RNA (2001), *7*:710–721. Cambridge University Press. Printed in the USA.

Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame

JOSÉ L. WALEWSKI,[1] TOBY R. KELLER,[1] DECHERD D. STUMP,[1] and ANDREA D. BRANCH[1,2]

[1]Division of Liver Diseases, Department of Medicine, Mount Sinai School of Medicine, New York, New York 10029, USA
[2]Recanati/Miller Transplantation Institute, Mount Sinai School of Medicine, New York, New York 10029, USA

ABSTRACT

Many viruses have overlapping genes and/or regions in which a nucleic acid signal is embedded in a coding sequence. To search for dual-use regions in the hepatitis C virus (HCV), we developed a facile computer-based sequence analysis method to map dual-use regions in coding sequences. Eight diverse full-length HCV RNA and polyprotein sequences were aligned and analyzed. A cluster of unusually conserved synonymous codons was found in the core-encoding region, indicating a potential overlapping open reading frame (ORF). Four peptides (A1, A2, A3, and A4) representing this alternate reading frame protein (ARFP), two others from the HCV core protein, and one from bovine serum albumin (BSA) were conjugated to BSA and used in western blots to test sera for specific antibodies from 100 chronic HCV patients, 44 healthy controls, and 60 patients with non-HCV liver disease. At a 1:20,000 dilution, specific IgGs to three of the four ARFP peptides were detected in chronic HCV sera. Reactivity to either the A1 or A3 peptides (both ARFP derived) was significantly associated with chronic HCV infection, when compared to non-HCV liver disease serum samples (10/100 versus 1/60; $p < 0.025$). Antibodies to A4 were not detected in any serum sample. Our western blot assays confirmed the presence of specific antibodies to a new HCV antigen encoded, at least in part, in an alternate reading frame (ARF) overlapping the core-encoding region. Because this novel HCV protein stimulates specific immune responses, it has potential value in diagnostic tests and as a component of vaccines. This protein is predicted to be highly basic and may play a role in HCV replication, pathogenesis, and carcinogenesis.

Keywords: alternate reading frame protein; hepatitis C virus; novel HCV antigen; RNA elements; western blot

INTRODUCTION

Hepatitis C virus (HCV), a plus sense RNA virus identified in 1989 (Choo et al., 1989), is estimated to chronically infect roughly 4,000,000 people in the United States (National Institutes of Health, 1997), often with serious consequences (for a review, see Branch et al., 2000). Because HCV poses a public health threat, it is important to identify all HCV RNA structural elements and expressed polypeptides to define all potential diagnostic markers, vaccine components, and targets for pharmaceutical agents. At the moment, HCV RNA is known to contain a single large open reading frame (ORF), about 9,000 nt in length, encoding a single polyprotein that is the source of 10 viral proteins: the core, E1, E2, P7, NS2, NS3, NS4a, NS4b, NS5a, and NS5b (Rice, 1996). This ORF is flanked by about 350 nt at its 5′ end and about 220 at its 3′ end. Although the full range of the functions provided by the flanking sequences is not yet clear, terminal structures are likely to play a role in replication, and the 5′ flanking sequence forms part of an internal ribosome entry site (IRES) that promotes the initiation of HCV polyprotein synthesis (Brown et al., 1992; Tsukiyama-Kohara et al., 1992; Reynolds et al., 1995; Lu & Wimmer, 1996).

As exemplified by the hepatitis B virus, viral genomes often contain overlapping genes. Thus, HCV RNA may contain regions where the main ORF is overlapped by another gene or by an RNA structural element. To seek these multifunctional regions, we carried out comparative sequence analysis on diverse HCV sequences retrieved from GenBank (Benson et al., 1996), locating synonymous codons in the standard HCV ORF in which the third position nucleotides are much more conserved than chance alone would pre-

Reprint requests to: Dr. Andrea D. Branch, One Gustave L. Levy Place, New York, New York 10029, USA; e-mail: ab8@doc.mssm.edu.

dict. This unusual third-base conservation is likely to occur in regions that have novel functions in addition to their known coding function (see Materials and Methods).

Previous studies identified some of the regions of HCV RNA that have unusual nucleotide conservation (Ina et al., 1994; Smith & Simmonds, 1997) and, in particular, they revealed that the RNA sequence of the core-encoding region is more conserved than would be necessary to maintain the observed level of conservation of the core protein. Ina and colleagues (Ina et al., 1994) suggested that an overlapping gene might constrain the sequence and proposed that translation of a second ORF might be initiated at the GUG codon at bases −41 to −39 and continue into the coding region. However, the reading frame that contains this GUG has an in-frame stop codon (bases +2 to +4) that terminates it at the start of the main ORF. This stop codon is present in all reported full-length core sequences; its presence reduces the likelihood that the GUG functions as the start codon for a protein that extends into the core-encoding region. Smith and Simmonds (1997) concluded that the reduced frequency of synonymous substitutions "cannot be accounted for by additional coding restraints" (p. 240). Recent studies indicate that the initial segment of the core-encoding region of the main HCV ORF contains features necessary for the efficient functioning of the IRES (Honda et al., 1996; Reynolds et al., 1996) providing a partial explanation for the unusual sequence conservation. However, the results of these studies do not account for the full extent of the sequence conservation. The current studies were performed to seek evidence of a novel antigen encoded in an alternate reading frame. This alternate reading frame protein (ARFP) is predicted to be highly basic and may play a role in HCV replication, pathogenesis, and carcinogenesis.

Regardless of whether this antigen is required for HCV replication, it could have clinical significance as an immunogen. "Cryptic" epitopes encoded in alternative reading frames and expressed as a result of a ribosomal scanning process stimulate cytotoxic T lymphocyte responses in the influenza virus system (Bullock & Eisenlohr, 1996). The general importance of "out of frame" proteins in the induction of immune responses has recently received attention as a result of studies carried out by Malarkannan et al. (1999). These investigators propose that cryptic peptides translated from noncoding mRNA sequences enhance immune surveillance by enormously increasing the variety of peptides presented on MHC class I molecules.

A preliminary report of the sequence analysis and development of the western blot assay described here was presented at the 1999 annual meeting of the American Association for the Study of Liver Diseases (Walewski et al., 1999).

RESULTS

Detection of a conserved open reading frame overlapping the core region of HCV

Many viruses, including ϕ X174 (Sanger et al., 1977), HBV (Valenzuela et al., 1980), and HEV (Tam et al., 1991) have overlapping genes and/or regions in which a nucleic acid signal is embedded in a coding sequence. The HCV genome almost certainly contains multifunctional regions and may have overlapping reading frames; however to date, no proteins encoded in overlapping genes have been described from this virus. We combined conventional sequence analysis methods and a new program called Framesplitter to search for multifunctional regions in the main ORF of HCV (see Materials and Methods). In an initial analysis, eight highly divergent HCV sequences were studied. Synonymous codons with exceptionally conserved third-position nucleotides were identified throughout the main ORF. Figure 1A shows the tabulated results of the analysis carried out on a small portion of the core-encoding region. In the first codon in this region (amino acid 33 of the main ORF), all eight sequences contain the same glycine codon, GGA (see Fig. 1B). This is noteworthy because the genetic code contains four glycine codons (GGA, GGC, GGG, and GGU). The probability that all eight sequences would contain the same glycine codon is one in 16,384 (i.e., 4^7). Three such unusually conserved codons occur in the segment of the HCV core-encoding region shown in Figure 1B. These codons form part of a larger cluster (#1; see Fig. 2). Clusters of such conserved codons very rarely occur by chance in a sequence with only a single ORF. Any region in which such a cluster occurs is likely to contain more genetic information than a single coding sequence.

Five clusters of codons with unusual conservation of third-position nucleotides were detected in the main ORF of HCV (Fig. 2): a prominent cluster occurs in the core-encoding region (amino acids 11–175); a small cluster is located near the E2 (amino acids 672–728); and three clusters are distributed across NS5b (amino acids 2565–3034). Earlier studies detected an unusual degree of conservation in the regions containing several of these clusters (Ina et al., 1994; Smith & Simmonds, 1997); however, the site near the terminus of E2 has not been noted previously. Based on our results, we reexamined the issue of possible reading frames overlapping the main ORF. Alternate open reading frames, defined here as 50 or more consecutive codons without an in-frame stop codon, were sought in and around the five clusters of excessive codon conservation noted in the main ORF.

A potential overlapping and open alternate reading frame (ARF) was found in the first cluster, in the +1 reading frame relative to the main ORF (see Fig. 2,

A

Posn	AA	AA=T	AA=P	AA=G	AA=A	AA=V	AA=RC	AA=LC	AA=SU	Hits
33	**gly**	0	0	**1**	0	0	0	0	0	**1**
34	val	0	0	0	0	0	0	0	0	0
35	try	0	0	0	0	0	0	0	0	0
36	leu	0	0	0	0	0	0	0	0	0
37	leu	0	0	0	0	0	0	0	0	0
38	pro	0	0	0	0	0	0	0	0	0
39	**arg**	0	0	0	0	0	**1**	0	0	**1**
40	arg	0	0	0	0	0	0	0	0	0
41	**gly**	0	0	**1**	0	0	0	0	0	**1**
42	pro	0	0	0	0	0	0	0	0	0
43	arg	0	0	0	0	0	0	0	0	0

B

AA #												Type
33	**gly**	val	tyr	leu	leu	pro	**arg**	arg	**gly**	pro	arg	1a
	GG**A**	GUU	UAC	UUG	UUG	CCG	CG**C**	AGG	GG**C**	CCU	AGA	
33	**gly**	val	tyr	leu	leu	pro	**arg**	arg	**gly**	pro	arg	4a
	GG**A**	GUU	UAC	UUG	UUG	CCG	CG**C**	AGG	GG**C**	CCC	AGA	
33	**gly**	val	tyr	leu	leu	pro	**arg**	arg	**gly**	pro	arg	1c
	GG**A**	GUU	UAC	UUG	UUG	CCG	CG**C**	AGG	GG**C**	CCC	AGA	
33	**gly**	val	tyr	val	leu	pro	**arg**	arg	**gly**	pro	arg	3a
	GG**A**	GUA	UAC	GUG	UUG	CCG	CG**C**	AGG	GG**C**	CCA	CGA	
33	**gly**	val	tyr	leu	leu	pro	**arg**	arg	**gly**	pro	arg	2b
	GG**A**	GUU	UAC	UUG	UUG	CCG	CG**C**	AGG	GG**C**	CCC	AGG	
33	**gly**	val	tyr	leu	leu	pro	**arg**	arg	**gly**	pro	arg	11
	GG**A**	GUA	UAC	UUG	UUG	CCG	CG**C**	AGG	GG**C**	CCU	CGA	
33	**gly**	val	tyr	val	leu	pro	**arg**	arg	**gly**	pro	lys	10
	GG**A**	GUU	UAC	GUA	UUG	CCA	CG**C**	AGG	GG**C**	CCC	AAG	
33	**gly**	val	tyr	leu	leu	pro	**arg**	arg	**gly**	pro	arg	2a
	GG**A**	GUA	UAC	UUG	UUG	CCG	CG**C**	AGG	GG**C**	CCC	AGG	

FIGURE 1. Details of the sequence analysis: three codons with unusual conservation of the third-position nucleotide in a small portion of the core-encoding region of HCV. Eight full-length sequences of HCV RNA were translated into amino acids, aligned and analyzed by Framesplitter linked to an Excel spreadsheet. Codons meeting the following criteria were identified: They specify one of the eight amino acids represented by at least four codons in the genetic code; they are members of a group of synonymous codons that contained four members; they specify the same amino acid; and they have the same nucleotide in the third position in all eight sequences. The random chance that a codon meeting the first three criteria will also satisfy the fourth is one in 16,384 (one in 4^7). **A** shows the printout of the spreadsheet for a small portion of the genome (codons 33–43). In the top row, the single letter amino acid code is used, with the following modifications, RC denotes arginine codons of the CG group, LC denotes leucine codons of the CU group, and SU denotes serine codons of the UC group. **B** shows the same portion of the genome. Amino acids are presented in lowercase letters, codons in uppercase letters, and excessively conserved third-position bases and the encoded amino acids are highlighted in boldface type.

lower insert; "X" identifies stop codons). The +1 reading frame is open for at least 124 codons in the vast majority (89%) of the 214 full-length core sequences available for analysis. This ARF is open for at least 142 codons in 93% of 90 genotype 1 sequences, and for at least 124 codons in 89% of 37 genotype 2 sequences, 84% of 25 genotype 3 and genotype 10 sequences, 80% of 15 genotype 4 sequences, 100% of 11 genotype 5 sequences, and 81% of 36 sequences of the genotypes 6, 7, 8, 9, and 11. In contrast, the +2 reading frame has numerous stop codons in this region of the genome. Figure 3 depicts the standard reading frame (the main ORF), the +1 reading frame, and the +2 reading frame of three HCV sequences that are known to be infectious. These infectious sequences have a potential coding region in the +1 frame that is from 124 to 160 codons in length (Kolykhalov et al., 1997; Yanagi et al., 1997; Beard et al., 1999). A coding region in the +1 reading frame could contribute to the excessive third-position conservation in the main ORF.

To determine whether a novel, immunogenic, alternate reading frame protein is expressed during HCV infection, patient sera were tested for specific antibodies. In these studies, the Protean package of Lasergene (DNASTAR, Madison, Wisconsin) was used to identify immunodominant domains in ARFP, the HCV core protein, and BSA. Peptides were selected that contained a roughly one-to-one-to-one ratio of aromatic, polar, and hydrophobic amino acids. Four peptides derived from a consensus amino acid sequence of ARFP (A1, A2, A3, and A4), two peptides from the core protein (C1 and C2; see Fig. 4), and one from BSA (BSA2) were individually conjugated to BSA and tested in both western blot and enzyme-linked immunoabsorbant assays (ELISA). However, because an unacceptably high background precluded the use of

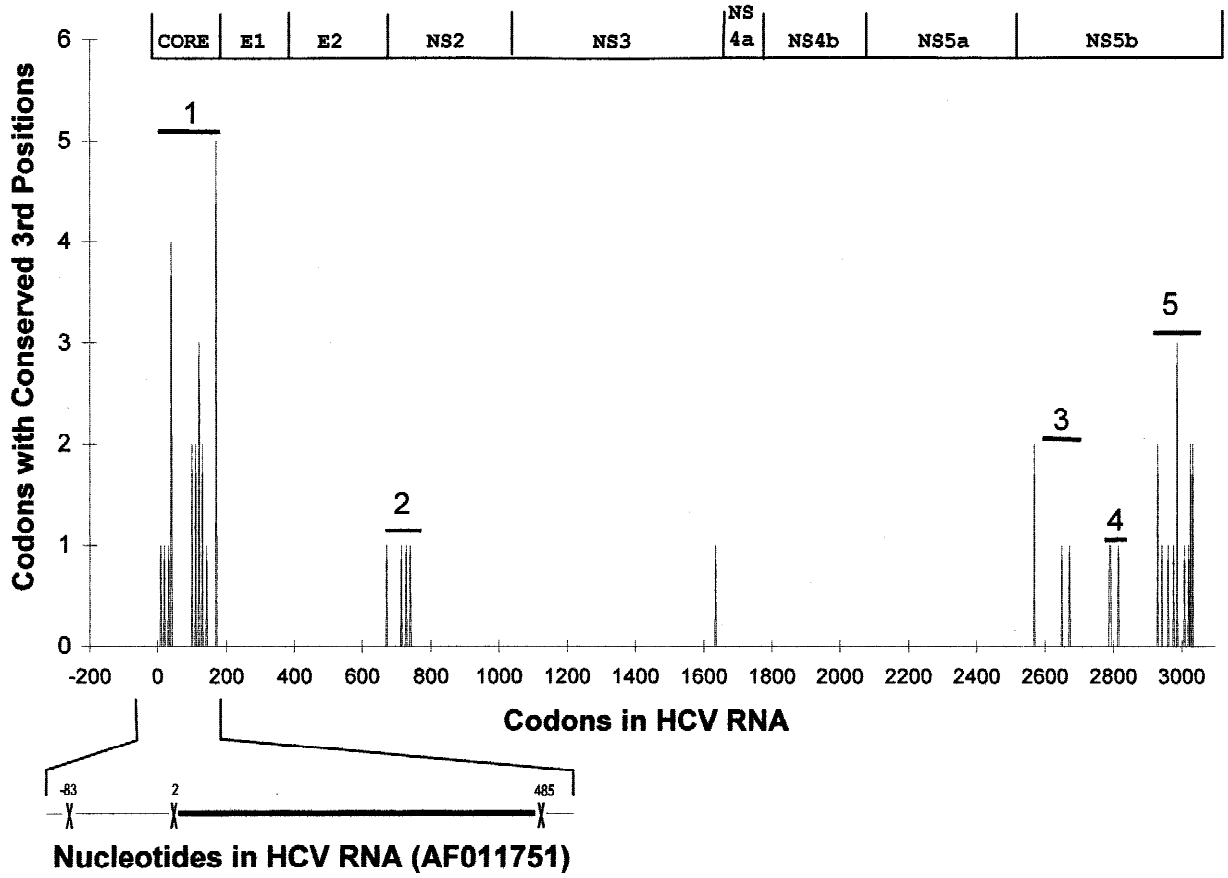

FIGURE 2. Clusters of conserved synonymous codons in the main ORF of HCV. The main ORFs of eight diverse HCV sequences were aligned, and analyzed as illustrated in Figure 1. Codons with conserved third-position nucleotides were identified and mapped onto the genome of HCV. Frequency was expressed per 10 consecutive codons. Five clusters (numbered 1–5) of highly conserved codons were found. A prominent cluster occurs in the core-encoding region. The insert at the top of the figure indicates the relative locations of the known HCV proteins of the main ORF. The insert below cluster 1 depicts the +1 reading frame of HCV RNA from base −83 to base +485 of AF011751, a sequence known to support HCV infection (Yanagi et al., 1997). The solid bar represents an alternate ORF (the ARF), which contains no stop codons for 160 consecutive amino acids. "X" represents the stop codons UAA, UAG, and UGA.

ELISA, all final studies were carried out using western blot analysis.

Sera from chronic HCV patients reacts with ARFP-derived peptides

Figure 5 depicts the results of a western blot assay of serum from an HCV-infected patient. After transfer, the membrane-bound conjugates were stained with the reversible dye Ponceau-S Red, and scanned (Fig. 5A). The resulting image provided a record of the protein bands transferred to the filter. As shown in Figure 5B, this patient's serum contained antibodies that recognize all three versions of the A1 peptide–BSA conjugates present on the filter (A1d, A1e, and A1f; lanes 1, 4, and 9). There was no reaction to the two negative control lanes BSA1 (full-length BSA, Fig. 5B, lane 5) and BSA2 (an immunogenic peptide of BSA conjugated to BSA, Fig. 5B, lane 10).

A dilution series of this patient's serum was tested at final dilutions ranging from 1/800 to 1/100,000. Each dilution was incubated with a separate membrane. The magnitude of the positive signal was directly dependent on the concentration of serum in the solution, confirming the linearity of the assay. Based on these results, we chose a 1/20,000 dilution for the screening of the serum samples.

Competition assays confirm the specificity of the western blots

In these assays, sera were preincubated with free peptides. Reactivity to conjugate A1 was reduced by prior incubation of the primary antibody solution with either 50 μg or 100 μg of free A1 peptide (Fig. 6, compare A to B and C). In contrast, preincubation with free A2 peptide (which resembles the A1 peptide in size and charge) did not decrease reactivity to the A1 conjugate (Fig. 6D). Similarly, for sera that were reactive to the A2 conjugate, binding was inhibited by preincubation with free A2 peptide, and not altered by preincubation with the A1 peptide. Analogous results were obtained

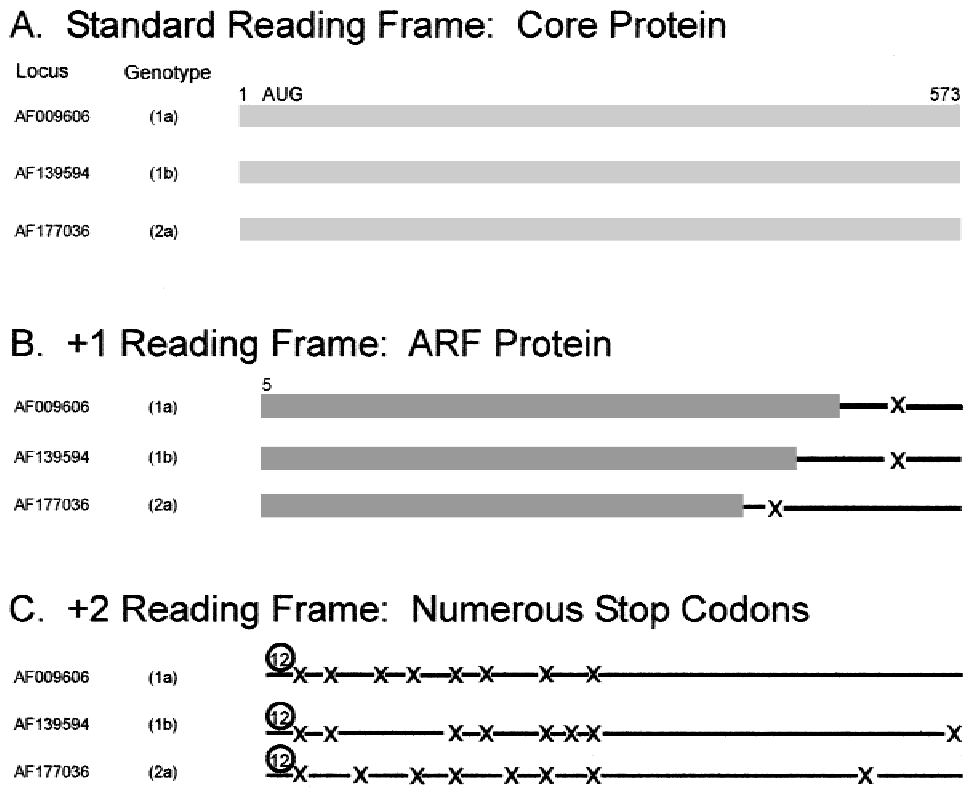

FIGURE 3. Open reading frames in the HCV core-encoding region. The ORFs in each of the three reading frames of the plus strand in the core-encoding region of three HCV infectious clones are depicted as filled rectangles. **A** presents open ORFs in the main reading frame; **B** indicates that the +1 frame (relative to the main ORF) also has a conserved ORF, which ranges in size from 160 amino acids to 125 amino acids in these infectious clones; **C** demonstrates that the +2 reading frame has numerous stop codons indicating that stop codons are not forbidden in this segment of the genome. Stop codons are designated X. Nucleotide positions from the main ORF are given by Arabic numbers.

with the A3 positive sera; preincubation with 25 or 50 μg of free A3 peptide blocked the signal, but the A1 peptide did not. All of the ARFP positive results were confirmed by competition assays.

Anti-ARFP antibodies are associated with HCV infection

Table 1 presents the results of Western blots carried out on one hundred serum samples from 99 chronic HCV patients (Panel A), 44 samples from 44 healthy volunteers (Panel B), 30 samples from patients with HBV infection (Panel C), and 30 samples from patients with noninfectious liver diseases (Panel D). In total, 204 serum samples were tested. All of the negative control lanes (i.e., lanes containing BSA [lane 5] or the BSA-BSA conjugate [lane 10]) showed either very faint reactivity or were completely blank. They were used to establish the negative standard for each blot. As indicated in Table 2A, eight of the serum samples from HCV patients tested positive for reactivity to A1; six to A2; two to A3, and none to A4. χ^2 analysis demonstrated that reactivity to the A1 peptide was significantly associated with HCV infection when compared to all of the samples from individuals who were not infected with HCV (8 of 100 versus 1 of 104; $p < 0.025$). Sera of 6% of the HCV patients reacted with the A2 peptide, compared to none of the 44 healthy controls, and 2 of 60 liver disease patients not infected with HCV. Reactivity to the A3 peptide was seen exclusively in chronic HCV serum samples; however, the number of positive samples was limited to two, and this association did not reach statistical significance. The A4 peptide did not react with any serum sample tested.

As shown in Table 2B, χ^2 analysis indicated that there is a significant correlation between chronic HCV infection and reactivity to at least one of the peptides from the ARFP: 13 of 100 sera from HCV patients compared to 3 of 104 sera from individuals without HCV infection ($p < 0.01$). Finally, as seen in Table 2C, reactivity to A1 and/or A3 was significantly associated with HCV infection ($n = 10/100$) when compared to individuals with other liver diseases ($n = 1/60$; $p < 0.025$), and to the total population of individuals who were not infected with HCV ($n = 1/104$; $p < 0.01$).

When tested with peptides representing the HCV core protein, 34 serum samples of chronic HCV patients reacted with C1, and 23 reacted with C2. One of the

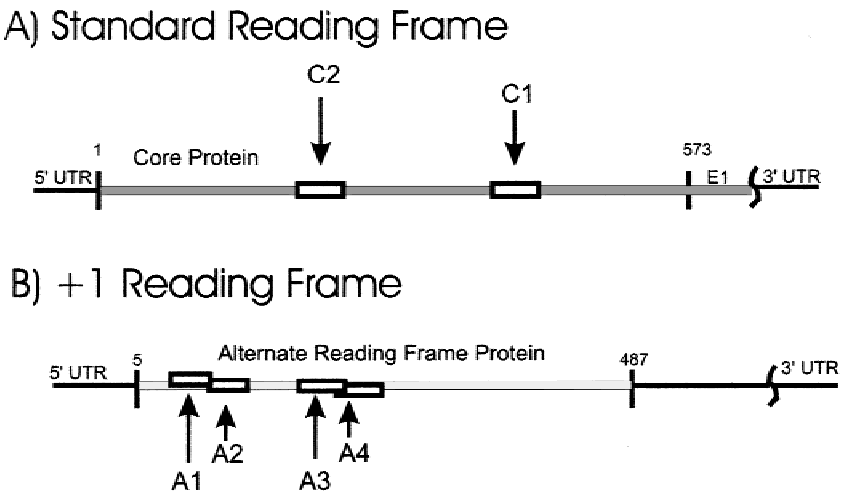

FIGURE 4. Map of the various antigenic peptides derived from HCV proteins that were used in the western blot assays. The relative locations of the two HCV core protein derived peptides (**A**) and the four ARFP derived peptides (**B**) are drawn on the shaded bars representing HCV gene products expressed from the standard open reading frame (**A**), and from the +1 reading frame (**B**). The individual peptides are represented by open rectangles. Noncoding portions of the genome are represented by a thin black line.

HBV serum samples reacted with the C1 conjugate. None of the samples from patients with noninfectious liver diseases and none of the healthy controls tested positive to either of the core peptide conjugates. The association between HCV infection and positivity to C1 and/or C2 was significant ($p < 0.01$). The results of Western blots with ARFP peptide conjugates are summarized and analyzed in Tables 2A–2C.

DISCUSSION

HCV sequence analysis

Conventional methods of nucleic acid sequence analysis, combined with a new software program, Framesplitter, identified a potential overlapping gene in the core-encoding region of HCV. The presence of this over-

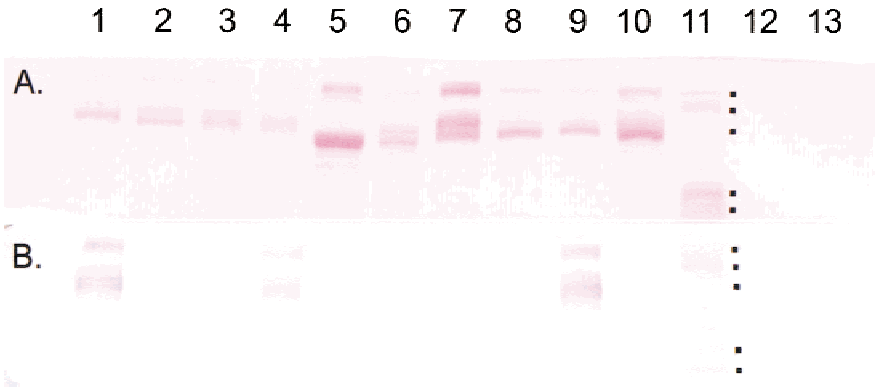

FIGURE 5. Ponceau-S red stain and western blot of various peptide–BSA conjugates incubated with serum of a patient with chronic HCV infection. Peptide–BSA conjugates were fractionated by electrophoresis, transferred to a PVDF filter, and reversibly stained with Ponceau-S Red to determine the distribution of membrane-bound proteins (**A**). After destaining, the blot was incubated with serum diluted 1/20,000, and later developed by reaction with a goat-antihuman secondary antibody conjugated to alkaline phosphatase (**B**). Lane 1: A1d, lane 2: A2d, lane 3: A2e, lane 4: A1e, lane 5: BSA1 (BSA), lane 6: A4, lane 7: A3, lane 8: A2f, lane 9: A1f, lane 10: BSA2 (BSA-BSA), lane 11: Kaleidoscope molecular weight markers, lane 12: C2, and lane 13: C1. Lanes 1 (A1d), 4 (A1e), and 9 (A1f) represent separately prepared BSA conjugates of ARFP-1; A2d, A2e, and A2f represent separately prepared conjugates of ARFP-2. A3 and A4 are BSA conjugates of ARFP-3 and ARFP-4, respectively. C1 and C2 are two different HCV core protein peptides conjugated to BSA. BSA1 represents full-length BSA and BSA2 represents an antigenic peptide of BSA conjugated to BSA. Lane 11: molecular weight markers (from top to bottom) kDa = 216, 132, 78, 46, and 33.

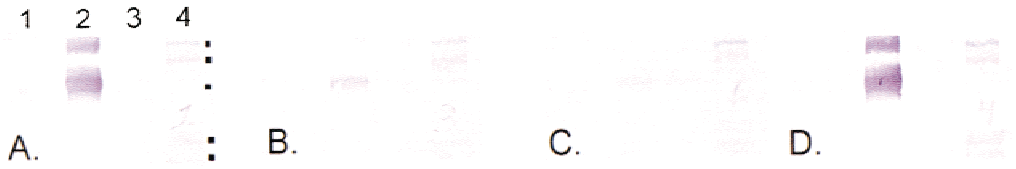

FIGURE 6. Competition assays of western blots. Competition assays were performed to confirm the specificity of the observed reactivity to the ARFP peptides. Four aliquots of patient sera (diluted 1/10,000) reactive to the A1 peptide–BSA conjugate were preincubated as described below, and then incubated overnight with individual membranes. The sample analyzed in **A** did not contain any free peptide; the sample in **B** was preincubated with 50 μg of free A1 peptide; the sample in **C** was preincubated with the 100 μg of the free A1 peptide; and the sample in **D** was preincubated with 50 μg of free A2 peptide. Lane 1: A2, lane 2: A1, lane 3: BSA2 (BSA–BSA), lane 4: Kaleidoscope molecular weight markers. BSA2 represents an antigenic peptide of BSA conjugated to BSA. Lane 4: highlighted molecular weight markers (from top to bottom) kDa = 216, 132, 78, 46, and 33.

lapping gene was initially suggested by an unusual degree of sequence conservation in this portion of the main ORF; certain unusual features of this gene segment were noted in previous studies of HCV sequences (Ina et al., 1994; Smith & Simmonds, 1997). To seek evidence that an overlapping gene accounted for the sequence conservation, we analyzed 214 full-length core sequences for regions lacking stop codons and found that the vast majority of these sequences (89%) contain a conserved second open reading frame that is a minimum of 124 codons in length. It is noteworthy that almost 90% of the HCV sequences in Genbank contain this ARF because stop codons are frequently detected even in the main HCV ORF. The stop codons within the main HCV ORF (Martell et al., 1992; Higashi et al., 1993) presumably reflect the quasi-species nature of HCV and the presence of defective HCV RNAs in the circulation.

Western blot assays

Western blot studies revealed that approximately 13% of HCV patient sera reacted with peptides representing the novel overlapping gene, suggesting that an antigen encoded by this gene is expressed during chronic infection and stimulates antibody responses in some patients. Statistical analysis indicated that there is an association between chronic HCV infection and reactivity to at least one of the peptides from the ARFP: 13 of 100 sera from HCV patients compared to 3 of 104 sera from individuals without HCV infection ($p < 0.01$). Reactivity to two of the ARFP peptide conjugates, A1 and A3, was significantly associated with HCV infection (10 of 100 versus 1 of 104; $p < 0.01$).

Our ARFP peptides A1 and A2 each contained only 13 amino acids of the novel protein. A3 and A4 contained 19 and 17 amino acids, respectively, but these were not expected to be as antigenic as the other two based on the primary amino acid sequence. Because the percentage of the HCV patients testing positive to the two core-derived peptides agrees with previous studies of core peptides of this size range (for which serum reactivity to each peptide ranged from 10 to 90% positivity, with an average of 34%; Kuo et al., 1989; Kotwal et al., 1992; Sallburg et al., 1992), the relatively low percentage of anti-ARFP antibody positive patients may reflect the use of short peptides. Also, the ARFP sequence is not as highly conserved as that of the HCV core protein across different genotypes. Antibodies to an ARFP from one genotype may not recognize the corresponding region in the ARFP of another genotype. To establish the actual percentage of HCV patients with antibodies to ARFP, further studies using recombinant proteins that match the sequence of the infecting strain and an overlapping series of peptides will be needed. The potential clinical usefulness of an ARFP screening assay remains to be explored. Serum assays for antibodies against conventional HCV proteins have provided valuable information about the virus, and continue to evolve (Seelig et al., 1994; Feucht et al., 1995; Caramelo et al., 1996; Huber et al., 1996).

Serum from one HBV patient reacted to the A1 peptide, and serum from another HBV patient reacted to the A2 peptide. A third HBV serum sample reacted with the C1 peptide. While these HBV patients tested negative for conventional markers of HCV infection, their risk factors for exposure to hepatitis viruses are not known and many of the risk factors for HBV are also risk factors for HCV. If the two HBV patients who were positive to A1 and A2 (one apiece) were exposed to HCV in the past, the lack of reactivity to conventional HCV antigens might be explained by the loss of anti-HCV antibodies, which can occur in HCV infected subjects who clear the infection (Takaki et al., 2000). Among the thirty noninfectious liver disease patients, none reacted with the A1 or A3 peptides, and one reacted with the A2 conjugate. Because this patient has no known markers for any viral hepatitis, reactivity to the A2 peptide in this patient may reflect nonspecific cross-reactivity. The four ARFP peptides used in this study were compared by tBLASTn (version 2.1.1) to the nearly complete human genome (NCBI). No significant sequence similarity was found.

TABLE 1. Western blot results of serum samples tested at a 1/20,000 dilution.[a]

A. HCV serum samples

ARF 1	ARF 2	ARF 3	Core 1	Core 2
-	-	-	-	-
*	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	*	-	**	**
-	-	-	-	-
-	-	-	*	*
-	-	-	***	***
-	-	-	*	*
*	-	-	**	*
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	*	*
-	-	-	*	-
-	-	-	**	*
-	**	-	-	-
-	-	-	-	-
-	-	-	-	-
*	**	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	**	*
-	-	-	**	*
*	*	-	-	-
*	*	-	**	*
-	-	-	-	-
-	-	-	-	-
-	-	*	*	*
-	-	-	**	-
-	-	-	***	-
-	**	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	**	-
-	-	-	**	-
-	-	-	-	-
*	-	-	-	***
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	*	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
***	-	-	-	-
-	-	-	-	-
-	-	-	*	*
-	-	-	-	-
-	-	-	**	*
-	-	-	*	-
-	-	-	-	-
-	-	-	***	**
-	-	-	-	-
-	-	-	**	**
**	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	**	*
-	-	-	-	-
-	-	-	*	-
-	-	-	-	-
-	-	-	**	*
-	-	-	*	-
-	-	-	**	*
-	-	-	**	*
-	-	-	**	***
-	-	-	-	-
-	-	-	*	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	*
-	-	-	-	-
-	-	**	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	*	-
-	-	-	*	-
-	-	-	*	*
-	-	-	-	-
-	-	-	*	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-

B. HBV serum samples

ARF 1	ARF 2	ARF 3	Core 1	Core 2
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	**	-
-	**	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
*	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-

C. Noninfectious liver disease samples

ARF 1	ARF 2	ARF 3	Core 1	Core 2
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	*	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-

D. Healthy control serum samples

ARF 1	ARF 2	ARF 3	Core 1	Core 2
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-

[a]For details of the Western blot, see Materials and Methods. Heavy shading indicates reactivity to one of the ARF-P peptides. Light shading indicates reactivity to one of the core peptides. The degree of reactivity is reflected in the number of stars (1–3) in the respective cell.

TABLE 2A. Analysis of positivity to individual ARFP peptides.

	A1	A2	A3	A4
HCV (100)	8/100	6/100	2/100	0/100
Non-HCV liver diseases (60)	1/60	2/60	0/60	0/60
Healthy controls (44)	0/44	0/44	0/44	0/44
Total non-TCV (104)	1/104	2/104	0/104	0/104
Comparison: HCV (100) vs. all non-HCV (104)	$p < 0.025$[a]	N.S.[b]	N.S.	N.S.

[a]A significant association was found between HCV infection and positivity for the A1 conjugate.
[b]N.S. denotes the absence of a significant association with HCV infection.

TABLE 2B. Analysis of positivity for at least one ARFP peptide conjugate.

Comparisons between groups	*p* value
HCV (13/100) vs. noninfectious liver disease (1/30)	N.S.
NCV (13/100) vs. HBV (2/20)	N.S.
HCV (13/100) vs. healthy (0/44)	<0.025[a]
HCV (13/100) vs. total non-HCV (3/104)	<0.01[a]

[a]A significant association was found between HCV infection and positivity for at least one ARFP peptide when compared to healthy controls, or all 104 non-HCV samples.

TABLE 2C. Analysis of positivity for A1 and/or A3.

Comparisons between groups	*p* value
HCV (10/100) vs. healthy (0/44)	<0.05[a]
HCV (10/100) vs. non-HCV L.D. (1/60)	<0.025[a]
HCV (10/100) vs. total non-HCV (1/104)	<0.01[a]

[a]A significant association was found between HCV infection and positivity for A1 and/or A3 when compared to 60 non-HCV liver disease patients (including 30 HBV patients), 44 healthy controls, and all (104) non-HCV samples.

ARFP properties and synthesis

Our studies allow tentative conclusions to be reached about the predicted size of the ARFP. Counting the A of the AUG of the main reading frame as 1, the ARFP peptide A1 corresponds to bases 14–52 and the A3 peptide corresponds to bases 113–169. To encompass the A1 and A3 peptides, the expressed ARFP must span at least 55 contiguous amino acids. Once translation in the ARFP reading frame began, it would be likely to continue until the first in-frame stop codon, which would yield a protein at least 124 amino acids in length from most HCV sequences. If translation were initiated at the codon containing bases 5–7, the ARFP of the infectious HCV cDNA clones of genotype 1a (AF009606) would contain 160 amino acids, and that of genotype 2a would contain 124 amino acids. The 160 amino acid long protein would lack all known nuclear localization signals and have an isoelectric point of 11.70 (for comparison, the pI of the core protein of this HCV sequence is 11.56). Detection and direct analysis of ARFP will be needed to determine its actual size and properties.

At the moment, we have no information about the molecular events that might lead to the initiation of ARFP synthesis; however, several possibilities can be considered. For example, the AUG of the main ORF, or an up-stream AUG, could be the start site of translation, with the reading frame slipping forward to bypass the stop codon at bases 2–4 (in the +1 reading frame). Frame shifts are known to occur during gene expression of many viruses (Brierley, 1995). Alternatively, translation of ARFP might involve a non-AUG initiator codon; in fact, the HCV IRES has the power to direct translation from codons other than AUG (Reynolds et al., 1995). Moreover, a recent report indicates that a novel translation initiation mechanism can lead to the synthesis of out-of-frame peptides (even some without upstream AUG initiation codons), which are then expressed by MHC class I molecules and stimulate immune responses (Malarkannan et al., 1999). This report describes a series of non-AUG codons that allow translation of out-of-frame peptides at levels that elicit robust T cell responses. One of these unusual start codons is AUC. This unusual start codon is the third in-frame codon in >90% of the +1 alternate reading frames of the reported HCV sequences. If this were the start codon for the ARFP, then the expected product would range from 122 to 158 amino acids in the majority of sequences.

Finally, RNA stuttering could modify the sequence of HCV RNA during RNA replication, generating a new RNA that has ARFP as its translation product. Stuttering gives rise to the RNA sequences required for the expression of proteins in most members of the Paramyxoviridae family (Liston & Briedis, 1995). If stuttering occurs during HCV replication, it could lead to the expression of chimeric proteins with amino terminal portions of the HCV core protein linked to carboxy-terminal portions of ARFP. HCV sequences that could give rise to such chimeric proteins have been detected in PCR products from hepatocellular carcinoma (Ruster et al., 1996) and the ascitic fluid of a patient with hepatocellular carcinoma (Yeh et al., 1997).

Our studies do not indicate whether ARFP plays a role in the HCV life cycle, or is synthesized “fortuitously.” However, the ORF is present in all reported HCV genotypes, and all the infectious clones described to date. If ARFP has an impact on cellular or viral function, its highly basic nature and lack of a known nuclear localization signal suggests an interaction with cytoplasmic RNA, perhaps HCV RNA itself. A possible role of ARFP in cellular transformation must be considered in light of studies showing chimeric proteins containing portions of ARFP might be expressed in HCC tissue and ascitic fluid (Ruster et al., 1996; Yeh et al., 1997).

Other cryptic elements in HCV RNA

In addition to ARFP, our sequence analysis also suggests the presence of RNA structural elements, at least some of which may modulate expression of HCV proteins. Within the core-encoding region we find sequence conservation that extends beyond the ARFP, indicating the potential presence of RNA regulatory elements. Interactions between RNA sequences in the 5′ nontranslated region and those in the core-encoding region have recently been reported by Honda and colleagues, and are associated with alterations in the efficiency of protein translation (Honda et al., 1999). One RNA element detected in our analysis that had not been described previously is located near the carboxyl terminus of E2 (see Fig. 2, cluster 2). We also found evidence of RNA elements in the NS5b-encoding regions (see Fig. 2, clusters 3, 4, and 5), in keeping with observations by Smith and Simmonds (1997). These elements, alone or in combination with RNA elements elsewhere in the genome, could function in HCV replication, virion assembly or translation modulation. Whatever their function, once the structures of these elements are known, they will become potential targets for RNA-binding pharmaceutical agents, similar to the aminoglycoside antibiotics, which interact with intricate three-dimensional RNA structures in (bacterial) ribosomal RNA (Moazed & Noller, 1987).

MATERIALS AND METHODS

Comparative sequence analysis

To locate synonymous codons in the standard HCV ORF whose third positions are much more highly conserved than chance would predict, eight full-length HCV sequences representing diverse genotypes and subgenotypes [AF011751 (genotype 1a), HCV4APOLY (4a), HPCEGS (3a), HPCJ8G (2b), HPCCGS (1c), HPCPOLP (2a), HPCJK046E2 (11), HPCJK049E1 (10)] were retrieved from GenBank (Benson et al., 1996) and aligned. A chain of third-position nucleotides was generated with a custom program called Framesplitter (written in Visual Basic by Dr. Michael Bradbury of the Mount Sinai School of Medicine; available upon request) and analyzed by an automated system on an Excel spreadsheet. This system identified codons that met the following criteria: They specified the same amino acid; they specified one of the eight amino acids represented by four or more codons in the genetic code; they were members of a group of synonymous codons that can have any of the four possible nucleotides at the third position; and they were identical in all eight HCV sequences. The probability of a codon that meets the first three criteria also meeting the fourth criterion is one in 16,384 (i.e., 4^7). Thus, the type of sequence conservation we identified occurs rarely by chance in a sequence containing only a single ORF. The positions of highly conserved codons were mapped onto the HCV ORF to locate "clusters." A cluster was defined as a region in which at least three highly conserved codons occurred within a 100-codon segment.

Comparative sequence analysis was also used in the second step of our search for overlapping genes. The regions in and around the clusters of conserved codons were examined for the presence of 50 or more consecutive codons without an in-frame stop codon in the five other reading frames.

Serum samples

Serum samples were obtained from either patients or volunteers under a protocol approved by the Institutional Review Board of the Mount Sinai Medical Center. Additional samples were purchased from BioClinical Partners (Franklin, Massachusetts), or were kindly provided by Dr. Jake Liang (National Institutes of Health), or Dr. Anna Lok (University of Michigan). All sera from HCV patients were positive by commercial antibody tests. All other sera were negative for HCV antibodies by commercial assay.

Antigenic peptide selection and conjugation to BSA

To select peptides predicted to be highly antigenic, consensus sequences of the proposed ARFP, the HCV core protein, and BSA were analyzed for immunogenic domains by the Protean program of LASERGENE (DNASTAR, Madison, Wisconsin). From the resulting list of candidates, peptides that had approximately equal proportions of polar, charged, and hydrophobic amino acids were chosen. The following peptides; core-1 (C1), PTDPRRRSRNLGKVIDTC; core-2 (C2), GCATRKTSERSQPRGRRAP; ARFP-1 (A1), LNLKEKPNVTP TAC; ARFP-2 (A2), AAHRTSSSRAVVRC; ARFP-3 (A3) RAGAPGWVCARLGRLPSGRC; ARFP-4 (A4) RLPSGRN LVEGDNLSPRC; (see Fig. 4) and bovine serum albumin (BSA2) ASLRETYGDMADCC were synthesized and conjugated to activated BSA by Biosynthesis (Lewisville, Texas) or Quality Control Biochemicals (Hopkinton, Massachusetts). To ensure that results could be reproduced, A1 and A2 were synthesized and conjugated to BSA by both companies.

Western blots

For SDS-PAGE, 5 μg total protein of each peptide–BSA conjugate (except C2, 50 ng) were loaded onto an 8% Tris-glycine polyacrylamide gel. Three independently produced A1 conjugates and two A2 conjugates were loaded per gel. Proteins were transferred to polyvinylidene fluoride (PVDF) membranes (Immobilon, 0.45 μm pore size; Millipore, Bedford, Massachusetts) in 10 mM CAPS buffer (3-[cyclohexylamino]-1-propane-sulfonic acid; Sigma, St. Louis, Missouri), pH 11.0, 10% methanol at 20 V, for 4.5 h, at 4 °C with constant stirring. After transfer, blots were temporarily stained with Ponceau-S Red (0.5% in 1% glacial acetic acid), and scanned to provide a record of each transfer. Blots were destained and blocked for 1 h at 30 °C, in 200 mL of 3% blotting-grade nonfat dried milk (BioRad, Hercules, California) in 1× TTBS (Tris-buffered saline with Tween-20 [0.05%]). Prior to first use, sera were diluted (1/20) in 1× BSA/Tween-20 (0.05%) and incubated for 1 h at 30 °C. For western blots, sera were further diluted to a final concentration of 1/20,000, in 2% BSA/1× TTBS. Membranes were incubated overnight

at 4 °C with constant rotation. After washing, the blots were then incubated with 5 μg of secondary antibody [the F(ab′)2 fragment of goat gamma-specific anti-human IgG (KPLabs) conjugated to alkaline phosphatase], for 1 h at 30 °C, with constant shaking. After the final wash, membranes were incubated in alkaline phosphatase substrate (AP Substrate, Bio-Rad, Hercules, California) for 30 min at room temperature.

The filters were read by three people, including a blinded evaluator who confirmed all determinations. Positives were scored according to the following criteria: (1) The recorded image of the Ponceau-S red stained filter was examined to determine the relative amount, shape, and position of the major protein bands. (2) The amount of background staining caused by the alkaline phosphatase reaction was assessed by examining the two BSA negative control lanes (one containing full-length BSA, the other containing BSA2), and the Kaleidoscope molecular weight markers on each blot. (3) A band was scored as positive if it was at least twice as intense as the corresponding bands in the negative control lanes, and occurred in a region of the blot known to contain a protein band. (4) Potential positives were confirmed by competition assays.

Competition assays

Competition assays were performed to confirm the specificity of each serum sample reacting positively with any ARFP conjugate. Prior to incubation with the membranes, 1/10,000 serum dilutions in 2% BSA were preincubated with 25, 50, or 100 μg free peptide (two concentrations chosen on the basis of the strength of the original signal), for 30 min at 30 °C, with agitation. Blots were then processed as usual.

Statistical analysis

Between group comparisons of positivity to each conjugate were done by χ^2 analysis (2×2 contingency table, 1 degree of freedom). Values of p less than 0.05 were considered to be significant.

ACKNOWLEDGMENTS

We thank Nora Bergasa (Columbia University School of Medicine) and our colleagues at Mount Sinai School of Medicine and the Recanati/Miller Transplantation Institute for their contributions, especially Jennifer Jones, who assisted with serum samples, Westyn Branch-Elliman, who helped in the laboratory, Averik Mosoian, who guided us in the use of immunoassays, Michael Bradbury, who developed Framesplitter (available upon request) and Nancy Bach, Paul Berk, Henry Bodenheimer, Jr., and Charles Miller. We thank Jake Liang (National Institutes of Health) and Anna Lok (University of Michigan) for serum samples from HBV patients. A computer was purchased with funds from the Aaron Diamond Foundation. This research was supported by R01 DK52071, P01 DK50795, project 2, Dean Rubenstein's Innovative Projects Fund, the Division of Liver Diseases, the Liver Transplantation Research Fund, and the Artzt Foundation.

Received January 23, 2001; accepted without revision February 23, 2001

REFERENCES

Beard MR, Abell G, Honda M, Carroll A, Gartland M, Clarke B, Suzuki K, Lanford R, Sangar DV, Lemon SM. 1999. An infectious molecular clone of a Japanese genotype 1b hepatitis C virus. *Hepatology 30*:316–324.

Benson DA, Boguski M, Lipman DJ, Ostell J. 1996. GenBank. *Nucleic Acids Res 24*:1–5.

Branch AD, Seeff LB, Berk PD (Eds). 2000. *Hepatitis C: State of the art at the millennium.* New York: Thieme.

Brierley I. 1995. Ribosomal frameshifting on viral RNAs. *J Gen Virol 76*:1885–1892.

Brown EA, Zhang H, Ping LH, Lemon SM. 1992. Secondary structure of the 5′ nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res 20*:5041–5045.

Bullock TN, Eisenlohr LC. 1996. Ribosomal scanning past the primary initiation codon as a mechanism for expression of CTL epitopes encoded in alternative reading frames. *J Exp Med 184*:1319–1329.

Caramelo C, Bartolome J, Albalate M, de Sequera P, Navas S, Bermejillo T, Oliva H, Marriott E, Ortiz A, Ruiz Tunon C, Casado S, Carreno V. 1996. Undiagnosed hepatitis C virus infection in hemodialysis patients: Value of HCV RNA and liver enzyme levels. *Kidney Int 50*:2027–2031.

Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science 244*:359–362.

Feucht HH, Zollner B, Polywka S, Laufs R. 1995. Study on reliability of commercially available hepatitis C virus antibody tests. *J Clin Microbiol 33*:620–624.

Higashi Y, Kakumu S, Yoshioka K, Wakita T, Mizokami M, Ohba K, Ito Y, Ishikawa T, Takayanagi M, Nagai Y. 1993. Dynamics of genome change in the E2/NS1 region of Hepatitis C virus in vivo. *Virology 197*:659–668.

Honda M, Ping LH, Rijnbrand RCA, Amphlett E, Clarke B, Rowlands D, Lemon SM. 1996. Structural requirements for initiation of translation by internal ribosome entry within genome length hepatitis C virus RNA. *Virology 222*:31–42.

Honda M, Rijnbrand R, Amphlett E, Abell G, Kim D, Lemon SM. 1999. Natural variation in translational activities of the 5′ nontranslated RNAs of hepatitis C virus genotypes 1a and 1b: Evidence for a long-range RNA–RNA interaction outside of the internal ribosomal entry site. *J Virology 73*:4941–4951.

Huber KR, Sebesta C, Bauer K. 1996. Detection of common hepatitis C virus subtypes with a third-generation enzyme immunoassay. *Hepatology 24*:471–473.

Ina Y, Mizokami M, Ohba K, Gojobori T. 1994. Reduction of synonymous substitutions in the core protein gene of hepatitis C virus. *J Mol Evol 38*:50–56.

Kolykhalov AA, Agapov EV, Blight KJ, Mihalik K, Feinstone SM, Rice CM. 1997. Transmission of hepatitis C by intrahepatic inoculation with transcribed RNA. *Science 277*:570–574.

Kotwal GJ, Baroudy BM, Kuramoto IK, McDonald FF, Schiff GM, Holland PV, Zeldis JB. 1992. Detection of acute hepatitis C virus infection by ELISA using a synthetic peptide comprising a structural epitope. *Proc Natl Acad Sci USA 89*:4486–4489.

Kuo G, Choo QL, Alter HJ, Gitnick GL, Redeker AG, Purcell RH, Miyamura T, Dienstag JL, Alter MJ, Stevens CE, Tegtmeier GE, Bonino F, Colombo M, Lee W-S, Kuo C, Berger K, Shuster JR, Overby LR, Bradley DW, Houghton M. 1989. An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science 244*:362–364.

Liston P, Briedis DJ. 1995. Ribosomal frameshifting during translation of measles virus P protein mRNA is capable of directing synthesis of a unique protein. *J Virol 69*:6742–6750.

Lu HH, Wimmer E. 1996. Poliovirus chimeras replicating under the translational control of genetic elements of hepatitis C virus reveal unusual properties of the internal ribosomal entry site of hepatitis C virus. *Proc Natl Acad Sci USA 93*:1412–1417.

Malarkannan S, Horng T, Shih PP, Schwab S, Shastri S. 1999. Presentation of out-of-frame peptide/MHC class I complexes by a novel translation initiation mechanism. *Immunity 10*:681–690

Martell M, Esteban JI, Quer J, Genescà J, Weiner A, Esteban R, Guardia J, Gómez. 1992. Hepatitis C virus circulates as a population of different but closely related genomes: Quasispecies nature of HCV genome distribution. *J Virol 66*:3225–3229.

Moazed D, Noller HF. 1987. Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature 327*:389–394.

National Institutes of Health. 1997. *Management of hepatitis C*. pp. 1–18.

Reynolds JE, Kaminski A, Carroll AR, Clarke BE, Rowlands DJ, Jackson RJ. 1996. Internal initiation of translation of hepatitis C virus RNA: The ribosome entry site is at the authentic initiation codon. *RNA 2*:867–878.

Reynolds JE, Kaminski A, Kettinen HJ, Grace K, Clarke BE, Carroll AR, Rowlands DJ, Jackson RJ. 1995. Unique features of internal initiation of hepatitis C virus RNA translation. *EMBO J 14*:6010–6020.

Rice CM. 1996. Flaviviridae: The viruses and their replication. In: Fields BN, Knipe DM, Howley PM, Chanock RM, Melnick JL, Monath TP, Roizman B, Straus SE, eds. *Fields Virology*, 3rd ed. Philadelphia: Lippincott-Raven. pp 931–959.

Ruster B, Zeuzem S, Roth WK. 1996. Hepatitis C virus sequences encoding truncated core proteins detected in a hepatocellular carcinoma. *Biochem Biophys Res Commun 219*:911–915.

Sallberg M, Ruden U, Wahren B, Magnius LO. 1992. Immunodominant regions within the hepatitis C virus core and putative matrix proteins. *J Clin Microbiol 30*:1989–1994.

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M. 1977. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature 265*:687–695.

Seelig R, Renz M, Bottner C, Seelig HP. 1994. Hepatitis C virus infections in dialysis units: Prevalence of HCV-RNA and antibodies to HCV. *Ann Med 26*:45–52.

Smith DB, Simmonds P. 1997. Characteristics of nucleotide substitution in the hepatitis C virus genome: Constraints on sequence change in coding regions at both ends of the genome. *J Mol Evol 45*:238–246.

Takaki A, Wiese M, Maertens G, Depla E, Seifert U, Liebetrau A, Miller JL, Manns MP, Rehermann B. 2000. Cellular immune responses persist and humoral responses decrease two decades after recovery from a single-source outbreak of hepatitis C. *Nat Med 6*:578–582.

Tam AW, Smith MM, Cuerra ME, Huang C-C, Bradley DW, Fry KE, Reyes GR. 1991. Hepatitis E virus (HEV): Molecular cloning and sequencing of the full-length viral genome. *Virology 185*:120–131.

Tsukiyama-Kohara K, Iizuka N, Kohara M, Nomoto A. 1992. Internal ribosome entry site within hepatitis C virus RNA. *J Virol 66*:1476–1483.

Valenzuela P, Quiroga M, Zaldivar J, Gray P, Rutter WJ. 1980. Animal virus genetics: The nucleotide sequence of the hepatitis B genome and identification of the major viral genes. In: Fields B, Jaenisch R, Fox CF, eds. New York: Academic Press. pp 57–70.

Walewski JL, Keller TR, Stump DD, Branch AD. 1999. A Western blot assay for detection of patient antibodies to an HCV alternate reading frame protein. *Hepatology 30*:41A.

Yanagi M, Purcell RH, Emerson SU, Bukh J. 1997. Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee. *Proc Natl Acad Sci USA 94*:8738–8743.

Yeh CT, Lu SC, Chu CM, Liaw YF. 1997. Molecular cloning of a defective hepatitis C virus genome from the ascitic fluid of a patient with hepatocellular carcinoma. *J Gen Virol 78*:2761–2770.