Group specific sequences and conserved secondary structures at the 3′ end of HCV genome and its implication for viral replication

Jang H.Han and Michael Houghton
Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA

Submitted April 28, 1992 GenBank accession nos*

Comparative sequence analysis on viral isolates revealed that the nucleotide sequence within the 3′ untranslated region of hepatitis C virus (HCV) RNA is least conserved (30% to 62%) within the HCV genome (1, 2, 3, 4); however, we find that the sequence in this region is an excellent group specific marker. When the sequences within the 3′ untranslated region of 8 isolates reported to date are aligned, it recreates the same group assignment made on the basis of the amino acid sequence homology (Figure 1a). For example, the nucleotide sequence identity is 96% between HCV-1 and HC-J1 (group I isolates), 95% between HCV-J and HC-J4 (group II isolates), and 82% between HC-J6 and HC-J7 (group III isolates) (HC-J7 may need to be classified as group IV). However the sequence identity is less than 60% in isolates of different group. Thus, taken together with the fact that the nucleotide sequence in the 5′ untranslated region is highly conserved (>93%) in all viral isolates ('signature sequence') (5), HCV RNA appears to have both virus specific and group specific nucleotide sequence at the 5′ and the 3′ untranslated region, respectively.

Viruses in general have a mechanism to preserve the ends of genome during replication. In positive strand RNA viruses, the 3′ untranslated region is an entry site for RNA polymerase for viral replication and it normally contains conserved sequence and/or a secondary structure as a recognition signal. However the observed sequence heterogeneity in the 3′ untranslated region leads us question as to how HCV initiates replication. In the absence of a conserved sequence except for a homopolymer tail at the 3′ end, which is poly A for HCV-1 (5) or poly U for Japanese isolates (4), we have searched by computer for secondary structure within a 300 nucleotide region from the 3′ end of the HCV RNA and found four stem and loop structures which are conserved in all viral groups (Figure 1b). Among these four structures, the size and the spacing of the first three are very similar in each group. The fourth structure is absent in HCV-1 and its size is different between groups II and III. We propose that the homopolymer tail and all or a part of these secondary structures are recognition signals for the viral RNA polymerase. Further, we anticipate that the viral replication efficiency might be different among different group-specific HCV isolates. This could be related to differences in viral pathogenicity among different HCV groups.

REFERENCES

1. Houghton,M., *et al.* (1991) *Hepatology* **14**, 381–388.
2. Okamoto,H., *et al.* (1992) *J. Gen. Virol.* **73**, 673–679.
3. Inchauspe,G., *et al.* (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10272–10276.
4. Kato,N., *et al.* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9524–9528.
5. Han,J.H., *et al.* (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1711–1715.

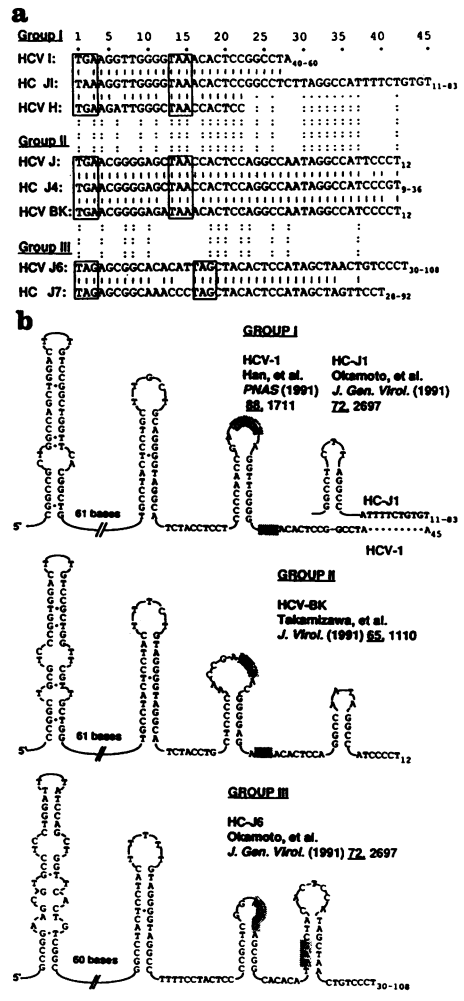

Figure. 1. a, Alignment of the 3′ untranslated region sequence. Matches and mismatches are indicated by vertical lines and dotted lines, respectively. b, Possible secondary structures at the 3′ end of HCV RNA. Two in-frame stop codons are boxed or shaded.

*GenBank accession nos M58407 and M67463; DDBJ accession nos D00825, D00826, D00944 and D00828