Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 1711–1715, March 1991
Biochemistry

Characterization of the terminal regions of hepatitis C viral RNA: Identification of conserved sequences in the 5′ untranslated region and poly(A) tails at the 3′ end

(5′ conservation/homology to pestiviruses/subgenomic poly(A)$^+$ RNA/evolution)

J. H. HAN*, V. SHYAMALA, K. H. RICHMAN, M. J. BRAUER, B. IRVINE, M. S. URDEA, P. TEKAMP-OLSON, G. KUO, Q.-L. CHOO, AND M. HOUGHTON

Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608

Communicated by William J. Rutter, November 5, 1990

ABSTRACT **We have determined the nucleotide sequence at the extreme 5′ and 3′ termini of the hepatitis C virus (HCV) genome. Our analyses of these sequences show (*i*) the nucleotide sequence in the 5′ untranslated region is highly conserved among HCV isolates of widely varying geographical origin, (*ii*) within this region, there are blocks of nucleotide sequence homology with pestiviruses but not with other viruses, (*iii*) the relative position of short open reading frames present in the same region of the HCV genome is similar to that of the pestiviral genome, (*iv*) RNAs truncated at the 5′ and 3′ ends are found, but the origin and functions of these RNAs are unknown, and (*v*) poly(A) tails appear to be present on 3′ subgenomic RNAs. These data differentiate HCV from the flaviviruses and indicate a closer evolutionary relationship of HCV with the pestiviruses. However, HCV also appears to be substantially different from other known pestiviruses. These data are consistent with the assignment of HCV to a separate viral genus.**

Hepatitis C virus (HCV) is a positive-strand RNA virus of ≈10 kilobases (kb) (1). Recently, this virus was cloned and its nucleotide sequence was determined from overlapping λ clones with the exception of the extreme 5′ and 3′ terminal regions (1, 2). The viral genome encodes a large polyprotein precursor of 3011 amino acids that has little overall sequence homology with other known viruses (2). However, the genetic organization of HCV resembles that of the flavi- and pestiviruses with structural proteins located in the N-terminal region and a variety of nonstructural proteins in the C-terminal region of the polyprotein (2). The amino acid sequences of some of the latter proteins show small but significant sequence similarities with the corresponding proteins encoded by pesti-, flavi-, poty-, and carmoviruses (2–5). A comparison of the hydrophobicity profiles of the polyproteins along with the relative location of the various proteins in HCV and related viruses shows that HCV is more closely related to flavi- and pestiviruses (2, 3, 5).

Typically, the noncoding region at the 5′ and 3′ termini of a viral genome contains regulatory elements important for viral gene expression and replication. We have determined and analyzed nucleotide sequences at both terminal regions of the HCV genome.† This reveals that a strongly conserved 5′ untranslated region of HCV is very similar to that of pestiviruses in terms of its size, blocks of homologous nucleotide sequence, and the organization of small open reading frames (ORFs). The HCV genome, however, displays singular characteristics at each terminus. We detect a hairpin structure at the 5′ end of the genome as well as 5′ and 3′ subgenomic RNAs, the latter of which are polyadenylated. These are consistent with a polyadenylylated 3′ terminus of the viral genome and perhaps of functional subgenomic RNAs. Our data provide insights into the organization of the HCV genome, which may have important ramifications regarding the replication strategy and evolution of the virus.

MATERIALS AND METHODS

RNA was extracted from a high-titer plasma of an experimentally infected chimpanzee (6) and plasma (or serum) from HCV-positive or negative blood donors by a low-temperature guanidinium thiocyanate method (7). Poly(A)$^+$ RNA was isolated from the liver of the same infected chimpanzee by the guanidinium thiocyanate/urea method (8). cDNA was synthesized from RNA according to Han and Rutter (9) and amplified by polymerase chain reaction (PCR) according to Saiki *et al.* (10). Briefly, RNA isolated from about 500 μl of plasma or 5 μg of poly(A)$^+$ liver RNA was converted into single-stranded cDNA by reverse transcriptase (BRL) using 150 pmol of the appropriate cDNA primer. For 5′ end characterization by primer extension (7), first-strand cDNA was precipitated by spermine (11) and tailed with dA (9). Tailed or untailed cDNA was converted into double-stranded cDNA using a second-strand cDNA primer (9). This double-stranded cDNA was amplified using the indicated HCV-specific PCR primers for 35 cycles (94°C, 1.5 min; 60°C, 2 min; 72°C, 3 min). PCR without cDNA template was routinely performed to check for possible contamination during PCR. The PCR product was analyzed by Southern blot hybridization using a ^{32}P-labeled oligonucleotide probe. The sequences and locations in the HCV genome of various cDNA and PCR primers are shown in Figs. 2 and 5. Most primers were designed to contain a *Not* I site for subsequent cloning of the PCR products into pUC18S, which contains a polylinker derived from λgt22 (9). DNA sequence was obtained by the supercoil sequencing (12) or the direct PCR sequencing method (13).

RESULTS

Based on the sequence of the major part of the HCV genome determined from overlapping λ clones (2), we devised a directed strategy for obtaining clones representing the remaining portion of the genome. Given the low titer of HCV and the absence of a cell culture system for the propagation

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HCV, hepatitis C virus; BVDV, bovine viral diarrhea virus; ORF, open reading frame; nt, nucleotide(s); PCR, polymerase chain reaction.

*To whom reprint requests should be addressed.

†The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M58406 and M58407).

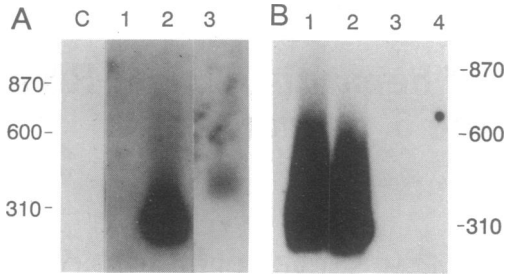

FIG. 1. (*A*) Analysis of the 5′ end of HCV RNA by primer extension. RNAs from serum of an uninfected (lane C) or infected donor (lanes 1–3) were converted into cDNA using primer 51 and tailed with dA. Untailed cDNA (lane 1) and tailed cDNA (lanes C, 2, and 3) were converted into double-stranded cDNA using primer 88, amplified by PCR using primer 52/11, and subjected to Southern blot analysis using probes 89 (lanes C, 1, and 2) and 90a (lane 3). (*B*) Analysis of the 3′ end by oligo(dT) priming. RNAs from uninfected (lane 4) or infected donor (lanes 1–3) were converted into cDNA using primer 88 (lanes 1 and 4), primer 79 (lane 2), and primer 11 (lane 3). Each cDNA was amplified by PCR using primers 32/11, 32/79, and 32/11, respectively, and analyzed with probe 34. Size markers are indicated in nucleotides.

of HCV, we utilized PCR for the characterization of the terminal regions of the HCV genome.

The 5′ Terminus of HCV RNA. To analyze the extreme 5′ end, we employed the combined methods of primer extension and PCR. Primer 51 (see Fig. 2), derived from a 5′ terminal clone, λ 18g (2), was used to prime cDNA synthesis on HCV RNA extracted from the plasma of an infected chimpanzee and a HCV-positive donor. After synthesis of the first strand, the cDNAs were dA-tailed by terminal deoxynucleotidyltransferase and converted to double-stranded cDNA using oligo(dT) primer-adapter 88, which contains 15 residues of dT, an SP6 phage promoter, and a *Not* I cloning site (dAATTCGCGGCCGCCATACGATTTAGGTGACACTATAGAAT$_{15}$) (9). The resultant cDNAs were amplified by PCR using primers 52 (see Fig. 2) and 11 (dAATTCGCGGCCGCCATACGA). Since primer 11 contains only the *Not* I site and a portion of the SP6 sequence, it can be used effectively under stringent conditions as a PCR primer for amplifying any 88-primed cDNA at the oligo(dT) priming site. Southern blot analysis of the PCR product with two probes, 89 and 90a (Fig. 1), indicated that two species of product were made by primer extension; the major product was small and hybridized to probe 89 (Fig. 1*A*, lane 2), and the minor product was larger and hybridized to probe 90a (Fig. 1*A*, lane 3). The PCR products were cloned into the *Not* I site of pUC18S. Upon screening, most clones hybridized to probe 89, but only a few also hybridized to 90a. The sequence analysis showed that all of the 30 clones positive to probe 89 precisely terminated at the same G residue (nucleotide 145) in the previously determined sequence (Fig. 2, arrow). In these clones, the length of the oligo(dT) tail before the G residue varied, suggesting that they were derived from independent cDNAs. On the other hand, probe 90a-positive clones contained inserts of variable length [250–270 (nt)], again preceded by a varying length of oligo(dT). The clone with the longest insert, J33, had an identical sequence to λ18g except for an additional 22 nt at the 5′ end (Fig. 2). This extended sequence is (G+C)-rich and could form a hairpin structure with a calculated ΔG value of −14.4 kcal (1 kcal = 4.18 kJ) (Fig. 2) (14). To confirm this result, we repeated the experiment described above except that primer 95 instead of primer 51 was used for the cDNA synthesis and PCR. Primer 95 binds to sequence upstream from nucleotide 145 (Fig. 2), thus avoiding amplification of the more abundant tailed cDNAs produced with primer 51. From this experiment, we obtained several clones, the sequences of which were identical to that of J33.

Our preliminary data suggest that the two species are present in approximately equal concentration in a given RNA

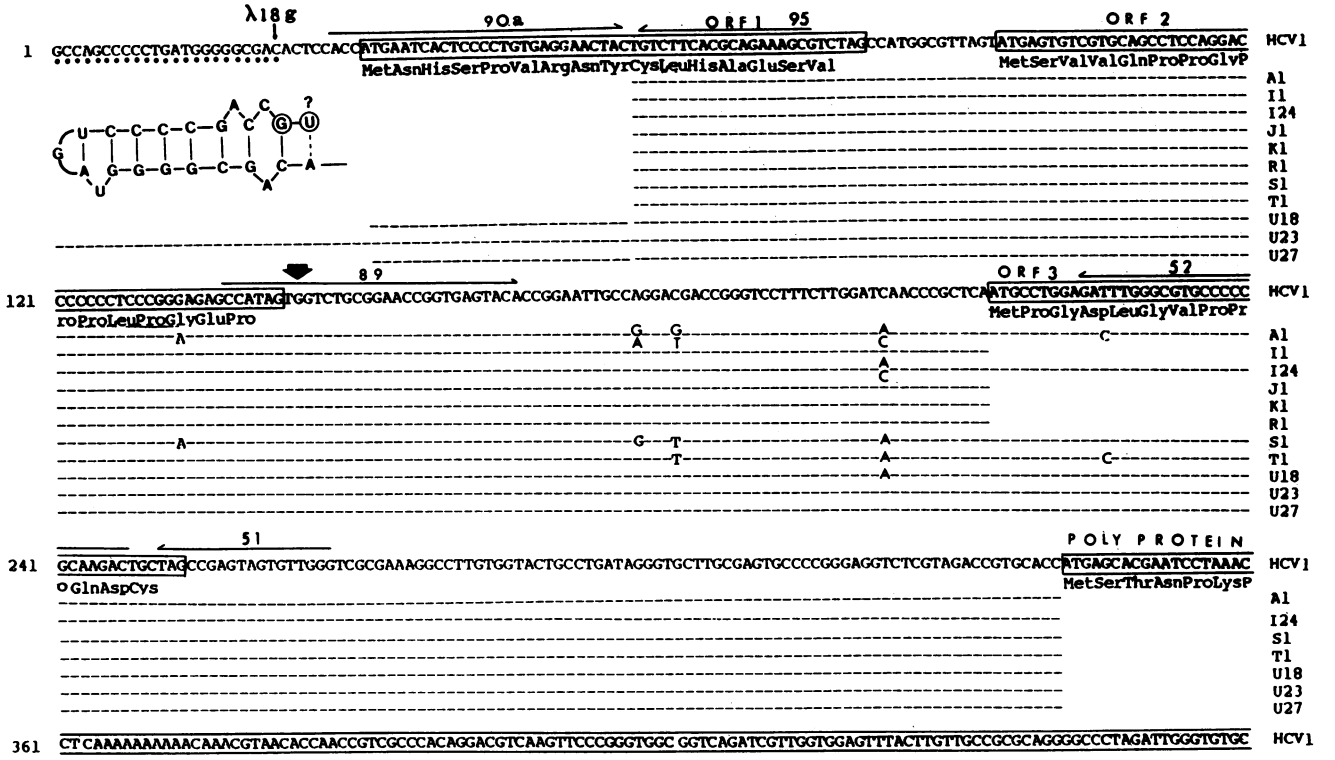

FIG. 2. Sequence at the 5′ untranslated region of the prototype HCV genome (HCV1) and its comparison to that of HCV isolates from Australia (A), Italy (I), Japan (J), Korea (K), Argentina (R), South Africa (S), Taiwan (T), and United States (U). Sequence is marked as dashed lines and nucleotide substitutions are given as letters. The 5′ hairpin structure formed by the terminal 22 nucleotides (nt) is shown. The 5′ end of the major primer extension product is indicated with an arrowhead. Three upstream ORFs and the polyprotein ORF are boxed. An internal poly(A) tract is underlined. Primers are overlined showing their polarities.

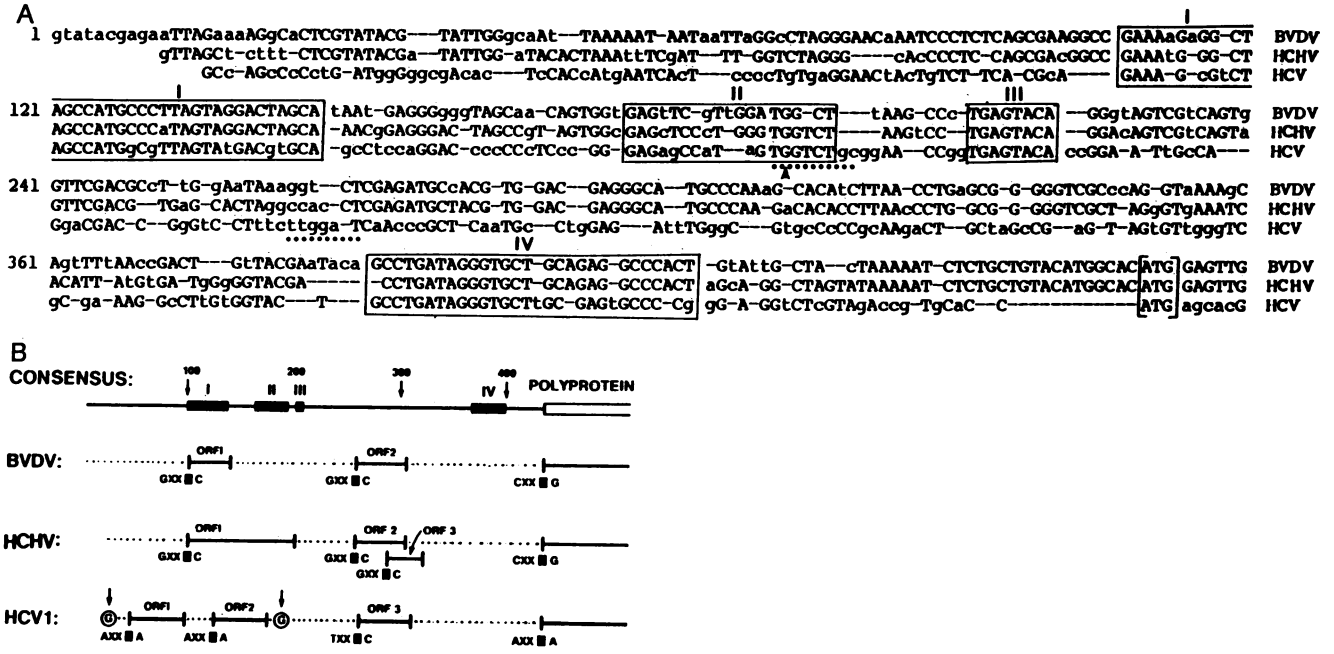

FIG. 3. (*A*) Sequence alignment of the 5′ untranslated region in HCV, BVDV, and hog cholera virus (HCHV). The nucleotide numbers are given from a consensus alignment. Dashes represent spaces between adjacent nucleotides introduced for maximum alignment. Uppercase and lowercase letters indicate matched nucleotides in two of the three viral sequences and no match, respectively. Four highly conserved regions are boxed. Two sequence matches between HCV and yellow fever virus are marked under HCV with a dotted line. An arrowhead in homology block II indicates the 5′ end of the major primer extension product from the HCV RNA. (*B*) Relative position of small ORFs in the 5′ untranslated region of the three viruses. The AUG codon in each ORF is marked as a closed square with the nucleotide sequence around it. Two possible cap sites in HCV RNA are indicated by arrows.

sample (data not shown). The difference in abundance of the two primer-extension products seen in Fig. 1 may be due to the 5′ hairpin structure present in the longer RNA. This feature probably affects the efficiency of dA-tailing by terminal deoxynucleotidyltransferase.

Sequence Conservations in the 5′ Untranslated Region of HCV. We determined the nucleotide sequence of the 5′ untranslated region of 11 HCV isolates collected from individuals from five continents (Fig. 2). All human isolates showed a high degree of sequence conservation with the prototype virus, HCV1. This 341-nt region was completely sequenced for one isolate and was identical to HCV1. Among 10 partially sequenced isolates, 6 were identical to HCV1 and the remaining 4 differed by between 1 and 5 nt. These mismatches were clustered at five positions (Fig. 2).

The nucleotide sequence of the 5′ untranslated region of HCV was compared with other viruses, including those that had previously been shown to be related to HCV (2–5, 15). A computer-assisted homology search in the data bank using the program ALIGN (16) did not reveal significant similarities between HCV and other viruses; however, it detected several tracts of perfect match, up to 14 nt, between HCV and two pestiviruses, bovine viral diarrhea virus (BVDV) (17) and hog cholera virus (18). In this search, the overall sequence homology in the 5′ untranslated region was 71% between BVDV and hog cholera virus, 49% between HCV and hog cholera virus, and 47% between HCV and BVDV. A similar comparison among the three sequences using the program MALIGN (19) with a window of 8 nt revealed three regions of pronounced homology, shown as regions I, III, and IV in Fig. 3*A*. When the same three sequences were aligned to allow for maximum homology, a total of four blocks of 8, 22, 30, and 37 nt showed sequence identity >70%. Furthermore, numerous small tracts of identical 2–5 nt are scattered throughout the entire region. In addition to the nucleotide sequence homology, the 5′ untranslated regions of these three viruses are similar in length. They also contain small ORFs upstream from the polyprotein ORF (Fig. 3*B*). The BVDV and hog cholera virus ORFs as well as HCV ORF 2 and 3 encode peptides rich in helix-breaking amino acids (30–50% Gly and Pro).

Exploring the 3′ Terminus of HCV. HCV RNA binds to oligo(dT)-cellulose (1). The viral RNA contains an internal tract of 10 A residues at nucleotide 23 from the polyprotein start codon (Fig. 2). To test whether a poly(A) sequence exists at the 3′ end of viral RNA as well, we synthesized cDNA using primer 88 and amplified by PCR using primers 11 and 32. The primer 32 (Fig. 4) was designed from λ6K (2), an extreme 3′ end clone derived from a randomly primed λgt11 library (1). This PCR product was positive to probe 34 (Fig. 1*B*, lane 1), indicating that primer 88 successfully primed cDNA synthesis. Since the cDNA primer 88 contains other sequences in addition to the oligo(dT) homopolymer tail, we investigated which of these components was used as the HCV-specific cDNA primer. cDNA was synthesized using three related primers—88, 79 (dAATTCGCGGC-$CGCT_{15}$), and 11. Primers 79 and 11 are deletion derivatives of primer 88; primer 79 lacks the SP6 sequence, whereas primer 11 lacks the oligo(dT) homopolymer tail. The cDNAs synthesized from primers 88, 79, and 11 were amplified by PCR using primer sets 32/11, 32/79, and 32/11, respectively. Whereas products from primers 88 and 79 were positive to probe 34, the product from primer 11 was negative (Fig. 1*B*, lanes 1–3). This result indicates that cDNA containing HCV-specific sequences was formed from the oligo(dT) portion of primers 88 and 79 and is consistent with a poly(A) or an A-rich sequence at the 3′ end of HCV RNA.

The cDNA formed from primer 88 and amplified by PCR with primers 11 and 32 was cloned into pUC18S and positive clones were identified with probe 34. The nucleotide sequence of the inserts from these clones, an example of which is clone J16, was identical to that of the original λ6K clone down to the C residue at nucleotide 9208 but was then followed by a poly(A) tail of variable length (Fig. 4). Since the sequence in

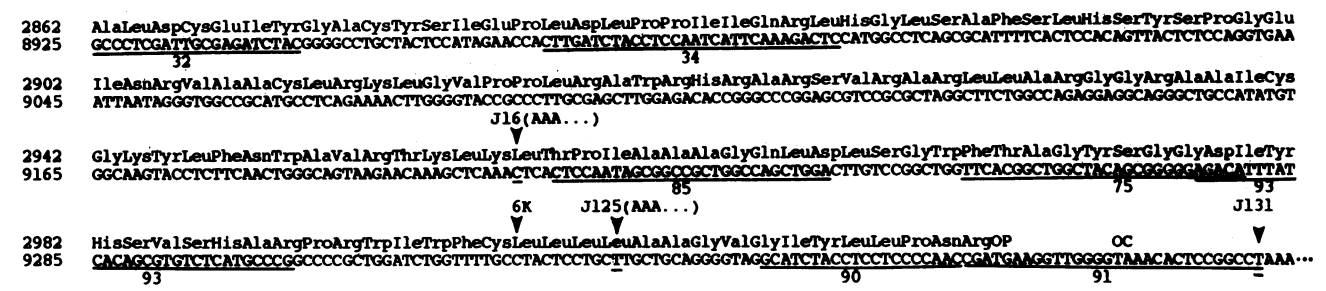

FIG. 4. Nucleotide and deduced amino acid sequences near the 3′ ends of HCV RNA. Positions of the 3′ end of a λ clone (λ6k) and the three major 3′ end clones (J16, J125, and J131) are indicated by arrowheads. The presence of poly(A)$^+$ tails in J16 and J125 are marked as (AAA . . .) over each poly(A) site. Primers are underlined.

J16 was smaller than anticipated, we presumed that HCV RNA containing sequences downstream from the 3′ end of J16 is found in molecules present at lower concentration.

Next, to avoid detecting the RNA species generating clone J16, we synthesized cDNA using primer 88 as before but used primers 75 and 11 to generate the PCR product. The PCR product was again cloned as before and the positive clones were selected using probe 93. Sequence analysis of these positive clones identified two additional species of RNA as exemplified by clones J125 and J131. Clone J125 had an additional 7 nt relative to the 3′ end of λ6K followed by a poly(A) tail of variable length (Fig. 4). Clone J131 carried a new sequence extending 67 nt 3′ from the poly(A) site of clone J125, followed by a poly(A) tail (Fig. 4). This new sequence contained two in-frame stop codons for the polyprotein ORF, 27 and 15 nt preceding the poly(A) tail (Fig. 4).

We estimated the relative concentration of the three species of poly(A)$^+$ RNA represented by clones J16, J125, and J131. Our initial attempt to measure the concentration of each species directly from the cDNA synthesized with primer 88 and then amplified with primers 32 and 11 was unsuccessful. Therefore, we selectively amplified cDNA formed by primer 88 using three sets of primers—32/11, 85/11, and 90/11. Since PCR products accumulate to saturation in the late phase of amplification, we measured the amount of each PCR product at different cycles in the amplification process by Southern analysis (Fig. 5). The ratio of the poly(A)$^+$ RNA species represented by clones J16, J125, and J131 was approximately 500:25:1 as determined by laser densitometric scanning, corrected for small differences in the measured priming efficiencies of the three sets of primers.

DISCUSSION

We have determined the nucleotide sequence of the 5′ and 3′ termini of the HCV genome using PCR techniques. This work confirms the earlier results of Choo *et al.* (2) and others (20, 21) and extends the HCV sequence at both termini, including a hairpin structure at the 5′ end and in-frame stop codons at the 3′ end. The combined HCV sequence now totals 9401 nt prior to the poly(A) tail, comprising 341 nt of 5′ untranslated region, a 9033-nt coding region, and only 27 nt in the 3′ untranslated region.

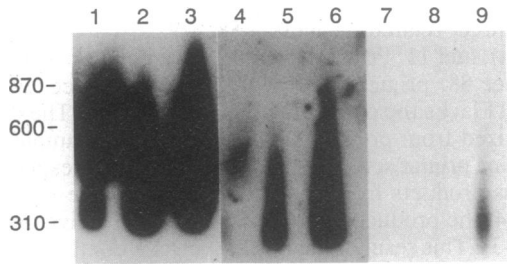

FIG. 5. Determination of the relative concentration of multiple poly(A)$^+$ HCV RNAs by differential PCR amplification. RNA from serum of an infected chimpanzee was converted into cDNA using primer 88. The cDNA was PCR amplified using primers 32/11 (lanes 1–3), 85/11 (lanes 4–6), and 90/11 (lanes 7–9). The PCR product was Southern blot analyzed using probes 34 (lanes 1–3), 75 (lanes 4–6), and 91 (lanes 7–9) at amplification cycles 20 (lanes 1, 4, and 7), 25 (lanes 2, 5, and 8), and 35 (lanes 3, 6, and 9). The blots were exposed to x-ray film for 12 hr (lanes 1–3) and 96 hr (lanes 4–9). Size markers are indicated in nucleotides.

Our experimental data show that the 5′ untranslated region of the HCV genome is highly conserved among viral isolates from around the world. This suggests that there is a strong evolutionary constraint operating on this nucleotide sequence, indicating a likely role for this sequence in viral replication. At a practical level, this "signature" sequence could serve as a HCV-specific DNA probe for the detection of all strains of the virus. We have developed a set of universal primers as well as a sensitive and highly reliable PCR protocol based on this conserved sequence (J.H.H., unpublished data).

The 5′ untranslated region of the HCV genome is remarkably similar to that of pestiviruses (Fig. 3). This homology is not found in any other known RNA viruses, including flaviviruses. The same region of the flaviviruses is smaller, exhibits little sequence homology with HCV, and does not contain small ORFs except for tick-borne encephalitis virus (23). Our analysis provides convincing evidence for a clear evolutionary relationship between HCV and pestiviruses. This confirms earlier observations showing that the helicase region of HCV is more similar to that of the pestiviruses than the flaviviruses (4, 5). Previously, Takeuchi *et al.* (21) also recognized partial nucleotide sequence identity between HCV and BVDV (block IV in Fig. 3*A*).

The 5′ terminal 22 residues of HCV RNA can form a hairpin structure that may be stable under physiological conditions (Fig. 2). The stability of this structure appears to be as great as that of a similar structure found at the 5′ termini of alphaviruses (24). This structure has been postulated to be a recognition signal for replicase or capsid protein in alphaviruses (24). We have not yet determined whether HCV RNA is capped at the 5′ terminus as reported for alphaviruses and flaviviruses (24, 25). If HCV has such a cap structure, the 5′ terminal G residue could be a cap site, although a U residue could also be assigned due to the tailing scheme we employed in our experiment.

The function of the small ORFs is currently not understood, but they could be negative translational control elements since HCV RNA is not efficiently translated *in vitro* unless they are removed (Q.-L.C., unpublished data). This result is consistent with Kozak's scanning hypothesis for upstream AUGs in mRNAs with a long 5′ untranslated region (26, 27). Inspection of the sequence reveals that the first and second ORFs and the polyprotein ORF have an identical initiation sequence of AXXAUGA, whereas the third ORF has UXXAUGC (Fig. 3*B*), which has often been found to be nonfunctional in translational initiation (26).

In addition to the presumptive genomic RNA, we have detected another prominent RNA species in RNA samples from different blood donors. Interestingly, the 5′ terminus of this RNA resides between the second and third ORF (Fig. 2); thus, it could be free of translational control applied by the first and/or second ORF. The origin and biological role of this truncated RNA is unknown at present. It could represent a viral mRNA or a cleavage product generated by a specific nuclease acting on genomic RNA.

By a PCR method, we have detected a poly(A) tail at the 3′ end of HCV RNA. In view of the unusual sensitivity of the method, PCR results always need to be carefully examined. It is however unlikely that the detection of 3′ poly(A) in HCV RNA is due to an experimental artifact for the following reasons. (*i*) An oligo(dT) sequence in the cDNA primer was necessary and sufficient for a HCV-specific priming. (*ii*) Inspection of the sequences occurring immediately after each poly(A) site of J16 and J125 revealed no internal A-rich tract (Fig. 4). (*iii*) Primer 88 exhibited a specific priming to yield a HCV cDNA from the internal A tract within the coding region (Fig. 2) (data not shown). (*iv*) The length of the poly(A) tail in each clone is much longer (average, 40) than that of the T tract (15 residues) in the cDNA primer 88, whereas the cDNA that was initiated on the internal A tract contained an A tract of 15 residues.

We have also detected heterogeneity in the 3′ end of HCV RNA. Three major species of RNA, all apparently polyadenylylated, have been detected in different concentrations in liver and serum. The two prominent RNAs, typified by J16 and J125 RNA, terminate at specific positions prior to the polyprotein stop codon. Although the biological role and the mode of polyadenylylation of these subgenomic RNAs are unknown, they could be derived from defective viruses that depend upon the intact virus for the functional RNA-dependent RNA polymerase. The longer RNA (J131) we detect contains two in-frame stop codons followed by 27 nt prior to a poly(A) tail. This RNA could represent genomic RNA or a subgenomic mRNA. However, we cannot confidently state that this represents the 3′ terminal sequence of the viral genome. In particular, the sequences associated with viral replication have not yet been identified. In Sindbis virus, poly(A) tails have been found at the 3′ end of genomic, defective viral, and mRNA (28), and the polyadenylylation seems to be a functional component of the replication and expression system. Whether the HCV genomic RNA is polyadenylylated and what fraction of total HCV RNA is polyadenylylated at the 3′ end remain to be determined.

HCV and pestiviruses share an additional similarity in genomic organization. Sequence (29) and expression data (R. Ralston and R. Spaete, personal communications) suggest that the flaviviral nonstructural protein 1 (NS1) homologue of HCV polyprotein is a second envelope protein. Pestiviruses are thought to have two or more envelope proteins (22, 23).

Despite the described similarities of HCV to pestiviruses, there are also significant differences between these viruses. (*i*) There is little overall amino acid sequence homology in the polyprotein of the two viruses (2–5). (*ii*) Neither subgenomic nor 3′ polyadenylylated RNA species have been found in pestiviruses (17, 18). (*iii*) The frequency of CG dinucleotide and codon usage in HCV genome is random, whereas the same is biased in pestiviruses (data not shown). These observations suggest that HCV is a unique viral type and should be classified as an independent genus in an appropriate viral family. The pestiviruses have traditionally been classified as a genus of Togaviridae but have recently been proposed for reclassification as a second genus of Flaviviridae (17, 30). In this context, HCV may deserve classification as a third separate genus within the Flaviviridae family as we have proposed (2, 3, 5). Intriguingly, HCV also resembles other RNA viruses that belong to alphavirus-and picornavirus-like superfamilies in some aspects. Whether these features reflect a direct evolutionary link of HCV to these RNA viruses or an independent origin of these motifs by genome shuffling remains to be elucidated.

We thank Drs. W. Rutter, E. Penhoet, and P. Valenzuela for encouragement, Dr. P. Barr for the synthesis of oligonucleotides, Drs. A. Weiner, T. Sanders, and R. Ralston for critical review and discussion, K. Crawford for assistance in computer analysis of sequences, and T. Jones and L. Clark for assistance in preparation of the manuscript. This work was supported by Chiron, CIBA-Geigy, and Ortho Diagnostics Systems.

1. Choo, Q.-L., Kuo, G., Weiner, A. J., Overby, L., Bradley, D. W. & Houghton, M. (1989) *Science* **244**, 359–362.
2. Choo, Q.-L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, A., Barr, P. J., Weiner, A. J., Bradley, D. W., Kuo, G. & Houghton, M. (1991) *Proc. Natl. Acad. Sci. USA*, in press.
3. Choo, Q.-L., Weiner, A. J., Overby, L. R., Kuo, G., Houghton, M. & Bradley, D. W. (1990) *Br. Med. Bull.* **46**, 423–441.
4. Miller, R. H. & Purcell, R. H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2057–2061.
5. Choo, Q.-L., Han, J. H., Weiner, A. J., Overby, L. R., Bradley, D. W., Kuo, G. & Houghton, M., in *Proceedings of 1989 International Symposium on Viral Hepatitis C, D, and E*, eds. Shikata, T., Purcell, R. H. & Uchida, T. (Elsevier, Amsterdam), in press.
6. Bradley, D. W. & Mayard, J. E. (1986) *Semin. Liver Dis.* **6**, 67–81.
7. Han, J. H., Stratowa, C. & Rutter, W. J. (1987) *Biochemistry* **26**, 1617–1625.
8. Cathala, G., Savouret, J., Mendez, B., West, B., Karin, M., Martial, J. & Baxter, J. (1985) *DNA* **2**, 329–335.
9. Han, J. H. & Rutter, W. J. (1988) in *Genetic Engineering*, ed. Setlow, J. K. (Plenum, New York), Vol. 10, pp. 195–219.
10. Saiki, R. K., Scarf, S., Faloonas, F., Mullis, K. B., Horn, J. T., Ehich, H. A. & Arnheim, N. (1985) *Science* **230**, 1350–1354.
11. Hoopes, B. C. & McClure, W. R. (1981) *Nucleic Acids Res.* **9**, 5493–5504.
12. Chen, E. & Seeburg, P. H. (1985) *DNA* **4**, 165–176.
13. Innis, M. A., Myambe, K. B., Gelfand, D. H. & Brow, M. A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 9436–9440.
14. Freir, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. & Turner, D. H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9373–9377.
15. Takeuchi, K., Kubo, Y., Boonmars, S., Watanabe, Y., Katayama, T., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I. & Miyamura, T. (1990) *Nucleic Acids Res.* **18**, 4626.
16. Meyers, E. & Miller, W. (1988) *Comput. Anal. Biosci.* **4**, 11–17.
17. Collett, M. S., Moennig, V. & Horzinek, M. D. (1989) *J. Gen. Virol.* **70**, 253–266.
18. Meyers, G., Rumenapf, T. & Thiel, H.-J. (1989) *Virology* **171**, 555–567.
19. Martinez, H. M. (1987) *Sequence Analysis Programs Manual* (Biomathematics Computation Lab., Univ. of California, San Francisco).
20. Okamoto, H., Okada, S., Sugiyama, Y., Yotsumoto, S., Tanaka, T., Yoshizawa, Y., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1990) *Jpn. J. Exp. Med.* **60**, 167–177.
21. Takeuchi, K., Yoshihiro, K., Boonmar, S., Watanabe, Y., Katayama, T., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I. & Miyamura, T. (1990) *J. Gen. Virol.* **71**, 3027–3033.
22. Collet, M. S., Larson, R., Gold, C., Strick, D. K. & Purchico, A. F. (1988) *Virology* **165**, 191–199.
23. Chambers, T. J., Hahn, C. S., Galler, R. & Rice, C. M. (1990) *Annu. Rev. Microbiol.* **44**, 649–688.
24. Strauss, E. G. & Strauss, J. H. (1986) in *Togaviridae and Flaviviridae*, eds. Schlesinger, S. & Schlesinger, M. J. (Plenum, New York), pp. 35–115.
25. Rice, C. M., Lenches, E. M., Eddy, S. R., Shin, S. J., Sheet, R. L. & Strauss, J. H. (1985) *Science* **229**, 726–733.
26. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8148.
27. Pelletier, J. & Sonenberg, N. (1988) *Nature (London)* **334**, 320–325.
28. Monroe, S. S., Ou, J.-H., Rice, C. M., Schlesinger, S., Straus, E. G. & Straus, H. J. (1982) *J. Virol.* **41**, 153–162.
29. Weiner, A. J., Brauer, M. J., Rosenblatt, J., Richman, K. H., Tung, J., Crawford, K., Bonino, F., Saracco, G., Choo, Q.-L., Houghton, M. & Han, J. H. (1991) *Virology*, in press.