JOURNAL OF VIROLOGY, May 1992, p. 3225–3229
0022-538X/92/053225-05$02.00/0

Vol. 66, No. 5

Hepatitis C Virus (HCV) Circulates as a Population of Different but Closely Related Genomes: Quasispecies Nature of HCV Genome Distribution

MARIA MARTELL,[1] JUAN I. ESTEBAN,[1]* JOSEP QUER,[1] JOAN GENESCÀ,[1] AMY WEINER,[2] RAFAEL ESTEBAN,[1] JAUME GUARDIA,[1] AND JORDI GÓMEZ[1]

Liver Unit, Department of Medicine, Hospital General Universitari Vall d'Hebron, 08035 Barcelona, Spain,[1] and Chiron Corporation, Emeryville, California 94608-2916[2]

Received 22 November 1991/Accepted 20 January 1992

Sequencing of multiple recombinant clones generated from polymerase chain reaction-amplified products demonstrated that the degree of heterogeneity of two well-conserved regions of the hepatitis C virus (HCV) genome within individual plasma samples from a single patient was consistent with a quasispecies structure of HCV genomic RNA. About half of circulating RNA molecules were identical, while the remaining consisted of a spectrum of mutants differing from each other in one to four nucleotides. Mutant sequence diversity ranged from silent mutations to appearance of in-frame stop codons and included both conservative and nonconservative amino acid substitutions. From the relative proportion of essentially defective sequences, we estimated that most circulating particles should contain defective genomes. These observations might have important implications in the physiopathology of HCV infection and underline the need for a population-based approach when one is analyzing HCV genomes.

Hepatitis C virus (HCV), a 10-kb positive-stranded RNA virus, has recently been shown to be the major causative agent of parenterally transmitted non-A, non-B hepatitis (1, 5, 11, 21). Despite little overall primary sequence identity, the genetic organization of HCV has been shown to be similar to those of flaviviruses and pestiviruses (6, 26, 33). Significant genetic heterogeneity has been reported among isolates from different geographic areas (17) and within single isolates from the same individual (17, 20, 33–35). Comparative sequence analysis of the different HCV isolates has shown, however, that the degree of variability is unevenly distributed throughout the HCV genome, with some very well conserved regions (6, 13, 17, 33–35) and some highly variable genes (17, 20, 33–35). In addition, the rate of fixation of mutations of the HCV genome has been estimated to be similar to that of other RNA viruses (approximately 10^{-3} to 10^{-4} base substitutions per genome site per year), and evidence suggesting that the HCV genome may rapidly evolve in vivo, with different rates of evolution for different regions of the viral genome, has been provided (27).

Since the pioneering studies by Batschelet et al. (3) providing evidence that RNA virus heterogeneity is a consequence of high error rates in RNA replication, data have accumulated which suggest that most RNA viruses consist of a heterogeneous mixture of circulating related genomes containing a master (most frequently represented) sequence and a large spectrum of mutants, a genomic distribution referred to as quasispecies (10, 32). This quasispecies model of mixed RNA virus populations implies a significant adaptation advantage because the simultaneous presence of multiple variant genomes (and the high rate of generation of new variants) allows for the rapid selection of the mutant(s) with better fitness for any new environmental condition. On the other hand, a quasispecies will remain in stable equilibrium with little evolution of its consensus or master sequence as long as conditions are unchanged (7, 15, 18, 22).

Many important biological implications predicted by the quasispecies model (8) have been found in several virus systems displaying such genomic structure, including vaccination failure through selection of neutralizing antibody escape mutants (19), establishment of persistent infection by selection of neutralizing antibody or cytotoxic T-lymphocyte escape mutants (23, 28, 30) or by generation of defective interfering particles (14), resistance to antiviral agents (29), and changes in cell tropism or virulence (9, 12).

In an attempt to define the degree of heterogeneity among circulating HCV RNA molecules within individual isolates, we analyzed a series of four sequential HCV isolates from an individual with transfusion-associated HCV and human immunodeficiency virus coinfection. Polymerase chain reaction (PCR)-amplified products of cDNA corresponding to fragments of the 5′ untranslated region (5′UTR) and nonstructural region 3 (NS3) of the HCV genome were cloned into a bacterial vector, and 6 to 20 recombinant clones from each sample were sequenced. A quasispecies distribution of sequences was observed in both genomic regions and in all sequential samples.

HCV RNA extracted from 75 μl of plasma by the acid guanidinium thiocyanate-phenol-chloroform method (4) was reverse transcribed into cDNA and PCR amplified in a single tube reaction (RT-PCR; kit N 808-007, Perkin Elmer) for 35 cycles (5 cycles of 94°C for 2 min, 50°C for 2 min, and 72°C for 3 min; 30 cycles of 94°C for 1.5 min, 60°C for 2 min, and 72°C for 3 min), using specific oligonucleotide primers of the 5′UTR (13) and NS3/NS4 regions (16). As represented in Fig. 1, the amplified products were 237 and 584 bp long, and their specificity was confirmed by Southern hybridization with ^{32}P probes. 5′UTR and NS3/NS4 products were cleaved with *Not*I and *Sac*I-*Nco*I, respectively, yielding restriction products of 177 and 240 bp, respectively, which were subsequently cloned into pSL1190 vector (Pharmacia-

* Corresponding author.

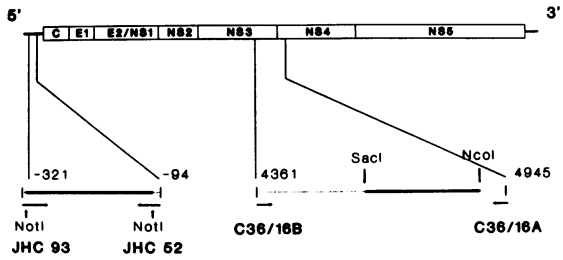

FIG. 1. Organization of the HCV genome. Putative structural genome regions: core (C), envelope (E1 and E2/NS1), nonstructural protein coding regions (NS1 to NS5). Amplified 5'UTR and NS3/NS4 products are depicted below, with their respective ends indicated by their corresponding nucleotide positions from the HCV-1 prototype (6, 13). Horizontal solid lines represent restriction fragments generated from each product for subsequent cloning and sequencing. Oligonucleotide primers JHC 52 (5'-AGTCTTGCGGC CGCACGCGCCCAAATC), JHC 93 5'-TTCGCGGCCGCACTC CACCATGAATCACTCCCC), C36/16A (5'-GCATGTCATGATGT AT), and C36/16B (5'-GCAATACGTGTGTCAC) used for cDNA synthesis and PCR amplification were synthesized according to the HCV-1 prototype sequence (13, 16).

LKB). Positive recombinant clones were subcloned into M13mp19. Six to twenty independent recombinant subclones were sequenced by the dideoxy-mediated chain termination method (31). The Maxam-Gilbert chemical degradation method (24) was used to resolve sequencing ambiguities.

When the sequences of 20 independent NS3 clones from the sample obtained 2 months postransfusion were compared (Table 1 and Fig. 2a), 9 sequences (45%) were identical (master sequence) and 11 (55%) differed from each other and from the master sequence in one to four nucleotides. Similarly, of the 20 5'UTR clones from the same sample, 12 sequences (60%) were identical and 8 (40%) contained single, nonrepetitive base substitutions (Table 1 and Fig. 2b). The predicted amino acid sequence encoded by the master NS3 sequence was identical to that of the corresponding HCV-1 prototype (16). However, of the 25 base substitutions observed (Fig. 2a) in the 11 mutant sequences, only 5 (20%), occurring in the third codon position, were silent mutations. Of the 20 nucleotide changes observed in the first and second codon positions, 6 (24%) led to conservative amino acid changes, 11 (44%) introduced drastic amino acid changes (five Ile → Thr, three Leu → Arg, one Asp → Val, one Met → Thr, and one Tyr → His), and 3 (12%) led to the appearance of in-frame stop codons in 2 of the 20 sequences. Although the functional implications of single amino acid substitutions (whether conservative or nonconservative) are unknown, it is obvious that mutants containing in-frame stop codons are essentially defective. It must be borne in mind that the sequenced fragment of the NS3 region encodes only 80 of a total of 3,011 residues (i.e., 2.8% of the coding capacity of the HCV genome). Assuming that the NS3 region is representative of the HCV open reading frame, the finding that 10% of these sequences include stop codons indicates that most circulating particle populations must contain defective genomes.

TABLE 1. Sequence complexity of HCV quasispecies

Sample	Postransfusion mo	Genomic region cDNA	No. of clones sequenced	No. of identical nucleic acid sequences	No. of different nucleic acid sequences
VL-1	2	NS3	20	9	11
		5'UTR	20	12	8
VL-2	3	5'UTR	10	7	3
VL-3	6	5'UTR	10	7	3
VL-4	29	5'UTR	6	4	2

To ensure that the observed heterogeneity was not due to nucleotide misincorporations introduced by the reverse transcriptase or the *Taq* DNA polymerase during the amplification reaction, two control experiments were carried out. First, one of the NS3 recombinant clones of known sequence was in vitro transcribed with SP6 RNA polymerase (Boehringer Mannheim). The RNA transcript was subjected to the original RT-PCR amplification procedure under identical conditions, and the amplified product was subcloned. Sequence analysis of five independent clones showed absolute identity with the parental clone from which the RNA transcript had been obtained. In a second experiment, the cDNA insert of one of the sequenced 5'UTR recombinant clones was subjected to 35 additional cycles of PCR amplification and subcloned. Within 23 such clones (5,451 bases sequenced), only one nucleotide change (A → G) was detected in one of the clones. Therefore, it seems that the observed heterogeneity was not an artifact, although some of the base substitutions might have been introduced during the amplification procedure.

With such a low level of misincorporation noise, the finding that two distinct and well-conserved regions of the viral genome (17) were composed of a mixture of cocirculating related genomes distributed as a predominant or master sequence and a spectrum of mutants (representing altogether about half of the genomic population) provides unequivocal evidence of the quasispecies structure of the HCV genome. Additional sequencing of 6 to 10 independent 5'UTR recombinant clones, generated from subsequent samples obtained later within the acute phase (at 3 months postransfusion) and during the chronic phase (at 6 and 29 months postransfusion), showed similar distribution and complexity of the quasispecies for that region again with randomly distributed single nonrepetitive base substitutions in each of the mutants around a conserved master sequence (Table 1). This finding demonstrates that the observed quasispecies distribution was not a phenomenon restricted to the acute phase of the infection.

Demonstration of the quasispecies structure of the HCV genome implies the need for a population-based approach when one is analyzing HCV genomes. Accordingly, the average or consensus sequence of each nucleic acid region should always be carefully defined when one is trying to determine the complete genomic sequence of a given isolate. Otherwise, reconstruction of the HCV genome by randomly sequencing overlapping clones of PCR products obtained from plasma pools of different subjects (or even from a single patient) would generate an artifactual genomic mosaicism.

Given our limited knowledge of the biological functions of most HCV genome regions and of their encoded products, besides the potential problems for vaccine development, it is not possible at present to anticipate which of the several relevant implications predicted by the quasispecies model will be applicable to HCV. One may speculate, however, on two possible mechanisms which might play a role in the two most striking features of HCV infection: its high tendency toward viral persistence and the characteristic smoldering and fluctuating course of chronic hepatitis C. On the one hand, the high degree of heterogeneity of the envelope-

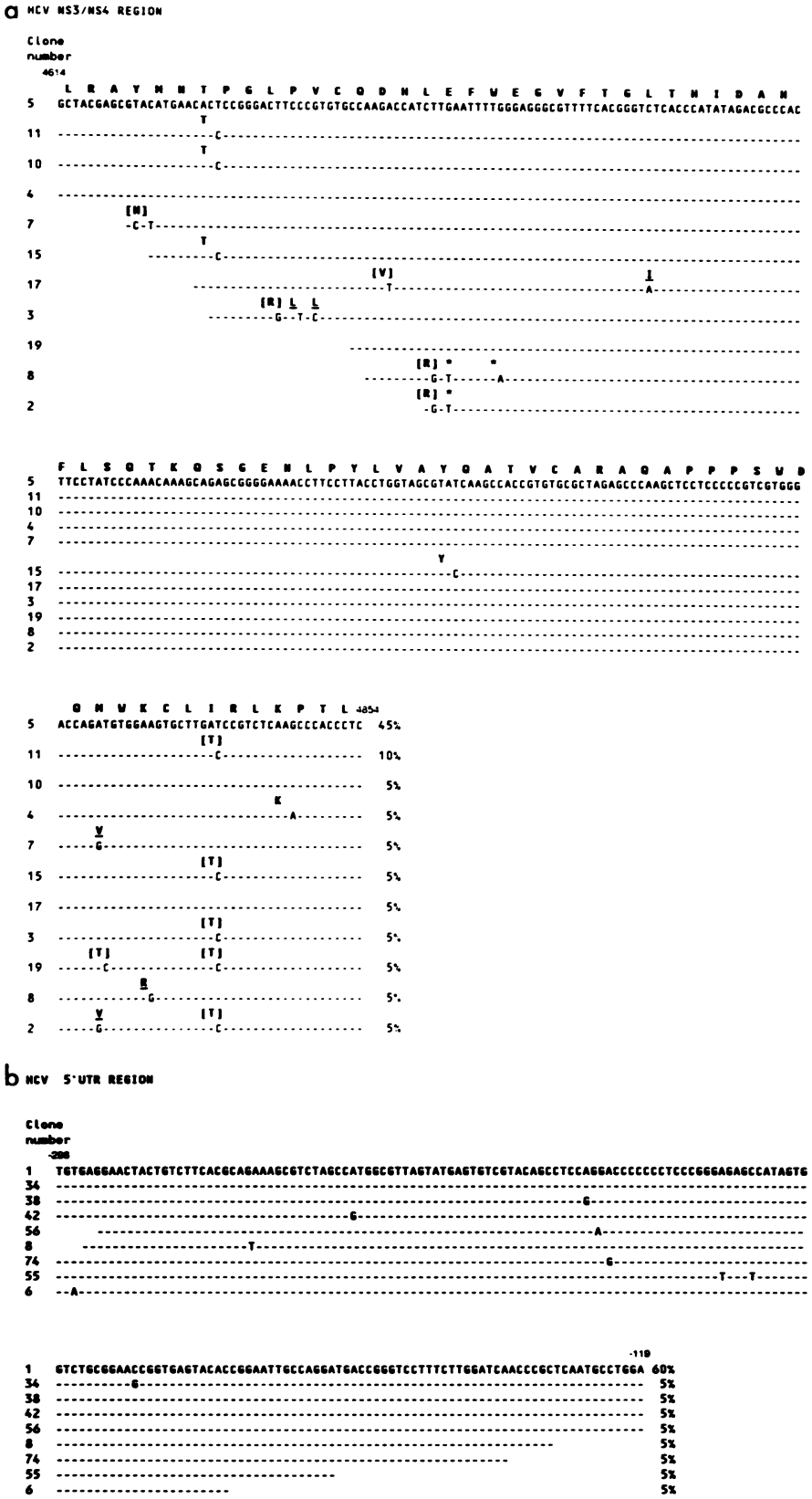

FIG. 2. Nucleotide and amino acid sequences of 20 independent NS3/NS4 (a) and 5'UTR (b) clones from sample VL-1 (see Table 1). Relative frequencies are shown on the right. Mutant sequences have been aligned to the master (most frequently represented) sequence for each region. The one-letter code has been used for protein sequences. Dots indicate sequence identity. Only nucleotides which differ from the master sequence are shown. Asterisks indicate in-frame stop codons. Conservative amino acid substitutions are underlined; those in brackets correspond to nonconservative changes.

encoding regions and its potential for rapid evolution (as anticipated by the quasispecies model) might lead to rapid antigenic changes in structural proteins likely to contain neutralizing antibody and/or cytotoxic T-cell-specific epitopes and hence provide the virus the opportunity to avoid the previous response of the host immune system. On the other hand, if our assumption that most circulating viral particles contain defective genomes is true, as it has been postulated for other viruses (25), this would provide an alternative or supplementary mechanism for HCV persistence. Indeed, there is a wealth of data indicating how defective RNA genomes may restrict viral replication in vivo, modulate the clinical expression of the disease, and lead to the establishment of persistent viral infection (2, 14, 18). Whether, or the extent to which, these mechanisms play a role in the natural history of HCV infection remains to be determined. The widespread nature of chronic HCV infection and its long-term clinical consequences, however, warrant further studies on any potential mechanism of HCV persistence.

We are indebted to E. Domingo (Centro de Biología Molecular, CSIC, Universidad Autónoma de Madrid) for continuous encouragement and support and useful discussion of the manuscript. We also thank P. Puigdomenech (Centro de Investigación y Desarrollo, CSIC, Barcelona, Spain) for useful scientific and technical support. We thank J. H. Han (Chiron Corporation, Emeryville, Calif.) for kindly providing primers JHC 93 and JHC 52.

This work was supported in part by grants SAL 90/0449 from the Comisión Interministerial de Ciencia y Tecnología and 91/0298 from the Fondo de Investigaciones Sanitarias.

REFERENCES

1. **Alter, H. J., R. H. Purcell, J. W. K. Shih, J. C. Melpolder, M. Houghton, Q. L. Choo, and G. Kuo.** 1989. Detection of antibody to hepatitis C virus in prospectively followed transfusion-recipients with acute and chronic non-A non-B hepatitis. N. Engl. J. Med. **321:**1494–1500.
2. **Barret, A. D. T., and N. J. Dimmock.** 1986. Defective interfering viruses and infections of animals. Curr. Top. Microbiol. Immunol. **128:**55–84.
3. **Batschelet, E., E. Domingo, and C. Weissmann.** 1976. The proportion of revertant and mutant phage in a growing population as a function of mutation and growth rate. Gene **1:**27–32.
4. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156–159.
5. **Choo, Q. L., G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradley, and M. Houghton.** 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science **244:**359–362.
6. **Choo, Q. L., K. H. Richman, J. H. Han, K. Berger, C. Lee, D. Dong, C. Gallegos, D. Coit, A. Medina-Selby, P. J. Bar, A. J. Weiner, D. W. Bradley, G. Kuo, and M. Houghton.** 1991. Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA **88:**2451–2455.
7. **Domingo, E.** 1989. RNA virus evolution and the control of viral disease. Prog. Drug Res. **33:**93–133.
8. **Domingo, E., E. Martínez-Salas, F. Sobrino, J. C. de la Torre, A. Portela, J. Ortín, C. López-Galindez, P. Pérez-Breña, N. Villanueva, R. Nájera, S. VandePol, N. DePolo, and J. Holland.** 1985. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. Gene **40:**1–8.
9. **Domingo, E., M. G. Mateu, M. A. Martínez, J. Dopazo, A. Moya, and F. Sobrino.** 1990. Genetic variability and antigenic diversity of foot-and-mouth disease virus, p. 233–266. *In* E. Kurstak, R. G. Marusyk, F. A. Murphy, and M. H. V. Van Regenmortel (ed.). Applied virology research, vol. 2. Plenum Publishing Corp., New York.
10. **Domingo, E., D. Sabo, T. Taniguchi, and C. Weissmann.** 1978. Nucleotide sequence heterogeneity of an RNA phage population. Cell **13:**735–744.
11. **Esteban, J. I., R. Esteban, L. Viladomiu, J. C. López-Talavera, A. González, J. M. Hernández, M. Roget, J. Genesca, M. Buti, G. Kuo, Q. L. Choo, M. Houghton, and J. Guardia.** 1989. Hepatitis C virus antibodies among risk groups in Spain. Lancet **ii:**294–297.
12. **Evans, D. M. A., G. Dunn, P. D. Minor, G. C. Schild, A. J. Cann, G. Stanway, J. W. Almond, K. Currey, and J. V. Maizel, Jr.** 1985. Increased neurovirulence associated with single nucleotide change in a noncoding region of the Sabin type 3 poliovaccine genome. Nature (London) **314:**548–550.
13. **Han, J. H., V. Shyamala, K. H. Richman, M. J. Brauer, B. Irvine, M. S. Urdea, P. Tekamp-Olson, G. Kuo, Q. L. Choo, and M. Houghton.** 1991. 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. Proc. Natl. Acad. Sci. USA **88:**1711–1715.
14. **Holland, J. J.** 1991. Defective viral genomes, p. 151–165. *In* B. N. Fields and D. N. Knipe (ed.), Fundamental virology. Raven Press, New York.
15. **Holland, J. J., J. C. de la Torre, D. K. Clark, and E. Duarte.** 1991. Quantitation of relative fitness and great adaptability of clonal populations of RNA viruses. J. Virol. **65:**2960–2967.
16. **Houghton, M., Q. L. Choo, and G. Kuo.** March 1989. European patent application 88 310 922.5; publication 318.216.
17. **Houghton, M., A. J. Weiner, J. H. Han, G. Kuo, and Q. L. Choo.** 1991. Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. Hepatology **14:**381–388.
18. **Huang, A. S.** 1988. Modulation of viral disease processes by defective interfering particles, p. 195–208. *In* E. Domingo, J. J. Holland, and P. Ahlquist (ed.). RNA genetics, vol. 3. CRC Press, Boca Raton, Fla.
19. **Kew, O. M., and B. K. Nottay.** 1984. Evolution of the oral polio vaccine strains in humans occurs by both mutations and intramolecular recombination, p. 357–367. *In* R. M. Chanock and R. A. Lerner (ed.), Modern approaches to vaccines. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Kremsdorf, D., C. Porchon, J. P. Kim, G. R. Reyes, and C. Bréchot.** 1991. Partial nucleotide sequence analysis of a French hepatitis C virus: implications for HCV genetic variability in the E2/NS1 protein. J. Gen. Virol. **72:**2557–2561.
21. **Kuo, G., Q. L. Choo, H. J. Alter, G. L. Gitnick, A. G. Redeker, R. H. Purcell, T. Miyamura, J. L. Dienstag, M. J. Alter, C. E. Stevens, G. E. Tegtmeier, F. Bonino, M. Colombo, W. S. Lee, C. Kuo, K. Berger, J. R. Schuster, L. R. Overby, D. W. Bradley, and M. Houghton.** 1989. An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. Science **244:**362–364.
22. **Martínez, M. A., C. Carrillo, F. González-Candelas, A. Moya, E. Domingo, and F. Sobrino.** 1991. Fitness alteration of foot-and-mouth disease virus mutants: measurement of adaptability of viral quasispecies. J. Virol. **65:**3954–3957.
23. **Mateu, M. G., M. A. Martínez, E. Rocha, D. Andreu, J. Parejo, E. Giralt, F. Sobrino, and E. Domingo.** 1989. Implications of a quasispecies genome structure: effect of frequent naturally occurring amino acid substitutions on the antigenicity of foot-and-mouth disease virus. Proc. Natl. Acad. Sci. USA **86:**5883–5887.
24. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74:**560–564.
25. **Meyerhans, A., R. Cheynier, J. Albert, M. Seth, S. Kwok, J. Sninsky, L. Morfeldt-Manson, B. Asjö, and S. Wain-Hobson.** 1989. Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. Cell **58:**901–910.
26. **Miller, R. H., and R. H. Purcell.** 1990. Hepatitis C virus shares amino acid sequence similarity with pestivirus and flavivirus as well as members of two plant virus supergroups. Proc. Natl. Acad. Sci. USA **87:**2057–2061.
27. **Ogata, N., H. J. Alter, R. H. Miller, and R. H. Purcell.** 1991. Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. Proc. Natl. Acad. Sci. USA **88:**3392–3396.

28. **Parry, N., G. Fox, D. Rowlands, F. Brown, E. Fry, R. Acharya, D. Logan, and D. Stuart.** 1990. Structural and serological evidence for a novel mechanism of antigenic variation in foot-and-mouth disease virus. Nature (London) **347**:569–572.
29. **Remington, K. M., B. Chesebro, K. Wehrly, N. C. Pedersen, and T. W. North.** 1991. Mutants of feline immunodeficiency virus resistant to 3′-azido-3′-deoxythymidine. J. Virol. **65**:308–312.
30. **Salvato, M., P. Borrow, E. Shimomaye, and M. B. A. Oldstone.** 1991. Molecular basis of viral persistence: a single amino acid change in the glycoprotein of lymphocytic choriomeningitis virus is associated with suppression of the antiviral cytotoxic T-lymphocyte response and establishment of persistence. J. Virol. **65**:1863–1869.
31. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
32. **Steinhauer, D. A., and J. J. Holland.** 1987. Rapid evolution of RNA viruses. Annu. Rev. Microbiol. **41**:409–433.
33. **Takamizawa, A., C. Mori, I. Fuke, S. Manabe, S. Murakami, J. Fujita, E. Onishi, T. Andoh, I. Yoshida, and H. Okayama.** 1991. Structure and organization of the hepatitis C virus genome isolated from human carriers. J. Virol. **65**:1105–1113.
34. **Takeuchi, K., Y. Kubo, S. Boonmar, Y. Watanabe, T. Katayama, Q. L. Choo, G. Kuo, M. Houghton, I. Saito, and T. Miyamura.** 1990. The putative nucleocapsid and envelope protein genes of hepatitis C virus determined by comparison of the nucleotide sequences of two isolates derived from an experimentally infected chimpanzee and healthy human carriers. J. Gen. Virol. **71**:3027–3033.
35. **Weiner, A. J., M. J. Brauer, J. Rosenblatt, K. H. Richman, J. Tung, K. Crawford, F. Bonino, G. Saracco, Q. L. Choo, M. Houghton, and J. H. Han.** 1991. Variable and hypervariable domains are found in the regions of HCV corresponding to the Flavivirus envelope and NS1 proteins and Pestivirus envelope glycoproteins. Virology **180**:842–848.