

The ‘cleavage’ activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring ‘2A-like’ sequences

Michelle L. L. Donnelly,^{1†} Lorraine E. Hughes,¹ Garry Luke,¹ Heidi Mendoza,¹ Edwin ten Dam,¹ David Gani² and Martin D. Ryan¹

¹ Centre for Biomolecular Sciences, School of Biology, Biomolecular Sciences Building, University of St Andrews, North Haugh, St Andrews KY16 9ST, UK

² The University of Birmingham, The School of Chemistry, Edgbaston, Birmingham B15 2TT, UK

The 2A/2B cleavage of aphtho- and cardiovirus 2A polyproteins is mediated by their 2A proteins ‘cleaving’ at their own C termini. We have analysed this activity using artificial reporter polyprotein systems comprising green fluorescent protein (GFP) linked via foot-and-mouth disease virus (FMDV) 2A to β -glucuronidase (GUS) – forming a single, long, open reading frame. Analysis of the distribution of radiolabel showed a high proportion of the *in vitro* translation products (~90%) were in the form of the ‘cleavage’ products GUS and [GFP2A]. Alternative models have been proposed to account for the ‘cleavage’ activity: proteolysis by a host-cell proteinase, autoproteolysis or a translational effect. To investigate the mechanism of this cleavage event constructs encoding site-directed mutant and naturally occurring ‘2A-like’ sequences were used to program *in vitro* translation systems and the gel profiles analysed. Analysis of site-directed mutant 2A sequences showed that ‘cleavage’ occurred in constructs in which all the candidate nucleophilic residues were substituted – with the exception of aspartate-12. This residue is not, however, conserved amongst all functional ‘2A-like’ sequences. ‘2A-like’ sequences were identified within insect virus polyproteins, the NS34 protein of type C rotaviruses, repeated sequences in *Trypanosoma* spp. and a eubacterial α -glucosiduronase sequence (*Thermatoga maritima aguA*). All of the 2A-like sequences analysed were active (to various extents), other than the eubacterial α -glucosiduronase 2A-like sequence. This method of control of protein biogenesis may well not, therefore, be confined to members of the *Picornaviridae*. Taken together, these data provide additional evidence that neither FMDV 2A nor ‘2A-like’ sequences are autoproteolytic elements.

Introduction

Polyprotein processing is, perhaps, synonymous with *proteolytic* processing – cleavage of the polyprotein at specific sites by either virus-encoded or host-cell proteinases. The primary 2A/2B cleavage of the aphtho- and cardioviruses is mediated by 2A ‘cleaving’ at its own C terminus. In aphthoviruses (the foot-and-mouth disease viruses and equine rhinitis A virus) the 2A region is very short (~18 aa) and, together with the N-terminal residue of protein 2B (a conserved proline residue), represents an autonomous element capable of

mediating ‘cleavage’ at its own C terminus. Since 2A has been used for a variety of biotechnological purposes its activity has been examined in a wide range of heterologous protein contexts and, to our knowledge, it is active in all eukaryotic systems analysed thus far (Ryan & Drew, 1994; Precious *et al.*, 1995; Schmidt & Rethwilm, 1995; Mattion *et al.*, 1996; Camon *et al.*, 1997; Roberts *et al.*, 1997; Cruz *et al.*, 1997; van der Ryst *et al.*, 1998; Collins *et al.*, 1998; Smolenska *et al.*, 1998; De Felipe *et al.*, 1999; Chaplin *et al.*, 1999; Smerdou & Liljestrom, 1999; Varnavski & Khromykh, 1999; Kokuhoh *et al.*, 1999; De Felipe & Izquierdo, 2000; De Rose *et al.*, 2000; Gopinath *et al.*, 2000; O’Brien *et al.*, 2000; Suzuki *et al.*, 2000; Thomas & Maule, 2000; Varnavski *et al.*, 2000).

We have shown that the C-terminal 19 aa of the longer cardiovirus 2A protein (together with the N-terminal proline of 2B) mediate ‘cleavage’ with an efficiency approximately

Author for correspondence: Martin Ryan.

Fax +44 1334 463400. e-mail martin.ryan@st-and.ac.uk

† Present address: Marie Curie Research Institute, The Chart, Oxted, Surrey RH8 OTL, UK.

Table 1. FMDV 2A constructs

The sequence of the FMDV 2A inserted between the *Xba*I and *Apal* restriction sites (underlined) is shown (pSTA1). Introduced mutations are shown in bold type (amino acids) and underlined (nucleotides). The 'cleavage' activity of each construct is also shown.

Construct	'Cleavage' (%)	
pSTA1	~90	
pSTA1	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 S R Q L L N F D L L K L A G D V E S N P G P <u>TCTAGACAGCTGTTGAATTTGACCTCTTAAGCTTGCGGGAGACGTCGAGTCCAACCCCGGGCCC</u> <i>Xba</i> I ApaI	
Aspartate 12;		
pSTA1/1	0	
pMD 2.7.15	0	
Glutamate 14;		
pSTA1/2	0	
pSTA1/3	~56	
Glutamate 14 / Asparagine 16;		
pSTA1/4	0	
pSTA1/5	0	
pSTA1/6	0	
pSTA1/7	0	
pSTA1/8	0	
pSTA1/9	0	
pSTA1/10	0	
Serine 15;		
pSTA1/11	~42	
pSTA1/12	~39	
Asparagine 16;		
pSTA1/13	~31	
pSTA1/14	~19	
pSTA1/15	~10	
Proline 17;		
pSTA1/16	0	
pSTA1/17	0	
pSTA1/18	0	
Glycine 18;		
pSTA1/19	0	
pSTA1/20	0	

Table 1 (cont.)**Proline 19;**

pSTA1 / 21	Q L L N F D L L K L A G D V E S N P G G CAGCTGTTGAATTTGACCTCTTAAGCTGCGGGAGACGTCGAGTCCAACCCCTGGAGGC	~ 11
pMD2.3.1	Q L L N F D L L K L A G D V E S N P G A CAGCTGTTGAATTTGACCTCTTAAGCTGCGGGAGACGTCGAGTCCAACCCCTGGGCT	0
pSTA1 / 22	Q L L N F D L L K L A G D V E S N P G S CAGCTGTTGAATTTGACCTCTTAAGCTGCGGGAGACGTCGAGTCCAACCCCTGGGAGC	0
pMD2.3.7	Q L L N F D L L K L A G D V E S N P G I CAGCTGTTGAATTTGACCTCTTAAGCTGCGGGAGACGTCGAGTCCAACCCCTGGGATT	0
pMD2.3.9	Q L L N F D L L K L A G D V E S N P G F CAGCTGTTGAATTTGACCTCTTAAGCTGCGGGAGACGTCGAGTCCAACCCCTGGGTTT	0

Insertion Mutants;

pSTA1 / 25	Q L L N F D L L K L P A G D V E S N P G P CAGCTGTTGAATTTGACCTCTTAAGCTCCCGCGGGAGACGTCGAGTCCAACCCCTGGGCC	0
pSTA1 / 26	Q L L N F D L L K L A A G D V E S N P G P CAGCTGTTGAATTTGACCTCTTAAGCTGCCCGGGAGACGTCGAGTCCAACCCCTGGGCC	0
pMD 3/11(a)	Q L L N F D L P L K L A G D V E S N P G P CAGCTGTTGAATTTGACCLTCCCTTAAGCTGCCGGAGACGTCGAGTCCAACCCCTGGGCC	0

N-Terminal Extensions / Deletions;

pSTA1 / 31 (+39aa 1D)	S R V T E L L Y R M K R A E T Y C P R P L L A I H TCTAGAGTCACCGAGTTGCTTTACGGATGAAGAGGGCGAAAACATACTGTCCAAGGCCCTTGCTGCCAATCCAC P T E A R H K Q K I V A P V K Q T L N F D L L K L CCAACTGAAGCCAGACAAACAGAAAATTGTGGCACCGGTGAAACAGACTTTGAATTTGACCTCTCAAGTTG A G D V E S N P G P GCCGGAGACGTCGAGTCCAACCCCTGGGCC	> 99
pSTA1 / 32 (+21aa 1D)	S R L L A I H P T E A R H K Q K I V A P V K Q T L TCTAGATTGCTGCCAATCCACCAACTGAAGCCAGACAAACAGAAAATTGTGGCACCGGTGAAACAGACTTTG N F D L L K L A G D V E S N P G P AATTTGACCTCTCAAGTTGGCGGGAGACGTCGAGTCCAACCCCTGGGCC	> 99
pSTA1 / 33 (+14aa 1D)	S R E A R H K Q K I V A P V K Q T L N F D L L K L TCTAGAGAAGCCAGACAAACAGAAAATTGTGGCACCGGTGAAACAGACTTTGAATTTGACCTCTCAAGTTG A G D V E S N P G P GCCGGAGACGTCGAGTCCAACCCCTGGGCC	> 99
pSTA1 / 34 (+5aa 1D)	S R A C A P V K Q T L N F D L L K L A G D V E S N P G P TCTAGAGCATGCCACGGTGAACAGACTTTGAATTTGACCTCTCAAGTTGGCGGGAGACGTCGAGTCCAACCTGGGCC	~ 96
pSTA1 / 35	S R - - - - - - - - - - L K L A G D V E S N P G P TCTAGA-----CTTAAGCTTGGCGGGAGACGTCGAGTCCAACCCCTGGGCC	~ 8
pSTA1 / 36	S R - - - - - - - - - - L A G D V E S N P G P TCTAGA-----CTTGCGGGAGACGTCGAGTCCAACCCCTGGGCC	0

equal to the aphthovirus (foot-and-mouth disease virus; FMDV) 2A sequence (Donnelly *et al.*, 1997). The characteristics of the 2A-mediated 'cleavage' in heterologous protein contexts are: (i) it occurs co-translationally, the small proportion of uncleaved translation product (~ 10%) not subsequently cleaving (Ryan & Drew, 1994); (ii) it functions in all eukaryotic expression systems tested thus far, but not in prokaryotes (Ryan & Drew, 1994; Donnelly *et al.*, 1997); (iii) the C-terminal region of the 2A protein from other picornaviruses (cardioviruses) functions in a similar manner (Donnelly *et al.*, 1997); (iv) upstream sequences are influential in, but not critical

for, 'cleavage' (Ryan *et al.*, 1991; Donnelly *et al.*, 1997); and (v) the 'cleavage' is not achieved by proteolysis of the polyprotein but by a translational effect [Ryan *et al.*, 1999; Donnelly *et al.*, 2001 (accompanying paper)].

To test the 'self-cleaving' hypothesis a range of synthetic peptides corresponding to FMDV 2A was synthesized and the potential autoproteolytic property tested under a wide range of incubation conditions without success (Ryan *et al.*, 1999). The C-terminal regions of the 2A proteins of the cardioviruses encephalomyocarditis virus (EMCV) and Theiler's murine encephalitis virus (TMEV) are similar to FMDV 2A in both

Table 2. 2A-like constructs

The sequence of the FMDV 2A sequence inserted between GFP and GUS is shown (pSTA1) together with the restriction sites used to clone the 2A-like sequences. The ds-oligonucleotide sequences used in this study are shown together with the 2A-like oligopeptides they encode.

Construct	Activity (%)	Sequence
pSTA1	~90	Q L L N F D L L K L A G D V E S N P G P <u>TCTAGA</u> CAGCTGTTGAATTGACCTCTTAAGCTTGCAGTCAGTCCACCC <u>GGGCC</u> XbaI ApaI
pSTA1-TMEV	~65	Y H A D Y Y K Q R L I H D V E M N P G P 5'-CTAGATACCATGCTGACTACTACAAACAGAGACTCATACATGATGATGAGAAATGAACCCGGGC-3' 3'- TATGGTACGACTGATGATGTTGCTCTGAGATGACTACATCTTACTTGGGC -5'
pSTA1-EMCV	~91	H Y A G Y F A D L L I H D I E T N P G P 5'-CTAGACACTACCGTGGTACTTGGGACCTACTGATTGATGACATTGAGACAAATCCGGGC-3' 3'- TGTGATGCGACCAATGAAACGCCCTGGATGACTAAGTACTGTAACTCTGTTAGGTC -5'
pSTA1-ERAV	~99	Q C T N Y A L L K L A G D V E S N P G P 5'-CTAGACAGTGTACTAAATTATGCTCTTGAATTGCTGGAGATGTTGAGAGCAACCCGGGC-3' 3'- TGTGACATGATTAATACGAGAGAACTTAAACCGACCTCTACAACTCTCGTTGGGC -5'
pSTA1-PTV1	~94	A T N F S L K Q A G D V E E N P G P 5'- CTAGAGCCACGAACTCTCTCTGTTAAAGCAAGCAGGAGATGTTGAGAAAACCCGGGC-3' 3'- TCGGTGCAAGAAGAGAGACAATTGCTGCTCTACAACTCTTTGGGC -5'
pSTA1-DrosC	~95	A A R Q M L L L S G D V E T N P G P 5'- CTAGAGCTGCACGTCAAGATGTTGCTCTTGAATTGAGAGATGTTGAGACTAACCTGGGC-3' 3'- TCGACGTGCAAGTCTACAACGAGAACATAGTCTCTACAACTCTGATTGGAC -5'
pSTA1-TaV	>99	R A E G R G S L L T C G D V E E N P G P 5'-CTAGAAAGGCCAGGGCAGGGGAAGTCTCTAAACATGCGGGAGCTGGAGGAAAATCCGGGC-3' 3'- TTCTCGGCTCCCTCTCAAGAAGATTGTAACGCCCCCTGCACCTCTTTAGGAC -5'
pSTA1-IFV	~63	T R A E I E D E L I R A G I E S N P G P 5'-CTAGAACGAGGGCGGAGATTGAGGATGAATTGATTGCTGAGGAATTGAATCAAATCCTGGGC-3' 3'- TTGCTCCCGCCTCTAACTCTACTTAACAACTAACGAGCTCTAACTTAGTTAGGAC -5'
pSTA1-IFV(D)	Inactive	T R A E I E D E L I R A D I E S N P G P 5'-CTAGAACGAGGGCGGAGATTGAGGATGAATTGATTGCTGAGACATTGAATCAAATCCTGGGC-3' 3'- TTGCTCCCGCCTCTAACTCTACTTAACAACTAACGAGCTCTAACTTAGTTAGGAC -5'
pSTA1-P.Rota	~31	A K F Q I D K I L I S G D V E L N P G P 5'-CTAGAGCTAAATTCAAATCGATAAAATTAAATTCTGGAGACGTCGAATTGAATCCTGGGC-3' 3'- TCGATTTAAGGTTAGCTATTAAATTAAAGACCTCTGCAGCTTAACCTAGGAC -5'
pSTA1-TSR1	~18	S S I I R T K M L V S G D V E E N P G P 5'-CTAGAACGAGTATCCGCACTAACAGATGCTGGTGCCTGATGTTGAGAACATCCGGGC-3' 3'- TTGCTCATAGTAGGCCGTATTCTACGACCAACAGGCCACTAACCTCTTTAGGAC -5'
pSTA1-APendo	~69	C D A Q R Q K L L S G D I E Q N P G P 5'-CTAGATGTGACCGCAGACGACAAAGCTACTGCTAACGCGAGACATTGAGCAGAACCCAGGGGC-3' 3'- TACACTGCGCGTGTGCTTTGATGACGATTGCGCTCTGTAACTCGTCTGGTC -5'
pSTA1-Therm	Inactive	Y I P D F G G F L V K A D S E F N P G P 5'-CTAGATATATTCCAGATTGAGGATTCTGTCAAAGCGATCTGAGTTCAATCCTGGGC-3' 3'- TATATAAGGTCTAAACCTCTAAAGAACAGTTCGGCTAAGACTCAAGTTAGGAC -5'

sequence and 'cleavage' activity (Donnelly *et al.*, 1997). Mutagenesis of this region of EMCV 2A showed that the motif common to EMCV and FMDV 2A proteins (-DxExNPG⁺P-) was very sensitive to substitution, only the glutamate to aspartate mutation showing some activity (Hahn & Palmenberg, 1996). Not surprisingly, residues which showed natural sequence variation between EMCV and FMDV proved to be more mutable. Dynamic molecular modelling studies had indicated that the majority of FMDV 2A (-NFDLLKLAGDVES-) could adopt an amphipathic helical conformation *ab initio*, whilst the sequence immediately N-terminal of the 'cleavage' site could adopt a tight-turn (-NPG⁺P-; Ryan *et al.*, 1999). In considering the autoproteolytic model we were particularly interested in the possible arrangements of residues known to act as nucleophiles within proteinases and the role of the asparagine, since this residue may also cleave peptide bonds (Geiger & Clarke, 1987; Klotz & Thomas, 1993).

There were, therefore, a number of both structural and mechanistic aspects to our models that we wished to test using site-directed mutagenesis of the FMDV 2A sequence. To this end a number of silent nucleotide substitutions were made within the 2A coding sequence to facilitate these mutagenic studies. Such changes did not alter the observed 'cleavage' activity between the native FMDV 2A sequence (Ryan *et al.*, 1991) and the (silently) mutated form (Ryan & Drew, 1994). Alignment of aphtho- and cardiovirus 2A sequences shows a conserved -DxExNPGP- motif which we, and others, have shown to be intimately involved in the 'cleavage' activity (Hahn & Palmenberg, 1996; Donnelly *et al.*, 1997). To augment our site-directed mutant database and to determine if this type of control of protein biogenesis is confined to the picornaviruses or is a more widely adopted strategy, we probed the databases for the occurrence of '2A-like' sequences, using the -DxExNPGP- motif as the probe. Indeed, other '2A-like' sequences were found to be present within the database:

picornaviruses other than aphtho- or cardioviruses, 'picorna-virus-like' insect viruses, type C rotaviruses, repeated sequences within *Trypanosoma* spp. and a bacterial sequence. A series of constructs was produced encoding a single open reading frame (ORF) comprising green fluorescent protein (GFP) linked to β -glucuronidase (GUS) via either a site-directed mutant form of FMDV 2A (Table 1) or a '2A-like sequence' (Table 2). These constructs were used to programme *in vitro* translation systems and the 'cleavage' activity of the mutated FMDV 2A or 2A-like sequences was assayed by the generation of discrete [GFP2A] and GUS translation products.

Methods

Plasmid constructs. All plasmids were constructed using standard methods and confirmed (or identified in the case of oligonucleotides with in-built sequence degeneracy) by nucleotide sequencing. All restriction enzymes and the coupled transcription/translation system (TNT) were purchased from Promega whilst oligonucleotides were obtained from a commercial supplier (Oswel DNA Service).

Site-directed mutagenesis of 2A

Plasmids pSTA1/1–18,25,26 and pMD3/11(a). Double-stranded (ds) oligonucleotide molecules encoding the 2A-like sequences were designed, when annealed, to form the appropriate *Xba*I and *Apal* 'sticky' ends – an 'adapter'. Plasmid pSTA1 was restricted with *Xba*I and *Apal* (restriction sites shown in Table 1) and the large restriction fragment gel purified. The ds-oligonucleotides (50 pmol ds-oligonucleotide) were ligated with this restriction fragment (100 ng), and thereby inserted between GFP and GUS, maintaining a single ORF.

Plasmids pSTA1/19–22. Sequences encoding the C-terminal region of protein GFP together with the 2A region were amplified by PCR using plasmid pSTA1 as the template and oligonucleotide GFP1 (5' TTAC-CAGACAACCATTAC 3') as forward primer and the following oligonucleotides as reverse primers: pSTA1/19 (5' GGTGGTGGGGCGCAGGGTTG 3'); pSTA1/20 (5' GGTGGTGGGGCACAGGGTTG 3'); pSTA1/21 (5' GGTGGTGGCCCCCCAGGGTTG 3') and pSTA1/22 (5' GGTGGTGGCTCCCAGGGTTG 3'). Mutagenic nucleotides are shown underlined. The gel-purified amplified cDNA products were then used as forward primers in a second round of amplifications using pcDNAPJ2.1 as the template and oligonucleotide Li1 (5' ATTAGGA-AAGGACAGTGGAGTGG 3') as reverse primer (these reactions amplified the mutated forms of 2A along with the GUS coding sequences). The cDNA product from the second round of PCRs was then restricted with *Xba*I and *Xho*I, gel purified, and ligated into similarly restricted pSTA1.

pSTA1/31. FMDV capsid protein 1D-coding sequences were amplified by PCR using plasmid pTG394 (Donnelly *et al.*, 1997) as the template: forward oligonucleotide primer OR394 (5' TTTTTTTCTAGAGTCAC-CGAGTTGTTTAC 3') and reverse primer SP6 (5' TATTAGGTG-ACACTATAG 3'). The amplified (Δ 1D_{39aa}-2A-GUS) cDNA product was restricted with *Xba*I and *Apal*, the small restriction fragment gel purified and ligated with pSTA1 similarly restricted.

pSTA1/32. FMDV capsid protein 1D-coding sequences were amplified by PCR using plasmid pTG394 (Donnelly *et al.*, 1997) as the template: forward oligonucleotide primer F26843 (5' TTTTTTTCTAGATTGC-TGGCAATCCACCAACT 3') and reverse primer SP6. The amplified (Δ 1D_{21aa}-2A-GUS) cDNA product was restricted with *Xba*I and *Apal*,

the small restriction fragment gel purified and ligated with pSTA1 similarly restricted.

pSTA1/33. FMDV capsid protein 1D-coding sequences were amplified by PCR using plasmid pTG394 (Donnelly *et al.*, 1997) as the template: forward oligonucleotide primer F26842 (5' TTTTTTTCTAGAGAAAG-CCAGACACAAACAGAAA 3') and reverse primer SP6. The amplified (Δ 1D_{14aa}-2A-GUS) cDNA product was restricted with *Xba*I and *Apal*, the small restriction fragment gel purified and ligated with pSTA1 similarly restricted.

pSTA1/34. Oligonucleotides TG5 (5' CTAGAGCATGCGCA 3') and TG6 (5' CCGGTGCGCATGCT 3') were annealed to form a ds-oligonucleotide adapter with *Xba*I and *Age*I 'sticky' ends (Donnelly *et al.*, 1997). The adapter was ligated with plasmid pSTA1/32, restricted with *Xba*I and *Age*I, as described above.

pSTA1/35. Oligonucleotides OR82 (5' CTAGACTTAAGCTTGGGG-GAGACGT 3') and OR83 (5' CTCCCGCAAGCTTAAGT 3') were annealed to form a ds-oligonucleotide adapter with *Xba*I and *Aat*II 'sticky' ends (Ryan & Drew, 1994). The adapter was ligated with plasmid pSTA1, restricted with *Xba*I and *Aat*II, as described above.

pSTA1/36. Oligonucleotides OR84 (5' CTAGACTTGGGGAGAC-GT 3') and OR85 (5' CTCCCGCAAGT 3') were annealed to form a ds-oligonucleotide with *Xba*I and *Aat*II 'sticky' ends (Ryan & Drew, 1994). The adapter was ligated with plasmid pSTA1, restricted with *Xba*I and *Aat*II, as described above.

pMD2 constructs

pMD2.7.15. Plasmid pMD2 (Donnelly *et al.*, 1997) was restricted with *Aat*II, treated with T4 DNA polymerase to remove overhangs, restricted with *Af*II and the large restriction fragment gel purified. Oligonucleotide OMD13 [5' TTAAGCTTGGGG(G/C)AGGT 3'] was annealed with oligonucleotide OMD14 [5' ACCT(G/C)TCCCGCAA-GC 3'] to form a ds-oligonucleotide adapter. The adapter was ligated with the pMD2 restriction fragment as described above.

pMD2.3/1/7/9. Plasmid pMD2 (Donnelly *et al.*, 1997) was doubly restricted with *Aat*II and *Af*II and the large restriction fragment gel purified. Oligonucleotide OMD5 (5' CGAGTCCAACCCTGGGNNN-TTTTTTTTTACTAGTA 3') was annealed with oligonucleotide OMD6 (5' G ATCTACTAGTAAAAAAAAANNCCAGGGTTGG-ACTCGACGT 3') to form a ds-oligonucleotide adapter. The adapter was ligated with the pMD2 restriction fragment as described above.

pHisGFP2AGUS. A His₆ affinity purification tag was introduced into the GFP coding sequences by amplification of the GFP coding sequence using the forward oligonucleotide primer HISGFP (5' CGCGCGGG-ATCCACCATGGGGCACCACCACCGTAAAGGAG-AACTT 3') and the reverse SP6 oligonucleotide primer plasmid pGFP2AGUS [Donnelly *et al.*, 2001 (accompanying paper)] as the template. The PCR product was doubly restricted with *Bam*HI and *Apal*, gel purified, and then ligated into pGFP2AGUS similarly restricted.

pPJ1. Plasmid pHisGFP2AGUS was doubly restricted with *Bam*HI and *Nol*I. The [His₆GFP2AGUS] cDNA insert was gel purified and ligated with the vector pYES (Invitrogen), similarly restricted.

pPJ2. A His₆ affinity purification tag was introduced into the GUS coding sequences by amplification of the GUS coding sequence using the forward oligonucleotide primer HISGUS (5' CGCGCGGGGCC-CACCACCAACCACCAACTACGTCTGTAGAAC 3') and SP6 as the reverse primer, plasmid pGFP2A-GUS being used as the

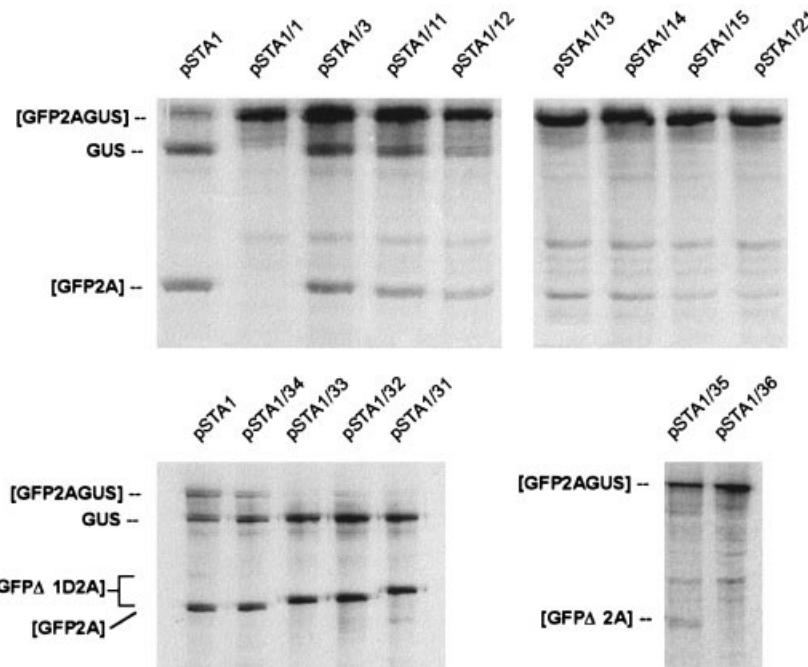


Fig. 1. Translation *in vitro*. Translation products derived from constructs encoding the wild-type 2A sequence (pSTA1) are shown together with mutated forms. The translation products derived from constructs encoding N-terminally extended forms of 2A (pSTA1/31–34) and deleted forms pSTA1/35,36 are also shown.

template. The PCR product was doubly restricted with *Apal* and *NoI*, gel purified, and then ligated into pPJ1 similarly restricted.

pcDNA3.1x. Prior to further cloning work the *Xba*I and *Apal* restriction sites were removed from the multiple cloning site of pcDNA3.1 (Invitrogen) by doubly restricting with *Xba*I and *Apal*, treatment with T4 DNA polymerase to produce blunt ends, and religation.

pcDNAPJ2.1. Plasmid pcDNAPJ2.1 contains the FMDV 2A sequence flanked by the His₆-tagged GFP and His₆-tagged GUS reporter genes assembled into the cellular expression vector pcDNA3.1x. Plasmid pPJ2 was doubly restricted with *Bam*HI and *Xho*I and the restriction fragment encoding the doubly His-tagged [GFP2AGUS] insert gel purified. The purified insert was then ligated with pcDNA3.1x, similarly restricted.

pSTA1. For other purposes we wished to remove a *Fok*I restriction site and alter the context of the *Xba*I restriction site (present at the 3' terminus of the GFP coding sequence) such that the *Xba*I site was not subject to Dam methylation. The His₆-GFP coding sequences were amplified by PCR using the forward oligonucleotide (T7) primer 5' TAATACGACTCACTATAGGG 3' and the reverse primer F4389 (5' GCGCGCTCTAGACCCGGACTTGTATAGTCGTCCATGCCAT-GTGTAAAT 3'). The PCR product was doubly restricted with *Bam*HI and *Xba*I, gel purified, and then ligated with pcDNAPJ2.1, similarly restricted.

■ 2A-like plasmid constructs. Ds-oligonucleotide adapter molecules (Oswel DNA Service) encoding the 2A-like sequences were designed, when annealed, to form the appropriate *Xba*I and *Apal* 'sticky' ends. Plasmid pSTA1 was restricted with *Xba*I and *Apal* (sites shown in Table 2) and the large restriction fragment gel purified. The ds-oligonucleotide adapters (50 pmol) were ligated with this restriction fragment (100 ng), and thereby inserted between GFP and GUS, maintaining a single ORF. The sequence for each ds-oligonucleotide, together with the oligopeptide encoded, is shown in Table 2.

■ Coupled transcription/translation *in vitro*. Rabbit reticulocyte lysates or wheat germ extracts (Promega) were programmed with unrestricted plasmid DNA (1 µg) and incubated at 30 °C for 45 min.

■ Distribution of radiolabel. Translation reactions were analysed by SDS-PAGE (10%) and the distribution of radiolabel determined either by autoradiography or by phosphorimaging using a Fujix BAS 1000. Incorporation of radioactivity into specific products was quantified directly by the latter method.

■ Calculation of 'cleavage' activity. The incorporation of radiolabel into the translation products [GFP2AGUS] (uncleaved form), and the 'cleavage products' GUS and [GFP2A] was determined by phosphorimaging (Fujix BAS 1000). The photo-stimulated luminescence (PSL) of each band was determined, and then divided by the methionine content of the appropriate translation product (PSL^{corr}). Cleavage activity (%) was calculated as:

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

■ Estimation of translational outcomes. The analysis of 'cleavage' activity described above was extended to calculate the proportion of ribosomes which synthesize a full-length translation product, the proportion which synthesize both GFP and then GUS, and those which synthesize GFP alone. The PSL^{corr} for GUS was subtracted from the GFP^{PSLcorr} value to estimate the proportion of ribosomes which ceased translation at the end of [GFP2A]. The remaining GFP^{PSLcorr} value was added to the GUS^{PSLcorr} value to estimate the proportion of ribosomes which 're-initiated' to synthesize GUS.

Results

Site-directed mutagenic forms of FMDV 2A

Translation profiles derived from the various site-directed FMDV 2A cDNA constructs (Table 1) are shown in Fig. 1,

Picornaviruses;

2A-Mediated 'Cleavage'			
EMCB	-119aa-	GIFNAHYAGYFADLLIHDIETNPG	P-
EMCD		GIFNAHYAGYFADLLIHDIETNPG	P-
EMCPV21		RIFNAHYAGYFADLLIHDIETNPG	P-
MENGO		HVFETHYAGYFSDLIHDVETNPG	P-
TMEGD7	-109aa-	KAVRGYHADYYKQRLIHDVEMNPG	P-
TMEDA		RAVRAYHADYYKQRLIHDVEMNPG	P-
TMEBEAN		KAVRGYHADYYRQRLIHDVETNPG	P-
3C-Mediated Cleavage			
		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18	19
FMDA10	-APAKQ	LLNFDLLKLAGDVEESNPG	P-
FMDA12	-APGKQ	LLNFDLLKLAGDVEESNPG	P-
FMDC1	-APAKQ	LLNFDLLKLAGDVEESNPG	P-
FMDO1K	-APVKQ	LTNF DLLKLAGDVEESNPG	P-
FMDV01G	-APVKQ	LLNFDLLKLAGDMEESNPG	P-
FMDSAT3	-KPDQK	MCNF DLLKLAGDVEESNPG	P-
FMDSAT2	-GVAKQ	LLNFDLLKLAGDVEESNPG	P-
ERAV	-NINKQ	CTNY ALLKLAGDVEESNPG	P-
ERBV	-TILSE	GATNFSLLKLAGDVEENPG	P-
PTV-1	-VMAFQ	GPGATNFSLLKQAGDVEENPG	P-

Insect Viruses;

DCV	-EAARQM LLLSGDVEETNPG	P-
CrPV	-FLRKRTQLLMSGDVEESNPG	P-
ABPV	-GWT D I L L L S G D V E T N P G	P-
TaV	-RAEGRGSLLTCGDVEENPG	P-
IFV	-TRAETEDELTRAGIESNPG	P-

Type C Rotaviruses;

Bovine Rotavirus	-SKFQIDRILISCDIELNPG	P-
Porcine Rotavirus	-AKFQIDKILISCDVELNPG	P-
Human Rotavirus	-SKFQIDKILISCDIELNPG	P-

Trypanosoma spp. Repeated Sequences;

TSR1	-SSI IRTKMLVSGDVEEENPG	P-
AP Endonuclease	-CDAQRQKLLLSGDI EQNPG	P-

Bacterial Sequence;

<i>T. maritima aguA</i>	-YIPDFGGFLVKADSEFNP GP-	
-------------------------	-------------------------	--

Fig. 2. 2A and 2A-like sequences. The sequences shown are not aligned by computer algorithms. The canonical motif used to probe databases is shown by inverse font. Potential 3C proteinase cleavage sites are shown for the aphtho-, erbo- and teschoviruses.

whilst profiles derived from naturally occurring 2A-like sequences (Table 2, Fig. 2) are shown in Fig. 3.

Point mutants

The D12E mutant shows no 'cleavage' activity: neither GUS nor [GFP2A] products were detectable. This was the case for 22 other mutations (indicated in Table 1) having no detectable 'cleavage' activity (data not shown). The 'cleavage' activities shown in Table 1 represent the proportion of radiolabel incorporated into the GUS and [GFP2A] products as compared to the total incorporation into the three forms 'uncleaved' [GFP2AGUS] plus GUS and [GFP2A]. Multiple phosphorimaging analyses (using alternative methods of band delineation and background subtraction) of the same translation profiles have shown us that the error in determining incorporation of label is $\pm 2\%$. The combined error in phosphorimaging analysis of translation profiles obtained from an identical construct, but analysed in multiple, independent, translation reactions, was $\pm 5\%$. A low level of internal

initiation resulted in N-terminally truncated forms migrating between [GFP2A]GUS and GUS. Phosphorimaging analyses of translation profiles obtained from the control construct [GFP2A]GUS and the mutants with no activity showed no incorporation of label into those regions of the gel corresponding to either GUS or [GFP2A]. Although GUS (Met = 12) has double the methionine content of [GFP2A] (Met = 6), it can be seen that in translation profiles obtained from mutants with low activity (e.g. pSTA1-13,14,15,21) that a [GFP2A] product is detectable, whilst GUS is not.

N-terminally extended/deleted forms of 2A

To more finely 'map' those sequences which are required for activity per se, and those sequences upstream of 2A which increase activity to wild-type levels, we analysed a series of constructs in which sequences N-terminal of 2A in the FMDV polyprotein (capsid protein 1D) were built-back into our artificial polyprotein system. N-terminal extension of 2A by 5 aa of 1D increased the activity from $\sim 90\%$ to $\sim 96\%$

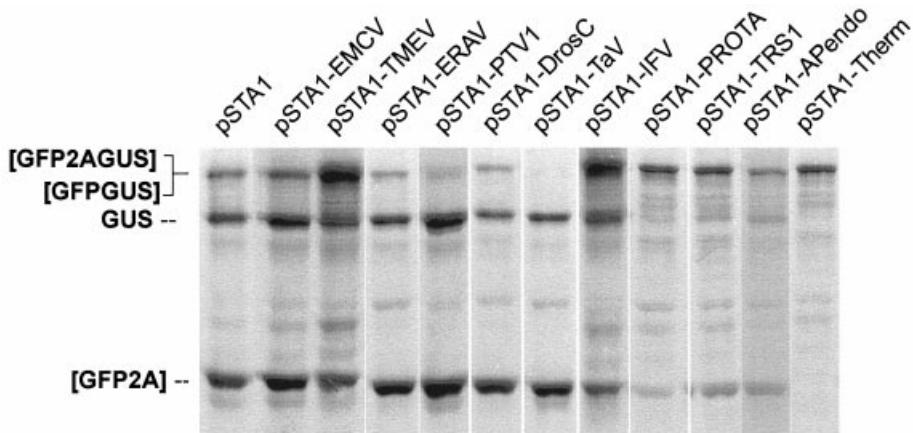


Fig. 3. Translation *in vitro*. Coupled transcription/translation wheat germ extracts were programmed with the plasmid constructs indicated. Full-length ('uncleaved') and 'cleavage' products are indicated.

(pSTA1/34; Fig. 1), whereas extension by 14 aa of 1D or longer increased the activity to > 99% (pSTA1/33,32,31; Fig. 1). In some cases a band corresponding to uncleaved [GFP2AGUS] was barely visible by prolonged autoradiography, incorporation could not be detected above the background level by phosphorimaging. In our previous [CAT2AGUS] constructs (Ryan & Drew, 1994), CAT sequences were juxtaposed with the N-terminally deleted forms of 2A by the deletion process. Since sequences immediately N-terminal of 2A were known to influence activity we wished to determine if these CAT sequences had perturbed our analysis. To confirm our previous findings we analysed the N-terminally truncated forms of 2A encoded by pSTA1/35 and pSTA1/36 in the [GFP2AGUS] system. Our data were entirely consistent with the earlier observations in the [CAT2AGUS] system that the minimal length required for activity was 12 aa, along with proline corresponding to the N-terminal residue of protein 2B.

Analysis of naturally occurring '2A-like' sequences

Analysis of the translation products showed that in all cases, other than the bacterial 2A-like sequence (*Thermatoga maritima aguA* gene) and a mutated form of the infectious flacherie virus 2A-like sequence, these 2A-like sequences had 'cleavage' activity (Fig. 3). Phosphorimaging analyses were performed to determine the relative 'cleavage' activities (Table 2).

(i) Picorna- and 'picornavirus-like' 2A sequences. We had previously reported that the C-terminal region (19 aa) of the cardiovirus 2A protein, together with the N-terminal residue of 2B, also mediated 'cleavage' (Donnelly *et al.*, 1997). Here our data show that this region of the cardiovirus EMCV 2A protein is as active (~ 91%) as FMDV 2A (~ 90%), but the equivalent region of the cardiovirus TMEV 2A protein was somewhat less active (~ 65%) than FMDV 2A (Table 2, Fig. 3). Since in infected cells the primary 2A/2B polyprotein

cleavage in all three cases is complete and a construct encoding the entire TMEV 2A protein linked to GUS showed complete 'cleavage' (Donnelly *et al.*, 1997), we assume that the length of the TMEV 2A C-terminal region we have analysed is suboptimal. Indeed, extending the FMDV 2A region by the inclusion of as little as 5 aa from FMDV protein 1D results in ~ 96% 'cleavage'.

Not surprisingly, the 2A sequence from equine rhinitis A virus (ERAV; formerly equine rhinovirus-1; Li *et al.*, 1996; Wutz *et al.*, 1996; accession nos L43052 and X96870, respectively) was highly active (~ 99% 'cleavage'; Table 2, Fig. 3): the polyprotein organization of ERAV is highly similar to the aphthoviruses such that it has recently been included in this genus. Similarly, equine rhinitis B virus (ERBV; formerly equine rhinovirus-2; Wutz *et al.*, 1996; accession no. X96871), the single member of the new *Erbovirus* genus, is similar in its organization to aphthoviruses and the 2A regions of both ERAV and ERBV are like that of FMDV. The recently sequenced porcine teschovirus-1 (PTV-1; formerly porcine enterovirus-1; Doherty *et al.*, 1999; accession no. AJ011380) shows a polyprotein organization in this region very similar to that of the aphtho- and erboviruses and the 2A sequence tested proved, also, to be highly active (~ 94% 'cleavage'; Table 2, Fig. 3).

(ii) Insect virus '2A-like' sequences. The insect viruses *Thosea asigna* virus (TaV; Pringle *et al.*, 1999; accession no. AF062037), infectious flacherie virus (IFV; Isawa *et al.*, 1998; accession no. AB000906), *Drosophila* C virus (DCV; Johnson & Christian, 1998; accession no. AF014388), acute bee paralysis virus (ABPV; Govan *et al.*, 2000; accession no. AF150629) and cricket paralysis virus (CrPV; Wilson *et al.*, 2000; accession no. AF218039) contain 2A-like sequences. Interestingly, the short TaV and DCV 2A-like sequences tested were even more active than FMDV 2A (TaV > 99%; DrosC ~ 95% cleavage; Table 2, Fig. 3) – in the case of TaV the uncleaved [GFP'2A'GUS] material was barely detectable. In

the case of the IFV 2A-like sequence the -DxExNPGP- motif is not conserved, but differs from the consensus by single change, -GxExNPGP- (Fig. 2, Table 2). When this glycine residue was mutated to an aspartate, to be consistent with what we believed to be the canonical motif, this sequence showed no cleavage activity (data not shown).

(iii) **Type C rotavirus 2A-like sequences.** A 2A-like sequence is present in bovine, porcine and human type C rotavirus non-structural protein 34 (NS34; gene 6; Jiang *et al.*, 1993; Qian *et al.*, 1991; James *et al.*, 1999; accession nos L12390, M69115 and AJ132203, respectively). Analysis of the porcine rotavirus 2A-like sequence showed much lower cleavage activity (~31%) than that observed for many other virus 2A-like sequences (Table 2, Fig. 3).

(iv) **Trypanosome repeated sequences.** 2A-like sequences are present within repeated sequence elements of *Trypanosoma* spp. In the case of *T. cruzi* the 2A-like sequence occurred in ORF1 of the non-LTR retrotransposon L1Tc (Martin *et al.*, 1995; accession no. X83098). This ORF encodes an AP endonuclease-like sequence (APendo; Fig. 4). The 2A-like sequence which was identified in *T. brucei* occurred, however, in the trypanosome repeated sequence TRS-1 (Hasan *et al.*, 1984; Murphy *et al.*, 1987; accession nos X05710, S28721, respectively; Fig. 4). Analysis of these 2A-like sequences showed low cleavage activities: for the *T. cruzi* APendo 2A-like sequence ~69% 'cleavage' was observed and ~18% for the *T. brucei* TRS1 2A-like sequence (Table 2, Fig. 3).

(v) **Bacterial '2A-like' sequence.** Probing databases for the presence of the conserved -DxExNPGP- motif revealed only one further occurrence. This motif is present within the thermophilic eubacterium *Thermatoga maritima aguA* gene product α -glucosiduronase (Ruile *et al.*, 1997; accession no. P96105). Analysis of this cellular 2A-like sequence showed no cleavage activity (Table 2, Fig. 3).

Discussion

When analysing the polyprotein processing of aphtho- and cardiovirus infected cell proteins or translation profiles obtained using *in vitro* systems, the 2A/2B polyprotein 'cleavage' resembles that of a 'primary' proteolytic processing event: it is extremely rapid, precursors spanning the site are not observed and it is insensitive to dilution. The remarkable aspect of the FMDV 2A/2B 'cleavage' is that it is mediated by the oligopeptide 2A. One of our original working hypotheses was that 2A was an autoproteolytic element. The activity of 2A in artificial polyprotein systems strongly indicated that, if this was the case, it would be an intrinsic property. The argument would run that a nucleophile within the 2A sequence would be oriented such that it could attack the putative glycyl-prolyl scissile peptide bond. We performed extensive molecular modelling and concluded that this was possible. A reverse turn comprising E14 to P17 oriented the putative

scissile G18–P19 bond towards potential nucleophiles. In this model the reverse turn was stabilized by hydrogen bonding interactions between the side-chains of E14 and N16. The scissile bond could be oriented such that D12 was some 5 Å distant, enough to accommodate a water molecule (Van der Waal's radius ~3.2 Å). Proteolysis would occur via a general-base hydrolytic mechanism. The pK of aspartate side-chains can be altered dramatically in hydrophobic environments: indeed, this is the proposed mechanism of proteolysis of the capsid protein of the nodavirus black beetle virus during capsid biogenesis (Zlotnick *et al.*, 1994). An alternative proteolytic mechanism would invoke D12 and E14 in a mechanism directly analogous to the aspartyl-, or acid, proteinases. The involvement of the other candidate nucleophile S15 was deemed unlikely since natural sequence variation (in the equivalent position in the cardiovirus TMEV 2A protein) includes a methionine (Table 2).

FMDV 2A site-directed mutants

Mutation of D12 (D12E, D12Q) abrogated activity – consistent with the observations of Hahn & Palmenberg (1996). Our analysis of a '2A-like' sequence from the insect virus IFV (Isawa *et al.*, 1998) showed, however, that D12 is not an absolute requirement for activity. Mutation of E14 showed the E14D mutant to be inactive, whilst activity was observed for the E14Q mutant, suggesting side-chain length rather than an acidic character was of more importance. Hahn & Palmenberg (1996) found, however, that the equivalent EMCV E12D mutant did show some activity. Our data showed that the general-base (D12) and aspartyl-proteinase (D12/E14) mechanisms could be discounted. It is interesting to note that the two constructs with a combination of a basic residue at 10 and an acidic residue at 12 [pSTA1-IFV(D) and pSTA1-Therm] were both inactive. This is consistent with the inactivity of constructs 42015 (-SRLLNFDLLHLDIETNPGP), 42016 (-SRLLNFDLLRLDIETNPGP) and pMD3/6(c) (-QLLN-FDLLHIDVESNPGP) that we observed previously (Ryan *et al.*, 1999).

To confirm the inference we made with regard to the natural sequence variation in positions equivalent to S15, two mutants were analysed (S15I, S15F), both of which showed activity. We have, therefore, analysed all of the potential nucleophiles within the highly conserved -D(V/I)E(S/T/M)NPG[‡]P- motif, and find none which are an absolute requirement for activity – which would be the case for a proteolytic mechanism.

The role of N16, completely conserved in all 2A and 2A-like sequences (Fig. 2), remains unanswered. Of particular interest here were two aspects: (i) the involvement of asparagine in protein deamidation, or even cleavage via the β -aspartyl shift mechanism (Geiger & Clarke, 1987; Klotz & Thomas, 1993) and (ii) the ability of this residue (*i*) to hydrogen bond with the (*i*+2) residue across a tight (Asx) turn (Wilmot & Thornton, 1988; Le Questel *et al.*, 1993). Of the limited

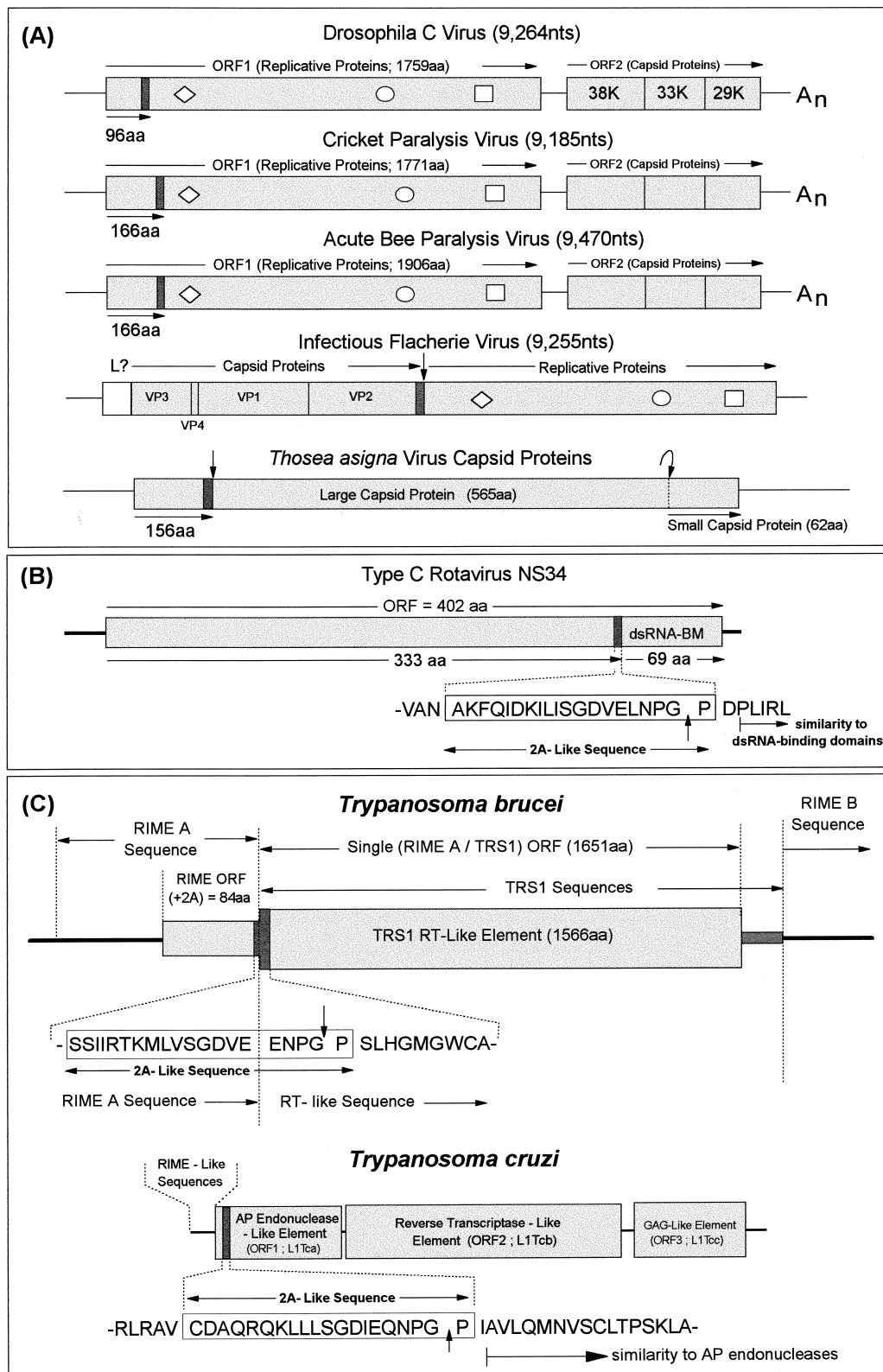


Fig. 4. Occurrence of 2A-like sequences. The positions of 2A-like sequences (dark rectangles) within insect virus polyproteins are shown. The positions of NTP-binding (picornavirus 2C-like), proteinase and RNA polymerase motifs are shown by open diamonds, circles and squares, respectively (A). The position and sequence of the porcine type C rotavirus 2A-like sequence are shown together with the predicted cleavage site and sizes of the 'cleavage' products (B). The positions and sequences of *T. brucei* TRS1 and *T. cruzi* APendo 2A-like sequences are shown together with the predicted cleavage sites (C).

number of N16 mutants we analysed (N16H, N16E, N16Q), all were active. These data are sufficient to show, however, that N16 is not involved in peptide bond cleavage, nor the formation of an Asx turn. The potential hydrogen bonding interaction of the side-chains of residues 14 and 16 was tested by the construction of a series of double E14/N16 mutants, none of which were active.

Proline-17 and glycine-18 are completely conserved amongst all active 2A and 2A-like sequences and we have extended the observations of Hahn & Palmenberg (1996) who found that mutation of the equivalent residues of EMCV 2A (P17L, P17R, P17Q, G18A, G18E, G18V, G18W) abrogated activity. Similarly, we found the identity of these residues to be critical for activity. In the same study mutation of proline 19 (P19L, P19R) also abrogated activity: our data show that whereas mutants P19A P19S, P19I and P19F were inactive, a low level of activity was observed for P19G. We have proposed a model of 2A activity in which the poor nucleophilic character of the residue in this position is an integral part of the proposed mechanism [Ryan *et al.*, 1999; Donnelly *et al.*, 2001 (accompanying paper)]. Interestingly, it has been shown that next to proline, glycine is the poorest of nucleophiles in the context of ribosomal peptidyltransferase activity (Nathans & Niedle, 1963; Rychlik *et al.*, 1970).

Our dynamic molecular modelling predicted a helical structure. Our translational model proposes that 2A mediates its effect upon the ribosomal peptidyltransferase centre whilst still in the ribosomal exit tunnel. Theoretical work (Lim & Spirin, 1986) predicted the likely conformation of the nascent peptide in this environment to be helical. Recent ultrastructural determination of the ribosomal large subunit showed that the dimensions (length \sim 100 Å; av. dia. \sim 15 Å) of the exit tunnel are entirely consistent with this notion (Ban *et al.*, 2000; Nissen *et al.*, 2000). We found that insertion of residues within this putative helical structure resulted in no activity. Our N-terminally extended and deleted forms showed that the minimum length for any activity was 12 aa and that restoring between 5 and 14 aa of the native sequence N-terminal of 2A increased the activity from \sim 90% to \sim 96% (5 aa extension); a 14 aa N-terminal extension restored complete 'cleavage'. The lengths of these oligopeptides are such that they could be accommodated entirely within the ribosome exit tunnel. Our site-directed mutagenic data provide a strong line of evidence arguing against a proteolytic mechanism for the 2A-mediated 'cleavage', and are consistent with our translational model of 2A activity.

2A-like sequences

In the analyses described above, we were primarily concerned with determining if the 2A-like sequences identified were active *per se*. It should be born in mind that the analysis of the FMDV 2A sequence showed that a relatively short N-terminal extension increased the overall 'cleavage' efficiency

and produced equimolar ratios of the cleavage products. It may very well be the case, therefore, that many of the lower cleavage efficiencies observed for the '2A-like' sequences could be also be affected substantially by the analysis of a (somewhat) longer sequence.

(i) Picornaviruses. Three 2A-like sequences were found in picornaviruses other than aphtho- or cardioviruses, although ERAV has recently been included within the aphthoviruses. The 2A-like sequences from ERAV and PTV-1 were highly active and we think it is quite reasonable to assume they perform the same primary 'cleavage' function in protein biogenesis as the aphtho- and cardiovirus 2A sequences, although detailed knowledge of the polyprotein processing is lacking for these viruses.

(ii) Insect viruses. In the case of IFV, we would propose that the 2A-like sequence could function as it does in picornaviruses – to bring about a primary cleavage between polyprotein domains comprising the capsid proteins and those comprising the replicative proteins (Fig. 4A). DCV, ABPV and CrPV all show a similar genome organization and the 2A-like sequence is conserved in both its sequence and position in the N-terminal region of the replicative ORF1 (Figs 2 and 4). In these cases, therefore, our translational model of 2A activity would predict that the translation of the replicative proteins (ORF1) would result in a 'primary' N-terminal cleavage product of 96 aa (DCV) and 166 aa (CrPV and ABPV). In the case of TaV the 2A-like sequence is present within the capsid protein precursor. The activity of the TaV 2A-like sequence has, however, been demonstrated by N-terminal sequencing of the capsid protein cleavage products (Pringle *et al.*, 1999).

(iii) Type C rotaviruses. Interestingly, type C rotavirus NS34 proteins may be aligned with the NS3 proteins of type A rotaviruses but have an additional dsRNA binding domain at their C terminus. The 2A-like sequence is conserved amongst all type C rotavirus NS34 sequences to date. Inspection of alignments of this domain with other dsRNA binding domains shows this domain to start immediately downstream of the 2A-like sequence (Fig. 4B). Alignments of NS3/34 and the dsRNA binding domain are available at <http://www.sanger.ac.uk/Software/Pfam/browse.shtml>. Our model would predict that the 'cleavage' activity of the type C rotavirus 2A-like sequence could serve to generate the NS34 protein lacking the dsRNA binding domain, plus the dsRNA binding domain as a discrete product or, perhaps, a mixture of the cleavage products together with the full-length NS34 protein. Whether the presence of a dsRNA binding domain in the type C rotaviruses represents a relative loss within other rotaviruses or a relative acquisition cannot be determined, but the *method* by which this extra domain is fused to NS3 is reflected in another instance of a 2A-like sequence.

(iv) Trypanosome repeated sequences. 2A-like sequences occur in repeated sequences of both *T. brucei* and *T. cruzi*. These 2A-

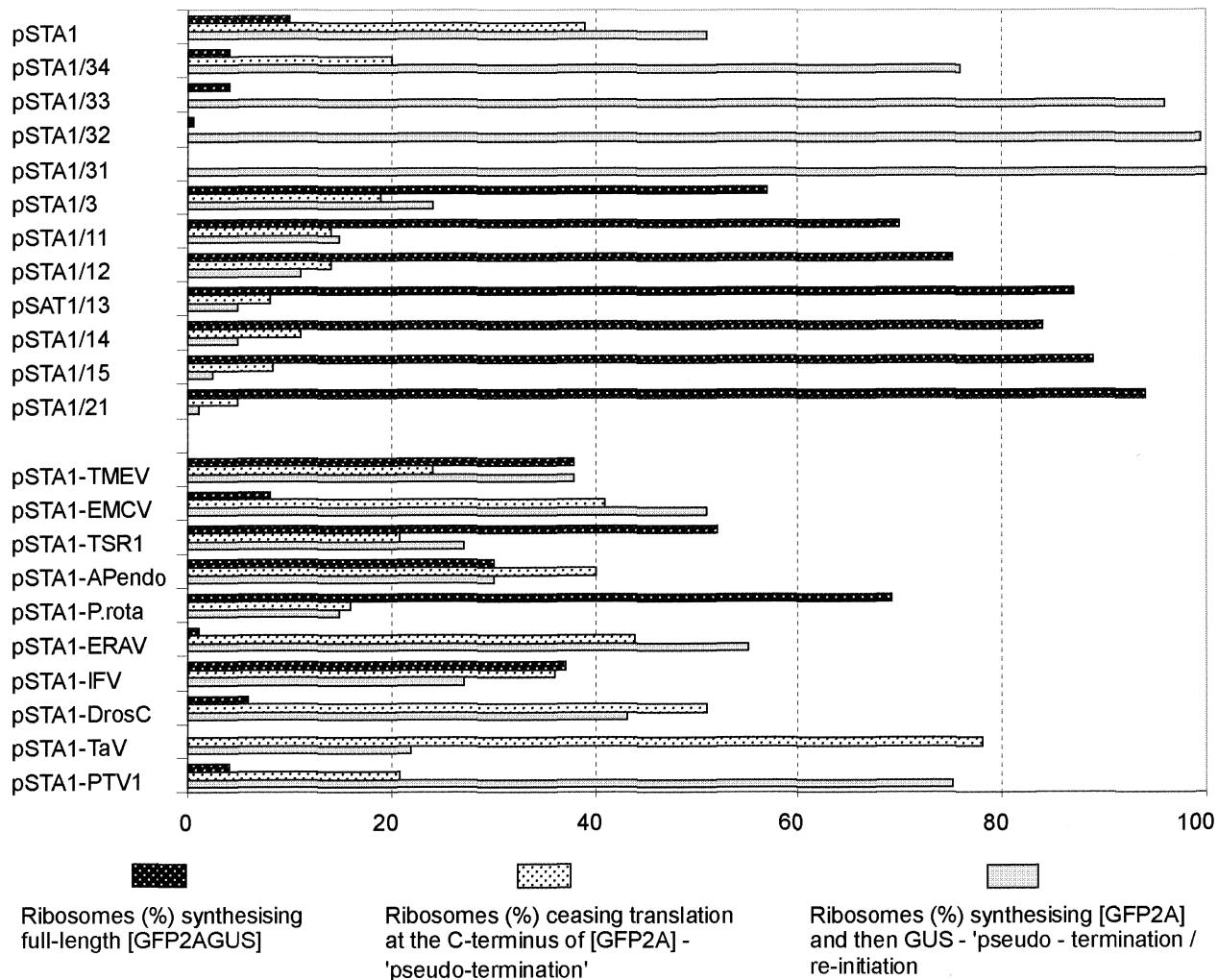


Fig. 5. Translational outcomes. The molar ratios of the translation products [GFP2AGUS], [GFP2A] and GUS were determined by phosphorimaging and used to calculate the proportion of ribosomes which would account for these ratios by either (i) the synthesis of the full-length product, (ii) ceasing translation at the 2A sequence or (iii) synthesizing [GFP2A], hydrolysis occurring to release [GFP2A], and then subsequently synthesizing GUS as a discrete translation product.

like sequences occur, however, in different types of insertion element. Trypanosome rDNA genes may be interrupted by the insertion of ribosomal insertion mobile elements (RIMEs). These elements, in turn, may themselves be disrupted by other insertions. In the case of *T. cruzi* a RIME may contain the insertion of a non-LTR retrotransposon (L1Tc). This element has three main ORFs: ORF1 (L1Tca) has significant similarity to the human AP endonuclease protein, ORF2 has significant similarity to retrotranscriptase-related sequences from non-LTR retrotransposons and ORF3 encodes a Gag-like protein (Fig. 4C). The *T. cruzi* 2A-like sequence is present in the N-terminal portion of the AP endonuclease-like sequence (L1Tca) and, interestingly, the similarity with other AP endonuclease protein family members starts immediately after the 2A-like sequence (Fig. 4C).

In *T. brucei*, however, the RIME is disrupted by the insertion of a different type of element with a single, long, ORF

encoding a reverse transcriptase (RT)-like protein (Fig. 4C). The 2A-like sequence is found at the junction of two ORFs during transposition: the N-terminal portion is derived from the RIME sequence and the C-terminal portion from the RT-like protein (Fig. 4C). We propose, therefore, that in both cases the trypanosome 2A-like sequence serves to generate either the 'mature' AP endonuclease-like protein (*T. cruzi*) or 'mature' RT-like protein (*T. brucei*) by 'cleaving' these proteins from their fusion partners. Presumably transcriptional control of both the AP endonuclease and the RT-like protein is still a function of the RIME. Whether the uncleaved forms of these proteins are active or inactive is only acquired upon 'cleavage' is an interesting question.

(v) Cellular sequences. Insertion of the eubacterial *Thermotoga maritima* 2A-like sequence (Fig. 2, Table 2) into our reporter system showed it to be inactive in our assay system. This

observation is consistent with both the mutant form of the IFV 2A-like sequence and our previous analyses of the N-terminally truncated forms of 2A: the presence of the -DxExNPGP- motif *alone* is not sufficient to confer self-'cleavage', but requires an appropriate upstream context to be active (Ryan *et al.*, 1991; Ryan & Drew, 1994; Donnelly *et al.*, 1997).

Viruses are known to 'manipulate' the molecular events occurring during the elongation cycle of protein synthesis. Programmed ribosomal frame-shifting and programmed ribosomal 'hopping' are two examples, although these effects are also harnessed in the expression of certain cellular genes (reviewed by Farabaugh, 1996). Although we have found active 2A-like sequences in organisms other than picornaviruses, it could be argued that the *Trypanosoma* 2A-like sequences may also have their origins in virus sequences and, as yet, that this particular method of controlling protein biogenesis has not been adopted by cellular sequences per se.

The translational model of 2A and 2A-like activity

We have proposed a translational, rather than proteolytic, model of 2A 'cleavage' activity [Ryan *et al.*, 1999; Donnelly *et al.*, 2001 (accompanying paper)]. When the translation profiles derived from the PTV-1 and TaV 2A-like sequences were examined an interesting difference was observed. In the case of the TaV 2A-like sequence a substantial molar excess of GFP2A was observed in comparison to the GUS product (Figs 3 and 5), whereas in the case of the PTV-1 2A-like sequence the proportion of the GUS translation product was much higher. We have observed similar effects in translation profiles of other such constructs and have eliminated different protein degradation rates or other properties of the translation system which could account for these different levels of accumulation (Donnelly *et al.*, 2001). We conclude from this and other analyses that this imbalance in the products was due to different levels of *synthesis* of GFP2A and GUS. In the TaV 2A-like sequence analysed, a high proportion of ribosomes (~78%) ceased translation at the C terminus of [GFP2A], only a small proportion subsequently going on to synthesize GUS (~22%). In the case of the PTV-1 2A-like sequence analysed, the opposite appears to be the case: ~21% of ribosomes ceased translation at the C terminus of [GFP2A] whereas ~75% of ribosomes then went on to synthesize GUS (Fig. 5). It should be noted that the initiation codon of the GUS sequence in all of our reporter protein polyprotein constructs was removed and that the first AUG codon in the GUS coding sequence is 111 codons downstream of the 2A sequence. The similarity between these 2A-like sequences is striking: the -GDVEENPGP- motif is common, and these two constructs only differ in the 11 aa that constitute the N-terminal half of the 2A-like sequences.

We have analysed a series of 2A-like sequences for 'cleavage' activity and found that all of the virus sequences were active together with the *Trypanosoma* spp. repeated

sequences. These analyses have enabled us to make predictions as to the molecular biological properties of the proteins in which these 2A-like sequences have been found and to formulate hypotheses as to the control of biogenesis of these proteins. We believe that these data have also provided another line of evidence in support of our translational model of 2A activity.

This work was supported by the Biotechnology and Biological Science Research Council and the Wellcome Trust (E. ten Dam).

References

Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920.

Camon, E. B., Chaplin, P. J., Howard, C. J. & Collins, R. A. (1997). Modulation of bovine T cell responses by IL-12 and the influence of IL-4 on interferon-gamma responses to respiratory syncytial virus. *Biochemical Society Transactions* **25**, S291.

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

Collins, R. A., Camon, E. B., Chaplin, P. J. & Howard, C. J. (1998). Influence of IL-12 on interferon-gamma production by bovine leucocyte subsets in response to bovine respiratory syncytial virus. *Veterinary Immunology and Immunopathology* **63**, 69–72.

Cruz, S. S., Chapman, S., Roberts, A. G., Roberts, I. M., Prior, D. A. M. & Oparka, K. J. (1997). Assembly and movement of a plant-virus carrying a green fluorescent protein overcoat. *Proceedings of the National Academy of Sciences, USA* **93**, 6286–6290.

De Felipe, P. & Izquierdo, M. (2000). Tricistronic and tetracistronic retroviral vectors for gene transfer. *Human Gene Therapy* **11**, 1921–1931.

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

De Rose, R., Scheerlinck, J. P. Y., Casey, G., Wood, P. R., Tennent, J. M. & Chaplin, P. J. (2000). Ovine interleukin-12: analysis of biologic function and species comparison. *Journal of Interferon and Cytokine Research* **20**, 557–564.

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

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

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

Farabaugh, P. J. (1996). Programmed translational frameshifting. *Microbiological Reviews* **60**, 103–134.

Geiger, T. & Clarke, S. (1987). Deamidation, isomerization, and racemization of asparaginyl and aspartyl residues in peptides: succinamide-linked reactions that contribute to protein degradation. *Journal of Biological Chemistry* **262**, 785–794.

Gopinath, K., Wellink, J., Porta, C., Taylor, K. M., Lomonosoff, G. P. & van Kammen, A. (2000). Engineering cowpea mosaic virus RNA-2 into a vector to express heterologous proteins in plants. *Virology* **267**, 159–173.

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

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

Hasan, G., Turner, M. J. & Cordingley, J. S. (1984). Complete nucleotide sequence of an unusual mobile element from *Trypanosoma brucei*. *Cell* **37**, 333–341.

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

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

Jiang, B., Tsunemitsu, H., Gentsch, J. R., Saif, L. J. & Glass, R. I. (1993). Nucleotide sequences of genes 6 and 10 of a bovine group C rotavirus. *Nucleic Acids Research* **21**, 2250.

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

Klotz, A. V. & Thomas, B. A. (1993). N⁵-Methylasparagine and asparagine as nucleophiles in peptides: main-chain and side-chain amide cleavage. *Journal of Organic Chemistry* **58**, 6985–6989.

Kokuho, T., Watanabe, S., Yokomizo, Y. & Inumaru, S. (1999). Production of biologically active, heterodimeric porcine interleukin-12 using a monocistronic baculoviral expression system. *Veterinary Immunology and Immunopathology* **72**, 289–302.

Le Questel, J.-Y., Morris, D. G., MacCallum, P. H., Poet, R. & Milner-White, E. J. (1993). Common ring motifs in proteins involving asparagine or glutamine amide groups hydrogen-bonded to main-chain atoms. *Journal of Molecular Biology* **231**, 888–896.

Li, F., Browning, G. F., Studdert, M. J. & Crabb, B. S. (1996). Equine rhinovirus 1 is more closely related to foot-and-mouth disease virus than to other picornaviruses. *Proceedings of the National Academy of Sciences, USA* **93**, 990–995.

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

Martin, F., Maranon, C., Olivares, M., Alonso, C. & Lopez, M. C. (1995). Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. *Journal of Molecular Biology* **247**, 49–59.

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

Murphy, N. B., Pays, A., Tebabi, P., Coquelet, H., Guyaux, M., Steinert, M. & Pays, E. (1987). *Trypanosoma brucei* repeated element with unusual structural and transcriptional properties. *Journal of Molecular Biology* **195**, 855–871.

Nathans, D. & Niedle, A. (1963). Structural requirements for puromycin inhibition of protein synthesis. *Nature* **197**, 1076–1077.

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

O'Brien, G. J., Bryant, C. J., Voogd, C., Greenberg, H. B., Gardner, R. C. & Bellamy, A. R. (2000). Rotavirus VP6 expressed by PVX vectors in *Nicotiana benthamiana* coats PVX rods and also assembles into viruslike particles. *Virology* **270**, 444–453.

Precious, B., Young, D. F., Birmingham, A., Fearn, R., Ryan, M. D. & Randall, R. E. (1995). Inducible expression of the P, V, and NP genes of the paramyxovirus simian virus 5 in cell lines and an examination of the NP–P and NP–V interactions. *Journal of Virology* **69**, 8001–8010.

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

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

Roberts, A. G., Santa Cruz, S., Roberts, I. M., Prior, D. A. M., Turgeon, R. & Oparka, K. J. (1997). Phloem unloading in sink leaves of *Nicotiana benthamiana*: Comparison of a fluorescent solute with a fluorescent virus. *Plant Cell* **9**, 1381–1396.

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

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

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

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

Rychlik, I., Cerna, J., Chaldek, S., Pulkabek, P. & Zemlicka, J. (1970). Substrate specificity of ribosomal peptidyl transferase. The effect of the nature of the amino acid side chain on the acceptor activity of 2'(3')-O-aminoacyladenosines. *European Journal of Biochemistry* **16**, 136–142.

Schmidt, M. & Rethwilm, A. (1995). Replicating foamy virus-based vectors directing high-level expression of foreign genes. *Virology* **210**, 167–178.

Smerdou, C. & Liljestrom, P. (1999). Two-helper RNA system for production of recombinant Semliki Forest virus particles. *Journal of Virology* **73**, 1092–1098.

Smolenska, L., Roberts, L. M., Learmonth, D., Porter, A. J., Harris, W. J., Wilson, T. M. A. & Cruz, S. S. (1998). Production of a functional single chain antibody attached to the surface of a plant virus. *FEBS Letters* **441**, 379–382.

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

Thomas, C. L. & Maule, A. J. (2000). Limitations on the use of fused green fluorescent protein to investigate structure–function relationships for the cauliflower mosaic virus movement protein. *Journal of General Virology* **81**, 1851–1855.

van der Ryst, E., Nakasone, T., Habel, A., Venet, A., Gomard, E., Altmeyer, R., Girard, M. & Borman, A. M. (1998). Study of the immunogenicity of different recombinant Mengo viruses expressing HIV1 and SIV epitopes. *Research in Virology* **149**, 5–20.

Varnavski, A. N. & Khromykh, A. A. (1999). Noncytopathic flavivirus replicon RNA-based system for expression and delivery of heterologous genes. *Virology* **255**, 366–375.

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

Wilmot, C. M. & Thornton, J. M. (1988). Analysis and prediction of the different types of β -turn in proteins. *Journal of Molecular Biology* **203**, 221–232.

Wilson, J. E., Powell, M. J., Hoover, S. E. & Sarnow, P. (2000).

Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Molecular and Cellular Biology* **20**, 4990–4999.

Wutz, G., Auer, H., Nowotny, N., Skern, T. & Kuechler, E. (1996). Equine rhinovirus serotypes 1 and 2: relationship to each other and to aphthoviruses and cardioviruses. *Journal of General Virology* **77**, 1719–1730.

Zlotnick, A., Reddy, V. S., Dasgupta, R., Schneemann, A., Ray, W. J., Jr, Rueckert, R. R. & Johnson, J. E. (1994). Capsid assembly in a family of animal viruses primes an autoproteolytic maturation that depends on a single aspartic acid residue. *Journal of Biological Chemistry* **269**, 13680–13684.

Received 29 September 2000; Accepted 26 January 2001