Cloning and characterization of a complete open reading frame of the hepatitis C virus genome in only two cDNA fragments

Kay Rispeter, Mengji Lu, Sabine Lechner, Andree Zibert and Michael Roggendorf

The synthesis of long cDNA molecules encoding the complete genome of RNA viruses has recently been demonstrated; this major improvement has numerous practical applications such as construction of infectious cDNA clones or study of sequence variability at the level of a single RNA molecule. Using hepatitis C virus (HCV) as a model, we established an RT–PCR technique for amplification of cDNA fragments with a length of about 5 kb. The RT reaction was carried out with a Moloney murine leukaemia virus reverse transcriptase lacking detectable RNase H activity. For PCR reactions an enzyme mix containing Taq and Pwo DNA polymerases was used. Hot start and addition of 5% DMSO were also important to efficiently achieve long PCR products.

About 10^6 HCV genome equivalents/ml in serum were needed in order to amplify the HCV genome in only two cDNA fragments covering about 98% of the complete genome. Analysis of the HCV quasi-species is also possible by this method as shown by sequencing of the hypervariable region 1 (HVR1) after cloning of cDNAs. The integrity of the long cDNA clones was proven by (1) restriction analyses, (2) partial sequencing and (3) expression of respective gene products. In vitro transcribed cDNAs were translated in rabbit reticulocyte lysate. Structural and nonstructural HCV proteins were identified by immunoprecipitation using patient serum. These results suggest that the two cDNA clones encode a complete and functional open reading frame of HCV.

Introduction

Hepatitis C virus (HCV) is the major agent of post-transfusion associated hepatitis (Choo et al., 1989). HCV infection often leads to virus persistence, resulting in severe chronic liver disease which can be followed by development of hepatocellular carcinoma (Saito et al., 1990). HCV has been classified on the basis of its genomic organization and the hydrophobicity profiles of its open reading frame (ORF) as a new genus in the family Flaviviridae, which includes two other genera, the flaviviruses and the pestiviruses (Francki et al., 1991). Characteristic for all members of the Flaviviridae is an envelope, a positive-strand RNA genome and a single ORF. The HCV genome is about 9·6 kb in length and contains a highly conserved 5’ nontranslated region (Kato et al., 1990; Choo et al., 1991; Bukh et al., 1992). The long ORF is followed by a 3’ nontranslated region. This region comprises a short sequence of 27–66 bases (Kato et al., 1990; Choo et al., 1991; Bukh et al., 1992) which is poorly conserved, a poly(U)- (Kato et al., 1990; Takamizawa et al., 1991) or a poly(A)-tract (Han et al., 1991), followed by a recently discovered highly conserved sequence of 98 bases (Tanaka et al., 1995; Kolykhalov et al., 1996).

The single ORF of HCV encodes a precursor polyprotein of about 3000 amino acids (Kato et al., 1990; Choo et al., 1991; Takamizawa et al., 1991; Okamoto et al., 1992a; Tanaka et al., 1992). HCV codes for three putative structural proteins: a nucleocapsid protein [C (core)] and two envelope proteins (E1 and E2). These proteins and at least six nonstructural proteins are generated after cleavage by host and viral proteases (Hijikata et al., 1991; Grakoui et al., 1993; Bartenschlager et al., 1994).

About 25 different complete nucleotide sequences of HCV have been published up to now (Bukh et al., 1995; Lamballerie et al., 1997). These complete sequences together with individual regions of the HCV genome have revealed that isolates of HCV differ substantially. HCV isolates have been grouped into at least six genotypes and a large number of subtypes according to their nucleotide sequences (Okamoto et
al., 1992b; Bukh et al., 1993; Simmonds, 1995). In addition to this heterogeneity of isolates found in different patients, a variety of different HCV nucleotide sequences (quasispecies) have been observed in the same patient (Martell et al., 1992), a phenomenon also known for other RNA viruses (Domingo et al., 1985).

Investigations on the HCV genome, e.g. detection of mutations within the same RNA molecule, are hampered by current techniques of RT–PCR. With conventional RT–PCR only short cDNA fragments (about 0.8–2.0 kb) can be amplified; these are probably generated from different genomes found in a given serum sample. Recently, similar difficulties for the DNA genome of hepatitis B virus have been resolved by the establishment of a full-length PCR (Günther et al., 1995).

Conventional RT–PCR using, e.g., Moloney murine leukaemia virus (MMLV) reverse transcriptase and Thermus aquaticus (Taq) DNA polymerase, is a widely used method for detection of HCV RNA in patient sera (Cuypers et al., 1991; Bukh et al., 1992). Recently, mixtures of thermostable Taq DNA polymerase and Pwo DNA polymerase containing a 3′–5′ proofreading activity have been used to perform PCR for the amplification of long DNA fragments of 20–40 kb from human genomic DNA or bacteriophage λ DNA as templates (Barnes, 1994; Cheng et al., 1994). We established an RT–PCR protocol employing this new approach for amplification of HCV cDNA fragments up to a length of about 5 kb. The integrity of these cDNAs was demonstrated by restriction analyses, partial sequencing and expression of structural and nonstructural HCV proteins.

Methods

**Samples.** Sera from five patients (P1–P5) with chronic hepatitis C were used for RT–PCR. These patients were infected in 1978/79 (Dittmann et al., 1991) by a well-characterized single HCV isolate (HCV-AD78) of genotype 1b present in an HCV-contaminated anti-D immunoglobulin preparation (Roggendorf et al., 1992; Zibert et al., 1995). Sera were obtained 10 years after infection. Virus titres were determined by a branched DNA assay (Quantiplex HCV-RNA Assay, version 1; Chiron).

**HCV RNA extraction.** This was done as described previously (Lu et al., 1995) with a few modifications. Sera (300 µl) were incubated in 50 mM Tris–HCl pH 8, 10 mM EDTA, 100 mM NaCl, 4% SDS and 3 mg/ml proteinase K (Boehringer Mannheim) in a total volume of 500 µl for 1 h at 37 °C. Yeast tRNA (1 µg) of was added to improve RNA yields. After extraction with phenol and phenol–chloroform, respectively, total RNA was precipitated by ethanol, dried in air and then dissolved in 50 µl of water.

**RT–PCR.** Primers were chosen and numbered according to the sequence of HCV isolate J (Kato et al., 1990) (Table 1). RT reactions were carried out by using antisense primers A3324, A5387 or A9412, respectively. A 10 µl volume of extracted HCV RNA was reverse transcribed for 60 min at 48 °C using 50 units of recombinant Expand RT (Boehringer Mannheim) and 2.5 mM primer. The samples were then heated for 5 min to 95 °C. cDNA fragments were amplified by nested PCR with the Expand High Fidelity PCR System (Boehringer Mannheim). The PCR reactions were performed with 5 µl of template in a total volume of 50 µl. Different cycling profiles were used depending on the target length. Generally, the PCR profiles consisted of a hot start at 94 °C, an initial denaturation at 94 °C for 2 min, a denaturation at 94 or 92 °C (with DMSO) for 20 s, an annealing temperature between 40 and 55 °C for 90 s and an extension step at 68 °C for 1 min per kb of amplified cDNA. After 30 cycles a final extension was carried out at 68 °C for 10 min. All amplifications were performed in a Thermoblock TR-1 (Biometra). PCR products were examined by agarose gel electrophoresis and staining with ethidium bromide.

**Cloning and DNA sequencing.** The two PCR products obtained with primers described in the text were cloned into vector pCRII (Invitrogen). Clones were isolated after transformation of E. coli (INV alpha F) and named p17-5310 and p4497-9401. Plasmid purification and agarose gel electrophoresis were carried out as described by Sambrook et al. (1989). DNA sequencing was performed by the dideoxy chain termination method (Sanger et al., 1977) with Sequenase (USB) after alkaline denaturation of plasmid DNA (Chen et al., 1985).

**In vitro transcription and translation.** Templates for in vitro transcription and translation were generated by amplification of sub-length PCR products using clones p17-5310 and p4497-9401 as templates. Each sense primer contained an identical sequence at its 5′ end including the sequence of the bacteriophage T7 promoter (underlined) (CAAGCTTCAATAAGACTCTACATAGGGACC) followed by an HCV-specific sequence at its 3′ end. The HCV-specific sequences were preceded by a start codon (ATG). The first HCV-specific nucleotides of these primers were at nt 330 (core), 903 (E1), 2757 (NS2), 3408 (NS3), 5300 (NS4A), 6246 (NS5A) and 7587 (NS5B). Antisense primers (see Fig.

### Table 1. Primers for amplification of the HCV genome

<table>
<thead>
<tr>
<th>Primer*</th>
<th>Sequence (5′–3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S6+</td>
<td>GGGCACAACCTCACCACATAGAT</td>
</tr>
<tr>
<td>S17+</td>
<td>ACCATAGATCCTCCCCITTG</td>
</tr>
<tr>
<td>S279</td>
<td>ACTGCGCTGATAGGCTGCTTGG</td>
</tr>
<tr>
<td>S330</td>
<td>ACCATAGCGCAAATCCTAAA</td>
</tr>
<tr>
<td>S1412+</td>
<td>CATTCCTCATGTTGGAAACTTG</td>
</tr>
<tr>
<td>S2524</td>
<td>GTCCCTGCTTGTGATGATGATG</td>
</tr>
<tr>
<td>S3240</td>
<td>GTAGAAGCCGCTTCTTTC</td>
</tr>
<tr>
<td>S4477+</td>
<td>AGGCCCACAAAGGCGGAG</td>
</tr>
<tr>
<td>S4497+</td>
<td>CATCTATCTCTGCACTTCAA</td>
</tr>
<tr>
<td>A9412+</td>
<td>AAAAAAAAAAGGGAATGGCATAT</td>
</tr>
<tr>
<td>A9401+</td>
<td>AGGGAATCCCTATTTGGCCTGG</td>
</tr>
<tr>
<td>A5387+</td>
<td>GATCTCCTGCCCCACGACAC</td>
</tr>
<tr>
<td>A5310+</td>
<td>CCCAGGTCCTAGTGACGAC</td>
</tr>
<tr>
<td>A3735</td>
<td>GCTCTGGAACAAAGTAAAAG</td>
</tr>
<tr>
<td>A3324</td>
<td>CCGAGATGATGCTCCCACGCC</td>
</tr>
<tr>
<td>A3287</td>
<td>CCGAGTATGAGAATTGACTTGG</td>
</tr>
<tr>
<td>A329</td>
<td>GTACGACCTGTCATAGGAC</td>
</tr>
<tr>
<td>A295</td>
<td>CAAAGCACCTATACAGCAGT</td>
</tr>
</tbody>
</table>

* The polarity of primers is indicated (S, sense; A, antisense).

Numbers indicate the nucleotide position of the respective 5′ end (Kato et al., 1990).
† Primers used for amplification of cDNA fragment nt 17–5310.
‡ Primers used for amplification of cDNA fragment nt 4497–9401.
were chosen according to isolate HCV-J (Kato et al., 1990). RNAs were transcribed from PCR products with T7 RNA polymerase (Gibco BRL) for 60 min at 37 °C as described by the manufacturer. The resulting RNAs were used for in vitro translation in rabbit reticulocyte lysate (Promega). In vitro translation was carried out at 30 °C for 60 min using 20 μCi $\left[^{35}S\right]$methionine (Amersham) in a total volume of 26 μl. Proteins were analysed on 12% SDS–polyacrylamide gels. The intensity of autoradiography was enhanced by fluorography using 1 M salicylic acid. Gels were dried and exposed to Kodak X-OMAT film.

Immunoprecipitation. In vitro translation mixtures (5–10 μl) were diluted with 250 μl of buffer A (2% Triton X-100, 50 mM Tris–HCl pH 8.0 and 5 mM EDTA). Serum (1 μl) obtained from a chronic HCV patient was added. Incubation time was 2 h at 4 °C. Protein A–Sepharose (Pharmacia) (20 μg bed volume) was then added to the immune complexes and incubated for 30 min with vigorous shaking. The protein A–Sepharose complexes were washed three times with buffer A and once with buffer A lacking Triton X-100. Retained immunocomplexes were analysed by SDS–PAGE followed by autoradiography.

Results

Generation of long cDNA transcripts

We first optimized the synthesis of cDNA by reverse transcription of HCV RNA extracted from sera of HCV-infected patients. Primer A9412, which starts within the poly(U)-stretch at the 3′ end of the HCV genome, was used for reverse transcription. Different reverse transcriptases and incubation temperatures were compared (Fig. 1). The size of the obtained cDNA was estimated by conventional amplification of short PCR products. These PCR products were generated with primers located at different distances towards

![Diagram](image)

**Fig. 1.** (a) Depiction of PCR products obtained after RT with primer A9412. Positions of PCR primers for the second round of nested PCR are given. (b) Analysis of amplified products by agarose gel analysis. RT was performed with MMLV (lanes 1 and 2) or with RNase H negative reverse transcriptase (lanes 3–8) at different temperatures as indicated. Expected PCR results: fragments of 814 bp (S4497/A5310; lanes 1–4), 496 bp (S3240/A3735; lanes 5 and 6) and 764 bp (S2524/A3287; lanes 7 and 8). M, marker DNA (bp).

<table>
<thead>
<tr>
<th>HCV genome nt 17–295</th>
<th>nt 330–3287</th>
<th>nt 17–5310</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient equivalents/ml</td>
<td>279 bp†</td>
<td>2959 bp§</td>
</tr>
<tr>
<td>P1+</td>
<td>$1 \times 10^6$</td>
<td>+</td>
</tr>
<tr>
<td>P2</td>
<td>$7 \times 10^5$</td>
<td>+</td>
</tr>
<tr>
<td>P3</td>
<td>$3 \times 10^5$</td>
<td>+</td>
</tr>
<tr>
<td>P4</td>
<td>$3 \times 10^5$</td>
<td>+</td>
</tr>
<tr>
<td>P5</td>
<td>$3 \times 10^5$</td>
<td>+</td>
</tr>
</tbody>
</table>

* HCV RNA titre was determined by branched DNA assay.
† RT was done with primer A329.
§ RT was done with primer A3324.
¶ RT was done with primer A5387.
+ Positive RT–PCR results are indicated by ‘+’, negative results by ‘−’.
the 5′ end of the cDNA transcript. Using a cDNA generated with RNase H negative reverse transcriptase at an incubation temperature of 48 °C, we were able to amplify fragments of the HCV genome comprising nt 3240–3735 (Fig. 1b, lane 6) and nt 2524–3287 (Fig. 1b, lane 8). This result indicates that the cDNA included nt position 2524 of the HCV genome and was therefore at least 6881 bp in length. The transcript generated with the same enzyme at an incubation temperature of 42 °C was shorter with a length of at least 4908 bases, since amplification was successful with primer pair S4497 and A5310 (Fig. 1b, lane 3) but not with primer pair S3240 and A3735 (Fig. 1b, lane 5). Using conventional MMLV reverse transcriptase, we were not able to amplify a fragment comprising nt 4497–5310 (Fig. 1b, lanes 1 and 2), indicating that the respective cDNA terminates before nt 4497 of the HCV genome and was less than 4908 bases in length. After reverse transcription with RNase H negative reverse transcriptase heating the sample for 5 min to 95 °C was necessary. Omission of this step resulted in shorter cDNAs (data not shown). The optimized conditions for improvement of cDNA synthesis, i.e. RNase H negative reverse transcriptase and a high incubation temperature of 48 °C, were used for all subsequent experiments.

**HCV RNA titre and long RT–PCR**

In the next experiment we addressed the question which titre of HCV RNA is needed to generate long cDNA fragments. A mixture of Taq and Pwo DNA polymerases was used for PCR. Five patient sera containing different titres of HCV RNA were tested for generation of PCR products of different lengths (Table 2). All patients were infected by the same HCV isolate (HCV-AD78), which is of genotype 1b (Zibert et al., 1995). The HCV RNA titres in these sera were determined by branched DNA assay and ranged from 3·0–10^5 to 1·0·10^6 genome equivalents/ml (Table 2). All five sera were tested HCV positive by amplification of a 279 bp fragment (nt 17–295) from the 5′ nontranslated region of the HCV genome. In contrast, an HCV cDNA fragment of 2959 bp (nt 330–3287) could be amplified from sera of two patients (P1 and P2) only. These two patients had the highest HCV RNA titres (1·0·10^5 and 7·4·10^5 HCV genome equivalents/ml, respectively). A longer cDNA fragment of about 5 kb (nt 17–5310) could be amplified only from serum of patient P1, which had the highest titre of viral RNA.

**Amplification of the complete HCV ORF in two PCR products and characterization of long cDNA fragments**

Two long cDNA fragments comprising regions nt 17–5310 and nt 4497–9401 of the HCV genome were reproducibly amplified (Fig. 2) using RNA extracted from serum of patient P1. RT reactions were carried out with primers A5387 and A9412, respectively. The long PCR products were generated using a combination of Taq and Pwo DNA polymerases but not with either Taq or Pwo alone (data not shown). Modifications of PCR conditions by addition of 5% DMSO and performing a ‘hot start’ improved the results of long PCR as judged by
Complete HCV ORF in two cDNA fragments

Fig. 4. Alignment of deduced amino acid sequences of the HVR1 (boxed) obtained from 12 independently cloned cDNAs (nt 17–5310) derived from serum of patient P1. Horizontal bars indicate sequence identity. The sequences are grouped into the variants HVR1.A, B, C, D and E. Amino acids are denoted by the standard single-letter code. Numbers at the top of the figure are according to HCV-J (Kato et al., 1990). Names of the cDNA clones are given as well as the sequence of HCV-J and HCV-AD78 (Höhne et al., 1994; Zibert et al., 1997).

ethidium bromide staining of agarose gels (Fig. 2, lanes 2–4). The amounts of PCR products increased approximately fivefold on average after addition of DMSO (Fig. 2, lanes 3 and 4). The two cDNA fragments were cloned, and the plasmids were named p17-5310 and p4497-9401, respectively.

The integrity of the two plasmids p17-5310 and p4497-9401 was first investigated by amplification of shorter PCR products using primers corresponding to different regions of the HCV genome. The two plasmids were used as templates. As a control, these short PCR fragments were generated in parallel by conventional RT–PCR using RNA from patient P1. The short PCR products obtained from p17-5310 and p4497-9401 did not show gross deletions when compared with cDNA fragments derived by conventional RT–PCR protocols. These results are exemplified for plasmid p17-5310 using three different primer pairs (Fig. 3). The cloned cDNA fragments were also characterized by digestion with different restriction enzymes and by partial sequencing (data not shown) and showed the expected results as compared to cDNAs obtained by conventional RT–PCR.

Quasispecies distribution of the cDNA fragments generated by long RT–PCR

We next addressed the question of whether the quasispecies nature of HCV is represented by results of long RT–PCR or whether this method leads to selection of single variants, e.g. the main variant. Therefore, we sequenced the hypervariable region 1 (HVR1) of 12 independently obtained cDNA clones (p17-5310/Q1–Q12) (Fig. 4). The deduced amino acid sequences could be grouped according to their homology into five variants named HVR1.A, B, C, D and E. Comparison at the nucleotide level revealed the same groups. HVR1.A was found six times and represented 50% of the sequenced clones. The sequence HVR1.B, with only one amino acid exchange, as well as the sequences HVR1.D and HVR1.E with eight and nine amino acid exchanges, respectively, were found only once. HVR1.C had four amino acid exchanges and was found three times representing 25% of the sequenced clones. HVR1.A could be identified as the main variant present in serum of patient P1 as confirmed by direct sequencing of a short PCR product obtained by conventional RT–PCR (data not shown). Interestingly, we did not find the insertion of two amino acids immediately in front of the HVR1 as determined for isolate HCV-AD78 and for other patient sera (Höhne et al., 1994; Zibert et al., 1997).

Immunoprecipitation of in vitro translated HCV proteins derived from single HCV genomes

*Taq* polymerase can introduce mutations into the amplified cDNA because of lack of proofreading during PCR. In order to detect, e.g., nonsense mutations in the cloned cDNA fragments we used the plasmids p17-5310 and p4497-9401 for *in vitro* translations of HCV proteins (Fig. 5a). p17-5310 and p4497-9401 served as templates to amplify shorter PCR fragments encoding HCV proteins. The sense primers each contained a T7 promoter followed by an in-frame ATG codon and individual HCV-specific sequences. The generated PCR products were used for *in vitro* transcription by T7 RNA polymerase, and the transcripts were *in vitro* translated in a cell free rabbit reticulocyte lysate. Seven proteins (p1–p7) representing almost the complete HCV ORF were translated (Fig. 5a). Immunoprecipitations using an HCV-positive patient serum were performed. Proteins with molecular mass of 37, 54, 28, 70, 35, 65 and 68 kDa were identified with individual transcripts (Fig. 5b). These molecular masses correspond to the predicted sizes of encoded HCV proteins.
Discussion

Recently, it has become possible to obtain PCR products with a size of about 40 kb after establishment of a method using a mix of different thermostable DNA polymerases (Barnes, 1994; Cheng et al., 1994). However, these protocols involved DNA templates from human genomes or from bacteriophage λ and not a low titre RNA template like that of HCV. We optimized the reverse transcription and PCR protocol for HCV and were able to generate HCV-specific cDNA transcripts with a length of at least 6±8 kb and PCR products of about 5 kb from a moderately high titre HCV patient serum (1±0¬10' genome equivalents/ml). Using this technique, a complete open reading frame could be cloned from only two cDNA fragments; different properties of the ORF have been analysed.

Several improvements of the RT–PCR protocol were necessary in order to generate these long cDNA fragments of HCV. Reverse transcription was optimized by using an enzyme without any detectable RNase H activity and by increasing the incubation temperature to 48 °C. These modifications led to cDNA transcripts of more than 6 kb in length. The high incubation temperature possibly reduced secondary structures within the HCV RNA. Secondary structures are predicted for the 5′ nontranslated (Brown et al., 1992; Smith et al., 1995) and the 3′ nontranslated (Kolykhalov et al., 1996) regions of the HCV RNA, and it is likely that the complete genome is also highly folded.

Previous studies have shown that the activity of Taq polymerase is reduced by an active reverse transcriptase (Sellner et al., 1992). It was recommended therefore to inactivate the reverse transcriptase, e.g. by heating the RT sample for 5 min to 95 °C before starting the PCR reaction. These results were obtained with a reverse transcriptase that had RNase H activity. The longest cDNA synthesized by us was obtained with a reverse transcriptase without any detectable RNase H activity. For this enzyme current protocols recommend not heating the RT sample immediately after cDNA synthesis when generating PCR products with a length of more than 3 kb. However, our results indicate that such a denaturation step is important for generation of long PCR products even with the RNase H free transcriptase used here.

A commercially available enzyme mix of Taq and Pwo DNA polymerases was used here. This enzyme mixture combines the advantages of both polymerases, the high processivity of Taq and the high fidelity of Pwo with its 3′–5′
Table 3. Comparison of primer sequences with isolate HCV-AD78

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>S4477*</td>
<td>AGGCCCATCAAGGGGGAAG</td>
</tr>
<tr>
<td>nt 4477–4496†</td>
<td>---A---------G---------------------</td>
</tr>
<tr>
<td>S4497*</td>
<td>CACTCCTAGTCCGGCATCCCA</td>
</tr>
<tr>
<td>nt 4497–5019†</td>
<td>---C---------T---------------------</td>
</tr>
<tr>
<td>A5310*</td>
<td>CCCCCGTTGCTAGTTGAC</td>
</tr>
<tr>
<td>nt 5292–5310†</td>
<td>------------G---------------------</td>
</tr>
<tr>
<td>A5387*</td>
<td>GATCTTGCCCAACTGAC</td>
</tr>
<tr>
<td>nt 5367–5387†</td>
<td>-------------------------------------</td>
</tr>
</tbody>
</table>

* Sequence of HCV-J (Kato et al., 1990).
† Sequence of HCV-AD78 (K. Rispeter, unpublished).

exonuclease activity. To achieve long PCR products we carried
out a hot start and added 5% DMSO to the reaction mixture.
The latter allowed us to lower the denaturation temperature
from 95 to 92 °C. This step may minimize damage to the DNA
template, e.g. by depurination (Cheng et al., 1994).

Our results suggest that synthesis of long HCV-specific
cDNA fragments is dependent on the concentration of HCV
RNA molecules (Table 2). The generation of longer PCR
products correlated with high HCV RNA titres in patient sera
although the overall difference of titres was low, varying by
a maximum factor of three. PCR products of about 3 kb could be
amplified only with sera that had an HCV RNA titre of at least
7.5 × 10⁶ HCV genome equivalents/ml while sera with a
lower HCV RNA titre did not result in such long PCR
products. In order to represent the HCV genome in only two
cDNA fragments it was necessary to use a serum with
1.0 × 10⁷ genome equivalents/ml. For generation of long
HCV cDNA fragments it is therefore recommended to
determine the HCV RNA titre and possibly to enrich a virus
preparation, e.g. by ultracentrifugation. Such an improvement
or the use of sera with more than 10⁸ genome equivalents/ml
may facilitate generation of a cDNA encoding almost the
complete genome of HCV. Our observation of a transcript of
at least 6.8 kb (Fig. 1) is promising in this respect.

In addition to the titre of HCV RNA other parameters
might also be important for long RT–PCR of HCV. The
variability of HCV genomes (Okamoto et al., 1992a; Bukh et
al., 1993; Simmonds, 1995) and the quasispecies nature of
HCV (Martell et al., 1992) may influence the results of long
RT–PCR. The choice of primers for long RT–PCR is therefore
critical. We carried out our investigations using sera from
patients obtained 10 years after infection with isolate HCV-
AD78. The sequence of the complete HCV-AD78 genome was
only recently determined (K. Rispeter, unpublished).
Comparison of the nucleotide sequences of four primers used for
generation of long cDNA fragments revealed several mis-
matches to HCV-AD78 (Table 3). Primer A5310 had one,
primer S4477 had two and primer S4497 showed three
mismatches. It is therefore possible that these three primers did
not match in all cases the HCV genome found in patient P1 and
that a further optimization of the RT–PCR protocol could be
achieved by selection of primers which are optimal for the
respective isolate.

We isolated 12 clones of the amplified cDNA fragment
covering nt 17–5310 of the HCV genome. Sequencing of the
HVR1 of these independently cloned cDNAs revealed five
different variants. One variant, HVR1.A, was present in 50%
of the clones and corresponds to the main variant in isolate P1
as confirmed by direct sequencing of a PCR product amplified
by conventional RT–PCR. This implies that no selection of
specific variants occurred during long RT–PCR and that the
quasispecies nature of HCV can also be analysed by this

There is concern about the value of PCR products for
determination of mutation rates or construction of infectious
clones, since nucleotide misincorporations are generated by
DNA Taq polymerases (Eckert et al., 1992). Such mutations are
difficult to recognize because of the heterogeneity of HCV
genomes. Our immunoprecipitations showed that there were
at least no nonsense codons present in the cloned HCV ORF
(Fig. 5b). The enzyme mixture used here has an error rate
estimated to be lower by a factor of approximately 5 in
comparison to Taq DNA polymerase alone (Frey et al., 1995).
Therefore, cDNA fragments generated by such an enzyme
mixture might result in lower rates of nucleotide misincor-
porations into the HCV genome.

Current estimates of the mutation rate of HCV during
chronic disease range from 1·4 to 1·9 × 10⁻³ base substitutions
per site per year in analysis of complete or partial genome
sequences, respectively (Ogata et al., 1991; Okamoto et al.,
1992a). However, such estimates are based on the analysis
of relative short cDNAs which probably represent different
genomes present in the same patient serum. The method
described here could be used to identify mutations present on
the same RNA genome and would therefore allow a more
reliable estimation of the variability of the HCV genome. In
addition, mutations within B- and T-cell epitopes have been
shown to be responsible for establishment of chronic disease
Whether such escape mutations are present on the same
genome can now be investigated. Also, single genes of the
HCV ORF, like NS5B, with a length critical for conventional RT–PCR technique, can now be easily amplified from single RNA molecules by our novel RT–PCR protocol.

Generation of a complete cDNA of the HCV genome which gives rise to infectious virus after transfection of tissue culture cells would be a major breakthrough in the field of HCV research. For other RNA viruses complete infectious cDNA clones have been established, e.g. for poliovirus (Racaniello et al., 1981), rhinovirus (Mizutani et al., 1985), coxsackievirus B (Kandolf et al., 1985), hepatitis A virus (Cohen et al., 1987) and foot-and-mouth disease virus (Zibert et al., 1990). Recently, it has been shown that application of the long RT–PCR technique for hepatitis A virus simplifies the construction of an infectious cDNA (Tellier et al., 1996). This has yet to be achieved for HCV. Replication of HCV after transfection of in vitro transcribed RNA has been reported but so far seems to be rather inefficient (Yoo et al., 1995). The cDNA used by the latter study consisted of small fragments derived from more than one virus variant, and this artificial combination of different variants might hamper establishment of an infectious HCV cDNA. Construction of a complete HCV cDNA could now be done with cDNAs derived from only one or two RNA molecules obtained by long RT–PCR.

We are grateful to Ulla Schmützer for sequence analysis and to Sergei Viazov for his help in preparing the manuscript.

This work was supported by Bundesministerium für Bildung und Forschung (BMBF) (Grant KJ9350).

References


Received 18 March 1997; Accepted 27 June 1997.