JOURNAL OF VIROLOGY, Sept. 1996, p. 6126–6135 Vol. 70, No. 9
0022-538X/96/$04.00+0

Translation Initiation in GB Viruses A and C: Evidence for Internal Ribosome Entry and Implications for Genome Organization

JOHN N. SIMONS,[1]* SURESH M. DESAI,[1] DERK E. SCHULTZ,[2] STANLEY M. LEMON,[2,3] AND ISA K. MUSHAHWAR[1]

Virus Discovery Group, Experimental Biology Research, Abbott Laboratories, North Chicago, Illinois 60064-4000,[1] and Departments of Medicine[2] and Microbiology and Immunology,[3] University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7030

Received 3 April 1996/Accepted 10 June 1996

GB viruses A and C (GBV-A and GBV-C) are two recently described RNA viruses which appear to be members of the *Flaviviridae*. Although both viruses appear to contain long 5′ nontranslated regions, the sites of polyprotein initiation and the presence of core-like proteins remain to be determined. Translation studies were undertaken to determine the mechanism and sites of polyprotein initiation in GBV-A and GBV-C. Rabbit reticulocyte lysates programmed with monocistronic RNAs containing 5′ ends of GBV-A or GBV-C fused in-frame with the chloramphenicol acetyltransferase (CAT) open reading frame generated GBV-CAT fusion proteins in vitro. Site-specific mutagenesis and N-terminal sequencing located the sites of translation initiation immediately upstream of the putative signal sequence for the GBV E1 envelope glycoproteins. Efficient translation of the monocistronic GBV-CAT RNAs required the inclusion of GBV coding sequences. This, coupled with the presence of at least 523 nucleotides of 5′ nontranslated RNA containing multiple AUG codons, suggests that translation initiation of these RNAs did not utilize a ribosome scanning mechanism. Translation of bicistronic RNAs containing 5′ nontranslated sequences within the intercistronic space was consistent with the presence of a weakly active internal ribosome entry site in both GBV-A and GBV-C. Secondary structure predictions indicate that the 5′ ends of these viruses assume similar complex structures distinct from those identified in the internal ribosome entry site-containing picornaviruses, pestiviruses, and hepatitis C viruses. The data indicate that GBV-A and GBV-C are unique members of the *Flaviviridae* that do not contain core-like proteins at the N termini of their putative polyproteins.

The GB viruses (GBV-A, -B, and -C) and hepatitis G virus (HGV) are recently discovered RNA viruses which infect primates. GBV-A and GBV-B were molecularly cloned from the serum of a GB agent-infected tamarin (31), while GBV-C was identified in human sera with degenerate PCR primers based on the homologous NS3 genes of hepatitis C virus (HCV), GBV-A, and GBV-B (30). Subsequently, Linnen et al. (13) reported the independent cloning of HGV. Sequence analysis of these four viruses revealed that HGV is related to GBV-A, GBV-B, and HCV (13) but is nearly identical to GBV-C (35). Because GBV-C and HGV are most likely genotypes or strains of the same virus (19), the GBV-C nomenclature will be used throughout this paper.

The GBVs appear to be members of the *Flaviviridae* family. They possess RNA genomes approximately 9.5 kb in size which contain a single long open reading frame (ORF). Structural and nonstructural proteins are contained in the N-terminal one-third and C-terminal two-thirds of the putative viral polyproteins, respectively. Phylogenetic analyses of the nonstructural helicase and replicase genes demonstrate that these viruses are related to, but distinct from, the HCV genus of the family *Flaviviridae* (12a, 18). Specifically, GBV-A and GBV-C appear to be most closely related because they share a common ancestor, while the GBV-A and GBV-C ancestor, GBV-B, and HCV all appear to be equally divergent from other members of the *Flaviviridae*. However, when the 5′ nontranslated regions (NTRs) and structural genes are examined, a more striking division between the GBVs and the other members of the *Flaviviridae* becomes apparent. GBV-B appears to be similar to the HCV and *Pestivirus* genera of the *Flaviviridae*. Conserved sequences present in the 5′ NTRs of HCV and pestiviruses are found in the 5′ NTR of GBV-B, and GBV-B and HCV share closely related RNA secondary structures within the 5′ NTR (6). Moreover, a basic (pI 11.1) core protein is present at the N terminus of the GBV-B putative polyprotein precursor, and two putative envelope glycoproteins with several potential N-linked glycosylation sites are located downstream of the core in GBV-B (18). These structural proteins appear in all members of the *Flaviviridae* examined to date (4, 17). In contrast to GBV-B, examination of GBV-A and GBV-C reveals marked differences between these viruses and other genera of the *Flaviviridae*. GBV-A and GBV-C contain long 5′ NTRs that have limited sequence identity to each other but no identity to GBV-B, HCV, or pestiviruses. GBV-A and GBV-C also encode putative envelope proteins that contain relatively few potential N-linked glycosylation sites. Most strikingly, clearly discernible basic core proteins are not found in the cDNA sequences cloned thus far from these viruses.

The absence of core proteins would distinguish GBV-A and GBV-C from other genera of the *Flaviviridae*. However, several important aspects of the structure of the GBV-A and GBV-C genomes remain undefined. Primary among these is the identification of the AUG codons at which translation of

* Corresponding author. Mailing address: Abbott Laboratories, Dept. 90D, L3, 1401 Sheridan Rd., North Chicago, IL 60064-4000. Phone: (847) 938-8688. Fax: (847) 938-6042. Electronic mail address: simonsj@apmac.abbott.com.

the viral polyproteins initiate. The sequence of GBV-A contains two potential in-frame initiator AUG codons 27 nucleotides (nt [9 amino acids]) and immediately upstream of the putative E1 signal sequence. Similarly, multiple GBV-C sequences possess two to three potential in-frame initiator AUG codons (12a, 13). However, none of these AUGs have been demonstrated to serve as the initiator codon, and initiation at any of these sites would result in a severely truncated core protein at best. It is conceivable that deletions during the cloning of these virus RNAs have resulted in the elimination of core sequences or a disruption of the true ORF in this region of the genome as suggested by Leary et al. (12a). However, multiple reverse transcriptase PCR products generated from the 5′ ends of GBV-A and GBV-C with a variety of primers, polymerases, and conditions (29a), in addition to determining the 5′ GBV-C sequences from 35 viremic individuals (19), provide no support for the existence of additional sequence missing from the previously described cDNA clones. Thus, it is possible that the 5′ ends of these viruses are complete (or nearly complete), and that GBV-A and GBV-C do not encode core proteins. Here, we describe experiments in which we sought additional evidence either supporting or refuting the existence of a core protein in GBV-A and GBV-C by functionally characterizing the 5′ ends of these viruses to determine the sites and mechanism of translation initiation.

Of the genera that make up the *Flaviviridae*, the viruses classified in the *Flavivirus* genus (e.g., yellow fever virus and dengue virus) contain relatively short 5′ NTRs of 97 to 119 nt. In these viruses, translational initiation is thought to utilize a conventional eukaryotic ribosome scanning mechanism in which ribosomes bind the RNA at a 5′ cap structure and scan in a 3′ direction until encountering an AUG codon in a favorable context for initiation (9, 10). In contrast to the *Flavivirus* genus, genomic RNAs from members of the pestiviruses and HCV genera contain relatively long 5′ NTRs of 341 to 385 nt, which in some ways are similar to those of picornaviruses. Extensive studies of the picornavirus 5′ NTRs reveal that translation initiation occurs through a mechanism of internal ribosome entry (8, 15). This internal entry requires a defined segment of the viral 5′ NTR known as an internal ribosome entry site (IRES) or ribosome landing pad. The RNA comprising the *cis*-acting IRES forms highly ordered structures which interact with *trans*-acting cellular translation factors to bind the 40S ribosome subunit at an internal site on the viral message, often many hundreds of nucleotides downstream of the 5′ end of the molecule. Such translation initiation functions in a 5′ cap-independent fashion and is generally not influenced by structure or sequence upstream of the IRES. Practically, the ability of a sequence to function as an IRES is assessed by insertion of the sequence between two cistrons of a bicistronic RNA. If the intercistronic sequence contains an IRES, there is significant translation of the downstream cistron, which is generally independent of the translational activity of the upstream cistron. Studies of the 5′ NTRs of HCV and pestiviruses with bicistronic mRNAs have demonstrated the presence of IRESs in these sequences (21, 23, 32, 33).

To functionally characterize the 5′ ends of GBV-A and GBV-C RNAs, we examined the sites and mechanism of translation initiation of both monocistronic and bicistronic RNAs in a cell-free in vitro translation system. Weak IRES elements were found in the 5′ RNAs of GBV-A and GBV-C, suggesting that these sequences are complete or nearly complete. In addition, the positions of the initiating AUG codons in the monocistronic RNAs, and presumably in the viral genomic RNA as well, demonstrate that GBV-A and GBV-C do not contain core proteins at the N termini of their polyproteins. Thus, GBV-A and GBV-C appear unique from other members of the *Flaviviridae* and may constitute a separate group within this family. Consistent with this hypothesis, we also show that the secondary structures of the 5′ ends of these viruses are different from the conserved structures present in the 5′ NTRs of the pestiviruses, HCV, and GBV-B.

MATERIALS AND METHODS

Plasmids. Various monocistronic and bicistronic plasmids were constructed with PCR-amplified sequences of GBV-A and GBV-C. PCRs utilized components of the GeneAmp PCR kit with AmpliTaq (Perkin-Elmer) as directed by the manufacturer, with final reaction concentrations of 1 μM for oligonucleotide primers and 2 mM $MgCl_2$. PCR products were digested with restriction endonucleases, gel purified, and cloned by standard procedures (25). Monocistronic fusions between GBV sequences and bacterial chloramphenicol acetyltransferase (CAT) were generated by replacement of the hepatitis A virus (HAV) *Hin*dIII-*Xba*I fragment of pHAV-CAT1 (34) with PCR-amplified cDNA from the 5′ ends of GBV-A and GBV-C. The bicistronic constructs were generated in pT7/CAT/ICS/Luc (14) in a two-step procedure. First, monocistronic fusions between GBV and luciferase (Luc) were constructed by insertion of GBV sequences into the *Hin*dIII-*Nco*I-cut plasmid pT7/CAT/ICS/Luc. Bicistronic vectors were constructed by cloning the *Hin*dIII-blunt-end-*Sac*I GBV fragment from these monocistronic vectors into pT7/CAT/ICS/Luc, which had been digested with *Sal*I-blunt-end-*Sac*I. The sequences of the cloned inserts and ligation junctions were confirmed by double-stranded DNA sequencing (Sequenase 2.0; U.S. Biochemical Corp., Cleveland, Ohio). The nomenclature used here (e.g., A15-707) describes the source (GBV-A) and range (nt 15 to 707) of sequence incorporated into the various vectors.

GBV-A sequences (GenBank accession no. U22303) were amplified from a plasmid clone. PCRs for the GBV-A monocistronic and bicistronic constructs utilized the sense primer 5′-TATAATAAGCTTGCCCCGGACCTCCCACCGAG-3′ (*Hin*dIII site underlined) coupled with 5′-GCTCTAGATCGGGAACAACAATTGGAAAG, 5′-GCTCTAGAGCACTGGTGCCGCGAGT, 5′-GCTCTAGAGAGGGGGAAGCAAACCA, and 5′-GCTCTAGACATGGTGAATGTGTCGACCAC (*Xba*I sites underlined) for the monocistronic vectors pA15-707/CAT, pA15-665/CAT, pA15-629/CAT, and pA15-596/CAT, respectively; and 5′-CCATAATCATGAGGGAACAACAATTGGAAAG, 5′-CCATAATCATGAGCCGCGAGTTGAAGAGCAC, and 5′-GCCAAGCCATGGTGAATGTG-3′ (*Bsp*HI or *Nco*I sites underlined) for the bicistronic vectors pCAT/A15-705/Luc, pCAT/A15-657/Luc, and pCAT/A15-596/Luc, respectively. In addition, a GBV-A sequence amplified with 5′-TATAATAAGCTTGCCGCGAGTTGAAGAGCAC and 5′-CCATAATCATGAGCCCCGGACCTCCCACCGAG was used to construct pCAT/A657-15/Luc, which contains GBV-A sequences in the antisense orientation.

GBV-C sequences were amplified from a plasmid generated during the cloning of GBV-C 5′ sequences (19). The sequence of this GBV-C cDNA (nt 1 to 631 [GenBank accession no. U62537]) corresponds to nt 30 to 659 of GenBank accession no. U44402, the longest GBV-C isolate reported to date. PCRs for the GBV-C monocistronic and bicistronic plasmids utilized the sense primer 5′-TATAATAAGCTTCACTGGGTGCAAGCCCCA (*Hin*dIII site underlined) coupled with 5′-GCTCTAGAGGCGCAACAGTTTGTGAGGAA, 5′-GCTCTAGAACAAGCGTGGGTGGCCGGGG, 5′-GCTCTAGAGACCACGAGAAGGAGCAGAAG, 5′-GCTCTAGACATGATGGTATAGAAAAGAG (*Xba*I site underlined) for the monocistronic vectors pC1-631/CAT, pC1-592/CAT, pC1-553/CAT, and pC1-526/CAT, respectively; and 5′-CATGCCATGGCGCAACAGTTTGTGAGGAA, 5′-GTATTGCGCCATGGCTCGACAAGCGTGGGTGGCCGGGG, and 5′-GGACTGCCATGGTGGTATAGAAAAGAG (*Nco*I sites underlined) were used for the bicistronic vectors pCAT/C1-629/Luc, pCAT/C1-596/Luc, and pCAT/C1-526/Luc, respectively. Additional GBV-C sequences were amplified with 5′-GCTCTAGACACTGGGTGCAAGCCCCA (*Xba*I site underlined) and 5′-TATAATAAGCTTGGCGCAACAGTTTGTGAG (*Hin*dIII site underlined) for the monocistronic pC631-1/CAT plasmid, and 5′-TATAATAAGCTTCTCGACAAGCGTGGGTGGCCGGGG 3′ (*Hin*dIII site underlined) and 5′-GTATTGCGCCATGGCACTGGGTGCAAGCCCCAGAA (*Nco*I site underlined) were used for the bicistronic pCAT/C596-1/Luc plasmid. Both of these plasmids contain GBV-C sequences in the antisense orientation.

HCV sequences were amplified from a plasmid clone of a genotype 1a isolate with the sense primer 5′-TATAATAAGCTTCACTCCCCTGTGAGGAACTAC (*Hin*dIII site underlined) coupled with 5′-GTATTGCGTCATGATGGTTTTTCTTTGGGGTTTAG or 5′-CCATAATCATGATGCACGGTCTACGAGACCT (*Bsp*HI sites underlined) to generate the bicistronic vectors pCAT/HCV39-377/Luc and pCAT/HCV39-345/Luc, respectively.

Site-specific nucleotide changes were generated in pA15-707/CAT and pC1-631/CAT with the MORPH site-specific plasmid DNA mutagenesis kit (5 Prime→3 Prime, Inc., Boulder, Colo.) as directed by the manufacturer. Nucleotide changes were confirmed by double-stranded DNA sequencing as described above.

IVTT. In vitro transcription translation (IVTT) reactions were performed with the TnT T7 coupled reticulocyte lysate system (Promega) according to the manufacturer's instructions. Reactions (25 μl) contained 20 U of rRNasin (Pro-

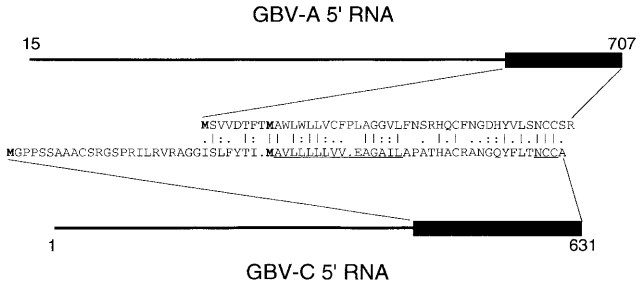


FIG. 1. Schematic of GBV-A and GBV-C 5′ sequences and amino acid alignment of their respective ORFs. The putative E1 signal sequence in GBV-C and the Asn-Cys-Cys motif (12a) are underlined.

mega), 20 μCi of [^{35}S]cysteine (1,000 Ci/mmol; Amersham), and 0.5 μg of plasmid template. After incubation at 30°C for 60 min, 5-μl aliquots were denatured (5 min, 99°C) in an equal volume of 2× sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer (125 mM Tris [pH 6.8], 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.2 mg of bromophenol blue per ml) and electrophoretically separated on 10 to 20% SDS–polyacrylamide gels (Bio-Rad). The gels were fixed in 10% methanol–20% acetic acid, dried, and analyzed with a PhosphorImager SI with ImageQuaNT software (Molecular Dynamics, Inc.). Image exposure time, white-black range, and product quantitations are presented in the corresponding figure legend.

Reporter gene enzymatic assays. Luc assays were performed by mixing 50 μl of 1× Luc assay reagent (Promega) with 1 μl of a 10-fold dilution of a rabbit reticulocyte lysate reaction. Activity was assayed immediately by a 5-s count in a Clinilumat LB9502 luminometer (Berthold Systems, Inc., Pittsburgh, Pa.). CAT assays were completed with a commercially available kit (Promega) according to the manufacturer's instructions. Briefly, 5 μl of lysate was incubated with [^{3}H]chloramphenicol and *n*-butyryl coenzyme A in a 125-μl reaction mixture for 1 h at 37°C. Butyrylated [^{3}H]chloramphenicol products were isolated by xylene extraction and quantitated by liquid scintillation counting.

Secondary RNA structure. A model of the secondary structure of the 5′ nontranslated RNA of the GBV-C genome was constructed by a combination of phylogenetic and thermodynamic approaches. A first-level phylogenetic analysis considered nucleotide sequences representing the 5′ RNA of GBV-C strains present in 35 different patient serum samples (19). These were aligned with the program PILEUP (Wisconsin Sequence Analysis Package, version 8, September 1994; Genetics Computer Group, Madison, Wis.) and subjected to a manual search for covariant nucleotide substitutions indicative of conserved helical structures. In addition to canonical Watson-Crick base pairs, G-U base pairs were considered acceptable for this analysis. Conserved helical structures identified by the presence of one or more covariant nucleotide substitutions were forced to base pair in the subsequent computer-based folding of the prototype GBV-C sequence (GenBank accession no. U36380), which was done with the program MFOLD. Separate MFOLD analyses were carried out with sequences representing nt 1 to 611, 43 to 522 (both closed at nt 273 to 418), 273 to 418, and 43 to 180. MFOLD predicts a series of alternative structures with different predicted folding energies. These were reviewed to determine which predicted structures were most permissive for covariant and noncovariant nucleotide substitutions present in the other GBV-C sequences. Where no predicted structure could accommodate most nucleotide substitutions, the sequence was left single stranded in the final model. A second-level phylogenetic analysis involved the alignment of GBV-C sequences with the 5′ RNA sequences of five separate GBV-A-like viruses (12, 27, 31), followed by a manual search for covariant substitutions indicative of similar structures in the 5′ sequences of these related viruses.

RESULTS

Translation of monocistronic transcripts containing 5′ GBV RNA. A common Asn-Cys-Cys motif homologous to the HCV E1 Asn-Ser-Cys motif is found near the N termini of the putative E1 proteins of GBV-A, GBV-B, and GBV-C (reference 12a and Fig. 1). Located near the N termini of the GBV-A and GBV-C large ORFs, this tripeptide sequence appears to be the 5′-most conserved motif between HCV and the GBVs. Because it is within the coding regions of GBV-B and HCV and is in frame with the long ORF, this sequence is likely to be translated in GBV-A and GBV-C as well. To determine whether the 5′ ends of GBV-A and -C could direct translation, nt 15 to 707 of GBV-A and nt 1 to 631 of GBV-C were cloned into plasmid vectors to create pA15-707/CAT and pC1-631/CAT, respectively. These vectors contain a T7 promoter driving transcription of the 5′ GBV sequences, which are ligated in frame (relative to the Asn-Cys-Cys motif) with the bacterial CAT gene (Fig. 2A).

IVTT reaction mixtures containing rabbit reticulocyte lysates were programmed with pA15-707/CAT, pC1-631/CAT, and a positive control plasmid, pHAV-CAT1, which contains the 5′ NTR of HAV inserted upstream of CAT. All three plasmid DNAs directed the translation of discreet products as detected by SDS-PAGE (Fig. 2B). The products derived from pA15-707/CAT and pC1-631/CAT were slightly larger than that derived from pHAV-CAT1, indicating that translation was initiating upstream of the site of GBV-CAT fusion. In contrast, no product was detected in IVTT reactions programmed with pC631-1/CAT, which contains the GBV-C sequences inserted in the antisense orientation relative to CAT. Only the pHAV-CAT1-programmed reaction possessed detectable CAT activity (data not shown). The absence of activity in the products of reactions programmed with pA15-707/CAT and pC1-631/CAT is likely to be due to the misfolding of the CAT protein as a result of its fusion with the N-terminal segment of the GBV polyprotein.

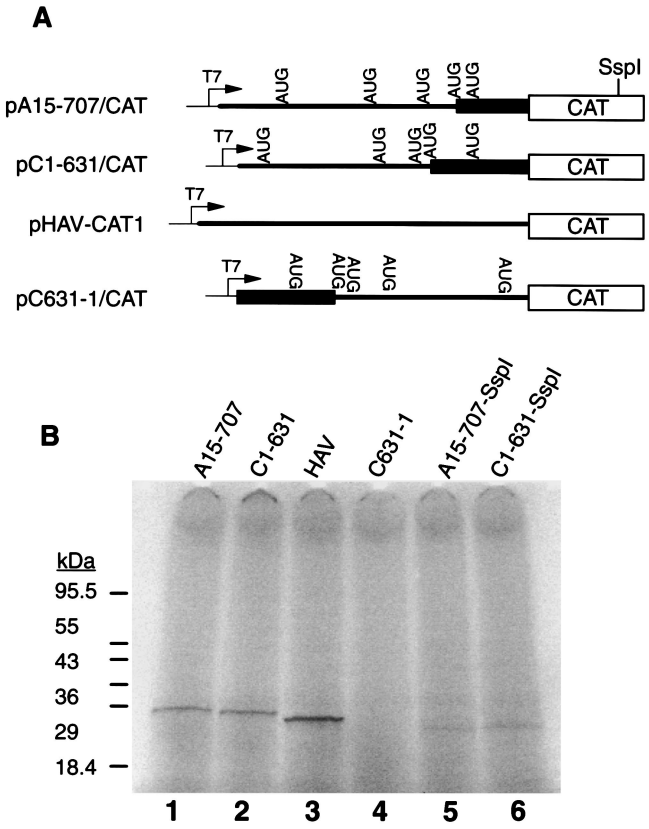


FIG. 2. (A) Schematic representation of monocistronic T7 templates. Viral RNA sequence is represented as a bold line. The positions of AUG codons (AUG) and the long ORFs (thick bars) are indicated for the GBV sequence. For GBV-C, only AUGs conserved in all isolates examined (19) are depicted. (B) PhosphorImager scan of products generated in IVTT reactions programmed with pA15-707/CAT (A15-707 [lane 1]), pC1-631/CAT (C1-631 [lane 2]), pHAV-CAT1 (HAV [lane 3]), pC631-1/CAT (C631-1 [lane 4]), *Ssp*I-linearized pA15-707/CAT (A15-707-SspI [lane 5]), and pC1-631/CAT (C1-631-SspI [lane 6]). The image was generated from a 16-h exposure with a linear range of 7 to 200. GBV-CAT products in lanes 1 and 2 are present at 26 to 27% of the level of the CAT product made from pHAV-CAT1 (lane 3) when the number of Cys residues have been normalized for each product.

To confirm that the products of the reactions programmed with pA15-707/CAT and pC1-631/CAT were in fact GBV-CAT fusion proteins, the pA15-707/CAT and pC1-631/CAT plasmids were digested with *Ssp*I prior to being used to program reactions. *Ssp*I linearizes these plasmids within the CAT coding region so that runoff transcripts produced from these plasmids lack sequences encoding the C-terminal 45 amino acids of CAT. As expected, reactions programmed with the *Ssp*I-digested pA15-707/CAT and pC1-631/CAT DNAs (Fig. 2B, lanes 5 and 6, respectively) contained products that were approximately 5 kDa smaller than those found in reactions programmed with undigested pA15-707/CAT and pC1-631/CAT plasmids (lanes 1 and 2, respectively).

Site of translation initiation in GBV-A and GBV-C. The apparent molecular masses of the GBV-CAT fusion proteins shown in Fig. 2B suggested possible sites of translation initiation. As indicated in Fig. 1, the GBV-A and GBV-C ORFs that were ligated to CAT in pA15-707/CAT and pC1-631/CAT each contain two in-frame AUG codons that might serve as potential sites of translation initiation within the sequence immediately upstream of CAT. These are the fourth and fifth AUG codons in each of the GBV-A and GBV-C sequences (Fig. 2A). If initiation occurs at the fourth AUG, the resultant fusion proteins would contain 46 amino acids of GBV-A (adding 5.1 kDa to the 24 kDa of CAT) or 67 amino acids of GBV-C (adding 7.5 kDa to CAT), respectively. In contrast, initiation at the fifth AUG in these transcripts would product CAT fusion proteins containing 38 and 36 amino acids of GBV-A- and GBV-C-encoded protein, respectively, adding 4.1 kDa to CAT. The apparent molecular mass of the ~28-kDa fusion proteins detected in the reactions programmed with pA15-707/CAT and pC1-631/CAT suggested that translation initiates at the fifth AUG in each transcript (i.e., the second in-frame Met codons in the long ORF, which are located at nt 594 of the GBV-A sequence and nt 524 of the GBV-C sequence). To identify the sites of translation initiation, the first and second in-frame AUG codons in GBV-A and GBV-C were changed to UAG stop codons producing pAmut1/CAT, pAmut2/CAT, pCmut1/CAT, and pCmut2/CAT (Fig. 3A). These plasmids were used to program IVTT reactions.

GBV-CAT fusion proteins were detected in reactions programmed with pAmut1/CAT (94% of pA15-707/CAT) and pCmut1/CAT (42% of pC1-631/CAT) (Fig. 3B, lanes 2 and 5, respectively). The lower level of protein expression by these mutants may reflect disruptions in RNA structure important for translation efficiency (described below). In contrast, reactions programmed with pAmut2/CAT and pCmut2/CAT (Fig. 3B, lanes 3 and 6, respectively) did not produce detectable quantities of fusion protein. Thus, because the 28-kDa GBV-CAT protein was detected when the first in-frame AUG codon (nt 570 in GBV-A and nt 431 in GBV-C) was replaced with a stop codon, initiation does not occur at this position. However, mutation of the second in-frame AUG codon (nt 594 in GBV-A and nt 524 in GBV-C) completely abrogated protein production directed by these constructs, consistent with the second in-frame AUG being the site of translation initiation in both GBV-A and GBV-C. In a related experiment, IVTT reactions programmed with a plasmid containing GBV-C sequence with an AUG-to-ACG change at the position of the second in-frame AUG (nt 524) produced protein of an identical size to pC1-631/CAT, although at a diminished level (data not shown). Because initiation has been found to occur with lower efficiency at ACG codons in other mRNAs (1), these data are consistent with translation of the GBV-C–CAT fusion protein initiating at the ACG codon.

The number and position of Leu residues immediately

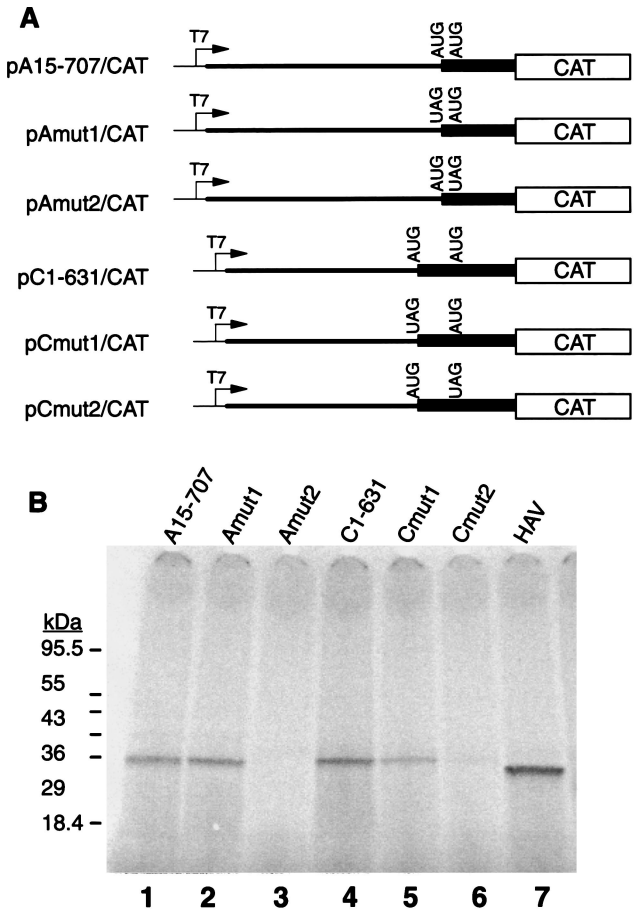


FIG. 3. (A) Organization of site-specific mutants of GBV-CAT monocistronic templates. (B) PhosphorImager scan of IVTT products generated from GBV-CAT mutant templates. Lanes 1, 4, and 7 are control reactions programmed with pA15-707/CAT, pC1-631/CAT, and pHAV-CAT1, respectively. Products generated from reactions programmed with the mutant templates are found in lanes 2 (pAmut1/CAT), 3 (pAmut2/CAT), 5 (pCmut1/CAT), and 6 (pCmut2/CAT). Image characteristics are identical to those in Fig. 2. The GBV-CAT proteins in lanes 1 and 4 are present at 35 to 41% of the level of CAT produced from pHAV-CAT1 template (lane 7).

downstream of the initiator Met in both GBV-A and GBV-C provided a biochemical method to confirm the position of the initiation site in the GBV-CAT fusion proteins. IVTT reactions containing [^{3}H]Leu were programmed with pA15-707/CAT and pC1-631/CAT. Reaction products were separated by SDS-PAGE and transferred onto a solid support, and the 28-kDa protein bands were excised. The N-terminal amino acids of the resultant GBV-CAT fusion proteins were sequentially removed by Edman degradation, and each fraction was analyzed by scintillation counting (Fig. 4). The [^{3}H]Leu profile obtained from the pA15-707/CAT product was consistent with the expected sequence of GBV-A downstream of the second in-frame AUG (Fig. 4A), assuming that the N-terminal Met residue is removed (29). Some trailing of the ^{3}H signal was noted, which may be attributed to incomplete removal of the N-terminal Met. However, for the pC1-631/CAT product, the [^{3}H]Leu profile exactly matched the expected amino acid sequence downstream of the second in-frame AUG for GBV-C (Fig. 4B). These experiments confirm that translation is initiated at nt 594 of the GBV-A sequence and nt 524 of the GBV-C sequence. The relative length of the 5′ nontranslated RNA segments and the multiple AUG codons (some of which

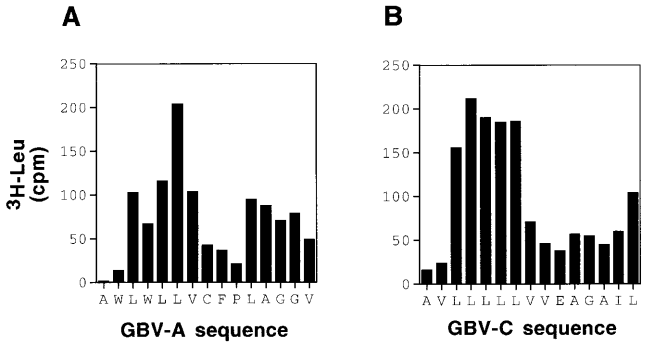


FIG. 4. Edman degradation of [^{3}H]Leu-labeled GBV-CAT fusion products. IVTT reactions programmed with pA15-707/CAT (A) and pC1-631/CAT (B) were performed in the presence of [^{3}H]Leu, separated on SDS-polyacrylamide gels, and blotted to polyvinylidene difluoride membranes. Labeled products were excised and subjected to Edman degradation. The number of counts per minute after each degradation cycle is plotted above the predicted N-terminal sequences (minus the initiator Met) of GBV-A and GBV-C.

are in good context for translation initiation) upstream of the authentic initiator AUG in these transcripts both suggest that translation is initiated on these RNAs by internal ribosomal entry, rather than by a conventional 5′ scanning mechanism. Thus, it is likely that the GBV-A and GBV-C 5′ sequences contain an IRES.

GBV coding sequence is required for efficient translation of monocistronic RNAs. The results of the in vitro translation reactions described above demonstrate that initiation begins at the Met residue positioned immediately upstream of the putative E1 signal sequence in both pA15-707/CAT and pC1-631/CAT. To determine the 3′ limits of the apparent IRES in GBV-A and GBV-C and whether any amount of GBV sequence downstream of the initiator AUG is necessary for protein production in the IVTT assays, several 3′ deletions were made which reduced the amount of GBV sequence in the GBV-CAT fusion proteins (Fig. 5A). Protein production was observed in reactions programmed with the deletion constructs pA15-665/CAT and pC1-592/CAT, which contain 72 and 69 nt of the GBV-A and GBV-C coding sequence fused to CAT, respectively (Fig. 5B, lanes 2 and 6). In contrast, no protein was detected in reactions programmed with the deletion constructs pA15-596/CAT, pC1-526/CAT, pA15-629/CAT, or pC1-553/CAT, which contain three (pA15-596/CAT and pC1-526/CAT), 36 (pA15-629/CAT), or 30 (pC1-553/CAT) nt of the GBV coding sequence ligated in frame with CAT. These results demonstrate, rather surprisingly, that sequences downstream of the predicted initiator AUG are necessary for efficient translation initiation in vitro. Given that the authentic initiator codons are in good context in both GBV-A and GBV-C, these data provide further evidence that translation is not initiated by a conventional 5′ scanning mechanism.

The quantity of CAT produced from the control plasmid, pHAV-CAT1 (Fig. 5B, lane 9), was considerably greater than that produced from either the GBV-A or GBV-C monocistronic constructs. This is of interest, because the HAV IRES directs the internal initiation of translation with very low levels of efficiency relative to those of other picornaviral IRES elements (34). The low level of production of GBV-CAT proteins is not likely to be due to differences in T7 transcriptional efficiency in these IVTT assays, because similar results were obtained with reactions programmed with equal amounts of RNA (data not shown). Thus, it appears that the level of GBV-CAT protein reflects the extremely low level of efficiency with which the GBV IRESs direct internal initiation in vitro.

Translation of bicistronic GBV RNAs. In an effort to formally demonstrate that the 5′ RNA sequences of GBV-A and GBV-C contain IRESs, these sequences were inserted between CAT and Luc genes to create bicistronic T7 transcriptional units (Fig. 6A). IVTT reactions programmed with the bicistronic constructs produced equivalent amounts of CAT activity and CAT protein (Fig. 6B). This confirmed that essentially equivalent amounts of RNA were being transcribed in each reaction. In contrast, the level of Luc activity and amount of Luc protein produced were dependent on the sequence cloned into the intercistronic space upstream of Luc. Although much lower than the level of Luc produced from two positive control plasmids containing the IRES of HCV in the intercistronic space (270,000 to 546,000 light units [Fig. 6B]), detectable levels of Luc activity were produced only in reactions programmed with GBV bicistronic constructs containing GBV-A and GBV-C sequences in the sense orientation (10,300 to 13,300 light units [Fig. 6B]). Although the quantities of Luc produced were barely detectable by SDS-PAGE, PhosphorImager analysis of these gels indicated that Luc enzymatic activity did not correlate with the protein detected in the IVTT assays (Fig. 6B, Luc-A versus Luc-P). This is most likely due

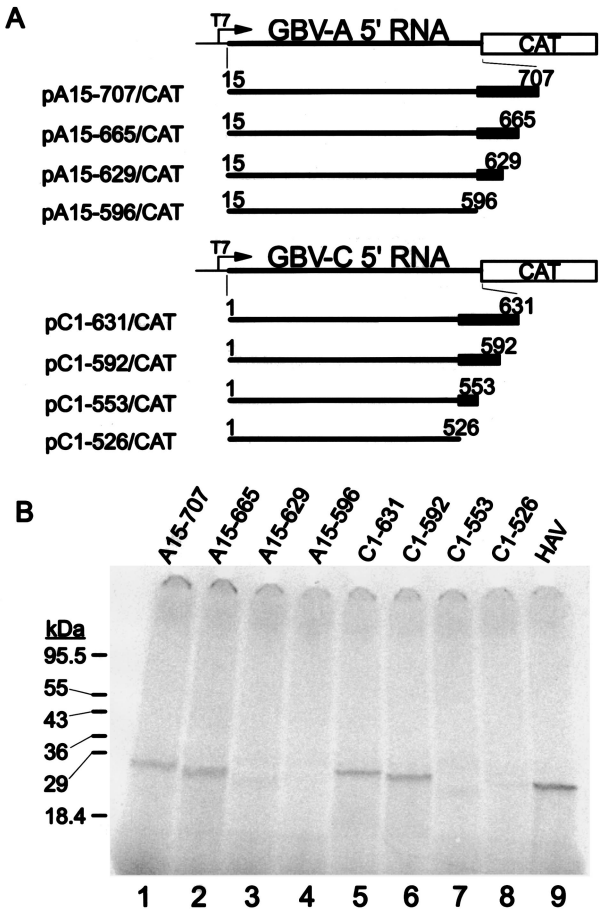
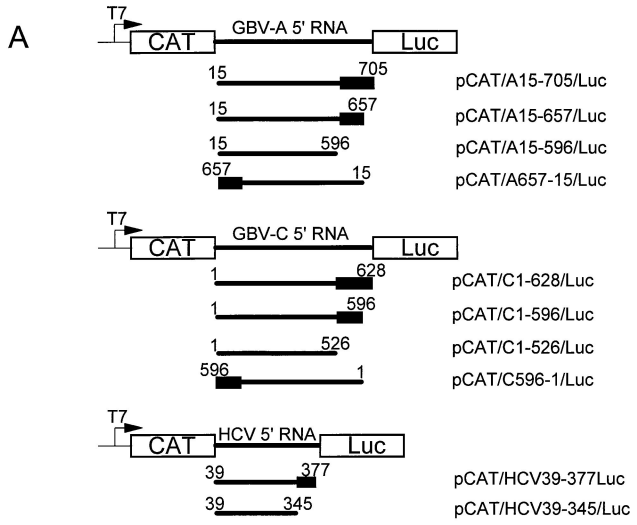


FIG. 5. Translation of monocistronic RNAs containing 3′ GBV deletions. (A) Schematic of the monocistronic templates. (B) PhosphorImager scan of IVTT products generated with pA15-665/CAT (lane 2), pA15-629/CAT (lane 3), pA15-596/CAT (lane 4), pC1-592/CAT (lane 6), pC1-553/CAT (lane 7), and pC1-526/CAT (lane 8). Control reactions are shown in lanes 1 (pA15-707/CAT), 5 (pC1-631/CAT), and 9 (pHAV-CAT1). Image characteristics are identical to those in Fig. 2. GBV-CAT protein (lanes 1, 2, 5, and 6) is present at 20 to 36% of the level of CAT produced from the pHAV-CAT1 template (lane 9).



B

pCAT/--/Luc	Luc-A	Luc-P
A15-705	10.2	25.2
A15-657	1.13	14.8
A15-596	10.3	13.8
A657-15	0.13	6.37
C1-629	4.09	15.4
C1-596	0.80	12.4
C1-526	13.5	14.7
C596-1	2.02	5.75
HCV39-377	546	403
HCV39-345	270	205

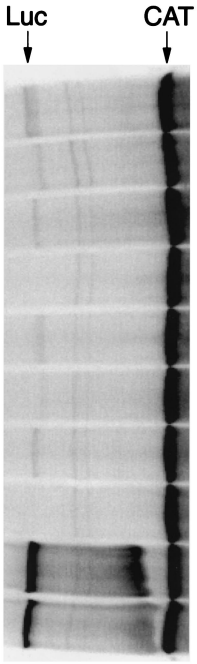


FIG. 6. Translation of bicistronic GBV and HCV vectors. (A) Schematic of the bicistronic T7 templates. (B) Luc activity (Luc-A, light units [10^3]), Luc protein (Luc-P, band volume [10^3]) and protein production of IVTTs programmed with the bicistronic vectors. CAT activities were equivalent in the reactions shown (157,000 ± 3,550 cpm). The PhosphorImager scan was generated from a 72-h exposure with a linear range of 25 to 600. Band volumes are reported without background subtraction.

to altered activity as a result of the GBV fusion. Of greater importance, however, is the fact that no detectable protein and only minimal Luc activities (130 and 2,020 light units) were produced in reactions programmed with bicistronic constructs containing GBV-A and GBV-C sequences in the antisense orientation. These results suggest that these viruses utilize internal ribosome entry for initiation of translation, but the extraordinarily low activities of the putative GBV IRES elements when placed in a bicistronic context raise a number of issues (see Discussion).

Secondary structure of the 5′ NTR of GBV-A and GBV-C. The results presented above suggest that translation of the GBV-A and GBV-C polyproteins is initiated by an unusual mechanism of internal ribosomal entry, which is likely to be controlled by RNA structures within the 5′ nontranslated RNA, and which is also dependent upon sequence downstream of the initiator AUG (Fig. 5). Thus, we attempted to characterize the secondary structure near the 5′ ends of GBV-A and GBV-C RNA by using a combination of phylogenetic analysis and thermodynamic predictions. Covariant nucleotide substitutions indicative of conserved base pair interactions were identified by manual searches of an alignment of 42 different GBV-C sequences or an alignment of GBV-A sequence with 5 different GBV-A-like sequences (12). These were used to constrain the folding of the RNA by the computer program MFOLD. Alternative structures were reviewed to determine which were the most permissive for observed variations in the nucleotide sequence, resulting in the models for secondary structure shown in Fig. 7. The predicted secondary structures of the 5′ NTR of GBV-C and GBV-A are very similar. However, these structures are quite different from that of HCV (2, 6), suggesting distinctly different evolutionary histories.

The models suggest that the 5′ RNA of GBV-C and GBV-A contains four major secondary structure domains upstream of the authentic initiator AUG at nt 594 and 524, respectively, which are conserved in all GBV-C and GBV-A sequences (domains II to V in Fig. 7). Domain I is missing from the 5′ end of the GBV-C sequence, suggesting that it may not include the authentic 5′ end of the genome. In addition, stem-loop IVc is present only in GBV-A, and stem-loop IVa′ is present only in GBV-C. Other structural features are highly conserved between these sequences, which share only ~50% nucleotide similarity in conventional alignments. Domain II consists of an extended stem-loop structure, which is highly conserved in nucleotide sequence between nt 68 and 152 in GBV-C but which contains several covariant nucleotide substitutions within the flanking RNA segments near its base (Fig. 7A, boxed base pairs). The predicted structure of the conserved sequence between nt 68 and 152 is confirmed by the presence of covariant nucleotide substitutions in alignments of GBV-C with GBV-A. Domain III contains two small stem-loops (IIIa and IIIb), both of which are supported by the presence of covariant substitutions in different GBV-C strains. Possible base pair interactions forming a classical RNA pseudoknot are noted in stem-loop IIIa. The existence of this putative tertiary structural feature is supported by a single covariant substitution in the sequence of additional GBV-A-like viruses recovered from tamarins (circled bases). The putative pseudoknot is present in both the GBV-C and GBV-A sequences, which are highly conserved in this region, although the loop spanning the deep groove of the RNA helix contains 3 nt in GBV-C but only 1 nt in GBV-A. The larger, complex stem-loops which comprise domains IV and V of the model structures are also well supported by covariant substitutions among different GBV-C isolates (Fig. 7A). Of particular interest, given the requirement for the inclusion of coding sequence for efficient translation of monocistronic GBV transcripts (Fig. 5), is evidence suggesting the existence of a very stable, conserved stem-loop containing 9 to 10 G-C bp within the ORF of GBV-C, downstream of the putative 5′ NTR (described below) (Fig. 7A). The existence of this stable helical structure is supported by the presence of a single covariant substitution among different GBV-C isolates. This stem-loop appears to be an extension of a larger, well-conserved structure (domain VI, Fig. 7) located 20 nt downstream of the putative initiator AUG. Importantly, a very similar structure is present near the 5′ end of the ORF of GBV-A (Fig. 7B).

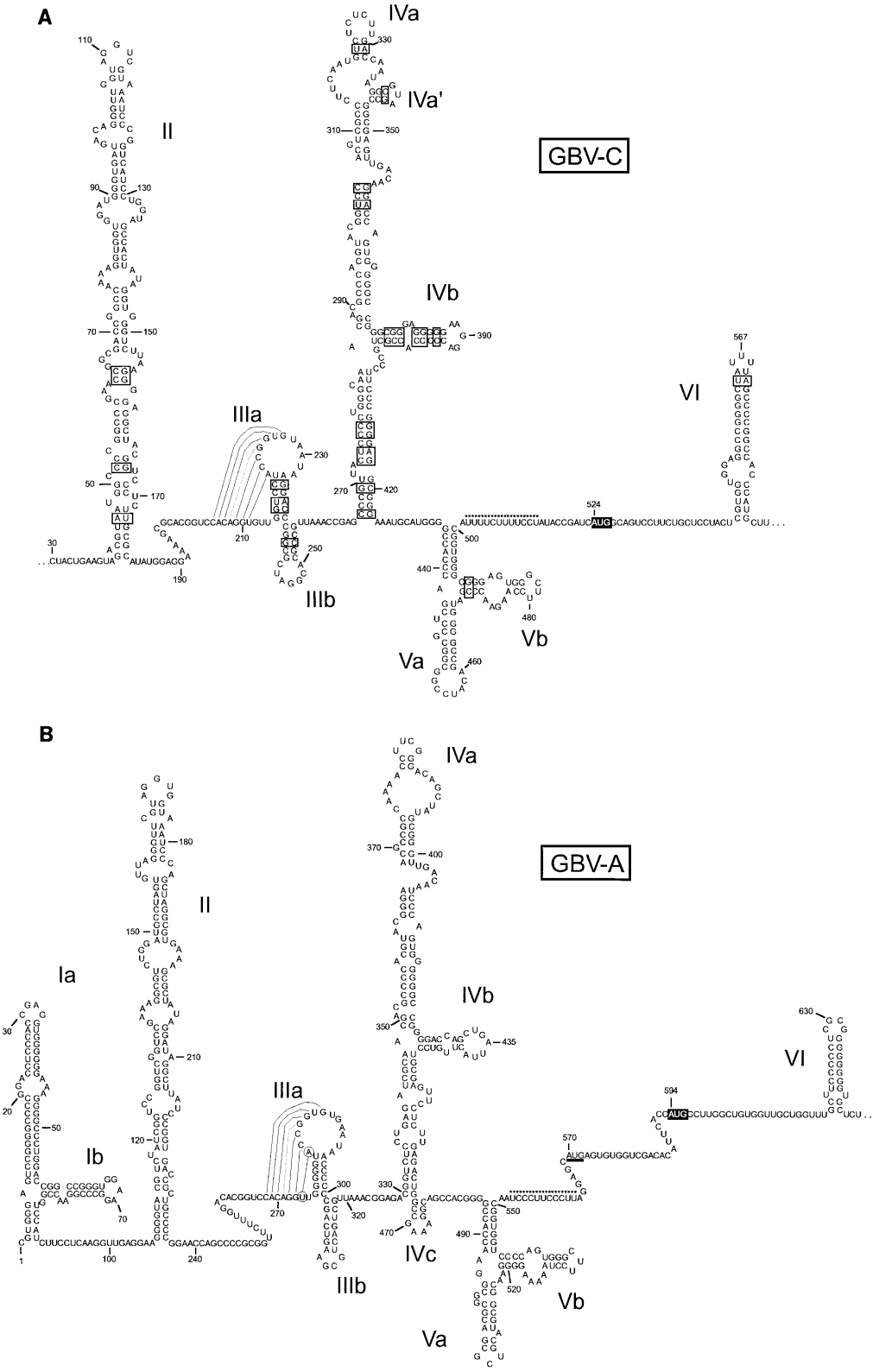


FIG. 7. Preliminary models of the secondary RNA structures which are present near the 5′ ends of GBV-C (nt 30 to 589; GenBank accession no. U36380) (A) and GBV-A (nt 1 to 645; GenBank accession no. U22303) (B). The model structures resulted from a combination of phylogenetic analysis and computational thermodynamic predictions (MFOLD). Major putative structural domains are labeled with roman numerals I to VI from the 5′ end, with individual stem-loops labeled alphabetically within each domain. The initiator AUG codons of GBV-C and GBV-A, as determined in the preceding experiments, are highlighted in black boxes (at nt 524 in GBV-C and at nt 594 in GBV-A). In both structures, a row of asterisks indicates the upstream oligopyrimidine tract, which is reminiscent of the box A motif present near the 3′ end of picornaviral IRESs (20). In panel A, boxed sequences indicate base pairs which are the sites of sequence covariation in the 41 different strains of GBV-C which were included in the phylogenetic analysis. With minor variation, the structure shown can be assumed by all available GBV-C sequences. In panel B, the in-frame AUG triplet at nt 570, which is not utilized for initiation of translation, is indicated by a double underline.

DISCUSSION

We found that monocistronic mRNAs containing the 5′ ends of the GBV-A and GBV-C genomic RNAs fused to CAT directed the production of GBV-CAT fusion proteins in IVTT reactions. Site-specific mutagenesis and Edman degradation of the translation products indicated that translation of these transcripts, and presumably GBV-A and GBV-C genomic RNAs as well, initiates immediately upstream of the putative E1 envelope signal sequence, at the AUG located at nt 594 in GBV-A and at nt 524 in the GBV-C sequence. The site of initiation identified in GBV-C is corroborated by analysis of 42 distinct 5′ RNA sequences obtained from 35 GBV-C-positive individuals (19). When these sequences are aligned, the only conserved AUG codon which is in-frame with the GBV-C polyprotein is the AUG at nt 524. Downstream of this AUG codon, nucleotide substitutions in the different GBV-C strains generally result in either silent or conservative amino acid changes. In contrast, upstream of this AUG codon, nucleotide substitutions, deletions, and insertions drastically change the encoded amino acid sequence in different strains. These data suggest that there is a selective pressure acting downstream of the AUG at nt 524 to maintain a protein coding sequence, while no selective pressure exists to maintain such a sequence upstream of this codon.

The fact that translation initiates at the fifth AUG codon in both viral RNAs, many hundreds of nucleotides from the 5′ end, is strongly reminiscent of translation in the picornaviruses and HCV and suggests that translation may be initiated by binding of the 40S ribosomal subunit at an internal site on the RNA. Thus, it seems likely that the 5′ NTRs of these viruses may contain an IRES. Because the functional activities of the IRES elements of HCV and the picornaviruses are known to be highly dependent on RNA secondary structure within the 5′ NTR, we sought evidence for conserved secondary RNA structures within the 5′ NTRs of these viruses. Although the 5′ nucleotide sequences of the GBV-C and GBV-A virus genomes have only ~50% nucleotide identity within the 500 nt preceding the initiator AUG of GBV-C, we found the secondary structures of these RNAs to be remarkably similar. Each of the major secondary structural domains shown for GBV-C is conserved in the structure of GBV-A with only minimal changes (Fig. 7). However, both the GBV-A and GBV-C 5′ NTR structures are very different from those of the pestiviruses, HCV, and GBV-B, despite the fact that these viruses share a common genome organization as well as multiple sequence motifs within their nonstructural proteins (12a, 18). While the 5′ NTRs of GBV-B, HCV, and the pestiviruses are particularly closely related to each other at the structural level (2, 6), the prominent domain III pseudoknot and complex stem-loop III structures of these viruses are completely lacking in GBV-C and GBV-A. In addition, there is no clear-cut structural relatedness to HCV or the pestiviruses in any of the upstream secondary structures of GBV-A and GBV-C. Thus, similar to the existence of two distinct types of 5′ NTR structures among the picornaviruses (one in the cardioviruses, aphthoviruses, and hepatoviruses and another in the enteroviruses and rhinoviruses [8]), there are two distinct types of 5′ NTR structures present in the flaviviruses. This has interesting implications for the evolution of these agents.

A prominent feature of the 5′ NTR sequences of GBV-C and GBV-A is the presence of a short oligopyrimidine tract located just upstream of the initiator AUG. While this tract is somewhat variable in sequence, it is present in all of the GBV-C sequences and is positioned approximately 21 nt upstream of the initiator AUG. Thus, this region of the 5′ NTR bears remarkable similarity to the "box A/box B" motif identified at the 3′ end of picornaviral 5′ NTRs by Pilipenko et al. (20), including the distance (20 to 25 nt) between the start of the pyrimidine tract and the initiator AUG, which Pilipenko et al. found to be critical to poliovirus IRES-directed translation. It is interesting that the segment intervening between the oligopyrimidine tract and the first downstream AUG is somewhat shorter in GBV-A (approximately 17 nt). By analogy with the picornaviruses (20), this might be expected to result in a preference for initiation of translation at the second in-frame AUG codon in GBV-A (nt +25 with respect to the first AUG). We confirmed this experimentally (Fig. 4A). The striking differences between the 5′ NTR structures of these viruses and that of HCV, coupled with these similarities between the translations of GBV-A and GBV-C and picornaviral 5′ NTRs, suggest that the mechanism of translation might be closer to that of picornaviruses than HCV. In HCV, relatively strong evidence supports the concept that the 40S ribosomal subunit binds RNA directly at the site of translation initiation (6, 7). In contrast, the 40S subunit appears to scan for a variable distance from an upstream primary binding site to the initiator AUG in some picornaviruses (8). Given the variable distances between the authentic initiator codons and the upstream oligopyrimidine tracts in GBV-A and GBV-C, this appears likely to be the case with GBV-A (and possibly also GBV-C).

Both GBV-A and GBV-C contain a very stable stem-loop structure within the translated ORF (domain VI, Fig. 7). This conserved structure is located about 20 nt downstream of the initiator AUG in GBV-C, although it is possible that additional, less-well-conserved base pair interactions may bring the base of this structure closer to the AUG. It is tempting to speculate that this stem-loop may function to enhance initiation by a scanning 40S ribosomal subunit, much as Kozak (11) has shown that stable stem-loops placed downstream of an AUG can result in a "pausing" of the ribosome over the AUG, enhancing the likelihood of initiation at that codon. This phenomenon may explain why the efficient translation of reporter proteins fused to the 5′ NTR requires inclusion of the most 5′ sequence of the GBV-C ORF. If so, this would provide a novel mechanism by which the sequence within the ORF can contribute to regulation of translation in flaviviruses. Both HCV and GBV-B differ from GBV-A and GBV-C in that their initiator AUG is located within the loop segment of a stem-loop which straddles the 5′ end of the ORF (6). Initiation of translation of these viral RNAs is thus dependent upon melting of this stem-loop, while, in the case of GBV-A and GBV-C, initiation of translation is likely to be dependent on maintenance of the integrity of the domain VI stem-loop.

The domain VI stem-loop, which is required for efficient translation of the monocistronic transcripts, does not appear to be required for efficient translation in the bicistronic transcripts (compare Fig. 5 and 6). This apparent discrepancy may be a result of the different reporter genes being utilized in these transcripts. Similar findings have been reported for HCV. Specifically, Reynolds et al. (22), using bicistronic vectors with the IRES-dependent reporter genes secreted alkaline phosphatase or a truncated influenza virus nonstructural protein, show that efficient translation directed by the 5′ end of HCV requires the inclusion of coding sequences. In contrast, Wang et al. (33), using monocistronic and bicistronic vectors with Luc as the IRES-dependent reporter gene, found that the inclusion of HCV coding sequences is not necessary for efficient translation. Addressing these conflicting results, Reynolds et al. hypothesize that the 5′ end of the luciferase gene may complement the function provided by the HCV coding sequences. A similar argument may explain the discordance

between the results obtained with the monocistronic GBV-CAT constructs and those obtained with the bicistronic GBV-Luc constructs.

Although all of these observations suggest the strong likelihood that GBV-A and GBV-C translation is initiated by internal ribosomal entry, only minimal translation of the downstream cistron was noted from bicistronic transcripts containing the 5′ NTRs of these viruses in the intercistronic space. Translation directed by the GBV-A and GBV-C 5′ NTRs within a bicistronic context was only 2 to 5% that of the HCV IRES in rabbit reticulocyte lysates in vitro (Fig. 6). The very low levels of activity of the GBV-A and GBV-C IRESs suggest several possibilities. First, it is possible that these viruses may in fact have IRES elements with extraordinarily low levels of activity. This is supported by a very low level of translation directed by monocistronic transcripts containing the 5′ ends of GBV-A and GBV-C in the in vitro system. Specifically, after adjustment for the number of Cys residues in each construct, GBV-CAT fusion proteins were translated from pA15-707/CAT and pC1-631/CAT transcripts at only 20 to 41% of the level produced by the IRES of HAV. The HAV IRES is known to have very low level of activity, in the range of 2% of the Sabin poliovirus type 1 IRES within HAV permissive cells (28, 34). Thus, the low level of GBV IRES activity noted in vitro may be a true reflection of the strength of these translation elements. Limiting production of viral proteins within an infected host might act to reduce recognition of the infection by the immune system and thus promote viral persistence. Alternatively, it is possible that the low level of IRES activity detected in reticulocyte lysates reflects a requirement for a specific host cell translation factor which is absent in reticulocyte lysates. The nuclear autoantigen La is an example of such a specific cellular factor. It is required for efficient translation directed by the poliovirus IRES, but it is not present in sufficient amounts in reticulocyte lysates (16). It is difficult to comment more specifically on this possibility, since the cellular tropisms of GBV-A and GBV-C are unknown. A third possibility is that the low level of translational activity of the GBV-A and GBV-C 5′ NTRs may reflect a requirement for additional, yet to be identified 5′ viral sequences that may be present in these viral genomes. It is also conceivable that translation is initiated by a mechanism distinct from both the classic 5′ scanning and IRES-directed translation initiation mechanism. For example, relatively efficient translation initiation at an internal site in monocistronic transcripts but low levels of translational activity in the bicistronic context could be explained by a mechanism involving ribosome shunting (5) after recognition of the 5′ end of the RNA by the 40S ribosome subunit. Further studies will be required to distinguish between these different possibilities.

The proteins located at or near the amino termini of the polyproteins of yellow fever virus (protein C), a flavivirus, bovine viral diarrhea virus (p14), a pestivirus, and HCV (core) are small and highly basic (3, 4, 17). Because GBV-A and GBV-C are phylogenetically related to these viruses (12a, 18), it was expected that such a protein would be encoded in these viruses. However, the position of the initiation codons in GBV-A and GBV-C eliminates the possibility of a basic core protein being located at the N termini of the viral polyproteins. The possibility that the core coding sequences may have been deleted during reverse transcriptase PCR amplification or cloning of the 5′ ends of GBV-A and GBV-C is unlikely for several reasons. First, identical deletions would have had to occur consistently in each of the several clones generated during the sequencing of GBV-A and GBV-C, in addition to the separate GBV-C isolates described by Muerhoff et al. (19) and the two HGV isolates described by Linnen et al. (13). This consistency, in addition to the correspondence between PCR and infective titers for GBV-A (26, 31), argues against GBV-A and GBV-C sequences being derived from defective interfering particles in the cloning sources. Second, the deletion of core sequences would have had to occur without disturbing the translational activity of the 5′ ends of these viruses. However, because proper initiation requires sequences located in the coding regions of GBV-A and GBV-C, the coupling between the translational activity and the coding regions appears to make this an impossibility. Finally, several reverse transcriptase PCR experiments with different virus isolates, different primer combinations, and different reverse transcriptase PCR conditions and polymerases provide no evidence for additional virus sequence (data not shown).

The lack of a core-like protein at the N terminus of the viral polyprotein distinguishes GBV-A and GBV-C from all other members of the *Flaviviridae*. In fact, searches of all six potential reading frames of the three full-length GBV-C sequences (12a, 13) or the GBV-A sequence (18) present in GenBank does not reveal a conserved ORF encoding a core-like protein. Thus, these viruses appear distinct from enveloped viruses in general because they do not appear to encode a basic protein which mediates the packaging of the viral nucleic acid into the virion envelope. Core-less infectious particles have been generated artificially with the vesicular stomatitis virus glycoprotein (24). Thus, it is possible that GBV-A and GBV-C may be truly core-less enveloped viruses. However, it is possible that a cellular RNA-binding protein has been appropriated by these viruses to facilitate the specific and efficient packaging of the virion RNA into the envelope. Determination of whether GBV-A and GBV-C contain core proteins and of the source of these cores awaits the biochemical characterization of these viruses.

ACKNOWLEDGMENTS

We thank Peter Sarnow for his gift of pT7/CAT/ICS/Luc; Carla Butzen, Maria Carinato, Michelle Chalmers, and James Erker for excellent technical assistance; Russell Henry for assistance with the Edman degradation; and Thomas Leary, Tami Pilot-Matias, and A. Scott Muerhoff for critical review of the manuscript.

This work was supported by Abbott Laboratories and a grant from the U.S. Public Health Service (RO1-A132599).

REFERENCES

1. **Böck, R., and D. Kolakofsky.** 1994. Positions +5 and +6 can be major determinants of the efficiency of non-AUG initiation codons for protein synthesis. EMBO J. **13:**3608–3617.
2. **Brown, E. A., H. Zhang, L.-H. Ping, and S. M. Lemon.** 1992. Secondary structure of the 5′ nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. Nucleic Acids Res. **20:**5041–5045.
3. **Choo, Q.-L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, A. Medina-Selby, P. J. Barr, A. J. Weiner, D. W. Bradley, and M. Houghton.** 1991. Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA **88:**2451–2455.
4. **Collett, M. S., D. K. Anderson, and E. Retzel.** 1989. Comparisons of the pestivirus bovine viral diarrhea virus with members of the *Flaviviridae*. J. Gen. Virol. **69:**2637–2643.
5. **Fütterer, J., Z. Kiss-László, and T. Hohn.** 1993. Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA. Cell **73:**789–802.
6. **Honda, M., E. A. Brown, and S. M. Lemon.** Stability of a stem-loop structure involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. Submitted for publication.
7. **Honda, M., L.-H. Ping, R. C. A. Rijnbrand, E. Amphlett, B. Clarke, D. Rowlands, and S. M. Lemon.** Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. Virology, in press.
8. **Jackson, R. J., S. L. Hunt, C. L. Gibbs, and A. Kaminski.** 1994. Internal initiation of translation of picornavirus RNAs. Mol. Biol. Rep. **19:**147–159.
9. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates the translation by eukaryotic ribosomes. Cell **44:**283–292.

10. **Kozak, M.** 1989. The scanning model for translation: an update. J. Cell Biol. **108:**229–241.
11. **Kozak, M.** 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA **87:** 8301–8305.
12. **Leary, T. P., S. M. Desai, J. Yamaguchi, M. L. Chalmers, G. G. Schlauder, G. J. Dawson, and I. K. Mushahwar.** Species-specific variants of GB virus A (GBV-A) in captive monkeys. Submitted for publication.
12a. **Leary, T. P., A. S. Muerhoff, J. N. Simons, T. J. Pilot-Matias, J. C. Erker, M. L. Chalmers, G. G. Schlauder, G. J. Dawson, S. M. Desai, and I. K. Mushahwar.** 1996. The sequence and genomic organization of GBV-C: a novel member of the *Flaviviridae* associated with human non-A-E hepatitis. J. Med. Virol. **48:**60–67.
13. **Linnen, J., J. Wages, Jr., Z.-Y. Zhang-Keck, K. E. Fry, K. Z. Krawczynski, H. Alter, E. Koonin, M. Gallagher, M. Alter, S. Hadziyannis, P. Karayiannis, K. Fung, Y. Nakatsuji, J. W.-K. Shih, L. Young, M. Piatak, Jr., C. Hoover, J. Fernandez, S. Chen, J.-C. Zou, T. Morris, K. C. Hyams, S. Ismay, J. D. Lifson, G. Hess, S. K. H. Foung, H. Thomas, D. Bradley, H. Margolis, and J. P. Kim.** 1996. Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. Science **271:**505–508.
14. **Macejak, D. G., S. J. Hambidge, L. Najita, and P. Sarnow.** 1990. eIF-4F-independent translation of poliovirus RNA and cellular mRNA encoding glucose-regulated protein 78/immunoglobulin heavy-chain binding protein, p. 152–157. *In* M. A. Brinton and F. X. Heinz (ed.), New aspects of positive-strand RNA viruses. American Society for Microbiology, Washington, D.C.
15. **Meerovitch, K., and N. Sonnenberg.** 1993. Internal initiation of picornavirus RNA translation. Semin. Virol. **4:**217–227.
16. **Meerovitch, K., Y. V. Svitkin, H. S. Lee, F. Lejbkowicz, D. J. Kenan, E. K. L. Chan, V. I. Agol, J. D. Keene, and N. Sonenberg.** 1993. La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulocyte lysate. J. Virol. **67:**3798–3807.
17. **Miller, R. H., and R. H. Purcell.** 1990. Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. Proc. Natl. Acad. Sci. USA **87:**2057–2061.
18. **Muerhoff, A. S., T. P. Leary, J. N. Simons, T. J. Pilot-Matias, G. J. Dawson, J. C. Erker, M. L. Chalmers, G. G. Schlauder, S. M. Desai, and I. K. Mushahwar.** 1995. Genomic organization of GB viruses A and B: two new members of the *Flaviviridae* associated with GB agent hepatitis. J. Virol. **69:**5621–5630.
19. **Muerhoff, A. S., J. N. Simons, T. P. Leary, J. C. Erker, M. L. Chalmers, M. E. Carinato, T. J. Pilot-Matias, G. J. Dawson, S. M. Desai, and I. K. Mushahwar.** Sequence of the 5′-terminal region of the hepatitis GB virus C genome: identification of conserved sequences and evidence for genotypes. J. Hepatol., in press.
20. **Pilipenko, E. V., A. P. Gmyl, S. V. Maslova, Y. V. Svitkin, A. N. Sinyakov, and V. I. Agol.** 1992. Procaryotic-like cis elements in the cap-independent internal initiation of translation on picornavirus RNA. Cell **68:**119–131.
21. **Poole, T. L., C. Wang, R. A. Popp, L. N. D. Potgieter, A. Siddiqui, and M. S. Collett.** 1995. Pestivirus translation initiation occurs by internal ribosome entry. Virology **206:**750–754.
22. **Reynolds, J. E., A. Kaminski, H. J. Kettinen, K. Grace, B. E. Clarke, A. R. Carroll, D. J. Rowlands, and R. J. Jackson.** 1995. Unique features of internal initiation of hepatitis C virus RNA translation. EMBO J. **14:**6010–6020.
23. **Rijnbrand, R., P. Bredenbeek, T. van der Straaten, L. Whetter, G. Inchauspé, S. Lemon, and W. Spaan.** 1995. Almost the entire 5′ nontranslated region of hepatitis C virus is required for cap-independent translation. FEBS Lett. **365:**115–119.
24. **Rolls, M. M., P. Webster, N. H. Balba, and J. K. Rose.** 1994. Novel infectious particles generated by expression of the vesicular stomatitis virus glycoprotein from a self-replicating RNA. Cell **79:**497–506.
25. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
26. **Schlauder, G. G., G. J. Dawson, J. N. Simons, T. J. Pilot-Matias, R. A. Gutierrez, C. A. Heynen, M. F. Knigge, G. S. Kurpiewski, T. P. Leary, A. S. Muerhoff, S. M. Desai, and I. K. Mushahwar.** 1995. Molecular and serologic analysis in the transmission of the GB hepatitis agents. J. Med. Virol. **46:** 81–90.
27. **Schlauder, G. G., T. J. Pilot-Matias, G. S. Gabriel, J. N. Simons, A. S. Muerhoff, G. J. Dawson, and I. K. Mushahwar.** 1995. Origin of GB-hepatitis viruses. Lancet **346:**447.
28. **Schultz, D. E., M. Honda, L. E. Whetter, K. L. McKnight, and S. M. Lemon.** 1996. Mutations within the 5′ nontranslated RNA of cell culture-adapted hepatitis A virus which enhance cap-independent translation in cultured African green monkey kidney cells. J. Virol. **70:**1041–1049.
29. **Sherman, F., J. W. Stewart, and S. Tsunasawa.** 1985. Methionine or not methionine at the beginning of a protein. Bioessays **3:**27–31.
29a. **Simons, J. N.** Unpublished data.
30. **Simons, J. N., T. P. Leary, G. J. Dawson, T. J. Pilot-Matias, A. S. Muerhoff, G. G. Schlauder, S. M. Desai, and I. K. Mushahwar.** 1995. Isolation of novel virus-like sequences associated with human hepatitis. Nature Med. **1:**564–569.
31. **Simons, J. N., T. J. Pilot-Matias, T. P. Leary, G. J. Dawson, S. M. Desai, G. G. Schlauder, A. S. Muerhoff, J. C. Erker, S. L. Buijk, M. L. Chalmers, C. L. Van Sant, and I. K. Mushahwar.** 1995. Identification of two flavivirus-like genomes in the GB hepatitis agent. Proc. Natl. Acad. Sci. USA **92:**3401–3405.
32. **Tsukiyama-Kohara, K., N. Iizuka, M. Kohara, and A. Nomoto.** 1992. Internal ribosome entry site within hepatitis C virus RNA. J. Virol. **66:**1476–1483.
33. **Wang, C., P. Sarnow, and A. Siddiqui.** 1993. Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. J. Virol. **67:**3338–3344.
34. **Whetter, L. E., S. P. Day, O. Elroy-Stein, E. A. Brown, and S. M. Lemon.** 1994. Low efficiency of the 5′ nontranslated region of hepatitis A virus RNA in directing cap-independent translation in permissive monkey kidney cells. J. Virol. **68:**5253–5263.
35. **Zuckerman, A. J.** 1996. Alphabet of hepatitis viruses. Lancet **347:**558–559.