



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

# Virology of Hepatitis C Virus

**Peter Simmonds, BM, PhD**

*Department of Medical Microbiology, The University of Edinburgh Medical School, Edinburgh, Scotland, United Kingdom*

## ABSTRACT

Hepatitis C virus (HCV) has been identified as the main causative agent of post-transfusion non-A, non-B hepatitis. Through recently developed diagnostic assays, routine serologic screening of blood donors has prevented most cases of post-transfusion hepatitis. The purpose of this paper is to comprehensively review current information regarding the virology of HCV. Recent findings on the genome organization, its relationship to other viruses, the replication of HCV ribonucleic acid, HCV translation, and HCV polyprotein expression and processing are discussed. Also reviewed are virus assembly and release, the variability of HCV and its classification into genotypes, the geographic distribution of HCV genotypes, and the biologic differences between HCV genotypes. The assays used in HCV genotyping are discussed in terms of reliability and consistency of results, and the molecular epidemiology of

HCV infection is reviewed. These approaches to HCV epidemiology will prove valuable in documenting the spread of HCV in different risk groups, evaluating alternative (nonparenteral) routes of transmission, and in understanding more about the origins and evolution of HCV.

## INTRODUCTION

Hepatitis C virus (HCV) has been identified as the main causative agent of post-transfusion non-A, non-B hepatitis.<sup>1,2</sup> The identification of HCV led to the development of diagnostic assays for infection, based either on detection of antibody to recombinant polypeptides expressed from cloned HCV sequences or direct detection of virus ribonucleic acid (RNA) sequences by polymerase chain reaction (PCR) using primers complimentary to the HCV genome. Routine serologic screening of blood donors now prevents most or all cases of posttransfusion hepatitis. Assays

for antibody also are important diagnostic tools and have been used to investigate the prevalence of HCV in different risk groups, such as intravenous drug users, patients with hemophilia, and other recipients of blood products, and to conduct epidemiologic studies of HCV transmission.

## GENOME ORGANIZATION

The complete genomic sequence of HCV has been determined for several isolates, revealing both its overall genome organization and its relationship to other RNA viruses. Deducing possible methods of replication by analogy with related viruses is possible, although such studies currently are hampered by the absence of a satisfactory *in vitro* culture method for HCV. As a consequence, most conventional virologic studies are difficult and artificial.

HCV contains a positive-sense RNA genome approximately 9400 bases in length. In overall genome organization and presumed method of replication, it is most similar to members of the family Flaviviridae, particularly in coding for a single polyprotein that is then cleaved into a series of presumed structural and nonstructural proteins (Figure 1).<sup>3</sup> The roles for these different proteins have been inferred by comparison with related viruses and by *in vitro* expression of cloned HCV sequences in prokaryotic and eukaryotic systems. These artificial systems allowed the investigation of protein expression, cleavage, and posttranslational modifications.

## RELATIONSHIP TO OTHER VIRUSES

There are numerous positive-stranded RNA virus families whose coding capacity is contained within a single open read-

ing frame (ORF) as is found in HCV, and with which it may be usefully compared (Table I). Among human viruses, these include both Picornaviridae (eg, poliovirus, coxsackievirus A and B, and hepatitis A virus) and Flaviviridae (eg, dengue fever and yellow fever virus). The genomes of those viruses have a similar organization with structural proteins at the 5' end and nonstructural proteins at the 3' end. However, virus families differ in genome size, the number of proteins produced, the mechanism by which the polyprotein is cleaved, and the detailed mechanism of genome replication. For example, the genome of the Picornaviridae is shorter than that of HCV (approximately 7200 to 8400 bases), contains four nucleocapsid proteins (compared with the single protein of HCV), is nonenveloped (and therefore contains no homologues of the two HCV-encoded HCV glycoproteins E1 and E2), and uses exclusively virus-encoded proteases to cleave its polyprotein. This is different from both HCV and the Flaviviridae, in which cleavage of the structural proteins is thought to be carried out by the host cell-derived signalase.

Members of the Flaviviridae have many features in common with HCV. They have a similar genome size (yellow fever virus has 10,862 bases<sup>4</sup> compared with 9379 for HCV<sup>3</sup>) and package a viral-encoded glycoprotein into the virus envelope (E1). The homologue of E2 in flaviviruses (a membrane-bound glycoprotein called NS1; "NS" stands for nonstructural) is expressed only on the infected cell surface. Like HCV, the polyprotein is cleaved by a combination of viral and host cell proteases. Although there is no close sequence similarity between HCV and other known viruses, at least two regions with conserved amino acid residues provide

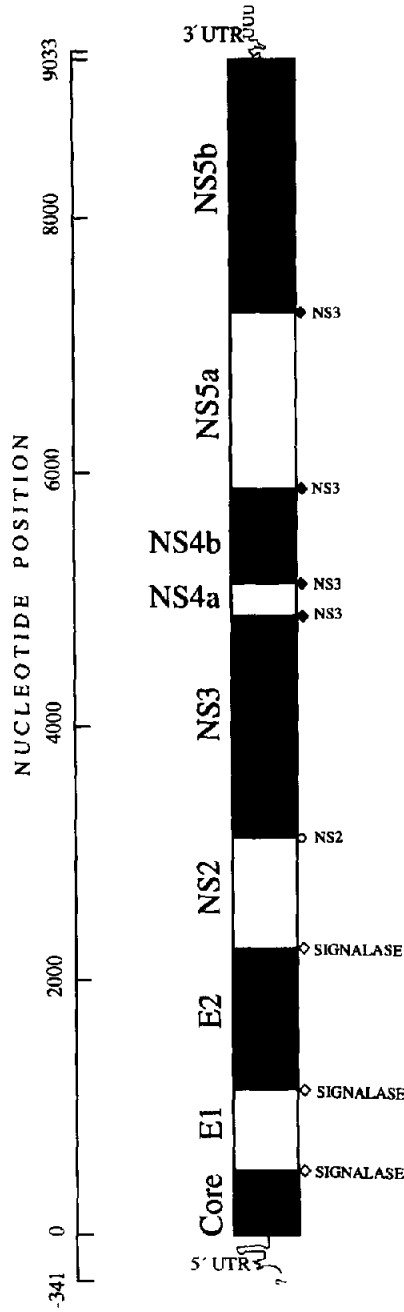


Figure 1. Organization of hepatitis C virus (HCV) genome, showing cleavage sites and relative sizes of HCV-encoded proteins. Nucleotides are numbered as in Choo et al.<sup>3</sup>

Table 1. Comparison of genome organization of hepatitis C virus with other positive-stranded ribonucleic acid (RNA) viruses.

Family	Genus	Example	Genome Size	Genome RNA Modifications	Structural Proteins	Nonstructural Proteins	Enveloped	Protease	Translation Initiation
Picornaviridae	Enteroviruses	Polio	7433	5':VpG, 3':poly(A)	VP1-VP4 (nucleocapsid)	2A-2C, 3A-3D	No	Viral: 2A, 3C	IRES element
	Hepamavirus	Hepatitis A	7478	5':VpG 3':poly(A)	VP1-VP4	2A-2C, 3A-3D	No	Viral: 2A, 3C	IRES element
Flaviviridae	Flavivirus	Yellow fever	10862	5':capped 3':no poly(A)	Core, matrix, E1	NS1-NS5	Yes, E1	Host: signalase Viral: NS3	Ribosomal scanning
	Pestivirus	Bovine viral diarrhea	12480	5':uncapped 3':no poly(A)	Core, gp48 gp25, gp53	p54, p80, p10 (38K), p54, p75	Yes, gp48 gp25, gp53	Host: signalase Viral: NS2, NS3	IRES element
Hepacivirus?	Hepatitis C	Hepatitis C	9379	5':uncapped 3':poly (U)/A	Core, E1 E2 (?)	NS2, 3, 4a, 4b NS5a, 5b	Yes, E1 E2 (?)	Host: signalase Viral: NS2, NS3	IRES element

IRES = internal ribosomal entry site; ? = likely but not demonstrated experimentally.

some information on possible protein functions. For example, the NS5b protein of HCV contains the GDD (glycine-aspartate-aspartate) motif associated with the active site in virus RNA-dependent RNA polymerases of positive-stranded RNA viruses.<sup>5,6</sup> Similarly, there is amino acid sequence similarity in part of NS3 to virus helicase polypeptides.<sup>5</sup>

Another fundamental aspect of genome organization that differs between the flavivirus and picornavirus families is the structure of the 5' and 3' untranslated regions (UTRs). These parts of the genome are involved in HCV replication and initiation of translation by cellular ribosomes of the virus-encoded polyprotein. Pestiviruses and HCV show evidence for a highly structured 5'UTR and 3'UTR, in which internal base-pairing produces a complex set of stem-loop structures that are thought to interact with various host cell and virus proteins during replication.<sup>7,8</sup> In particular, studies have shown that for the Picornaviridae and, more recently, for HCV and pestiviruses,<sup>7,9,10</sup> such structures are involved in internal initiation of translation, in which binding to the host cell ribosome directs translation to an internal methionine (AUG) codon. This contrasts strongly with translation of flavivirus genomes, which act much like cellular messenger RNA in which ribosomal binding initially occurs to the capped 5' end of the RNA, followed by scanning of the sequence in the 5' to 3' direction with translation commencing from the first AUG codon.

Structurally, HCV is also more similar to the pestiviruses than the flaviviruses, with an exceptionally low buoyant density in sucrose (1.08 to 1.11 g/cm<sup>3</sup>),<sup>11</sup> similar to that reported for pestiviruses and attributable in both cases to heavily gly-

cosylated external membrane glycoproteins in the virus envelope. By contrast, flavivirus envelope glycoproteins contain few sites for N-linked glycosylation, and the virion itself is relatively dense (1.2 g/cm<sup>3</sup>). The arrangement and number of cleavage sites of the HCV polyprotein are more similar to pestiviruses, particularly in the further cleavage of both NS4 and NS5 proteins into two subunits, in both cases with NS5b corresponding to the RNA polymerase.

Recently, two distinct RNA viruses have been discovered in New World primate tamarins (*Sanguinis*). This monkey species had previously been shown to harbor an infectious agent causing chronic hepatitis originally derived from inoculation with plasma from a surgeon (GB) in whom chronic hepatitis of unknown etiology had developed.<sup>12</sup> Parts of the genome of the two viruses (provisionally called GBV-A and GBV-B) show measurable sequence similarity to certain regions of HCV. For example, a 200-amino-acid sequence of part of NS3 of GBV-A and GBV-B shows 47% and 55% sequence similarity with the homologous region in HCV (positions 1298 to 1497 in the HCV polyprotein)<sup>3</sup> and 43.5% sequence similarity to each other. Similarly, in NS5, the region around the active site of the RNA polymerase (including the GDD motif and positions 2662 to 2761 in HCV) shows 36% and 41% sequence similarities and 43% between GBV-A and GBV-B.<sup>12</sup> In these nonstructural regions, these similarity values are greater than those between HCV and pestiviruses or flaviviruses, although little homology can be found on comparison of the regions of the genome-encoding structural proteins (ie, the core and envelope), nor with the normally highly conserved 5'UTR.

The degree of relatedness between HCV and other positive-stranded RNA viruses can be more formally analyzed by phylogenetic analysis of highly conserved parts of the genome, such as the NS5 region (and homologues in other viruses) encoding the RNA-dependent RNA polymerase, which invariably contains the canonical GDD motif necessary for the enzymatic activity of the protein. Comparisons of a 100-amino-acid sequence surrounding this motif indicate a close relationship between HCV and GBV-A and GBV-B, an intermediate degree of relatedness with the pestivirus bovine viral diarrhoea virus, and a much more distant relatedness to flaviviruses (Figure 2).<sup>6,13</sup> Remarkably, a series of plant viruses that are structurally distinct from each of the mammalian virus groups, and with different genome organizations, have RNA-dependent RNA polymerase amino acid sequences that are perhaps more similar to those of HCV than are the flaviviruses.

### REPLICATION OF HEPATITIS C VIRUS RIBONUCLEIC ACID

HCV replication has been studied using a variety of experimental techniques. However, little progress has been made toward the development of a practical HCV culture. HCV does not produce obvious cytopathology, and the amount of HCV released from cells infected *in vitro* often is low.<sup>14-18</sup> This might be because the cells used for culture are not representative of those infected *in vivo*, or because productive infection requires a combination of cytokines and growth factors that might be present in the liver but which cannot be recreated in cell culture. The observation that low levels of HCV replication might be detected in lymphocyte<sup>14,18</sup> and

hepatocyte cell lines<sup>16,17</sup> indicates that either the tropism of HCV for different cell types may be greater than first imagined or that the virus replication detected so far does not represent the full replicative cycle of HCV that occurs *in vivo*.

Transfection of full-length DNA sequences of the HCV genome might be expected to initiate the full replicative cycle of HCV, as it does when similar experiments are done in picornavirus sequences. However, only a low level of expression of virus proteins was observed when a complete HCV sequence was transfected into a transformed hepatocyte (hepatoma) cell line (Huh7).<sup>19</sup> Despite this, there was evidence of replication of the HCV genome and the production of low concentrations of progeny virus particles. Such models provide an important experimental system for future investigations of HCV replication.

In common with other positive-strand RNA viruses, HCV is presumed to replicate its RNA genome through the production of a replication intermediate (ie, an RNA copy of the complete genome) and is synthesized by the activity of a virally encoded RNA-dependent RNA polymerase. The minus-strand copy would then be used to generate positive-stranded copies. Because templates can be reused, several minus-strand copies can be synthesized from the infecting positive strand, and each of these transcripts can be used several times to produce positive-strand progeny sequences. In this way, a single input sequence may be amplified several thousandfold.

Although initiation of transcription is well understood for some positive-strand RNA viruses (such as the Picornaviridae), no information currently is available on how RNA synthesis of HCV or other fla-

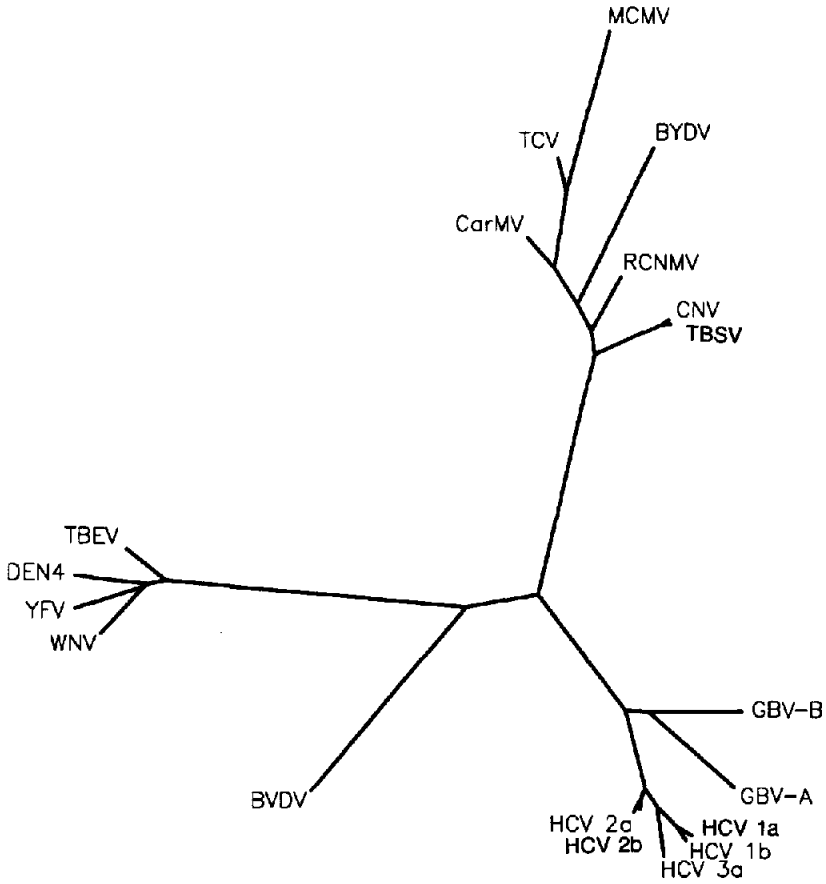


Figure 2. Phylogenetic analysis of amino acid sequences of the region of NS5 encoding the ribonucleic acid (RNA)-dependent RNA polymerase (positions 2662 to 2761 in hepatitis C virus [HCV]-1).<sup>4</sup> Homologous sequences are shown from flaviviruses (TBEV = tick-borne encephalitis virus; YFV = yellow fever virus; WNV = West Nile virus; DEN4 = dengue virus, serotype 4), plant viruses (CarMV = carnation mottle virus; TCY = turnip crinkle virus; MCMV = maize chlorotic mottle virus; BYDV = barley yellow dwarf virus; RCNMV = red clover necrotic mosaic virus; CNV = cucumber necrosis virus; TBSV = tomato bushy stunt virus), and pestiviruses (BVDV = bovine viral diarrhea virus); see Koonin<sup>6</sup> for sources of non-HCV sequences. GBV-A, GBV-B, and HCV (genotypes 1a, 1b, 2a, 2b, and 3a shown) were aligned using the program CLUSTAL, and phylogenetic analysis was done using the programs PROTDIST (PAM matrix), NEIGHBOR, and DRAWTREE in the PHYLIP package.<sup>13</sup>



viruses is primed. HCV lacks homopolymeric tracts (such as poly[U] in the picornaviruses) at the 5' end of the genome, whereas the 3' end is variable, containing either poly(U) or poly(A) tracts, or possibly neither, as now appears to be the case with the related pestiviruses. Furthermore, there appears to be no homologue of the VpG protein of Picornaviridae. For these reasons, it is likely that the mechanism of transcription initiation for HCV is different.

Using a strand-specific PCR method, antisense HCV RNA sequences have been detected in the liver of HCV-infected patients, confirming the presumed method of replication of HCV via a replication intermediate.<sup>20</sup> Such assays provide a valuable technique for detecting HCV replication, as both a sensitive method of monitoring HCV replication in virus culture experiments and a way of investigating the range of cell types and distribution of HCV infection in HCV-infected patients. In particular, the possibility of replication at extrahepatic sites has been proposed on the basis of such assays; these studies have been reviewed by Lau et al.<sup>21</sup>

### HEPATITIS C VIRUS TRANSLATION

The 5'UTR is thought to play a significant role in initiating and regulating translation of the large ORF of HCV. This region is approximately 341 to 344 bases long, and a combination of computer analysis, nuclease mapping experiments, and studies of covariance has led to a proposed secondary structure model for this part of the genome (Figure 3).<sup>22</sup> Using the same methods, researchers have predicted a remarkably similar structure for pestiviruses,<sup>8</sup> despite the virtual absence of nucleotide sequence similarities with

HCV, indicating the importance of the overall structure of this region in interactions with viral and cellular proteins or other RNA sequences.

Direct evidence for internal initiation of translation has been obtained from *in vitro* translation of reporter genes downstream from the 5'UTR sequence placed in mono- or dicistronic vectors.<sup>9,23,24</sup> The nonpaired tip of the stem-loop structure 3 is partially complementary to the 18S subunit of ribosomal RNA and may, therefore, be the site of binding during internal initiation.<sup>25</sup> The internal ribosomal entry site activity of 5'UTR is consistent with the hypothesis that translation is initiated from the AUG methionine codon at position 341.<sup>22</sup> There is no evidence for translation from any of the variable number of AUG triplets upstream from position 341, although production of the small proteins from these upstream potential ORFs may play some role in regulating expression of the large ORF.<sup>26</sup>

### HEPATITIS C VIRUS POLYPROTEIN EXPRESSION AND PROCESSING

In the absence of a cell culture system for HCV, most information available on the expression and processing of HCV proteins has been obtained from transfection experiments with cloned DNA sequences corresponding to the different proteins, and more recently by direct observations of the cellular distributions and properties of HCV proteins detected in liver or plasma *in vivo*. Transfection of prokaryotic or eukaryotic cells with DNA copies of different parts of the HCV genome under the control of artificial promoters allows expression of the encoded proteins, and provides a useful technique for studying their synthesis, biochemical properties, and

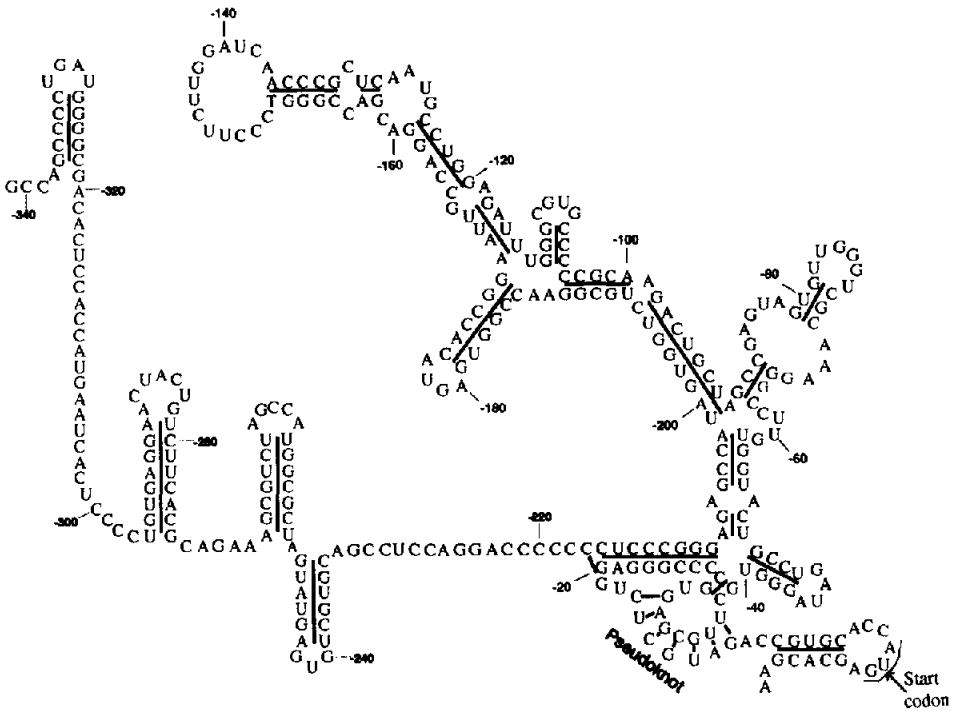


Figure 3. Model of the secondary structure of the 5' untranslated region of hepatitis C virus, based on models proposed by Brown et al<sup>8</sup> and Smith et al.<sup>22</sup>

posttranslational modifications and subcellular location. Alternatively, HCV sequences can be inserted into a vaccinia vector and used to infect mammalian cells.

As is found in the Flaviviridae and Picornaviridae, the proteins encoding the structural proteins of HCV are located at the 5' end of the genome and so are translated first (Figure 1; Table II). Expression of this part of the genome in cells<sup>27-29</sup> or reticulocyte-lysate-containing microsomal membranes<sup>30,31</sup> leads to the synthesis of a polyprotein and its cleavage into a series of proteins. The protein identified as the capsid protein on the basis of comparisons with related viruses is expressed as a protein of approximate size 21 to 22 kd. The

assignment of this protein as the nucleocapsid protein is supported by the presence of regions within the protein containing numerous basic (positively charged) amino acids that may have RNA-binding properties associated with the encapsidation of HCV RNA during virus assembly. Binding of core protein to ribosomal RNA has recently been reported.<sup>31</sup>

Using similar techniques, expression of the putative envelope proteins of HCV (E1 and E2) leads to the synthesis in mammalian cells of two heterogeneous proteins with sizes ranging from 31 to 35 kd and 68 to 72 kd, respectively.<sup>27-31</sup> Cleavage between the capsid protein and E1, E1 and E2, and E2 and NS2 depends on

Table II. Properties of hepatitis C virus (HCV) proteins.

Protein	Nucleotide Position*	Size†	In Virion	Antigen‡	Function
Core	1-573	191	Yes	c22-3	Thought to form the virus nucleocapsid and shows RNA-binding activity; highly conserved between HCV genotypes
E1	574-1149	192	Yes		Sequence predicts a membrane-anchored glycoprotein, with several potential N-linked glycosylation sites; highly variable between genotypes
E2/NS1	1150-2187	327 (?)	Yes (?)		Another membrane-bound glycoprotein, with prominent 'hypervariable' region; the homologue in pestiviruses is a virion component but is nonstructural in flaviviruses
NS2	2431-3078	316	No		Proteinase (Zn <sup>++</sup> -dependent)
NS3	3079-4971	631	No	c33	Multifunctional protein with protease activity and sequence motifs suggesting helicase activity
NS4a	4972-5133	54	No	c100-3	Cofactor for NS3 protease
NS4b	5134-5916	261	No		Unknown
NS5a	5917-7260	448	No	NS5	Unknown
NS5b	7261-9033	591	No		Probably the RNA-dependent RNA polymerase necessary for genome replication

RNA = ribonucleic acid; ? = likely but not demonstrated experimentally.

\*Amino acid positions numbered as in Choo et al.<sup>3</sup>

†Size expressed as number of amino acid residues.

‡Origin of HCV antigens used in current second- and third-generation assays for antibody to HCV.

the addition of microsomal membranes, implying that the host cell signalase has a role in these processing steps. The sizes of E1 and E2 are greater than could be explained by their amino acid sequences alone and support biochemical evidence for extensive glycosylation of both proteins after translation. Both E1 and E2 have a large number of potential N-linked glycosylation sites, although the details of which sites are used, the extent to which the glycoprotein moieties are modified, and whether there is also O-linked glycosylation await further biochemical analysis. Two cleavage sites between E2 and NS2 (both microsome dependent) have recently been identified,<sup>32</sup> leading to the production of E2 proteins differing in size by 80 amino acid residues.<sup>33</sup>

Evidence for intermolecular associations between E1 and E2 has been obtained through immunoprecipitation experiments, in which antibody to E1 or E2 could precipitate both proteins under non-denaturing conditions.<sup>28,29,34</sup> The nature or significance of this association is unclear, although current evidence suggests that the association is predominantly non-covalent<sup>34</sup> and does not occur simply through hydrophobic interactions between the membrane anchors of the two proteins.<sup>33,35</sup> Recently, monoclonal antibodies to either E1 or E2 were shown to co-precipitate NS2 and NS3,<sup>35</sup> and there also is evidence for associations between E2, NS2, and NS4b.<sup>33</sup>

In vitro translation of the rest of the genome leads to the production of proteins of sizes 23, 70 to 72, 4, 27, 56 to 58, and 66 kd, corresponding to NS2, NS3, NS4a, NS4b, NS5a, and NS5b, respectively (Figure 1; Table II). Proteolytic cleavage pathways that generate the non-structural proteins are mediated by NS2

and NS3<sup>36</sup> and have been extensively studied by several groups, as they represent possible targets of antiviral treatment. NS3 is a serine protease that catalyzes cleavage reactions between NS3/NS4a, NS4a/NS4b, NS4b/NS5a, and NS5a/NS5b.<sup>36-40</sup> NS2a is a metalloproteinase that cleaves the NS2/NS3 junction.<sup>36,41</sup> The NS3/NS4a cleavage reaction mediated by NS3 and the NS2/NS3 cleavage mediated by NS2 occur in *cis*, whereas other reactions can occur through intermolecular associations between NS3 and the rest of the polyprotein.

Accounts of the complex sequence of events and the interactions between non-structural proteins involved in cleavage reactions differ in detail depending on the experimental methods used. However, cleavage may be a sequential process modulated by the activities of other proteins, such as NS4a. NS2 protease activity is zinc dependent and contains an active site dependent on residues in NS3. Therefore, after the *cis* cleavage of the NS2/NS3 junction, the protease is inactivated and will not act in *trans* on other substrates. This cleavage reaction has been shown to be essential for activating NS3 protease, and that natural variation in the efficiency of the reaction may modulate the pathogenicity of HCV in vivo.<sup>42</sup>

When released, NS3 cleaves other sites with varying efficiencies. The active site of NS3 has been mapped by deletion experiments to lie at the amino terminus of the protein (residues 1409 to 1215).<sup>43</sup> The substrate specificity of the serine protease activity has been defined by sequence comparisons and mutagenesis experiments<sup>44-47</sup> and generally conforms to the consensus sequence D/E---C/T↓S/A in the target protein. There is some evidence for a less stringent requirement for spe-

cific amino acids around the *cis* cleavage site (NS3/NS4a) than for those cleaved in *trans*.<sup>46</sup> Several investigators have described the requirement for other protein cofactors for the activity of NS3. In particular, it appears that binding of NS4a to NS3<sup>48,49</sup> is necessary at least for the cleavage of NS4b/NS5a and may modulate the activity of NS3 in other ways.

### VIRUS ASSEMBLY AND RELEASE

Although there is now some information on the proteolytic cleavage steps used to process the HCV polyprotein, the difficulty associated with *in vitro* culture of HCV and production of infectious molecular clones of HCV so far has prevented a more detailed understanding of the sites of HCV replication in cells and the processes of virus assembly and release from the cell. Future research should reveal the nature of the interaction between the capsid protein and virus RNA and how this is packaged into the assembled provirion, the posttranslation modifications to the envelope proteins and where these occur in the cell, and the sites of budding of HCV through cellular membranes. To understand replication more fully, we must also identify the mechanism of priming of RNA synthesis from the ends of the genome, the nature of the primers, or whether circularization is necessary for transcription.

### VARIABILITY OF HEPATITIS C VIRUS

Because a cell culture system to investigate differences in neutralization and cytopathic properties of HCV is not available, nucleotide sequence comparisons and typing assays developed from se-

quence data have become the principal techniques for characterizing different variants of HCV. This type of analysis is fairly easy to perform, especially since virus sequences can be amplified by PCR directly from clinical specimens.

In common with other RNA viruses, variants of HCV show considerable sequence variability, many differing considerably from the prototype HCV (HCV-PT).<sup>3</sup> Differences of up to 29% have been found between the complete genomic sequences of the most extremely divergent variants analyzed to date,<sup>50</sup> comparable to those observed between serotypes of other human positive-strand RNA viruses such as poliovirus, coxsackievirus, and coronaviruses. Sequence variability is evenly distributed throughout all virus genes (Table III),<sup>50-57</sup> apart from the highly conserved nucleotide (and amino acid) sequence of the core (nucleocapsid) protein and 5'UTR and the greater variability of the envelope gene (Table III).

### CLASSIFICATION OF HEPATITIS C VIRUS INTO GENOTYPES

Nucleotide sequence comparison of complete genomes or subgenomic fragments between variants has shown that variants of HCV obtained from Japan are substantially different from the HCV-PT variant obtained in the United States.<sup>3</sup> Comparison of the complete genome sequence of HCV-J<sup>53</sup> and HCV-BK<sup>51</sup> from Japan showed 92% sequence similarity to each other but only 79% with HCV-PT. At that time, the former variants were classified as the "Japanese" type (or type II), while those from the United States (HCV-PT and HCV-H) were classified as type I.

Comparisons of subgenomic regions of HCV, such as E1,<sup>58</sup> core,<sup>59,60</sup> and NS5,<sup>61</sup>

Table III. Percent nucleotide (A) and percent amino acid (B) sequence divergence between the complete genomic sequence of the prototype hepatitis C virus (HCV)<sup>51</sup> with variants representing other HCV genotypes.

A. Type	5'NCR	Core	E1	E2/NS1	NS2	NS3	NS4	NS5	3'UTR	Total
1a	0	1	6	7	7	5	4	4	4	5
1b	2	9	25	28	29	20	21	21	33	21
1c	1	10	23	23	28	21	20	20	36	20
2a	6	19	38	33	43	30	33	34	70	32
2b	6	19	44	33	43	30	35	34	74	33
3a	9	16	35	33	42	31	31	32	50	32
3b	6	19	36	47	42	30	32	36	50	32

B. Type	Core	E1	E2/NS1	NS2	NS3	NS4	NS5	Total
1a	1	6	7	7	5	4	4	4
1b	2	21	22	23	9	13	16	15
1c	2	24	17	25	8	12	16	14
2a	10	40	29	42	19	27	30	28
2b	10	45	26	44	19	28	31	28
3a	9	34	28	41	19	20	28	25
3b	14	40	30	39	19	23	27	27

UTR = untranslated regions.  
 Sources of sequences: 1a = HCV-H<sup>52</sup>; 1b = HCV-J<sup>53</sup>; 1c = HC-J9<sup>54</sup>; 2a = HC-J6<sup>55</sup>; 2b = HC-J8<sup>50</sup>; 3a = NZL1<sup>56</sup>; 3b = Tr.<sup>57</sup> In all comparisons, the 5'NCR is the most conserved subgenomic region (maximum 9% nucleotide sequence divergence), whereas highly variable regions are found in parts of the genome encoding E1 and NS2 (35% to 44% nucleotide sequence and 34% to 45% amino acid sequence differences).

provide evidence of at least six major groupings of HCV sequences, each of which contains a series of more closely related clusters of sequences (Figure 4).<sup>62</sup> The current widely used nomenclature for HCV variants reflects this hierarchy of sequence relationships between different isolates. Based on previous suggestions,<sup>63,64</sup> the major branches in the phylogenetic tree are referred to as "types," while "subtypes" correspond to the more closely related sequences within most of the major groups (Figure 4). Although NS5 sequences are analyzed in Figure 4,

equivalent sequence relationships exist in other parts of the genome.  
 The types have been numbered 1 to 6 and the subtypes a, b, and c, in both cases in order of discovery. Therefore, the sequence cloned by Chiron<sup>3</sup> is assigned type 1a, HCV-J and HCV-BK are type 1b, HC-J6 is type 2a, and HC-J8 is type 2b. This nomenclature closely follows the schemes originally described by Enomoto et al<sup>63</sup> on the basis of sequence comparisons of the NS5 region (PT = 1a, K1 = 1b, K2a = 2a, K2b = 2b) and by Chan et al,<sup>64</sup> who provided evidence for a further HCV type

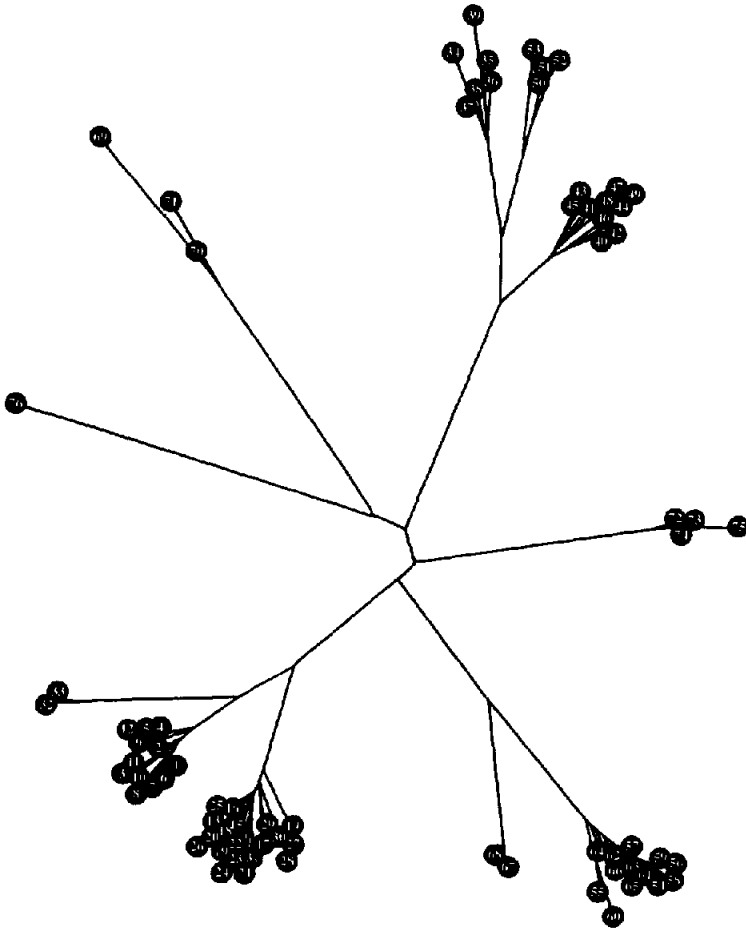


Figure 4. Phylogenetic analysis of nucleotide sequences in part of the hepatitis C virus (HCV) NS5 region. Individual sequences (numbered dots, identified in Simmonds et al<sup>61</sup>) are separated by branch lengths proportional to their reconstructed evolutionary distances. Six main groups of sequence variants are shown<sup>1-6</sup>; those labeled 1, 2, and 3 each comprise a number of more closely related sequences. The naming of the phylogenetic groups follows the proposed nomenclature for HCV variants.<sup>62</sup>

(type 3a) on the basis of phylogenetic analysis of sequences in the NS5, NS3, core, and 5'UTR noncoding regions. This approach avoids the inconsistencies of earlier systems and should be easier to extend when new genotypes are discovered.

#### **GEOGRAPHIC DISTRIBUTION OF HEPATITIS C VIRUS GENOTYPES**

Some genotypes of HCV (types 1a, 2a, and 2b) show a broad worldwide distribution, whereas others, such as types 5a

and 6a, are found only in specific geographic regions. Blood donors and patients with chronic hepatitis from countries in western Europe and the United States frequently are infected with genotypes 1a, 1b, 2a, 2b, and 3a, although the relative frequencies of each may vary.<sup>58,61,65-76</sup> There is a trend for more frequent infection with type 1b in southern and eastern Europe. In many European countries, genotype distributions vary with the age of the patients, reflecting rapid changes in genotype distribution with time within a single geographic area.

A striking geographic change in genotype distribution is apparent between southeast Europe and Turkey (both mainly type 1b) and several countries in the Middle East and parts of North and Central Africa where other genotypes predominate. For example, a high frequency of HCV infection is found in Egypt (20% to 30%),<sup>77-79</sup> of which almost all corresponds to type 4a.<sup>80,81</sup> HCV type 4 also is the principal genotype in countries such as Yemen, Kuwait, Iraq, and Saudi Arabia in the Middle East<sup>60</sup> and in Zaire, Burundi, and Gabon in Central Africa.<sup>58,69,71</sup> HCV genotype 5a is frequently found among patients with non-A, non-B hepatitis and blood donors in South Africa<sup>58,61,70,82</sup> but is found only rarely in Europe and elsewhere.<sup>58,72</sup>

In Japan, Taiwan, and some parts of China, genotypes 1b, 2a, and 2b are the most frequently found.<sup>63,83-90</sup> Infection with type 1a in Japan appears to be confined to patients with hemophilia who received commercial (US-produced) blood products, such as factor VIII and XI clotting concentrates.<sup>63,91</sup> The geographic distribution of type 3 varies; it is only rarely found in Japan<sup>92</sup> and is also infrequent in Taiwan, Hong Kong, and Macau.<sup>93</sup> How-

ever, this genotype is found with increasing frequency in countries to the west, frequently occurring in Singapore and accounting for most hepatitis infections in Thailand.<sup>93,94</sup> In a small sample, it was the only genotype found in Bangladesh and eastern India.<sup>60</sup> As with type 4 in Africa, there is now evidence of considerable sequence diversity within the type 3 genotype, with at least 11 different subtypes of type 3 identified in Nepal,<sup>95</sup> India, and Bangladesh.<sup>60</sup>

A genotype with a highly restricted geographic range is type 6a. This type was originally found in Hong Kong<sup>69,80</sup> and was shown to be a new major genotype by sequence comparisons in the NS5 and E1 regions.<sup>58,61</sup> Approximately one third of anti-HCV-positive blood donors in Hong Kong are infected with this genotype, as are an equivalent proportion in neighboring Macau<sup>81</sup> and Vietnam.<sup>96</sup> A series of novel genotypes has been found in Vietnam<sup>96</sup> and Thailand<sup>60</sup>; these genotypes are distinct from types 1 to 6 classified to date but are more closely related to type 6 than to other genotypes,<sup>96</sup> consistent with their overlapping geographic range with type 6 in Southeast Asia.

## BIOLOGIC DIFFERENCES AMONG HEPATITIS C VIRUS GENOTYPES

Numerous investigations are being conducted into possible differences in the course of disease associated with different HCV genotypes, such as the rate of development of cirrhosis and hepatocellular carcinoma, and whether certain genotypes are more or less likely to respond to interferon treatment. A large number of clinical investigations have documented severe and progressive liver disease in patients infected with each of the well-char-



acterized genotypes (types 1a, 1b, 2a, 2b, 3a, and 4a), so there is little evidence thus far of variants of HCV that are completely nonpathogenic. However, possible variation in the rate of disease progression, differences between genotypes in routes and frequency of person-to-person transmission, or differences in the probability of achieving a sustained response to antiviral treatment would indicate the potential usefulness of identifying the infecting genotype in certain clinical situations.

Several clinical studies have catalogued a variety of factors (including genotype) that correlate with the severity of liver disease and show predictive value for response to antiviral treatment. Factors that frequently have been shown to influence response to interferon treatment include age and duration of infection, presence of cirrhosis before treatment, genotype, and pretreatment level of circulating viral RNA in plasma.<sup>97</sup> A consistent finding reported by several different groups that used a variety of typing assays has been the greatly increased rate of long-term response found when treating patients infected with genotypes 2a, 2b, and 3a compared with type 1b.<sup>74,85,98-106</sup> For example, Chemello et al<sup>102</sup> found that long-term (>12 months) normalization of alanine aminotransferase levels was achieved in only 29% of patients infected with type 1 variants, compared with 52% of those infected with type 2 and 74% of those infected with type 3. In a study by Tsubota et al,<sup>106</sup> infection with type 1b, the presence of cirrhosis, and a high pretreatment virus load were each independently associated with a reduced chance of response (relative risks of 16, 5, and 4, respectively).

The mechanism by which different genotypes differ in response to treatment remains obscure. For treatments such as in-

terferon, we do not know whether the effect of the drug is directly antiviral or whether the inhibition of virus replication is secondary to increased expression of major histocompatibility complex class I antigens on the surface of hepatocytes and greater cytotoxic T-cell activity against virus-infected cells. Elucidating the mechanism of action of interferon and whether there are virologic differences between genotypes in sensitivity to antiviral agents awaits a cell-culture model for HCV infection.

### HEPATITIS C VIRUS GENOTYPING ASSAYS

Although determination of the nucleotide sequences is the most reliable method of identifying different genotypes of HCV, this method is not practical for large clinical studies. Many of the published methods for "genotyping" are based on amplification of viral sequences in clinical specimens, either by using type-specific primers that selectively amplify different genotypes, by analyzing the PCR product by hybridization with genotype-specific probes, or by using restriction fragment length polymorphisms (RFLP). The assays have different strengths and weaknesses. For example, methods based on amplification and analysis of 5'NCR sequences have advantages of sensitivity, because this region is highly conserved and can be more frequently amplified from HCV-infected patients than other parts of the genome. However, few nucleotide differences are found between different genotypes. Although reliably differentiating six major genotypes by using RFLP or by type-specific probes is possible, it is not always possible to reliably identify virus subtypes. Types 2a and 2b consistently differ at position -124, allowing them to be

differentiated by the restriction enzyme *ScrFI*<sup>67</sup> or by probes 10 to 13 in the INNO-LIPA (Innogenetics, Zwijnaarde, Belgium).<sup>71</sup> However, sequences of type 2c are indistinguishable from some of those of type 2a. Similarly, some of the novel subtypes of type 1 often show sequences identical to those of type 1a or 1b,<sup>54,60</sup> and a small proportion of type 1a variants are identical to type 1b and vice versa.<sup>22</sup>

Typing methods based on coding regions, such as core and NS5, can reliably identify subtypes as well as major genotypes because the degree of sequence divergence is much greater (Table III). However, amplifying sequences in coding regions of the genome generally is difficult because sequence variability in the primer-binding sites may reduce the effectiveness of sequence amplification by PCR. Nevertheless, the variation is exploited in a genotyping assay that uses type-specific primers complementary to variable regions in the core gene. Currently, this assay can identify and differentiate types 1a, 1b, 2a, 2b, and 3a,<sup>107,108</sup> although the method is technically complicated to perform reliably<sup>109</sup> and may be difficult to extend to the great range of HCV genotypes now described.

Serologic typing methods have advantages over PCR-based methods in terms of the speed and simplicity of sample preparation and the use of simple equipment found in any diagnostic virology laboratory. By careful optimization of reagents, such assays may show high sensitivity and reproducibility. For example, type-specific antibody to NS4 peptides can be detected in approximately 95% of patients with non-A, non-B hepatitis.<sup>110</sup> Furthermore, the assays can be readily extended to detect new genotypes. One NS4-based assay can reliably identify type-spe-

cific antibody to six major genotypes,<sup>110</sup> although the antigenic similarity between subtypes currently precludes the separate identification of types 1a and 1b and 2a and 2b using the NS4 peptides alone.

## IMMUNE ESCAPE

In contrast to the highly restricted sequence diversity of the 5'NCR and adjacent core region, the two putative envelope genes are highly divergent between different variants of HCV (Table III)<sup>111-114</sup> and show a three-to-four-times higher rate of sequence change with time in persistently infected patients.<sup>115</sup> Because these proteins are likely to lie on the outside of the virus, they would be the principal targets of the humoral immune response to HCV elicited on infection. Changes in the E1 and E2 genes may alter the antigenicity of the virus to allow "immune escape" from neutralizing antibodies,<sup>112</sup> therefore accounting for both the high degree of envelope sequence variability and the observed persistent nature of HCV infection. Supporting this model is the observation that much of the variability in the E1 and E2 genes is concentrated in discrete "hypervariable" regions,<sup>112-114</sup> possibly reflecting the pressures on the virus to evade immune recognition at specific sites where HCV may be neutralized.

Experimental evidence supporting this theory includes the observations that variants of HCV with changes in the E1 and E2 genes are antigenically distinct, and, in many cases, the *in vivo* appearance of variants with different sequences in the hypervariable region is followed by development of antibodies that specifically recognize the new variants.<sup>112,116-118</sup> In one report,<sup>119</sup> persistent HCV infection developed in a patient with deficient anti-

body responses (agammaglobulinemia), but without the development of sequence variability in E2 consistent with the role of antibody-driving variation in immunocompetent persons. On the other hand, envelope sequences obtained sequentially from persistently infected patients sometimes show no significant change,<sup>120-122</sup> whereas in others, variants coexist with antibodies that recognize the corresponding hypervariable region peptides.<sup>116</sup>

Cytotoxic T-cell responses also may play a protective role in HCV infection,<sup>123,124</sup> as they do in other virus infections for which they are more important in virus clearance than antibody response to infection. Although circumstantial evidence supports the theory of immune escape, additional studies are needed to confirm this as a plausible model of virus persistence. Many of the current uncertainties may be resolved when a satisfactory *in vitro* neutralization assay is developed for HCV that enables the effect of amino acid changes in the envelope gene to be directly investigated. Additional information also is needed on the relative importance of humoral and cell-mediated immunity to HCV and to determine which is more important in virus clearance and protection from reinfection.

## MOLECULAR EPIDEMIOLOGY OF HEPATITIS C VIRUS INFECTION

Persistent infection with HCV entails continuous replication of HCV over years or decades of infection in HCV carriers. The large number of replication cycles, combined with the relatively error-prone RNA-dependent RNA polymerase, leads to measurable sequence drift of HCV over time. For example, over an 8-year interval of persistent infection in a chim-

panzee, the rate of sequence change for the genome as a whole was 0.144% per site per year,<sup>115</sup> similar to the rate calculated for sequence change in the 5' half of the genome over 3 years observed in a human carrier (0.192%)<sup>125</sup> and in a cross-sectional study.<sup>126</sup> Using this "molecular clock," it is possible in principle to calculate times of divergence between HCV variants and therefore to establish their degree of epidemiologic relatedness. For example, the finding of relatively few sequence differences between variants infecting two individuals would provide evidence of recent HCV transmission between them. Sequence comparisons in variable regions, such as E2 and NS5, of the HCV genome have been used to document transmission between persons, either from mother to child,<sup>127</sup> within families,<sup>128</sup> by iatrogenic routes,<sup>129-132</sup> or by sexual contact.<sup>133-135</sup> In these studies, the possibility of transmission by different risk behaviors was assessed by measuring the degree of relatedness of HCV recovered from implicated persons.

Phylogenetic analysis of nucleotide sequences provides a more formal method of investigating relationships between sequences. Phylogenetic trees produced by such methods indicate the degree of relatedness between sequences, while the branching order of the different lineages shows the most likely evolutionary history of the sampled population. For example, clustering of HCV sequences into a single phylogenetic group among recipients of an HCV-contaminated blood product (anti-D immunoglobulin) was still apparent 17 years after infection (Figure 5). These approaches to HCV epidemiology will prove valuable in documenting the spread of HCV in different risk groups, evaluating alternative (nonparenteral) routes of trans-

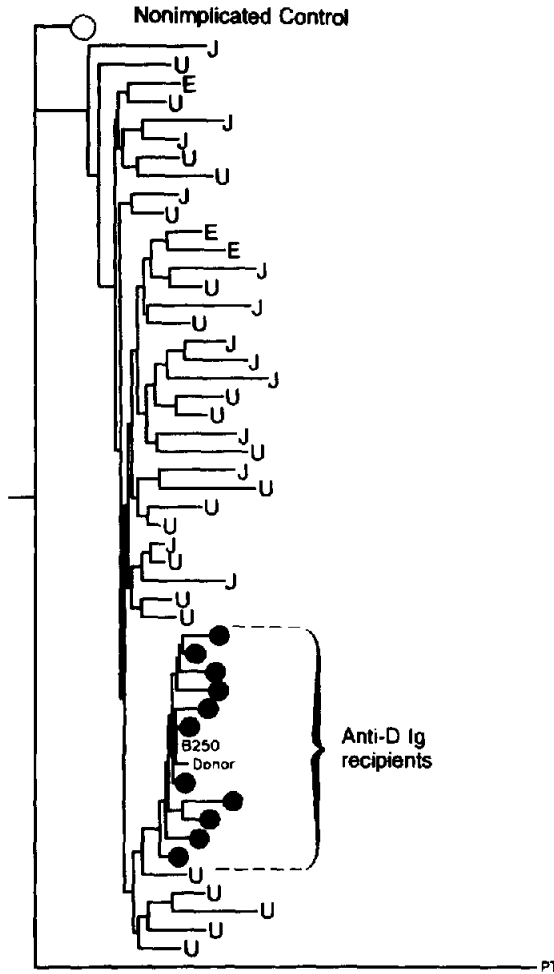


Figure 5. Phylogenetic relationships between sequences from the NS5 region of patients exposed to an implicated batch of anti-D immunoglobulin (Ig) in 1977 (●) and those of epidemiologically unrelated type 1b variants from Japan (J), the United States (U), and Europe (E). B250 = NS5 sequence of hepatitis C virus recovered from batch B250 of anti-D Ig; donor = sequence of variant infecting suspected donor to plasma pool used to manufacture batch B250. Phylogenetic analysis was done on a segment (222 base pairs; positions 7975 to 8196) of the NS5 gene that was amplified, sequenced, and analyzed as previously described.<sup>61</sup> Sequence distances were calculated using the program DNAML in a data set containing the prototype hepatitis C virus (type 1a) as an outgroup. Sequences were obtained from published sources.<sup>61,76</sup>

mission, and understanding more about the origins and evolution of HCV.

This paper attempts to review a rapidly expanding area of research. It is hoped that a combination of basic science and clinical studies may eventually lead to a greater understanding of the ways in which HCV infection may be prevented or cured by the use of antiviral vaccines. The information provided here will clearly form the basis of many of these developments.

## REFERENCES

1. Choo QL, Kuo G, Weiner AJ, et al. Isolation of a cDNA derived from a blood-borne non-A, non-B hepatitis genome. *Science*. 1989;244:359-362.
2. Kuo G, Choo QL, Alter HJ, et al. An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science*. 1989;244:362-364.
3. Choo QL, Richman KH, Han JH, et al. Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci USA*. 1991;88:2451-2455.
4. Rice CM, Lenches EM, Eddy SR, et al. Nucleotide sequence of yellow fever virus: Implications for flavivirus gene expression and evolution. *Science*. 1985;229:726-733.
5. Miller RH, Purcell RH. Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proc Natl Acad Sci USA*. 1990;87:2057-2061.
6. Koonin EV. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J Gen Virol*. 1991;72:2197-2206.
7. Tsukiyama Kohara K, Iizuka N, Kohara M. Internal ribosome entry site within hepatitis C virus RNA. *J Virol*. 1992;66:1476-1483.
8. Brown EA, Zhang H, Ping H-L, Lemon SM. Secondary structure of the 5' non-translated region of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res*. 1992;20:5041-5045.
9. Wang CY, Sarnow P, Siddiqui A. A conserved helical element is essential for internal initiation of translation of hepatitis C virus RNA. *J Virol*. 1994;68:7301-7307.
10. Poole TL, Wang CY, Popp RA, et al. Pestivirus translation initiation occurs by internal ribosome entry. *Virology*. 1995;206:750-754.
11. Miyamoto H, Okamoto H, Sato K, et al. Extraordinarily low density of hepatitis C virus estimated by sucrose density gradient centrifugation and the polymerase chain reaction. *J Gen Virol*. 1992;73:715-718.
12. Simons JN, Pilot-Matias TJ, Leary TP, et al. Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proc Natl Acad Sci USA*. 1995;92:3401-3405.
13. Felsenstein J. *PHYLIP Inference Package Version 3.5*. Seattle, Wash: Department of Genetics, University of Washington; 1993.
14. Shimizu YK, Iwamoto A, Hijikata M, et al. Evidence for in vitro replication of hepatitis C virus genome in a human T-cell line. *Proc Natl Acad Sci USA*. 1992;89:5477-5481.
15. Shimizu YK, Purcell RH, Yoshikura H. Correlation between the infectivity of hepatitis C virus in vivo and its infectivity in

- vitro. *Proc Natl Acad Sci USA*. 1993; 90:6037-6041.
16. Carloni G, Iacovacci S, Sargiacomo M, et al. Susceptibility of human liver cell cultures to hepatitis C virus infection. *Arch Virol*. 1993;8(Suppl):31-39.
  17. Shimizu YK, Yoshikura H. Multicycle infection of hepatitis C virus in cell culture and inhibition by alpha and beta interferons. *J Virol*. 1994;68:8406-8408.
  18. Kato N, Nakazawa T, Mizutani T, Shimotohno K. Susceptibility of human t-lymphotropic virus type i infected cell line MT-2 to hepatitis C virus infection. *Biochem Biophys Res Commun*. 1995; 206:863-869.
  19. Yoo BJ, Selby MJ, Choe J, et al. Transfection of a differentiated human hepatoma cell line (huh7) with in vitro-transcribed hepatitis C virus (HCV) RNA and establishment of a long-term culture persistently infected with HCV. *J Virol*. 1995; 69:32-38.
  20. Lanford RE, Sureau C, Jacob JR, et al. Demonstration of in vitro infection of chimpanzee hepatocytes with hepatitis C virus using strand-specific RT/PCR. *Virology*. 1994;202:606-614.
  21. Lau JYN, Krawczynski K, Negro F, Gonzalez-Peralta RP. In situ detection of hepatitis C virus: A critical appraisal. *J Hepatol*. In press.
  22. Smith DB, Mellor J, Jarvis LM, et al. Variation of the hepatitis C virus 5'-non coding region: Implications for secondary structure, virus detection and typing. *J Gen Virol*. 1995;76:174-176.
  23. Wang CY, Sarnow P, Siddiqui A. Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. *J Virol*. 1993;67:3338-3344.
  24. Fukushi S, Katayama K, Kurihara C, et al. Complete 5' noncoding region is necessary for the efficient internal initiation of hepatitis C virus RNA. *Biochem Biophys Res Commun*. 1994;199:425-432.
  25. Le SY, Sonenberg N, Maizel JV. Unusual folding regions and ribosome landing pad within hepatitis C virus and pestivirus RNAs. *Gene*. 1995;154:137-143.
  26. Yoo BJ, Spaete RR, Geballe AP, et al. 5' End-dependent translation initiation of hepatitis C viral RNA and the presence of putative positive and negative translational control elements within the 5' untranslated region. *Virology*. 1992;191: 889-899.
  27. Selby MJ, Choo QL, Berger K, et al. Expression, identification and subcellular localization of the proteins encoded by the hepatitis C viral genome. *J Gen Virol*. 1993;74:1103-1113.
  28. Ralston R, Thudium K, Berger K, et al. Characterization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia viruses. *J Virol*. 1993;67:6753-6761.
  29. Grakoui A, Wychowski C, Lin C, et al. Expression and identification of hepatitis C virus polyprotein cleavage products. *J Virol*. 1993;67:1385-1395.
  30. Hijikata M, Kato N, Ootsuyama Y, et al. Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. *Proc Natl Acad Sci USA*. 1991;88:5547-5551.
  31. Santolini E, Migliaccio G, Lamonica N. Biosynthesis and biochemical properties

- of the hepatitis C virus core protein. *J Virol.* 1994;68:3631–3641.
32. Mizushima H, Hijikata M, Tanji Y, et al. Analysis of n-terminal processing of hepatitis C virus nonstructural protein 2. *J Virol.* 1994;68:2731–2734.
  33. Selby MJ, Glazer E, Masiarz F, Houghton M. Complex processing and protein:protein interactions in the e2:NS2 region of HCV. *Virology.* 1994;204:114–122.
  34. Dubuisson J, Hsu HH, Cheung RC, et al. Formation and intracellular localization of hepatitis c virus envelope glycoprotein complexes expressed by recombinant vaccinia and sindbis viruses. *J Virol.* 1994;68:6147–6160.
  35. Matsuura Y, Suzuki T, Suzuki R, et al. Processing of e1 and e2 glycoproteins of hepatitis C virus expressed in mammalian and insect cells. *Virology.* 1994;205:141–150.
  36. Hijikata M, Mizushima H, Akagi T, et al. Distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. *J Virol.* 1993;67:4665–4675.
  37. Eckart MR, Selby M, Masiarz F, et al. The hepatitis C virus encodes a serine protease involved in processing of the putative nonstructural proteins from the viral polyprotein precursor. *Biochem Biophys Res Commun.* 1993;192:399–406.
  38. Grakoui A, McCourt DW, Wychowski C, et al. Characterization of the hepatitis C virus-encoded serine proteinase—determination of proteinase-dependent polyprotein cleavage sites. *J Virol.* 1993;67:2832–2843.
  39. Bartenschlager R, Ahlbornlaake L, Mous J, Jacobsen H. Nonstructural protein-3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. *J Virol.* 1993;67:3835–3844.
  40. Tomei L, Failla C, Santolini E, et al. NS3 is a serine protease required for processing of hepatitis C virus polyprotein. *J Virol.* 1993;67:4017–4026.
  41. Grakoui A, McCourt DW, Wychowski C, et al. A second hepatitis C virus-encoded proteinase. *Proc Natl Acad Sci USA.* 1993;90:10583–10587.
  42. Lin C, Pragai BM, Grakoui A, et al. Hepatitis C virus NS3 serine proteinase: Trans-cleavage requirements and processing kinetics. *J Virol.* 1994;68:8147–8157.
  43. Tanji Y, Hijikata M, Hirowatari Y, Shimotohno K. Identification of the domain required for trans-cleavage activity of hepatitis c viral serine proteinase. *Gene.* 1994;145:215–219.
  44. Komoda Y, Hijikata M, Sato S, et al. Substrate requirements of hepatitis C virus serine proteinase for intermolecular polypeptide cleavage in *Escherichia coli*. *J Virol.* 1994;68:7351–7357.
  45. Kolykhalov AA, Agapov EV, Rice CM. Specificity of the hepatitis C virus NS3 serine protease: Effects of substitutions at the 3/4a, 4a/4b, 4b/5a, and 5a/5b cleavage sites on polyprotein processing. *J Virol.* 1994;68:7525–7533.
  46. Bartenschlager R, Ahlbornlaake L, Yasargil K, et al. Substrate determinants for cleavage in cis and in trans by the hepatitis C virus NS3 proteinase. *J Virol.* 1995;69:198–205.

47. Leinbach SS, Bhat RA, Xia SM, et al. Substrate specificity of the NS3 serine proteinase of hepatitis C virus as determined by mutagenesis at the NS3/NS4a junction. *Virology*. 1994;204:163-169.
48. Failla C, Tomei L, Defrancesco R. An amino-terminal domain of the hepatitis C virus NS3 protease is essential for interaction with NS4A. *J Virol*. 1995;69:1769-1777.
49. Tanji Y, Hijikata M, Satoh S, et al. Hepatitis C virus-encoded nonstructural protein NS4a has versatile functions in viral protein processing. *J Virol*. 1995;69:1575-1581.
50. Okamoto H, Kurai K, Okada S, et al. Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: Comparative study of four distinct genotypes. *Virology*. 1992;188:331-341.
51. Takamizawa A, Mori C, Fuke I, et al. Structure and organization of the hepatitis C virus genome isolated from human carriers. *J Virol*. 1991;65:1105-1113.
52. Inchauspe G, Zebede S, Lee DH, et al. Genomic structure of the human prototype strain H of hepatitis C virus: Comparison with American and Japanese isolates. *Proc Natl Acad Sci USA*. 1991;88:10292-10296.
53. Kato N, Hijikata M, Ootsuyama Y, et al. Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc Natl Acad Sci USA*. 1990;87:9524-9528.
54. Okamoto H, Kojima M, Sakamoto M, et al. The entire nucleotide sequence and classification of a hepatitis C virus isolate of a novel genotype from an Indonesian patient with chronic liver disease. *J Gen Virol*. 1994;75:629-635.
55. Okamoto H, Okada S, Sugiyama Y, et al. Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: Comparison with reported isolates for conserved and divergent regions. *J Gen Virol*. 1991;72:2697-2704.
56. Sakamoto M, Akahane Y, Tsuda F, et al. The entire nucleotide sequence and characterization of a hepatitis C virus of genotype V/3a. *J Gen Virol*. 1994;75:1761-1768.
57. Chayama K, Tsubota A, Koida I, et al. Nucleotide sequence of hepatitis C virus (type 3b) isolated from a Japanese patient with chronic hepatitis C. *J Gen Virol*. 1994;75:3623-3628.
58. Bukh J, Purcell RH, Miller RH. At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. *Proc Natl Acad Sci USA*. 1993;90:8234-8238.
59. Bukh J, Purcell RH, Miller RH. Sequence analysis of the core gene of 14 hepatitis C virus genotypes. *Proc Natl Acad Sci USA*. 1994;91:8239-8243.
60. Mellor J, Holmes EC, Jarvis LM, et al. Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: Implications for virus classification. *J Gen Virol*. 1995;76:2493-2507.
61. Simmonds P, Holmes EC, Cha TA, et al. Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *J Gen Virol*. 1993;74:2391-2399.
62. Simmonds P, Alberti A, Alter HJ, et al. A proposed system for the nomenclature of hepatitis C viral genotypes. *Hepatology*. 1994;19:1321-1324.



63. Enomoto N, Takada A, Nakao T, Date T. There are two major types of hepatitis C virus in Japan. *Biochem Biophys Res Commun.* 1990;170:1021-1025.
64. Chan S-W, McOmish F, Holmes EC, et al. Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants. *J Gen Virol.* 1992;73:1131-1141.
65. Chan SW, Simmonds P, McOmish F, et al. Serological responses to infection with three different types of hepatitis C virus. *Lancet.* 1991;338:1391. Letter.
66. Li J, Tong S, Vitvitski L, et al. Two French genotypes of hepatitis C virus: Homology of the predominant genotype with the prototype American strain. *Gene.* 1991;105:167-172.
67. McOmish F, Chan S-W, Dow BC, et al. Detection of three types of hepatitis C virus in blood donors: Investigation of type-specific differences in serological reactivity and rate of alanine aminotransferase abnormalities. *Transfusion.* 1993;33:7-13.
68. Lee C, Cheng C, Wang J, Lumeng L. Identification of hepatitis C viruses with a non-conserved sequence of the 5' untranslated region. *J Clin Microbiol.* 1992;30:1602-1604.
69. Bukh J, Purcell RH, Miller RH. Sequence analysis of the 5' noncoding region of hepatitis C virus. *Proc Natl Acad Sci USA.* 1992;89:4942-4946.
70. Cha TA, Beall E, Irvine B, et al. At least five related, but distinct, hepatitis C viral genotypes exist. *Proc Natl Acad Sci USA.* 1992;89:7144-7148.
71. Stuyver L, Rossau R, Wyseur A, et al. Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. *J Gen Virol.* 1993;74:1093-1102.
72. Kleter GEM, Van Doorn LJ, Brouwer JT, et al. Sequence analysis of the 5' untranslated region in isolates of at least four genotypes of hepatitis C virus in the Netherlands. *J Clin Microbiol.* 1994;33:306-310.
73. Murphy D, Willems B, Delage G. Use of the 5' non-coding region for genotyping hepatitis C virus. *J Infect Dis.* 1994;169:473-475.
74. Tisminetzky SG, Gerotto M, Pontisso P, et al. Genotypes of hepatitis C virus in Italian patients with chronic hepatitis C. *Int Hepatol Commun.* 1994;2:105-112.
75. Qu D, Hantz O, Gouy M, et al. Heterogeneity of hepatitis C virus genotypes in France. *J Gen Virol.* 1994;75:1063-1070.
76. Mahaney K, Tedeschi V, Maertens G, et al. Genotypic analysis of hepatitis C virus in American patients. *Hepatology.* 1994;20:1405-1411.
77. Saeed AA, al Admawi AM, al Rasheed A, et al. Hepatitis C virus infection in Egyptian volunteer blood donors in Riyadh. *Lancet.* 1991;338:459-460.
78. Darwish MA, Raouf TA, Rushdy P, et al. Risk factors associated with a high seroprevalence of hepatitis C virus infection in Egyptian blood donors. *Am J Trop Med Hyg.* 1993;49:440-447.
79. Kamel MA, Ghaffar YA, Wasef MA, et al. High HCV prevalence in Egyptian blood donors. *Lancet.* 1992;340:427. Letter.
80. Simmonds P, McOmish F, Yap PL, et al. Sequence variability in the 5' non coding region of hepatitis C virus: Identification

- of a new virus type and restrictions on sequence diversity. *J Gen Virol.* 1993; 74:661–668.
81. McOmish F, Yap PL, Dow BC, et al. Geographical distribution of hepatitis C virus genotypes in blood donors: An international collaborative survey. *J Clin Microbiol.* 1994;32:884–892.
82. Ohno T, Mizokami M, Tibbs CJ, et al. New genotype of hepatitis C virus in South-Africa. *J Med Virol.* 1994;42: 409–413.
83. Nakao T, Enomoto N, Takada N, et al. Typing of hepatitis C virus (HCV) genomes by restriction fragment length polymorphisms. *J Gen Virol.* 1991;72: 2105–2112.
84. Kato N, Ootsuyama Y, Ohkoshi S, et al. Distribution of plural HCV types in Japan. *Biochem Biophys Res Commun.* 1991; 181:279–285.
85. Takada N, Takase S, Enomoto N, et al. Clinical backgrounds of the patients having different types of hepatitis C virus genomes. *J Hepatol.* 1992;14:35–40.
86. Liu K, Hu Z, Li H, et al. Genomic typing of hepatitis C viruses present in China. *Gene.* 1992;114:245–250.
87. Wang Y, Okamoto H, Mishiro S. HCV genotypes in China. *Lancet.* 1992;339: 1168. Letter.
88. Takada N, Takase S, Takada A, Date T. HCV genotypes in different countries. *Lancet.* 1992;339:808. Letter.
89. Takada N, Takase S, Takada A, Date T. Differences in the hepatitis C virus genotypes in different countries. *J Hepatol.* 1993;17:277–283.
90. Wang Y, Okamoto H, Tsuda F, et al. Prevalence, genotypes, and an isolate (HC-C2) of hepatitis C virus in Chinese patients with liver disease. *J Med Virol.* 1993; 40:254–260.
91. Kinoshita T, Miyake K, Okamoto H, Mishiro S. Imported hepatitis C virus genotypes in Japanese hemophiliacs. *J Infect Dis.* 1993;168:249–250.
92. Chayama K, Tsubota A, Arase Y, et al. Genotypic subtyping of hepatitis C virus. *J Gastroenterol Hepatol.* 1993;8:150–156.
93. Davidson F, Simmonds P, Ferguson JC, et al. Survey of major genotypes and subtypes of hepatitis C virus using restriction fragment length polymorphism of sequences amplified from the 5' non-coding region. *J Gen Virol.* 1995;76:1197–1204.
94. Mori S, Kato N, Yagyu A, et al. A new type of hepatitis C virus in patients in Thailand. *Biochem Biophys Res Commun.* 1992;183:334–342.
95. Tokita H, Shrestha SM, Okamoto H, et al. Hepatitis C virus variants from Nepal with novel genotypes and their classification into the third major group. *J Gen Virol.* 1994;75:931–936.
96. Tokita H, Okamoto H, Tsuda F, et al. Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups. *Proc Natl Acad Sci USA.* 1994;91:11022–11026.
97. Davis GL. Prediction of response to interferon treatment of chronic hepatitis C. *J Hepatol.* 1994;21:1–3.
98. Kanai K, Kako M, Okamoto H. HCV genotypes in chronic hepatitis C and response to interferon. *Lancet.* 1992;339: 1543. Letter.

99. Yoshioka K, Kakumu S, Wakita T, et al. Detection of hepatitis C virus by polymerase chain reaction and response to interferon-alpha therapy: Relationship to genotypes of hepatitis C virus. *Hepatology*. 1992;16:293-299.
100. Tsubota A, Chayama K, Arase Y, et al. Factors useful in predicting the response to interferon therapy in chronic hepatitis C. *J Gastroenterol Hepatol*. 1993;8:535-539.
101. Dusheiko G, Schmilovitzweiss H, Brown D, et al. Hepatitis C virus genotypes—an investigation of type-specific differences in geographic origin and disease. *Hepatology*. 1994;19:13-18.
102. Chemello L, Alberti A, Rose K, Simmonds P. Hepatitis C serotype and response to interferon therapy. *NEJM*. 1994;330:143. Letter.
103. Yamada G, Takahashi M, Miyamoto R, et al. Prediction of interferon effect in chronic hepatitis C by both quantification and genotyping of HCV-RNA. *Dig Dis Sci*. 1994;39:441. Letter.
104. Hino K, Sainokami S, Shimoda K, et al. Genotypes and titers of hepatitis C virus for predicting response to interferon in patients with chronic hepatitis C. *J Med Virol*. 1994;42:299-305.
105. Chayama K, Arase Y, Koida I, et al. Antiviral effect of lymphoblastoid interferon-alpha on hepatitis C virus in patients with chronic hepatitis type C. *J Gastroenterol Hepatol*. 1994;9:128-133.
106. Tsubota A, Chayama K, Ikeda K, et al. Factors predictive of response to interferon-alpha therapy in hepatitis C virus infection. *Hepatology*. 1994;19:1088-1094.
107. Okamoto H, Sugiyama Y, Okada S, et al. Typing hepatitis C virus by polymerase chain reaction with type-specific primers: Application to clinical surveys and tracing infectious sources. *J Gen Virol*. 1992;73:673-679.
108. Okamoto H, Tokita H, Sakamoto M, et al. Characterization of the genomic sequence of type V (or 3a) hepatitis C virus isolates and PCR primers for specific detection. *J Gen Virol*. 1993;74:2385-2390.
109. Lau JYN, Mizokami M, Kolberg JA, et al. Application of six hepatitis C virus genotyping systems to sera from chronic hepatitis C patients in the United States. *J Infect Dis*. 1995;171:281-289.
110. Bhattacharjee V, Prescott LE, Pike I, et al. Use of NS-4 peptides to identify type-specific antibody to hepatitis C virus genotypes 1, 2, 3, 4, 5 and 6. *J Gen Virol*. 1995;76:1737-1748.
111. Kato N, Ootsuyama Y, Ohkoshi S, et al. Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus. *Biochem Biophys Res Commun*. 1992;189:119-127.
112. Weiner AJ, Geysen HM, Christopherson C, et al. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: Potential role in chronic HCV infections. *Proc Natl Acad Sci USA*. 1992;89:3468-3472.
113. Kato N, Ootsuyama Y, Tanaka T, et al. Marked sequence diversity in the putative envelope proteins of hepatitis C viruses. *Virus Res*. 1992;22:107-123.
114. Hijikata M, Kato N, Ootsuyama Y, et al. Hypervariable regions in the putative glycoprotein of hepatitis C virus. *Biochem Biophys Res Commun*. 1991;175:220-228.

115. Okamoto H, Kojima M, Okada SI, et al. Genetic drift of hepatitis-c virus during an 8.2-year infection in a chimpanzee—variability and stability. *Virology*. 1992;190:894–899.
116. Kato N, Sekiya H, Ootsuyama Y, et al. Humoral immune response to hypervariable region-1 of the putative envelope glycoprotein (gp70) of hepatitis C virus. *J Virol*. 1993;67:3923–3930.
117. Lesniewski RR, Boardway KM, Casey JM, et al. Hypervariable 5'-terminus of hepatitis C virus E2/NS1 encodes antigenically distinct variants. *J Med Virol*. 1993;40:150–156.
118. Taniguchi S, Okamoto H, Sakamoto M, et al. A structurally flexible and antigenically variable n-terminal domain of the hepatitis C virus e2/NS1 protein—implication for an escape from antibody. *Virology*. 1993;195:297–301.
119. Kumar U, Monjardino J, Thomas HC. Hypervariable region of hepatitis C virus envelope glycoprotein (e2 NS1) in an agammaglobulinemic patient. *Gastroenterology*. 1994;106:1072–1075.
120. Okada S-I, Akahane Y, Suzuki H, et al. The degree of variability in the amino terminal region of the E2/NS1 protein of hepatitis C virus correlates with responsiveness to interferon therapy in viraemic patients. *Hepatology*. 1992;16:619–624.
121. Kumar U, Brown J, Monjardino J, Thomas HC. Sequence variation in the large envelope glycoprotein (E2/NS1) of hepatitis C virus during chronic infection. *J Infect Dis*. 1993;167:726–730.
122. Higashi Y, Kakumu S, Yoshioka K, et al. Dynamics of genome change in the E2/NS1 region of hepatitis C virus in vivo. *Virology*. 1993;197:659–668.
123. Koziel MJ, Dudley D, Wong JT, et al. Intrahepatic cytotoxic T lymphocytes specific for hepatitis-c virus in persons with chronic hepatitis. *J Immunol*. 1992;149:3339–3344.
124. Koziel MJ, Dudley D, Afdhal N, et al. Hepatitis C virus (HCV)-specific cytotoxic T lymphocytes recognize epitopes in the core and envelope proteins of HCV. *J Virol*. 1993;67:7522–7532.
125. Ogata N, Alter HJ, Miller RH, Purcell RH. Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proc Natl Acad Sci USA*. 1991;88:3392–3396.
126. Cuypers HTM, Winkel IN, Vander Poel CL, et al. Analysis of genomic variability of hepatitis C virus. *J Hepatol*. 1991;13(Suppl):S15–S19.
127. Weiner AJ, Thaler MM, Crawford K, et al. A unique, predominant hepatitis C virus variant found in an infant born to a mother with multiple variants. *J Virol*. 1993;67:4365–4368.
128. Honda M, Kaneko S, Unoura M, et al. Risk of hepatitis C virus infections through household contact with chronic carriers—analysis of nucleotide sequences. *Hepatology*. 1993;17:971–976.
129. Chou WH, Lin SF, Sheu SH, et al. Comparison of hepatitis C virus strains obtained from hemodialysis patients. *Jpn J Med Sci Biol*. 1993;46:191–202.
130. Watanabe M, Ohkoshi S, Tawaraya H, et al. Hepatitis C viral markers in patients who received blood that was positive for hepatitis C virus core antibody, with genetic evidence of hepatitis C virus transmission. *Transfusion*. 1994;34:125–129.

131. Allander T, Medin C, Jacobson SH, et al. Hepatitis C transmission in a hemodialysis unit: Molecular evidence for spread of virus among patients not sharing equipment. *J Med Virol.* 1994;43:415–419.
132. Suzuki K, Mizokami M, Lau JYN, et al. Confirmation of hepatitis C virus transmission through needlestick accidents by molecular evolutionary analysis. *J Infect Dis.* 1994;170:1575–1578.
133. Rice PS, Smith DB, Simmonds P, Holmes E. Heterosexual transmission of hepatitis C virus. *Lancet.* 1993;342:1052–1053.
134. Setoguchi Y, Kajihara S, Hara T, et al. Analysis of nucleotide sequences of hepatitis C virus isolates from husband-wife pairs. *J Gastroenterol Hepatol.* 1994;9:468–471.
135. Healey CJ, Smith DB, Walker JL, et al. Acute hepatitis C infection after sexual exposure. *Gut.* 1995;36:148–150.