ORF2. The N-terminus encodes for a third protein of 17 kDa (p17) of unknown function and not incorporated in the virion. In TaV and EeV, the 2A sequence is found between p17 and the capsid precursor (Zeddam et al., 2010). Providence virus has a monopartite genome but encodes three ORFs (Walter et al., 2010). The first ORF encodes for a protein of 130 kDa (p130) of unknown function. The second ORF encodes for a 40 kDa protein (p104 and in figure 1.2 the replicase polyprotein) in a +1 frame relative to p130. The third ORF begins 4 nts after p104 stop codon. ORF3 encodes for the capsid polyprotein p81 from a read-through event at the second ORF stop codon. PrV has three 2A elements. PrV-2A1 is at the N-terminus of p130 and produces fragments of 17 and 113 kDa. PrV-2A2 and PrV-2A3 are at the N-terminus of p81. The two 2A elements process p81precursors into fragments of 7, 8 and 68 kDa. The capsid precursor is the 68 kDa which undergoes auto-proteolysis to yield the peptides of 60 and 7.4 kDa composing the virion (Walter et al., 2010). The 7 and 8 kDa segments resulting from PrV-2A2 and 2A3 processing have not been studied.

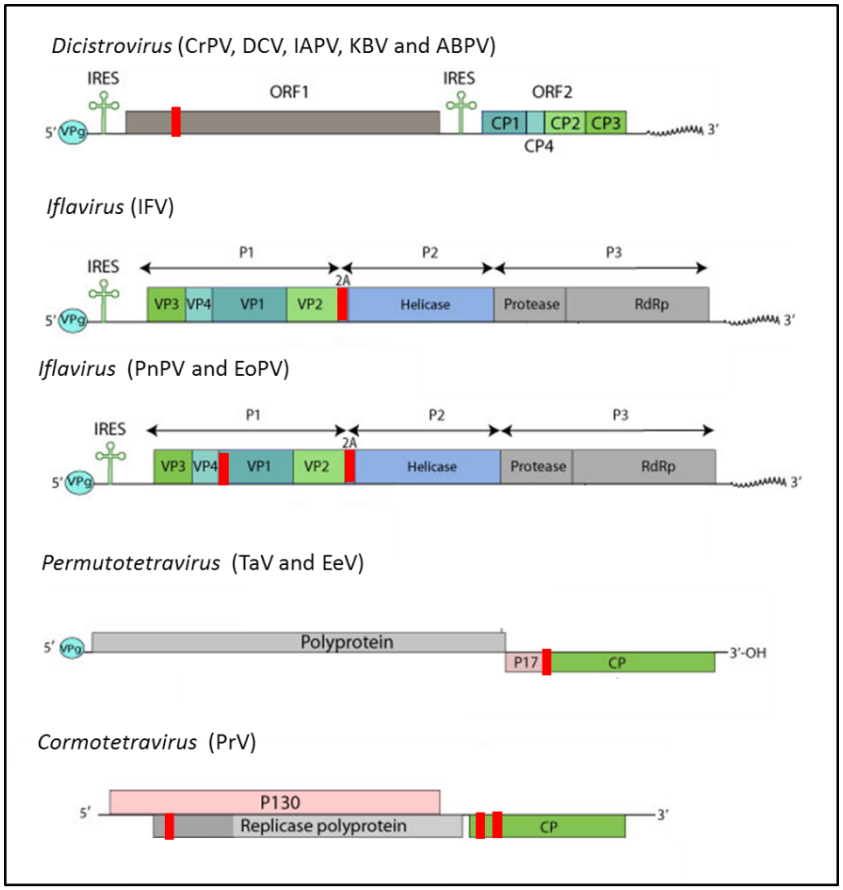


Figure 1.2: Location of 2A coding sequences in the genomes of other ssRNA (+) viruses

The 2A is marked in red. The key features of the genomes are provided. The proteins sequences associated with virion formation (VP or CP) are coloured in shades of green, the grey boxes represent the replicative domains, blue and pink represent other coded proteins. (Graphics are reproduced and adapted from http://www.viralzone.expasy.org, using information from Luke et al., 2008).

In addition two families of double-stranded RNA viruses, *Reoviridae* and *Totiviridae*, have members coding for a 2A sequence.

Reoviruses have segmented genomes. Figure1.3 gives an illustration of the position of the 2A motif in the relevant genomic segment. Reoviruses are organised in two subfamilies: *Sedoreovirinae* and *Spinareovirinae*.

Sedoreovirinae consists of six genera. It includes the genus *Rotavirus* further organised in five species: *Rotavirus A, B, C, D,* and *E*. Rotaviruses have eleven segments encoding six virion forming proteins and five replicative proteins (James et al., 1999). Species of the genus *Rotavirus C*: *Bovine rotavirus C* (BoRV), *Porcine rotavirus C* (PoRV) and *Human rotavirus C* (HuRV) encode a 2A sequence in their segment 6 (NSp3). In addition, 2A was identified in segment 5 (NSp1) of the rotavirus type non -A,B, C *novel adult diarrhea virus* (ADRV) (Luke et al., 2008).
Rotaviruses mRNAs lack 3' poly(A) endings. NSp3 protein enhances rotavirus translation by its ability to circularise the viral mRNAs (Jayaram et al., 2004). When the NSp3 segment of porcine rotavirus C was expressed in COS-1 cells, three products were observed by SDS-PAGE analysis a 45 kDa full-length product, a 38 kDa and a 8kDa product (Langland et al., 1994). Further analysis of the NSp3 aa sequences for type C rotaviruses revealed three regions: a ssRNA binding protein region at the N-terminus, a initiation factor 4G binding region in the mid region, the 2A motif followed by a dsRNA binding protein region (Luke et al., 2008). In other types of rotaviruses, the NSp3 protein is shorter and lacks the 2A-dsRNA binding domain of type C rotavirus (James et al., 1999).
NSp1 is the most variable protein in rotaviruses and targets the host interferon response (Arnold and Patton, 2011). Although NSp1 of ADRV is unrelated to NSp3 of type C rotaviruses, both of these segments contain downstream of 2A a dsRNA binding protein.

The subfamily *Spinareovirinae* is organised in nine genera and includes the genus *Cypovirus*, which is further categorised in sixteen species, *Cypovirus 1* to *16*. Four viruses *Bombyx mori* cypovirus 1 (BmCPV-1), *Lymantria dispar* cypovirus 1 (LdCPV-1), *Dendrolimus punctatus* cypovirus 1(DcpCPV-1) and *Operophtera brumata* cypovirus 18 (OpbuCPV-18), encode a 2A in their segment 5. The segment 5 for BmCPV-1 is 881 aa long and was analysed *in vitro* (Hagiwara et al., 2001). Three bands were detected by SDS-PAGE analysis corresponding to proteins of 107 kDa (p107), 80 kDa (p80) and 23 kDa (p23). The fragments of lesser molecular weight constitute the processed products of p107. Hagiwara and colleagues (2001) identified a 2A motif between residues 219 and 235 but failed to identify any other domains in p80 and p23.

The family of *Totiviridae* comprises five genera, amongst these in the genus *Giardavirus*, the virus *Penaeid shrimp infectious myonecrosis virus* (IMNV) has two 2A sequences at the N-terminus of ORF1(Luke et al., 2008). Totiviruses have non segmented dsRNA genomes, with two ORFs that overlap (Nibert, 2007). ORF1 encodes the capsid protein and ORF2 the polymerase. The polyprotein processing of ORF1 results in the production of three products of different sizes, 93, 284 and 1228 amino acids (Poulos et al., 2006). The extreme 93 aa N-terminal fragment is a dsRNA binding protein terminated with a 2A sequence (Nibert, 2007). Totiviruses were originally isolated from fungi but in 2006 IMNV was isolated in aquacultures of shrimps in Brazil and Indonesia (Poulos et al., 2006). In IMNV, the capsid protein starts half-way through the ORF1 and is preceded by the dsRNA binding protein, which is an unusual feature for totiviruses. New totiviruses were isolated from drosophila cultures and mosquitoes. These new isolates retain the dsRNA binding protein-2A feature previously observed in IMNV (Isawa et al., 2010).

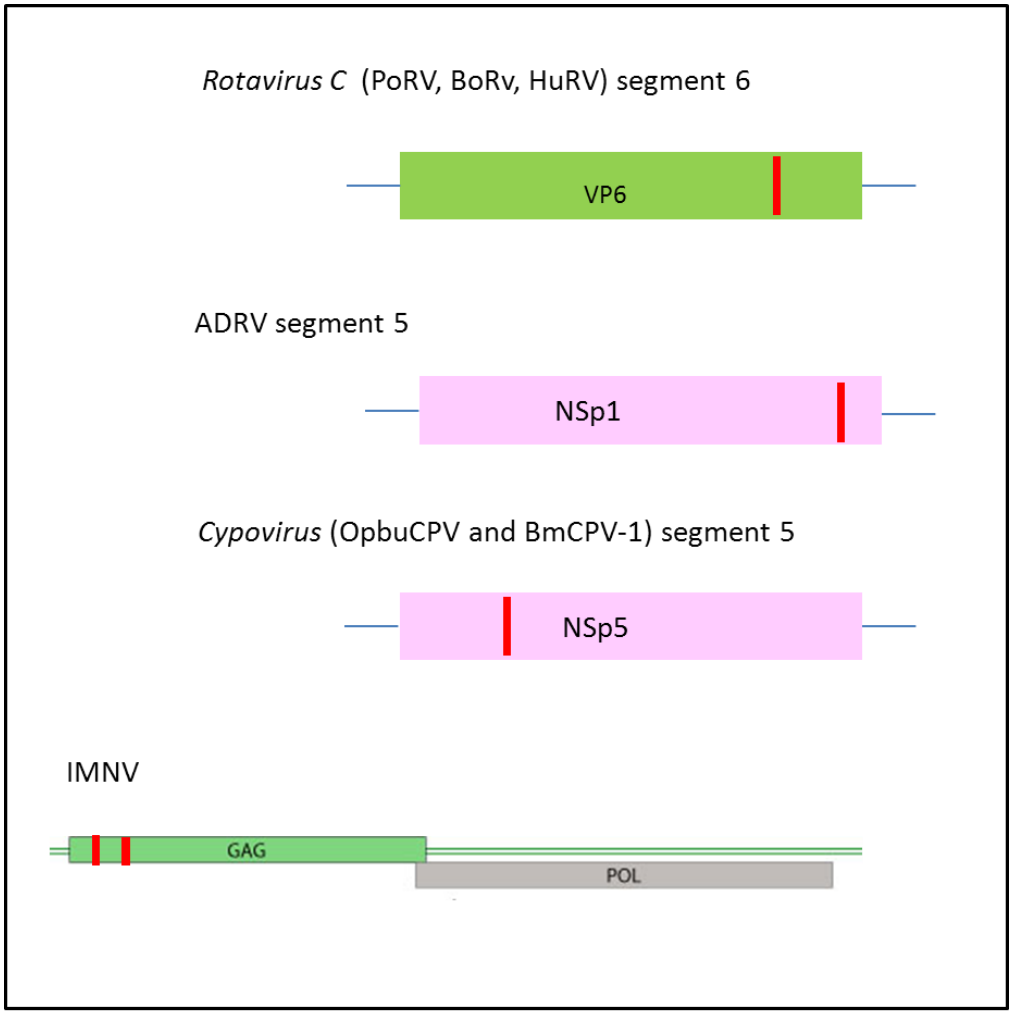


Figure 1.3: Location of the 2A coding sequence in the genome of dsRNA viruses

2A is marked in red, the rotaviruses, and cypoviruses genomes are segmented and only the relevant segment is shown. The capsid protein is represented in green, the replicative protein in grey and other proteins in pink, the ORF in boxes.

Sequence alignments proved that the only conserved part was the 2A motif used for identification, as the upstream context is very diverse. Previous analyses of the residues by positions ruled out any regularity or patterns (Luke et al., 2008). The table 1.2 was created with a representative 2A-like sequence for each virus previously identified with a 2A motif in their genome and provides the 2A sequences tested to date and their efficiencies.

The evolutionary analyses performed by Luke an co-workers (2008) suggested that 2A could have emerged independently or could have been acquired vertically. Luke and co-workers (2008) used the polymerase domain to construct a phylogenetic tree relating the various viruses which were identified with a 2A sequence (figure 1.4). They included in their analysis the sequence of closely related viruses. The analysis showed that for cypoviruses, rotaviruses and totiviruses, 2A was probably acquired independently. The authors argued that given the short length of the motif, and high mutational rate of the viruses, multiple acquisition of 2A is likely to occur. Members of the cardioviruses and aphthoviruses genera however may have acquired their 2A from a common ancestor.

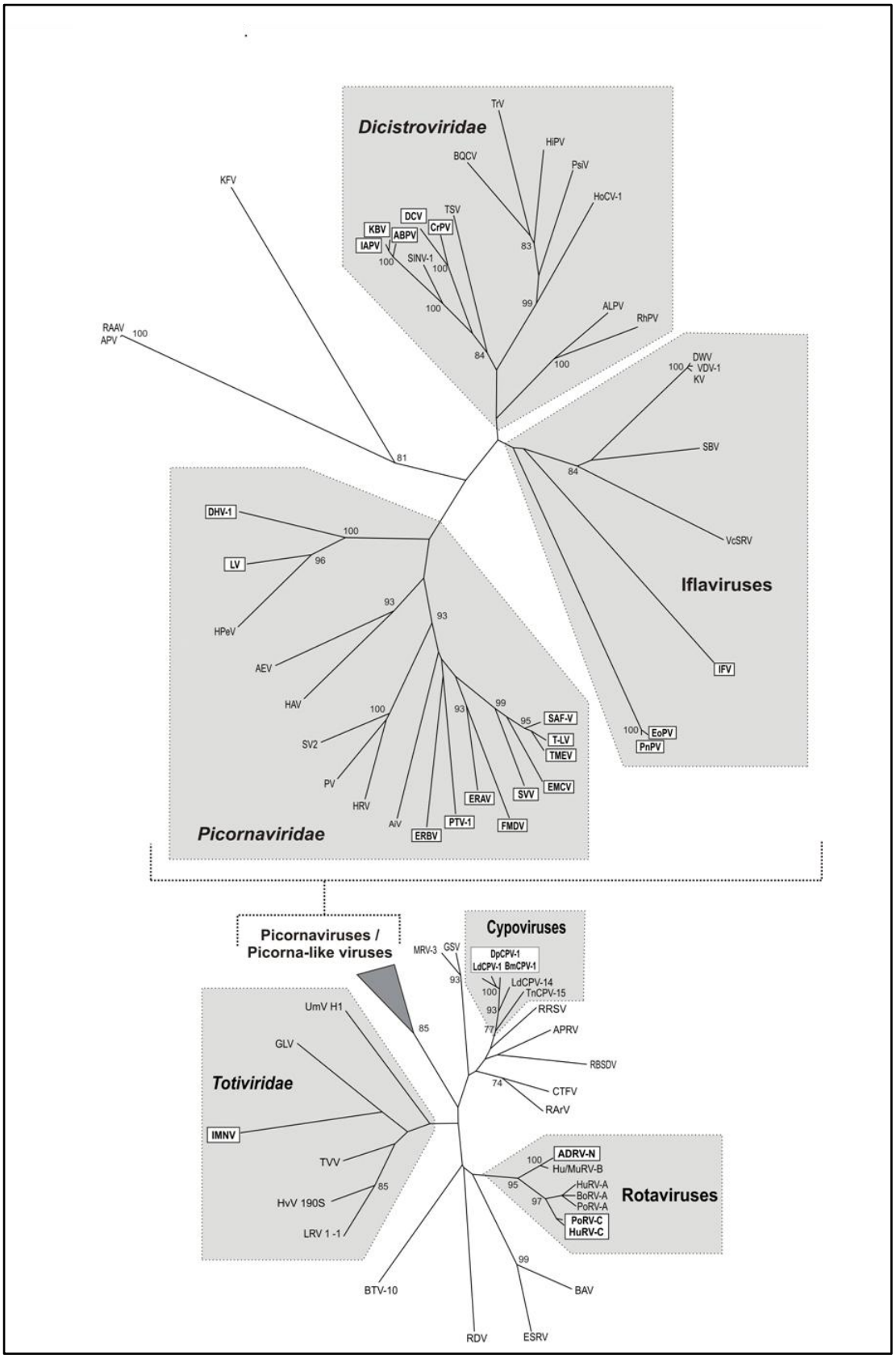


Figure 1.4: Phylogenetic tree showing occurrence of 2A in diverse virus groups

(provided by Prof. Ryan as published in Luke et al., 2008.)

The analysis was based on alignment of the RNA-dependent RNA polymerase domain. The viruses with a 2A element in their genome are indicated in boxes and the virus group to which they belong highlighted in grey.

Table 1.3: The stalling efficiency of the viral 2A motifs tested to date

The table features the reported efficiency for viral 2A-sequences previously tested in the *in vitro* expression system (Compiled from Luke et al., 2008 and Donnelly et al., 2001b).

virus	sequence tested	proportion (%) of processed proteins	reference
Positive stranded RNA viruses			
Picornaviruses			
	FMDV		
FMDV	NFDLLKLAGDVESNPG/P	75	Ryan and Drew, 1994
	Other picornaviruses		
TMEV	FREFFKAVRGYHADYYKQRLIHDVEMNPG/P	98	Donnelly *et al.*, 1997
EMCV	VFGLYRIFNAHYAGYFADLLIHDIETNPG/P	99	Donnelly *et al.*, 2001*b*
ERAV	QCTNYALLKLAGDVESNPG/P	99	
PTV	ATNFSLLKQAGDVEENPG/P	94	
SAF-V	FTDFFKAVRDYHASYYKQRLQHDVETNPG/P	99	Luke *et al.*, 2008
ERBV	EATLSTILSEGATNFSLLKLAGDVELNPG/P		
LV	YFNIMHSDEMDFAGGKFLNQCGDVETNPG/P		
Other viruses			
	Iflaviruses		
IFV	TRAEIEDELIRAGIESNPG/P	63	Donnelly *et al.*, 2001*b*
Longer version of IFV	PSIGNVARTLTRAEIEDELIRAGIESNPG/P	99	Luke *et al.*, 2008
EoPV-2A1	GQRTTEQIVTAQGWAPDLTQDGDVESNPG/P		
PnPV-2A1	GQRTTEQIVTAQGWVPDLTVDGDVESNPG/P		
EoPV-2A2	TRGGLQRQNIIGGGQRDLTQDGDIESNPG/P		
PnPV-2A2	TRGGLRRQNIIGGGQKDLTQDGDIESNPG/P		
	Tetraviruses		
TaV	RAEGRGSLLTCGDVEENPG/P	99	Donnelly *et al.*, 2001*b*
EeV	RRLPESAQLPQGAGRGSLVTCGDVEENPG/P	99	Luke *et al.*, 2008
PrV-2A1	LEMKESNSGYVVGGRGSLLTCGDVESNPG/P		
PrV-2A2	NSDDEEPEYPRGDPIEDLTDDGDIEKNPG/P	94	
PrV-2A3	TIMGNIMTLAGSGGRGSLLTAGDVEKNPG/P	99	
	Dicistroviruses		
CrPV	LVSSNDECRAFLRKRTQLLMSGDVESNPG/P	88	Luke *et al.*, 2008
ABPV	TGFLNKLYHCGSWTDILLLLSGDVETNPG/P	94	
Double stranded RNA viruses			
	Animals rotaviruses		
PoRV	AKFQIDKILISGDVELNPG/P	31	Donnelly *et al.*, 2001*b*
BoRV-C	GIGNPLIVANSKFQIDRILISGDIELNPG/P	89	Luke *et al.*, 2008
HuRV-C	GAGYPLIVANSKFQIDKILISGDIELNPG/P	82	
ADRV-N	FFDSVWVYHLANSSWVRDLTRECIESNPG/P	97	
	Insect cypoviruses		
BmCPV-1	RTAFDFQQDVFRSNYDLLKLCGDIESNPG/P	99	Luke *et al.*, 2008
OpbuCPV-18	IHANDYQMAVFKSNYDLLKLCGDVESNPG/P		
	Totiviruses		
IMNV-2A1	WDPTYIEISDCMLPPPDLTSCGDVESNPG/P	99	Luke *et al.*, 2008
IMNV-2A2	RDVRYIEKPEDKEEHTDILLSGDVESNPG/P		

1.3 Discovery of the 2A NPGP sequence

The ribosome-stalling hypothesis was derived by a process of elimination. The 2A gene product for poliovirus (belonging to the genus *Enterovirus*) is $2A^{pro}$, an enzyme of 142 residues (Toyoda et al., 1986), distantly related to $3C^{pro}$ and which performs a cleavage at its N-terminus. Nicklin and co-workers (1987) constructed various plasmids to demonstrate the proteolytic activity of poliovirus $2A^{pro}$. They showed that the cleavage site consisted of a tyrosine-glycine pair. $2A^{pro}$ catalytic triad was later determined by mutational and structural studies (Hellen et al., 1992, Sommergruber et al., 1997) and consists of Cys106, His18 and Asp35. Compounds such as iodoacetamide and N-ethylmaleimide, inhibitors of thiol proteases, were active against $2A^{pro}$ (Konig and Rosenwirth, 1988).

The mature 2A protein of *Hepatovirus* escapes detailed molecular analyses as it so far failed to be purified, suggesting that the protein is either unstable or simply not released. The predicted translated 2A presents no similarities to *Aphthovirus*, *Cardiovirus* or *Enterovirus* counterparts and deletion of 60 % of the C-terminal 2A sequence did not impair replication. For *Hepatovirus*, cleavage at the 2A/2B junction is carried out by $3C^{pro}$ (Martin et al., 1995) whereas the cleavage between VP1/2A occurs in the course of virion formation and is likely mediated by a cellular protease (Martin et al., 1999, Cohen et al., 2002).

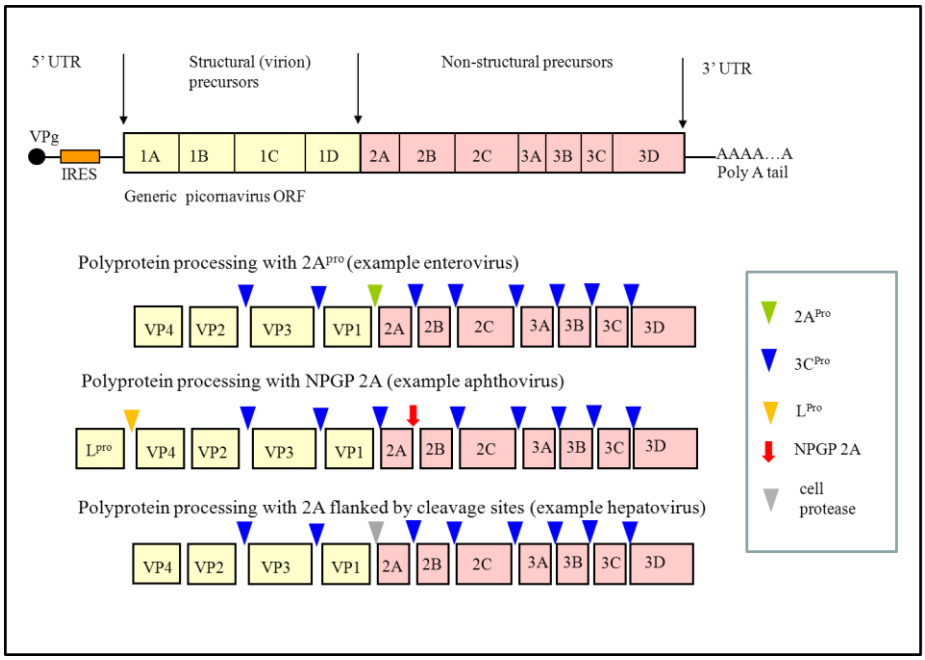


Figure 1.5: Picornavirus organisation and schematic representation of the three possible types of activity at the 2A region

$3C^{pro}$ carries out most of the proteolytic processing. At the 2A/2B junction, $2A^{pro}$ cleaves at its own N-terminus (example of *Enterovirus*, where the processing is mediated by two virally encoded enzymes, $2A^{Pro}$ and $3C^{Pro}$). 2A NPGP results in segregation of capsid and replicative precursors (example *Aphthovirus* genus characterised by the NPGP 2A and the leader

protease (L^{pro})). 2A is flanked by cleavage sites (example of *Hepatovirus*) (Derived from information available at http://www.picornaviridae.com).

Theiler's murine encephalomyelitis virus (TMEV) transcripts lacking $3C^{pro}$ sequence retained the ability to operate the separation at the 2A/2B junction (Roos et al., 1989). Inhibition studies involving encephalomyocarditis virus (EMCV) showed that the 2A/2B processing occurred very quickly and was not influenced by all the inhibitors used (Jackson, 1986). Sequence alignment showed no possible correlation to rhinoviruses and poliovirus 2A. No protease domains were found in the 2A sequences of cardioviruses EMCV and TMEV (Palmenberg, 1990). In a separate study, EMCV 2A could tolerate large deletions (60 %) at the N-terminal half and still result in the 2A/2B separation. Deletions to the C-terminal of 2B produced a similar result. A sequence alignment showed that the C-terminus of 2A is absolutely conserved amongst all cardioviruses. Subsequent directed mutagenesis identified the cleavage site. For cardioviruses, the cleavage occurred between the last glycine of 2A and the next proline of 2B of the absolutely conserved four amino acid sequence [NPGP] (Palmenberg et al., 1992).

L^{pro} or $3C^{pro}$ do not cleave the 2A/2B junction of FMDV (Ryan et al., 1989, Belsham et al., 1990). Alignment of the 2A regions between FMDV and cardioviruses EMCV and TMEV showed that the C-terminal [NPGP] motif was also conserved. FMDV 2A was later introduced internally in a synthetic reporter system flanked by virally unrelated sequences and translated in rabbit reticulocyte lysates and showed to retain activity. The processing event at the 2A [NPGP] motif did not rely on any other viral elements (Ryan et al., 1991).

This synthetic *in vitro* approach (Ryan et al., 1991) has been adopted since and consists of two genes encoding for different molecular weight proteins which could then easily be visualised on an acrylamide gel. The 2A sequence is inserted in-frame between these two genes creating one single ORF on a plasmid harbouring a T7 promoter. The plasmid is then introduced in cell-free systems such as rabbit reticulocyte lysates or wheat germ assays supplemented with [^{35}S]-methionine.
In the current expression system (figure 1.6), 2A is inserted between green fluorescent protein (GFP) and beta-glucuronidase (GUS) coding sequences. Three bands can be visualised. One band corresponding to the full-length 'unprocessed' polyprotein, and two further bands corresponding to the segregated proteins.

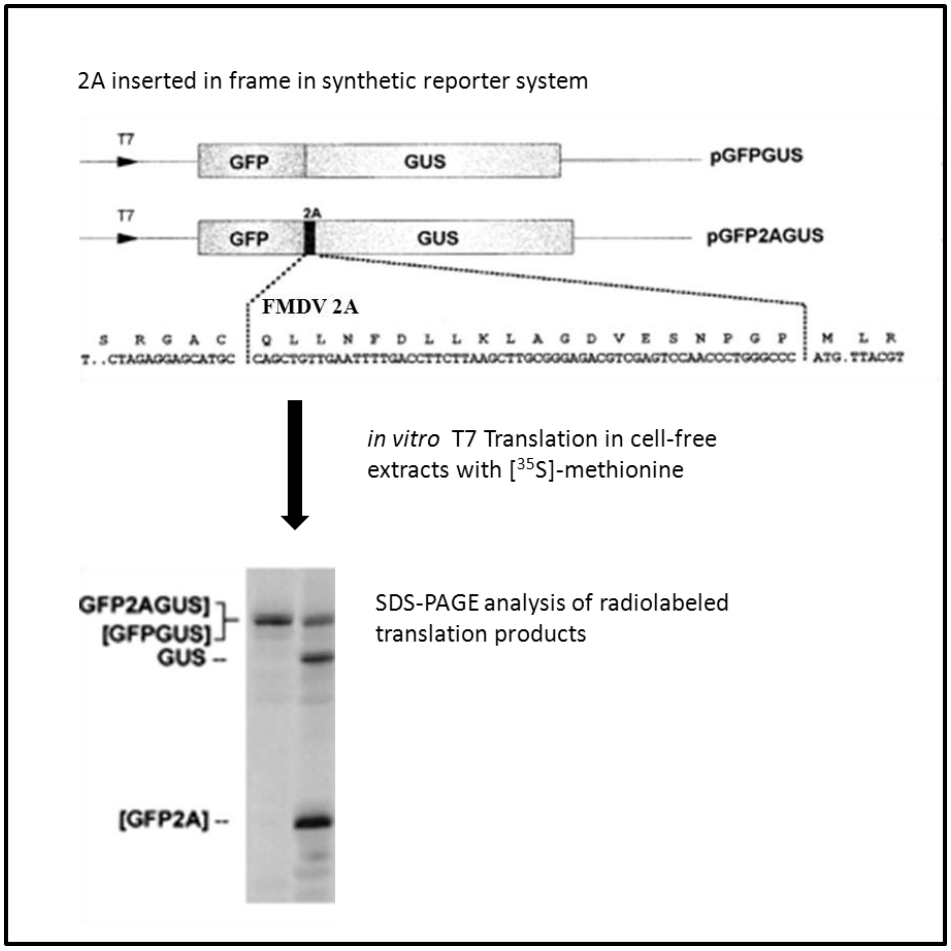


Figure 1.6: The bicistronic expression system for analysis of 2A activity

Cloning involves insertion of the 2A sequence in-frame between GFP (27 kDa) and GUS (70 kDa) coding sequences. The plasmid is introduced in a cell-free system supplemented with [^{35}S]-methionine and translation is driven by T7 promoter. The distribution of *de novo* radiolabeled translation products are analysed on SDS-PAGE gels. Three translation products can be observed: the full-length non-processed polyprotein (100 kDa) and the two segregated products GFP-2A (~30 kDa) and GUS (top drawings provided by Prof. Ryan, figure adapted from Donnelly et al., 2001*b*).

Several common features were reported between FMDV 2A and the cardioviruses TMEV, EMCV and mengovirus (mengovirus is a strain of EMCV) 2As, despite their difference in size. The 2A peptide is eighteen amino acids long in FMDV and 150 amino acids long in cardioviruses. In addition to the common [NPGP] conserved motif at the end of their 2A and at the beginning of 2B, there are three other conserved amino acids [D(V/I)E] (figure 1.7). In addition, their processing involves a C-terminus event between 2A and 2B rather than an N-terminal cleavage between VP1 and 2A as is the case for the other proteases in picornaviruses (Donnelly et al., 1997). Donnelly and colleagues (1997) therefore proposed that the mechanism of action for these separate members of *Picornaviridae* were exactly the same. The motif [D(V/I)ExNPGP] was proposed as the critical amino acid sequence involved in the unusual activity of 2A.

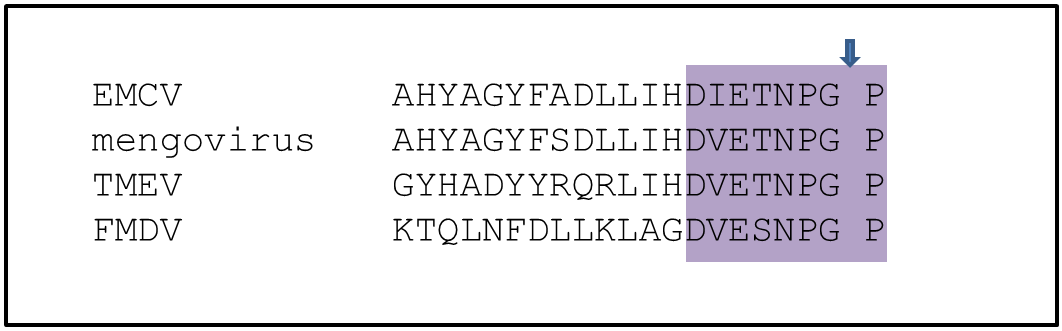


Figure 1.7: C-terminal sequences at the 2A/2B region of cardioviruses (EMCV, TMEV and mengovirus) and aphthovirus (FMDV)

The violet box highlights the conserved residues, the separation between 2A and 2B occurs between the last glycine of 2A and the first proline of 2B, shown with a blue arrow (adapted from Donnelly et al., 1997).

It was therefore concluded that the picornavirus 2A region can either encode a proteinase ($2A^{pro}$) as for enteroviruses, or the translated 2A can be flanked with protease cleavage sites such as for hepatoviruses and parechoviruses, or it can induce a ribosome stalling event characterised by the C-terminal residues [D(V/I)ExNPGP], as is the case for aphthoviruses and cardioviruses (summarised in figure 1.5).

Ryan and Drew (1994) focussed first on understanding how much of the already small peptide (eighteen amino acids) was necessary to maintain 2A activity in an artificial context. It seemed that the critical threshold was thirteen amino acids, although at that length the amount of polyprotein processed was reduced from 75 % to about 65 %. On the other hand adding more amino acids upstream of 2A altered that activity favourably to the point where almost all the polyproteins (>99 %) were effectively divided into the two expected distinct proteins. Similarly, truncating the EMCV and TMEV cardioviruses 2A to only the last N-terminal eighteen amino acids (Donnelly et al., 1997) resulted in decrease in the proportion of polyproteins processed. It seemed therefore that the upstream context played an important role in increasing 2A efficiency. It was not, however, what explained the mechanism of action.

A noted feature of FMDV 2A activity was a molar excess of the upstream versus the downstream translated proteins *in vitro*. These suggested a differential synthesis of the coding mRNA such as the first part of the mRNA up to 2A was more translated than the second part (Donnelly et al., 1997). Similar translation profiles were observed with shorter periods of translation, meaning that the odd ratio was not attributed to proteolytic degradation of the downstream product (Donnelly et al., 2001a). There was no post-translational proteolytic degradation of the full-length product observed (Ryan and Drew, 1994).

Based on these observations and inspired by the early work published about *Escherichia coli* (*E. coli*) SecM sequence, Ryan and co-workers (1999) proposed that NPGP 2A influences the translating ribosome and affects its ability to perform the formation of a peptide bond between the C-terminal glycine and proline. In *E. coli*, SecM and SecA form an operon. SecA protein is essential to *E. coli* and is a translocase (Sarker and Oliver, 2002). SecM has a C-terminal motif FxxxxWIxxxxGIRAGP able to stall the prokaryotic ribosome (Nakatogawa and Ito, 2002). The intergenic area between SecM and SecA contains a Shine-Dalgarno motif, normally inaccessible, and which becomes available to a ribosome only if the mRNA is re-structured. This change in secondary structure is introduced by the presence nearby of a stalled ribosome on the SecM motif (Butkus et al., 2003). SecM is translated normally in *E. coli*. When the bacterial cell has an abundance of SecA, SecA interacts with the SecM N-terminal sequence which extends outside the ribosome tunnel. SecA translocase activity acts like a pull and thus un-stalls the ribosome. There is a stop codon following the SecM stalling motif where the ribosome is able to terminate translation. In these conditions the stalled ribosome does not have the time to re-organise the intergenic mRNA structure. When the cellular level of SecA decreases, SecM stalled- ribosomes can re-structure the mRNA, allowing a new ribosome to start translation of SecA downstream of SecM (Yap and Bernstein, 2011).

1.4 Model for 2A Activity

The first model to explain 2A activity was published by Ryan and co-workers (1999) and suggested that translation of the ORF started and progressed normally from a start codon up to the 2A [NPGP] stalling motif. The glycyl-prolyl peptide bond was not being made, the ribosome thus could terminate translation and the imbalance in products suggested that a proportion of the ribosomes stopped while others could carry translation further into the downstream sequence. In this study, computer generated modelling of several C-terminal 2A peptides revealed a tight turn at the [NPG] while the short eighteen amino acids FMDV 2A was predicted to adopt an α-helix.

Ryan and co-workers (1999) mentioned that proline is structurally different from other amino acids. The nitrogen is engaged within a ring and is sterically constrained. The nitrogen of proline is a poor nucleophile. There are two prolines in the [NPG/P] motif therefore presumably allowing very little flexibility in the structural organisation of the 2A peptide into the ribosome exit tunnel. The glycine is also an unfavorable amino acid as its nitrogen is considered the second poorest nucleophile (Ryan et

al., 1999). These features are thought to play a critical role in 2A activity. At the peptidyl transferase centre (PTC) within the ribosome, the 2A nascent peptide linked to glycyl-tRNA is located in the P site. When prolyl-tRNA occupies the ribosome A site, the conformation of 2A in the ribosomal exit tunnel prevents peptide bond formation.

Toeprint analyses confirmed that 2A acted upon translating ribosomes, and mapped the activity to be at the Gly-Pro pair (Doronina et al., 2008). Toeprint analysis also confirmed the hypothesis of a pause in translation. In this study, yeasts (strain *sup45-2*) with reduced release factor 1 (eRF1) activity at 34 ^{o}C, led to an increase in full-length products. In yeasts and yeast extracts with impaired release factor 3 (eRF3) activity, there was a large reduction in the production of downstream and full-length product. This implied a rescue function for factors eRF1 and eRF3 (Doronina et al., 2008).

Doronina and co workers (2008) proposed that the pause in elongation observed is related to dissociation of prolyl-tRNA from the A site. They argued that the unusual dissociation of prolyl-tRNA from the A site following failure to generate the peptide bond, would leave a structural conformation, similar to a natural stop codon, that can be a substrate for binding of the termination factors.

Following the release of the nascent peptide, eRF1/3 would exit the A site, prolyl-tRNA re-enters the A site and the ribosome ingresses assisted by eEF2. One possible outcome is the dissociation of the ribosomal subunits which could also explain the imbalance in ratio between the product upstream of 2A and the product downstream of 2A (the model is summarised in figure 1.8).

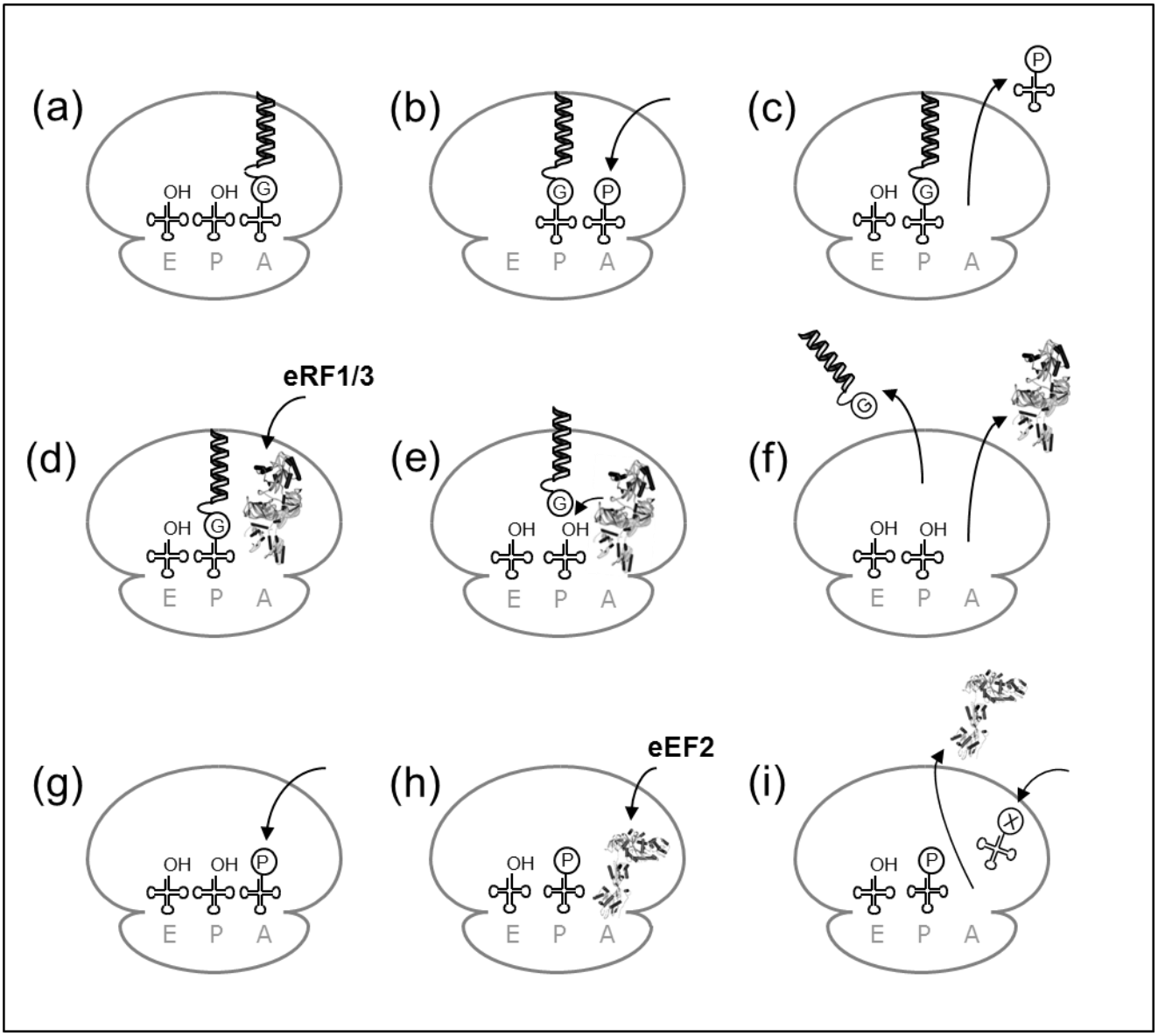


Figure 1.8: Model of the mechanism of 2A-induced ribosome stalling

The assembled ribosome large and small subunits are represented containing the aminoacyl site (A), peptidyl site (P) and exit site (E), fitted with tRNA deacylated (-OH) or aminoacylated with Pro or Gly. The cartoon shows the possible step by step scenario for when a eukaryotic ribosome encounters the last Gly-Pro residues of the NPGP stalling motif of 2A.
(Step a) peptidyl-tRNA is in A site and
(step b) is translocated to the P site, prolyl-tRNA occupies A site. Interaction of the 2A structure within the ribosome exit tunnel and the tight turn at the last NPG residues, precludes peptide bond formation.
(Step c) the prolyl-tRNA exits the ribosomal unit and
(step d) the eRF1/3 complex enters the A sites and hydrolyses the ester bond linking the nascent peptide to tRNA in P site.
(Step f) eRF1/3 leaves the A site. The nascent peptide is released.
(Step g) prolyl-tRNA re-enters A site and
(step h) is translocated to P site by eEF2.
(step i) the next aa-tRNA enters the A site, peptide bond is formed and the mRNA sequence downstream of 2A is translated.
(Figure provided by Prof. Ryan)

1.5 The translating ribosome, implications for 2A activity

1.5.1 Elongation and peptide bond formation

1.5.1.1 The ribosome

The ribosome is a cellular organelle assembled from three (prokaryotes) or four (eukaryotes) molecules of RNA and proteins for the process of translating mRNAs into proteins (summary provided in table 1.4 and the structure of the assembled complex is provided in figure 1.9). It is assembled from two subunits: a large and a small. The number of proteins associated with the ribosome varies. However, the core of the ribosome where peptide bond formation occurs is conserved across all life and is composed of rRNA only (Nissen et al., 2000). It is thought that peptide bond formation is a conserved mechanism. The reactive centre or the peptidyl transferase centre (PTC) has been located in a deep cleft in domain V, the central loop (and equivalent in eukaryotes) of 23S rRNA (Nissen et al., 2000).

Table 1.4: Summative table of relevant prokaryotic and eukaryotic ribosome and translational features.

(Adapted from Taylor et al., 2009, Harish and Caetano-Anollés, 2012 and Dever and Green, 2012).

	Ribosomal element	prokaryotic	Eukaryotic	
			homolog	No homologies to prokaryotes
Large subunit	rRNA	23 S (2900 nts)	28 S (4800 nts)	
		5 S (120 nts)	5 S (120nts)	
				5.8 S (160nts)
	Number of ribosomal proteins	31	50	
	Proteins associated with exit tunnel	L22	L17	
		L4	L4	
		L23	L25	
		L16	L10e	
				L38e
	Size	50 S	60 S	
Small subunit	rRNA	16 S (1540 nts)	18 S (1900 nts)	
	Number of ribosomal proteins	21	33	
	size	30 S	40 S	
Assembled ribosome size		70 S	80 S	
Elongation proteins		EF-Tu	EF1A	
		EF-G	eEF2	
Termination factors Class I		RF1 and 2		eRF1
Termination factors Class II		RF3		eRF3

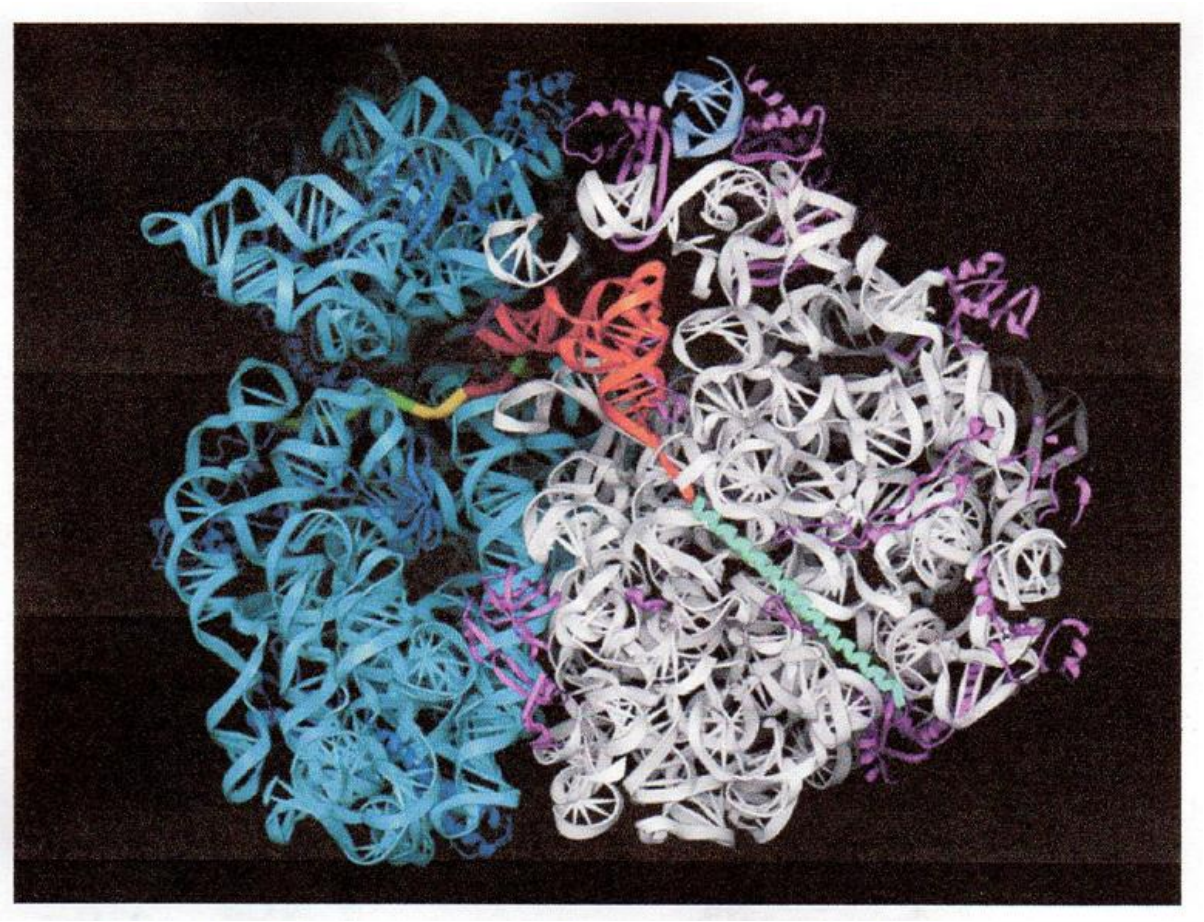


Figure 1.9: Crystal structure of *Thermus thermophilus* 70S ribosome

The 30S subunit is on the left, the 50S subunit is on the right. Peptidyl-tRNA is in orange and the mRNA is in green-yellow-red, and is shown wrapped around the neck of the 30S subunit. The cross section also depicts an α-helical nascent peptide chain (green) in the exit tunnel. 23S is in grey and 16S in cyan (sourced from Noller, 2012 based on his early work Yusupov et al., 2001).

The ribosome has three sites where the tRNA substrates can bind: the amino-acyl site at the decoding centre (A site) where new amino-acyl-tRNA (aa-tRNA) are fitted based upon their matching sequence to the mRNA codon triplet; the peptidyl site (P site) where the tRNA moves to after peptide bond formation has occured, thus leaving an empty A site ready to accommodate the next aa-tRNA; and the exit site (E site) where the deacylated tRNA is moved to before being ejected from the ribosome-mRNA complex (Noller, 2012). The figure 1.10 shows tRNA features relevant to this project: the anticodon arm that interacts with specific mRNA codons and at the 3' end The $C_{74}C_{75}A_{76}$ arm (the CCA acceptor arm) which is the site of aminoacylation.

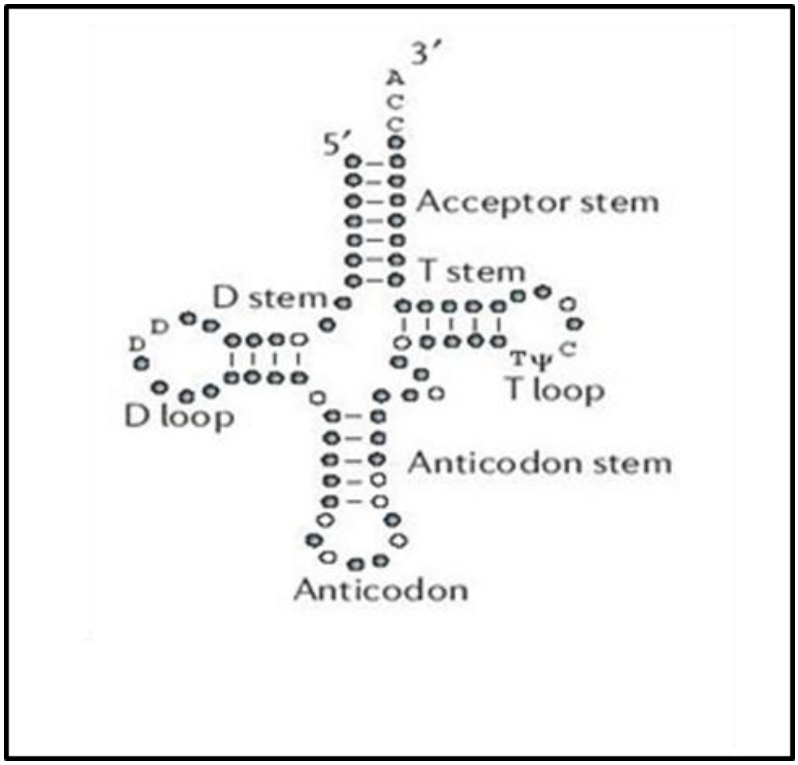


Figure 1.10: Schematic representation of features of the tRNA molecule

Nucleotides are represented in dots. Of special interest for this study are: the anticodon arm that interacts with specific mRNA codons and the 3' end, characterised by the CCA tri-nucleotides (the acceptor arm). The $C_{74}C_{75}A_{76}$ arm is the site of aminoacylation by dedicated cellular tRNA synthesases and also plays a critical role in positioning amino acids for peptide bond formation (taken from Kazantsev and Pace, 2006).

The elongation step on the ribosome involves a series of induced fit re-arrangements. Much of the elucidation of molecular mechanism comes from structural, cross-linking, antibiotics and mutational studies of the prokaryotic 70S ribosome. The movements have been characterised as 'classical and hybrid states' based upon the configurations of the tRNA species during elongation (figure 1.12). Upon entry in the A site the incoming aa-tRNA associated with EF-Tu adopts the A/T state where the CCA arm faces away from the PTC. After release of EF-Tu, the CCA arm swings towards the PTC. The aa-tRNA as well as the P site tRNA are then fully positioned in the 'classical' A/A, P/P and E/E state. After peptide bond formation, the tRNAs adopt the 'hybrid' state where their anticodon arms are still positioned in the A or P sites but their CCA arm has moved to the P or E sites (Noller, 2012).

1.5.1.2 Accommodation and selection of incoming aa-tRNA on the A site.

Based on data from crystallography and FRET analyses, Rodnina and co-workers (2005) proposed a model for the aa-tRNA selection on the A site. The process involves two steps: the initial selection step and the proofreading step separated by GTP hydrolysis of EF-Tu (in bacteria) or EF1A (in eukaryotes).

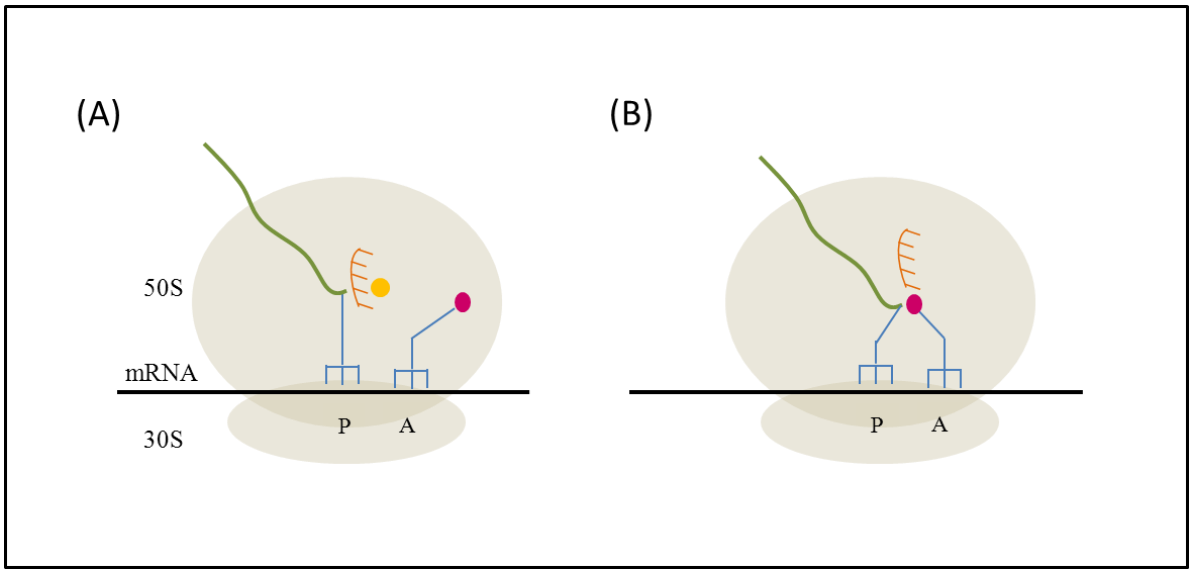


Figure 1.11: Summary cartoon depicting activity at the ribosomal PTC

(A)- The A site is selecting for a cognate aa-tRNA specified by an mRNA codon. The nascent peptide (in green) attached to the P site tRNA is protected (represented by spikes) from hydrolysis by water (orange dot) by a specific configuration of the PTC.

(B)- When the cognate aa-tRNA is accommodated in the A site, the PTC is re-organised to allow peptide bond formation. The A and P sites tRNAs are represented in blue, the incoming amino acid in pink. (Created using information from Schmeing et al., 2005b).

Aa-tRNAs are transferred to the ribosome A site in complex with the GTPase EF-Tu or EF1A. Single molecule FRET was used in a study to monitor the incorporation of amino acids from tRNA. The study provided evidence that the charged tRNA was at first labile in the A site (Blanchard et al., 2004). Crystal structures provided complementary details. Discrimination of matching codon anti-codon pairing is based on recognition of structural features. A suitable match triggers 30S subunit closure (Ogle et al., 2002) and a closed conformation activates EF-Tu for GTP hydrolysis (Ogle et al., 2003). The GDP- bound factor looses affinity for its tRNA, dissociates and this in turn allows the aa-tRNA to be fully accommodated in the A site and the PTC. At proofreading step, a mismatch can still be detected. The incorrect tRNA dissociates from the ribosome due to low stability or low rate of

accommodation whereas a correct charged-tRNA will be accommodated to the PTC (Rodnina et al., 2005).

From entering the ribosome coupled to EF-Tu, to adopting the right position for peptide bond formation, requires that the aa-tRNA CCA arm moves 70Å towards the PTC (Blanchard et al., 2004). A simulated study, aimed at retracing movements of aa-tRNA adaptation to the PTC from the A/T state to the A/A state suggested that the tRNA moved through a corridor made of twenty conserved nucleotides of the 23S rRNA. The authors also argue that a 'gate' of 23S rRNA nucleotides: U2492/C2556/C2573 configures the aa-tRNA CCA arm to facilitate bonding of the tRNA C75 to the 23S rRNA base G2553 (Sanbonmatsu et al., 2005) (depicted in figure 1.10). The accommodation stage of the elongation step is considered the rate limiting step (Rodnina et al., 2005).

Before peptidyl transfer, the A and P sites tRNAs are held in place by 23S rRNA. The CCA arms form interactions with the ribosome. In the P site, C74 is paired to G2251 and C75 to G2252 and the 3' A76 of P site tRNA interacts with the base-pair A2450-C2501 of 23S rRNA. The C75 of the CCA arm of A site tRNA is fixed by G2253 while the 3' A76 form interactions with G2583 (Nissen et al., 2000, Bashan et al., 2003, Schmeing et al., 2005). The core of the peptidyl centre (PTC) contains the following highly conserved nucleotides: A2451, U2506, U2585, C2452 and A2602 (Bashan et al., 2003, Schmeing et al., 2005).

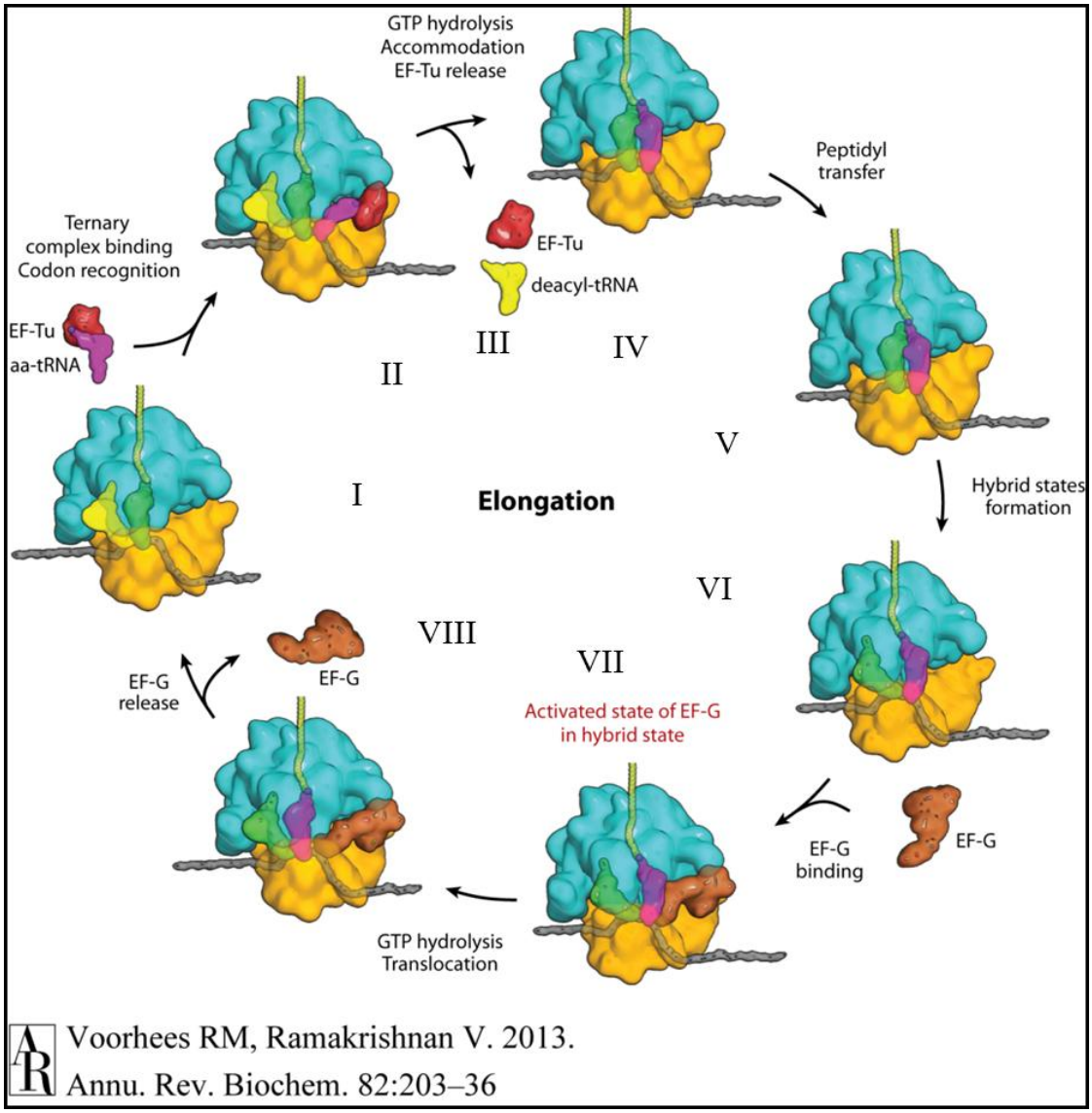


Figure 1.12: Overview of the elongation step on the 70S ribosome, the central role of peptidyl transfer in translocation

The positions of the tRNAs on the ribosome are characterised by the position of their anticodon arms (paired to mRNA codon on the small subunit) in relation to their acceptor arms located within the large ribosomal subunit). The ribosome large subunit is in turquoise and the small subunit in gold.

Before peptide bond formation, (**I)** the ribosome is in the classical state characterised by peptidyl tRNA (in green) in P/P and exit tRNA (in yellow) in E/E configuration. (**II**) New aa-tRNAs (in pink) delivered by EF-Tu (in red) are probed on the A site. The tRNA adopts the A/T state on the A site. Following accommodation, (**III**) EF-Tu GTPase activity is activated EF-Tu departs. (**IV)** The tRNAs adopt the classical A/A and P/P configuration and (**V**) the peptide is transferred to A site tRNA. The peptide transfer to the A site tRNA is the essential step to trigger the ratcheting motion of the ribosome (**VI**), where the small subunit moves counter-clockwise to the large subunit. The tRNAs enter the hybrid A/P and P/E state. (**VII**) EF-G-GTP (in brown) binds the complex and is thought to stabilize and facilitate translocation. (**VIII**) Translocation is accompanied by the clockwise motion of the 30S subunit. After GTP hydrolysis, EF-G dissociates and the ribosome re-enters the classical P/P and E/E state.

(Images sourced from Voorhees and Ramakrishnan, 2013)

1.5.1.3 Peptide bond formation and elongation

Once the aa-tRNA is positioned in the PTC, peptide bond formation between NH_2 group of the A site amino acid and the COOH group of the P site amino acid, occurs rapidly (Sievers et al., 2004). The elongation and peptidyl transfer steps are summarised in figure 1.12 and 1.13.

The creation of the peptide bond between the amino group and the carboxy group of the last amino acid depends exclusively on the proximity of the groups. Many theories were put forward to explain the ribosome function. It was thought that the peptide bond formation could be attributed to a chemical reaction involving the ribosome nucleotides in domain V of 23S rRNA. Mutations of the nucleotides of the PTC A2451, U2506, U2585 and A2602 retained reaction to puromycin, a tRNA analogue (Polacek et al., 2001, Thompson et al., 2001, Hesslein et al., 2004, Youngman et al., 2004, Beringer et al., 2005). It was thought that ribosome proteins could participate in peptide bond formation, but mutagenesis, as well deproteination proved this theory wrong, also the theory of an acid base reaction involving a proton relay across the ribosome could not be demonstrated (Polacek et al., 2001, Beringer et al., 2003). It was however important that the interaction between C75 of the A site tRNA and G2553 in 23S rRNA stayed intact (Youngman et al., 2004, Brunelle et al., 2006).

Based on crystallographic results, Schmeing and co-workers (2005) discovered that binding of a suitable substrate to the A site rearranges residues G2583, U2506 and U2585 such that the carbonyl carbon of the peptidyl-tRNA is oriented favourably for attack by an incoming nucleophile in the A site. The PTC structural re-arrangement is supported by X-ray data collected using analogues of tRNA acceptor arm in 50S subunits (Schmeing et al., 2005, Voorhees et al., 2009).

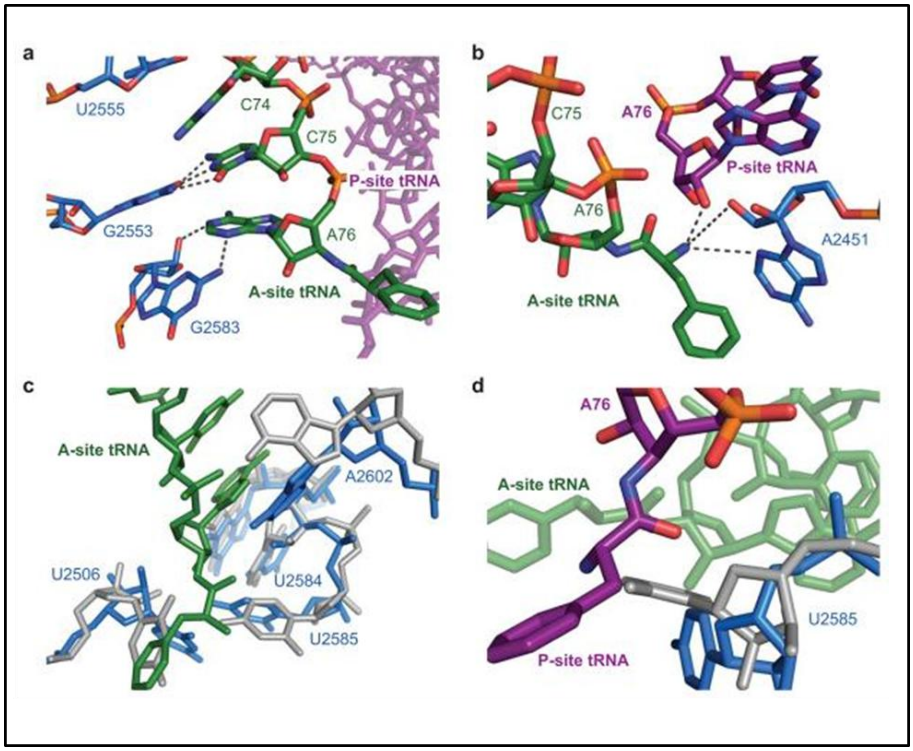


Figure 1.13: Activity at the catalytic pocket of the PTC

In (a), the diagram shows the stabilising interactions between the CCA arm of aa-tRNA in the A site (shown in green) and 23S rRNA (in blue). In (b), the interaction of the aa of A site tRNA to the 23S rRNA (in blue) and to the peptidyl-tRNA (in violet) is shown. In (c), the empty (in grey) and pre-peptidyl transfer (in blue) movements of the A sites nucleotides are characterised by shifts of A2584, U2585, U2506 and A2602. In (d), upon binding of the A site tRNA, the residue U2585 exposes the P site tRNA ester for nucleophilic attack. (Images reproduced from Voorhees et al., 2009).

Peptide bond formation proceeds through a nucleophilic attack of the A site aa-tRNA on the carbonyl carbon of the peptidyl-tRNA (figure 1.13 and 1.14). The reaction involves two intermediary steps: a tetrahedral intermediate (T+/-) and (T-) (figure 1.14).

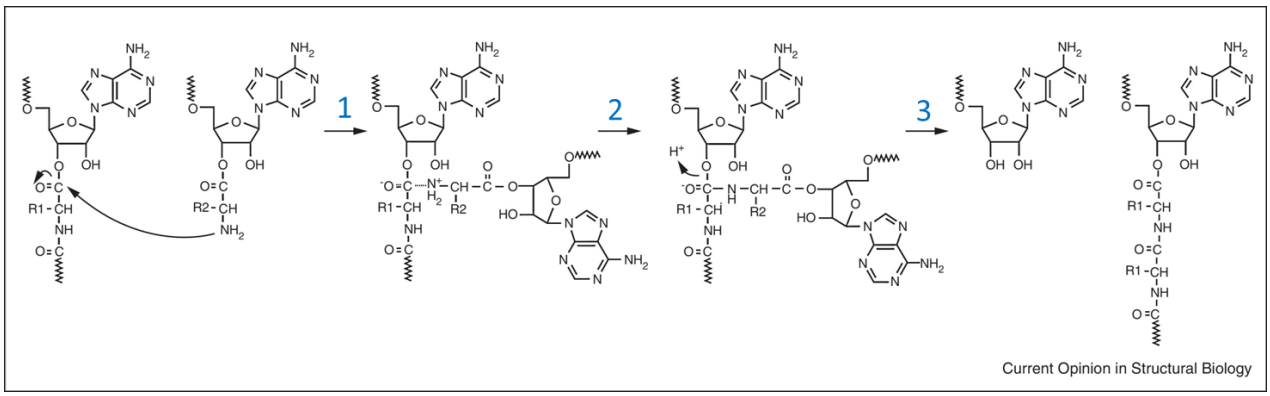


Figure 1.14: Peptide bond formation

Attack of the amino group of the aa-tRNA at the A site to the carbonyl carbon of the peptidyl-tRNA at the P site, which results in formation of the intermediate (T+/-) (step 1). The intermediate is deprotonated to yield the (T-) intermediate (step 2) and (step 3) the intermediate is broken down into the final product leaving a deacetylated tRNA in the P site and the peptidyl-tRNA with an additional amino acid in the A site (Picture reproduced from Rodnina, 2013).

Peptide bond formation results in the P site tRNA being deacylated. The peptide is transferred to the A site tRNA. This pre-translocational complex is the substrate for elongation factor G (or eEF2 in eukaryotic systems). EF-G derives energy from GTP hydrolysis and was thought to catalyse the translocation of the tRNAs, the mRNA and also promote conformational changes in the ribosome. Previous studies showed that the stimulation of the GTP hydrolysis of EF-G is accompanied by the clockwise rotation of the 30S and translocation (Zavialov and Ehrenberg, 2003). The exact participation of EF-G in translation is being re-assessed currently (Ermolenko and Noller, 2011).

Ribosome subunits undergo a ratchet motion where the two subunits are in a counter-clockwise or clockwise rotation relative to each other (Franck and Agrawal, 2000). A recent publication by Ermolenko and Noller (2011) using FRET to monitor the rotation of the subunits in *E. coli* proved that the translocation occurs in two steps. First the translocation involves a counter-clockwise rotation that favours the hybrid A/P, P/E state for the tRNAs. Previous work also showed that the hybrid state adopted by the acceptor arms of the tRNA, the A/P and P/E state, is spontaneous in nature and does not require any other element (Moazed and Noller, 1989). Ermolenko and Noller (2011) however proved that this first spontaneous step is followed in a second time by the clockwise rotation of the subunits, independent of the EF-G and GTP hydrolysis and responsible for the movements of the tRNAs and the mRNA. The study does not volunteer much discussion about the actual role of EF-G. The body of the article does mention, however, that EF-G may be necessary to stabilise the hybrid state (Ermolenko and Noller, 2011).
The current model for translocation however still retains the theory that the movement one codon forward is dictated by the formation of the peptide bond, as the ratcheting (counter-clockwise transition) does not occur without deacylated P site (Zavialov and Ehrenberg, 2003, and Valle et al., 2003).

1.5.2 Nascent peptides that influence the elongating ribosome

Several nascent peptides (leader peptides or stop-peptides) form interactions with the ribosome exit tunnel and halt the translating ribosome. The mechanism of translation arrest was discovered to involve the regulation of expression of the downstream ORF:

- for secretion (SecM, Nakatogawa and Ito, 2002)
- amino acid metabolism (AAP, Gong et al, 2007)
- antibiotic resistance (*ermC*, Vasquez-Laslop et al., 2008).

Translation arrest can be brought about by mRNA secondary structure, co-effectors or not require any other elements outside the nascent peptide sequence. This section provides an up to date profile for four well studied stop peptides: *ermC*, SecM, *TnaC* and AAP.

1.5.2.1 Antibiotic resistance

ermC, *catA-86* and *cmlA* are resistance genes to macrolide antibiotics in bacteria. A well studied example in this category is resistance to erythromycin mediated by the *ermCL* operon. The *ermCL* cassette has two ORFs: the leading ORF *ermL* and a downstream regulated ORF *ermC*. The *ErmC* translation product is a methyltransferase able to dimethylate the amino group of nucleotide A2058, and therefore provides resistance. The erythromycin binding site is near ribosomal nucleotides A2058, G2057, C2611 (Vasquez-Laslop et al., 2008). Interaction of the antibiotic with ribosome translating the *ermL* mRNA causes translation arrest at a C-terminal specific peptide, nine amino acids long in a nineteen codon long ORF. Mutational analyses of the nine residues showed that the C-terminal IFVI motif was critical for arrest (Ramu et al., 2009). Conservative substitution of the last isoleucine to valine eliminated arrest. Erythromycin cannot mediate arrest by itself, but the combined effect antibiotic and stalling peptide leads to a retardation of the ribosome on the motif. In the absence of antibiotic the mRNA structure between the two ORFs prevents translation of the ORF 60bp downstream of *ermL* (see figure1.15). In the presence of antibiotic, ribosome translating the critical nine aa of *ermL* are stalled. This results in mRNA re-conformation and exposure of the ribosome binding site for *ermC* ORF. The nucleotide A2062 proved to be the key ribosomal element which changed from open to closed upon stalling, and mutations of this residue to C or U resulted in no induction by erythromycin (Vasquez-Laslop et al., 2008).

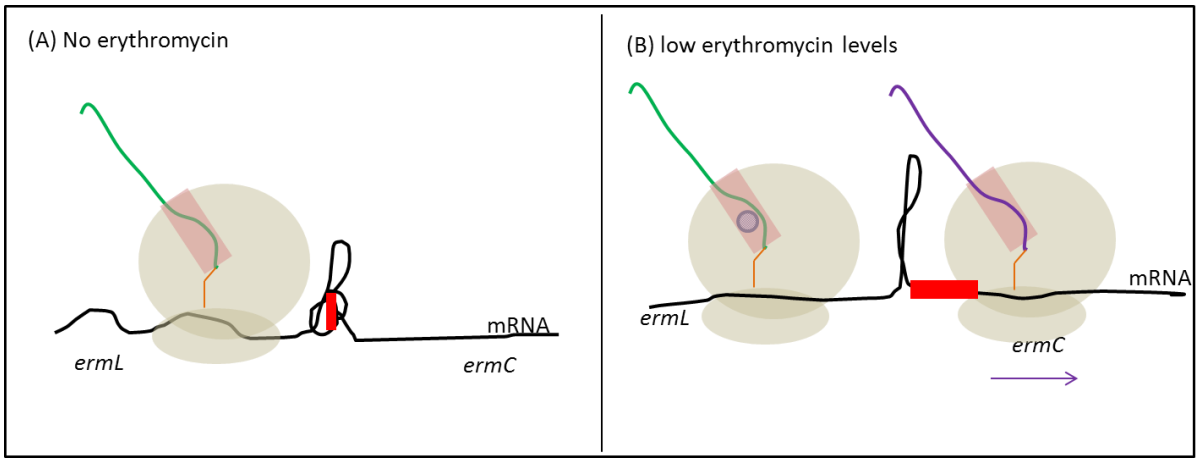


Figure 1.15: Expression of *ermCL*

In the absence of antibiotic erythromycin (panel A), the ribosome (represented in grey spheres with the exit tunnel in pink) translates *ermL*. The ribosome binding site (red rectangle) for *ermC* is shielded by mRNA secondary structure. In the presence of erythromycin (panel B), stalling leads to the re-arrangement of the mRNA secondary structure; exposing the ribosome binding site and allowing translation of *ermC*.

Analogous to *ermCL*, the *catA* and *cmlA* resistance operons are induced by the antibiotic chloramphenicol. This antibiotic inhibits translation at random, but the specificity is provided by a nascent peptide, five amino acids long (MVKTD) for *catA* and nine amino acids long for *clmA* (Rogers and Lovett, 1994 and Lovett and Rogers, 1996). Chloramphenicol binds in the vicinity of 23S rRNA C2452, A2062, and G2505 part of the PTC and the exit tunnel (Dunkle et al., 2010).

1.5.2.2 SecM regulation of SecA expression

SecM and SecA form a bicistronic operon. SecA is involved in protein secretion in *E. coli*. The leader region SecM encodes a 170aa protein constitutively expressed. At the N-terminal region SecM has a signal peptide recognised by the secretion machinery (by SecA) and at the C-terminal part a nascent stalling sequence. The SecA ORF translation is repressed by a secondary structure shielding the ribosome binding site. Cellular deficiency of SecA results in retardation of the ribosome on the SecM stalling motif resulting in melting of the secondary structure, exposure of the ribosome binding site and translation of the SecA ORF (Huber et al., 2011 and Yap and Bernstein, 2011). Mutational analysis of SecM showed that the critical residues were F150xxxxWIxxxxGIRAGP166. In this sequence x corresponds to residues where substitutions to alanine did not alter the activity of the sequence. The exact spacing between these residues is critical. In addition, SecM stalling involves a critical prolyl-tRNA in the A site (Nakatogawa and Ito, 2002). The arrest is caused by the ribosome's inability to synthesise the Gly165 to Pro166 peptide bond, with the stalled complex attached to the tRNA in the P site (Muto et al., 2006) (summarised in figure 1.16). Using cryoEM, cross-linking and FRET studies, Woolhead and colleagues (2006) concluded that the SecM conformation within the tunnel is an essential element for stalling. The peptide adopts a compacted conformation, and the interaction between the nascent sequence and the tunnel leads to changes in both, that result in altered PTC activity. The absence of compaction in the C-terminal sequence results in decreased efficiencies to complete abolition of activity. Woolhead and co-workers (2006) concluded that the compaction could be regulated by the nature of the residues in the sequence and that the spacial location of key residues to face ribosome exit tunnel nucleotides was the essential element for SecM activity. A comparison of SecM residues in the genome of different bacteria demonstrated that Arg163 and Pro166 were the only absolutely necessary residues; Gly165, Ile 156 and Trp 155 would increase the arrest efficiency. Based on a sequence alignment the authors concluded that other variable residues in the SecM sequence served to create the conformation that fixes Arg163 to the tunnel (Yap and Bernstein, 2009). The ribosomal nucleotides required for SecM stalling are A2058, derived from

mutagenesis studies (Ito and Nakatogawa, 2002) A2062 and A2503 derived from cryoEM (Bushan et al., 2011).

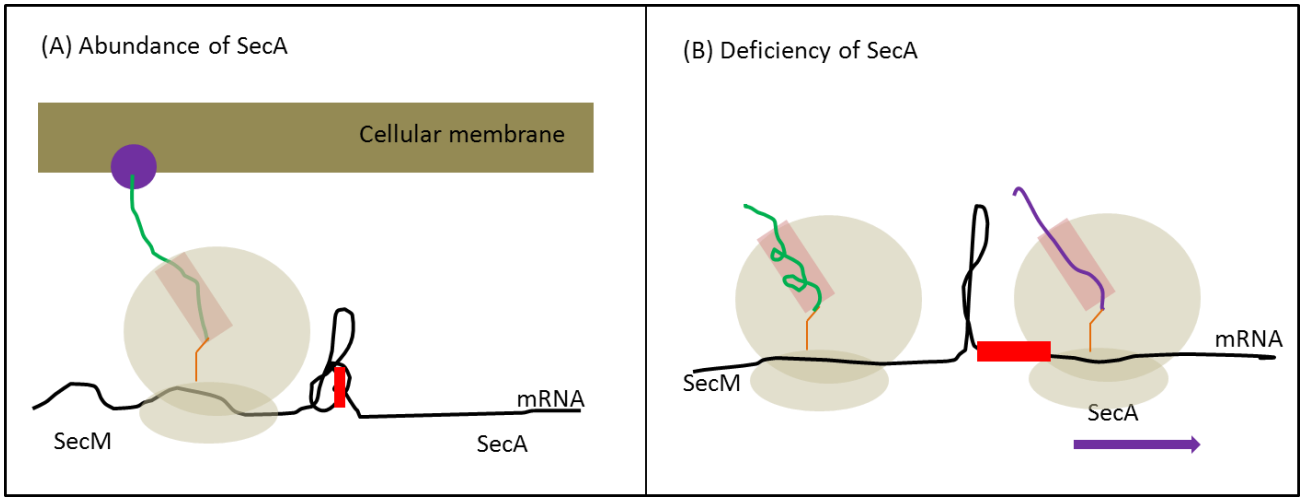


Figure 1.16: SecM regulation of SecA expression

The ribosome is represented by two spheres and the tunnel appears in red. In (panel A), translation of SecM (green line) introduces a pause which normally is alleviated quickly by the pulling force exerted by the translocase secA (represented by a violet dot), the ribosome binding site (red rectangle) is shielded and unavailable to a translating ribosome. When the cellular level of SecA is limited (in panel (B)), the SecM-paused ribosome melts the mRNA secondary structure and exposes the ribosome binding site for SecA. SecA is translated (violet line and arrow).

1.5.2.3 Regulation of tryptophanase expression

Tryptophan (Trp) is a source of carbon and nitrogen for many bacteria, including *E. coli*. Tryptophanase is the main operating enzyme able to degrade tryptophan. The Trp operon consists of a leader sequence *TnaC* and two genes: *TnaA* and *TnaB.* Tryptophanase is the translational product of *TnaA* and its expression is under the regulation of the leader peptide *TnaC*. In *E. coli*, transcription and translation are coupled. The leading sequence *TnaC* is twenty-four codons long (translated to MNILHICVTSKWFNIDNKIVDHRP, red letters are the critical amino acids). The purpose of the stalling function is to mask or expose the *rut* site downstream of the translating *TnaC* ribosome. Since the *rut* site is the binding site for Rho termination factor, its availability determines the transcription of downstream *TnaA* and *TnaB.* There are two essential elements necessary for stalling, free tryptophan and the *TnaC* coding sequence (summarised in figure 1.17). When tryptophan is not abundant, *TnaC* translation stops at the stop codon, the *rut* site is free and Rho termination factor can bind and terminate transcription activity of DNA polymerase.With high levels of intracellular tryptophan, the coupled action of free tryptophan binding to the ribosome translating the *TnaC* motif

leads to ribosomal arrest. The *rut* site is sequestered and DNA polymerase carries transcription in the *TnaA* and *TnaB* regions of the operon (Yanofsky, 2007).

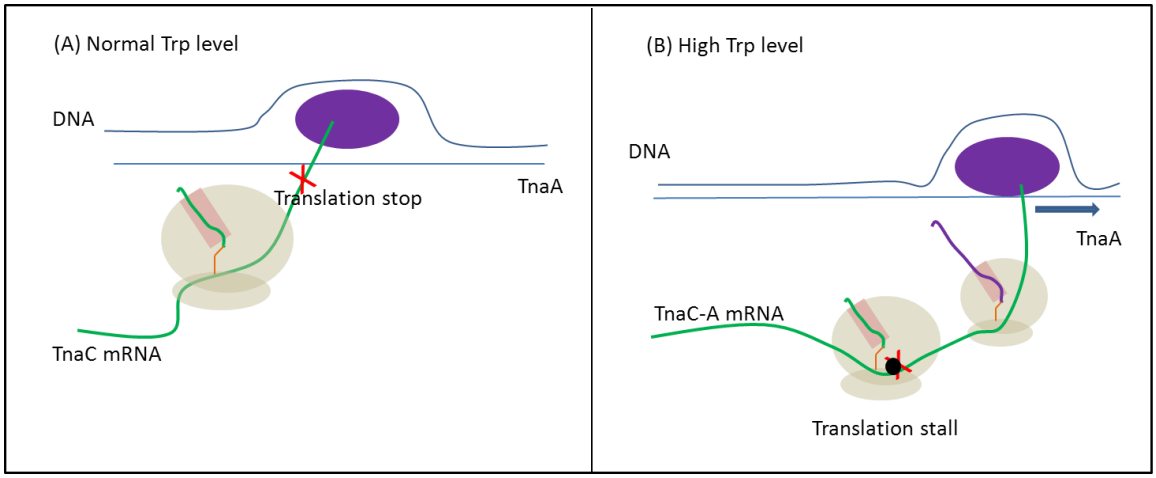


Figure 1.17: Expression of the *TnaA* under regulation of *TnaC* and free Trp

In normal levels of Trp in (panel (A)), the ribosome translating the *TnaC* mRNA terminates translation at the stop codon (represented by a red cross). Rho factor terminates transcription of the remaining operon *TnaA* and *B.* In panel (B), Trp (black dot) binds to the A site at the stop codon (red cross) of ribosome translating *TnaC* mRNA. The stalling complex hides the *rut* site so Rho termination factor can not bind. Transcription carries into the *TnaA* downstream region.

The residues Trp12, Lys11 could be covalently cross-linked to nucleotide A750 and Trp12 protected against methylation of A788. Mutational analyses showed that ribosomal nucleotides U2609, A752 and aa K90 of protein L22 were critical for stalling (Cruz-Vera et al., 2005). Puromycin as well as sparsomycin could not bind to the A site in the presence of free Trp for ribosome translating *TnaC* (Gong et al, 2007). The authors concluded that free Trp binds to the ribosome A site and induces an altered structure in the PTC. Deletions and insertions in the N-terminal region of *TnaC* have no effect. Trp12 is absolutely required, as no induction was observed with substitution of Trp12 for Arg. The correct spacing between Trp12 and Pro24 is also critical (Cruz-Vera et al., 2007). CryoEM study of the *TnaC*-ribosome interaction showed that although *TnaC* is unfolded, the nascent peptide makes contacts with the exit tunnel at many places, and Trp12 is located at the constriction point opposite the Arg92 of L22 protein and nucleotide A751 of 23S rRNA (Martinez et al., 2011).

1.5.2.4 Fungal arginine attenuator peptide (AAP)

The stalling mechanism leads to repression of the expression of genes involved in the biosynthesis of arginine in *Saccharomyces cerevisiae* and *Neurospora crassa* in response to high levels of intracellular Arg. The *CPA1* gene of *S. cerevisiae* and *arg-2* gene of *N. crassa* are involved in synthesis of Arg. The 5' end of their mRNA has a conserved ORF of twenty-four codons specifying the AAP leader peptide. In normal conditions, the ribosome translates the uORF and terminates at the stop codon. In high levels of free Arg, the leader peptide AAP and the binding of free Arg prevent termination and further elongation in the main ORF (Wang and Sachs, 1997). The stalling is linked to degradation of the mRNA (Gaba et al., 2001). The leader peptide can stall the eukaryotic ribosome in response to free Arg when placed into a polypeptide. The binding site for Arg has not yet been elucidated, but recent cryoEM studies (without occupied Arg) demonstrated that the C-terminus of AAP interacts with the ribosome exit tunnel as well as with the proteins L4 and L22. The PTC conformation had an altered geometry, which the authors concluded would be adversely affecting the release factor activity (Bhushan et al., 2010).
The stalling sequence is: MNGRPSVFTSQDYLSDHLWRALNA, (the letters in red represent the critical aa and the blue letters aa for which substitution reduces activity) (Delbecq et al., 2000).
The ribosomal nucleotides in proximity to the important residues are: U2585/A2062, U2609, A2058 and A751 as well as the residue Arg136 from protein L17 (Bhushan et al., 2010)

1.5.3 Implications for 2A activity

The model for 2A activity predicts that a peptide bond cannot be formed and suggests the formation of a α-helix in the tunnel, yet no two 2A sequences are alike and some do not have a predicted helical conformation; this is the subject of chapter 4. Why is the peptide bond not formed? In the light of the ribosome dynamic nature, and compaction of some of the other stop peptides, the theory of a structural pattern adopted to stall the ribosome could be true.
There are several aspects of translation which could be the 'target' for 2A activity:

- the local organisation of the mRNA at the A site and the abundance of the tRNA specified
- the conformation of the nascent peptide in the tunnel
- the presence of elongation and release factors
- and the identity of the amino acid in the ribosome A and P site.

1.6 Outline of this thesis

Nascent peptides able to stall the translating ribosome have been characterised by mutational analysis of the stalling peptide (successive alanine substitution, spacing between important residues and conservative/ non-conservative mutations), mutational analysis of the ribosome residues, the reactivity of the peptidyl-tRNA in the P site to puromycin and structural resolution generally by cryoEM of the ribosome bound peptide. In addition other relevant and applicable molecular analyses have also been performed. The aim of this project was to carry out some of these characterisations.

Although many studies were focused on identifying new 2A sequences, scanning mutagenesis has never been performed. In chapter 4, mutation was carried out on the consensus motif of TaV and the upstream context of FMDV, TaV and DHV 2As. This study revealed several features of 2As. The upstream context immediately preceeding the consensus motif is critical for stalling. The results suggest that 2A likely adopts different stalling mechanisms. One mechanism precludes a helical structure and favours the hypothesis of interactive partnership with the tunnel, whereas for the second mechanism a compaction or helical structure may be part of the mechanism.
Chapter 5 presents other *in vitro* characterisations of 2A-induced ribosome stalling. The main characteristic of 2A is an altered PTC leading to a restrictive A site.
It was previously shown that eRF1 and 3 were somehow linked to the re-initiation step after 2A-mediated stalling at the glycine codon. A method employing a visual reporter for cellular stress was tested, and was not successful and summarised in chapter 3.The discussion of this chapter focusses on the experimental conditions necessary to study the function eRF1 and 3 may have in rescuing the stalled ribosome.
And finally, in chapter 6, evidence is provided for the presence of 2A in two cellular organisms: *Saccoglossus kowalevskii* and *Branchiostoma floridae*. Notably, the 2A element was identified at the N-terminus of two clades of non-LTR retrotransposons in *B. floridae*.

Chapter 2 Materials and generic methods

2.1 Solutions, bacterial strains and enzymes

For preparation of media as well as reagents, water was deionised and sterilised. Heat-sensitive solutions were filtered using Millex 0.2 µm syringe filters (Millipore, PES membrane). Chemicals were purchased from various sources, mainly Sigma-Aldrich and were at least of analytical grade.

2.1.1 Solutions, media and other reagents

Standard molecular stock solutions were prepared following standard protocols (Sambrook et al., 1989).

- 50× TAE (2 M Tris-base, 0.05 M EDTA pH 8.0, adjusted to pH 7.8 with glacial acetic acid
- 10 and 40 %(w/v) glycerol stock filtered sterilised.
- DNA gel loading buffer type III (30 %(w/v) glycerol, 0.25 %(w/v) bromophenol blue and 0.25 %(w/v) of xylene cyanol, stored at −20°C).
- 100 mg/ml ampicillin stock filtered and stored at −20 °C and added to media to a working concentration of 1 µg/ml.
- 0.1 M IPTG, working concentration of 0.1 mM (*E. coli* blue/white screening), 0.4 mM or 1 mM (for bacterial protein expression).
- 50 mg/ml XGal, dissolved in DMSO, filtered, wrapped in foil and stored at −20 °C, added to final concentration of 40 µg/ml.
- 1 %(w/v) agarose gel for routine DNA analysis was sourced from Biogene and for DNA gel extraction low melting point Seakem LE agarose from Cambrex was used.
- 2× SDS loading dye (0.5 M Tris-HCl pH 6.8, 4.4 %(w/v) SDS, 20 %(v/v) glycerol, 2 %(v/v) 2-mercaptoethanol, and 20 %(w/v) bromophenol blue).
- 1 M potassium phosphate adjusted to pH 6 (132 ml of 1M K_2HPO_4 + 868 ml 1M KH_2PO_4)
- 20 %(w/v) biotin (made fresh and filtered sterilised).
- 20 %(w/v) dextrose, filter-sterilised.
- 13.4 %(w/v) yeast nitrogen base with ammonium sulfate and without amino acids, filter-sterilised.

- 1 %(w/v) yeast extract -2 %(w/v) peptone for *Pichia pastoris* culture was disolved in water and autoclaved. Yeast extract and peptone were purchased from Oxoid.
- Luria Bertani (LB) broth and 1 % (w/v) agar for bacterial culture were purchased from the university store.
- dNTPs were purchased from Promega and aliquoted in 10 µl volume, stored at −20 ^{o}C and used only once when thawed. Molecular weight markers used were 1 kb marker and 100 bp step ladder from Promega. Ethidium bromide at 10 mg/ml for DNA visualisation was also supplied by Promega.
- DMEM, fetal calf serum (FCS), trypsin and PBS were purchased form the university stores.
- Optimem was sourced from Gibco.
- Cell transfection reagent PEI (40 000 Da, from polysciences) was diluted in HEPES pH 7 at 1 mg/ml at −20^{o}C, and filter-sterilised in sterile environment and stored at −20^{o}C. Other transfection reagents were lipofectamine 2000 (Invitrogen) and fugene 6 (Promega).
- CelLytic for cell lysis was purchasd from Sigma.
- Restore western blot stripping buffer was purchased from Thermo Scientific.
- Protein marker precision plus protein standard was purchased from Biorad.
- PBST solution for western blot 0.5 % (v/v) tween 20 dissolved in PBS.

2.1.2 Bacterial strains and cell lines

Escherichia coli strain used for all routine bacterial transformation was JM109, genotype: *end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F´ *tra*D36, *pro*AB, *laq*I^{q}ZΔM15].

DE3 strains for protein expression (from this laboratory stock or kind gift from Prof. Naismith laboratory)

BLl21: F$^-$ *ompT gal dcm lon hsdS$_B$ (r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]*.

BL21*: F$^-$*ompT hsdS$_B$ (r_B^-, m_B^-) gal dcm rne131 (DE3)*

C43: F$^-$ *ompT hsdS$_B$* (r_B^- m_B^-) *gal dcm* (DE3)

Origami: *Δ(ara-leu)7697 ΔlacX74 ΔphoA Pvu*II *phoR araD139 ahpC galE galK rpsL*F′[*lac*$^+$ *lacI*^{q} *pro*] (DE3)*gor522*::Tn*10 trxB* (Kan^R, Str^R, Tet^R)

Rosetta: F$^-$ *ompT hsdS$_B$(R_B^- m_B^-) gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) pLysSRARE (Cam^R)*

TUNR: F$^-$ *ompT hsdS$_B$* (r_B^- m_B^-) *gal dcm lacY1*(DE3)

BLR: F$^-$ *ompT hsdS$_B$*(r_B^- m_B^-) *gal dcm* (DE3) Δ(srl-recA)306::Tn*10* pLysS (Cam^R, Tet^R)

Mammalian cells: fibroblast cells baby hamster kidney cells BHK21, human tumour cells HeLa (kindly provided by Prof. Elliott's group) and human embryonic kidney cells 293T (a kind gift from Prof Randall) and bovine lung cell (EBL) from this laboratory stock were revived and maintained through continuous passing.

2.1.3 Enzymes, antibodies and kits

Routine PCR were performed with GoTaq DNA polymerase (Promega), amplification of fragments for cloning purposes were performed with platinum DNA polymerase (Invitrogen) and PCR for mutagenesis purposes were carried out using the KOD Hot Start polymerase mix (Novagen).

Enzymes for DNA restriction digests were purchased from Fermentas and New England Biolabs. Primers were ordered individually from IDTDNA and resuspended in sterile water to a concentration of 100 pM for PCR for cloning purpose, 100 ng/µl for mutagenesis purpose and 32 pM for sequencing reactions. The relevant primers for each study are mentioned in the related chapters. Sequencings were outsourced from Dundee sequencing services.

For routine DNA extraction, QIAprep Spin miniprep Kit and HiSpeed Qiagen maxi prep.
For purification of PCR fragments from gel the Wizard SV gel and PCR clean up (Promega) was selected. For thymine-adenine cloning, the pGEM-T system was selected (Promega). All other ligations were carried out using the T4 DNA ligase kit (Promega).

Cell-free systems for protein expression: T7 quick coupled transcription- translation system was used for routine analysis of novel 2A and mutants (Promega). The T7 coupled version of this system was used to incorporate analogs of proline. T7 insect cell-free system (Promega) was used to test for the viability of constructs for bacterial expression.

The *Pichia pastoris* PichiaPink expression system (Invitrogen) was selected for expression of 2A.

SDS-PAGE protein analyses were run with the Nu-PAGE 10 % Bis-Tris 1.0 mm, 12 wells Gel and MES buffers from Invitrogen or the RunBlue pre-cast 4-20 % , 12 wells and Tris-tricine-SDS buffer from Expedeon. ECL solutions kit for chemiluminescence was purchased from Amersham Biosciences. BCA (Bicinchoninic acid) protein assay kit for microscale determination of protein concentration was purchased from Novagen.

Table 2.1: List of antibodies

Antibody	Target protein	source
Anti- His	His tagged expressed proteins	Sigma
Anti V5	V5 tagged expressed proteins	Kindly provided by Prof. Randall
Anti eEF2	Elongation factor 2	Cell signalling
Anti 2A	2A	This laboratory stock
Anti βtubulin	β tubulin	Roche Diagnostic
Secondary antibodies: Anti mouse or rabbit HRP	Mouse or rabbit primary antibody	Dako

Table 2.2: List of equipment

Function	Equipment	Supplier
PCR	GeneAmpPCR system 9700	Applied biosystems
DNA gel electrophoresis	Horizon 11-14	Life Technologies
	Allegra 21^R centrifuge	Beckman Coulter
Cell/ bacterial cultures	Sterile 96 and 6 -well plates	Greiner Bio-one
	Sterile flasks	
	Sterile Petri dishes	
	Incubator: Hera cell 150	Thermo Electron Corporation
Microscopy	Evos *fl*	AMG micro
	Delta vision	Applied Precision
Electroporation	Gene Pulser Xcell	Biorad
Protein transfer to membrane	iBlot transfer	Invitrogen
microplate reader	Infinite M200 Pro	Tecan
Protein electrophoresis and gel analysis	XCell SureLock™ Novex Mini-Cell	Invitrogen
	Model 583 gel drier	Biorad
	Kodak X-OMAT 1000 processor	Kodak

2.2 Protocols

2.2.1 Polymerase chain reaction (PCR)

PCR was routinely used to amplify small quantities of a segment of DNA, insert restriction sites or create novel 2A sequences at the 3' or 5' end of a DNA template. The specific pairs of primers are detailed in the material and method sections of each of the result chapters.

For cloning purposes, PCR were performed using platinum DNA polymerase, which has 5'3' exonuclease activity, as well terminal transferase activity and uses the following 10× buffer (600 mM Tris-SO_4 (pH 8.9), 180 mM $(NH_4)_2SO_4$). The PCR amplified products carrying additional adenosine triphosphate residues are suitable for cloning to vectors with thymine-overhangs such as pGEM-T.

For routine analytical procedures, the GoTaq polymerase was used with the proprietary buffer supplied.

The annealing temperature was estimated by subtracting 10 ^{o}C from the melting temperature (supplied by the manufacturer) of a primer pair. The reaction mixtures were set in 200 µl eppendorfs in a total volume of 50 µl, using 200 ng of plasmid DNA as template. Each reaction contained also, 5 µl of 10× reaction buffer, 1 µl forward primer at 100 pM, 1µl reverse primer at 100 pM, 1 µl dNTP mix, 1 µl of DNA polymerase at 2.0 U/µl, and sterile water to a final volume of 50 µl.

The samples were subjected to the following cycling programme: 94 ^{o}C for 2 min, 25 cycles of 94 ^{o}C (30 s), annealing temperature (55 to 68 ^{o}C) (30 s) and 72 ^{o}C (1 min for 1kb of product size) and a final prolonged extension of 7 min at 72^{o}C.

For mutagenesis, the reaction mixtures were set in a total volume of 20 µl using 10 ng of plasmid DNA as template, 1 µl of 100 ng/µl of forward and reverse primers, and 0.04 U of KOD (10 µl of mix). KOD creates blunt end fragments and has high fidelity and high processivity. The PCR programme was reduced to 18 cycles and the extension time adjusted to the size of the plasmid on the basis of 25 s/Kbp. The template DNA was then digested with restriction enzyme 1 µl of Dpn1 for 1h at 37 ^{o}C, the restriction enzyme was subsequently inactivated at 70 ^{o}C for 10 min.

The PCR products were then visualised on 1 %(w/v) agarose gel.

2.2.2 DNA gel electrophoresis and gel extraction

DNA samples were analysed by gel electrophoresis. Agarose to give a 1 %(w/v) solution was dissolved in TAE buffer. Routinely, gels were run at 100 V in horizontal gel tanks. For visualisation of DNA under UV light, 7 µl of ethidium bromide was added to 100 ml of agarose solution before pouring. 2 µl of 6× DNA gel loading buffer was added to 10 µl of sample before loading onto gel. To perform a DNA agarose gel extraction, the same protocol as previously described was followed using a low melting point agarose gel instead. The DNA segment of interest was excised using a clean disposable scalpel. Gel slices were placed in a previously weighed eppendorf tube and dissolved according to the protocol for Wizard, SV gel and PCR clean up. The DNA was resuspended in 50 µl of water and stored at −20 ^{o}C.

2.2.3 Enzymatic restriction digestions

Restriction enzymes were employed following the recommendations specified by suppliers, for buffers and active temperatures.

Routinely, 1 µg of DNA was digested with 1 U of enzyme. 5 to 10 µg of DNA were used for preparative double digestions (set in 20 µl) and 500 ng for routine analysis (set in 10 µl). The reactions were incubated for 2 h at optimum temperature (generally at 37 ^{o}C), before analysis by gel electrophoresis.

2.2.4 TA cloning in pGEM-T-easy

pGEM-T vector contains a thymine overhang able to anneal to the adenine overhang added to the PCR product by DNA polymerases such as *Taq*. In addition the pGEM-T vector has T7 and SP6 RNA polymerase promoters flanking a multiple cloning region with also the α-peptide coding region for β-galactosidase. Insertional inactivation of the α-peptide allows for white/blue selection of recombinants.

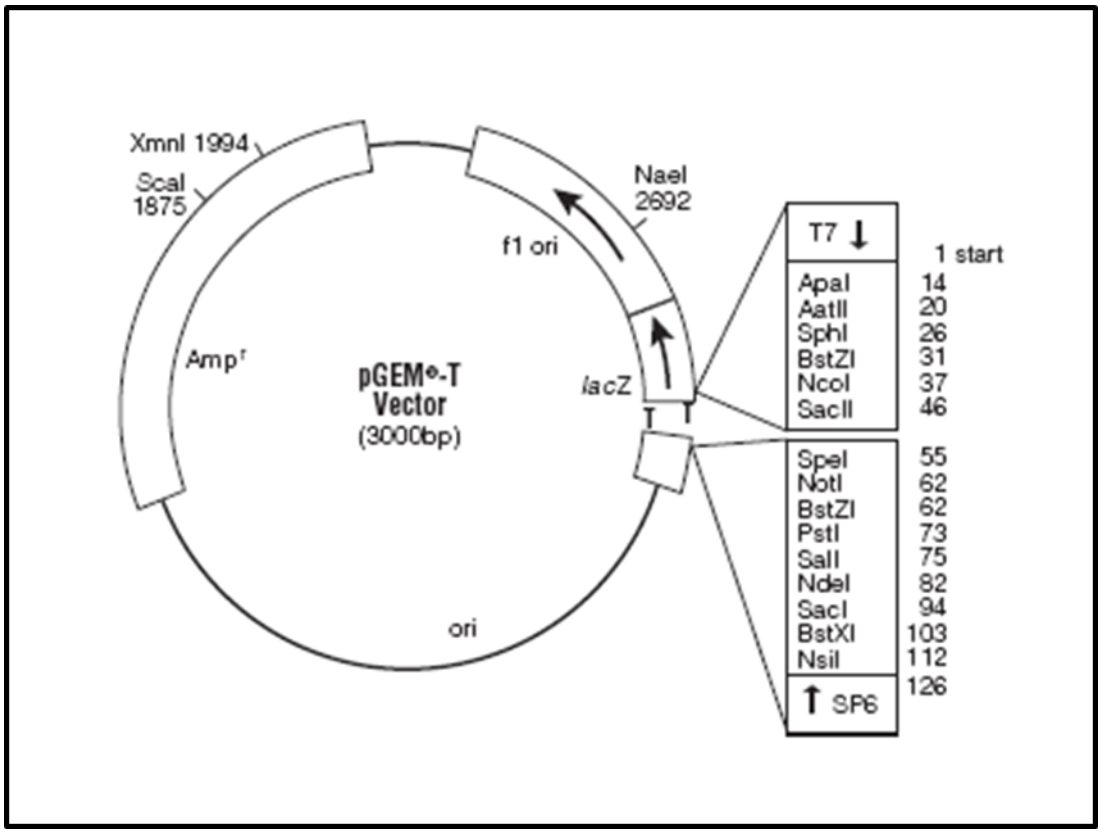


Figure 2.1: Map of pGEM-T vector. (Sourced from http://www.promega.com)

3 µl of the PCR product gel purified was added to a ligation mixture. A positive control (using a segment of DNA provided by manufacturer) as well as a negative control (with no insert) was set up in parallel. Ligations were set to a final volume of 10 µl. The reactions were incubated o/n at 4 ^{o}C to maximise ligation yields.

The *E. coli* JM109 transformation protocol is described in section 2.2.6, when using pGEM-T mixtures, LB agar - ampicillin plates were also supplemented with IPTG and X-Gal, before incubation at 37 ^{o}C o/n. The transformed clones, white in colour, were picked and incubated in LB broth-ampicillin o/n.

2.2.5 T4 DNA Ligation

Vectors and inserts concentrations were estimated visually by gel electrophoresis. A ratio of 1:3, vector: insert was used. The ligation mixtures containing vector and insert DNAs, 10 U of T4 DNA ligase and 10× ligase buffer were set up in a total volume of 10 µl and incubated o/n at 4 ^{o}C.

2.2.6 Transformation by heat shock of competent *E. coli* strains.

Transformation of competent *E. coli* strains were carried out by adding plasmid DNA or ligation mixtures to 50 µl cells. After 20 min of incubation on ice, cells were heat-shocked at 42 ^{o}C for 50 s and returned to ice for 2 min before addition of 950 ml LB broth. The cells were incubated at 37 ^{o}C for 90 min. 100 µl of cells were spread onto LB agar plate containing ampicillin and incubated o/n at 37 ^{o}C. Several transformed colonies were transferred with disposable loops to 10 ml LB broth-ampicillin and incubated o/n at 37 ^{o}C, shaking at 200 rpm. For the purpose of DNA extraction, the bacteria were harvested by centrifugation at 3000 rpm for 10 min.

For the purpose of preservation, at −80 ^{o}C as 20 %(v/v) glycerol stocks, 500 µl of culture was aliquoted into a sterile eppendorf containing 500 µl of 50 % (v/v) glycerol.

2.2.7 Plasmid DNA extraction and sequencing

Plasmid DNA was extracted using either Qiagen plasmid mini kit for routine analysis or Qiagen Hi-speed maxi kit. DNA extractions were carried out following the manufacturers' protocols.
Purity and concentration of DNA were assessed spectrophotometrically with absorbencies of samples measured at 260 and 280 nm.
Plasmid DNA were extracted and restricted with appropriate enzymes to identify correct clones. Sequencing reactions were outsourced to Dundee sequencing services, using either T7 or SP6 promoter to initiate the thermal cycling. Sequencing was performed with 600 ng of sample plasmid DNA and 3.2 µM primers. Once available the sequence of each clone was aligned against the expected construct using DNAMAN software.

2.2.8 Protein expression in *E. coli*.

Several human translation factors: eEF2, elongation factor 2 kinase (eEF2-K), eRF1 and eRF3 were previously cloned in TOPO system by previous members of this laboratory.
The TOPO vector has C-terminal V5 and polyhistidine tags and is suitable for IPTG induction of protein expression in DE3 bacterial strains. They were transformed in *E. coli* following the protocol in section 2.2.6.

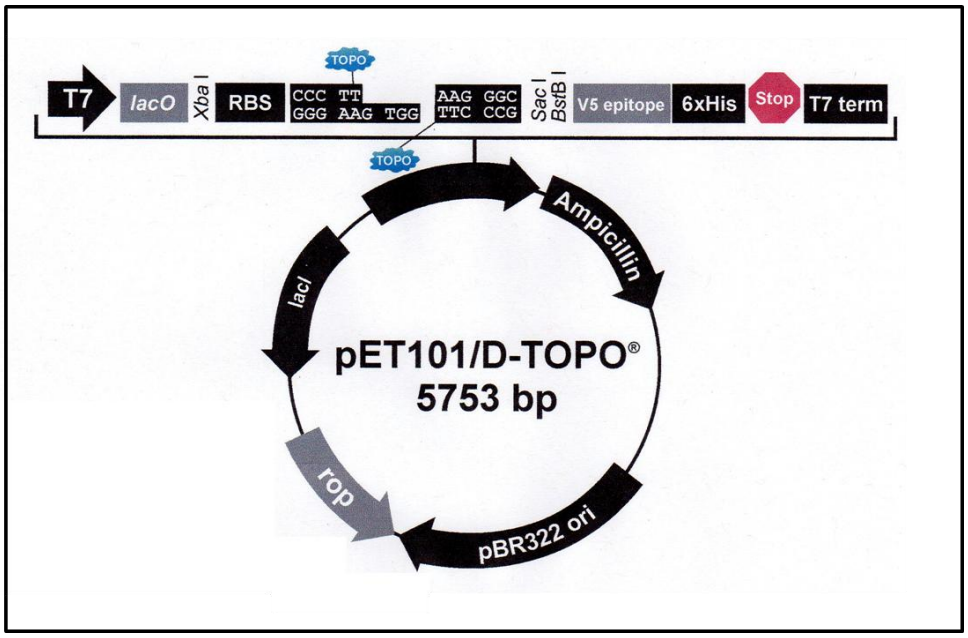


Figure 2.2: Map of TOPO vector.

T7 promoter, Lac operator (lacO), ribosome binding site (RBS), TOPO cloning sites for directional cloning of PCR products with CACC overhang, V5 epitope, polyhistidine (6xHis) region for affinity purification, T7 transcription termination region, ampicillin resistance gene. (Vector map sourced from http://www.lifetechnologies.com)

10 ml of LB- ampicillin was inoculated with one colony and cultured o/n at 37 ^{o}C. 10 ml of LB-ampicillin was inoculated with 100 µl of the o/n culture and further incubated until the OD_{600nm} reached 0.6. Expression was induced with either 0.4 or 1 mM IPTG. Negative controls were preprared the same way and run in parallel but not supplemented with IPTG.
The bacterial suspensions were cultivated further for 4 h and subsequently harvested at 3000 rpm for 10 min. The supernatants were discarded and the cells resuspended in 1 ml of PBS to which was previously added a protease inhibitor tablet (Roche). The cells were sonicated for 10 s twice and centrifuged at 16 000 rpm for 10 min at 4 ^{o}C to separate the cellular debris from the soluble fraction. The protocol was repeated for three induction temperatures: 16 ^{o}C, 24 ^{o}C and 37 ^{o}C.
3 µl of supernatant were analysed for protein expression on Coomassie-stained SDS-PAGE gels and western blot analysis.

2.2.9 Protein expression in eukaryotic cell-free systems

The rabbit reticulocyte lysate system (Promega) provides an *in vitro* eukaryotic expression system for proteins. The quick coupled transcription translation uses plasmid DNA with a T7 promoter. The coupled system version offers the flexibility for modifying the amino acid content. The T7 insect cell-free (ICE) system (Promega) also provides a simplified eukaryotic system for expression of proteins. The reaction mixtures were set up following the manufacturer instructions. For rabbit reticulocytes lysates, proteins were radiolabelled using [^{35} S]- methionine (10 µCi/µl). 200 ng of a plasmid preparation was added to the TnT reaction mixture to a final volume of 11.5 µl and incubated at 30 ^{o}C for 60 min. For T7 ICE reactions, 2 µg DNA was added and incubated for 4 h at 30 ^{o}C, without radiolabels to a final reaction volume of 50 µl. 5 µl of the translation reactions were resolved by SDS-PAGE and western blot (section 2.2.16).

2.2.10 2A expression in *Pichia pastoris*

The growth media were prepared according to the manufacturer's protocols, the kit included some of the media, and the others were purchased separately from Invitrogen.
The YPD medium to culture *P. pastoris* strains and the selection medium PAD (pichia adenine drop-out) lacking adenine were part of the kit. After dissolution in water and autoclaving, the media were supplemented with 100 ml of 20 %(w/v) dextrose.
1 l of starter medium BMGY was made with 700 ml sterilised 1 %(w/v) yeast extract and 2 %(w/v) peptone, 100 ml 1 M potassium phosphate at pH 6, 100 ml 10 %(v/v) glycerol, 20 ml biotin solution and 100 ml of YNB without amino acids.

1 l of induction medium was made with 700 ml sterilised 1 %(w/v) yeast extract and 2 %(w/v) peptone, 100 ml 1 M potassium phosphate at pH 6, 100 ml 5 %(v/v) methanol (filter-sterilised), 20 ml of biotin solution and 100 ml YNB without amino acids.

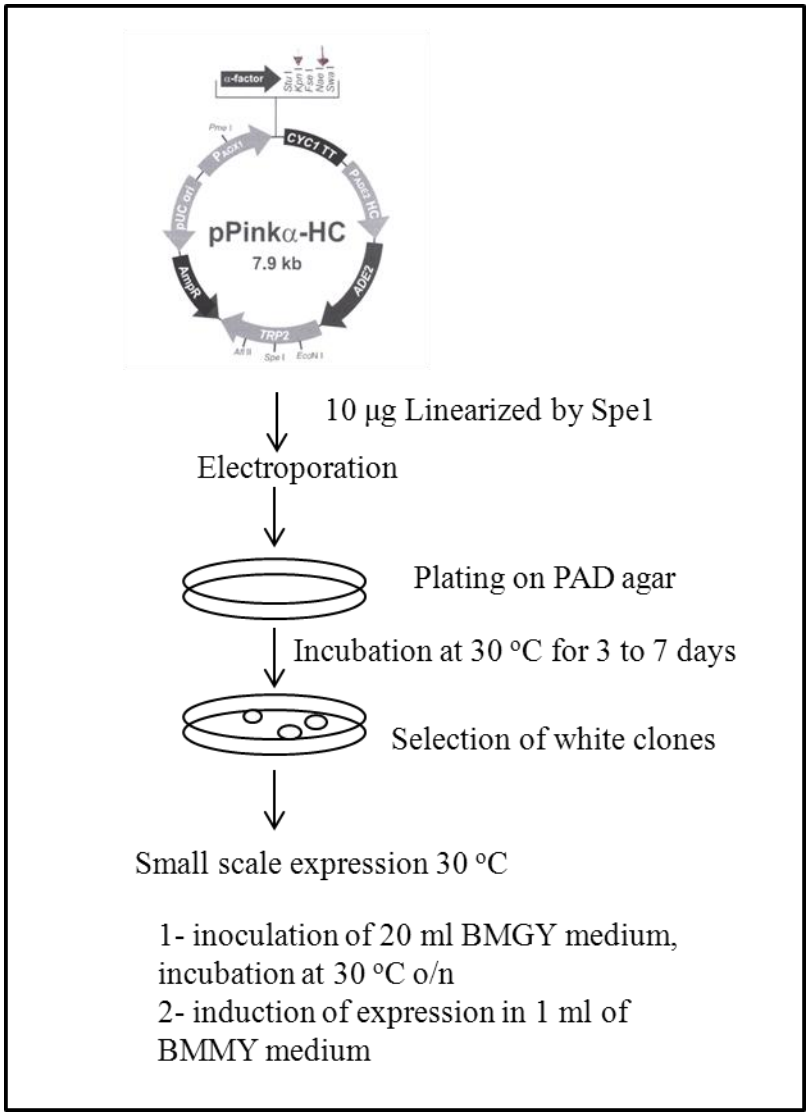


Figure 2.3: Experimental outline for protein expression in Invitrogen PichiaPink system.

Preparing electrocompetent PichiaPink strains, electroporation and selection

YPD agar plates were inoculated with PichiaPink strain 4 and incubated at 30 ^{o}C for 3 days until colonies were formed. 10 ml of YPD media was inoculated with 1 colony and incubated in a shaking incubator for 24 h at 300 rpm. The starter culture was transferred to 100 ml of YPD media in a 1 litre flask to a OD_{600nm} of 0.2. When the OD reached 1.3- 1.5, the yeasts were harvested at 4500 rpm at 4 ^{o}C for 5 min, the supernatant was discarded and the cells re-suspended in 250 ml ice-cold sterile water.

The cells were subjected to several centrifugations at 4500 rpm for 5 min and at 4 ^{o}C. The supernatant each time was discarded and cells were resuspended serially in 50 ml of ice-cold sterile water, 10 ml ice-cold 1 M sorbitol and finally in ice-cold 1 M sorbitol to a final volume of 500 µl.

PichiaPink strains were prepared freshly prior to electroporation. 80 µl of electrocompetent cells were mixed with 10 µg Spe1 linearized vector, transferred to a 0.2 cm electroporation cuvette pre-chilled and incubated on ice for 5 min. The cuvettes were pulsed in Biorad gene pulser Xcell with the following settings: 2 kV, 25 Ω, 200 µF. 1 ml of the pre-chilled YPD was added and the cells were further incubated at 30 ^{o}C for 2 h without shaking. 100 and 300 µl were spread onto a PAD selection plate and incubated at 30 ^{o}C for 3 to 7 days until individual colonies started to form. Single white colonies were selected and re-plated on fresh PAD agar and cultured under the same conditions.

Extraction of DNA and PCR of transformants

Extraction of DNA was performed using the YeaStar Genomic DNA kit, following the instructions for protocol 1. Single colonies were cultured in 1 ml YPD medium for 5 h, the cells were collected by centrifugation at 4500 rpm for 2 min. The supernatant was discarded. 120 µl of digestion buffer and 8 µl of R-zymolase were added to the pellet. The mixture was vortexed and incubated for 1 h at 37 ^{o}C. 250 µl of chloroform was added and mixed for 1 min before proceeding further to the routine DNA extraction protocol, with the columns provided. The genomic DNA was subjected to a PCR and electrophoresis as described in section 2.2.1.

Small scale expression of recombinant strains

A single colony was used to inoculate 10 ml of BMGY medium in a 500 ml sterile flask. The cultures were grown for 24 h in a 24 ^{o}C incubator set to 300 rpm. The cells were harvested at 4500 rpm for 5 min at room temperature, and resuspended in 1 ml of induction media BMMY and cultivated o/n. 100 µl was removed from the samples and 100 µl of 40 %(v/v) methanol added and let to continue to grow for another 16 h. Cells were centrifuged for 10 min at 4500 rpm. The supernatant fractions were analysed for protein expression by western blot analysis.

2.2.11 Mammalian cell culture

DMEM-10 %(v/v) FCS media, trypsin and PBSA were stored at 4 ^{o}C, these were then pre-warmed to 37 ^{o}C for 1 h before usage. The old media from the cells were aspirated under sterile conditions. The cells were then washed with PBSA, trypsin was added (1 ml or 2 ml depending on the size of the container) and the flask placed in incubator until cells were seen floating. To ensure that these were individually separated, the flasks were tapped against a hard surface. 3 ml for T25 or 8 ml for a T75 of DMEM-10 %(v/v) FCS media were added and pipetted up and down a few times. The cells were either grown further in a flask or in a sterile 96-well, or 6-well plates.
1 ml of the cells was used to seed 9 ml of DMEM-10 %(v/v) FCS, and cultured in a new T25 or T75 sterile flask for maintenance. The culture was incubated for 3 days at 37 ^{o}C, 5 % CO_2, in a humidified incubator. The cell maintenance protocol was repeated every 3 days up to 30 times, before new cells were thawed from a −80 ^{o}C stock.
96-well plates were seeded with 20×10^4 cells per well, 6-well plates were seeded with 60×10^4 cells for functional assays and were incubated for 24 h before transfection.

2.2.12 Transient transfections of mammalian cells

Transfection enables the introduction of negatively charged molecules such as DNA into cells without using a virus. Confluency of cells was assessed visually under microscope. The following protocols required cells to be about 50 % confluent.

Using PEI, 5 µg of plasmid DNA of interest was mixed with 600 µl DMEM (serum free) pre-warmed for 1 h at 37 ^{o}C. 15 µl of 1 mg/ml PEI solution (ratio of 1:3 of DNA: PEI) was added and the mixture vortexed for 5 s and incubated at room temperature for 30 min to form the miscelleous complex PEI-DNA. 100 µl of PEI-DNA mix was then added drop-wise to each well to a six-well plate, and swirled to gently mix.
Lipofectamine 2000 required 1.5 µg of DNA per well for a 60 mm petri dish. 200 µl of Opti-MEM was mixed with 5 µl of lipofectamine, incubated for 5 min. In parallel 200 µl of Opti-MEM was mixed with 1.5 µg of plasmid DNA of interest. The two mixtures were combined and incubated for 15 min at RT, before being added drop-wise to a 60 mm petri dish.
Using Fugene 6, 6 µl of Fugene 6 was diluted into 90 µl of serum free DMEM and incubated for 5 min at RT. 2 µg of plasmid DNA was added and incubated for a further 15 min.
The plates were returned to the 37 ^{o}C, 5 % CO_2 humidified incubator for 24 h and the outcome of transfection assessed by microscopy. Alternatively, the cells were also harvested for western blot analysis.

2.2.13 Establishing HeLa stable cell lines with lentiviral vectors.

Lentiviruses allow for the long term study of transfected cells as they rapidly and stably integrate the genome of a broad range of cells.

Three vectors were co-transfected in 293T cell, cultured in 75 cm^2 flask 50 % confluent. The co-transfection was carried out as described in section 2.2.13. For a 75 cm^2 flask, 33 µl of Fugene 6 was added to 467 µl of serum-free DMEM. 3 µg of the envelop vector (VSV-G), 3 µg of the packaging vector (CMV) and 5 µg of the transfer vector (carrying antibiotic selection and the insert of interest) were added to the mix. Post-infection, the medium from the cells were collected and another 10 ml of DMEM-10 %(v/v) FCS added. After 24 h, the medium was collected. The cellular debris were removed by centrifugation for 15 min at 3500 rpm. The supernatant was filter-sterilised and aliquoted in 1 ml vials and stored at −80 ^oC.

Infection by lentiviruses was performed on 30 to 50 % confluent HeLa cells. For a 25 cm^2 flask, 1 ml of lentivirus aliquot was added to 1 ml serum-free DMEM, supplemented with polybrene to a working concentration of 8 µg/ml. The mixture was pipetted in a pre-emptied flask of HeLa cells and left to rock for 2 h at 37 ^oC. 2 ml of 10 %(v/v) FCS DMEM was subsequently added and the cells incubated for 2 days before puromycin was added to select for transformants.

2.2.14 Fixing cells for microscopy and imaging

Sterile coverslips were placed at the bottom of individual wells of a six-well plate. After 24 h the cells were treated or transfected and returned to their experimental conditions. After a further 24 h, the media was discarded, the well washed twice with pre-warmed PBS and 3 ml of 5 %(v/v) formaldehyde was added to fix the cells to the coverslip. The formaldehyde was aspirated and the well washed several times with PBS. With a pair of tweezers, the coverslips were removed and placed onto glass slides so that the cell surface was in contact with 15 µl of Mowiol-Dapi mounting solution. The coverslips were left in the dark on a flat surface. The fixed cells were visualised with the DeltaVision microscope (Applied Precision). This system consists of a light source (X-cite 120Q lamp), an inverted microscope, a ×100 oil immersion objective and the softWoRx software package for capturing and processing high resolution static images. The DeltaVision permits the observation of micro-anatomical details and the use of multiple filters. Fluorescences were detected with an excitation wavelength of 488 nm for GFP (FITC filter), 587 nm for cherryFP (RFP filter) and 350 nm for Dapi (DAPI filter).

2.2.15 Mammalian cell lysis for protein analysis

The media were aspirated from the cell cultures and the cells were washed with pre-warmed PBS 3 times. The cells were scrapped and incubated for 15 min on a rocking shaker with 200 µl (for a six-well plate) or 400 µl (for 60 mm dish) of celLytic buffer (Sigma) which was supplemented with PhosStop and mini-complete tablets inhibitors for phosphatases and proteases (Roche). The cellular debris were removed by centrifugation at 7.500 rpm for 15 min at 4 ^{o}C. The supernatant was stored at −80 ^{o}C and analysed by western blot.

2.2.16 Protein analysis

2.2.16.1 SDS-PAGE analysis

5 µl of samples were mixed with an equivalent volume of 2× SDS-loading buffer. The samples were further incubated at 94 ^{o}C for 2 min and returned to room-temperature before being loaded onto the SDS-PAGE gel. Cell extracts and bacterial cultures supernatants were loaded onto a NuPage 10 % SDS gel (Invitrogen). Rabbit reticulocyte lysates reaction samples were run onto a 4-20 % RunBlue SDS-PAGE gel (Expedeon).
The gels were Coomassie stained with InstantBlue (Expedeon), following the manufacturer's recommendation. Alternatively the gels were dried for 30 min and radioactively labelled proteins exposed to a photographic film o/n.

2.2.16.2 Western blot analysis

Proteins resolved by SDS-PAGE electrophoresis were transferred to a nitrocellulose membrane using the iBlot transfer system (Invitrogen). Following transfer, the membranes were blocked for 1h with a solution of 2.5 %(w/v) skimmed milk-PBST. The primary antibodies were diluted 1:2000 in 2.5 %(w/v) skimmed milk- PBST solution. Membranes were then incubated for 4 h with the primary antibodies, washed three times in PBST for 5 min. The membranes were then incubated with the secondary-HRP bound antibody for 2 h and following three further 5 min washes in PBST, the bound antibodies were detected by chemiluminescence using the ECL kit (Amersham Biosciences)

2.2.16.3 Total protein quantification by BCA assay

Where required (chapter 3) the total protein concentration was determined by BCA assay (Novagen) following the manufacturer's protocol. The BCA protein assay allows micro-scale determination (in the Tecan microplate reader) of total proteins from cell lysates and is based on the reduction of Cu^{2+} by protein in alkaline conditions. The Bicinchoninic acid (BCA) chelates the reduced copper and produces a purple complex absorbing at 562 nm. The protein standard bovine serum albumin was diluted to final concentrations of 0, 25, 125, 250, 500 and 1,000 µg/ml. 25 µl of each standard was introduced in individual wells of a 96-well plate. 200 µl of the BCA-cupric sulfate reagent was added to each well, and incubated in the dark for 2 h before the absorbance at 562 nm was measured. A standard curve was generated from the absorbance of the protein standards, and the concentration of protein in the samples interpolated from the plot.

Chapter 3 Analysis of re-initiation following ribosome stalling

3.1 Introduction

These experiments were inspired by an early study by Svitkin and Agol (1983). The authors showed that eEF2 modulated the translation of the picornavirus EMCV RNA at the 2A/2B junction. They used a Krebs cell-free extract which had limited amounts of eEF2 to translate EMCV RNA, and supplemented the cell extracts with purified eEF2 and at time intervals withdrew aliquots. The results proved that on SDS-PAGE, the radiolabeled polypeptide corresponding to the RNA segment after the 2A region started to appear after 40 min of translation for the sample, whereas for the negative control (without additional eEF2), only a small amount of this polypeptide was apparent after 180 min. The synthesis of sequences downstream of the 2A sequence therefore depended on the availability of active eEF2.

Additionally, The toeprint analysis performed on yeast extracts translating 2A provided compelling but not direct evidence for the involvement of the release factors (eRF1 and 3) in rescuing the 2A-stalled ribosome (Doronina et al., 2008). Peptide release involves hydrolysis of the nascent chain-tRNA ester linkage by a conserved [GGQ] motif from the release factor class 1 (Frolova et al., 1999). The eRF1 has three domains, the N (amino) domain interacts with mRNA and contains a highly conserved [NIKS] motif, and a [YxCxxxF] motif. The M domain (middle) extends in the PTC and contains the [GGQ] motif (Song et al., 2000). The C domain (carboxy) interacts with class 2 release factor: eRF3 (Salas-Marco and Bedwell, 2004). Although release factors from bacteria and eukaryotes are not direct homologues, they both contain the [GGQ] motif, and mutational studies in bacteria showed that substitutions of glutamine did not affect peptide release but substitution of the glycine residues resulted in losses of catalytic properties (Shaw and Green, 2007). Shaw and Green (2007) concluded that mutation of the glycines in the motif results in a change in polypeptide conformation that alleviate the catalysis ability and proposed that the combination of glycines and glutamine provides enough space and the correct configuration for a water molecule to reach the ribosome PTC and hydrolyse the peptidyl-tRNA ester linkage. In the model of translation termination, the [GGQ] motif is proposed to create a channel structure for water. The [GGQ] motif enters the PTC, the two glycines are thought to provide local flexibility as well as just enough space for a water molecule to

reach the PTC. Substitutions of glycine for other amino acids, occludes passage of the water molecule, therefore prevents hydrolysis (Shaw and Green, 2007).

The 2A model proposes that following eRF1/3 rescue, the nascent chain would be released and the ribosome subunits would dissociate, which the model predicts will be the inevitable outcome in conditions where limited amounts of eEF2 are present to support the next translocation step. However, in conditions where there is enough active eEF2 available the dynamics will be driven towards re-initiation of translation. Potentially the time delay between rescue and re-initiation is shortened and more ribosomes can re-initiate translation. The model predicts a linear relationship between re-initiation of translation after 2A-stalled ribosome rescue and availability of eEF2 (Doronina et al., 2008). eEF2 activity is regulated by elongation factor 2 kinase (eEF2K) and protein phosphatase 2A (PP2A). The kinase phosphorylates eEF2 and prevents its binding to the ribosome. PP2A reverses the eEF2K effect. Both of these enzymes are regulated by various cellular stress-related signalling pathways. Under conditions of stress, the cell globally inhibits protein synthesis, through repression of the initiation process, and through phosphorylation of eEF2 (Redpath et al., 1993).

There were two aims to this study.
First, this study aimed to visually demonstrate the link between translation capabilities of the cell and re-initiation of translation following the 2A motif, under conditions of stress, where the general available pool of active eEF2 is reduced.
This project and the vector pHE27 were previously developed in this laboratory using cyan fluorescent protein (cyanFP) as a fluorescent reporter protein. The experimental system uses a V5-tagged cyanFP reporter protein upstream of 2A and downstream of 2A the single-chain Fv antibody (scFv) which has a strong binding affinity for V5. Immunoprecipitation of pHE27 products translated in 293T cells proved that *de novo* V5 tag and the scFv chain could bind intracellularly. However, the cyanFP fluorescence signal was not satisfactory. At the beginning of this thesis, it was therefore decided to replace cyanFP with with cherryFP for an improved signal, transfect mammalian cells with the modified plasmids and introduce several stresses in an attempt to modulate the concentration of active cellular eEF2.
By augmenting the size of the reporter protein relying on the strong affinity of *de novo* synthesised V5 for scFV, the hypothesis is that the accumulation of the reporter would be limited to the cytoplasm only. In contrast, the cherryFP control without the V5 tag would localize in both cytoplasm and nucleus of transfected cells. In normal conditions of cell culture, a balanced ratio of V5-cherryFP-2A and the scFv should be produced. Following the application of a stress, a higher proportion of V5-cherryFP-2A should be produced and diffuse to the nucleus.

The second purpose of this study was to express the translation-regulating factors in bacteria and supplement the purified proteins to rabbit reticulocyte lysates translating a bicistronic vector carrying a 2A motif. A mutant eRF1 was created where the [GGQ] conserved motif was mutated to [RGQ]. eRF1 RGQ is unable to perform hydrolysis of peptidyl-tRNA ester linkage.

The results obtained for this thesis revealed that a visual approach employing cherryFP did not provide adequate discrimination to demonstrate the role of eEF2 in re-initiation of translation in HeLa cells. Similarly attempts to study the re-initiation of translation after 2A-induced stalling by supplementing a cell-free reticulocyte lysate with additional affinity-purified factors were not successful. The expression of translation-regulating factors (eEF2, eEF2K, eRF1 and eRF3 and eRF1 RGQ mutant) failed using various *E. coli* strains.

3.2 Materials and methods

3.2.1 Plasmids and experimental procedures for a visual method

3.2.1.1 Plasmids and cloning

The pHE27 plasmid (from the laboratory collection of clones) was modified to encode cherryFP instead of the cyanFP. The resulting plasmid pSTU1 carried the V5-CherryFP-FMDV2A-scFv insert and the negative control plasmid pSTU2 was created without the V5 epitope tag.

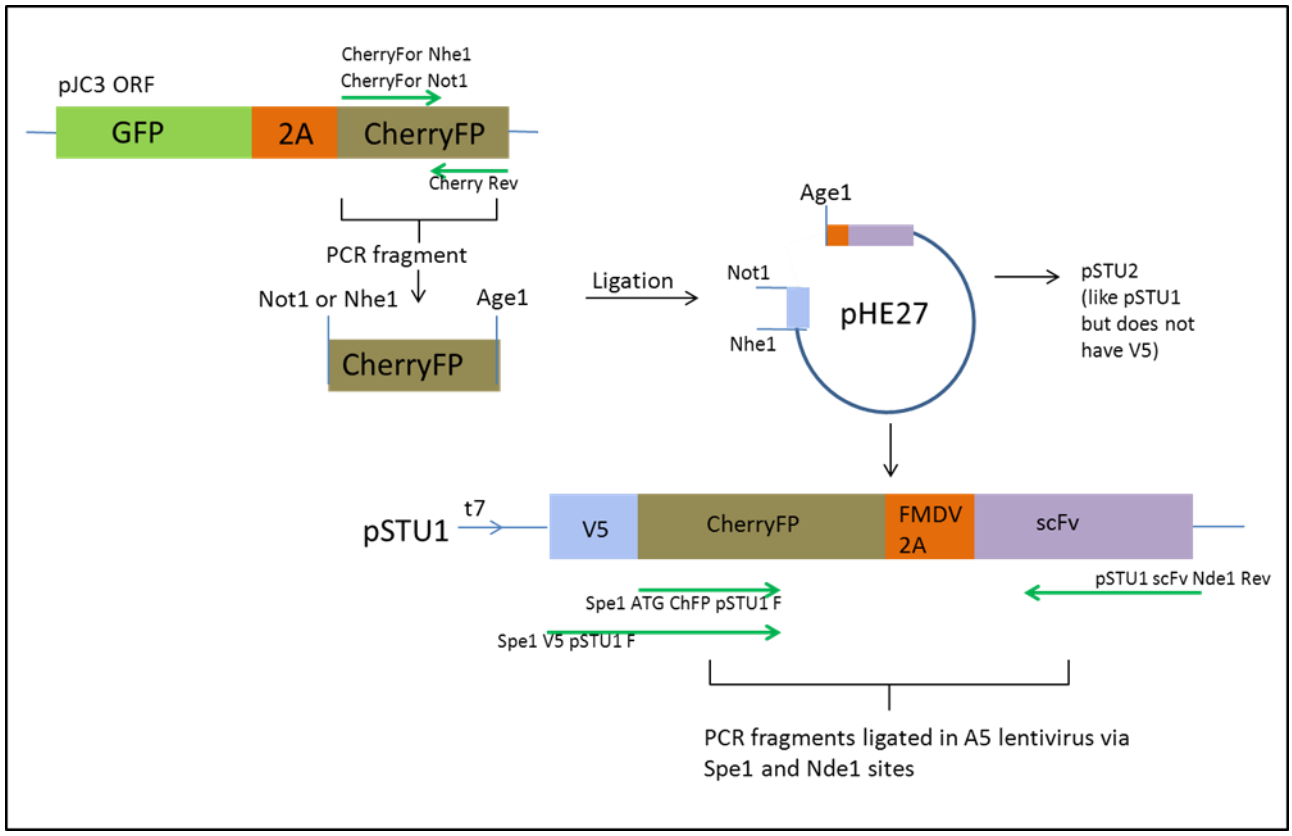


Figure 3.1: Cloning overview for the creation of pSTU1 and lentiviruses A5 for stress study

The primers are represented in green arrows, the V5 tag is blue, cherryFP in brown, FMDV in red and the scFv in violet. cherryFP was amplified from pJC3 with either Nhe1 or Not1 at the 5' end and Age1 at the 3' end. The cyanFP or V5cyanFP were extracted from pHE27 and replaced with cherryFP. This created the pSTU1and pSTU2 plasmids. The pSTU1 plasmid was amplified to add the restriction sites for insertion in the lentivirus A5.

Primers annealing to cherryFP

CherryForNot1 5’-GCGGCCGCGGTGAGCAAGGGCGAGGAGGAT-3’

CherryFor Nhe1 5’-GCTAGCACCATGGTGAGCAAGGGCGAGGAGGAT- 3’and reverse primer

Cherry Rev 5’-CACCGGTGCCTTGTACAGCTCGTCCATGCC-3’

The pSTU1 and pSTU2 plasmids have the resistance gene for selection by Neomycin, G418 (geneticin) is an analogue and is generally used to select for neomycin resistance.

To modify the A5 lentivirus vector: pSTU1 plasmid carrying V5-cherryFP-2A-scFV served as template.

forward primers:

Spe1 V5 pSTU1F 5’-ACTAGTGCTAGCACCATGGGAAAGCC-3’ to create the V5 tagged version A5V5,

Spe1 ATG ChFP pSTU1F 5’-ACTAGTATGGCGGCCGCGGTGAGCAAG -3’ to create the non tagged version A5ChFP,

and reverse primer: pSTU1 ScFV Nde1Rev 5’- CATATGAAGCTTTTACCGTCTTATTTC-3’

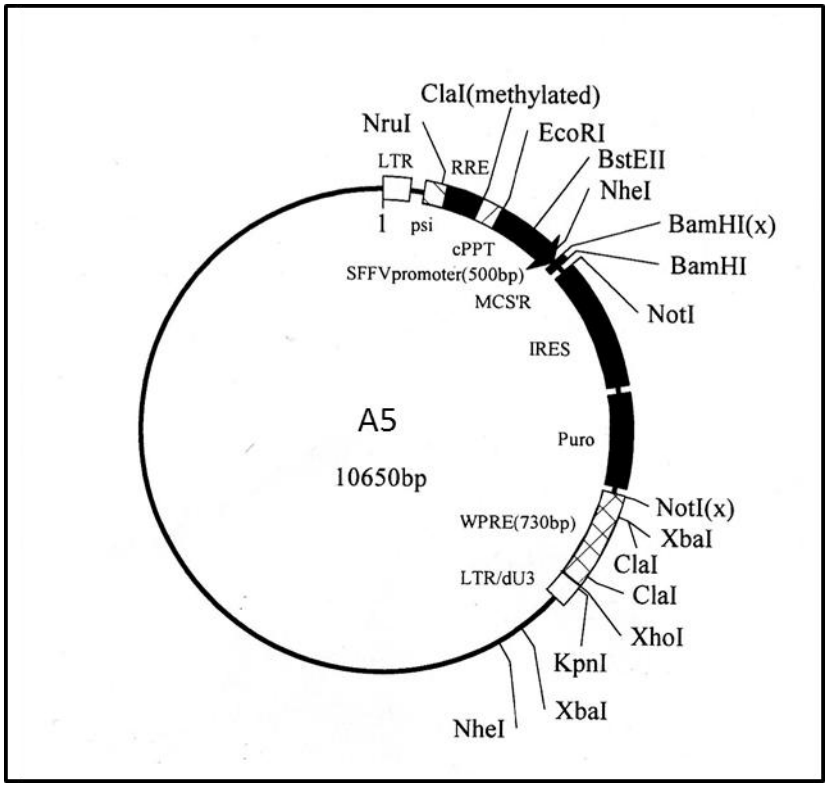


Figure 3.2: Map of A5 lentivirus expression system

The SFFV promoter (spleen-focus-forming virus) was used to drive expression of the insert cloned in the multiple cloning site (MCS’R) between the Spe1 and Nde1 sites. A5 is a bicistronic vector with EMCV IRES (IRES) driving expression of puromycin (puro) resistance gene (Map provided by Prof. Randall Laboratory).

3.2.1.2 Experimental procedures

Propagation and extraction of plasmid DNA were carried according to the protocol in section 2.2.7. The mammalian cells were maintained, transfected and mounted for microscopy following the protocols in section 2.2.11, 2.2.12 and 2.2.14. For total protein quantification, a BCA assay was carried out following the protocol in section 2.2.16.3. For western blot analyses the cells were prepared and analysed following protocol in section 2.2.15 and 2.2.16. Lentiviruses were produced and stored as mentioned in section 2.2.13 and used to infect HeLa cells. To determine the selection conditions with geneticin and puromycin, untransfected cells were cultured in presence of increasing concentration of the antibiotic of interest, from 0 to 1200 mg/ml of geneticin for BHK21 cells. BHK21 cells are resistant to puromycin and could not be used with the A5 lentivirus vector. HeLa cells were cultured in presence of up to 2.5 mg/ml of puromycin.

BHK21 cells were originally transiently transfected with pSTU1 and pSTU2 and assessed by fluorescent microscopy as described in section 2.2.14. Stable HeLa cell lines expressing V5-cherryFP-2A- scFv (A5V5 HeLa) and cherryFP-2A-scFv (A5ChFP HeLa) were created by lentivirus. Stress trials: heat-shock at 42 ^{o}C, nutrient deprivation, ethanol stress and EMCV infection were performed on the A5V5 HeLa cell line and the cells were harvested at time intervals.

The cells incubated under mild heat-shock at 42 ^{o}C were harvested after 15 min, 1, 2, 4 and 6 h but were cultured with the normal medium (DMEM-10 %(v/v) FCS).

For the nutrient deprivation experiment, DMEM was supplemented with 0, 1, 2.5, 5, 7.5 or 10 %(v/v)FCS. After the HeLa cells were seeded and cultured for 24 h in normal conditions, the media were aspirated and the wells washed with pre-warmed PBS. The cells were then incubated at 37^{o}C with nutrient deficient media for 24 h before being harvested.

For the A5V5 HeLa cells subjected to alcohol stress, DMEM-10 %(v/v) FCS was modified with increasing amount of ethanol 0 to 5 %(v/v). After the cells were seeded and left to adhere for 24 h in the six-well plates, the media were aspirated and the wells washed with pre-warmed PBS. The cells were then incubated with the ethanol modified media for 12 h before being harvested.

For EMCV infection experiment, the HeLa cells were cultured for 24 h in the six-well plates in normal condition, then infected with EMCV(multiplicity of infection [MOI]= 10^{-3}). The cells were harvested at 0, 1, 2, 4, 6, and 8 h post-infection.

The cells were seeded in six-well plates for western blot analysis or fluorescent microscopy and seeded in parallel (five replicates for each condition) in 96-well plates for the MTT protocol for cell viability. The water soluble MTT reagent (Sigma) is yellow and is metabolised to formazan (purple in colour) by viable cells. Formazan is stable; the conversion is directly dependent on the cells viability and can be quantified spectrophotemetrically at 550 nm. MTT was dissolved at 5 mg/ml in PBS and filter-sterilised, the media from the wells aspirated and 150 µl of MTT solution added to the well. The 96-well plate was incubated wrapped in foil on a shaker for 2 h at 37 ^{o}C.

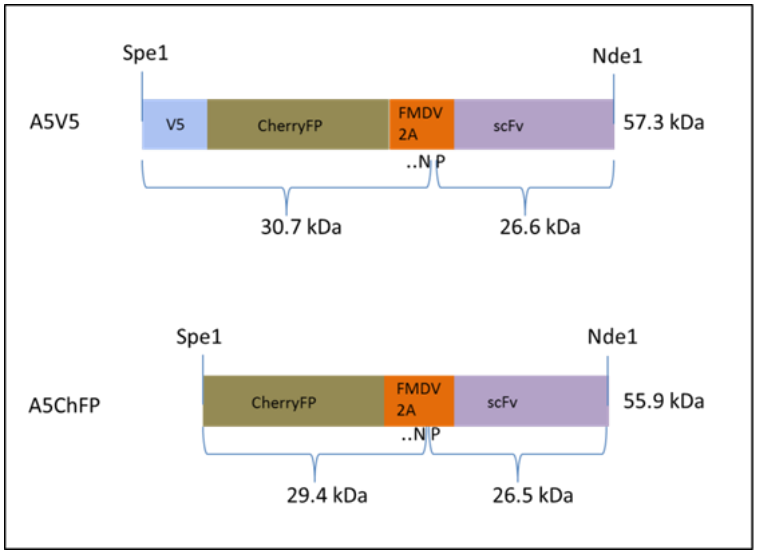


Figure 3.3: Molecular weight details of A5V5 and A5ChFP lentiviruses inserts.

3.2.2 Bacterial expression of translation-regulating factors

Six DE3 strains of *E. coli* suitable for protein expression (genotype in section 2.1.2) were heat-shocked with TOPO vectors where the coding sequences for eEF2, eEF2K, eRF1 and eRF3 were inserted. The TOPO clones were sourced from this laboratory collection of plasmid.
In addition the eRF1 [GGQ] conserved motif was mutated to [RGQ] following the protocol in section 2.2.1.
The mutagenesis primer pair for eRF1 RGQ was
RGQ eRF1 Fw 5'-GAAACACGGTAGACGAGGTCAGTCAGCCTTGCGTTTTGC-3'
RGQ eRF1 Rev 5'- GGCTGACTGACCTCGTCTACCGTGTTTCTTTGGG-3'

The positive control for the protein expression experiment was pEHISGFPTEV (~30 kDa), which confers kanamycin resistance and has a T7 promoter driving expression of N-terminally polyhistidine-tagged green fluorescent protein (kindly provided by the Prof Naismith laboratory). The negative control was the empty vector without GFP (Liu and Naismith, 2009).

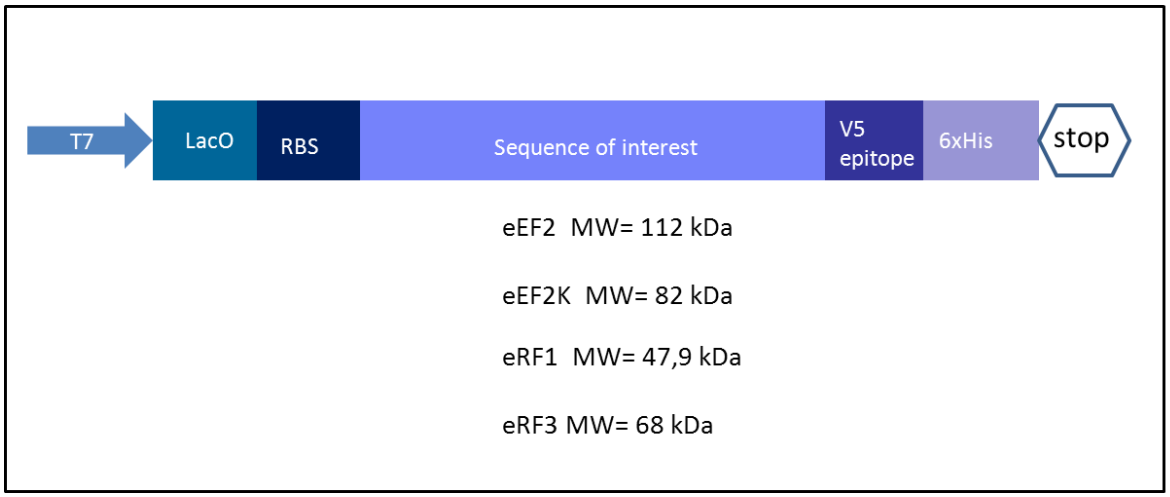


Figure 3.4: Representation of the TOPOpET101/D-TOPO vector and the molecular weight of the translated proteins

The TOPO vector has a T7 promoter, lac operator (lacO), ribosome binding site (RBS) and the C-terminal V5 epitope and poly-histidine (6xHis).

After initial failures at expressing all five proteins in BL21*, further optimisations were performed with eEF2, eRF1 and eRF1 RGQ mutant. BL21, BL21*, BLR, Origami, TUNR and Rosetta and C43 were cultured, induced and analysed as described in section 2.1.2 and 2.2.8, induced with 0.4 as well as 1mM IPTG and incubated at 16 ^{o}C, 24 ^{o}C and 37 ^{o}C. After sonication, the soluble fractions (the supernatants) were retained for further analyses. The expression was determined by Coomassie-stained gels and western blots using an anti-his primary antibody.

The TOPO vectors were also translated in the ICE system (insect cell-free extract from Promega) following the protocol described in section 2.2.9. The T7 insect cell-free extract (Promega) is made from *Spodoptera frugiperda* Sf21 cell line and is a pre-mixed all-inclusive extract which contains all the required components for protein expression in a single mix. The synthesis is initiated with 2 µg of template DNA, and requires a T7 promoter sequence. The pre-mixed extract and the template DNA were incubated for 4 h at 30 ^{o}C without radioactive amino acids. The western blot analysis was carried out using V5 primary antibody.

3.3 Results

3.3.1 Visual method exploiting cellular stress to study re-initiation of translation

3.3.1.1 Transient transfection

The first set of experiments involved transiently transfecting BHK21 cells with the constructs pSTU1 and pSTU2 conferring resistance to geneticin. The rationale was to culture BHK21 cells with geneticin to select for transfected cells and obtain a homogenous population, before applying a cellular stress.

The expected results were that under normal growth conditions, where no stress was applied, cherryFP expression would be localised to the cytoplasm only for cells transfected with pSTU1. For the control plasmid pSTU2, cherryFP however should be diffused throughout the cell. Under conditions of stress, the only difference to the above was that cherryFP should also diffuse to nuclei for cells transfected with pSTU1.

To estimate the concentration of geneticin necessary to select for transformants, 60×10^4 BHK21 cells were seeded in six-well plates and transfected with pSTU1 and incubated for 24 h. In parallel, control BHK21 (not transfected) were cultured under the same conditions. Geneticin was introduced at increasing concentrations (0 to 1200 μg/ml) to both cultures and the percentage of alive cells was assessed under light microscopy 24, 48 and 72 h after addition of geneticin (table 3.1). At suitable selective concentration of antibiotic, there should be few to no alive cells for non-transfected cells versus a higher proportion of alive cells for transfected BHK21.

Table 3.1: Visual determination of the suitable concentration of geneticin for selection of BHK21 transiently transfected with pSTU1

Geneticin was introduced at different working concentrations to six-well plates and the result visually assessed under light microscopy at 24, 48 and 72 h. the results represent an estimation of the average alive cells for non-transfected versus transfected cells.

Concentration geneticin (G418) (µg/ml)	Transfected with pSTU1 (neomycin Resistant)			Normal BHK21		
time	72h	48h	24h	72h	48h	24h
0		All alive			All alive	
150		All alive			All alive	
300		All alive			~ 60% alive	
600		All alive		None alive	~20% alive	
900		~30%alive		None alive	~10% alive	~20% alive
1200	<10% alive	~20% alive	~20% alive		None alive	

Based on the results presented in table 3.1, the selection conditions established were 600 µg/ml of geneticin for three days and 900 µg/ml afterwards.

After defining the selective concentrations necessary, the next step was to determine a suitable stress required to produce a decreasing concentration of eEF2. Transient transfection and geneticin selection were discarded because of two major limitations. The selection process using geneticin required three weeks but BHK21 cells transiently transfected lost the expression of the fluorescent marker within several days post-transfection. The second problem encountered was that fluorescent microscopy results (figure 3.5) for pSTU1 cells, to which no stress was applied, were not as expected. In figure 3.5, a population of BHK21 transfected with pSTU1 shows cherryFP expression throughout the cytoplasm and nuclei, while for others fluorescence is limited to the cytoplasm. To demonstrate that the plasmids used to transfect BHK21 were correct, a test translation in rabbit reticulocyte lysates, analysed by SDS-PAGE (figure 3.6) was carried out. The figure 3.6 shows the translation profile expected for these two vectors and demonstrates that the 2A sequences for pSTU1 and pSTU2 were functional.

Because transient transfection may act as a cellular stress and limit the expression of the transgene in time, it was decided that a lentiviral system might be more adequate. Modified lentiviruses for the purpose of cloning retain the sequences necessary for insertion of the transgene in the host genome, thus allowing long term expression. The lentivirus A5 plasmid (provided by Prof Randall laboratory) contains a suitable multiple cloning site for the inserts of pSTU1 and pSTU2 and was therefore selected.

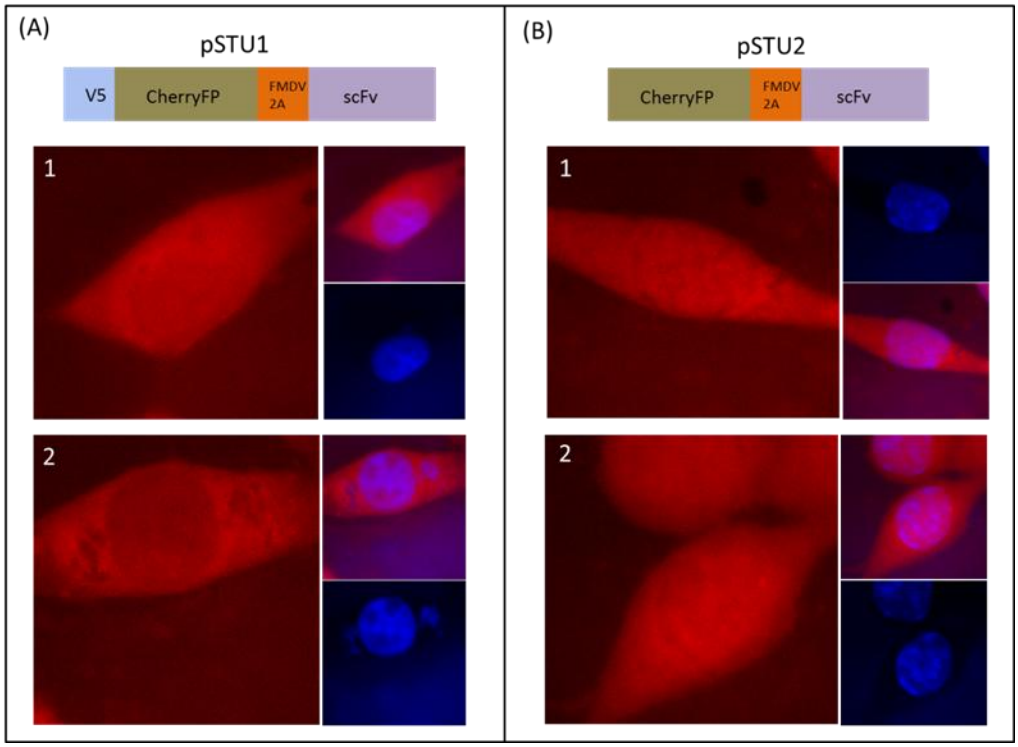


Figure 3.5: Results showing localisation of cherryFP expression by fluorescent microscopy of BHK21 cells transfected with plasmids pSTU1 (A) and pSTU2 (B)

As expected, cherryFP expression was detected throughout nucleus and cytoplasm for pSTU2 (no V5 tag, (B)). Unexpectedly, a population of cells ((A), cell 1) showed cherryFP expression throughout cytoplasm and nucleus for pSTU1 transformants. The panels consist of the red fluorescence image contrasted to the nucleus only picture stained with DAPI (bottom right for each cell) and the merged version (top right for each cell).

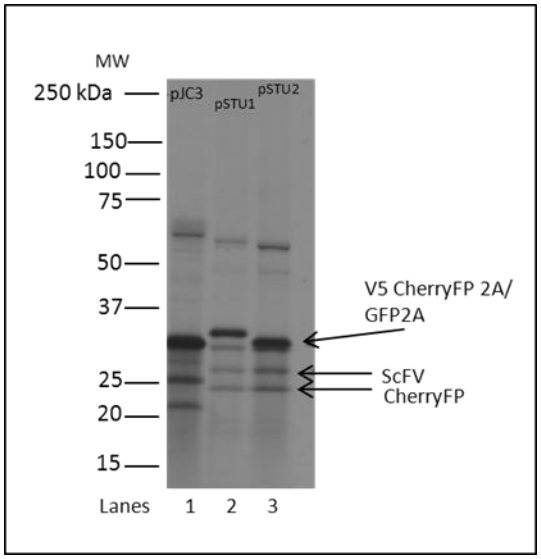


Figure 3.6: Cell-free control translation of pSTU1 and pSTU2 plasmids

Plasmids were translated in rabbit reticulocytes with [^{35}S]-Methionine. This was carried out as a control to demonstrate that the plasmids used to generate figure 3.7 were functional. The positive control lane 1 was pJC3.

3.3.1.2 Stable transfection

The coding sequences of pSTU1 and pSTU2 were inserted into the A5 plasmid for lentiviral transfection (figure 3.2). The A5 plasmid confers resistance to puromycin, it is not suitable to BHK21 cells which are puromycin resistant, but adequate for mammalian cells such as HeLa cells.

The lentiviral particles were generated in 293T cells and used to infect HeLa cells. The resulting cell lines are referred to as A5V5 HeLa for cells expressing the pSTU1 coding sequence and A5ChFP HeLa for cells expressing the control pSTU2 coding sequence.

A titration of puromycin using non-transfected HeLa cells was conducted, it was found that after addition of 1.5 µg/ml of puromycin ~ 20 % of cells survived but with a concentration of 2 µg/ml no cells survived. After lentivirus infection, cells were selected for two weeks with 2 µg/ml of puromycin before any cellular stress was applied.

The defining parameter for this assay was that the cellular stress would decrease the level of active eEF2 within the cells. Preliminary stresses were applied to A5V5 HeLa cell line to determine a stress that would correspond to a drop in active eEF2. Western blot analyses were performed with anti-eEF2 antibody and the phosphorylated anti-eEF2 antibody. After introduction of stresses, the cells were lysed with CelLytic buffer (Sigma) and their total protein content quantified using the BCA assay (Pierce) against the calibration standards (bovine serum albumin) in micro-plates. 20 µg of total protein was loaded in each well of the SDS-PAGE gel for western blot analysis.

The four stresses introduced were: addition of ethanol (0 to 5%) to the culture for 12 h, nutrient deprivation (0 to 10 % FCS) over 24 h, temperature rise to 42 ^{o}C over 6 h and viral infection (EMCV) over 8 h. A5V5 HeLa cells were seeded in six-well plates and allowed to adhere for 24 h before the media or the culture conditions were altered. In parallel, for nutrient deprivation and addition of ethanol, a cell viability test (MTT) carried out in 96-well plates (using 20×10^4 cells per well) was also performed to assess the relative survival of cells compared to control cells. The tetrazolium in the MTT reagent is intracellularly reduced to the purple formazan by oxidoreductase enzymes of metabolically active mitochondria. The more formazan is formed the more viable the cells are, and inversely, a drop in the conversion of formazan from the MTT reagent indicates loss of viability.This assay was not suitable for 42 ^{o}C temperature rise, because of the incubation step at 37 ^{o}C in the MTT protocol. The cell viability test was also not performed for EMCV infection of HeLa cells because the 96-well plate result would not necessarily correlate to the experiment in the six-well plate. The figure 3.7 shows the western blot analyses and relevant cell viability assays for A5V5 HeLa cells subjected to heat-shock at 42 ^{o}C (A), viral infection (B), serum deprivation (C) and ethanol stress (D).

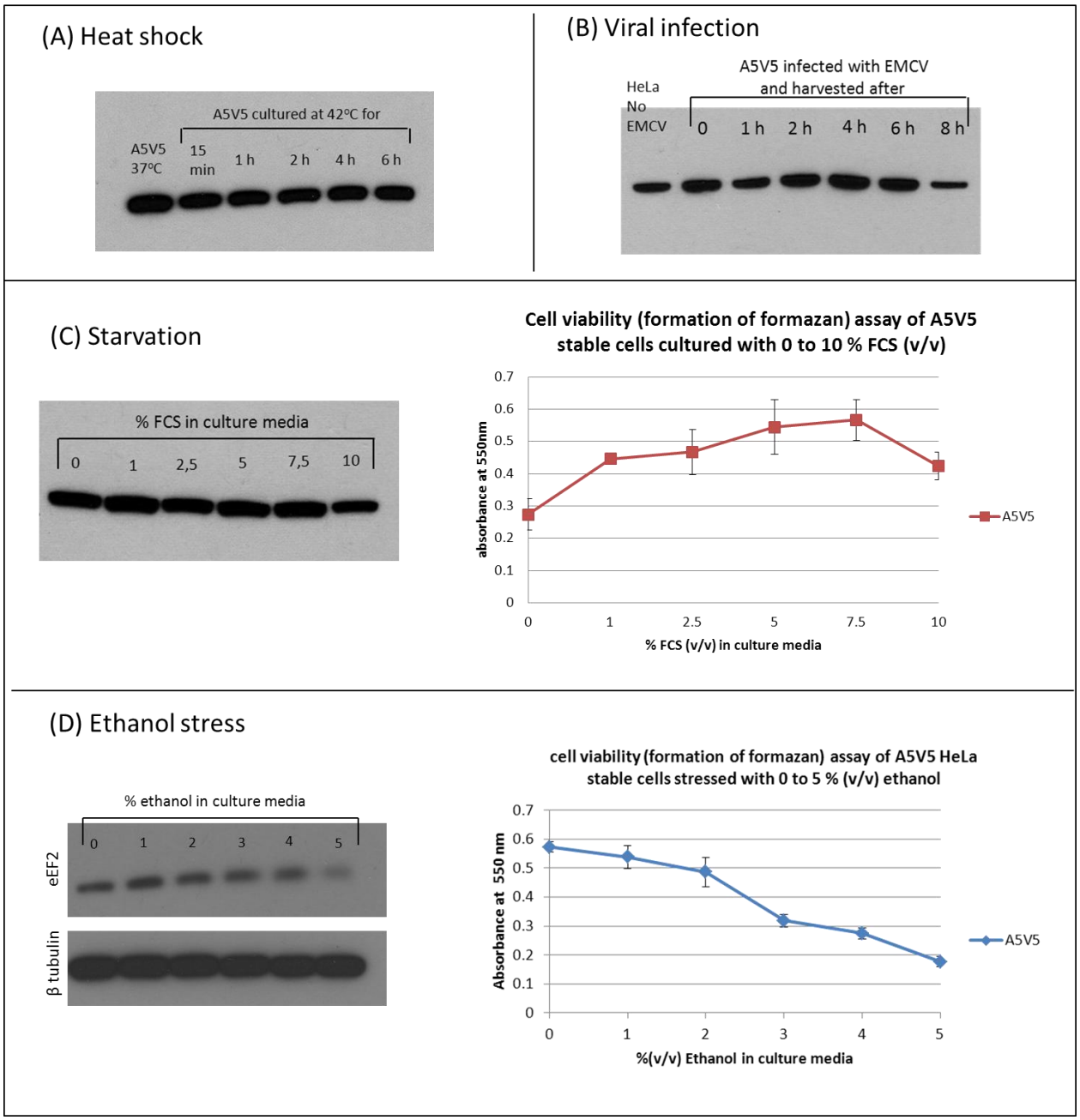


Figure 3.7: Effect of different stresses on the viability and cellular concentration of eEF2 in A5V5 HeLa

A5V5 HeLa cell lysates were examined by western blot analysis probing for eEF2 (panels (A), (B), (C) and (D)) and β-tubulin (panel D- lower gel). The legend above the gels corresponds to the experimental condition tested. (A) Cells were grown at 42 °C and lysed at various times and compared to cells cultured at 37 °C. (B) Cells were infected with EMCV over 8 h and analysed at different time points. (C) Cells were cultured with decreasing concentration of serum. (D) Cells were cultured with increasing levels of ethanol.

MTT cell viability assays measuring the formation of formazan by live cells, at the wavelength of 550 nm were performed for starvation (panel C) and ethanol stress (panel D) experiments. The same experimental conditions as for western blot were respected and are indicated on the x-axis of the graphs. The data are presented as the average value +/- the standard deviation for five replicates.

Western blot analyses using the phosphorylated form of eEF2 were not successful. The antibody is designed for immunofluorescence and most likely not suitable for western blot analysis where the protein is denatured.

In figure 3.7 (A), the western blot analysis with anti-eEF2 antibody shows that the concentration of eEF2 remains constant when the HeLa cells were incubated at 42 ^{o}C between 15 min to 6 h. In (B), during viral infection, the concentration of eEF2 does not decline over time and the variations observed reflect the individual cellular deviations in the susceptibility to viral infection. In (C), A5V5 HeLa cells did not appear to be stressed when starved; in the graph, there is no decline in formation of formazan after 24 h without nutrients. This result correlates with the western blot showing no decline in the level of eEF2 when cells were cultivated with 0 to 10 % (v/v) FCS.

There is however a decrease in the amount of eEF2 when the cells were stressed with increasing levels of ethanol (figure 3.7 (D)). The stress status is confirmed by the cell viability assay, showing a regular decline in formation of formazan with increasing levels of alcohol (from an initial OD_{550nm} of 0.6 to 0.2 at 5 %(v/v) ethanol). The nitrocellulose membrane was stripped with the Restore western blot stripping buffer (Thermo scientific) and re-probed for β-tubulin which is constitutively expressed. This confirmed that the decline in eEF2 observed in the western blot is not due to a difference in the amount of samples loaded onto the wells. The experimental condition selected for this experiment was therefore an increasing level of ethanol applied for 12 h before mounting cells for fluorescent microscopy.

A5V5 and A5ChFP cell lines were seeded in six-well plates with a sterile coverslip and left to adhere for 24 h. The medium was aspirated and the cells were incubated for 12 h with different concentrations of ethanol: 0, 2, 4 and 5 %(v/v). The cells were then prepared and mounted for electron microscopy. The result for this experiment is provided in figure 3.8, where the A5V5 and A5ChFP expression is shown side by side.

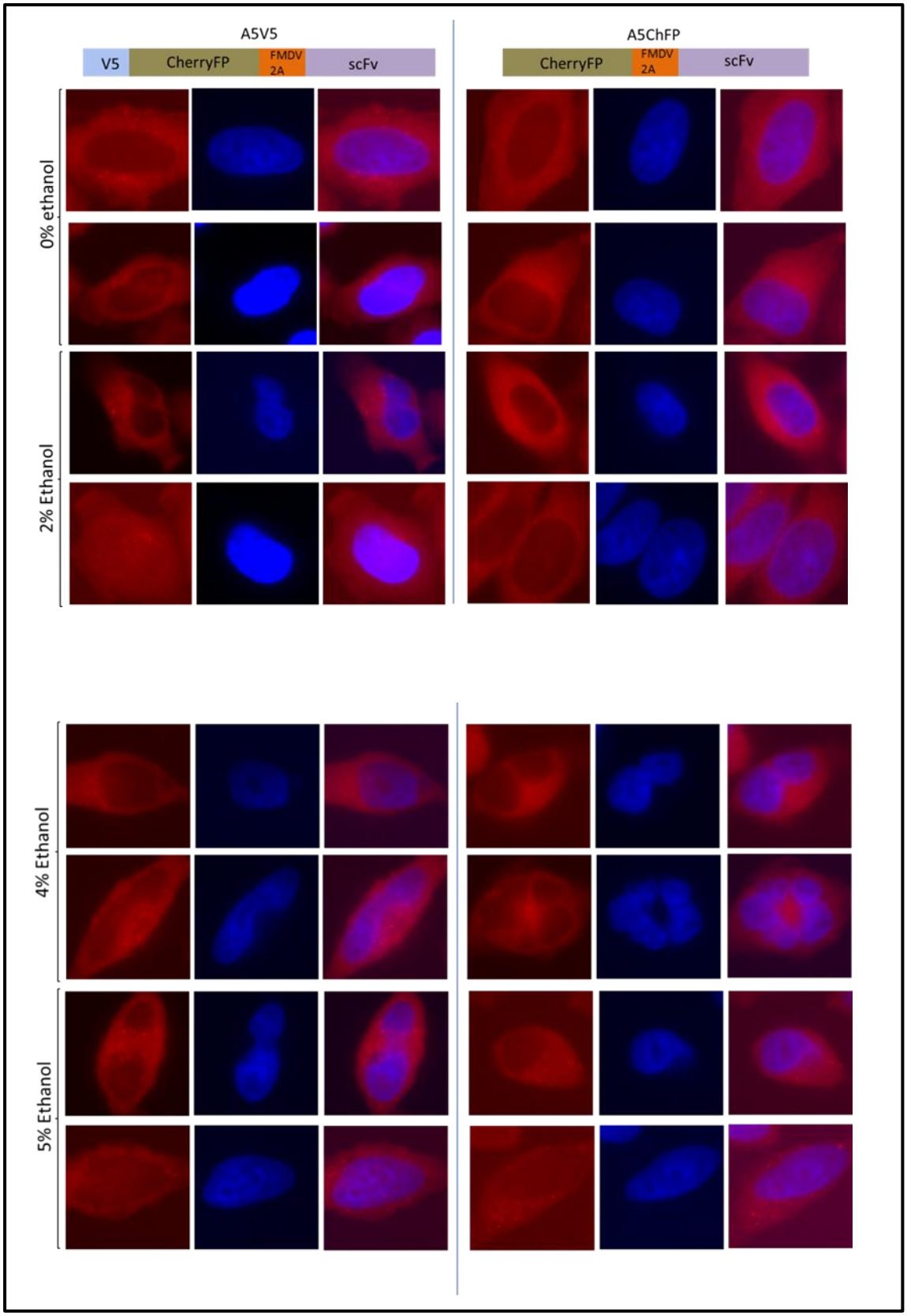


Figure 3.8: CherryFP-2A with (left) or without V5 (right) did not enter the nucleus

HeLa stable cell lines expressing A5V5 or A5CherryFP were subjected to ethanol stress and viewed by electron microscopy. Results are provided for two cells for each condition tested: 0, 2, 4 and 5% ethanol. The panel shows red fluorescence, nucleus stained with DAPI (blue) and merged image.

In figure 3.8, for each condition tested, without ethanol or with increasing level of ethanol, cherryFP was localised only in the cytoplasm of cells for both A5V5 and A5ChFP cell lines. The expected result was that cherryFP 2A should be found in the cell cytoplasm and nuclei for A5ChFP. This is however, not the case. With increasing levels of stress and therefore less available eEF2, more cherryFP should be diffusing to the nuclei for the A5V5 cell line. This is also not the observed result. To demonstrate that the result observed is not due to a 'faulty' sequence, the sequencing results for the A5V5 and A5ChFP vectors are provided in figure 3.10. These vectors were used for the viral packaging step in 293T cells, figure 3.10 shows integrity of the V5 epitope for A5V5 and its absence for the A5ChFP. The 2A sequences were not mutated for both of these vectors.

To provide a reference for cherryFP expression, the plasmid pJC3 which has a GFP-TaV 2A-cherryFP insert, was transiently expressed in HeLa cells, prepared and mounted for microscopy. In this construct, cherryFP is expressed as a discrete protein and not fused to the 2A peptide. The image in figure 3.9 shows that cherryFP is diffused throughout the HeLa cell.

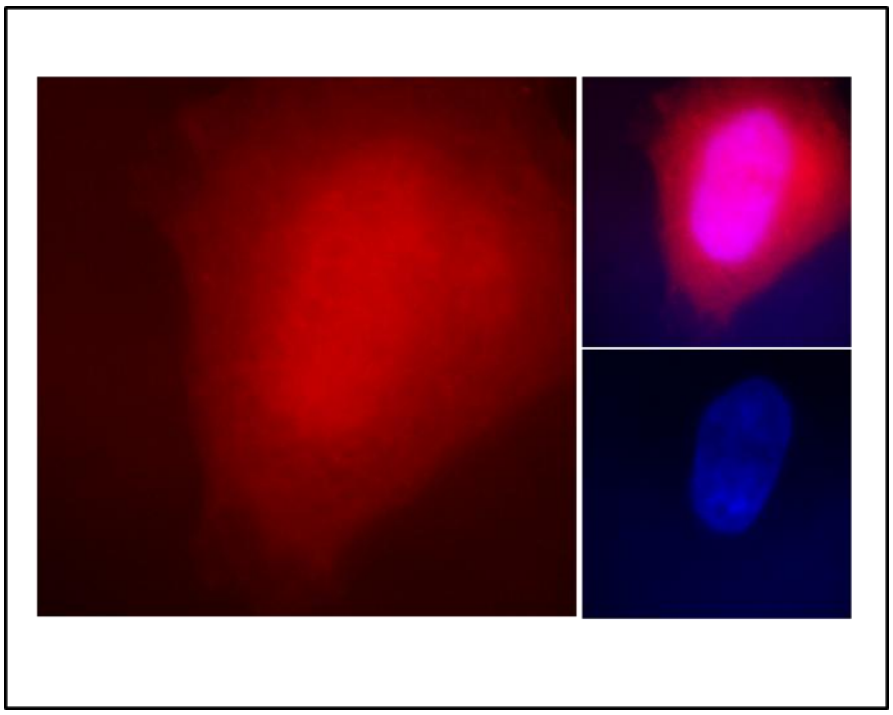


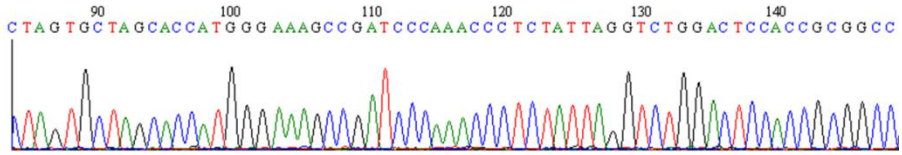
Figure 3.9: CherryFP expression in transiently transfected HeLa cell

The panel consists of red fluorescence image (left) and nucleus stained with DAPI in blue (bottom right) and merged figure (upper right). In pJC3, the cherryFP-coding sequence is downstream of 2A and therefore translation of this plasmid should give rise to discrete cherryFP.

(A) lentivirus A5V5 ChFP

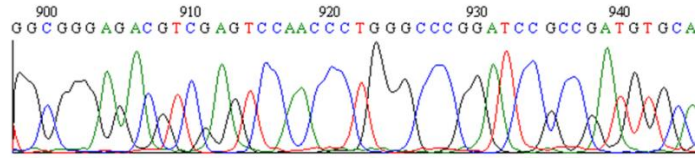
V5 epitope beginning of CherryFP

```
1       ATGGGAAAGCCGATCCCAAACCCTCTATTAGGTCTGGACTCCACCGCGGC…
        ||||||||||||||||||||||||||||||||||||||||||||||||||
98      ATGGGAAAGCCGATCCCAAACCCTCTATTAGGTCTGGACTCCACCGCGGC
```



2A sequence for A5V5 ChFP clone

```
1       GACGTCGAGTCCAACCCTGGGCCCGGATCCGCCGA
        |||||||||||||||||||||||||||||||||||
905     GACGTCGAGTCCAACCCTGGGCCCGGATCCGCCGA
```

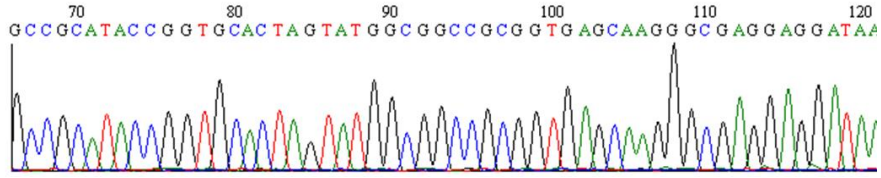


(B) lentivirus A5ChFP

Aligned against A5V5 sequence

```
27      ATTAGGTCTGGACTCCACCGCGGCCGCGGTGAGCAAGGGCGAGGAGGATAACATGGCCAT
        ||    |           |  ||||||||||||||||||||||||||||||||||||||||
71      ATACCGGTGCACTAGTATGGCGGCCGCGGTGAGCAAGGGCGAGGAGGATAACATGGCCAT
```

Beginning of ChFP



2A sequence for A5ChFP

```
1       GACGTCGAGTCCAACCCTGGGCCCGGATCCGCCGA
        |||||||||||||||||||||||||||||||||||
852     GACGTCGAGTCCAACCCTGGGCCCGGATCCGCCGA
```

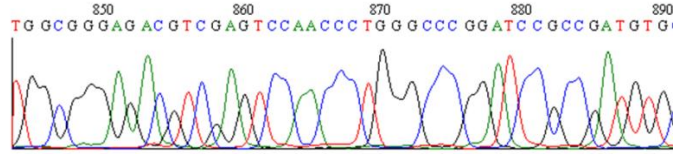


Figure 3.10: Alignment of lentivirus inserts (bottom line) against expected sequences (top line)

CherryFP beginning is edited in blue text, 2A in pink text. The chromatograph for the area of interest is presented underneath the sequence alignments. The lack of result in the previous figure was not due to a lack of V5 in A5V5 construct or a mutation in the 2A sequence.

3.3.2 Bacterial expression of translation-regulating factors

The translation-regulating factors eRF1, eRF1 RGQ, eRF3, eEF2 and eEF2K in TOPO plasmids were expressed in six bacterial strains, at three temperatures and with two concentrations of IPTG. The bacterial expression was not successful for all conditions tested. An example of a western blot is shown in figure 3.12 (A) for expression in BL21 where 5 µl of the supernatant fractions were probed with anti-His antibody. The TOPO plasmids were compatible with the T7 insect cell-free system (Promega). A test expression in 50 µl of extract was carried out for all translation factors, aiming at testing the viability of the plasmid sequences. 5 µl of the translation reactions, figure 3.12 (B) were analysed by western blot using ScFv antibody which probes for the V5 tag.

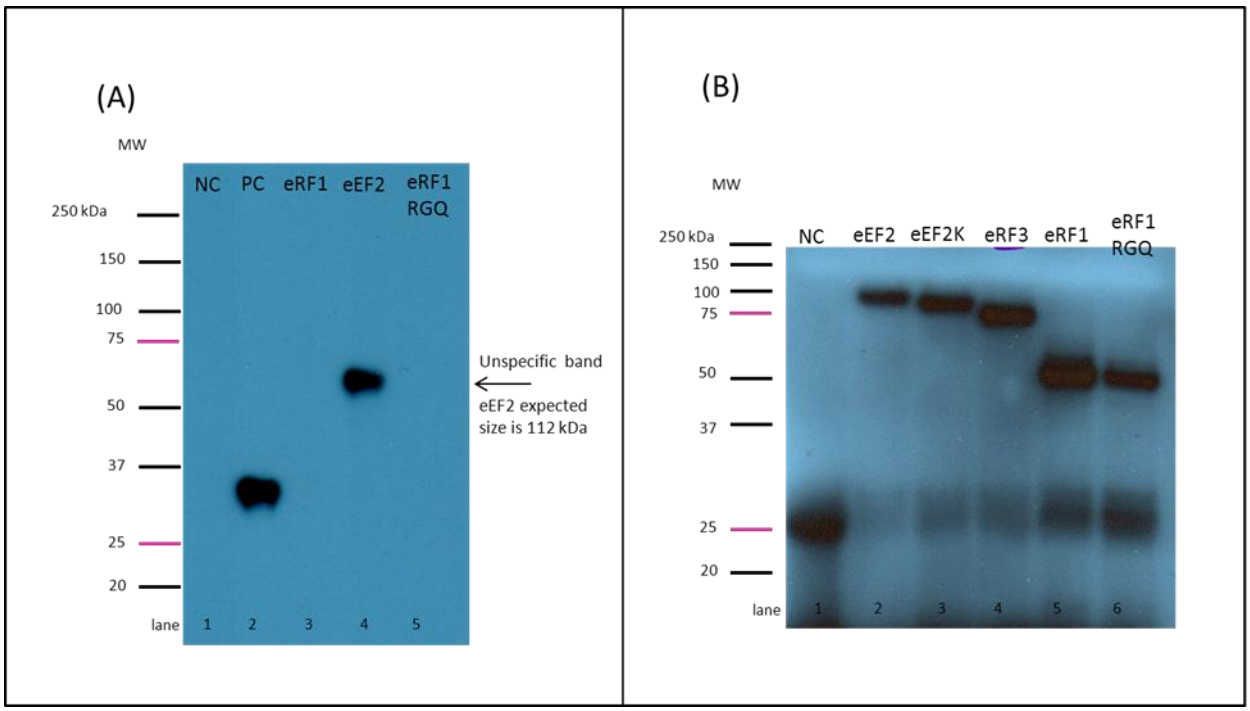


Figure 3.11: Western blot analyses of expression of translation-regulating factors in BL21 (A) and in Insect cell-free system (B)

In (A), BL21 cels were cultivated at 37 ^{o}C and induced with 1mM IPTG, western blotting was performed with anti-His antibody, the positive control (PC) pEHISGFPTEV (lane2) expresses a tagged GFP at the expected size (30kDa), eRF1 and mutant (lanes 3 and 5) did not show expression. eEF2 expression (lane 4) resulted in an unspecific binding to the anti-His antibody. Control test of the plasmids performed with T7 insect cell-free system (B) shows successful expression of all the proteins. In (B) the western was carried out with anti-V5 antibody.

The above result showed that the protocol (developed in section 2.2.8) is functional and that the lack of expression is not a loss of the reading frame, since the V5 tag would not have been recognised by anti-V5 antibody. The T7 promoter was therefore also intact and functional. The induction phase of the protocol may have failed for the TOPO plasmids or the His-tag is not exposed to the antibody, the

latter is however unlikely since the denaturating step before SDS-PAGE resolution should overcome this problem. The IPTG used was also functional since the positive control in (A) did show expression. The proteins may also have been insoluble. It was decided that for future experiments re-cloning of the genes of interest in another vector could be the remedy to the expression problems. Alternatively, adopting an insect-cell system may prove more successful (figure 3.12 (B)).

3.4 Discussion

Results for transient transfection of cells with the constructs pSTU1 and pSTU2 in BHK21 were not reproducible, while stable integration of the inserts in HeLa cells failed to produce conclusive results.

Lentiviruses integrate heterologous DNA into the host genome (Buchschacher and Wrong-Staal, 2000). The transcript is capped and the regulation of its translation is subjected to cellular status. Translation shut off for capped mRNA during conditions of stress (Holick and Sonenberg, 2005) would explain the little difference observed in cherryFP localization between the controls and samples for increasing stress levels. Translation of the insert depends largely upon the site of integration of the heterologous DNA in the host genome and the promoter driving the expression (Buchschacher and Wrong-Staal, 2000). It is unlikely for the site of integration to be located after a promoter up-regulated under conditions of stress. Transient transfection has several drawbacks including that heterologous DNA is not integrated in the genome and not stably expressed (Geisse and Voedish, 2009). Therefore the reporter protein is not expressed long enough for selection and stress to be applied.

Surprisingly when transfected with pSTU1 (figure 3.7), for some cells cherryFP was localised throughout the cytoplasm and nucleus, for others cherryFP seemed limited to the cytoplasm. A recent short communication by David and colleagues (2013) brought doubts to the feasibility of the assay developed here. The authors showed that in HeLa cells ribosome translation also occurs in the nucleus. Coupling puromycin, the antibiotic emetine and a fluorescent anti-puromycin antibody, they showed nuclear localisation of actively translating ribosomes. In another experiment still involving HeLa cells, Iborra and colleagues (2001) permeabilised the cells in a physiological buffer and allowed translation to proceed for 15 residues in presence of BODIPY-lysine-tRNA^{lys}. While most of the products were detected in the cytoplasm, 14 % of the signal originated from the nuclei and was reduced by cycloheximide (a ribosome inhibitor). One possibility is that the result presented here is not only representative of the diffusing ability of cherryFP but also of its localised expression. The result may also be related to the brightness of the reporter protein. In addition, the assay does not at

the moment provide a true representation of the cellular status at a given time, cherryFP is photostable compared to other fluorescent proteins and especially developed for long term imaging (Shaner at al., 2005). Since the marker accumulates, the visualisations do not reflect *de novo* synthesis of translation products, which for this assay is a particularly sensitive point since the aim is to differentiate at different stress levels, the amount of proteins produced after 2A. The destabilisation domain (DD domain)/ shield ligand supplied by Clontech might be a solution, as it allows to reversibly destabilise in a dose dependent manner a protein tagged with the DD domain.

These results highlight the experimental considerations necessary to achieve the challenging task of proving directly the role release factors play in rescuing 2A-induced ribosome stalling. For the bacterial expression of translation-regulating factors; the straightforward conclusion is that the coding sequences should be cloned into another vector. Purifying eEF2 from cellular cultures may be a more successful approach since there is no commercial test readily available to assert the functionality of the recombinant eEF2. In the literature for the purpose of reconstituting a translation assay, eEF2 was not over-expressed but purified while eRF1 has routinely been expressed from BL21 cultures (Alkalaeva et al., 2006 and Pisareva et al., 2011). The experimental rationale here was to modify a rabbit reticulocyte by over-riding the translating factors present in the system with additional mutant eRF1 or other translation-regulating factors. Since the reticulocyte system has been optimised for efficient translation, modulating its content requires either depletion of the factors or addition of drastic amounts of additional proteins. Pisarev and colleagues (2011), in an effort to study release factors activity, have developed a eukaryotic system fully reconstituted from a mixture of purified and overexpressed proteins and ribosome subunits. The system has been used extensively to introduce relevant mutations to eRF1 and eRF3 and adopting such a system will provide clear and direct evidence for the involvement of eRF1/3 in rescuing 2A- paused ribosome. Moreover, it will enable the introduction of multiple mutations so as to test what allows the factor this previously uncharacterised ability.

Chapter 4 Sequence requirements for 2A activity

'What saves a man is to take a step. Then another step. It is always the same step, but you have to take it.'
- Antoine de Saint-Exupery

4.1 Introduction

This chapter summarises the successive mutagenesis performed on hybrids and truncated constructs made for selected model 2A sequences in an effort to identify the important residues leading to ribosome stalling. Another aim was to gather evidence to answer the questions: Is the 2A-induced ribosomal stalling likely the result of an interaction with the ribosome exit tunnel? Or is it the result of the nascent 2A adopting a conformation unfavourable for peptide bond formation?

Viral 2A sequences tested *in vitro* and shown to be active, display an intriguing non-conserved upstream context preceding the general consensus motif [D(V/I)ExNPGP]. The lack of conservation would imply the conclusion that the upstream context to the consensus motif is auxiliary. However, previous studies with FMDV 2A showed that the upstream context leading to the signature motif is essential. The IFV sequence [TRAEIEDELIRAGIESNPG/P] is 63 % efficient, but a single point mutation [TRAEIEDELIRADIESNPG/P] renders it inactive (Donnelly *et al*., 2001*b*). For FMDV 2A, an extended sequence leads to higher levels of stalling. A FMDV 2A with seventeen amino acids [NFDLLKLAGDVESNPG/P] is 75 % efficient, the efficiency decreases with shorter sequences. FMDV2A efficiency drops to 65 % when only thirteen amino acids [LKLAGDVESNPG/P] of the sequence were included in the insert (Ryan and Drew, 1994). Adding thirty-nine amino acids of the 1D region of FMDV at the N-terminus of FMDV 2A, increased the efficiency to nearly 100 % (Donnelly et al., 1997). In the context of FMDV, these results led to the conclusion that efficient 2A stalling required a minimum of thirty amino acids. TaV2A which is naturally only twenty amino acids long (Donnelly *et al*., 2001*b*) and shown to be highly efficient at stalling the eukaryotic ribosome is not altered by mutations and alteration at the C-terminus of the preceding protein (Minskaia, publication in preparation). Systematic mutagenesis of 2A sequences have not previously been carried out.

The reactive centre (PTC) of the ribosome, where peptide bond formation occurs, is situated in a cavity deep in the large 60S ribosomal subunit. As residues are added, the nascent peptide enters and progresses through an exit tunnel, which originates from the PTC (Yonath et al., 1987). The exit

tunnel is 100 Å long in eukaryotes and varies in width, hosting between thirty to forty amino acids of an elongating peptide. The first 20 Å from the PTC are considered to be made of rRNA only. The tunnel is then constricted by the protruding loops of two ribosomal proteins, L4 and L17 in eukaryotes (and L4 and L22 in prokaryotes). Protein contribution to the wall is thought to be limited to these 2 proteins over a length of 10 Å and to a stretch of residues nearer the exit port from protein L23 or L25. The tunnel wall is assembled from domains I, II, III, IV, V of 23S rRNA in prokaryotes or 28S rRNA in eukaryotes. Near the PTC the tunnel is 15 Å wide, in the constriction 10 Å wide and further in the lower tunnel it enlarges to 25 Å. The L22 protein is involved in the constriction and also at the exit port where together with L39 (eukaryotes) or L23 (prokaryotes) frame the exit port (Nissen et al., 2000 and Voss et al., 2006).

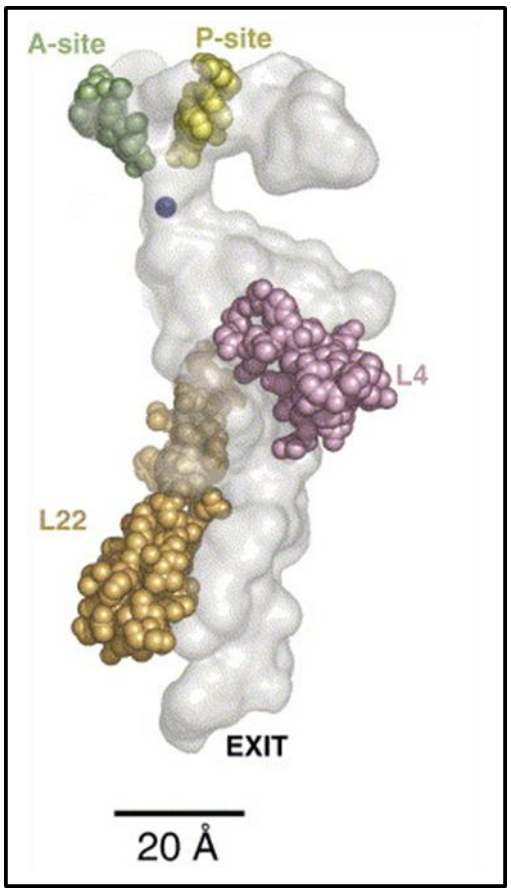


Figure 4.1: The ribosome exit tunnel

The prokaryotic ribosome exit tunnel is used as a model for this structure in all domains of life. Diagram show the A and P site tRNAs in green and khaki, the L4 and L22 ribosomal protein are represented in violet and gold. The blue dot corresponds to the entrance to the tunnel (image reproduced from Voss et al., 2006).

Theosa asigna virus 2A sequence (TaV2A) was selected for the purpose of this study. TaV virus has two ORFs, 2A was originally identified preceding the capsid proteins on the second ORF. Luke and co-workers (2008) cloned and measured the activity of a twenty aa TaV 2A sequence. TaV 2A is considered an optimal 2A since this minimal sequence is sufficient to induce nearly 100 % stalling. In

order to answer the question what made TaV 2A so efficient, in the context of this thesis, several mutants were created. Other 2A sequences were also employed to provide supplementary data. From the picornavirus family: FMDV 2A and Duck Hepatitis virus 2A (DHV 2A) were selected. FMDV 2A was included in this experiment since it has been the object of many studies (as detailed in the general introduction section) related to the mechanism of action of 2A. DHV 2A was selected because it has a deviating consensus motif [EGVENPGP]. Other 2A like elements shown previously to be potent ribosomal stalling sequences (Luket et al., 2008) were also selected based on their deviating consensus motifs. The adult diarrhoea virus C (ADRV 2A) consensus sequence is [CIESNPGP]. The insect iflavirus 2A (IFV 2A) sequence resembled DHV 2A [GIESNPGP].

There were several questions driving the mutational studies:

What part does the consensus motif play in ribosome stalling? To this effect, several hybrid sequences were created interchanging the upstream contexts of 2As. The TaV 2A consensus motif was also altered to match the ten different other consensus sequences listed to date and finally the residues side chains in TaV 2A consensus motif were successively substituted to alanine.

Since there is much diversity in upstream contexts, is there any residue or type of residue at a particular position that is required for efficient stalling? To answer this question, successive alanine substitutions were carried out in the upstream context of three model 2As: TaV, FMDV and DHV.

Is the activity of 2A purely structural as previously proposed (Doronina et al., 2008) or is it the result of an interaction with the ribosome exit tunnel or both? An attempt was made to answer this question by mutations of TaV residues to glycine. TaV arginine residues were also substituted to isoleucine (hydrophobic), lysine (increasing ability to react with RNA), glutamine (no charge) and glutamic acid (negative charge). IFV, DHV and ADRV 2As were truncated. Double proline mutations and non-conservative mutation of leucine residues to asparagine (similar MW, but hydrophilic) were also introduced in the FMDV, DHV and TaV 2A sequences.

4.2 Materials and methods

4.2.1 General procedures

The DHV 2A sequence was identified using the basic local alignment algorithm at NCBI online server (http://www.blast.ncbi.nlm.nih.gov). Viral 2A sequences were sorted by consensus sequences and aligned with ClustalX. The weblogo online server (http://www.weblogo.berkeley.edu) was used for multiple alignments of 2A sequences. Secondary structure was predicted using the Jpred3 online server (http://www.compbio.dundee.ac.uk/jpred).

Plasmid pSTA1 coding for GFP-2A-GUS in one single ORF was used and modified. pSTA1 vectors containing 2A sequences for TaV 2A, IFV 2A, FMDV 2A or ADRV 2A were from the laboratory collection of clones, DHV was created as part of this body of work. **DHV2A** (thirty amino acids long) was inserted at the 5' end of GUS coding sequence with two successive PCRs. Primers for PCR1: forward 5'-TCTAGAGCATTTGAACTCAATTTGGAAATTGAATCTGACCAAATT AGAAACAAGAAAGATCTCACTACTGAAGGAGTCGAGCCAAACCCCGGGCCCCACC-3' and reverse 5'-ACGGGCCTCGAGCGGCCGCACTA GATTGTTT-3'. Primers for PCR2: forward 5'-TAATACGACTCACTATAGGG-3' and reverse 5'_GCTTTGGTGGGGCCCGGGGTTTGGCTCGACTCCTTCAGTAGTGAGATC TTTCTTGTTTCTAAT-3'.
For directional cloning, an Xba1 site was engineered at the 5' end of all the 2As. The nucleotide sequence corresponding to Gly-Pro at the C-terminal motif of the 2A forms a restriction site for Apa1 (GGGCCC). PCR products were first cloned in pGEM-T *via* A overhangs and inserted in pSTA1 preferably digested with Xba1-Xho1 (2A-GUS) or Xba1- Apa1 (insertion of 2A only). The experimental procedures summarised underneath are fully detailed in the section 2.2.1-7.

Primers were designed to introduce a mutation of a specific amino acid (introduced mutation is highlighted in blue in the forward primers sequences in this section). PCR were performed using the KOD Hot start enzyme from Novagen, following manufacturer's instructions. The protocol utilised is detailed in section 2.2.1. The integrity of the mutants was assessed by big Dye terminator sequencing (ABI) outsourced from Dundee sequencing services (sequences in appendix 2). *In vitro* translation was performed in 12.5 µl Rabbit reticulocyte extracts supplemented with [^{35}S]-methionine and carried out for ninety minutes. The translation products were resolved on 4-20 % tris-bis SDS-PAGE denaturing gels (Expedeon). The gels were Coomassie stained and [^{35}S]-labelled proteins were visualised using photographic film and their size determined against the dual colour marker precision standard from Biorad.

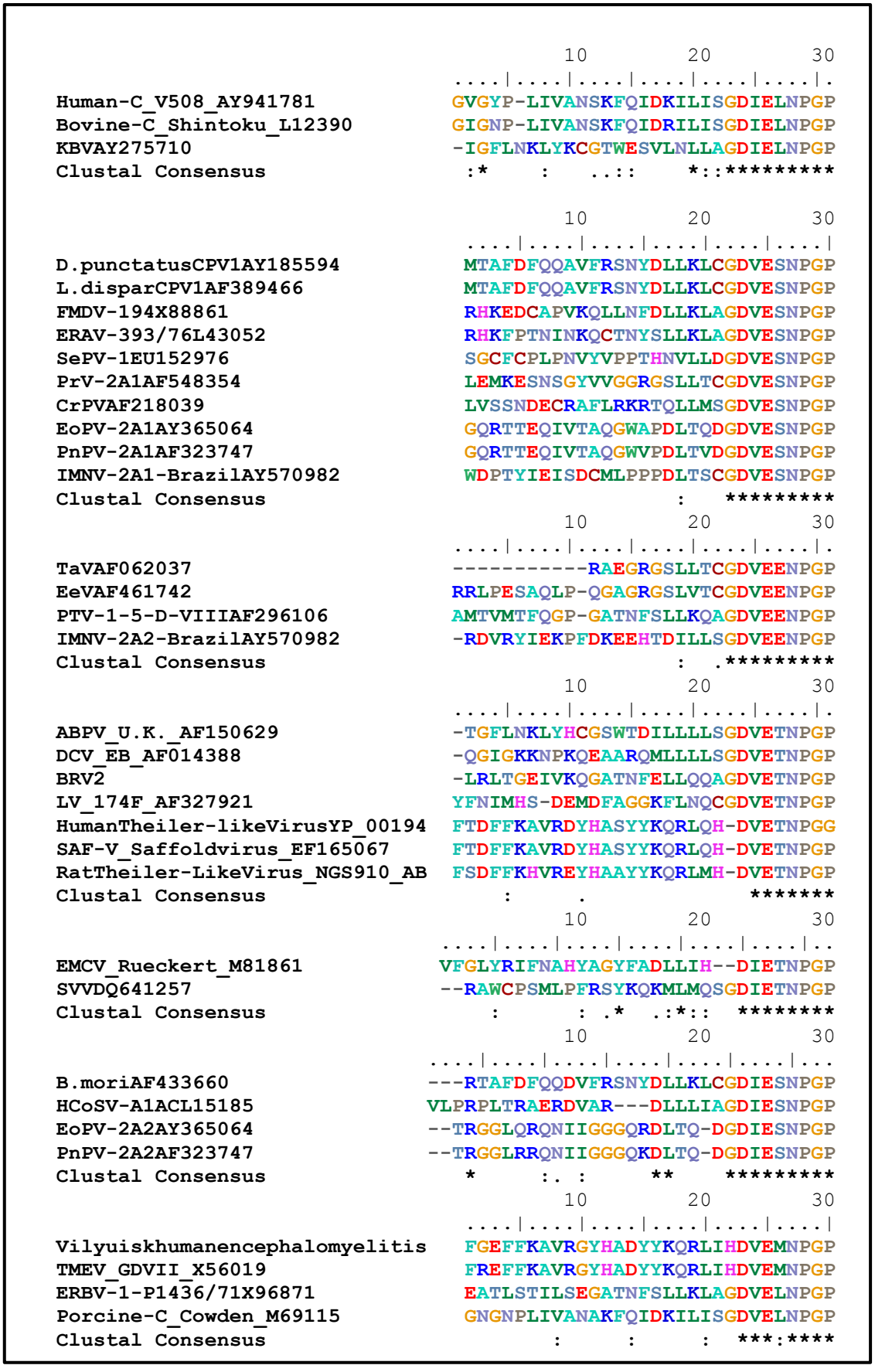


Figure 4.2: Representative virus 2A sequences aligned by their consensus motifs

Alignment was performed with ClustalX

4.2.2 Primers for mutations and alterations of 2A consensus motifs

4.2.2.1 TaV 2A modified with other viral consensus motifs

Alignment of the viral 2A were performed with ClustalX, and resulted in seven groups of consensus sequences. The TaV 2A consensus motif [DVEENPGP] was swapped for [DVETNPGP], [DVEMNPGP], [DVELNPGP], [DVESNPP], [DIELNPGP], [DIETNPGP] and [DIESNPGP] in order to test for their efficiencies in ribosomal stalling. The PCRs were performed with T7 (forward) and reverse primers underneath with modified sequences. The BamH1-Apa1 segment was re-inserted in pSTA1.

DVETNPGP, 5'- GGTGGTGGTGGGGCCCGGGATTTGTCTCCACGTCC- 3'
DVEMNPGP, 5'- GGTGGTGGTGGGGCCCGGGATTCATCTCCACGTCC- 3'
DVELNPGP, 5'- GGTGGTGGTGGGGCCCGGGATTAAGCTCCACGTCC- 3'
DVESNPGP, 5'- GGTGGTGGTGGGGCCCGGGATTAGACTCCACGTCC- 3'
DIELNPGP, 5'- GGTGGTGGTGGGGCCCGGGATTTAGCTCAATGTCC- 3'
DIETNPGP, 5'-GGTGGTGGTGGGGCCCGGGATTCGTCTCAATGTCC- 3'
DIESNPGP, 5'- GGTGGTGGTGGGGCCCGGGATTCGACTCAATGTCC- 3'

4.2.2.2 Hybrid 2A sequences type 1

The upstream context sequences of ADRV, DHV and IFV, were replaced with that of TaV. Primers annealing at the desired position on these consensus motifs were engineered to create Xba1 and TaV upstream context for directional cloning in pSTA1 in conjunction with the reverse primer beginning at the end of GUS and creating a Xho1 restriction site. Templates for PCR were pSTA1 ADRV 2A, DHV 2A and IFV 2A.

DHV30 2A, nucleotide and amino acid sequences

```
1          GCATTTGAACTCAATTTGGAAATTGAATCTGACCAAATTAGAAACAAGAAAGATCTCACT
1           A   F   E   L   N   L   E   I   E   S   D   Q   I   R   N   K   K   D   L   T

61         ACTGAAGGAGTCGAGCCAAACCCCGGGCCC
21          T   E   G   V   E   P   N   P   G   P
```

IFV30 2A, nucleotide and amino acid sequences

```
1          CCCTCAATTGGTAATGTCGCGCGGACTCTGACGAGGGCGGAGATTGAGGATGAATTGATT
1           P   S   I   G   N   V   A   R   T   L   T   R   A   E   I   E   D   E   L   I

61         CGTGCAGGAATTGAATCAAATCCTGGGCCC
21          R   A   G   I   E   S   N   P   G   P
```

ADRV30 2A, nucleotide and amino acid sequences

```
1          TTCTTCGATTCGGTTTGGGTGTACCACTTGGCAAACAGCTCTTGGGTTCGAGATTTAACT
1           F   F   D   S   V   W   V   Y   H   L   A   N   S   S   W   V   R   D   L   T

61         AGAGAATGCATTGAATCTAACCCTGGGCCC
21          R   E   C   I   E   S   N   P   G   P
```

TaV 2A nucleotide and amino acid sequences

```
1          CGCGCCGAGGGCAGGGGAAGTCTTCTAACATGCGGGGACGTGGAGGAAAATCCCGGGCCC
1           R   A   E   G   R   G   S   L   L   T   C   G   D   V   E   E   N   P   G   P
```

Forward primers:

11TaV 9ADRV, 5'-GGTCTAGACGCGCCGAGGGCAGGGGAAGTCTTCTAACAT GCGAATGCATTGAATCTAACCCTGG-3'

12TaV 8ADRV, 5'- GGTCTAGACGCGCCGAGGGCAGGGGAAGTCTTCTAAC ATGCGGGTGCATTGAATCTAACCCTGG-3'

12TaV 9ADRV, 5'- GGTCTAGACGCGCCGAGGGCAGGGGAAGTCTTCTAACATGCGG GGAATGCATTGAATCTAACCCTGG-3'

12TaV 8IFV, 5'- GGTCTAGACGCGCCGAGGGCAGGGGAAGTCTTCTAACATGCGGG GGAATTGAATCAAATCCTG-3'

11TaV 9DHV, 5'- GGTCTAGACGCGCCGAGGGCAGGGGAAGTCTTCTAACA TGCGAAGGAGTCGAGCCAAACCCCG-3'

12TaV 8DHV, 5'- GGTCTAGACGCGCCGAGGGCAGGGGAAGTCTTCTAACATGCGG GGGAGTCGAGCCAAACCCCG-3'

12TaV 9DHV, 5'- GGTCTAGACGCGCCGAGGGCAGGGGAAGTCTTCTAACATGC GGGGAAGGAGTCGAGCCAAACCCCG-3'

Reverse primer for the above 5'-ACGGGCCTCGAGCGGCCGCACTAGATTGTTT-3'

4.2.2.3 Hybrid 2A sequences type 2

The second type of hybrid 2As carrying the upstream context of ADRV, DHV or IFV and the consensus motif of TaV were created by site-directed mutagenesis of these sequences to match the TaV motif [DVEENPGP].

Modification of the end of DHV, IFV and ADRV 2As with TaV 2A consensus motif

Mutant	Forward primer	Reverse primer
ADRV-TaV	5'GTTCGAGATTTAACTAGAGAAGACGTGGAGGAAAATCCCGGGCCCCACCACCACCACCACCACTTACG -3'	5'GGTGGTGGTGGGGCCCGGGATTTTCCTCCACGTCTTCTCTAGTTAAATCTCGAACCCAAG-3'
IFV-TaV	5'-GATGAATTGATTCGTGCAGACGTGGAGGAAAATCCCGGGCCCCACCACCACCACCACCACTTACG -3'	5'-GGTGGTGGGGCCCGGGATTTTCCTCCACGTCTGCACGAATCAATTCATCCTCAATAATCTCCG -3'
DHV-TaV	5'-GAAAGATCTCACTACTGAAGACGTGGAGGAAAATCCCGGGCCCCACCACCACC -3'	5'-GGTGGTGGTGGGGCCCGGGATTTTCCTCCACGTCTTCAGTAGTGAGATCTTTCTTG -3'

11TaV-9 DHV were successively mutated back to TaV 2A

Mutant	Forward primer	Reverse primer
TvDHV G	5'-GGAAGTCTTCTAACATGCGGAGGAGTCGAGCCAAACCCCGGG -3'	5'-GTTTGGCTCGACTCCTCCGCATGTTAGAAGACTTCCCCTGC -3'
TvDHV D	5'-GTCTTCTAACATGCGAAGACGTCGAGCCAAACCCCGGGCCCC -3'	5'-GGTTTGGCTCGACGTCTTCGCATGTTAGAAGACTTCCCC -3'
TvDHV E	5'-GCGAAGGAGTCGAGGAGAACCCCGGGCCCCACCACCACC -3'	5'-GGTGGGGCCCGGGGTTCTCCTCGACTCCTTCGCATGTTAG -3'
TvDHV GD	5'-GGAAGTCTTCTAACATGCGGAGACGTCGAGCCAAACCCCGGGCCCCAC -3'	5'-CGGGGTTTGGCTCGACGTCTCCGCATGTTAGAAGACTTCCCCTGCC -3'
TvDHV GE	5'-GAAGTCTTCTAACATGCGGAGGAGTCGAGGAGAACCCCGGGCCCCACCACCACC -3'	5'-GTGGGGCCCGGGGTTCTCCTCGACTCCTCCGCATGTTAGAAGACTTCCCCTG -3'
TvDHV DE	5'-CTTCTAACATGCGAAGACGTCGAGGAGAACCCCGGGCCCCACCACCAC -3'	5'-GGGGCCCGGGGTTCTCCTCGACGTCTTCGCATGTTAGAAGACTTCCCCTG -3'

4.2.2.4 TaV 2A distance between the [GDV] and [NPGP] motifs

Mutants were created with 1, 2, 3 or 4 residues between TaV: [DVE] and [NPGP] motifs

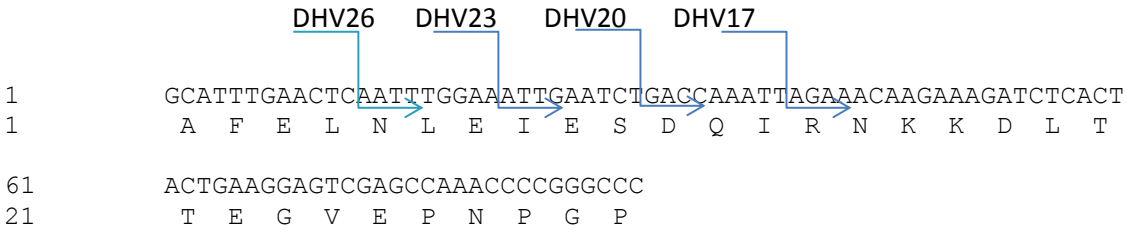
Mutant	Forward primer	Reverse primer
1 AA- A	5'-CGGGGACGTGGCAAATCCCGGGCCCCACC-3'	5'-GGCCCGGGATTTGCCACGTCCCCGCATGTTAG -3'
1AA- E	5'-GCGGGGACGTGGAAAATCCCGGGCCCCACCAC -3'	5'-GGCCCGGGATTTTCCACGTCCCCGCATGTTAG -3'
3AA- EAE	5'-GGGACGTGGAGGCAGAAAATCCCGGGCCCC -3'	5'-CGGGATTTTCTGCCTCCACGTCCCCGCATG -3'
4AA- EAAE	5'-GGGACGTGGAGGCAGCAGAAAATCCCGGGCCCCAC -3'	5'-GCCCGGGATTTTCTGCTGCCTCCACGTCCCCGCATG -3'

4.2.3 Primers for truncations of 2As and alterations of residues side chains

4.2.3.1 N-terminal truncation of ADRV, IFV and DHV

N-terminally truncated 2As were constructed where specific sections in the N-terminal region of ADRV, DHV and IFV sequences were deleted. The primer annealing at the desired position on the ORF was engineered to create a Xba1 site for directional insertion in pSTA1 and to maintain the reading frame. The deletions created 2As of 26, 23, 20 or 17 amino acids. The forward primers are detailed underneath; the reverse primer was 5'-ACGGGCCT CGAGCGGCCGCACTA GATTGTTT-3'.

DHV 2A full-length amino acid and nucleotide sequences and the priming sites (blue arrows) for truncated products are shown underneath.



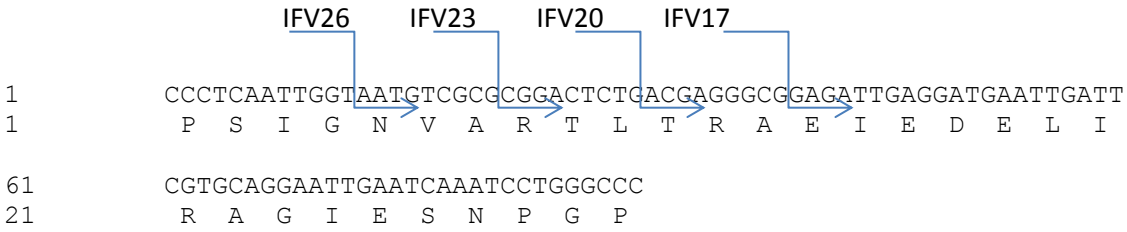
DHV26 forward 5'- TCTAGAAATTTGGAAATTGAATCTGACC-3'

DHV23 forward 5'- TCTAGAATTGAATCTGACCAAATTAGAAAC-3'

DHV20 forward 5'- TCTAGAGACCAAATTAGAAACAAG-3'

DHV17 forward 5'- TCTAGAAGAAACAAGAAAGATCTC-3'

Following is the IFV 2A full-length amino acid and nucleotide sequences with the blue arrows showing the priming site to generate truncated sequences.



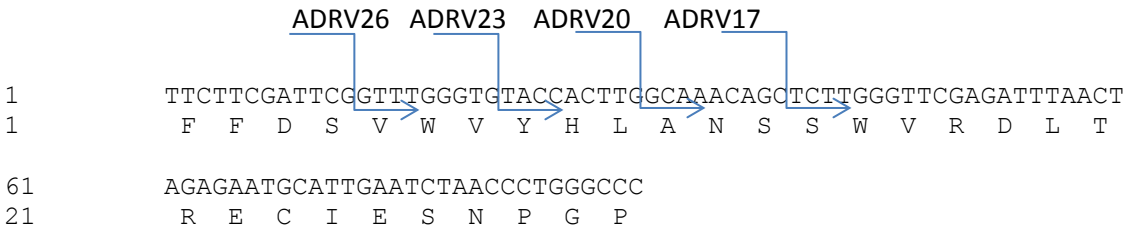
IFV26 forward 5'- TCTAGAAATGTCGCGCGGACTCTGACGAGG -3'

IFV23 forward 5'- TCTAGACGGACTCTGACGAGGGCGGAG -3'

IFV20 forward 5'- TCTAGAACGAGGGCGGAGATTGAGG -3'

IFV17 forward 5'- TCTAGAGAGATTGAGGATGAATTG -3'

Following is the ARDV 2A full-length amino acid and nucleotide sequences with the blue arrows showing the priming site for each truncated sequence.



ADRV26 forward 5'- TCTAGAGTTTGGGTGTACCACTTGGCAAAC -3'

ADRV23 forward 5'- TCTAGATACCACTTGGCAAACAGCTCTTGG -3'

ADRV20 forward 5'- TCTAGAGCAAACAGCTCTTGGGTTCGAGATTTAAC -3'

ADRV17 forward 5'- TCTAGATCTTGGGTTCGAGATTTAACTAG -3'

4.2.3.2 Site-directed mutations to TaV 2A

Target TaV 2A nucleotides and corresponding amino acid sequences:

```
1          CGCGCCGAGGGCAGGGGAAGTCTTCTAACATGCGGGGACGTGGAGGAAAATCCCGGGCCC
1            R   A   E   G   R   G   S   L   L   T   C   G   D   V   E   E   N   P   G   P
```

Primers for successive alanine substitutions

Mutant	Forward primer	Reverse primer
R1A	5‘-CCGGGTCTAGAGCGGCCGAGGGCAGGGG-3‘	5‘-CTGCCCTCGGCCGCTCTAGACCCGGACTT-3‘
E3A	5‘-GTCTAGACGCGCCGCGGGCAGGGGAAGTCTT-3‘	5‘-CTTCCCCTGCCCGCGGCGCGTCTAGACCC-3‘
G4A	5‘-GACGCGCCGAGGCGAGGGGAAGTCTTC-3‘	5‘-GACTTCCCCTCGCCTCGGCGCGTCTAGAC-3‘
R5A	5‘-CGCCGAGGGCGCGGGAAGTCTTCTAACATG-3‘	5‘-GAAGACTTCCCGCGCCCTCGGCGCGTC-3‘
G6A	5‘-CGAGGGCAGGGCGAGTCTTCTAACATGC-3‘	5‘-GTTAGAAGACTCGCCCTGCCCTCGGCG-3‘
S7A	5‘-GGGCAGGGGAGCGCTTCTAACATGCGGG-3‘	5‘-CATGTTAGAAGCGCTCCCCTGCCCTCGGC-3‘
L8A	5‘-CAGGGGAAGTGCGCTAACATGCGGGGAC-3‘	5‘-CGCATGTTAGCGCACTTCCCCTGCCCTC-3‘
L9A	5‘-GGGAAGTCTTGCGACATGCGGGGACGTG-3‘	5‘-CCCCGCATGTCGCAAGACTTCCCCTGCC-3‘
T10A	5‘-GAAGTCTTCTAGCGTGCGGGGACGTGGAG-3‘	5‘-CGTCCCCGCACGCTAGAAGACTTCCCC-3‘
C11A	5‘-GAAGTCTTCTAACAGCGGGGGACGTGGAGGAAAATC-3‘	5‘-CCACGTCCCCCGCTGTTAGAAGA CTTCCCCTG-3‘
G12A	5‘-CTTCTAACATGCGCGGACGTGGAGGAAAATCC-3‘	5‘-CCTCCACGTCCGCGCATGTTAGAAGACTTCC-3‘
D13A	5‘-CTAACATGCGGGGCCGTGGAGGAAAATCCCG-3‘	5‘-GATTTTCCTCCACGGCCCCGCATGTTAGAAG-3‘
V14A	5‘-CATGCGGGGACGCGGAGGAAAATCCCGGGCC-3‘	5‘-GGATTTTCCTCCGCGTCCCCGCATGTTAGAAG-3‘
E15A	5‘-CGGGGACGTGGCGGAAAATCCCGGGCCCCAC-3‘	5‘-CGGGATTTTCCGCCACGTCCCCGCATGTTAG-3‘
E16A	5‘-GGGGACGTGGAGGCAAATCCCGGGCCCCACCACCAC-3‘	5‘-GCCCGGGATTTGCCTCCACGTCCCCGCATGT TAG-3‘
N17A	5‘-GACGTGGAGGAAGCTCCCGGGCCCCACCACCACC-3‘	5‘-GTGGGGCCCGGGAGCTTCCTCCACGTCCCCGCA TG-3‘
P18A	5‘-GGAGGAAAATGCCGGGCCCCACCACCACCACCAC-3‘	5‘-GTGGTGGGGCCCGGCATTTTCCTCCACGTCC CCG-3‘
G19A	5‘-GAGGAAAATCCCGCGCCCCACCACCACCACCACCAC-3‘	5‘-GTGGTGGTGGGGCGCGGGATTTTCCTCCAC GTCC-3‘
P20A	5‘-GAAAATCCCGGGGCCCACCACCACCACCACCACTTAC-3‘	5‘-GGTGGTGGTGGTGGGCCCCGGGATTTTCCTCCA CGTC-3‘

Primers for mutations of TaV 2A arginine 1 and 5

Mutant	Forward primer	Reverse primer
R1E	5‘-GAACTATACAAGTCCGGGTCTAGAGAAGCCGAGGGCAG GGGAAGTCTTCTAACATG -3‘	5‘-GACTTCCCCTGCCCTCGGCTTCTCTAGACCCGGAC TTGTATAGTTCG -3‘
R5E	5‘-GGTCTAGACGCGCCGAGGGCGAAGGAAGTCTTCTAACATG CGG -3‘	5‘-GCATGTTAGAAGACTTCCTTCGCCCTCGGCGCGT CTAGACCCG -3‘
R(1+5)E	5‘-GAACTATACAAGTCCGGGTCTAGAGAAGCCGAGGG CGAAGGAAGTCTTCTAACATGCGGGG -3‘	5‘-GCATGTTAGAAGACTTCCTTCGCCCTCGG CTTCTCTAGACCCGGACTTGTATAGTTCG -3‘
R1I	5‘-CTATACAAGTCCGGGTCTAGAATAGCCGAGGGCAGGGGAA GTCTTC -3‘	5‘-GACTTCCCCTGCCCTCGGCTATTCTAGACCCGGA CTTGTATAGTTCG -3‘
R5I	5‘-GTCTAGACGCGCCGAGGGCATAGGAAGTCTTCTAACATG CGG -3‘	5‘-GCATGTTAGAAGACTTCCTATGCCCTCGGCGCG TCTAGACCCG -3‘
R(1+5)E	5‘-GAACTATACAAGTCCGGGTCTAGAATAGCCGAGGGCATAG GAAGTCTTCTAACATGCGG -3‘	5‘-GTCCCCGCATGTTAGAAGACTTCCTATGCCCTCG GCTATTCTAGACCCGGACTTGTATAG -3‘
R1K	5‘-GAACTATACAAGTCCGGGTCTAGAAAAGCCGAGGGCAGG GGAAGTCTTCTAACATG -3‘	5‘-GACTTCCCCTGCCCTCGGCTTTTCTAGACCCGGAC TTGTATAGTTCGTCCATG -3‘
R5K	5‘-GTCTAGACGCGCCGAGGGCAAAGGAAGTCTTCTAACATGC GGGGACG -3‘	5‘-GCATGTTAGAAGACTTCCTTTGCCCTCGGCGCGT CTAGACCCGGACTTG -3‘
R(1+5)K	5‘-GAACTATACAAGTCCGGGTCTAGAAAAGCCGAGGGCAAAG GAAGTCTTCTAACATGCGGG -3‘	5‘-GTCCCCGCATGTTAGAAGACTTCCTTTGCCCTCGG CTTTTCTAGACCCGGACTTGTATAGTTCGTCCATG -3‘
R1Q	5‘-GAACTATACAAGTCCGGGTCTAGACAAGCCGAGGGCAGG GGAAGTCTTCTAACATG-3‘	5‘-GACTTCCCCTGCCCTCGGCTTGTCTAGACCCGGA CTTGTATAGTTCG -3‘
R5Q	5‘-GTCTAGACGCGCCGAGGGCCAAGGAAGTCTTCTAACATGC GGGGAC -3‘	5‘-GCATGTTAGAAGACTTCCTTGGCCCTCGGCGCGT CTAGACCCG -3‘
R(1+5)Q	5‘-GAACTATACAAGTCCGGGTCTAGACAAGCCGAGGGCCAAG GAAGTCTTCTAACATGCGGGGACG -3‘	5‘-GTCCCCGCATGTTAGAAGACTTCCTTGGCCCTCG GCTTGTCTAGACCCGGACTTGTATAGTTCG -3‘
Q rescue	5‘-GACGAACTATACAAGTCCGGGTCTAGACAAAGGGAGGGCC AAGGAAGTCTTCTAACATGCGGGGACGTG -3‘	5‘-GTCCCCGCATGTTAGAAGACTTCCTTGGCCCTCCC TTTGTCTAGACCCGGACTTGTATAGTTCGTCCATG -3‘

4.2.3.3 Other selected mutations on the upstream context

Following are the primers used for mutations performed on the C-terminal region of FMDV, DHV and ADRV 2As.

Target FMDV nucleotides and corresponding amino acid sequences:

```
16          CAACTGTTGAATTTTGACCTTCTTAAGCTTGCGGGAGACGTCGAGTCCAACCCCGGGCCC
6             Q  L  L  N  F  D  L  L  K  L  A  G  D  V  E  S  N  P  G  P
```

FMDV 2A successive Ala substitution from residues 6 to 17 (in grey in above sequence)

Mutant	Forward primer	Reverse primer
Q6A	5'-GGAGCATGCGCACTGTTGAATTTTGACCTTC -3'	5'-GTCAAAATTCAACAGTGCGCATGCTCCTCTAG -3'
L7A	5'-GGAGCATGCCAGGCCTTGAATTTTGACCTTC-3'	5'-CAAAATTCAAGGCCTGGCATGCTCCTC -3'
L8A	5'-CATGCCAGCTGGCCAATTTTGACCTTC-3'	5'-GTCAAAATTGGCCAGCTGGCATGCTCC -3'
N9A	5'-CCAGCTGTTGGCCTTTGACCTTCTTAAG-3'	5'-GAAGGTCAAAGGCCAACAGCTGGCATGC -3'
F10A	5'-GCTGTTGAATGCCGACCTTCTTAAGCTTG-3'	5'-CTTAAGAAGGTCGGCATTCAACAGCTGGCATGC -3'
D11A	5'-GTTGAATTTTGCCCTTCTTAAGCTTGCGG-3'	5'-GCTTAAGAAGGGCAAAATTCAACAGCTGG -3'
L12A	5'-GAATTTTGACGCCCTTAAGCTTGCGGGAG-3'	5'-CAAGCTTAAGGGCGTCAAAATTCAACAGC -3'
L13A	5'-GAATTTTGACCTTGCCAAGCTTGCGGGAGACGTCGAG-3'	5'-CGCAAGCTTGGCAAGGTCAAAATTCAAC -3'
K14A	5'-GACCTTCTTGCCCTTGCGGGAGACGTC-3'	5'-CTCCCGCAAGGGCAAGAAGGTCAAAATTC -3'
L15A	5'-CCTTCTTAAGGCCGCGGGAGACGTCGAGTC-3'	5'-CGTCTCCCGCGGCCTTAAGAAGGTCAAAATTC -3'
A16S	5'-CTTCTTAAGCTTTCGGGAGACGTCGAGTCCAACCC -3'	5'-GACCTTCTTAAGCTTTCGGGAGACGTCG -3'
G17A	5'-CTTAAGCTTGCGGCCGACGTCGAGTCCAACCCC**GG**-3'	5'-GACTCGACGTCGGCCGCAAGCTTAAGAAGGTC -3'

Target DHV20 nucleotides and corresponding amino acid sequences:

```
1           GACCAAATTAGAAACAAGAAAGATCTCACTACTGAAGGAGTCGAGCCAAACCCCGGGCCC
1             D  Q  I  R  N  K  K  D  L  T  T  E  G  V  E  P  N  P  G  P
```

DHV 2A successive Ala substitution from residues 1 to 12 (in grey in above sequence)

Mutant	Forward primer	Reverse primer
D1A	5'-GTCCGGGTCTAGAGCCCAAATTAGAAACAAG -3'	5'-GTTTCTAATTTGGGCTCTAGACCCGGACTTG -3'
Q2A	5'-CGGGTCTAGAGACGCCATTAGAAACAAGAAAG -3'	5'-CTTGTTTCTAATGGCGTCTCTAGACCCGGAC -3'
I3A	5'-GTCTAGAGACCAAGCCAGAAACAAGAAAGATCTC -3'	5'-CTTTCTTGTTTCTGGCTTGGTCTCTAGACCCGG AC -3'
R4A	5'-CTAGAGACCAAATTGCCAACAAGAAAGATCTCACTAC -3'	5'-GATCTTTCTTGTTGGCAATTTGGTCTCTAGACC CG -3'
N5A	5'-GACCAAATTAGAGCCAAGAAAGATCTCACTAC -3'	5'-GAGATCTTTCTTGGCTCTAATTTGGTCTCTAGAC -3'
K6A	5'-CAAATTAGAAACGCCAAAGATCTCACTACTG-3'	5'-GTAGTGAGATCTTTGGCGTTTCTAATTTGGTCTC TAG -3'
K7A	5'-CAAATTAGAAACAAGGCCGATCTCACTACTGAAG GAG-3'	5'-CAGTAGTGAGATCGGCCTTGTTTCTAATTTG GTC -3'
D8A	5'-GAAACAAGAAAGCCCTCACTACTGAAGGAG -3'	5'-CTTCAGTAGTGAGGGCTTTCTTGTTTCTAATTTG -3'
L9A	5'-GAAACAAGAAAGATGCCACTACTGAAGGAGTC-3'	5'-CTCCTTCAGTAGTGGCATCTTTCTTGTTTCTAAT TTG -3'
T10A	5'-CAAGAAAGATCTCGCCACTGAAGGAGTCGAGCC -3'	5'-GACTCCTTCAGTGGCGAGATCTTTCTTGTTTC -3'
T11A	5'-GAAAGATCTCACTGCCGAAGGAGTCGAGCCAA ACC-3'	5'-CTCGACTCCTTCGGCAGTGAGATCTTTCTTGT TTC -3'
E12A	5'-GATCTCACTACTGCCGGAGTCGAGCCAAACC -3'	5'-GGCTCGACTCCGGCAGTAGTGAGATCTTTCTTG -3'

Other mutations performed on ADRV 2A and DHV 2A

Mutant	Forward primer	Reverse primer
ADRV E22G	5‘-GTTCGAGATTTAACTAGAGGATGCATTGAATCT AACCCTGGG -3‘	5‘-GGTTAGATTCAATGCATCCTCTAGTTAA ATCTCGAACCCAAG -3‘
DHV E12D	5‘-GAAAGATCTCACTACTGATGGAGTCGAGCCAAAC CCCGGG -3‘	5‘-GCCCGGGGTTTGGCTCGACTCCATCAGT AGTGAGATCTTTCTTGTTTCTAATTTG -3‘
DHV E12G	5‘-GAAAGATCTCACTACTGGAGGAGTCGAGCCAAAC CCCGGG -3‘	5‘-GGTTTGGCTCGACTCCTCCAGTAGTGAG ATCTTTCTTGTTTCTAATTTGG -3‘
DHV G13D	5‘-GAAAGATCTCACTACTGAAGATGTCGAGCCAAAC CCCGGG -3‘	5‘-GCCCGGGGTTTGGCTCGACATCTTCAGT AGTGAGATCTTTCTTG -3‘

4.2.4 Primers for probing 2A-secondary structure.

Gly and Pro mutants of selected TaV 2A residues

Mutant	Forward primer	Reverse primer
R1G	5‘-CCGGGTCTAGAGGGGCCGAGGGCAGGGG -3‘	5‘-CTGCCCTCGGCCCCTCTAGACCCGGACTTG -3‘
A2G	5‘-GTCCGGGTCTAGAGGGGAGGGCAGGGGAAG -3‘	5‘-CCCCTGCCCTCCCCTCTAGACCCGGACTTG -3‘
E3G	5‘-GTCTAGACGCGCCGGGGGCAGGGGAAGTCTTC -3‘	5‘-CTTCCCCTGCCCCCGGCGCGTCTAGACCC -3‘
R5G	5‘-CGCCGAGGGCGGGGGAAGTCTTCTAACATG -3‘	5‘-GAAGACTTCCCCCGCCCTCGGCGCGTC -3‘
G6P	5‘-CGAGGGCAGGCCCAGTCTTCTAACATGC -3‘	5‘-GTTAGAAGACTGGGCCTGCCCTCGGCG -3‘
S7G	5‘-GGGCAGGGGAGGGCTTCTAACATGCGGG -3‘	5‘-CATGTTAGAAGCCCTCCCCTGCCCTCGGC -3‘
L8G	5‘-CAGGGGAAGTGGGCTAACATGCGGGGAC -3‘	5‘-CGCATGTTAGCCCACTTCCCCTGCCCTC -3‘
L9G	5‘-GGGAAGTCTTGGGACATGCGGGGACGTG -3‘	5‘-CCCCGCATGTCCCAAGACTTCCCCTGCC -3‘
T10G	5‘-GAAGTCTTCTAGGGTGCGGGGACGTGGAG -3‘	5‘-CGTCCCCGCACCCTAGAAGACTTCCCC -3‘
G12P	5‘-CTTCTAACATGCCCCGACGTGGAGGAAAATCC -3‘	5‘-CCTCCACGTCGGGGCATGTTAGAAGACTTCC -3‘

Disruption of secondary structure by proline substitutions, mutations performed on TaV, DHV20 and FMDV 2As

Mutant	Forward primer	Reverse primer
TaV R1P+A2P	5‘-CAAGTCCGGGTCTAGACCACCGGAGGGCAGGGGA AGTCTTC -3‘	5‘-CTTCCCCTGCCCTCCGGTGGTCTAGACC CGGACTTGTATA -3‘
TaV T10P +C11P	5‘-GGAAGTCTTCTACCACCCGGGGACGTGGAGGAAA ATC -3‘	5‘-CTCCACGTCCCCGGGTGGTAGAAGACTTC CCCTGCC -3‘
FMDV A16P	5‘-CTTAAGCTTCCGGGAGACGTCGAGTCC -3‘	5‘-CGACGTCTCCCGGAAGCTTAAGAAGGTC -3‘
FMDV S21P	5‘-GAGACGTCGAGCCCAACCCCGGGCCCCACCACC -3‘	5‘-GCCCGGGGTTGGGCTCGACGTCTCCCGCAAG -3‘
FMDV D11P+ L12P	5‘-CTGTTGAATTTTCCGCCACTTAAGCTTGCGGGAG ACGTC -3‘	5‘-CCGCAAGCTTAAGTGGCGGAAAATTCAA CAGCTGGCATGC -3‘
FMDV A4P+C5P	5‘-CCGGGTCTAGACCACCACCCCAGCTGTTGAATTT TGACC -3‘	5‘-CAACAGCTGGGGTGGTGGTCTAGACCC GGACTTGTATAG -3‘
DHV20 P16A	5‘-GAAGGAGTCGAGGCAAACCCCGGGCCCCACC -3‘	5‘-GCCCGGGGTTTGCCTCGACTCCTTCAGT AGTG-3‘
DHV20 K6P+ K7P	5‘-CAAATTAGAAACCCGCCAGATCTCACTACTGAAG GAG -3‘	5‘-GTAGTGAGATCTGGCGGGTTTCTAATTT GGTCTCTAG -3‘
DHV20 T10P +T11P	5‘-GAAAGATCTCCCTCCTGAAGGAGTCGAGCCAA ACC -3‘	5‘-CTCGACTCCTTCAGGAGGGAGATCTTT CTTGTTTCTAATTTG-3‘

Non-conservative mutations of leucine for asparagine

Mutant	Forward primer	Reverse primer
TaV L8N	5‘-GCAGGGGAAGTAATCTAACATGCGGGGACGTG-3‘	5‘-CCGCATGTTAGATTACTTCCCCTGCCCTCG-3‘
FMDV L15N	5‘-CCTTCTTAAGAATGCGGGAGACGTCGAGTCC-3‘	5‘-GACGTCTCCCGCATTCTTAAGAAGGTCAAAATTC-3‘
FMDV L7N+L8N	5‘-CTAGAGGAGCATGCCAGAATAACAATTTTGACCTTCTTAAGC-3‘	5‘-GAAGGTCAAAATTGTTATTCTGGCATGCTCCTCTAGACCC-3‘
FMDV L12N+ L13N	5‘-GAATTTTGACAATAATAAGCTTGCGGGAGACGTCGAG-3‘	5‘-CTCCCGCAAGCTTATTATTGTCAAAATTCAACAGCTGGC-3‘
DHV20 L9N	5’-CAAGAAAGATAACACTACTGAAGGAGTCGAG-3’	5’-CCTTCAGTAGTGTTATCTTTCTTGTTTC-3’

4.3 Results

The ribosome stalls upon encountering the nascent 2A, a peptide bond is not made and this is followed by re-initiation of translation. The re-initiation step depends on the ability of 2A to stall the ribosome. The results in this section are expressed in terms of arrest or stalling ability/ activity of a 2A sequence.

The standard method previously published to identify the residues responsible for the elongation arrest in SecM, AAP and TnaC, consists of non/conservative and scanning mutagenesis and a qualitative assessment of each residue (Nakatogawa and Ito, 2002, Cruz-Vera and Yanofski, 2008, Spevak et al., 2010). In these publications, each mutated sequence was visually compared to their respective wild-type sequence. The authors categorise the residues as essential (mutation alleviate stalling), partially required (little to no change in stalling activity) or important (mutation affect acitivity but does not result in complete loss of activity). This standard method has been adopted in the context of this thesis.

4.3.1 Mutations of the consensus sequences

4.3.1.1 The [DV/IExNPGP] consensus sequences are interchangeable

The experiment aimed at testing the consensus sequences present at the C-terminal of viruses 2A. Figure 4.2 shows the alignment of the 2A sequences from representative viruses. Only eight consensus motifs were identified and they represent the only conserved region of the 2A sequences listed. In the consensus motif most aa are conserved, differences exist at position two (occupied by Val or Ile) and at position four. Consensus motifs beginning with DVor DI are followed by Thr, Leu or Ser in position four. However, two more aa (Glu and Met) in position four can be identified with consensus motif beginning with DV.

In this experiment, the TaV consensus motif [DVEENPGP] was swapped for [DVETNPGP], [DVEMNPGP], [DVELNPGP], [DVESNPGP], [DIESNPGP], [DIELNPGP] and [DIETNPGP].

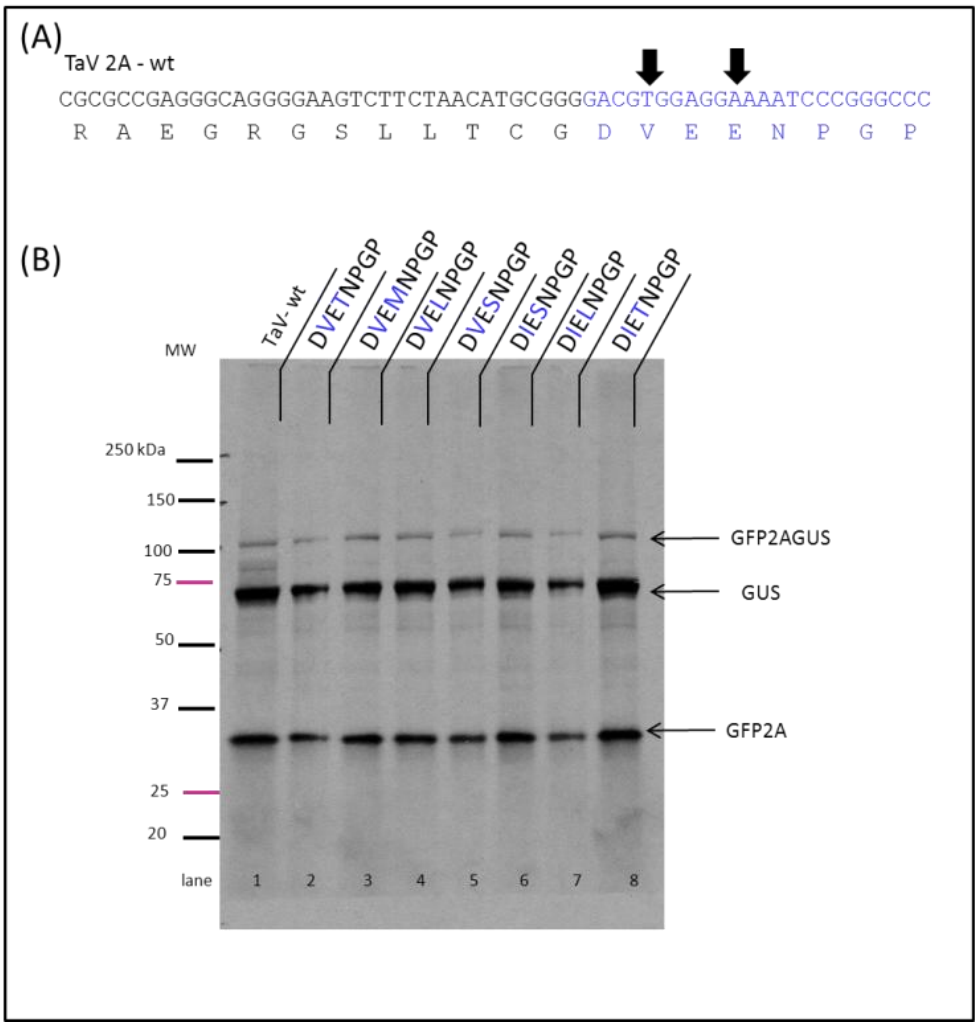


Figure 4.3: Testing the consensus motif variants
SDS-PAGE analysis of the *in vitro* translation of TaV 2A and derivatives. The modified segment of TaV 2A (A) is in blue. The residues that characterised each mutant are in blue in (B) and indicated by a black arrow in (A).

Given the considerable variation and the lack of conservation for the sequences in the upstream context, the straightforward conclusion is that the consensus motif contributes much to the stalling mechanism. This experiment aimed at comparing the eight consensus motifs identified to determine whether the two variable positions are simply plastic or whether these residues are linked to efficiency of stalling and fitted to complement a specific upstream context. No difference was observed between the translation profiles of the wild-type TaV 2A and consensus motif mutants in figure 4.3, all the mutants performed equally efficiently, and therefore not fitted to an upstream context. The conclusion is that the 2A consensus motif beginning with Asp [D(V/I)ExNPGP] exhibits some plasticity at position two and four in its consensus sequence and a large degree of plasticity in its upstream context.

4.3.1.2 The consensus motifs have different tolerances

Hybrid 2As were created. ADRV [CIESNPGP], IFV [GIESNPGP] and DHV [GVEPNPGP] have a mutation on the first aa of the 2A consensus sequence. The upstream context of ADRV, IFV and DHV were joined to the consensus motif of TaV, and the TaV upstream context was placed preceding the ADRV, IFV and DHV consensus sequence.

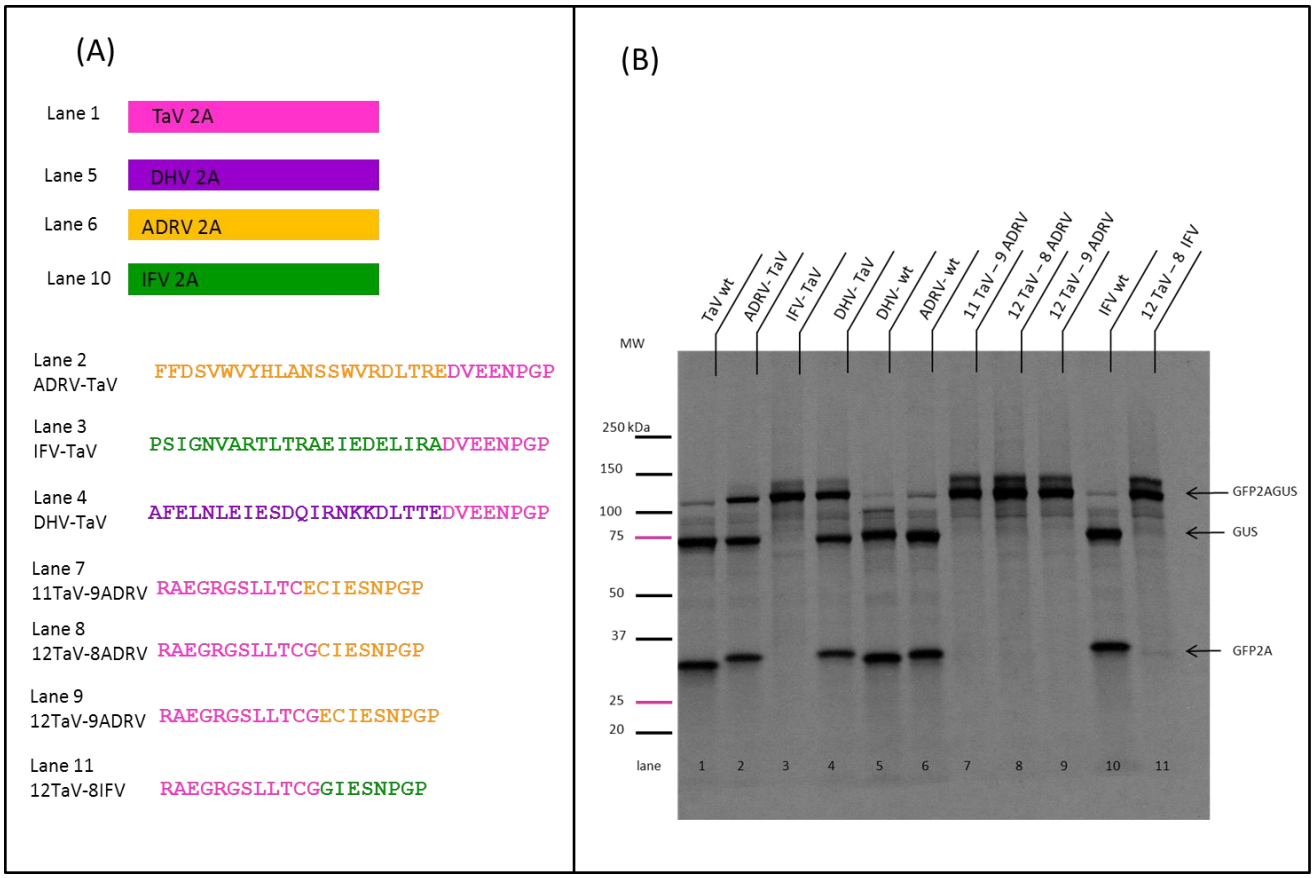


Figure 4.4: The translation of hybrids
The sequences tested are listed in panel (A), and are colour coded. TaV sequence iss represented in pink, DHV in violet, ADRV in orange and IFV in green. The indicated mutants were translated in rabbit reticulocytes lysates with [^{35}S]-methionine and analysed by SDS-PAGE (B).

TaV, DHV and ADRV were highly efficient at stalling the ribosome. DHV-TaV and ADRV-TaV hybrids had a canonical consensus motif, and were still able to stall the ribosome although less efficiently than their respective wild type (lanes 2 and 4). The IFV-TaV hybrid (lane 3) however lost that function. Inverting the sequences and using the TaV upstream context with several consensus motifs, compromised stalling in all cases.

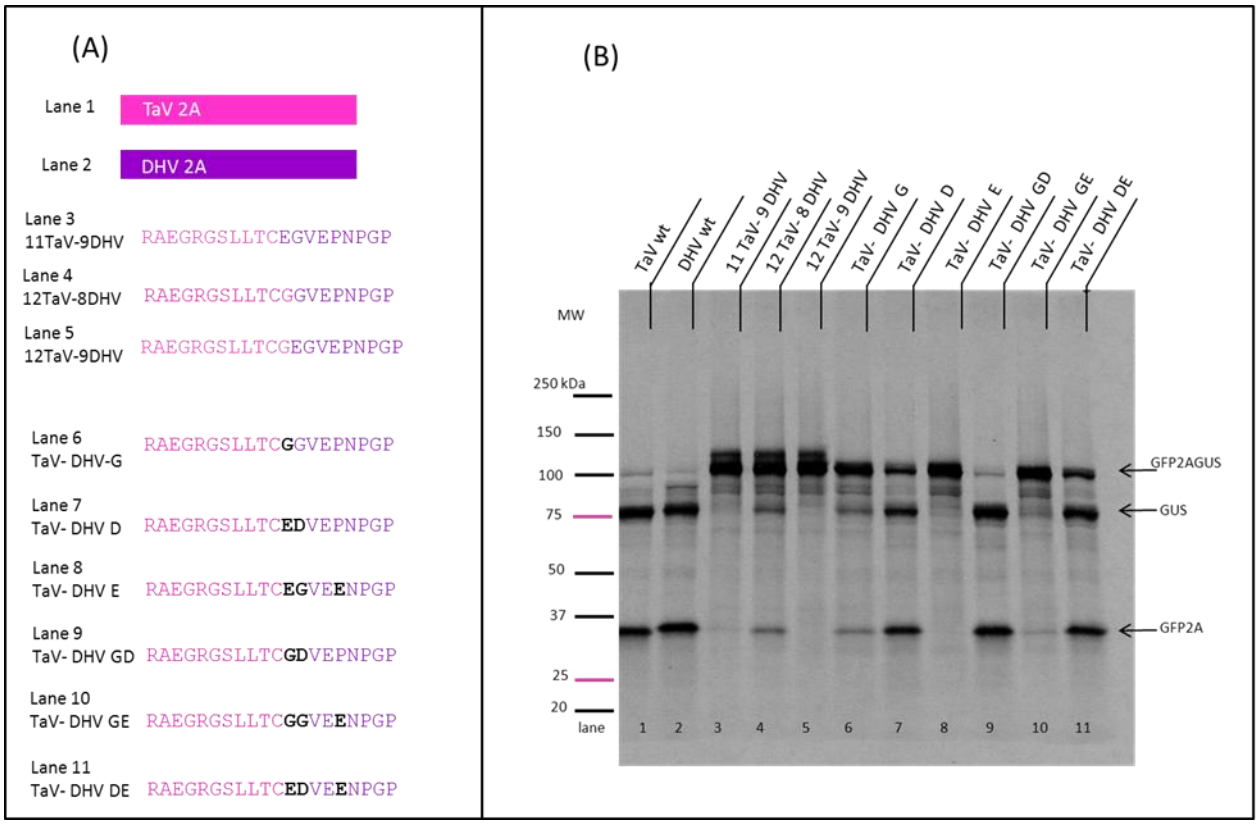


Figure 4.5: Translation of TaV and DHV hybrids
The sequences tested are listed in panel (A). TaV sequence is represented in pink and DHV in violet, and mutations highlighted in black. The sequences indicated in (A) were translated in rabbit reticulocytes lysates with [^{35}S]-methionine and resolved by SDS-PAGE (B).

TaV with an unsual [GGVEPNPGP] motif retained weak stalling abilities 12TaV-8DHV (figure 4.5, lane 4). The four residues [CGGV] were able, although very inefficiently, to generate stalling in the context of TaV 2A (figure 4.5, lane 6).

These results demonstrate that consensus sequences have different requirements from their upstream context. The canonical consensus motif [DV/IExNPGP] could be used with different upstream context as it was testified by the range of sequences identified preceding this motif (figure 4.2) and the results presented in figures 4.3 to 4.5. Non-canonical consensus motifs such as [GIESNPGP] and [CIESNPGP] seemed to have stricter requirements to perform efficiently.

Interestingly, between 11TaV-9DHV (unable to stall the ribosome) and TaV 2A wild type, there were only three amino acids difference. Step mutations to reverse this sequence back to TaV 2A showed that only a few residues were critical to resume activity. Comparing by pairs: lane 6 [GGVEPNPGP]

with 10 [GGVEENPGP], lanes 7 [EDVEPNPGP] with 11 [EDVE ENPGP] and lanes 8 [EGVEENPGP] and 5 [EGVEENPGP], the results were similar within pairs and consistent with previous results (figure 4.2). The identity of the amino acid just preceding [NPGP] did not influence 2A activity. The varying abilities of the mutants seemed to be attributed to the nature of the two residues at the N-terminus of the consensus motif, within the context of this hybrid. The consensus sequence in lane 8 [EGVEENPGP] had no activity. The consensus sequence in lane 10 [GGVEENPGP], had very weak stalling abilities. In lane 11, [EDVEENPGP] 2A could stall the ribosome in an estimated 50% of cases. In lane 9, which corresponded to a mutation back to the generic motif [GDVEPNPGP], 2A activity was completely recovered. From these results, it was concluded that the beginning of the consensus motifs were also to an extent plastic. The other residues which could be accommodated were GG, ED and IC instead of GD. The residues found at these positions seemed to be somehow involved in the efficiency of stalling.

4.3.1.3 A hierarchy of importance in the consensus amino acids.

The consensus motif of TaV was subjected to successive alanine substitutions, in order to uncover the essential residues. The previous studies proved that the x in [DV/IExNPGP] could be substituted for M, L, S, and T without interfering with 2A activity (figure 4.3). The motif [GD(V/I)] could also be substituted for [GGV], [EDV] and [ICE], however in the context of TaV 2A these latter resulted in loss of efficiency (figures 4.4 and 4.5). Based on these results, further mutations were introduced in the consensus motif of TaV 2A to modify the distance between the two segments [GDV] and [NPG] (figure 4.6). A single amino acid was inserted; Ala [GDVANPGP], Glu [GDVENPGP], and one and two extra Ala residues were inserted to yield [GDVEAENPG] and [GDVEAAENPGP].

In figure 4.6, mutations of the glutamic acid residues to alanine (E15A and E16A, panel (A), lane 4 and 5) did not alter the stalling capacity of TaV 2A; insertion or deletions of amino acids in this segment (panel (B) lanes 2 to 6) completely abolished arrest activity. Therefore the exact spacing of [GDV] in relation to [NPG] was critical. Mutations of aspartic acid (D13A in panel (A), lane 2) as well as mutations of proline and glycine (P18A, G19A in panel (A), lanes 7 and 8) completely abolished 2A activity. Mutations of glycine, valine and asparagine (G12A, V14A and N17A in panel (A) lanes 1, 3 and 6) impaired activity. Interestingly, substitution of proline (P20A in panel (A) lane 9) resulted in only 50 % loss of stalling ability. Proline substitutions at this position were further explored in chapter 5.

From these results it was concluded that [NPG] was the segment of the 2A required for stalling and its exact positioning in relation to the next stretch of important residues was essential. There was a hierarchy of importance in residues. [NPG] were considered essential residues, [GDV] important and EE partially required.

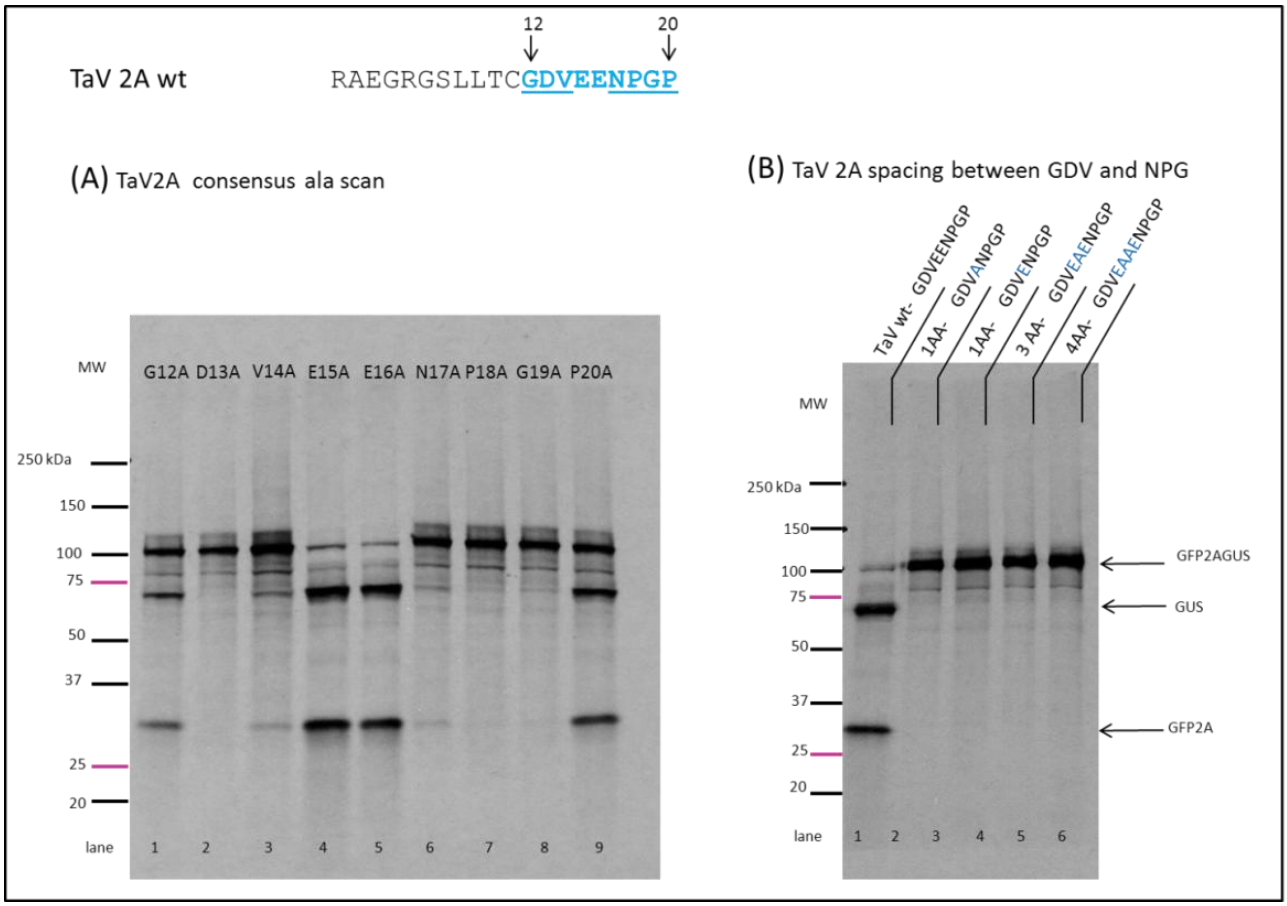


Figure 4.6: Mutations to TaV 2A consensus sequence

The target sequences are written in blue, the black arrows indicates the residue positions and the underlined sequences are the two areas of the consensus motif between which residues were inserted or removed. (A) Translation profile for successive alanine substitution of the TaV 2A consensus motif from residues 12 to 20. In (B), results of the activity assay with varying spaces between the consensus segments [GDV] and [NPG], the mutations in (B) are indicated in blue.

4.3.2 Requirements from the upstream context

4.3.2.1 The sufficient sequence required

It was previously proven that FMDV 2A required thirty amino acids to be optimal (100 % stalling) but the minimum length required was thirteen residues (Ryan and Drew, 1994). This was the arbitrary reason for cloning a length of thirty residues for each novel 2A. For the purpose of rationalising this experiment ADRV, IFV and DHV 2As were subjected to a series of deletions to yield 26, 23, 20 and 17 aa truncated versions.

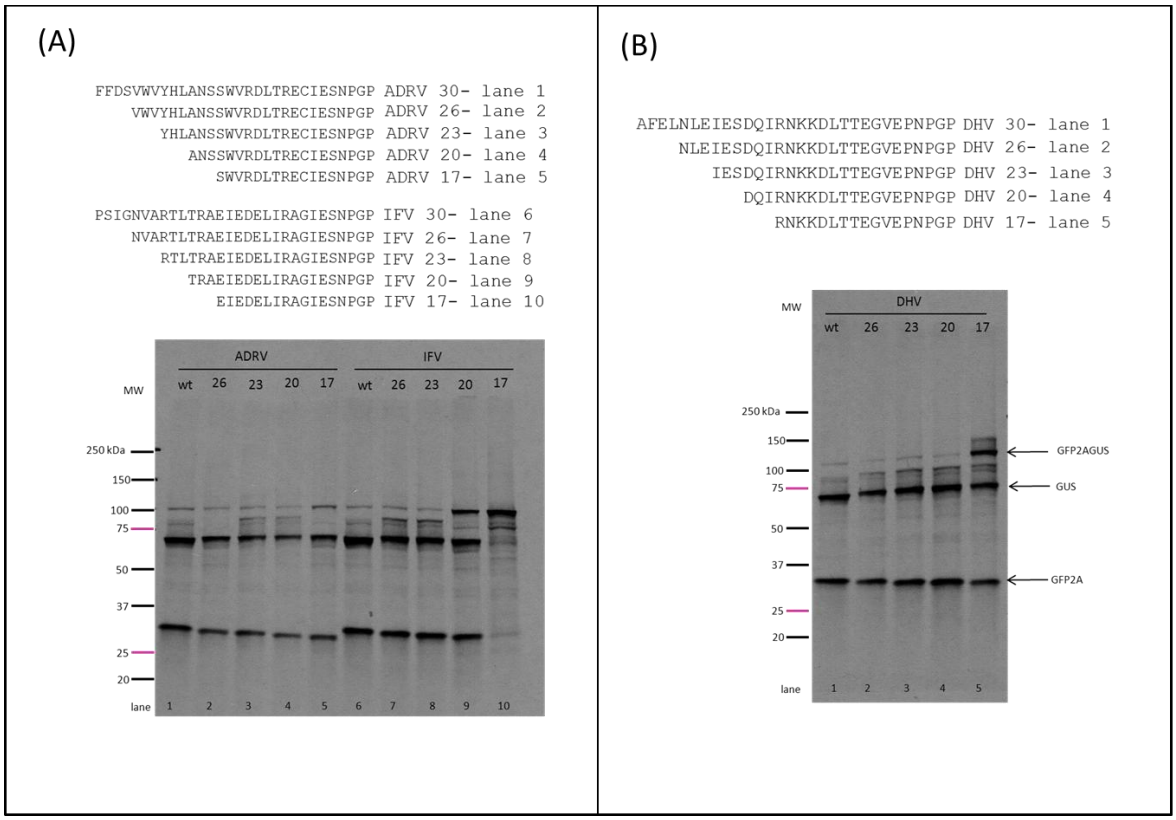


Figure 4.7: N-terminal truncation of ADRV, IFV and DHV 2A

The sequences tested were above the gel in (panel A) and (panel B). The N-terminal regions were deleted and analysed in the I*n vitro* activity assay and resolved by SDS-PAGE gels.

In figure 4.7, comparing ADRV truncated 2As to the wild type, the activity was not affected until the sequence was truncated to seventeen aa (comparing lane 5 to lane 1). IFV 2A activity was not affected until the sequence was truncated to twenty aa and a further deletion to seventeen aa abolished stalling (lanes 9 and 10). Shorter versions of DHV mediated more full-length products with seventeen C-terminal residues.

From this experiment it was concluded that the minimum length sufficient to maintain wild type 2A activity was twenty-three aa for IFV and twenty aa for DHV and ADRV. TaV 2A was twenty aa long. FMDV 2A was naturally eighteen aa long and defined by the $3C^{pro}$ cleavage site at the 1D/2A junction. It was therefore decided to target the twenty C-terminal residues for the further site directed mutageneses on DHV, FMDV and TaV 2As.

4.3.2.2 Mutational tolerance of the upstream context

TaV, FMDV and DHV 2A upstream contexts were subjected to successive alanine mutations. Testing the derivatives *in vitro* allowed visualizing the contribution of each residue to 2A activity. The TaV 2A residue 2 (alanine) mutant to glycine is included in the glycine scan of TaV 2A in figure 4.14.

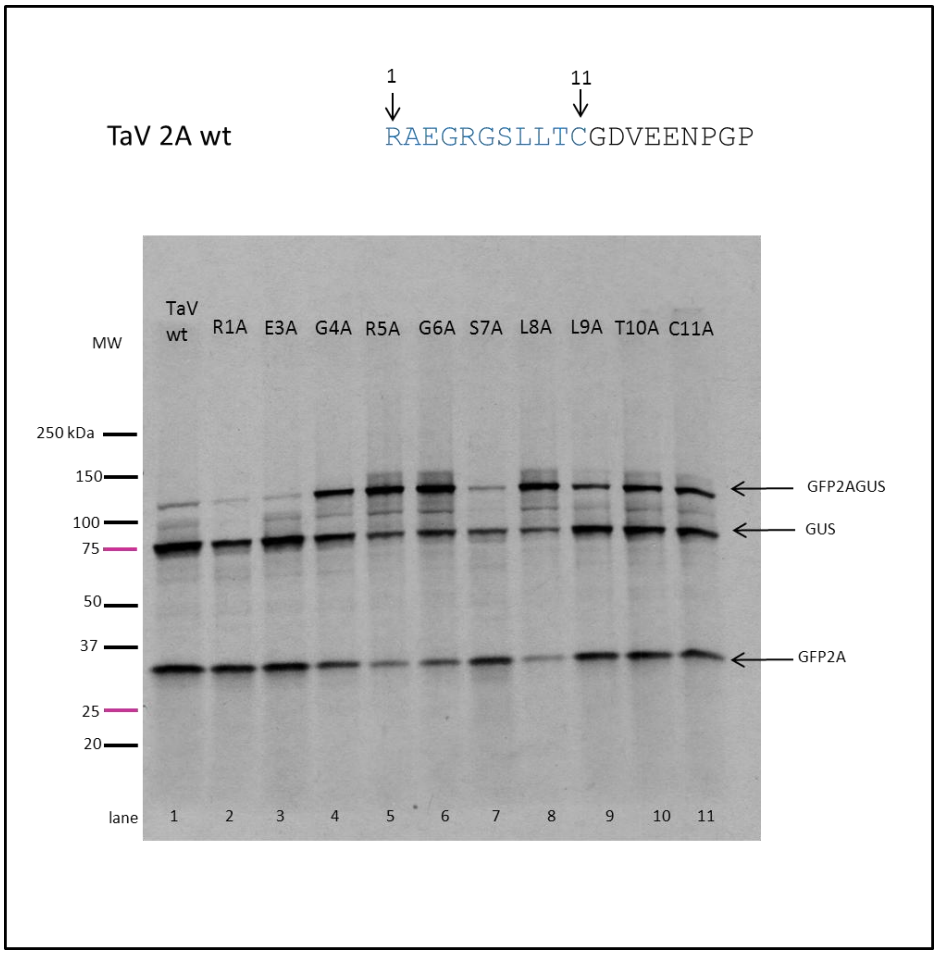


Figure 4.8: Alanine scan of TaV 2A upstream context

TaV 2A target sequence (residues 1 to 11, in blue) was subjected to alanine scanning mutagenesis, tested *in vitro* and resolved by SDS-PAGE. The residues number and mutations are indicated for each lane.

In figure 4.8, altering the side chains of the residues Arg1 (lane 2), Glu3 (lane 3) and Ser7 (lane 7) did not compromise the activity of TaV 2A. Any other substitution across the length of the upstream contest was arrest impairing. However no substitution completely abolished TaV 2A activity.

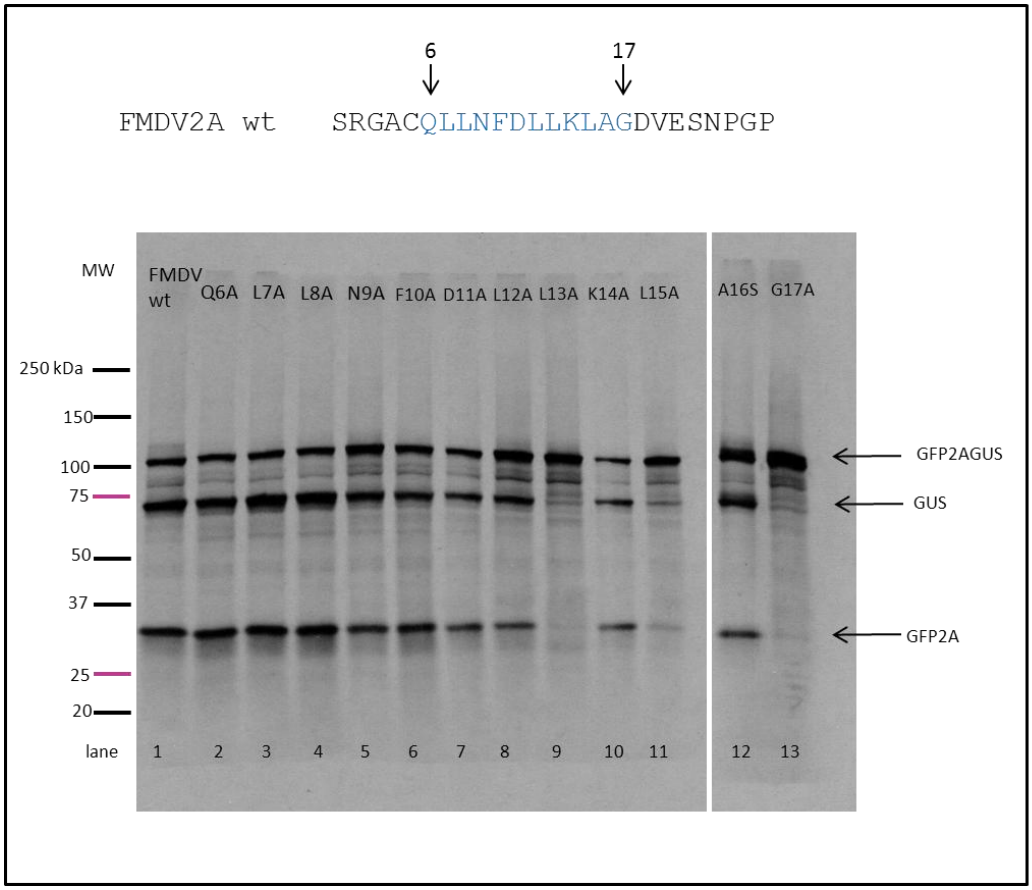


Figure 4.9: Alanine scan of FMDV 2A upstream context

The target sequence (residues 6 to 17) is written in blue, with black arrows pointing the residues number. The mutated 2As were subjected to the activity assay and resolved by SDS-PAGE. The residue number and the mutation performed are indicated at the top of each lane.

In figure 4.9, FMDV 2A could not, or was barely able, to mediate the ribosome stalling when Leu13 (lane 9), Leu15 (lane 11) side chains were altered. Substitution of Gly17 (lane 13) to alanine had a similar affect. Modification of Ala16 to serine (lane 12) resulted in more full-length products compared to wild type (lane 1). However, substitutions of any other residues did not alter the translation profile as derivatives were as efficient as the wild type 2A.

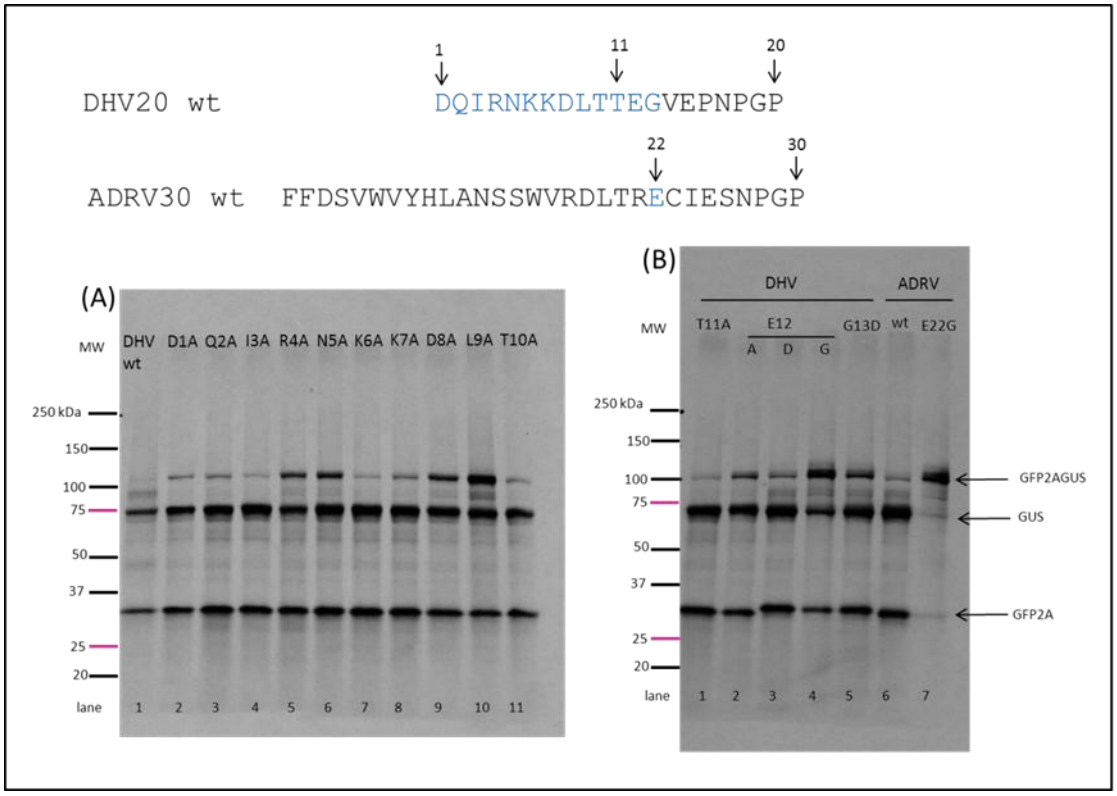


Figure 4.10: Other mutations of the upstream contexts

The target sequence (residues 6 to 17) is written in blue, with black arrows pointing the residues number. DHV 2A upstream context was subjected to successive Ala mutations (panel (A)). Panel (B) shows non-conservative mutations performed on DHV Glu12 and ADRV 2A Glu22. The mutated 2As were subjected to the activity assay and resolved by SDS-PAGE.

In figure 4.10, substitutions of DHV20 2A Glu12 (E12) for aspartic acid, alanine or glycine (panel (B) lanes 3,2 and 4) resulted in more full-length products but did not abolish the 2A stalling ability, adding to the list of residues employable in that region of the consensus motif. Successive alanine substitutions was performed on DHV20 2A. Mutations of the first three residues (lanes 2, 3 and 4- panel (A)) and of the residues Lys6 (lane 7), Lys7 (lane 8), Thr10 (lane 11) and Thr11 (lane 1 panel (B)) did not cause significant alterations in the ratio of full-length versus GFP2A and GUS products. Mutations of Arg4 (lane 5), Asn5 (lane 6), Asp8 (lane 9) and Leu9 (lane 10) however impaired 2A activity and increased the amount of full-length products.

A comparison of the results in figures 4.8, 4.9 and 4.10 showed that the residues in the upstream contexts exhibited different tolerances to substitution. The conclusion from this set of experiments is that residues in 2A sequences are partially required, important or essential.The above findings are summarised in figure 4.11.

The alanine-scanning mutagenesis served as the basis upon which the residues were classified in three categories. The effect of each susbtitution was assessed against the activity of their respective wild

type 2A sequences. The activity of the 2A variants was estimated by the amount of full-length products formed. In figure 4.11, residues marked in brown tolerated alanine substitution, and none to little alteration to the activity was observed. The residues side chains at these positions may therefore not directly be involved in elongation arrest. For the residues attributed the colour orange or red, alanine substitution at these positions impaired 2A activity (orange colour) or abolished 2A activity (red colour). Residues at these positions may therefore be critical for 2A-induced ribosome stalling. It is interesting to note that although the Asp residue of the consensus motif is marked red since mutation to Ala completely abolished stalling (figure 4.6 lane 2), its mutation to Gly (figure 4.5, lane 6) preserved weak stalling abilities.

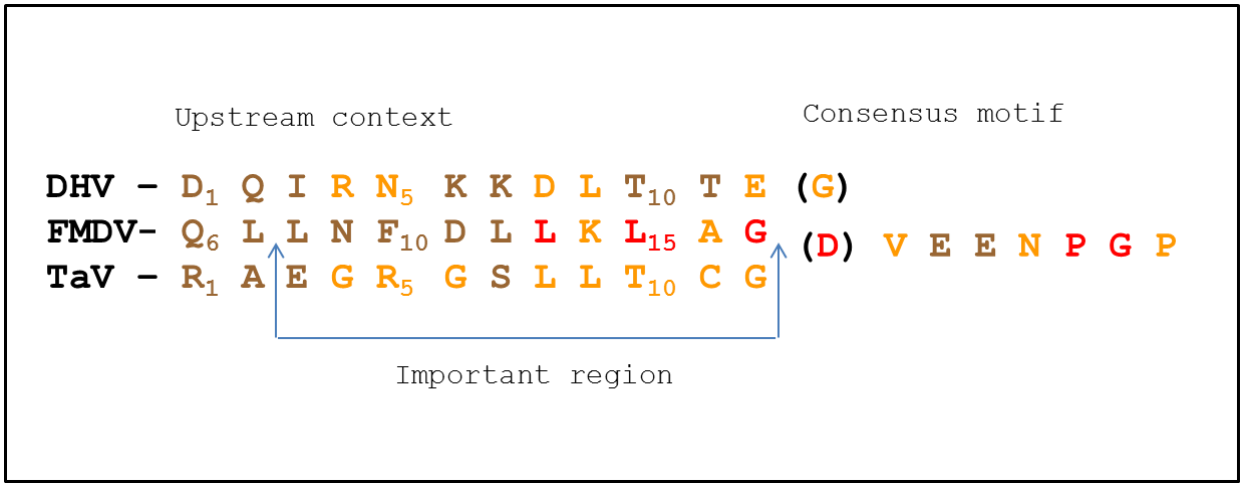


Figure 4.11: Summary showing importance of residues for efficiency of 2A sequences

The residues marked red were found to alleviate stalling and are judged essential, the residues written in orange, are important. The stalling ability was compromised when these residues were substituted. And finally the brown colour is attributed to residues that were required but did not seem to play an active role in the stalling mechanism. The first consensus motif residue Asp is in bracket as this position is occupied by Gly for DHV 2A.

The observation from the figure 4.11 is that for all 2As there is a stretch of important residues (red and orange residues) at the C-terminus of the upstream context leading to the consensus motif. This section of the 2A sequences is therefore termed the ‘important region’. TaV and DHV 2As ‘important region’ encompasses nine residues. FMDV 2A ‘important region’ is shorter and comprises the five residues preceding the consensus motif.

Next, the sequences deposited in NCBI for FMDV 2A and DHV 2A were aligned using the weblogo web server (http://www.weblogo.berkeley.edu) and the result is shown in figure 4.12. A similar analysis was not possible for TaV 2A. TaV is not an economically important virus and there are only two entries in the public databases for its genome.

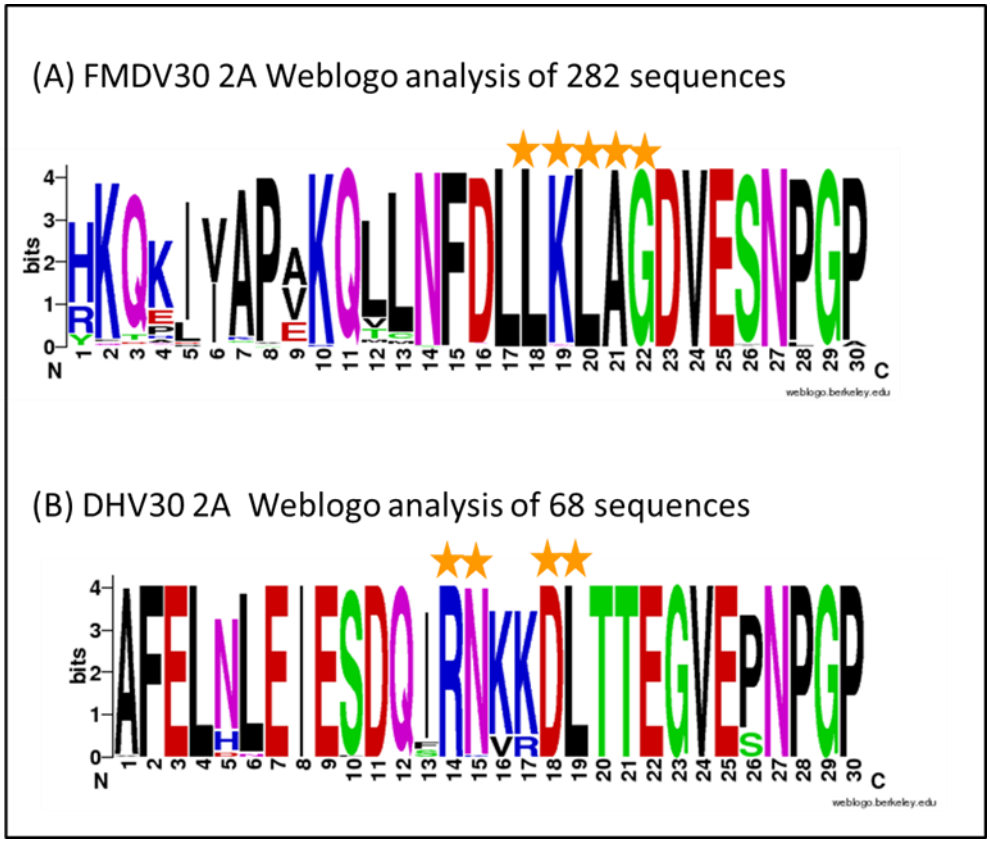


Figure 4.12: Frequency plot of the residues conservation of FMDV and DHV 2A

The letter size corresponds to the frequency of occurrence for this residue; the numbering is shown underneath. In (A) the plot was generated from 282 available FMDV 2A sequences and in (B) from 68 available DHV 2A sequences. The gold star indicated residues in the upstream context for which alanine substitution changed the translation profile for these 2A.

The frequency plots in fugure 4.12 showed that residues that were either important or essential were conserved across sequences for each 2A. Surprisingly the conservation extended to the other residues: [NFDL] for FMDV 2A and to two stretches of residues [EIESDQ] and [AFEL] for DHV 2A. Referring to the frequency plot for FMDV, Pro28 as well as Pro30 of the consensus motif are not completely conserved. It was previously assumed that these were sequencing errors. Pro28 is in a critical position; it is therefore likely that the deviating sequence Pro28 to Leu is due to a sequencing error. In contrast, Pro30 to Ala mutant was viable (figure 4.6). It is therefore likely that FMDV 2A sequences containing a C-terminal Ala do exist naturally.

4.3.3 Probing interactions with the ribosome exit tunnel

Argine and lysine are residues able to interact with RNA (Lu et al., 2007). An early experiment involved a series of non-conservative and conservative mutations performed on TaV 2A. TaV 2A has two arginine residues in its upstream context. These were mutated in turn and together to glutamic acid, isoleucine, lysine or glutamine.

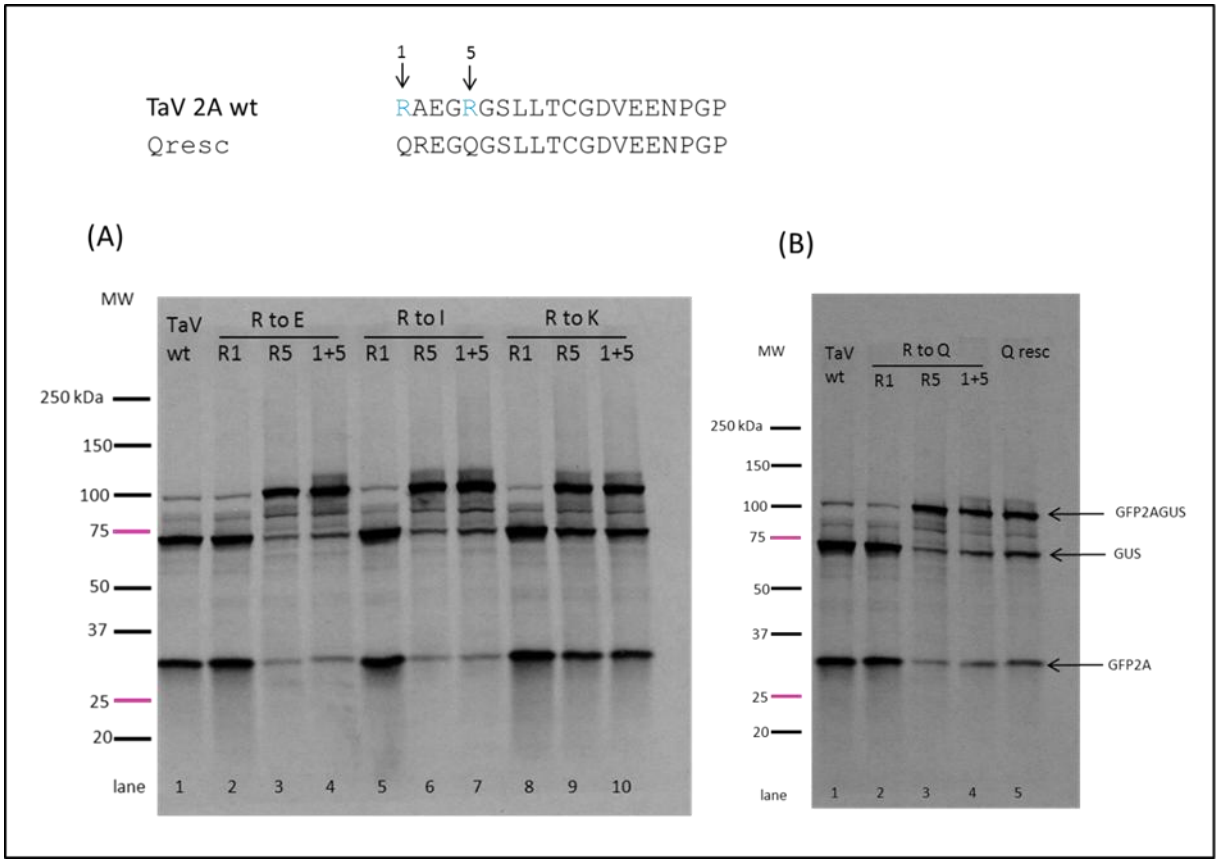


Figure 4.13: Mutations of TaV Arg 1 and/or 5

The target residues were written in blue and indicated with a black arrow at the top of the figure. The relevant mutations and residues number were indicated for each lane. Panel (A) shows mutation of argine residues R1 and/or R5 to glutamic acid, isoleucine or lysine. Panel (B), mutation to glutamine. The mutants 2As were subjected to the activity assay and resolved by SDS-PAGE.

Mutations of the first residue did not alter the activity of the 2As (figure 4.13). In panel (A) lane 2 (R1E), lane 5 (R1I) and lane 8 (R1K), in panel (B) lane 2 (R1Q) had the same translation profile as TaV 2A wild type (panel (A) lane 1). Comparing pair-wise in panel (A) lane 3 with 4, lane 6 with 7, and lane 9 with 10, in panel (B) lane 3 and 4, the double mutations had a similar translation profile to the single point mutation of Arg5 only.

Mutations of Arg5 to glutamic acid (panel (A) lane 3), or isoleucine (panel (A) lane 6) or glutamine (panel (B) lane 3) proved arrest impairing. Mutation of Arg5 to lysine (panel (A) lane 9) enabled stalling with reduced efficiency compared to wild type (lane 1). The results suggested that a conservation of charge in that position was necessary for TaV 2A.

Proline and glycine residues disrupt the secondary structure of a protein. If the peptide adopted an impairing conformation in the ribosome exit tunnel, the hypothesis was that substituting residues successively to glycine, and glycine residues to proline would completely abolish 2A activity.

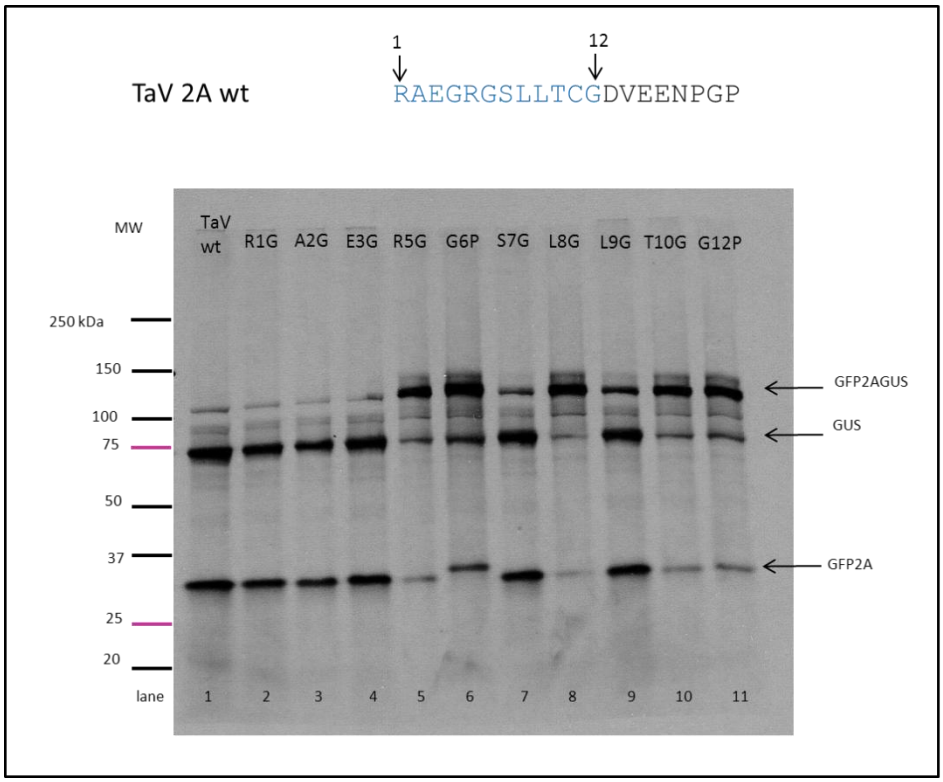


Figure 4.14: Glycine and proline mutations of the upstream context of TaV 2A

The target residues are written in blue and indicated with a black arrow. The relevant mutations and residues number are indicated for each lane. The mutants 2As were translated in rabbit reticulocyte lysates with [^{35}S]-methionine and resolved by SDS-PAGE.

The results shown in figure 4.14 were similar to those previously obtained for the alanine scan of the upstream context of TaV 2A (figure 4.8). Weak stalling ability is retained for Arg5 (lane 6) Gly6 (lane 7), Leu8 (lane 8), Thr10 (lane 10) and Gly12 (lane 11). The results do not support a model where TaV 2A would adopt a helical conformation within the tunnel which is disrupted by introduction of glycine or proline. The results demonstrate that specific amino acids cannot be altered and strongly suggest a model where TaV interacts with the ribosome exit tunnel at these specific residues.

Woolhead and colleagues (2006), using FRET found that the seventeen C-terminal aa of SecM adopt a compacted conformation in the exit tunnel. They reasoned that proline substitutions should abolish the compaction by constraining the polypeptide backbone. They substituted in turn T152, S157, Q158, A159 and Q160 to proline and these abolished SecM activity. The resulting FRET analysis revealed an extended conformation. Addition of multiple proline residues in 2A would constrain the movements of the peptide chain. The conformation and therefore the positioning of the nascent 2A in the ribosome exit tunnel should be compromised. The next set of mutations was carried out with DHV20, FMDV and TaV 2As. Double proline mutations were introduced to the partially required residues of DHV upstream context: Lys6-Lys7 then Thr10-Thr11. Presuming 2A activity relied on its conformation; DHV 2A proline mutants would lose activity. If however there was no structural pattern and these residues were partially required for providing the correct spacing for example, then addition of proline at these residues would not alter DHV 2A activity. Similarly, double mutants were introduced in the upstream context of TaV: Arg1-Ala2, which were not necessary to activity, and Thr10-Cys11 which were important residues for 2A activity. The same strategy was also applied to FMDV. Proline mutations for residues Ala4-Cys5 (outside of the minimal length required) were introduced, also partially required residues Asp11-Leu12 were mutated to proline.

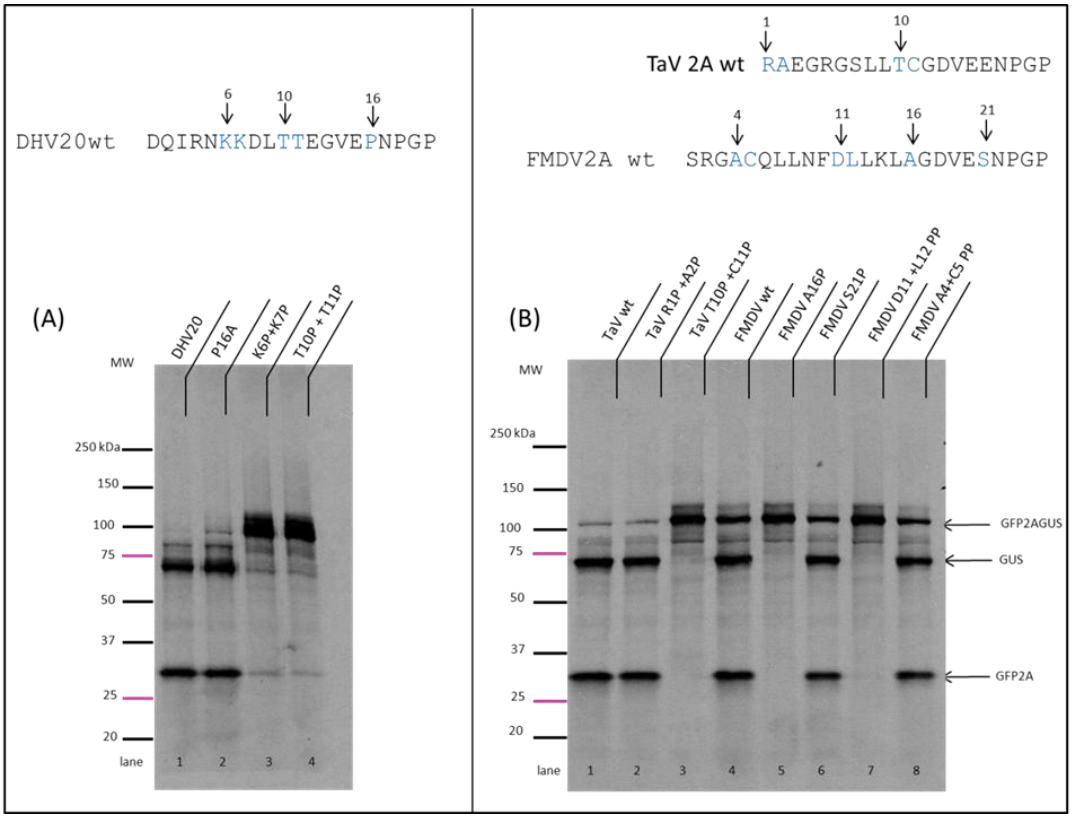


Figure 4.15: Proline disruption of DHV20, FMDV and TaV 2As

The target residues are written in blue and indicated with a black arrow and numbered. The relevant mutations and residues number are indicated for each lane. The mutants 2As (in panel (A) DHV and panel (B) TaV and FMDV) were translated in rabbit reticulocytes lysates with [^{35}S]-methionine and resolved by SDS-PAGE.

Introducing multiple proline in the segment of 2A (figure 4.15) outside the 'important region', did not change the translation profile for TaV and FMDV (panel (B) lanes 2 and 8 respectively). Introducing proline within that 'important region', completely compromised 2A activity, whether the residues were previously shown to be essential (panel (B) lane 5), partially required (panel (A) lanes 3 and 4 and panel (B) lane 7) or important (panel (B) lane 3).

From this study it was concluded that TaV and DHV 2As would most likely adopt a conformation in the tunnel that favoured the partnering of several of its residues with nucleotides of the exit tunnel. Since both of these 2As are not predicted to have helical propensity (table 4.1), it is possible these would adopt an extended conformation within the exit tunnel.

Referring to figure 4.11, it is noticeable that all the 2As have a leucine residue or sometimes even two in the region just upstream of their consensus motif. Leucine and asparagine are of similar size, but asparagine is hydrophilic. In the next experiment Leu was substituted with Asn.

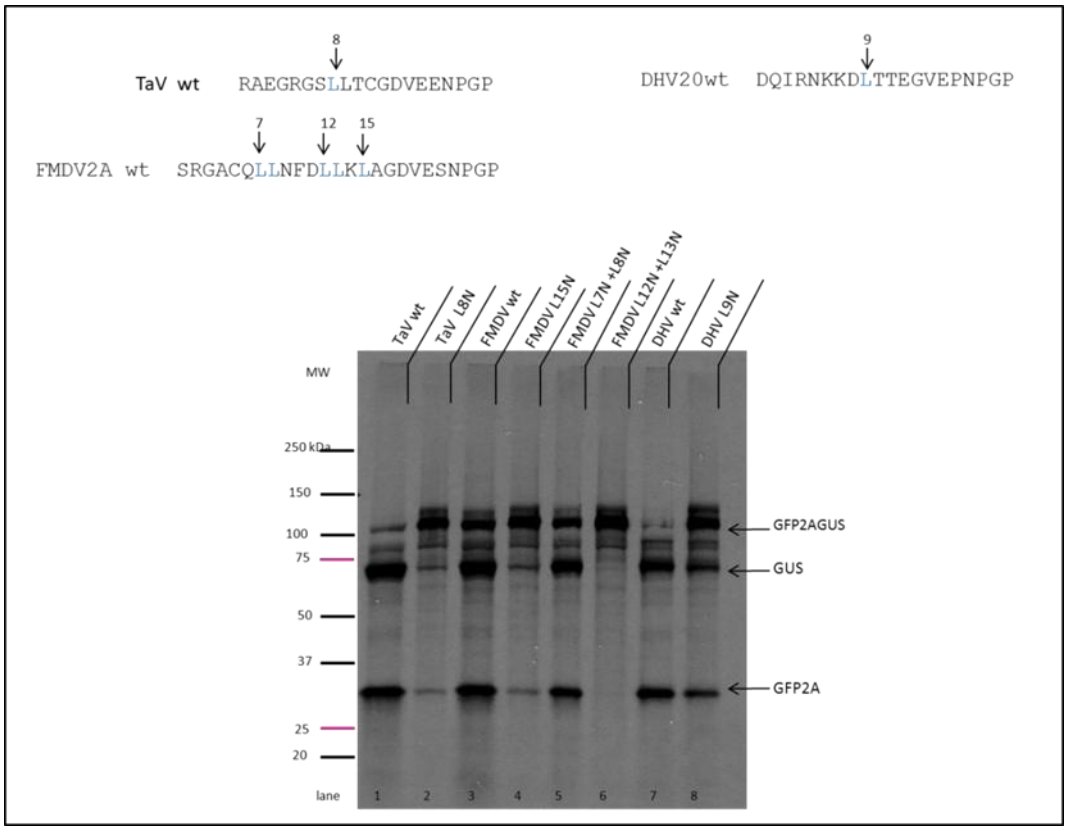


Figure 4.16: Leucine substitutions in the upstream context of DHV, TaV and FMDV 2As

The target residues are indicated in blue and indicated with a black arrow and numbered. The relevant mutations and residues number are indicated for each lane. The mutants 2As were translated in rabbit reticulocyte extracts with [^{35}S]-methionine and resolved by SDS-PAGE.

Mutations of leucine residues (figure 4.16) in the regions of FMDV 2A partially required (lane 5), did not impair 2A activity. But substituting the residues 12 and 13 (lane 6), and 15 (lane 4) for asparagine created inactive 2As. Similarly, substitution of leucine residues for a hydrophilic amino acid resulted in decrease in the ability of the sequence to stall the ribosome for DHV (lane 8) and TaV 2A (lane 2), a result similar to substitution with the hydrophobic residue alanine. These results proved that a leucine in that position is critical for 2A activity.

From the mutational analysis and truncation studies carried out so far, the 'important region' is critical for stalling and precedes the consensus motif. How diverse are the 2A sequences when only the 'important region' is analysed? Figure 4.17 was created by manually sorting the 2A sequences into groups based on sequence similarities of the five C-terminal residues. The results show that viruses from a family share very similar and sometimes identical 'important regions'.

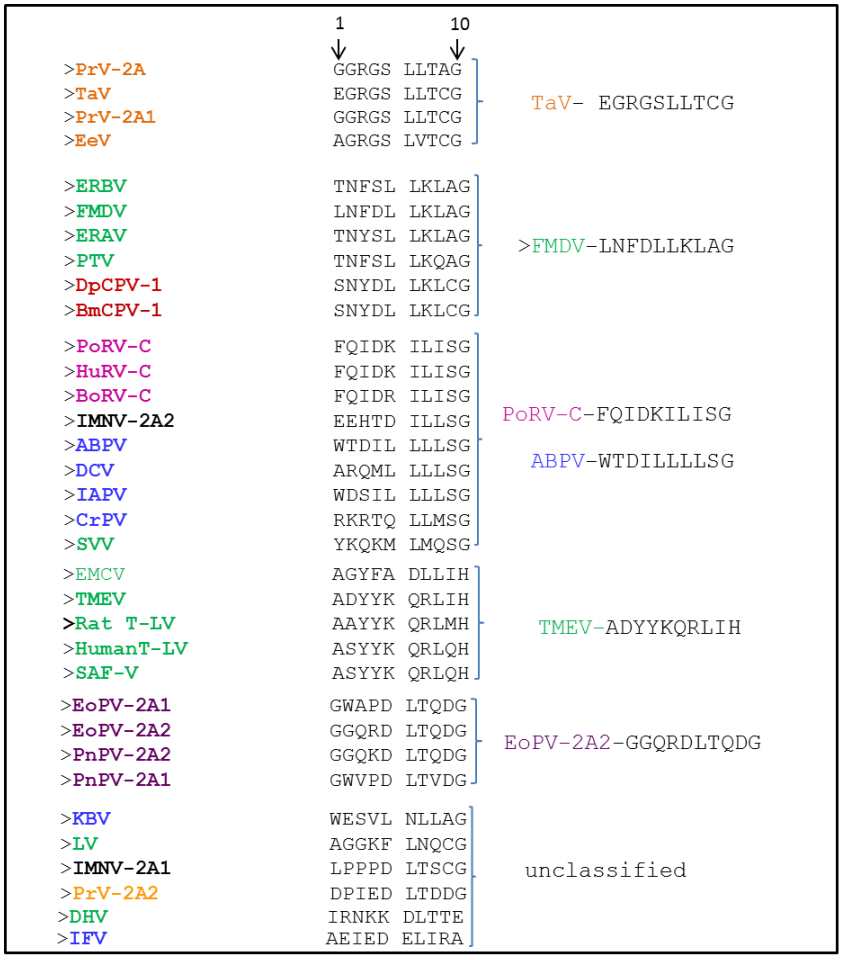


Figure 4.17: Subjective grouping of 2A based on sequence similarities of their important region

For clarity, the consensus motifs were omitted and the residues spanning the important region (preceding the consensus motif) were manually grouped by sequence similarity based on the five C-terminal residues. Each family of virus was assigned a colour. Tetravirus members are shown in orange, picornavirus in green, cypovirus in red, rotavirus in pink,

totivirus in black, cripavirus in blue and iflavirus in violet. This alignment showed high level of conservation of 2A residues within a virus family. On the right, a motif was selected to represent a group of 2A. Based on this study, six groups of 2A are proposed. This does not account for unclassified 2As, grouped in a category at the bottom of the figure.

By sorting the 2A sequences according to the primary sequence of their important regions, it was found that the majority could be aligned and produce six groups (figure 4.17). The group 1 model was TaV [GRGSLLTCGDVEENPGP] which has been the object of this body of work.
The group 2 of 2A could be modelled by FMDV 2A. This group was characterised by threonine or serine in position 1, a completely conserved asparagine in position 2, an amino acid with an aromatic ring phenylalanine or tyrosine in position 3, a hydrophilic residue aspartic acid, glutamic acid or serine followed by completely conserved residues leucine, lysine and leucine position 5, 6 and 7. The next residue was either alanine or cysteine and the last residue of the important region of the upstream context was glycine.
The group 3 and 4 shared the key motif from residue 6 to 10: [(I/L)LISG] and was mainly hydrophobic. Group 3 can be modelled by PoRV-C [FQIDKILISG] and the group 4 modelled by ABPV [WTDILLLLSG].
The group 5 modelled by TMEV consisting of cardioviruses had a highly conserved upstream context: [ADYYKQRLIH] and employed histidine instead of glycine in position 10.
Iflaviruses 2As were categorised separately in group 6. These sequences had no propensity for helix formation. The model selected was EoPV [GGQRDLTQDG].

This categorisation serves the purpose of reducing the population of 2A sequences to several models. This could be the starting material to future mutagenesis and a way to compare and measure the activity of each variable region and understand the contribution of each amino acid on the efficiency of the stalling. The numbers of groups may evolve according to the understanding of the mechanism of action.

Ryan and colleagues (1999) proposed that FMDV 2A activity was directly linked to the α-helical propensity of the peptide. The author proposed a model of action where the α-helix and the tight turn introduced by the [NPG] residues would re-orient the carboxy moiety of P site Gly-tRNA. In this model the α-helix is a critical contributing factor to stalling. It has also been suggested that every 2A peptide adopts the same mechanism of action (Donnelly et al., 2001).
The table 4.1 summarises the prediction for α-helix formation in solution for the model 2As. The ribosome exit tunnel is not a homogenous environment (this is further elaborated in the discussion section). The prediction presented in table 4.1 can not be extrapolated to infer the conformation the 2A elements would adopt in the ribosome tunnel. It is not known if FMDV 2A is helical in the exit tunnel. It is equally not known what conformation nascent peptides adopt in the upper exit tunnel.

The prediction in table 4.1 shows that the 2A sequences do not have the same propensities to form α-helices. FMDV 2A is predicted to be α-helical but DHV and TaV 2As, however are not. If the α-helical propensity is a critical element of 2A activity, then DHV and TaV 2As may not share a similar intra-ribosomal mechanism. This does not rule out the hypothesis that the [NPG] motif contributes to the stalling activity, but it implies that the α-helical hypothesis may not explain the activity for every 2A.

Table 4.1 α-helical propensity of the model 2As

The regions in the 2A sequences predicted by Jpred3 to form α-helix are edited in blue, the Jpred3 score representing the reliability of the prediction accuracy is given underneath and ranges from 0 to 9, higher score corresponds to a higher confidence in the prediction. The prediction used the thirty amino acids sequences presented in this table.

Group and model 2A	2A sequence and α-helix propensity
Group 1 TaV 2A	RGPRPQNLGVRAEGRGSLLTCG**DVEENPGP** No α-helix predicted
Group 2 FMDV 2A	RHKEDCAPVKQLLNFDLLKLAG**DVESNPGP** 566323899885
Group 3 PoRV 2A	GNGNPLIVANAKFQIDKILISG**DVELNPGP** No α-helix predicted
Group 4 ABPV 2A	TGFLNKLYHCGSWT**DIL**LLLSG**DVETNPGP** 87367 5777664
Group 5 TMEV 2A	FREFFKAVRGYHADYYKQRLIH**DVEMNPGP** 9058999986000699999998
Group 6 EoP V 2A	GQRTTEQIVTAQGWAPDLTQDG**DVESNPGP** 344
Relevant unclassified 2A DHV 2A	AFELNLEIESDQIRNKKDLTTEGVEPNPGP No α-helix predicted

4.4 Discussion

In this section, 2A activity was subjected to site-directed mutagenesis and truncation. The activity of 2A was assessed by its ability to mediate the production of two proteins from a single ORF. Scanning mutagenesis identified a hierarchy of importance for the residues comprising each sequence.

All the [D(V/I)E xNPGP] are interchangeable in TaV 2A, in other words, the [DV/IExNPGP] consensus motif allows for flexibility in the upstream context. For the [DV/IExNPGP] consensus motif, the exact spacing between [NPG] motif (which is an absolute requirement) and the [GDV/I] motif is critical. There is flexibility allowed in the [GDV/I] motif, but the resulting consensus sequences are not as efficient as [GDV/I] motif in the context of TaV. Other consensus sequences [CIExNPGP] and [EGVxNPGP] have stricter requirements from their upstream context to be highly efficient.

The minimal length required to maintain wild type activity is twenty amino acid for DHV 2A, twenty-three for IFV and seventeen aa for TaV 2A. It is however not possible to ascertain the wild type efficiency of FMDV 2A. The various 2A sequences tested exhibit different efficiencies. TaV and DHV 2As are highly efficient but FMDV 2A is sub-optimal.

The alanine scan mutagenesis identified an ‘important region’. It consists of the C-terminus of the upstream context adjacent to the consensus motif and stretches over five residues for FMDV 2A and nine residues for TaV and DHV. Alignment for FMDV and DHV sequences prove the residues in this region are highly conserved. This finding was extrapolated to other 2A sequences. The region preceding the consensus sequence (ten residues in total) was aligned, since this region is highly conserved within a family of virus and it was possible to categorise the viral 2A into six subjective groups and an unclassified seventh group. Twenty-eight out of thirty-four representative 2A sequences could be organised into six models. Some have predicted helical propensities in solution, others not, surprisingly the most efficient 2As, TaV and DHV do not.

There are two mutagenesis profiles. DHV and TaV 2A are highly efficient. Any point mutations across their variable region alter, but do not result in complete abolition of activity.
FMDV 2A however tolerates substitutions at all amino acids except for three specific residues Leu13, Leu15 and Gly17; for these residues substitution results in completely inactive sequences.
The results from the double proline mutations imply that in active 2As, the important region has a spatial relationship to the exit tunnel and promote a mechanism of action where the residues that participate in the activity are positioned to interact with their partnering residue or nucleotide.

The mutations performed here are consistent with previously reported point mutations for FMDV 2A. [DLLNFLLKLAGDVESHPG/P] is 35% and [DLLNFLLKLAGDVESQPG/P] is only 17% efficient compared to the wild type sequence which is 65% efficient (Heras et al., 2006). This is consistent with the result presented here: asparagine is an important residue and strongly influences efficiency of stalling. [DLLNFLLKLAGDVESNLG/P] is not active (Luke et al., 2008), this result is consistent with the Pro18 substitution and suggests that proline is critical.

Computer simulations of FMDV 2A structure predicted an α-helix, a tight turn at the [NPG] motif and an i-i+4 main-chain main-chain interaction that would stabilise the 2A and create the conformation that would reorient peptidyl-tRNA (Ryan et al., 1999). All prediction is created using parameters defined for protein in solution and in a uniform environment. The fact that [NPG] substitutions abolished 2A activity suggests that the tight turn predicted might be originating the predicted re-orientation of the peptidyl-tRNA. There is at the moment no available data regarding the conformation nascent peptides take in the upper tunnel. How to explain the mechanism of action of 2A?

There are three main points to discuss:

(i) interaction of 2A with the ribosome
(ii) for FMDV 2A, the efficiency of stalling increases with the length of the upstream context. Double proline mutations proved that residues outside the 'important region' may employ an additional mechanism.
(iii) some residues are conserved when they are not necessary

Interaction with the exit tunnel

There are several lines of evidence to prove 2A interacts with the exit tunnel. The most important one is that localised substitutions limit activity. The nature of the residue in these positions confers 2A efficiency: Arg5, Leu8 for TaV, [GDV] versus [EGV] in the 2A consensus motif, Leu13, Leu15 and Gly17 for FMDV 2A and the residues preceding [NPG] are not required. Another proof is that introducing double proline in the important region result in no activity, a result that is not observed outside this area. Although the hypothesis of no interaction to the tunnel can not be ruled out until cryoEM resolution and ribosome mutagenesis are carried out, this hypothesis seems very unlikely. If the question is to be asked differently, then it will become, what stabilizes the peptide long enough for stalling, rescue and re-initiation to occur if the ribosome is not involved?

There could be two options: 2A adopts a stabilised helical structure (or perhaps a compacted structure) jamming the tunnel or an extended conformation interacting with partnering residues in the tunnel halting translation. Either of these will not be possible without the involvement of the ribosome. The rate of integration of amino acids to the growing chain varies with the nature of the residue in the A and P site but the ribosome can process lysine at a rate of 100 s^{-1} (Wohlgemuth et al., 2008). The upper tunnel is not a uniform environment, it has bends (Nissen et al., 2000) and a varying electrostatic environment (Lu et al., 2007), and to induce stalling will most likely require stabilisation by the ribosome to override the fluctuations. In the translating ribosome, the important region and the consensus motif of the 2A sequence would be localised in the upper tunnel, in the area of the exit tunnel considered to be flexible. By comparing ribosomes in various states for four bacteria, it was discovered that there are flexible areas within the exit tunnel. The most flexible areas are the regions adjacent to the PTC and the constriction point where the proteins L22 and L4 protrude in the tunnel lumen, in that several nucleotides (especially A2062) or protein residues are seen flat against the tunnel wall or protruding in the lumen. The last region of flexibility consists of a small group of residues at the exit port (Fulle and Gohlke, 2009).

Could 2A fit in the models elaborated for other stalling peptides? As detailed in the introduction, nascent peptides that stall elongation exist both for prokaryotes and eukaryotes. Their primary sequence is diverse. Extensive site directed mutagenesis were carried out and uncovered important residues at their C-terminus. The length responsible for the stalling activity varies from nine residues for *ermC* (Tenson et al., 2003) to twenty-one residues for fungal arginine attenuator peptide (AAP) (Lovett and Rogers, 1996). In every case, the stalling involved an interaction between the important residues and the ribosome exit tunnel. For SecM, Yap and Bernstein (2011) concluded that the critical residues were facilitated in their placement by neighbouring amino acids. For all nascent stalling sequences, mutational analyses of the ribosome exit tunnel coupled with cryoEM analysis uncovered nucleotides in the vicinity of the PTC and residues in the protruding loops of protein L17/L22 and L4 to be involved in stalling. Most importantly cryoEM of the nascent stalling sequences showed multiple contacts with the tunnel. *TnaC* nascent peptide contacts: A751, U2609, U2602, L4 and L17 (Cruz-Vera et al., 2005). AAP nascent peptide contacts U2585, G2061, A2062, A2058, A2059, and U2609 (*E. coli* numbering) (Bhushan et al., 2010). Nascent SecM interacts with A2062, A2058, L22, A2503 and A751 (Nakatogawa and Ito, 2002). *ermC* contacts A2062, A2503, L4 and A2058 (Vasquez-Laslop et al., 2008). The authors all agree on a signalling relay between the nascent peptide and the ribosomal nucleotides leading to a compromised PTC. Two possible routes of signalling *via* the nucleotide A2058 (Vasquez-Laslop, 2010) or via the nucleotide A751 (Martinez et al., 2012) have been proposed. The relay mechanism is at present unknown and SecM is thought to be sensed by the ribosome at two locations in the vicinity of the PTC and at the L22 loop (Bhushan et al., 2011).

The folding zone of the ribosome exit tunnel (corresponding to Fulle and Gohlke (2009) flexible zone, corresponding to the upper tunnel) can accommodate α-helices (Lu and Deutsch, 2010). Lim and Spirin (1986) theorised that nascent peptides would adopt an α-helical conformation when growing through the exit tunnel. The nascent stalling peptides such as AAP and SecM have a compacted formation in the tunnel in the vicinity of the PTC (Woolhead et al., 2006 and Bhushan et al., 2011). *TnaC* however has an extended conformation throughout the tunnel (Seidelt et al., 2009). CryoEM studies proved that polypeptides could adopt an extended or a compacted conformation and even a helical structure. If a nascent peptide was helical, there would be sixty amino acids protected at any time by the ribosome, and if the structure was completely extended, this number will fall to twenty-five residues only. Very early studies established by proteolysis of nascent chains that thirty to forty amino acids are protected by the eukaryotic ribosome (Malkin and Rich, 1967 and Blobel and Sabatini, 1970). The logical conclusion was therefore that peptides would adopt partly compacted and partly extended conformation in transit. Woolhead and co-workers (2004) used FRET analysis to demonstrate that a transmembrane segment (α-helical) as opposed to a control peptide (without that propensity) adopted a compact conformation which they localised at the upper tunnel (near the PTC). Lu and Deutsch (2010) also demonstrated α-helices in the eukaryotic tunnel. They used blocks of five, ten and fifteen alanines at different positions in a growing peptide. The inserted a single cysteine residue at the N-terminus of the peptide and measured the ability of this residue to interact with maleimide. The reactant chosen for this experiment was too large to enter the tunnel, and the reaction could only occur if the cysteine residue was exposed in the solution. They found that α-helices could be located in the upper tunnel and the exit port, and failed to observe any helices for control peptides. A third study also identified α-helix using cryoEM (Bhushan et al., 2010b). They inserted five times the sequence [EAAAK] successively at two locations in a growing peptide. The rationale for their choice was that the alanine motif would form an α-helix and the glutamic acid and lysine would stabilise the structure with a salt bridge. Their sequence was also known to prefer a helical conformation in solution. They found α-helix in the lower tunnel, and in the upper tunnel but not in the central region. Collectively, these results prove that α-helix can be accommodated and stabilised in the tunnel. But none of these studies report a helical conformation for a peptide that does not have a helical propensity. And the result obtained for the stalling sequence *TnaC* prove that stalling does not necessarily require a helical conformation. The literature tends to suggest that a peptide with helical propensity may well adopt it as early as the upper tunnel, and for that reason the hypothesis that FMDV 2A helical propensity plays a part in its activity, can not be ruled out.

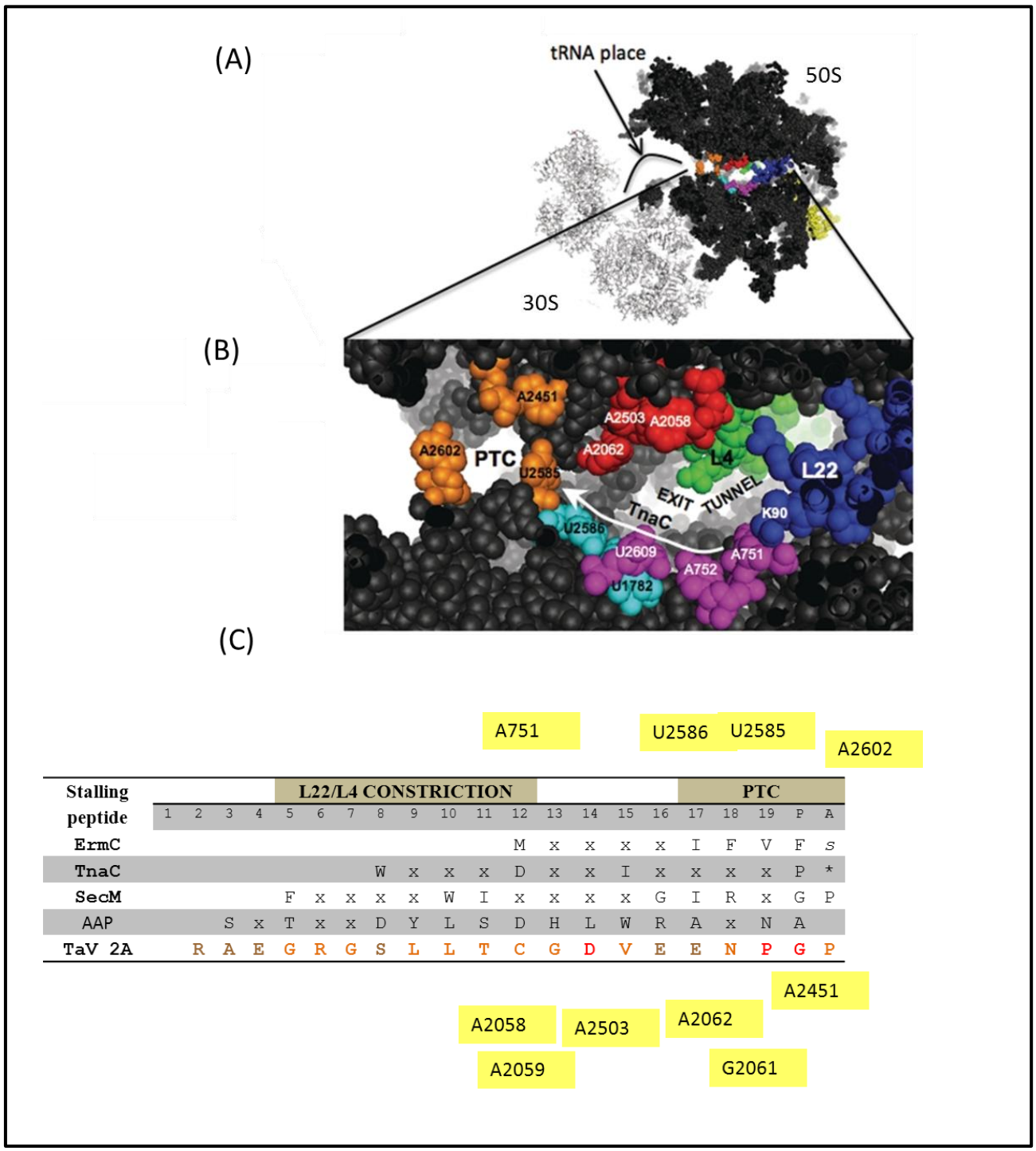
What would explain the activity of TaV and DHV? The results prove TaV 2A adopts a localised conformation, involving interactions with the tunnel elements. From the mutagenesis, Leu8 side chains and Arg5 are important for TaV 2A. DHV 2A has a hydrophilic context, and outside the important leucine residue, has especially an arginine and also two lysines in its important region. A

possibility is that DHV 2A positively charged context contribute to pausing via electrostatic interactions. Several studies demonstrate the ability of positively charged residues to interact and slow and even pause the elongating ribosome.

Dimitrova and colleagues (2008) inserted twelve consecutive lysines between two reporter proteins and as a control inserted a stem loop instead of the poly-Lysine. Employing a yeast system, they found that poly-lysine caused translation arrest, the downstream peptide was not being synthesised. The mRNA was also degraded. Insertion of a stem loop did not result in mRNA degradation. They concluded that lysine positive charge could interact with the rRNA from the tunnel, a result which mimics poly-A tail at the end of eukaryotic mRNA. The poly-A tail (eighty or so adenines) regardless of reading frame encodes for poly-lysine and serves the purpose of stalling the ribosome which has read through their termination codon (Akimitsu et al., 2007). In a recent study (Charneski ad Hurst, 2013), it was reported that charged amino acids regulate the ribosome velocity and the authors argue that in their studies they found no evidence that codon usage and mRNA structure would alter the rate of elongation. The study consisted of a re-analysis of experimental data collected using ribosomal footprinting of *Saccharomyces cerevisiae* transcriptome. The data set was generated by obtaining, reverse transcribing, sequencing and identifying the sequences of the mRNA protected by the ribosome. The analysis repeated at time intervals provides the total occupancy for the ribosomes at various times points and therefore allows by derivation to assess the velocity of the ribosome. It was found that a single positive charge within thirty amino acids without any other positively charged residue could slow the ribosome. This ability was not linked to codon usage, or the structure of the mRNA and they established that histidine caused a weaker slowing compared to arginine and lysine. The authors concluded that positive charge slowing of the translating ribosome is a normal feature of translation. A noteworthy detail is that ribosomal protein L22 residue Lys90 is involved in stalling of SecM (Nakatogawa and Ito, 2002). Based on the literature, the logical conclusion is that one mechanism involved in DHV and TaV 2A stalling is a slowing of translation induced by positively charged residues. TaV2A Arg5 may serve also an additional function. Its substitution for lysine resulted in less activity. This amino acid would reside in the area of the tunnel constricted by proteins L17 and L4. The hypothesis is that Arg5 may interact with a residue from one of these proteins.

Out of the six identified 2A groups, each have a lysine or an arginine residue except for the ABPV model (sequence is: WTDILLLLSG) which is leucine rich. 2A without helical propensity can rely on the positive charge forming weak interaction with the rRNA as it egresses through the exit tunnel. Leucine residues may orient favourably the sequence in the tunnel.

The following figure (4.18) provides an illustration for the upper tunnel and the table underneath provides a hypothetical interpretation for the location of TaV 2A residues in the tunnel.



Stalling peptide	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	P	A
					L22/L4 CONSTRICTION												PTC				
ErmC												M	x	x	x	x	I	F	V	F	*s*
TnaC								W	x	x	x	D	x	x	I	x	x	x	x	P	*
SecM					F	x	x	x	x	W	I	x	x	x	x	G	I	R	x	G	P
AAP			S	x	T	x	x	D	Y	L	S	D	H	L	W	R	A	x	N	A	
TaV 2A		R	A	E	G	R	G	S	L	L	T	C	G	D	V	E	E	N	P	G	P

Figure 4.18: Illustration of the ribosome exit tunnel and interpretative table showing TaV 2A and other stalling peptides

(A) an illustration of a bacterial ribosome 50S and 30S subunits and the position of the tunnel in the 50S subunit. **(B)** Amplification of the first part of the exit tunnel. Several nucleotides in the wall of the tunnel and at the PTC are represented as well as the proteins L22 (blue) and L4 (green). The PTC nucleotides are represented in orange. Red residues are involved in SecM and *ermC*. Nucleotides in pink and residue K90 from the protein L22 are critical for SecM and *TnaC*. The white arrow indicated the possible relay to the PTC (via the nucleotides represented in cyan) for TnaC. In **(C)** the table provide an interpretative residue by residue alignment showing their approximate locations in the ribosome (top line highlighted in grey, ribosomal residues in yellow boxes). The table shows the critical residues for stalling. x represents residues that can be

substituted. The residue in lower case and in italic at the A site of *ermC* is not required for the arrest. * represents the UGA stop codon for TnaC. The last line provides the hypothetical organisation of TaV2A in the tunnel the colouring for TaV residues shows their importance: in brown partially required, orange important and red essential (Adapted from Martinez et al., 2012 and Ito et al., 2010).

The relay that other nascent peptides engaged in is thought to involve the residues communicating with the PTC. It is critical for peptide bond formation that the two tRNA substrates are positioned correctly (Schmeing et al., 2005), and the suggested way that this activity is impaired is through a disturbance of the position of critical elements as subtle as a hydrogen bond which takes part in the positioning of the P site tRNA. The folding of the 23/28 S rRNA domain V involves base pairing and base stacking (Kim and Green, 1999) and it is reasonable to envisage that a simple disturbance in this arrangement might be the cause for altered PTC geometry.

Length and double proline mutations.

From early work and as already cited numerous times, the length of the FMDV 2A peptide is proportional to its efficiency. The longer it is the more efficiently it stalls. Recent work in this laboratory proved that modification of the N terminus of TaV, both single point mutations and insertion and deletions of six residues in the region preceding TaV 2A does not alter in any way TaV 2A ability to stall the ribosome. This study was conducted without taking into account that positively charged residues can slow the ribosome, as a result it is not possible to rule out the involvement of the lower tunnel in TaV 2A activity (Minskaia, manuscript in preparation). Additional deletions to the six groups of 2A identified will greatly elucidate what role the lower tunnel plays in the mechanism of action. The lower tunnel has a larger diameter (25Å) and allows potentially more freedom of movement to the growing peptide. One hypothesis is that the N-terminus in the lower tunnel adds to the stability of the structure/ conformation and providing that the sequence supports that structure/conformation, the 2A will be efficient. Another hypothesis is that the lower tunnel is also imbued with sensing abilities like the upper tunnel and intereacts with the growing peptide. The argument in favour of an extra sensing mechanism is that introducing double proline in the region preceding the important region for different 2As did not alter the stalling ability. The residues involved were located eighteen (for TaV 2A) to twenty amino acids (for FMDV 2A) away from the Gly-Pro pair, it is likely that the residues mutated would be at the very least located in the constriction area or perhaps even further depending on the structure the 2A adopts. If the lower tunnel only stabilised the 2A, double Pro would have decreased the activity of FMDV 2A.

The N-terminal half of FMDV 2A sequence is very variable but it is noticeable (figure 4.12) that some positively charged residues are favoured. In position 1 histidine and arginine are preferred, the

next amino acid is a completely conserved lysine, and eight amino acids downstream there is also a completely conserved lysine (Lys10).

Conservation of residues not necessary to 2A activity.
What is puzzling is that residues that were believed to be the signature motif for 2A: [DV/IE] are shown to be replaceable by other residues. [GGV], [AGV], [DGV] and [ICV]. However, in FMDV (sub-optimal) these residues are not mutated, and Glu is completely conserved in all viral sequences when it could be replaced by alanine, why? In fact several residues not required for activity are still conserved. The [NFD] (residues 14 to 16) motif of FMDV 2A, a large part of the N-terminal region of DHV too not necessary for activity (since it could be truncated and not affect stalling efficiency) are completely conserved.

Aphthoviruses share the FMDV type of 2A. Similarly, TaV and related viruses PV and EeV share the same 2A sequence. The result is consistent with previous analyses (Luke et al, 2008). A virus finds itself with a certain type of 2A by phylogenetic relationship. There are no available data correlating the level of activity of 2A with the virus strategies and fitness. Considering the high level of conservation of the sequences, the viruses most likely have a vital reason for maintaining this precise sequence and outside the observation that FMDV optimised its codon usage to adapt to that of its host (Zhou et al., 2000), it is very likely that FMDV has utility for a sub-optimal 2A and TaV for an optimum 2A. The reason is unknown at present.

In summary, the highly efficient 2A makes contacts/ interaction with the ribosome tunnel elements. The strength or the stability of this interaction dictates the level of efficiency of the 2A sequence and dictates plasticity in the consensus sequence. 2A with less efficacies have limited abilities to either interact or stabilise that interaction.

In this chapter it was shown that the 2A sequences could be reduced to several models but there is at least two mechanisms by which 2A induces the ribosome stalling, to uncover the 2A rule, future steps could involve trimming the model sequences to twenty amino acids and measure their activity paying particular attention to the role charges could play in their activity and carrying out non-conservative mutagenesis. Understanding the 2A-induced ribosome stalling would bring a further insight into the interaction between nascent peptides and ribosome exit tunnel.

Chapter 5 Inhibition of peptide bond formation

5.1 Introduction

The last chapter explored the possible interaction between the nascent peptide in the exit tunnel and 2A-induced stalling. In this chapter the question adressed is: why is the peptide bond not made, in other words, why is the peptide not transferred to the A site tRNA?

For all other stalling peptides listed to date, the PTC geometry is altered and the most straightforward way to test for this is to introduce puromycin in the ribosome A site. Puromycin is a tRNA analogue and has been employed and modified extensively in ribosomal studies (Pestka et al., 1973). It is particularly useful because it circumvents the need for the accommodation stage and reacts to prokaryotic and eukaryotic ribosomes. Puromycin reactivity to the peptide drops when the PTC geometry is altered.

The abundance of a tRNA species has been demonstrated to play a role in the rate of elongation. The choice of codons may be another contributing factor to 2A-induced stalling. The hypothesis was that rarer codons would amplify 2A activity since the ribosome would have to stall longer to accommodate a rarer species of tRNA. For example, Woolstenhulme and colleagues (2013) identified stalling motifs related to cluster of rare codons in *E. coli*. For this experiment FMDV 2A coding sequence was inserted between luciferase and renilla genes. The C-terminal proline codon (CCC) was mutated to the synonymous codons CCG, CCA and CCT. In addition the following controls were included: a stop or a sense codon between the two genes. The experiment aimed at measuring the levels of renilla and luciferase activity using the Dual-Glo system (Promega) in bovine lung cells transiently transfected.

An unexpected result was that the Pro20 mutant to Ala (figure 4.6) of TaV 2A still retained good stalling activity. The model for 2A activity explained the stalling with the poor reactivities of glycine and proline. The experiment carried out was to substitute theTaV 2A A site proline for glycine, valine, serine, threonine and cysteine.

The *Pichia pastoris* system enables the overexpression of milligrammes of heterologous proteins. The pPinkα-HC vector has the secretion signal α-mating factor from *Saccharomyces cerevisiae* that enables secretion of the protein of interest in the media. The *P. pastoris* PichiaPink system from Invitrogen is designed for high level production of recombinant proteins. The expression system relies on the strong inducible *AOX1* promoter in the presence of methanol. For the purpose of this study, it was decided to use the PichiaPink strain 4 (genotype: *ade2, prb1, pep4*). *pep4* knockout prevents the strain from synthesising proteinase A and *prb1* knockout, proteinase B. *ade2* auxotroph strains can not grow without adenine and purine precursors accumulate in vacuoles, the colonies appear small and red. In the pPinkα-HC vector the ADE2 gene is driven by a truncated 13bp ADE2 promoter. ADE2 encodes phosphoribosylaminoimidazole carboxylase involved in the *de novo* synthesis of purines. The successful transformation of the pichiaPink strain with a copy of the ADE2 gene allows the strain to thrive in adenine-deprived media and appear white. Selection of transformants is based on colours. Very white colonies have integrated several copies of the plasmid and potentially indicate high level of expression. The strategy was to secrete 2A for future NMR resolution and comparison of their structure with inactive 2As. This first attempt at expressing 2A in the *P. pastoris* system was carried out with TaV 2A.

5.2 Materials and methods

Plasmid pSTA1 (coding for GFP-2A-GUS) and pJC3 (coding for GFP-2A-CherryFP) were modified. Primers were designed to introduce a mutation in at a specific residue (induced mutation was highlighted in blue in the forward primers sequences in this section). Alternatively blunt-ended DNA fragments were created for puromycin experiment. PCR were performed using the KOD Hot start enzyme from Novagen, following manufacturer instructions. The protocols utilised were detailed in sections 2.2.1 to 7. The DNA were gel purified using Wizard SV gel clean up kit (Promega) and vectors were extracted using Qiagen spin miniprep kit. The integrity of the mutants was assessed by big Dye terminator sequencing (ABI) outsourced from Dundee sequencing services. *In vitro* translation were performed in 12,5 µl Rabbit reticulocytes supplemented with [^{35}S]-Met and incubated for ninety minutes. The translation products were resolved on 4-20 % tris-bis SDS-PAGE denaturating gels (Expedeon). The gels were Coomassie-stained and [^{35}S]-labelled proteins were visualised onto photographic film and size determined against the dual colour marker precision standard from Biorad.

5.2.1 Puromycin test

To generate the blunt ended PCR fragments, the PCR templates were pSTA1 TaV 2A and pSTA1 TaV 2A mutant G12A, which is inactive. The PCRs amplified the GFP-2A (samples) and the GFP segment terminated before TaV 2A (negative control). The translations were carried out in rabbit reticulocyte lysates using 1 µl of purified and sequenced PCR fragments, [^{35}S]-Met, in a total volume of 25 µl. The primers were designed so that two alternative translated C-terminus were tested for the active TaV 2A : [NPG] and [NPGP]. The inactive TaV mutant fragment tested ended with [NPG]. Forward primer pSTA1 beg T7 Fw: 5'-CTGGCTTATCGAAATTAATACGAC-3' and reverses:
NPG Rev : 5'-CCCGGGATTTTCCTCCACGTC-3'
NPGP rev: 5'-GGGCCCGGGATTTTCCTCCACGTC-3'
GFP Rev2: 5'-CCCGGACTTGTATAGTTCGTC-3'

After translation for 30 min, puromycin was added, incubated for 1 min, the reactions were stopped on ice, a 10 µl aliquot of each samples was treated with 5 U of RNAse A/T1 and incubated at 37 ^{o}C for 30 min. the samples were then resolved on 4-20 % SDS-PAGE.

5.2.2 A site proline mutations

A dual renilla-2A-luciferase insert was created in the backbone of pSTA1 vector. Renilla was amplified to create a Hind3 site in the 5' region of the gene and a Xba 1site at the 3' end. The modified renilla was inserted in pSTA1 instead of GFP.

Primers:

Ren Hind3 f: 5'-AAGCTTATGACTTCGAAAGTTTATGATC-3'

Ren Xba1 rev2: 5'-TCTAGATTGTTCATTTTTGAGAACTC-3'

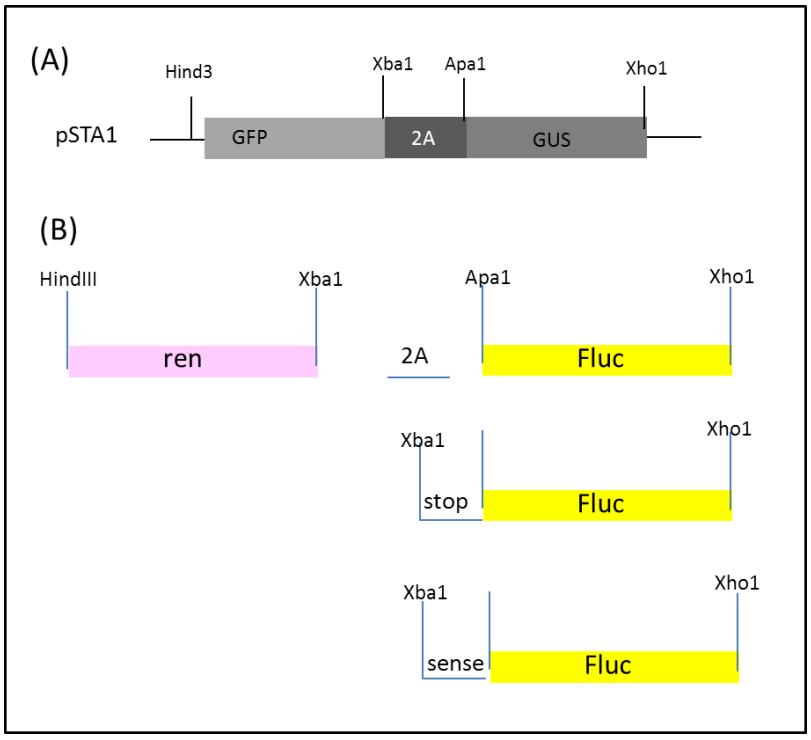


Figure 5.1: Details of the dual renilla-luciferase insert for expression in bovine cell line

In (A) the schema shows the cloning sites used to modified pSTA1 FMDV 2A with renilla and luciferase genes. In (B) Renilla (in pink) replaced GFP at the 5' end and luciferase (in yellow) replaced GUS at the 3' end of FMDV 2A. Two controls were also created where 2A was replaced by a stop codon TAA or the sense codon TAC.

The firefly luciferase gene was amplified to create a Xho1 site at the 3' end. The 5' end contained either an Apa1 site (for ligation to 2A sequence in pSTA1) or a Xba1 site (for ligation to the renilla sequence in pSTA1). Stop codon was TAA and sense codon was TAC coding for Tyr.

Forward primers:

Apa1Fluc F1: GGGCCCATGAAGAGATACGCCCTGGTTC

Xba1StopFluc F2:TCTAGATAAATGAAGAGATACGCCCTGGTTC

Xba1senseFluc F3: TCTAGATACATGAAGAGATACGCCCTGGTTC
Reverse primer, Fluc xho1 rev: 5'- CTCGAGTTACAATTTGGACTTTCCGCCC-3'

The A site proline (FMDV 2A Pro) was mutated to the synonymous codons: CCA, CCT and CCG using the following primers. The substitutions are in blue in the sequences underneath.
CCA Fw: 5'-GTCCAACCCCGGGCCAATGAAGAGATACGCCCTGG-3'
CCA R:5'-GCGTATCTCTTCATTGGCCCGGGGTTGGACTCGAC-3'
CCG Fw: 5'-GTCCAACCCCGGGCCGATGAAGAGATACGCCCTGG-3'
CCG R:5'-GCGTATCTCTTCATCGGCCCGGGGTTGGACTCGAC-3'
CCT Fw: 5'-GTCCAACCCCGGGCCTATGAAGAGATACGCCCTGG-3'
CCT R:5-GCGTATCTCTTCATAGGCCCGGGGTTGGACTCGAC-3'

Bovine lung cells were cultured and transfected as described in sections 2.2.11 and 12.

TaV 2A A site proline (Pro20) was mutated to glycine, serine, valine, threonine and cysteine. In the mutation primers underneath, the mutated nucleotides are written in blue.

mutant	Sequence 5'-3'	Primers names
TaV P20G	Fw: GGAAAATCCCGGGGGACACCACCACCACCACCAC R:GGTGGTGGTGGTGTCCCCCGGGATTTTCCTCCACG	TaV P20G F TaV P20G R
TaV P20S	Fw: GGAAAATCCCGGGAGCCACCACCACCACCACCAC R: GGTGGTGGTGGTGGCTCCCGGGATTTTCCTCCACG	TaV P20S F TaV P20S R
TaV P20V	Fw: GGAAAATCCCGGGGTACACCACCACCACCACCAC R: GGTGGTGGTGGTGTACCCCGGGATTTTCCTCCACG	TaV P20V F TaV P20V R
TaV P20T	Fw: GGAAAATCCCGGGACCCACCACCACCACCACCAC R: GGTGGTGGTGGTGGGTCCCGGGATTTTCCTCCACG	TaV P20T F TaV P20T R
TaV P20C	Fw: GGAAAATCCCGGGTGCCACCACCACCACCACCAC R: GGTGGTGGTGGTGGCACCCGGGATTTTCCTCCACG	TaV P20C F TaV P20C R

5.2.3 2A expression using the PichiaPink system.

TaV 2A was created from pJC3 plasmid (appendix 1) and inserted in the pPinkα-HC vector.

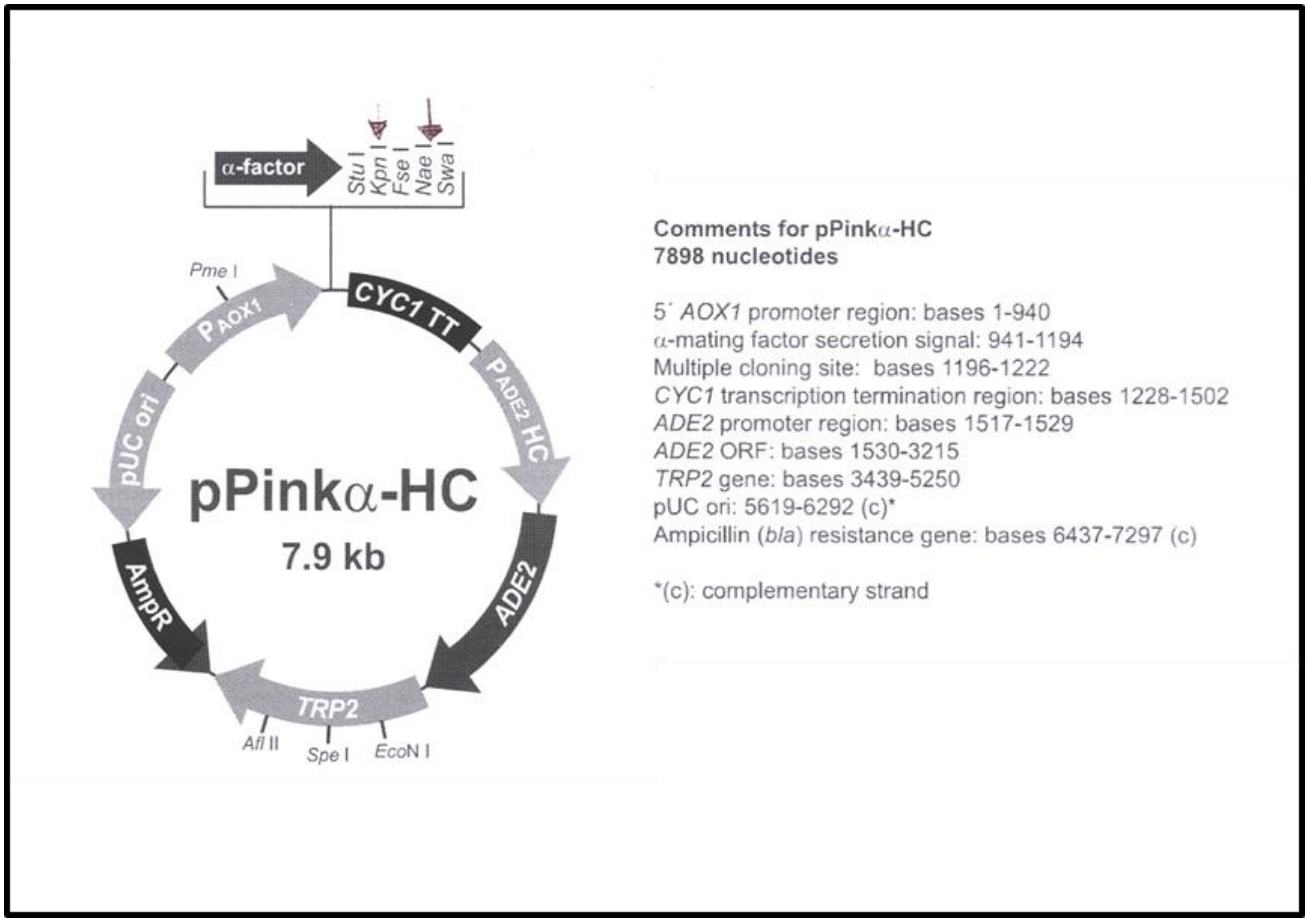


Figure 5.2: Map of pPinkα-HC

The vector contains the α-mating factor for secretion of the recombinant protein. The AOX1 promoter drives expression when induced by high levels of methanol. The vector also expresses the ADE2 gene for selection of transformants. (vector map sourced from http://www.lifetechnologies.com)

Table 5.1: Details of sequences forming the TaV 2A insert for expression in the PichiaPink system

The histidine tag was inserted for affinity purification; the TeV site for proteolytic cleavage of the purified 2A and 2A was inserted at the N-terminus of cherryFP. The insert was cloned in pPinkα vector *via* Xho1and Kpn1sites.

	motif	Nucleotides and amino acid sequences
Beginning of ORF	Xho1	**CTCGAG**
	α vector sequence	**AAAAGGCCT**
	Start codon and polyHis	**ATGGGCCATCATCATCATCATCATCATCAT** M G H H H H H H H H
	TeV cleavage sequence	**GAGAATAAATATTTTCAGGGGATGCAT** E N K Y F Q G M H
	Xba1+ TaV2A	**TCTAGAGGCTCCGGAGAGGGCAGGGGAAGTCTT** S R G S G E G R G S L **CTAACATGCGGGGACGTGGAGGAAAATCCCGGGCCC** L T C G D V E E N P G P
End of ORF	cherryFP	…… **GATGAATTGTACAAATAA**
	Kpn1	**GGTACC**

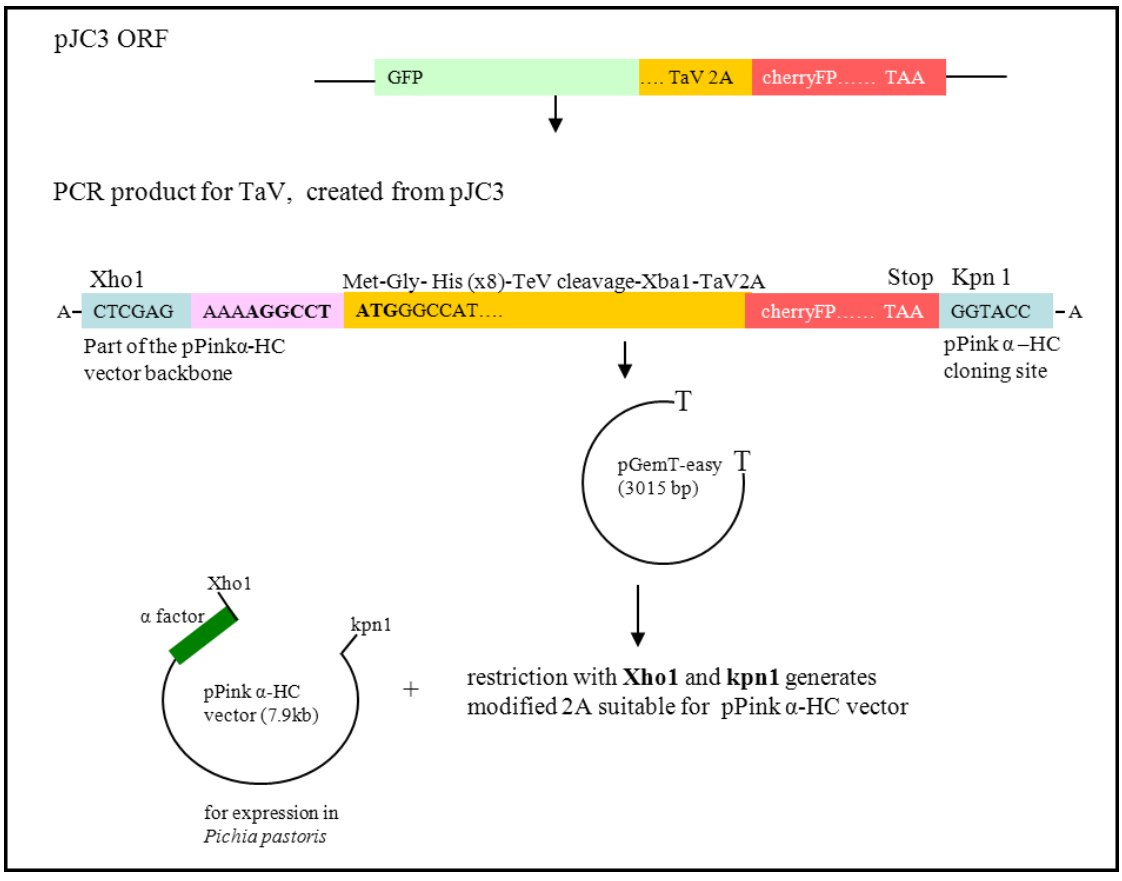


Figure 5.3: Creation of the TaV 2A insert for pPinkα vector

Cloning sites for pPinkα-HC, as well as polyhistidine for affinity purification and a protease site were created are the 5' end of the TaV 2A sequence, ligated and sequenced in pGemT-easy and further cloned in the final vector via Xho1 and Knp1 restriction sites.

Primers for cloning were,
Xho1 Pink αHTaV F : 5'-CTCGAGAAAAGGCCTATGGGCCATCATCATCAT-3'
TaV linker pich F: 5'-GGATCCATGGGCCATCATCATCATCATCATCATCATGAGAATAAAT ATTTTCAGGGGATGCATTCTAGAGGCTCCGGA-3'
and reverse Kpn1 chFP R: 5'- GGTACCTTATTTGTACAATTCATC-3'.The insert was created in 2 successive PCRs.

Forward primer used for sequencing was
pink seq F: 5'-ATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCT-3'.

Primers used for PCR analysis of transformants were
Forward, Pichia 2A ana F: 5'- ATGGGCCATCATCATCATCATC- 3' and
Reverse Pichia 2A ana R: 5'-TTATTTGTACAATTCATCCATG-3'
Expected proteins sizes after translation: cherryFP total amino acid number is 237, MW=26771 and MGHistag-Tev-TaV2A total amino acid number is 41, MW is 4575 Da.

P. pastoris media preparation and culture was carried out as described in section 2.2.10.
Briefly, competent yeast was electroporated with linearised pPinkα vector. They were then cultured on PAD media for several days until white colonies were formed. These were sub-cultured again on PAD. The transformants were grown in BMGY media containing glycerol for 24 h, the cells were collected by centrifugation and resuspended in the induction media BMMY which contained methanol. The cells and supernatant fractions were analysed by PCR, western blot and microscopy.

5.3 Results

5.3.1 Puromycin test revealed an altered PTC

The toe print analysis carried out by Doronina and colleagues (2008) showed that in the 2A-stalled complex, the peptide bond between the last glycine and the proline does not take place. The glycine residue remains the last residue added to the nascent chain before translation arrest. Puromycin is a tRNA analogue, is able to form peptide bond indiscriminatingly and is therefore able to release nascent peptides. It has been used extensively as an indirect assessment of the PTC geometry in studies involving other nascent stalling-peptides (Wei et al., 2012, Woolhead et al., 2006, Cruz-Vera et al., 2007). To determine whether 2A-induced ribosome stalling also features and altered PTC, several truncated PCR products were prepared and amplified from the GFP-2A-GUS insert of pSTA1, as depicted in figure 5.4. The DNA products were terminated after the last glycine (TaV NPG) or after the last proline (TaV NPGP). In addition, two controls were included in the experiment: an arrest-impaired sequence which had the G12A mutation and terminated at the last glycine (inactive TaV NPG) as well as an unrelated sequence to TaV 2A amplified from GFP.

Because the PCR products lacked a stop codon, ribosomes reaching the 3' end of the truncated mRNA should remain complexed to the mRNA with the peptidyl-tRNA in the P site. In the following experiment, the PCR products were added to the cell-free system and translated in presence of [^{35} S]-Met. After translation for 30 min, 1 µg of puromycin (an excess of puromycin) was added. The puromycin should then enter the A site and readily release the nascent peptide. However, if the PTC geometry is altered, the stalled complex is refractory to a reaction to puromycin. The difference between the two is observable by gel electrophoresis, as a refractory stalled complex generates an additional 'heavier' band corresponding to the nascent chain-tRNA. Treatment of this sample with RNAse should degrade the tRNA and release the nascent chain and only one band corresponding to the nascent chain molecular weight should then be visible on the gel. The molecular weights for the expected translated products are 27 kDa for GFP, 30 kDa for TaV NPG, Tav NPGP and inactive TaV NPG. tRNA adds 23 kDa to the nascent chain.

In figure 5.4, the results prove that TaV induces an altered geometry of the PTC. In lane 2 and 4 (TaV NPG and TaV NPGP) puromycin does not react with some of the stalled ribosomes, as a result a heavier band corresponding to the peptidyl-tRNA is visible. In lane 6 and 8 for the negative controls, however that extra band is not apparent; puromycin has reacted to all the stalled ribosomes. In lane 3, and 5 the treatment with RNAse resulted in the complete restriction of the tRNAs and release of the

nascent chains. TaV therefore alters the PTC such that the peptide bond is not made because one or several ribosomal elements in the PTC are not positioned correctly.

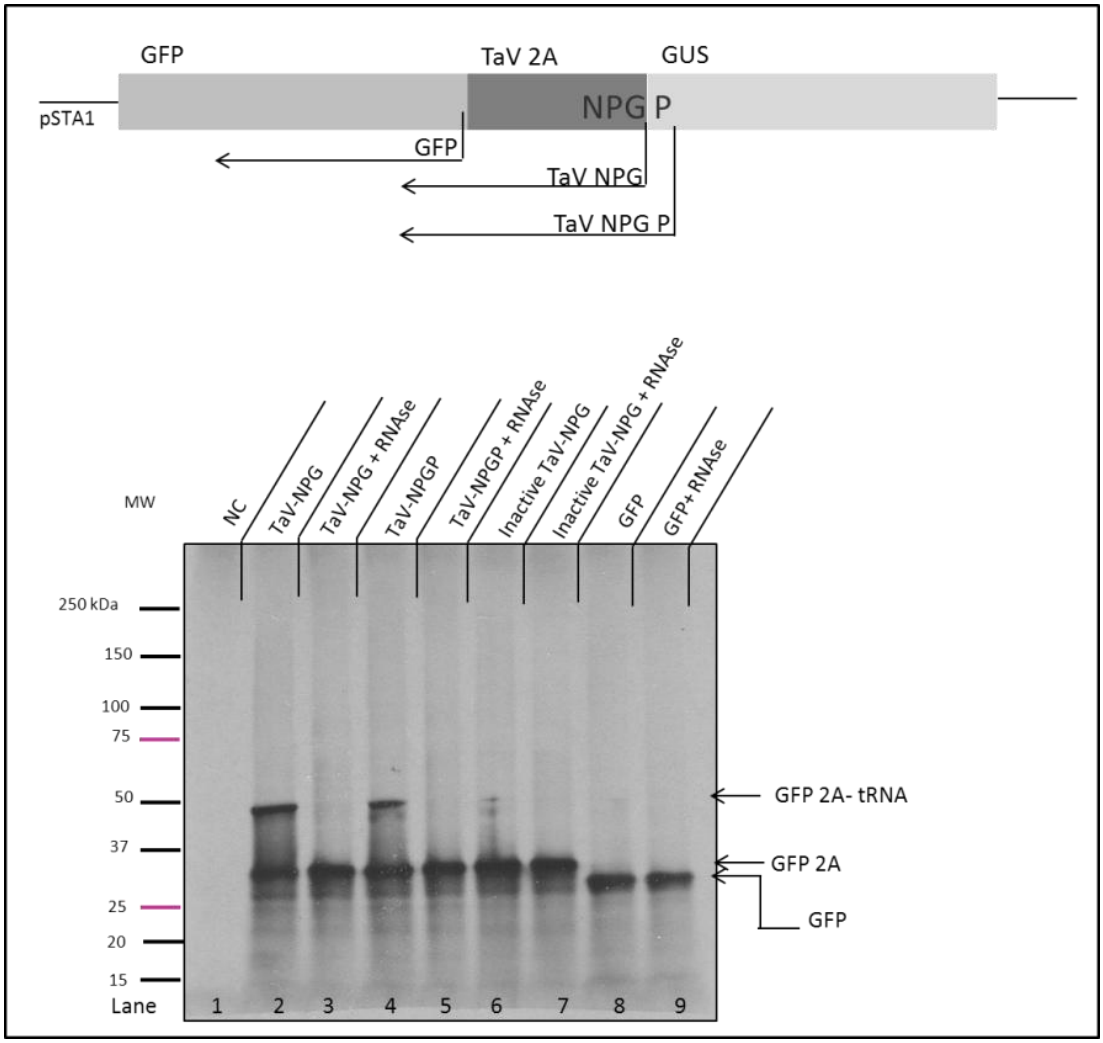


Figure 5.4: Puromycin test for TaV 2A activity

SDS-PAGE showing translation of truncated TaV 2A with 1 µM puromycin in rabbit reticulocyte lysates supplemented with [^{35}S]- Met. In lane 1, the translation assay without DNA.

In lane 2 and 3, TaV sequence was stopped at NPG and in lane 3 digested with RNAseA/T1 for 30 min.

In lane 4 and 5, TaV sequence was stopped at NPGP and in lane 5 digested with RNAseA/T1 for 30 min.

Negative controls included: In lane 6 and 7 mutant TaV (mutant inactive) sequence was truncated at NPG and in lane 7 digested with RNAseA/T1 for 30 min. Truncated GFP in lane 8 and 9, without (lane 8) and with digestion with RNAseA/T1.

5.3.2 2A renders the ribosome A site restrictive

In the previous chapter it was inadvertently discovered that mutation of Pro20 did not alleviate stalling. Previously it was thought that proline at that position was mandatory. TaV2A Pro20 substitutions with residues with smaller or slightly bigger molecular weight amino acids are presented: glycine (75 Da), serine (105 Da), valine (117) threonine (119 Da) and cysteine (121), proline (115Da).

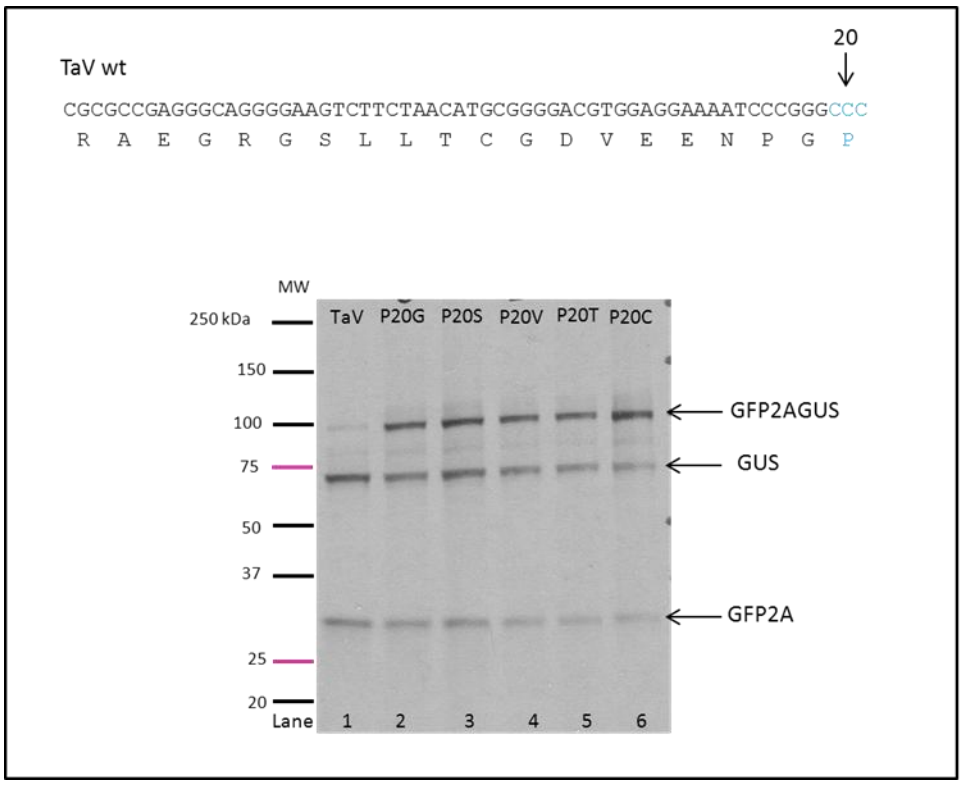


Figure 5.5: Mutation of Pro20 of pSTA1 TaV 2A

SDS-PAGE showing translation of TaV 2A and mutants in rabbit reticulocyte lysates supplemented with [^{35}S]-methionine. TaV 2A sequence is written above the gel and the black arrow and blue text show the residue mutated. The point mutation tested is indicated at the top of each lane, Pro20 is substituted for glycine (lane 2), serine (lane 3), valine (lane 4), threonine (lane 5) and cysteine (lane 6). The gel shows more full-length product with increasing amino acid MW.

The 2A sequence was able to mediate the production of discrete products with all the substitutions tested (figure 5.5). This resulted in more full-length products as opposed to wild type with increasing MW of the A site amino acid. These results demonstrated that the stalling ability of TaV 2A altered the geometry at the PTC so that peptide transfer to a range of amino acid is inhibited. It is observable however that proline added greatly to the efficiency of the stalling.

FMDV 2A coding sequence was inserted between luciferase and renilla genes. C-terminal proline codon (CCC) was mutated to the synonymous codons CCG, CCA and CCT, Unfortunately, test transfection with pJC3 (GFP-2A-CherryFP) to bovine lung cells using three different carriers: lipofectamine 2000, PEI and Fugene 6 were not successful. The results underneath showed the translation of the dual renilla-FMDV2A-Luciferase in rabbit reticulocyte lysates.

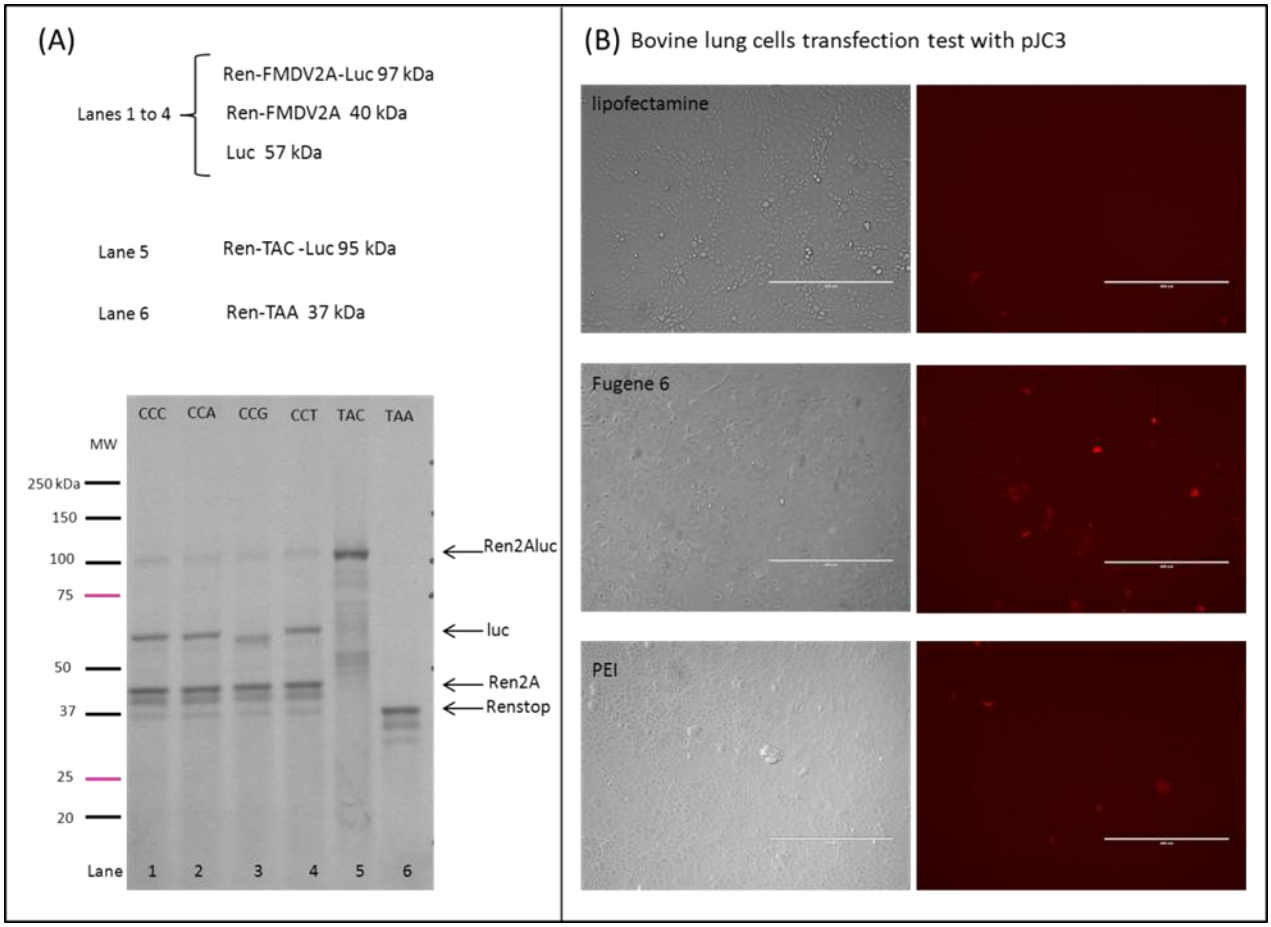


Figure 5.6: Effect of proline synonymous codons on activity of FMDV 2A

In (A) FMDV 2A last Pro (occupying the A site) was mutated to other synonymous codons, CCA, CCG and CCT (lanes 1 to 4). Negative controls are shown in lane 5 and 6. The text above gave the MW of the proteins expected and the reference lane for the gel underneath. The SDS-PAGE gel (panel (A) bottom) resolved samples translated in rabbit reticulocyte lysates supplemented with [^{35}S]-Met. Panel (B) showed poor cherryFP expression (on the right) of bovine lung cells transfected with different reagents: lipofectamine, Fugene 6 and PEI. The bright field pictures of confluent cells were provided on the left.

The expression in bovine cell lung (figure 5.6) did not allow correlating the level of translation of downstream luciferase to abundance of prolyl-tRNA isoforms, perhaps re-cloning the inserts in a lentivirus system may overcome the limitations of transient transfection relying on cationic lipids. Although, cell-free systems are optimised for efficient translation of any T7 driven inserts, the

translation profile would have reflected a problematic structure linked to synonymous codons. For all proline synonymous codons tested, the same result was obtained. The conclusion derived was that the mRNA structural pattern associated to codon identity did not influence the overall translation profile *in vitro*.

5.3.3 *P. pastoris* for resolution of 2A structure required further cloning

2A is active only in eukaryotic ribosomes. Because of its small size, 2A can be subjected to NMR analysis, however several milligrammes of the purified product may be required and for these reasons *P. pastoris* seemed an ideally suited expression system. TaV 2A was inserted in frame fused to the pre-pro α-mating factor. This latter consists of a nineteen amino acids signal sequence (pre- sequence) targeting the protein being expressed to the endoplasmic reticulum and of a sixty amino acids pro-sequence transported to the Golgi (figure 5.7). During translocation the signal peptidase cleaves the nineteen amino acid pre-sequence, later at the golgi the Kex2 protease recognises basic residues such as Lys-Arg or Arg-Arg and cleaves the pro-sequence from the downstream protein. To address the purification needs, a poly-histidine tag was cloned upstream of 2A followed by the commonly used protease site for TeV. Integration in the genome occurs via homologous recombination and results in multiple integrations. Because clones with multiple integrations express higher levels of proteins, the PichiaPink system was ideally suited since it allowed a visual assessment of multiple integrations by the colour of the transformants. White colonies thrived on the selection media, four white colonies were selected for further molecular analyses. The PCR results (figure 5.8) demonstrated that the heterologous DNA was stably integrated in the genome of the yeast. *P. pastoris* transformants were subsequently cultured in BMGY and induced in BMMY. The figures 5.10 and 5.11 show the analytical results from the supernatant and cellular fractions for four samples. The western blot analyses were performed with anti-2A and anti-cherryFP antibodies. The polyclonal anti-2A antibody was not commercially available and acquired from this laboratory stock. Figure 5.9 provides a western blot test using this antibody to probe FMDV and TaV 2As.
In the supernatant fractions (figure 5.10) no 2A was detected but unprocessed TaV 2A- cherryFP could be detected using the anti-cherryFP antibody. The cell pellets were probed for cherryFP (figure 5.11). CherryFP could not be detected.

The marked difference in fluorescence between induced and non-induced transformant, demonstrated that the induction was also successful. In addition, the western blot analysis of the supernatant fraction identified unprocessed cherryFP in the media, and this corresponds to the fraction of recombinant proteins which escaped the Kex2 protease activity and the 2A-stalling mechanism. It is not clear from these results if 2A can be processed from CherryFP in this system, or too diluted for a successful

analysis or simply too small to be adequately seen on the gel. Inserting a stretch of nucleotides corresponding to 10 kDa protein upstream of 2A will address the possible size issue and give 2A more traceability, while screening for a large amount of colonies will likely identify one or several suitable transformants providing good yields.

The second limitation experienced here and the reason why secretion was selected over intracellular accumulation was the difficulty associated with breaking the yeast cell wall. Three methods were tested: microwaving, enzymatic digest by zymolase and incubation in a breaking buffer. Out of these zymolase proved more successful for DNA extraction, it was subsequently applied to the cell fractions for western blot analyses, without success. There are other reported methods to explore: sonication, glass beads, other enzymatic preparations and freezing-grinding.

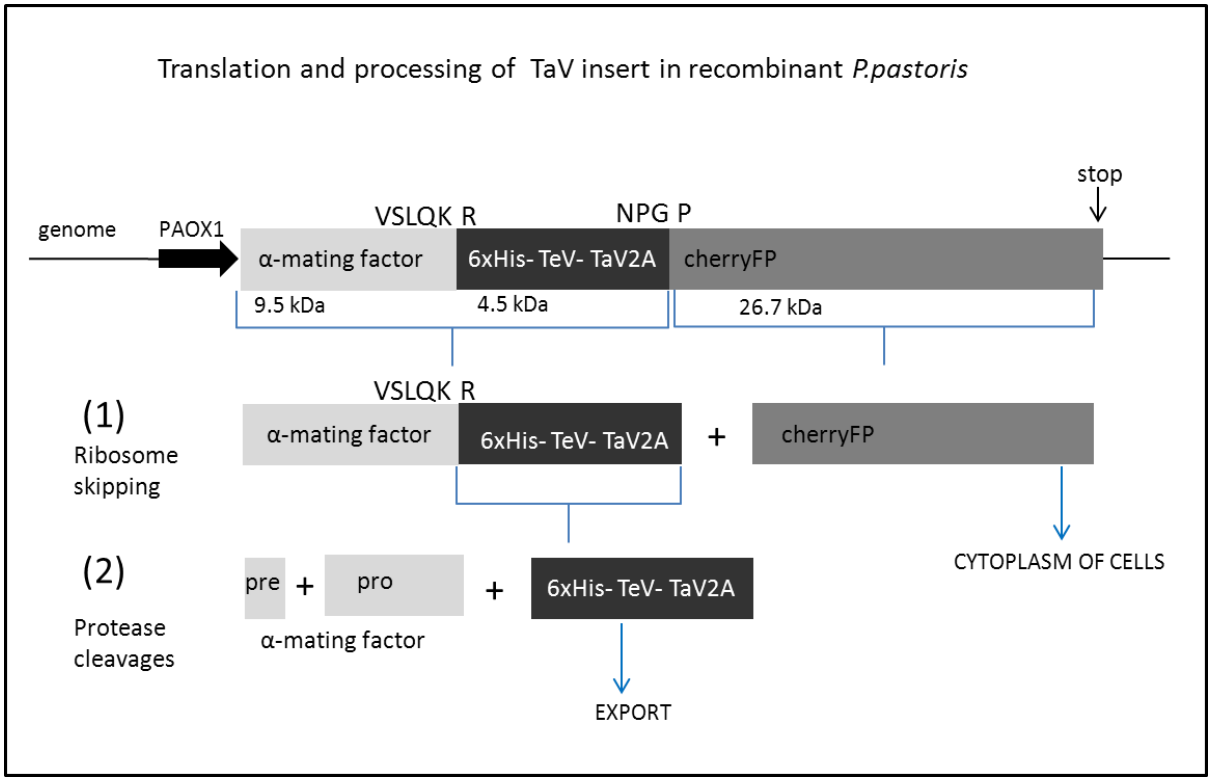


Figure 5.7: Diagram of the processing of the TaV-cherryFP insert in *P.pastoris*

The expression of the TaV insert fused to α-mating type was driven by *AOX1* promoter and induced with methanol. During translation (1), the α-factor and the TaV insert should be separated by the stalling mechanism at the 2A [NPGP] motif. (2) Subsequently, the signal peptidase cleaves the signal during translocation, and the protease Kex2 that recognises the cleavage site [VSLQKR] should cleave the α-mating factor. The TaV 2A insert should then be secreted to the media and could be purified by affinity via the poly-His tag. CherryFP should be localised in the cytoplasm of the yeast.

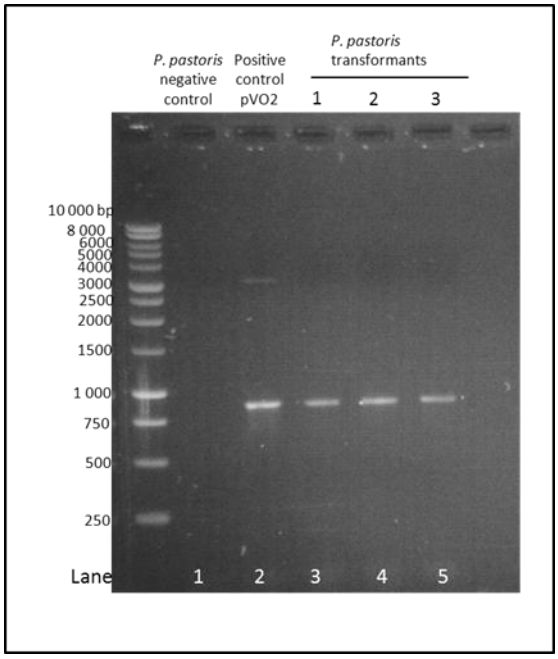


Figure 5.8: DNA gel electrophoresis of PCR of *P. pastoris* transformants

Lanes 3 to 5 contain examples of three positive samples (TaV 1,2 and 3). Lane 1 the negative control set with un-transfected yeast and the lane 2 used pVO2 as a positive control, which contains only the TaV-CherryFP insert. The primers annealed at the beginning of TaV and at the end of cherryFP and the resulting PCR showed the expected insert of 834 bp.

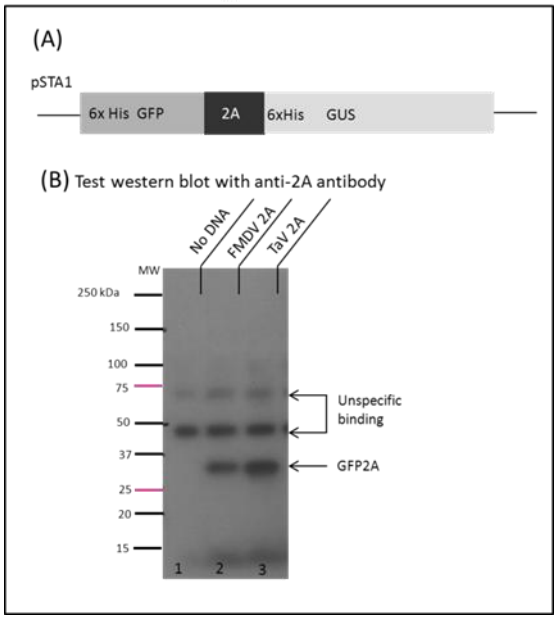


Figure 5.9: Western blot test of 2A antibody

The pSTA1 containing TaV 2A and FMDV 2A were translated in rabbit reticulocyte lysates without [^{35} S]-Methionine and analysed by western blot using anti-2A antibody. The lane 1 had contained no DNA and lanes 2 and 3 showed detection of GFP2A product for pSTA1 FMDV 2A (lane 2) and pSTA1 TaV 2A (lane 3). The antibody was able to detect GFP2A, but not the full-length product (GFP2AGUS). However it also reacted with other constituents of the cell-free extract.

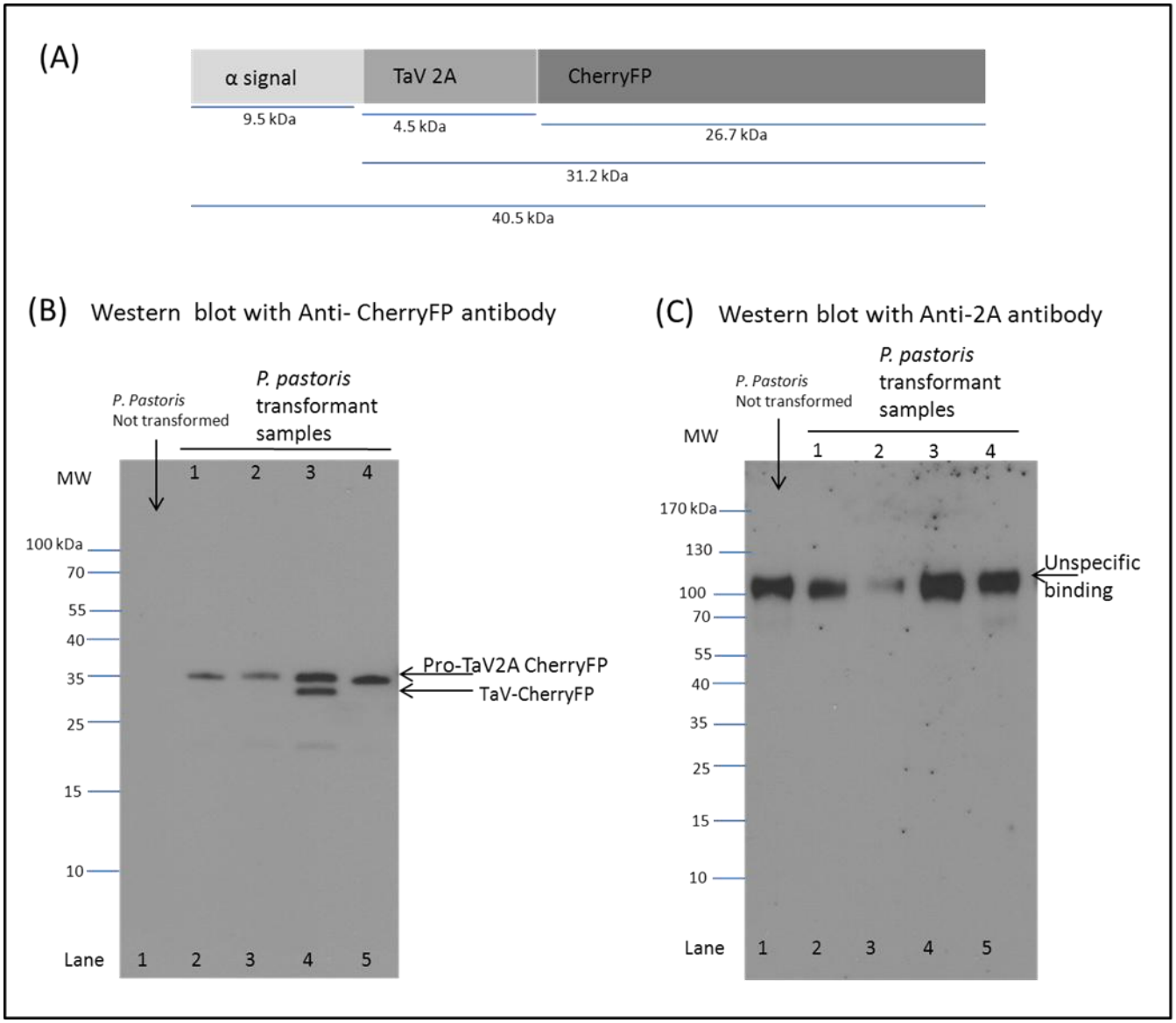


Figure 5.10: Western blot analysis of the supernatant fractions for four *P. pastoris* transformants

The cartoon in (A) indicates the MW of the proteins (underlined blue) constituting the TaV 2A inserts. In (B) the western blot analysis was carried out with anti-cherryFP antibody, and in (C) with anti-2A antibody. No TaV 2A could be detected in the supernatant (C) but unprocessed CherryFP in (B) could.

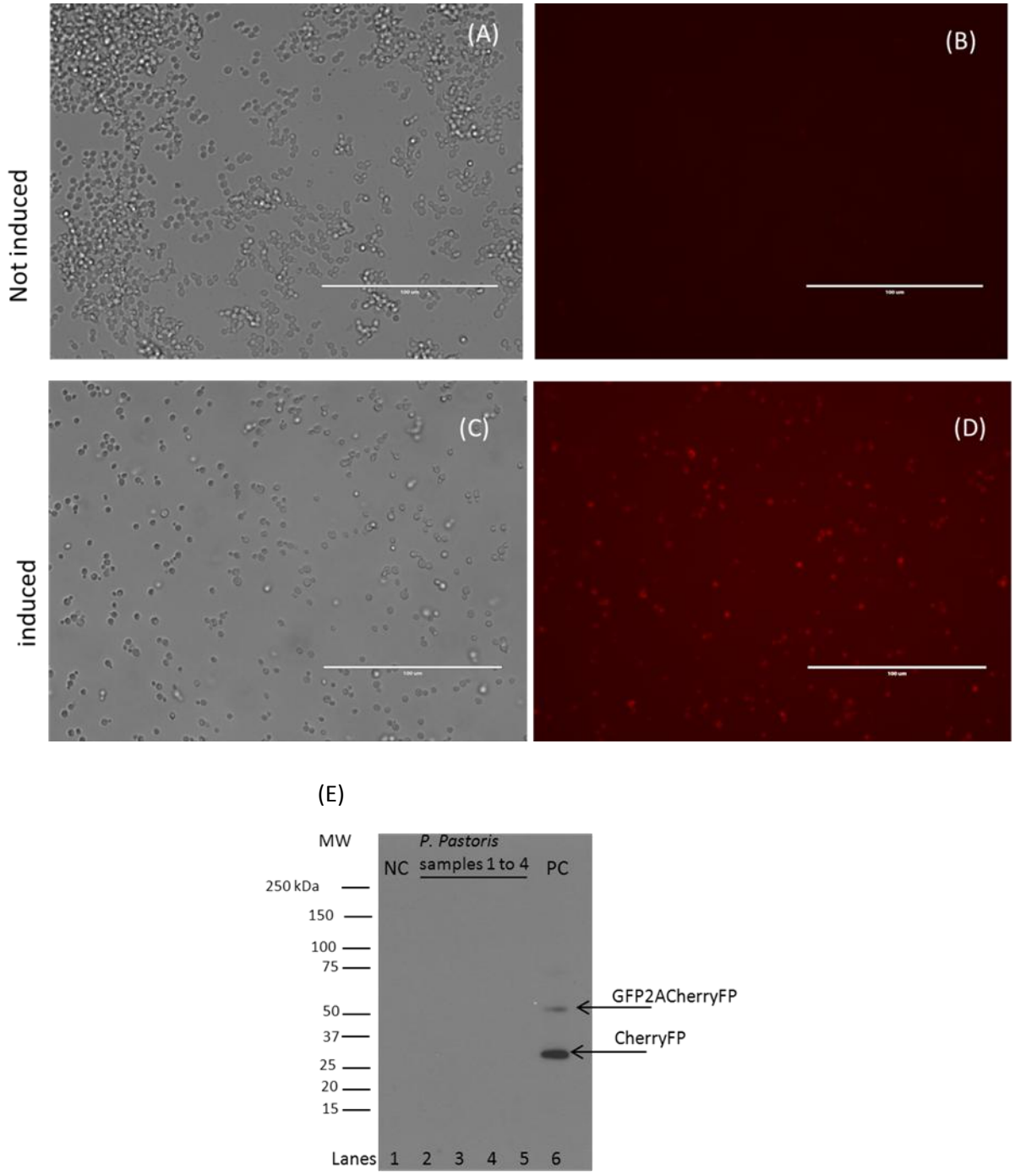


Figure 5.11: Analysis of cell fractions of *P. pastoris* transformants, western blot and fluorescence microscopy

Bright field pictures are on the left and the corresponding fluorescence showing cherryFP expression on the right. (A) and (B) were from the transformant TaV1 after culture in BMGY for 24 h. (C) and (D) demonstrated fluorescence of this sample after 24 h of induction by methanol in the BMMY medium. (E) Using zymolase to disrupt the yeast membranes, proteins were extracted from cell fractions of all four TaV transformants as well as untransfected yeast and examined by anti-cherryFP antibody. CherryFP was detected for the positive control (PC in lane 6) plasmid pJC3 translated in rabbit reticulocyte lysates without radioactive aa. CherryFP was not detected in the *P. pastoris* transformants (lane 2-5).

5.4 Discussion

In summary, the puromycin test proves that 2A induces an altered geometry at the PTC. Most importantly, substitutions of the A site proline shows that 2A renders the ribosome A site restrictive. The structure of the A site mRNA differs considerably in the Pro20 TaV 2A mutagenesis, yet the translation of the downstream protein implies that eRF1/3 performed the rescue activity as for TaV wild-type. This study proves that the potential rescue activity of eRF1/3 can accommodate several codon on the ribosome A site. The N domain of eRF1 carries the motifs for stop codon recognition (NIKS, YxCxxxF and GTx) (Song et al., 2000, Kryushkova et al., 2013). In the current model, eRF1is thought to be specific for stop codon decoding, it enters the A site complexed with eRF3 and GTP. This introduces a +2 nucleotide forward shift on toe print analysis. eRF1 undergoes a conformational change following GTP hydrolysis to accommodate the [GGQ] motif in the PTC. The activity of eRF1/3 results in peptide hydrolysis and the dissociation of the ribosome subunits (Alkalaeva et al., 2006, pisareva et al., 2011). Further studies are required to characterise this activity for eRF1 and 3.

Translation arrest by 2A results in a peptide bond not being formed. There are two scenarios; the peptide bond is not created because of an unfavourable chemistry involving proline or by steric hindrances. It is known that proline residues both in the P and the A site are poorly reactive. A study in *E. coli,* reported that even an efficient stop codon preceded with proline would recruit the tmRNA rescue system and induce tagging of the YbeL protein for degradation. The modification of another protein to match the end of YbeL, also induced tmRNA tagging (Hayes et al., 2002). Woolstenhulme and colleagues (2013) set out to uncover all possible stalling motifs for *E. coli*. They engineered the tmRNA system to link rescue of stalled ribosome to cell survival. They found that proline before a termination codon always arrested ribosome, a strategy which has been exploited by TnaC stalling peptide. Muto and Ito (2008) used the first 153 aa of the OmpA protein and modified the last proline with glycine or alanine. They translated the construct using the *in vitro* fully reconstituted translation system (PURE system), and reacted the translated products to puromycin. Unsurprisingly, more peptidyl-tRNA was formed with proline. After 3 min of incubation, 72 % of peptidyl-tRNA remained for the protein ended by proline. In another study, the time of transfer of dipeptide (fMet-xtRNA^{x}), to puromycin was quantified by quench flow technique (Wohlgemuth et al., 2008), and it was established that prolyl-tRNA ^{pro} was poorly reactive when in the P site (0.14 s $^{-1}$) compared to lysine (100 s^{-1}). The glycyl-prolyl peptide bond at the C-terminus of secM is also not effected. And it was proposed that in that context, prolyl-tRNA played the role of effector of translation arrest (Nakatogawa and Ito, 2002). In *E. coli*, strings of three consecutive prolines, whether naturally present or engineered stall the translating ribosome and it is argued that this is likely the consequence of the imino nature of the residue. Stalling occurs also independently of proline codon usage and in different

proteins. Two studies revealed that a special elongation factor EF-P is designed to assist with poly-proline stretches in *E. coli* (Ude et al., 2013 and Doerfel et al., 2013).
Although all nascent stalling peptides differ in their upstream context and also in their precise mechanisms in the exit tunnel, they contain a proline at their C-terminus. It is therefore not surprising that in the context of 2A, a proline in the A site is linked to higher efficiency of stalling. But it is surprising to find that 2A renders the A site restrictive. The result implies that the imino nature of the proline residue is not the principal factor in stalling, but the steric constraint that its pyrrolidine ring imposes on the A site is. There are a multitude of possible targets in modulating the PTC activity, any nucleotides or H-bondings maintaining the correct position for the tRNA substrates could affect the reactivity.

There are no studies involving a systematic quantification of the A site residue in accepting the nascent peptide, but two studies reported a quantification for a peptide transfer to the A site tRNA. The first study reacted di-peptide to puromycin in the A site (Wohlgemuth et al., 2008). The quench flow technique consists of a pretranslocation ribosome/di-peptide mixed with puromycin or Phenylalanyl-tRNA^{Phe}, reacted and quenched with KOH before HPLC analysis. The amino acid reacted to puromycin at diverse speed. Only eight residues in the P site were tested, and their order of reactivity was: lysine (100 s^{-1}), arginine (90 s^{-1}), alanine (57 s^{-1}), serine (44 s^{-1}), phenylalanine (16 s^{-1}), valine (16 s^{-1}), aspartic acid (8 s^{-1}) and proline (0.14 s^{-1}). When the authors misacylated tRNAs (in the P site) to carry Phe the reactivity to puromycin (in the A site) decreased to match that of Phenylalanyl-tRNA^{Phe}. When the di-peptides fMet-Arg, fMet-Asp, fMet-Phe and fMet-Pro (in the P site) were reacted to Phenylalanyl-tRNA^{Phe} (in the A site) the authors found a uniform velocity of synthesis regardless of the nature of the last amino acid. These last data demonstrate that the A site tRNA, is an important determinant of the efficiency of peptide transfer to the A site (Wohlgemuth et al., 2008). Pavlov and colleagues (2009) compared the formation of di-peptide between fMet and Ala-tRNA^{Phe}, fMet and Phe-tRNA^{Phe} and FMet and prolyl-tRNA^{Pro} or prolyl-tRNA^{Phe}. They demonstrated that it takes 11 ms to transfer fMet to Phe, and 23 ms to transfer fMet to alanine misacylated to tRNA^{Phe}. The alanyl-tRNA^{Ala} result was not provided. Surprisingly, the reaction of fMet to proline misacylated to tRNA^{Phe} occurred in 555 ms, but that time was decreased to 50 ms when the proline cognate tRNA was employed. They interestingly also reported a time difference between the four synonymous codons. Proline codon CCG reacted in 50 ms to fMet. Proline codons CCA and CCU, reacted in 47 ms and CCC reacted in 66 ms. So the logical conclusion is that tRNA somehow compensates for the poor reactivity of the residue. Zhou and colleagues (2011) calculated the codon usage for FMDV flanking the cleavages and processing sites along the polyprotein and found that the viral choice of codons matched the codon usage for cattle and pigs, suggesting that the virus uses this strategy to make efficient use of their host translational machinery. At the 2A/2B junction, the AAU (Asn) and UUA (Leu) codons are under-represented. However, at the Gly-Pro pair, the viral codon

usage pattern resembles that of the host. It is therefore unlikely that *in vivo* the codon usage at the 2A/2B Gly-Pro bond would affect the efficiency of stalling.

Other stalling sequences also render the ribosome A site selective. The *ermAL* (an example of *erm*) stalling peptide can not be transferred to A site: aspartic acid, glutamic acid, glycine and tryptophan (Ramu et al., 2011). Woolstenhulme and colleagues (2013) demonstrated that peptides finishing with Pro-Pro motifs could not be transferred to A sites: asparagine, aspartic acid, glutamine, glycine and tryptophan. In a breakthrough article, summarising the results from a ribosome profiling study carried out on mouse stem cells, it was reported that stalling sequences in the eukaryotic proteome is also common. The authors found 1500 pauses in a set of 4994 expressed genes. They reported that the majority of the ribosomes carried out their translation following pauses. They did not provide the full list of the stalling motifs however mentioned that the PPD and PPE were the strongest pause sites with aspartic acid or glutamic acid in the A site (Ingolia et al., 2011). The -2 residue also influenced the level of tmRNA tagging for the YbeL protein. Substitution of the -2 residue for glutamic acid, proline, valine, aspartic acid and isoleucine led to high levels of tagging, interestingly substitution for asparagine led to low levels of tagging (Hayes et al., 2002). Referring back to the example of *ermC* and *ermAL*, it is interesting to note that two peptides very similar in sequence do not induce the same restrictivities. *ermC* (IFVI) stalls regardless of A site amino acid. But *ermAL* (IAVV) stalls only when the A site is occupied by amino acids that can not accept the nascent peptide: aspartic acid, glutamic acid, glycine and tryptophan (Ramu et al., 2011). The interaction of the nascent stalling peptides to the exit tunnel influences the activity at the PTC and the reactivity of the tRNA/aa at the P and A site. *ermC*, SecM and AAP interacts with the ribosomal nucleotide A2062. A2062 transitions from an open (towards the lumen) to a closed (against the tunnel wall) conformation and interacts with the -2 amino acid (Phe7) of *ermC* (Ramu et al., 2011) and the compulsory -2 amino acid Arg163 for SecM (Yap and Bernstein, 2009). A2062 is flanked at both sides with nucleotides that participate in the PTC activity. Collectively, these examples demonstrate that the penultimates residues preceding the stalling residues play a major role in the stalling efficiency. The -2 amino acid for 2A would be Asn18. The mutagenesis performed for the previous chapter resulting in no 2A activity with mutated [NPG], supports the theory that a ribosome residue within the vicinity of the PTC is responsible for the stalling mechanism, and through a relay to PTC residues induces a subtle disturbance of the PTC geometry.

In the context of 2A, mutations of the A site residues would therefore be a perfect test to categorise the nascent 2As. Any nascent 2A interacting in a similar way to the exit tunnel, would produce the same results with A site proline substitutions.If the stalling is induced by a different mechanism it is likely that this will result in different restrictivities in the A site amino acid.

Chapter 6 The 2A collection expands to cellular organisms

6.1 Introduction

2A presence has firmly been established as part of the genome of several viruses.The motif has also been reported at the N-terminal of non-LTR retrotransposons of *Trypanosoma cruzi* (L1Tc) and *T. brucei* (*igni*) (Donnelly et al., 2001*a* and Heras et al., 2006). In the past three years, more than one hundred 2A sequences were predicted to be part of the sea urchin *Strongylocentrotus purpuratus* genome based on a BLAST search using the [DV/IExNPGP] motif and are being analysed at the time of writing. 2A in the Sea Urchin was identified at the N-terminus of innate immunity CATERPILLER proteins or associated with non-LTR retrotransposons. Several of the sequences were tested in the expression system and shown to be active. If these sequences were translated, they would confer the stalling mechanism and create discrete products from a single ORF.

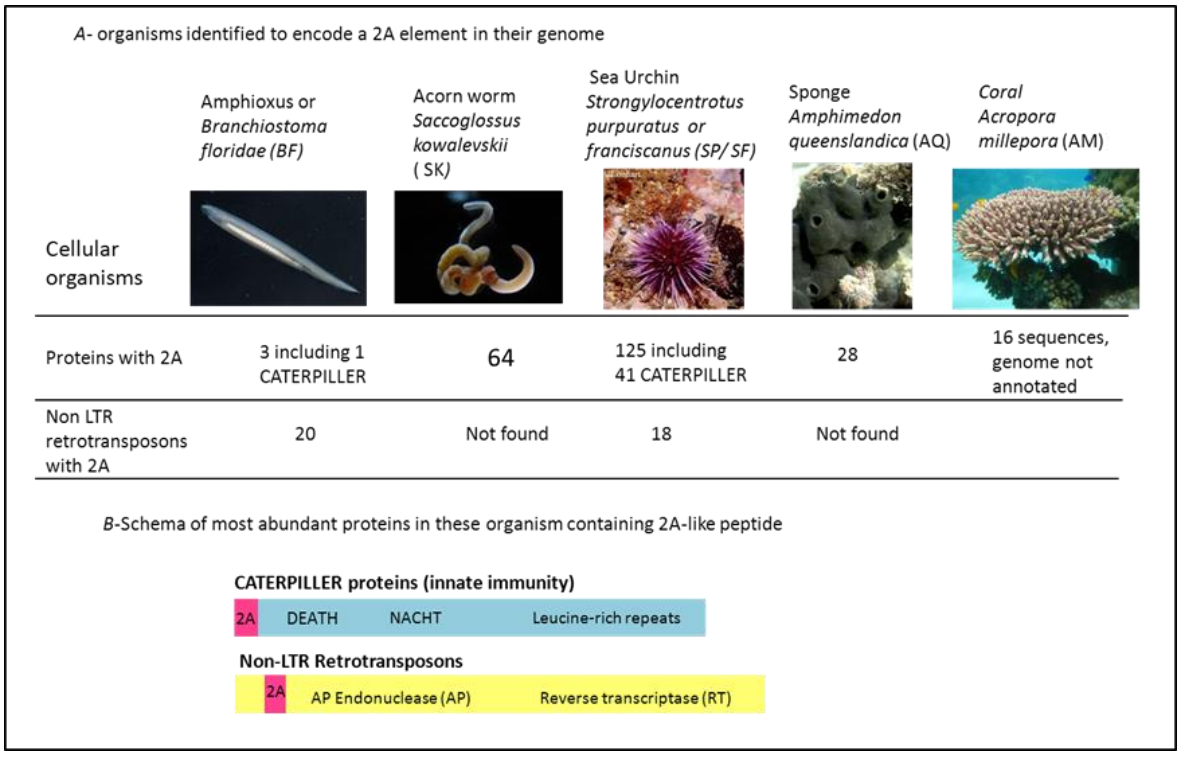


Figure 6.1: Summary of 2A –like elements in marine organisms
The summary shows in ***A*** abundance and ***B*** the typical localisation of 2A in protein sequences. Images were reproduced from various sources online. Amphioxus image is from http://www.gsite.univ-provence.fr, acorn worm image is from http://www.hermes.mbl.edu, sea urchin image is from http://www.bio.classes.ucsc.edu, *Amphimedon queenslandica* image is from http://www.blogs.sciencemag.org and *Acropora millepora* image is from http://www.lemur.amu.edu.pl.

This section consists of a listing of newly discovered 2A in the genome of several organisms. 2A sequences with a deviating motif were observed in the genome of viruses *Armigeres subalbatus virus* (AsTV), *Omono river virus* (OMRV1 and 2) and *Drosophila megalonaster totivirus* (DTV) (Isawa et al., 2011). Similarly, two sequences were also identified in *Bovine rotavirus B* and *Human rotavirus A*. 2A motifs were identified in several marine cellular organisms. The analytical task was divided between several workers and the results for amphimedon, acropora and *S. purpuratus* / *franciscanus* will not be presented here as they did not constitute my primary focus of work. This chapter summarises the results for *Branchiostoma floridae* (amphioxus), and *Saccoglossus kowalevskii* (acorn worm). The acorn worm bioinformatics analyses were carried out in majority by Dr Sukkhodub in this laboratory, but I carried out part of these and tested four 2A sequences. However, all the bioinformatic analyses and subsequent testing for amphioxus 2A sequences were exclusively my work.

It is not certain that 2A is part of the transcribed genome for the marine organisms; but its presence as part of specific non-LTR retrotransposon clades suggests a horizontal transfer. It is now established that 2A is not restricted to the consensus motif previously identified, however searches revealed that [DV/IExNPGP] is the most abundant and probably still the only reliable query for identification of new sequences.

6.2 Materials and methods

6.2.1 Identification of 2A candidates- search for homologies

BLAST searches used the 2A motif [DV/IExNPGP] and [GVEPNPGP] to identify potential new 2A sequences. The consensus motif was entered into the ProtBlast search online facility from NCBI (http://www.blast.ncbi.nlm.nih.gov) with a maximum target of 500. Protein sequences were entered into the ProDom (http://www.prodom.prabi.fr) online facility to identify domains. Thirty amino acids leading to the consensus motifs were considered for the purpose of testing the activity of the potential 2A, and used to derive the corresponding nucleotide sequence. For non-LTR retrotransposons in amphioxus, sequences were identified from clustalX alignments to the collection of non-LTR retrotransposon available at girinst (repbase) (http://www.girinst.org). Phylogenetic trees were built with FigTree v1.3.1 (http://www.tree.bio.ed.ac.uk/software/figtree). For *S. kowalevskii*, contigs coding for potential 2A sequences were identified using the online blast facility at Baylor college of Medicine (http://www.blast.hgsc.bcm.tmc.edu), the GenScan programme (http://www.genes.mit.edu/GENSCAN.html) was used to predict coding sequences and fastPCR to perform *in silico* translation.

6.2.2 Selection and cloning

Representatives 2A for each organism/ non-LTR retrotransposon clade were selected on the basis of their deviation to the [DV/IExNPGP] 2A motif. Gene blocks of the sequences spaced with unique restriction sites were ordered from Dundee cell products and each 2A sequence was excised, gel purified and further restricted with Xba1 and Apa1 and ligated to pSTA1. The modified pSTA1 were tested in rabbit reticulocyte lysates following the protocol in section 2.2.9.

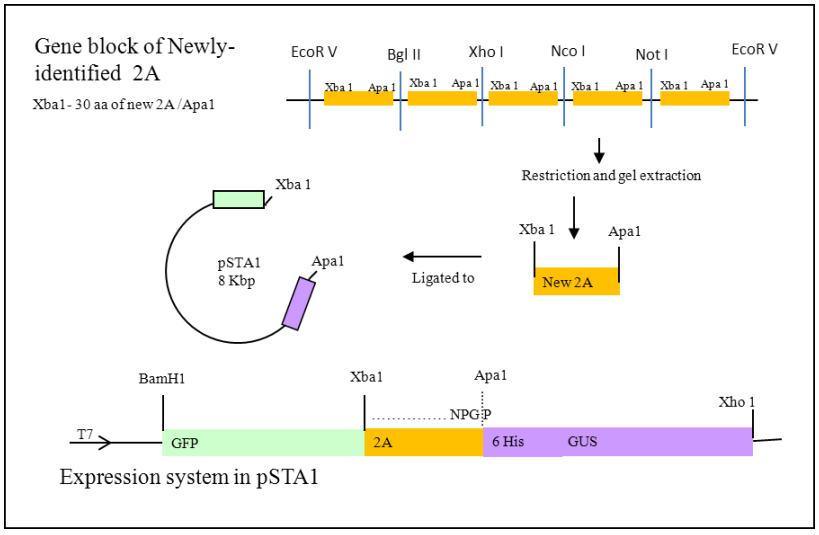


Figure 6.2: Modification of pSTA1 for the cloning of candidate 2As.

Table 6.1: Novel amphioxus and totivirus/rotavirus 2A sequences tested.

Sequences were identified by local alignment at NCBI, the table features the accession number, the nucleotide and protein sequence for all the 2A tested.

Organism and NCBI accession number	Experiment reference+ Protein sequence (bottom) and corresponding nucleotide sequences (top)
Branchiostoma floridae (XP_002204853.1)	BF53 TGGGCGGAGGCCTCCTGTCTACTTGTGTGGATGGTCATTTCCCAACTCATGCTCAAACTT W A E A S C L L V W M V I S Q L M L K L GCGGGAGACGTCGAGGAGAACCCCGGGCCC A G D V E E N P G P
B. floridae (XP_002224141)	BF41 TCTAGATACGTCATGATGTCCTGCCTCCTCGTCTGGTTCATGATGGTTCATAAACTTCTT S R Y V M M S C L L V W F M M V H K L L CTGCAAGCGGGAGACATCGAGCCCAACCCCGGGCCC L Q A G D I E P N P G P
B. floridae (XP_002229629)	BF29 CGCGTCTGCTCCCCCGACGCCACAGCCACCAAGAACTGCGCTATGTACATGCTTCTGCTT R V C S P D A T A T K N C A M Y M L L L TCTGGAGACGTCGAGACCAACCCCGGGCCC S G D V E T N P G P
B. floridae (XP_002234262)	BF62 CGCGTCTGCTCCCCCGACGCCACAACCACCAAGAACTGCGCTATGTACATGCTTCTGCTT R V C S P D A T T T K N C A M Y M L L L TGTGGAGACGTCGAGCCCAACCCCGGGCCC C G D V E P N P G P
Armigeres subalbatus virus (ACH85915)	AsTV CCTGAGCTTAATGGAGATCAAAGAGCTACTCTTTCCGCATGGACTCGAGATCTCACCAAA P E L N G D Q R A T L S A W T R D L T K GATGGCGATGTAGAATCCAATCCCGGGCCC D G D V E S N P G P
Omono river virus (BAJ21512.1)	OMRV1 TTTGATGAGACCAAGAAGCTCGTCTGTCTCGGCACCTATCCCTTTAAGAAGGATCTCACC F D E T K K L V C L G T Y P F K K D L T CGCGAGGGCGTCGAGTCCAATCCCGGGCCC R E G V E S N P G P
Omono river virus (BAJ21510.1)	OMRV2 TTTGATGAGAATGCCAAGATCGTCCTCCTCGGCACCTTTCCCTTTAAGAAGGATCTCACC F D E N A K I V L L G T F P F K K D L T AAGGAGGGCGTCGAGCCCAATCCCGGGCCC K E G V E P N P G P
Drosophila megalonaster totivirus (YP_003289292.1)	DTV CTCGCCTATCTCAATAATGATCAGAAGACCACCCTCTCCGCCTGGGTCCGCGGCCTCACC L A Y L N N D Q K T T L S A W V R G L T GTCGAGGGCGTCAAGCCCAATCCCGGGCCC V E G V K P N P G P
Human rotavirus A (CAC01682.1)	HuRV A CTCACCAATTCCTATACCGTCGAGCTCTCCGATGAGATCAATACCATCGGCTCCGAGAAG L T N S Y T V E L S D E I N T I G S E K GGCGAGAATGTCACCATCAATCCCGGGCCC G E N V T I N P G P
Bovine rotavirus B (AAA79028)	HuRV B CTCGCCAATTCCTATACCTCCGATCTCCAGGATACCATCGATGACATCTCCGCCCAGAAG L A N S Y T S D L Q D T I D D I S A Q K ACCGAGAATGTCACCGTCAATCCCGGGCCC T E N V T V N P G P

Table 6.2: Proteins and nucleotide sequences tested for 2A elements identified in amphioxus non-LTR retrotransposons.

Sequences were identified in repbase database, through similarities to 2A motif and N-terminal position in the organisation of the non-LTR retrotransposon

Non-LTR	Protein sequence (bottom) and corresponding nucleotide sequences (top)
Crack-15_BF	CACTCAGTACTAGTATGCGACCACTGCGTAACAGTATTCGTAGTAATTCTACTACTACTA H S V L V C D H C V T V F V V I L L L L TGCGGAGACATTCACAACAACCCCGGGCCC C G D I H N N P G P
Crack-17_BF	GCAGTAACATCAACATCAGTAAACTGCGTACACCTATGCTTCCACACACTACTAATTCTA A V T S T S V N C V H L C F H T L L I L TCAGGAGACGTAGCAGTAAACCCAGGGCCC S G D V A V N P G P
CR1-2_BF	CGAACATCAGACCGACTATTCACATGCCTACTATACCTATGCTCAGTACTAATGTCACAA R T S D R L F T C L L Y L C S V L M S Q GCAGTAGACCTAGAAACAAACCCCGGGCCC A V D L E T N P G P
CR1-10_BF	GGAACAGACAACGTATCAGCAGAATTCACACAATGGAAACCAGCAATTGACCTAACACAA G T D N V S A E F T Q W K P A I D L T Q CACTACGACGTACACCCAAACCCAGGGCCC H Y D V H P N P G P
CR1-11_BF	CTAGCACCACACTGCCGACCAAAATTCACACTATTCTCACTAACACTAATTATTCTACTA L A P H C R P K F T L F S L T L I I L L GCAGGAGACGTAGAACTAAACCCCGGGCCC A G D V E L N P G P
CR1-31_BF	TACCTAATGTCACGACAACGACTAGTACTACTATACCTAACAATGCTACTAATTAGCAAA Y L M S R Q R L V L L Y L T M L L I S K TCATACTCACCAGAACCAAACCCCGGGCCC S Y S P E P N P G P
CR1-53_BF	CACTTCGACATTTTCCTACTATTCTTCCCACTACCAGTACTAGTAGTACTATCACTAATT H F D I F L L F F P L P V L V V L S L I GCAGGAGACATTCACCCAAACCCAGGGCCC A G D I H P N P G P

.

Table 6.3: Protein and nucleotides sequences tested for 2A elements identified in *S. kowalevskii* (SK).

2A were identified by blast search at Baylor College of Medicine, the resulting nucleotide sequence was tested for the above contigs.

Contig number	Experiment reference+ Protein sequence (bottom) and corresponding nucleotide sequences (top)
XP_002731766 (accession number in NCBI)	**SK1** ATGTGGTTTCTCGTACTACTAAGCTTTATTCTATCAGGGGATATTGAAGTCAATCCCGGG M W F L V L L S F I L S G D I E V N P G CCC P
Contig 36012	**SK6** TGTGACCAACACGGGCGCATTCACGTCAGCCTCTTTATTTTTGGAATCCTACTACTACTA C D Q H G R I H V S L F I F G I L L L L AGCGGGGATATTGAAGTAAATCCCGGGCCC S G D I E V N P G P
contig 76190	**SK61** ATGTTTTCTATCCAGCCCCTCTGGACTTCTAAACTCATCATCCTCTGTGGGGATGTAGAA M F S I Q P L W T S K L I I L C G D V E TCTAATCCCGGGCCC S N P G P
contig 10214	**SK47** ATGTTTGCACTCTATCTCTATCATTGTCGTATGCGTAGTAACCTCATACTCATACGTGAC M F A L Y L Y H C R M R S N L I L I R D ATCGAAACCAACCCCGGGCCC I E T N P G P

6.3 Results

Using the [D(V/I)ExNPGP] motif, several databases were screened: the non-LTR retrotransposon database, NCBI and the genome sequence of *S. kowalevskii* at Baylor college of medicine. The searches identified potential new 2A encoded by the genomes of the cellular organisms: amphioxus and the acorn worm and several totiviruses. Further bioinformatics analyses characterised the type of sequences where the 2A motif was discovered.

Table 6.4: Summary table of abundance and type of sequences identified with a 2A motif.

B. floridae sequences were accessed at NCBI, the genome of the *S. kowalevskii* was accessed at Baylor College of Medicine. These are further detailed in figures 6.4 to 6.9 for *B. floridae* and figure 6.10 and appendix 4 for *S. kowalevskii*.

Organism	Type of sequence	Abundance
Branchiostoma floridae	Predicted protein	3
	Non-LTR retrotransposon	20
Saccoglossus kowalevskii	contigs	64

6.3.1 2A in *B. floridae*

In amphioxus, 2A is part of two types of non-LTR retrotransposon clades: the Crack and the CR1 clades. The figure 6.5 provides the full amino acid sequences of all the amphioxus non-LTRs with a 2A and their identity. Several of these were tested in the *in vitro* bicistronic expression test. The result provided in figure 6.3 shows that the non-LTRs are mostly unable or inefficient at performing the separation between GFP and GUS, compared to FMDV. It is notable that they have a deviating consensus motif [GDIHPNPGP] instead of [GDIExNPGP]. Following the standard analysis established by Malik and colleagues (1999), the phylogenetic trees (figure 6.7 and 6.8) were created by alignment of the reverse transcriptase domain of the non-LTR retrotransposons. In figure 6.7 the phylogenetic tree shows that the Crack elements with a 2A motif cluster, a result which is also reproduced in figure 6.8 for CR1 elements with a 2A motif. Thirteen CR1 and seven Crack non-LTR retrotransposons were found to have a 2A motif. The repbase database provides the details of the domains identified in each non-LTR retrotransposon. CR1 and Crack elements share a common organisation (Kapitonov et al., 2009), but clades must share a 40 % homology within their reverse transcriptase domain to be grouped (Malik et al., 1999). A canonical CR1 element represented in figure 6.6 has two ORFs. ORF1 generally encodes a plant homeodomain (PHD) or an esterase. ORF2

encodes a protein with endonuclease and reverse transcriptase properties (kapitonov et al., 2009). An examination of the CR1 elements with a 2A sequence (CR1-2A) shows that these CR1 may be organised differently (figure 6.6). In one subgroup, the 2A sequence is at the N-terminus of the endonuclease (CR1-17, 10, 36, 46, 26 and 53) and therefore potentially in ORF2. In that subgroup, it is known that CR1-26 and CR1-53 have two ORFs. In the process of successive retrotranspositions, the non-LTR undergoes many deletions at the 5'end, and so it is not known if the ORF1 is simply missing for the other non-LTR retrotrnaposons of this subgroup of CR1-2A. However, another subgroup of CR1-2A (CR1-1, 2, 3, 11, 12 and 18) corresponding to a cluster in figure 6.8 features the 2A sequence at the N-terminus of the PHD domain, but the classical ORF1 and ORF2 could not be identified.

In amphioxus, 2A was also identified at the N-terminus of three predicted proteins shown in figure 6.4. An analysis of the proteins domains showed no homologies between these. Their 2A sequences however can operate on eukaryotic ribosomes and cause stalling with varying efficiencies (figure 6.3 (A)).

The search for the 2A motif was extended to related amphioxus species. The query for *Branchiostoma belcheri*, *B. lancealeatum*, *B. malayanum* and *B. japonicum* did not return any hits in NCBI or repbase.

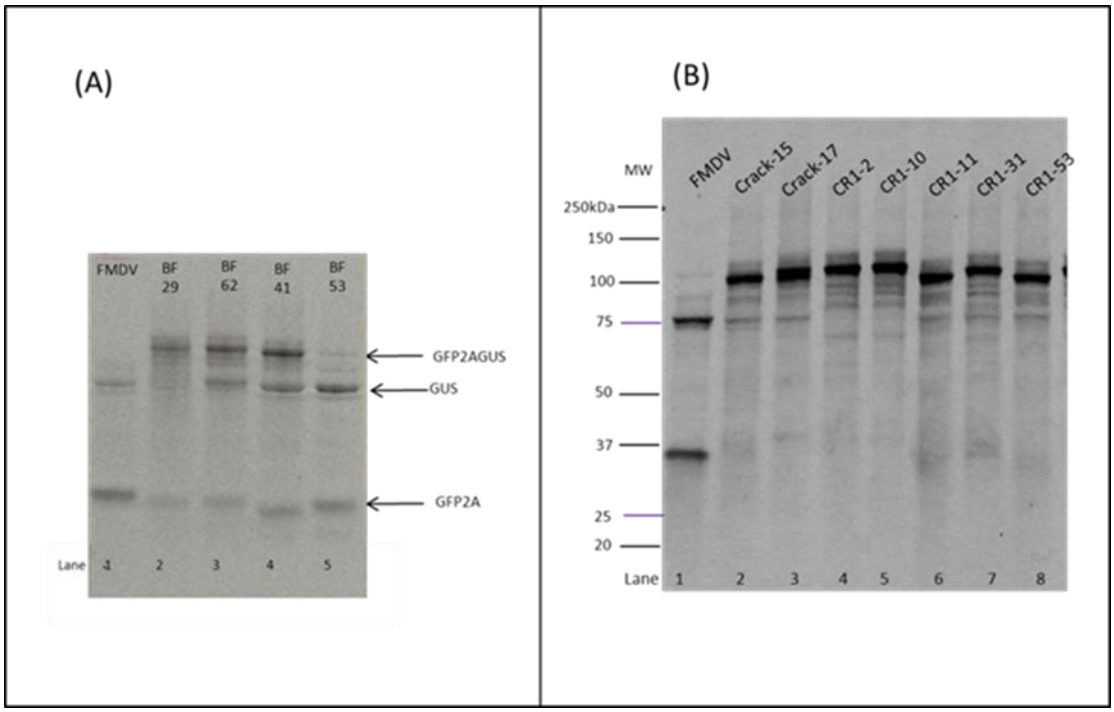


Figure 6.3: SDS-PAGE analysis of test expression for amphioxus 2As

In (A), all the amphioxus 2A sequences predicted to precede a protein were tested (sequences in table 6.1). They exhibit the typical FMDV 2A (lane 1) expression profile: a full-length product at 100 kDa, and a disproportionate amount of upstream

GFP2A (at 32 kDa) compared with GUS. Panel (B) shows the translation profile of 2A identified at the N-terminal of non-LTR retrotransposons (CR1 and Crack clades) in amphioxus (sequences in table 6.2). The non-LTR retrotransposons 2A are mostly unable or inefficient at performing the separation between GFP2A and Gus (lanes 2 to 8) compared to FMDV (lane 1).

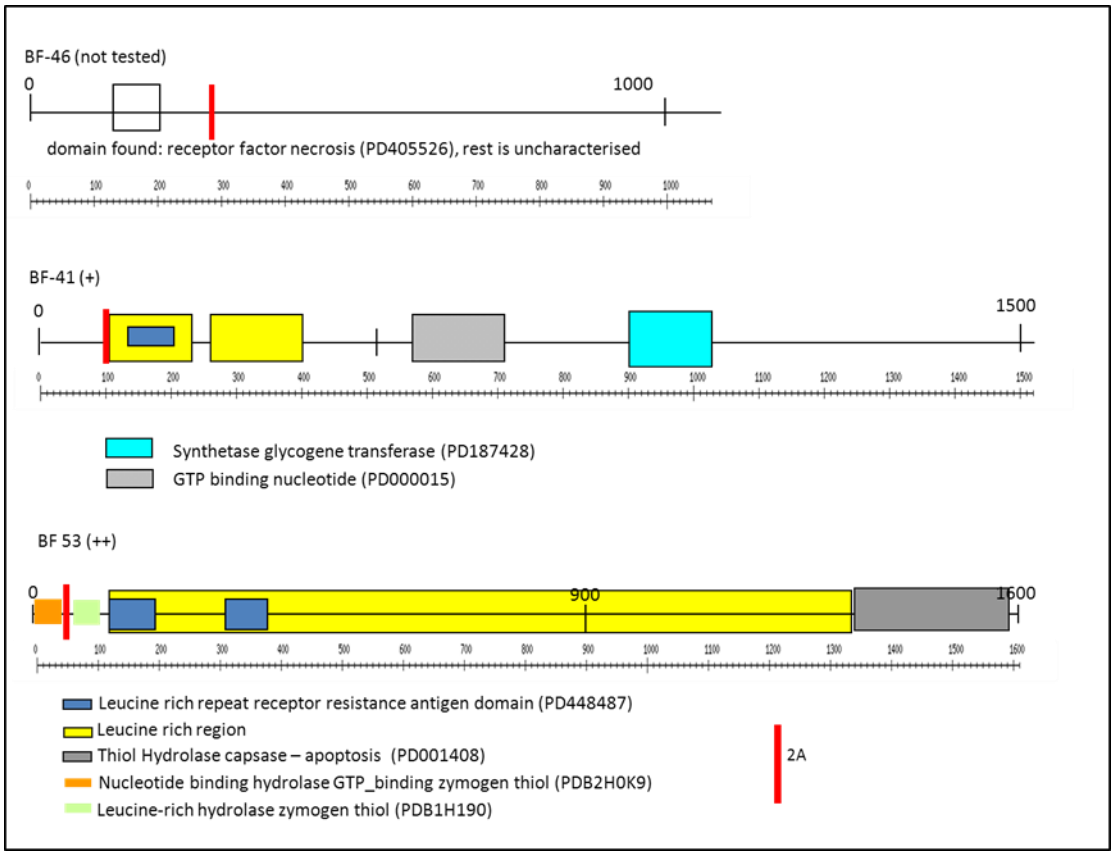


Figure 6.4: Visual representation of key domains in the 2A-containing amphioxus proteins

Each domain is assigned a coloured box positioned where it has been identified within the protein of interest, the protein domain number (PD) is cited in bracket. This is a summary of search done using the ProDom facility available online. The position of 2A is indicated as a red bar. The efficiency of 2A is indicated in bracket. These are hypothetical proteins.

```
Crack-9      -IHVKTSVNLAHLCIHTLLLLSGDVACNPGPNQSEDNVVANVDWQPLFERI-  (NT)
Crack-10     -LYHKNLLTEQCNDQVNLICLAFDIHPNPGPISSTCGTCSKRVTNKQRAIC-  (NT)
Crack-11     -CHVETRVNVVHLCLHTLLLLSGDVASNPGPKDPCGVCTKSVRNNQKGICC-  (NT)
Crack-15     -HSVLVCDHCVTVFVVILLLLCGDIHNNPGPARLNLPQKGLHIGHLNICSW-  (+)
Crack-16                          DIQTNPGPDLHLPSKGLHVGHVNINSLR-  (NT)
Crack-17     -AVTSTSVNCVHLCFHTLLILSGDVAVNPGPKDPCGICNKCVRKNQKGICC-  (+)
Crack-28     -TCTERTERTLNLLVCATLLLAGDVSPNPGPDTGGLPVWRKGIVYAFYNVV-  (NT)

CR1-1        -KKTMIHNDSTKLSLIMILLLSGDIEINPGPRPPKCPCGSCNKAVQNKHAA-  (NT)
CR1-2        -RTSDRLFTCLLYLCSVLMSQAVDLETNPGPRPPKYPCGSCGKAVTFKHKG-  -
CR1-3        -YLRTSDRLCLLYICSVLMAQAVDLETNPGPRPPKYPCGCCGKAVTFKHKG-  (NT)
CR1-10       -GTDNVSAEFTQWKPAIDLTQHYDVHPNPGPDLSELLLSTDFRSKGLLTIA-  -
CR1-11       -LAPHCRPKFTLFSLTLIILLAGDVELNPGPRAPKYPCGVCHRAVRWEKVD-  -
CR1-12       -PRNPLKSISVSIALLVMLTQSGDVHPNPGPYKPKFPCLLCGKAAKWNQRA-  (NT)
CR1-17       -TISFILSIFYSNFLLLLIVLSNDIHPNPGPIQPTGTSKCLNIFHANVNSL-  (NT)
CR1-18       -YLRTSDRLCLLYICSVLMAQAVDLETNPGPRPPKYPCGSCGKAVTFKHKA-  (NT)
CR1-26       -NLDIFLSYTTVFISFVVILVAGDVHPNPGPVCRKQFNVMHLNVNSLVAGT-  (NT)
CR1-31       -YLMSRQRLVLLYLTMLLISKSYSPEPNPGPLLDQCPNHTCTNDSSSSQSH-  (+)
CR1-36       -DKDYGIVIQFMLPFFVLFLICGDIHPNPGPQQNELIVRFTNIRGLRTNLT-  (NT)
CR1-46       -TLTICPQCILIFISLIMIILAGDIHPNPGPFFRKEINFMHINVNSLVAGS-  (NT)
CR1-53       -HFDIFLLFFPLPVLVVLSLIAGDIHPNPGPSTMYTSFKYLNILHANVNSL-  -

FMDV 2A      -ELYKSGSGACQLLNFDLLKLAGDVESNPGPHHHHHHLRPVETPTREIKKL-  ++++
```

Figure 6.5: The non-LTR retrotransposons with 2A in amphioxus

Amino acid window of all non-LTRs with a 2A element. The activity for those tested was indicated in bracket and the 2A consensus motif in bold.

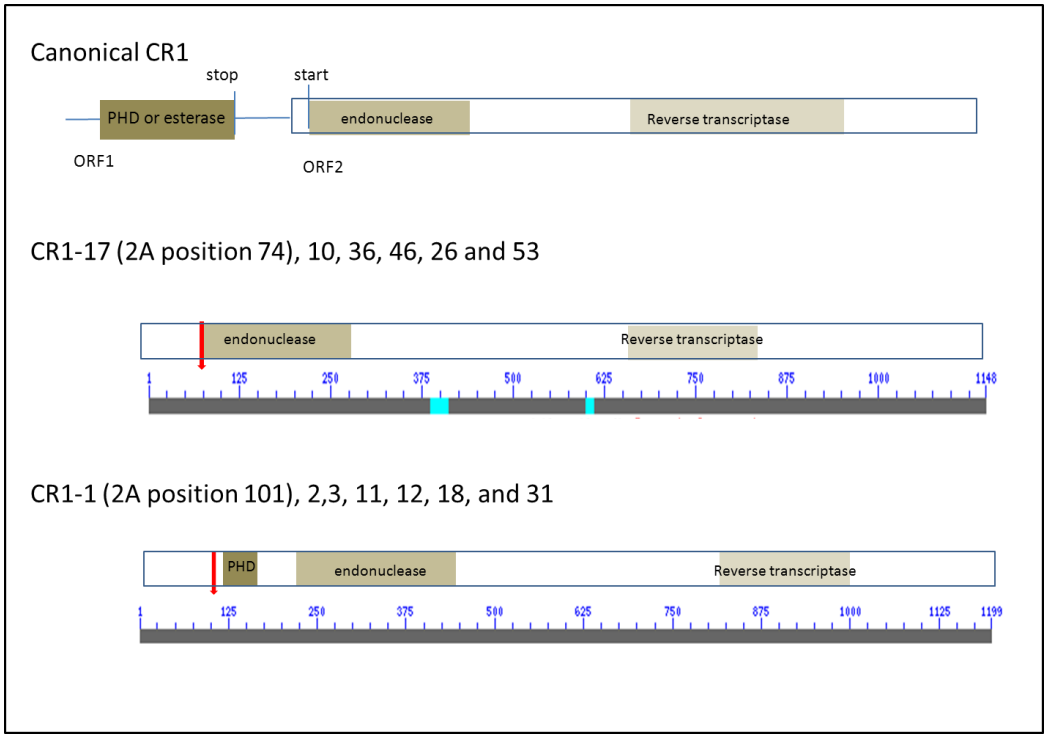


Figure 6.6: Visual representation of the non-LTR CR1 organisation in amphioxus

The canonical CR1 non-LTR retrotransposon has two ORFs, CR1-2As however only possess one ORF and is followed by either the canonical endonuclease (CR1-17) or the PHD domain (CR1-1) normally present in ORF1. The canonical CR1 organisation was reproduced according to the schematic structure in Kapitonov et al., 2009.

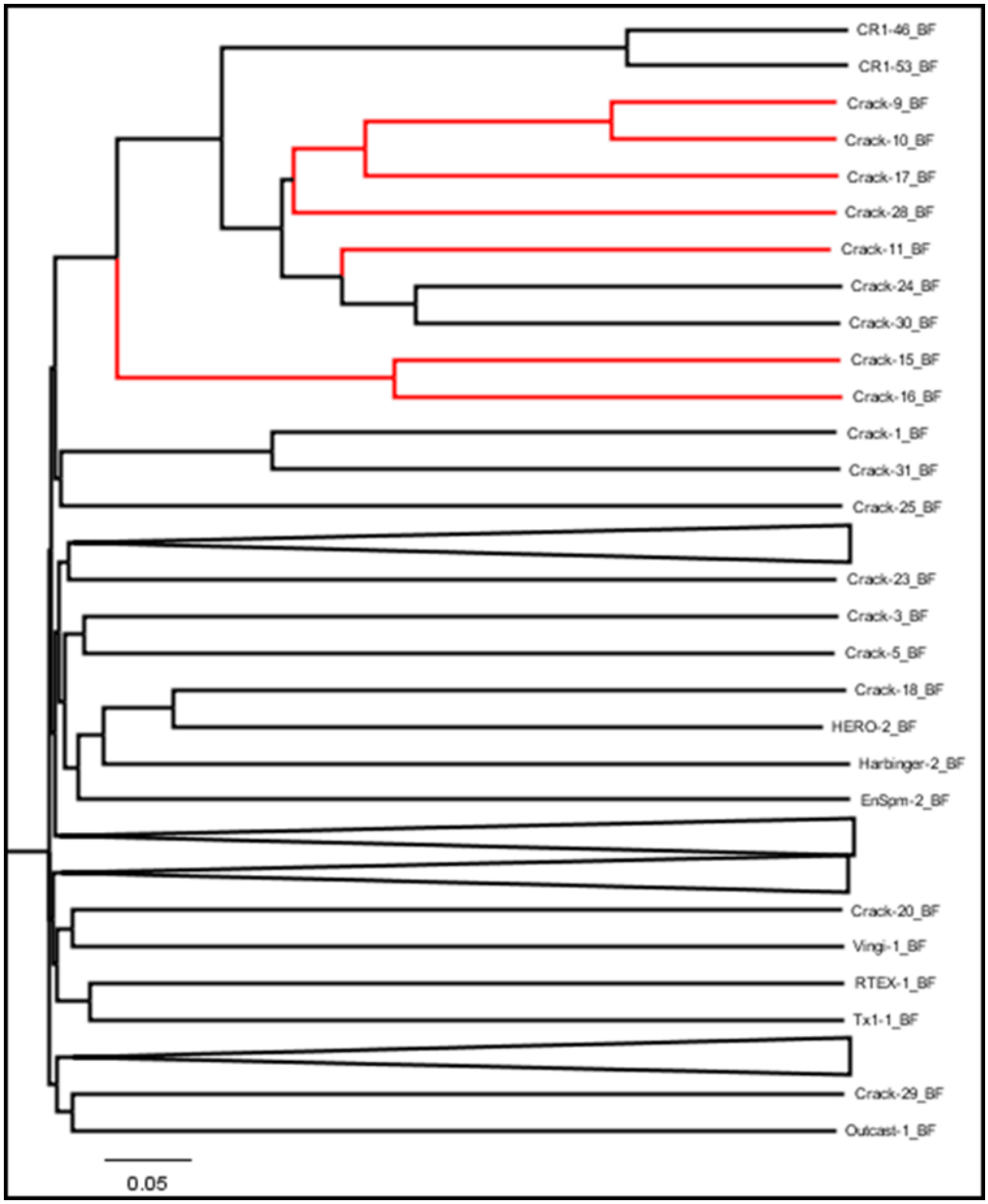


Figure 6.7: Tree showing clustering of Crack non-LTR retrotransposons containing a 2A motif

The Crack non-LTR retrotransposons containing a 2A motif are highlighted in red. The alignment includes a representative of the other clades of non-LTR retrotransposons present in the genome. The alignment was created based on the reverse transcriptase domains.

To date, the amphioxus genome is believed to host a total of thirty-two Crack elements (identified from repbase). Seven of these contain a 2A sequence at the N-terminus.

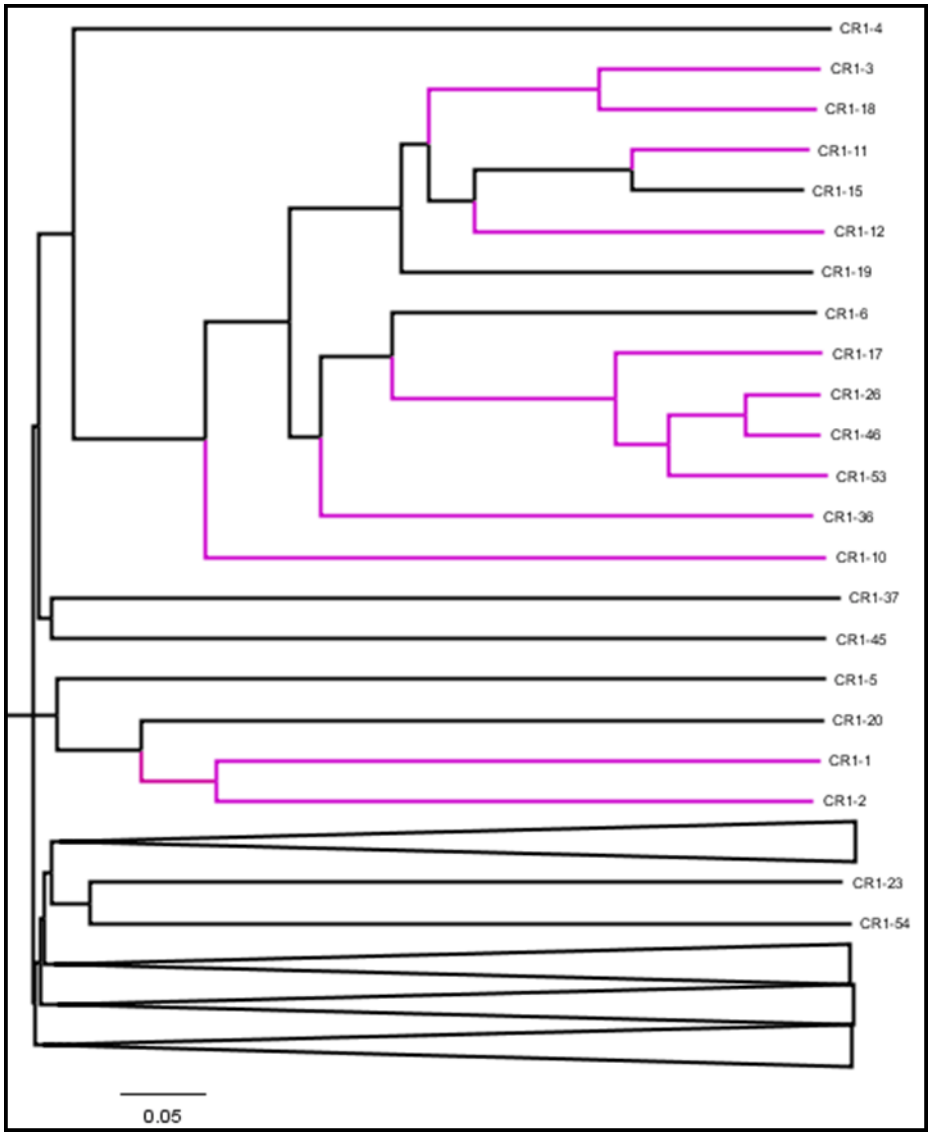


Figure 6.8: Tree showing clustering of CR1 non-LTR retrotransposons containing a 2A motif

The branches for the CR1s with a 2A sequence are highlighted in pink.

A total of fifty-six CR1 elements were identified for amphioxus at repbase. Twelve of these have a 2A element at their N-terminus detailed in figures 6.5 and 6.6.

6.3.2 Recently identified 2A in other organisms and viruses.

DTV, AsTV and OMRV1 and 2 were newly identified totiviruses. Their consensus motif resembled DHV 2A and as expected performed equally well (figure 6.9). However, the consensus motif for rotavirus A and B, which deviated from the usual motif did not exhibit any stalling ability.

For *S. kowalevskii*, only four out of the sixty-four contigs were predicted to have an identifiable protein domain following the 2A motif. One of them was tested (SK 1). Out of these four contigs, two sequences were predicted to precede a transposase and therefore be a transposable element. Searching of repbase (transposable elements database) for *S. kowalevskii* transposable elements failed to identify any transposon with a 2A sequence. The *in vitro* assay demonstrated 2A activity for SK6 and SK61 which were not predicted to precede any domains at present.

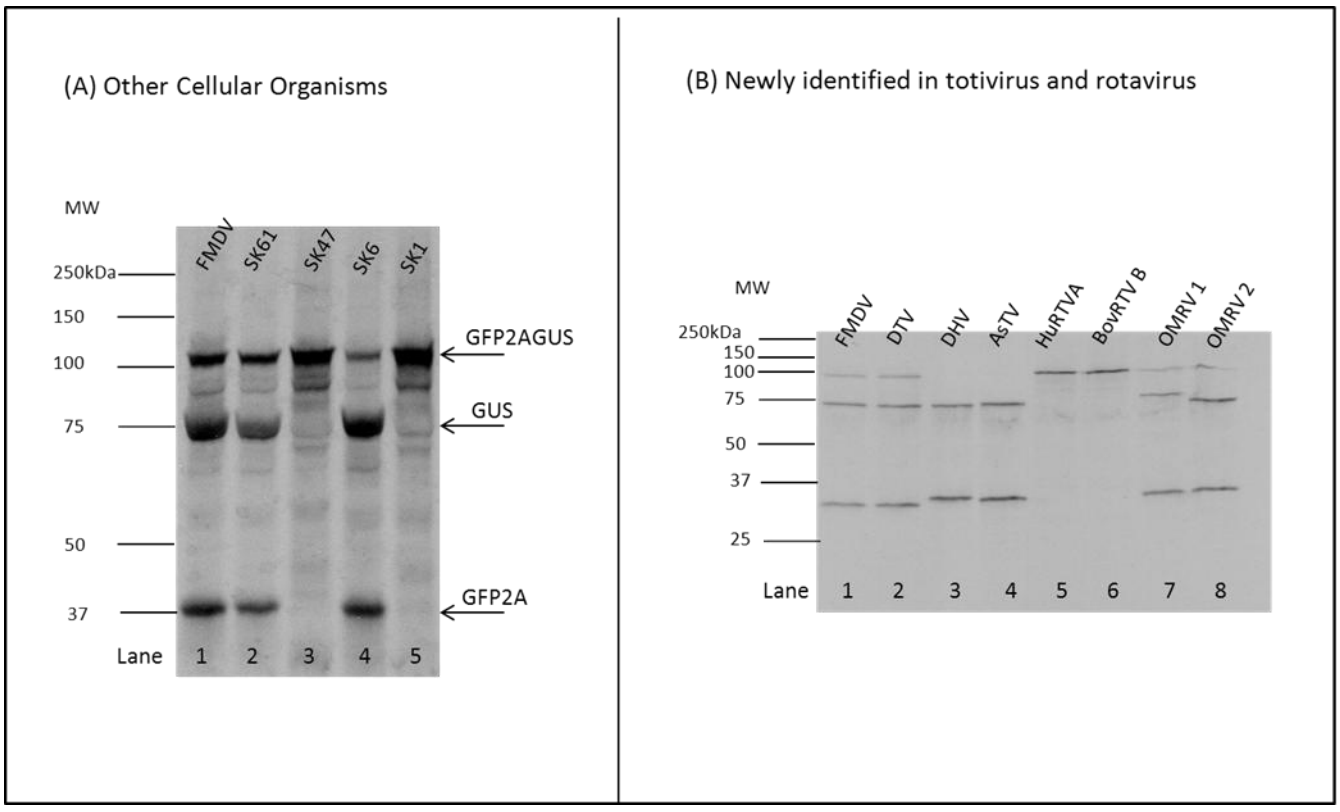


Figure 6.9: Activity assay for newly identified 2A in other cellular organisms and viruses

In (A), *S. kowalevskii* (SK, lanes 2,3,4,5) are shown. In (B), all the 2A motif (lanes 2, 4, 7 and 8) resembling DHV consensus motif (lane 3) were highly efficient, in contrast candidate motif in the genome of rotavirus A and B (lanes 5 and 6) could not perform the ribosome stalling mechanism.

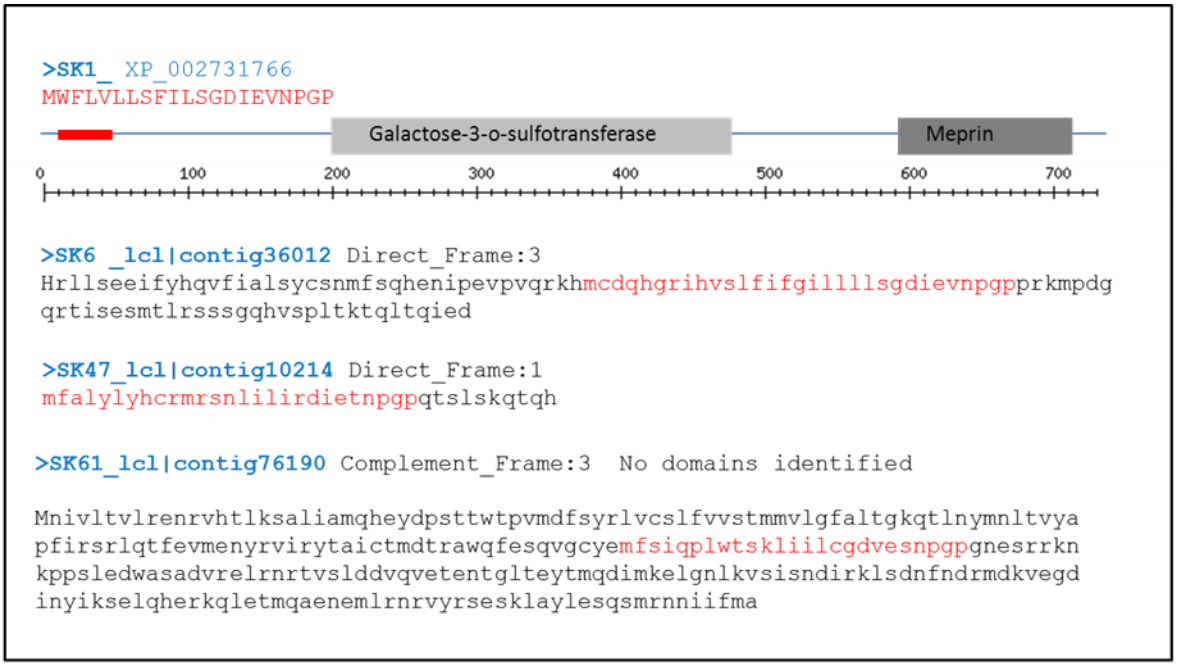


Figure 6.10: Details for the selected *S. kowalevskii* contigs with a 2A sequence tested *in vitro*

The 2A sequence is marked in red, and the accession number in blue. SK1 is the only sequence identified in NCBI (accession number XP_002731766) and the others were derived from bioinformatic analyses of contig sequences from Baylor College of Medicine.

6.4 Discussion

Identification of new sequences is based on comparison of an amino acid or nucleotide sequence in a search for homologues. This is followed by alignment of homologues and finally by a reconstruction of the history by phylogenetic analysis. In this section, the result for search by homologies resulted in the discovery of 2A motifs in cellular organisms, *B. floridae* and *S. kowalevskii*. For the vast majority of *S. kowalevskii* sequences, no protein domains were predicted following the motif; it is at present not known if these sequences are translated. The result relied on an algorithm to reverse transcribe a protein motif into the various nucleotides sequences possible and matching these to the genome of interest. Therefore the number of 2A really present in the genomes of these two organisms might be over estimated. For amphioxus, three hypothetical proteins have 2A at their N-terminus. The 2A sequences were authenticated in the activity assay, but it is not known at present if these hypothetical proteins are translated. They present no homology, unlike for the sea urchin *S. purpuratus* where all the 2A were attributed to a family of proteins: the CATERPILLER proteins involved in innate immunity (Lich and Ting, 2007).

The 2A motif was also identified at the N-terminus of two clades of non-LTR retrotransposons: Crack and CR1 in amphioxus. Non-LTR retrotransposons replicate via an RNA intermediate which they reverse transcribe and integrate in the host genome at new locations. Non-LTR retrotransposons are identified by their reverse transcriptase domains and classified into clades according to the type of additional domains they carry and the presence of one or two ORF (Malik et al., 1999). The repbase facility online is a repository of transposable elements identified in a multitude of organisms. There are currently six groups of non-LTR retrotransposons (R2, Randl, L1, RTE, I, and Jockey), further organised as clades. Of relevance to this study is the Jockey group, which consists of the clades jockey, rex1, CR1, L2, L2A, L2B, Daphne and Crack (Kapitonov et al., 2009). Although non-LTR retrotransposons are found in every eukaryotic genome sequenced so far, the type and number of copies vary greatly. Certain clades such as crack are host specific. Crack is found in sea urchin, sea squirts and amphioxus. Other clades are widely distributed, such as the CR1 clade (Kapitonov et al., 2009). It was previously established that in amphioxus the number of non-LTR retrotransposons account for less than 1% of the genome (Parmanyer et al., 2006). This study also estimated a low copy number (up to forty-two) for all the transposable elements. The 5' end of the non-LTR retrotransposon is often truncated and mutated as the reverse transcriptase often fails the transcription process at the 5' end, resulting in copies of various lengths (Kapitonov and Jurka, 2003). The CR1 clade however is considered to have two ORFs. ORF2 encodes for a reverse transcriptase and the APE domain. Kapitonov and Jurka (2003) analysed a variety of CR1 elements and concluded that the ORF1 consisted of a PHD domain (involved in protein-protein interaction) or an esterase (lipolytic acetylhydrolase). They did not conclude on the function that these two domains could serve in the

retrotransposition process and data are still lacking in that domain. What is interesting is that the 2A segment is found at the N-terminal of the retrotransposons, before the PHD domain. The CR1-2A elements have a single ORF.CR1-26 and 53 have two ORFs and the 2A is found between ORF1 and ORF2. It is likely that mutations resulted in loss of the two ORFs for the CR1-1 type of 2A non-LTR retrotransposon. It is possible that the truncation/ error prone transcription resulted in the loss of the ORF1 for the CR1-17 type of 2A non-LTRs and therefore these did not appear on the bio-informatic analyses. It is also possible for that last subgroup that the 2A element specifies a distinct type of CR1 element.

How has *B. floridae* acquired a 2A non-LTR retrotransposon? Non-LTR retrotransposons are thought to be transmitted strictly vertically (Malik et al., 1999). There is very little divergence in the 2A sequences, such as the motif and several residues preceding [NPGP] are conserved. The fact that most are not active is arguably the result of the error prone reverse transcriptase. The identified 2A sequences could be the early copies of a non-LTR retrotransposon which subsequently underwent serial deletions/mutations. Alternatively, this could constitute a late acquisition of the 2A sequence during the history of this genome. The scenario of an early acquisition which then resulted in loss seems less probable than the scenario of a recent acquisition, because CR1-2A would have been more widespread and is not present in related amphioxus species. One could also argue that the acquisition of the 2A is the result of the error prone reverse transcriptase. Acquisition however, likely involved a horizontal transfer, which presumably occurred twice in the history of that species for Crack and for CR1. The results presented here proved that these recoding elements were present in two clades of non-LTR retrotransposons in *B. floridae* however any conclusions about the origin and mode of transmission of 2A would only be hypothetical; perhaps uncovering 2A in greater number in diverse organisms as sequencing projects progress will answer some of these questions.

In terms of consensus sequences, probing databases with deviating sequences such as [EGVENPGP] identified only a few extra sequences. Other 2A motfs might exist, but it is very likely the most abundant form of 2A would be the consensus sequence [DV/IExNPGP], as previously shown deviating sequences have stricter requirements from their upstream context. Searching for 2A in transcriptomes alone will under estimate the nature and amount of 2A sequences present in organisms and would have not resulted in for instance the discovery of this element in classes of non-LTR retrotransposon. Screening genomic databases can result in a large amount of false sequences as is the case for *S. kowalevskii*. Several 2A sequences in the sea urchin CATERPILLER family have at their N-terminus a signal sequence, several are being tested and preliminary results showed a localised expression of reporter proteins depending on whether the 2A sequence is removed by stalling or remains attached. *S. purpuratus* is, therefore, a perfect model to study the expression of cellular 2As, and acquire information about the ribosomal and cellular response to its repeated use.

Concluding remarks and future work

All the stalling peptides that have been referred to throughout this thesis have unique sequences and a direct comparison is not possible. They do not necessarily employ the same route of intra-ribosomal signalling. There are however several features that are common to all of them.

- They act in cis
- They use proline at their C-terminus
- They disturb the PTC activity by interacting with the exit tunnel
- They show divergent sequences which thwarted any bioinformatics analyses aimed at identifying and predicting novel stalling sequences
- They have a limited occurence.

To acquire a better understanding of 2A-induced ribosome stalling, TaV 2A was chosen as a model for its optimum activity. Several mutations and molecular analyses were carried out and it is demonstrated here that following translation of TaV 2A, the organisation of the ribosome PTC is disturbed (chapter 5). Proline in the A site which is more constraint in its movement than any other amino acid accounts for some of the efficiency of stalling, although the disturbed PTC can not make the peptide bond between glycine and several other small amino acids in the context of TaV (chapter 5).

The understanding of the mechanism of action of TaV 2A-induced ribosome stalling then becomes the understanding of why the PTC is disturbed. Alanine scan (chapter 4) identified an 'important region' as well as several residues within the consensus motif that are critical for stalling. Beginning with the consensus sequence, what transpires is that the penultimate amino acids: [NPG] (figure 4.11) are critical. Substitutions can be tolerated at other positions within the consensus motif, for example [GDV] can be replaced by [EGV]. Alanine scan of two other 2A sequences FMDV and DHV also identified an 'important region' just preceding the consensus motif as critical for stalling. The proline experiment (figure 4.17) shows that free rotation of the 'important region' is an essential element to stalling for these three 2As.

What would be the link between a flexible nascent chain in the upper tunnel and a disturbed PTC? It is very reasonable to suggest that 2A will interact with one or several ribosome nucleotides and/or residues. The discussion in chapter 4 provides many arguments to support this hypothesis. It is especially notable that no other nascent-stalling sequences bring about arrest without several

interactions with the ribosome exit tunnel. It is therefore proposed that the flexibility required for 2A serves the purpose of bringing in proximity the key elements to their ribosomal counterparts.

The results suggest that TaV may form at least two interactions with the exit tunnel to produce the altered PTC.

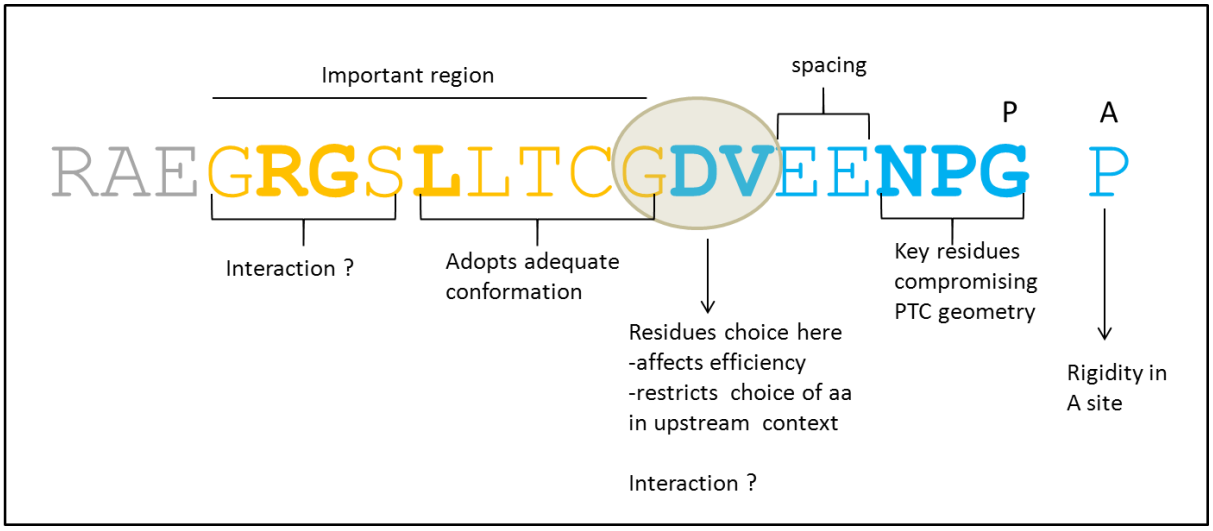
- One interaction may involve TaV arginine 5 and the tunnel. In the context of TaV, a positively charged residue is required in that position, although substitution for lysine results in a less efficient stalling sequence.

- Another interactions may involve the [GDV] region. The fact that the spacing between the [GDV] and [NPG] regions is critical indicates that the positioning of [GDV] in relation to [NPG] is critical. It is therefore proposed that [GDV] interacts with an element of the ribosome.

There are several possible scenarios; one is that the interaction(s) of TaV 2A with the ribosome induces a placement of the penultimate residues [NPG] such as to disturb one of the PTC neighbouring nucleotides like U2585 or A2062, which are known to be flexible and to influence the activity of the PTC. Alternatively, the critical position of TaV 2A in the exit tunnel is triggered by the addition of [NPG] residues to the growing peptide, in this case the intra-ribosomal partnering and positioning is introduced by the penultimate amino acids.

Do we have a single 2A mechanism or is it that each 2A induces stalling *via* a different intra-ribosomal relay? Because there is little conservation of critical residues between three 2A elements (figure 4.11): DHV, FMDV and TaV, one conclusion is that these may adopt a similar structure within the exit tunnel and arrive at that structure *via* different interactions. It may also be that they adopt different intra-ribosomal signalling routes and therefore position and partner with different ribosomal elements. The data suggest that there are at least two intra-ribosomal mechanisms one involving positively charged residues (TaV and DHV) and the other not involving that kind of residues but perhaps a α-helix stabilised by the ribosome (FMDV). The work presented here favours the hypothesis that each 2A adopts a unique conformation within the tunnel.

Several components contribute to the 2A-induced ribosome stalling (summarised in final figure). The choice of proline in the P site adds efficiency, the key [NPG] motif may directly compromise the PTC but requires upstream a sequence which could make interaction to the tunnel wall. When the consensus sequence begins with the motif [GDV], much plasticity is tolerated in the upstream context. Aside interacting residues, the neighbouring residues may simply provide contextual information to promote the appropriate placement of the 2A element within the exit tunnel.



Final figure: Dissection of TaV 2A

The consensus motif is in blue, and the upstream context divided between the important region (in orange) and the residues not involved in 2A activity (in grey). The residues in bold are critical for 2A activity.

As previously mentioned an alteration of the PTC geometry means that an element of the PTC is not positioned appropriately for peptide transfer to occur. The intra-ribosomal signalling system resulting in altered PTC geometry could be compared to a domino effect. The disturbance of a nucleotide from the tunnel displaces several residues along the way up to the PTC. For SecM, it has been proposed that physical force alleviates stalling. The N-terminus of SecM has a unique signal sequence that preferentially binds to SecA. The distance between this region to the stalling motif is critical for SecM activity. A shorter length abolished SecM activity and a longer length resulted in more release and less stalling (Yap and Bernstein, 2011). The authors concluded that the hydrophobic signal sequence acts as a timer. Therefore their theory favours a physical movement of the peptide within the tunnel to reset the ribosome function after stalling, it is also notable that the ribosome encounters a stop peptide after the Gly-Pro pair. The signal sequence of SecM serves three functions, to recruit preferentially SecA, to localize the SecA synthesis (in the ORF2) to the site where it will be needed and to reset the ribosome after SecM stalling. The 2A motif is always found at the N-terminus of all the cellular sequences listed to date containing it. It is also an interesting finding that some Sea Urchin 2A sequences have a signal sequence at their N-terminus (Dr Sukhodub, unpublished). Whether the signal sequence also serves other purposes than protein targeting for these 2As has not yet been tested. What could reset the ribosome so it can resume its translation past the 2A-stalling? Can eRF1 and 3 whose duties is to lyse the ester bond be the only factor that enables the correct repositioning of the PTC and the disturbed residues in the exit tunnel? All the A site residues tested for TaV 2A retained stalling abilities and re-initiation of translation. Therefore eRF1 and 3 succesfully rescued the stalled complex with structuraly different A site mRNAs, how can eRF1 and 3 bind to very structurally

different sequences? These are two of the many unanswered questions regarding the role eRF1 and 3 play in relation to 2A-induced ribosome stalling. Because their presumed function in the context of 2A goes against the well established notion that these factors have evolved to accommodate only stop codons, it is critical that this is investigated in a reconstituted translation system.

Two aims were not achieved: the structural resolution of 2A and proof for eRF1 and 3 direct interventions in rescuing the 2A-stalled ribosome. The work carried on these two instances served as negative results to pave the way to more successful experiments. In the light of the dynamic nature of the ribosome, the un-uniform environment within the upper exit tunnel, and the fact that 2A sequences may adopt at least two mechanisms to stall the ribosome, it will be more informative to resolve the structure of a 2A-bound ribosome by cryoEM than to purify and study these in a uniform solution. To dissect the molecular aspects of the 2A mechanism, a synthetic eukaryotic translation system will provide the necessary platform to carry out cross-linking studies, mutations, additions and omissions of various factors. Mutagenesis of the six proposed model 2As and their comparison will complete this study as well as further non conservative mutagenesis of TaV 2A important region.

2A and other stalling sequences prove that translation is also modulated by the sequence of amino acid progressing in the ribosome exit tunnel. The stalling sequences can therefore be a primary platform to study and map the type of interaction possible between the nascent peptide and the ribosome exit tunnel.

In memory of my beloved Maman,
The gentlest star in my Night-Sky

Acknowledgements

I would like to thank the following scientists for contributing reagents and protocols.

In the picornavirus laboratory, Dr Luke contributed plasmids pSTA1 plasmids and pJC3 and protocols for cloning and rabbit reticulocyte lysates expression. Miss Sarah Conner offered a protocol and training for mammalian cell cultures, and western blot analysis. Dr Minskaia offered a protocol for mounting cells for microscopy.

Prof. Naismith laboratory kindly contributed BL21, BL21*, C43, TUNR, Origami and Rosetta bacterial strains as well as plasmid pEHISGFPTEV for protein expression. Thank you to Su-Yin Kan and Prof Naismith laboratory for the protocols for protein expression and valuable discussions.

Prof. Randall laboratory kindly provided the lentivirus vectors A5, pVSVG and pCMV. 293T cells, V5 antibody, EMCV virus stock, as well as a protocol for lentivirus infection of mammalian cells and MTT assay.

Prof. Elliott laboratory gifted HeLa cells. Prof. Bedwell contributed a dual renilla-luciferase plasmid.

I would like to thank the following scientists for discussions and ideas.

Prof. Ryan designed the visual system to explore the effect of stress on 2A activity, and suggested expressing the translation-regulating factors in bacteria and expressing 2A in *Pichia pastoris*. Prof Ryan laboratory contributed the translation factors TOPO plasmids, BHK21, EBL cells, pHE27 and all other consumables.

Dr Ulrich Schwarz-Linek assisted with valuable suggestions regarding *P. pastoris* culture and with critical reviews of my work.

My studentship was financed by BBSRC.

A personal acknowledgement

I would like to express my most sincere gratitude to my grand-ma Alexina Lia Roset, who worked more than is humanely acceptable to lift her children out of poverty. And I would like to thank my mum Denise Marie-Felicia Roset for her devoted love.What a privilege it has been for me to be part of her family. Whoever may believe this will….. but I am convinced that wherever their Souls may be, these women smile approvingly. And to my aunt Jeanine Roset, thank you for staying behind!

References

ADAMS, M. J., LEFKOWITZ, E. J., KING, A. M. & CARSTENS, E. B. 2013. Recently agreed changes to the International Code of Virus Classification and Nomenclature. *Arch Virol,* 158**,** 2633-9.

AKIMITSU, N., TANAKA, J. & PELLETIER, J. 2007. Translation of nonSTOP mRNA is repressed post-initiation in mammalian cells. *EMBO J,* 26, 2327-38.

ALKALAEVA, E. Z., PISAREV, A. V., FROLOVA, L. Y., KISSELEV, L. L. & PESTOVA, T. V. 2006. In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell,* 125, 1125-36.

ARNOLD, M. M. & PATTON, J. T. 2011. Diversity of interferon antagonist activities mediated by NSP1 proteins of different rotavirus strains. *J Virol,* 85, 1970-9.

BALTIMORE, D. 1971. Expression of animal virus genomes. *Bacteriol Rev,* 35, 235-41.

BARCO, A., FEDUCHI, E. & CARRASCO, L. 2000. Poliovirus protease 3C(pro) kills cells by apoptosis. *Virology,* 266, 352-60.

BASHAN, A., AGMON, I., ZARIVACH, R., SCHLUENZEN, F., HARMS, J., BERISIO, R., BARTELS, H., FRANCESCHI, F., AUERBACH, T., HANSEN, H. A., KOSSOY, E., KESSLER, M. & YONATH, A. 2003. Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression. *Mol Cell,* 11, 91-102.

BELSHAM, G. J., BRANGWYN, J. K., RYAN, M. D., ABRAMS, C. C. & KING, A. M. 1990. Intracellular expression and processing of foot-and-mouth disease virus capsid precursors using vaccinia virus vectors: influence of the L protease. *Virology,* 176, 524-30.

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. *J Virol,* 74, 272-80.

BERINGER, M., BRUELL, C., XIONG, L., PFISTER, P., BIELING, P., KATUNIN, V. I., MANKIN, A. S., BOTTGER, E. C. & RODNINA, M. V. 2005. Essential mechanisms in the catalysis of peptide bond formation on the ribosome. *J Biol Chem,* 280, 36065-72.

BHUSHAN, S., GARTMANN, M., HALIC, M., ARMACHE, J. P., JARASCH, A., MIELKE, T., BERNINGHAUSEN, O., WILSON, D. N. & BECKMANN, R. 2010b. alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat Struct Mol Biol,* 17, 313-7.

BHUSHAN, S., HOFFMANN, T., SEIDELT, B., FRAUENFELD, J., MIELKE, T., BERNINGHAUSEN, O., WILSON, D. N. & BECKMANN, R. 2011. SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biol,* 9, e1000581.

BHUSHAN, S., MEYER, H., STAROSTA, A. L., BECKER, T., MIELKE, T., BERNINGHAUSEN, O., SATTLER, M., WILSON, D. N. & BECKMANN, R. 2010. Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol Cell,* 40, 138-46.

BIELING, P., BERINGER, M., ADIO, S. & RODNINA, M. V. 2006. Peptide bond formation does not involve acid-base catalysis by ribosomal residues. *Nat Struct Mol Biol,* 13, 423-8.

BLAIR, W. S., PARSLEY, T. B., BOGERD, H. P., TOWNER, J. S., SEMLER, B. L. & CULLEN, B. R. 1998. Utilization of a mammalian cell-based RNA binding assay to characterize the RNA binding properties of picornavirus 3C proteinases. *RNA,* 4, 215-25.

BLANCHARD, S. C., GONZALEZ, R. L., KIM, H. D., CHU, S. & PUGLISI, J. D. 2004. tRNA selection and kinetic proofreading in translation. *Nat Struct Mol Biol,* 11, 1008-14.

BLOBEL, G. & SABATINI, D. D. 1970. Controlled proteolysis of nascent polypeptides in rat liver cell fractions. I. Location of the polypeptides within ribosomes. *J Cell Biol,* 45, 130-45.

BRUNELLE, J. L., YOUNGMAN, E. M., SHARMA, D. & GREEN, R. 2006. The interaction between C75 of tRNA and the A loop of the ribosome stimulates peptidyl transferase activity. *RNA,* 12, 33-9.

BUCHSCHACHER, G. L., JR. & WONG-STAAL, F. 2000. Development of lentiviral vectors for gene therapy for human diseases. *Blood,* 95, 2499-504.

HAYES, C. S., BOSE, B. & SAUER, R. T. 2002. Proline residues at the C terminus of nascent chains induce SsrA tagging during translation termination. *J Biol Chem,* 277, 33825-32.

HELLEN, C. U., LEE, C. K. & WIMMER, E. 1992. Determinants of substrate recognition by poliovirus 2A proteinase. *J Virol,* 66, 3330-8.

HESSLEIN, A. E., KATUNIN, V. I., BERINGER, M., KOSEK, A. B., RODNINA, M. V. & STROBEL, S. A. 2004. Exploration of the conserved A+C wobble pair within the ribosomal peptidyl transferase center using affinity purified mutant ribosomes. *Nucleic Acids Res,* 32, 3760-70.

HOLCIK, M. & SONENBERG, N. 2005. Translational control in stress and apoptosis. *Nat Rev Mol Cell Biol,* 6, 318-27.

HUBER, D., RAJAGOPALAN, N., PREISSLER, S., ROCCO, M. A., MERZ, F., KRAMER, G. & BUKAU, B. 2011. SecA interacts with ribosomes in order to facilitate posttranslational translocation in bacteria. *Mol Cell,* 41, 343-53.

HUGHES, P. J. & STANWAY, G. 2000. The 2A proteins of three diverse picornaviruses are related to each other and to the H-rev107 family of proteins involved in the control of cell proliferation. *J Gen Virol,* 81, 201-7.

IBORRA, F. J., JACKSON, D. A. & COOK, P. R. 2001. Coupled transcription and translation within nuclei of mammalian cells. *Science,* 293, 1139-42.

INGOLIA, N. T., LAREAU, L. F. & WEISSMAN, J. S. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell,* 147, 789-802.

ISAWA, H., ASANO, S., SAHARA, K., IIZUKA, T. & BANDO, H. 1998. Analysis of genetic information of an insect picorna-like virus, infectious flacherie virus of silkworm: evidence for evolutionary relationships among insect, mammalian and plant picorna(-like) viruses. *Arch Virol,* 143, 127-43.

ISAWA, H., KUWATA, R., HOSHINO, K., TSUDA, Y., SAKAI, K., WATANABE, S., NISHIMURA, M., SATHO, T., KATAOKA, M., NAGATA, N., HASEGAWA, H., BANDO, H., YANO, K., SASAKI, T., KOBAYASHI, M., MIZUTANI, T. & SAWABE, K.

2011. Identification and molecular characterization of a new nonsegmented double-stranded RNA virus isolated from Culex mosquitoes in Japan. *Virus Res,* 155, 147-55.

JACKSON, R. J. 1986. A detailed kinetic analysis of the in vitro synthesis and processing of encephalomyocarditis virus products. *Virology,* 149, 114-27.

JAMES, V. L., LAMBDEN, P. R., DENG, Y., CAUL, E. O. & CLARKE, I. N. 1999. Molecular characterization of human group C rotavirus genes 6, 7 and 9. *J Gen Virol,* 80 (Pt 12), 3181-7.

JAYARAM, H., ESTES, M. K. & PRASAD, B. V. 2004. Emerging themes in rotavirus cell entry, genome organization, transcription and replication. *Virus Res,* 101, 67-81.

JEN, G. & THACH, R. E. 1982. Inhibition of host translation in encephalomyocarditis virus-infected L cells: a novel mechanism. *J Virol,* 43, 250-61.

JENNER, L., DEMESHKINA, N., YUSUPOVA, G. & YUSUPOV, M. 2010. Structural rearrangements of the ribosome at the tRNA proofreading step. *Nat Struct Mol Biol,* 17, 1072-8.

KAPITONOV, V. V. & JURKA, J. 2003. The esterase and PHD domains in CR1-like non-LTR retrotransposons. *Mol Biol Evol,* 20, 38-46.

KAPITONOV, V. V., TEMPEL, S. & JURKA, J. 2009. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. *Gene,* 448, 207-13.

KAZANTSEV, A. V. & PACE, N. R. 2006. Bacterial RNase P: a new view of an ancient enzyme. *Nat Rev Microbiol,* 4, 729-40.

KERKVLIET, J., EDUKULLA, R. & RODRIGUEZ, M. 2010. Novel roles of the picornaviral 3D polymerase in viral pathogenesis. *Adv Virol,* 2010, 368068.

KIM, D. F. & GREEN, R. 1999. Base-pairing between 23S rRNA and tRNA in the ribosomal A site. *Mol Cell,* 4, 859-64.

KOLUPAEVA, V. G., PESTOVA, T. V., HELLEN, C. U. & SHATSKY, I. N. 1998. Translation eukaryotic initiation factor 4G recognizes a specific structural element within the internal ribosome entry site of encephalomyocarditis virus RNA. *J Biol Chem*, 273, 18599-604.

KÖNIG, H. R., B. 1988 Purification and partial characterization of poliovirus protease 2A by means of a functional assay. *J Virol*, 62, 1243-1250.

KRYUCHKOVA, P., GRISHIN, A., ELISEEV, B., KARYAGINA, A., FROLOVA, L. & ALKALAEVA, E. 2013. Two-step model of stop codon recognition by eukaryotic release factor eRF1. *Nucleic Acids Res*, 41, 4573-86.

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. *J Gen Virol*, 79 (Pt 8), 1911-21.

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. *J Virol*, 68, 3821-9.

LICH, J. D. & TING, J. P. 2007. CATERPILLER (NLR) family members as positive and negative regulators of inflammatory responses. *Proc Am Thorac Soc*, 4, 263-6.

LIM, V. I. & SPIRIN, A. S. 1986. Stereochemical analysis of ribosomal transpeptidation. Conformation of nascent peptide. *J Mol Biol*, 188, 565-74.

LIU, H. & NAISMITH, J. H. 2009. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr Purif*, 63, 102-11.

LOVETT, P. S. & ROGERS, E. J. 1996. Ribosome regulation by the nascent peptide. *Microbiol Rev*, 60, 366-85.

LU, J., KOBERTZ, W. R. & DEUTSCH, C. 2007. Mapping the electrostatic potential within the ribosomal exit tunnel. *J Mol Biol*, 371, 1378-91.

LUKE, G. A., DE FELIPE, P., LUKASHEV, A., KALLIOINEN, S. E., BRUNO, E. A. & RYAN, M. D. 2008. Occurrence, function and evolutionary origins of '2A-like' sequences in virus genomes. *J Gen Virol*, 89, 1036-42.

MALIK, H. S., BURKE, W. D. & EICKBUSH, T. H. 1999. The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol*, 16, 793-805.

MALKIN, L. I. & RICH, A. 1967. Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *J Mol Biol,* 26, 329-46.

MARTIN, A. B., D.; CHAO, S.F.; COHEN,L.M.; & LEMON, S. M. 1999. Maturation of the Hepatitis A Virus Capsid Protein VP1 Is Not Dependent on Processing by the 3Cpro Proteinase. *J Virol,* 73, 6220–6227.

MARTIN, A. E., N.; CHAO,S.F.; LEMON,S.M.; GIRARD,M.; WYCHOWASKI, 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 transcripts. *Virology*, 213, 213-222.

MARTINEZ, A. K., SHIROLE, N. H., MURAKAMI, S., BENEDIK, M. J., SACHS, M. S. & CRUZ-VERA, L. R. 2012. Crucial elements that maintain the interactions between the regulatory TnaC peptide and the ribosome exit tunnel responsible for Trp inhibition of ribosome function. *Nucleic Acids Res,* 40, 2247-57.

MOAZED, D. & NOLLER, H. F. 1989. Intermediate states in the movement of transfer RNA in the ribosome. *Nature,* 342, 142-8.

MUTO, H. & ITO, K. 2008. Peptidyl-prolyl-tRNA at the ribosomal P-site reacts poorly with puromycin. *Biochem Biophys Res Commun,* 366, 1043-7.

MUTO, H., NAKATOGAWA, H. & ITO, K. 2006. Genetically encoded but nonpolypeptide prolyl-tRNA functions in the A site for SecM-mediated ribosomal stall. *Mol Cell,* 22, 545-52.

NAKATOGAWA, H. & ITO, K. 2002. The ribosomal exit tunnel functions as a discriminating gate. *Cell,* 108, 629-36.

NAPTHINE, S., LEVER, R. A., POWELL, M. L., JACKSON, R. J., BROWN, T. D. & BRIERLEY, I. 2009. Expression of the VP2 protein of murine norovirus by a translation termination-reinitiation strategy. *PLoS One,* 4, e8390.

NIBERT, M. L. 2007. '2A-like' and 'shifty heptamer' motifs in penaeid shrimp infectious myonecrosis virus, a monosegmented double-stranded RNA virus. *J Gen Virol*, 88, 1315-8.

NICKLIN, M. J., KRAUSSLICH, H. G., TOYODA, H., DUNN, J. J. & WIMMER, E. 1987. Poliovirus polypeptide precursors: expression in vitro and processing by exogenous 3C and 2A proteinases. *Proc Natl Acad Sci U S A*, 84, 4002-6.

NISSEN, P., HANSEN, J., BAN, N., MOORE, P. B. & STEITZ, T. A. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289, 920-30.

NOLLER, H. F. 2012. Evolution of protein synthesis from an RNA world. *Cold Spring Harb Perspect Biol*, 4, a003681.

NOVOA, I. & CARRASCO, L. 1999. Cleavage of eukaryotic translation initiation factor 4G by exogenously added hybrid proteins containing poliovirus 2Apro in HeLa cells: effects on gene expression. *Mol Cell Biol*, 19, 2445-54.

OGLE, J. M., CARTER, A. P. & RAMAKRISHNAN, V. 2003. Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem Sci*, 28, 259-66.

OGLE, J. M., MURPHY, F. V., TARRY, M. J. & RAMAKRISHNAN, V. 2002. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell*, 111, 721-32.

PALMENBERG, A. C. 1990. Proteolytic processing of picornaviral polyprotein. *Annu Rev Microbiol*, 44, 603-23.

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, 754-62.

PALMENBERG, A. C., SPIRO, D., KUZMICKAS, R., WANG, S., DJIKENG, A., RATHE, J. A., FRASER-LIGGETT, C. M. & LIGGETT, S. B. 2009. Sequencing and analyses of all known human rhinovirus genomes reveal structure and evolution. *Science*, 324, 55-9.

PAVLOV, M. Y., WATTS, R. E., TAN, Z., CORNISH, V. W., EHRENBERG, M. & FORSTER, A. C. 2009. Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc Natl Acad Sci U S A*, 106, 50-4.

PERMANYER, J., ALBALAT, R. & GONZALEZ-DUARTE, R. 2006. Getting closer to a pre-vertebrate genome: the non-LTR retrotransposons of Branchiostoma floridae. *Int J Biol Sci*, 2, 48-53.

PESTKA, S., VINCE, R., DALUGE, S. & HARRIS, R. 1973. Effect of puromycin analogues and other agents on peptidyl-puromycin synthesis on polyribosomes. *Antimicrob Agents Chemother*, 4, 37-43.

PISAREVA, V. P., SKABKIN, M. A., HELLEN, C. U., PESTOVA, T. V. & PISAREV, A. V. 2011. Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. *EMBO J*, 30, 1804-17.

POLACEK, N., GAYNOR, M., YASSIN, A. & MANKIN, A. S. 2001. Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide. *Nature*, 411, 498-501.

POULOS, B. T., TANG, K. F., PANTOJA, C. R., BONAMI, J. R. & LIGHTNER, D. V. 2006. Purification and characterization of infectious myonecrosis virus of penaeid shrimp. *J Gen Virol*, 87, 987-96.

PRINGLE, F. M., KALMAKOFF, J. & WARD, V. K. 2001. Analysis of the capsid processing strategy of Thosea asigna virus using baculovirus expression of virus-like particles. *J Gen Virol*, 82, 259-66.

RAMU, H., MANKIN, A. & VAZQUEZ-LASLOP, N. 2009. Programmed drug-dependent ribosome stalling. *Mol Microbiol*, 71, 811-24.

RAMU, H., VAZQUEZ-LASLOP, N., KLEPACKI, D., DAI, Q., PICCIRILLI, J., MICURA, R. & MANKIN, A. S. 2011. Nascent peptide in the ribosome exit tunnel affects functional properties of the A-site of the peptidyl transferase center. *Mol Cell*, 41, 321-30.

REDONDO, N., SANZ, M. A., WELNOWSKA, E. & CARRASCO, L. 2011. Translation without eIF2 promoted by poliovirus 2A protease. *PLoS One*, 6, e25699.

REDPATH, N. T., PRICE, N. T., SEVERINOV, K. V. & PROUD, C. G. 1993. Regulation of elongation factor-2 by multisite phosphorylation. *Eur J Biochem*, 213, 689-99.

RIBIERE, M., OLIVIER, V. & BLANCHARD, P. 2010. Chronic bee paralysis: a disease and a virus like no other? *J Invertebr Pathol*, 103 Suppl 1, S120-31.

RODNINA, M. V. 2013. The ribosome as a versatile catalyst: reactions at the peptidyl transferase center. *Curr Opin Struct Biol*, 23, 595-602.

RODNINA, M. V., GROMADSKI, K. B., KOTHE, U. & WIEDEN, H. J. 2005. Recognition and selection of tRNA in translation. *FEBS Lett*, 579, 938-42.

RODRIGUEZ, P. L. & CARRASCO, L. 1993. Poliovirus protein 2C has ATPase and GTPase activities. *J Biol Chem*, 268, 8105-10.

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

ROOS, R. P., KONG, W. P. & SEMLER, B. L. 1989. Polyprotein processing of Theiler's murine encephalomyelitis virus. *J Virol*, 63, 5344-53.

RYAN, M. D., BELSHAM, G. J. & KING, A. M. 1989. Specificity of enzyme-substrate interactions in foot-and-mouth disease virus polyprotein processing. *Virology*, 173, 35-45.

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. & DREW, J. 1994. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J*, 13, 928-33.

RYAN, M. D., KING, A. M. & THOMAS, G. P. 1991. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J Gen Virol*, 72 (Pt 11), 2727-32.

SALAS-MARCO, J. & BEDWELL, D. M. 2004. GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. *Mol Cell Biol*, 24, 7769-78.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. 1989. *Molecular Cloning. A Laboratory Manual (2nd ed.)*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).

SANBONMATSU, K. Y., JOSEPH, S. & TUNG, C. S. 2005. Simulating movement of tRNA into the ribosome during decoding. *Proc Natl Acad Sci U S A,* 102, 15854-9.

SANCHEZ-MARTINEZ, S., MADAN, V., CARRASCO, L. & NIEVA, J. L. 2012. Membrane-active peptides derived from picornavirus 2B viroporin. *Curr Protein Pept Sci,* 13, 632-43.

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. *J Virol,* 71, 4679-93.

SARKER, S. & OLIVER, D. 2002. Critical regions of secM that control its translation and secretion and promote secretion-specific secA regulation. *J Bacteriol,* 184, 2360-9.

SCHEIN, C. H., OEZGUEN, N., VOLK, D. E., GARIMELLA, R., PAUL, A. & BRAUN, W. 2006. NMR structure of the viral peptide linked to the genome (VPg) of poliovirus. *Peptides,* 27**,** 1676-84.

SCHMEING, T. M., HUANG, K. S., KITCHEN, D. E., STROBEL, S. A. & STEITZ, T. A. 2005. Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. *Mol Cell,* 20, 437-48.

SCHMEING, T. M., HUANG, K. S., STROBEL, S. A. & STEITZ, T. A. 2005b. An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature,* 438, 520-4.

SEIDELT, B., INNIS, C. A., WILSON, D. N., GARTMANN, M., ARMACHE, J. P., VILLA, E., TRABUCO, L. G., BECKER, T., MIELKE, T., SCHULTEN, K., STEITZ, T. A. & BECKMANN, R. 2009. Structural insight into nascent polypeptide chain-mediated translational stalling. *Science,* 326, 1412-5.

SHANER, N. C., STEINBACH, P. A. & TSIEN, R. Y. 2005. A guide to choosing fluorescent proteins. *Nat Methods,* 2, 905-9.

SHAW, J. J. & GREEN, R. 2007. Two distinct components of release factor function uncovered by nucleophile partitioning analysis. *Mol Cell,* 28, 458–467.

SIEVERS, A., BERINGER, M., RODNINA, M. V. & WOLFENDEN, R. 2004. The ribosome as an entropy trap. *Proc Natl Acad Sci U S A,* 101, 7897-901.

SOMMERGRUBER, W., SEIPELT, J., FESSL, F., SKERN, T., LIEBIG, H. D. & CASARI, G. 1997. Mutational analyses support a model for the HRV2 2A proteinase. *Virology,* 234, 203-14.

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, 311-21.

SOTO-RIFO, R. & OHLMANN, T. 2013. The role of the DEAD-box RNA helicase DDX3 in mRNA metabolism. *Wiley Interdiscip Rev RNA,* 4, 369-85.

SPEVAK, C. C., IVANOV, I. P. & SACHS, M. S. 2010. Sequence requirements for ribosome stalling by the arginine attenuator peptide. *J Biol Chem,* 285, 40933-42.

STEIL, B. P., KEMPF, B. J. & BARTON, D. J. 2010. Poly(A) at the 3' end of positive-strand RNA and VPg-linked poly(U) at the 5' end of negative-strand RNA are reciprocal templates during replication of poliovirus RNA. *J Virol,* 84, 2843-58.

SVITKIN, Y. V. & AGOL, V. I. 1983. Translational barrier in central region of encephalomyocarditis virus genome. Modulation by elongation factor 2 (eEF-2). *Eur J Biochem,* 133, 145-54.

TAYLOR, D. J., DEVKOTA, B., HUANG, A. D., TOPF, M., NARAYANAN, E., SALI, A., HARVEY, S. C. & FRANK, J. 2009. Comprehensive Molecular Structure of the Eukaryotic Ribosome. *Structure,* 17, 1591.

TENSON, T., LOVMAR, M. & EHRENBERG, M. 2003. The mechanism of action of macrolides, lincosamides and streptogramin B reveals the nascent peptide exit path in the ribosome. *J Mol Biol,* 330, 1005-14.

THOMPSON, J., KIM, D. F., O'CONNOR, M., LIEBERMAN, K. R., BAYFIELD, M. A., GREGORY, S. T., GREEN, R., NOLLER, H. F. & DAHLBERG, A. E. 2001. Analysis of

mutations at residues A2451 and G2447 of 23S rRNA in the peptidyltransferase active site of the 50S ribosomal subunit. *Proc Natl Acad Sci U S A,* 98, 9002-7.

TOYODA, H., NICKLIN, M. J., MURRAY, M. G., ANDERSON, C. W., DUNN, J. J., STUDIER, F. W. & WIMMER, E. 1986. A second virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein. *Cell,* 45, 761-70.

TU, L. W. & DEUTSCH, C. 2010. A folding zone in the ribosomal exit tunnel for Kv1.3 helix formation. *J Mol Biol,* 396, 1346-60.

UDE, S., LASSAK, J., STAROSTA, A. L., KRAXENBERGER, T., WILSON, D. N. & JUNG, K. 2013. Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science,* 339, 82-5.

VALLE, M., ZAVIALOV, A., SENGUPTA, J., RAWAT, U., EHRENBERG, M. & FRANK, J. 2003. Locking and unlocking of ribosomal motions. *Cell,* 114, 123-34.

VANCE, L. M., MOSCUFO, N., CHOW, M. & HEINZ, B. A. 1997. Poliovirus 2C region functions during encapsidation of viral RNA. *J Virol,* 71, 8759-65.

VAZQUEZ-LASLOP, N., RAMU, H., KLEPACKI, D., KANNAN, K. & MANKIN, A. S. 2010. The key function of a conserved and modified rRNA residue in the ribosomal response to the nascent peptide. *EMBO J,* 29, 3108-17.

VAZQUEZ-LASLOP, N., THUM, C. & MANKIN, A. S. 2008. Molecular mechanism of drug-dependent ribosome stalling. *Mol Cell,* 30, 190-202.

VOORHEES, R. M., WEIXLBAUMER, A., LOAKES, D., KELLEY, A. C. & RAMAKRISHNAN, V. 2009. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat Struct Mol Biol,* 16, 528-33.

VOSS, N. R., GERSTEIN, M., STEITZ, T. A. & MOORE, P. B. 2006. The geometry of the ribosomal polypeptide exit tunnel. *J Mol Biol,* 360, 893-906.

WALTER, C. T., PRINGLE, F. M., NAKAYINGA, R., DE FELIPE, P., RYAN, M. D., BALL, L. A. & DORRINGTON, R. A. 2010. Genome organization and translation products of Providence virus: insight into a unique tetravirus. *J Gen Virol,* 91, 2826-35.

WANG, Z. & SACHS, M. S. 1997. Ribosome stalling is responsible for arginine-specific translational attenuation in Neurospora crassa. *Mol Cell Biol,* 17, 4904-13.

WANG, X., ZHANG, J., LU, J., YI, F., LIU, C. & HU, Y. 2004. Sequence analysis and genomic organization of a new insect picorna-like virus, Ectropis obliqua picorna-like virus, isolated from Ectropis obliqua. *J Gen Virol,* 85, 1145-51.

WEI, J., WU, C. & SACHS, M. S. 2012. The arginine attenuator peptide interferes with the ribosome peptidyl transferase center. *Mol Cell Biol,* 32, 2396-406.

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, 260-71.

WOHLGEMUTH, I., BRENNER, S., BERINGER, M. & RODNINA, M. V. 2008. Modulation of the rate of peptidyl transfer on the ribosome by the nature of substrates. *J Biol Chem,* 283, 32229-35.

WOOLHEAD, C. A., JOHNSON, A. E. & BERNSTEIN, H. D. 2006. Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol Cell,* 22, 587-98.

WOOLSTENHULME, C. J., PARAJULI, S., HEALEY, D. W., VALVERDE, D. P., PETERSEN, E. N., STAROSTA, A. L., GUYDOSH, N. R., JOHNSON, W. E., WILSON, D. N. & BUSKIRK, A. R. 2013. Nascent peptides that block protein synthesis in bacteria. *Proc Natl Acad Sci U S A,* 110, E878-87.

WU, C. Y., LO, C. F., HUANG, C. J., YU, H. T. & WANG, C. H. 2002. The complete genome sequence of Perina nuda picorna-like virus, an insect-infecting RNA virus with a genome organization similar to that of the mammalian picornaviruses. *Virology,* 294, 312-23.

YANOFSKY, C. 2007. RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. *RNA,* 13, 1141-54.

YAP, M. N. & BERNSTEIN, H. D. 2009. The plasticity of a translation arrest motif yields insights into nascent polypeptide recognition inside the ribosome tunnel. *Mol Cell,* 34, 201-11.

YAP, M. N. & BERNSTEIN, H. D. 2011. The translational regulatory function of SecM requires the precise timing of membrane targeting. *Mol Microbiol*, 81, 540–553.

YONATH, A., LEONARD, K. R. & WITTMANN, H. G. 1987. A tunnel in the large ribosomal subunit revealed by three-dimensional image reconstruction. *Science,* 236, 813-6.

YOUNGMAN, E. M., BRUNELLE, J. L., KOCHANIAK, A. B. & GREEN, R. 2004. The active site of the ribosome is composed of two layers of conserved nucleotides with distinct roles in peptide bond formation and peptide release. *Cell,* 117, 589-99.

YUSUPOV, M. M., YUSUPOVA, G. Z., BAUCOM, A., LIEBERMAN, K., EARNEST, T. N., CATE, J. H. & NOLLER, H. F. 2001. Crystal structure of the ribosome at 5.5 A resolution. *Science,* 292, 883-96.

ZAVIALOV, A. V. & EHRENBERG, M. 2003. Peptidyl-tRNA regulates the GTPase activity of translation factors. *Cell,* 114, 113-22.

ZHOU, J. H., ZHANG, J., CHEN, H. T., MA, L. N., DING, Y. Z., PEJSAK, Z. & LIU, Y. S. 2011. The codon usage model of the context flanking each cleavage site in the polyprotein of foot-and-mouth disease virus. *Infect Genet Evol,* 11, 1815-9.

Guide to appendices

The attached CD contains appendices organised in four folders.

Appendix 1a- pSTA1 vector

Appendix 1b- pJC3 vector

Appendix 2- provide the chromotagram results for each 2A mutant featured in chapter 4

A sub-folder was created for each figure of chapter 4. The mutant identity is indicated first, the second line provides the expected nucleotide sequence. The chromatogram for the 2A region was inserted underneath.

For each 2A, the Xba1 site (TCTAGA), the first codon of GUS following 2A (CAC), as well as the mutation introduced are highlighted in the nucleotide sequence. The mutations are encased in a black rectangle in the chromatograms.

Appendix 3- provides the chromatograms for puromycin experiment, the TaV P20 mutants and FMDV codon study.

Appendix 4a- the amphioxus protein sequences with NCBI accession number

Appendix 4b- the list of contigs where 2A was identified in *Saccoglossus kowalevskii*.