Genetic heterogeneity of hepatitis G virus isolates from different parts of the world

S. Viazov,1,2 M. Riffelmann,2 Y. Khoudyakov,3 H. Fields,3 C. Varenholz4 and M. Roggendorf2

1 Ivanovsky Institute of Virology, Gamaleya St 16, Moscow 123098, Russia
2 Institute of Virology, University of Essen, Robert Koch Haus, Hufelandstr. 55, 45122 Essen, Germany
3 Hepatitis Branch, Centers for Disease Control, Atlanta, GA 30333, USA
4 Institute of Hygiene and Occupational Medicine, University of Essen, Hufelandstr. 55, 45122 Essen, Germany

Comparative sequence analysis of a 354 nt fragment of the NS5 region of hepatitis G virus (HGV) isolates was performed to assess two levels of HGV genomic variability: (1) heterogeneity of HGV within an infected individual, and (2) heterogeneity of different HGV isolates. Comparison of nucleotide sequences of DNA clones from two virus isolates demonstrated that in each infected individual HGV is represented by a population of virions with closely related but heterogeneous genomes (quasispecies). Phylogenetic analysis of nucleotide sequences of 42 isolates collected from 14 countries revealed less significant genome variability of HGV as compared to hepatitis C virus. Sequences of all HGV isolates fell into one group of distribution of evolutionary distances. On a phylogenetic tree all HGV sequences segregated into numerous branches. All sequences of isolates from Africa, South and South-East Asia, however, were clustered together and were separated from those of other isolates collected in Europe, North America and Central Asia.

Recently, two independent groups reported the identification of new infectious agents of humans: GBV-C (Simmons et al., 1995b) and HGV (Linnen et al., 1996). Sequence analysis data demonstrated that these viruses are different isolates of the same virus, in this paper provisionally termed hepatitis G virus (HGV). The HGV genome is a positive-strand RNA of about 9400 bases and contains a large open reading frame that encodes a polypeptide precursor of about 2900 amino acids.

Comparative sequence analysis of the genome and encoded polyprotein suggested that HGV is a member of the family Flaviviridae, and is classified into a new distinctive group that also includes hepatitis C virus (HCV), GBV-A and GBV-B viruses (Simmons et al., 1995a). Within this group HGV is a sole representative of its own subgroup.

HGV has been found in several populations: (1) in some patients with acute, fulminant and chronic hepatitis; (2) in individuals at high risk of contracting blood-borne infections (multiple transfused patients, haemophiliacs, intravenous drug users); and (3) in blood donors (Simmons et al., 1995b; Yoshiba et al., 1995; Linnen et al., 1996; Aikawa et al., 1996; Masuko et al., 1996; Nübling & Löwer, 1996; Schmidt et al., 1996). The causative role of HGV in any form of hepatitis has not been demonstrated, and the clinical significance of HGV infection remains unknown. Additional epidemiological and clinical studies are required to clarify this critical question.

HGV is an RNA virus capable of establishing long time persistence in an infected host (Masuko et al., 1996). HGV is related to HCV, which is characterized by substantial genome variability (Bukh et al., 1995; Simmonds, 1995). Since many RNA viruses have a high genome instability attributed to the lack of proof reading functions of the virus-encoded RNA polymerase, one could expect that HGV isolates from different geographical regions would also demonstrate relatively low conservation of at least some parts of the viral genome. In the present paper we report comparative sequence analysis of a 354 nt fragment from the NS5 region of 42 HGV isolates collected in 14 countries.

More than 200 blood samples were tested for the presence of HGV by RT–PCR. In some cases, only the presence of HCV infection was considered as an indication to search for HGV RNA. Overall, 39 HGV RNA positive patients were identified and enrolled in this study: one fulminant hepatitis patient from Moldova; twenty chronic hepatitis patients from Germany, Russia, Belorussia, Moldova, Uzbekistan, Brazil, Indonesia and Vietnam; two intravenous drug users from Spain; four blood donors from Russia; four human immunodeficiency virus infected persons from Cote D'Ivoire, Sierra Leone and Zaire;
Fig. 1. Nucleotide sequences of five clones of the HGV NS5B region derived from the same chronic hepatitis patient (HGV isolate d22). Nucleotides 7271 to 7624; numbering according to Leary et al. (1996).

two haemophiliacs from Russia, and six multiply transfused patients from Germany.

Current genotyping methods used for many viruses including HCV are based mainly on a comparison of partial sequences in specific regions of the viral genome. We aligned three published sequences of GBV-C and HGV (Simmons et al., 1995b; Linnen et al., 1996) and six partial HGV sequences (Y. Khudyakov and others, unpublished). This analysis suggested a variable region flanked by conserved regions within NS5 as suitable for amplification by RT–PCR. RNA was prepared from plasma by guanidine thiocyanate–phenol–chloroform extraction and used for cDNA synthesis as described previously.
Genome heterogeneity of HGV

(Viazov et al., 1994). The primer sequences for the NS5 region of HGV were as follows (derived from the HGV1 sequence; Linnen et al., 1996):

- **YK-877**: nt 7765–7742, 5’ ACCGACACCTAGATCCC-CAGCCC (RT and PCR I)
- **YK-874**: nt 7234–7257, 5’ CTGATGTGTAGCCTGT-GTGAGA (PCR I)
- **YK-1183**: nt 7266–7289, 5’ CAGAACCATACAGCCT-ATTGTGAC (PCR II)
- **YK-876**: nt 7727–7704, 5’ CCTTACAGTCCTTATTGC-TTCCTC (PCR II).

Both rounds of PCR were run with Taq polymerase (Promega) for 35 cycles, each consisting of 1 min at 94 °C, 1 min at 60 °C and 2 min at 72 °C; the cycles were followed by an elongation step at 72 °C for 7 min.

PCR products were purified from agarose gel with the QIAquick Gel Extraction kit (QIAGEN) and cloned into the plasmid vector pCRII (Invitrogen). Nucleotide sequences were determined using an AutoRead Sequencing kit (Pharmacia). Alternatively, purified PCR products were subjected to direct sequencing from both directions using a Prizm Ready Reaction DyeDeoxy Terminator Cycle Sequencing kit (Perkin Elmer). Phylogenetic analysis of sequences was accomplished using the package PHYLIP, version 3.5c (Felsenstein, 1993). Heteroduplex analysis in a temperature gradient gel electrophoresis (TGGE) was described by us earlier (Lu et al., 1995).

PCR products from two individuals, one patient with chronic hepatitis (HGV isolate d22) and one blood donor (HGV isolate d18), were used to investigate the micro-heterogeneity of the HGV genome. These fragments were cloned and 12 randomly chosen clones from each specimen were subjected to heteroduplex analysis in TGGE. All clones tested demonstrated different electrophoretic mobilities (data not shown), which reflect their sequence heterogeneity. Six randomly chosen clones of isolate d22 and five clones of isolate d18 were further sequenced and compared. In both cases very similar results were obtained. DNA clones from each HGV positive subject differed from each other by 1 to 9 nucleotides. Fig. 1 illustrates the nucleotide variations in six DNA clones of HGV isolate d22. In a series of preliminary experiments (Lu et al., 1995) the error rate of our standard technique was established. The rate of nucleotide misincorporation was approximately 10⁻⁵/base/PCR cycle. This low error rate cannot account for a number of nucleotide changes identified in all clones tested. Thus, as expected, in each infected subject HGV was represented by a population of virions with closely related but heterogeneous genomes (quasispecies).

To assess the second level of genetic heterogeneity of HGV, PCR fragments, obtained from 39 isolates of HGV, were subjected to direct sequencing. The resulting nucleotide sequences along with aforementioned individual clones from two HGV isolates (d22 and d18), and published nucleotide sequences of the GBV-C and HGV prototype strains (overall 51 sequences), were studied by phylogenetic analysis. Pairwise comparison revealed a wide range of evolutionary distances; however, the distribution of distances was confined to two separate and non-overlapping peaks (Fig. 2). The first ranged from 0 to 0.05 and reflects the differences between clones of

---

**Fig. 2.** Distribution of evolutionary distances upon pairwise comparison of 51 nucleotide sequences of HGV isolates in the NS5 region. Number of calculated evolutionary distance measurements is recorded on the x-axis. Number of observed sequence similarities is recorded on the y-axis.
the same isolates d22 and d18, while the second showed a range of 0.09 to 0.23. Distribution of sequences of all HGV isolates fell into the last group suggesting that all isolates belong to the same type (subtype). Nucleotide changes were distributed evenly throughout the DNA fragments and most of these changes were located in the third position of each codon.

There are several explanations for these results. First, HGV circulating in nature is represented predominantly or exclusively by one virus type or closely related subtypes. The phylogenetic tree (Fig. 3), constructed on the basis of established genetic relatedness, illustrates this notion. HGV sequences segregated into numerous branches. All sequences from Africa and South and South-East Asia, however, are clustered together and are separated from sequences of other isolates collected in Europe, North America and Central Asia. Considering that phylogenetic analysis was performed after bootstrap re-sampling (1000 replicates), this difference is significant and might suggest separate evolutionary lineages. Of importance is the fact that the established HGV genome diversity (0.09 to 0.23) is somewhat intermediate between variability of isolates (0±0 to 0±12) and subtypes (0±16 to 0±32) of HCV in the NS5 region (Simmonds et al., 1993). Reasons for such a difference remain unclear.

The second explanation could be that HGV types or subtypes may exist but the chosen fragment from the NS5 gene is not optimal for a study of HGV genome variability. Sequence analysis of HCV, which is closely related to HGV, demonstrated that distinct genotype specific differences could be found throughout the HCV genome and that any coding region could be used for identification of genotypes (Bukh et al., 1995; Simmonds, 1995). Recently, we have sequenced a 226 bp region of the 5'UTR of 16 of our HGV isolates, collected in Europe, USA, Africa and South Asia. Evolutionary
distances established for these isolates fell into the range (0.01 to 0.14) which is characteristic of HCV isolates (Simmonds et al., 1993; Simmonds, 1995). Thus, these preliminary results support our observations for the NS5 region of the HGV genome and suggest the absence of HGV genotypes.

Third, the PCR primers used in this study are suboptimal and could only identify a particular HGV type (subtype). This explanation, however, seems unlikely as (1) all isolates tested were collected in different parts of the world on a random basis, and (2) in our laboratories more than 200 samples were tested for HGV in parallel by PCR with primers from the NS5, NS3 and 5′UTR regions of the HGV genome. Not a single specimen was found to be positive with primers from the 5′UTR or NS3 and negative with primers from the NS5 region. On the other hand, no more than 70% of the samples identified with the NS5 set were found to be positive with NS3 or 5′UTR primers (Y. Khoudyakov and others, unpublished).

Presently, it is difficult to choose between these three explanations. This uncertainty may be resolved by increasing the number of samples from different parts of the world and by a comparative analysis of variability of different fragments of the HGV genome including envelope protein genes. Still, our data suggest that HGV most probably is not characterized by a genome variability as high as that of HCV. This dissimilarity could arise from peculiarities of infections caused by these two viruses. HCV infection in a significant proportion of cases is associated with chronic liver disease. Data available for HGV (Alter, 1996; Masuko et al., 1996; Wang et al., 1996) suggest that most subjects infected with this virus usually do not experience a liver disease. Numerous attempts to detect antiviral antibodies in HGV infected individuals have been unsuccessful, which may reflect a diminished or impaired immune response to HGV proteins. In other words, most probably the mechanisms of HCV and HGV interactions with the infected host are different, which in turn may lead to different evolutionary pathways for these two viruses.

We are very grateful to Dr Lieven Stuyver for his help in phylogenetic analysis and Mrs Susanne Standar for help in DNA sequencing. We gratefully acknowledge the helpful discussion and critical comments of Dr Peter Simmonds.

References


Received 21 October 1996; Accepted 11 November 1996