Highly structured genomic sequences mediate the function of a RNA virus polymerase.

Joseph C. Ward[a], Lidia Lasecka-Dykes[b], Samuel J. Dobson[a], Sarah Gold[b], Natalie J. Kingston[a], Morgan R. Herod[a], Donald P. King[b], Tobias Tuthill[b], David J. Rowlands[a#] and Nicola J. Stonehouse[a#]

a School of Molecular and Cellular Biology, Faculty of Biological Sciences and Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, United Kingdom.

b The Pirbright Institute, Ash Road, Pirbright, Surrey, GU24 ONF, United Kingdom.

Running Head: The S fragment in FMDV replication

#Address correspondence to Nicola J. Stonehouse, N.J.Stonehouse@leeds.ac.uk

David J. Rowlands, d.j.rowlands@leeds.ac.uk

Key Words: picornavirus, S fragment, SHAPE, replication, 5′ UTR, $3D^{pol}$

Abstract word count: 213

Text word count: 5731

Abstract

Secondary and tertiary RNA structures play key roles in genome replication of single stranded positive sense RNA viruses. Complex, functional structures are particularly abundant in the untranslated regions of picornaviruses, where they are involved in initiation of translation, priming of new strand synthesis and genome circularisation. The 5′ UTR of foot-and-mouth disease virus (FMDV) is predicted to include a *c.* 360 nucleotide-long stem-loop, termed the short (S) fragment. This structure is highly conserved and essential for viral replication, but the precise function(s) are unclear. Here, we used selective 2′ hydroxyl acetylation analysed by primer extension (SHAPE) to experimentally-determine aspects of the structure, alongside comparative genomic analyses to confirm structure conservation from a wide range of field isolates. To examine its role in virus replication, we introduced a series of deletions to the distal and proximal regions of the stem loop. These truncations affected genome replication in a size-dependent and, in some cases, host cell-dependent manner. Furthermore, during passage of viruses incorporating the largest tolerated deletion from the proximal region of the S fragment stem loop, an additional mutation was selected in the viral RNA-dependent RNA polymerase, $3D^{pol}$, which influenced the function of the enzyme. These data suggest that the S fragment and $3D^{pol}$ interact in the formation of the FMDV replication complex.

Introduction

The genomes of single-stranded positive sense RNA viruses include both coding and regulatory regions. Picornaviruses are amongst the smallest, with genomes approximately 7.5 - 8 kb in length. Typically, the genome comprises a major single open reading frame flanked by 5′ and 3′ untranslated regions (UTRs), however, additional small open reading frames have been identified in some cardioviruses and enteroviruses (1, 2). The large 5′ UTRs present in most picornaviruses range from ~800 to ~1300 nucleotides (nts) and have been predicted *in silico* by energy minimisation algorithms to comprise several distinct highly structured domains. Some of these RNA structures are common for all picornaviruses and have been well characterised, for instance an internal ribosome entry site (IRES) responsible for the initiation of translation, which comprises ~450 nts for most picornaviruses, although smaller IRES elements resembling those found in some flaviviruses were also found (3–5). The number and organisation of these structural domains varies amongst picornaviruses. For example, a small stem-loop, the cis-active replicative element (*cre*), which is essential for replication, is found in the 5′ UTR of aphthoviruses, but within the coding region in enteroviruses (6–8). All picornavirus genomes include a structured domain at the 5′ end, although the size, conformation and sequence varies in viruses from different genera. Enterovirus 5′ UTRs terminate in a ~80 nucleotide cloverleaf structure, while in some other picornaviruses the 5′ end forms an extended stem-loop (9–11). The size of this stem-loop varies markedly; 40 nts in hepatoviruses and ~80 nts in cardioviruses, whereas the largest (~360 nts) is present in the aphthoviruses (12). The functions of some of these RNA structures remain unknown.

Here, we focus on the long 5′ terminal stem-loop (termed S fragment) of foot-and-mouth disease virus (FMDV), an apthovirus which infects cloven-hoofed animals. It is the causative

agent of foot-and-mouth disease, a highly contagious and economically important infection posing a constant thread to the global livestock industry. In endemic countries FMD is controlled by vaccination and by movement restrictions, while movement restrictions and slaughter were used when outbreaks occurred in non-endemic countries (13). Control by vaccination is complicated by high antigenic variability and the occurrence of asymptomatic carrier animals (14–16). FMDV has an 8.5 kb RNA genome organised into L, P1-2A, P2 and P3 coding regions, flanked by the 5′ UTR described above and a short 3′ UTR. L is a protease responsible for separating itself from the polyprotein and cleaving key cellular proteins, P1 encodes the structural (capsid) proteins, while P2 and P3 encode non-structural proteins involved in genome replication. The latter include the RNA-dependent RNA polymerase (RdRp, also termed $3D^{pol}$) and the protease $3C^{pro}$, as well as less well-characterised proteins including 2C, a potential helicase (17–19). At approximately 1.3 kb, the 5′ UTR of FMDV is ~1/7th of the entire genome and is the longest 5′ UTR within the picornavirus family (10). Its sequence comprises at least five structurally and functionally distinct domains. The S fragment stem-loop is located at the 5′ terminus and is followed by a long poly-C-tract of variable length (about 70 to 250 nts), 2-4 tandem repeated sequences predicted to form pseudoknots, the small stem-loop *cre* involved in uridylation of the replication primer peptide VPg and finally the IRES element responsible for protein translation initiation (20). The functions of several of these domains are poorly understood (21, 22), although we have recently described the importance of a sequence in the pseudoknot region for virus assembly (23).

Whilst the S fragment is known to be essential for replication, its precise function remains unclear. The secondary structure of the S fragment appears to be conserved, although its sequence and length varies between different clades of FMDV (10, 24). Truncated S

fragments have been observed for some field isolates, each time retaining its long stem-loop structure (10). It has been reported that while distal portions of this stem-loop structure are not essential for genome replication, they play a role in the modulation of host-cell innate immune responses (25, 26). A corresponding structure present at the 5′ end of poliovirus (PV) RNA, a highly structured region called cloverleaf , is involved in the switch from protein translation to RNA replication, in addition to recruiting host and viral proteins which contribute to RNA circularisation during genome replication (9, 27). It is possible that the S fragment in FMDV has similar roles and it is known to interact with the 3′ UTR, suggesting a role in genome circularisation (28).

We used selective-2′ hydroxyl acylation analysed by primer extension (SHAPE) to provide direct evidence to complement comparative bioinformatic analyses and *in silico* predictions of the structure of the S fragment. We confirm that substantial deletions can be made to the distal region of the S fragment stem-loop without seriously compromising *in vitro* replication, although host cell differences were observed. We also show that proximal deletions are less well tolerated, which correlates with our observation that the proximal part of the S fragment shows higher conservation of nucleotide pairings. Furthermore, viruses reconstructed to include the maximal viable proximal deletion can, during sequential passages, select a compensatory mutation in the $3D^{pol}$. This mutation is located in a highly conserved position known to interact with template RNA, and granted faster replication of a mutant replicon carrying this proximal S fragment deletion.

Material and Methods

Cells lines.

Baby hamster kidney (BHK)-21 and Madin-Darby bovine kidney (MDBK) cells were obtained from the American Type Culture Collection (ATCC) (LGC Standard) and maintained in Dulbecco's modified Eagle's Medium with glutamine (Sigma-Aldrich) supplemented with 10% foetal calf serum (FCS), 50 U/ml penicillin and 50 µg/ml streptomycin as previously described (29).

Plasmid construction.

The FMDV ptGFP replicon plasmids along with the equivalent $3D^{pol}$ knock-out controls $3D^{pol}$ -GNN have already been described (30, 31).

A variant of the ptGFP replicon was produced to allow for easy mutagenesis of the S-fragment. An *Eag*I site was introduced at each end of the S-fragment by PCR mutagenesis, altering the wild-type sequence from AAAGGGGGCATTA to AAACGGCCGATTA at the 5′ end and GCGCCCGCCTTT to GCGCGGCCGTTT at the 3′ end. S-fragment sequences containing deletions were chemically synthesised (Thermofisher) and inserted into the replicon vector using complementary *Eag*I sites.

Mutation of residue I189L of $3D^{pol}$ in the ptGFP replicon was achieved by PCR mutagenesis resulting in A7023C substitution. All primer sequences are available on request.

Protein purification.

The I189L mutation was introduced into the His-tagged $3D^{pol}$ expression clone pET28a-3D, and recombinant protein was expressed and purified as previously described (32, 33).

***In vitro* transcription.**

In vitro transcription reactions were performed as described previously (34, 35). Briefly, 5 µg of replicon plasmid was linearised with *Asc*I (NEB), purified by phenol-chloroform extraction, ethanol precipitated, redissolved in RNase-free water and used in a T7 *in vitro* transcription reaction. Reactions were incubated at 32°C for 4 hours, treated with 1.25 units of RQ1 DNase for 30 minutes at 37°C and RNA recovered using an RNA Clean & Concentrator-25 spin column kit (Zymo Research), following manufacturer's instructions. All transcripts were quantified by NanoDrop 1000 (Thermo Scientific) and RNA integrity assessed by MOPS-formaldehyde gel electrophoresis.

Replication assays.

BHK-21 and MDBK cell based replicon replication assays were performed in 24-well plates with 0.5 µg/cm^2 of RNA using Lipofectin transfection reagent (Life Technologies) as previously described (35). For every experiment, transfection was performed in duplicate and experiments biologically repeated as indicated. Replicon replication was assessed by live cell imaging using an IncuCyte Zoom Dual colour FLR, an automated phase-contrast and fluorescent microscope within a humidifying incubator. At hourly intervals up to a defined end point, images of each well were taken and used to enumerate the ptGFP positive cell count per well, using our established protocols (34, 35). Data are shown at 8 hours post transfection (when maximum replication was observed) on a linear scale.

For ribavirin sensitivity assays, BHK-21 cells were pre-treated with ribavirin (Sigma) for 7 hours prior to transfection.

SHAPE analysis.

RNA transcripts representing the entire 5′ UTR of FMDV (nucleotides 1 – 1581) were prepared as above. A sample containing 12 pmol of transcribed RNA was heated to 95°C for 2 minutes before cooling on ice. RNA folding buffer (100 mM HEPES, 66 mM $MgCl_2$ and 100 mM NaCl) was added to the RNA and incubated at 37°C for 30 minutes. The folded RNA was treated with 5 mM *N*-methyl isatoic anhydride (NMIA) or DMSO for 50 minutes at 37°C. The chemically modified RNA was ethanol precipitated and resuspended in 0.5x Tris-EDTA (TE) buffer.

Hex or FAM fluorescent primers were bound to modified RNA by heating the reaction to 85°C for 1 minute, 60°C for 10 minutes and 35°C for 10 minutes in a thermocycler. Reverse transcription was continued using Superscript III (Invitrogen) following manufacturers protocol with incubation at 52°C for 30 minutes.

Post-extension, cDNA:RNA hybrids were disassociated by incubation with 4M NaOH at 95°C for 3 minutes before neutralisation with 2M HCl. Extended cDNA was ethanol precipitated and resuspended in deionized formamide (Thermo Fisher). Sequencing ladders were similarly produced using 6 pmol of RNA with the inclusion of 10 mM ddCTP in the reverse transcription mix and using a differentially labelled fluorescent primer (either Hex or FAM). A sequencing ladder was combined with NMIA or DMSO samples and dispatched on dry ice for capillary electrophoresis (Dundee DNA seq) (23).

NMIA reactivity was used as constraints for RNA secondary structure prediction of the S fragment using the Vienna RNA probing package and based on sequence of the replicon used in this study (36).

Construction of recombinant viruses.

Replicons used here are based on plasmid pT7S3 which encodes a full-length infectious copy of FMDV O1 Kaufbeuren (O1K) (37). To generate infectious viral genomes the reporter sequence was removed from replicons by digestion with flanking *Psi*I and *Xma*I restriction enzymes and replaced with the corresponding fragment from pT7S3 encoding the capsid proteins. Full length viral RNA was transcribed using a T7 MEGAscript kit (Thermo Fisher Scientific), DNase treated using TurboDNase (Thermo Fisher Scientific) and purified using a MEGAclear Transcription Clean-Up kit (Thermo Fisher Scientific). RNA quality and concentration were determined by denaturing agarose gel electrophoresis and Qubit RNA BR Assay Kit (Thermo Fisher Scientific).

Virus recovery.

BHK-21 cells were transfected in 25 cm^2 flasks with 8 µg per flask of infectious clone-derived RNA using TransIT transfection reagent (Mirus) as described previously (35). At full cytopathic effect (CPE) or 24 hours post-transfection (whichever was earlier) cell lysates were freeze-thawed and clarified by centrifugation. The clarified lysate (1 ml) was blind passaged onto naïve BHK-21 cells and this was repeated for five rounds of passaging.

Sequencing of recovered virus.

Recovered viruses at passage 3, 4 and 5, were sequenced with an Illumina Miseq (Illumina) using a modified version of a previously described PCR-free protocol (38). Total RNA was extracted from clarified lysates using TRizol reagent (Thermo Fisher Scientific) and residual

genomic DNA was removed using DNA-free DNA removal Kit (Thermo Fisher Scientific) prior to ethanol precipitation. Purified RNA was used in a reverse transcription reaction as previously described (38, 39). Following reverse transcription, cDNA was purified and quantified using a Qubit ds DNA HS Assay kit (Thermo Fisher Scientific) and a cDNA library prepared using Nextera XT DNA Sample Preparation Kit (Illumina). Sequencing was carried out on the MiSeq platform using MiSeq Reagent Kit v2 (300 cycles) chemistry (Illumina) and paired-end sequencing.

FastQ files were quality checked using FastQC with poor quality reads filtered using the Sickle algorithm (40). Host cell reads were removed using FastQ Screen algorithm and FMDV reads assembled *de novo* into contigs using IDBA-UD (41). Contigs that matched the FMDV library (identified using Basic Local Alignment Search Tool (BLAST)) were assembled into consensus sequences using SeqMan Pro software in the DNA STAR Lasergene 13 package (DNA STAR) (42). Finally, the filtered fastQ reads were aligned to the *de novo* constructed consensus sequences using BWA-MEM algorithm incorporated into Burrows-Wheeler Aligner (BWA), and mutations were visualised using the Integrative Genomics Viewer (IGV) (43–46).

Cell killing assays.

Virus titre was determined by plaque assay on BHK-21 cells as described before (47). BHK-21 cells were seeded with 3 x10^4 cells/well in 96 well plates and allowed to settle overnight. Cell monolayers were inoculated with each rescued virus at MOI of 0.01 PFU for 1 hour, inoculum was removed and 150 µl of fresh GMEM (supplemented with 1% FCS) was added to each well. Appearance of CPE was monitored every hour by live cell microscopy (contrast

phase) using the IncuCyte S3. CPE was observed as rounding of the cells and progress of infection was monitored as a drop in confluency compared to the mock treated cells (i.e., treatment with uninfected cell lysate). Serial images of cells were analysed using IncuCyte S3 2018B software.

Sym/Sub polymerase activity assays

Sym/Sub RNA with sequence GCAUGGGCCC was synthesised (Sigma Aldrich) and 5′ end labelled using γ^{32}P UTP (Perking Elmer) and T4 polynucleotide kinase (NEB) following manufacturer's protocol. Labelled RNA was purified by ethanol precipitation and resuspended in nuclease free water.

0.5 μM of end labelled RNA was incubated in polymerisation buffer (30 mM MOPS pH 7.0, 33 mM NaCl, 5 mM MgAc) and heated to 95°C for 2 minutes before chilling on ice. 2 μM of recombinant 3D^{pol} was added to the RNA and incubated at 37°C for 10 minutes to promote annealing. After annealing, 50 μM of rNTP was added to the reaction, addition of nucleotide varied depending on experiment. Aliquots were taken periodically, and reactions stopped by addition of 2x TBE-urea RNA loading dye (Thermo Fisher).

Extension products in loading dye were heated to 70°C for 5 minutes before loading on a 23% denaturing polyacrylamide gel containing 7 M urea. Samples were electrophoresed until sufficiently separated. After electrophoresis gels were fixed for 30 minutes in fixative solution and exposed onto a phosphoscreen (48, 49).

Bioinformatic analysis

Full genome sequences of 118 FMDV isolates, representing all seven serotypes (Supplementary Table S1), were downloaded from GenBank. The sequences were chosen (based on the region encoding the VP1 protein) to represent the known genomic diversity of FMDV across all seven serotypes. VP1 is the most variable part of the FMDV genome and frequently used to calculate phylogenetic relationship of FMDV isolates.

The RNA structure of the S fragment was predicted as described before using the RNAalifold program implemented in the ViennaRNA package (36, 50). For the covariance analysis, also calculated using RNAalifold program, only those isolates were included which contained complete sequence at the 5′ end of the genome. Data representing covariance were superimposed onto the schematics of the S fragment RNA structure and visualised using Forna tool implemented in the ViennaRNA Web Services (51). The sequence logo representing amino acid conservation within positions 185 – 194 of the $3D^{pol}$ was prepared using a WebLogo 3.7.4 server and using 1123 FMDV $3D^{pol}$ sequences obtained from fmdbase.org (an FMDV sequence database generated by the FAO World Reference Laboratory for FMD at The Pirbright Institute) (52, 53).

Results

SHAPE analysis of the S fragment.

Based on computational folding, such as mFOLD algorithms, the S fragment is predicted to form a single large hairpin stem-loop comprising approximately 360 nucleotides. Here, we use selective 2′ hydroxyl acylation analysed by primer extension (SHAPE) to further investigate the structure of this part of the FMDV genome. This approach relies on the formation of 2′-O-adducts in single stranded and accessible regions of the RNA, which can then be detected by reverse transcription.

RNA transcripts representing the full FMDV 5′ UTR were folded before incubation with 5 mM NMIA. The modified RNAs were then used in reverse-transcription reactions containing fluorescently end-labelled primers. Reverse transcription is terminated when the enzyme reaches a nucleotide with a NMIA adduct, thus creating a series of cDNA fragments of different lengths. These cDNA fragments were analysed by capillary gel electrophoresis alongside a sequencing ladder to identify the sites of termination, indicating the locations of unpaired and accessible nucleotides. The proximity of the poly-C-tract to the 3′ end of the S fragment limited potential primer binding sites, resulting in unreliable data for the last 60 nucleotides of the S fragment and such was excluded from the analysis (shown in grey in Figure 1A). The rest of the S fragment was well covered and the SHAPE reactivities from 6 independent experiments were used to complement the mFOLD algorithm analyses to generate a structural prediction of this region (Figure 1).

In general, the SHAPE mapping data fitted well with the *in silico* predictions with the most reactive residues coinciding with bulges within the predominately double stranded stem-loop structure (Figure 1A). The locations of reactive residues can be seen in the NMIA reactivity graph, where groups of nucleotides within bulges showed high reactivity (Figure 1B). However, some predicted bulges showed little NMIA reactivity, suggesting steric hindrance possibly due to higher order tertiary structure. Similarly, some predicted base-paired nucleotides were NMIA reactive. These mostly occurred at the tops and bottom of bulges and may result from nucleotides becoming transiently available for modification during 'breathing' (54). The 5′ terminal two nucleotides (uridines at position 1 and 2) were highly reactive and therefore likely to be non-base paired.

The effects of deletions to the distal or proximal regions of the S fragment stem-loop on replicon replication.

The SHAPE data and the structural prediction described in Figure 1 were used to design a series of truncations to the S fragment, and the modified sequences were introduced into a replicon. The replicon was based on an infectious clone of FMDV O1K in which the P1 region of the genome is replaced with a ptGFP reporter (55) (Figure 2A). Deletions were designed to end at either the top or bottom of bulges, helping ensure that the stem-loop hairpin structure was maintained. To facilitate replacement of the S fragment with modified sequences *Eag*I restriction sites were introduced at each end of the S fragment sequence. These sites required minimal modification of the original sequence and showed no predicted change in RNA structure, and the modified genome sequence is hereafter referred to as WT. Addition of the *Eag*I site at the 5′ end of the S fragment resulted in a large drop in replication, however when paired with the 3′ EagI site, restoring base pairing, replicative potential was

drastically improved, although still below that of the original O1K WT sequence (Figure S1). Deletions of 70, 148, 246 and 295 nucleotides were introduced at the distal end (top) of the S fragment stem-loop (constructs named D-70, D-148, D-246 and D-295, respectively) and deletions of 48, 97, 195 and 271 nucleotides were made to the proximal end (bottom) of the S fragment (constructs referred to as P-48, P-97, P-195 and P-271, respectively) (Figure 2B). The truncated sequences were chemically synthesised and introduced into the WT replicon via the *Eag*I sites, meaning that the highly conserved 5′ and 3′ regions and essential UU nucleotides at positions 1 and 2 were maintained in all truncation mutants. The eight new replicon clones were transcribed into RNAs and transfected into BHK-21 or MDBK cells, alongside WT and a $3D^{pol}$-GNN negative control. The $3D^{pol}$-GNN replicon cannot replicate and so the ptGFP signal produced represents translation of the input RNA. BHK-21 and MDBK cells are continuous cell lines known to support replication of FMDV: BHK-21 cells are used for FMDV vaccine production, whereas MDBK cells originate from a natural host of FMDV. Replication was monitored by measuring reporter expression using an IncuCyte ZOOM live cell imaging system, with analysis by our established protocols (performed 8 hours post transfection, when replication has reached its peak (Figure S2)) (Figure 2).

As described previously and consistent with sequence information from natural virus isolates (10, 26), the D-70 mutation caused no significant drop in replication in either cell line. However, the replicon bearing the D-148 deletion, which removes the distal half of the S fragment stem-loop, showed reduced replication in BHK-21 cells and none in MDBK cells (Figure 2C,D). Innate immunity is compromised in BHK-21 cells but intact in MDBK cells (56, 57), suggesting that the distal portion of the S fragment stem-loop plays a role in modulating a competent host immune response, in agreement with previously published data

(25, 58). Replicons with the largest distal deletions, D-246 and D-295, did not replicate in either cell line.

While the consequences of deleting sequences from the distal region of the S fragment stem-loop have been reported (25), the effects of proximal deletions have not been investigated. The two smallest proximal deletions (P-48 and P-97), showed modest but significant decreases in replication compared to WT (1.7 and 1.4-fold, respectively) (Figure 2E,F. However, the larger deletions, P-195 and P-271, completely ablated replication and reporter gene expression was reduced to a level comparable to the GNN control. Interestingly, there was no significant difference between the replication of the P-48 and P-97 deletions in either BHK-21 or MDBK cells (Figure 2). Overall, the proximal part of the S fragment stem-loop appears to be more sensitive to deletion in a cell type-independent manner, with truncations as small as 48 nucleotides having a significant effect on replication.

Earlier studies indicated that the proximal part of the S fragment stem-loop is more conserved than the distal part (10). In recent years, advances in high throughput sequencing (HTS) have led to a great increase in the number of FMDV full genome sequences available on public domains, including sequences of viruses of SAT serotypes which were previously underrepresented. This increase in the range of sequences available prompted us to revisit the subject of the S fragment variability and we carried out a covariance analysis based on 118 FMDV isolates representing all seven serotypes. Figure 3 shows that there is more base-pairing conservation in the proximal than in the distal parts of the S fragment stem-loop (Figure 3). The base-pairing conservation in the proximal part of the otherwise very variable S fragment stem-loop (low nucleotide identity of approximately 47%) suggests evolutionary

pressure to maintain the structure of this portion of the stem-loop and complements our observations from the replicon experiments (Figure 2) (10, 59).

Viruses with proximal deletions of the S fragment replicate more slowly than WT and can select for a mutation in 3D^{pol} during serial passage.

We investigated the consequences of proximal deletions to the S fragment stem-loop further by introducing them into an FMDV O1K infectious clone. This enabled evaluation of replication in the context of the entire genome and investigation of the possible selection of compensatory or adaptive mutations during serial passage. Replicons bearing proximal deletions of the S fragment stem-loop were converted into infectious clones by replacing the ptGFP reporter sequence with the original O1K structural protein sequence. RNA transcripts of the truncated S fragment clones and WT were transfected into BHK-21 cells. Supernatants containing recovered virus were harvested and used to infect naïve BHK-21 cells for 5 continuous passages. Recovery of infectious virus from each construct was assessed by the ability of 5th passage samples to induce CPE. As expected, CPE was induced and virus recovered from the WT construct. Infectious virus was also recovered from P-48 and P-97 but no virus was recovered from either P-195 or P-271 constructs.

The consequences of the P-48 and P-97 S fragment deletions for the dynamics of virus replication were investigated by assessing the time taken to induce CPE in BHK-21 cell monolayers after infection with the mutated viruses or WT at a low MOI of 0.01 PFU. The integrity of the cell monolayers was assessed at hourly intervals over 63 hours using live cell imaging monitored by Incucyte S3 (Figure 4A). Both P-48 and P-97 viruses developed CPE

more slowly than WT, as anticipated from the earlier replicon experiments (Figure 4A). For comparison, while a D-148 virus bearing a larger deletion from the distal end of the S fragment stem-loop replicated at a slower rate than the WT virus, D-148 replicated significantly faster than viruses carrying deletions at the proximal end of the S fragment (i.e., P-48 and P-97; Figure 4A); further confirming replicon experiments. Ability of the viruses to form plaques was also assessed. Both, P-48 and P-97 mutants generated significantly smaller plaques when comparing to WT and/or D-148 mutant virus (Figure 4B).

Sequencing of recovered viruses revealed one isolate with mutations in 2C, which were not investigated further. In addition, a single mutation in the $3D^{pol}$ sequence was found in a high proportion of the viral progeny of one of three P-97 viral replicates. The A7203C change resulted in I189L mutation in the $3D^{pol}$ protein amino acid sequence. The evolution of this mutation was investigated further by sequencing viruses from earlier passages. The proportion of the I189L mutation in the population increased from 40% at passage 3 to 51% at passage 4, and 61% at passage 5, suggesting it confers a selective advantage in the context of the proximal S fragment stem-loop deletion and so is sequentially enriched. With four exceptions (two isolates containing valine and two isolates containing threonine at this position) out of 1123 sequences of $3D^{pol}$ available, all known FMDV isolates have isoleucine at this position (Figure S3), making the amino acid substitution for leucine novel. . This residue is known to interact with template RNA (Figure 5A), playing a crucial role during viral replication (60). No such mutations were observed in the WT or the P-48 mutant viruses after passage.

Effect of the 3D^{pol} I189L compensatory mutation on replicon replication.

The 3D^{pol} I189L mutation was introduced into WT, GNN and P-97 replicons by site-specific mutagenesis and the consequence for replication assessed by reporter gene expression following transfection into BHK-21 cells, as above. Introduction of 3D^{pol} I189L into the WT replicon had no significant effect on replication, however, a small but significant increase in replication was observed when it was introduced into the P-97 replicon (Figure 5B). As expected, I189L was unable to restore the activity of the GNN negative control.

I189L modifies 3D^{pol} activity.

The influence of the 3D^{pol} I189L mutation on polymerase function was investigated using Sym/Sub assay; a highly sensitive method for interrogating polymerase activity. Sym/Sub assays have been used to study PV replication and enable extension of a radiolabelled template to be measured at the single nucleotide level (48). The method uses a 10-nucleotide template oligonucleotide which base-pairs to create a small double stranded region with a four nucleotide 5′ overhang, thus allowing 3′ nucleotide addition. In the presence of rNTPs and recombinantly expressed 3D^{pol}, the addition of single or multiple nucleotides can be assessed over time by separating input template and elongation products by PAGE (Figure 6A). Initially, an assay using ATP only was undertaken to probe the ability of the enzyme to add a single nucleotide. WT and I189L 3D^{pol} proteins behaved similarly in the assay with no significant differences in the rate of addition of the single nucleotide (Figure 6B & C).

3D^{pol} activity was examined further by including all four rNTPs into the Sym/Sub assay. The design of the Sym/Sub oligonucleotide allows for elongation by a maximum of +4

nucleotides to produce a fully double stranded product (Figure 7A). Both the WT and the I189L 3D^{pol} enzymes produced the expected +4 product and, in addition, a longer product equivalent to +12 nucleotides was produced by both polymerases (as calculated by Rf value for migration travelled) (Figure 7B). This is, reminiscent of the larger products seen in recombination assays reported in the investigation of PV 3D^{pol} (48, 61). Although there was no measurable difference in rate of addition of the four nucleotides by WT or mutant enzyme, I189L 3D^{pol} produced the larger +12 product in greater quantity and faster (Figure 7C & D).

The functions of WT and I189L 3D^{pol} enzymes were dissected in finer detail by examining incremental addition of rNTP combinations to Sym/Sub assay in the order ATP, UTP, GTP and CTP. The reactions were incubated for 300 seconds (as above) before separation by electrophoresis. Addition of ATP alone produced a distinct +1 band, as previously described (Figure 8A). The addition of ATP and UTP together produced a +2 product in addition to lower intensity bands corresponding to +3 and +4 nucleotides. Inclusion of ATP, UTP and GTP resulted in a strong band equivalent to +3 product in addition to a minor +4 band. All four rNTPs produced a +4 band, as expected (Figure 8A). The appearance of the additional bands in reactions where some rNTPs were missing indicates misincorporation into the growing strand. We examined the error-rates of WT and I189L 3D^{pol} by measuring the relative amount of misincorporated +4 product and found no significant difference between the two enzymes (Figure 8B).

Differences in misincorporation by the WT and I189L mutant polymerase were investigated further by assessing the effects of ribavirin on replicons bearing either of the two enzymes as increased misincorporation rates should result in increased sensitivity to the nucleoside

analogue. However, no differences were seen in the sensitivity of the WT, P-97 and P-97 $3D^{pol}$ I189L replicons in the presence of up to 1 mM ribavirin (Figure 8C). Therefore, it appears that differences in error rate were unlikely to be responsible for the production of increased amounts of the larger +12 product by the I189L $3D^{pol}$ (Figure 7C and Figure 8D).

Discussion

In comparison to other picornaviruses, the 5′ UTR of FMDV is uniquely large and complex, comprising several distinct structural domains, the functions of most of which are poorly understood. The first *c.* 360 nucleotides are predicted to form a single, long stem-loop called the S fragment (22). Although supported by functional studies, to our knowledge the secondary structure of the S fragment has not been determined biochemically (25). Here, using SHAPE chemistry applied to a sequence of a well-studied FMDV isolate (O1 Kaufbeuren) (35), we confirm that the S fragment folds into a single, long stem-loop containing several bulges formed by unpaired nucleotides. The importance of some of these bulges for viral viability was determined previously (25).

Despite the high sequence and length variability of the S fragment observed in FMDV field isolates, most viruses maintain the full-length stem-loop (10). This suggests a strong evolutionary pressure to maintain the structure. The S fragment is known to interact with viral and cellular proteins and the 3′ UTR, likely facilitating circulation of the viral genome during replication (64–66). However, there is evidence that some deletions to the distal part of the S fragment are tolerated and a number of unrelated field isolates from different FMDV serotypes (e.g., A, C and O) have been shown to carry deletions in their S fragments. These deletions arose independently and range in length from few to 76 nucleotides (10, 26, 62).

Unlike deletions found in other parts of the FMDV genome (e.g., deletions found in the region encoding 3A), those found in the S fragment do not appear to be host specific, with viruses carrying deletions within the S fragment being isolated from both cattle and pig (10, 26, 62, 63). A more recent study showed that deletion of 164 nucleotides from the distal part of the S fragment can produce a viable, although attenuated virus, suggesting a role in evasion of innate immunity, at least partially explaining the evolutionary trend to maintain the full structure of the S fragment by field viruses (25). Here, we show that although deletion of 246 nucleotides or more from the distal region on the S fragment is not tolerated mutants carrying deletions D-70 and D-148 remain replication-competent.

The proximal part of the S fragment stem-loop shows higher sequence and nucleotide pairing conservation compared to the distal region (Figure 3). Surprisingly, deletions to the proximal part of the stem-loop (i.e., up to 97 nucleotides) are also tolerated, although to a lesser extent than distal deletions, resulting in both attenuated replicons and viruses (cf P-48 and P-97 mutants versus D-148 mutant, Figure 2 and 4) (10). Interestingly, while the replication of D-148 mutant replicon was more impaired in MDBK cells than in the BHK-21 cells, this cell-specific difference was not seen for the viable replicons carrying deletions to the proximal part of the S fragment stem-loop. Since, unlike BHK-21 cells, MDBK cells appear to have a functional interferon pathway this might suggest that, although required for viral replication, this proximal part of the S fragment stem-loop does not play a part in the evasion of the innate immunity (56, 57).

Of three viral isolates carrying the maximum viable proximal deletion to the S fragment (P-97), two resulted in selective enrichment of a compensatory mutation over serial passages. In

both cases, the mutations showed a trend towards fixation. One isolate developed a number of mutations in 2C (data not shown), while another developed a single I189L mutation in a highly conserved site of $3D^{pol}$ known to form hydrophobic contact with a nucleotide base within template RNA during viral replication (60). Given the essential role of isoleucine 189 in viral RNA replication, we investigated the effect of the I189L mutation on replication and RNA strand synthesis. In fact, introduction of I189L into the P-97 replicon enhanced its replication, suggesting an advantageous effect of this mutation. Although further biochemical analysis of the I189L $3D^{pol}$ mutant did not show an altered rate of nucleotide incorporation when compared to the WT polymerase, it did produce more of a 12-nucleotide long product. Such long products were observed in similar reactions for the $3D^{pol}$ of PV and were the result of a nucleotide mis-incorporation and template switch (48, 67, 68). While the 12-nucleotide long product most likely occurred due to nucleotide misincorporation, there is no evidence for altered fidelity of the I189L mutant. Whether this product occurred as the result of strand slippage, or a template switch remains to be investigated. Nevertheless, development of the compensatory mutation over a serial passage in a virus carrying maximal viable proximal deletion to the S fragment structure, and the advantageous effect of this mutation on replicon replication suggests the possibility of (direct or indirect) interaction between the S fragment and the viral polymerase (or its precursor). Further studies are required to verify this interplay.

In conclusion, using biochemical methods, we confirmed previously proposed secondary structure of the S fragment of FMDV. Despite its higher sequence and nucleotide pairing conservation, small deletions to the proximal part of the S fragment stem-loop are viable, although resulting in attenuated replicons and viruses. An advantageous compensatory mutation within a highly conserved site of $3D^{pol}$ appeared during serial passage of a viral

isolate bearing the largest viable mutation to the proximal part of the S fragment, suggests potential interplay between the S fragment and the viral polymerase.

Acknowledgements

This work was supported by funding from the Biotechnology and Biological Sciences Research Council (BBSRC) of the United Kingdom (research grant BB/K003801/1). Additionally, The Pirbright Institute receives grant-aided support from the BBSRC (projects BBS/E/I/00007035, BBS/E/I/00007036 and BBS/E/I/00007037) and the UK Department for the Environment, Food and Rural Affairs (Defra project SE2945).

1. Lulla V, Dinan AM, Hosmillo M, Chaudhry Y, Sherry L, Irigoyen N, Nayak KM, Stonehouse NJ, Zilbauer M, Goodfellow I, Firth AE. 2019. An upstream protein-coding region in enteroviruses modulates virus infection in gut epithelial cells. Nat Microbiol 4:280–292.
2. Roos,' RP, Kong W-P, Semler2 BL. 1989. Polyprotein processing of Theiler's murine encephalomyelitis virus. J Virol 63:5344.
3. Martínez-Salas E, Francisco-Velilla R, Fernandez-Chamorro J, Lozano G, Diaz-Toledano R. 2015. Picornavirus IRES elements: RNA structure and host protein interactions. Virus Res 206:62–73.
4. Lopez de Quinto S. 2002. IRES-driven translation is stimulated separately by the FMDV 3'-NCR and poly(A) sequences. Nucleic Acids Res 30:4398–4405.
5. Chard LS, Kaku Y, Jones B, Nayak A, Belsham GJ. 2006. Functional analyses of RNA structures shared between the internal ribosome entry sites of hepatitis C virus and the picornavirus porcine teschovirus 1 Talfan. J Virol 80:1271–1279.
6. Mason PW, Bezborodova S V, Henry TM. 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 76:9686–94.
7. Goodfellow IG, Kerrigan D, Evans DJ. 2003. Structure and function analysis of the poliovirus cis-acting replication element (CRE). RNA 9:124–137.
8. Goodfellow I, Chaudhry Y, Richardson A, Meredith J, Almond JW, Barclay W, Evans DJ. 2000. Identification of a cis-acting replication element within the poliovirus coding region. J Virol 74:4590–600.
9. Barton DJ, O'Donnell BJ, Flanegan JB. 2001. 5' cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis. EMBO J 20:1439–48.
10. Carrillo C, Tulman ER, Delhon G, Lu Z, Carreno A, Vagnozzi A, Kutish GF, Rock DL.

2005. Comparative genomics of foot-and-mouth disease virus. J Virol 79:6487–504.

11. Carocci M, Bakkali-Kassimi L. 2012. The encephalomyocarditis virus. Virulence 3:351–67.

12. Drexler JF, Corman VM, Lukashev AN, Van Den Brand JMA, Gmyl AP, Brunink S, Rasche A, Seggewi N, Feng H, Leijten LM, Vallo P, Kuiken T, Dotzauer A, Ulrich RG, Lemon SM, Drosten C, Akoua-Koffi CG, Ameneiros RS, Goodman SM, Godlevska E V., Tungaluna GCG, Görföl T, Hassanin A, Koivogui L, Krüger DH, Lakim MB, Lavrenchenko L, Leroy EM, Matthee S, ter Meulen J, Adu-Sarkodie Y, Oppong S, Preiser W, Pereira MJR, Reusken CBEM, Schneider J, Schmidt-Chanasit J, Fichet-Calvet E, Seebens-Hoyer A, Gloza-Rausch F, Setién AA, Tschapka M, Wells K, Wegner T, Wilkinson DA. 2015. Evolutionary origins of hepatitis A virus in small mammals. Proc Natl Acad Sci U S A 112:15190–15195.

13. Knight-Jones TJD, Rushton J. 2013. The economic impacts of foot and mouth disease – What are they, how big are they and where do they occur? Prev Vet Med 112:161–173.

14. Mahapatra M, Parida S. 2018. Foot and mouth disease vaccine strain selection: current approaches and future perspectives. Expert Rev Vaccines 17:577–591.

15. Park J-H. 2013. Requirements for improved vaccines against foot-and-mouth disease epidemics. Clin Exp Vaccine Res 2:8–18.

16. Stenfeldt C, Eschbaumer M, Rekant SI, Pacheco JM, Smoliga GR, Hartwig EJ, Rodriguez LL, Arzt J. 2016. The Foot-and-Mouth Disease Carrier State Divergence in Cattle. J Virol 90:6344–64.

17. Sweeney TR, Cisnetto V, Bose D, Bailey M, Wilson JR, Zhang X, Belsham GJ, 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.

18. Teterina NL, Gorbalenya AE, Egger D, Bienz K, Rinaudo MS, Ehrenfeld E. 2006. Testing the modularity of the N-terminal amphipathic helix conserved in picornavirus 2C proteins and hepatitis C NS5A protein. Virology 344:453–467.

19. Bienz K, Egger D, Troxler M, Pasamontes L. 1990. Structural organization of poliovirus RNA replication is mediated by viral proteins of the P2 genomic region. J Virol 64:1156–1163.

20. Black DN, Stephenson P, Rowlands DJ, Brown F. 1979. Sequence and location of the poly C tract in aphtho- and cardiovirus RNA. Nucleic Acids Res 6:2381–2390.

21. Mellor EJC, Brown F, Harris TJR. 1985. Analysis of the Secondary Structure of the Poly(C) Tract in Foot-and-Mouth Disease Virus RNAs. J Gen Virol 66:1919–1929.

22. Clarke BE, Brown AL, Currey KM, Newton SE, Rowlands DJ, Carroll AR. 1987. Potential secondary and tertiary structure in the genomic RNA of foot and mouth disease virus. Nucleic Acids Res 15:7067–7079.

23. Ward JC, Lasecka-Dykes L, Neil C, Adeyemi OO, Gold S, McLean-Pell N, Wright C, Herod MR, Kealy D, Warner E, Jackson T, King DP, Tuthill TJ, Rowlands DJ, Stonehouse NJ. 2022. The RNA pseudoknots in foot-and-mouth disease virus are dispensable for genome replication, but essential for the production of infectious virus. PLOS Pathog 18:e1010589.

24. Lasecka-Dykes L, Wright CF, Di Nardo A, Logan G, Mioulet V, Jackson T, Tuthill TJ, Knowles NJ, King DP. 2018. Full Genome Sequencing Reveals New Southern African Territories Genotypes Bringing Us Closer to Understanding True Variability of Foot-and-Mouth Disease Virus in Africa. Viruses 10.

25. Kloc A, Diaz-San Segundo F, Schafer EA, Rai DK, Kenney M, de los Santos T, 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.

26. Valdazo-González B, Timina A, Scherbakov A, Abdul-Hamid N, Knowles NJ, King DP. 2013. Multiple introductions of serotype O foot-and-mouth disease viruses into East Asia in 2010–2011. Vet Res 44:76.

27. Xiang W, Harris K, Alexander L, Wimmer E. 1995. Interaction between the 5'-terminal cloverleaf and 3AB/3CDpro of poliovirus is essential for RNA replication. J Virol 69:3658–3667.

28. Serrano P, Pulido MR, 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' end region. J Gen Virol 87:3013–22.

29. Forrest S, Lear Z, Herod MR, Ryan M, Rowlands DJ, Stonehouse NJ. 2014. Inhibition of the foot-and-mouth disease virus subgenomic replicon by RNA aptamers. J Gen Virol 95:2649–57.

30. Tulloch F, Pathania U, Luke GA, Nicholson J, Stonehouse NJ, Rowlands DJ, Jackson T, Tuthill T, Haas J, Lamond AI, Ryan MD. 2014. FMDV replicons encoding green fluorescent protein are replication competent. J Virol Methods 209C:35–40.

31. Herod MR, Ferrer-Orta C, Loundras E-A, Ward JC, Verdaguer N, Rowlands DJ, Stonehouse NJ. 2016. Both cis and trans Activities of Foot-and-Mouth Disease Virus 3D Polymerase Are Essential for Viral RNA Replication. J Virol 90:6864–6883.

32. 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 279:47212–21.

33. Bentham M, Holmes K, Forrest S, Rowlands DJ, Stonehouse NJ. 2012. Formation of Higher-Order Foot-and-Mouth Disease Virus 3Dpol Complexes Is Dependent on Elongation Activity. J Virol 86:2371–2374.

34. Herod MR, Tulloch F, Loundras E-A, Ward JC, Rowlands DJ, Stonehouse NJ. 2015. Employing transposon mutagenesis to investigate foot-and-mouth disease virus replication. J Gen Virol 96:3507–3518.

35. Herod MR, Gold S, LaseckaDykes L, Wright C, Ward JC, McLean TC, Forrest S, Jackson T, Tuthill TJ, Rowlands DJ, Stonehouse NJ. 2017. Genetic economy in picornaviruses: Foot-and-mouth disease virus replication exploits alternative precursor cleavage pathway. PLOS Pathog 13:e1006666.

36. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. Algorithms Mol Biol 6:1–14.

37. King AMQ, Blakemore WE, Ellard FM, Drew J, Stuart DI. 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.

38. Logan G, Freimanis GL, King DJ, Valdazo-González B, Bachanek-Bankowska K, Sanderson ND, Knowles NJ, King DP, Cottam EM. 2014. A universal protocol to generate consensus level genome sequences for foot-and-mouth disease virus and other positive-sense polyadenylated RNA viruses using the Illumina MiSeq. BMC Genomics 15:828.

39. Acevedo A, Andino R. 2014. Library preparation for highly accurate population sequencing of RNA viruses. Nat Protoc 9:1760–1769.

40. GitHub - najoshi/sickle: Windowed Adaptive Trimming for fastq files using quality.

41. Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428.

42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

43. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26:589–595.

44. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nat Biotechnol 29:24–26.

45. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

46. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

47. Jackson T, Sheppard D, Denyer M, Blakemore W, King AMQ. 2000. The Epithelial Integrin αvβ6 Is a Receptor for Foot-and-Mouth Disease Virus. J Virol 74:4949.

48. Arnold JJ, Cameron CE. 2000. Poliovirus RNA-dependent RNA polymerase (3D(pol)). Assembly of stable, elongation-competent complexes by using a symmetrical primer-template substrate (sym/sub). J Biol Chem 275:5329–36.

49. de la Higuera I, Ferrer-Orta C, Moreno E, de Ávila AI, Soria ME, Singh K, Caridi F, Sobrino F, Sarafianos SG, Perales C, Verdaguer N, Domingo E. 2018. Contribution of a Multifunctional Polymerase Region of Foot-and-Mouth Disease Virus to Lethal Mutagenesis. J Virol 92.

50. Lasecka-Dykes L, Tulloch F, Simmonds P, Luke GA, Ribeca P, Gold S, Knowles NJ, Wright CF, Wadsworth J, Azhar M, King DP, Tuthill TJ, Jackson T, Ryan MD. 2021. Mutagenesis Mapping of RNA Structures within the Foot-and-Mouth Disease Virus Genome Reveals Functional Elements Localized in the Polymerase (3Dpol)-Encoding Region. mSphere 6.

51. Kerpedjiev P, Hammer S, Hofacker IL. 2015. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. Bioinformatics 31:3377–3379.

52. Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display consensus

sequences. Nucleic Acids Res 18:6097.

53. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14:1188–1190.

54. Homan PJ, Favorov O V, Lavender CA, Kursun O, Ge X, Busan S, Dokholyan N V, Weeks KM. 2014. Single-molecule correlated chemical probing of RNA. Proc Natl Acad Sci U S A 111:13858–63.

55. Tulloch F, Pathania U, Luke GA, Nicholson J, Stonehouse NJ, Rowlands DJ, Jackson T, Tuthill T, Haas J, Lamond AI, Ryan MD. 2014. FMDV replicons encoding green fluorescent protein are replication competent. J Virol Methods 209:35–40.

56. Conzelmann KK. 2004. Reverse Genetics of Mononegavirales. Curr Top Microbiol Immunol 283:1–41.

57. Schlender J, Bossert B, Buchholz U, Conzelmann K-K. 2000. Bovine Respiratory Syncytial Virus Nonstructural Proteins NS1 and NS2 Cooperatively Antagonize Alpha/Beta Interferon-Induced Antiviral Response. J Virol 74:8234.

58. Luna VER, Luk ADH, Tyring SK, Hellman JM, Lefkowitz SS. 1984. Properties of bovine interferons. Exp 1984 4012 40:1410–1412.

59. Lasecka-Dykes L, Wright CF, Di Nardo A, Logan G, Mioulet V, Jackson T, Tuthill TJ, Knowles NJ, King DP. 2018. Full genome sequencing reveals new southern african territories genotypes bringing us closer to understanding true variability of foot-and-mouth disease virus in Africa. Viruses 10.

60. 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.

61. Woodman A, Arnold JJ, Cameron CE, Evans DJ. 2016. Biochemical and genetic analysis of the role of the viral polymerase in enterovirus recombination. Nucleic

Acids Res 44:6883–95.

62. Mohapatra JK, Pawar SS, Tosh C, Subramaniam S, Palsamy R, Sanyal A, Hemadri D, Pattnaik B. 2011. Genetic characterization of vaccine and field strains of serotype A foot-andmouth disease virus from India. Acta Virol 55:349–352.

63. Wright CF, Knowles NJ, Di Nardo A, Paton DJ, Haydon DT, King DP. 2013. Reconstructing the origin and transmission dynamics of the 1967–68 foot-and-mouth disease epidemic in the United Kingdom. Infect Genet Evol 20:230–238.

64. 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 83:11356–66.

65. Rodríguez-Pulido M, Borrego B, Sobrino F, Sáiz M. 2011. RNA Structural Domains in Noncoding Regions of the Foot-and-Mouth Disease Virus Genome Trigger Innate Immunity in Porcine Cells and Mice. J Virol 85:6492.

66. Rodríguez Pulido M, Serrano P, Sáiz M, Martínez-Salas E. 2007. Foot-and-mouth disease virus infection induces proteolytic cleavage of PTB, eIF3a,b, and PABP RNA-binding proteins. Virology 364:466–474.

67. Woodman A, Arnold JJ, Cameron CE, Evans DJ. 2016. Biochemical and genetic analysis of the role of the viral polymerase in enterovirus recombination. Nucleic Acids Res 44:6883–6895.

68. Arnold JJ, Cameron CE. 1999. Poliovirus RNA-dependent RNA polymerase (3D(pol)) is sufficient for template switching in vitro. J Biol Chem 274:2706–2716.

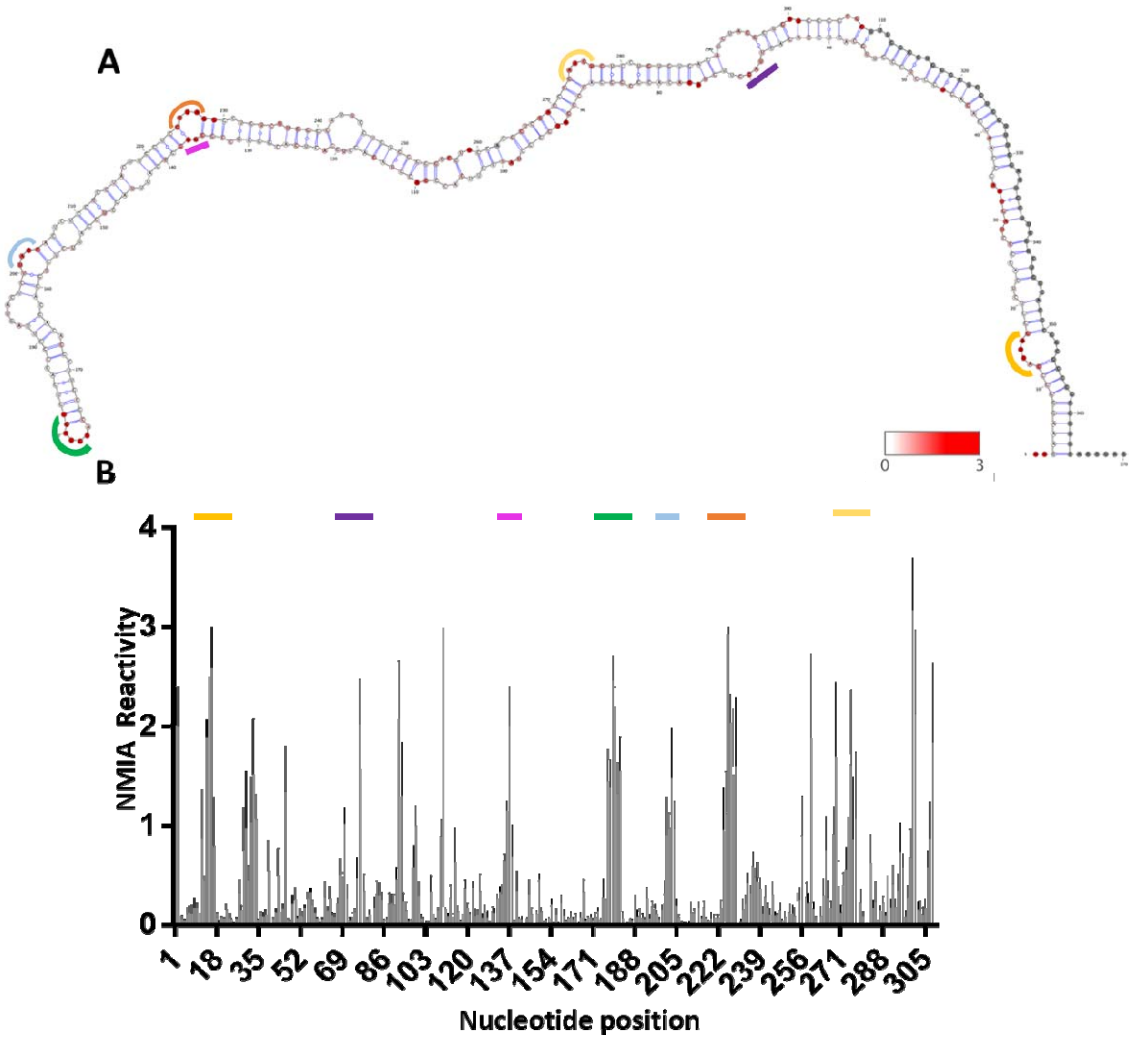


Figure 1. SHAPE analysis of the S fragment. (A) Superimposition of SHAPE reactivity on the *in silico* predicted secondary structure WT S fragment. Secondary structure was predicted using the Vienna RNA probing package and visualised using VARNA. NMIA reactivity is overlaid and represented on a colour scale from low (white) to high (red). Nucleotides for which there is no data are represented as grey. Coloured regions corresponding to the peaks in NIMIA reactivity are shown for ease of interpretation and localisation. **(B)** Individual nucleotide NMIA reactivity as analysed by SHAPE reactions and capillary electrophoresis. High reactivity indicates high probability of single-stranded regions i.e. non base-paired nucleotides. Data was analysed using QuSHAPE. Corresponding coloured regions are shown for some regions across **A** and **B** for ease of interpretation and localisation of SHAPE data to RNA structure. (n = 6, error bars represent SEM).

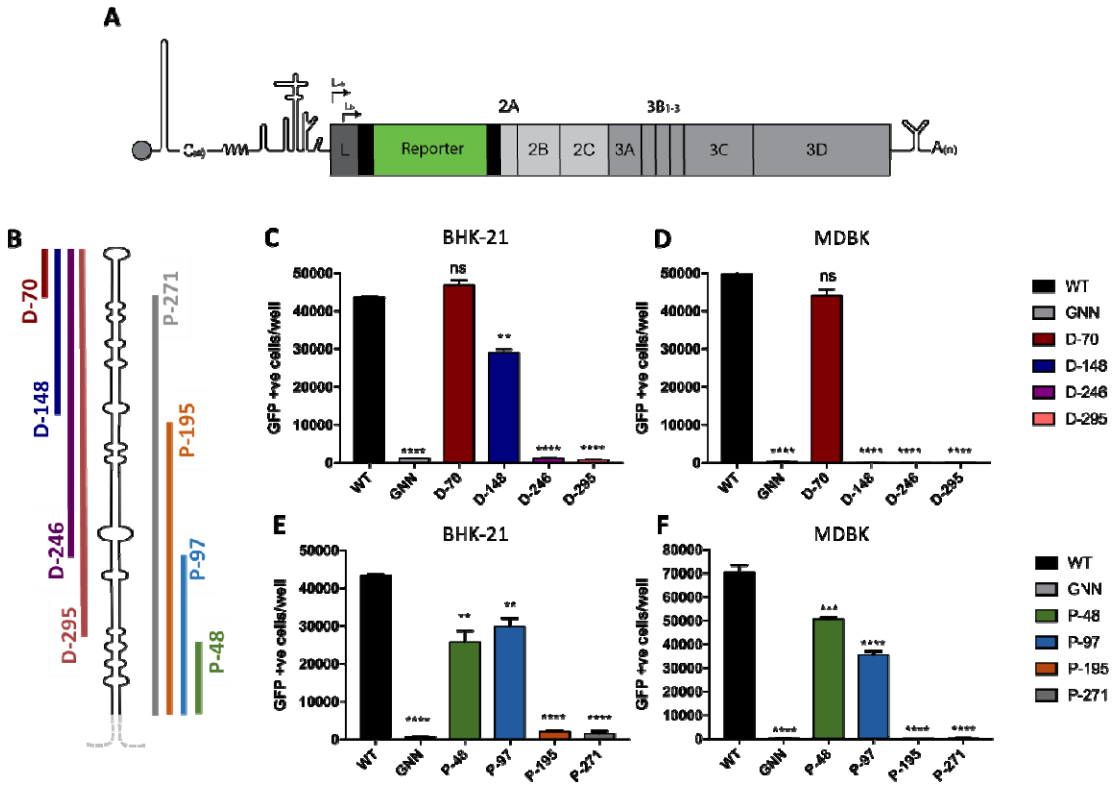


Figure 2. Truncations to the distal and proximal regions of the S fragment can impair replication. **(A)** Cartoon schematic of FMDV O1K replicon with the region encoding the structural proteins replaced with a GFP reporter. **(B)** Schematic representation of S fragment deletions. The maintained 3′ and 5′ proximal (P) regions are represented by the dotted grey line and was constant in all deletions **(C-D)** Replication of replicons with 70, 148, 246 and 295 nucleotides removed from the distal (D) region of the S fragment was measured following transfection into BHK-21 cells or MDBK cells. Replication was monitored by GFP expression using an IncuCyte, shown at 8 hours post-transfection alongside WT and $3D^{pol}$ inactivating mutant 'GNN' acting as a positive and negative control, respectively. **(E-F)** Replication of replicons with 48, 97, 195 and 271 nucleotides removed from the proximal region of the S fragment and measured as in (C). (n = 3, error bars represent SEM) ** P<0.01, **** P<0.0001.

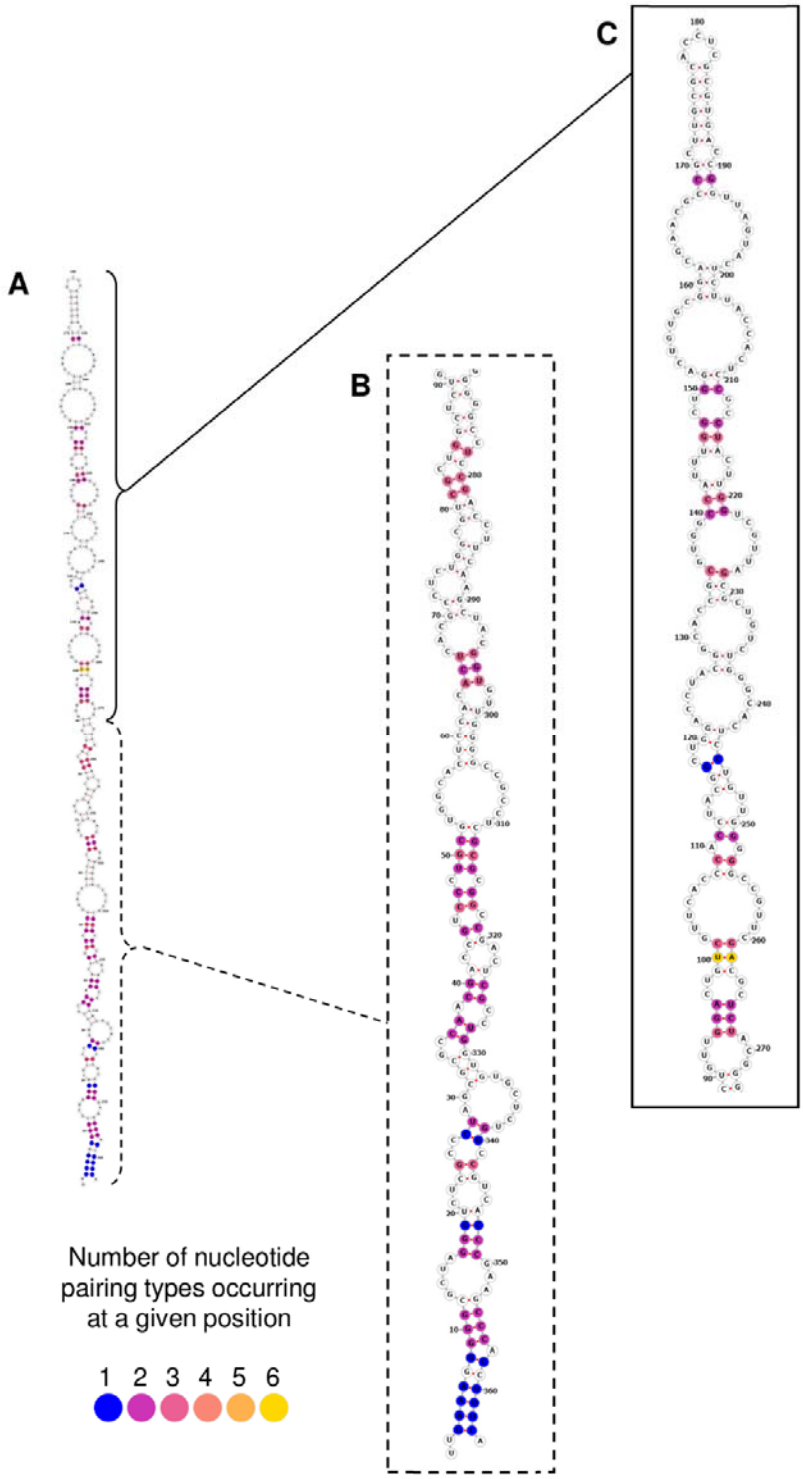


Figure 3. The extent of nucleotide and pairing conservation within the S fragment of the FMDV. (A) A schematic representation of conserved RNA structure of the S fragment in all 118 FMDV isolates. Nucleotide positions which form conserved pairing in 116/118 FMDV isolates were colour-coded according to number of pairing types (purple for 1 nucleotide pairing type at a given position and yellow for 6 nucleotide pairing types; see the colour matrix included at the left bottom corner of the

figure). There are six possible nucleotide pairings: A-U, G-C, G-U, U-A, C-G and U-G. Nucleotide pairings which were not conserved in three or more FMDV isolates, remained white. Due to the length of the S fragment stem-loop, the resolution of the image was not sufficient for detailed view of the individual nucleotide pairings and so the S fragment was artificially divided into two parts, proximal and distal, and the images of their schematic RNA structures enlarged (**B-C**). **(B)** The proximal part of the S fragment included nucleotide positions 1 to 90 and 272 to 364, **(C)** while the distal part included nucleotide positions 90 - 272. Numbers represent nucleotide positions of the S fragment sequence.

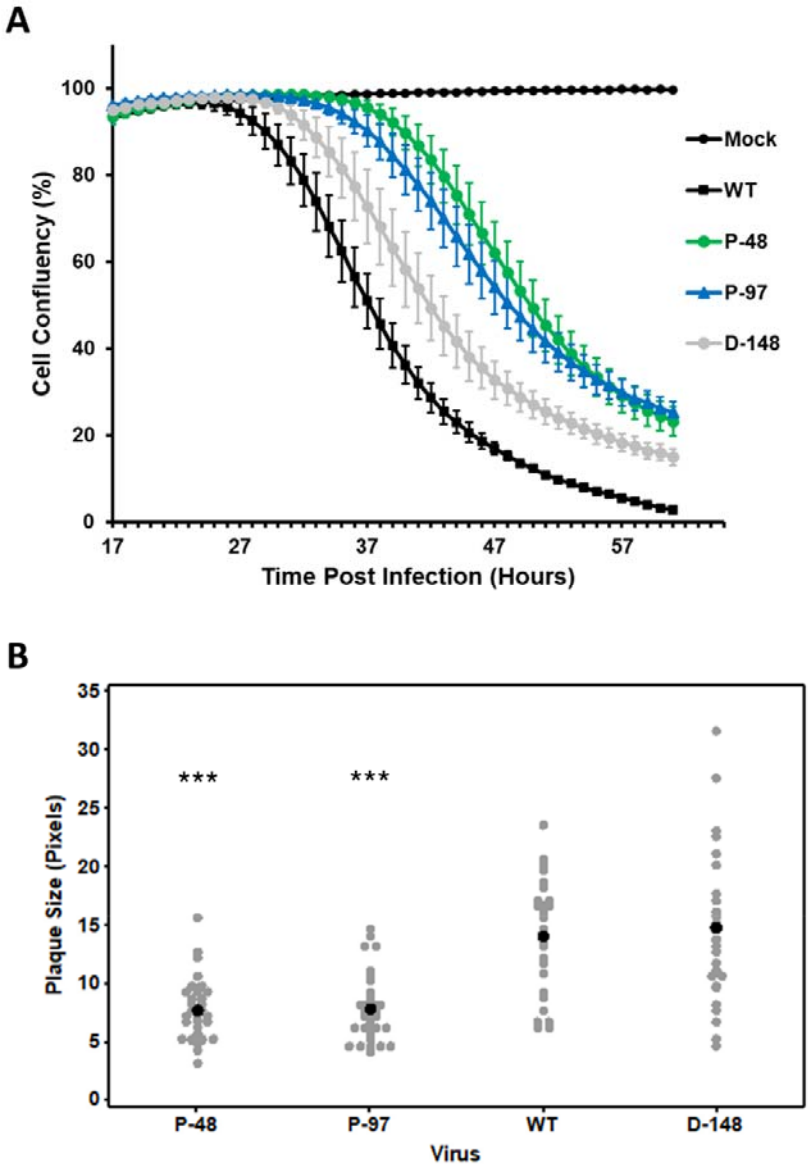


Figure 4. Fitness of FMDV containing a truncation at the proximal end of the S fragment. (A) Recovered WT, P-48, P-97 and D-148 virus populations were used to infect BHK-21 cells at an MOI of 0.01 PFU. CPE was monitored over time using live an IncuCyte S3 live cell imaging and is shown as a drop in cell confluency; (n =3 error bars represent SEM). **(B)** Plaque sizes of the WT, P-48, P-97 and D-148 variants grown on BHK-21 cells. Each wells containing plaques was scanned and the plaque size was estimated in pixels. All plaques were counted to avoid bias in plaque selection. Grey dots represent individual plaques, while black dot represents the mean; *** $P < 0.001$ when comparing to plaque size induced by the WT variant.

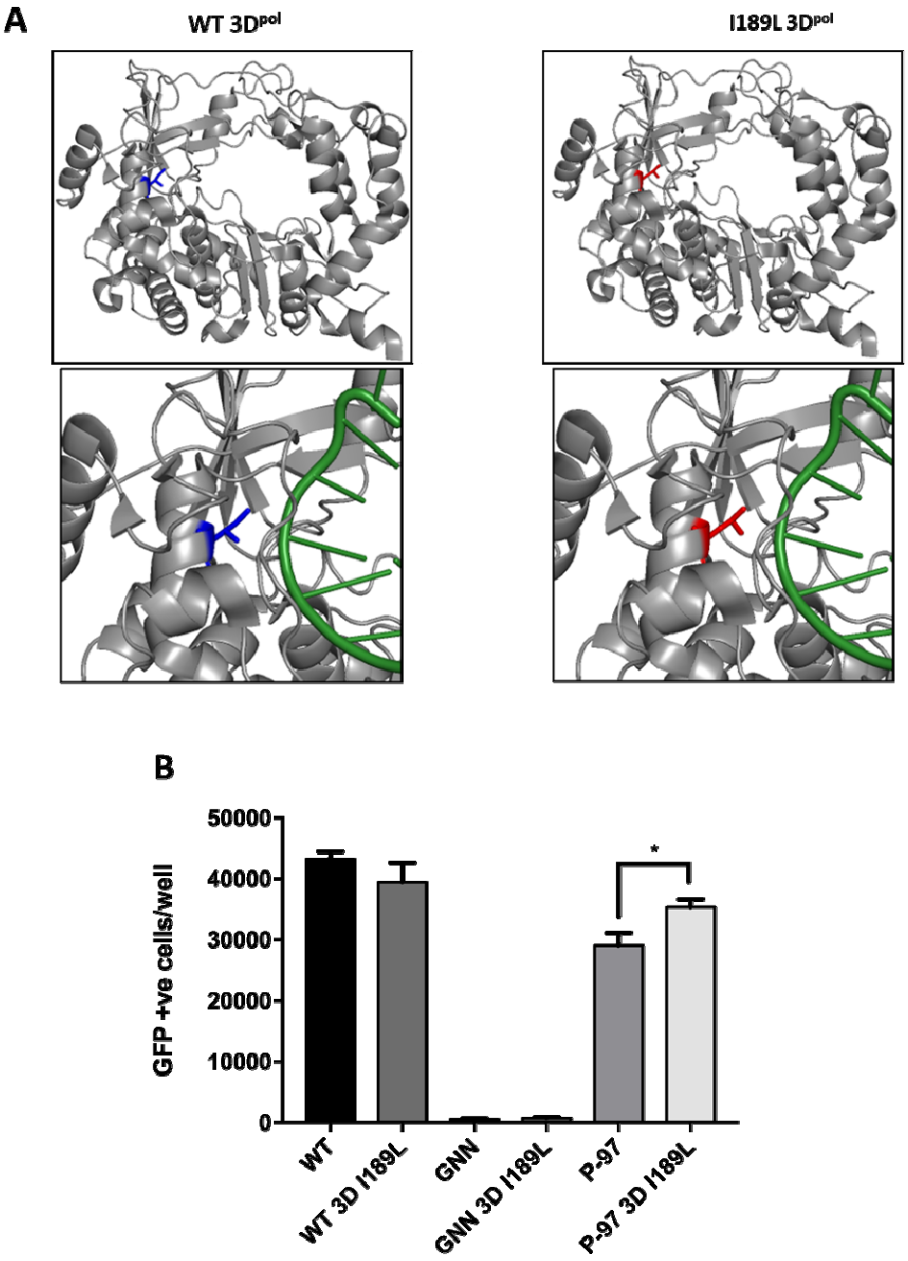


Figure 5. Compensatory mutation I189L in the structure of 3D^{pol} and its effect on replication of FMDV replicons. (A) WT 3D^{pol} crystal structure with I189 residue highlighted in blue, shown with (bottom, left) and without bound RNA (top, left). The I189L residue mutation was modelled onto the WT 3D^{pol} crystal structure and highlighted in red (right). Structure shown with (bottom, right) and without (top, right) bound RNA. PDB file 1WNE. **(B)** The 3D^{pol} I189L mutation was introduced into WT, GNN and P-97 replicons, termed WT I189L, GNN I189L and P-97 I189L respectively. Replicon RNA was transcribed and transfected into BHK-21 cells. WT and 3D^{pol} GNN replicons were included as controls, with the latter for the level of input translation (n =3 error bars represent SEM, * $P < 0.05$).

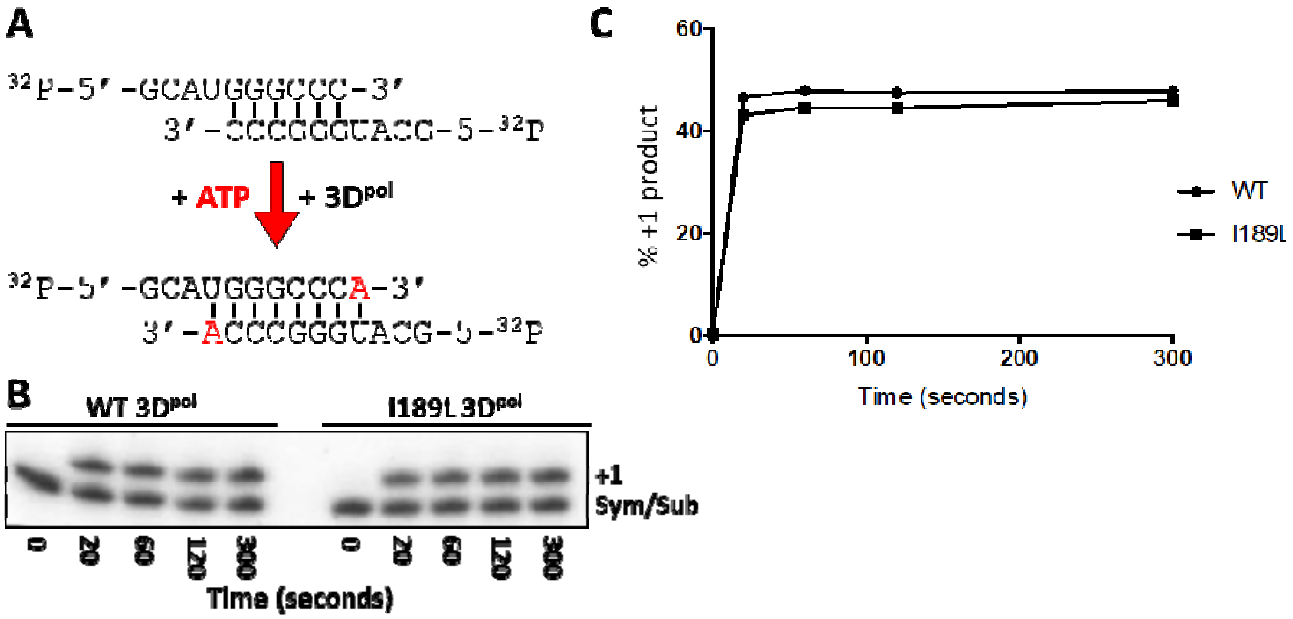


Figure 6. **Single nucleotide addition to a Sym/Sub substrate by WT and I189L 3Dpol.** **(A)** Schematic of the Sym/Sub experimental protocol. Radioactively end labelled RNA oligonucleotides are annealed before addition of rNTPs and recombinant $3D^{pol}$. **(B)** Extension of the 10mer Sym/Sub template with a single nucleotide (ATP) using either WT or I189L FMDV $3D^{pol}$. Aliquots of the reactions were taken at 20, 60, 120 and 300 seconds, RNA fragments separated by electrophoresis and visualised using a phosphoimager. **(C)** Densitometry of +1 product is plotted as rate of addition of a single nucleotide shown over time, %+1 product refers to the total amount of input Sym/Sub template that was extended by 1 nucleotide.

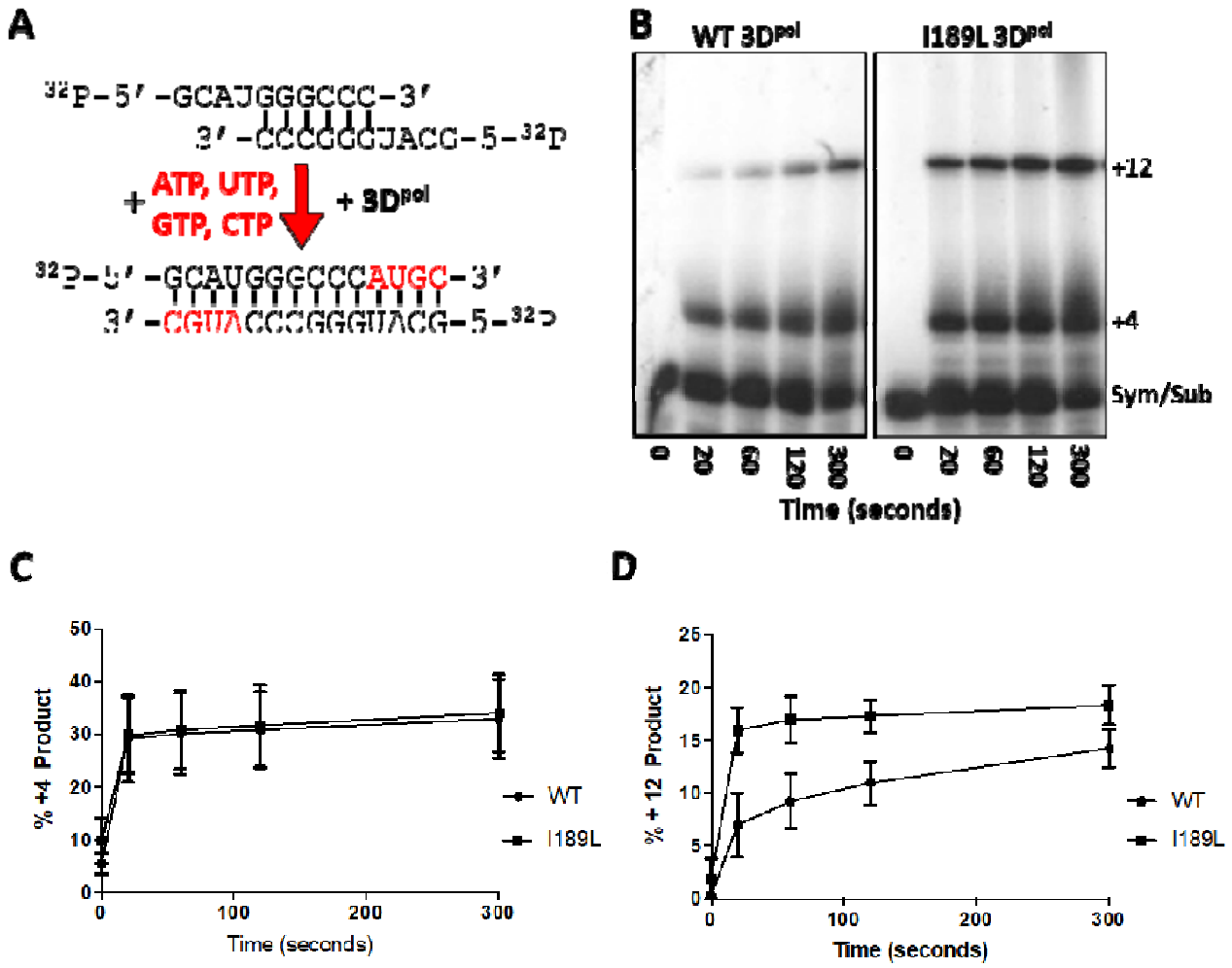


Figure 7. **Sym/Sub assays showing differences between WT and I189L 3D^{pol} in the production of a ~ +12 nucleotide product. (A)** Schematic of the Sym/Sub experimental protocol using all four rNTPs. **(B)** Extension of the 10mer Sym/Sub template with all four nucleotides using either WT or I189L recombinant FMDV 3D^{pol}. Aliquots of the reactions were taken at 20, 60, 120 and 300 seconds, RNA fragments separated by electrophoresis and visualised using a phosphoimager. **(C)** Densitometry of the production of the +4 product by either WT or I189L 3D^{pol} over time. **(D)** Densitometry of the production of the +12 product by either WT or I189L 3D^{pol}. Data are shown as the % of the input Sym/Sub template elongating to the +4 or +12 products (n = 3, error bars represent SEM).

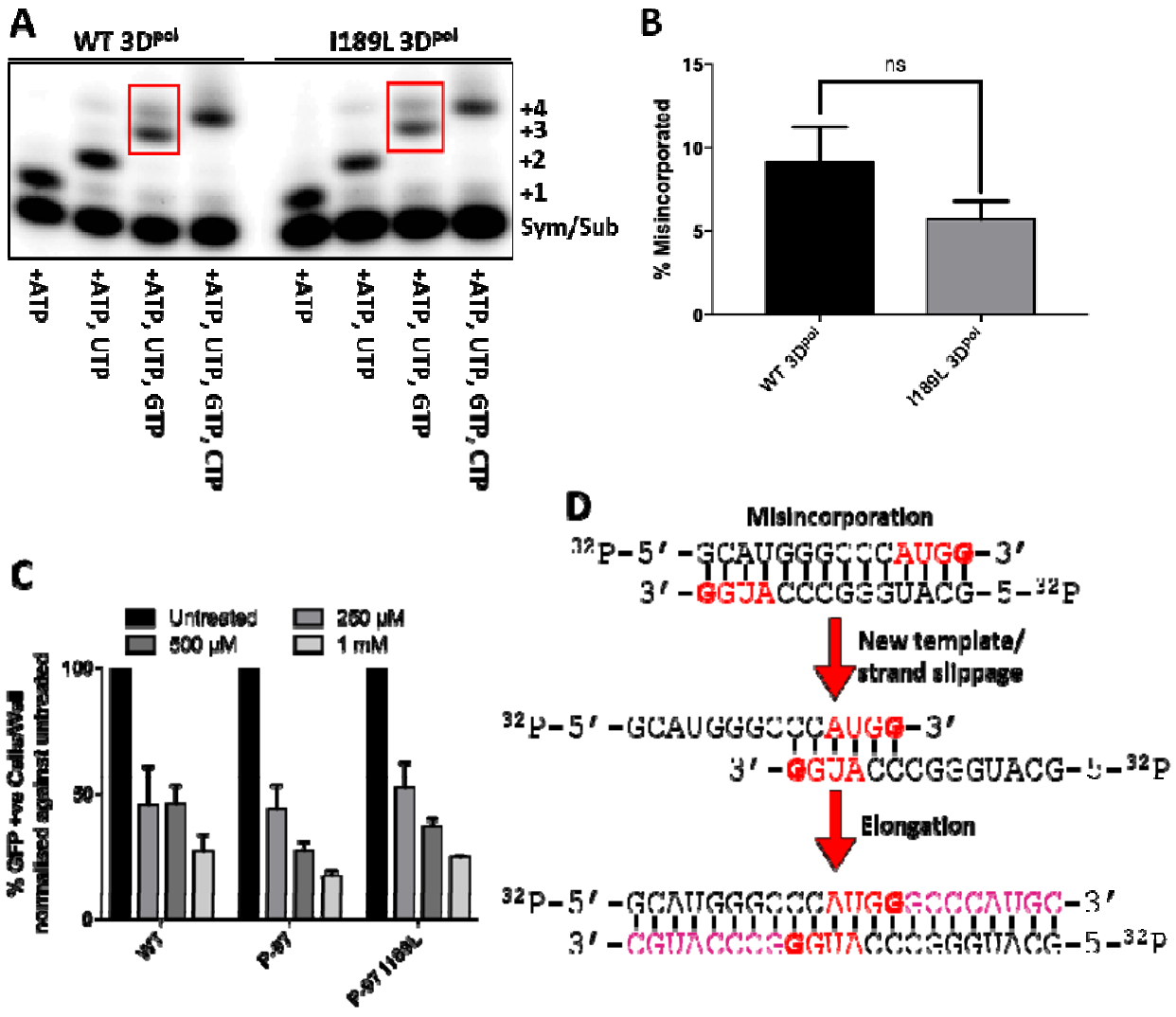
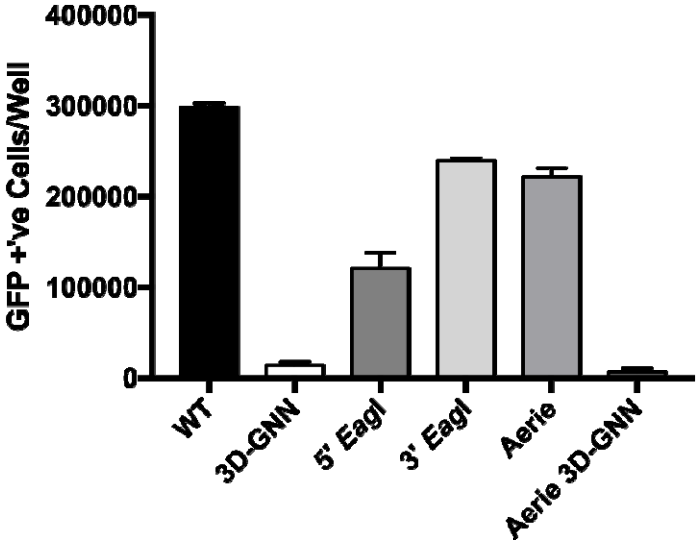
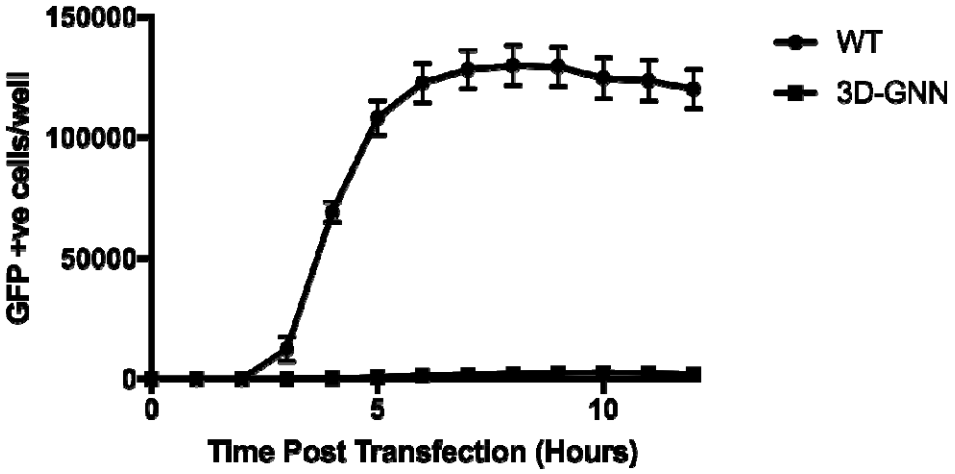


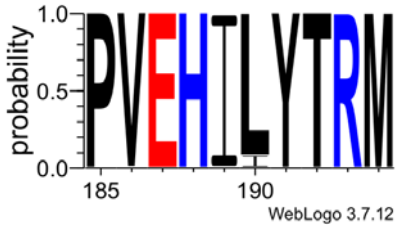
Figure 8. Misincorporation by WT and I189L 3D^{pol}. (A) Sym/Sub experiments were repeated with the addition of either ATP, ATP and UTP, ATP and UTP and GTP or all four rNTPs together. Reactions were continued for 300 seconds before electrophoresis and visualization. Misincorporation of GTP to generate a +4 product is highlighted by the red box. **(B)** Densitometry of the misincorporated GTP shows no difference in the rate of misincorporation by WT or I189L 3D^{pol}. **(C)** WT, P-97 and P-97 I189L replicons were transfected into BHK-21 cells in the presence of up to 1 mM of ribavirin. Replication was monitored by expression of the GFP reporter and shown here at 8 hours post-transfection. **(D)** Schematic for potential mechanism for the generation of a +12 product. Since misincorporation can occur at the fourth position allowing for the addition of a second guanosine residue, if templates were to separate and reanneal by either strand slippage or annealing of a new template, extension could result in the addition of 12 nucleotides (n = 3, error bars represent SEM).



Supplementary Figure S1. Effect of introduction of *Eag*I sites on replication. Number of GFP positive cells/well detected at 8 hours post-transfection of BHK-21 cells. WT represents the original O1K replicon sequence. 3D-GNN is the polymerase defective negative control. 5′ *Eag*I replicon possesses mutations to the sequence to create an *Eag*I site at the 5′ end of the S fragment. 3′ *Eag*I possesses mutations to the sequence to create and *Eag*I site at the 3′ end of the S fragment. Aerie shows replicon with both 5′ and 3′ sites modified and Aerie 3D-GNN comprises both *Eag*I sites and 3D-GNN mutation. n=2, error bars represent SEM.



Supplementary Figure S2. Time course of FMDV replicon replication. Number of GFP positive cells/well detected at hourly time points post-transfection of WT and 3D-GNN replicons in BHK-21 cells. n=2, error bars represent SEM.



Supplementary Figure S3. Conservation of the isoleucine at the position 189 (I189) within the 3D^{pol} protein of FMDV. Sequence logo based on 1123 sequences of FMDV 3D^{pol} available on fmdbase.org and prepared using WebLogo 3 server. The x-axis represents amino acids positions of the 3D^{pol} protein of FMDV, while the y-axis shows probability of a particular amino acid being present at a given position. Amino acids are colour-coded according to their charge.

Supplementary Table S1

FMDV sequences selected from GenBank

Serotype:	Number of isolates:	GenBank accession numbers:
A	19	AY593788, MH053305, JF749843, HM854024, HQ832580, MH053306, KM268896, AY593802, KJ608371, MH053307, AY593751, AY593754, AY593761, AY593764, AY593766, AY593767, HM854022, AY593791, AY593794
Asia 1	12	AY593795, AY687334, DQ533483, DQ989306, DQ989315, DQ989319, EF149010, EF614458, HQ632774, JF739177, KM268898, MF782478
C	6	MH053308, KM268897, MH053309, AJ133357, MH053310, AJ007347
O	21	AY593819, MH053313, MH053311, MH053312, KF112885, KJ206909, HQ632769, HQ632771, KU291242, KR401154, GU384683, KF694737, AJ539140, MH053315, JX040491, MH053317, MH053318, MH053316, KJ560291, DQ404170, KU821591
SAT 1	21	AY593838, AY593845, MH053319, AY593844, JF749860, MH053321, AY593846, AY593839, AY593842, AY593841, AY593840, MH053322, AY593843, KM268899, MH053323, MH053324, MH053325, MH053326, MH053327, MW355668, MW355669
SAT2	19	MH053330, MH053332, MH053328, MH053329, JX014255, MH053333, AY593849, JX014256, MW355670 - MW355673， AY593847, MH053335, KM268900, JF749862, MH053336, MH053337, KU821592,
SAT3	20	AY593853, AY593851, MH053339, MH053340, MH053344, MH053343, AY593850, KJ820999, MH053341, MH053351, KX375417, KM268901, MH053350, MW355674 - MW355680