Vol. 175, No. 1, 1991 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 28, 1991 Pages 220-228

HYPERVARIABLE REGIONS IN THE PUTATIVE GLYCOPROTEIN OF HEPATITIS C VIRUS[1]

Makoto Hijikata, Nobuyuki Kato, Yuko Ootsuyama, Masanori Nakagawa, Shogo Ohkoshi, and Kunitada Shimotohno

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

Received January 11, 1991

SUMMARY: A comparison of the sequences of the putative glycoprotein region in three independent cDNA clones of hepatitis C virus and of sequences of four other clones revealed extensive genetic variation clustered and interspersed with highly conserved amino acid sequences. We obtained evidence for two hypervariable regions in the putative envelope glycoprotein, one region was assumed to be a potential antigenic site, as deduced from the hydrophilicity and analyses of secondary structures. These observations suggest the existence of a large pool of antigenic variants of hepatitis C virus, in Japan.

Hepatitis C virus (HCV), the genome of which has been molecularly cloned (1, 2), is thought to be one of the causative agents of post-transfusion non-A, non-B, hepatitis (3). HCV is apparently an enveloped virus with a positive, single stranded RNA genome encoding a large polyprotein precursor (1). This virus seems to be related to pestiviruses and to flaviviruses, as well as the carmovirus and potyvirus families of plant viruses, as deduced from amino acid sequence similarities (4). Although a viral structural protein has not been identified, the envelope glycoprotein(s) of HCV was roughly estimated to locate in amino acid positions 190 to 700 on the viral polyprotein precursor (2). The amino acid sequence of the region from amino acid positions 350 to 500 was extensively diverged between the HCV genomes in the United States (HCV-US) and in Japan (HCV-J)(35% difference), while amino acid positions 500 to 750 were relatively conserved (16% difference)(2). These results suggested polymorphism of the putative envelope protein(s) of HCV. For purposes of elucidation, cDNA clones encoding the amino-terminal portion of the putative

[1] The nucleotide sequences of clones 63, 64, and 168 have been submitted to the DDBJ/EMBL/GenBank DNA databases with accession number(s) D00689, D00690, and D00691, respectively.

0006-291X/91 $1.50

envelope protein from a cDNA library originating from plasma RNAs of 9 Japanese patients with PT-NANBH (2), were screened, sequenced, and compared. Four sequences already reported (2, 5, 6) were also examined for purposes of comparison.

MATERIALS AND METHODS

Cloning of cDNA sequences containing the HCV structural protein region. A λgt11 cDNA library was constructed from plasma RNAs from 9 patients with NANBH, using a HCV specific primer, 5'-GGGCTCGGAGTGAAGCAATA-3' (nucleotide positions 1848 to 1867 in HCV genome)(2). The cDNA library was screened using as a probe a fragment of clone 4 (nucleotide position 1718 to 2098 in HCV genome)(2). After tertiary screening, positive plaques were purified and cDNA inserts of these clones were purified and subcloned into the pTZ19R vector (United States Biochemical Corp.).

DNA sequence analysis. Recombinants were sequenced as double-stranded DNA by the dideoxy chain termination method (7), using Sequenase DNA sequencing kits (United States Biochemical Corp.). Sets of 5′ and 3′ deletions of the inserts were prepared using exonuclease III (Takara Shuzo) for sequencing (8).

Protein analysis. Hydrophilicity analysis of the deduced amino acid sequences of cloned cDNA clones was performed according to Hopp and Woods (9), using a window of 7 consecutive residues. The secondary structures of the proteins were analysed according to Chou and Fasman (10).

RESULTS AND DISCUSSION

Five cDNA clones, named clone 63, 64, 158, 168, and 171, were finally obtained. Since nucleotide sequences of clones 158 and 171 were identical to clones 63 and 3 (2), respectively, clone 63, 64, and 168 were selected for further analyses. Comparisons of the nucleotide and amino acid sequences of the 3 isolated clones, as well as the 4 reported HCV sequences (2, 5, 6), are shown in Fig. 1. The differences in the amino acid sequences are shown in Table 1.

These 7 clones differed from each other in 8.3 to 30.4% of nucleotides (8.4 to 27.9% of amino acids), except between clone 3 and 63 (Fig. 1). Clone 63 differed from clone 3 in only 1.6% of nucleotides (3.2% of amino acids). Clones 63 and 3 are likely to have originated from a single patient with PT-NANBH, for the following reasons: The region of the putative non-structural protein 5 (NS5) region (2), one of the most conserved regions between HCV genomes, showed less than a 2.5% nucleotide sequence diversity in HCV-J clones from 19 independent patients (11). In many cases, variation in part of the putative non-structural region of HCV-J genomes from a single patient with PT-NANBH is in the range from 1.0% to 2.0% differences, at the nucleotide sequence level (unpublished data). However, it cannot be

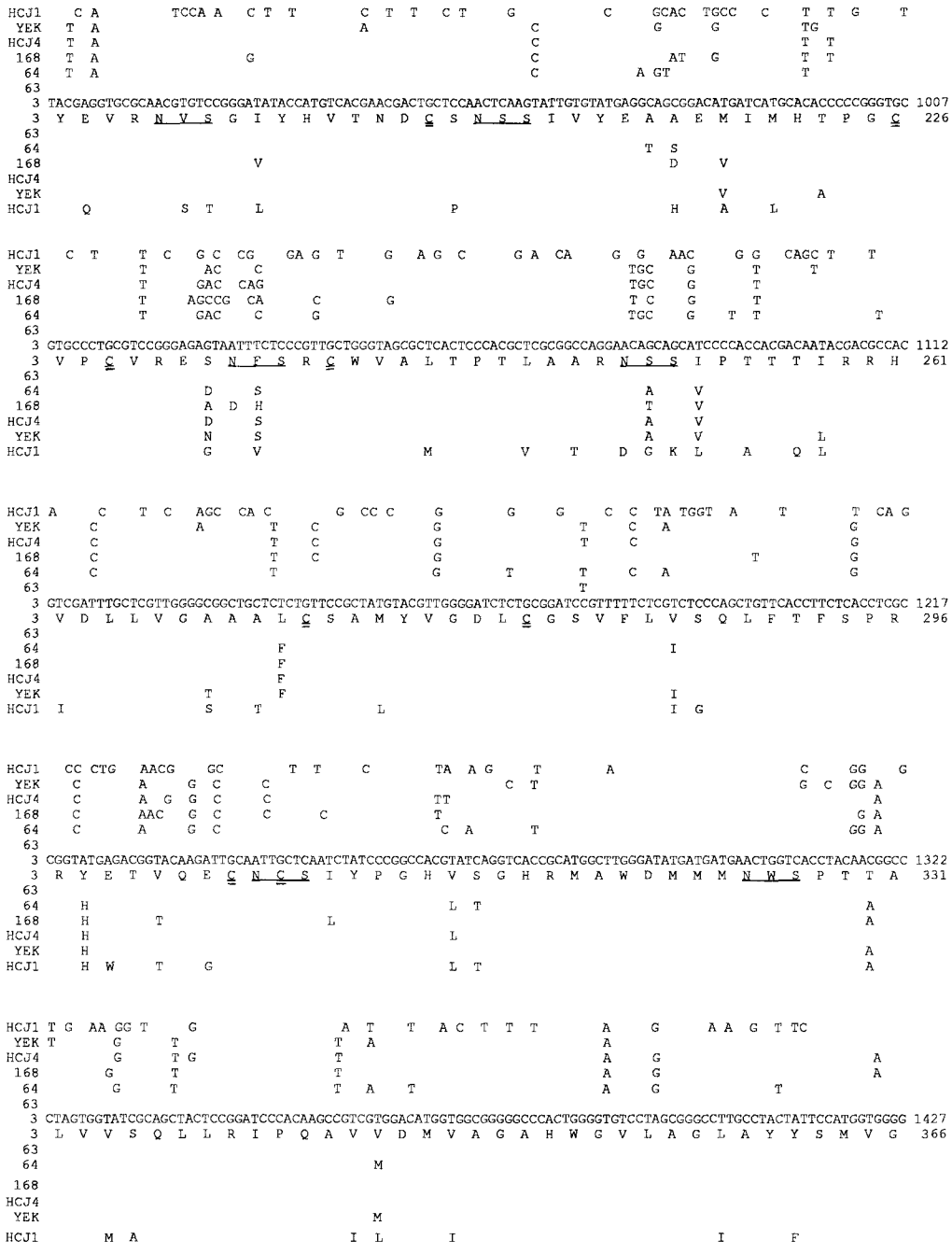

Fig. 1. Alignment of nucleotide and amino acid sequences (single-letter code) within the putative envelope region of 7 HCV genomes (clone 3 (2), 63, 64, 168, HCJ1 (6), HCJ4 (6), and tentatively named YEK (5)). Sequence of clone 3 is used as the reference. cDNA and amino acid sequences are aligned above and below the reference, respectively. Nucleotides and amino acids are according to the entire sequence of the HCV-J genome (2). Only different residues from the reference sequence are shown. Reported range of the YEK sequence in this region is indicated in parenthesis. Potential N-glycosylation sites (N-X-S/T) are underlined. Invaried 14 cysteine residues are double-underlined. Hypervariable regions 1 and 2 are boxed.

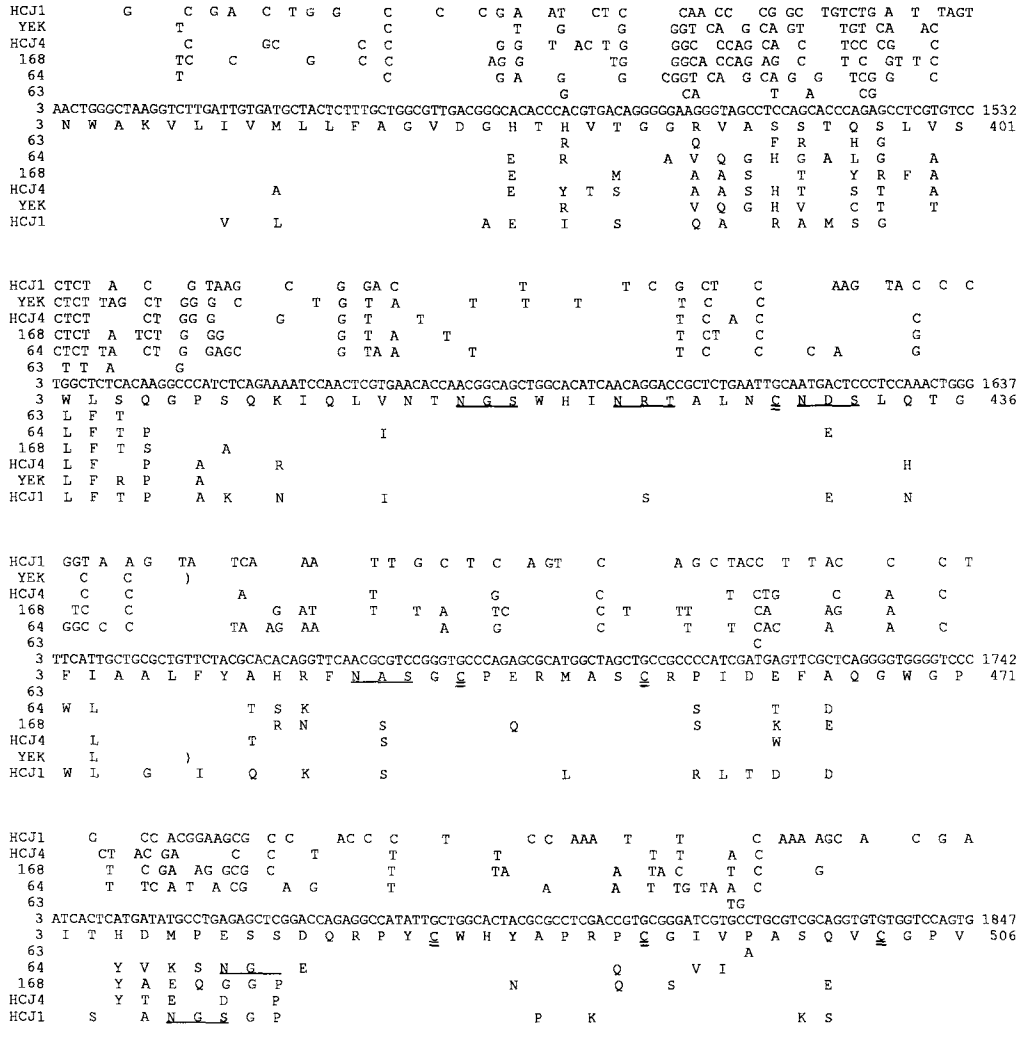

Fig. 1-Continued

excluded that the highly homologous HCV clones have infected some patients with NANBH. On the other hand, there was a high average of amino acid difference (26.4%) between HCJ1 and others (Table 1). This diversity of HCJ1 probably means that HCJ1 belongs to the HCV-US, as suggested (2, 6). Japanese hemophiliacs transfused with imported clotting factors may have been infected with HCV-US (12).

Extensive variation in putative envelope glycoproteins of HCV isolates is apparent and like that of HIV envelope proteins (13, 14). Differences among the 7 sequences appeared as single nucleotide changes and insertion and deletion were absent, in contrast to envelope genes of the HIV isolates (13, 14)(Fig. 1). About 42% of the nucleotide changes lead to amino acid substitutions, the result being numerous amino acid variations.

Table 1. Differences in amino acid sequences in the putative envelope region between seven HCV genomes[a]

clone	region[b]	63	64	168	HCJ4	YEK[c]	HCJ1
3	1	10/ 3.2[d]	49/15.6	47/14.9	35/11.1	28/11.2	88/27.9
	2	5/50.0	8/80.0	8/80.0	8/80.0	8/80.0	7/70.0
	3	0/0	6/85.7	7/100.0	5/71.4		6/85.7
63	1		45/14.3	47/14.9	35/11.1	25/10.0	84/26.7
	2		8/80.0	8/80.0	8/80.0	8/80.0	5/50.0
	3		6/58.7	7/100.0	5/71.4		6/85.7
64	1			49/15.6	41/13.0	21/ 8.4	82/26.0
	2			9/90.0	7/70.0	5/50.0	8/80.0
	3			5/71.4	6/85.7		6/85.7
168	1				40/12.7	29/11.6	87/27.6
	2				4/40.0	9/90.0	9/90.0
	3				4/57.1		4/57.1
HCJ4	1					23/ 9.2	79/25.1
	2					6/60.0	7/70.0
	3						6/58.7
YEK	1						62/24.8
	2						9/90.0

a. Based on the aligned amino acid sequences shown in Fig.1.
b. 1. Entire sequence (315 amino acids, amino acid positions from 192 to 506). 2. Hypervariable region 1 (10 amino acids, amino acid positions from 391 to 400). 3. Hypervariable region 2 (7 amino acids, amino acid positions 474 to 480).
c. Reported amino acid sequences of YEK covers 250 amino acids in the region 1 (amino acid positions from 192 to 471) and does not contain the hypervariable region 2, as shown in Fig. 1.
d. Number of amino acid differences/% diversity of amino acid sequence.

However, despite the overall extensive sequence variation, distribution in the variation is not uniform throughout this region which is clustered and interspersed with highly conserved amino acid sequences as seen in the envelope genes of HIV (13, 14)(Fig. 2). In particular, regions from amino acid positions 391 to 400 (hypervariable region 1) and from amino acid positions 474 to 480 (hypervariable region 2) showed a 40.0 to 90.0% and

57.1 to 100.0% diversity, respectively, between 7 HCV clones, except between clone 3 and 63 (Table 1). Even here, a 35.3% difference in amino acids was observed in the hypervariable region 1, although amino acid sequences in hypervariable region 2 of each clone were identical (Table 1). However, a large part of the other region of the envelope glycoprotein was highly conserved, and 14 invariant cysteine residues and most of the potential N-glycosylation sites were located in these conserved regions (Fig. 2). These features may facilitate conservation of a higher-order structure of envelope glycoproteins among the HCV isolates.

The polymorphic nature of viral sequences probably arises as a result of replication errors and is followed by a selection of the mutated population. This raises the possibility that the fidelity of the RNA-dependent RNA polymerase of HCV, possibly

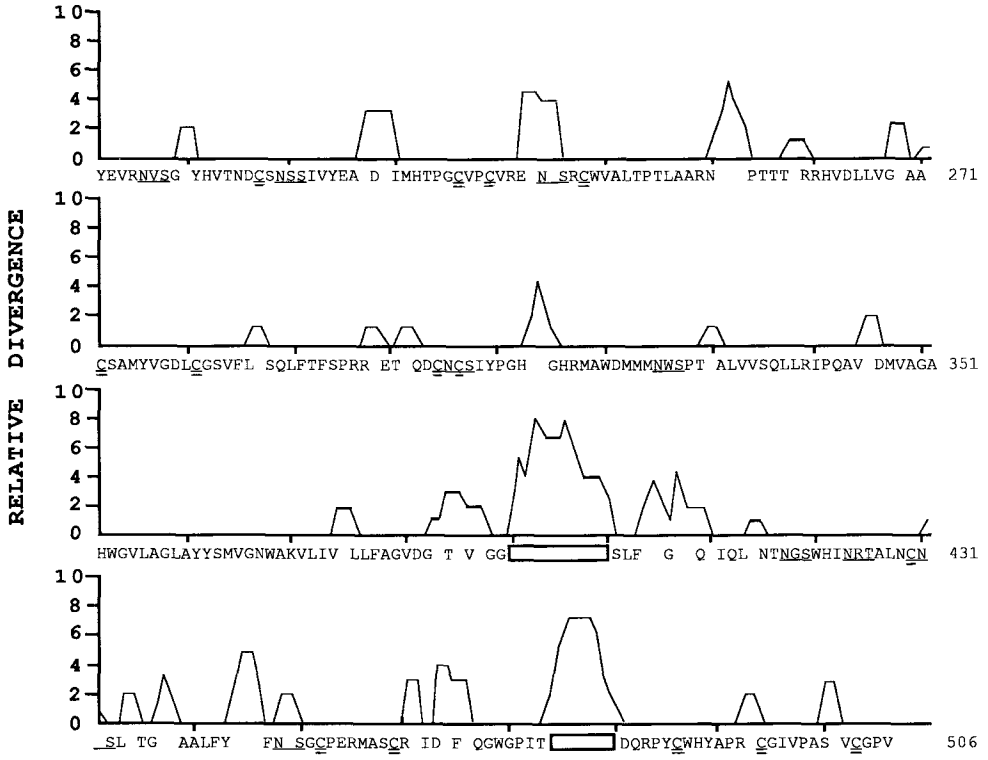

Fig. 2. Schematic diagram of relative variation, or divergence in the putative envelope region of 7 HCV-J clones. Relative degrees of variation were calculated and plotted on the ordinate from 0 (minimum divergence) to 10 (maximum divergence)(13). Single letters indicate the conserved amino acids, defined by identity in seven out of seven or six out of seven of the sequences shown beneath the abscissa. The amino acid position number is indicated in accordance with the entire sequence of HCV-J genome (2). Invaried potential N-glycosylation sites and 14 cysteine residues are underlined and double-underlined, respectively. Hypervariable regions 1 and 2 are boxed.

encoded in the NS5 region (2), is quite low, as is the case with the reverse transcriptase of HIV (15, 16). It seems highly probable that the heterogeneity of putative envelope glycoprotein serves as the antigenic variation of HCV, as indicated for the visna virus (17, 18), equine infectious anemia virus (19), and HIV (20, 21), and/or may cause changes in the viral tropism, as in the case of HIV (22), both of which make way for the persistent infection of HCV. The possibility that these highly variable residues bear no functional importance cannot be excluded. The region surrounding the hypervariable region 2 of all HCV clones was extensively hydrophilic and had a predicted β turn structure, as noted in a hydrophilicity analysis (9) and protein secondary structure analysis (10), respectively (Fig. 3). These characteristics meet criteria for probable antigenic epitopes (23, 24, 25).

As there is sequence variability in the putative envelope region, the possibility of a large pool of antigenic variants of HCV in Japan and elsewhere would need to be considered. When developing effective therapeutic strategies for HCV infection, account must be taken of this and related factors.

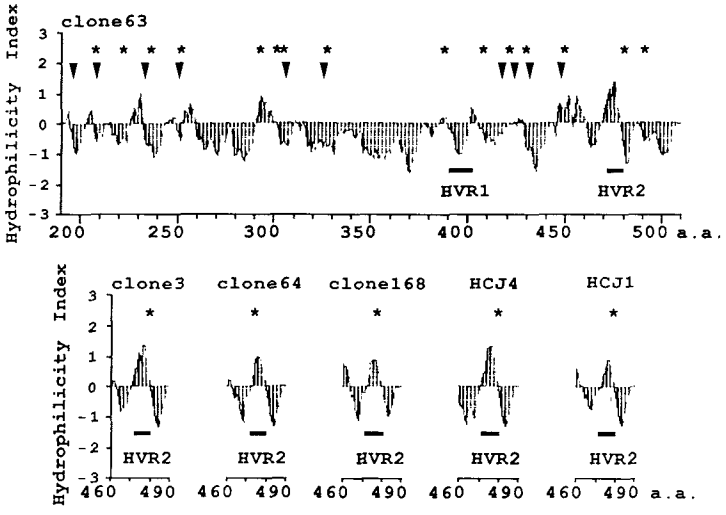

Fig. 3. Structural features of the putative envelope proteins of six HCV clones. Hydrophilicity profile of the protein of clone 63 is shown. Results of the region from residues 192 to 510 of clone 63 and regions surrounding HVR2 (residues 460 to 490) of clone 3, 64, 168, HCJ4, and HCJ1 are also shown. Hydrophilicity profiles of each clone showed a similar pattern. Locations of predicted β turn structures are indicated by asterisks. Potential N-glycosylation sites are indicated by arrowheads. Hypervariable region 1 (HVR1) and 2 (HVR2) are indicated by thick bars.

ACKNOWLEDGMENTS: We thank M. Ohara for comments. This work was supported in part by a Grant in-Aid from the Ministry of Health and Welfare for a Comprehensive 10-Year Strategy for Cancer Control and a Grant in-Aid from the Ministry of Science and Culture, Japan. M.H. is a receipient of a Research Resident fellowship from the Foundation for Promotion of Cancer Research, Japan.

REFERENCES

1. Choo, Q.-L., Kou, G., Weiner, A. J., Overby, L. R., Bradley, D. W., and Houghton, M., (1989) Science **244**, 359-362.
2. Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T., and Shimotohno, K., (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 9524-9528.
3. 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.
4. Miller, R. H., and Purcell, R. H., (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 2057-2061.
5. Takeuchi, K., Kubo, Y., Boonmar, S., Watanabe, Y., Katayama, T., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I., and Miyamura, T., (1990) Nucleic Acids Res. **18**, 4626-4626.
6. Okamoto, H., Okada, S., Sugiyama, Y., Yotsumoto, S., Tanaka, T., Yoshizawa, H., Tsuda, F., Miyakawa, Y., and Mayumi, M., (1990) Japan J. Exp. Med. **60**, 167-177.
7. Sanger, F., Nicklen, S., and Coulson, A. R., (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463-5467.
8. Henikoff, S., (1984) Gene, **23**, 351-359.
9. Hopp, T. P., and Woods, K. R., (1981) Proc. Natl. Acad. Sci. U.S.A. **81**, 3824-3828.
10. Chou, P. Y., and Fasman, G. D., (1974) Biochemistry **13**, 222-245.
11. Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., and Shimotohno, K., Mol. Biol. Med. in the press.
12. Hijikata, M., Kato, N., Mori, S., Ootsuyama, Y., Nakagawa, M., Sugimura, T., Ohkoshi, S., Kojima, H., Meguro, T., Taki, M., Takayama, S., Yamada, K., and Shimotohno K., (1990) Japan J. Cancer Res. **81**, 1191-1197.
13. Starcich, B. R., Hahn, B. H., Shaw, G. M., McNeely, P. D., Modrow, S., Wolf, H., Parks, E. S., Parks, W. P., Josephs, F., Gallo, R. C., Wong-Staal, F., Cell **45**, 637-648 (1986).
14. Alizon, M., Wain-Hobson, S., Montagnier, L., and Sonigo, P., (1986) Cell **46**, 63-74.
15. Preston, B. D., Poiesz, B. J., and Loeb, L. A., (1988) Science, **242**, 1168-1171.
16. Roberts, J. D., Bebenek, K., and Kunkel, T. A., (1988) Science, **242**, 1171-1173.
17. Scott, J. V., Stowring, L., Haase, A. T., Narayan, O., and Vigne, R., (1979) Cell **18**, 321-327.
18. Clements, J. E., Narayan, O., Griffin, D. E., and Johnson, R. T., Proc. Natl. Acad. Sci. U.S.A. **77**, 4454-4458 (1980).
19. Montelaro, R. C., Parekh, B., Orrego, A., and Issel, C. J., (1984) J. Biol. Chem. **259**, 10539-10544.
20. Modrow, S., Hahn, B. H., Shaw, G. M., Gallo, R. C., Wong-Staal, F., and Wolf, H., (1987) J. Virol. **61**, 570-578.

21. Reitz, Jr., M. S., Wilson, C., Naugle, C., Gallo, R. C., Robert-Guroff, M., (1988) Cell **54**, 57-63.
22. Koyanagi, Y., Miles, S., Mitsuyasu, R. T., Merrill, J. E., Vinters, H. V., Chen, I. S. Y., (1987) Science **236**, 819-822.
23. Cohen, G. H., Dietzschold, B., Ponce DE Leon, M., Long, D., Golub, E., Varrichio, A., Pereira, L., and Eisenberg, R. J., (1984) J. Virol. **49**, 102-108.
24. Eisenberg, R. J., Long, D., Ponce DE Leon, M., Matthews, J. T., Spear, P. G., Gibson, M. G., Laskey, L. A., Berman, P., Golub, E., and Cohen, G. H., (1985) J. Virol. **53**, 634-644.
25. Modrow, S., and Wolf, H., (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 5703-5707.