Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus genome

Gary Loughran1, Andrew E. Firth2,3, and John F. Atkins4,6

1The Department of Pathology, University of Cambridge, Cambridge CB2 1QP, United Kingdom; 2Department of Human Genetics, University of Utah, Salt Lake City, UT 84112-5330

The genus Cardiovirus (family Picornaviridae) currently comprises the species Encephalomyocarditis virus (EMCV) and Theilovirus. Cardioviruses have a positive-sense, single-stranded RNA genome that encodes a large polyprotein (L-1ABCD-2ABC-3ABCD) that is cleaved to produce approximately 12 mature proteins. We report on a conserved ORF that overlaps the 2B-encoding sequence of EMCV in the +2 reading frame. The ORF is translated as a 128–129 amino acid transframe fusion (2B*) with the N-terminal 11–12 amino acids of 2B, via ribosomal frameshifting at a conserved GGUUUYU motif. Mutations that knock out expression of 2B* result in a small-plaque phenotype. Curiously, although theiloviruses also utilize frameshifting here, albeit into a different reading frame (3). Like many picornaviruses, cardioviruses use an unusual proteolysis-independent but ribosome-dependent mechanism, termed “StopGo” or “stop-carry on,” to separate the L-1ABCD-2A segment of the polyprotein from the 2BC-3ABCD segment during polyprotein translation (4–7). StopGo is mediated by the amino acid motif D(V/I) ExNPG|P (where the “X” represents the junction between 2A and 2B), which, together with upstream amino acids that have a strong alpha-helical propensity, prevents formation of a peptide bond between Asn-Pro-Gly and Pro but allows the continuation of translation with up to near-100% efficiency (7–13). Recently, SAFV has been shown to be a common and widespread human virus—causing infection mainly in early childhood—although the pathogenicity and clinical significance of SAFV infection is currently unclear (14, 15).

Many viruses harbor sequences that induce a portion of translating ribosomes to shift −1 nt and continue translating in the new reading frame (16, 17). The eukaryotic −1 frameshift site typically consists of a “slippery” heptanucleotide fitting the consensus motif X XXX YYZ, where XXX represents any three identical nucleotides; YYZ represents AAA or UUU; Z represents A, C, or U; and spaces separate zero-frame codons (18). In the tandem slippage model, the P-site anticodon re-pairs from XXY to XXX, whereas the A-site anticodon re-pairs from YYZ to YYY, thus allowing for perfect re-pairing except at the wobble position (19). Because the codon:anticodon duplex in the P site is not monitored so strictly as that in the A site, certain deviations from the canonical XXX of the slippery site are tolerated, including UCC in some members of the Japanese encephalitis serogroup of flaviviruses, GGU in Bean leafroll luteovirus, GUU in Equine arteritis arterivirus, and GGA in Culex flavivirus (20–24). The relative efficiencies of a large selection of slippery heptanucleotides have been tested extensively in reticulocyte lysates (18). The efficiency of frameshifting depends on the identity of the slippery site nucleotides but is typically <1% in the absence of additional stimulatory elements. Thus, most known instances of eukaryotic −1 frameshifting are stimulated (typically to a level of between 1% and 50%, depending on the particular system) by the presence of a 3′ stable RNA secondary structure, such as a pseudoknot or stem-loop, that is separated from the slippery heptanucleotide by a “spacer” region of 5–9 nt (16, 17). Structures of this type are thought to be located at the mRNA unwinding site of the mRNA entrance channel when their stimulatory effect is exerted, with at least part of the relevant consequence being a ribosomal pause. However, another class of 3′ frameshift stimulators appear not to exert their effect via intra-mRNA structure but, instead, are likely to involve direct interaction between the mRNA and tRNA or other translation components (25, 26).

Until recently, most known cases of phylogenetically conserved eukaryotic −1 ribosomal frameshifting involved frameshifting between two long ORFs. The vast majority of known cases came from viral genomes, and very often the frameshift ORF encodes the viral polymerase that is required in lower quantities than the products of the upstream ORF (27). One reason for this bias is simply that long ORFs are easily detected by sequence analysis. With the recent massive increase in sequence data available for many viruses, powerful comparative computational analyses have revealed a number of cases where the frameshift ORF is relatively, and sometimes very, short (28–30). Where the function of frameshifting is simply to produce a truncated form of a protein, the frameshift ORF may be almost nonexistent (e.g., a frameshift site followed immediately by a stop codon), as in the dnaX gene of Escherichia coli and related bacteria (reviewed in ref. 31). Here, we describe a short ORF accessed via −1 ribosomal frameshifting in the cardioviruses. Frameshifting occurs at a conserved GUU UUY motif just downstream of the StopGo cassette. The authors declare no conflict of interest. This article is a PNAS Direct Submission. Freely available online through the PNAS open access option.

Author contributions: G.L., A.E.F., and J.F.A. designed research; G.L. and A.E.F. performed research; G.L., A.E.F., and J.F.A. analyzed data; and G.L., A.E.F., and J.F.A. wrote the paper.

Supporting Information online: www.pnas.org/cgi/doi/10.1073/pnas.1102932108/DCSupplemental.
and, in EMCV, results in the translation of a 128–129 amino acid transframe fusion protein, 2B*.

Results

Computational Analysis Reveals a Conserved Overlapping ORF in EMCV. The polyprotein coding sequences from the 21 full- or near full-length EMCV sequences available in GenBank were extracted, translated, aligned, and back-translated to a nucleotide sequence alignment. Next, the alignment was analyzed for conservation at polyprotein-frame synonymous sites, as described in ref. 32. Briefly, beginning with pairwise sequence comparisons, conservation at synonymous sites (only) was evaluated by comparing the observed number of base substitutions with the number expected under a neutral evolution model. The procedure takes into account whether synonymous site codons are 1-, 2-, 3-, 4-, or 6-fold degenerate and the differing probabilities of transitions and transversions. Statistics were then summed over a phylogenetic tree as described in ref. 33, and averaged over a sliding window.

The analysis revealed a striking and highly statistically significant (\( p \sim 10^{-32} \)) increase in synonymous site conservation in a region covering approximately 120 codons beginning shortly downstream of the junction between the 2A and 2B coding regions (Fig. 1A). Within this region, the mean synonymous substi-
tution rate was only approximately 27% of the genomic average (Fig. 1A). Such peaks in synonymous site conservation are generally indicative of functionally important overlapping elements—either coding or noncoding. In this case, however, an inspection of the positions of stop codons in the +1 and +2 reading frames relative to the polyprotein reading frame, in all 21 sequences, revealed a conserved absence of stop codons in the +2 reading frame in a region that corresponds precisely to the region of enhanced conservation (Fig. 1A), thus suggesting an overlapping coding sequence in the +2 reading frame as a possible explanation for the enhanced conservation at polyprotein-frame synonymous sites. One other explanation for extended conservation could be recombination, a phenomenon that is relatively common in the Picornaviridae, though analysis with a variety of recombination detection algorithms in the RDP3 package failed to reveal any putative recombination events that could explain the conservation in this region (35, 36).

At the 5′ end of the conserved region there is a conserved G GUU UUY motif (where Y represents either a C or a U), which we inferred to be the site of frameshifting (Fig. 2A). This heptanucleotide has been implicated previously as the slippery sequence for −1 ribosomal frameshifting in certain luteoviruses, such as Bean leafroll and Soybean dwarf (22). Moreover, although the G at position one is conserved, the corresponding polyprotein-frame codon is not conserved, with CCG (Pro), CAG (Gln), and AAG (Lys) all being represented in different EMCV polyprotein-frame codons relative to the polyprotein reading frame, in all 21 sequences, revealed a conserved absence of stop codons in the +2 reading frame in a region that corresponds precisely to the region of enhanced conservation (Fig. 1A), thus suggesting an overlapping coding sequence in the +2 reading frame as a possible explanation for the enhanced conservation at polyprotein-frame synonymous sites. One other explanation for extended conservation could be recombination, a phenomenon that is relatively common in the Picornaviridae, though analysis with a variety of recombination detection algorithms in the RDP3 package failed to reveal any putative recombination events that could explain the conservation in this region (35, 36).

However, we were also unable to definitely rule out such a structure, because long-range base pairings and/or structures involving complex tertiary interactions can be difficult to detect based on computational analysis alone (39, 40).

A number of other peaks in synonymous site conservation were detected, including one corresponding to the cis-active RNA element (CRE) (41) that plays a role in replication, as well as several in the region of the genome that encodes 3D, the viral polymerase (Fig. 1A). In fact, such conservation peaks are common in the 5′- and/or 3′-terminal regions of many viral polyprotein coding sequences and often correspond to overlapping noncoding elements, such as RNA secondary structures involved in replication and transcription, translation enhancement, and/or packaging (42). In contrast to these elements, the conservation peak in the +2B region of EMCV extends to a greater degree over a greater region.

Computational Analysis of Theiloviruses. Next, we applied the same analyses to other members of the Cardiovirus genus. Phylogenetically, VHEV clusters with TMEV, whereas SAFV forms another clade that is more closely related to the TMEV–VHEV clade than to EMCV (14, 43). The divergent RTV clade clusters ambiguously with either SAFV or TMEV, depending on the genomic region analyzed, but we chose to analyze RTV alongside TMEV because of the common presence of the L+ ORF (Fig. 1B). When we analyzed the conservation at synonymous sites and the positions of +2-frame stop codons in alignments of either TMEV/RTV (Fig. 1B; 12 sequences) or SAFV (Fig. 1C; 22 sequences) we still observed a striking peak in synonymous site conservation just downstream of the 2A–2B junction. However, the extent of the peak was greatly reduced compared to that seen in EMCV. Moreover, the extent of the corresponding +2-frame ORF was similarly reduced—covering just 8 codons in the majority of sequences (Fig. 2A). Nonetheless, an inspection of the sequence data revealed a conserved G GUU UUU motif at the 5′ end of the conserved region and, again, the functionality of this motif was further supported by conservation of the G at position one despite both UCG (Ser) and UUG (Leu) codons being present at that position (recall that there are six possible codons for each of Ser and Leu, yet only three of the 12 codons end in G; Fig. 2A). Analysis of the 3′ sequence with RNA-folding algorithms again failed to reveal a 3′ RNA structure at the canonical (5–9 nt) distance for stimulation of −1 frameshifting. However, enhanced conservation at polyprotein-frame synonymous sites extended for approximately 5 codons 3′ of the +2-frame stop codon, suggesting that the 3′ nucleotides may nonetheless be involved in the stimulation of frameshifting.

Besides the conservation peak in the sequence encoding 2B, our analysis also recovered the known overlapping gene in the 5′-terminal region of the TMEV/RTV polyprotein coding sequence that encodes L+, a protein essential for the progression of TMEV infection to persistence (Fig. 1B) (44–46) and—as for EMCV—the CRE, and several conservation peaks in the 3′ region of the genome.

There is currently only one VHEV sequence in GenBank with coverage of the 2B region, and because this sequence covers only

---

**Fig. 2.** Representative nucleotide and peptide sequences. (A) Nucleotide sequences flanking the proposed site of frameshifting (confirmed in EMCV) in representative cardiovirus sequences (GenBank accession numbers shown at left). The codons that encode NPGP of the StopGo cassette are highlighted in pale blue. The slippery heptanucleotide where −1 frameshifting is facilitated is highlighted in pale orange. The stop codon that terminates the very short +2-frame ORF is much longer; see Fig. 1A. (B) Amino acid sequence of the transframe protein 2B+ in mengovirus DQ294633. The two antigens used for Abs anti-N and anti-C are highlighted in yellow. Peptide sequences detected by mass spectrometric analysis of 2B+ purified from infected cells are indicated with underscores (anti-N IP) and overscores (anti-C IP). The amino acids, V, corresponding to the site of frameshifting are indicated in bold. (C) Amino acid sequences of the much shorter predicted transframe peptide in representative TMEV and SAFV sequences.
part of the polypeptide, it was not used in the synonymous site
conservation analysis. However, it maintains the G GUU UUU
motif and an 8-codon +2-frame ORF, as observed for TMEV.

Conservation of a More Distal 3'-RNA Structure. Although conserved
RNA structures were not found at the canonical spacing (5–9 nt)
for stimulating ribosomal frameshifting at the G GUU UUY
motif, alignment folding did uncover a potential structure at a
larger spacing (13–14 nt) from the slippery site (Fig. 3). Because
of very high nucleotide conservation in this region, the structure
—although conserved—has no support from compensatory
substitutions within the EMCV alignment. However, a similar
structure was also predicted for TMEV, RTV, and SAFV, where
it is supported by an A-U to U-A substitution between TMEV
and RTV, besides many differences in the primary nucleotide se-
quence between these viruses and EMCV (Fig. 3). Indeed, this
structure was one of just six EMCV–TMEV phylogenetically con-
served potential RNA structures predicted within the cardiovirus
coding region in a computational survey of the Picornaviridae
(47). The structure bears a resemblance—possibly coincident-
—to the cardiovirus CRE, a structure essential for VPg uridylyla-
tion in all picornaviruses (48) (Fig. 3). Using the EMCV and
TMEV CREs defined in ref. 41 as seeds, alignment folding pre-
dicts very similar core CRE structures in EMCV, TMEV, RTV,
and SAFV: a 6-bp stem, 1-nt 3’ bulge, a potential 2-bp stem-ex-
tension, and a 16-nt loop. The predicted stems are supported by
a number of compensatory substitutions, but certain nucleotides in
the loop are completely conserved, most notably the 5’-most 6 nt,
A AAC AC. In fact AAC AC loop motifs are a widespread feature
of Picornaviridae CREs, and this motif is also present in the
loop of the EMCV shift-site 3’-proximal stem-loop structure (48)
(Fig. 3).

Immunoblotting Demonstrates Expression of 2B*. We chose to test
for expression of the predicted transframe protein 2B* in meng-
govirus (a strain of EMCV) rather than TMEV because the
length of 2B* in mengovirus (129 codons) is more amenable to
experimental analysis. Two separate antibodies (Abs) were raised
against predicted antigens within mengovirus 2B*: one against
the amino-terminal 12 aa (anti-N) and a second against the car-
boxy-terminal 14 aa (anti-C). The amino-terminal 12 aa of 2B*
are encoded by the zero-frame sequence 5’ of, and including, the
G GUU UUU motif so that anti-N, but not anti-C, was expected to
react with both 2B and 2B* (Fig. 2B). Additional Abs were also
obtained specifically for 2B but proved ineffective.

Western blots with anti-N detected virus-specific proteins mi-
grating at approximately 13 and 16 kDa in lysates from BHK-21
cells infected for 5 h with wild-type (WT) mengovirus (MOI 10; Fig. 4A). To determine whether these products might correspond to
2B and/or 2B*, separate in vitro translation reactions were pro-
grammed with constructs expressing 2B and 2B*. Although 2B*
has a predicted molecular mass of 14.2 kDa (assuming –1 frame-
shifting at the G GUU UUU motif), in vitro expressed 2B* migrates at a position corresponding to approximately 16 kDa
(Fig. 4A). Meanwhile, in vitro expressed 2B, which has a pre-
dicted molecular mass of 16.5 kDa, migrates at a position corre-
sponding to approximately 13 kDa (Fig. 4A). Thus, we identified
the more slowly migrating product as 2B*. Interestingly, in
infected cells, 2B* was much more readily detected with anti-N
than 2B (Fig. 4A). Possible reasons for this include the unknown
WT frameshifting efficiency, differences in 2B and 2B* stability,
or possible 3C protease cleavage between phylogenetically con-
served Gin and Gly residues just 3 aa downstream of the shift
site (conserved also in TMEV, RTV, and SAFV), which could
remove the N-terminal antigen from a portion of 2B. Although the anti-N

Fig. 3. RNA structures predicted from alignment folding. The G GUU UUU
frameshift site in EMCV and corresponding predicted frameshift site in TMEV
are boxed. A potential lower extension to the stem in EMCV is indicated by
dashes, but this extension is not completely conserved throughout the
EMCV alignment. The EMCV CRE is also shown, with the 5’-most loop nucleo-
tides, AAAAC, that are conserved throughout the cardioviruses highlighted
in bold. Similar nucleotides in the shift-site 3’-proximal stem-loops are also
highlighted in bold. Numbers represent genomic coordinates of selected nu-
cleotides within the sequences indicated by GenBank accession numbers.

Fig. 4. Immunodetection of 2B* and 2B. (A) 2B/2B* anti-N Western blot
(WB) of protein lysates from BHK-21 cells either mock-infected (mock) or
infected (MOI 10) with WT mengovirus (vMC0) for 5 h. As controls for the
apparent molecular masses of 2B and 2B*, in vitro translated (IVT) 2B and
2B*—as well as a negative control (Co) with no input RNA added—were also
immunoblotted. The right-hand panel shows radiolabeled in vitro translated
2B and 2B* (8). (B) Western blots of anti-N and anti-C immunoprecipitates
prepared from mock- or vMC0-infected (MOI 1.0) BHK-21 cells as indicated by
+ or −. Cell lysates were prepared 8 h p.i. and incubated with IgG beads a either
anti-N or anti-C as indicated. Control cell lysates removed before the addition
of beads or Abs are shown in the far right lanes marked “lysates.” The 2B/2B*
N-terminal and 2B* C-terminal antigens are indicated in Fig. 2B.
Western blots of the 2B and 2B⁺ in vitro translations suggested otherwise, it is also possible that 2B suffered from a reduced transfer efficiency or reduced Ab affinity (e.g., due to the differing amino acid sequences immediately adjacent to the N-terminal epitope) relative to 2B⁺.

Western blots with anti-C also detected a virus-specific product in lysates from virus-infected cells that migrates at approximately 16 kDa but, as expected, did not detect a product migrating at approximately 13 kDa (Fig. 4B). To determine whether the approximately 16-kDa product detected by both anti-N and anti-C is the same product, immunocomplexes were prepared from BHK-21 cells infected with WT mengovirus (MOI 1.0) for 8 h. Infected cells were lysed in RIPA buffer to dissociate proteins that could coimmunoprecipitate with the approximately 16-kDa product. As shown in Fig. 4B, the approximately 16-kDa protein immunoprecipitated by anti-N was detected by anti-C, and similarly the approximately 16-kDa protein immunoprecipitated by anti-C was detected by anti-N. This strongly suggests that the antigens recognized by anti-N and anti-C are on the same protein.

Mass Spectrometric Confirmation of 2B⁺ Expression and the Site of Frameshifting. To further establish that the approximately 16-kDa protein detected by both anti-N and anti-C is in fact the predicted 2B⁺, we resolved SDS-PAGE anti-N and anti-C immunoprecipitates prepared from both mengovirus-infected and mock-infected BHK-21 cells, followed by total protein staining with Coomassie blue and excision of gel slices containing a protein migrating at approximately 16 kDa that was present only in those immunoprecipitates prepared from virus-infected cells. Gel slices were digested with trypsin and the resulting peptides analyzed by liquid chromatography–tandem mass spectrometry (LC/MS/MS) [with Fourier transform–ion cyclotron resonance (FT-ICR)]. Peptides corresponding to 75% and 60% of the predicted 2B⁺ sequence were identified, respectively, from the anti-N and anti-C immunoprecipitates (Fig. 2B). Three different sized peptides from both anti-N and anti-C immunoprecipitates spanned the predicted frameshift site. These data confirm that a portion of ribosomes translate 2B⁺ via +1 frameshifting on the G GUU UUY motif during translation of the mengovirus transcript (raw data in Fig. S1).

Frameshifting Is Stimulated by Virus Infection and Requires at Least 50 nt 3′ of the Slip Site. Eukaryotic +1 frameshifting is usually stimulated by a stable 3′ RNA secondary structure beginning 5–9 nt 3′ of the shift site. However, as mentioned previously, no such conserved structures were predicted, but we did identify a potential structure beginning 13–14 nt 3′. The frameshifting efficiency and possible role of sequences 3′ of the slip site were investigated using the dual luciferase reporter assay (49, 50). WT (G GUU UUU) and mutant (A GUG UUU) slip-site sequences, as well as up to 125 nt 3′, were cloned between the two luciferase genes in vector pDluc so that the downstream firefly luciferase gene was in the −1 frame relative to the upstream renilla luciferase gene. Each frameshift construct had a corresponding in-frame control (IFC) construct in which the firefly gene was in the same frame as the renilla gene. Frameshifting efficiencies were determined by comparing the ratio of firefly to renilla enzymatic activities in parallel cell cultures transfected with either the frameshift or IFC construct. Constructs were transfected into BHK-21 cells 18 h before either mock-infection or infection with WT mengovirus (MOI 2.0). Frameshifting efficiencies of approximately 3.5% and 6.6% were observed in virus-infected cells transfected with constructs containing the WT shift site and, respectively, 50 or 125 nt 3′ (Fig. 5). In contrast, the same constructs promoted only low levels (<1%) of frameshifting in the absence of virus infection (Fig. 5). When the shift site was mutated, frameshifting was reduced to background levels. The potential RNA structure in EMCV ends

![Image](https://example.com/image.png)

#### Fig. 5.
Extent of the 3′ sequence that contributes to frameshift stimulation. Dual luciferase assay showing frameshifting efficiencies following transfection of WT and slip-site mutated (SS) pDluc constructs with 45, 50, or 125 nt 3′ of the slip site. Relevant IFCs were also transfected in parallel. Eighteen hours later, transfecants were either mock-infected or infected with vMCO (MOI 2.0) for 7 h. Cells were then lysed and luciferase activities measured. Error bars represent the standard deviation of four independent experiments, in each of which each construct was transfected in quadruplicate.

48 nt 3′ of the G GUU UUY motif (Fig. 3) and, consistent with this structure having a stimulatory effect on frameshifting, constructs with 50 nt of 3′ sequence stimulated frameshifting in virus-infected cells to a level 3-fold higher than constructs with just 45 nt of 3′ sequence (Fig. 5). Targeted mutations that disrupted the potential structure also reduced frameshifting to low levels; however, mutations designed to restore the potential stem but with reversed base pairings did not restore frameshifting to WT levels, suggesting that, if the structure does indeed stimulate frameshifting, then both the structure and primary sequence elements within the structure are important for this effect (Fig. S2). Because of the potential for additional stimulatory factors, and the effect of virus infection, it is not clear at this stage whether these values are closely representative of the actual level of frameshifting in the virus itself.

Inactivating 2B⁺ Results in a Small-Plaque Phenotype. To investigate whether 2B⁺ is an essential protein, we introduced mutations into an infectious clone of mengovirus. Three mutants were generated that are synonymous with respect to the polyprotein reading frame but that are unable to express full-length 2B⁺: (i) a G GUU UUU to A GUG UUU mutation to prevent frameshifting at the slip site (SS), (ii) a CCG AAC GAC to CCU AAI GAC mutation at nt positions +48 and +51 from the shift site to introduce two consecutive +2 frame premature termination codons just 3′ of the slip site (PTC1), and (iii) a CUU UCG GAU to CUU AGU GAU mutation at nt positions +153 – 156 from the shift site to introduce an alternative set of +2 frame stop codons (PTC2). Although StopGo separation is not known to be affected by downstream elements, to avoid this as a potential confounding factor we tested whether any of the mutants might affect StopGo. Parts of the mengovirus genome (specifically, from the start of 1D to the end of 2B) from WT, SS, PTC1, and PTC2 viruses were cloned into a plasmid suitable for cell-free transcription and translation (Fig. 6A). As a control for impaired StopGo, a construct in which the critical NGP motif was mutated to NPLV was also prepared. Similar to WT, failure of StopGo separation occurred at just a very low level in the SS, PTC1, and PTC2 mutants (Fig. 6B).

RNA transcribed from WT, SS, PTC1, and PTC2 infectious clones was transfected into BHK-21 cells to generate virus stocks. Although end-point titers were similar between all viruses, plaque sizes were significantly reduced when either the slip site was mutated or premature termination codons were introduced into the 2B⁺ reading frame (Fig. 6C). We also consistently observed a delayed cytopathic effect (CPE) when BHK-21 cells were infected
with any of the three mutant viruses [up to approximately 24 h postinfection (p.i.) at MOI 1.0] compared to WT virus infections (CPE within approximately 10 h p.i. at MOI 1.0). Host cell shut-off was also delayed in the mutants, with many host cell proteins expressed at higher levels after 8 h of mutant virus infection than after 8 h of infection with WT virus (arrowheads in Fig. 6D). This may be due to slower virus multiplication within the cell, rather than a specific role for 2B* in shut-off.

Immunoprecipitation (IP) experiments using lysates from BHK-21 cells infected with WT mengovirus for 8 h or the three mutant viruses for 22 h detected the approximately 16-kDa product (2B*) for WT virus but failed to detect any specific product from the mutant viruses when probed with either anti-N or anti-C (Fig. 6C). Immunoblotting with anti-N detected the approximately 13-kDa product (2B) was detected for WT mengovirus (MOI 1.0) for 8 h or each indicated mutated mengovirus (MOI 1.0) for 22 h. (D) Pulse labeling of vMC0, vMC0-SS, vMC0-PTC1, vMC0-PTC2 (MOI 10), and mock-infected BHK-21 cells at 8 and 22 h p.i.. The positions of the major mengovirus proteins are shown to the left. Arrowheads indicate positions of host proteins that are absent or decreased in cells infected with vMC0 when compared to mock-infected cells. There were no labeled proteins in the WT lane at 22 h p.i. because cells infected with vMC0 for 22 h had already undergone CPE.

Given that 2B* mutants have an attenuated phenotype in cell culture, one might expect there to be selection for reversion to WT sequence. To test this, the SS, PTCP1, and PTCP2 mutants were serially passaged in cell culture, and progeny from each passage were sequenced. The shift-site mutant (SS) partially reverted (A GUG UUU to A GUU UUU) from passage 3 in one experiment and from passage 8 in another, but was not observed to fully revert to WT G GUU UUU by passage 20. Indeed (see the Introduction), the G at position 4 in SS is expected to have a much greater inhibitory effect on frameshifting than the A at position 1, and the partial revertant A GUU UUU is likely to allow some frameshifting (cf. ref. 18). In fact, 2B* expression is detectable via Western blotting using lysates prepared from cells infected with the partial revertant A GUU UUU via the transframe 2B* is translated. (C) Mean plaque diameter determined by measuring 100 plaques formed by standard plaque assay after infection with each WT or mutated mengovirus as indicated. Error bars indicate standard deviations. Inset: anti-N and anti-C Western blots of anti-N immunocomplexes on lysates from BHK-21 cells infected with either WT mengovirus (MOI 1.0) for 8 h or each indicated mutated mengovirus (MOI 1.0) for 22 h. (D) Pulse labeling of vMC0, vMC0-SS, vMC0-PTC1, vMC0-PTC2 (MOI 10), and mock-infected BHK-21 cells at 8 and 22 h p.i.. The positions of the major mengovirus proteins are shown to the left. Arrowheads indicate positions of host proteins that are absent or decreased in cells infected with vMC0 when compared to mock-infected cells. There were no labeled proteins in the WT lane at 22 h p.i. because cells infected with vMC0 for 22 h had already undergone CPE.

This is not surprising because (unlike the shift-site mutant) single
nucleotide changes are not expected to even partially improve fitness for these mutants.

Discussion

We have demonstrated the existence of programmed ribosomal frameshifting and a 128–129 aa transframe gene, 2B*, in EMCV. Evidence includes (i) a 117-codon overlapping ORF conserved in all available EMCV sequences, (ii) a statistically highly significant enhancement in the conservation at polyprotein-frame synonymous sites coinciding with the position of the ORF, (iii) a well-defined and conserved translation mechanism via programmed ribosomal frameshifting at the 5' end of the ORF, (iv) a product of the expected size was specifically detected by two separate Abs raised against N- and C-terminal peptides, (v) mass spectrometry of this product purified from infected cells confirmed expression of 2B* and the predicted site of frameshifting, and (vi) three separate 2B* knockout mutants, each differing from WT virus by just 2–4 mutations synonymous in the polyprotein frame, exhibited an attenuated small-plaque phenotype.

Interestingly, a frameshifting stimulatory RNA structure was not apparent at the canonical distance of 5–9 nt 3′ of the shift site. The reporter assays with 45 and with 50 nt 3′ of the shift site show the importance of the longer sequence for efficient frameshifting. Fifty nucleotides, but not 45 nt, is adequate to allow formation of the potential stem-loop structure identified in EMCV, TMEV, and SAFV (Fig. 3). As the 5′ end of this potential structure is 13–14 nt 3′ of the shift site, if it forms it would be at the leading edge of the mRNA entrance tunnel at the time the shift occurs. There are other candidates for a frameshifting stimulator at this position (albeit for +1 not −1 frameshifting), as well as evidence for 3′ stimulators that do not act via intra-mRNA base pairing (25, 26, 53). The predicted 3′ RNA structure bears some resemblance to the cardiovirus CRE element that binds 3CD (41, 48) (Fig. 3). If the shift-site 3′-proximal stem-loop also binds 3CD, then it is plausible that the RNA-protein complex could act as an unusual stimulator for frameshifting (with implications for the distance between the shift site and the stimulator), and this would also explain why efficient frameshifting was only observed in virus-infected cells. Such a mechanism has a precedent in an artificial system where eukaryotic –1 frameshifting stimulated by a 3′-adjacent iron-responsive element (IRE) stem-loop structure was shown to be greatly enhanced by stimulation of IRE-binding proteins (54). However, this is just one of many possible explanations: virus infection could also conceivably modulate frameshifting efficiency via other virus or host trans-acting protein factors, changes in tRNA populations, ribosome modifications, and other changes in cellular environment and/or architecture (cf. refs. 55–57). It is also possible that non-Watson–Crick intra-mRNA interactions extend the structure into the spacer region in such a way that allows it to affect frameshifting in a manner more analogous to typical frameshift-stimulating structures (albeit still, unusually, apparently dependent on virus infection). Establishing the nature and mode of action of the 3′ stimulator, as well as the influence of virus-infection, is, however, outside the scope of this report.

Although we have not tested the functionality of the proposed site of frameshifting in SAFV and TMEV, it seems probable that this is an attractive hypothesis given that the efficiency via other virus or host trans-acting proteins, in particular the polymerase 3D, freeing up more ribosomes to translate the 5′-encoded structural proteins; though, to be effective, such a sink would require significant levels of frameshifting (e.g., 50%). Interestingly, StopGo has also been proposed as a candidate for a ribosome sink that might be regulated during the course of virus infection (7), though a possible advantage of frameshifting over StopGo as a ribosome sink is that frameshifting and termination at an out-of-frame stop codon cleanly removes ribosomes from the message, whereas discontinuation of translation at the StopGo cassette may result in a stalled ribosome. A third explanation might be that frameshifting acts as a ribosome sink to reduce the number of ribosomes translating the 3′-encoded replicase proteins, in particular the polymerase 3D, freeing up more ribosomes to translate the 5′-encoded structural proteins; though, to be effective, such a sink would require significant levels of frameshifting (e.g., 50%). Interestingly, StopGo has also been proposed as a candidate for a ribosome sink that might be regulated during the course of virus infection (7), though a possible advantage of frameshifting over StopGo as a ribosome sink is that frameshifting and termination at an out-of-frame stop codon cleanly removes ribosomes from the message, whereas discontinuation of translation at the StopGo cassette may result in a stalled ribosome. A third explanation might be that frameshifting plays some sort of regulatory role in the viral lifecycle (cf. ref. 39). This is an attractive hypothesis given that the efficiency of frameshifting in EMCV appears to be modulated by viral infection.

For whatever benefit, however, it seems that programmed frameshifting evolved at this location in the ancestral cardiovirus. Then in EMCV, but not TMEV or SAFV, the ancestral frameshift site was co-opted (perhaps while also retaining the original function) as a suitable site to begin the evolution of a long overlapping gene via incremental elongation by substitutions of intervening +2-frame stop codons with sense codons and selection on the encoded amino acid sequence (cf. ref. 34). This finding adds to a small number of known cases of phylogenetically conserved overlapping genes that are internal to longer coding sequences and accessed via programmed ribosomal frameshifting. Other examples include potyvirus P1PO, alphavirus TF, and flavivirus NS1 (21, 28, 29). Such overlapping genes are difficult to identify and are often overlooked. However, it is important to be aware of these genes as early as possible. Undetected overlapping genes are a source of confusion and confound mutagenic analysis of the genes they overlap, in this case 2B. Furthermore, only once
it has been identified can the functions of an overlapping gene be investigated in their own right.

Materials and Methods
Computational Analysis. Cardiovascular nucleotide sequences in Genbank with complete or near-complete coverage of the polyprotein coding sequence were identified by applying National Center for Biotechnology Information tblastn (58) to the polyprotein sequences derived from GenBank cardiovascular reference sequences NC_001479, NC_001366, NC_009448, and NC_010810. The following sequences were retrieved. EMCV: DQ464062, HG186697, DQ464063, FJ604852, FJ604853, FJ897755, EU780148, EU780149, DQ517424, AF358282, M181861, X00463, DQ288856, X74312, AY296731, M22457, M22458, M37858, X87335, DQ294363, L22089, and DQ385155. TMVE: DQ401688, M16200, EU718732, EU723238, M20562, X56019, M20031, EU810554, and X03377. TBEV: J02435, M11133, M16020, EU718732, EU723238, M20562, X56019, M20031, EU810554, and X03377. VHEV: DQ464063, FJ604852, FJ604853, FJ897755, EU780148, EU780149, DQ517424, AF358282, M181861, X00463, DQ288856, X74312, AY296731, M22457, M22458, M37858, X87335, DQ294363, L22089, and DQ385155. TMEV: DQ401688, M16200, EU718732, EU723238, M20562, X56019, M20031, EU810554, and X03377. VMENTS: BM71861 but has six single-nucleotide deletions that cause local changes in reading frame in the 1A/1B coding region, was omitted from the analysis. Within each clade, the polyprotein-encoding sequences were extracted, translated, aligned, and back-translated to produce nucleotide sequence alignments using EMBOSS and Clustal (59, 60). Besides the full-or near full-length sequences above, one additional partial sequence with coverage of the 2B region was retrieved (namely, EU723237, which was the only VHEV length sequences above, one additional partial sequence with coverage of 2B*). Rabbits were injected with one of the two peptides, and Abs were affinity-purified from immune sera. A third Ab was raised against peptide sequence CAWENVKGTLNNPEF (60). Anti-C was raised against peptide sequence CRDHKPDKPVRRNSS (61). Anti-C was raised against peptide sequence CRDHKPDKPVRRNSS (80) against peptide sequence CRDHKPDKPVRRNSS (61) overnight at 4 °C with gentle rocking. The beads were adsorbed overnight in 2% milk with a dilution of anti-N or anti-C (3 μg of each antibody) overnight at 4 °C.

Antibodies. Polyclonal Abs to two predicted antigens within 2B* were prepared by GenScript. Anti-N was raised against peptide sequence PFTFPK-RQRPVFC (N-terminal 12 aa shared by 2B and 2B* + C-terminal 14 aa of 2B*). Rabbits were injected with one of the two peptides, and Abs were affinity-purified from immune sera. A third Ab was raised against peptide sequence CAWENVKGTLNNPEF (60) against peptide sequence CRDHKPDKPVRRNSS (61) overnight at 4 °C with gentle rocking.

Vaccines and Cell Culture. BHK-21 cells (ATCC) were maintained in DMEM supplemented with 10% FBS, 1 mM L-glutamine, and antibiotics. Virus VMC0 is a recombinant mengovirus that is identical to GenBank accession number DQ294633 except in the 5′-terminal 12 nt where vMC0 has the poly-C tract deleted (62). For virus infection, BHK-21 cells were washed with PBS and then overlaid with VMC0 diluted in PBS/1% FBS at the multiplicities of infection (MOI) indicated throughout the text. After adsorption (60 min at 37 °C) with gentle agitation, virus was removed and cells overlaid with DMEM/2% FBS for the times p.i. indicated.

Immunodetection and IP. Infected cells were lysed in RIPA buffer plus protease inhibitors on ice for 20 min. Cell debris were removed from cell lysates by centrifugation at 15,000 × g at 4 °C for 15 min. Proteins were resolved by 15% SDS-PAGE (Bio-Rad) and transferred to nitrocellulose membranes (Protran) by electrotransfer in a tris/glycine buffer with 20% methanol at 1 mA/cm² (constant amps) for 55 min. Membranes were incubated at 4 °C overnight in 2% milk with a 1:1000 dilution of anti-N or a 1:500 dilution of anti-C. Immunoreactive bands were detected on membranes after incubation with appropriate fluorescein-labeled secondary Abs using a LI-COR Odyssey® infrared Imaging Scanner (LI-COR Biosciences). For IP, cell lysates were incubated with 20 μl of protein G Agarose beads plus anti-N or anti-C overnight at 4 °C. Supernatants were collected then aliquotted and stored at −80 °C.

In Vitro Transcription and Transfection. Plasmids were linearized with BamHI and transcribed with T7 RNA polymerase (Ambion) for 3 h at 37 °C as recommended by the manufacturer. RNA integrity was examined by electrophoresis, and RNA was transcribed into BHK-21 cells using DMRIE-C (Invitrogen) according to the manufacturer’s recommendations.

Cell-Free Translation. We chose not to use rabbit reticulocyte lysates for cell-free translation because of potential distortions in the apparent molecular masses of small proteins caused by large amounts of endogenous β-globin. In addition, Western blotting with rabbit reticulocyte lysates subsequently probed with Abs raised in rabbit often results in high background. Therefore, cell-free translations were performed in human lysates (Determined) determined by using the Dual Luciferase Stop & Glo® Reporter Assay System (Promega). Relative light units were measured on a Veritas Microplate Luminometer with two injectors (Turner Biosystems). Transfected cells were washed once with PBS and then lysed in 12.6 μl of 1× passive lysis buffer, and light emission was measured following injection of 25 μl of either renilla or firefly luciferase substrate. Firefly luciferase activity was calculated relative to the activity of renilla luciferase, and frameshifting efficiencies were determined by comparing the ratio of firefly to renilla enzymatic activities in parallel cell cultures transfected with either the test construct or an IFC.

Plaque Assays. Virus titers were quantified by plaque assay. Twenty-four hours after transfection of each infectious RNA, cells were subjected to three cycles of freezing and thawing followed by centrifugation at 1,800 × g to pellet cell debris. Supernatants were collected then aliquotted and stored at −80 °C. BHK-21 cells were infected with 10-fold dilutions of each supernatant for 1 h at 37 °C. The virus inoculums were then removed and cells were overlaid with 1.0% agarose containing MEM. After 48 h of incubation, cells were fixed at 37 °C for 18 h before virus infection as described above. Relative luciferase activity was calculated relative to the activity of renilla luciferase, and frameshifting efficiencies were determined by comparing the ratio of firefly to renilla enzymatic activities in parallel cell cultures transfected with either the test construct or an IFC.

Plaque Assays. Virus titers were quantified by plaque assay. Twenty-four hours after transfection of each infectious RNA, cells were subjected to three cycles of freezing and thawing followed by centrifugation at 1,800 × g to pellet cell debris. Supernatants were collected then aliquotted and stored at −80 °C. BHK-21 cells were infected with 10-fold dilutions of each supernatant for 1 h at 37 °C. The virus inoculums were then removed and cells were overlaid with 1.0% agarose containing MEM. After 48 h of incubation, cells were fixed with 2.5% formalin, and plaques were visualized with crystal violet.

Metabolic Labeling. BHK-21 cells at 90–100% confluency in 35-mm petri dishes were infected with WT or mutant mengovirus at MOI 10. After 60 min of adsorption at 37 °C, the virus inoculums were removed by aspiration. The cells were washed once with methionine-free DMEM and incubated in 0.25 ml of methionine-free DMEM. The cells were labeled with [35S] methionine (50 μCi/ml) at 37 °C for 30 min. Cells were then washed once in PBS and lysed by being suspended in 0.2 ml of sample buffer and heated at 95 °C for 5 min. Lysates were analyzed by SDS-PAGE (15%). After electrophoresis, the gels were processed for fluorography with Amplify™ (Amersham).

ACKNOWLEDGMENTS. We thank Chad Nelson for performing the mass spectrometry and analysis. We are also very grateful to Martina Scanlan for facilitating use of the virus laboratory. We thank Mike Howard for providing vector pDiuc and Robert Fujinami for providing a copy of the
mengovirus cDNA (pMCO) developed by Ann Palmenberg. We also thank Ann Palmenberg for providing other materials and for stimulating discussions. This work was supported by Science Foundation Ireland Grant 08/N1/1B1889 (to J.F.A.) and Wellcome Trust Grant 088789 (to A.E.F.). J.F.A. was personally supported by US National Institutes of Health Grant R01 GM079523.