Conserved elements within the genome of foot-and-mouth disease virus; their influence on viral replication

Kjær, Jonas

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Conserved elements within the genome of foot-and-mouth disease virus; their influence on viral replication

PhD Thesis · 2017

Jonas Kjær

Technical University of Denmark

National Veterinary Institute

Lindholm

Supervisors

Graham J. Belsham, Professor

National Veterinary Institute, Technical University of Denmark, Kalvehave, Denmark

Thomas B. Rasmussen, Senior Researcher

National Veterinary Institute, Technical University of Denmark, Kalvehave, Denmark

Assessment committee

Lars E. Larsen, Professor

National Veterinary Institute, Technical University of Denmark, Lyngby, Denmark

Martin D. Ryan, Professor

School of Biology, University of St. Andrews, St. Andrews, United Kingdom

Jens Bukh, Professor

Department of Immunology and Microbiology, University of Copenhagen, Copenhagen, Denmark

Conserved elements within the genome of foot-and-mouth disease virus; their influence on viral replication

Contents

Preface I

Acknowledgements II

Objectives and Summary IV

Formål og Sammenfatning VI

Part 1 · Introduction 1

Foot-and-mouth disease **2**

Foot-and-mouth disease: from past to present 2

Genome organization of foot-and-mouth disease virus **5**

Genome organisation 5

The 5′- and 3′-untranslated regions 5

Leader protein 6

The structural proteins VP1-VP4 7

2A 7

2B 8

2C 8

3A 9

3B 9

3C 10

3D 10

Quasispecies and genetic diversity 11

Reverse genetics system 12

Strategies to study viral replication 12

Replication of FMDV **16**

FMDV life cycle 16

Adsorption and entry 16

Translation and polyprotein processing....18
Replication of the viral genome....19
Assembly and release of the FMDV virion....21
2A mediated ribosomal skipping at the 2A/2B junction....22
Functional structural RNA elements within the FMDV genome....28
The multi-functional RNA molecule....28
RNA structure and function....28
RNA secondary structure probing....29
Conserved RNA structures in FMDV....31
SHAPE probing of the FMDV genome....35
In silico prediction of RNA secondary structures....36
Part 2 · Manuscripts....37
Manuscript 1: Modifications to the foot-and-mouth disease virus 2A peptide; influence on polyprotein processing and virus replication....38
Manuscript 2: Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias....74
Manuscript 3: The influence of conserved RNA structures in the 2B, 3C and 3D coding regions of the foot-and-mouth disease virus genome on virus replication....93
Part 3 · Conclusions and perspectives....110
Conclusions and perspectives....111
Abbreviations....117
References....119
Appendix 1....141

Preface

This PhD thesis is based on the experimental work performed at the island of Lindholm which is a part of the National Veterinary Institute at the Technical University of Denmark (DTU-Vet). It was conducted from December 2014 to December 2017 under the supervision of Professor Graham J. Belsham with co-supervision from Senior Researcher Thomas B. Rasmussen. The studies were supported by core funds within DTU-Vet.

The thesis is organized in three parts: **Part 1** consist of a general introduction which provides a review surrounding the different aspects of the molecular biology of FMDV with a focus on the viral life cycle and the conserved elements within the viral genome which influence it. **Part 2** describes the results achieved during this study presented as three manuscripts listed below:

I. **Kjær J**., Belsham G. J. (2018) Modifications to the foot-and-mouth disease virus 2A peptide; influence on polyprotein processing and virus replication. *Submitted*.

II. **Kjær J**., Belsham G. J. (2017). Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias. *RNA*. DOI: 10.1261/rna.063339.117

III. **Kjær J**., Belsham G. J. The influence of conserved RNA structures in the 2B, 3C and 3D coding region of the foot-and-mouth disease virus on virus replication. *Manuscript in preparation*.

Part 3 provides a Discussion of the achieved results and future perspectives for continued research.

Acknowledgements

The last three years have included so many memorable moments. Extreme joy, when my experiments worked but also a lot of frustrations, when hours of optimizations did not pay off. Nevertheless, it has been three very interesting years where I have learned so much, both scientific and personal. I have been very lucky to meet so many amazing people who I need to thank for helping me create this PhD thesis.

First, I would like to express my sincere gratitude to my main supervisor Professor Graham J. Belsham for giving me the opportunity to carry out this project. Thanks for your scientific guidance and discussions and for your supernaturally fast feedback, be it on posters, presentations or this thesis. Thank you for always keeping your door open and for always taking the time to explain it one more time. I approach science in a different manner due to you.

In addition, special thanks goes to my co-supervisor Senior Scientist Thomas B. Rasmussen for taking me on as a master student and subsequently as research assistant all those years ago. We have also had good scientific discussions and you have taught me the important lesson that preparation is half the work.

I wish to thank all the amazing people on the island of Lindholm for all the discussions, laughter and the overall fantastic work environment you have helped create. It made me feel welcome from day one. You are tough people, who despite many years of occupational insecurity continue to show up at work with a positive mind-set. You have my deepest respect and appreciation. I would like to thank lab technician Preben Normann for his vast technical knowledge, which I have benefited from and for his willingness to help me when needed. A “thank you” also goes to my good friend and former PhD colleague Johanne Hadsbjerg. We have had so many wonderful moments together, both in office, at conferences and after work. Your advice has always been much appreciated. I wish to thank Thea Kristensen, my fellow PhD student in the FMDV group. We have had countless laughs and scientific discussions where we far from always agreed. Thanks go to PhD student Ann-Sofie Olesen for your always-positive mind and for sharing the office with me.

A special thanks goes to my friend Louise Møller Slaaby who even though she did not understand my work helped me create the poster for the Europic2016 conference.

Thanks to Professor Nicola Stonehouse, Professor David Rowlands and the rest of the Stonehouse group for hosting me during my academic stay at the University of Leeds, United Kingdom. I will never forget the passion for research I experienced there.

Finally, my deepest thanks goes to my friends and family for their moral support during the last three years. To my husband Nicki Dam Nielsen, thank you for your everlasting patience and for letting me fulfil this amazing journey.

Jonas Kjær

Fuglebjerg, December 2017

Objectives and Summary

Foot-and-mouth disease (FMD) is caused by a highly contagious virus infection of cloven-hoofed farm animals with high financial importance for countries, such as Denmark, which rely on a significant trade in agricultural products. At present, FMD is exotic to the cloven hoofed livestock in Denmark. However, outbreaks of FMD have previously been detected in other parts of Europe (United Kingdom) during the last decade, thus an introduction of FMD remains a continuing threat to the highly industrialized pig production in Denmark.

Several conserved elements within the genome of the foot-and-mouth disease virus (FMDV) have been identified, e.g. the IRES. Such elements can be crucial for the efficient replication of the genomic RNA. A better understanding of the influence of these elements is required to identify currently unrecognized interactions within the viruses which may be important for the development of anti-viral agents.

SHAPE analysis of the entire FMDV genome (Poulsen, 2015) has identified three conserved RNA structures within the coding regions for 2B, 3C and 3D (RNA-dependent RNA polymerase) which might have an important role in virus replication. The FMDV 2A peptide, another conserved element, is responsible for the primary "cleavage" at its own C-terminus (2A/2B junction). It is believed that this "cleavage" is achieved by a protease-independent event termed ribosomal skipping or StopGo, in which the 2A peptide prevents the ribosome from linking the next amino acid to the growing polypeptide. The nature of this "cleavage" has so far not been investigated in the context of the full-length FMDV RNA within cells.

The focus of this PhD thesis has been to characterize these elements and their influence on the FMDV replication. In order to fulfil the aims of this thesis a series of studies were performed and the scientific work is presented as three manuscripts.

Manuscript 1 investigated the impact of 2A modifications within the conserved C-terminal D(V/I)E(S/T)NPG$^{\downarrow}$P motif on FMDV protein synthesis, polyprotein processing and virus viability. Certain amino acid substitutions (E14Q, S15I, S15F and N16H) that have been shown to severely (50-70%) reduce the cleavage activity at the 2A/2B junction compared to the wt, using *in vitro* assays, have been found to be tolerated within infectious FMDVs. In contrast, substitutions (N16A, P17A, G18A, P19A, and P19G) that inhibit cleavage by 89-100% *in vitro*, reverted to the wt sequence in the rescued viruses. The 2A substitutions impaired the replication

of a FMDV replicon, however surprisingly the viable 2A mutant viruses did not exhibit an attenuation of virus growth in cell culture. Expression of cDNAs encoding a truncated FMDV polyprotein, without any viral proteinases showed that certain amino acid substitutions at residues E^{14}, S^{15}, N^{16} and P^{19} resulted in partial "cleavage" indicating that these specific residues are not essential for co-translational "cleavage". This shows that the StopGo function at the 2A/2B junction is necessary for efficient virus replication. However, maximal cleavage activity does not appear to be essential for the viability of FMDV.

Manuscript 2 sought to identify which codons at each position of the coding sequence for the conserved NPG$^{\downarrow}$P motif at the 2A/2B junction would produce viable progeny viruses. This was achieved using a reverse genetics system with highly degenerate primers with all possible codons present for each of the amino acids individually within the NPG$^{\downarrow}$P motif. This generated pools of modified plasmids for each codon position from which RNA transcripts were made which were subsequently introduced into cells in culture. The rescued viruses all encoded the NPG$^{\downarrow}$P motif, confirming the importance of this amino acid sequence. However, at passage two, these four residues were found to be encoded by all possible codons (14 in total) within the rescued pools. Subsequent passages in cell culture revealed a distinct codon bias. Remarkably, this bias matches the codon bias observed within naturally occurring FMDV strains. Interestingly, the codons selected are different for P^{17} and P^{19}. Residue P^{17} is preferentially encoded by CCU while P^{19} is preferentially encoded by CCC. However, a single prolyl-tRNA species recognizes both of these two codons in cattle and pigs, which are the major hosts for FMDV, and suggests a role for the RNA sequence itself.

Manuscript 3 examines the influence of three conserved RNA structures within the genome of FMDV on viral protein synthesis and virus viability. Poulsen, (2015) previously identified these RNA structures within the coding regions for the FMDV 2B, 3C and 3D proteins using SHAPE probing. Interestingly, the structures had significant lower synonymous substitution rates compared to the remainder of the genome. However, introduction of synonymous substitutions which disrupted the structures but did not alter the amino acid sequence are tolerated and were retained after three passages in cell culture. The same substitutions did not impair the replication of a FMDV replicon either. This shows that the identified structures located in the coding region of 2B, 3C and 3D are not required for FMDV replication in BHK cells.

Formål og Sammenfatning

Mund- og klovesyge (MKS) skyldes en meget smitsom virusinfektion hos klovbærende dyr, som har stor økonomisk betydning for lande som Danmark, der er afhængige af en betydelig handel med landbrugsprodukter. MKS betegnes i dag som eksotisk for besætninger med klovbærende dyr i Danmark. Udbrud af mund- og klovesyge har dog fundet sted i Europa (Storbritannien) i løbet af det sidste årti. Introduktionen af MKS er derfor en fortsat trussel mod den stærkt industrialiserede svineproduktion i Danmark.

Flere konserverede elementer er blevet identificeret i genomet af mund- og klovesyge virus (MKSV), f.eks. IRES. Sådanne elementer kan være afgørende for en optimal replikation af det genomiske RNA. En bedre forståelse af betydningen af disse elementer er nødvendig for at identificere hidtil ukendte interaktioner i viruset, som kan være vigtige for udviklingen af antivirale midler.

RNA strukturanalyse af hele MKSV genomet (Poulsen, 2015) har identificeret tre konserverede RNA strukturer i de kodende regioner for proteinerne 2B, 3C og 3D (den RNA-afhængige RNA-polymerase), som kan have en vigtig rolle i virus replikationen. MKSV 2A peptidet, et andet konserveret element, er ansvarlig for den primære "kløvning" i sin egen C-terminus (2A/2B-kløvningssitet). Det antages, at denne "kløvning" sker ved en protease uafhængig hændelse betegnet ribosomal skipping eller StopGo, hvor 2A peptidet forhindrer ribosomet i at forbinde den næste aminosyre med det voksende polyprotein. Denne "kløvning" har hidtil ikke været undersøgt ved brug af fuldlængde MKSV RNA i celler.

Fokus for denne ph.d. afhandling har været at karakterisere disse elementer og deres indflydelse på MKSV replikationen. For at opfylde disse mål blev der udført en række af undersøgelser og det videnskabelige arbejde er præsenteret som tre manuskripter.

Manuskript 1 undersøger effekten af modifikationer i 2A, i det konserverede C-terminale D(V/I)E(S/T)NPG$^{\downarrow}$P motiv, på polyprotein-syntese og -bearbejdning samt levedygtighed af virus. Visse aminosyresubstitutioner (E14Q, S15I, S15F og N16H), der tidligere er vist at reducere (50-70%) kløvningsaktiviteten ved 2A/2B kløvningssitet, sammenlignet med vildtypen i et *in vitro* assay, blev tolereret i infektiøse MKSV. I modsætning til dette førte andre substitutioner (N16A, P17A, G18A, P19A og P19G), der hæmmer kløvningen med 89-100% *in vitro*, til reversion til vildtype sekvensen i de genererede virus. 2A

substitutionerne forringede replikationen af et MKSV replikon, men overraskende viste de levedygtige 2A mutanter ikke en svækkelse af væksten i cellekultur. Ekspression af cDNAer, der koder for et trunkeret MKSV polyprotein, uden de virale proteinaser, viste at visse aminosyre substitutioner ved E^{14}, S^{15}, N^{16} og P^{19} resulterede i delvis "kløvning" og indikerer at disse specifikke aminosyre ikke er essentielle for co-translationel "kløvning". Dette viser at StopGo funktionen ved 2A/2B kløvningssitet er nødvendig for optimal virus replikation. Den maksimale kløvningsaktivitet synes imidlertid ikke at være afgørende for levedygtighed af MKSV.

Manuskript 2 forsøger at identificere hvilke kodoner, i hver position af den kodende sekvens for det konserverede NPG$^{\downarrow}$P motiv ved 2A/2B kløvningssitet, som producerer levedygtige virus. Dette blev opnået ved anvendelse af et revers genetiksystem og degenererede primere, som kodede for alle potentielle kodoner for hver individuel aminosyre i NPG$^{\downarrow}$P motivet. Disse dannede en pool af modificerede plasmider for hver kodon position, hvorfra RNA transkripter blev fremstillet, som efterfølgende blev introduceret i cellekultur. De genererede virus kodede alle for NPG$^{\downarrow}$P motivet, hvilket bekræfter betydningen af denne aminosyresekvens. Efter to cellekulturpassager blev det vist at alle potentielle kodoner (14 i alt) blev fundet, som kodende for de fire aminosyrer. Efterfølgende passager i cellekultur viste en tydelig kodon-bias. Bemærkelsesværdigt matcher denne bias den kodon-bias, der observeres i naturligt forekommende MKSV stammer, hvor f.eks. kodoner for henholdsvis P^{17} og P^{19} er forskellige. Aminosyre P^{17} er fortrinsvis kodet af CCU hvorimod P^{19} hovedsagelig er kodet af CCC. En enkelt prolyl-tRNA enhed genkender imidlertid begge disse to kodoner hos både kvæg og svin, som er de vigtigste værter for MKSV, hvilket antyder at RNA sekvensen også spiller en rolle.

Manuskript 3 undersøger indflydelsen af tre konserverede RNA strukturer i genomet for MKSV på polyproteinsyntese og levedygtighed. Poulsen, (2015) har tidligere identificeret disse RNA strukturer i de kodende regioner for MKSV 2B, 3C og 3D proteinerne ved anvendelse af RNA strukturanalyse. Interessant nok havde strukturerne betydelige lavere synonyme substitutionsrater sammenlignet med den øvrige del af virusgenomet. Indsættelsen af synonyme substitutioner som ophævede strukturerne, men ikke ændrede aminosyresekvensen blev tolereret og blev bibeholdt efter tre passager i cellekultur. De samme substitutioner svækkede heller ikke replikationen af et MKSV replikon. Dette viser, at de identificerede konserverede RNA strukturer, der er lokaliseret i den kodende region af 2B, 3C og 3D ikke er nødvendige for MKSV replikation i BHK celler.

Part 1 · Introduction

Foot-and-mouth disease

Foot-and-mouth disease: from past to present

Foot-and-mouth disease virus (FMDV) is an animal pathogen responsible for the highly contagious and economically devastating foot-and-mouth disease (FMD). It affects cloven-hoofed domestic and wild animals such as cattle, sheep, pigs, goats and about 70 species of wildlife e.g. African buffalo (reviewed in Jamal & Belsham, 2013). The origin of FMDV remains unclear; however, an Italian monk, Hieronymus Fracastorius first described the symptoms of a disease similar to that of FMD in cattle in Venice in 1514. Four centuries later, in 1897, Loeffler and Frosch demonstrated that the infectious causative agent of FMD was a virus by transmitting the disease from FMDV infected calves to other susceptible cloven-hoofed animals through the use of filtered material (reviewed in Grubman & Baxt, 2004).

The FMDV is the prototypic member of the *Aphthovirus* genus within the family *Picornaviridae*, which has approximately 30 genera including many important human and animal pathogens such as poliovirus (PV) and coxsackie B viruses (e.g. CVB3) (see Figure 1). The *Aphthoviruses* consist of FMDV, bovine rhinitis B virus (BRBV), equine rhinitis A virus (ERAV) and the recently identified bovine rhinitis A virus (BRAV) (Knowles *et al.*, 2010). FMDV exists in seven serologically distinct serotypes (A, O, C, Asia 1 and Southern African Territories 1, 2 and 3), but with numerous subtypes within each serotype (Rweyemamu *et al.*, 2008).

An outbreak of FMD in areas with high-density pig and cattle farming can cause great socio-economic challenges as is evident from the 2001 and 2007 outbreaks of FMD in the United Kingdom (UK) (Cottam *et al.*, 2008; Thompson *et al.*, 2002). Even though the average mortality rate from FMD is only about 1 %, the morbidity within a herd can be close to 100% due to a very high transmission rate. The FMDV infected animals develop fever and vesicles in the mouth, muzzle, feet and/or teats, which has a severe effect on the productivity of the livestock. Dairy cattle not only experience a substantial fall in milk production of up to 40 % and loss of weight but also experience reduced mobility because of the lesions (reviewed in Grubman & Baxt, 2004). In the recent 2001 FMD UK epidemic, 2030 infected premises were identified across the country. This resulted in the culling of 6.5 million animals out of which only 1.3 million of these were on infected premises. These circumstances, as well as the costs associated with tourism and trade restrictions, contributed to the substantial financial losses.

The total financial cost of the UK FMD epidemic in 2001 was calculated to be 12.3-13.8 billion USD out of which 5.8 billion was associated with agriculture and food production and $5.1 to $6.0 billion in tourism-related business (Thompson *et al.*, 2002).

The preferred method for controlling a FMD outbreak in non-endemic countries is through restriction of animal movement, stamping out of all diseased and potentially infected herds with the subsequent incineration of their carcasses followed by quarantine (Paton *et al.*, 2009). However, the pre-emptive culling of susceptible animals can, in countries with high cattle or pig populations (like the UK or Denmark), lead to the widespread slaughter of healthy animals. Emergency vaccination or systematic FMDV vaccination programs can therefore be an additional tool for disease control (Doel, 2003). Nevertheless, there are complex problems associated with the use of vaccination as a control measure. The current FMD vaccine is an inactivated virus, which does not provide long-lasting immunity and ideally needs to be administered every 6 months (Doel, 2003). Furthermore, the high mutation rate of FMDV provides the chance of antigenic changes, potentially producing a new subtype, which can reduce the effectiveness of the vaccine. In addition, the use of vaccination hinders the differentiation between infected and vaccinated animals. This blocks trade with other countries, as detection of anti-FMDV antibodies, although due to vaccination, is an admission of the existence of a FMD outbreak. This lead the EU, which until 1991 (except in Denmark, UK and Ireland) systematically vaccinated against FMDV, to stop this practice (Sutmoller *et al.*, 2003). Countries with a status as "FMD-free without vaccination" have commercial advantages as they face less trade restrictions. In FMD endemic countries, primarily found in Asia and Africa, little or no effort has been made to control the spread of the disease.

The worldwide distribution of FMDV mirrors economic development. Industrialised countries in Europe, North American and Australia have eradicated the disease and all of South America has been free of clinical disease since 2013 (but with vaccination). Developing countries, primarily found in Africa and Asia, lack the resources and infrastructure to initiate proper eradication programs and FMD has, during the last decades, been reported as either endemic or with sporadic outbreaks in domestic or wild cloven-hoofed animals species (Paton et al., 2009; Rweyemamu et al., 2008). The seven FMDV serotypes are not distributed equally around the world. The serotype O, A and C viruses have had the widest distribution and have been responsible for outbreaks in Europe, America, Asia and Africa, whereas the SAT serotypes are normally restricted to sub-Saharan Africa. In Denmark, FMD has not been reported since the last outbreak in 1982 (Jamal & Belsham, 2013).

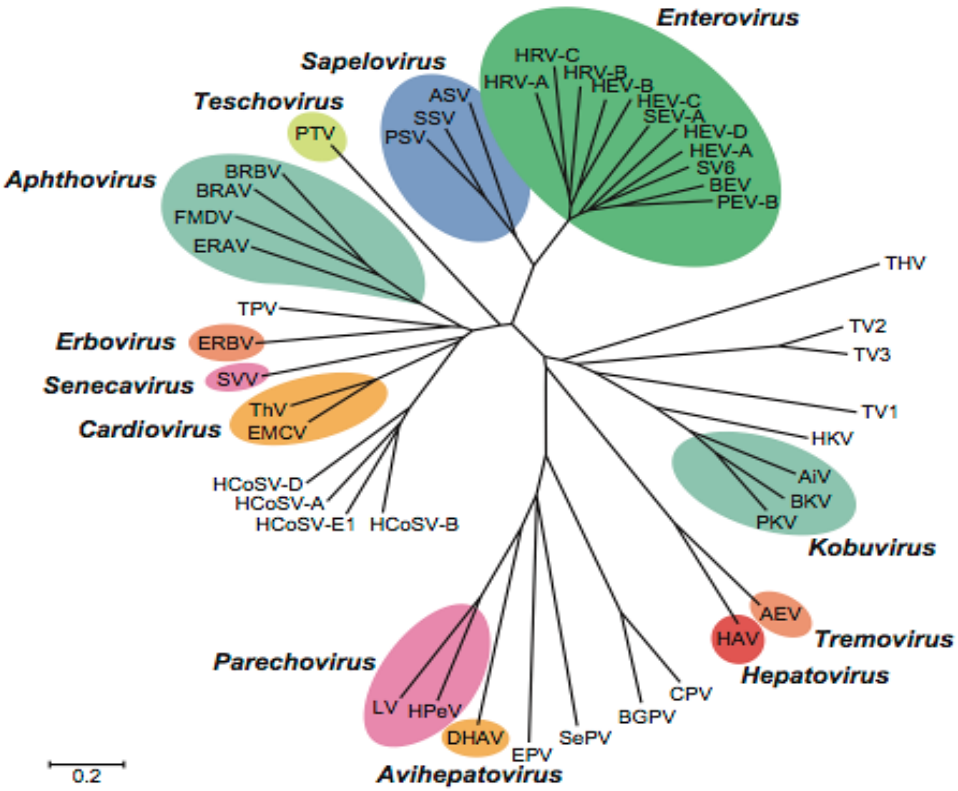

Figure 1: Phylogenetic tree of the *Picornaviridae* family. The tree depicts the genetic relationship based on the analysis of the P1 capsid precursor. The neighbour joining tree was produced using MEGA (Tamura et al., 2011). The branch lengths are proportional to the genetic difference where the distance scale corresponds to amino acid substitutions per position. Adapted from the ICTV 9th Report: Virus Taxonomy: The Classification and Nomenclature of Viruses.

Genome organization of foot-and-mouth disease virus

Genome organisation

The FMDV particle has an icosahedral protein shell (capsid) about 30 nm in diameter (Acharya *et al.*, 1989; Fry *et al.*, 2005). It consist of 60 copies of each of the four structural proteins VP1, VP2, VP3, and VP4 enclosing the viral genome. The capsid serves to protect the genome before the infection of the host cell, whereupon it is released into the cytoplasm. The FMDV genome consists of a positive-sense, single-stranded (+ss) RNA molecule approximately 8.3 kb long and includes a single, long, open reading frame (ORF) of about 7 kb (Martinez-Salas & Belsham, 2017). The RNA genome shares many characteristics with eukaryotic mRNAs as these also usually contain a single ORF followed by a 3´-untranslated region (UTR) and a poly(A) tail. However, in contrast to cellular mRNAs, the viral RNA does not possess a 7-methyl-G cap (m^7GpppN) structure at its 5´-terminus which is normally required for translation initiation. The large ORF within the genome encodes a large polyprotein that is rapidly processed, largely by virus-encoded proteases, to generate 15 distinct mature proteins (Belsham, 2005). The FMDV viral polyproteins, with four primary precursor proteins is divided into a leader (L^{pro}) protein, the structural capsid proteins (P1-2A), and non-structural proteins (P2 and P3) associated with the replication of the viral genome (see Figure 2).

The 5′- and 3′-untranslated regions

The approximately 1,300 nucleotide 5′-UTR of FMDV RNA accounts for ~16 % of the genome. The large size of this UTR is in contrast to the ~50-100 nt normally observed in eukaryotic mRNAs. This region contains extensive secondary structures and can be divided into five functional elements termed the S-fragment, poly(C) tract and pseudoknots, *cis*-acting replication element (*cre*) and the Internal Ribosome Entry Site (IRES) (Belsham, 2005; Grubman & Baxt, 2004). The 3′-UTR of FMDV consists of a sequence of ca. 100 nt that is predicted to fold into two separate stem-loops (Witwer *et al.*, 2001) followed by a poly(A) tract of varying length. These structures and their influence on the viral replication and translation are described in more detail in the section: Functional structural RNA elements within the FMDV genome.

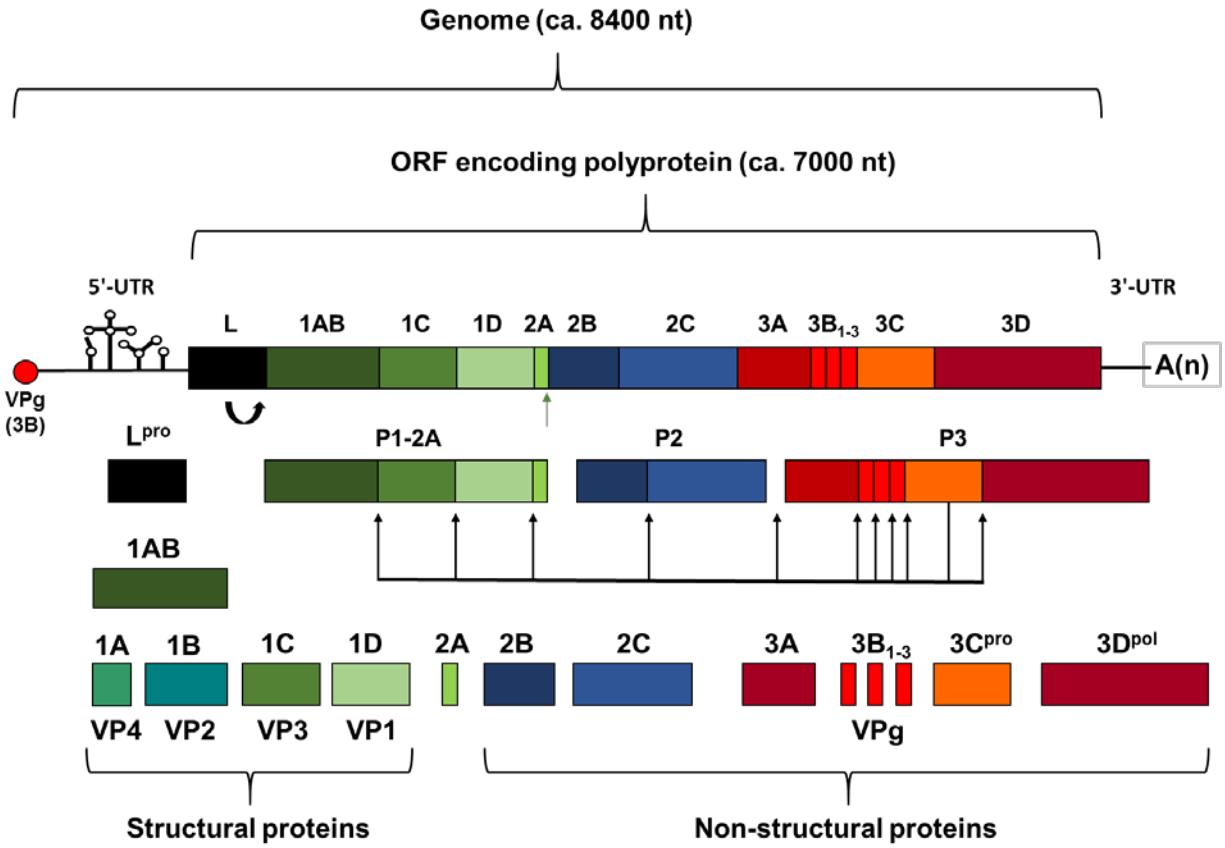

Figure 2: Schematic diagram of the genome organization of FMDV. The FMDV genome RNA contains a single large ORF of about 7 kb. The ORF is flanked by a long 5′-UTR and a short 3′-UTR. 3B (VPg) is covalently linked to the 5′-terminus of the RNA. The ORF encodes a polyprotein which is never observed due to rapid co-translational processing by viral proteinases (L^{pro} and $3C^{pro}$, indicated by black arrows) and by the translational recoding event (StopGo, ribosomal skipping) at the 2A/2B junction (indicated by a green arrow). The ORF is divided into four functional regions L, P1-2A (structural proteins), P2 and P3 (non-structural proteins) due to their distinct functions in the viral life cycle. Adapted from Jamal & Belsham, (Submitted).

Leader protein

Members of the *Aphthovirus* genus, are unusual in having a protease as the first component of the viral polyprotein. The FMDV Leader (L^{pro}) protein is a papain-like cysteine protease (Strebel & Beck, 1986). It cleaves itself from the polyprotein between its own C-terminus and the N-terminus of VP4 and it is also involved in repression of cap-dependent protein synthesis (host cell translation) by inducing cleavage of the translation initiation factor eIF4G (Devaney *et al.*, 1988; Kirchweger *et al.*, 1994; Medina *et al.*, 1993). The L^{pro} protein is produced in two distinct forms, termed Lab and Lb. This is a result of two in-frame start codons, 84 nts apart, which are conserved in all natural FMDV isolates (Carrillo *et al.*, 2005; Sangar *et al.*, 1987). Despite the Lab proteins N-terminal extension of an additional 28 residues compared to Lb it has been demonstrated that each form of proteinase has indistinguishable proteolytic activity

and specificity (Medina et al., 1993). Interestingly, deletion of the Lb region does not affect virus viability within BHK cells (Piccone *et al.*, 1995). Conversely, deletion of the entire Lab coding sequence is not tolerated (Belsham, 2013). Viable virus can also be produced from mutant FMDV genomes if the Lab start site is modified, however modification of the Lb initiation site hinders the production of viable progeny (Cao *et al.*, 1995). Previously, it was suggested that the "spacer" region was required for activity unrelated to its coding function. However, Belsham, (2013) demonstrated that deletion of the "spacer" region between the two initiation codons can be tolerated.

The structural proteins VP1-VP4

The FMDV capsid is made up of 60 copies of the four structural proteins VP1 (1D), VP2 (1B), VP3 (1C) and VP4 (1A). VP1, VP2 and VP3 are surface-exposed and fold into eight-stranded antiparallel β-barrels that are a characteristic of picornavirus capsid proteins (Acharya *et al.*, 1989). VP4 is found within the interior of the mature virus particle and can until virion maturation be considered an N-terminal myristoylated extension of VP2. The stability of virions is dependent on the myristoylation of the N-terminal glycine residue generated during cleavage at the L/P1 junction (Chow *et al.*, 1987). The virus encoded proteinase $3C^{pro}$ is responsible for processing of the P1/2A precursor. This generates protomers, consisting of one copy each of VP1, VP3, and VP0. During assembly, five protomers, associate into a pentamer, and 12 pentamers interact with a newly transcribed RNA molecule to form a mature infectious FMDV particle.

2A

The large polyprotein precursor produced during translation of FMDV RNA is proteolytically processed by the virus-encoded L^{pro} and $3C^{pro}$ proteins to produce multiple precursors (including P1-2A) and some 15 mature proteins (Belsham, 2005). Separation of the capsid precursor P1-2A from P2 at the 2A/2B junction occurs in a co-translational manner and is mediated by the short (18 residues long) 2A peptide (Belsham, 1993; Palmenberg *et al.*, 1992; Ryan *et al.*, 1989). The 2A peptide lacks any protease motifs but has a characteristic conserved C-terminal motif (D^{12}(V/I)E(S/T)NPG$^{\downarrow}$$P^{19}_{2B}$) shared with some other picornaviruses (e.g. encephalomyocarditis virus (EMCV) and Theiler's murine encephalitis virus (TMEV)). The FMDV 2A-mediated cleavage occurs at its C-terminus using a non-proteolytic translational

recoding mechanism termed “ribosome skipping” or “StopGo” (Ryan *et al.*, 1989, 1991; Palmenberg *et al.*, 1992). Mutational analysis has shown that the amino acid sequence, especially within the conserved NPG$^{\downarrow}$P motif, are critical for optimal ribosomal skipping (Donnelly *et al.*, 2001a, b; Sharma *et al.*, 2012). The function of 2A is described in more detail in the section: FMDV replication.

2B

2B proteins from some other picornaviruses (enterovirus 71, PV) have been characterized as hydrophobic transmembrane proteins termed viroporins (Agirre *et al.*, 2002; Xie & Wang, 2011). The hydrophobic domain of such proteins has been found to interact with the phospholipid bilayer of membranes. They are found to increase cell membrane permeability and promote the release of viral particles from the host cell (Nieva *et al.*, 2012). Presently, little is known about the exact function of the FMDV 2B. The enterovirus 2B protein has been found to impair Golgi complex trafficking by decreasing the endoplasmic reticulum (ER) luminal Ca^{2+} content (Van Kuppeveld *et al.*, 2005). Within FMDV infected cells and cells made to express 2B, the protein has been mainly located in the ER. The Ca^{2+} concentration in cells expressing the 2B protein increased resulting in autophagy in host cells (Ao *et al.*, 2015). Expression of both 2B and 2C within BHK cells has shown the transport between the ER and Golgi apparatus to become blocked by the 2BC precursor protein or by 2B and 2C together but not by either 2B or 2C alone indicating a synergistic effect (Moffat *et al.*, 2005, 2007).

2C

The exact biological function of the 2C protein is still unknown. Sequence analysis has shown that it is the most conserved viral protein within the FMDV genome (Gao *et al.*, 2016). 2C-like proteins have also been observed in other RNA viruses (Gorbalenya & Koonin, 1989). It has been suggested that FMDV 2C shares the same function as other picornavirus 2C proteins and along with 2B is required for the facilitation of membrane rearrangement. An ATPase domain, with documented ATPase activity, has been identified upstream of a predicted N-terminal amphipathic helix (Sweeney et al., 2010). Although, a helicase domain also has been identified, no evidence for helicase activity has so far been documented (Gorbalenya & Koonin, 1989). It has been established that guanidine inhibits the initiation of negative-strand PV synthesis within a cell-free replication system (Barton & Flanegan, 1997). Guanidine resistant mutants

of FMDV have amino acid substitutions within the 2C protein (Belsham & Normann, 2008; Saunders & King, 1982) and it is thus inferred that 2C has a role in this process.

3A

The FMDV 3A protein is considerable larger than other 3A proteins from within the picornavirus family. The N-terminus of 3A is highly conserved in all FMDV isolates and includes both hydrophilic and hydrophobic regions, by which the latter is capable of anchoring the protein to cellular membranes (Knowles *et al.*, 2001). Immunofluorescence staining of FMDV-infected BHK cells using anti-3A antibodies showed that 3A associates with both the ER and the Golgi apparatus (García-Briones *et al.*, 2006). PV 3A has been found to inhibit the protein trafficking from the ER to the Golgi, and it has been suggested that FMDV 3A could exhibit a similar function. However, as previously described, 2BC exerts this role in FMDV infected cells (Moffat *et al.*, 2005). 3A is a potential virulence and host range determinant. Several passages of FMDV in chicken embryos produced attenuated strains with deletions in the less conserved C-terminus of 3A (He *et al.*, 2011). FMDV isolated from outbreaks in Taiwan in 1997, which interestingly only infected swine and had an unusually high mortality rate, were also found to carry similar deletions in 3A (Beard & Mason, 2000). In addition, this strain also displayed restricted growth in bovine cells *in vitro* and significant attenuation in cattle *in vivo* (Pacheco *et al.*, 2003, 2013). Núñez *et al.*, (2001) also demonstrated that the mutation of a single residue within 3A mediated adaptation of FMDV to the guinea pig.

3B

The 3B (VPg) is covalently attached to the 5′-terminus of both positive- and negative-sense FMDV RNA genomes. 3B serves as a primer for RNA synthesis and is attached to most viral RNA transcripts, although not those associated with polysomes. In contrast to other members of the picornavirus family that only encode a single copy of 3B, the FMDV genome encodes three non-identical copies of the 3B (termed $3B_1$,$3B_2$, and $3B_3$) in tandem (Forss & Schaller, 1982). Each distinct form of the 3B protein has been found attached to genomic RNA, and is believed to be functionally equivalent (Nayak *et al.*, 2005). So far, all natural FMDV isolates have been reported to contain three copies of 3B, albeit not all three copies of FMDV 3B are required to maintain infectivity (Pacheco *et al.*, 2003). Deletion of one or two FMDV 3B sequences can produce viable viruses, but the mutant viruses have a lower replication

efficiency than the wt virus. Using a replicon system it has been established that deletion of $3B_3$ prevents replication but it could be restored by introducing substitutions at the C-terminus of $3B_2$ equivalent to that of the original sequence at the $3B_3$-3C cleavage site (Herod *et al.*, 2017). *In vitro* translation of plasmids expressing the wild-type FMDV P3 polyprotein showed that deletions of the $3B_3$ produced 3CD but no 3D.

3C

The FMDV $3C^{pro}$ is responsible for most of the polyprotein cleavages thereby generating most of the mature and intermediate protein products. It can exercise its function alone, in contrast to the PV 3C protease (Ypma-Wong *et al.*, 1988), which requires the 3D sequences for some of its processing activities. $3C^{pro}$ is a member of the chymotrypsin-like family of serine proteases, although the active site serine is replaced by a cysteine (Birtley *et al.*, 2005). The predominant function of FMDV $3C^{pro}$ is polyprotein processing but it has also been associated with inhibition of host cell transcription and translation (Belsham *et al.*, 2000), RNA binding and it is required (as part of 3CD) for the uridylylation of VPg *in vitro* (Nayak *et al.*, 2006). Like L^{pro}, $3C^{pro}$ is involved in eIF4G cleavage, although much later in the infection cycle and at an alternative site. It has also been reported to cleave eIF4A which is another component of the cap-binding complex with the function of an RNA helicase. Moreover, $3C^{pro}$ has been identified as the agent responsible for cleavage of histone H3 following FMDV infection (Falk *et al.*, 1990). Precursors containing $3C^{pro}$ (P3, 3ABC, 3CD) have also been shown to retain catalytic activity in PV (Spear *et al.*, 2015).

3D

The 3D protein is an RNA-dependent RNA polymerase (RdRp) catalysing the synthesis of positive- and negative-sense FMDV RNA genomes during viral replication (Robertson *et al.*, 1983a). The overall structure of the 3D resembles those of other RdRps by taking the appearance of a cupped "right hand" consisting of "palm", "fingers", and "thumb" subdomains (Ferrer-Orta *et al.*, 2006). The catalytic site of all RdRps is contained in the palm domain and includes a GDD motif which is conserved within the picornavirus family (Ferrer-Orta *et al.*, 2004). Uridylylated 3B (VPg) is required as a primer by 3D to initiate replication of either strand (Paul *et al.*, 1998) and will be described in more detail in the section: FMDV replication.

Quasispecies and genetic diversity

RNA viruses and, in particular FMDV, have extremely high mutation rates due to the lack of proofreading capabilities by their respective RdRps (Drake, 1993). The low replication fidelity leads to differences between FMDV replicated genomes compared to the original parental genome with 10^{-4} to 10^{-5} substitutions per nucleotide synthesized (Domingo & Holland, 1997; Drake & Holland, 1999). This is approximately 10^5 times higher than the mutational rate for the human DNA polymerase. For FMDV, the consequence of this significant high mutational rate is a 0.5 - 1.0% change in the consensus strain per year corresponding to 40-80 nucleotides/year (Haydon *et al.*, 2001a). Continuous error-prone nucleotide substitutions ultimately lead to a very heterogeneous viral population within the host cell, called a viral quasispecies (Manrubia & Lázaro, 2006). The significant genetic diversity allows the virus to rapidly adapt to dynamic environments thereby enabling it to evade the immune response of the host or evolve resistance to antiviral drugs. The quasispecies nature of the FMDV is a well-known characteristic of the virus and has been studied in detail at the nucleotide level (Domingo *et al.*, 1980, 2012). Studies have shown that non-synonymous mutations, which can lead to changes in the FMDV viral phenotype, represent about 50% of the total substitutions introduced during replication of the viral genome (Escarmis *et al.*, 1996). The highest occurrence of codon change is observed in the capsid-coding region of the genome (P1) and can result in antigenic variation. Studies have shown that antigenic variation within field isolates increases with time and most likely arises as a result of immunological pressure placed on the virus by the infected host's immune system (Domingo *et al.*, 2003; Haydon *et al.*, 2001b). Interestingly, antigenic variation in FMDV has also been reported in cell cultures incapable of producing any immunological pressure, suggesting that antigenic sites on the capsid could be involved in other functions (Diez *et al.*, 1989; Domingo *et al.*, 1993; Fares *et al.*, 2001). Non-synonymous mutations also occur within the P2 and P3 protein-coding regions of the genome but are observed less frequently possibly because these proteins are necessary for the viral replication and changes are more likely to result in deleterious phenotypes. A PV variant with an RdRp derivative possessing a 2-3-fold higher error rate has previously been found to move towards extinction more easily than wt in cell culture. The same virus also had a reduction in virulence in mouse, correlating with its inability to sustain replication in different cell types in which the wt virus can grow (Korboukh *et al.*, 2014).

Recombination is an alternative process, which can contribute to the genetic diversity in FMDV. It has been demonstrated that FMDV is able to undergo RNA recombination in cell

culture and that it was most likely to occur within the regions of the genome coding for the NS proteins (King *et al.*, 1985; McCahon *et al.*, 1985). However, Tosh *et al.*, (2002) suggested that RNA recombination within the capsid-coding P1 region of the genome may contribute to genetic diversity in FMDVs isolated from the field. Regardless of the mechanism, analysis of both synonymous and non-synonymous mutations has been a valuable tool in epidemiological studies of FMDV outbreaks.

Reverse genetics system

It has been known for years that the genome of various + ssRNA viruses, including FMDV, is infectious when introduced into susceptible cells (Baltimore, 1971; Belsham & Bostock, 1988). Reverse genetics systems exploits this principle enabling the generation of genetically modified viruses and is today a cornerstone of molecular virus research. Full-length complementary DNA (cDNA) of a positive-strand RNA virus is inserted into a bacterial DNA plasmid under the influence of a suitable promoter, most often T7. These plasmids can be genetically manipulated either by restriction enzyme cloning or PCR mutagenesis after which the genomic viral RNA can be produced by *in vitro* transcription and introduced into cells. If the mutated RNA is infectious, then modified viruses are produced. This allows for determination of important viral properties such as virulence/attenuation, cell penetration, replication, host range and functions of coding or non-coding genomic regions. To date, few infectious cDNA clones of FMDV has been established (Ellard *et al.*, 1999; Liu *et al.*, 2004; Rieder *et al.*, 1993; Zibert *et al.*, 1990) which are based on the O1 Kaufbeuren and A12 strains respectively.

Strategies to study viral replication

Quantification of virus replication represents an important step in the characterization of the viral life cycle of wt or genetically modified virus. Accurate determination of virus concentration (virus titre) can be archived through various methods (see Figure 3), of which the tissue culture infective dose assay ($TCID_{50}$) or virus plaque assay has been considered the "gold standard" for many years. In the last decade, RT-qPCR has also been increasingly used. The methods are well established but have disadvantages and limitations. The measurement of growth kinetics of viruses using $TCID_{50}$ does not differentiate between the different elements of the viral life cycle. A decrease in growth characterises can thus be a result of other altered processes in the viral life cycle than only the replication. RT-qPCR which can be laborious

determines RNA production, and is dependent on primer specificity and an accurate standard curve (reference point). Recently, the use of fluorescently-tagged viruses as tools to characterize the viral life cycle has been investigated for various RNA viruses. The methods utilize the tagging of viral proteins which subsequently can be monitored over time. This has been demonstrated for various viruses within the picornavirus family such as FMDV, PV and coxsackievirus B3 (Mueller & Wimmer, 1998; Seago *et al.*, 2013; Tong *et al.*, 2011). The GFP coding sequence (700 nt) and *Renilla* luciferase (800 nt) for FMDV was inserted downstream of the structural proteins coding region, but was rapidly removed, presumably by recombination, following serial passage. In addition, insertion of GFP into the PV genome was shown to be detrimental for viral replication efficiency. This suggested that packaging limits of the viral capsid and stability of the RNA genome following GFP coding sequence insertion were contributing factors leading to the results seen by both groups. Insertion of the coding sequence for a smaller, alternative, fluorescent protein termed iLOV (400 nt) resulted in the production of a virus that displayed only a slightly reduced rate of growth compared to the parental virus in cell culture. The fluorescent-tagging of influenza virus and Marburg virus (Albariño *et al.*, 2013) which both have negative-sense RNA genomes has also been demonstrated (Manicassamy *et al.*, 2010).

Advances in molecular biology, especially within the field of reverse genetics have enabled the development of replicon systems, which can be suitable for characterization of RNA replication. Replicons are defined as self-replicating but non-infectious RNAs. FMDV replicons can be generated using in-frame deletions within the region encoding the structural proteins (Khromykh, 2000). This abrogates capsid assembly making the virus unable to produce new infectious FMDV particles. In certain members of the picornaviruses family, this approach is not possible due to the presence of an essential RNA replication element, termed the *cre,* in the coding region for the structural proteins (McKnight & Lemon, 1996). This element is located in the non-coding region upstream of the IRES within the FMDV genome (Mason *et al.*, 2002). Thus, a reporter protein in place of, or in frame with, viral coding sequences allows quantitative measurements of the viral replication efficiencies. The reporter protein can be either fluorescent or enzymatic in nature and both have previously been applied in FMDV replicon systems (Mason *et al.*, 2002; McInerney *et al.*, 2000; Tulloch *et al.*, 2014). The use of FMDV replicon systems has provided an opportunity for research into FMDV replication, which had previously been restricted to high-security containment facilities. This is because they are unable to produce infectious virus particles and recombination with

circulating FMDV can be neglected if the laboratory is located within a FMD free region and otherwise does not conduct work with constructs harbouring the capsid coding region.

The first reported FMDV replicon system with a detectable reporter protein was based on the genome of FMDV O1K and contained the chloramphenicol acetyl-transferase (CAT) protein (McInerney *et al.*, 2000). It was a derivative of the infectious cDNA of FMDV, pT7S3, in which a majority of the L^{pro} (Lab and Lb) and structural protein coding sequences (VP4-VP3) were deleted and replaced by a sequence encoding the CAT protein. The deletion of most of the structural proteins was not detrimental for RNA replication consistent with later observations (that the *cre* is located in the 5′-UTR and the L^{pro} is not essential for FMDV replication). Recently, a FMDV replicon encoding the *Gaussia* Luciferase (Gluc) reporter protein instead of the VP1-VP3 sequence has been generated (Kjær & Belsham, submitted, see Manuscript 1). This replicon was used to demonstrate the detrimental effect of amino acid substitutions within the 2A peptide on the translation/replication of the FMDV genome. Gluc is naturally secreted and thus has advantages to *Renilla* based replicons which require cell lysis prior to detection of the reporter protein. Fluorescent proteins like GFP, mCherry or DsRed can also be used in the investigation of FMDV replication. They possess several advantages over their enzymatic counterparts, since the intracellular accumulation can be directly observed *in vivo* over time and their expression can be measured for individual cells. A GFP replicon system has been generated for FMDV (Herod *et al.*, 2015, 2016; Tulloch *et al.*, 2014). The FMDV CAT replicon system reported by (McInerney *et al.*, 2000) was modified by deleting the sequence encoding CAT and replacing it with the coding regions for a functional L^{pro} together with the fluorescent protein GFP linked to the puromycin resistance protein. The GFP replicon was replication competent and it was easily detected. However, the replication-incompetent forms of the genome surprisingly produced a signal only 2-fold lower. Interestingly, a derivative of the GFP replicon, which did not contain the L^{Pro}, showed a significantly stronger fluorescence signal, which has been suggested to be the result of cleavage of the GFP protein by the L^{pro}. The co-transfection of a helper-replicon containing a fluorescent reporter protein detectable at another wavelength has been shown to identify intergenomic trans-complementation of elements such as RNA structures or viral proteins (Herod *et al.*, 2015, 2016, 2017)

.

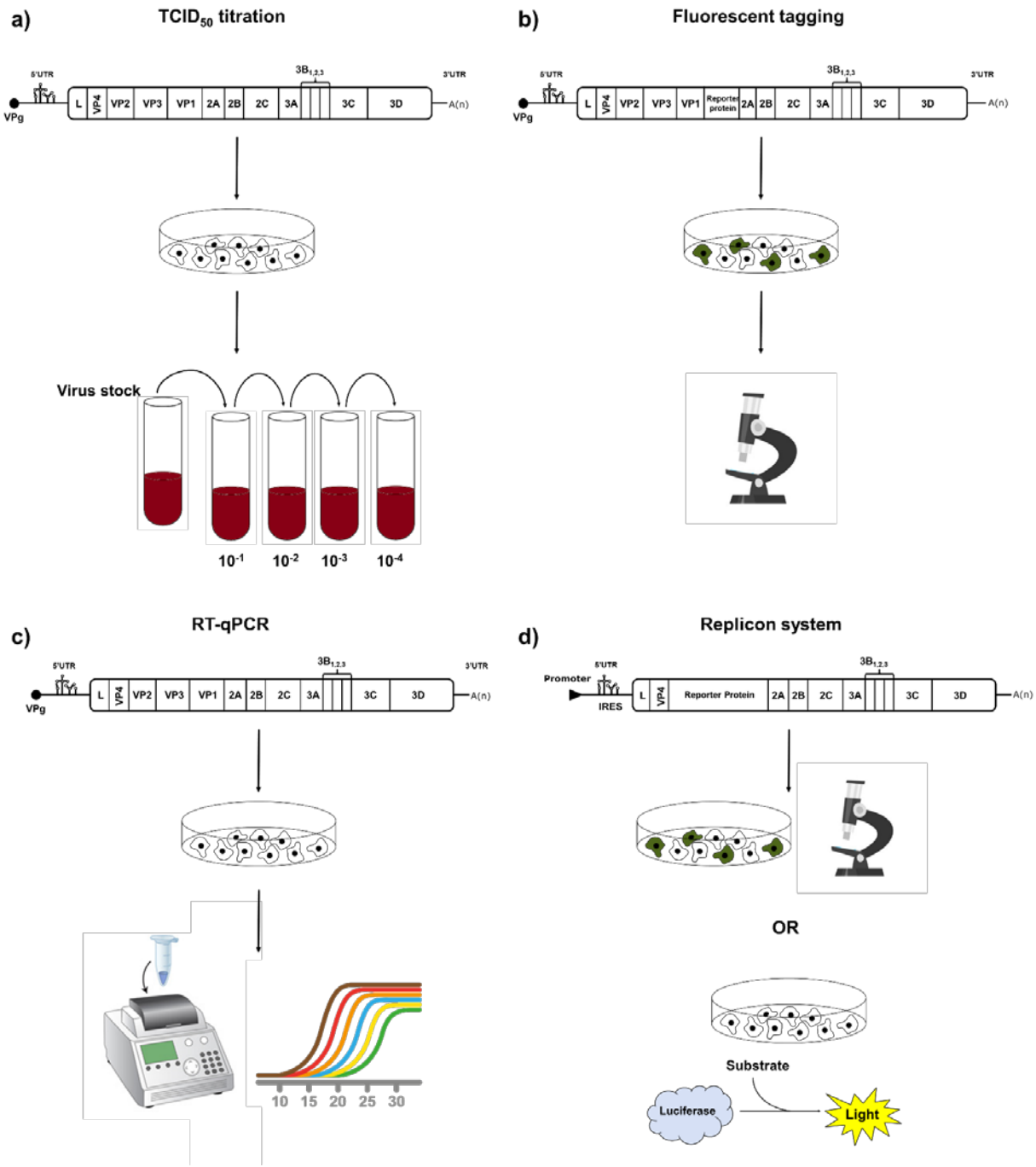

Figure 3: Schematic representation of methods used to study viral replication. Four methods to study viral replication including: **a)** $TCID_{50}$ titration **b)** fluorescent tagging of viruses and **c)** RT-qPCR. All use infectious viruses and require high containments facilities. The replicon system **(d)** removes all or the majority of the capsid proteins of the virus and replaces them with a reporter protein which can be enzymatic (e.g. Gluc or Rluc) or fluorescent (GFP, mCherry) in nature. This prevents the production of new infectious particles and does not require a high containment facility. Expression of the reporter protein is directly proportional to the replication.

Replication of FMDV

FMDV life cycle

FMDV has, like other members of the picornavirus family, a relative short life cycle in cultured cells. Depending on the initial virus titre and multiplicity of infection, newly generated infectious virus particles are detectable between 4 and 6 hours after infection (Grubman & Baxt, 2004). FMDV is cytocidal, resulting in morphological alterations referred to as cytopathic effects within infected cells. This includes a rounding of the cells and redistribution of internal cellular membranes. The basic components of the infectious cycle of FMDV are:

- Adsorption and entry
- Uncoating and genome release into the cytosol
- Translation of the genomic viral RNA
- Assembly of RNA replication complexes
- Genome replication
- Virion assembly, maturation and release

An overview of the FMDV infectious life cycle can be seen in Figure 4.

Adsorption and entry

The first step in the virus replication cycle is represented by the initial attachment of the virion to the host cell surface. For non-enveloped viruses like FMDV, this process typically involves an interaction between the viral capsid proteins and a host cell receptor molecule present on the surface of the cell membrane. Binding of the virion to the cell surface will eventually lead to several processes whereby the virus enters the cell. It is well established that FMDV receptors, as well as other viral receptors, play a part in determining the host range of the virus and its tissue tropism defining its pathogenic potential (Evans & Almond, 1998; Schneider-Schaulies, 2000). The binding of the FMDV particles to the cell surface in cell culture has been demonstrated to occur to a limited number of receptor sites (Baxt & Bachrach, 1980). Through the use of cross-competition studies it was revealed that all seven serotypes appeared to bind to a common primary class of receptor *in vitro*, whereas some of the serotypes bound to an alternate second class of receptors which are found to be present at the cell surface in a higher

number (Sekiguchi *et al.*, 1982). Limited trypsin digestion of FMD virions was shown to result in non-infectious particles incapable of binding to cells in culture (Moore & Cowan, 1978). Additional studies of non-infectious trypsin-inactivated FMDV particles showed a single cleavage of the VP1 protein at Arg144 which subsequently has been found to be located within the βG-βH surface loop of the VP1 (Robertson *et al.*, 1983b). It was suggested that this region of the VP1 interacted with the cellular receptor. Arg144 has since been found to be part of the tripeptide sequence Arg-Gly-Asp (RGD), and has been identified to be a recognition sequence for the transmembrane glycoproteins called integrins (Fox *et al.*, 1989). Integrins consist of two different subunits, α and β, and are involved in cell adhesion, cell migration, thrombosis, and lymphocyte interactions. Currently, it is known that integrins are composed of eighteen α and eight β subunits, which can combine to generate twenty-four different αβ combinations. However, only eight out of these twenty-four integrins use the RGD motif as a recognition sequence and four ($\alpha_v\beta_1$, $\alpha_v\beta_3$, $\alpha_v\beta_6$, $\alpha_v\beta_8$) of these have so far been associated with FMDV attachment (Duque & Baxt, 2003; Jackson *et al.*, 2000, 2002, 2004). FMDV strains which have adapted for growth in cultured cell lines have been found to acquire a high affinity for heparan sulfate proteoglycans (HSPGs) receptors for attachment (Jackson *et al.*, 1996). However, the use of HSPGs as a facilitator for host cell entry of field viruses has not been proven. The plasmids containing the O1K cDNA (Ellard et al., 1999) are based on a cell-culture-adapted B64 strain of the O1 Kaufbeuren virus (O1K B64). Viruses rescued from O1K B64-derived infectious cDNA have so far not been reported to cause disease in natural host animals. However, replacing the surface-exposed capsid proteins VP2, VP3 and VP1 within this virus with the field strain sequences of O-UKG and A-TUR have been shown to restore pathogenicity in cattle (Bøtner et al., 2011). Recently, it has also been identified that cell culture adaption of FMDV allowed for integrin- and HS-independent infection of CHO cells due to the substitution of glutamine (Q) to lysine (K) at amino acid position VP1-110 (Berryman et al., 2013). In addition, Chamberlain et al., (2015) have identified a novel site on the capsid within VP2 of a FMDV strain where substitutions resulted in cell culture adaptation. Interestingly, it remained infectious for cattle, despite having a major deletion in the G-H loop of VP1 that results in the loss of the integrin-binding RGD. FMDV binds to the cell surface through interaction with these receptors and enters the cell through clathrin mediated endocytosis (Burman et al., 2006). This is followed by a decrease in the pH of the endosome which breaks down the pH-sensitive capsid releasing the viral RNA into the cytoplasm (Carrillo & Giacheiti, 1984).

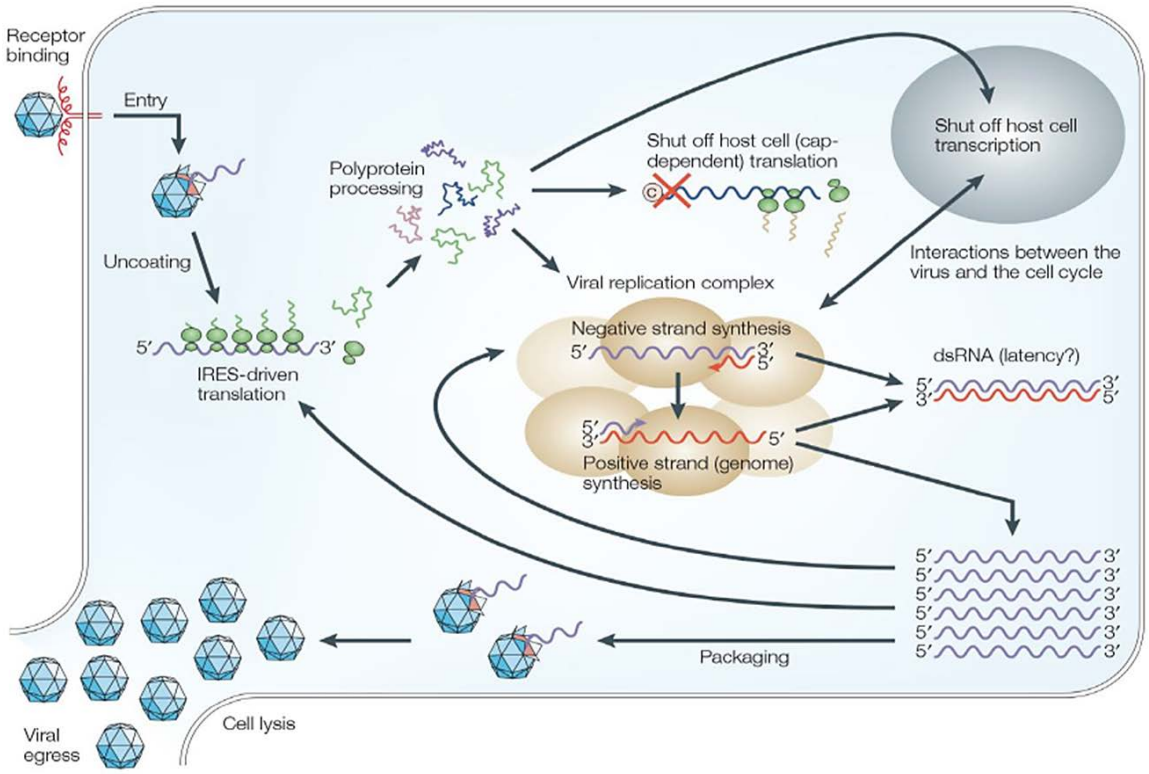

Figure 4: Overview of the FMDV life cycle. The virus enters the host cell through attachment of the virion to specific host receptors (e.g. integrins or heparan sulphate) which facilitates internalization through clathrin dependent endocytosis. The capsid dissociates and releases the positive-sense single stranded RNA genome into the cytoplasm. The cap-independent translation of the single large ORF within the FMDV genome is driven by the internal ribosome entry site (IRES) element that recruits the ribosomal subunits. This result in a polyprotein which is rapidly processed into structural and non-structural proteins, 15 in total. The viral proteases L^{pro} and $3C^{pro}$ shut off host cell translation (cap-dependent). The non-structural proteins form viral replication complexes through the rearrangement of the internal cellular membranes. Viral replication is initiated in these complexes and newly formed +ssRNA is released and can either be used in translation and replication or be encapsidated to form new infectious virions which subsequently are released. Adapted from Whitton *et al.*, (2005).

Translation and polyprotein processing

Following uncoating of the +ssRNA FMDV genome inside the endosome, it is released into the cystosol. The FMDV does not bring any viral proteins capable of replicating the genomic RNA inside the cell nor are there any cellular RNA polymerases present to perform this task. Accordingly, the first major event for the genomic RNA is to associate with the cellular translational system. Unlike most host-cell mRNAs, the FMDV RNA genome is missing a standard m^7GpppN cap-structure at its 5′-end which eukaryotic cellular translation relies on. Cellular cap-dependent translation is initiated when the 5′-cap structure is recognized by eIF4E, a cap-binding protein, which also binds to the N-terminal domain of eIF4G (reviewed in Sonenberg & Hinnebusch, 2009). The C-terminal domain of eIF4G also binds to eIF4A and

the 40S ribosomal subunit via eIF3. The translation initiation complex then scans along the mRNA until it encounters a start codon at which the 60S ribosomal subunit is recruited and translation begins. In contrast, cap-independent protein synthesis is dependent on the FMDV IRES element and requires almost every cellular translation initiation factors except eIF4E to bind to its three-dimensional structures (Martinez-Salas & Belsham, 2017) including the poly(rC) binding protein PCBP2. This protein has been found to stimulate the activity of the PV IRES and binds to domain I of the FMDV IRES. However, depletion of this protein does not block FMDV IRES activity *in vitro* (Stassinopoulos & Belsham, 2001). Translation of the FMDV genome will eventually lead to the shutoff of cap-dependent mRNA translation, since the viral encoded L^{pro} and $3C^{pro}$ induce cleavage of eIF4G (Belsham *et al.*, 2000; Devaney *et al.*, 1988; Gradi *et al.*, 2004). However, cap-independent translation remains unaffected as cleavage of eIF4G only removes the N-terminal portion of the protein which is essential for the binding site for the eIF4E. The C-terminal cleavage product of eIF4G, which contains the binding sites for eIF4A and eIF3 is still functional for IRES-mediated translation initiation. Previously, it has been suggested that the poly(A) tract and the 3′ stem-loop help facilitate FMDV translation, since deletion of either produced non-infectious FMDV with a lower translation efficiency in *in vitro* translation reactions (Serrano *et al.*, 2006). Polypeptide chain formation can commence at two distinct in-frame start codons (see Genome Organization) located in the L^{pro} coding sequence. The single large ORF encodes a polyprotein which is never observed in FMDV infected cells because of a series of co- and post-translationally proteolytic cleavages by the viral encoded proteinases (L^{pro} and $3C^{pro}$) together with the translational recoding mechanism (StopGo or ribosomal skipping) at the 2A/2B junction (Donnelly *et al.*, 2001a). The cleavage products are L^{pro} (cleaves at its own C- terminus), P1-2A, P2 and P3. Each of the P1-2A, P2 and P3 precursor products are processed by the $3C^{pro}$ to generate the mature structural and non-structural proteins.

Replication of the viral genome

Replication of positive-stranded RNA viruses occurs within membranous replication complexes, derived from host-cellular membranes i.e. the ER and Golgi apparatus. These membranous structures have previously been extensively studied in PV (Bienz *et al.*, 1990; Schlegel *et al.*, 1996). These membranes are embedded with numerous PV non-structural proteins encoded by both the P2 (2B, 2BC, and 2C) and P3 (3A, 3B, 3C, and 3D) regions (Van Kuppeveld & Ehrenfeld, 2010). Electron microscopy has demonstrated that the location of the

FMDV replication complex forms next to the nucleus and contains membranous structures (Monaghan *et al.*, 2004). Knox *et al.*, (2005) showed using immunofluorescence staining that FMDV structural and non-structural proteins co-localize in close proximity to the Golgi complex. In contrast to other picornaviruses, FMDV-infected cells experience a collapse of the cellular organelles to one side of the cell. The membranous replication complexes have been reported to appear distinct from those induced by other picornaviruses such as PV, as they appear in fewer numbers and do not aggregate into densely packed clusters (Monaghan *et al.*, 2004).

A negative-stranded RNA generated from the incoming + ssRNA genome serves a template for the replication of additional positive-stranded RNAs (Gamarnik & Andino, 1998). Translation of the viral genome proceeds in the 5'-3' direction whereas the replication of the genome mediated by the RdRp moves in a 3'-5' direction. The simultaneous events of translation and replication are apparently not compatible and translation of the positive-strand RNA must presumably cease before negative-strand synthesis can begin. Switching between these two mechanisms has in PV been proposed to be initiated by the accumulation of the 3CD precursor in the cytoplasm which binds to the cloverleaf structure in the 5′-UTR thereby forming a ribonucleoprotein (RNP) complex (Gamarnik & Andino, 1998). This will hinder interaction between the PCBP2 and the IRES otherwise required for translation. It remains unknown whether 3CD plays a similar role in FMDV RNA replication as 3CD is rapidly cleaved to $3C^{pro}$ and 3D. In addition, PCBP2 does not appear to affect the activity of the FMDV IRES *in vitro.* At the 3'-terminus of the PV viral RNA, the poly(A) binding protein (PABP) binds to the poly(A) tract which subsequently interacts with the RNP complex at the 5'-terminus essentially circularizing the genome (reviewed in Grubman & Baxt, 2004; Herold & Andino, 2000). For FMDV, this interaction has been suggested to include the S-fragment and the poly(C) tract (Serrano *et al.*, 2006). Negative-strand RNA synthesis begins with the uridylylation of 3B (VPg) which acts as a primer for the RdRp. VPgs bind to the first two A's in the conserved AAACA motif of the *cre*, serving as a template for the synthesis of VPgpU and VPgpUpU catalysed by the 3D (Nayak *et al.*, 2005). It is unclear whether free VPg is normally utilized in the initiation step or whether a precursor (either 3AB, 3ABC, or 3BCD) is needed. Elongation of the negative-strand catalysed by the FMDV RdRp begins once the uridylylated VPg (VPgpU or VPgpUpU) binds to the poly(A) tail located at the 3′-end of the positive-strand template. The elongation of the nascent strands has been suggested to result in a double-stranded replicative form (RF) consisting of positive-strand template and newly-

synthesised negative RNA (Paul & Wimmer, 2015; Wimmer *et al.*, 1993). Free negative-strands are not detectable *in vivo*, further supporting this claim. The ratio of positive- to negative-strands is about 50:1 (Novak & Kirkegaard, 1991), indicating that a single negative-strand serves as a template for the synthesis of many positive-strands which occurs through a partially double-stranded RNA molecule. The Qβ bacteriophage (another + ssRNA virus) has been found to use an alternative model of replication which does not involve a double-stranded RF (reviewed in Tomita, 2014). The 3′-end of the RNA template (either positive- or negative-sense RNA) is recruited to the active site of the Qβ RdRp (Qβ replicase) though the tunnel-II where it initiates RNA replication *de novo*. The nucleotide triphosphates (NTP) enter through the NTP entrance tunnel and base-pair with the template RNA to form the new elongating RNA strand. When the positive:negative duplex exceeds 9 bp, a change in the orientation of the interior of the RdRp results in a separation of the positive- and negative-strands. This model is independent of a RF intermediate, but it has not been investigated whether it is applicable to the picornavirus family.

Assembly and release of the FMDV virion

In order to generate mature FMD virions capable of infecting new cells, encapsidation of the FMDV RNA and the cleavage of VP0 to VP2 and VP4 to form the mature virion must occur. The mechanisms involved in FMDV RNA encapsidation and the maturation cleavage are not well studied and remain largely unknown. Picornaviruses have been shown only to encapsidate VPg-linked positive-strand RNA thereby ensuring that unwanted cellular RNA is not incorporated into the mature capsid. Furthermore, it appears that only newly synthesized positive-strand RNAs are encapsidated, suggesting that there is a connection between the ongoing RNA replication and the encapsidation process (Nugent *et al.*, 1999). It is believed that the positive-strand RNA may contain a *cis*-acting packaging signal which helps to facilitate encapsidation (Grubman & Baxt, 2004). Disruption of conserved secondary RNA structures within the coding region of the FMDV genome has been shown to results in attenuation of virus growth in cell culture as a consequence of fewer available mature virions (Logan *et al.*, 2017). This was interpreted as evidence for the involvement of these RNA structures in the packaging of FMD virions.

The FMDV capsid is formed from 60 copies of each of the four different capsid proteins (VP1 to VP4), whereby five protomers, each containing one copy of VP0, VP1, and VP3, assemble into a pentamer, and 12 pentamers assemble to make the complete particle. It still remains

unknown whether the viral RNA is inserted into the capsid following assembly or if there is interaction between the capsid pentamers enabling virus assembly around the RNA genome. Mature virions are formed when VP0 is autocatalytically cleaved to VP2 and VP4, which is believed to occur upon encapsidation of the RNA but also happens following empty capsid assembly (Curry *et al.*, 1995; Gullberg *et al.*, 2013a). Cleavage of VP0 is primarily required for particle stability, since the interactions between the newly formed VP2 and VP4 help reinforce the connection between capsid subunits. The influence of this cleavage on the FMDV virion maturation has been investigated by introducing substitutions within the VP0 precursor near the cleavage site. Some substitutions resulted in a loss of infectivity, but the mutant viruses were still capable of recognizing the receptor on susceptible cells emphasising the importance of this cleavage for the generation of infectious virus (Knipe *et al.*, 1997). Following assembly and maturation of the virions, they are released from the infected host cells.

2A mediated ribosomal skipping at the 2A/2B junction

The separation of the P1-2A precursor from P2 of the FMDV polyprotein occurs at the 2A/2B junction (Belsham, 2005). In the majority of picornaviruses, cleavage at the junction between the structural and non-structural proteins is mediated by a 2A protein, but the size and function of the 2A differs between the different genera (Luke et al., 2008). In entero- and rhinoviruses, the 2A protein is a thiol proteinase, designated $2A^{pro}$, which cleaves the polyprotein at its own N-terminus, i.e. the P1/2A junction (Sommergruber et al., 1989; Toyoda et al., 1986). In contrast, the cleavage of the capsid precursor (P1-2A) from P2 (2BC) within the cardio- and aphthovirus polyproteins occurs at the 2A/2B junction, i.e. at the C-terminus of 2A. The 2A subsequently remains attached to P1 until removed by the $3C^{pro}$. The 2A region of the FMDV is only 18 amino acids in length compared to the 150 residues of the 2A protein of cardioviruses, but both lack any characteristic protease motifs. In addition, no sequence similarity is observed between the 2A of cardio- and aphthoviruses other than the conserved C-terminal sequence motif $D(V/I)E(x)NPG_{2A}{\downarrow}P_{2B}$ at the junction between 2A and 2B (Ryan et al., 1989, 1991, Palmenberg et al., 1992). The first residue of the adjacent protein 2B, Proline or P, is a key part of the “cleavage” site and is likewise completely conserved. Thus these 2A proteins can be defined as having two-parts, one comprising a non-conserved sequence of amino acids followed by the conserved $D(V/I)E(x)NPG_{2A}{\downarrow}P_{2B}$ motif. Early studies into the mechanism of this cleavage demonstrated that neither the FMDV encoded proteases L^{pro} and $3C^{pro}$ nor host cellular proteases were responsible for the rapid cleavage (Palmenberg et al.,

1992; Ryan et al., 1989, 1991). These results suggested that "cleavage" mediated by the FMDV 2A occurred co-translationally through a novel protease independent system (this process will simply be described as cleavage subsequently, for convenience). Ryan et al., (1991) constructed multiple recombinant FMDV polyproteins in which sequences either upstream or downstream of the 2A peptide were removed. Only the polyproteins with deletions extending into the 2A peptide abrogated cleavage activity. This suggested that the 2A sequence was solely responsible for the co-translational cleavage. To demonstrate this, artificial polyprotein systems were generated encoding reporter sequences flanking the 2A peptide together with the N-terminal proline of 2B (Ryan & Drew, 1994). Analysis of these artificial polyprotein systems using rabbit reticulocyte lysates exhibited high-level cleavage activity confirming that cleavage at the 2A/2B junction was an autonomous event. The current model for cleavage at the 2A/2B junction (see Figure 5) is that the 2A peptide induces an unusual proteolysis-independent co-translational cleavage event that is referred as "ribosomal skipping" (Donnelly *et al.*, 2001a) or, alternatively, "stop-carry on" or "StopGo" (Atkins *et al.*, 2007; Doronina *et al.*, 2008). Prediction of the geometry of the amino acids within 2A suggest that the non-conserved upstream amino acids possess a strong propensity to form an α-helix over most of its length with a tight reverse turn motif (ESNPG) at its C-terminus (Donnelly *et al.*, 2001a; Ryan *et al.*, 1999). This particular structure is believed to interact with the exit tunnel of the elongating ribosome and bring about a stereo-chemical reorientation of the turn motif thereby influencing the events within the peptidyl transferase centre (PTC) of the ribosome. This subtle shift of the peptidyl-tRNA^{Gly} in the PTC, precludes peptide bond formation to the prolyl-tRNA^{Pro} (Doronina et al., 2008). Peptide bond formation is achieved by the nucleophilic attack on the electrophilic carbonyl group of the tRNA^{Gly} located in the P-site by the tRNA^{Pro} amino group located in the A-site. Proline has a distinctive cyclic structure which results in an exceptional conformational rigidity. It has also been documented that proline is a poor nucleophile in peptide synthesis compared to other amino acids (Rychlík *et al.*, 1970). It is believed that these characteristics are a contributing factor to 2A activity and explains the conservation of a P at the C-terminus of the NPG$^{\downarrow}$P motif in all 2A proteins. These events are believed to bring about a pause in translation resulting in inhibition of the formation of a peptide bond between the C-terminal amino acid, glycine (gly, or G), of 2A and the first residue, proline (P), of 2B. The pause in translation leads to dissociation of tRNA^{Pro} from the A-site of the ribosome. There are two alternative outcomes: I) release of the nascent amino acid chain and reinitiation of translation with proline as the first residue or II) dissociation of the ribosome complex terminating translation. This is interestingly in accordance with previous results which found a

substantial molar excess of the upstream product in several polyproteins comprising two reporter proteins linked via the 2A peptide (CAT-2A-GUS, GUS-2A-CAT, GFP-2A-GUS or GUS-2A-GFP) (Donnelly *et al.*, 2001a).

During classical translation, eukaryotic release factors eRF1 and eRF3 are responsible for stimulating the hydrolysis (release) of completed proteins from tRNA during termination by formation of a complex with the elongating eukaryotic ribosomes. Doronina et al., (2008) demonstrated that by depleting eRF1 in yeast, a greater proportion of ribosomes translate through the 2A coding sequence without cleavage and thus yield the full-length protein. Impaired eRF3 activity lead to many ribosomes failing to translate beyond 2A. These results suggested that stalled ribosome-2A complexes promote eRF entry without reading the specific mRNA, as the eRFs must recognize the A-site proline codon as, in effect, the termination signal. This must be followed by a dissociation of the eRF complex to facilitate the entry of a charged tRNA to the A site of the ribosome. The tRNA located in the A-site is then translocated to the P-site to become the initiating N-terminus tRNA of the downstream nascent protein chain. It is suspected that the level of eukaryotic elongation factors may determine whether, after hydrolysis, the ribosome dissociates or continues to translate the downstream context (Brown & Ryan, 2010). In contrast, a recent study using a fully reconstituted system, using recombinant factors based on human cell extracts, showed that EMCV 2A-mediated translational recoding functions in the absence of eRFs (Machida *et al.*, 2014). Thus the role of these factors is still controversial.

Specific amino acid substitutions within the FMDV 2A sequence, and especially within the highly conserved $D^{12}(V/I)E(S/T)NPG_{2A}{\downarrow}P^{19}{}_{2B}$ motif, drastically reduce the cleavage efficiency *in vitro* and can even block it entirely (Donnelly et al., 2001b; Sharma et al., 2012). Gao *et al.*, (2014) assayed four different synonymous codons for residue G^{18} of the 2A peptide in the context of an artificial reporter polyprotein within CHO cells. They were all shown to function with very similar apparent cleavage efficiencies at the 2A/2B junction. However, it is noteworthy that the cleavage efficiency of the system was only 88-89% and thus not optimal. These results indicate that it is the amino acid residues and not the nucleotide sequence within the 2A region which is instrumental for 2A activity.

The different 2A sequences can be transposed into new contexts and still remain active meaning that they are autonomous elements. However, despite being shown to work in a variety of eukaryotic settings (including plant, insect and mammalian cell systems) they do not

work in bacteria (Donnelly et al., 1997). The context of the 2A sequence is important. An appropriate upstream sequence has been demonstrated to be essential for 2A function. Sharma et al., (2012) generated hybrid 2A peptides by replacing the N- or C-terminal portions of the 2A region from FMDV, TMEV and EMCV with another 2A variant within the context of an artificial polyprotein system. The hybrid 2A variants showed low or no activity suggesting that each 2A variant is optimized for function in its natural state. By elongating the FMDV 2A, in an artificial polyprotein system, with upstream VP1 residues, thereby increasing the chain length to up to 30 amino acids, its activity was enhanced (Donnelly *et al.*, 1997, 2001b; Minskaia *et al.,* 2013). Similar results were observed for other picornavirus 2As, in which the 2A peptide was elongated with sequence from their respective polyprotein precursors (Luke *et al.*, 2008). Moreover, an extensive alanine (A), glycine (G) and proline (P) scanning mutagenesis of the entire FMDV 2A peptide showed that the majority of substitutions and especially those within the highly conserved D(V/I)E(S/T)NPG$_{2A}$↓P^{19}$_{2B}$ region reduced the cleavage activity (Donnelly *et al.*, 2001b; Sharma *et al.*, 2012). It is, however, difficult to distinguish critical residues which interact with the ribosome tunnel from flanking "space-filling" residues. Sharma *et al.,* (2012) suggested that the 2A peptides are fine-tuned to function as a single unit within their natural polyprotein.

Much of the research into the function of 2A has been conducted using *in vitro* experiments with mRNAs encoding artificial polyproteins comprising two reporter proteins linked via the 2A peptide. It has been considered that the cleavage mediated by the 2A peptide was essential for the viability of cardioviruses and FMDV. However not much information is available on 2A modifications in the context of the native polyprotein and its influence on the replication and translation. Loughran *et al.,* (2013) studied the effect of substitutions in the 2A coding sequences within the context of full-length TMEV and FMDV genomes on virus viability and polyprotein processing. Modification of the SNPG↓P$_{2B}$ sequence to SNPL↓V$_{2B}$ at the 2A/2B junction blocked polyprotein cleavage. Surprisingly, this modification had no observable effect on the growth of TMEV. However, interestingly, it was detrimental for the replication of mengovirus (another cardiovirus) and apparently lethal for FMDV. This suggested that the StopGo function is not essential for virus viability for certain cardioviruses but is critical for FMDV.

The FMDV 2A peptide is located at the boundary between the structural proteins and the non-structural proteins involved in RNA replication. It has been speculated the 2A peptide has dual functions (Tulloch et al., 2017). Firstly, to achieve cleavage of the polyprotein at the 2A/2B

junction but secondly to act as a translational regulator to down-regulate translation of the non-structural proteins in the later stages of infection. Such a mechanism has not yet been confirmed in FMDV but 2A has been found to act as a translational regulator in the translation of the EMCV genome. Recently, it was discovered that a programmed -1 ribosomal frameshifting motif was present within the coding region for the 2B protein. It directs a proportion of ribosomes into a short overlapping ORF resulting in the production of a "trans-frame" protein, $2B^{*}$ (Loughran *et al.*, 2011). The coding sequence for $2B^{*}$ harbours a stop codon at its C-terminus. The ribosome slip site (GGUUUUY) is also conserved in TMEV and other cardioviruses. Napthine et al., (2017) showed that the 2A protein, at 6-8 hours after infection (late stages of infection), interacts with an RNA structure beginning 14 nt downstream of the ribosome slip site, forming a RNA-protein complex that induces highly efficient programmed -1 ribosomal frameshifting. As a result, the frameshifting efficiency increases from 0 to 70% over the course of infection, thus downregulating the translation of the viral non-structural proteins.

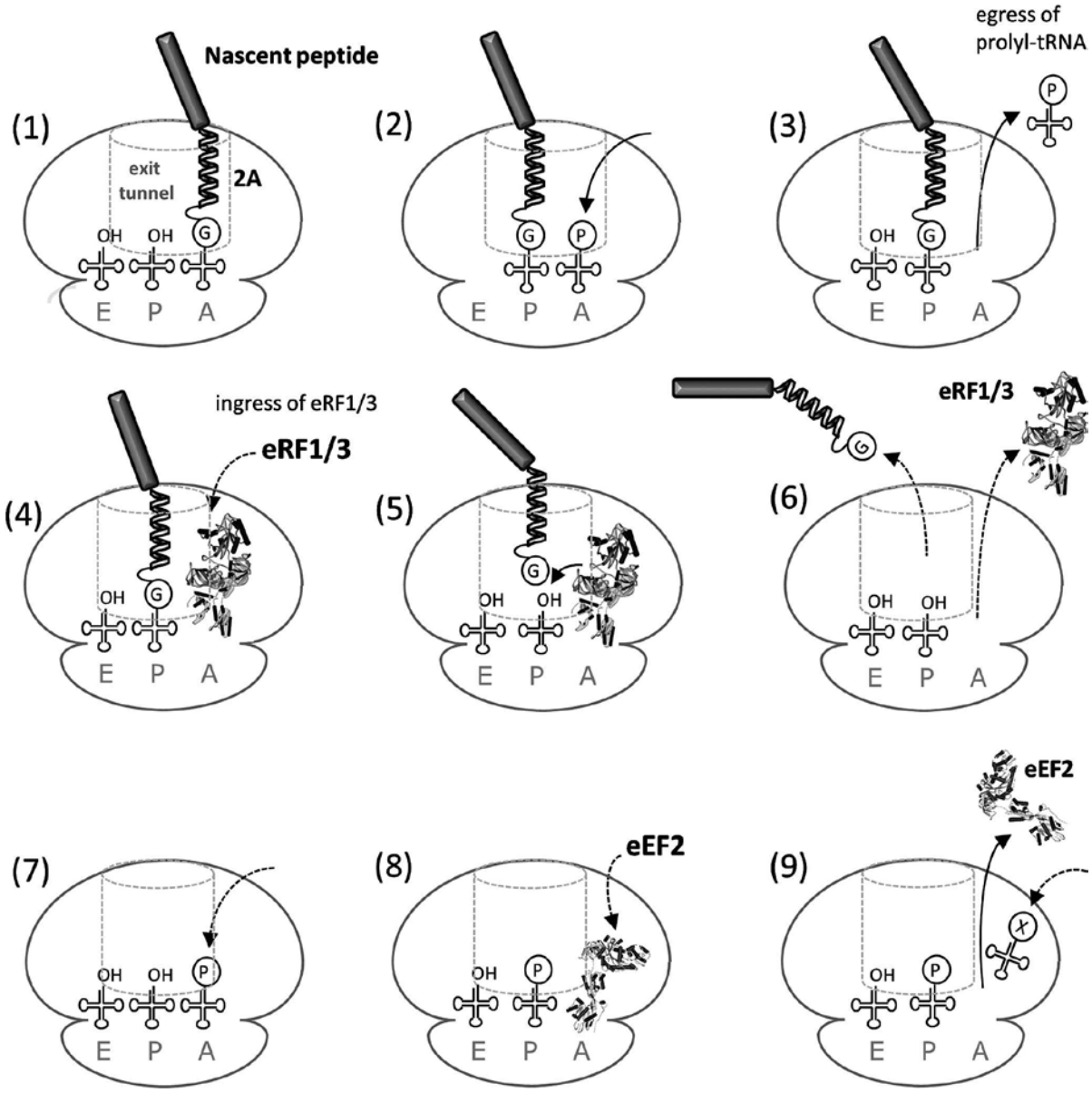

Figure 5: Model of the 2A mediated translational skipping at the 2A/2B junction. (1) The growing nascent polyprotein chain upstream of the 2A/2B junction exits the ribosome through the ribosomal exit tunnel. **(2)** The polyprotein-2A bound to the glycyl-tRNA is translocated from the A- to P-site after which proline charged prolyl-tRNA enters the A-site. **(3)** The 2A region in the growing polyprotein interacts with the ribosomal exit tunnel leading to a stereo-chemical reorientation of the turn motif of the C-terminal portion (ESNPG) within the peptidyl-transferase centre (PTC) of the ribosome. The peptide bond formation (nucleophilic attack) between 2A and the proline attached to the prolyl-tRNA in the A-site is inhibited thereby pausing the translation. The prolyl-tRNA leaves the A-site of the ribosome. **(4)** Release factors eRF1/3 binds to the A-site **(5)** eRF1/3 hydrolyse the ester linkage between the polyprotein and the glycyl-tRNA **(6)** The nascent protein and the eRF1/3 release factors leave the ribosome **(7)** The proline charged prolyl-tRNA enters the A-site **(8)** The translocation to the P-site is mediated by the elongation factor 2 (eEF2) **(9)** The next amino-acyl tRNA enters the A-site and protein synthesis of the proteins downstream of 2A continues without the need of a new reinitiation event. Adapted from Roulston *et al.*, (2016).

Functional structural RNA elements within the FMDV genome

The multi-functional RNA molecule

The central dogma has been a cornerstone in molecular biology since it first was defined in 1958 (Crick, 1970). It states that the irreversible transfer of genetic information from DNA to protein flows through RNA. RNA had previously only been considered a messenger for translation (mRNA) and had been divided into either coding or non-coding species (Jacob & Monod, 1960). The understanding of the role of RNA in biological systems has since expanded and today RNA is known to also act as regulatory, enzymatic and structural molecules (Garst *et al.*, 2011; Vandivier *et al.*, 2016). This has led to the formulation of the RNA World hypothesis (Gilbert, 1986). It postulates that RNA was crucial for the evolutionary history of life on Earth and that self-replicating RNA molecules with catalytic properties preceded DNA. To accomplish these divergent tasks, RNA forms functional structures through a variety of modular, intramolecular interactions through base pairs via Watson-Crick, Hoogsteen or sugar-edge patterns of hydrogen bonds in conjunction with co-factors and RNA-binding proteins. These structures define the catalytic active sites or modulate the function of regulatory motifs. RNA can fold into several distinct three-dimensional secondary structures including stem-loops, bulges, internal loops, hairpins, stems, terminal loop and multiloops (see Figure 6).

RNA structure and function

The overall structure, which RNA assumes after transcription, is essential for correct function of many RNAs. tRNAs are a fundamental component of the translational machinery and the most abundant of all small non-coding RNA molecules, constituting 4–10% of all cellular RNA. It has been shown that their overall tertiary structure is shared among the different tRNAs and that this structure is essential for tRNAs to become charged with their cognate amino acid (Giegé *et al.*, 2012). Ribosomal RNA (rRNA) is the most abundant class of RNA in both eukaryotic and prokaryotic cells. It is well established that the three-dimensional structure is an integral part of the function of ribosomes (Wilson & Cate, 2012). Other RNA molecules including non-coding miRNAs or small interfering RNA (siRNAs) also adopt distinct tertiary structures which enable them to function as post-transcriptional regulators of gene expression

or interfere with the expression of specific genes (Filipowicz *et al.*, 2005). The structures within the mRNAs found in both prokaryotes and eukaryotes can also influence translational efficiency. Insertion of stable hairpin loops around the Shine-Dalgarno sequence can alter the kinetics of translation and reduce expression levels (Kubo & Imanaka, 1989). Svitkin *et al.*, (2001) demonstrated that translation of mRNAs with stable secondary structures in the 5′-UTR were more sensitive to inhibition by an eIF4A mutant. eIF4A is believed to unwind the secondary structures located in the mRNAs 5′-terminus to facilitate ribosome binding (Rogers *et al.*, 1999; Sonenberg & Gingras, 1998). Furthermore, it has shown that the IRES elements from PV, EMCV and FMDV also were inhibited by the mutant.

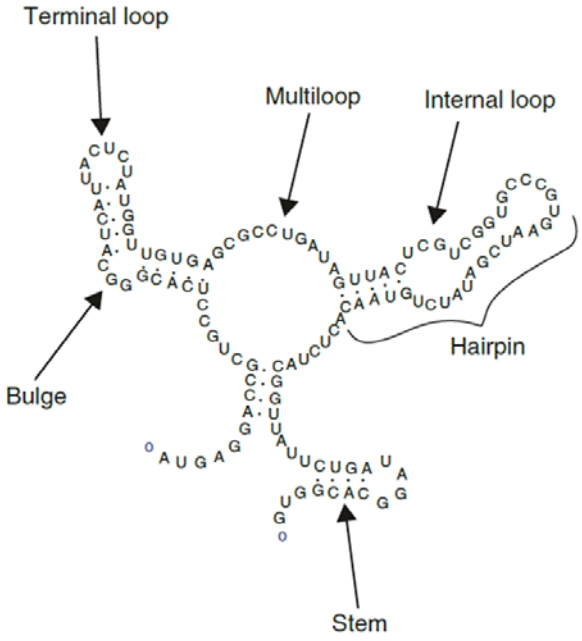

Figure 6: Components of RNA secondary structures. The single stranded RNA molecule generates elaborate secondary structures formed by base pairing between self-complementary sequences from different parts of the RNA chain**.** Adapted from Fariza *et al.*, 2017.

RNA secondary structure probing

A better understanding of the fundamental cellular processes, which these RNA are an essential part of, will depend on knowledge of the structures of the RNAs. To determine the function of either cellular or viral RNA molecules various chemical, physical, and computational methods have been developed to investigate their structure. Investigations into RNA folding were promoted by the characterization of the tertiary structure of both prokaryotic and eukaryotic tRNAs through the use of X-ray crystallography (Kim *et al.*, 1974; Robertus & Ladner, 1974). However, this technique requires transcripts to readily form crystals, which limits the technique

to the study of a few classes of RNA. There are, to date, a variety of structural probing methods available, either chemical or enzymatic in nature. The common feature for many of these methods is that they rely on modification of the RNA either through RNase digestion, RNA crosslinking, binding of complementing nucleotides to single-stranded RNA regions, nuclease treatments to determine scissions specific for single stranded/double stranded regions (Vandivier *et al.*, 2016). These methods can have nucleotide biases, preferentially cleaving/binding to some nucleotides more than others, potentially generating unreliable RNA structures. More accurate probing without nucleotide bias was achieved using selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) (Deigan *et al.*, 2009) (see Figure 7). This method uses electrophilic reagents such as isatoic anhydride (IA), N-methyl isatoic anhydride (NMIA), 1-methyl-7-nitro-isatoic anhydride (1M7), or benzoyl cyanide (BzCN) which react with the 2′-hydroxyl group of the nucleotide resulting in the formation of covalent adducts. The methods are therefore independent of the purine or pyrimidine nature of the bases. The reactivity is strongly dependent on the nucleotide flexibility since flexible nucleotides are more prone to form covalent adducts compared to constrained nucleotides (Weeks & Mauger, 2011). SHAPE reagents undergo hydrolysis when in contact with water, thus inactivating them and making an additional quenching step redundant. Adducts can be detected using reverse transcriptase-mediated primer extension. The presence of the 2′-O adduct hinders the elongation of the DNA strand and DNA synthesis by reverse transcriptase is terminated one nucleotide prior to the position of an adduct. The length of each cDNA reports the site of a SHAPE modification in the original RNA. A reaction using unmodified RNA is performed to determine the sites of natural reverse transcriptase pausing or pre-existing RNA degradation. The cDNA is subsequently sequenced using either conventional Sanger sequencing or next-generation sequencing (NGS) and peaks from the modified and no-reagent analyses are assigned by comparison with sequencing ladders. Subtracting the intensity of the “no-reagent” control from the modified RNA intensities results in a reactivity profile. Nucleotides constrained by base-pairing and tertiary interactions have low SHAPE reactivities whereas single-stranded and unconstrained nucleotides have higher reactivities.

In the last decade, the introduction of NGS technologies has revolutionized nucleotide sequencing and caused a shift away from first generation sequencing, i.e. Sanger sequencing, in the study of biological systems (van Dijk *et al.*, 2014). The major advance offered by NGS technologies are their abilities to produce, at a low cost, an enormous amount of data (thousands to millions of reads) which would otherwise be impossible for traditional DNA sequencing.

Recently, the coupling of probing SHAPE with NGS has allowed for higher throughput, increased accuracy, and the ability to perform experiments on a mixture of RNAs (multiplex).

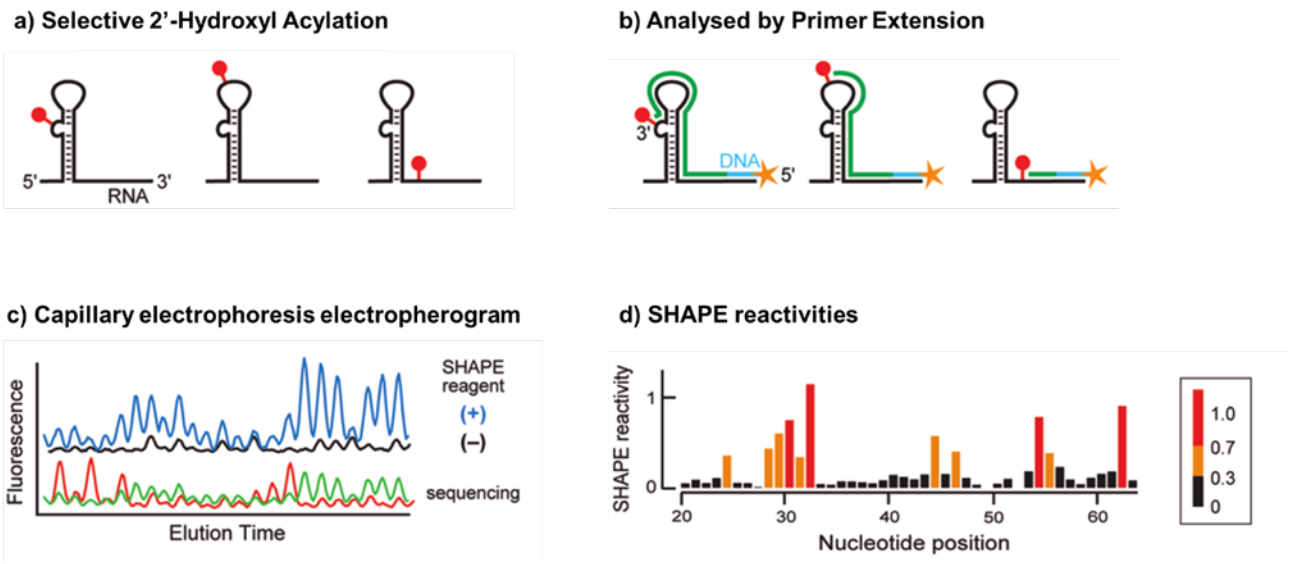

Figure 7: Schematic representation of the strategy of the Selective 2′-Hydroxyl Acylation analysed by Primer Extension probing. (a) The RNA is probed with a suitable SHAPE reagent. Flexible nucleotides are more prone to interact with the reagent forming adducts. **(b)** Reverse transcription primers are extended by reverse transcriptase which stops one nucleotide prior to the position of an adduct, thus revealing the positions of adduct formation. (**c)** Each generated cDNA fragment is sequenced to determine the site of the SHAPE modification. A no-reagent control and sequencing markers is included. (**d)** Subtraction of the intensity of the no-reagent control from the modified RNA yields a normalized reactivity profile. Adapted from Weeks & Mauger, (2011).

Conserved RNA structures in FMDV

The FMDV genome harbours multiple RNA structures, which have been demonstrated to be crucial for both genome replication and translation. These structures are primarily found in the 5′-UTR and 3′-UTR regions of the genome. In the 5′-UTR (see Figure 8) they are divided into the five functional elements termed the S-fragment, poly(C) tract and pseudoknots, *cre* and IRES. The 3′-UTR contains two stem-loops (Belsham, 2005).

The S-fragment is approximately 360 nucleotides in length and its sequence is predicted to fold into a long stem-loop (Clarke *et al.*, 1987; Witwer *et al.*, 2001). The exact function of the S-fragment is not known. It was been reported that the S-fragment can interact with the cellular proteins RNA helicase A, PABP and PCBP, together with the viral proteins 2C, 3A and $3C^{pro}$, which are all involved in genome replication (Lawrence & Rieder, 2009). Furthermore, it has been established that the S-fragment binds to structural motifs in the 3′-UTR which are

presumed to lead to circularization of the viral genome (Lawrence & Rieder, 2009; Serrano *et al.*, 2006). Recently, Kloc *et al.*, (2017) have demonstrated that FMDV mutants with deletions in the S-fragment triggered an increase in type I interferon mRNA levels (innate immune response). Furthermore, the deletion of a 164 nucleotide sequence in the upper S-fragment loop proved to be highly attenuated in an FMDV mouse model system. This suggest that the S-fragment is involved in modulating the host innate immune response.

Following the S-fragment is a long stretch of C-residues termed the poly(C) tract, which is common to the aphthoviruses and most cardioviruses. The poly(C) tract exhibits considerable variation in length among the variety of different FMDV isolates, from 80 nt to 450 nt. The short tracts are most often found within laboratory strains whereas field strains of FMDV, in general, contain poly(C) tracts of around 200 nt in length. The length of this region has previously been shown to significantly influence pathogenicity within cardioviruses (Duke *et al.*, 1990). Although some early studies suggested that the length of the poly(C) tract was associated with virulence of FMDV (Harris & Brown, 1977; Rieder *et al.*, 1993), others have not been able to directly associate the two (Costa Giomi *et al.*, 1984). It has been shown that the poly(C) lengths of a genetically modified FMDV genome requires a minimum length to function optimally. Genetically modified FMDV genomes which contained just six C-residues were selected against. The rescued viruses, after a few passages in cell culture, contained poly(C) tracts with at least 80 C-residues. In contrast, the tract did not increase in length when the modified FMDV genomes only contained two C-residues. This virus remained as pathogenic in suckling mice as the wt, albeit with a lower replication efficiency in tissue culture (Rieder *et al.*, 1993). Modified mengoviruses with short poly(C) tracts were viable in cell culture, but attenuated in mice. The rescued viruses were genetically stable after several passages (Duke *et al.*, 1990). In contrast, the function of the poly(C) tract appear to be different in the closely related EMCV. Modified EMC viruses with different sized poly(C) tracts were shown to possess the same level of pathogenicity in animals as the wt and also demonstrated the same replication rate using growth-curves (Hahn & Palmenberg, 1995). Downstream of the poly(C) tract in FMDV RNA is a region, which is predicted to fold into a series of RNA pseudoknot structures (Escarmís *et al.*, 1995). The function of these elements are still unknown, but interestingly the poly(C) tract is located on the 3′ side of the pseudoknots within the genomes of cardioviruses (Martin & Palmenberg, 1996) thereby making it plausible that these elements have some form of interaction.

RNA structures, now termed cis-acting replication elements (*cre*), were first characterized by McKnight & Lemon (1996) who demonstrated that a sequence within the P1 coding region of the human rhinovirus 14 genome was required for viral RNA replication. The sequence was necessary in the form of RNA and is predicted to form a stem-loop structure containing a conserved AAAC motif in the loop region. This motif serves as the template for the RdRp mediated uridylylation of 3B (VPg) to produce VPgpU and/or VPgpUpU, collectively termed VPgpU(pU) (Paul *et al.*, 1998). VPgpU(pU) serves as a primer for the initiation of viral RNA synthesis. Interestingly, the location of the *cre* within other picornavirus genomes has been identified in various different places within the coding region of the respective genomes (Goodfellow *et al.*, 2000). FMDV is unique is this regard in having the *cre* structure embedded outside of the coding region and within the 5′-UTR (Mason *et al.*, 2002). The importance of this region is demonstrated by the impaired replication of FMDV replicons when the AAACA motif in the stem region was modified. Furthermore, it was also shown that the position of the *cre* within the 5′-UTR of the genome was not essential for replication, since insertion of the structure into the 3′-UTR could restore replication ability in replicons and viruses lacking the original 5′ *cre*. Tiley *et al.*, (2003) demonstrated that the *cre* can be complemented in *trans* by co-infecting a temperature-sensitive FMDV with a mutation within the *cre* with another ts mutant FMDVs with a defects in other regions of the genome. The *cre* has, therefore, been suggested to be renamed a 3B-uridylylation site (*bus*).

The region between the *cre* and ORF is termed the internal ribosome entry site (IRES) and is approximately 450 nt in length and contains a series of highly conserved complex stem-loop structures essential for the cap-independent internal initiation of translation on the viral RNA (Kühn *et al.*, 1990; Belsham & Brangwyn, 1990). IRES-driven translation initiation depends on the recognition of the IRES by both translation initiation factors and other cellular proteins necessary for translation of cellular mRNAs. These proteins interact with the five highly conserved domain structures, H-L, and with a pyrimidine-rich region at its 3′ end immediately preceding the first AUG translation initiation codon to promote translation (Pilipenko *et al.*, 1989). Domain H contains a polypyrimidine tract (UCUUU) that provides a binding site for PTB. Domain I has a distinctive cruciform structure that harbours three conserved essential motifs, GNRA and RAAA at the apical region, and a conserved C-rich loop. Modifications of either the GNRA or RAAA motif led to a reorganization of the entire domain and significantly diminished the IRES activity (López de Quinto *et al.*, 2002). Domains J and K are arranged as two stem-loop structures in a Y-shape containing A residue-rich internal bulges and act as the

binding site for the essential translation initiation factor eIF4G. Domain L consists of a short conserved hairpin-loop with a polypyrimidine tract at 20 nt upstream of the initiation codon AUG. This provides the binding site for eIF4B, PTB and other cellular proteins necessary for IRES driven translation (Belsham, 2005; Belsham & Sonenberg, 1996; Martinez-Salas & Belsham, 2017).

The 3′-UTR of FMDV consists of a unique sequence of ca. 100 nt folded into two separate stem-loops (Witwer *et al.*, 2001) followed by a poly(A) tract of varying length. Studies have demonstrated the 3′-UTR of FMDV is essential for viral replication, since deletion of this region renders viral replication impossible (Sáiz *et al.*, 2001). In contrast, PV replication can still proceed without this region, although at a reduced level (Todd *et al.*, 1997). The 3′-UTR can bind to both the S-fragment and the IRES domains within the 5′-UTR through long-range RNA-RNA interactions (Serrano *et al.*, 2006). Interaction between the 3′-UTR and the IRES requires both stem–loop structures of the 3′-UTR whereas only one stem-loop is required for the S-fragment interaction.

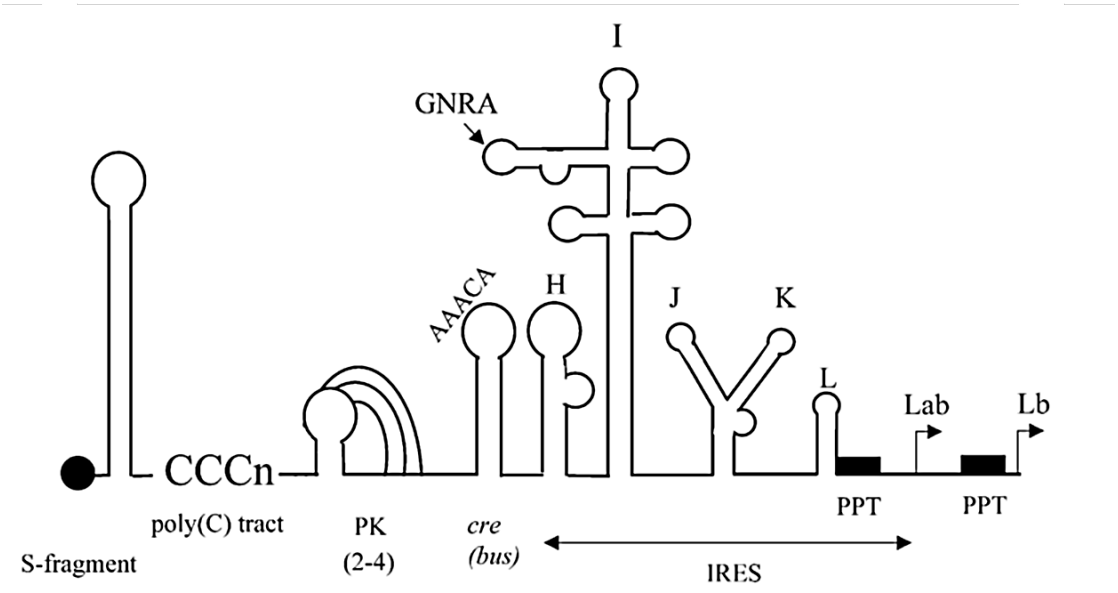

Figure 8: Schematic representation of the structural elements within the 5′-UTR of FMDV. The 5′-UTR has been shown to contain five secondary RNA structures termed the S-fragment, poly(C) tract, cis-acting replication element (*cre*) or alternatively the 3B-uridylylation site (*bus*), pseudoknots (PKs), and internal ribosome entry site (IRES). A detailed description of the investigations into the exact function of these elements is described in the text. The IRES is further divided into individual domains (H-L). The conserved GNRA motif and the polypyrimidine tracts (PPTs) are indicated. The conserved and essential AAACA motif in the *cre*/*bus* is also indicated. The two different initiation sites (Lab and Lb) are shown. Adapted from Belsham, (2005).

SHAPE probing of the FMDV genome

Despite the well-established influence of the previously described functional RNA elements on FMDV replication and translation the nature of many of these RNA structures and the effects of sequence variation on their structure are poorly understood. Recently, Diaz-Toledano *et al.,* 2017 investigated the RNA local flexibility of the FMDV IRES and the 3′-UTR in the context of RNA transcripts *in vivo*. It has been suggested that long-range RNA-RNA interactions between the IRES and the 3′-UTR exist, but whether the interaction induces a conformational change in the structures has remained unknown. This study used a novel approach by using SHAPE in living cells thus providing better and more accurate structure models of the RNA in its natural environment. It was demonstrated that the SHAPE results using the FMDV IRES in its natural state showed significant differences with *in vitro* data using RNA transcripts, revealing protected or exposed positions within the IRES and the 3′-UTR. These data have provided new information concerning the local flexibility of the IRES and 3′-UTR confirming long-range interactions between these distant elements of the viral genome and proving a conformational change during interaction.

IRES function is tightly linked to the RNA structure and substitutions which affect the structure significantly diminish function. The GNRA motif located within region I adopts a tetraloop conformation and has been shown to mediate the overall structure of the region by a possible RNA-RNA interaction. Fernández *et al.,* (2011) confirmed, using SHAPE probing, that the conformation of region I depends on specific nucleotides of the GNRA loop and that this loop most likely are involved in tertiary interactions within region I. SHAPE probing of the FMDV genome has so far been focused on the IRES and research into the structure of the remaining structures in the 5′-UTR are still to be conducted. In addition, little is known about the RNA structures in the coding region of the FMDV. Some picornaviruses have been found to also harbour RNA structures within the coding sequence of the genome which are essential for viral RNA replication and translation. McKnight & Lemon, (1996) demonstrated that a sequence within the capsid coding sequence is required for efficient replication of human rhinovirus 14. A similar element has been identified in the 2C coding region of PV (Goodfellow *et al.,* 2000). A global RNA structure map of the PV genome generated using SHAPE probing has recently also demonstrated several conserved RNA structures in the ORF of the viral genome (Burrill *et al.,* 2013). In particular, a structure located in the region encoding the RdRp has been shown to be required for viral replication and infectivity. The study further suggested a novel interaction between an RNA structure in the RdRp coding region and the viral protein(s) $3C^{pro}$

and/or its precursor $3CD^{pro}$. Similar global RNA structure maps of the FMDV genome have also been predicted (Davis *et al.*, 2008; Simmonds *et al.*, 2004), however the functional role of these structures has not been determined, but have been suggested to be implicated in the evasion of cell defences (Simmonds *et al.*, 2008). Kloc *et al.*, (2017) have recently demonstrated that the S-fragment within the 5′-UTR of FMDV is involved in the regulation of the host innate immune response.

In silico prediction of RNA secondary structures

Chemical structure probing analysis can provide reliable structural data, but the methods can be laborious. In contrast, computational prediction (*in silico*) of RNA secondary structures has shown to be fast. *In silico* predications are based on either thermodynamics of the molecules (minimum free energy) (Mathews *et al.*, 1999), conservation of RNA structures between organisms or a combination of the two (James *et al.*, 1989). The thermodynamic approach relies on the principle that the real structure must be the most thermodynamically stable, i.e., the one with the minimum free energy. The mfold software for RNA folding was developed in the late 1980s (Zuker, 2003), and has been used in numerous papers. The program predicts a minimum free energy for an input sequence. However, there are limitations to this model including that RNA structures are influenced by the environment such as other macromolecules (other RNAs, protein, etc.) which can give rise to new RNA structures with lower energy levels through interactions with the original. Additionally, the fidelity of *in silico* predictions decreases as a function of increasing RNA complexity (sequence length). The comparative approach is only applied when homologous RNA sequences are available.

Part 2 · Manuscripts

Manuscript 1

Modifications to the foot-and-mouth disease virus 2A peptide; influence on polyprotein processing and virus replication

Jonas Kjær & Graham J. Belsham

Submitted to Journal of Virology

2017

Modifications to the foot-and-mouth disease virus 2A peptide; influence on polyprotein processing and virus replication

Jonas Kjær[1] and Graham J. Belsham[1*].

1: DTU National Veterinary Institute, Technical University of Denmark, Lindholm, DK-4771, Kalvehave, Denmark

*Corresponding author: Graham J. Belsham, email: grbe@vet.dtu.dk, phone: +45 3588 7985

Figures: 7

Tables: 2

Abstract

Foot-and-mouth disease virus (FMDV) has a positive-sense ssRNA genome that includes a single, large, open reading frame encoding a polyprotein. The co-translational “cleavage” of this polyprotein at the 2A/2B junction is mediated by the 2A peptide (18 residues in length) using a non-proteolytic mechanism termed “ribosome skipping” or “StopGo”. Multiple variants of the 2A polypeptide with this property among the picornaviruses share a conserved C-terminal motif (D(V/I)E(S/T)NPG↓P). The impact of 2A modifications within this motif on FMDV protein synthesis, polyprotein processing and virus viability were investigated. Amino acid substitutions can be tolerated at residues E^{14}, S^{15} and N^{16} within the 2A sequence of infectious FMDVs despite their reported “cleavage” efficiencies at the 2A/2B junction of only ca. 30-50% compared to wt. In contrast, no viruses were rescued containing substitutions at residues P^{17}, G^{18} or P^{19} that displayed little or no “cleavage” activity *in vitro*, but wt revertants were obtained. The 2A substitutions impaired the replication of a FMDV replicon. Using transient expression assays, it was shown that certain amino acid substitutions at residues E^{14}, S^{15}, N^{16} and P^{19} resulted in partial “cleavage” of a protease-free polyprotein indicating that these specific residues are not essential for co-translational “cleavage”. Immunofluorescence studies, using full-length FMDV RNA transcripts encoding mutant 2A peptides, indicated that the 2A peptide remained attached to adjacent proteins, presumably 2B. These results show that efficient “cleavage” at the 2A/2B junction is required for optimal virus replication. However, maximal StopGo activity does not appear to be essential for the viability of FMDV.

Importance

Foot-and-mouth disease virus (FMDV) causes one of the most economically important diseases of farm animals. Co-translational “cleavage” of the FMDV polyprotein precursor at the 2A/2B junction, termed StopGo, is mediated by the short 2A peptide through a non-proteolytic mechanism which leads to release of the nascent protein and continued translation of the downstream sequence. Improved understanding of this process will not only give a better insight into how this peptide influences the FMDV replication cycle but may also assist the application of this sequence in biotechnology for the production of multiple proteins from a single mRNA. Our data show that single amino acid substitutions in the 2A peptide can have a major influence on viral protein synthesis, virus viability and polyprotein processing. It also indicates that efficient “cleavage” at the 2A/2B junction is required for optimal virus replication. However, maximal StopGo activity is not essential for the viability of FMDV.

Introduction

Foot-and-mouth disease virus (FMDV) is the causative agent of foot-and-mouth disease, a highly contagious disease of domestic and wild cloven-hooved animal species. The virus has been successfully eradicated from Europe but is still endemic in many regions of the world (in Asia, Africa and the Middle East) and can potentially cause major outbreaks in domestic livestock elsewhere with severe economic losses (reviewed in [1]). FMDV is the prototypic member of the *Aphthovirus* genus within the family *Picornaviridae*. These viruses are small (ca. 25-30nm) and have a positive-sense ssRNA genome [2]. The FMDV genome is ~8.5 kb in length and includes a single, large, open reading frame (ORF) encoding a long polyprotein of over 2300 residues [3]. However, this polyprotein is never observed within infected cells due to rapid co- and post-translational processing to produce, initially, the mature Leader protein (L^{pro}) and the precursor proteins P1-2A, P2, and P3. The L^{pro} is a papain-like protease and cleaves the polyprotein at its own C-terminus; that is the junction between L^{pro} and the capsid precursor P1-2A [4, 5]. The 3C protease ($3C^{pro}$) is responsible for proteolytic cleavage of P1-2A to produce the structural proteins VP0, VP3 and VP1 plus the 2A peptide. The P2 and P3 precursors are also processed by $3C^{pro}$ to generate the non-structural proteins, required for the replication of the viral genome. The processing of VP0 to VP4 and VP2 occurs during encapsidation of the viral RNA although it can also occur on assembly of empty capsid particles [6, 7].

The separation of the P1-2A precursor from 2B (within P2) is achieved by yet another mechanism. There is considerable heterogeneity among the picornaviruses with respect to the 2A peptide/protein that is located on the C-terminal side of the capsid protein precursor. Both the size and the function of the different 2A species differ between different picornavirus genera [8]. The entero- and rhinovirus 2A proteins, termed $2A^{pro}$, are thiol proteinases of ~150 amino acids which catalyse the proteolytic cleavage of the junction between the P1 and P2 precursors at their own N-termini, i.e. at the P1/2A junction [9, 10]. In contrast, in the aphthoviruses and cardioviruses (e.g. encephalomyocarditis virus (EMCV) and Theiler's murine encephalitis virus (TMEV)), the separation of the capsid precursor (P1-2A) from P2 (2BC) occurs at the 2A/2B junction, i.e. at the C-terminus of 2A. In FMDV, the 2A peptide is only 18 amino acids long and lacks any characteristic protease motifs [11–13]. Earlier studies have demonstrated that the "cleavage" at the 2A/2B junction is not dependent on the FMDV proteases L^{pro} or $3C^{pro}$ nor on host cell proteases [12, 14, 15]. The FMDV 2A peptide contains a highly conserved amino acid sequence at its C-terminus, $D^{12}(V/I)E(S/T)NPG_{2A}{\downarrow}P^{19}{}_{2B}$. This

sequence induces a co-translational "cleavage" event that is referred to as "ribosomal skipping" [16] or, alternatively, "stop-carry on" or "StopGo" [17, 18]. The first residue of 2B (pro or P) is referred to as P^{19} as it is a key part of the "cleavage" site. This motif, together with upstream amino acids that form an α-helix over most of its length, are believed to interact with the exit tunnel of the translating ribosome and prevent the formation of a peptide bond between the C-terminal amino acid, glycine (gly, or G), of 2A and the first residue, proline (P), of 2B [16, 19]. This produces a break in the growing amino acid chain, but the process of protein synthesis continues, without the requirement for a new translation initiation event. The same conserved motif is also present at the C-terminus of the cardiovirus 2A proteins that are substantially larger; the FMDV 2A sequence appears to be the minimal functional entity to break the growing polypeptide chain (this process will simply be described as cleavage subsequently, for convenience) and other functions have been assigned to the cardiovirus 2A protein as well [20]. The majority of studies on the function of the FMDV 2A peptide have been conducted using *in vitro* experiments with mRNAs encoding artificial polyproteins comprising two reporter proteins linked via the 2A peptide, thus generating two separate translation products (e.g. [16, 21, 22]). These previous studies have shown that specific amino acid substitutions within the FMDV 2A sequence, and especially within the highly conserved $D^{12}(V/I)E(S/T)NPG_{2A}{\downarrow}P^{19}{}_{2B}$ motif, drastically reduce the apparent cleavage efficiency and can even block it entirely. These results indicate that these amino acid residues are critical for optimal ribosomal skipping [16, 22]. Furthermore, in the context of a synthetic reporter polyprotein, assayed within CHO cells, four different synonymous codons for residue G^{18} of the 2A peptide were shown to function with very similar apparent cleavage efficiencies at the 2A/2B junction but the cleavage efficiency was not optimal, only 88-89% complete, in this system [23]. These results were interpreted as showing that it is this amino acid residue rather than the nucleotide sequence which is critical for achieving cleavage [23]. The 2A peptide has been shown to mediate cleavage in all eukaryotic translation systems tested, whereas a number of artificial polyproteins containing this sequence have been examined in prokaryotic systems and no detectable cleavage products were observed [22].

The less conserved part of the 2A sequence, located upstream of the $D(V/I)E(S/T)NPG_{2A}{\downarrow}P_{2B}$ motif, has also been shown to be important for optimal 2A function. Chimeric FMDV, TMEV and EMCV 2A peptides were generated by replacing the N- or C-terminal portions with another 2A variant and then assayed within artificial polyprotein systems where they showed little or no activity [24]. In addition, when the FMDV 2A, in an artificial polyprotein system, was elongated by the addition of up to 30 amino acids, from the upstream VP1, then its apparent

cleavage activity was enhanced [16, 22, 25]. Thus, the context of the 2A sequence is important. The 2A peptides from other picornaviruses exhibited similar increases in activity when elongated with 30 amino acids from their respective polyprotein precursors [8]. Moreover, an extensive alanine (A), glycine (G) and proline (P) scanning mutagenesis of the entire FMDV 2A peptide showed a decrease in apparent cleavage activity for all mutants [24]. This supports the view that the specific identity of the amino acid at nearly all positions within the 2A peptide is important for activity and that 2A peptides are fine-tuned to function as a single unit within their natural polyprotein.

In the studies of Loughran et al. [26], a number of mutations in the 2A coding sequences within the full-length TMEV and FMDV genomes were tested for their effects on virus viability and polyprotein processing. Modification of the SNPG$^{\downarrow}$P$_{2B}$ sequence to SNPL$^{\downarrow}$V$_{2B}$ at the 2A/2B junction blocked polyprotein cleavage. However, this modification had no significant effect on the growth of Theiler's murine encephalomyelitis virus (TMEV) whereas it was detrimental for the replication of mengovirus (another cardiovirus) and apparently lethal for FMDV. Thus, it was concluded that the 2A/P2 cleavage event is not essential for virus viability for certain cardioviruses but is critical for FMDV.

In this study, we have re-investigated the effect of 2A modifications in the context of the native FMDV polyprotein and its effect on virus protein synthesis and replication, virus viability and on polyprotein processing. In contrast to earlier studies, mutant infectious FMDVs having certain amino substitutions within the 2A peptide have been obtained but such changes do adversely affect virus replication and polyprotein processing to some degree.

Results

Effect of single amino acid substitutions in 2A on FMDV viability

Several studies using artificial polyprotein systems have demonstrated that nearly all positions of the 2A peptide are important for the "StopGo" activity and modifications can severely impair cleavage [22, 24]. To establish whether the StopGo activity plays a crucial role in FMDV viability, this study set out to investigate the constraints on the 2A sequence within the context of the full-length FMDV genome (see Fig. 1a).

To determine the viability of FMD viruses with single amino acid substitutions within the 2A peptide, selected modifications that were previously found to impair, to different extents, the StopGo activity in artificial polyproteins systems [22, 24] were introduced into the plasmid, pT7S3, that contains the full length FMDV cDNA [27], using site-directed mutagenesis (see

Methods). The resultant plasmids were linearized and RNA transcripts, prepared *in vitro*, were introduced into BHK cells by electroporation. Unexpectedly, all of the FMDV 2A mutants (Table 1) produced viable progeny viruses, with full CPE detectable after the second passage. The rescued viruses were then sequenced to identify possible adaptations or reversions (Table 1). After three passages of the 2A mutant viruses in cells, the viruses rescued from the transcripts encoding the N16H, E14Q, S15F and S15I modifications, had each retained the plasmid-derived amino acid substitutions. In contrast, the rescued viruses derived from the N16A, G18, P17A, P19G and P19A mutant transcripts all matched the wt sequence (i.e., the rescued viruses were not mutant) even when 2 nt changes were required to achieve this (e.g., see Fig. 2 for P19A mutant). To determine whether this reflected reversion to the wt sequence or some form of contamination/carryover of the wt sequence, three synonymous substitutions were inserted ca. 20 nucleotides downstream of the 2A/2B junction in the wt and the N16A, P17A, G18A and P19A mutant plasmids (see Fig. 2) after which the "marked" RNA transcripts were introduced into BHK cells. After three passages, the rescued mutant viruses had lost the 2A modification in each case but had, like the "marked" wt virus, each retained the three synonymous substitutions in the 2B coding region (see Fig. 2); this provides strong evidence that the presence of the wt sequence in the rescued viruses reflects reversion.

The growth characteristics of the wt and the viable 2A mutant viruses in BHK cells were examined in more detail by determining growth curves using a multiplicity of infection (m.o.i.) of 0.1. Surprisingly, both the wt and the viable 2A mutants grew with similar kinetics (Fig. 3a). Analysis of the FMDV capsid proteins within cells infected with the wt and the viable 2A mutant viruses, as determined by immunoblotting using anti-FMDV, is shown in Fig. 3b. As expected, the production of the capsid proteins was similar for the wt and the 2A mutants in each of the infected cell extracts (Fig. 3b, lanes 1-5).

Requirements for efficient 2A/2B "cleavage" in its native context

To examine the effects of the 2A mutants on the StopGo cleavage at the 2A/2B junction in its natural context and in cells, a plasmid encoding a truncated FMDV polyprotein termed the P1-2A-2BC-FLAG protein with a FLAG epitope at its C-terminus was generated (see Fig. 1d). The transient expression of this truncated viral polyprotein (without any proteases) was designed to permit the simultaneous assessment of the production of the uncleaved P1-2A-2BC-FLAG (ca. 135 kDa) and of the "cleavage" product 2BC-FLAG (ca. 54 kDa). The coding sequences for the wt or mutant P1-2A-2BC-FLAG products were under the control of the T7

promoter. The plasmids, were transfected into BHK cells that had been infected with the recombinant vaccinia virus vTF7-3 [28] which expresses the T7 RNA polymerase. The expression and processing of the proteins generated from these plasmids was visualized in immunoblots using anti-FLAG antibodies. Expression of the wt cassette led to apparently complete cleavage of the P1-2A-2BC-FLAG polyprotein as expected (Fig. 4a, lane 1) and thus only the 2BC-FLAG product was observed. In contrast, the E14Q mutant generated a mixture of both uncleaved and cleaved products (Fig. 4a, lane 5). Unexpectedly (c.f. [22, 24]), in the system used here, the S15A, S15F and S15I mutant proteins were each efficiently cleaved (Fig. 4a, lanes 6-8). The mutants N16C, N16H, P19A, P19G, P19V and P19S all produced a mixture of cleaved and uncleaved products (Fig. 4b, lanes 2, 3 and Fig. 4c, lanes 1-4). However, the D12A, V13A, E12A, N16A, N16V, N16W, P17A and G18A substitutions resulted in the production of only the uncleaved product and hence these mutant 2A peptides were all inactive in this system (Fig. 4a, lanes 2, 3, 4 and Fig. 4b, lanes 1, 4, 5, 6, 7). Overall, there is partial agreement between the results described here, using assays of the 2A in its near native context within cells, and those described previously [22, 24]. The main discrepancies concern the S15A, S15F and S15I mutants which resulted in essentially complete cleavage (≥90%) here but gave rather sub-optimal cleavage (42 and 39% of wt, respectively) *in vitro* [22] while the P19A, P19G, P19V and P19S mutants resulted in detectable, but low level, cleavage (8-20%) here but completely abrogated cleavage *in vitro* [22]. The same cell lysates were also analysed using an anti-FMDV capsid protein antibody to detect the intact polyprotein and the P1-2A product (data not shown). The pattern of results was fully consistent with those obtained using the anti-FLAG to detect the intact polyprotein and the 2BC-FLAG product. Thus it seems that the efficiency of cleavage detected in this assay system is higher than that observed using cell-free translation systems *in vitro*.

Influence of the amino acid substitutions in FMDV 2A on FMDV RNA replication efficiency assessed using a replicon that expresses the *Gaussia* luciferase

To evaluate the impact of the 2A mutants on the replication of viral RNA, nine different substitutions within the 2A coding sequence were introduced into a FMDV replicon (see Fig. 1b). In this replicon, the coding sequences for the FMDV structural proteins (VP1-VP3) have been replaced by the sequence encoding the *Gaussia* luciferase (Gluc) reporter protein, thus allowing replication to be readily monitored via measurement of Gluc expression. RNA transcripts were produced *in vitro* from the linearized plasmids and introduced into BHK cells using electroporation. As a negative control, a derivative of the wt-Gluc replicon was produced

which lacks a portion of the coding sequence for 3Dpol (the RNA dependent RNA polymerase) and is termed wt-GlucΔ3D (see Fig. 1c). Lysates were prepared from cells at various times after electroporation with the different transcripts and assayed for Gluc activity (see Fig. 5). The wt-Gluc-Δ3D transcript produced Gluc initially, that was already detectable at 1 h post-electroporation, but no further increase in luciferase activity was observed after 2 h. This expression presumably represents the translation of the input RNA. In contrast, the replication-competent wt-Gluc, while generating an initially similar level of Gluc activity at 2 h, showed a sustained increase in expression at later time points. Interestingly, all of the 2A mutants expressed low levels of Gluc activity initially, almost 10-fold less than the wt-GlucΔ3D at 2 h. However, the expression increased to some degree at later time points; the level of Gluc expression first surpassed the polymerase knockout mutant after 6 h. It is noteworthy that the mutant transcripts with the E14Q, S15F, S15I and N16H changes, which were retained in the rescued viruses, did not have better RNA replication efficiencies than the other 2A mutants. This may reflect sub-optimal cleavage at the 2A/2B junction due to the absence of the upstream VP1 coding sequences in these replicons (see [25]).

Influence of the StopGo function on the correct processing of the FMDV P1-2A precursor

Hahn & Palmenberg [29] demonstrated that amino acid substitutions within the conserved D(V/I)E(S/T)NPG$_{2A}$↓P$_{2B}$ motif at the C-terminus of the 2A protein of EMCV not only severely reduced or abrogated the StopGo function but also impaired the subsequent cleavage of L-P1-2A by 3C^{pro}. The effects of substitutions in 2A on the FMDV P1-2A processing in cells has now been assayed using the truncated FMDV polyprotein termed P1-2A-2BC-FLAG (as above, see Fig. 1d) which was co-expressed with the FMDV 3C^{pro}. The wt and mutant P1-2A-2BC-FLAG plasmids encoding the N16A, P17A, G18A and P19A substitutions (shown in Fig. 4 to abrogate or impair (P19A) cleavage) were transfected, alone or together with a plasmid that expresses the 3C^{pro} (as in [7]), into vTF7-3-infected BHK cells. Analysis of the FMDV P1-2A processing, was determined by immunoblotting using anti-VP2 antibodies and is shown in Fig. 6a. Expression of the wt plasmid alone led to complete cleavage at the 2A/2B junction of the P1-2A-2BC-FLAG polyprotein, to yield P1-2A, as expected (Fig. 6a, lane 1). Furthermore, co-expression of the wt product with the 3C^{pro} produced VP0 (from the P1-2A) also as expected (Fig. 6a, lane 2). When the mutant plasmids, with defective cleavage at the 2A/2B junction, were expressed alone then the larger, intact, P1-2A-2BC-FLAG product was detected (Fig. 6a, lanes 3, 5, 7, 9), as above (see Fig. 3). In the presence of the 3C^{pro}, the production of VP0, derived from P1-2A, was still readily apparent in each case (Fig. 6a, lanes

4, 6, 8, 10). These results were confirmed by immunoblotting using anti-FMDV antibodies (Fig. 6b). Co-expression of the wt and mutant plasmids with the $3C^{pro}$ produced a very similar pattern of detectable capsid proteins in each case (Fig. 6b, lanes 2, 4, 6, 8, 10). Thus, abrogating cleavage at the 2A/2B junction did not block the processing of the capsid precursor by $3C^{pro}$ in this system. It should be noted that this is in contrast to some earlier studies [14], which showed that a truncated version of FMDV P1-2A (lacking the C-terminus of VP1) could not be processed at all by $3C^{pro}$.

Detection of a novel FMDV 2A-2B fusion protein using immunofluorescence

The FMDV capsid protein precursor, P1-2A, is normally processed by the $3C^{pro}$ to VP0, VP3, VP1 and 2A. In previous studies, it has been shown that when the cleavage of the VP1/2A junction is impaired, then the presence of FMDV 2A (still attached to VP1, as VP1-2A) can be detected in BHK cells by immunofluorescence using anti-2A antibodies [7, 30]. When the 2A is released from the VP1 then the 2A is no longer detectable (presumably it is either degraded or not fixed in the procedure). Thus, it seemed possible that substitutions within the 2A peptide that impair the 2A/2B cleavage activity (and prevent formation of viable, mutant, viruses), would result in the formation of detectable 2A-2B fusion proteins. Full-length FMDV RNA transcripts, with or without modifications in 2A, were introduced into BHK cells by transfection and after 8 hrs, the cells were stained with either anti-2A antibodies or anti-FMDV capsid protein antibodies. The FMDV VP1 K210E mutant, as previously described [7], which produces an uncleaved VP1-2A protein, was included as a positive control for the detection of 2A attached to an adjacent protein. FMDV capsid proteins could be detected in cells transfected with each of the RNA transcripts, as expected (see Fig. 7b-g). In contrast, no signal for the 2A peptide was observed in cells transfected with the wt O1K RNA (Fig. 7b) or in untransfected cells (Fig. 7a). However, the presence of FMDV 2A (still attached to VP1) was detected in cells transfected with the VP1 K210E mutant RNA (Fig. 7c), consistent with previous results [7, 30]. Furthermore, using the transcripts with the mutant 2A/2B junctions, the presence of FMDV 2A, presumably attached to 2B (and maybe VP1), could be detected in the transfected cells (Fig. 7d-g). It should be noted that it is not possible to detect the free 2A peptide by immunoblotting due to its small size (ca. 2 kDa) and attempts to identify the presence of the 2A fused to other proteins in extracts from these RNA transfected cells were unsuccessful (c.f. detection of VP1-2A within cells infected with the VP1 K210E mutant virus [7, 30]), presumably because the 2A could be attached to a number of different proteins, e.g. within 2A-

2B, 2A-2BC, VP1-2A-2B and VP1-2A-2BC and not all cells take up and replicate the RNA transcripts.

Discussion

The 2A peptide plays a significant role in the FMDV life cycle as it is required for the co-translational cleavage of the growing polyprotein into two separate entities at the junction between 2A and 2B. Related 2A peptide sequences are found in a variety of other members of the picornavirus family; this suggests that they contribute significantly to the correct production and function of the viral proteins.

Using artificial polyprotein systems, it has been well documented [18, 22, 24, 26] that point mutations in the highly conserved $D^{12}(V/I)E(S/T)NPG_{2A}{}^{\downarrow}P^{19}{}_{2B}$ motif, located at the C-terminus of FMDV 2A, can either severely reduce or completely abrogate cleavage activity. In this study, we have extended these observations and investigated the effects of single amino acid substitutions in 2A on FMDV RNA replication, on virus viability and on polyprotein processing in its natural context within cells. The results presented here clearly demonstrate that certain 2A mutants previously found to greatly impair the StopGo activity in artificial polyproteins systems [22, 24] were still able to produce infectious viruses and thus the wt sequence and maximal cleavage activity is not essential for virus viability. It was anticipated that some mutations might have resulted in lethal phenotypes since earlier mutagenesis studies using FMDV and EMCV did not produce any viable progeny when the C-terminal 2A sequence was changed from $SNPG^{\downarrow}P_{2B}$ to $SNPL^{\downarrow}V_{2B}$ even after several passages [26]. Interestingly, we were able to rescue viruses from all of the RNA transcripts. When the apparent cleavage activity of the mutant 2A was low (<31% of wt activity) then it was found that reversions to the wild type sequence had occurred. This indicates that some RNA replication must have occurred (to allow the formation of wt revertants) despite the presence of a defective 2A peptide. In contrast, mutants with a higher level of cleavage activity (≥31% of wt) retained, in each case, the introduced amino acid substitutions in the rescued viruses. These results clearly indicate that efficient 2A mediated cleavage activity is advantageous for the virus but that optimal efficiency is not essential. This raises the question of why the separation of the capsid proteins from the non-structural proteins is so advantageous for some picornaviruses? It seems necessary for these viruses to have a 3C-independent mechanism to break the polyprotein. Some members of the picornavirus family (e.g. enteroviruses) possess a 2A protease to achieve the separation of the capsid protein precursor from the rest of the polyprotein and the StopGo

mechanism that occurs at the 2A/2B junction is clearly a distinct mechanism but one that is used by many members (e.g. aphthoviruses, cardioviruses, sapeloviruses, teschoviruses) of this virus family [8].

It has been speculated [13] that 2A can act as a translational regulator to modify the amount of the different parts of the polyprotein that are produced. In FMDV, the 2A peptide is located at the boundary between the upstream capsid proteins and the non-structural proteins involved in RNA replication. There could be two distinct functions for the 2A peptide. One primary function of 2A could be to achieve the cleavage of the polyprotein but it may also down-regulate downstream translation. Potentially, this could prove beneficial to the virus as the assembly of the FMDV capsid requires up to sixty copies of each of the four structural proteins whereas fewer copies of the proteins involved in the replication process are required. On the other hand, it could be considered that in the early stages of the virus infection, then it would seem advantageous to produce more of the proteins required for replication and processing than the capsid proteins. It is also noteworthy that most members of the picornavirus family that use a different mechanism for separation of the capsid proteins from the non-structural proteins do not apparently have any mechanism for modifying the ratio of proteins produced, thus the need for such a mechanism within the picornaviruses, in general, is not established. However, recently, Napthine et al.,[32] have demonstrated that in EMCV a programmed -1 ribosomal frameshift occurs within the 2B coding region, just downstream of the 2A coding region. This frameshift results in the production of a distinct protein, termed 2B*, and then termination of translation. The level of ribosomal frame shifting increases dramatically late in infection and thus the production of the non-structural proteins involved in virus replication is reduced at this time. The process requires the interaction of the EMCV 2A protein (ca. 16 kDa) with a stem-loop structure some 14 nt downstream of a “slip site” (GGUUUUU) within the 2B coding region. Although a U-rich motif (UUCUUUUUCU) is present just downstream of the coding region for the 2A/2B junction in the FMDV genome, certain other elements of this process appear to be absent. As indicated above, the FMDV 2A is only 18 residues long and it lacks the cluster of basic residues (R95-R97) that appear to be important for the interaction of the EMCV 2A protein to the stem-loop structure that is critical for the high frameshift efficiency. Thus, currently, there is no evidence for such a process within FMDV.

Assessment of the RNA replication efficiency, using a replicon system, demonstrated that alterations in the 2A peptide have a clear, negative, effect on either the replication of the viral RNA or on the translation of the polyprotein. Clearly, the processes of translation and replication are linked since when translation of the polyprotein is reduced, then the levels of

protein available to replicate the RNA are also reduced resulting in a lower level of RNA replication. As indicated above, it may be that the detrimental effect of the changes in 2A were accentuated by the absence of the VP1 coding sequence in the replicons. In the context of the full-length viral polyprotein, it was shown (Fig. 7) that blocking the cleavage at the FMDV 2A/2B junction produced fusion proteins containing 2A (presumably as 2A-2B or possibly VP1-2A-2B, before or after the cleavage of the VP1/2A junction by $3C^{pro}$). However, the addition of just 18 amino acids to the N-terminus of the 2B protein may be considered to be unlikely to cause this decrease in replication efficiency (indeed it has been shown that leaving the 2A peptide fused to the C-terminus of VP1 has no apparent effect on virus viability [7, 30]. It should be remembered, however, that the VP1/2A cleavage is the slowest of the 3C-mediated processing events within P1-2A [7, 14, 31]. Previously it has been found that cleavage at the VP1/2A junction in poliovirus appears to a have a role in processing of the capsid precursors since amino acid substitutions that prevented cleavage resulted in a P1 capsid precursor which was resistant to $3C^{pro}$ processing [10]. Furthermore, Hahn & Palmenberg demonstrated, using *in vitro* translation assays, that a mutation in the EMCV 2A impaired the processing of the L-P1-2A precursor by $3C^{pro}$ [29]. This may suggest a critical role for the 2A cleavage to allow proper folding of the (L)-P1-(2A) precursor to permit efficient cleavage by $3C^{pro}$. However, in our studies, no adverse effect of blocking cleavage at the 2A/2B junction on processing of P1-2A by $3C^{pro}$ was apparent (see Fig. 6). It was also observed with TMEV that normal capsid protein processing occurred in mutant viruses in which the 2A/2B processing was blocked [26]. The *Gluc* replicon, as used here, lacks the coding sequences for the structural proteins except for VP4, however, the replication / translation is still impaired in the 2A mutants compared to the wild type (Fig. 5). It is, therefore, conceivable that the possible cleavage restrictions that could govern the processing of the structural proteins also apply to the non-structural proteins. This may mean that correct processing of these proteins, which are required for RNA replication, is impaired, thereby resulting in lower RNA synthesis. Although the FMDV P1-2A processing by $3C^{pro}$ appears to be unaffected when the 2A peptide is mutated (Fig. 6), this does not rule out the possibility of a negative effect on the 2B-2C (or P3) processing. Surprisingly, there was relatively little difference in the growth characteristics between the viable 2A mutant viruses (E14Q, S15F, S15I and N16H) and the wt (Fig. 3a), which contrasts with the decrease in replication efficiency observed in the context of a FMDV replicon. This could suggest that the changes in the 2A peptide influence the initial rate of viral RNA replication but not the final virus yield.

Investigation of the effect of 2A mutations on the StopGo mechanism revealed that certain amino acid substitutions are severely detrimental for the proper function of the 2A whereas others only moderately impair the cleavage resulting in a mixture of products (some cleaved and others not, see Fig. 4). Previous studies have suggested that the 2A geometry is the determining factor for its function [19, 22]. The current hypothesis is that the N-terminal portion of 2A (in an α-helical conformation) interacts with the ribosomal exit tunnel to confer specific constraints required for the turn motif (ESNPG) to be in a position to influence events within the peptidyl transferase centre of the ribosome. Some amino acid substitutions could severely change the conformation of the 2A, thereby preventing the disruption of the peptide bond formation between the G and P residues, and hence result in an uncleaved polyprotein. The substitutions N16C and N16H were found to result in cleavage although with decreased efficiency (both cleaved and uncleaved products were observed, see Fig 4). The function of residue N16 within 2A has not yet been determined, however it has been suggested that the N16 forms a hydrogen bond with E14 to stabilise the right turn [22]. The substitutions S15A, S15I and S15F were found to result in essentially complete cleavage in contrast to earlier studies [22] that reported a reduction in the cleavage activity. Comparison of the 2A sequence from different picornavirus species has shown that a variety of amino acids are allowed at this position within the C-terminus of 2A suggesting that this particular amino acid is of low importance for the StopGo function. However, Sharma et al. [24] demonstrated that substitution of S15 by glycine (G) (in the FMDV sequence), which influenced the peptide secondary structure, impaired function more significantly than Ala or Pro substitutions, suggesting that increased backbone flexibility imposed by the Gly residue at this position was especially detrimental [24].

Interestingly, the substitutions P19A, P19G, P19V and P19S greatly reduced the level of cleavage but did not abolish it (see Fig. 4). This is in contrast to previous studies [22] which have reported that these amino acid substitutions resulted in no apparent cleavage activity in an artificial polyprotein system. A model for the mechanism of 2A mediated cleavage developed by Donnelly et al. [16] suggests that the P19 residue (at the N-terminus of 2B) is an absolute requirement for cleavage as a poor nucleophilic character in this position is an integral part of the proposed mechanism. However, our data clearly shows that ala (A), ser (S) and val (V) residues are also functional at this position albeit with reduced activity. Rychlík et al. [33] demonstrated that A, S, and V are, in fact, also poor nucleophiles in the context of ribosomal peptidyl transferase activity, however not to the same extent as P and G. This could explain

why these amino acids are able to support the cleavage activity to some degree although not at a level compatible with virus viability. Although this does not account for the reduced cleavage activity observed for the P19G mutants, suggesting that another, not yet identified, characteristic of residue 19 must apply.

The study by Gao et al. [23] found that the codon usage for the NPGP motif is conserved among the seven FMDV serotypes. Through the use of mRNAs encoding artificial polyproteins comprising two reporter proteins, assayed within CHO cells, the study investigated the role of synonymous codons for the G^{18}. It was concluded that the different synonymous codon usage for G^{18} did not influence the cleavage efficiency in that system. However, in separate studies, we have provided evidence that a clear codon bias operates to encode the NPG/P motif at the 2A/2B junction within FMDV-infected cells [34]. This raises the interesting possibility that the RNA sequence itself contributes to the cleavage event at the 2A/2B junction.

Materials and methods

Construction of plasmids containing full-length mutant FMDV cDNAs

The plasmid pT7S3 [27] contains the full-length cDNA for the O1Kaufbeuren B64 strain of FMDV. Modification of the coding sequence around the 2A/2B junction was achieved by a 2-step site-directed mutagenesis procedure, a variation of the QuickChange protocol (Stratagene), using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). The first round of PCR, using forward mutagenic 2A PCR primers (Table 2) with a single reverse primer 10PPN10 (Table 2) and the plasmid pT7S3 as template, generated an amplicon (ca. 450 bp) specifying particular amino acid substitutions within 2A. The primary PCR products were gel purified (GeneJet gel extraction kit, Thermo Fisher Scientific) and used as primers for a second round of PCR with plasmid pT7S3 as template. The *Dpn-I* resistant full-length products were selected in chemically competent *Escherichia coli* (*E.coli*) TOP10 cells (Thermo Fisher Scientific), amplified, then the plasmid DNA was purified (Midiprep kit; QIAGEN) and verified by sequencing of the 2A coding region with a BigDye Terminator v. 3.1 Cycle Sequencing kit and a 3500 Genetic Analyzer (Applied Biosystems).

The generation of plasmids with three synonymous mutations ca. 20 bp downstream of the modified 2A/2B junction was achieved essentially as described above. The first round of PCRs, used the forward mutagenic 2A_Synonymous_Fwd primer (Table 2) with a single reverse primer 10PPN10 (Table 2) and plasmid pT7S3 as template. The primary PCR products were gel purified (GeneJet gel extraction kit, Thermo Fisher Scientific) and used as primers for a

second round of PCR with modified versions of the pT7S3, with the codons for N^{16}, P^{17}, G^{18} or P^{19} changed to encode an alanine (A) residue in each case, as templates.

Construction of plasmids containing a FMDV replicon containing *Gaussia* Luciferase

The *Gaussia* luciferase (Gluc) FMDV replicon was constructed by replacement of the coding region for VP2, VP3, VP1 and 2A from pT7S3-NheI [35] with the coding region for Gluc fused to FMDV 2A (as used in [36]). The Gluc-2A sequence was amplified by PCR using primers 13APN1 and 13APN4 (see Table 1) using the rPad2GL BAC (see [34]) as template. The amplicon was inserted into the vector pCR-XL-TOPO (Invitrogen), the *NheI-ApaI* fragment was excised and inserted between the same sites within the ca. 5kb *XbaI*-fragment from pT7S3-*NheI* (essentially as described previously [35]). The modified *XbaI* fragment (now containing the Gluc-2A sequence) was reconstructed into the backbone of the O1K FMDV cDNA within the *XbaI*-digested pT7S3 [27] and the orientation established by restriction digestion (using *EcoRI* and *HpaI*). The Gluc FMDV replicon was termed wt-Gluc.

The replication-defective Gluc FMDV replicon termed wt-GlucΔ3D was prepared by digesting the wt-Gluc plasmid with *BamHI* and *HpaI* to liberate a fragment of ca. 770 bp corresponding to the 3'-terminus of the FMDV genome (including part of the 3Dpol coding region, see Fig. 1). The large residual fragment was gel purified, blunt ended, self-ligated and transformed into *E. coli*. The wt-GlucΔ3D plasmid DNA was purified (Midiprep kit; QIAGEN) and verified by sequencng of the 3Dpol coding region, as above.

Construction of plasmids containing FMDV P1-2A-2BC-FLAG cDNA cassettes

The FMDV cDNA cassette, in the plasmid pP1-2A-2BC-FLAG, was prepared by PCR using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). Briefly, the coding region for P1-2A-2BC from O1K FMDV cDNA (as in pT7S3, [27]) was amplified with a forward primer ATG_P1_fwd which incorporates an initiation codon and the reverse primer 2C_FLAG_Stop_rev that includes the sequence for a FLAG epitope tag followed by a termination codon (see Table 2). The blunt-end amplicon (ca. 3670 bp) was ligated into the pJET1.2 vector (Thermo Fisher Scientific) according to the manufacturer's instructions. Sequencing revealed an unwanted initiation codon between the T7 promoter and the insert, which was then removed. A 2-step site-directed mutagenesis PCR using Phusion High-Fidelity DNA Polymerase as previously described with mutagenic PCR 2A primers (Table 2) and 10PPN10 was used to produce the following plasmids encoding the indicated single amino acid

substitutions within 2A: pP1-2C-FLAG D12A, V13A, E14A, E14Q, S15A, S15F, S15I, N16A, N16C, N16W, P17A, G18A, P19A, P19G, P19V and P19S. All plasmids were propagated in *E. coli* TOP10 cells (Thermo Fisher Scientific), purified (Midiprep kit; QIAGEN), and verified by sequencing.

In vitro transcription

Briefly, 5 µg of replicon plasmid or full-length FMDV plasmid were linearized by digestion with *HpaI* (Thermo Fisher Scientific), purified (GeneJET PCR Purification Kit, Thermo Fisher Scientific) and eluted in RNase-free water. Both replicon and full-length FMDV RNA transcripts were prepared using the Megascript T7 kit (Ambion). Reaction mixtures were incubated at 37°C for 4 h and treated with 2 units of TURBO DNase for 30 min after which the RNA was purified using the MEGAclear Transcription Clean-Up Kit according to the manufacturer's instructions. RNA integrity was assessed by electrophoresis using an ethidium bromide-stained agarose gel (1%), in TBE buffer, and quantified by spectrophotometry (NanoDrop 1000, Thermo Fisher Scientific).

Rescue of virus from full-length cDNA plasmids

For rescue and passage of infectious FMDV, 5 µg full-length FMDV RNA was introduced into BHK cells by electroporation (as described previously [37]). The cells were then transferred to one well of a 6-well plate and incubated for 1-3 days at 37°C after which the viruses were harvested by freezing. The rescued viruses were then amplified using additional passages (P2 and P3) using fresh BHK cells. After the third passage (P3), viral RNA was extracted (RNeasy Mini Kit, Qiagen) and converted to cDNA using ready-to-go you-prime first-strand beads (GE Healthcare Life Sciences) with random hexamer primers. Amplicons (ca. 660 bp), including the 2A coding region, were amplified by PCR (AmpliTaq Gold DNA polymerase, Thermo Fischer Scientific) using the primers 8APN206 and 8APN203 (see Table 2). Control reactions, without RT, were used to ensure that the analysed products were derived from RNA and not from the presence of carryover plasmid DNA template. The amplicons were visualized in 1% agarose gels, purified (GeneJET gel extraction kit, Thermo Fischer Scientific) and sequenced as above. Sequences were analysed using Geneious 7.2 (Biomatters, Auckland, New Zealand).

Gaussia luciferase assay

BHK cells suspended in cold PBS were transferred to a 4 mm cuvette after which 2 µg replicon RNA was added, briefly mixed, and the cells were electroporated (25 ms and 240 V; one pulse)

on a Gene Pulser X-Cell (Bio-Rad). Following incubation for 10 min at room temperature, the cells were transferred to 5 wells of a 24-well plate (140 µl per well with 500 µl DMEM containing 5% FCS). Following incubation, at 37°C for the required time, the medium was removed and the BHK cells were lysed by adding 100 µl of *Renilla* luciferase assay lysis buffer (Promega) to the cells in each well (24-plate well) and incubated at room temperature for 15 min. The luciferase activity was quantified in a Luminometer (Titertek-Berthold) by addition of this lysate (20 µl) to Renilla Luciferase Assay reagent (100 µl) according to the manufacturer's instructions.

Virus growth kinetics

Virus titres for the wt and the 2A mutant viruses: E14Q, S15I, S15F and N16H were determined in BHK cells as $TCID_{50}$/ml, as described previously [38]. Monolayers of BHK cells, grown in 96-well plates were infected with either wt or mutant FMDV at an m.o.i of 0.1 at 37°C. At 0, 2, 5, 10 and 24 hours post infection the infected cells were harvested by freezing (at -80°C) to determine the virus yield as $TCID_{50}$/ml.

Transient expression assays

BHK cells (in 35mm wells) were grown to 90 % confluency and infected with vTF7-3, a recombinant vaccinia virus that expresses the T7 RNA polymerase [28], as described previously [38]. Briefly, following the infection, plasmid DNA (pP1-2A-2BC-FLAG and its derivatives, 2µg) was transfected alone or, when indicated, with pSKRH3C (50 ng) [38] that expresses FMDV $3C^{pro}$, using FuGene6 (Roche), into the infected BHK cells and incubated overnight at 37°C.

Western blotting

Cell lysates for immunoblotting were prepared by addition of cold Buffer C (0.125 M NaCl, 20 mM Tris/HCl (pH 8.0), 0.5% NP-40) to the cells. After incubation (on ice, for at least 5 mins), the cell extracts were clarified by centrifugation (20000 g for 10 min) and Laemmli sample buffer (with 25 mM DTT) was added (as described previously [39]) . Following heating to 98°C for 5 min, samples were resolved by SDS-PAGE (4-15% polyacrylamide) and transferred to a PVDF membrane (Bio-Rad) and blocked for 1 h in 5% PBS-Tween (PBS, 0.1% Tween) with 5% non-fat milk. The membranes were incubated overnight at 4°C with either goat anti-FLAG antibodies (Abcam), guinea pig anti-FMDV O1 Manisa serum (to detect FMDV capsid proteins) or mouse anti-FMDV VP2 (4B2) monoclonal antibody [40], as used

previously [7]. The membranes were washed 3x with PBS-Tween and incubated for 3 hours at room temperature with either HRP-conjugated anti-goat IgG (Dako), HRP-conjugated anti-guinea pig IgG (Dako) or HRP-conjugated anti-mouse IgG (Dako), respectively. The membranes were then washed 3x with PBS-Tween and bound proteins were detected using a chemiluminescence detection kit (ECL Prime, Amersham) with a Chemi-Doc XRS system (Bio-Rad). The intensities of the signals for the FLAG-tagged polyproteins were, when necessary, quantitated using ImageJ software (v1.50).

Immunofluorescence assay

Monolayers of BHK cells were grown on glass coverslips in 6 well plates and immediately prior to transfection, cells were washed briefly in PBS and the medium replaced with DMEM without serum. FMDV RNA transcripts were introduced into BHK cells using Lipofectin transfection reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. After 8 h, the cells were fixed, stained and mounted as described previously [31] using rabbit anti-FMDV O serum or rabbit anti-2A (ABS31, Merck) followed by a donkey Alexa Fluor 568-labelled anti-rabbit IgG (A10042, Life Technologies). The slides were washed in PBS after which they were mounted with Vectashield (Vector Laboratories) containing DAPI and images were captured using an epifluorescence microscope.

Acknowledgements

We wish to thank Li Yu (Chinese Academy of Agricultural Sciences, China) for providing us with the anti-FMDV VP2 antibody. We would also like to acknowledge the excellent technical assistance of Preben Normann and helpful advice from Thea Kristensen.

References

1. **Jamal SM, Belsham GJ**. 2013. Foot-and-mouth disease: past, present and future. Vet Res 44:116–129.

2. **Martinez-Salas E, Belsham GJ**. 2017. Genome Organisation, Translation and Replication of Foot-and-mouth Disease Virus RNA, p. 13–42. *In* Sobrino, F, Domingo, E (eds.), Foot-and-Mouth Disease: Current Research and Emerging Trends. Caister Academic Press.

3. **Belsham GJ**. 2005. Translation and replication of FMDV RNA. Curr Top Microbiol Immunol 288:43–70.

4. **Medina M, Domingo E, Brangwyn JK, Belsham GJ**. 1993. The two species of the foot-and-mouth disease virus leader protein, expressed individually, exhibit the same activities. Virology 194:355–359.

5. **Strebel K, Beck E**. 1986. A second protease of foot-and-mouth disease virus. J Virol 58:893–899.

6. **Curry S, Fry E, Blakemore W, Abu-Ghazaleh R, Jackson T, King A, Lea S, Newman J, Stuart D**. 1997. Dissecting the roles of VP0 cleavage and RNA packaging in picornavirus capsid stabilization: the structure of empty capsids of foot-and-mouth disease virus. J Virol 71:9743–9752.

7. **Gullberg M, Polacek C, Bøtner A, Belsham GJ**. 2013. Processing of the VP1/2A junction is not necessary for production of foot-and-mouth disease virus empty capsids and infectious viruses: characterization of “self-tagged” particles. J Virol 87:11591–11603.

8. **Luke GA, de Felipe P, Lukashev A, Kallioinen SE, Bruno EA, Ryan MD**. 2008. Occurrence, function and evolutionary origins of “2A-like” sequences in virus genomes. J Gen Virol 89:1036–1042.

9. **Sommergruber W, Zorn M, Blaas D, Fessl F, Volkmann P, Maurer-Fogy I, Pallai P, Merluzzi V, Matteo M, Skern T, Kuechler E**. 1989. Polypeptide 2A of human rhinovirus type 2: Identification as a protease and characterization by mutational analysis. Virology 169:68–77.

10. **Toyoda H, Nicklin MJH, Murray MG, Anderson CW, Dunn JJ, Studier FW, Wimmer E**. 1986. A second virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein. Cell 45:761–770.

11. **Belsham GJ**. 1993. Distinctive features of foot-and-mouth disease virus, a member of the picornavirus family; aspects of virus protein synthesis, protein processing and structure. Prog Biophys Mol Biol 60:241–260.

12. **Palmenberg AC, Parks GD, Hall D, Ingraham RH, Seng TW, Pallal P V.** 1992. Proteolytic Processing of the Cardioviral Cleavage in Clone-Derived P2 Region : Primary 2A / 2B Precursors. Virology 190:754–762.

13. **Tulloch F, Luke GA, Ryan MD**. 2017. Foot-and-mouth disease virus proteinases and polyprotein processing, p. 43–59. *In* Sobrino, F, Domingo, E (eds.), Foot-and-Mouth Disease: Current Research and Emerging Trends. Caister Academic Press.

14. **Ryan MD, Belsham GJ, King AMQ**. 1989. Specificity of enzyme-substrate interactions in foot-and-mouth disease virus polyprotein processing. Virology 173:35–45.

15. **Ryan MD, King AMQ, Thomas GP**. 1991. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. J Gen Virol 72:2727–2732.

16. **Donnelly M, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, Ryan MD**. 2001. Analysis of the aphthovirus 2A/2B polyprotein “cleavage” mechanism indicates not a proteolytic reaction, but a novel translational effect: A putative ribosomal “skip.” J Gen Virol 82:1013–1025.

17. **Atkins JF, Wills NM, Loughran G, Wu C-Y, Parsawar K, Ryan MD, Wang C-H, Nelson CC**. 2007. A case for “StopGo”: reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go). RNA 13:803–810.

18. **Doronina VA, Wu C, de Felipe P, Sachs MS, Ryan MD, Brown JD**. 2008. Site-specific release of nascent chains from ribosomes at a sense codon. Mol Cell Biol 28:4227–4239.

19. **Ryan MD, Donnelly M, Lewis A, Mehrotra AP, Wilkie J, Gani D**. 1999. A Model

for Nonstoichiometric, Cotranslational Protein Scission in Eukaryotic Ribosomes. Bioorg Chem 27:55–79.

20. **Groppo R, Palmenberg AC**. 2007. Cardiovirus 2A protein associates with 40S but not 80S ribosome subunits during infection. J Virol 81:13067–13074.

21. **Donnelly M, Gani D, Flint M, Monaghan S, Ryan MD**. 1997. The cleavage activities of aphthovirus and cardiovirus 2A proteins. J Gen Virol 78:13–21.

22. **Donnelly MLL, Luke GA, Hughes LE, Luke G, Mendoza H, Dam E, Gani D, Ryan MD**. 2001. The "cleavage" activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring "2A-like" sequences. J Gen Virol 82:1027–1041.

23. **Gao Z, Zhou J, Zhang J, Ding Y, Liu Y**. 2014. The silent point mutations at the cleavage site of 2A/2B have no effect on the self-cleavage activity of 2A of foot-and-mouth disease virus. Infect Genet Evol 28:101–106.

24. **Sharma P, Yan F, Doronina VA, Escuin-Ordinas H, Ryan MD, Brown JD**. 2012. 2A peptides provide distinct solutions to driving stop-carry on translational recoding. Nucleic Acids Res 40:3143–3151.

25. **Minskaia E, Nicholson J, Ryan MD**. 2013. Optimisation of the foot-and-mouth disease virus 2A co-expression system for biomedical applications. BMC Biotechnol 13:67.

26. **Loughran G, Libbey JE, Uddowla S, Scallan MF, Ryan MD, Fujinami RS, Rieder E, Atkins JF**. 2013. Theiler's murine encephalomyelitis virus contrasts with encephalomyocarditis and foot-and-mouth disease viruses in its functional utilization of the StopGo non-standard translation mechanism. J Gen Virol 94:348–353.

27. **Ellard FM, Drew J, Blakemore WE, Stuart DI, King AMQ**. 1999. Evidence for the role of His 142 of protein 1C in the acid induced disassembly of foot and mouth disease virus capsids. J Gen Virol 80:1911–1918.

28. **Fuerst TR, Niles EG, Studier FW, Moss B**. 1986. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. Proc Natl Acad Sci U S A 83:8122–8126.

29. **Hahn H, Palmenberg AC**. 1996. Mutational analysis of the encephalomyocarditis virus primary cleavage. J Virol 70:6870–6875.

30. **Napthine S, Ling R, Finch LK, Jones JD, Bell S, Brierley I, Firth AE**. 2017. Protein-directed ribosomal frameshifting temporally regulates gene expression. Nat Commun 8:15582.

31. **Kristensen T, Normann P, Gullberg M, Fahnøe U, Polacek C, Rasmussen TB, Belsham GJ**. 2017. Determinants of the VP1/2A junction cleavage by the 3C protease in foot-and-mouth disease virus-infected cells. J Gen Virol 98:385–395.

32. **Gullberg M, Muszynski B, Organtini LJ, Ashley RE, Hafenstein SL, Belsham GJ, Polacek C**. 2013. Assembly and characterization of foot-and-mouth disease virus empty capsid particles expressed within mammalian cells. J Gen Virol 94:1769–1779.

33. **Rychlík I, Černá J, Chládek S, Pulkrábek P, Žemlička J**. 1970. Substrate Specificity of Ribosomal Peptidyl Transferase The Effect of the Nature of the Amino Acid Side Chain. Eur J Biochem 16:136–142.

34. **Kjær J, Belsham GJ**. 2017. Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias. RNA (in press). e-pub October 2017 doi:10.1261/rna.063339.117

35. **Bøtner A, Kakker NK, Barbezange C, Berryman S, Jackson T, Belsham GJ**. 2011. Capsid proteins from field strains of foot-and- mouth disease virus confer a pathogenic phenotype in cattle on an attenuated, cell-culture-adapted virus. J Gen Virol 92:1141–1151.

36. **Risager PC, Fahnøe U, Gullberg M, Rasmussen TB, Belsham GJ**. 2013. Analysis of classical swine fever virus RNA replication determinants using replicons. J Gen Virol 94:1739–1748.

37. **Nayak A, Goodfellow IG, Woolaway KE, Birtley J, Curry S, Belsham GJ**. 2006. Role of RNA structure and RNA binding activity of foot-and-mouth disease virus 3C protein in VPg uridylylation and virus replication. J Virol 80:9865–9875.

38. **Reed LJ, Muench H**. 1938. A simple method of estimating fifty percent endpoints. Am J Hyg 27:493–497.

39. **Belsham GJ, Nielsen I, Normann P, Royall E, Roberts LO**. 2008. Monocistronic mRNAs containing defective hepatitis C virus-like picornavirus internal ribosome entry site elements in their 5’ untranslated regions are efficiently translated in cells by a cap-

dependent mechanism. RNA 14:1671–1680.

40. **Belsham GJ, McInerney GM, 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 74:272–280.

41. **Polacek C, Gullberg M, Li J, Belsham GJ**. 2013. Low levels of foot-and-mouth disease virus 3C protease expression are required to achieve optimal capsid protein expression and processing in mammalian cells. J Gen Virol 94:1249–58.

42. **Yu Y, Wang H, Zhao L, Zhang C, Jiang Z, Yu L**. 2011. Fine mapping of a foot-and-mouth disease virus epitope recognized by serotype independent monoclonal antibody 4B2. J Microbiol 49:94–101.

Tables

Table 1: Amino acid sequences in the encoded 2A peptide within rescued viruses following three passages in BHK cells. Sequence differences from wt are shown whereas identical amino acids are indicated by a (**.**). The wt and mutant 2A cleavage activities previously determined by Donnelly et al. [22] and Sharma et al.[24], *in vitro,* are indicated.

Plasmid	"Cleavage" *in vitro* (%)[1]	Residue encoded in rescued viruses					
				2A			2B
wt	**90**	**E14**	**S15**	**N16**	**P17**	**G18**	**P19**
E14Q	**56**	**Q**	**.**	**.**	**.**	**.**	**.**
S15F	**39**	**.**	**F**	**.**	**.**	**.**	**.**
S15I	**42**	**.**	**I**	**.**	**.**	**.**	**.**
N16A	**0**	**.**	**.**	**.**	**.**	**.**	**.**
N16H	**31**	**.**	**.**	**H**	**.**	**.**	**.**
P17A	**0**	**.**	**.**	**.**	**.**	**.**	**.**
G18A	**0**	**.**	**.**	**.**	**.**	**.**	**.**
P19A	**0**	**.**	**.**	**.**	**.**	**.**	**.**
P19G	**11**	**.**	**.**	**.**	**.**	**.**	**.**

1: Data from Donnelly et al. and Sharma et al. [22, 24].

Table 2: PCR primers used to create and sequence mutant FMDV cDNAs. Mutagenic nucleotides are shown underlined.

Primer name	Sequence (5′-3′)
Fwd_2A_D12A	AAGTTGGCGGGAGCCGTCGAGTCCAACCCTGG
Fwd_2A_V13A	AAGTTGGCGGGAGACGCCGAGTCCAACCCTGG
Fwd_2A_E14A	AAGTTGGCGGGAGACGTCGCGTCCAACCCTG
Fwd_2A_E14Q	GATGTCCAGTCCAACCCTGG
Fwd_2A_S15A	TTGGCGGGAGACGTCGAGGCCAACCCTG
Fwd_2A_S15F	GATGTCGAGTTTAACCCTGC
Fwd_2A_S15I	GATGTCGAGATTAACCCTGG
Fwd_2A_N16A	GTC CGC CCC TGG GCC CTT C
Fwd_2A_N16C	CGAGTCCTGCCCTGGGCCCTTCTTTTTCTCCGA
Fwd_2A_N16H	CGAGTCCCACCCTGGGCCCTTCTTTTTCTCCGA
Fwd_2A_N16V	CGAGTCCGTCCCTGGGCCCTTCTTTTTCTCCGA
Fwd_2A_N16W	CGAGTCCTGGCCTGGGCCCTTCTTTTTCTCCGA
Fwd_2A_P17A	GTCCAACGCTGGGCCCTTC
Fwd_2A_G18A	GTCCAACCCTGCGCCCTT C
Fwd_2A_P19A	CAACCCTGGGGCTTTCTT
Fwd_2A_P19G	CAACCCTGGGGGCTTCTT
Fwd_2A_P19S	CGAGTCCAACCCTGGGTCCTTCTTTTTCTCCGA
Fwd_2A_P19V	CGAGTCCAACCCTGGGGTCTTCTTTTTCTCCGA
8APN203	CTCCTTCAACTACGGTGCC
8APN206	CACCCGAAGACCTTGAGAG
10PPN10	CTTTGACCAACCCGGCCA
13APN1	CCGGGCCCAGGGTTGGACTCGAC
13APN4	CCGGATCCGCTAGCCATGGGAGTCAAAGTTCTGTTTGC
ATG_P1_fwd	ATGAATACTGGCAGCATAATAAACAACTAC
2C_FLAG_Stop_rev	CTATTACTTGTCGTCATCGTCTTTGTAGTCCTGCTTGAAGATC GGGTGACTCGACAC
2A_Synonomous_Fwd	TCTCCGACGTAAGATCAAACTTCTCCA

Figures

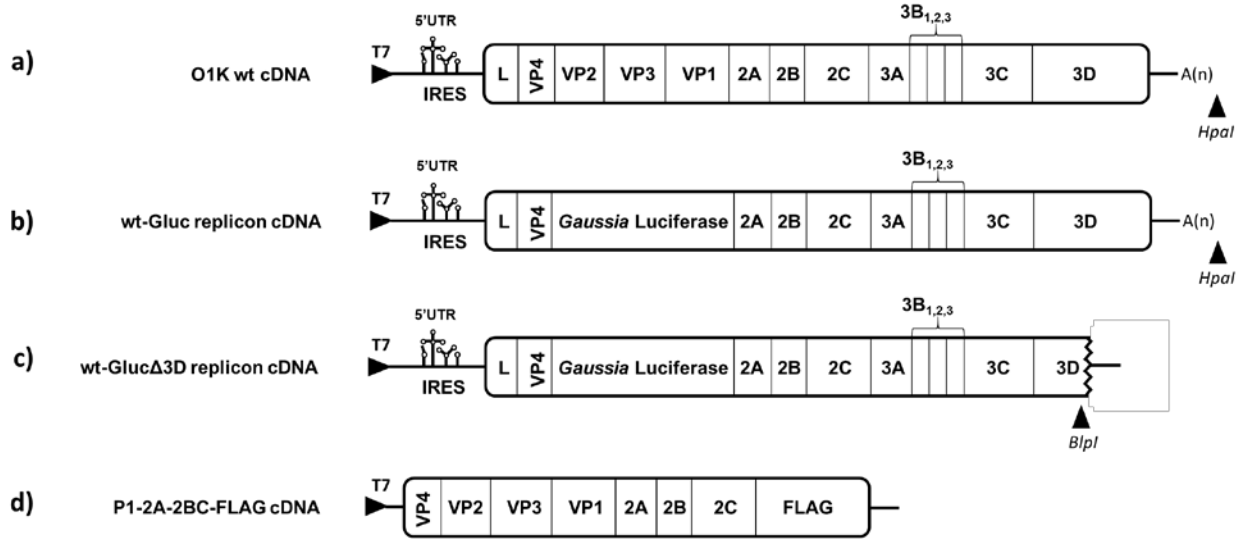

Fig. 1: Structure of the plasmids used in this study. These include: (a) full-length FMDV O1K cDNA, (b) *Gluc* replicon cDNA and (c) RNA polymerase defective *Gluc* replicon cDNA. The plasmids were linearized using *HpaI* or *BlpI* prior to *in vitro* transcription. Panel (d): Schematic representation of the P1-2A-2BC-FLAG cDNA cassette expressed in transient expression assays (as described in Material and Methods).

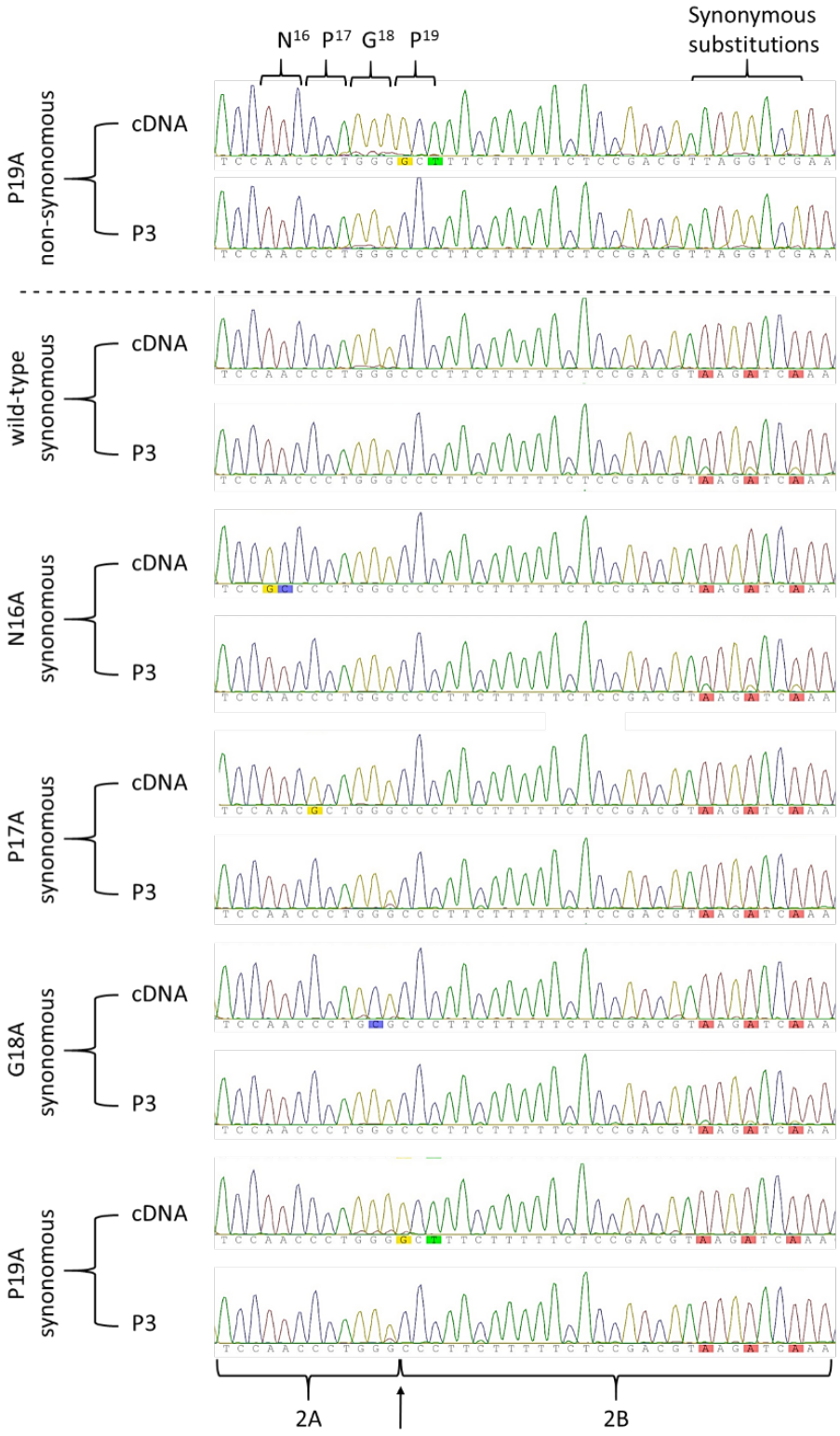

Fig. 2: FMDVs rescued from mutants N16A, P17A, G18A and P19A had reverted to the wt sequence. Three synonymous mutations downstream of the 2A/2B junction were introduced into the wt and mutant N16A, P17A, G18A and P19A plasmids. The resultant RNA transcripts were introduced into BHK cells. The rescued viruses were analysed after 3 passages

in BHK cells. The region of the FMDV genome including that encoding the 2A peptide was amplified by RT-PCR and the PCR products were sequenced. The chromatograms are shown, note the retained synonymous mutations ca. 20 nt downstream of the 2A/2B junction, that had been introduced as a marker.

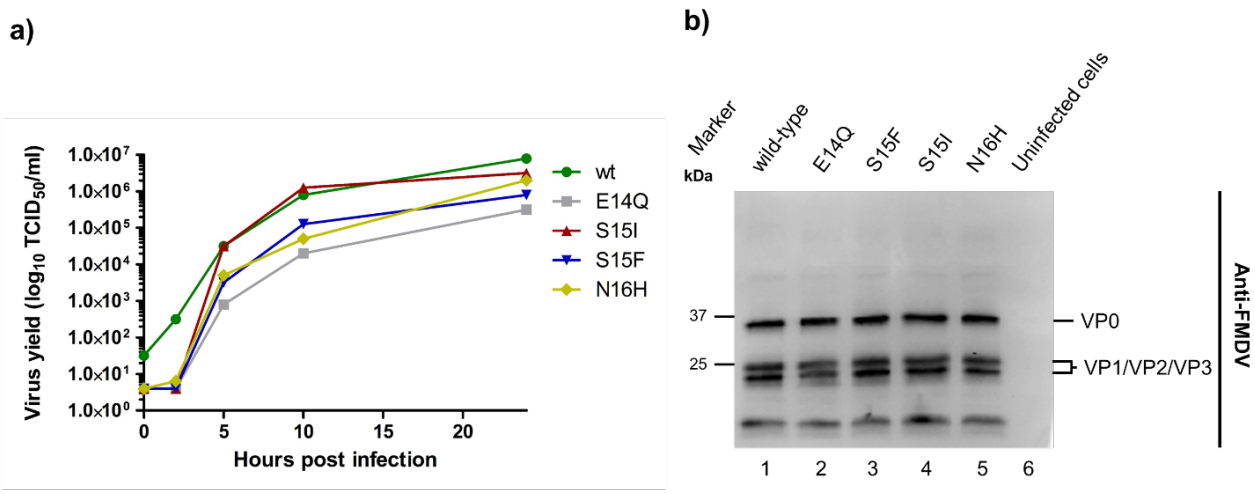

Fig. 3: Growth curves and assessment of the production of FMDV capsid proteins in BHK cells infected with wt and viable 2A mutant viruses. (a) BHK cells were infected with wt and the indicated 2A mutants at an m.o.i. of 0.1 and virus was harvested by freezing at 0, 2, 5, 10 and 24 hours post-infection. Virus yields were determined as $TCID_{50}$ by titration in BHK cells. **(b)** Uninfected or FMDV-infected BHK (m.o.i. 0.1) cell lysates were analysed by SDS-PAGE and immunoblotting with antibodies specific for FMDV capsid proteins (anti-FMDV sera). Uninfected cells were used as a negative control. Molecular mass markers (kDa) are indicated on the left.

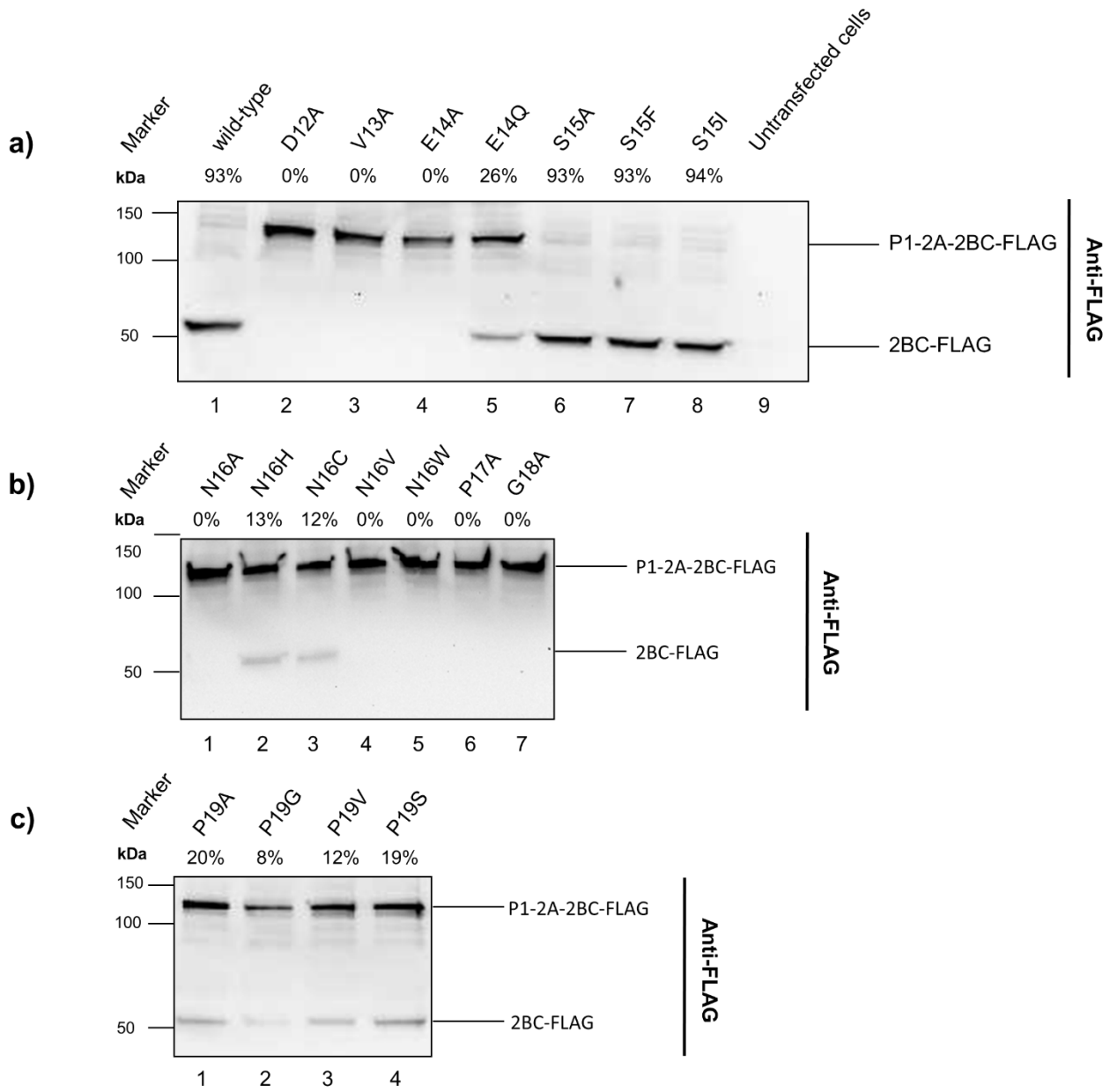

Fig. 4: Transient expression assays to determine 2A/2B "cleavage" induced by the wt and mutant FMDV cDNAs. The indicated plasmids were transfected into vTF7-3 infected BHK cells as described in Materials and Methods. After 24 hours, cell extracts were prepared and analysed by SDS-PAGE and immunoblotting using an anti-FLAG antibody. The uncleaved P1-2A-2BC-FLAG and the cleavage product (2BC-FLAG) are marked. Molecular mass markers (kDa) are indicated on the left. The cleavage activities (percentage of cleaved product) of the wt and each 2A mutant were determined by quantifying the intensity of the signal for the FMDV capsid proteins using ImageJ (v1.50) and are indicated above each lane.

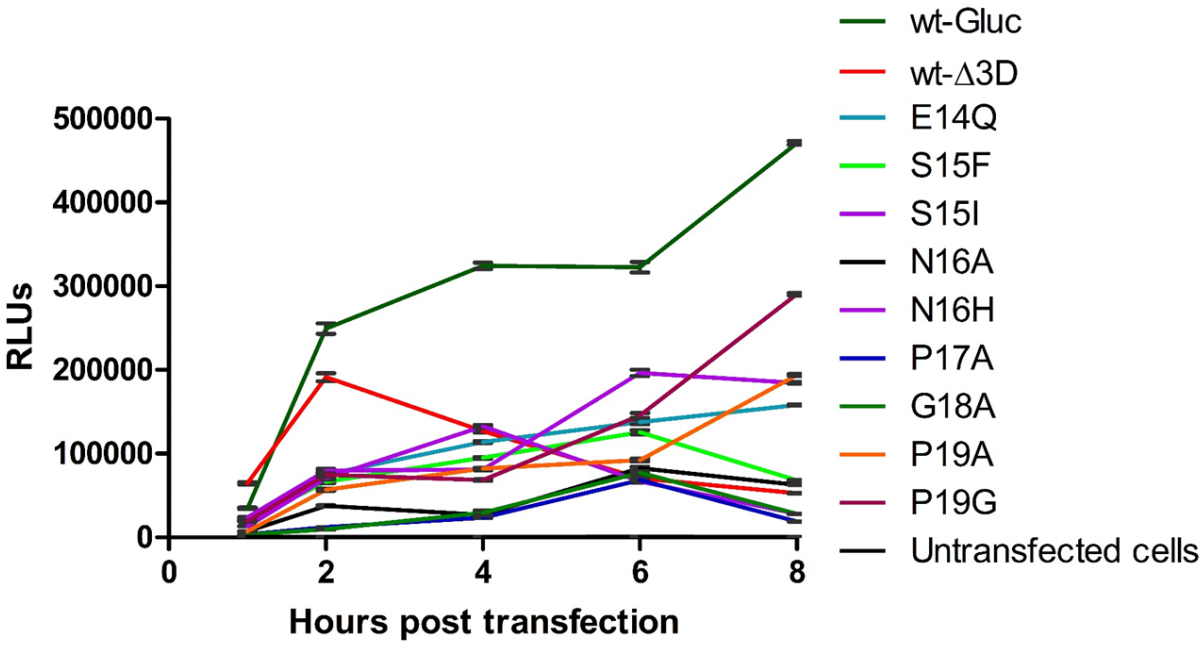

Fig. 5: Expression of the luciferase reporter protein, Gluc, by a FMDV replicon. BHK cells were electroporated with wt or mutant RNA transcripts derived from the indicated cDNAs and, at the indicated times, cell lysates were prepared and assayed for Gluc activity. RLUs = Relative light units. Data are presented as mean ± standard deviation (SD) RLUs from samples (n=3) harvested at the indicated times.

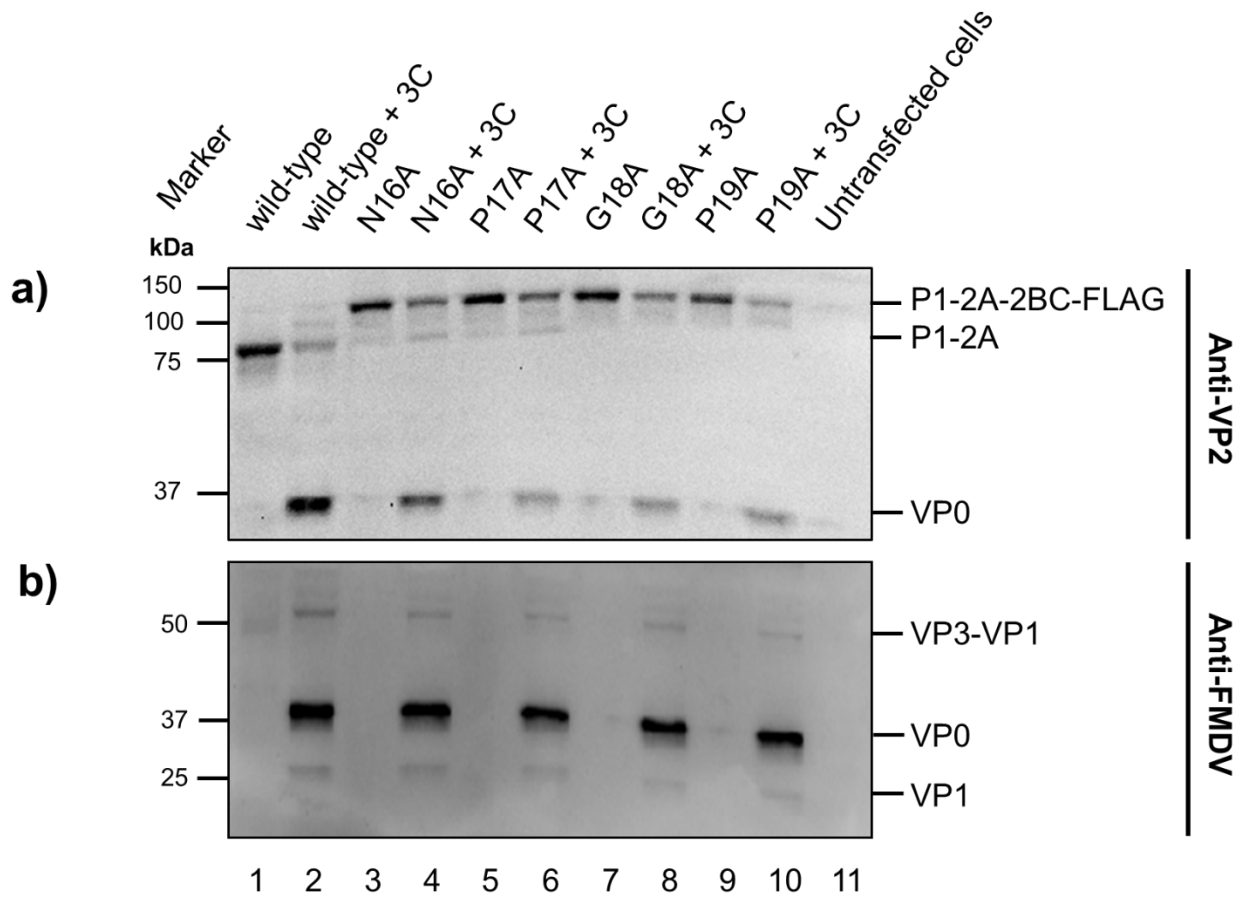

Fig. 6: Transient expression assays to determine the influence of 2A substitutions on the processing of the FMDV capsid precursor P1-2A. The wt and mutant P1-2A-2BC-FLAG plasmids were transfected alone or with pSKRH3C [40] (which expresses FMDV $3C^{pro}$), as indicated, into vTF7-3 infected BHK cells as described in Methods. After 24 h, cell extracts were prepared and analysed by SDS-PAGE and immunoblotting using anti-FMDV VP2 (panel a) and anti-FMDV antisera (to detect all FMDV capsid proteins) (panel b) as indicated. Molecular mass markers (kDa) are indicated on the left.

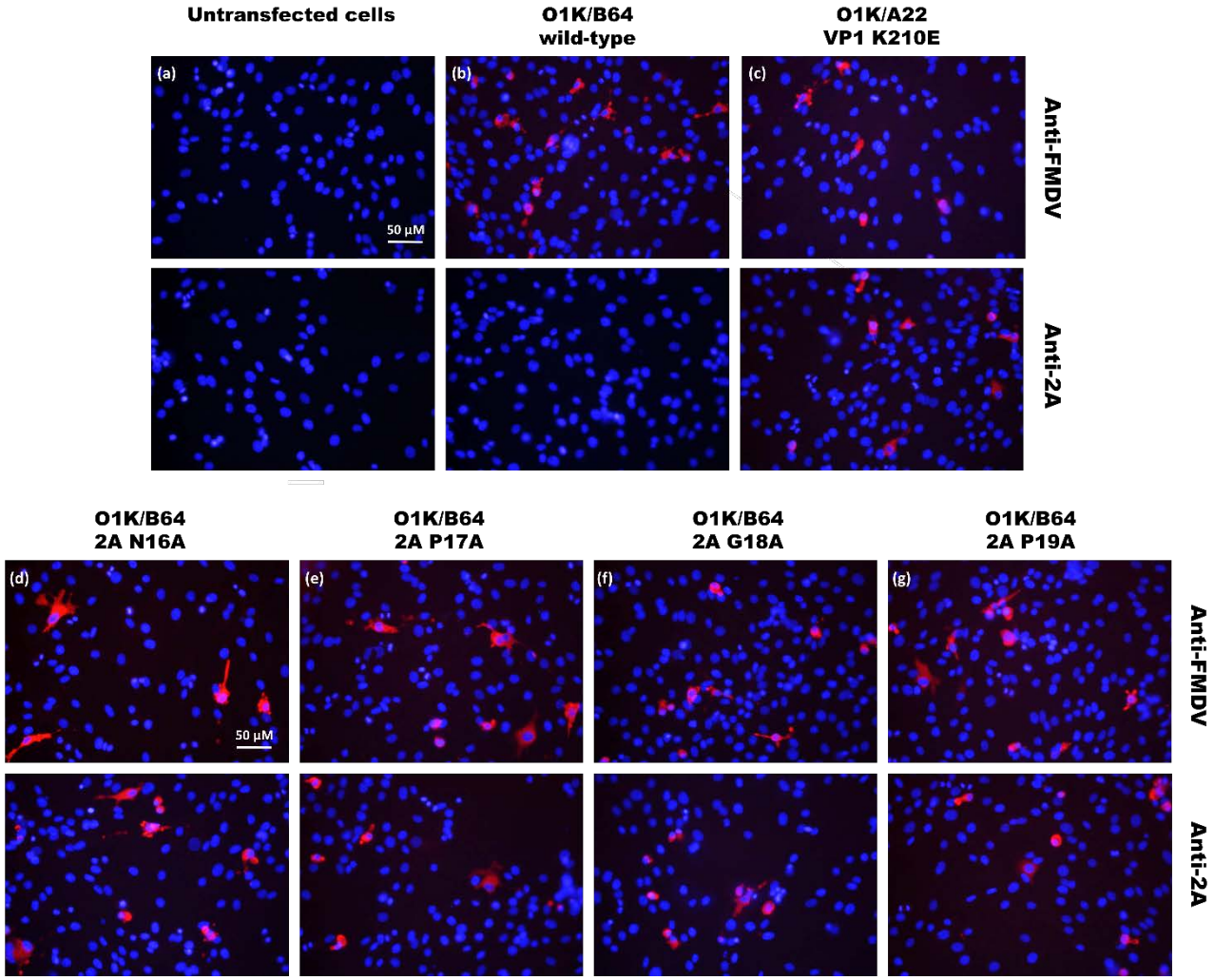

Fig. 7: Detection of FMDV 2A fusion proteins by IF staining within cells. BHK cells were untreated or transfected with wt or mutant FMDV RNA transcripts. At 8 h post-transfection, the cells were fixed. FMDV capsid proteins or the FMDV 2A peptide were detected using anti-FMDV O1K polyclonal antibodies (upper panels) or anti-2A antibodies (lower panels), respectively, plus a secondary antibody labelled with Alexa Fluor 568 (red). The 2A substitutions are indicated. Untransfected cells were used as a negative control whereas the O1K VP1 K210E mutant, described previously [7], in which the 2A remains joined to VP1, served as a positive control. The cellular nuclei were visualized with DAPI (blue). Bar, 50 µm.

Manuscript 2

Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias

Jonas Kjær & Graham J. Belsham

Published in RNA

2017

http://dx.doi.org/10.1261/rna.063339.117

Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias

Jonas Kjær[1] and Graham J. Belsham[1*].

1: DTU National Veterinary Institute, Technical University of Denmark, Lindholm, DK-4771, Kalvehave, Denmark

Short title: Codon bias for NPGP motif in FMDV 2A

Keywords: Picornavirus; Synonymous codon; Codon bias; translation

Corresponding author: Graham J. Belsham, grbe@vet.dtu.dk, phone: +45 3588 7985

Figures: 1

Tables: 3

Abstract

Foot-and-mouth disease virus (FMDV) has a positive-sense ssRNA genome including a single, large, open reading frame. Splitting of the encoded polyprotein at the 2A/2B junction is mediated by the 2A peptide (18 residues long) which induces a non-proteolytic, co-translational, “cleavage” at its own C-terminus. A conserved feature among variants of 2A is the C-terminal motif $N^{16}P^{17}G^{18}/P^{19}$ where P^{19} is the first residue of 2B. It has been shown previously that certain amino acid substitutions can be tolerated at residues E^{14}, S^{15} and N^{16} within the 2A sequence of infectious FMDVs but no variants at residues P^{17}, G^{18} or P^{19} have been identified. In this study, using highly degenerate primers, we analysed if any other residues can be present at each position of the NPG/P motif within infectious FMDV. No alternative forms of this motif were found to be encoded by rescued FMDVs after 2, 3 or 4 passages. However, surprisingly, a clear codon preference for the wt nucleotide sequence encoding the NPGP motif within these viruses was observed. Indeed, the codons selected to code for P^{17} and P^{19} within this motif were distinct; thus the synonymous codons are not equivalent.

Introduction

Foot-and-mouth disease virus (FMDV) is the prototypic member of the genus *Aphthovirus* within the family *Picornaviridae*. This virus is the causative agent of the highly contagious and economically important disease of cloven-hoofed animals, foot-and-mouth disease. The positive-sense ssRNA genome of around 8400 nt includes a single, large, open reading frame (ORF), ca. 7000 nt, encoding a polyprotein (Belsham, 2005). The full-length viral polyprotein is never observed since it is rapidly processed during and after synthesis mainly by the virus-encoded proteases (primarily $3C^{pro}$) to produce 15 distinct mature proteins plus multiple precursors (reviewed in Martinez-Salas and Belsham 2017). Interestingly, FMDV, like many (but by no means all) other picornavirus (e.g. cardioviruses, erboviruses, teschoviruses etc.) employs a co-translational, protease-independent mechanism for the "cleavage" of the polyprotein at the 2A/2B junction (the boundary between the capsid proteins and the non-structural proteins) (Donnelly *et al.*, 2001a). This mechanism has been referred-to as "ribosomal skipping" or, alternatively, "stop-carry on" or "StopGo" (Atkins *et al.*, 2007; Donnelly *et al.*, 2001a; Doronina *et al.*, 2008; Tulloch *et al.*, 2017). The 2A peptide lacks characteristic protease motifs and only mediates "cleavage" during translation. It has been demonstrated that the 2A sequence is able to mediate "cleavage" in all eukaryotic translation systems tested whereas a number of artificial polyproteins containing this sequence have been examined in prokaryotic systems and no detectable cleavage products were observed (Donnelly et al., 1997).

The 2A peptide contains a highly conserved $D^{12}(V/I)E(S/T)NPG_{2A}{}^{\downarrow}P^{19}{}_{2B}$ motif at its C-terminus which is critical for its function (Donnelly *et al.*, 1997; Ryan & Drew, 1994). This motif, together with upstream amino acids, is believed to interact with the ribosomal exit tunnel. This prevents the formation of a peptide bond between the C-terminal glycine (G^{18}) of 2A and the N-terminal proline of 2B, referred to here, as P^{19} since it is an important part of the cleavage mechanism (see also Donnelly et al. 2001a; Ryan et al. 1999). However, remarkably, protein synthesis continues without the requirement for a re-initiation event.

Investigations into the activity of the 2A sequence have mainly been performed using *in vitro* experiments. Typically, these have either used mRNAs with single ORFs encoding artificial polyproteins comprising two reporter proteins linked via the 2A peptide (Donnelly *et al.*, 2001b; Ryan *et al.*, 1991; Ryan & Drew, 1994) or by expressing cDNAs encoding a truncated viral polyprotein including the StopGo coding region (Palmenberg *et al.*, 1992). Alterations to

the conserved D^{12}(V/I)E(S/T)NP$G_{2A}{}^{\downarrow}P^{19}{}_{2B}$ motif reduced or abrogated the StopGo function (Donnelly *et al.*, 2001b; Sharma *et al.*, 2012), thereby showing that these amino acids are important for the correct StopGo "cleavage". Furthermore, Hahn and Palmenberg (1996) demonstrated that alterations to this motif also influenced the viability of encephalomyocarditis virus (EMCV, a cardiovirus) as they resulted in lethal phenotypes. Subsequently, Loughran et al. (2013) reported a similar observation for FMDV, as modification of the $S^{15}NPG^{18}{}_{2A}{}^{\downarrow}P^{19}{}_{2B}$ sequence to $S^{15}NPL^{18}{}_{2A}{}^{\downarrow}V^{19}{}_{2B}$ or $S^{15}NPA^{18}{}_{2A}{}^{\downarrow}P^{19}{}_{2B}$ also gave rise to a lethal phenotype.

However, recently, certain amino acid substitutions (e.g. 2A S^{15} to F/I and 2A N^{16} to H) that have been shown to severely (60-70%) impair "cleavage" at the 2A/2B junction, using *in vitro* assays (Donnelly *et al.*, 2001b), have been found to be tolerated within infectious FMDVs (Kjær J and Belsham GJ, submitted). In contrast, other substitutions (e.g. P^{19} to A and P^{19} to G) that inhibit cleavage more severely (by 89-100%) *in vitro*, were not found within rescued viruses. Indeed, viruses rescued from these mutant transcripts had sequences that exactly matched the wt sequence (i.e. the rescued viruses were not mutant). In these studies, we also determined a critical role for the StopGo mechanism for the overall level of replication/translation of FMDV RNA. FMDV replicons with a defective 2A sequence had a markedly lower replication efficiency compared to the wt replicon (Kjær J and Belsham GJ, submitted).

It is, therefore, apparent that some amino acid substitutions can be tolerated within the FMDV 2A peptide whereas other changes are not compatible with viability. To identify if any alternative residues can be accepted within the critical $N^{16}P^{17}G^{18}{}_{2A}{}^{\downarrow}P^{19}{}_{2B}$ motif, degenerate sequences, encoding all possible amino acid substitutions at each of these positions individually, were introduced into a full-length FMDV cDNA, as used previously (Gullberg *et al.*, 2013b; Kristensen *et al.*, 2017). In principle, this should result in the production of RNA transcripts encoding 2A peptides with a wide spectrum of "cleavage" activities. This was achieved by generating a large pool of plasmids, using site-directed mutagenesis with highly degenerate oligonucleotides, to change each of the individual codons corresponding to the amino acid residues within this conserved motif to NNN (where N is a mixture of all 4 bases). Using each pool of plasmids, RNA transcripts were prepared, *in vitro*, and introduced into baby hamster kidney (BHK) cells. Infectious viruses were rescued and characterized.

Results and Discussion

The expected generation of a pool of StopGo cDNA mutants that could potentially result in all possible single amino substitutions in place of the N^{16}, P^{17}, G^{18} and P^{19} residues (see Figure 1A) was analysed by sequencing (see Figure 1B). The heterogeneity at the expected positions was clear in each case (this does not prove that each of the possible codons was present but indicates it is likely).

Full-length RNA transcripts were produced, *in vitro*, and introduced into BHK cells. Infectious virus was generated and passaged in fresh cells. RNA was then extracted from the virus harvests and the sequence encoding the 2A peptide was amplified by RT-PCR. The pool of amplicons was introduced into the pCR-XL-TOPO vector and then the sequence of the inserts in 20 individual colonies was determined for each virus harvest. It was found that all of the rescued viruses analysed after passages p2, p3 and p4 encoded the wt amino acid sequence at the NPGP motif in 2A. Interestingly, the complete spectrum of the possible synonymous codons for each of the residues N^{16}, P^{17}, G^{18} and P^{19} was present in the rescued viral genomes at p2 (see Table 1). These results indicated that the approach had indeed generated a diverse pool of codons within the viruses. Furthermore, the very restricted range of nucleotide sequences encoding 2A observed within the rescued viruses strongly suggests that the specific amino acid sequence (NPGP), encoded by these nucleotide sequences, is critical for FMDV viability.

However, it was also apparent that the utilization of the different codons for the conserved amino acid residues varied. At p2, 55% of the sequences analysed had the wt codon for residue N^{16} (AAC) while the synonymous AAT codon was present in the remaining 45% of the rescued sequences. In the subsequent passages, the proportion of the AAC codon within the sequences increased to 75% and 95% by p3 and p4 respectively while the incidence of the AAT codon declined (Table 1). For residue P^{17}, at p2, the codon CCT was present in 55% of the colonies analysed and increased to 100% by p4. Each of the three other possible codons for P^{17} (CCC, CCA and CCG) were also observed at p2 but each declined as the wt codon became dominant. For residues G^{18} and P^{19}, the wt codons (GGG and CCC respectively) were in the minority (10 or 20%) at p2 and each of the synonymous codons were also present. However, interestingly, by p3 the wt codons had markedly increased to 50% abundance and by p4 were dominant ($\geq$90% abundance). For G^{18}, the GGA codon was the most abundant at p2 but declined during further passages to be only 10% of the sequences at p4. Similarly, for P^{19} the CCT codon was

present in 50% of the sequences at p2 but declined to just 5% by p4. Strikingly, by p3, the wt codon was present in 50-75% of the population at each of the 4 residues and by p4 the wt codon was present in 90-100% of the virus population in each case (Table 1). Thus, it appears that selection occurs for the wt nucleotide sequence during passage of the rescued viruses in cell culture.

The wt GGGCCC nt sequence encoding residues G^{18} and P^{19} is recognized in DNA by the restriction enzyme *ApaI* (see Figure 1A). Hence, it was possible to deplete the cDNA amplicons generated by RT-PCR, of the wt sequence from the rescued viruses by digesting them with *ApaI* prior to the cloning step (it was anticipated that this should enhance the detection of non-wt nucleotide sequences). The residual, full-length, 650bp amplicons were inserted into the pCR-XL-TOPO vector, as described above, and the plasmid DNA from individual colonies was sequenced. As expected, the wt codons for G^{18} and P^{19} were no longer observed in the cloned fragments (Table 2) and the G^{18} (GGA) and P^{19} (CCT) codons were predominant in these enriched populations. These results are consistent with those obtained without the *ApaI* digestion (since the GGA and CCT codons were also present in 50% of the fragments at p2 without this treatment, see Table 1) but clearly the apparent abundance of these non-wt codons is enhanced following the *ApaI* digestion (Table 2), as anticipated. The enrichment for non-wt sequences did not result in the detection of codons for alternative amino acids within the virus population. It had been anticipated that some amino acid substitutions at residue N^{16} might be rescued since a mutant (with N^{16} changed to H) has been shown to be viable (Kjær J and Belsham GJ, submitted) but, presumably, it was outcompeted by the wt virus.

It is interesting to note that the G^{18} (GGA) and P^{19} (CCT) codons have previously been found to be the second most abundant codons found in FMDV genomes from all seven serotypes (see Gao et al. 2014). This comparison of FMDV sequences also indicated that the alternate codon for N^{16} (AAC) is present in only a small minority of FMDV genomes and CCC is also a minor population of the codons used for residue P^{17}. The results presented in Table 1 clearly indicate that infectious FMDVs with these synonymous changes can be obtained but these viruses do not appear to be stably maintained in cell culture and are apparently selected against.

The evidence presented here strongly suggests that there is a distinct selection, within the virus when grown in cell culture, for codon AAC for N^{16}, CCT for P^{17}, GGG for G^{18} and CCC for P^{19}; thereby indicating that synonymous codon usage for this conserved motif is biased in these rescued viruses. It is particularly noteworthy that the codon preference for P^{17} and P^{19} is

different (CCT and CCC respectively). This raises the question of why does the virus select some codons over others? Various studies have demonstrated that synonymous codon usage bias plays an important role in the translation of certain mRNAs (Akashi 2001; Bulmer 1991; Novoa and Ribas de Pouplana 2012; Mauro and Chappell 2014). It is therefore conceivable that synonymous codons may influence the cleavage efficiency through the FMDV StopGo mechanism. As indicated above, a marked codon bias within the FMDV genome is apparent from the alignment of diverse FMDV 2A sequences as described by Gao et al. (2014). However, in the context of a synthetic reporter polyprotein, assayed within CHO cells, use of the four different synonymous codons for residue G^{18} of the 2A peptide resulted in very similar apparent "cleavage" efficiencies at the 2A/2B junction. This was interpreted as showing that it is the amino acid residue rather than the nt sequence which is critical for achieving cleavage (Gao *et al.*, 2014). However, using that assay system, the "cleavage" efficiency was only about 88-89% while essentially 100% cleavage occurs within the native context, as in the virus. The results obtained here (see Table 1) indicate that two separate selection effects may be operating. There is a clear selection for the NPGP motif at the amino acid level. However, in addition, there is a distinct codon bias within the context of the rescued infectious viruses and a significant selection pressure appears to exist for the wt sequence. This effect is fully consistent with the codon bias observed in the analysis of natural FMDV genomic sequences (Gao et al. 2014). This suggests that the FMDV RNA sequence itself (rather than just the encoded amino acid sequence) affects the "cleavage" process (StopGo mechanism) at the 2A/2B junction. Such an effect could be achieved through a direct interaction of the RNA sequence itself or potentially through interactions with the specific charged tRNAs involved in the translation process. In the case of the P^{17} and P^{19} codons, it is interesting to note that the same type of prolyl tRNA (with an IGG anticodon) has been reported to be used for decoding of the CCC and CCU codons in human cells (no gene for a tRNA that is cognate for CCC was identified, see Mauro and Chappell, 2014). However, in the current database of tRNA sequences from the Lowe laboratory, it appears that in humans, 1 of 23 genes for prolyl tRNAs has a GGG anticodon with 10 copies having an AGG anticodon. In the mouse genome, 1 of 20 genes for the prolyl tRNAs has the GGG anticodon and 8 genes have the AGG anticodon (see the gtrnadb.ucsc.edu database described in Chan and Lowe (2009)). Interestingly, in cattle and pigs (major hosts for FMDV) and also in the rat, there is no gene for a prolyl tRNA with a GGG anticodon. Thus, it is not clear whether a single, post-transcriptionally modified prolyl tRNA recognizes these two Pro codons (at least some of the time) or if different tRNAs are involved in the hamster cells used here. If a single tRNA is involved in recognizing both codons (as in

cattle, pigs and rats), then it seems that the RNA sequence itself must be influencing the StopGo process; it seems unlikely that this effect is mediated through some secondary or tertiary RNA structure, as this would presumably be lost on the ribosome during the process of translation. It will clearly be important to analyse the effect of the presence of the non-optimal synonymous codons on "cleavage" at the 2A/2B junction in its native context.

Materials and Methods

Construction of plasmids containing full-length mutant FMDV cDNAs

Pools of StopGo cDNA mutants that potentially result in all possible single amino substitutions in place of the N^{16}, P^{17}, G^{18} and P^{19} residues, respectively, were constructed. This was achieved using a 2-step site-directed mutagenesis procedure. This is a variation of the QuickChange protocol (Stratagene), using Phusion High-Fidelity DNA Polymerase (Thermo Scientific) with modified versions of the plasmid pT7S3 (Ellard *et al.*, 1999) as template. The wt pT7S3 contains the full-length cDNA for the O1Kaufbeuren B64 strain of FMDV. To eliminate the possibility of carrying over some residual wt template from the PCR, the templates used were modified versions of the pT7S3 with the codons for N^{16}, P^{17}, G^{18} or P^{19} changed to encode an alanine (A) residue in each case (see Figure 1A). These substitutions have been reported previously to result in a complete loss of apparent cleavage activity (Donnelly *et al.*, 2001b; Sharma *et al.*, 2012) and it has not been possible to rescue infectious virus containing these substitutions (Kjær J and Belsham GJ, submitted). The first round of PCRs used the forward mutagenic 2A PCR primers (Table 3), with a single reverse primer 8APN206 (Table 3) plus the four different modified pT7S3 plasmids as templates and generated amplicons of ca. 450 bp. These primary PCR products were then used as megaprimers for a second round of PCR with the respective mutant pT7S3 plasmids as templates to produce full-length plasmids. Following *DpnI* digestion, the products from each reaction were introduced into *E. coli* and grown as separate pools. The plasmid pools were sequenced using a BigDye Terminator v. 3.1 Cycle Sequencing Kit and a 3500 Genetic Analyzer (Applied Biosystems).

Rescue of virus from full-length cDNA plasmids

Plasmid DNA isolated from each pool was linearized by digestion with *HpaI* and RNA transcripts were prepared using T7 RNA polymerase (Ambion T7 MEGAscript) at 37°C for 4

hours. The integrity of the transcripts was assessed on agarose gels and quantified by spectrophotometry (NanoDrop 1000, Thermo Scientific) after which they were introduced into BHK cells by electroporation, as described previously (Nayak *et al.*, 2005). The BHK cells were maintained in Dulbecco's Modified Eagles Medium (DMEM) supplemented with 5% fetal calf serum, and incubated at 37°C with 5% CO_2. At 2 days post-electroporation, the viruses were harvested by freezing and then amplified through three passages (p2, p3 and p4) in BHK cells.

Characterization of viruses following multiple passages

After each passage, viral RNA was extracted from a sample of the virus harvest (using the RNeasy Mini Kit, Qiagen) and converted to cDNA using ready-to-go you-prime first-strand beads (GE Healthcare Life Sciences). FMDV cDNA, which included the whole 2A coding region, was amplified in PCRs (AmpliTaq Gold DNA polymerase, Thermo Scientific) using primers 8APN206 and 8APN203 (see Figure 1 and Table 3). Control reactions, without RT, were used to ensure that the analysed products were derived from RNA and not from the presence of carryover plasmid DNA template. The amplicons (ca. 650 bp) were visualized on 1% agarose gels and purified (GeneJET gel extraction kit, Thermo Scientific). These amplicons should be representative of the heterogeneity present in the rescued virus populations. The resulting collections of fragments were inserted into pCR-XL-TOPO (Thermo Scientific) and the sequence of the cDNA fragment present in individual bacterial clones (20 colonies for each of the 4 residues) was determined using the same reverse primer as used for the PCR. The fragments from codon mutants G^{18} and P^{19} were also enriched for the non-wt sequence populations by digestion of the cDNA with *ApaI* prior to gel purification and insertion into the pCR-XL-TOPO vector as described above.

Acknowledgements

We thank Preben Normann for excellent technical assistance. The studies were supported by core funds within DTU-Vet.

References

Akashi H. 2001. Gene expression and molecular evolution. *Curr Opin Genet Dev* **11**: 660–666.

Atkins JF, Wills NM, Loughran G, Wu C-Y, Parsawar K, Ryan MD, Wang C-H, Nelson CC. 2007. A case for "StopGo": reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go). *RNA* **13**: 803–810.

Belsham GJ. 2005. Translation and replication of FMDV RNA. *Curr Top Microbiol Immunol* **288**: 43–70.

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**: 897–907.

Chan PP, Lowe TM. 2009. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucl Acids Res* **37**:D93-D97.

Donnelly ML, Gani D, Flint M, Monaghan S, Ryan MD. 1997. The cleavage activities of aphthovirus and cardiovirus 2A proteins. *J Gen Virol* **78**: 13–21.

Donnelly ML, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, Ryan MD. 2001a. Analysis of the aphthovirus 2A/2B polyprotein "cleavage" mechanism indicates not a proteolytic reaction, but a novel translational effect: A putative ribosomal "skip." *J Gen Virol* **82**: 1013–1025.

Donnelly ML, Luke GA, Hughes LE, Luke G, Mendoza H, Dam E, Gani D, Ryan MD. 2001b. The "cleavage" activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring "2A-like" sequences. *J Gen Virol* **82**: 1027–1041.

Doronina VA, Wu C, de Felipe P, Sachs MS, Ryan MD, Brown JD. 2008. Site-specific release of nascent chains from ribosomes at a sense codon. *Mol Cell Biol* **28**: 4227–4239.

Ellard FM, Drew J, Blakemore WE, Stuart DI, King AMQ. 1999. Evidence for the role of His 142 of protein 1C in the acid induced disassembly of foot and mouth disease virus capsids. *J Gen Virol* **80**: 1911–1918.

Gao ZL, Zhou JH, Zhang J, Ding YZ, Liu YS. 2014. The silent point mutations at the cleavage site of 2A/2B have no effect on the self-cleavage activity of 2A of foot-and-mouth disease virus. *Infect Genet Evol* **28**: 101–106.

Gullberg M, Polacek C, Bøtner A, Belsham GJ. 2013. Processing of the VP1/2A junction is not necessary for production of foot-and-mouth disease virus empty capsids and infectious viruses: characterization of “self-tagged” particles. *J Virol* **87**: 11591–11603.

Hahn H, Palmenberg AC. 1996. Mutational analysis of the encephalomyocarditis virus primary cleavage. *J Virol* **70**: 6870–6875.

Kristensen T, Normann P, Gullberg M, Fahnøe U, Polacek C, Rasmussen TB, Belsham GJ. 2017. Determinants of the VP1/2A junction cleavage by the 3C protease in foot-and-mouth disease virus-infected cells. *J Gen Virol* **98**: 385–395.

Loughran G, Libbey JE, Uddowla S, Scallan MF, Ryan MD, Fujinami RS, Rieder E, Atkins JF. 2013. Theiler’s murine encephalomyelitis virus contrasts with encephalomyocarditis and foot-and-mouth disease viruses in its functional utilization of the StopGo non-standard translation mechanism. *J Gen Virol* **94**: 348–353.

Martinez-Salas E, Belsham GJ. 2017. Genome Organisation, Translation and Replication of Foot-and-mouth Disease Virus RNA. In *Foot-and-Mouth Disease: Current Research and Emerging Trends* (eds. F. Sobrino and E. Domingo), pp. 13–42, Caister Academic Press.

Mauro VP, Chappell SA. 2014. A critical analysis of codon optimization in human therapeutics. *Trends Mol Med* **20**: 604–613.

Nayak A, Goodfellow IG, Belsham GJ. 2005. Factors required for the Uridylylation of the foot-and-mouth disease virus 3B1, 3B2, and 3B3 peptides by the RNA-dependent RNA polymerase (3Dpol) in vitro. *J Virol* **79**: 7698–7706.

Novoa EM, Ribas de Pouplana L. 2012. Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends Genet* **28**: 574–581.

Palmenberg AC, Parks GD, Hall D, Ingraham RH, Seng TW, Pallal P V. 1992. Proteolytic processing of the cardioviral cleavage in clone-derived P2 region : primary 2A / 2B precursors. *Virology* **190**: 754–762.

Ryan MD, Donnelly M, Lewis A, Mehrotra AP, Wilkie J, Gani D. 1999. A Model for Nonstoichiometric, Cotranslational Protein Scission in Eukaryotic Ribosomes. *Bioorg Chem* **27**: 55–79.

Ryan MD, Drew J. 1994. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J* **13**: 928–33.

Ryan MD, King AMQ, Thomas GP. 1991. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J Gen Virol* **72**: 2727–2732.

Sharma P, Yan F, Doronina VA, Escuin-Ordinas H, Ryan MD, Brown JD. 2012. 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Res* **40**: 3143–3151.

Tulloch F, Luke GA, Ryan MD. 2017. Foot-and-mouth disease virus proteinases and polyprotein processing. In *Foot-and-Mouth Disease: Current Research and Emerging Trends* (eds. F. Sobrino and E. Domingo), pp. 43–59, Caister Academic Press.

Tables

Table 1: Codon utilization encoding the "NPGP" motif at the 2A/2B junction within rescued FMDVs.

Residue	Codon		p2 %[1]	p3 %[1]	p4 %[1]
N16	*AAT*		45	25	5
N16	AAC	wt	55	75	95
P17	CCT	wt	55	70	100
P17	*CCC*		10	15	0
P17	*CCA*		20	5	0
P17	*CCG*		15	10	0
G18	*GGT*		15	5	0
G18	*GGC*		15	5	0
G18	*GGA*		50	40	10
G18	GGG	wt	20	50	90
P19	*CCT*		50	25	5
P19	CCC	wt	10	50	95
P19	*CCA*		10	5	0
P19	*CCG*		30	20	0

1: From sequencing of plasmid DNA isolated from separate 20 colonies in each case, the proportion (%) of each codon present in the rescued FMDVs is indicated at the different passage (p) numbers. Codon frequency values of 50-70% are highlighted in light grey whereas values from 75-100% are highlighted in dark grey.

Table 2: Enrichment for non-wt sequences encoding residues G[18] and P[19] within the "NPGP" motif within rescued FMDVs.

Residue	Codon		Pretreatment	p2 %[2]	p3 %[2]
G18	*GGT*		*ApaI*[1]	5	0
G18	*GGC*		*ApaI*[1]	15	5
G18	*GGA*		*ApaI*[1]	80	95
G18	GGG	wt	*ApaI*[1]	0	0
P19	*CCT*		*ApaI*[1]	60	55
P19	CCC	wt	*ApaI*[1]	0	0
P19	*CCA*		*ApaI*[1]	15	0
P19	*CCG*		*ApaI*[1]	25	45

1: Following RT-PCR, the 650bp amplicons were digested with *ApaI* to enrich the population in non-wt sequences and the residual intact products were inserted into the pCR-XL-TOPO vector (see text).

2: From sequencing of plasmid DNA isolated from separate 20 colonies in each case, the proportion (%) of each codon present in the rescued FMDVs is indicated at the different passage (p) numbers. Codon frequency values of 50-70% are highlighted in light grey whereas values from 75-100% are highlighted in dark grey.

Table 3: Primers used to create and sequence mutant FMDV cDNAs.

Primer	Sequence (5'-3')
Fwd_2A_N16A_degen	GGAGTCCNNNCCTGGGCCCTTC
Fwd_2A_P17A_ degen	GTCCAACNNNGGGCCCTTC
Fwd_2A_G18A_ degen	GACGTCGAGTCCAACCCTNNNCCCTTCTTTTTCTCCGA CGTTA
Fwd_2A_P19G_ degen	TCG AGTCCAACCCTGGGNNNTTCTTTTTCTCCGACGTTAGG
8APN206	CACCCGAAGACCTTGAGAG
8APN203	CTCCTTCAACTACGGTGCC

Figure

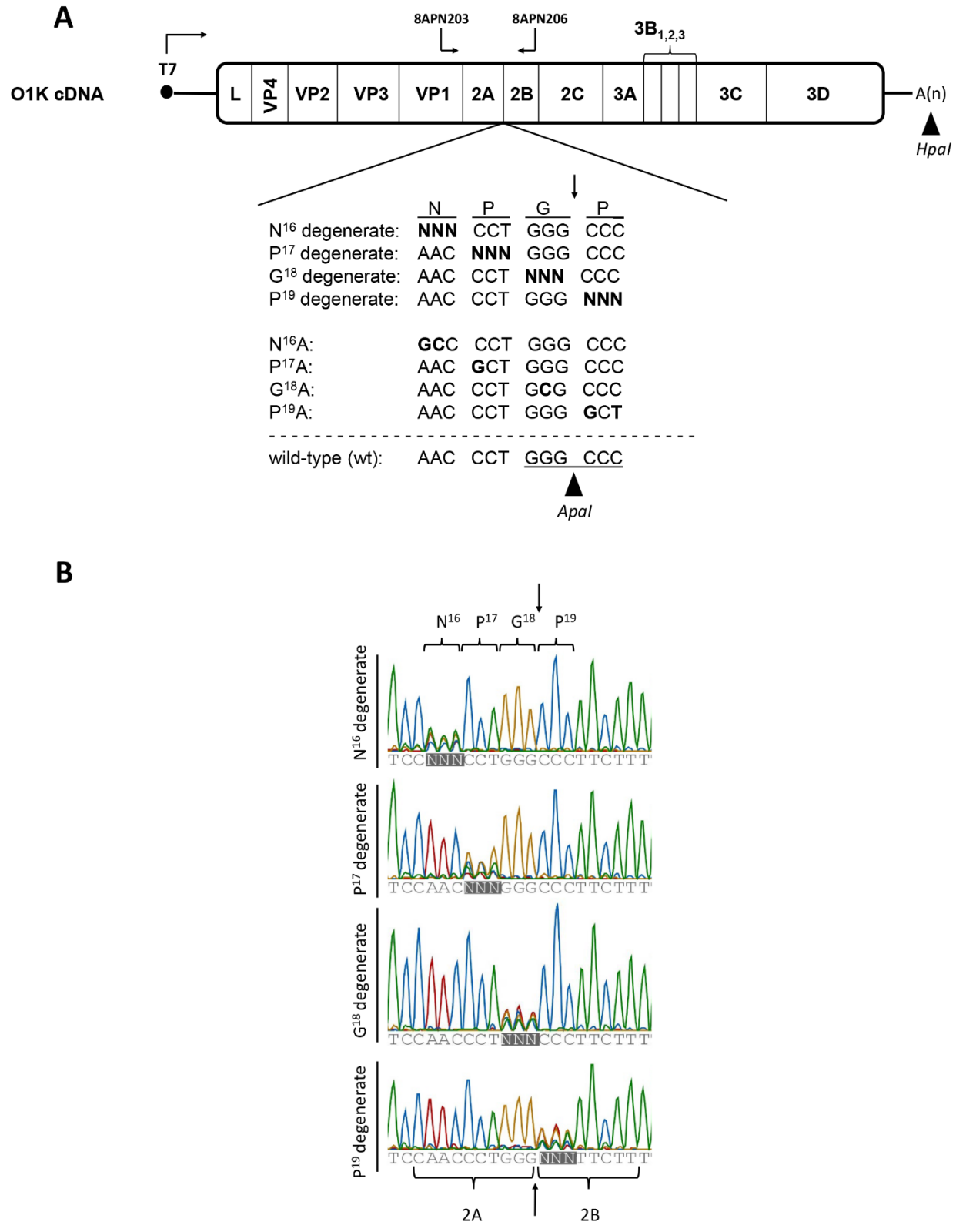

Figure 1: Structure of the FMDV O1 Kaufbeuren (O1K) cDNA and its derivatives. (A) The plasmid-encoded amino acids and the corresponding nucleotide sequences at the 2A/2B

junction are shown. The FMDV O1K degenerate codon mutants were produced as described in the text using the mutant pT7S3 plasmids encoding the $N^{16}A$, $P^{17}A$, $G^{18}A$ and $P^{19}A$ substitutions as templates. The full-length plasmid pools were linearized using *HpaI* prior to *in vitro* transcription and virus rescue. The locations of the *HpaI* and *ApaI* restriction sites that were used are marked. N = a mixture of the 4 nucleotides. (B) Chromatograms and sequences of the FMDV cDNA corresponding to the NPGP motif at the 2A/2B junction. Degenerate positions showing the presence of multiple nucleotides are marked with an N (in bold type). The colour code in the chromatograms is as follows: A (red), T (green), G (yellow), C (blue).

Manuscript 3

The influence of conserved RNA structures in the 2B, 3C and 3D coding regions of the foot-and-mouth disease virus genome on virus replication

Jonas Kjær & Graham J. Belsham

Manuscript in preparation

2017

The influence of conserved RNA structures in the 2B, 3C and 3D coding regions of the foot-and-mouth disease virus genome on virus replication

Jonas Kjær[1] and Graham J. Belsham[1*].

1: DTU National Veterinary Institute, Technical University of Denmark, Lindholm, DK-4771, Kalvehave, Denmark

*Corresponding author: Graham J. Belsham, email: grbe@vet.dtu.dk, phone: +45 3588 7985

Figures: 3

Tables: 1

Abstract

The positive-sense ssRNA genome of foot-and-mouth disease virus (FMDV) has multiple conserved RNA secondary and higher-order structures which play fundamental roles in the life cycle of the virus. A previous study using SHAPE probing found three conserved RNA structures within the coding region for 2B, 3C and 3D with significantly lower synonymous substitution rates compared to the remainder of the genome. In this study, we aimed to investigate the influence of these structures on viral protein synthesis and virus viability. It was found that introduction of synonymous nucleotide sequence changes that disrupt the structures but do not result in amino acids changes can be tolerated and are retained after three passages in BHK cells. The same substitutions did not impair the replication of a FMDV replicon in these cells. These results show that the RNA structures located in the coding region of 2B, 3C and 3D are not essential for RNA replication or viability of FMDV in BHK cells.

Introduction

Foot-and-mouth disease virus (FMDV) is the causative agent of foot-and-mouth disease, a highly economically important infectious disease of cloven- hooved animals. The virus is the prototypic member of the genus *Aphthovirus* within the family *Picornaviridae* and exists in seven different serotypes (O, A, C, SAT 1–3 and Asia-1) (Knowles & Samuel, 2003). The FMDV particle includes a single copy of the positive-sense RNA genome (~8300 nt) which has a single large ORF (~7000 nt) encoding a large polyprotein generating 15 distinct mature proteins (Belsham, 2005). The genome is encapsidated by an icosahedral capsid composed of 60 copies of four different capsid proteins (VP1, VP2, VP3, and VP4). The ORF is flanked by untranslated regions (UTRs) termed the 5′-UTR and the 3′-UTR which contain multiple secondary structures. The long 5′-UTR contains five distinct elements, including a type II internal ribosome entry site (IRES) and the *cis*-acting replicative element (*cre*) in addition to the S-fragment, a polypyrimidine tract of variable length and multiple copies of a pseudoknot (Escarmís *et al.*, 1995). The smaller 3′-UTR (ca. 90 nt) contains two RNA stem-loops upstream of a 3′-poly(A) tail (Witwer *et al.*, 2001). These structures can participate in inter- and intramolecular interactions such as RNA/RNA interactions but also binding to viral and cellular proteins. Besides the local folding displayed by the some of these regions, long-range interactions between distantly separated RNA elements are also suggested to regulate FMDV replication (Diaz-Toledano *et al.*, 2017; Sáiz *et al.*, 2001). Studies have demonstrated that the

structured 3′-UTR directly binds to the S-fragment and IRES elements at two distinct positions in the 5′-UTR through specific long-range RNA–RNA contacts (Diaz-Toledano *et al.*, 2017). Elaborate RNA structures are not limited to the UTRs of the genome. Functional structures embedded within protein-coding sequence have been found in rhinovirus and poliovirus (PV) genomes, both within the picornavirus family (McKnight & Lemon, 1996; Yang *et al.*, 2008). Such structures are crucial for replication of the viral genome and may play additional roles in other processes e.g. genome packaging and alterations of the host antiviral defence. Capsid assembly and the encapsidation of the viral genome are suggested to involve interactions between capsid proteins and secondary structures in the viral genome as found for the bacteriophage MS2 (Stockley *et al.*, 2007). Recently, Logan *et al.*, (2017) identified several predicted RNA secondary structures located within the ORF of FMDV that may be involved in genome packaging. Synonymous substitutions which disrupted the structures did not influence viral replication but resulted in attenuation of virus growth in cell culture.

Selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) has previously been successfully used to identify new potential RNA structures within the genomes of important human pathogens as PV and HIV (Watts *et al.*, 2009). Burrill *et al.*, (2013) used SHAPE probing to generate a global RNA structure map of the PV genome and found a conserved structure with the coding region of the 3D, the RNA-dependent RNA polymerase (RdRp), which was involved in viral replication and infectivity. Poulsen, (2015) used a similar approach with a modified version of SHAPE (Poulsen *et al.*, 2015) to probe the RNA structure of the entire FMDV genome in order to identify regions with stable secondary RNA structures. This study aimed to investigate whether these structures play an essential role in viral protein synthesis, viral RNA replication and viability of FMDV.

Materials and methods

Construction of plasmids containing full-length mutant FMDV cDNAs and FMDV replicons containing *Gaussia* Luciferase.

The structures of the plasmids containing full-length FMDV cDNAs used in this study are indicated in Figure 1. The pT7S3 plasmid (Ellard et al., 1999) containing the full-length cDNA corresponding to the O1Kaufbeuren B64 strain of FMDV was used as a template in a 2-step site-directed mutagenesis procedure, a variation of the QuickChange protocol (Stratagene), to generate full-length mutant FMDV cDNAs with synonymous mutations in the 2B, 3C and 3D

coding regions. Briefly, amplicons specifying particular synonymous mutations in the 2B, 3C and 3D coding regions were generated in a first round of PCR using forward mutagenic PCR primers (Table 1) with appropriate reverse primers (Table 1) and the plasmid pT7S3 as template. The primary PCR products of varying size (2B: 212 bp, 2C: 605 bp and 3D: 326 bp) were gel purified (GeneJet gel extraction kit, Thermo Fisher Scientific) and used as primers for a second round of PCR with plasmid pT7S3 as template. The *Dpn-I* resistant full-length products (plasmids) were amplified in Escherichia coli (*E.coli*) TOP10 cells (Thermo Fisher Scientific), purified (Midiprep kit; Thermo Scientific) and verified by sequencing of the 2B, 3C and 3D coding region respectively with a BigDye Terminator v. 3.1 Cycle Sequencing kit and a 3500 Genetic Analyzer (Applied Biosystems).

The FMDV replicon plasmid and RNA polymerase knockout control have been described previously (Kjær & Belsham, submitted, see Manuscript 1) and the mutants were generated as described above.

In vitro transcription

Replicon plasmid or full-length FMDV plasmid (5µg) DNA was linearized with *HpaI* (Thermo Fisher Scientific), purified (GeneJET PCR Purification Kit, Thermo Fisher Scientific) and eluted in RNase-free water. The linear DNA was used in a 20 µl *in vitro* transcription reaction using the Megascript T7 kit (Ambion). Reaction mixtures were incubated at 37°C for 4 h and with 2 units of TURBO DNase for 30 min after which the RNA was purified using the MEGAclear Transcription Clean-Up Kit according to the manufacturer's instructions. All transcripts were quantified by NanoDrop 1000 spectrophotometry (Thermo Scientific), and RNA integrity was assessed by electrophoresis using an ethidium bromide-stained agarose gel (1%) in TBE buffer.

Rescue of virus from full-length cDNA plasmids

Full-length linearized FMDV RNA transcripts (5 µg) were introduced into BHK cells by electroporation as described previously (Nayak et al., 2005). At 48 h after electroporation, the rescued viruses were harvested by freezing and amplified in two subsequent passages (P2 and P3) in fresh BHK cells at 37°C. After these passages, RNA was extracted (RNeasy Mini Kit, Qiagen) and reverse transcribed into cDNA using Ready-To-Go You-Prime First-Strand Beads with random hexamer primers. Amplicons including either the 2B, 3C or 3D coding region,

were amplified by PCR (AmpliTaq Gold DNA polymerase, Thermo Fischer Scientific) using appropriate primers (see Table 1). Control reactions, lacking reverse transcriptase, were used to ensure that the PCR products were derived from RNA and not from residual DNA template. The amplicons were visualized in 1% agarose gels, purified (GeneJET gel extraction kit, Thermo Fischer Scientific) and sequenced as previously described. Sequences were analysed using Geneious 7.2 (Biomatters, Auckland, New Zealand).

Gaussia luciferase assay

The replicon replication assay was performed in 24-well plates using monolayers of BHK cells. Replicon RNA (2 µg) was introduced into BHK cells by transfection using the Lipofectin transfection reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. The plates were then incubated at 37°C in 5 % CO_2 . After 1, 2, 3, 4, 6 and 8 hours, the medium was removed and the BHK monolayers were washed twice. The cells were then harvested by adding 100 µl of *Renilla* luciferase assay lysis buffer (Promega) to each well and incubated at room temperature for 15 min. The luciferase expression was quantified in a Luminometer (Titertek-Berthold) by addition of this lysate (20 µl) to Renilla Luciferase Assay reagent (100 µl) according to the manufacturer's instructions.

RT-PCR and full-genome sequencing

cDNA was generated from extracted RNA (RNeasy Mini Kit, Qiagen) using SuperScript III (Invitrogen) with a T27 reverse primer according to the manufacturer's protocol. Two PCR products were amplified from the viral cDNA using Phusion DNA polymerase (Thermo Fisher) with the primers 13NPN2 and 10PPN30 (Table 1) and separately 8APN200 +13NPN3 (Table 1). The two fragments of ca. 4 and 4.2 kb respectively are overlapping and correspond to most of the FMDV genome (downstream of the poly(C) tract). The fragments were visualized on an agarose gel, purified (GeneJET gel extraction kit, Thermo Fischer Scientific and analysed by NGS essentially as described previously (Kristensen *et al.*, 2017). Geneious 7.2 (Biomatters, Auckland, New Zealand) was used for mapping the reads to the sequence of the FMDV strain O1K sequence (accession no. X00871).

RNA secondary structure prediction

Thermodynamic RNA structural predictions were carried out using the Zuker energy minimization algorithm (Zuker, 2003) implemented in the Mfold program using default settings through the web interface at http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form.

Results

Production of modified FMDVs with synonymous substitutions within the coding region for 2B, 3C and 3D

A recent study by Poulsen, (2015) conducted a genome-wide SHAPE probing of the FMDV genome. The SHAPE probing identified 17 novel secondary structures within the protein coding region. Interestingly, three of these structures, which are located within the coding regions for 2B, 3C and 3D (see Figure 1 and 2), were found to have synonymous substitution rates significantly lower than the rest of the genome suggesting that the nucleotide structure in addition to the amino acid sequence is of importance.

To investigate whether these structures are important for viral RNA replication and viability, specific substitutions within the coding regions for 2B, 3C and 3D which would disrupt the secondary structures without altering the encoded amino acid sequence were introduced into a FMDV replicon and a full-length FMDV cDNA respectively (see Figure 1). Three mutants with synonymous substitutions covering the entire sequence length of the 2B, 3C and 3D structures termed $2B_{syn\_mut}$, $3C_{syn\_mut}$ and $3D_{syn\_mut}$ respectively were generated. The 2B structure also has non-paired nucleotides in the terminal loop and bulge (see Figure 2). Such non-paired nucleotides have previously been reported to be involved in long-range RNA-RNA interactions. Two further mutants, termed $2B_{syn\_mut\_loop}$ (terminal loop) and $2B_{syn\_mut\_bugle}$ (bugle) were generated in which the motif in both elements was changed. The structure identified by SHAPE within the coding region of 3D spans 128 nt and is larger and more elaborate compared to those identified within the coding sequence of 2B and 3C. To identify the regions within this structure which potentially were of significance for virus replication, synonymous substitutions were designed, in addition, to disrupting the entire structure also aimed at disrupting just the 5′ and 3′ regions of this structure termed $3D_{syn\_mut\_5'}$ and $3D_{syn\_mut\_3'}$. The 5′ and 3′ substitutions were also both incorporated in the same construct and termed $3D_{syn\_mut\_5'+3'}$. Modifications to the FMDV *Gluc* replicon and the full-length cDNA

corresponding to the O1 Kaufbeuren strain of FMDV were performed and an illustration of the modifications is shown in Figure 2.

The effect of synonymous substitutions within the conserved structures found in 2B, 3C and 3D on RNA replication assessed using a FMDV replicon

A FMDV replicon encoding a reporter protein provides a quantitative measurement of FMDV RNA replication. To monitor the influence of the RNA structure modifications on replication of viral RNA, eight different changes within the coding regions for 2B, 3C and 3D were introduced into a FMDV replicon. This replicon has the coding sequences for the FMDV structural proteins (VP1-VP3) replaced by the sequence encoding the *Gaussia* luciferase (Gluc) reporter protein. The plasmids were linearized and RNA transcripts were prepared *in vitro* and introduced into BHK cells (see Methods). Lysates were prepared from cells at various times after electroporation with the different transcripts and assayed for Gluc activity. At early times post-transfection of BHK cells with wt-Gluc RNA transcripts, Gluc expression increased and reached maximal expression by 8 hours (see Figure 3). In contrast, wt-GlucΔ3D, a mutant replicon which lacks a portion of the coding sequence for 3D expressed Gluc maximally only at 2 hours post-transfection (Figure 3), consistent with translation of the input RNA followed by RNA turnover. Unexpectedly, all of the RNA structure mutants expressed the same level of Gluc activity as the wt-Gluc at each time point. The same mutants also exhibited the same level of replication efficiency when tested in a GFP replicon system as previously described (Herod *et al.*, 2016; Tulloch *et al.*, 2014) (data not shown).

Effect of synonymous substitutions in 2B, 3C and 3D on viral viability and genetic stability

To determine the viability and genetic stability of FMD viruses with disrupted RNA structures within the 2B, 2C and 3D coding regions, the mutant plasmids were linearized and RNA transcripts, prepared *in vitro*, were introduced into BHK cells by electroporation. All eight FMDV mutants were indeed viable within BHK cells with the production of progeny viruses and full CPE detectable after the third passage. Full CPE for the wt was detectable at the second passage. The rescued viruses were sequenced to identify possible adaptations or reversions. After three passages in BHK cells, the rescued viruses had each retained the plasmid-derived synonymous nucleotide substitutions. To determine whether adaptation to overcome the

defects had appeared elsewhere in the genomes of the 2B mutants, we examined the rescued viral genomes throughout the near-complete genome sequence. To achieve this, RNA was extracted from the rescued 2B mutant viruses (see Methods) at the third passage. Two separate, but overlapping, cDNA fragments including the near-complete genome (downstream of the poly(C) tract) were produced by reverse transcription (RT)-PCR, mixed and then sequenced using next-generation sequencing (NGS). No adaptations were observed within the modified 2B coding region or in the remainder of the genome.

Discussion

The genomes of RNA viruses contain RNA structures that are crucial for translation and RNA replication and may play additional, uncharacterized roles during the viral replication cycle. In recent years, advances been made in identifying and charactering such RNA structures in positive-sense RNA viruses through the use of SHAPE probing. Within the field of FMDV biology, SHAPE has given new insights into the dynamic RNA–RNA interactions between the IRES and the 3′-UTR (Diaz-Toledano *et al.*, 2017). Recently, Poulsen, (2015) probed the RNA structures within the entire FMDV genome, also using SHAPE probing, thereby generating a genome-wide RNA structure model. Three structures present in the coding sequences for 2B, 3C and 3D showed significantly lower synonymous substitution rates than observed in the remainder of the genome, suggesting a selective pressure to maintain these structures. Synonymous mutations were designed to disrupt the structures in these regions without altering the encoded amino acid sequence and these were introduced into a FMDV replicon and full-length FMDV cDNAs. The site-directed mutagenesis of these structures within a replicon system revealed that none of the RNA elements played an important role in viral replication in BHK cells as the mutants exhibited about the same level of RNA replication as wt in this system, as judged by the reporter protein expression. The mutants were likewise also able to produce infectious viruses in the context of the full-length FMDV suggesting that the structures are not essential for virus viability. Interestingly, the synonymous mutations proved to be genetically stable after three passages in BHK cells. In addition, no other adaptations were present elsewhere in the genome of the 2B mutants suggesting that the structure was not involved in long-range RNA interaction with other parts of the genome. Previous studies of other picornavirus RNA structures have found *cre's* which are essential for the viral life cycle, within the coding region (McKnight & Lemon, 1996; Yang *et al.*, 2008). The understanding of

the role RNA structures play in biological systems has expanded in recent years and although the structures tested in the FMDV genome appear to have no observable influence on either translation or replication of the FMDV genome in BHK cells they can still possess additional, uncharacterized roles during the viral replication cycle, e.g. within the context of the host response to infection as seen by Kloc *et al.*, (2017) with the S-fragment. Logan *et al.*, (2017) recently demonstrated a novel technique to identify possible RNA secondary structures involved in packaging of the FMDV genome. Through the use of NGS and structure prediction software they identified several RNA structures with a common motif at the terminal loops which are suggested to be involved in FMDV packaging. Disruption of these structures through synonymous substitutions resulted in attenuation of virus growth in BHK cells probably due to a reduction in assembly of mature virions. In this connection, the conserved loop sequences (AAUCCA) that are presumed to be part of the packaging mechanism were not found in any of the three structures tested here. Simmonds *et al.*, (2004) suggested that RNA structure could be implicated in the evasion of host cell defences by facilitating escape from innate responses induced by certain structured RNAs. Large cellular RNA structures such as rRNA and tRNA have been found not to induce a dsRNA defence response. It is speculated that RNA structures within the viral genome have evolved as a way to mimic cellular structured RNAs thereby circumventing the cellular dsRNA response. It could be interesting to study the growth kinetics of all the structural mutant viruses within primary bovine thyroid cells (pBTY) as they have an intact interferon system in contrast to BHK cells.

Certain RNA structures are important for optimal function of the virus and SHAPE probing has provided an excellent way for determining RNA structures at a single nucleotide resolution. The combination of SHAPE analysis of a viral genome along with evolutionary analysis and functional biochemical assays should help to find additional structures located inside or outside of the coding sequence which influence the life cycle of FMDV. Further investigations into the influence of these structures on the viral life cycle are needed.

References

Belsham, G. J. (2005). Translation and replication of FMDV RNA. *Curr Top Microbiol Immunol* **288**, 43–70.

Burrill, C. P., Westesson, O., Schulte, M. B., Strings, V. R. & Segal, M. (2013). Global RNA Structure Analysis of Poliovirus Identifies a Conserved RNA Structure Involved in Viral Replication and Infectivity **87**, 11670–11683.

Diaz-Toledano, R., Lozano, G. & Martinez-Salas, E. (2017). In-cell SHAPE uncovers dynamic interactions between the untranslated regions of the foot-and-mouth disease virus RNA. *Nucleic Acids Res* **45**, 1416–1432.

Ellard, F. M., Drew, J., Blakemore, W. E., Stuart, D. I. & King, A. M. Q. (1999). Evidence for the role of His 142 of protein 1C in the acid induced disassembly of foot and mouth disease virus capsids. *J Gen Virol* **80**, 1911–1918.

Escarmís, C., Dopazo, J., Dávila, M., Palma, E. L. & Domingo, E. (1995). Large deletions in the 5′-untranslated region of foot-and-mouth disease virus of serotype C. *Virus Res* **35**, 155–167.

Herod, M. R., Ferrer-Orta, C., Loundras, E.-A., Ward, J. C., Verdaguer, N., Rowlands, D. J. & Stonehouse, N. J. (2016). Both *Cis*- and *Trans*- Activities of the Foot-and-Mouth Disease Virus 3D Polymerase are Essential for Viral RNA Replication. *J Virol* **90**, JVI.00469-16.

Kloc, A., Segundo, F. D., Schafer, E. A., Rai, D. K., Kenney, M., Santos, T. D. L. & Rieder, E. (2017). Foot-and-mouth disease virus 5’ -terminal S fragment is required for replication and modulation of the innate immune response in host cells. *Virology* **512**, 132–143.

Knowles, N. J. & Samuel, A. R. (2003). Molecular epidemiology of foot-and-mouth disease virus. *Virus Res* **91**, 65–80.

Kristensen, T., Normann, P., Gullberg, M., Fahnøe, U., Polacek, C., Rasmussen, T. B. & Belsham, G. J. (2017). Determinants of the VP1/2A junction cleavage by the 3C protease in foot-and-mouth disease virus-infected cells. *J Gen Virol* **98**, 385–395.

Logan, G., Newman, J., Wright, C. F., Lasecka-Dykes, L., Haydon, D. T., Cottam, E. M. & Tuthill, T. J. (2017). Deep sequencing of foot-and-mouth disease virus reveals RNA

sequences involved in genome packaging. *J Virol* . (in press). doi: e-pub October 2017 10.1128/JVI.

McKnight, K. L. & Lemon, S. M. (1996). Capsid Coding Sequence Is Required for Efficient Replication of Human Rhinovirus 14 RNA. *J Virol* **70**, 1941–1952.

Nayak, A., Goodfellow, I. G. & Belsham, G. J. (2005). Factors required for the Uridylylation of the foot-and-mouth disease virus 3B1, 3B2, and 3B3 peptides by the RNA-dependent RNA polymerase (3Dpol) in vitro. *J Virol* **79**, 7698–7706.

Poulsen, L. D. (2015). Investigation of RNA structure by high-throughput SHAPE-based probing methods. PhD thesis. University of Copenhagen.

Poulsen, L. D., Kielpinski, L. J., Salama, S. R., Krogh, A. & Vinther, J. (2015). SHAPE Selection (SHAPES) enrich for RNA structure signal in SHAPE sequencing-based probing data. *RNA* **21**, 1042–52.

Sáiz, M., Gómez, S., Martínez-Salas, E. & Sobrino, F. (2001). Deletion or substitution of the aphthovirus 3′ NCR abrogates infectivity and virus replication. *J Gen Virol* **82**, 93–101.

Simmonds, P., Andrew, T. & Evans, D. J. (2004). Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. *RNA* **10**, 1337–1351.

Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J., Homans, S. W. & Ashcroft, A. E. (2007). A Simple, RNA-Mediated Allosteric Switch Controls the Pathway to Formation of a T = 3 Viral Capsid. *J Mol Biol* **369**, 541–552.

Tulloch, F., Pathania, U., Luke, G. A., Nicholson, J., Stonehouse, N. J., Rowlands, D. J., Jackson, T., Tuthill, T., Haas, J. & other authors. **(2014).** FMDV replicons encoding green fluorescent protein are replication competent. *J Virol Methods* **209**, 35–40. Elsevier B.V.

Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess Jr, J. W., Swanstrom, R., Burch, C. L. & Weeks, K. M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716.

Witwer, C., Rauscher, S., Hofacker, I. L. & Stadler, P. F. (2001). Conserved RNA

secondary structures in Picornaviridae genomes. *Nucleic Acids Res* **29**, 5079–5089.

Yang, Y., Yi, M., Evans, D. J., Simmonds, P. & Lemon, S. M. (2008). Identification of a conserved RNA replication element (cre) within the 3Dpol-coding sequence of hepatoviruses. *J Virol* **82**, 10118–10128.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406–3415.

Table 1: Oligonucleotides used to create and sequence mutant FMDV cDNAs.

Primer name	Sequence (5′-3′)	Function
$2B_{syn\_mut\_loop\_fwd}$	GTCCAAGGACCCGGTCCTTGTGG	Forward 2B mutagenesis
$2B_{syn\_mut\_bugle\_fwd}$	CTTGTGGCCATTATGCTGGCCGACACC	Forward 2B mutagenesis
$2B_{syn\_mut\_few}$	AGCTCCTAAGCCGCCTGTCGTGCATGGCAG CGGTAGCGGCGCGATCAAAAGATCCGGTA CTAGTAGCAATAATGCTGGCCGACACCGGT CTCGAGATTCT	Forward 2B mutagenesis
$3C_{syn\_mut\_fwd}$	CGGTAGCCATCTGCTGCGCTACTGGAGTCT TCGGAACGGCGTATCTAGTACCGCGGCACC TATTTGCGGAAAAATACGATAAGATCATGG TGGACGGCAGAGCCAT	Forward 3C mutagenesis
$3D_{syn\_mut\_fwd}$	AGATACATACACAATGATTTCATATGGCGA TGATATAGTTGTAGCGAGCGACTACGACTT AGATTTTGAAGCGCTAAAACCTCATTTCAA GTCGCTGGGA	Forward 3D mutagenesis
$3D_{syn\_mut\_5'}$	CCACCACGATGTCGTCTCCGTATGAAATCA TTGTGTATGTATCTAGTTCTACTCCCTCATA GTGTCTACG	Forward 3D mutagenesis
$3D_{syn\_mut\_3'}$	GTCAGCTGGAGTGATGGTTTGGCCAAGGGA CTTGAAATGAGGTTTTAGAGCCTCGAAGTC CAAATC	Forward 3D mutagenesis
1-I TBR2	ACGAGTACCGGCGTCTCTT	Reverse 3D mutagenesis
8APN206	CACCCGAAGACCTTGAGAG	Reverse 2B mutagenesis
10PPN4	CGGTGGGTGCAAGCTTGG	Reverse 3C mutagenesis
13-NPN2	AAGTTTTACCGTCGTTCCCGACGTAAAAGG GAGGTAACCACAAGCTTGAA	Forward NGS primer
10-PPN30	TCTGGACAGCACCTTTGTCG	Reverse NGS primer
8-APN200	GAGACGTTGAGTCCAACCC	Forward NGS primer
13-NPN3	CCGTAGGAGTGAAAATCCCGAAAGGGTTTT TCCCGCTTCCTTAATCCAAA	Reverse NGS primer

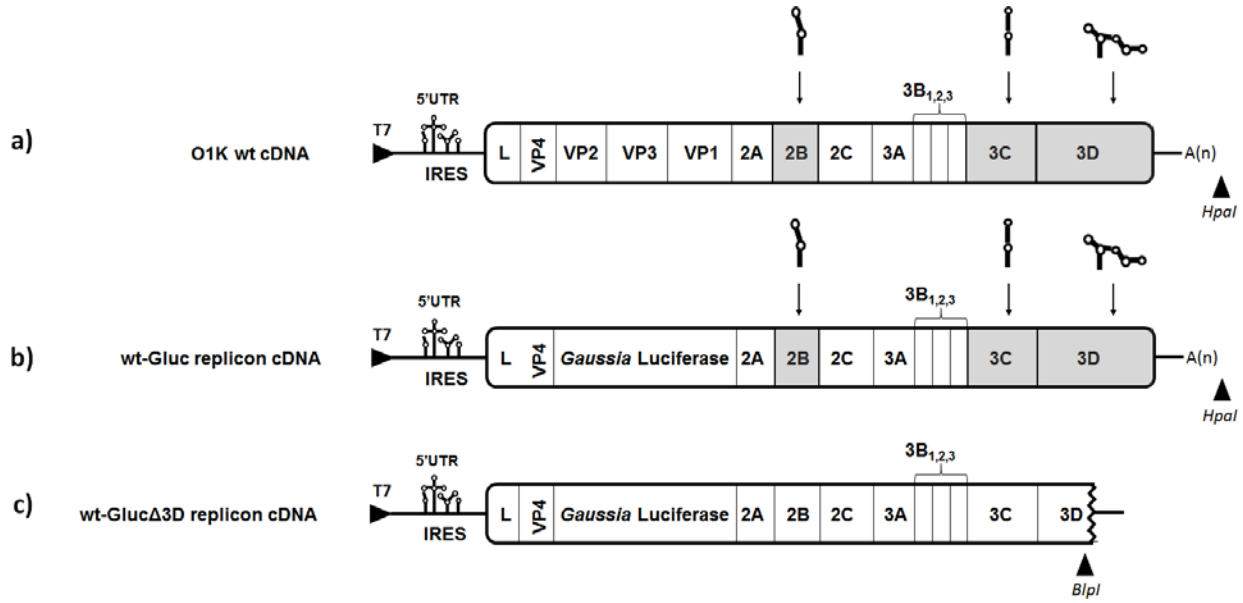

Figure 1: Schematic representation of the plasmids used in this study. These include: (a) full-length FMDV O1K cDNA, (b) *Gluc* replicon cDNA and (c) RNA polymerase defective *Gluc* replicon cDNA. All plasmids contain the T7 promoter and were linearized using either *HpaI* or *BlpI* prior to *in vitro* transcription and virus rescue. The locations of restriction sites used are marked *HpaI* and *BlpI.* The location of the three conserved RNA structures within the coding region of 2B, 3C and 3D (marked in grey) are indicated.

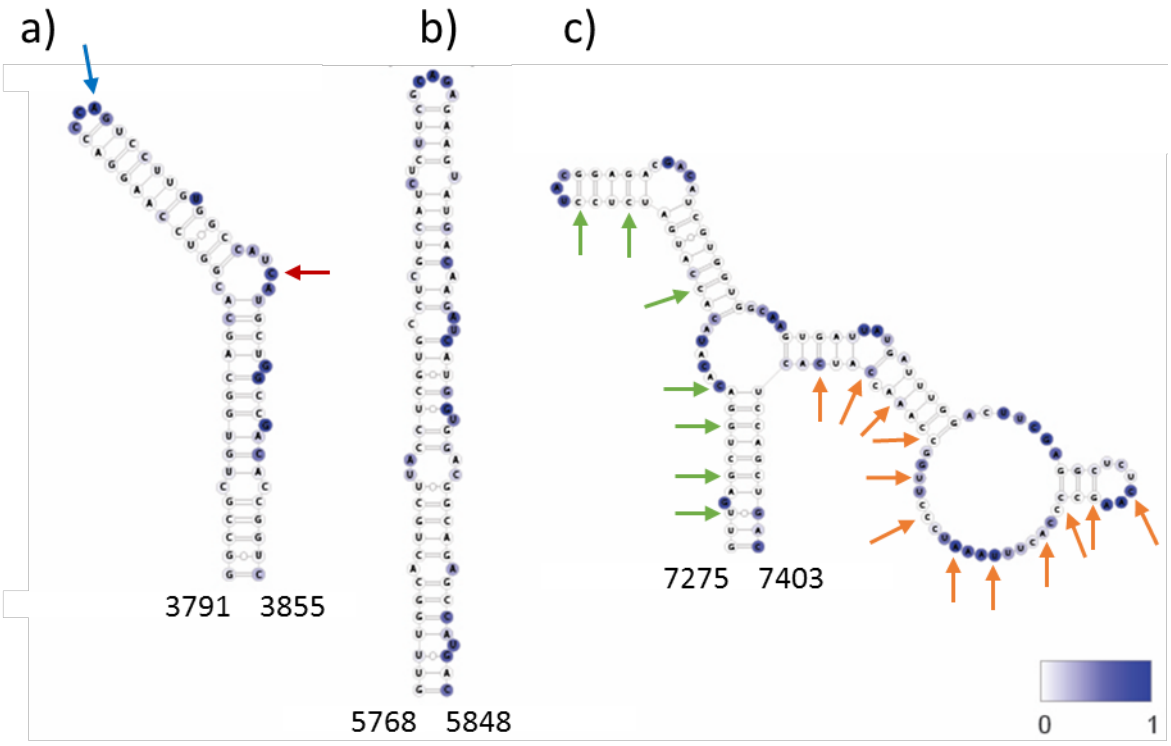

Figure 2: Previously predicted RNA structures, following SHAPE analysis, with low synonymous substitution rates (Poulsen, 2015) showing the introduced synonymous substitutions. These are located within the protein coding region of a) 2B, b) 3C and c) 3D. The SHAPE predicted model for each structure is shown with the nucleotide position in the genome (accession no. X00871). The nucleotides are coloured compared to their individual SHAPE selection reactivity. The arrows indicate the synonymous substitutions in the terminal loop (blue) and bulge (red) of the 2B structure and in the 5′- (green) and 3′-terminal end (orange) of 3D structure. $2B_{syn\_mut}$, $3C_{syn\_mut}$ and $3D_{syn\_mut}$ introduced synonymous substitutions covering the entire sequence length of each structure and are not marked. The nucleotides are coloured by SHAPE selection reactivity. Adapted from Poulsen, (2015).

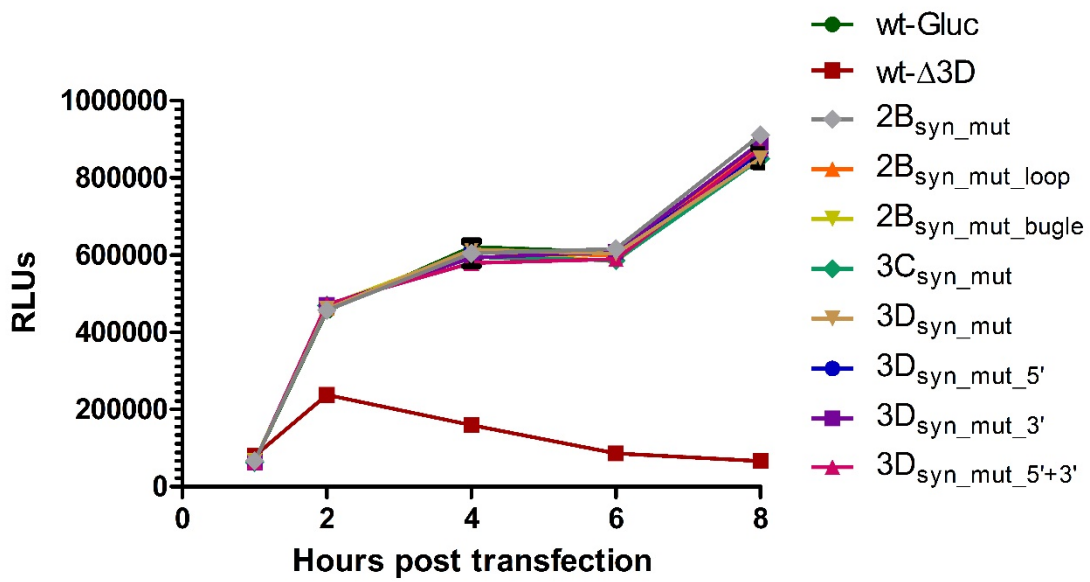

Figure 3: Expression of the secreted luciferase reporter protein, *Gluc*, by a FMDV replicon. BHK cells were transfected with RNA transcripts derived from the indicated cDNAs and, at the indicated times, lysates were prepared and assayed for *Gluc* activity. RLU = Relative light units. Data are represented as mean RLUs from a subset of samples (n=3) ± standard deviations (SD) over 8 hours.

Part 3 · Conclusions and perspectives

Conclusions and perspectives

The FMDV has a relatively small positive-sense single-stranded RNA genome like the other members of the *Picornaviridae* family. Many aspects of the FMDV viral life cycle remain unknown, but it has similarities to that of PV and relies on interactions with many cellular and viral elements (e.g. proteins and RNA structures). The FMDV genome includes a single, large, ORF encoding a polyprotein that is rapidly processed by the viral proteinases L^{pro} and $3C^{pro}$ and the StopGo inducing peptide 2A to generate multiple precursors and the mature proteins. This is a strategy which enables the virus to facilitate regulation of the polyprotein processing and several aspects of the FMDV life cycle. The RNA genome of FMDV folds into distinct three dimensional RNA structures, e.g. the IRES, which also have been found to regulate the life cycle of the virus. Many, if not all of these regulatory elements, are highly conserved due to their important role. Several conserved elements within the genome of FMDV have been identified. The aim of this thesis has been to investigate the influence of the conserved 2A peptide and three conserved RNA structures within the coding region for 2B, 3C and 3D on viral RNA replication and virus viability.

The FMDV 2A peptide located at the boundary between the structural proteins (capsid proteins) and the non-structural proteins (involved in RNA replication and polyprotein processing) is only 18 amino acid in length but is found to induce an unusual co-translational "cleavage" of the two regions. The majority of the reported work regarding the 2A peptide has been focused on characterizing the function of the peptide and identifying the residues essential for cleavage. This has primarily been achieved using mRNAs encoding artificial polyproteins proteins comprising two reporter proteins linked via the 2A sequence. Recently it has been established that for certain cardioviruses this cleavage event is not essential (Loughran *et al.*, 2013). The modification of the conserved $S^{15}N^{16}PG^{18\downarrow}P^{19}$ motif to $SNPL^{18}V^{19}$ at the 2A/2B junction blocks the cleavage but, surprisingly, it has no effect on the growth of TMEV while it is highly deleterious for replication of mengovirus (another cardiovirus) and lethal for FMDV. Manuscript 1 investigated the influence of the 2A peptide on polyprotein processing and virus replication of FMDV. Using a reverse genetics system based on using a full-length FMDV cDNA, it was established that certain amino acids substitutions in the 2A peptide that severely (60-70%) inhibit cleavage at the 2A/2B junction, as measured in *in vitro* assays (Donnelly *et al.*, 2001b; Sharma *et al.*, 2012), are tolerated within infectious FMDVs. In contrast, other

substitutions, previously found to inhibit cleavage by 89-100% *in vitro*, were not found in rescued viruses as they had reverted to the wt sequence. This was confirmed by marking the RNA transcripts with synonymous substitutions downstream of the 2A/2B junction. These results indicate that there is a selection pressure towards efficient cleavage activity at the 2A/2B junction. Furthermore, the absolute conservation of the 2A/2B cleavage motif in FMDV is not essential for viability as previously reported. The replication or translation of the FMDV genome was also found to be impaired in all the tested 2A mutants in the context of a FMDV replicon. The analysis of the 2A/2B "cleavage" in a transient expression assay within BHK cells using cDNA that encodes a truncated FMDV polyprotein, P1-2A-2BC-FLAG, thus free of the viral proteases showed that the 2A wt, S15A, S15I and S15F sequence results in complete cleavage at the 2A/2B junction. Some mutants were severely defective in "cleavage" by allowing the P1-2A-2BC-FLAG product to be detected while others generated a mix of cleaved and uncleaved product.

Manuscript 2 investigated which codons were tolerated at each of the four positions encoding the NPG$^{\downarrow}$P motif. Using highly degenerate primers, pools of modified plasmids were generated with all possible codons present for each of the amino acids in the NPG$^{\downarrow}$P motif individually. RNA transcripts were made from these pools and introduced into BHK cells and subsequently infectious virus was rescued. Sequencing of the recued viruses showed that they all encoded the NPG$^{\downarrow}$P motif, thus confirming the importance of this amino acid sequence. However, viruses with each of the 14 possible codons encoding these four residues were found within the rescued pool. Interestingly during subsequent passages in BHK cells, a distinct codon bias became apparent. This was in accordance with the codon bias previously identified within naturally occurring FMDVs (Gao *et al.*, 2014). Surprisingly, the codons selected are different for the two P residues; P^{17} is preferentially encoded by CCU while P^{19} is preferentially encoded by CCC. Interestingly, a single tRNA species recognizes both of these two codons, in both cattle and pigs which are the major hosts for FMDV. In addition, no gene for a prolyl tRNA with a GGG anticodon has found in the rat (close relationship with hamster). Gao et al., (2014) demonstrated that the synonymous codon usage for G^{18} did not influence the cleavage efficiency at the 2A/2B junction in an artificial polyprotein construct. It has been shown that the rate of peptide bond formation differs significantly depending on the Pro codon. The kinetics at codons CCA, CCU (P^{17}) and CCG are similar, but dipeptide synthesis at CCC (P^{19}) is twice as slow (Pavlov *et al.*, 2009) and could explain, at least in part, the codon bias. The generation of constructs with all four codons having synonymous (non wt) substitutions in the

context of a full-length FMDV cDNA and a P1-2A-2BC-FLAG expression cassette would help reveal whether the nucleotide composition at the 2A/2B junction also influences the cleavage activity and virus viability.

The eukaryotic translation factor eIF5A (eukaryotic initiation factor 5A) which contains the unique amino acid hypusine was originally identified as stimulating the peptide bond formation between a methionine charged tRNA and puromycin. It was thus concluded that eIF5A acted as a translation initiation factor. More recently, studies have identified eIF5A that facilitates polypeptide elongation, more precisely as a factor that overcomes stalling of ribosomes during synthesis of polyproline stretches (Gutierrez *et al.*, 2013; Schuller *et al.*, 2016). Furthermore, eIF5A has also been linked to a range of different processes such as tumorigenesis, cancer and virus replication. Some amino acids, proline in particular, are poor substrates for peptide bond formation and thus result in a lower rate of peptide bond formation (Pavlov *et al.*, 2009; Wohlgemuth *et al.*, 2008). It has been established that ribosomes stall when translating polyproline sequences and certain other "stalling motifs". eIF5A have been found to be important for the translation of proline rich motifs, e.g. PPP, (c.f. the NPG$^{\downarrow}$P motif in the 2A). The eIF5A serves to promote polypeptide elongation on stalled ribosomes that occurs when P-rich or G-rich sequences are being synthesized. It is suggested that eIF5A stimulates the peptidyl transferase activity of the ribosome and facilitates the reactivity of poor substrates like Pro. Hence, it can be hypothesized that eIF5A may be involved in the continuation of protein synthesis at the NPG$^{\downarrow}$P motif at the junction of FMDV 2A/2B. Doronina et al., (2008) demonstrated that the eukaryotic release factors eRF1 and eRF3 are essential for the StopGo function in a yeast system. However, Machida *et al.*, (2014) showed using a translation system reconstituted with human factors that the cleavage at the 2A/2B junction in EMCV does not require eRF1 and eRF3. Thus the role of these factors is still unclear. It could be interesting to investigate the effect of depleting cells of specific cellular proteins on the 2A/2B cleavage. In particular the level of eIF5A, eRF1 and eRF3 could be reduced within cells through the use of appropriate siRNAs or by the CRISPR/Cas9 method. Alternatively, the level of the functional eIF5A protein can be reduced by blocking the production of hypusine, a critical and unique component of eIF5A. Schuller *et al.*, (2016) also identified various tripeptide motifs associated with ribosome pausing in eIF5A depleted cells. Some 29 tripeptide motifs were calculated to have high propensity for ribosomal stalling in eIF5A depleted cells, but surprisingly, Pro or Asp were favoured in the E and the P sites and Pro in the A site, which is accordance with the PGP motif found in the conserved 2A NPG$^{\downarrow}$P motif. Clearly, the amino acid composition at

the NPG↓P motif is important. But it has been established in this thesis that some changes still permit cleavage. It would be interesting to investigate whether any of the other identified stall frameshifting motifs could induce cleave in the context of the 2A peptide.

Loughran *et al.*, (2011) have demonstrated a programmed -1 ribosomal frameshift into a conserved overlapping ORF within the 2B region of EMCV, just downstream of the 2A coding region. This result in the use of a distinct "trans-frame" protein termed 2B* and then termination of translation. Recently, Napthine *et al.*, (2017) showed that the frameshifting efficiency increases significantly during the later stages of infection and as such the production of the non-structural proteins associated with the virus replication is reduced. It was demonstrated that the EMCV 2A protein (ca. 16 kDa) serves as a trans-activator of this frameshifting by interacting with a stem-loop structure some 14 nt downstream of a slip site (GGUUUUU). Indeed, a U-rich motif (UUCUUUUUCU) can be found downstream of the 2A/2B junction in the FMDV genome, however other requirements necessary for this process to function appear to be absent. The EMCV is 150 amino acids long and possess a cluster of positively charged residues (R95-R97) which is suggested to bind to the stem-loop structure that is critical for the high frameshift efficiency. In contrast, the FMDV 2A peptide is only 18 amino acids long and lacks any such basic cluster. At present, there is no evidence for such a process within FMDV. Interestingly, is has previously been suggested that the 2A in FMDV could serve as a regulator to decrease the production of the non-structural proteins (Tulloch *et al.*, 2017). Sixty copies of each of the four structural proteins (VP1-VP4) are needed for the generation of a mature virion, albeit the FMDV genome only encodes a single copy of each capsid protein. An excess of the capsid proteins could prove beneficial to the virus during the late stages of the viral infection where new infectious virions assemble. Ribosomal profiling (as used by Napthine et al., 2017) which identifies translated RNA regions of wt and 2A mutant FMDV genomes could help to determine whether the FMDV 2A functions as a regulator of the expression levels of the structural and non-structural proteins.

For the past two decades over 100 recombinant proteins with therapeutic application have been successfully approved by the United States Food and Drug Administration (FDA) for the treatment of several human diseases (Zhu, 2012). The majority of these protein include immunoglobulins, cell receptors, interleukins, enzymes and transcription factors which primarily all consist of multimeric complexes which require the need for co-expression of multiple proteins from a single mRNA. This can be achieved through several approaches e.g. use of alternative initiation codons, proteolytic processing, ribosomal frame shifting, multiple

internal promoters, fusion proteins or the use of dicistronic mRNAs containing an internal ribosome entry site (IRES). These methods are each associated with several limitations and disadvantages. The use of multiple promoters can result in interference decreasing the level of expressed protein, whereas fused proteins can lead to loss of function due to steric effects. Post-translational cleavage can help remedy this, but the protease must be co-expressed in the same subcellular site. IRES elements vary in length and efficiency, within different cell types, due to their requirements for specific cellular trans-acting proteins. The presence of such cellular factors varies amongst different cell types. The two proteins produced can have their native sequences but they may not be made in equimolar amounts. In some IRES elements the downstream product is only expressed to ~10% of that of the upstream. The small FMDV 2A peptide could bypass this as it allows the near stoichiometric production of multiple, discrete, protein products from a single mRNA.

It has become increasingly evident that viral RNA genomes, like that of FMDV, form elaborate three dimensional structures through base pairing between complementary sequences. These structures can influence important aspects of the viral life cycle through long-range RNA-RNA interactions or through binding sites for viral or cellular proteins. The influence of three distinct RNA structures within the coding regions for 2B, 3C and 3D which previously has been determined through the use of SHAPE probing has been investigated in Manuscript 3. It was shown that synonymous substitutions within these sequences, designed to modify the structures, did not affect the replication of a FMDV replicon within BHK cells. The structure mutants were able to form viable infectious viruses and sequencing of rescued viruses (third passage in BHK cells) showed that the mutants had retained their mutations. Furthermore, NGS of the near complete genome of the 2B mutants did not show any secondary mutations indicating that the structures are not involved in RNA-RNA interactions.

Various distinct functional RNA structures within the genomes of viruses from different viral families have been described during the last two decades. The majority have so far been identified within the untranslated region of the genome where they are involved in intra-genomic interactions. Extensive RNA structures have also been identified within the coding regions. However, the exact function of these structures are still unknown. Some could be required for virus replication, like the cis-acting replication element, or for RNA packaging within new virus particles. Others could be important for protein binding to either separate RNA structures, for bringing the RNA structures together or to maintain RNA structure stability. Recently Logan *et al.*, (2017) found secondary structures within the coding region of FMDV with a similar conservation of the structure and a common motif at the loop. Disruption

of these structures resulted in attenuation of virus growth in BHK cells due to a reduction in assembly of mature virions, thus suggesting that these predicted RNA structures are involved in FMDV packaging. Simmonds *et al.*, (2004) suggested that RNA structures could be implicated in the evasion of cell defences and Kloc *et al.*, (2017) have provided evidence that changes in the FMDV S-fragment exert an effect in this way. Viability and replication studies in Manuscript 3 have been carried out in BHK cells which lack an intact interferon signalling system. Growth kinetics and the study of additional mutations using pBTY cells could reveal if these structures are important in this regard.

Abbreviations

BHK	Baby Hamster Kidney
bp	Base pair
CAT	Chloramphenicol acetyl-transferase
cDNA	Complimentary DNA
CHO	Chinese hamster ovary
cre	Cis-acting RNA element
DNA	Deoxyribonucleic acid
eIF4A	Eukaryotic translation initiation factor 4 A
eIF4E	Eukaryotic translation initiation factor 4E
eIF4G	Eukaryotic translation initiation factor 4 G
eIF5A	Eukaryotic translation initiation factor 5A-1
EMCV	Encephalomyocarditis virus
ER	Endoplasmic reticulum
eRF1	Eukaryotic release factor 1
eRF3	Eukaryotic release factor 3
Foot-and-mouth disease	FMD
Foot-and-mouth disease virus	FMDV
GFP	Green fluorescent protein
Gluc	Gaussia luciferase
HSPG	Heparan sulfate proteoglycan
HIV	Human immunodeficiency virus
IRES	Internal ribosome entry site
Kb	Kilobase
mRNA	Messenger RNA
NGS	Next generation sequencing
Nucleotide triphosphates	NTP
ORF	Open reading frame
PCR	Polymerase Chain Reaction
pfu	Plaque forming units
PCBP2	Poly(rC) binding protein 2
PV	Poliovirus
RdRp	RNA-dependent RNA polymerase
Ribonucleoprotein	RNP
Ribosomal RNA	rRNA
RLU	Relative light units
RNA	Ribonucleic acid
RT-qPCR	Reverse transcription quantitative PCR
SHAPE	Selective 2'-hydroxyl acylation analysed by primer extension
siRNA	Small interfering RNA
$TCID_{50}$	Tissue culture Infective Dose 50%

TMEV	Theiler’s murine encephalitis virus
tRNA	Transfer RNA
UK	United Kingdom
UTR	Untranslated region
wt	wild type
(+) ssRNA	Positive single stranded RNA

References

Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D. & Brown, F. (1989). The three-dimensional structure of foot-and-mouth disease virus at 2.9 A resolution. *Nature* **337**, 709–16.

Agirre, A., Barco, A., Carrasco, L. & Nieva, L. (2002). Viroporin-mediated Membrane Permeabilization. *J Biol Chem* **277**, 40434–40441.

Akashi, H. (2001). Gene expression and molecular evolution. *Curr Opin Genet Dev* **11**, 660–666.

Albariño, C. G., Uebelhoer, L. S., Vincent, J. P., Khristova, M. L., Chakrabarti, A. K., McElroy, A., Nichol, S. T. & Towner, J. S. (2013). Development of a reverse genetics system to generate recombinant Marburg virus derived from a bat isolate. *Virology* **446**, 230–237.

Ao, D., Guo, H. C., Sun, S. Q., Sun, D. H., Fung, T. S., Wei, Y. Q., Han, S. C., Yao, X. P., Cao, S. Z. & other authors. **(2015).** Viroporin activity of the foot-and-mouth disease virus non-structural 2B protein. *PLoS One* **10**, 1–18.

Atkins, J. F., Wills, N. M., Loughran, G., Wu, C.-Y., Parsawar, K., Ryan, M. D., Wang, C.-H. & Nelson, C. C. (2007). A case for 'StopGo': reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go). *RNA* **13**, 803–810.

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

Barton, D. J. & Flanegan, J. B. (1997). Synchronous Replication of Poliovirus RNA : Initiation of Negative-Strand RNA Synthesis Requires the Guanidine- Inhibited Activity of Protein 2C. *J Virol* **71**, 8482–8489.

Baxt, B. & Bachrach, H. L. (1980). Early interactions of foot-and-mouth disease virus with cultured cells. *Virology* **104**, 42–55.

Beard, C. W. & Mason, P. W. (2000). Genetic Determinants of Altered Virulence of Taiwanese Foot-and-Mouth Disease Virus. *J Virol* **74**, 987–991.

Belsham, G. J. (1993). Distinctive features of foot-and-mouth disease virus, a member of the

picornavirus family; aspects of virus protein synthesis, protein processing and structure. *Prog Biophys Mol Biol* **60**, 241–260.

Belsham, G. J. (2005). Translation and replication of FMDV RNA. *Curr Top Microbiol Immunol* **288**, 43–70.

Belsham, G. J. (2013). Influence of the Leader protein coding region of foot-and-mouth disease virus on virus replication. *J Gen Virol* **94**, 1486–95.

Belsham, G. J. & Bostock, C. J. (1988). Studies on the Infectivity of Foot-and-Mouth Disease Virus RNA using Microiujection. *J Gen Virol* **69**, 265–274.

Belsham, G. J. & Normann, P. (2008). Dynamics of picornavirus RNA replication within infected cells. *J Gen Virol* **89**, 485–493.

Belsham, G. J. & Sonenberg, N. (1996). RNA-Protein Interactions in Regulation of Picornavirus RNA Translation. *Microbiol Rev* **60**, 499–511.

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**, 272–280.

Belsham, G. J., Nielsen, I., Normann, P., Royall, E. & Roberts, L. O. (2008). Monocistronic mRNAs containing defective hepatitis C virus-like picornavirus internal ribosome entry site elements in their 5’ untranslated regions are efficiently translated in cells by a cap-dependent mechanism. *RNA* **14**, 1671–1680.

Berryman, S., Clark, S., Kakker, N. K., Silk, R., Seago, J., Wadsworth, J., Chamberlain, K., Knowles, N. J. & Jackson, T. (2013). Positively charged residues at the 5-fold symmetry axis of cell-culture adapted foot-and-mouth disease virus permit novel receptor interactions. *J Virol* **87**, 8735–8744.

Bienz, K., Egger, D., Troxler, M. & Pasamontest, L. (1990). Structural Organization of Poliovirus RNA Replication Is Mediated by Viral Proteins of the P2 Genomic Region. *J Virol* **64**, 1156–1163.

Birtley, J. R., Knox, S. R., Jaulent, A. M., Brick, P., Leatherbarrow, R. J. & Curry, S. (2005). Crystal structure of foot-and-mouth disease virus 3C protease: New insights into catalytic mechanism and cleavage specificity. *J Biol Chem* **280**, 11520–11527.

Brown, J. D. & Ryan, M. D. (2010). Ribosome 'Skipping': 'Stop-Carry On' or 'StopGo' Translation. In *Recoding Expans Decod Rules Enriches Gene Expr*, pp. 110–122. Edited by J. F. Atkins & R. F. Gesteland. Springer.

Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**, 897–907.

Burman, A., Clark, S., Abrescia, N., Fry, E., Stuart, D. & Jackson, T. (2006). Specificity of the VP1 GH Loop of Foot-and-Mouth Disease Virus for αv Integrins. *J Virol* **80**, 9798–9810.

Burrill, C. P., Westesson, O., Schulte, M. B., Strings, V. R. & Segal, M. (2013). Global RNA Structure Analysis of Poliovirus Identifies a Conserved RNA Structure Involved in Viral Replication and Infectivity **87**, 11670–11683.

Bøtner, A., Kakker, N. K., Barbezange, C., Berryman, S., Jackson, T. & Belsham, G. J. (2011). Capsid proteins from field strains of foot-and- mouth disease virus confer a pathogenic phenotype in cattle on an attenuated, cell-culture-adapted virus. *J Gen Virol* **92**, 1141–1151.

Cao, X. M., Bergmann, I. E., Fullkrug, R. & Beck, E. (1995). Functional analysis of the two alternative translation initiation sites of foot-and-mouth disease virus. *J Virol 69 p560-563* **69**, 560–563.

Carrillo, C., Tulman, E. R., Delhon, G., Lu, Z., Carreno, a, Vagnozzi, a, Kutish, G. F. & Rock, D. L. (2005). Comparative genomics of Foot-and-Mouth Disease Virus. *J Virol* **79**, 6487–6504.

Carrillo, E. C. & Giacheiti, C. (1984). Effect of lysosomotropic agents on the foot-and-mouth disease virus replication. *Virology* **545**, 542–545.

Chamberlain, K., Fowler, V. L., Barnett, P. V., Gold, S., Wadsworth, J., Knowles, N. J. & Jackson, T. (2015). Identification of a novel cell culture adaptation site on the capsid of foot-and-mouth disease virus. *J Gen Virol* **96**, 2684–2692.

Chow, M., Newman, J. F. E., Filman, D., Hogle, J. M., Rowlands, D. J. & Brown, F. (1987). Myristylation of picornavirus capsid protein VP4 and its structural significance. *Nature* **327**, 482–486.

Clarke, B. E., Brown, A. L., Currey, K. M., Newton, S. E., Rowlands, D. J. & Carroll, A.

R. (1987). Potential secondary and tertiary structure in the genomic RNA of foot and mouth disease virus. *Nucleic Acids Res* **15**, 7067–7079.

Costa Giomi, M. P., Bergmann, I. E., Scodeller, E. A., Auge de Mello, P., Gomez, I. & La Torre, J. L. (1984). Heterogeneity of the polyribocytidylic acid tract in aphthovirus: biochemical and biological studies of viruses carrying polyribocytidylic acid tracts of different lengths. *J Virol* **51**, 799–805.

Cottam, E. M., Wadsworth, J., Shaw, A. E., Rowlands, R. J., Goatley, L., Maan, S., Maan, N. S., Mertens, P. P. C., Ebert, K. & other authors. **(2008).** Transmission Pathways of Foot-and-Mouth Disease Virus in the United Kingdom in 2007. *PLOS Pathog* **4**, 1–8.

Crick, F. H. C. (1970). Central Dogma of Molecular Biology. *Nature* **227**, 561–563.

Curry, S., Abrams, C. C., Fry, E., Crowther, J. C., Belsham, G. J., Stuart, D. I. & King, A. M. Q. (1995). Viral-RNA Modulates the Acid Sensitivity of Foot-and-Mouth-Disease Virus Capsids. *J Virol* **69**, 430–438.

Curry, S., Fry, E., Blakemore, W., Abu-Ghazaleh, R., Jackson, T., King, A., Lea, S., Newman, J. & Stuart, D. (1997). Dissecting the roles of VP0 cleavage and RNA packaging in picornavirus capsid stabilization: the structure of empty capsids of foot-and-mouth disease virus. *J Virol* **71**, 9743–9752.

Davis, M., Sagan, S. M., Pezacki, J. P., Evans, D. J. & Simmonds, P. (2008). Bioinformatic and Physical Characterizations of Genome-Scale Ordered RNA Structure in Mammalian RNA Viruses. *J Virol* **82**, 11824–11836.

Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. (2009). Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci* **106**, 97–102.

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

Diaz-Toledano, R., Lozano, G. & Martinez-Salas, E. (2017). In-cell SHAPE uncovers dynamic interactions between the untranslated regions of the foot-and-mouth disease virus RNA. *Nucleic Acids Res* **45**, 1416–1432.

Diez, J., Mateu, M. G. & Domingo, E. (1989). Selection of antigenic variants of foot-and-mouth disease virus in the absense of antibodies, as revealed by an in situ assay. *J Gen*

Virol **70**, 3281–3289.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet* **30**, 418–426.

Doel, T. R. (2003). FMD vaccines. *Virus Res* **91**, 81–99.

Domingo, E. & Holland, J. J. (1997). RNA Virus Mutations and Fitness for Survival. *Annu Rev Microbiol* **51**, 151–78.

Domingo, E., Dfivila, M. & Ortin, J. (1980). Nucleotide sequence heterogeneity of the RNA from a natural population of Foot-and-mouth disease virus. *Gene* **11**, 333–346.

Domingo, E., Diez, J., Martinez, M. A., Herndndez, J., Holguin, A., Borrego, B. & Mateu, M. G. (1993). New observations on antigenic diversification of RNA viruses. Antigenic variation is not dependent on immune selection. *J Gen Virol* 2039–2045.

Domingo, E., Escarmís, C., Baranowski, E., Ruiz-Jarabo, C. M., Carrillo, E., Núñez, J. I. & Sobrino, F. (2003). Evolution of foot-and-mouth disease virus. *Virus Res* **91**, 47–63.

Domingo, E., Sheldon, J., Perales, C., Fitness, V., Gain, F., Equilibrium, P., Stasis, A., Selection, P. & Tropism, C. (2012). Viral Quasispecies Evolution **76**, 159–216.

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

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

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

Doronina, V. A., Wu, C., de Felipe, P., Sachs, M. S., Ryan, M. D. & Brown, J. D. (2008). Site-specific release of nascent chains from ribosomes at a sense codon. *Mol Cell Biol* **28**, 4227–4239.

Drake, J. W. & Holland, J. J. (1999). Mutation rates among RNA viruses. *Proc Natl Acad*

Sci U S A **96**, 13910–3.

Drake, J. W. (1993). Rates of spontaneous mutation among RNA viruses. *Proc Natl Acad Sci* **90**, 4171–4175.

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

Duque, H. & Baxt, B. (2003). Foot-and-Mouth Disease Virus Receptors : Comparison of Bovine αv Integrin Utilization by Type A and O Viruses. *J Virol* **77**, 2500–2511.

Ellard, F. M., Drew, J., Blakemore, W. E., Stuart, D. I. & King, A. M. Q. (1999). Evidence for the role of His 142 of protein 1C in the acid induced disassembly of foot and mouth disease virus capsids. *J Gen Virol* **80**, 1911–1918.

Escarmis, C., Davila, M., Charpentier, N., Bracho, A., Moya, A. & Domingo, E. (1996). Genetic lesions associated with Muller's ratchet in an RNA virus. *J Mol Biol* **264**, 255–267.

Escarmís, C., Dopazo, J., Dávila, M., Palma, E. L. & Domingo, E. (1995). Large deletions in the 5′-untranslated region of foot-and-mouth disease virus of serotype C. *Virus Res* **35**, 155–167.

Evans, D. J. & Almond, J. W. (1998). Cell receptors for picornaviruses as determinants of cell tropism and pathogenesis. *Trends Microbiol* **6**, 198–200.

Falk, M. M., Grigera, P. R., Bergmann, I. E., Zibert, A., Multhaup, G. & Beck, E. (1990). Foot-and-mouth disease virus protease 3C induces specific proteolytic cleavage of host cell histone H3. *J Virol* **64**, 748–756.

Fares, M. A., Moya, a, Escarmís, C., Baranowski, E., Domingo, E. & Barrio, E. (2001). Evidence for positive selection in the capsid protein-coding region of the foot-and-mouth disease virus (FMDV) subjected to experimental passage regimens. *Mol Biol Evol* **18**, 10–21.

Fariza, T., Van, T. D. T. & Boucheham, A. (2017). In Silico Prediction of RNA Secondary Structure. In *Promot Assoc RNA*, pp. 145–168. Edited by S. Napoli. Springer Nature.

Fernández, N., García-Sacristán, A., Ramajo, J., Briones, C. & Martínez-Salas, E. (2011). Structural analysis provides insights into the modular organization of picornavirus IRES.

Virology **409**, 251–261.

Ferrer-Orta, C., Arias, A., Perez-Luque, R., Escarmís, C., Domingo, E. & Verdaguer, N. (2004). Structure of foot-and-mouth disease virus RNA-dependent RNA polymerase and its complex with a template-primer RNA. *J Biol Chem* **279**, 47212–47221.

Ferrer-Orta, C., Arias, A., Escarmís, C. & Verdaguer, N. (2006). A comparison of viral RNA-dependent RNA polymerases. *Curr Opin Struct Biol* **16**, 27–34.

Filipowicz, W., Jaskiewicz, L., Kolb, F. A. & Pillai, R. S. (2005). Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr Opin Struct Biol* **15**, 331–341.

Forss, S. & Schaller, H. (1982). A tandem repeat gene in a picornavirus. *Nucleic Acids Res* **10**, 6441–6450.

Fox, G., Parry, N. R., Barnett, P. V., McGinn, B., Rowlands, D. J. & Brown, F. (1989). The Cell Attachment Site on Foot-and-Mouth Disease Virus Includes the Amino Acid Sequence RGD (Arginine-Glycine-Aspartic Acid). *J Gen Virol* **70**, 625–637.

Fry, E. E., Newman, J. W. I., Curry, S., Najjam, S., Jackson, T., Blakemore, W., Lea, S. M., Miller, L., Burman, A. & other authors. **(2005).** Structure of Foot-and-mouth disease virus serotype A1061 alone and complexed with oligosaccharide receptor: Receptor conservation in the face of antigenic variation. *J Gen Virol* **86**, 1909–1920.

Fuerst, T. R., Niles, E. G., Studier, F. W. & Moss, B. (1986). Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proc Natl Acad Sci U S A* **83**, 8122–8126.

Gamarnik, A. V. & Andino, R. (1998). Switch from translation to RNA replication in a positive-stranded RNA virus. *Genes Dev* **12**, 2293–2304.

Gao, Y., Sun, S.-Q. & Guo, H.-C. (2016). Biological function of Foot-and-mouth disease virus non-structural proteins and non-coding elements. *Virol J* **13**, 107.

Gao, Z., Zhou, J., Zhang, J., Ding, Y. & Liu, Y. (2014). The silent point mutations at the cleavage site of 2A/2B have no effect on the self-cleavage activity of 2A of foot-and-mouth disease virus. *Infect Genet Evol* **28**, 101–106.

García-Briones, M., Rosas, M. F., González-Magaldi, M., Martín-Acebes, M. A., Sobrino, F. & Armas-Portela, R. (2006). Differential distribution of non-structural proteins of

foot-and-mouth disease virus in BHK-21 cells. *Virology* **349**, 409–421.

Garst, A. D., Edwards, A. L., Batey, R. T., Garst, A. D., Edwards, A. L. & Batey, R. T. (2011). Riboswitches: Structures and Mechanisms. *Cold Spring Harb Perspect Biol* **6**, 1–14.

Giegé, R., Jühling, F., Pütz, J., Stadler, P., Sauter, C. & Florentz, C. (2012). Structure of transfer RNAs: Similarity and variability. *WiREs RNA* **3**, 37–61.

Gilbert, W. (1986). The RNA world. *Nature* **319**, 618.

Goodfellow, I., Chaudhry, Y., Richardson, A., Meredith, J., Almond, J. W., Barclay, W. & Evans, D. J. (2000). Identification of a cis-acting replication element within the poliovirus coding region. *J Virol* **74**, 4590–600.

Gorbalenya, A. E. & Koonin, E. V. (1989). Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Res* **17**, 8413–8438.

Gradi, A., Foeger, N., Strong, R., Yuri, V., Sonenberg, N., Skern, T., Graham, J., Svitkin, Y. V & Belsham, G. J. (2004). Cleavage of Eukaryotic Translation Initiation Factor 4GII within Foot-and-Mouth Disease Virus-Infected Cells : Identification of the L-Protease Cleavage Site In Vitro Cleavage of Eukaryotic Translation Initiation Factor 4GII within Foot-and-Mouth Disease. *J Virol* **78**, 3271–3278.

Groppo, R. & Palmenberg, A. C. (2007). Cardiovirus 2A protein associates with 40S but not 80S ribosome subunits during infection. *J Virol* **81**, 13067–13074. American Society for Microbiology.

Grubman, M. J. & Baxt, B. (2004). Foot-and-Mouth Disease. *Clin Microbiol Rev* **17**, 465–493.

Gullberg, M., Muszynski, B., Organtini, L. J., Ashley, R. E., Hafenstein, S. L., Belsham, G. J. & Polacek, C. (2013a). Assembly and characterization of foot-and-mouth disease virus empty capsid particles expressed within mammalian cells. *J Gen Virol* **94**, 1769–1779.

Gullberg, M., Polacek, C., Bøtner, A. & Belsham, G. J. (2013b). Processing of the VP1/2A junction is not necessary for production of foot-and-mouth disease virus empty capsids and infectious viruses: characterization of ‘self-tagged’ particles. *J Virol* **87**, 11591–11603.

Gutierrez, E., Shin, B. S., Woolstenhulme, C. J., Kim, J. R., Saini, P., Buskirk, A. R. & Dever, T. E. (2013). eIF5A promotes translation of polyproline motifs. *Mol Cell* **51**, 35–45.

Hahn, H. & Palmenberg, A. C. (1995). Encephalomyocarditis viruses with short poly(C) tracts are more virulent than their mengovirus counterparts. *J Virol* **69**, 2697–9.

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

Harris, T. J. R. & Brown, F. (1977). Biochemical analysis of a virulent and an avirulent strain of foot-and-mouth disease virus. *J Gen Virol* **34**, 87–105.

Haydon, D. T., Samuel, A. R. & Knowles, N. J. (2001a). The generation and persistence of genetic variation in foot-and-mouth disease virus. *Prev Vet Med* **51**, 111–124.

Haydon, D. T., Bastos, D., Knowles, N. J. & Samuel, R. (2001b). Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. *Genetics* **157**, 7–15.

He, D. S., Li, K. N., Lin, X. M., Lin, S. R., Su, D. P. & Liao, M. (2011). Genomic comparison of foot-and-mouth disease virus R strain and its chick-passaged attenuated strain. *Vet Microbiol* **150**, 185–190.

Herod, M. R., Loundras, E.-A., Ward, J. C., Tulloch, F., Rowlands, D. J. & Stonehouse, N. J. (2015). Employing transposon mutagenesis to investigate foot-and-mouth disease virus replication. *J Gen Virol* **96**, 3507–3518.

Herod, M. R., Ferrer-Orta, C., Loundras, E.-A., Ward, J. C., Verdaguer, N., Rowlands, D. J. & Stonehouse, N. J. (2016). Both *Cis*- and *Trans*- Activities of the Foot-and-Mouth Disease Virus 3D Polymerase are Essential for Viral RNA Replication. *J Virol* **90**, JVI.00469-16.

Herod, M. R., Gold, S., Lasecka-dykes, L., Wright, C., Ward, J. C., Mclean, T. C., Forrest, S., Jackson, T., Tuthill, T. J. & other authors. (2017). Genetic economy in picornaviruses : Foot-and- mouth disease virus replication exploits alternative precursor cleavage pathways. *PLOS Pathog* 1–25.

Herold, J. & Andino, R. (2000). Poliovirus requires a precise 5' end for efficient positive-strand RNA synthesis. *J Virol* **74**, 6394–6400.

Jackson, T., Ellard, F. M., Ghazaleh, R. A. B. U., Brookes, S. M., Blakemore, W. E., Corteyn, A. H., Stuart, D. I., Newman, J. W. I. & King, A. M. Q. (1996). Efficient Infection of Cells in Culture by Type O Foot-and-Mouth Disease Virus Requires Binding to Cell Surface Heparan Sulfate. *J Virol* **70**, 5282–5287.

Jackson, T., Sheppard, D., Denyer, M., Blakemore, W. & King, A. M. Q. (2000). The Epithelial Integrin αvβ6 Is a Receptor for Foot-and-Mouth Disease Virus. *J Virol* **74**, 4949–4956.

Jackson, T., Mould, A. P., Sheppard, D. & King, A. M. Q. (2002). Integrin αvβ1 Is a Receptor for Foot-and-Mouth Disease Virus. *J Virol* **76**, 935–941.

Jackson, T., Clark, S., Berryman, S., Burman, A., Cambier, S., Mu, D., Nishimura, S. & King, A. M. Q. (2004). Integrin αvβ8 Functions as a Receptor for Foot-and-Mouth Disease Virus : Role of the β-Chain Cytodomain in Integrin-Mediated Infection. *J Virol* **78**, 4533–4540.

Jacob, F. & Monod, J. (1960). Genetic Regulatory Mechanisms in the Synthesis of Proteins. *J Mol Biol* **3**, 318–35.

Jamal, S. M. & Belsham, G. J. Molecular epidemiology, evolution and phylogeny of foot-and-mouth disease virus. *Infect Genet Evol* . (Submitted).

Jamal, S. M. & Belsham, G. J. (2013). Foot-and-mouth disease: past, present and future. *Vet Res* **44**, 116–129.

James, B. D., Olsen, G. J. & Pace, N. R. (1989). Phylogenetic comparative analysis of RNA secondary structure. *Methods Enzymol* **180**, 227–239.

Khromykh, A. (2000). Replicon-based vectors of positive strand RNA viruses. *Curr Opin Mol Ther* **2**, 555–559.

Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L. & Wang, A. H. J. (1974). Three-Dimensional Tertiary Structure of Yeast Phenylalanine Transfer RNA. *Am Assoc Adv Sci Sci* **185**, 435–440.

King, A. M. Q., McCahon, D., Saunders, K., Newman, J. W. I. & Slade, W. R. (1985). Multiple sites of recombination within the RNA genome of foot-and-mouth disease virus. *Virus Res* **3**, 373–384.

King, A. M. Q., Adams, M. J., Carstens, E. B. & Lefkowitz, E. J. (Eds.). **(2011).** Virus Taxonomy Classification and Nomenclature of Viruses. Report of the International Committee on Taxonomy of Viruses. ICTV.

Kirchweger, R., Ziegler, E., Lamphear, B. J., Waters, D., Liebig, H. D., Sommergruber, W., Sobrino, F., Hohenadl, C., Blaas, D. & Rhoads, R. E. (1994). Foot-and-mouth disease virus leader proteinase: purification of the Lb form and determination of its cleavage site on eIF-4 gamma. *J Virol* **68**, 5677–5684.

Kjær, J. & Belsham, G. J. (2017). Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias. *RNA* **(in press)**. e-pub October 2017 doi:10.1261/rna.063339.117

Kloc, A., Segundo, F. D., Schafer, E. A., Rai, D. K., Kenney, M., Santos, T. D. L. & Rieder, E. (2017). Foot-and-mouth disease virus 5' -terminal S fragment is required for replication and modulation of the innate immune response in host cells. *Virology* **512**, 132–143.

Knipe, T., Rieder, E., Baxt, B., Ward, G. & Mason, P. W. (1997). Characterization of Synthetic Foot-and-Mouth Disease Virus Provirions Separates Acid-Mediated Disassembly from Infectivity. *J Virol* **71**, 2851–2856.

Knowles, N. J. & Samuel, A. R. (2003). Molecular epidemiology of foot-and-mouth disease virus. *Virus Res* **91**, 65–80.

Knowles, N. J., Davies, P. R., Henry, T., O'Donnell, V., Pacheco, J. M. & Mason, P. W. (2001). Emergence in Asia of foot-and-mouth disease viruses with altered host range: characterization of alterations in the 3A protein. *J Virol* **75**, 1551–1556.

Knowles, N. J., Hovi, T., King, D. P. & Stanway, G. (2010). Overview of Taxonomy. In *The Picornaviruses*, pp. 19–32. Edited by E. Ehrenfeld & E. Domingo. Wasingtong DC: American Society for Microbiology (ASM).

Knox, C., Moffat, K., Ali, S., Ryan, M., Wileman, T. & Wileman, T. (2005). Foot-and-mouth disease virus replication sites form next to the nucleus and close to the Golgi apparatus, but exclude marker proteins associated with host membrane compartments. *J Gen Virol* **86**, 687–696.

Korboukh, V. K., Lee, C. A., Acevedo, A., Vignuzzi, M., Xiao, Y., Arnold, J. J., Hemperly, S., Graci, J. D., August, A. & other authors. **(2014).** RNA Virus Population Diversity,

an Optimum for Maximal Fitness and Virulence. *J Biol Chem* **289**, 29531–29544.

Kristensen, T., Normann, P., Gullberg, M., Fahnøe, U., Polacek, C., Rasmussen, T. B. & Belsham, G. J. (2017). Determinants of the VP1/2A junction cleavage by the 3C protease in foot-and-mouth disease virus-infected cells. *J Gen Virol* **98**, 385–395.

Kubo, M. & Imanaka, T. (1989). mRNA secondary structure in an open reading frame reduces translation efficiency in Bacillus subtilis. *J Bacteriol* **171**, 4080–4082.

Van Kuppeveld, F. J. M. & Ehrenfeld, E. (2010). Remodelling Cellular Membranes. In *The Picornaviruses*, pp. 181–193. Edited by E. Domingo & R. Roos. American Society for Microbiology (ASM).

Van Kuppeveld, F. J. M., De Jong, A. S., Melchers, W. J. G. & Willems, P. H. G. M. (2005). Enterovirus protein 2B po(u)res out the calcium: A viral strategy to survive? *Trends Microbiol* **13**, 41–44.

Kühn, R., Luz, N. & Beck, E. (1990). Functional analysis of the internal translation initiation site of foot-and-mouth disease virus. *J Virol* **64**, 4625–31.

Lawrence, P. & Rieder, E. (2009). Identification of RNA helicase A as a new host factor in the replication cycle of foot-and-mouth disease virus. *J Virol* **83**, 11356–11366.

Liu, G., Liu, Z., Xie, Q., Chen, Y., Bao, H., Chang, H. & Liu, X. (2004). Generation of an infectious cDNA clone of an FMDV strain isolated from swine. *Virus Res* **104**, 157–164.

Logan, G., Newman, J., Wright, C. F., Lasecka-Dykes, L., Haydon, D. T., Cottam, E. M. & Tuthill, T. J. (2017). Deep sequencing of foot-and-mouth disease virus reveals RNA sequences involved in genome packaging. *J Virol* . (in press). doi: e-pub October 2017 10.1128/JVI.

López de Quinto, S., Sáiz, M., Morena, D., Sobrino, F. & Martínez-Salas, E. (2002). IRES-driven translation is stimulated separately by the FMDV 3'NCR and poly(A) sequences. *Nucleic Acids Res* **30**, 4398–4405.

Loughran, G., Libbey, J. E., Uddowla, S., Scallan, M. F., Ryan, M. D., Fujinami, R. S., Rieder, E. & Atkins, J. F. (2013). Theiler's murine encephalomyelitis virus contrasts with encephalomyocarditis and foot-and-mouth disease viruses in its functional utilization of the StopGo non-standard translation mechanism. *J Gen Virol* **94**, 348–353.

Loughran, G., Firth, A. E. & Atkins, J. F. (2011). Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus genome. *Proc Natl Acad Sci* **108**, 1111–1119.

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**, 1036–1042.

Machida, K., Mikami, S., Masutani, M., Mishima, K., Kobayashi, T. & Imataka, H. (2014). A translation system reconstituted with human factors proves that processing of encephalomyocarditis virus proteins 2A and 2B occurs in the elongation phase of translation without eukaryotic release factors. *J Biol Chem* **289**, 31960–31971.

Manicassamy, B., Manicassamy, S., Belicha-Villanueva, A., Pisanelli, G., Pulendran, B. & Garcia-Sastre, A. (2010). Analysis of in vivo dynamics of influenza virus infection in mice using a GFP reporter virus. *Proc Natl Acad Sci* **107**, 11531–11536.

Manrubia, S. C. & Lázaro, E. (2006). Viral evolution. *Phys Life Rev* **3**, 65–92.

Martin, L. R. & Palmenberg, A. C. (1996). Tandem mengovirus 5' pseudoknots are linked to viral RNA synthesis, not poly(C)-mediated virulence. *J Virol* **70**, 8182–6.

Martinez-Salas, E. & Belsham, G. J. (2017). Genome Organisation, Translation and Replication of Foot-and-mouth Disease Virus RNA. In *Foot-and-Mouth Dis Curr Res Emerg Trends*, pp. 13–42. Edited by F. Sobrino & E. Domingo. Caister Academic Press.

Mason, P. W., Bezborodova, S. V, Tina, M. & Henry, T. M. (2002). Identification and Characterization of a cis-Acting Replication Element (cre) Adjacent to the Internal Ribosome Entry Site of Foot-and-Mouth Disease Virus. *J Virol* **76**, 9686–9694.

Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**, 911–940.

Mauro, V. P. & Chappell, S. A. (2014). A critical analysis of codon optimization in human therapeutics. *Trends Mol Med* **20**, 604–613.

McCahon, D., King, A. M., Roe, D. S., Slade, W. R., Newman, J. W. & Cleary, A. M. (1985). Isolation and biochemical characterization of intertypic recombinants of foot-and-mouth disease virus. *Virus Res* **3**, 87–100.

McInerney, G. M., King, A. M. Q., Ross-Smith, N. & Belsham, G. J. (2000). Replication-competent foot-and-mouth disease virus RNAs lacking capsid coding sequences. *J Gen Virol* **81**, 1699–1702.

McKnight, K. L. & Lemon, S. M. (1996). Capsid Coding Sequence Is Required for Efficient Replication of Human Rhinovirus 14 RNA. *J Virol* **70**, 1941–1952.

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

Minskaia, E., Nicholson, J. & Ryan, M. D. (2013). Optimisation of the foot-and-mouth disease virus 2A co-expression system for biomedical applications. *BMC Biotechnol* **13**, 67.

Moffat, K., Knox, C., Howell, G., Clark, S. J., Yang, H., Belsham, G. J., Ryan, M. & Wileman, T. (2007). Inhibition of the Secretory Pathway by Foot-and-Mouth Disease Virus 2BC Protein Is Reproduced by Coexpression of 2B with 2C, and the Site of Inhibition Is Determined by the Subcellular Location of 2C. *J Virol* **81**, 1129–1139.

Moffat, K., Howell, G., Knox, C., Graham, J., Monaghan, P., Ryan, M. D., Belsham, G. J., Wileman, T., Monaghan, P. & other authors. (2005). Effects of Foot-and-Mouth Disease Virus Nonstructural Proteins on the Structure and Function of the Early Secretory Pathway : 2BC but Not 3A Blocks Endoplasmic Reticulum-to-Golgi Transport Effects of Foot-and-Mouth Disease Virus Nonstructural Proteins on. *Society* **79**, 4382–4395.

Monaghan, P., Cook, H., Jackson, T., Ryan, M. & Wileman, T. (2004). The ultrastructure of the developing replication site in foot-and-mouth disease virus-infected BHK-38 cells. *J Gen Virol* **85**, 933–946.

Moore, D. M. & Cowan, K. M. (1978). Effect of trypsin and chymotrypsin on the polypeptides of large and small plaque variants of foot-and-mouth disease virus: Relationship to specific antigenicity and infectivity. *J Gen Virol* **41**, 549–562.

Mueller, S. & Wimmer, E. (1998). Expression of foreign proteins by poliovirus polyprotein fusion: analysis of genetic stability reveals rapid deletions and formation of cardioviruslike open reading frames. *J Virol* **72**, 20–31.

Napthine, S., Ling, R., Finch, L. K., Jones, J. D., Bell, S., Brierley, I. & Firth, A. E. (2017).

Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nat Commun* **8**, 15582.

Nayak, A., Goodfellow, I. G. & Belsham, G. J. (2005). Factors required for the Uridylylation of the foot-and-mouth disease virus 3B1, 3B2, and 3B3 peptides by the RNA-dependent RNA polymerase (3Dpol) in vitro. *J Virol* **79**, 7698–7706.

Nayak, A., Goodfellow, I. G., Woolaway, K. E., Birtley, J., Curry, S. & Belsham, G. J. (2006). Role of RNA structure and RNA binding activity of foot-and-mouth disease virus 3C protein in VPg uridylylation and virus replication. *J Virol* **80**, 9865–9875.

Nieva, J. L., Madan, V. & Carrasco, L. (2012). Viroporins: structure and biological functions. *Nat Rev Microbiol* **10**, 563–574.

Novak, J. E. & Kirkegaard, K. (1991). Improved Method for Detecting Poliovirus Negative Strands Used To Demonstrate Specificity of Positive-Strand Encapsidation and the Ratio of Positive to Negative Strands in Infected Cells. *J Virol* **65**, 3384–3387.

Novoa, E. M. & Ribas de Pouplana, L. (2012). Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends Genet* **28**, 574–581.

Nugent, C. I., Johnson, K. L., Sarnow, P. & Kirkegaard, K. (1999). Functional coupling between replication and packaging of poliovirus replicon RNA. *J Virol* **73**, 427–435.

Núñez, J. I., Baranowski, E. & Molina, N. (2001). A Single Amino Acid Substitution in Nonstructural Protein 3A Can Mediate Adaptation of Foot-and-Mouth Disease Virus to the Guinea Pig A Single Amino Acid Substitution in Nonstructural Protein 3A Can Mediate Adaptation of Foot-and-Mouth Disease Virus to th. *J Virol* **75**, 3977–3983.

Pacheco, J. M., Henry, T. M., Donnell, V. K. O., Gregory, J. B. & Mason, P. W. (2003). Role of Nonstructural Proteins 3A and 3B in Host Range and Pathogenicity of Foot-and-Mouth Disease Virus Role of Nonstructural Proteins 3A and 3B in Host Range and Pathogenicity of Foot-and-Mouth Disease Virus. *J Virol* **77**, 13017–13027.

Pacheco, J. M., Gladue, D. P., Holinka, L. G., Arzt, J., Bishop, E., Smoliga, G., Pauszek, S. J., Bracht, A. J., Donnell, V. O. & other authors. **(2013).** A partial deletion in non-structural protein 3A can attenuate foot-and-mouth disease virus in cattle. *Virology* **446**, 260–267. Elsevier.

Palmenberg, A. C., Parks, G. D., Hall, D., Ingraham, R. H., Seng, T. W. & Pallal, P. V.

(1992). Proteolytic Processing of the Cardioviral Cleavage in Clone-Derived P2 Region : Primary 2A / 2B Precursors. *Virology* **190**, 754–762.

Paton, D. J., Sumption, K. J. & Charleston, B. (2009). Options for control of foot-and-mouth disease: knowledge, capability and policy. *Philos Trans R Soc London* **364**, 2657–2667.

Paul, A. V & Wimmer, E. (2015). Initiation of protein-primed picornavirus RNA synthesis. *Virus Res* **206**, 12–26.

Paul, A. V, van Boom, H. J., Filippov, D. & Wimmer, E. (1998). Protein-primed RNA synthesis by purified poliovirus RNA polymerase. *Nature* **393**, 280–284.

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**, 50–54.

Piccone, M. E., Rieder, E., Mason, P. W. & Grubman, M. J. (1995). The foot-and-mouth disease virus leader proteinase gene is not required for viral replication. *J Virol* **69**, 5376–5382.

Pilipenko, E. V., Blinov, V. M., Chernov, B. K., Dmitrieva, T. M. & Agol, V. I. (1989). Conservation of the secondary structure elements of the 5’ untranslated region of cardio- and aphthovirus RNAs. *Nucleic Acids Res* **17**, 5701–5711.

Polacek, C., Gullberg, M., Li, J. & Belsham, G. J. (2013). Low levels of foot-and-mouth disease virus 3C protease expression are required to achieve optimal capsid protein expression and processing in mammalian cells. *J Gen Virol* **94**, 1249–58.

Poulsen, L. D. (2015). Investigation of RNA structure by high-throughput SHAPE-based probing methods. PhD thesis. University of Copenhagen.

Poulsen, L. D., Kielpinski, L. J., Salama, S. R., Krogh, A. & Vinther, J. (2015). SHAPE Selection (SHAPES) enrich for RNA structure signal in SHAPE sequencing-based probing data. *RNA* **21**, 1042–52.

Reed, L. J. & Muench, H. (1938). A simple method of estimating fifty percent endpoints. *Am J Hyg* **27**, 493–497.

Rieder, E., Bunch, T., Brown, F. & Peter Mason, And. W. (1993). Genetically Engineered Foot-and-Mouth Disease Viruses with Poly(C) Tracts of Two Nucleotides Are Virulent

in Mice. *J Virol* **67**, 5139–5145.

Risager, P. C., Fahnøe, U., Gullberg, M., Rasmussen, T. B. & Belsham, G. J. (2013). Analysis of classical swine fever virus RNA replication determinants using replicons. *J Gen Virol* **94**, 1739–1748.

Robertson, B. H., Morgan, D. O., Moore, D. M., Grubman, M. J., Card, J., Fischer, T., Weddell, G., Dowbenko, D. & Yansura, D. (1983a). Identification of amino acid and nucleotide sequence of the foot-and-mouth disease virus RNA polymerase. *Virology* **126**, 614–623.

Robertson, B. H., Moore, D. M., Grubman, M. J. & Kleid, D. G. (1983b). Identification of an exposed region of the immunogenic capsid polypeptide VP1 on foot-and-mouth disease virus. *J Virol* **46**, 311–6.

Robertus, J. D. & Ladner, J. E. (1974). Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **185**, 435–440.

Rogers, G. W., Richter, N. J. & Merrick, W. C. (1999). Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J Biol Chem* **274**, 12236–12244.

Roulston, C., Luke, G. A., de Felipe, P., Ruan, L., Cope, J., Nicholson, J., Sukhodub, A., Tilsner, J. & Ryan, M. D. (2016). '2A-Like' Signal Sequences Mediating Translational Recoding: A Novel Form of Dual Protein Targeting. *Traffic* **17**, 923–939.

Rweyemamu, M., Roeder, P., Mackay, D., Sumption, K., Brownlie, J., Leforban, Y. & Valarcher, J. (2008). Epidemiological Patterns of Foot-and-Mouth Disease Worldwide. *Transbound Emerg Dis* **55**, 57–72.

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

Ryan, M. D., Donnelly, M., Lewis, A., Mehrotra, A. P., Wilkie, J. & Gani, D. (1999). A Model for Nonstoichiometric, Cotranslational Protein Scission in Eukaryotic Ribosomes. *Bioorg Chem* **27**, 55–79.

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

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

Rychlík, I., Černá, J., Chládek, S., Pulkrábek, P. & Žemlička, J. (1970). Substrate Specificity of Ribosomal Peptidyl Transferase The Effect of the Nature of the Amino Acid Side Chain. *Eur J Biochem* **16**, 136–142.

Sáiz, M., Gómez, S., Martínez-Salas, E. & Sobrino, F. (2001). Deletion or substitution of the aphthovirus 3′ NCR abrogates infectivity and virus replication. *J Gen Virol* **82**, 93–101.

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

Saunders, K. & King, A. M. Q. (1982). Guanidine-Resistant Mutants of Aphthovirus Induce the Synthesis of an Altered Nonstructural Polypeptide , P34. *J Virol* **42**, 389–394.

Schlegel, A., Giddings, T. H., Ladinsky, M. S. & Kirkegaard, K. (1996). Cellular origin and ultrastructure of membranes induced during poliovirus infection. *J Virol* **70**, 6576–88.

Schneider-Schaulies, J. (2000). Cellular receptors for viruses: Links to tropism and pathogenesis. *J Gen Virol* **81**, 1413–1429.

Schuller, A. P., Chih-Chien Wu, C., Dever, T. E., Buskirk, A. R. & Green, R. (2016). eIF5A Functions Globally in Translation Elongation and Termination. *Mol Cell* **66**, 194–205.

Seago, J., Juleff, N., Moffat, K., Berryman, S., Christie, J. M., Charleston, B. & Jackson, T. (2013). An infectious recombinant foot-and-mouth disease virus expressing a fluorescent marker protein. *J Gen Virol* **96**, 1517–1527.

Sekiguchi, K., Franke, A. J. & Baxt, B. (1982). Competition for cellular receptor sites among selected aphthoviruse. *Arch Virol* **64**, 53–64.

Serrano, P., Pulido Rodriguez, M., Sáiz, M. & Martínez-Salas, E. (2006). The 3′ end of the foot-and-mouth disease virus genome establishes two distinct long-range RNA-RNA interactions with the 5′ and region. *J Gen Virol* **87**, 3013–3022.

Sharma, P., Yan, F., Doronina, V. A., Escuin-Ordinas, H., Ryan, M. D. & Brown, J. D.

(2012). 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Res* **40**, 3143–3151.

Simmonds, P., Andrew, T. & Evans, D. J. (2004). Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. *RNA* **10**, 1337–1351.

Simmonds, P., Karakasiliotis, I., Bailey, D., Chaudhry, Y., Evans, D. J. & Goodfellow, I. G. (2008). Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses. *Nucleic Acids Res* **36**, 2530–2546.

Sommergruber, W., Zorn, M., Blaas, D., Fessl, F., Volkmann, P., Maurer-Fogy, I., Pallai, P., Merluzzi, V., Matteo, M. & other authors. **(1989).** Polypeptide 2A of human rhinovirus type 2: Identification as a protease and characterization by mutational analysis. *Virology* **169**, 68–77.

Sonenberg, N. & Gingras, A.-C. (1998). The mRNA 5′ cap-binding protein eIF4E and control of cell growth. *Curr Opin Cell Biol* **10**, 268–275.

Sonenberg, N. & Hinnebusch, A. G. (2009). Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* **136**, 731–745.

Spear, A., Ogram, S. A., Morasco, B. J., Smerage, L. E. & Flanegan, J. B. (2015). Viral precursor protein P3 and its processed products perform discrete and essential functions in the poliovirus RNA replication complex. *Virology* **485**, 492–501.

Stassinopoulos, I. A. & Belsham, G. J. (2001). A novel protein-RNA binding assay: functional interactions of the foot-and-mouth disease virus internal ribosome entry site with cellular proteins. *RNA* **7**, 114–122.

Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J., Homans, S. W. & Ashcroft, A. E. (2007). A Simple, RNA-Mediated Allosteric Switch Controls the Pathway to Formation of a T = 3 Viral Capsid. *J Mol Biol* **369**, 541–552.

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

Sutmoller, P., Barteling, S. S., Casas, R. & Sumption, K. J. (2003). Control and eradication

of foot-and-mouth disease. *Virus Res* **91**, 101–144.

Svitkin, Y. V., Pause, A., Haghighat, A., Pyronnet, S., Witherell, G., Belsham, G. J. & Sonenberg, N. (2001). The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5′ secondary structure. *RNA* **7**, 382–394.

Sweeney, T. R., Cisnetto, V., Bose, D., Bailey, M., Wilson, J. R., Zhang, X., Belsham, G. J. & Curry, S. (2010). Foot-and-mouth disease virus 2C is a hexameric AAA+ protein with a coordinated ATP hydrolysis mechanism. *J Biol Chem* **285**, 24347–24359.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731–2739.

Thompson, D., Muriel, P., Russell, D., Osborne, P. & Bromley, A. (2002). Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2001. *Rev Sci Tech* **21**, 675–687.

Tiley, L., King, A. M. Q. & Belsham, G. J. (2003). The Foot-and-Mouth Disease Virus cis-Acting Replication Element (cre) Can Be Complemented in trans within Infected Cells The Foot-and-Mouth Disease Virus cis-Acting Replication Element (cre) Can Be Complemented in trans within Infected Cells. *J Virol* **77**, 2243–2246.

Todd, S., Towner, J. S., Brown, D. M. & Semler, B. L. (1997). Replication-competent picornaviruses with complete genomic RNA 3’ noncoding region deletions. *J Virol* **71**, 8868–8874.

Tomita, K. (2014). Structures and Functions of Qβ Replicase: Translation Factors beyond Protein Synthesis. *Int J Mol Sci* **15**, 15552–15570.

Tong, L., Lin, L., Zhao, W., Wang, B., Wu, S., Liu, H., Zhong, X., Cui, Y., Gu, H. & other authors. **(2011).** Destabilization of coxsackievirus B3 genome integrated with enhanced green fluorescent protein gene. *Intervirology* **54**, 268–275.

Tosh, C., Hemadri, D. & Sanyal, A. (2002). Evidence of recombination in the capsid-coding region of type A foot-and-mouth disease virus. *J Gen Virol* **83**, 2455–2460.

Toyoda, H., Nicklin, M. J. H., 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**, 761–770.

Tulloch, F., Pathania, U., Luke, G. A., Nicholson, J., Stonehouse, N. J., Rowlands, D. J., Jackson, T., Tuthill, T., Haas, J. & other authors. **(2014).** FMDV replicons encoding green fluorescent protein are replication competent. *J Virol Methods* **209**, 35–40. Elsevier B.V.

Tulloch, F., Luke, G. A. & Ryan, M. D. (2017). Foot-and-mouth disease virus proteinases and polyprotein processing. In *Foot-and-Mouth Dis Curr Res Emerg Trends*, pp. 43–59. Edited by F. Sobrino & E. Domingo. Caister Academic Press.

Vandivier, L. E., Anderson, S. J., Foley, S. W. & Gregory, B. D. (2016). The Conservation and Function of RNA Secondary Structure in Plants RBP: RNA-binding protein. *Annu Rev Plant Biol* **67**, 463–488.

Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess Jr, J. W., Swanstrom, R., Burch, C. L. & Weeks, K. M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716.

Weeks, K. M. & Mauger, D. M. (2011). Exploring RNA structural codes with SHAPE chemistry. *Acc Chem Res* **44**, 1280–1291.

Whitton, L. J., Cornell, C. & Feuer, R. (2005). Host and virus determinants of picornavirus pathogenesis and tropism. *Nat Rev Microbiol* **3**, 1–9.

Wilson, D. N. & Cate, J. H. D. (2012). The Structure and Function of the Eukaryotic Ribosome. *Cold Spring Harb Perspect Biol* **4**, 1–18.

Wimmer, E., Hellen, C. U. & Cao, X. (1993). Genetics of poliovirus. *Annu Rev Genet* **27**, 353–436.

Witwer, C., Rauscher, S., Hofacker, I. L. & Stadler, P. F. (2001). Conserved RNA secondary structures in Picornaviridae genomes. *Nucleic Acids Res* **29**, 5079–5089.

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**, 32229–32235.

Xie, S. & Wang, Y. (2011). DIDS blocks a chloride-dependent current that is mediated by the 2B protein of enterovirus 71. *Cell Res* **21**, 1271–1275.

Yang, Y., Yi, M., Evans, D. J., Simmonds, P. & Lemon, S. M. (2008). Identification of a conserved RNA replication element (cre) within the 3Dpol-coding sequence of hepatoviruses. *J Virol* **82**, 10118–10128.

Ypma-Wong, M. F., Dewalt, P. G., Johnson, V. H., Lamb, J. G. & Semler, B. L. (1988). Protein 3CD is the major poliovirus proteinase responsible for cleavage of the p1 capsid precursor. *Virology* **166**, 265–270.

Yu, Y., Wang, H., Zhao, L., Zhang, C., Jiang, Z. & Yu, L. (2011). Fine mapping of a foot-and-mouth disease virus epitope recognized by serotype independent monoclonal antibody 4B2. *J Microbiol* **49**, 94–101.

Zhu, J. (2012). Mammalian cell protein expression for biopharmaceutical production. *Biotechnol Adv* **30**, 1158–1170.

Zibert, A., Maass, G., Strebel, K., Falk, M. M. & Beck ', E. (1990). Infectious Foot-and-Mouth Disease Virus Derived from a Cloned Full-Length cDNA. *J Virol* 2467–2473.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406–3415.

Appendix 1

Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias. In press: e-pub October 2017 doi:10.1261/rna.063339.117

Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias

JONAS KJÆR and GRAHAM J. BELSHAM

DTU National Veterinary Institute, Technical University of Denmark, Lindholm, DK-4771 Kalvehave, Denmark

ABSTRACT

Q1 **Foot-and-mouth disease virus (FMDV) has a positive-sense ssRNA genome including a single, large, open reading frame. Splitting of the encoded polyprotein at the 2A/2B junction is mediated by the 2A peptide (18 residues long), which induces a nonproteolytic, cotranslational, "cleavage" at its own C terminus. A conserved feature among variants of 2A is the C-terminal motif $N^{16}P^{17}G^{18}/P^{19}$, where P^{19} is the first residue of 2B. It has been shown previously that certain amino acid substitutions can be tolerated at residues E^{14}, S^{15}, and N^{16} within the 2A sequence of infectious FMDVs, but no variants at residues P^{17}, G^{18}, or P^{19} have been identified. In this study, using highly degenerate primers, we analyzed if any other residues can be present at each position of the NPG/P motif within infectious FMDV. No alternative forms of this motif were found to be encoded by rescued FMDVs after 2, 3, or 4 passages. However, surprisingly, a clear codon preference for the wt nucleotide sequence encoding the NPGP motif within these viruses was observed. Indeed, the codons selected to code for P^{17} and P^{19} within this motif were distinct; thus the synonymous codons are not equivalent.**

Keywords: picornavirus; synonymous codon; codon bias; translation

INTRODUCTION

Foot-and-mouth disease virus (FMDV) is the prototypic member of the genus *Aphthovirus* within the family *Picornaviridae*. This virus is the causative agent of the highly contagious and economically important disease of cloven-hoofed animals, foot-and-mouth disease. The positive-sense ssRNA genome of around 8400 nt includes a single, large, open reading frame (ORF), ~7000 nt, encoding a polyprotein (Belsham 2005). The full-length viral polyprotein is never observed since it is rapidly processed during and after synthesis mainly by the virus-encoded proteases (primarily $3C^{pro}$) to produce 15 distinct mature proteins plus multiple precursors (for review, see Martinez-Salas and Belsham 2017). Interestingly, FMDV, like many (but by no means all) other picornavirus (e.g., cardioviruses, erboviruses, teschoviruses etc.) uses a cotranslational, protease-independent mechanism for the "cleavage" of the polyprotein at the 2A/2B junction (the boundary between the capsid proteins and the nonstructural proteins) (Donnelly et al. 2001a). This mechanism has been referred to as "ribosomal skipping" or, alternatively, "stop-carry on" or "StopGo" (Donnelly et al. 2001a; Atkins et al. 2007; Doronina et al. 2008; Tulloch et al. 2017). The 2A peptide lacks characteristic protease motifs and only mediates "cleavage" during translation. It has been demonstrated that the 2A sequence is able to mediate "cleavage" in all eukaryotic translation systems tested whereas a number of artificial polyproteins containing this sequence have been examined in prokaryotic systems and no detectable cleavage products were observed (Donnelly et al. 1997).

The 2A peptide contains a highly conserved D^{12}(V/I)E(S/T)$NPG_{2A}{}^{\downarrow}P^{19}{}_{2B}$ motif at its C terminus, which is critical for its function (Ryan and Drew 1994; Donnelly et al. 1997). This motif, together with upstream amino acids, is believed to interact with the ribosomal exit tunnel. This prevents the formation of a peptide bond between the C-terminal glycine (G^{18}) of 2A and the N-terminal proline of 2B, referred to here, as P^{19} since it is an important part of the cleavage mechanism (see also Ryan et al. 1999; Donnelly et al. 2001a). However, remarkably, protein synthesis continues without the requirement for a reinitiation event.

Investigations into the activity of the 2A sequence have mainly been performed using in vitro experiments. Typically, these have either used mRNAs with single ORFs encoding artificial polyproteins comprising two reporter proteins linked via the 2A peptide (Ryan et al. 1991; Ryan and

Corresponding author: grbe@vet.dtu.dk

Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.063339.117.

Drew 1994; Donnelly et al. 2001b) or by expressing cDNAs encoding a truncated viral polyprotein including the StopGo coding region (Palmenberg et al. 1992). Alterations to the conserved $D^{12}(V/I)E(S/T)NPG_{2A}{\downarrow}P^{19}{}_{2B}$ motif reduced or abrogated the StopGo function (Donnelly et al. 2001b; Sharma et al. 2012), thereby showing that these amino acids are important for the correct StopGo "cleavage". Furthermore, Hahn and Palmenberg (1996) demonstrated that alterations to this motif also influenced the viability of encephalomyocarditis virus (EMCV, a cardiovirus) as they resulted in lethal phenotypes. Subsequently, Loughran et al. (2013) reported a similar observation for FMDV, as modification of the $S^{15}NPG_{2A}{\downarrow}^{18}P^{19}{}_{2B}$ sequence to $S^{15}NPL_{2A}{\downarrow}^{18}V^{19}{}_{2B}$ or $S^{15}NPA_{2A}{\downarrow}^{18}P^{19}{}_{2B}$ also gave rise to a lethal phenotype.

However, recently, certain amino acid substitutions (e.g., 2A S^{15} to F/I and 2A N^{16} to H) that have been shown to severely (60%–70%) impair "cleavage" at the 2A/2B junction, using in vitro assays (Donnelly et al. 2001b), have been found to be tolerated within infectious FMDVs (J Kjær and GJ Belsham, in prep.). In contrast, other substitutions (e.g., P^{19} to A and P^{19} to G) that inhibit cleavage more severely (by 89%–100%) in vitro, were not found within rescued viruses. Indeed, viruses rescued from these mutant transcripts had sequences that exactly matched the wt sequence (i.e., the rescued viruses were not mutant). In these studies, we also determined a critical role for the StopGo mechanism for the overall level of replication/translation of FMDV RNA. FMDV replicons with a defective 2A sequence had a markedly lower replication efficiency compared to the wt replicon (J Kjær and GJ Belsham, in prep.).

It is, therefore, apparent that some amino acid substitutions can be tolerated within the FMDV 2A peptide whereas other changes are not compatible with viability. To identify if any alternative residues can be accepted within the critical $N^{16}P^{17}G_{2A}{\downarrow}^{18}P^{19}{}_{2B}$ motif, degenerate sequences, encoding all possible amino acid substitutions at each of these positions individually, were introduced into a full-length FMDV cDNA, as used previously (Gullberg et al. 2013; Kristensen et al. 2017). In principle, this should result in the production of RNA transcripts encoding 2A peptides with a wide spectrum of "cleavage" activities. This was achieved by generating a large pool of plasmids, using site-directed mutagenesis with highly degenerate oligonucleotides, to change each of the individual codons corresponding to the amino acid residues within this conserved motif to NNN (where N is a mixture of all four bases). Using each pool of plasmids, RNA transcripts were prepared, in vitro, and introduced into baby hamster kidney (BHK) cells. Infectious viruses were rescued and characterized.

RESULTS AND DISCUSSION

The expected generation of a pool of StopGo cDNA mutants that could potentially result in all possible single amino substitutions in place of the N^{16}, P^{17}, G^{18}, and P^{19} residues (see Fig. 1A) was analyzed by sequencing (see Fig. 1B). The heterogeneity at the expected positions was clear in each case (this does not prove that each of the possible codons was present but indicates it is likely).

Full-length RNA transcripts were produced, in vitro, and introduced into BHK cells. Infectious virus was generated and passaged in fresh cells. RNA was then extracted from the virus harvests and the sequence encoding the 2A peptide was amplified by RT-PCR. The pool of amplicons was introduced into the pCR-XL-TOPO vector and then the sequence of the inserts in 20 individual colonies was determined for each virus harvest. It was found that all of the rescued viruses analysed after passages p2, p3, and p4 encoded the wt amino acid sequence at the NPGP motif in 2A. Interestingly, the complete spectrum of the possible synonymous codons for each of the residues N^{16}, P^{17}, G^{18}, and P^{19} was present in the rescued viral genomes at p2 (see Table 1). These results indicated that the approach had indeed generated a diverse pool of codons within the viruses. Furthermore, the very restricted range of nucleotide sequences encoding 2A observed within the rescued viruses strongly suggests that the specific amino acid sequence (NPGP), encoded by these nucleotide sequences, is critical for FMDV viability.

However, it was also apparent that the utilization of the different codons for the conserved amino acid residues varied. At p2, 55% of the sequences analysed had the wt codon for residue N^{16} (AAC) while the synonymous AAT codon was present in the remaining 45% of the rescued sequences. In the subsequent passages, the proportion of the AAC codon within the sequences increased to 75% and 95% by p3 and p4 respectively while the incidence of the AAT codon declined (Table 1). For residue P^{17}, at p2, the codon CCT was present in 55% of the colonies analysed and increased to 100% by p4. Each of the three other possible codons for P^{17} (CCC, CCA, and CCG) were also observed at p2 but each declined as the wt codon became dominant. For residues G^{18} and P^{19}, the wt codons (GGG and CCC, respectively) were in the minority (10 or 20%) at p2 and each of the synonymous codons were also present. However, interestingly, by p3 the wt codons had markedly increased to 50% abundance and by p4 were dominant ($\geq$90% abundance). For G^{18}, the GGA codon was the most abundant at p2 but declined during further passages to be only 10% of the sequences at p4. Similarly, for P^{19} the CCT codon was present in 50% of the sequences at p2 but declined to just 5% by p4. Strikingly, by p3, the wt codon was present in 50%–75% of the population at each of the four residues and by p4 the wt codon was present in 90%–100% of the virus population in each case (Table 1). Thus, it appears that selection occurs for the wt nucleotide sequence during passage of the rescued viruses in cell culture.

The wt GGGCCC nt sequence encoding residues G^{18} and P^{19} is recognized in DNA by the restriction enzyme ApaI (see Fig. 1A). Hence, it was possible to deplete the cDNA amplicons generated by RT-PCR, of the wt sequence from the

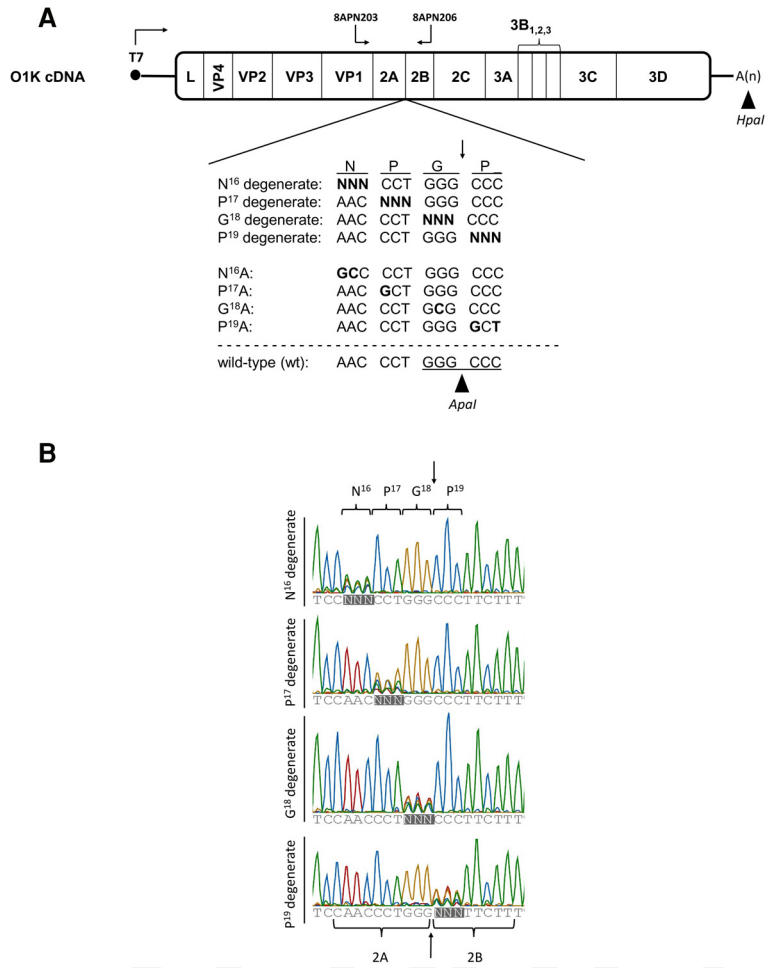

FIGURE 1. Structure of the FMDV O1 Kaufbeuren (O1K) cDNA and its derivatives. (*A*) The plasmid-encoded amino acids and the corresponding nucleotide sequences at the 2A/2B junction are shown. The FMDV O1K degenerate codon mutants were produced as described in the text using the mutant pT7S3 plasmids encoding the N^{16}A, P^{17}A, G^{18}A, and P^{19}A substitutions as templates. The full-length plasmid pools were linearized using HpaI prior to in vitro transcription and virus rescue. The locations of the HpaI and ApaI restriction sites that were used are marked. (*N*) A mixture of the 4 nucleotides. (*B*) Chromatograms and sequences of the FMDV cDNA corresponding to the NPGP motif at the 2A/2B junction. Degenerate positions showing the presence of multiple nucleotides are marked with an N (in bold type). The color code in the chromatograms is as follows: A (red), T (green), G (yellow), C (blue).

rescued viruses by digesting them with ApaI prior to the cloning step (it was anticipated that this should enhance the detection of non-wt nucleotide sequences). The residual, full-length, 650-bp amplicons were inserted into the pCR-XL-TOPO vector, as described above, and the plasmid DNA from individual colonies was sequenced. As expected, the wt codons for G^{18} and P^{19} were no longer observed in the cloned fragments (Table 2) and the G^{18} (GGA) and P^{19} (CCT) codons were predominant in these enriched populations. These results are consistent with those obtained without the ApaI digestion (since the GGA and CCT codons were also present in 50% of the fragments at p2 without this treatment, see Table 1) but clearly the apparent abundance of these non-wt codons is enhanced following the ApaI digestion (Table 2), as anticipated. The enrichment for non-wt sequences did not result in the detection of codons for alternative amino acids within the virus population. It had been anticipated that some amino acid substitutions at residue N^{16} might be rescued since a mutant (with N^{16} changed to H) has been shown to be viable (J Kjær and GJ Belsham, in prep.) but, presumably, it was outcompeted by the wt virus.

It is interesting to note that the G^{18} (GGA) and P^{19} (CCT) codons have previously been found to be the second most abundant codons found in FMDV genomes from all seven serotypes (see Gao et al. 2014). This comparison of FMDV sequences also indicated that the alternate codon for N^{16} (AAC) is present in only a small minority of FMDV genomes and CCC is also a minor population of the codons used for residue P^{17}. The results presented in Table 1 clearly indicate that infectious FMDVs with these synonymous changes can be obtained but these viruses do not appear to be stably maintained in cell culture and are apparently selected against.

The evidence presented here strongly suggests that there is a distinct selection, within the virus when grown in cell culture, for codon AAC for N^{16}, CCT for P^{17}, GGG for G^{18}, and CCC for P^{19}; thereby indicating that synonymous codon usage for this conserved motif is biased in these rescued viruses. It is particularly noteworthy that the codon preference for P^{17} and P^{19} is different (CCT and CCC, respectively). This raises the question of why does the virus select some codons over others? Various studies have demonstrated that synonymous codon usage bias plays an important role in the translation of certain mRNAs (Bulmer 1991; Akashi 2001; Novoa and Ribas de Pouplana 2012; Mauro and Chappell 2014). It is therefore conceivable that synonymous codons may influence the cleavage efficiency through the FMDV StopGo mechanism. As indicated above, a marked codon bias within the FMDV genome is apparent from the alignment of diverse FMDV 2A sequences as described by Gao et al. (2014). However, in the context of a synthetic reporter polyprotein, assayed within CHO cells, use of the four different synonymous codons for residue G^{18} of the 2A peptide resulted in very similar apparent "cleavage" efficiencies at the 2A/2B junction. This was interpreted as showing that it is the amino acid residue

TABLE 1. Codon utilization encoding the "NPGP" motif at the 2A/2B junction within rescued FMDVs

Residue	Codon		p2 %[a]	p3 %[a]	p4 %[a]
N16	*AAT*		45	25	5
N16	AAC	wt	55	75	95
P17	CCT	wt	55	70	100
P17	*CCC*		10	15	0
P17	*CCA*		20	5	0
P17	*CCG*		15	10	0
G18	*GGT*		15	5	0
G18	*GGC*		15	5	0
G18	*GGA*		50	40	10
G18	GGG	wt	20	50	90
P19	*CCT*		50	25	5
P19	CCC	wt	10	50	95
P19	*CCA*		10	5	0
P19	*CCG*		30	20	0

[a]From sequencing of plasmid DNA isolated from separate 20 colonies in each case, the proportion (%) of each codon present in the rescued FMDVs is indicated at the different passage (p) numbers. Codon frequency values of 50%–70% are highlighted in light gray, whereas values from 75%–100% are highlighted in dark gray.

rather than the nt sequence which is critical for achieving cleavage (Gao et al. 2014). However, using that assay system, the "cleavage" efficiency was only about 88%–89% while essentially 100% cleavage occurs within the native context, as in the virus. The results obtained here (see Table 1) indicate that two separate selection effects may be operating. There is a clear selection for the NPGP motif at the amino acid level. However, in addition, there is a distinct codon bias within the context of the rescued infectious viruses and a significant selection pressure appears to exist for the wt sequence. This effect is fully consistent with the codon bias observed in the analysis of natural FMDV genomic sequences (Gao et al. 2014). This suggests that the FMDV RNA sequence itself (rather than just the encoded amino acid sequence) affects the "cleavage" process (StopGo mechanism) at the 2A/2B junction. Such an effect could be achieved through a direct interaction of the RNA sequence itself or potentially through interactions with the specific charged tRNAs involved in the translation process. In the case of the P^{17} and P^{19} codons, it is interesting to note that the same type of prolyl tRNA (with an IGG anticodon) has been reported to be used for decoding of the CCC and CCU codons in human cells (no gene for a tRNA that is cognate for CCC was identified, see Mauro and Chappell 2014). However, in the current database of tRNA sequences from the Lowe laboratory, it appears that in humans, 1 of 23 genes for prolyl tRNAs has a GGG anticodon with 10 copies having an AGG anticodon. In the mouse genome, 1 of 20 genes for the prolyl tRNAs has the GGG anticodon and 8 genes have the AGG anticodon (see the gtrnadb.ucsc.edu database described in Chan and Lowe (2009)). Interestingly, in cattle and pigs (major hosts for FMDV) and also in the rat, there is no gene for a prolyl tRNA with a GGG anticodon. Thus, it is not clear whether a single, post-transcriptionally modified prolyl tRNA recognizes these two Pro codons (at least some of the time) or if different tRNAs are involved in the hamster cells used here. If a single tRNA is involved in recognizing both codons (as in cattle, pigs, and rats), then it seems that the RNA sequence itself must be influencing the StopGo process; it seems unlikely that this effect is mediated through some secondary or tertiary RNA structure, as this would presumably be lost on the ribosome during the process of translation. It will clearly be important to analyse the effect of the presence of the non-optimal synonymous codons on "cleavage" at the 2A/2B junction in its native context.

MATERIALS AND METHODS

Construction of plasmids containing full-length mutant FMDV cDNAs

Pools of StopGo cDNA mutants that potentially result in all possible single amino substitutions in place of the N^{16}, P^{17}, G^{18}, and P^{19} residues, respectively, were constructed. This was achieved using a two-step site-directed mutagenesis procedure. This is a variation of the QuickChange protocol (Stratagene), using Phusion High-Fidelity DNA Polymerase (Thermo Scientific) with modified versions of the plasmid pT7S3 (Ellard et al. 1999) as template. The wt pT7S3 contains the full-length cDNA for the O1Kaufbeuren B64 strain of FMDV. To eliminate the possibility of carrying over some residual wt template from the PCR, the templates used were modified versions of the pT7S3 with the codons for N^{16}, P^{17}, G^{18}, or P^{19} changed to encode an alanine (A) residue in each case (see Fig. 1A). These substitutions have been reported previously to result in a complete loss of apparent cleavage activity (Donnelly et al. 2001b; Sharma

TABLE 2. Enrichment for non-wt sequences encoding residues G^{18} and P^{19} within the "NPGP" motif within rescued FMDVs

Residue	Codon		Pretreatment	p2 %[b]	p3 %[b]
G18	*GGT*		ApaI[a]	5	0
G18	*GGC*		ApaI[a]	15	5
G18	*GGA*		ApaI[a]	80	95
G18	GGG	wt	ApaI[a]	0	0
P19	*CCT*		ApaI[a]	60	55
P19	CCC	wt	ApaI[a]	0	0
P19	*CCA*		ApaI[a]	15	0
P19	*CCG*		ApaI[a]	25	45

[a]Following RT-PCR, the 650-bp amplicons were digested with ApaI to enrich the population in non-wt sequences and the residual intact products were inserted into the pCR-XL-TOPO vector (see text).
[b]From sequencing of plasmid DNA isolated from separate 20 colonies in each case, the proportion (%) of each codon present in the rescued FMDVs is indicated at the different passage (p) numbers. Codon frequency values of 50%–70% are highlighted in light gray, whereas values from 75%–100% are highlighted in dark gray.

TABLE 3. Primers used to create and sequence mutant FMDV cDNAs

Primer	Sequence (5′–3′)
Fwd_2A_N16A_degen	GGAGTCCNNNCCTGGGCCCTTC
Fwd_2A_P17A_degen	GTCCAACNNNGGGCCCTTC
Fwd_2A_G18A_degen	GACGTCGAGTCCAACCCTNNNCCCTTCTTTTTCTCCGACGTTA
Fwd_2A_P19G_degen	TCG AGTCCAACCCTGGGNNNTTCTTTTTCTCCGACGTTAGG
8APN206	CACCCGAAGACCTTGAGAG
8APN203	CTCCTTCAACTACGGTGCC

et al. 2012) and it has not been possible to rescue infectious virus containing these substitutions (J Kjær and GJ Belsham, in prep.). The first round of PCRs used the forward mutagenic 2A PCR primers (Table 3), with a single reverse primer 8APN206 (Table 3) plus the four different modified pT7S3 plasmids as templates and generated amplicons of approximately 450 bp. These primary PCR products were then used as megaprimers for a second round of PCR with the respective mutant pT7S3 plasmids as templates to produce full-length plasmids. Following DpnI digestion, the products from each reaction were introduced into *E. coli* and grown as separate pools. The plasmid pools were sequenced using a BigDye Terminator v. 3.1 Cycle Sequencing Kit and a 3500 Genetic Analyzer (Applied Biosystems).

Rescue of virus from full-length cDNA plasmids

Plasmid DNA isolated from each pool was linearized by digestion with HpaI and RNA transcripts were prepared using T7 RNA polymerase (Ambion T7 MEGAscript) at 37°C for 4 h. The integrity of the transcripts was assessed on agarose gels and quantified by spectrophotometry (NanoDrop 1000, Thermo Scientific) after which they were introduced into BHK cells by electroporation, as described previously (Nayak et al. 2005). The BHK cells were maintained in Dulbecco's Modified Eagles Medium (DMEM) supplemented with 5% fetal calf serum, and incubated at 37°C with 5% CO_2. At 2 d post-electroporation, the viruses were harvested by freezing and then amplified through three passages (p2, p3, and p4) in BHK cells.

Characterization of viruses following multiple passages

After each passage, viral RNA was extracted from a sample of the virus harvest (using the RNeasy Mini Kit, QIAGEN) and converted to cDNA using ready-to-go you-prime first-strand beads (GE Healthcare Life Sciences). FMDV cDNA, which included the whole 2A coding region, was amplified in PCRs (AmpliTaq Gold DNA polymerase, Thermo Scientific) using primers 8APN206 and 8APN203 (see Table 3; Fig. 1). Control reactions, without RT, were used to ensure that the analysed products were derived from RNA and not from the presence of carryover plasmid DNA template. The amplicons (approximately 650 bp) were visualized on 1% agarose gels and purified (GeneJET gel extraction kit, Thermo Scientific). These amplicons should be representative of the heterogeneity present in the rescued virus populations. The resulting collections of fragments were inserted into pCR-XL-TOPO (Thermo Scientific) and the sequence of the cDNA fragment present in individual bacterial clones (20 colonies for each of the 4 residues) was determined using the same reverse primer as used for the PCR. The fragments from codon mutants G^{18} and P^{19} were also enriched for the non-wt sequence populations by digestion of the cDNA with ApaI prior to gel purification and insertion into the pCR-XL-TOPO vector as described above.

ACKNOWLEDGMENTS

We thank Preben Normann for excellent technical assistance. The studies were supported by core funds within DTU-Vet.

Received August 3, 2017; accepted October 14, 2017.

REFERENCES

Akashi H. 2001. Gene expression and molecular evolution. *Curr Opin Genet Dev* **11:** 660–666.

Atkins JF, Wills NM, Loughran G, Wu CY, Parsawar K, Ryan MD, Wang CH, Nelson CC. 2007. A case for "StopGo": reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go). *RNA* **13:** 803–810.

Belsham GJ. 2005. Translation and replication of FMDV RNA. *Curr Top Microbiol Immunol* **288:** 43–70.

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129:** 897–907.

Chan PP, Lowe TM. 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucl Acids Res* **37:** D93–D97.

Donnelly ML, Gani D, Flint M, Monaghan S, Ryan MD. 1997. The cleavage activities of aphthovirus and cardiovirus 2A proteins. *J Gen Virol* **78:** 13–21.

Donnelly ML, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, Ryan MD. 2001a. Analysis of the aphthovirus 2A/2B polyprotein "cleavage" mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal "skip." *J Gen Virol* **82:** 1013–1025.

Donnelly ML, Luke GA, Hughes LE, Luke G, Mendoza H, Dam E, Gani D, Ryan MD. 2001b. The "cleavage" activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring "2A-like" sequences. *J Gen Virol* **82:** 1027–1041.

Doronina VA, Wu C, de Felipe P, Sachs MS, Ryan MD, Brown JD. 2008. Site-specific release of nascent chains from ribosomes at a sense codon. *Mol Cell Biol* **28:** 4227–4239.

Ellard FM, Drew J, Blakemore WE, Stuart DI, King AMQ. 1999. Evidence for the role of His 142 of protein 1C in the acid induced disassembly of foot and mouth disease virus capsids. *J Gen Virol* **80:** 1911–1918.

Gao ZL, Zhou JH, Zhang J, Ding YZ, Liu YS. 2014. The silent point mutations at the cleavage site of 2A/2B have no effect on the self-cleavage activity of 2A of foot-and-mouth disease virus. *Infect Genet Evol* **28:** 101–106.

Gullberg M, Polacek C, Bøtner A, Belsham GJ. 2013. Processing of the VP1/2A junction is not necessary for production of foot-and-mouth

disease virus empty capsids and infectious viruses: characterization of "self-tagged" particles. *J Virol* **87:** 11591–11603.
Hahn H, Palmenberg AC. 1996. Mutational analysis of the encephalomyocarditis virus primary cleavage. *J Virol* **70:** 6870–6875.
Kristensen T, Normann P, Gullberg M, Fahnøe U, Polacek C, Rasmussen TB, Belsham GJ. 2017. Determinants of the VP1/2A junction cleavage by the 3C protease in foot-and-mouth disease virus-infected cells. *J Gen Virol* **98:** 385–395.
Loughran G, Libbey JE, Uddowla S, Scallan MF, Ryan MD, Fujinami RS, Rieder E, Atkins JF. 2013. Theiler's murine encephalomyelitis virus contrasts with encephalomyocarditis and foot-and-mouth disease viruses in its functional utilization of the StopGo non-standard translation mechanism. *J Gen Virol* **94:** 348–353.
Martinez-Salas E, Belsham GJ. 2017. Genome organisation, translation and replication of foot-and-mouth disease virus RNA. In *Foot-and-mouth disease: current research and emerging trends* (ed. Sobrino F, Domingo E), pp. 13–42. Caister Academic Press, Poole, UK.
Mauro VP, Chappell SA. 2014. A critical analysis of codon optimization in human therapeutics. *Trends Mol Med* **20:** 604–613.
Nayak A, Goodfellow IG, Belsham GJ. 2005. Factors required for the Uridylylation of the foot-and-mouth disease virus 3B1, 3B2, and 3B3 peptides by the RNA-dependent RNA polymerase ($3D^{pol}$) in vitro. *J Virol* **79:** 7698–7706.
Novoa EM, Ribas de Pouplana L. 2012. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet* **28:** 574–581.
Palmenberg AC, Parks GD, Hall D, Ingraham RH, Seng TW, Pallal PV. 1992. Proteolytic processing of the cardioviral cleavage in clone-derived P2 region: primary 2A/2B precursors. *Virology* **190:** 754–762.
Ryan MD, Drew J. 1994. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J* **13:** 928–33.
Ryan MD, King AMQ, Thomas GP. 1991. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J Gen Virol* **72:** 2727–2732.
Ryan MD, Donnelly M, Lewis A, Mehrotra AP, Wilkie J, Gani D. 1999. A model for nonstoichiometric, cotranslational protein scission in eukaryotic ribosomes. *Bioorg Chem* **27:** 55–79.
Sharma P, Yan F, Doronina VA, Escuin-Ordinas H, Ryan MD, Brown JD. 2012. 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Res* **40:** 3143–3151.
Tulloch F, Luke GA, Ryan MD. 2017. Foot-and-mouth disease virus proteinases and polyprotein processing. In *Foot-and-mouth disease: current research and emerging trends* (ed. Sobrino F, Domingo E), pp. 43–59. Caister Academic Press, Poole, UK.

KjaerRNA063339

Queries

Jonas Kjær and Graham J. Belsham

Q1 As outlined in our Instructions to Authors, it is the journal's style to set genes, alleles, and loci in italic, and proteins in Roman type. Please verify that all have been properly set throughout the manuscript.