Abstract. Hepatitis C virus (HCV) is an enveloped, single-stranded RNA virus that has been classified in the Flaviviridae family. The genome of 9400 nucleotides comprises two non-coding regions in 5' and 3' flanking a large reading frame which codes for a polyprotein of 3000 amino acids; this polyprotein is further cleaved into structural (C, E1, E2) and non-structural (NS1, NS2, NS3, NS4, NS5) proteins. The positive RNA acts as a cap-independent messenger; the transcription is mediated by the NS5 RNA polymerase. After the maturation step, the virion is liberated by budding through the cytoplasmic membrane. As for many other RNA viruses, the HCV genome exhibits a high degree of variability, especially in the E2/NS1, E1, NS3 and NS5b regions. Conversely the 5' non-coding region is highly conserved, at least in part, and can be used for diagnostic purposes by PCR technique. Six genotypes of HCV have already been reported, numbered from 1 to 6 in Simmonds’ classification. The same genotype can be divided into subtypes (for instance, genotype 1 comprises three subtypes: 1a, 1b and 1c). Various minor variants of the same strain, called quasispecies, are commonly present in the blood of the same patient. Strains of genotype 1b—which is the most widespread worldwide—are correlated with more severe clinical manifestations, greater viral loads and lower response to interferon treatment. The high variability of the HCV genome contributes greatly to the difficulty of designing potent vaccines.

Introduction

The Hepatitis C virus (HCV) was identified in 1988 by Houghton et al. [1]. It was the first time in virology history that a virus was characterized by genetic cloning and not by isolation. Some six years after the initial publications, knowledge on HCV has dramatically progressed with regard to the genomic structure and variability of the virus. In contrast, morphological and viral growth data are still in their early years [2].

HCV structure

HCV is a single-stranded RNA virus. Its genome has marked similarities with that of pestiviruses and flaviviruses. Pending more accurate structural description, HCV has been classified as a new genus, sometimes called heparnavirus (hepatic RNA virus) due to symmetry with the hepatitis B virus, within the Flaviviridae family. It is an enveloped virus; its resistance in the environment nevertheless is still not fully known.

The virus is spherical in shape; the size of the viral particles is between 36 and 62 nm and its density has been estimated to be 1.08 g/ml. It is frequently complexed with antibodies in blood circulation.

Genomic organization and protein constitution

The organization of the HCV RNA is shown in Figure 1. The genome includes approximately 9400 nucleotides; it consists in two non-coding regions in 5' and 3' (5' NC and 3' NC) flanking a long reading frame coding for a 3000-amino-acid polyprotein.

The 5' non-coding region is the most conserved part of the genome. This explains why the primers used to
amplify the HCV RNA for diagnostic purposes are usually selected from that region. It plays a crucial role in viral genome replication.

The 5′ coding region corresponds to the capsid (C) and envelope (E1 and E2) protein genes (structural region). The capsid protein (p22), is widely used for serological diagnosis. Proteins E1 (gp33) and E2/NS1 (gp72) are glycosylated. Their involvement in envelope constitution is very likely, although not totally proven. As for gp120 of human immunodeficiency virus (HIV), they probably exhibit transmembrane anchoring and extramembrane extension. Their diagnostic use is still limited.

The 3′ coding region codes for the non structural or enzymatic proteins:

- Protein NS2 (p23) is a metalloprotease involved in NS2–NS3 cleavage;
- Protein NS3 (p72) is a serine-protease responsible for cleavage between NS3 and NS4a, NS4a and NS4b, NS4b and NS5a, and NS5a and NS5b; it also has an helicase activity necessary for viral RNA replication. Like the capsid protein, it is one of the major antigens of the virus and is widely used for serological diagnosis;
- Protein NS4 features two subunits a (p10) and b (p27); its role is still unknown. Protein C 100-3, the first ever product cloned from the virus, was used as immunological base for first-generation serological tests; it is an artificial antigen composed of the C-terminal part of protein NS3, protein NS4a and the N-terminal part of protein NS4b (Fig. 1);
- Protein NS5 is also composed of two subunits a (p58) and b (p70); subunit b plays the role of RNA polymerase; it is antigenic and was included in third generation serological tests.

The 3′ non coding region varies in length and sequence from one isolate to the other. It may play a major role in RNA replication.

**HCV multiplication**

Because of the absence of viral culture models, the multiplication cycle of HCV is little known. For instance, the cellular structure which acts as a receptor for the virus is totally unknown. However, some points can be clarified to better understand the pathogenicity of HCV.

As other Flaviviridae, HCV replicates according to a direct RNA-RNA mechanism, without any intermediate passage by a DNA stage and without genomic integration:

- Heparnavirus (HCV): RNA + RNA- RNA + RNA- DNA
- Hepadnavirus (HBV): DNA (integration) => RNA => DNA
- Retrovirus (HIV, HTLV): RNA => DNA (integration) => RNA

Protein NS5 acts as a polymerase. RNA negative strands have been detected in liver, and more questionably in plasma from subjects with chronic infection. Detection of large amounts of negative genomic RNA in liver is astonishing because replication intermediaries are usually highly unstable and very short-lived. Although that observation should be studied further before any definitive conclusions can be drawn, it is tempting to relate it to similar occurrences seen during chronic infection by picornaviruses and which were incriminated as factors of viral persistence. Indeed, in the absence of viral integration in the cellular genome, the mechanism by which a RNA virus persists in the organism is still a mystery. These replication abnormalities, with excess negative strand production, may provide the beginning of an explanation to the persistence of HCV or picornaviruses.

The HCV genomic RNA is of the positive or messenger type, i.e. it can be directly translated at the ribosome level without being transcribed into complementary RNA. It is worth noting that, as was shown for poliovirus and Picornavirus in general, the mRNA of HCV does not need a cap to initiate ribosomal translation, as is the rule for other mRNA in eukaryotic and prokaryotic systems alike. A very well conserved region, localized in the proximal part of the 5′ non coding region and whose secondary structure has the shape of a hairpin, would permit direct binding of the messenger to the ribosome, according to a cap-independent mechanism similar to that demonstrated for poliovirus.

The coding part of the genome is translated into a polyprotein composed of some 3000 amino acids, which is cleaved partly by cellular proteases (especially in the structural part) and partly by the protease activities of proteins NS2 and NS3, into structural and non structural proteins, as described above. By analogy with other Flaviviridae, the newly formed RNA molecules wraps around the procapsid and the new virions leave the cell by budding through the cell membrane.

**Genetic variability of HCV**

As many RNA viruses, HCV exhibits a dramatic genetic variability, apparently related to the high level of transcription errors made by the viral polymerase. In man and in chimpanzee, the rate of spontaneous mutations has been assessed as \( 10^3 \) base substitutions per site per year.

The distribution of mutations on the HCV genome is very heterogeneous. This is due to the fact that the mutations that occur in regions crucial to viral replication are lethal. Also, the immune response by the host exerts high selective pressure, especially on the genes which code for the most exposed antigenic determinants. Many HCV strains have been entirely sequenced and a very large number of isolates have been partially sequenced by polymerase chain reaction (PCR). Together, these results revealed that the viral genomic regions most affected by this variability are, in decreasing order: (i) the N-terminal part of the E2/NS1 region (re Fig. 1) which may be considered as a hypervariable region; (ii) the E1 region, and (iii) the NS3 and NS5b regions. The 5′ NC and NS4b regions are the most
conserved. Nonetheless, the 5' NC region is widely used in practice to study HCV variability. Indeed, within a well-conserved structure, there are two short nucleotide sequences, partially complementary, with a hairpin-like structure, which make it possible to differentiate isolates. It is thus possible to amplify, using PCR, any HCV RNA sequence by using primers selected in the very conserved part of region 5' NC and then, from the same PCR product, to subtype the strain by using specific probes localized in the variable region.

Four hierarchical levels of genetic variability have been identified for HCV: (i) the type (>68% homology), (ii) the subtype (>80% homology), (iii) the isolate, and (iv) the quasi-species. The concept of quasi-species corresponds to the fact that the same viral population is in fact composed of heterogeneous RNA molecules which are nonetheless very close to one another (homology <92% in the region coding for envelope proteins); such differences can be evidenced by sequencing the hypervariable regions.

Several HCV type and subtype classifications have been proposed. Those by Okamoto [3,4], Simmonds [5], and reviewed by Stuyver [6,7] are the most comprehensive. The Simmonds classification is the one most widely used; we shall describe it in detail. Six HCV phenotypes have been reported, numbered 1 to 6. Type I includes three subtypes: 1a (Okamoto's group I), 1b (Okamoto's Group II), and 1c. Type 2 includes at least four subtypes: 2a (Okamoto's Group III), 2b (Okamoto's Group IV), 2c, and 2d. Type 3 includes several subtypes designated as 3a (Okamoto's Group V) to 3f. Type 4 includes many subtypes (4a to 4k). So far, type 5 has been limited to only one subtype (5a). Type 6 is very heterogeneous both genetically and geographically, and gathers strains which Okamoto classified into 5 different genotypes [8].

From the technical standpoint, various methods can be used to type HCV strains [6]. From the most complex to the simplest (several of which are commercially available), there are: (i) direct sequencing of complete genomes or fragments amplified by PCR; (ii) PCR using type- and subtype-specific primers; (iii) PCR using universal primers, followed by digestion of the PCR product by restriction enzymes (restriction fragment length polymorphism, RFLP); (iv) PCR using universal primers, followed by hybridization with type- or subtype-specific probes; and (v) serological typing, which reveals type-specific antibodies.

It should be stressed that the same individual can be simultaneously or successively infected by several types of HCV. This is frequent in haemophiliacs.

Practical consequences of genetic variability [9]

Geographical distribution of HCV genotypes

Genotype 1a, initially described as an American variant, is distributed worldwide. So is genotype 1b (Japanese variant), the most widespread worldwide, and so are types 2 and 3. Genotype 4 is mainly found in Africa and the Middle-East. Type 5 has mainly been isolated in South Africa and Central Africa. Type 6 strains are essentially distributed in the Far East [10].

In France, most isolates belong to subtypes 1a, 1b, 2, and 3a [11]. Genotype 1b is predominant in subjects infected by nosocomial contamination, whereas other subtypes are more commonly found in drug addicts [12].

Clinical correlations

An increasing number of studies have established correlations between the detection of a virus belonging to genotype 1b and (i) more severe clinical manifestations (active chronic hepatitis, cirrhosis), (ii) greater viral concentrations and (iii) weaker response to interferon treatment. Although this brief synopsis does not permit detailed analysis of these interesting findings, it can at least be noted that the origin of these differences is probably multifactorial. The role of epidemiological factors cannot be ruled out, such as the virus inoculation route, the duration of contamination or other hepatic risk factors (alcohol, hepatitis B). However virological (liver tropisms of strains, replication rate) and immunological factors would deserve to be analysed in depth.

Epidemiological implications

Analysing the hypervariable regions (E1, E2/NS1) provided a comparison of isolates and made it possible to establish epidemiological filiation between strains, such as for example: (i) mother-to-child transmission (now well established), (ii) sexual transmission (infrequent but possible), (iii) nosocomial transmission (transfusion, transplantation, haemodialysis, dental surgery, endoscopic investigations...), (iv) intrafamilial transmission, and (v) strain recurrence after liver transplantation.

Vaccinal implications

The development of an anti-HCV vaccine is a target for the next decade. Besides the low immunogenicity of the virus and the difficulty to prevent re-infection, the wide variability of the virus constitutes the main problem for the development of an effective vaccination strategy.

References


8. Tokita H, Okamoto H, Iizuka H, Kishimoto J, Tsuda F, Lesmana LA, Miyakawa Y, Mayumi M. Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2p), tenth (10a) and eleventh (11a) genetic groups. *J Gen Virol* 1996; 77: 293-301.

