Hepatitis C Virus Genotypes

Malek Hossein Ahmadipour1, Seyed-Moayed Alavian2,3, Safie Amini1, Kayhan Azadmanesh 1
1Department of Hepatitis and AIDS, Pasteur Institute of Iran.
2Tehran Hepatitis Center
3Baqiyatallah University, Faculty of Medicine, Tehran, Iran

Introduction

Hepatitis C virus (HCV) is an important cause of chronic liver disease. HCV causes 20% of acute hepatitis cases, 70% of all chronic hepatitis cases, 40% of all cases of liver cirrhosis, 60% of hepatocellular carcinomas, and 30% of liver transplants in Europe. It is also recognized as the leading cause of liver transplantation in the world(2). Only 20% of infected individuals will recover from this viral infection, while the rest become chronically infected(3). While the majority of chronically infected individuals never exhibit symptoms, approximately 10-30% of these patients will eventually develop cirrhosis or hepatocellular carcinoma, both of which are associated with significant morbidity and mortality(4).

More than 170 million people worldwide are chronically infected with HCV. According to WHO report in 2002, chronic liver diseases were responsible for 1.4 million deaths, including 796,000 due to cirrhosis and 616,000 due to primary liver cancer. At least 20% of these deaths are probably attributable to HCV infection- more than 280,000 deaths(5, 6). The prevalence of chronic HCV infection in general population varies greatly in different parts of the world, being estimated between 0.1 and 5%, with a peak prevalence of 20-25% in Egypt. HCV prevalence seems to be less than 1% in Iran, which is much lower than most of the neighboring countries(7). HCV was the first virus discovered by molecular cloning method without the direct use of biologic or biophysical methods. This was accomplished by extracting, copying into cDNA, and cloning all the nucleic acid from the plasma of a chimpanzee infected with non-A, non-B hepatitis by contaminated factor XIII concentrate(8).

The HCV genome is a positive-sense, single-stranded RNA genome approximately 10 kb long. It has marked similarities to those of members of the genera Pestivirus and Flavivirus. Different HCV isolates from around the world show substantial nucleotide sequence variability throughout the viral genome(9).

Genomic Organization of HCV

This virus has a positive-sense single-stranded RNA genome of about 10 kb containing one long open reading frame (ORF). All its proteins are encoded in a single open reading frame that encodes a polyprotein of approximately 3010 amino acids(10) (Fig. 1), with genes coding for structural proteins situated towards the N-terminus of the genome and non-structural genes located near the C-terminus.

The structural genes code for the capsid protein (C, or Core) and the envelope glycoproteins (E1, E2). The first 27 amino acids of the E2 gene constitute the hypervariable region 1 (HVR1), which is the most variable region of the genome and appears to be involved in virus evasion of the immune system and disease progression. The non-structural genes code for a protease (NS2, NS3) and its cofactor (NS4A), a helicase (NS3), a protein of unknown function (NS4B), a phosphoprotein (NS5A), and an RNA-dependent RNA polymerase (NS5B). In addition, the HCV genome has 5'-UTR and 3'-UTR untranslated regions (UTRs) that are involved in control of viral translation.

Correspondence:
Malek H. Ahmadipour, Department of Hepatitis and AIDS, Pasteur Institute of Iran, Tehran, Iran

Tel: +98 21 6696 92 91
Fax: +98 21 6646 51 32
E-mail: m_ahmadipour@yahoo.com
It has a complicated secondary structure comprised of at least seven stem-loops important for ribosome entry and presumably, for viral RNA replication\(^{(11, 12)}\). The translation of HCV RNA begins at the internal ribosome entry site (IRES) and at the 40S ribosomal subunit in the absence of external factors, which makes the HCV translation efficient\(^{(13)}\). Thus, the secondary structure of IRES plays an important role in HCV replication\(^{(14)}\), and single nucleotide changes within this region could alter viral replication characteristics.

HCV replication via RNA-dependent RNA-polymerase is very error-prone and generates mutations at an estimated rate of 10-5 mutations per nucleotide per replication. This high mutation rate is the ultimate source of the virus's genetic diversity. HCV circulates as a heterogeneous population of genetically different but closely related genomes known as the quasispecies\(^{(15)}\).

As only 30-35% of nucleotides actually differ, there is obviously considerable heterogeneity in evolutionary rates among nucleotide sites in the genome. This heterogeneity is the result of variable evolutionary constraints. The 5'-UTR contains extensive secondary RNA structure and is correspondingly the slowest evolving genomic region\(^{(16)}\). The next slowest region is the C (Core) gene, which evolves three times faster than the 5' UTR. The envelope genes E1 and E2 constitute the most diverse genome region and evolve about nine times faster than the 5'-UTR\(^{(16)}\), probably as a result of their presumed role in evading the host immune response.

**Genomic Heterogeneity and Classification Systems**

Shortly after its discovery in 1989, it became clear that HCV had substantial nucleotide sequence diversity, with only 66 to 80% overall sequence similarity among strains belonging to different genotypes or subtypes\(^{(17)}\).

HCV isolates show four levels of genomic variations: types, subtypes, isolates, and quasispecies. The overall sequence similarities over complete genomic sequences are at least 91% within quasispecies, approximately 79% (range, 77 to 80%) between subtypes, and about 68% (range, 66 to 69%) between different types. This quasispecies is composed of a group of heterogeneous RNA sequences centered around a dominant nucleotide sequence that changes, throughout the course of the infection, under the selective pressure of the host immune system\(^{(18)}\). More than one genotype can be found in the circulations of some HCV-infected patients, particularly in individuals who have received multiple transfusions and intravenous drug users. These are referred to as mixed-genotype infections\(^{(19, 20)}\).

The lack of a routinely available cell culture system and an easily available animal model has rendered classification of HCV, based on its phenotype according to its disease pattern or cytopathology, and serotype based on a panel of cross-neutralization antibodies, impossible\(^{(21)}\). Phylogenetic analysis may aid in the separation of sequences into distinct types\(^{(22)}\). So far six major genotypes (HCV-1 to HCV-6) have been described, each containing multiple subtypes (e.g., 1a, 1b, etc.). More than 50 subtypes have been reported to date and are normally identified on the bases of partial gene sequences from E1 and NS5B. The isolates formerly published as genotypes 7 to 11 are now considered subtypes within genotypes 3 (subtype 10) and 6 (subtypes 7, 8, 9, and 11)\(^{(22)}\).

In general, sequence characteristics of a particular subtype are found throughout the HCV genome. Thus, the HCV genotype has been determined primarily based on analysis of partial genome sequences. The most extensive database exists for the 5'-UTR, core, E1, and NS5B (23-28). Whereas the 5'-UTR is highly conserved and therefore preferred for diagnosis, the core, the envelope, and the NS5B regions are less conserved and therefore highly discriminative and preferred for subtyping. Although the 5'-UTR contains characteristic sequence motifs of some genotypes, analysis of this region may not accurately predict all genotypes or subtypes.

**HCV Genotyping Methods**

**Molecular Genotyping**

Because differences in geographical distribution, disease outcome, and response to therapy among HCV genotypes have been suggested, reliable

<table>
<thead>
<tr>
<th>1</th>
<th>573</th>
<th>1149</th>
<th>2187</th>
<th>3078</th>
<th>4971</th>
<th>5916</th>
<th>9033</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>E1</td>
<td>E2</td>
<td>P7</td>
<td>NS2</td>
<td>NS3</td>
<td>NS4A</td>
<td>NS4B</td>
</tr>
<tr>
<td>NS5A</td>
<td>NS5B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1.** The HCV genome contains a single open reading frame (ORF). The genes for structural proteins (C, E1, E2, P7) are situated towards the N-terminus of the ORF. Genes coding for proteins necessary for viral replication are found towards the C-terminus of the ORF.
methods for determining the HCV genotype may become an important clinical test. The HCV genotype can be determined by nucleotide sequencing of a specific PCR-amplified portion of the HCV genome obtained from the patient, followed by composition of a phylogenetic tree, which is presently the ‘gold standard’ for the detection and identification of the various HCV genotypes and subtypes(29). Investigators of HCV genotyping have used sequence analysis of HCV NS5, core, E1, and 5’-UTRs. This approach, however, is cumbersome and regarded as impractical for routine clinical laboratory settings(30).

More convenient methods focus on the amplification of defined regions of the HCV genome by reverse transcription (RT)-PCR followed by digestion with restriction enzymes and restriction length polymorphism analysis (RFLP)(31), amplification with genotype-specific primers(32), hybridization of genotype-specific probes with the amplified products(33) heteroduplex mobility assay(34), melting curve analysis with fluorescence resonance energy transfer probe(35) and DNA enzyme assays(36). HCV typing can also be done by analyzing the sequences of several regions of the virus genome(37, 38). A commercial kit (InnoLipa) for HCV genotyping has been introduced in Europe by Innogenetics (Zwijndre, Belgium), which is based on hybridization of 5’-UTR amplification products with genotype specific probes(39). It has been shown that genotyping methods using 5’-UTR, including InnoLipa, may not discriminate subtype 1a from 1b in 5 to 10% of cases and also may not distinguish between subtypes 2a and 2c(39). Others have used restriction enzymes to determine a restriction fragment length polymorphism. In this method, a PCR amplified DNA fragment is digested into fragments with different lengths by enzymes (restriction endonucleases) that recognize cleavage sites specific for each genotype(40). Investigators have used different regions of the HCV genome for restriction fragment length polymorphism, including NS5 and the 5’-UTR(41, 42). Although the 5’-UTR region is sufficiently variable to allow HCV genotypes in most clinical situations, it cannot always characterize HCV subtypes(43) and may fail to trace the origin of a de novo infection. The NS5b region contains a subtype-specific motif which makes it suitable for epidemiological applications(44).

It is difficult to assess the actual prevalence of mixed-genotype infections by currently available assays, including direct DNA sequencing, since they are designed to identify only the HCV genotype dominant in the population(45, 46).

Serologic Genotyping

More recently, investigators identified genotype-specific antibodies that could be used as indirect markers for the HCV genotype(47-49). Serological typing uses enzyme immunoassays to detect the antigenic properties of several specific epitopes encoded by the NS-4 or the core regions of the HCV genome(50-52). Serologic genotyping has several advantages that make it suitable for large epidemiologic studies. These advantages include the low risk of contamination and the simplicity of the assay. However, serologic typing seems to lack specificity and sensitivity, which limits its usefulness.

Clinical Relevance of HCV Genotypes

Genotype, viral load, and liver histology are important parameters used in selecting an antiviral therapy with the greatest chance of success. Genotyping and subtyping of HCV is relevant to the epidemiology of HCV, vaccine development, clinical management, and assessment of the risk-benefit ratio of therapeutic measures against chronic HCV infection(53, 54). It has been postulated that differences in nucleotide sequence could result in differential activity of HCV proteins that could alter the rate of HCV replication, sensitivity to the antiviral activity of interferon, or pathogenicity of the virus(55). In recent years, substantial evidence has emerged indicating that typing and subtyping for HCV is important clinically; genotype 1 in particular cannot be treated efficiently with IFN-alfa, while genotypes 2 and 3 respond favorably(65, 57).

The causes of variation in treatment response are not well understood. Studies of Japanese patients infected with subtype 1b indicated that the outcome of interferon therapy was correlated with genetic variability in a portion of the NS5A gene (the interferon sensitivity determining region, ISDR)(58), although subsequent studies of European patients did not confirm this result(59). Moreover, genotype 1 infection may proceed more rapidly to severe forms of chronic hepatitis, cirrhosis and hepatocellular carcinoma, when compared with genotype 2 and 3(60, 61).

Geographic Distribution of HCV Genotypes

HCV types and subtypes exhibit complex patterns of geographic distribution, relative prevalence and modes of transmission that can be best understood by categorizing them into three groups. The ‘epidemic’ group which contains subtypes 1a, 1b, 2a, 2b and 3a are distributed
globally and account for the majority of HCV infections worldwide\(^{(62, 63)}\). The rapid spread and global dissemination of these subtypes arises from their efficient transmission via certain transmission routes, namely, infected blood products and injecting drug use. Subtypes 1b and 2a are more strongly associated with the former route and the relative prevalence of these subtypes has decreased in recent years due to improved blood screening\(^{(64, 65)}\). Subtypes 1a and 3a most often infect injecting drug users and appear to be increasing in prevalence\(^{(66, 67)}\).

The ‘endemic’ group of HCV strains is less prevalent than the epidemic subtypes and tends to have restricted geographic distributions. For example, the subtypes of type 6 are found only in Southeast Asia. The high genetic diversity of endemic strains points to a long period of infection in these areas, where transmission is thought to be maintained by a variety of relatively inefficient social and domestic routes, including sexual transmission\(^{(68)}\). As HCV was identified in 1989, differences in the long-term transmission dynamics of the endemic and epidemic strains seem to be impossible to discover. However, using methods based on coalescent theory\(^{(69)}\), the epidemic history of different HCV strains can be reconstructed from observed viral genetic diversity\(^{(70)}\). The third group is the ’local epidemic’ strains of HCV that are found at high prevalence but only in specific locations and risk groups. The best example is subtype 4a which infects more than 10\% of the Egyptian population but is rare outside the Middle East. Epidemiological studies suggest that this strain was widely transmitted in Egypt during the twentieth century by mass injectable drug treatment campaigns against schistosomiasis\(^{(71)}\).

Subtype 1a and 3a were predominant accounting for 47 and 36\%, whereas 1b and 4 accounted for 8 and 7\%\(^{(72)}\).

References

3. Cohen J. A coefficient of agreement for nominal scales.

*Hepatitis C Virus Genotypes*


64. Mondelli MU and Silini E. Clinical significance of hepatitis C virus genotypes. *J Hepatol* 1999; 31(Suppl. 1): 65-70