Volume 280, number 2, 325–328 FEBS 09522 March 1991

ADONIS 0014579391002298

Molecular structure of the Japanese hepatitis C viral genome

Nobuyuki Kato, Makoto Hijikata, Masanori Nakagawa, Yuko Ootsuyama, Kanae Muraiso, Showgo Ohkoshi* and Kunitada Shimotohno

Virology Division, National Cancer Center Research Institute, 5-1-1, Tsukiji, Chuo-ku, Tokyo 104, Japan

Received 24 January 1991; revised version received 25 January 1991

The amino acid sequence of the polyprotein deduced from the nucleotide sequence of the Japanese hepatitis C virus genome (N. Kato et al. (1990) Proc. Natl. Acad. Sci. USA 87, 9524–9528) indicated that this virus is a member of a new class of positive-stranded RNA viruses. Several domains of this polyprotein also showed weak homology with those of flaviviruses and pestiviruses including the chymotrypsin-like serine proteinase, NTPase and RNA-dependent RNA polymerase.

Non-A,non-B hepatitis; RNA virus; HCV-J; Amino acid sequence.

1. INTRODUCTION

Hepatitis C virus (HCV) is the major causative agent of post-transfusional non-A, non-B hepatitis throughout the world [1]. Choo et al. [2] were the first to clone part of the genome of HCV and reported that this genome was a positive-stranded RNA molecule of about 10 kb. A homology search using the partial sequence of the original isolate of the hepatitis C virus (HCV-US) suggested that HCV-US is related to flaviviruses and pestiviruses [3].

Recently, we have molecularly cloned almost the whole genome of a Japanese isolate of hepatitis C virus (HCV-J), which consists of 9413 nucleotides [4]. On comparison of about 8 kb lengths of the two genomes, lacking their two ends, we found that HCV-J showed 23% difference in nucleotide sequence and 15% difference in amino acid sequence from HCV-US [5,6]. These findings suggest that HCV-J and HCV-US are different subtypes of HCV [4]. In this communication, we report that HCV is a member of a new class of positive-stranded RNA viruses, which shows weak homology with flaviviruses and pestiviruses.

2. EXPERIMENTAL

Nucleotide and amino acid sequences of the HCV-J genome were analyzed with a DNASIS DNA sequence analysis system (Hitachi) and compared with published sequences of GeneBank (R64.0, June 1990), EMBL (R23.0, May 1990), SWISS-PROT (R14.0, April 1990), or Protein Identification Resource (R25.0, June 1990).

Correspondence address: K. Shimotohno, Virology Division, National Cancer Center Research Institute, 5-1-1, Tsukiji, Chuo-ku, Tokyo 104, Japan

**Present address:* Third Department of Internal Medicine, School of Medicine, Niigata University, 1-757 Asahimachi, Niigata-City 951, Japan

3. RESULTS AND DISCUSSION

To clarify the relationship of the HCV-J genome with those of other animal RNA viruses and to obtain information on the biological functions of the virus-encoded proteins, we carried out homology searches using the nucleotide sequence of the HCV-J genome and deduced amino acid sequence of the HCV-J polyprotein. At the nucleotide level, we could not find any significant homologies of the sequence with other animal viruses. However, at the amino acid level, we found several putative functional domains which showed similar hydropathy profiles and sequences to those of flaviviruses or pestiviruses in addition to a unique structure in the HCV-J polyprotein.

The hydropathy profile of amino acids 190–730 of the HCV-J polyprotein was similar to that of amino acids 480–1050 of the polyprotein of bovine viral diarrhea virus (BVDV) [7], as shown in Fig. 1. The hydropathy profiles of other regions were quite different from those of BVDV, Dengue type 2 virus (DV2) [8] and other flaviviruses or pestiviruses (Fig. 1 and data not shown). Amino acids 480–1050 of the BVDV polyprotein corresponds to the sequences of two envelope glycoproteins (gp25 and gp53) [9], so a putative envelope protein(s) of HCV-J may be located in amino acids 190–730 of the HCV-J polyprotein. In fact, we recently demonstrated that amino acids 192–730 of HCV-J encoded two glycoproteins (gp35 and gp70) as predicted (Hijikata et al., manuscript in preparation). We could not find any significant homology between the amino acid sequences of the envelope regions of HCV-J and BVDV, but we found a

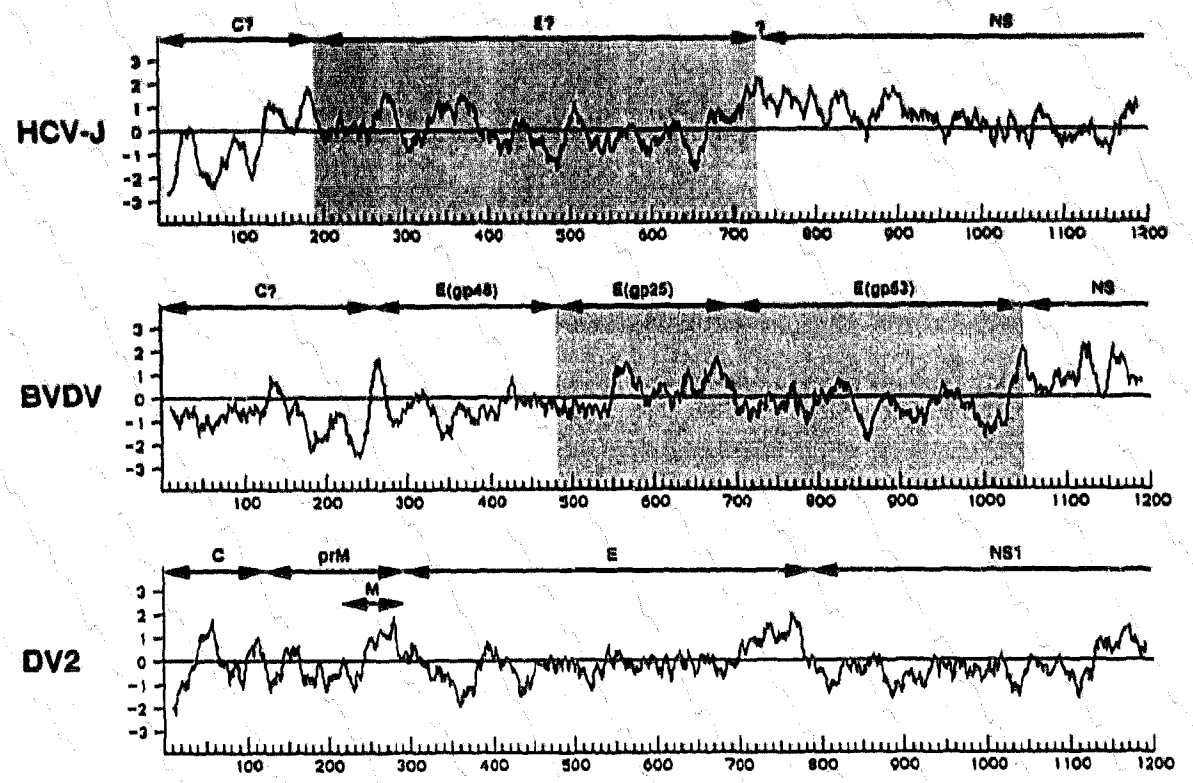

Fig. 1. Hydropathy plots of polyprotein sequences (N-terminal 1200 amino acids) of HCV-J, BVDV and DV2. The program of Kyte and Doolittle [17] with a search length of 21 amino acids was used. The degree of hydrophobicity increases with distance above the horizontal line; hydrophilicity increases with distance below the horizontal line. Shaded areas show regions with similar hydropathy profiles in HCV-J and BVDV.

unique, regular amino acid sequence, Pro-$(X)_5$-Pro-$(X)_6$-Pro-$(X)_6$-Pro-$(X)_6$-Pro$(X)_6$-Pro-$(X)_5$-Pro(amino acids 471–511), in the putative HCV-J envelope protein (gp70). This regular sequence is conserved in HCV-US [5] (except for the second Pro residue) and other HCV-J strains (unpublished data). Interestingly, a cysteine residue is present in each of the last 4 intervening (X)-sequences, and these cysteine residues are in the same positions in HCV-J and HCV-US (data not shown). This unique sequence has not been found in any of the polyproteins of flaviviruses or pestiviruses or the cellular proteins sequenced so far. The sequence most closely homologous is that of chicken elastin [Pro-$(X)_6$-Pro-$(X)_6$-Pro-$(X)_6$-Pro-$(X)_6$-Pro-$(X)_5$-Pro] [10], but in elastin the (X)-sequences between the Pro residues are all(Gly-Val-Gly-Gly-Leu-Val) [10] unlike in HCV-J. This unique sequence may be important for the specific secondary structure of the protein.

A putative non-structural protein 3(NS3) region of HCV-J contains a consensus sequence of a trypsin-like serine proteinase, as predicted for the NS3 regions of flaviviruses and pestiviruses [11]. Recently, the N-terminal domain of the NS3 protein, which contains the consensus sequences of chymotrypsin-like serine proteinase, of yellow fever virus was demonstrated to have proteinase activity which is responsible for site-specific cleavages in the viral polyprotein [12]. This consensus sequence is located at amino acids 1075–1185 in the HCV-J polyprotein. The amino acid sequences of chymotrypsin-like serine proteinase domains in HCVs, flaviviruses and pestiviruses are compared in Fig. 2a. Residues His-57, Asp-102, and Ser-195 (chymotrypsin numbering system), which constitute the catalytic site [13] in flaviviruses and pestiviruses, are well conserved in HCV-J as well as HCV-US. However, the amino acid residues that might contribute to substrate binding (* in Fig. 2a) are different from those of flaviviruses and pestiviruses. These differences may explain the difference in substrate specificities, because the polyproteins of HCVs do not contain any putative cleavage sites (dibasic Arg/Lys residues) that the flavivirus-encoded proteinase preferentially recognizes. The region immediately C-terminal to the putative serine proteinase domain has the sequence, Gly-Ser-Gly-Lys (amino acids 1233–1236) in the HCV-J polyprotein and Asp-Glu (amino acids 1316–1317) in the HCV-J polyprotein. These sequences well match the NTP-binding sequence [Gly-X-Gly-Lys-$(X)_{39-105}$-Asp-Glu], which is evolutionarily conserved [14]. These sequences are conserved in HCV-US and in a number of other viruses, including flaviviruses and pestiviruses, as shown in Fig. 2b. The distance of 80 amino acids between the Gly-Ser-Gly-Lys

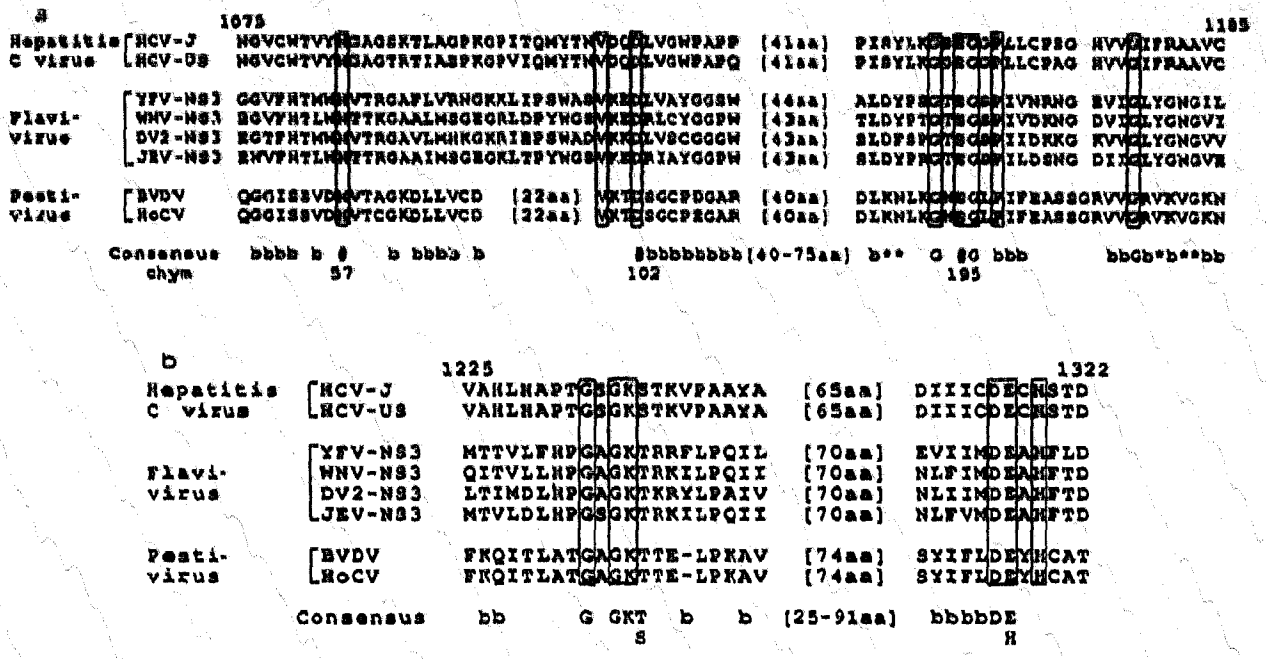

Fig. 2. Amino acid sequence alignment of functionally important regions in HCV, flaviviruses and pestiviruses. (a) Putative serine proteinase domain. The chymotrypsin numbering system (chym) is used. (#) indicate a conserved catalytic triad of His (chym 57), Asp (chym 102) and Ser (chym 195). (*) indicate the residues that contribute to substrate binding. b means a hydrophobic amino acid residue. Identical residues are boxed. The consensus sequence was obtained from the sequences of flaviviruses, pestiviruses and cellular Ser proteinase [11]. Numbers at the top show amino acid positions in HCV-J [4]. (b) NTPase (NTP-binding domain). Identical residues are boxed. The consensus sequence was obtained from [13]. b means a hydrophobic amino acid residue. Numbers at the top show amino acid positions in HCV-J [4]. Additional abbreviations used: YFV, yellow fever virus; WNV, West Nile virus; JEV, Japanese encephalitis virus; HoCV, hog cholera virus.

and Asp-Glu sequences is also similar to those of 84 amino acids in flaviviruses and 88 amino acids in pestiviruses. In flaviviruses, this NTP-binding domain is located in the NS3 region and the viral protein containing this amino acid sequence is proposed to be an NTPase involved in duplex unwinding during replication, transcription, recombination, repair, etc. [14].

The HCV-J polyprotein contains the conserved motifs [15] of RNA-dependent RNA polymerase. The amino acid sequence from residues 2631–2740 of the HCV-J polyprotein, [Asp-$(X)_4$-Asp-$(X)_{56}$-Ser-Gly-$(X)_3$-Thr-$(X)_3$-Asn-$(X)_{25}$-Gly-Asp-Asp-X-Val] well matches the conserved motifs of flaviviruses and pestiviruses, although the 4th motif [15] is not present in the HCV-J polyprotein (data not shown). Since these motifs are located in the C-terminal region as in flaviviruses and pestiviruses, the overall gene organization of HCV-J may be similar to those of flaviviruses and pestiviruses. However, the nucleotide sequence of the HCV-J genome shows no significant homology with those of flaviviruses and pestiviruses. The 3′-noncoding region of the HCV-J genome is only 63 nucleotides long ([4], unpublished data), which is much shorter than those of flaviriruses and pestiviruses (several hundreds nucleotides). Furthermore, this region contains a U stretch, which gives a unique structure in the 3′-terminus of a viral genome [4] in contrast to the strong secondary structures found in flaviviruses [16]. The pattern of codon usage for the polyprotein was also different from those of flaviviruses and pestiviruses. The codon usage of the HCV-J genome for 3010 amino acids was nearly random, although the UUA, GAA and AGA codons were somewhat infrequently used (data not shown). The low CG doublet frequency pointed out in other flaviviruses [16] was not observed. The HCV-US genome showed similar characteristics to the HCV-J genome [5]. Therefore, we propose that HCV should be classified into a new virus family distinct from the *Flaviviridae* including flaviviruses and pestiviruses.

Acknowledgements: The authors are indebted to Dr K. Yamamoto for helpful comments. This study was supported by grants-in-aid for Cancer Research and for a Comprehensive 10-Year Strategy of Cancer Control from the Ministry of Health and Welfare of Japan. M.H. and K.M. are recipients of Research Resident Fellowships from the Foundation for Promotion of Cancer Research.

REFERENCES

[1] Kuo, G., Choo, Q.-L., Alter, H.J., Gitnick, G.L., Redeker, A.G., Purcell, R.H., Miyamura, T., Dienstag, J.L., Alter, M.J., Stevens, C.E., Tegtmeier, G.E., Bonino, F., Colombo, M., Lee, W.-S., Kuo, C., Berger, K., Shuster, J.R., Overby, L.R., Bradley, D.W. and Houghton, M. (1989) Science 244, 362–364.

[2] Choo, Q.-L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W. and Houghton, M. (1989) Science 244, 359–362.
[3] Miller, R.H. and Purcell, R.H. (1990) Proc. Natl. Acad. Sci. USA 87, 2057–2061.
[4] Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T. and Shimotohno, K. (1990) Proc. Natl. Acad. Sci. USA 87, 9524–9528.
[5] Houghton, M., Choo, Q.-L. and Kuo, G. (1988) Eur. Patent Appl. 88, 310, 922.5 and publ. 318, 216.
[6] Takeuchi, K., Boonmer, S., Kubo, Y., Katayama, T., Harada, H., Ohbayashi, A., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I. and Miyamura, T. (1990) Gene 91, 287–291.
[7] Collett, M.S., Larson, R., Gold, C., Strick, D., Anderson, D.K. and Purchio, A.F. (1988) Virology 165, 191–199.
[8] Deubel, V., Kinney, R.M. and Trent, D.W. (1988) Virology 165, 234–244.
[9] Collett, M.S., Larson, R., Belzer, S.K. and Retzel, E. (1988) Virology 165, 200–208.
[10] Bressan, G.M., Argos, P. and Stanley, K.K. (1987) Biochemistry 26, 1497–1503.
[11] Bazan, J.F. and Fletterick, R. (1989) Virology 171, 637–639.
[12] Chambers, T.J., Weir, R.C., Grakoui, A., McCourt, D.W., Bazan, J.F., Fletterick, R.J. and Rice, C.M. (1990) Proc. Natl. Acad. Sci USA 87, 8898–8902.
[13] Sprang, S., Standing, T., Fletterick, R.J., Stroud, R.M., Finer-Moore, J., Xuong, N.-H., Hamlin, R., Rutter, W.J. and Craik, C.S. (1987) Science 237, 905–909.
[14] Gorbalenya, A.E. and Koonin, E.V. (1989) Nucleic Acids Res. 17, 8413–8440.
[15] Poch, O., Sauvaget, I., Delarue, M. and Tordo, N. (1989) EMBO J. 8, 3867–3874.
[16] Rice, C.M., Strauss, E.G. and Strauss, J.H. (1985) In: The Togaviridae and Flaviviridae (Schlesinger, S. and Schlesinger, M.J. eds) Plenum, pp. 279–326.
[17] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.