Molecular Biology of Picornaviruses
Editors, B. L. Semler and E. Wimmer

The Aphtho- and Cardiovirus "Primary" 2A/2B Polyprotein "Cleavage"

MARTIN D. RYAN, GARRY LUKE, LORRAINE E. HUGHES, VANESSA M. COWTON, EDWIN TEN DAM, XUEJUN LI, MICHELLE L. L. DONNELLY, AMIT MEHROTRA, AND DAVID GANI

18

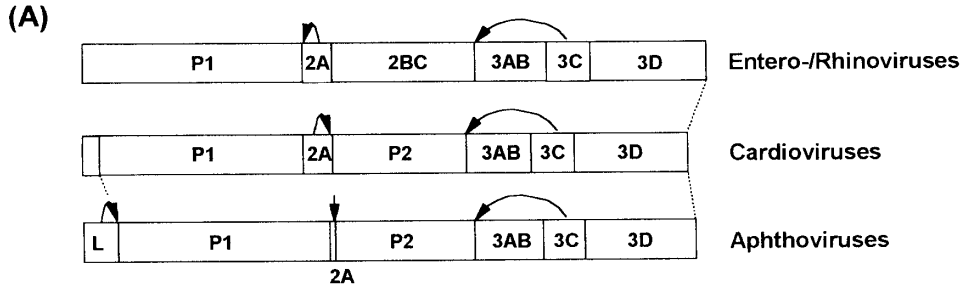
SEQUENCES INVOLVED IN THE PRIMARY 2A/2B CLEAVAGE

Early work on poliovirus polyprotein processing revealed the existence of two proteinases within the polyprotein and that the primary cleavage, which served to separate the capsid protein region of the polyprotein (P1) from the replicative protein polyprotein domain (P2), was mediated by the 2A proteinase ($2A^{pro}$) cleaving at its own N terminus (Fig. 1A) (44). In the case of the aphtho- and cardioviruses, however, the primary cleavage in this region of the polyprotein was known to be different, occurring at the C terminus of 2A (Fig. 1A) (11, 24, 31). In the aphtho- and cardioviruses, therefore, this primary cleavage results in the 2A protein remaining as a C-terminal extension of P1 until it is cleaved from P1 during secondary processing (33). Precursor forms spanning the 2A/2B junction are not observed in aphtho- or cardiovirus polyprotein processing. Comparison of the sequences of the 2A regions of different picornaviruses showed that the entero- and rhinoviruses possessed highly similar 2A proteinases. While the cardiovirus 2A proteins (between 142 and 157 amino acids [aa]) were of a size similar to that of entero- and rhinovirus $2A^{pro}$, no sequence similarity was apparent. Indeed, the 2A region of aphthoviruses was thought to be only 16 aa in length. The extremely short aphthovirus 2A region did, however, show sequence similarity with the C-terminal region of the longer cardiovirus 2A proteins (Fig. 1B).

Experiments analyzing the endogenous processing properties of recombinant aphthovirus (foot-and-mouth disease virus [FMDV]) polyproteins in which either the upstream- or downstream contexts of 2A were deleted showed that the 2A oligopeptide region did not function as part of a larger precursor form. Deletion of sequences immediately upstream did have an effect in that "cleavage" became "highly efficient" (~90%) rather than complete. Deletions downstream of 2A (although maintaining the N-terminal proline residue of protein 2B) did not appear to affect cleavage (34). These studies indicated that the cleavage activity could be a property of the 2A oligopeptidic region alone. Consistent with this notion, studies on the endogenous processing properties of domains of the cardiovirus Theiler's murine encephalomyelitis virus (TMEV) polyprotein localized the 2A/2B cleavage activity within the 2AB region (2). The 2A/2B cleavage activity of the cardiovirus encephalomyocarditis virus (EMCV) was mapped to the C-terminal third of 2A plus the N-terminal half of 2B (14). In this study the importance of the conservation of the sequence at the cleavage site, and the cleavage site itself (-NPG ⇓ P-), was confirmed. Deletions within the TMEV 2A protein, although leaving the C-terminal region intact, resulted in genomes that were competent, although impaired, in RNA replication and showed reduced virus titers (22, 49). Interestingly, the TMEV 2A protein could not substitute for mengovirus 2A protein (49).

These observations were extended by analyzing the self-processing properties of artificial polyprotein systems. Polyproteins comprised two reporter proteins (chloramphenicol acetyltransferase [CAT]; β-glucuronidase [GUS]) flanking 2A (together with the N-terminal proline residue of protein 2B) forming [CAT2AGUS] were encoded by a single open reading frame (ORF) (35). Translation reactions in vitro were programmed with transcripts derived from either the control construct pCATGUS (encoding CAT and GUS in a single ORF) or pCAT2AGUS. The translation profiles showed that the FMDV 2A region mediated a cleavage at its own C terminus, as in native FMDV processing, and that this cleavage occurred with high efficiency, but not to completion (~90%) (Fig. 2A)—as was observed in the recombinant FMDV polyproteins in which sequences N terminal of 2A were deleted. A nested set of deletions extending from the N terminus into the FMDV 2A sequence showed that cleavage occurred with only 13 aa remaining, but not 11 aa (Fig. 2B). Furthermore, a precursor-product relationship was shown not to exist between uncleaved [CAT2AGUS] and the GUS and [CAT2A] cleavage products. Following the arrest of protein

Martin D. Ryan, Garry Luke, Lorraine E. Hughes, Vanessa M. Cowton, Edwin ten Dam, and Xuejun Li ■ Centre for Biomolecular Sciences, School of Biology, University of St. Andrews, Biomolecular Sciences Building, North Haugh, St. Andrews, Fife KY16 9ST, Scotland. *Michelle L. L. Donnelly* ■ Marie Curie Research Institute, The Chart, Oxted, Surrey RH8 OTL, United Kingdom. *Amit Mehrotra and David Gani* ■ The School of Chemistry, The University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom.



(B)

Picornaviruses;

2A-Mediated 'Cleavage'

EMCB	-119aa-	GIFNAHYAGYFADLLIHDIETNPG P-
EMCD		GIFNAHYAGYFADLLIHDIETNPG P-
EMCPV21		RIFNAHYAGYFADLLIHDIETNPG P-
MENGO		HVFETHYAGYFSDLLIHDVETNPG 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	LLNFDLLKLAGDVESNPG P-
FMDA12	-APGKQ	LLNFDLLKLAGDVESNPG P-
FMDC1	-APAKQ	LLNFDLLKLAGDVESNPG P-
FMDO1K	-APVKQ	LTNFDLLKLAGDVESNPG P-
FMDVO1G	-APVKQ	LLNFDLLKLAGDVESNPG P-
FMDSAT3	-KPDKQ	MCNFDLLKLAGDVESNPG P-
FMDSAT2	-GVAKQ	LLNFDLLKLAGDVESNPG P-
ERAV	-NINKQ	CTNYALLKLAGDVESNPG P-
ERBV	-TILSE	GATNFSLLKLAGDVELNPG P-
PTV-1	-VMAFQ	GPGATNFSLLKQAGDVEENPG P-

Insect Viruses;

DCV	-EAARQMLLLLSGDVETNPG P-
CrPV	-FLRKRTQLLMSGDVESNPG P-
ABPV	-GSWTDILLLLSGDVETNPG P-
TaV	-RAEGRGSLLTCGDVEENPG P-
IFV	-TRAEIEDELIRAGIESNPG P-

Type C Rotaviruses;

Bovine Rotavirus	-SKFQIDRILISGDIELNPG P-
Porcine Rotavirus	-AKFQIDKILISGDVELNPG P-
Human Rotavirus	-SKFQIDKILISGDIELNPG P-

***Trypansoma* spp. Repeated Sequences;**

TSR1	-SSIIRTKMLVSGDVEENPG P-
AP Endonuclease	-CDAQRQKLLLSGDIEQNPG P-

Bacterial Sequence;

T. maritima aguA	-YIPDFGGFLVKADSEFNPGP-

FIGURE 1 Picornavirus polyproteins. (A) The polyprotein organizations of entero-, rhino-, cardio-, and aphthoviruses are shown together (boxed areas) with the sites of primary polyprotein cleavage. (B) The sequences of the C-terminal region of cardioviruses and the 2A region of aphthoviruses are shown together with 2A-like sequences from other virus and cellular sequences.

synthesis, the [CAT2AGUS] translation product was stable and did not subsequently cleave into [CAT2A] and GUS. Cleavage occurred, therefore, cotranslationally, but not posttranslationally. With such artificial reporter polyprotein systems the 2A/2B cleavage activity of both EMCV and TMEV was subsequently mapped to the C-terminal 18 aa of their 2A proteins (together with the N-terminal proline of 2B)—these cardiovirus sequences being as efficient as the FMDV 2A in mediating cleavage. In addition, the influence on cleavage activity of the upstream sequences proximal to FMDV 2A was more finely mapped to within the C-terminal 5 aa of protein 1D (5). Restoration of this 5-aa sequence increased the cleavage activity to >99% (Fig. 2B) (5, 6). Interestingly, expression of these artificial

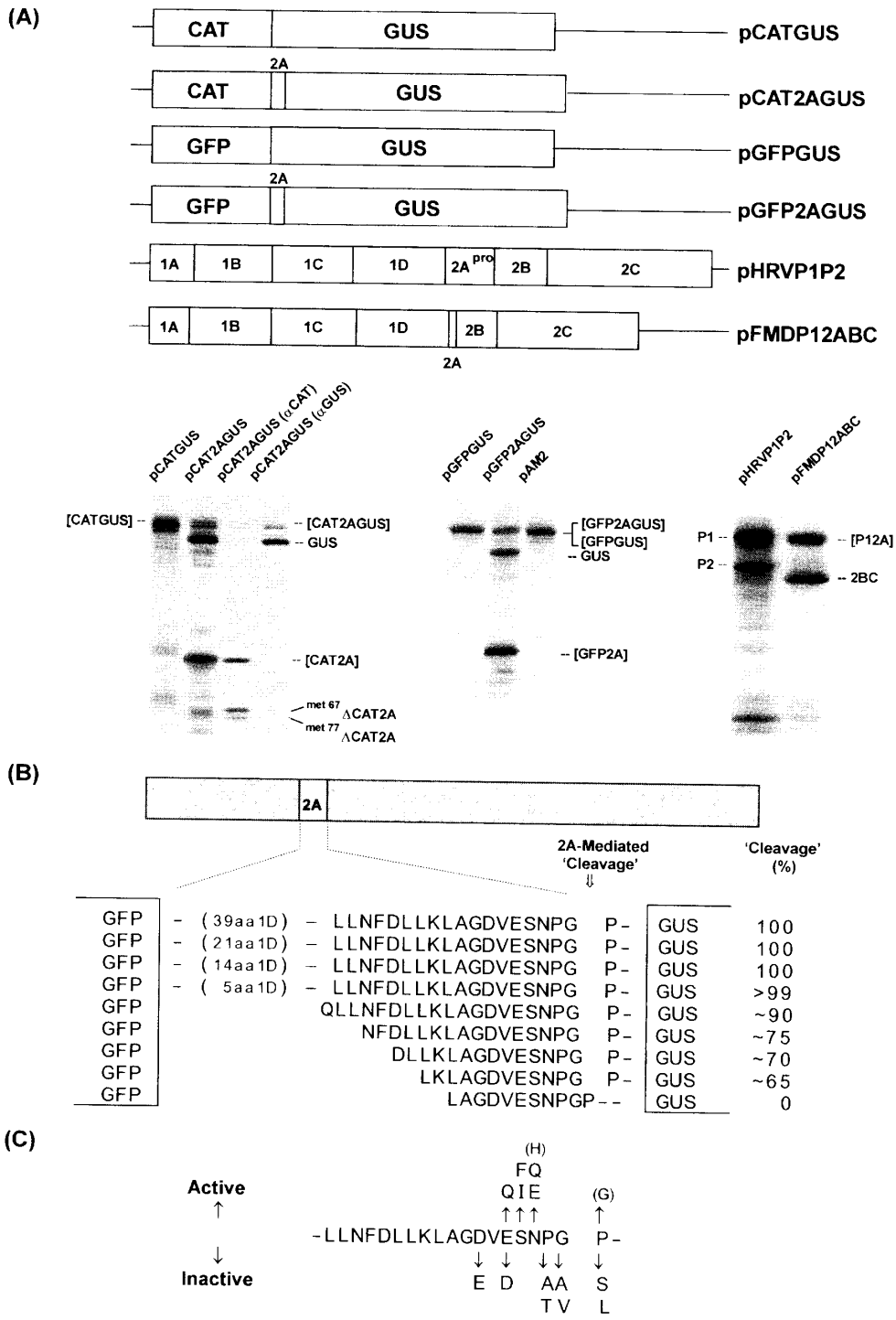


FIGURE 2 Translational analyses. Artificial reporter polyproteins (boxed areas) used to program in vitro translation systems are shown together with translation profiles obtained from rabbit reticulocyte lysates. (A) The FMDV region was either N-terminally extended by the incorporation of FMDV 1D sequences or by stepwise deletion. (B) Cleavage activities (%) are shown. (C) Site-directed mutants of FMDV 2A were constructed and the cleavage activities analyzed with in vitro translation systems.

polyproteins in a prokaryotic system did not reveal any cleavage activity. The presence of a 3C proteinase cleavage site conserved among FMDVs (also present in the same position in newly sequenced related viruses) led us to propose that FMDV 2A is 18 rather than 16 aa long (Fig. 1B).

We found significant levels of internal initiation occurred within CAT using these in vitro translation systems, which compromised our quantitative analyses (see below). Internal initiation sites within CAT were identified by N-terminal deletion and immunoprecipitation studies (Fig. 2A). We now use green fluorescent protein (GFP), rather than CAT, in our artificial polyprotein cleavage assay systems. Translation profiles obtained from [GFP2AGUS] constructs show very little internal initiation (Fig. 2A). Data obtained with this system are entirely consistent with the [CAT2AGUS] polyprotein, but translation profiles may be quantified much more accurately.

To determine whether the RNA sequence of this region, rather than the peptide sequence which it encoded, was responsible for cleavage, we engineered a construct (pAM2) that contained the FMDV 2A sequence in a [GFP2AGUS] construct, although out of frame (+2 frame) with respect to the GFP and GUS sequences. Translation profiles obtained from this construct showed no cleavage activity (Fig. 2A). Our conclusions from these data were that (i) the cleavage activity per se was a function of the aphthovirus 18-aa 2A peptide sequence, (ii) the C-terminal 18 aa of the cardiovirus 2A protein is functionally equivalent, (iii) proximal upstream sequences were influential in but not critical for this activity, and (iv) cleavage did not occur in *Escherichia coli*.

MUTAGENETIC ANALYSES OF 2A-MEDIATED CLEAVAGE

For purposes of this discussion, the numbering scheme shown in Fig. 1B is used. Inspection of picornavirus (and nonpicornavirus, see below) 2A sequences shows that only the -DxExNPG ⇓ P- motif is conserved in nature. Sequences immediately upstream of this motif that were shown to be either critical or very important for the activity are, however, not conserved. In a study in which this motif was subjected to site-directed mutagenesis, Hahn and Palmenberg found mutation of the (completely) conserved residues of this motif to either abrogate or very severely affect activity (14). The single exception to this was an E14D mutant in which partial activity was observed. We have performed site-directed mutagenesis on the FMDV 2A sequence, and our data are summarized in Fig. 2C. We find that, at variance with the EMCV mutagenesis, the FMDV E14D mutant is not active, although the E14Q mutant showed partial activity. Residue S15 shows natural sequence variation (Fig. 1B), and mutant forms not found in nature (S15F, S15I) are partially active. Residue N16 can be mutated while retaining partial cleavage activity, but mutation of both P17 and G18 abrogates activity. Interestingly, P19 (the N-terminal residue of protein 2B) shows very slight activity when mutated to glycine, but not to serine or leucine (6).

While we have shown the sequences immediately upstream of the conserved -DxExNPGP- motif are critical for activity, inspection of the available sequences in this region reveals that there are a number of different "solutions" that nature has adopted in different groups of picornaviruses and other viruses described below. Mutants of the conserved motif uniformly show either substantially lower or no cleavage activity.

"2A-LIKE" SEQUENCES

Probing the databases for sequences containing this conserved motif shows the presence of 2A-like sequences in other virus polyproteins (Fig. 1B). We have tested these 2A-like sequences (as indicated in Fig. 1) by insertion of the 2A-like sequences into our [GFP-GUS] reporter system (6). Not surprisingly, the picornavirus 2A sequences from the equine rhinoviruses types A and B and PTV-1 are active.

Insect Viruses

In the case of the insect viruses Drosophila C virus (DCV; accession no. AF014388), cricket paralysis virus (CrPV; accession no. AF218039), acute bee paralysis virus (ABPV; accession no. AF150629), and *Thosea asigna* virus (TaV; accession no. AF062037), their 2A-like sequences, when inserted between GFP and GUS, showed high cleavage activity. In the case when cleavage was >99%, uncleaved [GFP'2A'GUS] material was only barely detectable. In the case of infectious flacherie virus (IFV; accession no. AB000906), however, the 2A-like sequence was only partially active in our assay system (~50%). It can be seen that the -DxExNPGP- motif is not conserved but differs from the consensus by D12G (Fig. 1B).

In the case of IFV, we propose that the 2A-like sequence functions 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. 3A). In the case of DCV, CrPV, and ABPV, however, the 2A-like sequence occurs toward the beginning of ORF1 (replicative proteins; Fig. 3A), whereas in TaV, the 2A-like sequence is present within the capsid protein precursor. In this case the activity of the 2A-like sequence has been demonstrated by N-terminal sequencing of the capsid protein cleavage products (29).

Type C Rotaviruses

The 2A-like sequence is present in human, bovine, and porcine type C rotavirus nonstructural protein 34 (NS34; gene 6; accession no. AJ132203, L12390, and M69115, respectively). Analysis of these 2A-like sequences in our [GFP'2A'GUS] system showed much lower cleavage activity than that observed for other 2A-like sequences (6). Interestingly, the type C rotavirus NS34 protein may be aligned with the NS3 protein of type A rotaviruses but has an additional double-stranded RNA binding domain at its C terminus. Inspection of alignments of this domain with other dsRNA binding domains shows this domain to start immediately downstream of the 2A-like sequence (Fig. 3B).

Trypanosoma Repeated Sequences

2A-like sequences are present within repeated sequences in *Trypanosoma cruzi* (accession no. X83098) and *Trypanosoma brucei* (accession no. X05710 and S28721). Surprisingly, these 2A-like sequences are in different types of insertion element. Ribosomal insertion mobile elements (RIMEs) insert themselves into trypanosome rDNA genes. These elements, in turn, are themselves disrupted by insertions. In the case of *T. cruzi* a RIME may contain the insertion of a non-long terminal repeat (LTR) retrotransposon (L1Tc) (19). This element has three main ORFs: ORF1 (L1Tca) has sig-

nificant 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. 3C). The 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. 3C).

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-like protein (Fig. 3C) (19). The 2A-like sequence is formed by the juxtaposition of two ORFs during transposition: the N-terminal portion being derived from the RIME sequence and the C-terminal portion being derived from the reverse transcriptase-like protein (Fig. 3C). We have inserted these 2A-like sequences into our [GFP'2A'GUS] system and found both to be active (6). We propose, therefore, that in both cases the 2A-like sequence serves to generate either the "mature" AP endonuclease-like protein (*T. cruzi*) or "mature" reverse transcriptase-like protein (*T. brucei*) by cleaving these proteins from their fusion partners.

Cellular Sequences

Probing the databases for the presence of the conserved -DxExNPGP- motif reveals (to date) only one further occurrence. This motif is present within the thermophilic eubacterium *Thermatoga maritima augA* gene product α-glucuronidase (accession no. P96105). Insertion of this 2A-like sequence (Fig. 1B) into our reporter system shows, however, this 2A-like sequence to be inactive (6). This observation is consistent with our previous analyses of the N-terminally truncated forms of 2A: the -DxExNPGP- motif alone is not sufficient to confer self-cleavage but requires an appropriate upstream context.

STRUCTURE OF 2A

We have used a range of secondary structural prediction algorithms on 2A and 2A-like sequences. The consensus that emerges is that of a helical structure followed by a tight turn (-NPG-). The prediction of an α-helical structure was supported by dynamic molecular modeling performed upon the FMDV 2A sequence (36). In this structural model residues D5, K8, D12, and N16 align along one side of the proposed α-helical segment, an arrangement that remained stable in dynamic simulations. Two salt bridges, D5 and K8 (in an i, $i + 3$ arrangement) and K8 and D12 (i, $i + 4$), and a hydrogen bonding interaction between D12 and N16 (i, $i + 4$) could serve to stabilize this structure.

MECHANISM OF CLEAVAGE ACTIVITY

In our first report of 2A-mediated cleavage of an artificial polyprotein system we put forward three hypotheses to explain these data: (i) that FMDV served as a substrate for a host-cell proteinase, (ii) that 2A represented a novel type of proteolytic element, or (iii) that 2A in some way interfered in peptide bond formation (35).

The "Substrate" Hypothesis

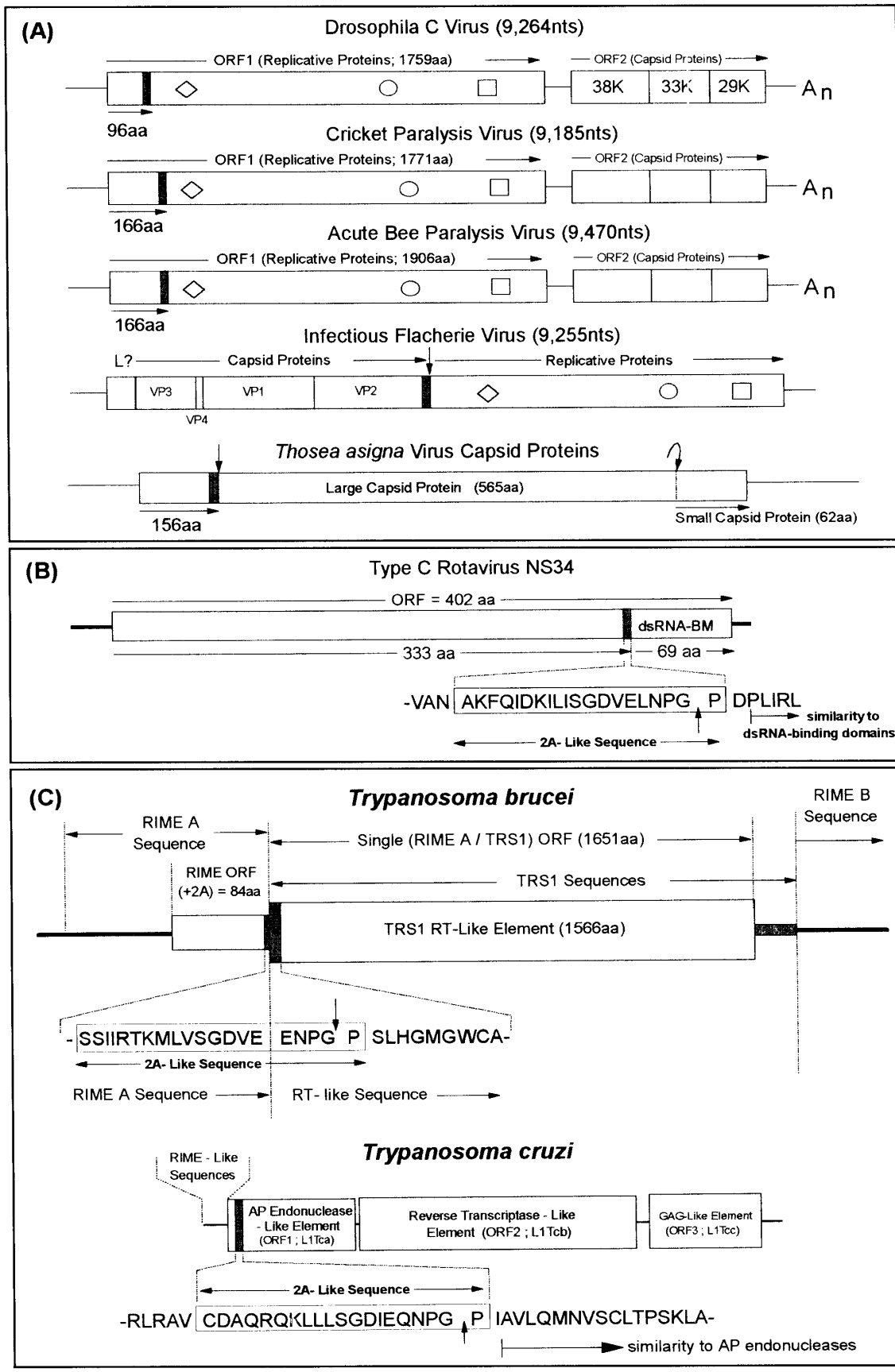
In this model 2A represents a substrate for a cellular proteinase that would need to be both very efficient and very tightly coupled to translation, since cleavage occurs cotranslationally, but not posttranslationally (35). At present there are no reports of such a ribosome-associated proteinase. 2A has been used for many biotechnological purposes and, as such, a wide range of recombinant self-processing systems have been expressed in a wide range of cell types, including mammalian, insect, plant, and fungal (3, 4, 10, 15, 18, 20, 25, 28, 38–40, 42, 45, 46). Such a proteinase would, therefore, need both to be present in this range of cell types and to have conserved its substrate specificity. Our site-directed mutagenesis data (only part of which is presented here) do not resolve this issue since the different mutant forms could represent, for example, substrates with different binding affinities for the hypothetical cellular proteinase. It should be noted here, however, that some point mutants we have constructed (which show no cleavage) are over 10 residues N terminal of the cleavage site. Similarly, it could be argued that our observations regarding the influence of the upstream sequences reflect their role in protein folding vis-à-vis the "presentation" of the site of cleavage to this hypothetical proteinase.

The "Proteinase" Hypothesis

Perhaps dispensing with the protein "architecture" required by a proteinase to recognize a substrate in *trans*, to cleave a peptide bond, then regenerate the nucleophile, could enable such a short sequence to function as a single-turnover *cis*-acting proteolytic element. In considering a novel *cis*-acting proteolytic activity for 2A, the presence of conserved potential nucleophiles (D, E, S/T) within the canonical motif was of great interest. Indeed, asparagine residues are known to mediate peptide bond cleavage via a mechanism similar to the β-aspartyl shift (reviewed in reference 48). Here our site-directed mutagenesis data are more informative. Mutation of D12 either abrogates cleavage (D12H, see reference 14; D12E, Fig. 2C) or very severely affects it (D12N, see reference 14). The IFV 2A-like sequence is active and has a glycine residue in this position (Fig. 1B). The E14Q mutation retains (reduced) activity (Fig. 2C). These two "candidate" nucleophilic residues do not appear, therefore, to function in the manner of an acidic proteinase. Although serine or threonine (both nucleophiles in proteinases) are present in all aphthoviruses and many cardioviruses, natural sequence variation in this position (Fig. 1B) precludes S15/T15 functioning in the manner of a serine or threonine proteinase. Mutation of N16 and retention of activity exclude peptide bond cleavage via a mechanism similar to the β-aspartyl shift. On the basis of the site-directed mutagenesis and natural sequence variation data sets, a proteolytic mechanism seems improbable.

The "Translational" Hypothesis

Analysis of the in vitro translation profiles obtained from both [CAT2AGUS] and [GFP2AGUS] showed a common, remarkable feature. Proteins were radiolabeled with ^{35}S-methionine and the distribution of radioactivity in the accumulated products was quantified by phosphorimaging (normalized for methionine contents). The upstream translation product (either [CAT2A] or [GFP2A]) was present in a molar excess over the downstream translation product, GUS. This effect varied between rabbit reticulocyte lysates, where ratios were commonly 5:1, and wheat germ extracts, where ratios were commonly 15:1. Interestingly, extension of the N-terminal FMDV protein 2B residues present immediately downstream of 2A in our reporter polyproteins (from just -P- to -PFFF-) increased the observed imbalance dramatically (data not shown).


(A)
Drosophila C Virus (9,264nts)
ORF1 (Replicative Proteins; 1759aa)
ORF2 (Capsid Proteins)
38K
33K
29K
An
96aa
Cricket Paralysis Virus (9,185nts)
ORF1 (Replicative Proteins; 1771aa)
ORF2 (Capsid Proteins)
An
166aa
Acute Bee Paralysis Virus (9,470nts)
ORF1 (Replicative Proteins; 1906aa)
ORF2 (Capsid Proteins)
An
166aa
Infectious Flacherie Virus (9,255nts)
L?
Capsid Proteins
Replicative Proteins
VP3
VP1
VP2
VP4
Thosea asigna Virus Capsid Proteins
Large Capsid Protein (565aa)
156aa
Small Capsid Protein (62aa)
(B)
Type C Rotavirus NS34
ORF = 402 aa
dsRNA-BM
333 aa
69 aa
-VAN AKFQIDKILISGDVELNPG P DPLIRL
similarity to dsRNA-binding domains
2A- Like Sequence
(C)
Trypanosoma brucei
RIME A Sequence
Single (RIME A / TRS1) ORF (1651aa)
RIME B Sequence
RIME ORF (+2A) = 84aa
TRS1 Sequences
TRS1 RT-Like Element (1566aa)
-SSIIRTKMLVSGDVE ENPG P SLHGMGWCA-
2A- Like Sequence
RIME A Sequence
RT- like Sequence
Trypanosoma cruzi
RIME - Like Sequences
AP Endonuclease - Like Element (ORF1 ; L1Tca)
Reverse Transcriptase - Like Element (ORF2 ; L1Tcb)
GAG-Like Element (ORF3 ; L1Tcc)
2A- Like Sequence
-RLRAV CDAQRQKLLLSGDIEQNPG P IAVLQMNVSCLTPSKLA-
similarity to AP endonucleases

The products we are measuring are those that have accumulated: a function of both protein synthesis and degradation. Protein degradation studies showed this explanation could not account for our observed product imbalances (7). An alternative explanation for this product imbalance could be that significant premature termination of transcription or translation was occurring, at random, in these coupled systems. This would give rise to a polar effect resulting in a greater level of synthesis of sequences N terminal to the polyprotein system. In our system this effect coupled with cleavage at a specific site could explain this imbalance. To address this question two types of approaches were adopted: first, to reverse the gene order—to place the (longer) GUS sequence upstream of GFP by making a [GUS2AGFP] polyprotein. Second, a control construct encoding polyprotein encoding the [P1P2] from human rhinovirus 14 (HRV14) was made—in this case P1 and P2 are cleaved by a known, characterized proteinase, $2A^{pro}$. Analyses of the [GUS2AGFP] construct showed molar excess of the [GUS2A] cleavage product over GFP, whereas analysis of the HRV14 [P1P2] construct showed the *proteolytic* cleavage products P1 and P2 to be present in ratios of between 1.3 and 1.2 to 1 (Fig. 2A) (7). This showed that our translation systems were able to synthesize long ORFs (HRV14 [P1P2] is some 55% longer than [GFP2AGUS]) and observe ratios of ~1:1 when this should be the case. Interestingly, translation of a construct (pFMDP12ABC) encoding the [P12A2BC] region of the FMDV polyprotein, in which 2A is in its entirely native protein context, showed the cleavage products [P12A] and [2BC] also to be present in ratios of between 1.2 and 1.1 to 1 (Fig. 2A) (7).

From this work we concluded that the molar excess of the translation product N terminal of 2A over that C-terminal of 2A was a product of inserting the 2A sequence into our artificial polyprotein systems: its functioning in a suboptimal polyprotein context. This is supported by the data whereby sequences immediately N terminal of 2A affect its activity. These effects have, however, given us a new insight into the mechanism of 2A-mediated cleavage.

A TRANSLATIONAL MODEL OF 2A CLEAVAGE ACTIVITY

Given that the substrate and proteinase hypotheses are improbable, taken together with our observations on the imbalance of the translation products, how can cleavage be explained? We have proposed a mechanism whereby two specific translation products may be obtained from a single ORF by modification of the elongation stage of translation. In our model the upstream translation product ([P12A], aphthoviruses; [L-P1-2A], cardioviruses; [GFP2A], artificial systems) is released from the translation complex by the cleavage of the ester linkage between the nascent peptide and the tRNA to which it is attached (7, 36).

The scheme we have proposed is shown in Fig. 4. The individual steps are outlined here and discussed below. Step (i): The peptide bond between P17 and G18 is synthesized; deacylated $tRNA^{gly}$ is present in the P site and (2A)peptidyl-tRNA is present in the A site of the ribosome. Step (ii): Following translocation, the 2A-peptidyl-tRNA is translocated to the P site, the deacylated $tRNA^{gly}$ is translocated to the E site, and deacylated $tRNA^{pro}$ exits the complex. An interaction between the putative helical portion of 2A and the exit pore of the eukaryotic ribosome serves to fix the orientation of the base of the helix within the peptidyl-transferase center of the ribosome. The tight turn (-NPG-) repositions the peptide-tRNA ester linkage away from a conformation that would (normally) lead to peptide bond formation. Step (iii): Prolyl-tRNA enters the A site. The normal nucleophilic attack performed by prolyl-tRNA upon the electrophilic center (the glycyl carbonyl carbon atom) is inhibited due to the reorientation of this center—peptide bond formation is inhibited. Step (iv): The 2A-peptidyl-$tRNA^{gly}$ ester bond is hydrolyzed. Step (v): The nascent polypeptide is released from the translation complex—cleavage. The P site is now occupied by deacylated $tRNA^{gly}$, the A site being occupied by prolyl-tRNA. Step (vi): Deacylated $tRNA^{pro}$ exits the complex, deacylated $tRNA^{gly}$ is translocated from the P to E site, prolyl-tRNA is translocated from the A to P site, and the next aminoacyl-tRNA enters the A site to continue translation—synthesizing a discrete downstream translation product.

This proposed translational model of 2A cleavage activity is novel and represents yet another method whereby viruses modify the host cell's translation process for their own purposes. Such modifications include leaky scanning, reinitiation, suppression of termination, ribosomal frameshifting, internal ribosome entry, and ribosomal "hopping." Indeed, it has emerged that many of these strategies used by viruses to control protein biogenesis are, in fact, also used by cells themselves. Here various aspects of our model will be discussed in light of our data and observations others have made on the process of translation.

2A:Ribosome Interactions

An early observation was that the 2A protein of EMCV was associated with ribosomes prepared from EMCV-infected Krebs-II cell extracts (21). Presumably the components that comprise the exit pore of ribosomes have evolved to minimize any potential interactions with nascent polypeptides. An increasing body of literature now shows, however, that certain peptide sequences may interact with this ribosome structure and in doing so can bring about a "pause" in the elongation cycle and, indeed, inhibit peptidyl-transferase activity (12, 16, 32). During the translation of bacteriophage T4 *gene* 60 a ribosomal "hop" occurs from codon 47 to a matched "landing" codon 50 nucleotides downstream (47). Of specific relevance to our model is the finding that a requirement for this translational bypass is a 16-aa *cis*-acting sequence (residues 17 to 30) in the nascent chain, probably acting in the exit channel of the ribosome (47). This type of translational control

FIGURE 3 2A-like sequences. (A) Insect virus polyproteins are shown together with the location of the 2A-like sequence (shaded rectangle), picornavirus protein 2C-like domain (open circles), proteinase domain (open squares), and polymerase domains (open diamonds). (B) The position of the type C rotavirus 2A-like sequence (shaded rectangle) is shown. (C) The sequence and position of the 2A-like sequences from *Trypanosoma* spp. repeated sequences are shown (shaded rectangle).

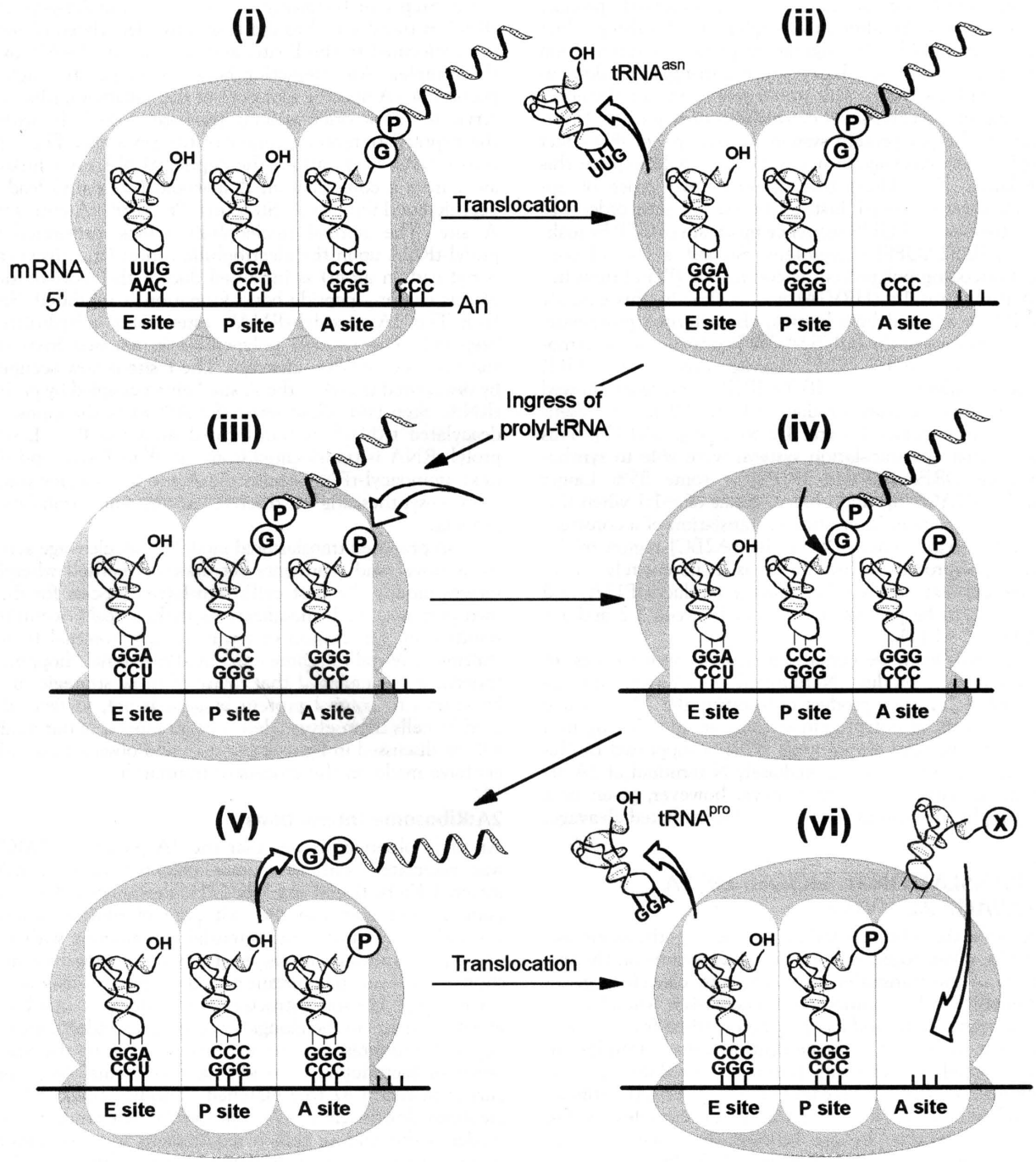


FIGURE 4 Translational model of 2A-mediated "cleavage." Step (i): The synthesis of the 2A peptide sequence is completed with the 2A-peptidyl-tRNA complex in the ribosomal A site. Step (ii): This complex is translocated from the A to P site by eEF2. Step (iii): Prolyl-tRNA is bound to the A site. Step (iv): Cleavage of the peptide-tRNA ester linkage occurs. Step (v): The nascent peptide is released from the ribosome. Step (vi): Prolyl-tRNA is translocated from the A to P site, the next aminoacyl-tRNA is bound to the A site, and translation of the downstream product continues.

has also been implicated in the expression of cellular genes (reviewed in references 8 and 9).

The structural model of 2A we have developed is one of a helix with a tight turn at its C terminus. In this model the helical portion of 2A would, at a specific stage in the elongation/translocation cycle, interact with the exit pore of the ribosome. This would result in 2A adopting a specific, fixed orientation and it conferring a strong spatial constraint upon the tight turn at the base of the helix, reorienting the peptide-tRNA ester linkage away from its normal conformers. The length of the proposed helical portion of the 2A sequence is some 27 Å, which could be entirely accommodated within the exit tunnel of the ribosome, some 100 Å long (1).

Inhibition of Peptidyl-Transferase Activity

During peptide bond formation the peptidyl-tRNA electrophilic center (P site) must occupy a space within the peptidyl-transferase site, which is accessible by the nucleophilic center of the aminoacyl-tRNA (A site). Our site-directed mutagenetic data show that if this incoming nucleophile is prolyl-tRNA, then cleavage occurs extremely efficiently; if the nucleophile is glycyl-tRNA, then cleavage occurs extremely inefficiently; and if the nucleophile is any other aminoacyl-tRNA, then cleavage does not occur at all. It is known that prolyl-tRNA is the poorest nucleophile of all the possible aminoacyl-tRNAs for two reasons: first, that the nitrogen is a secondary amine but, perhaps more important, that this secondary amino group is sterically hindered due to its location in a five-membered pryolidone ring structure. When analogues of the antibiotic puromycin were tested for inhibition of translation, the poorest analogue was that containing proline (23). Furthermore, 3′-O-prolyl adenosine was found to be the worst 3′-O-aminoacyl adenosine substrate for peptidyl-transferase activity (37). Interestingly, in both studies glycine proved to be the next poorest nucleophile—consistent with our site-directed mutagenesis data.

Recent analyses of the function of the putative catalytic bases within the peptidyl-transferase center has led to the proposal that transpeptidation is promoted not through chemical catalysis but simply by the proper positioning of the substrates (27)—the essence of our model is that this precise positioning is subverted by the nascent 2A sequence, thereby inhibiting transpeptidation.

Cleavage of the 2A Peptidyl-tRNA Ester Linkage

Given that peptide bond formation between G18 and P19 is inhibited by either or both of the mechanisms described above, how could this ester bond be cleaved? Presumably the intrinsic rate of ester bond cleavage is not sufficient to account for the observed rapid rate of cotranslational cleavage, since (i) peptidyl-tRNA hydrolase is required to "scavenge" peptidyl-tRNAs arising from abortive termination, and (ii) a system has evolved in bacteria (tmRNA) to rescue "stalled" translation complexes where truncation of the mRNA has occurred. This implies that the peptidyl-tRNA ester linkage within the cytoplasm or within such a translational complex is relatively stable under physiological conditions. We have proposed a mechanism in which the ester moiety is activated as an electrophile precisely in the same manner as in peptidyl transfer but that the nucleophile is an activated water molecule. Magnesium ions are known to be required for peptidyl-transferase activity, and we have suggested that a Mg^{2+} ion can bind at the base of the 2A helix axis (36), positioning the Mg^{2+} ion in the negatively charged field of the helix dipole, together with a coordinated water molecule, perfectly for attack on the peptidyl-tRNA ester carbonyl group.

Synthesis of Full-Length [GFP2AGUS] Translation Products

One notable difference between 2A activity in its native polyprotein context and in our artificial polyprotein systems is that ~10% of the translation products are full length. When these constructs were modified by the addition of just 5 aa of the native polyprotein upstream context of 2A, the amount of uncleaved translation product was much reduced (5, 7). The FMDV 2A is suboptimal in heterologous protein contexts and requires sequences from the C terminus of FMDV capsid protein 1D for complete cleavage. If the function of the helical portion of 2A is to interact with the exit pore of the ribosome, thereby "fixing" the reverse turn at its base, then the 18-aa 2A sequence, by itself, is too short to accomplish this completely. The increased conformational space of the carbonyl carbon would permit the prolyl-tRNA to access this electrophilic center and form a peptide bond. Increasing the length by 5 aa could, however, have the effect of increasing the binding to the ribosomal exit pore and restricting the conformational freedom of the carbonyl carbon. Consistent with this notion is the observation that N-terminal truncation of 2A sequence resulted in increased peptide bond formation (35). The lack of conservation immediately upstream of the conserved -DxExNPG P- motif suggests that there are a (limited) number of ways in which 2As can interact with the ribosomal exit pore to achieve the correct orientation of the reverse-turn-ester linkage within the peptidyl-transferase center.

Synthesis of the Discrete Downstream Translation Product

Following cleavage of the ester bond, a deacylated tRNA is present in the P site and a prolyl- (rather than peptidyl-) tRNA occupies the A site—a situation analogous, but not identical, to normal peptide bond synthesis. Translocation of the prolyl-tRNA into the P site would permit ingress of the next aminoacyl-tRNA and synthesis of the downstream translation product to occur. Were the A site to be unoccupied at this stage, we suggest termination of translation might well occur. The kinetics of the hydrolysis of the ester bond would need to be such that it be slower than the combined rate of (i) the translocation of (2A)peptidyl-tRNA from the A to P site plus (ii) the overall process of binding prolyl-tRNA into the A site.

The Imbalance in Synthesis—Cleavage in the Ribosomal A Site?

Immediately after the addition of G18 to the growing nascent chain the (2A)peptidyl-tRNA complex is present within the A site of the ribosome. One can envisage that were this complex to reside in the A site for relatively prolonged periods, then cleavage could occur in this site. This would result in deacylated tRNAs being present in both P and A sites—directly analogous to termination of translation. This would, therefore, result in the synthesis of the upstream translation product alone. It was shown that during the course of a picornavirus infection, perhaps not surprisingly, the rate of ribosome processivity decreases (13, 30, 41), although the rate-limiting steps are not known. Interestingly, translation studies on cardiovirus RNA using Krebs-2 cell-free extracts showed a "transla-

tional barrier" in the central region of the genome that could be overcome by the addition of eEF2 (43).

The Imbalance in Synthesis—Cleavage in the Ribosomal P Site?

Cleavage of the ester bond while the peptidyl-tRNA complex was in the P site could occur in two states: the A site occupied by prolyl-tRNA or the A site being unoccupied. If the A site were unoccupied, then, presumably, this situation mimics that which occurs in the termination of translation and the ribosome subunits would dissociate. Were cleavage of the ester bond to occur while the A site was occupied by prolyl-tRNA (Fig. 4, step v), then the closest analogous situation that would occur during normal translation is that of dipeptidyl-tRNA in the P site during the early stages of protein synthesis. Translation complexes with short nascent peptides are less stable than those with longer peptides (17), an effect exacerbated by hydrophobic residues (26).

In summary, we and others have shown the aphtho- and cardiovirus 2A/2B cleavage is mediated by an oligopeptidic region, representing either the whole (aphthoviruses) or part (cardioviruses) of the 2A region. We provide evidence that this method of protein biogenesis is not confined to the picornaviruses, but is used by other RNA viruses and virus-like mobile genetic elements. We have, however, analyzed only the 2A-like tracts that corresponded to the 18-aa region of FMDV. The figures given for the cleavage activities may very well, therefore, not represent the true activities of these sequences in their native protein contexts, but be the product of the (suboptimal) lengths of the 2A-like sequences we analyzed. The mechanism that is consistent with all of our translational and site-directed mutagenetic data is one of modification of the host-cell translational process at a specific site within the polyprotein. In our model of cleavage, 2A acts as a *cis*-acting hydrolase element (not an enzyme sensu stricto), rather than a proteinase, and we propose the term "chysel" to refer to its intriguing activity.

We gratefully acknowledge the Biotechnology and Biological Sciences Research Council and The Wellcome Trust for supporting this work.

REFERENCES

1. **Ban, N., P. Nissen, J. Hansen, M. Capel, P. B. Moore, and T. A. Steitz.** 1999. Placement of protein and RNA structures into a 5Å-resolution map of the 50S ribosomal subunit. *Nature* **400:**841–847.
2. **Batson, S., and K. Rundell.** 1991. Proteolysis at the 2A/2B junction in Theiler's murine encephalomyelitis virus. *Virology* **181:**764–767.
3. **Chaplin, P. J., E. B. Camon, B. Villarreal-Ramos, M. Flint, M. D. Ryan, and R. A. Collins.** 1999. Production of interleukin-12 as a self-processing polypeptide. *J. Interferon Cytokine Res.* **19:**235–241.
4. **de Felipe, P., V. Martín, M. L. Cortés, M. D. Ryan, and M. Izquierdo.** 1999. Use of the 2A sequence from foot-and-mouth disease virus in the generation of retroviral vectors for gene therapy. *Gene Ther.* **6:**198–208.
5. **Donnelly, M. L. L., D. Gani, M. Flint, S. Monoghan, and M. D. Ryan.** 1997. The cleavage activity of aphtho- and cardiovirus 2A proteins. *J. Gen. Virol.* **78:**13–21.
6. **Donnelly, M. L. L., L. E. Hughes, G. Luke, X. Li, H. Mendoza, E. ten Dam, D. Gani, and M. D. Ryan.** 2001. The 'cleavage' activities of FMDV 2A site-directed mutants and naturally-occurring '2A-like' sequences. *J. Gen. Virol.* **82:**1027–1041.
7. **Donnelly, M. L. L., G. Luke, A. Mehrotra, X. Li, L. E. Hughes, D. Gani, and M. D. Ryan.** 2001. Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J. Gen. Virol.* **82:**1013–1025.
8. **Farabaugh, P. J.** 1996. Programmed translational frameshifting. *Microbiol. Rev.* **60:**103–134.
9. **Gesteland, R. F., and J. F. Atkins.** 1996. Recoding: dynamic reprogramming of translation. *Annu. Rev. Biochem.* **65:**741–768.
10. **Gopinath, K., J. Wellink, C. Porta, K. M. Taylor, G. P. Lomonossoff, and A. van Kammen.** 2000. Engineering cowpea mosaic virus RNA-2 into a vector to express heterologous proteins in plants. *Virology* **267:**159–173.
11. **Grubman, M. J., and B. Baxt.** 1982. Translation of foot-and-mouth disease virion RNA and processing of the primary cleavage products in a rabbit reticulocyte lysate. *Virology* **116:**19–30.
12. **Gu, Z., R. Harrod, E. J. Rogers, and P. S. Lovett.** 1994. Anti-peptidyl transferase leader peptides of attenuation-regulated chloramphenicol-resistance genes. *Proc. Natl. Acad. Sci. USA* **91:**5612–5616.
13. **Hackett, P. B., E. Egberts, and P. Traub.** 1978. Translation of ascites and mengovirus RNA in fractionated cell free systems from uninfected and mengovirus-infected Ehrlich-ascites-tumor cells. *Eur. J. Biochem.* **83:**341–352.
14. **Hahn, H., and A. C. Palmenberg.** 1996. Mutational analysis of the encephalomyocarditis virus primary cleavage. *J. Virol.* **70:**6870–6875.
15. **Halpin, C., S. E. Cooke, A. Barakate, A. El Amrani, and M. D. Ryan.** 1999. Self-processing polyproteins—a system for co-ordinate expression of multiple proteins in transgenic plants. *Plant J.* **17:**453–459.
16. **Harrod, R., and P. S. Lovett.** 1995. Peptide inhibitors of peptidyltransferase alter the conformation of domains IV and V of large subunit rRNA: a model for nascent peptide control of translation. *Proc. Natl. Acad. Sci. USA* **92:**8650–8654.
17. **Karimi, R., M. Y. Pavlov, V. Heurgue-Hamard, R. H. Buckingham, and M. Ehrenberg.** 1998. Initiation factors IF1 and IF2 synergistically remove peptidyl-tRNAs with short polypeptides from the P-site of translating *Escherichia coli* ribosomes. *J. Mol. Biol.* **281:**241–252.
18. **Kokuho, T., S. Watanabe, Y. Yokomizo, and S. Inumaru.** 1999. Production of biologically active, heterodimeric porcine interleukin-12 using a monocistronic baculoviral expression system. *Jap. Vet. Immunol. Immunopathol.* **72:**289–302.
19. **Martin, F., C. Maranon, M. Olivares, C. Alonso, and M. C. Lopez.** 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. *J. Mol. Biol.* **247:**49–59.
20. **Mattion, N. M., E. C. Harnish, J. C. Crowley, and P. A. Reilly.** 1996. Foot-and-mouth disease virus 2A protease mediates cleavage in attenuated Sabin 3 poliovirus vectors engineered for delivery of foreign antigens. *J. Virol.* **70:**8124–8127.
21. **Medvedkina, O. A., I. V. Scarlat, N. O. Kalinina, and V. I. Agol.** 1974. Virus-specific proteins associated with ribosomes of Krebs-II cells infected with encephalomyocarditis virus. *FEBS Lett.* **39:**4–8.
22. **Michiels, T., V. DeJong, R. Rodrigus, and C. Shaw-Jackson.** 1997. Protein 2A is not required for Theiler's virus replication. *J. Virol.* **71:**9549–9556.

23. **Nathans, D., and A. Niedle.** 1963. Structural requirements for puromycin inhibition of protein synthesis. *Nature* **197:**1076–1077.
24. **Palmenberg, A. C., G. D. Parks, D. J. Hall, R. H. Ingraham, T. W. Seng, and P. V. Pallai.** 1992. Proteolytic processing of the cardioviral P2 region: primary 2A/2B cleavage in clone-derived precursors. *Virology* **190:**754–762.
25. **Percy, N., W. S. Barclay, A. Garcia-Sastre, and P. Palese.** 1994. Expression of a foreign protein by influenza A virus. *J. Virol.* **68:**4486–4492.
26. **Picking, W. D., O. W. Odom, T. Tsalkova, I. Serdyuk, and B. Hardesty.** 1991. The conformation of nascent polylysine and polyphenylalanine peptides on ribosomes. *J. Biol. Chem.* **266:**1534–1542.
27. **Polacek, N., M. Gaynor, A. Yassin, and A. S. Mankin.** 2001. Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide. *Nature* **411:** 498–501.
28. **Precious, B., D. F. Young, A. Bermingham, R. Fearns, M. D. Ryan, and R. E. Randall.** 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. *J. Virol.* **69:**8001–8010.
29. **Pringle, F. M., K. H. J. Gordon, T. N. Hanzlik, J. Kalmakoff, P. D. Scotti, and V. K. Ward.** 1999. A novel capsid expression strategy for *Thosea asigna* virus (*Tetraviridae*). *J. Gen. Virol.* **80:**1855–1863.
30. **Ramabhadran, T. V., and R. E. Thatch.** 1981. Translational elongation rate changes in encephalomyocarditis virus-infected and interferon-treated cells. *J. Virol.* **39:** 573–583.
31. **Robertson, B. H., M. J. Grubman, G. N. Weddell, D. M. Moore, J. D. Welsh, T. Fischer, D. J. Dowbenko, D. G. Yansura, B. Small, and D. G. Kleid.** 1985. Nucleotide and amino acid sequence coding for polypeptides of foot-and-mouth disease virus type A12. *J. Virol.* **54:**651–660.
32. **Rogers, E. J., and P. S. Lovett.** 1994. The *cis*-effect of a nascent peptide on its translating ribosome: influence of the cat-86 leader pentapeptide on translation at leader codon 6. *Mol. Microbiol.* **12:**181–186.
33. **Ryan, M. D., G. J. Belsham, and A. M. Q. King.** 1989. Specificity of substrate-enzyme interactions in foot-and-mouth disease virus polyprotein processing. *Virology* **173:** 35–45.
34. **Ryan, M. D., A. M. Q. King, and G. P. Thomas.** 1991. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J. Gen. Virol.* **72:**2727–2732.
35. **Ryan, M. D., and J. Drew.** 1994. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J.* **13:**928–933.
36. **Ryan, M. D., M. L. L. Donnelly, A. Lewis, A. P. Mehrotra, J. Wilkie, and D. Gani.** 1999. A model for non-stoichiometric, co-translational protein scission in eukaryotic ribosomes. *Bioorganic Chem.* **27:**55–79.
37. **Rychlik, I., J. Cerna, S. Chladek, P. Pulkrabek, and J. Zemlicka.** 1970. Substrate specificity of ribosomal peptidyl transferase. *Eur. J. Biochem.* **16:**136–142.
38. **Schmidt, M., and A. Rethwilm.** 1995. Replicating foamy virus-based vectors directing high-level expression of foreign genes. *Virology* **210:**167–178.
39. **Smerdou, C., and P. Liljestrom.** 1999. Two-helper RNA system for production of recombinant Semliki Forest virus particles. *J. Virol.* **73:**1092–1098.
40. **Smolenska, L., L. M. Roberts, D. Learmonth, A. J. Porter, W. J. Harris, T. M. A. Wilson, and S. S. Cruz.** 1998. Production of a functional single chain antibody attached to the surface of a plant virus. *FEBS Lett.* **441:**379–382.
41. **Summers, D. F., J. V. Maizel, and J. E. Darnell.** 1967. Decrease in size and synthetic activity of poliovirus polysomes late in the infectious cycle. *Virology* **31:**427–435.
42. **Suzuki, N., L. M. Geletka, and D. L. Nuss.** 2000. Essential and dispensable virus-encoded replication elements revealed by efforts to develop hypoviruses as gene expression vectors. *J. Virol.* **74:**7568–7577.
43. **Svitkin, Y. V., and V. I. Agol.** 1983. Translational barrier in central region of encephalomyocarditis virus genome. *Eur. J. Biochem.* **133:**145–154.
44. **Toyoda, H., M. J. H. Nicklin, M. G. Murray, C. W. Anderson, J. J. Dunn, F. W. Studier, and E. Wimmer.** 1986. A second virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein. *Cell* **45:**761–770.
45. **van der Ryst, E., T. Nakasone, A. Habel, A. Venet, E. Gomard, R. Altmeyer, M. Girard, and A. M. Borman.** 1998. Study of the immunogenicity of different recombinant mengo viruses expressing HIV1 and SIV epitopes. *Res. Virol.* **149:**5–20.
46. **Varnavski, A. N., and A. A. Khromykh.** 1999. Noncytopathic flavivirus replicon RNA-based system for expression and delivery of heterologous genes. *Virology* **255:** 366–375.
47. **Weiss, R., W. Huang, and D. Dunn.** 1990. A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60. *Cell* **62:**117–126.
48. **Wright, H. T.** 1991. Nonenzymatic deamidation of asparaginyl and glutaminyl residues in proteins. *Crit. Rev. Biochem. Mol. Biol.* **26:**1–52.
49. **Zoll, J., F. J. M. van Kuppeveld, J. M. D. Galama, and W. J. G. Melchers.** 1998. Genetic analysis of mengovirus protein 2A: its function in polyprotein processing and in virus production. *J. Gen. Virol.* **79:**17–25.