Full-length open reading frame of a recombinant hepatitis C virus strain from St Petersburg: proposed mechanism for its formation

Olga Kalinina,1,2 Helene Norder1 and Lars O. Magnius1

1Department of Virology, Swedish Institute for Infectious Disease Control, 171 82 Solna, Sweden
2St Petersburg Pasteur Institute, 14 Mira Street, St Petersburg 197101, Russia

The full-length ORFs for the hepatitis C virus recombinant RF1_2k/1b (N687) and the non-recombinant 1b strain N589 were sequenced. A single recombination point was found and the sizes of the genes (C, E1, E2, p7, NS2, NS3, NS4 and NS5) were according to the parental subtypes. The PKR-eIF2α phosphorylation site homology domain sequence of the E2 protein was identical to those of genotype 2 strains, while the IFN-α-sensitivity-determining region of the NS5A protein was identical to those of interferon-resistant 1b strains. For the parental strains, two hairpin structures, HS1 and HS2, were predicted for the plus-strand up- and downstream of the crossover site, which were not present in the recombinant strain. HS2 shared similarity with the motif1 hairpin of turnip crinkle virus RNA that binds to the RNA-dependent RNA polymerase and facilitates 3’-terminal extension during recombination. This study suggests that RF1_2k/1b has emerged by homologous recombination during minus-strand synthesis via template switching because of constraints imposed by the HS1 hairpin of the 3’-parental genome.

Hepatitis C virus (HCV) is an RNA virus belonging to the family Flaviviridae. It has a single-stranded plus-sense genome of approximately 9.6 kb with a single open reading frame (ORF) and three structural (C, E1 and E2) and seven non-structural genes (p7, NS2, NS3, NS4A, NS4b, NS5A and NS5B) (de Francesco, 1999). HCV exhibits a high genetic variability, enabling its classification into four hierarchical strata: clades, genotypes, subtypes and isolates (Robertson et al., 1998; Simmonds, 1999). A high mutation rate is considered to be a major factor behind its genetic diversity (Simmonds, 1999, 2001; Farci & Purcell, 2000). Recombination provides another mechanism for creating genetic variation and is known to be widespread among non-segmented RNA viruses, including other members of Flaviviridae (Lai, 1992; Figlerowicz et al., 1997; Worobey et al., 1999; Guilhot et al., 2000; Twiddy & Holmes, 2003).

Some viruses, such as coronaviruses, bromoviruses and carmoviruses, for example turnip crinkle virus (TCV), may undergo both homologous and non-homologous recombination, while some non-segmented plus-strand RNA viruses, such as picornaviruses, may only undergo homologous recombination (Lai, 1992). Recombination was not recognized as a likely mechanism for creating diversity of HCV until our recent finding of a spontaneous HCV recombinant in St Petersburg, RF1_2k/1b (Kalinina et al., 2002).

The virus genotype is one factor influencing the outcome of interferon (IFN) therapy (Zein, 2000). Infections with genotype 1 are more resistant to IFN therapy than infections with genotypes 2 and 3 (Taylor et al., 2000). Two domains, the PePHD (PKR-eIF2α phosphorylation site homology domain) in the E2 protein and the ISDR (IFN-α-sensitivity-determining region) in the NS5A protein, are claimed to mediate IFN resistance through inhibition of the PKR (an IFN-α-induced, double-stranded RNA-activated protein kinase), although the importance of the ISDR is still under debate (Enomoto et al., 1996; Gerotto et al., 1999; Taylor, 2001; Taylor et al., 1999, 2000, 2001). There is evidence that only the E2 protein of genotype 1 can bind and inhibit PKR in vitro because of the high similarity that the PePHD sequence of this genotype has to the PKR sequence (Taylor et al., 1999; Polyak et al., 2000). The genetic diversity of HCV is therefore critical in our understanding of therapeutic modules against hepatitis C (Farci & Purcell, 2000).

We have determined the full-length ORFs of two HCV strains, the recombinant RF1_2k/1b (N687) and the most similar non-recombinant subtype 1b strain N589, previously published in part (Kalinina et al., 2002), and predicted the secondary RNA structures around the crossover point for the parental strains to reconstruct the mechanism of recombination.

RNA extraction, cDNA synthesis, strategy of amplification...
and sequencing of four overlapping fragments covering the E2–NS3, NS3–NS5A, NS3–NS5B and NS5B regions, at positions 2500–5157, 4036–6815, 5002–8638 and 8168–9396, respectively, were performed mainly as described previously (Kalinina et al., 2002). Overlapping PCR products were sequenced in both directions with internal primers and the primers used in the PCR (Kalinina et al., 2002). PePHD and ISDR domains for strains N687 and N589 were compared with six strains from GenBank, k1-s2 (subtype 1b; ‘intermediate type’), HC-J4 (subtype 1b; ‘wild-type’), VAT96 (subtype 2k), HC-J6 (subtype 2a), CB (subtype 3a) and NZL1 (subtype 3a). Similarity and bootscanning analyses were performed with the SIMPLOT program, version 1.2.2 for Windows 95/NT, available from the author S. C. Ray (http://sray.med.som.jhmi.edu/RaySoft/SimPlot/; Lole et al., 1999) and according to Salminen & Cobb (1998).

The deduced polyprotein of the recombinant strain, N687, consisted of 3014 aa with protein sizes according to the parental subtypes (C, 191 aa; E1, 192 aa; E2-HVR, 367 aa; p7, 63 aa; NS2, 217 aa; NS3, 631 aa; NS4A, 54 aa; NS4B, 261 aa; NS5A, 447 aa; and NS5B, 591 aa) without any deletion or insertion related to the recombination site. As previously reported, the recombination point mapped within the NS2 region, most likely at positions 3175–3176 (Kalinina et al., 2002). This was the only recombination point found in the full-length ORF.

The genetic organization of N589 was similar to that of other HCV subtype 1b strains. As for HC-J4, its putative polyprotein consisted of 3010 aa with protein sizes according to the parental subtypes (C, 191 aa; E1, 192 aa; E2-HVR, 363 aa; p7, 63 aa; NS2, 217 aa; NS3, 631 aa; NS4A, 54 aa; NS4B, 261 aa; NS5A, 447 aa; and NS5B, 591 aa).

PePHD sequences of the recombinant, N687, and 1b strain, N589, were similar to those of genotype 2 and subtype 1b strains, respectively (Fig. 1). Thus, the E2 protein of the recombinant strain should not be able to bind PKR and confer IFN resistance. The ISDR sequence of the recombinant strain was identical to that of the ‘wild-type’ subtype 1b strain, N589, and 1b strain, N589, had an ISDR sequence corresponding to that of the ‘intermediate type’ of Enomoto (1996). The 1b strain, N589, had an ISDR sequence corresponding to the ‘intermediate type’ of Enomoto et al. (1996). So far, patients infected with the recombinant have not been subject to IFN therapy, which would contribute to the understanding of the relative importance of the PePHD and the ISDR regions for the outcome of such therapy.

To investigate the presence of stable secondary RNA structures close to the crossover junction, RNA folding predictions were performed for the first 6000 nt of the genomes and within the NS2–NS3 regions of eight HCV strains (N687, N589, k1-s2, HC-Jta, MD4-1, VAT96, HC-J6 and CB) using the MFOLD program version 3.1 with standard parameters via the MFOLD server provided by Michael Zuker, Rensselaer Polytechnic Institute (http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi; Mathews et al., 1999). The NS2–NS3 region was analysed in 800 nt fragments with 400 nt overlap. Regions around the crossover point were also examined using a series of 200 nt fragments increasing in length stepwise by 20–30 nt. Additionally, RNA folding predictions were performed as described above for the minus-strand sequences of three HCV strains, N687, N589 and VAT96.

A stable hairpin structure, HS1, upstream of the crossover point was predicted for three subtype 1b strains and one la strain but not for the recombinant or for other subtypes (Fig. 2a). This structure involves 58 to 88 nt depending on the strain. Two substitutions, G3110A and T3149C, present in the N589 strain compared with k1-s2 abolished two internal loops, whereby a longer stem was formed with a low p-num value (Fig. 2a). Another hairpin structure, HS2, was predicted downstream of the crossover point, formed by 32 or 60 nt for the subtype 2a and 2k strains, respectively (Fig. 2b). Five substitutions, G3218, C3224, T3239, A3245 and T3251, in the sequence of subtype 2k strain, VAT96, were critical for the formation of a longer stable hairpin structure of 60 nt with a low p-num value (Fig. 2b). The hairpin structure of 60 bases contained two bulges (I and II) and one asymmetrical interior loop or U-bulge (III) with a C opposite four U residues (Fig. 2b). No hairpin structures similar to HS1 and HS2 were predicted close to the crossover junction for the minus-strand sequences of strains N687, N589 or VAT96.

<table>
<thead>
<tr>
<th>(a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b)</td>
</tr>
</tbody>
</table>

**Fig. 1.** Alignment of amino acid sequences within the E2 (a) and NS5A (b) regions. HCV sequences retrieved from GenBank are indicated by the name of the strain or clone; HC-J4 (AF054247), k1-s2 (D50485), VAT96 (AB031663), HCJ6 (D00944), CB (AF046866), NZL1 (D14306). Recombinant strains are indicated by the letter R and shown in bold. Dotted squares show the PePHD (a) and the ISDR regions (b). Nucleotide positions are numbered according to the subtype 2a strain, pJ6CF.
The preserved genomic structure of the recombinant shows that it has resulted from a homologous recombination, which occurs when two related RNA molecules recombine at corresponding sites. Most homologous recombinants are formed by a template switching mechanism in two steps. (i) Synthesis of a nascent RNA with the donor RNA as template, the transcriptional pausing of the RNA-dependent RNA polymerase (RdRp) and the dissociation of the transcriptional machinery from the donor RNA template; and (ii) transfer of the transcriptional complex with the nascent RNA and the binding of the RdRp to the acceptor RNA followed by the elongation of the nascent RNA in the acceptor RNA (Lai, 1992; White & Morris, 1995; Figlerowicz et al., 1997; Nagy & Simon, 1998a; Nagy et al., 1998). RdRps usually pause at regions with strong secondary structure situated in the template or sometimes in the nascent RNA (Lai, 1992).

In this study, HS1 and HS2 were predicted around the crossover junction for both parental strains but not for the recombinant. Thus, transcriptional pausing of RdRp might be caused by the secondary structure of either parental template. To achieve template switching, the RdRp should be able to bind to the acceptor RNA template. In the TCV system, an efficient 3’-terminal extension on the acceptor RNA was found to depend on several factors. (i) A hairpin structure (designated the motif1 hairpin) on the acceptor genome; (ii) a short base-paired region formed between the acceptor genome and the nascent RNA designated the priming stem; (iii) a spacer region and a U-rich sequence located 5’ of the motif1 hairpin that increases the efficacy of 3’-terminal extensions; and (iv) the inability of the RdRp to re-initiate synthesis within the double-stranded region (White & Morris, 1995; Nagy & Simon, 1998a, b; Nagy et al., 1998). The motif1 hairpin binds to the RdRp and may act as an enhancer to recruit the RdRp to the acceptor genome (Nagy & Simon, 1998a, b; Nagy et al., 1998). The possible role of the priming stem is to keep the 3’-terminus near the RdRp bound to the hairpin structure and to facilitate 3’-terminal extension. However, in the absence of the motif1 hairpin, the recognition of the priming stem by TCV RdRp is inefficient (Nagy et al., 1998). HS2 predicted herein for subtype 2k matches the motif1 hairpin predicted for TCV (Fig. 2B). Similar to HS2, the motif1 hairpin has a tetra-loop cap and a U-bulge, which may play a role in recombination (Nagy & Simon, 1998a). Deletion of three U residues was found to reduce 3’-terminal extension activity and eliminate in vivo recombination, although it was also shown that the TCV RdRp is likely to recognize the secondary structure of the motif1 hairpin, where individual nucleotides play a minor role (Nagy & Simon, 1998a; Nagy et al., 1998). Recently, it was revealed that the RdRps of HCV, bovine viral diarrhea virus and GB virus-B and the replicase complexes of plant-infecting RNA viruses, such as TCV or brome mosaic bromovirus, may initiate RNA synthesis by a common de novo mechanism, although with distinct preferences for template and substrate initiation nucleotides (Ranjith-Kumar et al., 2002). This suggests that HS2 has an equal role in HCV recombination as the motif1 hairpin and is involved in the recruitment of the RdRp as an enhancer. The sequence between HS1 and HS2 (nucleotide positions 3175–3202) did not form a double-stranded stem in any of the templates. This fragment has a short stretch (nucleotide positions 3177–3188) that is conserved for all genotypes and might serve as a primer during re-initiation of RNA synthesis on the acceptor RNA. HCV RdRp is shown to be capable of non-sequence-specific primer extension.

**Fig. 2.** Secondary structures HS1 (a) and HS2 (b) predicted by MFOLD. Circles, squares and triangles show positions with high influence on the stability of long stems. A dotted circle shows a substitution in the Md4-1 sequence resulting in the formation of an additional internal loop.
independent of the presence of promoter sequences (Behrens et al., 1996). In addition, there was a poly(A) stretch (nucleotide positions 3270–3275) in the subtype 2k strain template that could play a role as a promoter. Fig. 3 shows the proposed mechanism for formation of the recombinant, RFI_2k/1b.

Despite the fact that HCV synthesis of the plus-strand RNA occurs more frequently than that of the minus-strand RNA (Gastaldi et al., 1995; Nouri-Aria et al., 1993), it seems likely that this recombinant has occurred during minus-strand synthesis, where the subtype 1b parental strain provided the donor template and the subtype 2k strain the acceptor template. This is supported by experimental data showing that extension by HCV RdRp starting at the 3’ UTR minus-strand RNA is more efficient than extension starting at the 3’ UTR plus-strand RNA (Reigadas et al., 2001). Heterogeneous products shorter than the template were formed with initiation at the 3’ UTR of the plus-strand RNA, while products the same size as templates were formed with initiation at the 3’ UTR of the minus-strand RNA (Reigadas et al., 2001). The prediction of a hairpin structure with a U-bulge downstream of the crossover junction only for subtype 2k strain offers a plausible explanation for why this rare subtype with restricted distribution was one of the parental strains.

Fig. 3. Schematic representation of a plausible mechanism for HCV recombination. (a) Synthesis of nascent RNA on the RNA-donor template and transcriptional pausing of the RdRp caused by HS1. (b) Dissociation of the transcriptional complex including the nascent RNA from the RNA-donor template. (c) Binding of RdRp with the transcriptional complex to HS2 on the acceptor RNA followed by elongation of the nascent RNA on the RNA-acceptor template. The RdRp is shown as a filled oval. Stable hairpin structures are indicated by HS. The conserved stretch at nucleotide positions 3177–3188 serving as primer during re-initiation of RNA synthesis is shown in bold.

**Acknowledgements**

This work was supported by grant G-126 from Östeuropa Kommitetén.

**References**


