Published in final edited form as:

J Virol Methods. 2010 August ; 167(2): 158–164. doi:10.1016/j.jviromet.2010.03.030.

Enhanced Protocol for Determining the 3' Terminus of Hepatitis C Virus

Xiaoan Zhang^{a,c}, Xiaofeng Fan^{a,b,*}, Yanjuan Xu^a, and Adrian M. Di Bisceglie^{a,b,*}

^a Division of Gastroenterology and Hepatology, Department of Internal Medicine, Saint Louis University School of Medicine, St. Louis, Missouri 63104, U.S.A

^b Saint Louis University Liver Center, Saint Louis University School of Medicine, St. Louis, Missouri 63104, U.S.A

^c Department of Histology and Embryology, College of Basic Medical Sciences, Zhengzhou University, Zhengzhou 450001, Henan, China

Abstract

The determination of viral 3' ends is a routine practice in molecular biology. However, this has been a challenging task for hepatitis C virus (HCV), an enveloped single-stranded, positive-sense RNA virus classified into the *Flaviviridae* family. The extreme end of HCV 3' untranslated region (3'UTR), the so-called 3' X tail, was not identified at the time of HCV discovery. Complete HCV 3'UTR sequences occupy a very small percentage of the exponentially growing HCV sequence databases. Although commercial kits and experimental protocols are available, these methods are both tedious and not reproducible. A stepwise optimization procedure was developed as a simple and robust protocol for determining the complete HCV 3'UTR from clinical samples. The availability of abundant authentic sequence information for the complete HCV 3'UTR will allow full investigation of its biological role in the life cycle of HCV.

Keywords

hepatitis C virus; 3'UTR; RT-PCR; ligation

1. Introduction

Hepatitis C virus (HCV), a major etiological agent responsible for most post-transfusional and community-acquired hepatitis (Alter et al., 1999), is an enveloped single-stranded, positive-sense RNA virus classified into the *Flaviviridae* family (Choo et al., 1989). The viral genome comprises an approximately 9600 bp RNA molecule which is divided into three regions: 5' untranslated region (UTR), a single large open reading frame and a 3'UTR. Like most RNA viruses, genetic variability is a hallmark of HCV and is a key factor in both basic HCV biology and clinical practice. Thus, one of major tasks in HCV research is to decipher viral sequences from infected individuals. In the past 20 years, a large number of

*Corresponding authors: Division of Gastroenterology and Hepatology, Department of Internal Medicine, Saint Louis University School of Medicine, 3635 Vista Avenue, St. Louis, MO 63110, Phone: +1 314-977-7833, Fax: +1 314-577-8125, fanx@slu.edu (X. Fan), dibiscam@slu.edu (A.M. Di Bisceglie).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

HCV sequences have been generated and deposited in GenBank as well as in several HCV sequence databases (Kuiken et al., 2005; Combet et al., 2007). However, it is interesting to note that sequence abundance is not distributed evenly along with the entire HCV genome. For instance, the Los Alamos HCV database has collected approximate 50,000 HCV sequences of which there are only a dozen complete HCV 3'UTR sequences (0.03%). Such a paucity of HCV 3'UTR sequences can be attributed to several factors. First, HCV is not cultured readily *in vitro* and therefore the only source of HCV sequences is clinical samples. It is difficult to obtain a large quantity of high-quality HCV RNA for conventional determination of viral ends. Second, HCV 3'UTR is composed of three distinct regions, an about 40 bp variable region immediately downstream of the stop codon, a poly(U/UC) tract with heterogeneous lengths, and the highly conserved 98 bp terminal sequence, the so-called 3' X tail (Figure 1). These regions are highly structured and may reduce the efficiency of the reverse transcription-polymerase chain reaction (RT-PCR) that is a necessary step in the determination of HCV 3'UTR (Tuplin et al., 2002, 2004; Simmonds et al., 2004). Through a stepwise optimization procedure, a simple and robust protocol has been developed for determining the complete HCV 3'UTR from clinical samples.

2. Materials and Methods

2.1. Samples

The optimization of experimental protocols was conducted directly with serum samples collected in 1995 from four patients infected with HCV genotype 1a, referred to as #069, #0273, #1099 and #1564, respectively. A large volume of serum stored at -80°C was available from these patients, which allowed repeated and detailed optimization of experimental protocols. After optimization, additional serum samples were used for the estimation of protocol sensitivity and robustness. HCV RNA levels were quantitated by bDNA assay (Bayer VERSANT HCV 3.0) immediately prior to the start of this study.

2.2. RNA extraction

Total RNA was extracted from serum by using QIAamp Viral RNA Mini Kit (Qiagen, Valencia, CA) according to the instructions provided. RNA was extracted from 280 μL of serum and finally eluted into 60 μL of Tris buffer containing 20 U/mL of RNasein Ribonuclease Inhibitor (Promega, Madison, WI). Vigorous vortexing was avoided to prevent shearing of RNA templates.

2.3. Amplification of the homopolymer repeat of HCV 3'UTR

HCV 3'UTR contains a poly(U/UC) tract of variable length, and such a homopolymer repeat is sometimes difficult to amplify. Thus the processing activity and fidelity were investigated first for several common DNA polymerases, including AmpliTaq (Applied Biosystems, Foster City, CA), AmpliTaq360 (Applied Biosystems), AmpliTaqGold360 (Applied Biosystems), rTth XL (Applied Biosystems) and DyNAzyme EXT DNA Polymerase (New England Biolabs, Ipswich, MA). The plasmid p90H/FL pol- consisting of the full-length HCV genome was used as the PCR template (Kolykhalov, et al., 1997). Two primers, p90CircleF2 and p90R9643, were designed to amplify a region covering a ~ 130 bp poly(T/TG) domain (Table 1). Briefly, 2 μL of 10-fold dilution of plasmid, starting at 2 ng/ μL , was mixed with 48 μL of PCR matrix containing 1 mM MgCl_2 [1 mM $\text{Mg}(\text{OAc})_2$ for rTth XL only], 1x polymerase-matched PCR buffer, 0.8 mM dNTPs (Invitrogen, Carlsbad, CA), 0.4 μM each of forward and reverse primers and 1 U each of polymerases. For both AmpliTaq360 and AmpliTaqGold360 polymerases, 1 μL of 360GC enhancer was also added. The PCR matrix with rTth XL also included 0.4 μM Trnc-21 as described previously (Fan et al., 2006). Cycle parameters were programmed on DNA Thermal Cycler 480 as 94°C for 1 min (9 min for AmpliTaqGold360) followed by the first 5 cycles of 94°C for 30

sec, 60°C for 30 sec and 72°C for 1 min and the final 20 cycles in which the annealing temperature was reduced to 55°C. The reaction was ended with 7 min incubation at 72°C. The expected bands were gel-purified using QIAEX II Gel Extraction Kit (Qiagen) and subjected to direct sequencing with both forward and reverse primers. The reaction was prepared with ABI PRISM dye terminator cycle sequencing ready reaction kit and run on ABI 373A automated sequencer (Applied Biosystems).

2.4.1. RT-PCR amplification of HCV poly(U/UC) tract and 3' X tail—After estimation of the processing activity and fidelity of DNA polymerases in the amplification of homopolymer repeats, selected polymerases were used to amplify HCV 3'UTR directly from serum samples. In a RT-“nested” PCR strategy, two forward primers, CircleF11 and CircleF269.5, were located in the 3' end of HCV NS5B gene (Table 1). Three reverse primers, 3HBVR1UTR1, 3HBVR1UTR2 and 3HBVR1UTR3, were designed to be located at different positions of HCV 3' X tail (Table 1). Each of them contained an identical sequence from hepatitis B virus genome, which was used as the reverse primer in the second round of PCR amplification (Table 1).

In brief, 5µL of extracted RNA was mixed with 15µL of RT matrix consisting of 2.5 mM MgCl₂, 1x AmpliTaq PCR buffer, 5 mM DTT, 1 µM of reverse primer, 1.5 mM dNTPs (Invitrogen), 16 U of Rnasein Ribonuclease Inhibitor (Promega) and 200 U of SuperScript III (Invitrogen). The reaction was incubated at 50°C for 45 min and inactivated by heating at 70°C for 15 min. For the first round of PCR, the entire RT reaction was mixed with 30 µL of PCR matrix containing 1x AmpliTaq PCR buffer, 0.2 µL of 100 mM dNTPs, 0.4 µL of 50 µM forward primer CircleF11 and 1.25 U AmpliTaq polymerase. Cycle parameters were programmed as 94°C for 1 min connected by the first 10 cycles of 94°C for 45 sec, 60° for 45 sec and 72°C for 1 min, followed by final 20 cycles in which the annealing temperature was reduced to 55°C. The reaction was ended with a 7 min incubation at 72°C. A 2 µL aliquot of the first round of PCR product was used for the second round of amplification by added it into 48 µL of PCR matrix consisting of 1.2 mM Mg(OAc)₂, 1x rTth XL PCR buffer, 0.8 mM dNTPS, 0.4 µM Trnc-21, 0.4 µM each of primers (CircleF269.5 and 3HBVR1) and 2 U of rTth XL polymerase. Cycle parameters were programmed as 94°C for 1 min followed by the first 10 cycles of 94°C for 30 sec and 70°C for 1 min and final 20 cycles in which the annealing/elongation temperature was reduced to 66°C. The reaction was ended with a 5 min incubation at 72°C. A 5µL aliquot of the PCR product was run on a 1% agarose gel to examine the size of the product. When necessary, the PCR product was gel-purified and subjected to direct sequencing as described above.

2.4.2. The optimization of RT-PCR amplification of HCV 3'UTR—Using the RT-PCR protocol described above, one of the reverse primers, 3HBVR1UTR1, located at the uttermost part of HCV 3' X tail, resulted in a product with a much smaller size than expected (see Results for detail). To solve this issue, extensive optimization at multiple levels was performed. First, in the RT step, the incubation temperature was increased to 55°C. Second, the RNA template was mixed first with sodium hydroxide at concentrations ranging from 0.3 to 100 mM in the 20 µL RT reaction. The mixture was incubated at either room temperature or 90°C for 3 min and then quickly chilled on ice. Third, various additives were tested in the RT reaction, including DMSO (Sigma) (10~20%), 360 GC enhancer (Applied Biosystems) (10~20%), betaine (Sigma) (1~2 M), Fc400 (Sigma) (4~8 mg/mL) and extreme thermostable single-stranded DNA binding protein (ET SSB) (New England Biolabs) (15 ng/µL). Fourth, the first round of PCR was also tested with the additives used in the RT reaction at similar concentrations. Fifth, the first round of PCR amplification was tried with different DNA polymerases, including rTth XL, DyNAzyme EXT, Vent (New England Biolabs), DeepVent (New England Biolabs) and Phusion High-Fidelity (New England Biolabs). Based on their characteristics and recommended usages, other components and

cycle parameters were adjusted accordingly. Finally, in the first round of PCR, the forward primer CircleF11 was replaced with CircleF269.5, which had a higher melting temperature and allowed a 5°C increase of the annealing temperature. The forward primer in the second round of PCR was changed to CircleF368.8 (Table 1). It should be noted the above modifications were tested either independently or in combination, resulting in the evaluation of more than 100 experimental protocols.

2.5. The addition of an adaptor to HCV 3' X tail

For determining the complete HCV 3' UTR, the 3' X tail of HCV must become an internal domain during RT-PCR amplification. The current study explored three different approaches: self-ligation of HCV 5' and 3' ends, polyadenylation of HCV 3' extreme end and adaptor ligation. In the self-ligation approach, 10.5 µL of extracted RNA was mixed with 40 U Rnasein Ribonuclease Inhibitor (Promega), incubated at 65°C for 10 min and chilled on ice. A 4.5 µL aliquot of ligation matrix was then added into the mixture to make a final 15 µL reaction consisting of 1x T4 RNA ligase 1 buffer and 20 U T4 RNA ligase 1 (New England Biolabs). The reaction was incubated at 16°C overnight or at room temperature for 4 hr, followed by inactivation at 65°C for 15 min. Two additional adjustments were also tried by either the addition of 10% DMSO in the RNA template incubation or the inclusion of 10 U T4 polynucleotide kinase (New England Biolabs). A 5 µL aliquot of ligation reaction was used for RT-PCR amplification with two sets of primers, raceF1/raceR1 for the first round and raceF2/raceR2 for the second round (Table 1). The protocol was basically the same as the one described above in which AmpliTaq and rTth XL polymerases were used for the first and second round amplifications, respectively.

In the second approach, the poly (A) tailing was conducted by using poly(A) polymerase (New England Biolabs). In brief, 14 µL aliquot of extracted RNA was mixed with 2 µL of 10x polymerase buffer, 2 µL of 10 mM ATP (Applied Biosystems), 40 U Rnasein Ribonuclease Inhibitor (Promega) and 5 U poly(A) polymerase. The reaction was incubated at 37°C for 45 min and then inactivated at 70°C for 15 min. A 10 µL aliquot of reaction was utilized for RT-PCR amplification as described above except for the use of dT35HBVR1.23 as the reverse primer in the RT and first PCR round, and HBVR1T3 in the second round of PCR (Table 1). The poly(U) tailing was also estimated using poly(U) polymerase (New England Biolabs), followed by RT-PCR amplification in which primers dT35HBVR1.23 and HBVR1T3 were replaced by dA35HBVR1.23 and HBVR1A3, respectively (Table 1).

The final approach used conventional adaptor ligation into the HCV 3' X tail. In this approach, four adaptors were estimated. The first adaptor was a 77-bp oligonucleotide designed from HBV genome, named HBV77. The HBV77 was phosphorylated at the 5' end during commercial synthesis (Invitrogen). Its 3' end was then blocked by the addition of ddATP (Sigma). The blocking reaction consisted of 1x TdT buffer, 5 mM CoCl₂, 50 µM ddATP, 2.5 µg HBV77 and 400 U terminal transferase (Roche Applied Science, Indianapolis, IN). The reaction was incubated at 37°C for 60 min, followed by the inactivation by adding 2 µL of 0.2 M EDTA. The blocked HBV77 (HBV77b) was purified from a 4% low-melting agarose gel using QIAEX II Gel Extraction Kit (Qiagen) and the concentration was determined with a NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE). The second adaptor was adopted from a universal miRNA cloning linker (New England Biolabs), named linker 1 in this study (Table 1). The final two adaptors, respectively named linker 2 and linker 3, were hybrid oligonucleotides designed from the HBV genome. These two linkers were characterized by a phosphorylated 5' end, initial four RNA bases and a blocked 3' end preventing from self-ligation (Table 1). All linkers were synthesized from Integrated DNA Technologies (IDT, Coralville, IA). The ligation reaction included 14 µL extracted RNA, 1x T4 RNA ligase 1 buffer, 40 U Rnasein Ribonuclease Inhibitor, 100 pmol of adaptor (either HBV77b, linker 1, linker 2 or linker 3)

and 20 U T4 RNA ligase 1 (New England Biolabs). The reaction was incubated at room temperature for 4 hr and then inactivated at 70°C for 20 min. For the ligation of HBV77b, possible enhancement of the ligation efficiency was tested by the addition of PEG8000 (Sigma) (20% final concentration) and/or hexamine cobalt chloride (Sigma) (1 mM final concentration). A 5 µL aliquot of the ligation product was directly used for RT-PCR amplification as described above. The reverse primers for HBV77b, linker 1, linker 2 and linker 3 were 77R1, HBVR1linker1, HBVR1linker2 and HBVR1linker3, respectively (Table 1). These reverse primers were also used for the first round PCR amplification in combination with forward primer CircleF11. The second round PCR was done with forward primer CircleF269.5 and reverse primer 77R2 for HBV77b (Table 1). The reverse primer for linkers 1, 2 and 3 was 3HBVR1 (Table 1).

2.6. Molecular cloning of RT-PCR product

The cloning procedure was essentially the same as that described previously (Chambers et al., 2005). Briefly, after the determination of optimized protocols, the PCR product was gel-purified using QIAEX II Gel Extraction Kit (Qiagen) and ligated into pUC19 vector (Invitrogen). *E. coli* DH5α bacteria (Invitrogen) were used for transformation and recovery of recombinant clones. Approximately five clones for each PCR product were sequenced with universal primer M13R. The consensus sequence for each PCR product was determined by using the program Clustal W (version 1.74) (Higgins et al., 1988) and Molecular Evolutionary Genetics Analysis software package (MEGA, version 3.0) (Kumar et al., 2004).

3. Results

3.1. DNA polymerases capable to amplify the homopolymer repeat

A total of 5 DNA polymerases were estimated for their capability to amplify the poly (T/TG) domain from the plasmid p90H/FL pol-. As shown in Figure 2, rTth XL, AmpliTaq and DyNAzyme EXT gave DNA bands with expected sizes. Directed amplicon sequencing showed complete identity to the template without any insertions or deletions in the poly(T/TG) domain. In contrast, AmpliTaq360 and AmpliTaqGold360 were failed for the amplification, suggesting the limited processing activity on these homopolymer repeats. The polymerases rTth XL, AmpliTaq and DyNAzyme were then used for subsequent optimization of RT-PCR protocols.

3.2. RT-PCR amplification of HCV poly(U/UC) tract and 3' X tail

Amplification of the poly(U/UC) tract and the 3' X tail utilized a RT-“nested” PCR strategy in which reverse primers for the RT and the first round PCR contained a distinct sequence from the HBV genome that was used for the second round PCR, a so-called step-out PCR strategy (Matz et al., 1999). Three reverse primers were designed in the 3' X tail, including 3HBVR1UTR1, 3HBVR1UTR2 and 3HBVR1UTR3. The RT-PCR was successful with 3HBVR1UTR2 and 3HBVR1UTR3 as indicated by DNA bands with expected sizes (Figure 3). However, the RT-PCR with primer 3HBVR1UTR1 resulted in the generation of DNA bands much smaller than expected (Figure 3). These results were very reproducible when using patient samples in addition to samples #069, #0273, #1099 and #1564. Direct sequencing of amplicons obtained with 3HBVR1UTR1 showed that the entire poly(U/UC) tract and 3' X tail were missing. It was speculated that such shortened products resulted from highly structured characteristic of the target domain. Thus the experiment was repeated with extensive optimization, including the partial denaturation of the RNA template, the use of different DNA polymerases, changes of cycle parameters, and the inclusion of various additives in either RT and/or the first round PCR. However, none of these approaches was successful. Finally, a series of overlapped primers was designed to locate a possible domain

displaying strong RT-PCR impairment. Primers with 5' end over position 9614 (numbering according to HCV H77 strain, AF009606) consistently resulted in either negative amplification or a shortened product (Figure 4).

3.3. Efficient ligation of short hybrid oligonucleotides to HCV 3' X tail

The determination of the complete HCV 3'UTR tried 3 different approaches, including self-ligation of HCV 5' and 3' ends, poly (A) tailing and adaptor ligation. The first two approaches consistently gave negative amplification. The poly (U) tailing repeatedly obtained a smaller RT-PCR product, which resulted from the priming of the reverse primer in the poly (U/UC) tract as confirmed by direct sequencing. Thus, the third approach, conventional adaptor ligation to the 3' X tail became the focus. Among four adaptors, successful amplification was achieved with linker 1, linker 2 and linker 3. Linker 2 gave the best results (Figure 5). However, as encountered in the amplification of the poly (U/UC) tract and the 3' X tail, all ligation-based RT-PCR generated a shortened product. Similar optimization used in the RT-PCR did not provide any improvement.

3.4. The optimized protocol for the determination of the complete HCV 3' UTR

Due to impaired RT-PCR in the 3' X tail, the complete HCV 3' UTR was divided into two overlapped regions in the optimized protocol. The first domain from NS5B to the middle part of the 3' X tail was amplified by regular RT-PCR. In brief, 5 μ L of extracted RNA was mixed with 15 μ L of RT matrix consisting of 2.5 mM MgCl₂, 1x DyNAzyme EXT buffer (detergent-free), 5 mM DTT, 1 μ M of reverse primer 3HBVR13x20, 2 mM dNTPs, 16 U of Rnasein Ribonuclease Inhibitor (Promega) and 200 U of SuperScript III (Invitrogen). The reaction was incubated at 50°C for 60 min and inactivated by heating at 70°C for 15 min. The entire RT reaction was then mixed with 30 μ L of PCR matrix containing 1x DyNAzyme EXT buffer (detergent-free), 0.4 μ L of 50 μ M forward primer CircleF11 and 1 U DyNAzyme EXT. Cycle parameters were programmed as 94°C for 1 min connected by the first 10 cycles of 94°C for 45 sec, 65° for 45 sec and 72°C for 45 sec, followed by final 20 cycles in which the annealing temperature was reduced to 60°C. The reaction was ended with a 7 min incubation at 72°C. A 2 μ L aliquot of the first round of PCR product was used for the second round amplification by added into 48 μ L of PCR matrix consisting of 1.2 mM Mg(OAc)₂, 1x rTth XL PCR buffer, 0.8 mM dNTPS, 0.4 μ M Trnc-21, 0.4 μ M each of primers (CircleF269.5EcoRI and 3HBVR1HindIII) and 2 U of rTth XL polymerase. Cycle parameters were programmed as 94°C for 1 min followed by the first 10 cycles of 94°C for 30 sec and 70°C for 1 min and final 20 cycles in which the annealing/elongation temperature was reduced to 66°C. The reaction was ended with a 5 min incubation at 72°C.

The second domain containing the 3' X tail was determined by ligation-based RT-PCR. Briefly, 20 μ L of ligation reaction contained 14 μ L of extracted RNA, 1x T4 RNA 1 ligase buffer, 40 U Rnasein Ribonuclease Inhibitor, 5 μ M linker 2 and 40 U T4 RNA ligase 1 (New England Biolabs). The reaction was incubated at 14°C overnight or at room temperature for 4 hr, and inactivation was not necessary. A 5 μ L aliquot of the ligation product was then subjected to RT-PCR as described above except for the replacement of primers, 3UTRF1/HBVR1linker2 for RT/first PCR round and 3UTRF2EcoRI/3HBVR1hindIII for the second round of PCR (Table 1).

3.5. Sensitivity and reproducibility of the optimized protocol

The sensitivity and robustness of the optimized protocol was estimated in 15 serum samples with HCV RNA levels ranging from 1×10^5 to 3×10^7 copies/mL as quantitated by HCV bDNA assay (Bayer VERSANT HCV 3.0). Total RNA was extracted from 140 μ L of serum. Two overlapped fragments covering complete HCV 3'UTR were successfully amplified from all 15 samples (Figure 6). All fragments were gel-purified and cloned into pUC19

vector. About 5 independent clones per fragment were picked and sequenced. Consensus sequences were then generated by alignment and assembly. No additional sequences were observed in the 3' extreme end (Figure 7).

4. Discussion

The determination of viral 3' ends is a routine practice in molecular biology. However, this has been a challenging task for HCV, and the 3' X tail was not identified in the initial HCV clone (Choo et al., 1989). Although commercial kits and experimental protocols are available, these methods are both tedious and not reproducible (Tanaka et al., 1996; Kolykhalov et al., 1996; Wakita et al., 2009). As a result, complete HCV 3'UTR sequences occupy a very small percentage in the exponentially growing HCV sequence databases (Kuiken et al., 2005; Combet et al., 2007). In the present study, a protocol with adequate sensitivity and robustness has been developed for determining the complete HCV 3'UTR from clinical serum samples.

The optimization procedures consist of three major steps: the selection of DNA polymerases capable of going through a homopolymer repeat, the amplification of HCV poly (U/UC) and 3' X tail and the determination of complete HCV 3'UTR through ligation-based RT-PCR. The final success of protocol optimization depends on 3 major observations. First, priming sites over position 9614 in RT-PCR result in a shortened product. As the HCV 3'UTR is well known for its highly structured nature, this finding is not surprising. A shortened RT-PCR product is generated frequently when the templates hold strong secondary structure, such as hepatitis E virus and some cellular genes (Wei et al., 2003; Zhang et al., 2001). There are also several advanced studies on its intrinsic mechanism, including the concept of replication slippage (Viguera et al., 2001). However, such replication slippage in the HCV 3'UTR cannot be eliminated by existing methods, including the use of polymerases showing strand-displacement activity (Viguera et al., 2001), the alleviation of template structure by reported additives (Spiess et al., 2002; Henke et al., 1997; Hengen et al., 1997; Chou et al., 1992) and the emulation of a crowded reaction environment (Lareu et al., 2007). Only very limited success was achieved in a few freshly-collected HCV serum samples in which single ligation-based RT-PCR has given the expected amplification of the entire HCV 3'UTR (data not shown). Therefore, it is concluded that no current protocol reliably amplifies the complete HCV 3'UTR in a single step. Similar frustration has also been encountered by other groups (van Leeuwen et al., 2006; Fan and van Leeuwen, personal communication). To circumvent this limitation, the HCV 3'UTR has been divided into two overlapped fragments in the final optimized protocol.

The second observation is the highly efficient ligation of short hybrid oligonucleotides to the 3' X tail of HCV. Less productive ligation was observed with HBV77b, a long single-strand DNA fragment. The use of PEG8000 and/or hexamine cobalt chloride did not result in any significant improvement. Furthermore, inefficient ligation is one of the major factors associated with the appearance of high background or smears in subsequent RT-PCR (data not shown). Homopolymer tailing did not provide any meaningful results in this study.

Another important finding was that RT and the first round PCR could be combined into a single tube. The entire RT reaction was used directly for the first round PCR amplification. Any additional steps, such as the purification of the RT reaction, were detrimental to PCR amplification (data not shown). In the final protocol, DyNAzyme EXT was chosen for the first round PCR amplification due to its superior compatibility in different buffer conditions (Fan and New England Biolabs, personal communication). The reverse transcriptase SuperScript III functioned well in a tailored RT reaction using DNA polymerase buffer.

Such a combination not only increases the protocol sensitivity but also minimizes experimental processing.

The reproducibility and sensitivity of the optimized protocol were demonstrated in 15 serum samples with various HCV RNA levels (all HCV genotype 1a). Considerable polymorphism was observed in the poly(U/UC) tract in both composition and length, ranging from 32 to 97 bp (average 62 bp). Its potential effect on HCV replication warrants further investigation.

In conclusion, the enhanced protocol represents the most convenient and efficient approach to determine the complete HCV 3'UTR from clinical serum samples. Although the protocol has been developed with HCV genotype 1a, it should be applicable for other HCV genotypes or subtypes due to similar genome structures of HCV 3'UTR. The availability of abundant authentic sequence information about the complete HCV 3'UTR should facilitate the understanding with respect to its evolutionary origin as well as its essential role in the modulation of HCV replication and assembly. Finally, this protocol can be adapted easily for the determination of the ends of other difficult RNA templates.

Acknowledgments

This work was supported by NIH grants R01 DK80711 (to XF) and R21 AI076834 (to AMD). We thank Hans C. van Leeuwen (Leiden University Medical Center, The Netherlands) for clarifying issues regarding his work.

References

- Alter MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, Gao F, Moyer LA, Kaslow RA, Margolis HS. Prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med* 1999;341:556–562. [PubMed: 10451460]
- Chambers TJ, Fan X, Droll DA, Hembrador E, Slater T, Nickells MW, Dustin LB, Di Bisceglie AM. E1/E2 quasispecies heterogeneity as a pretreatment variable during pegylated interferon therapy of chronic HCV infection. *J Virol* 2005;79:3071–3083. [PubMed: 15709027]
- Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 1989;244:359–362. [PubMed: 2523562]
- Chou Q. Minimizing deletion mutagenesis artifact during Taq DNA polymerase PCR by E. coli SSB. *Nucleic Acids Res* 1992;20:4371. [PubMed: 1508731]
- Combet C, Garnier N, Charavay C, Grando D, Crisan D, Lopez J, Dehne-Garcia A, Geourjon C, Bettler E, Hulo C, Mercier PL, Bartenschlager R, Diepolder H, Moradpour D, Pawlotsky JM, Rice CM, Trepo C, Penin F, Deléage G. euHCVdb: the European hepatitis C virus database. *Nucleic Acids Res* 2007;35:D363–D366. [PubMed: 17142229]
- Fan X, Xu Y, Di Bisceglie AM. Efficient amplification and cloning of near full-length hepatitis C virus genome from clinical samples. *Biochem Biophys Res Commun* 2006;346:1163–1172. [PubMed: 16793008]
- Hengen PN. Optimizing multiplex and LA-PCR with betaine. *Trends Biochem Sci* 1997;22:225–226. [PubMed: 9204710]
- Henke W, Herdel K, Jung K, Schnorr D, Loening SA. Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res* 1997;25:3957–3958. [PubMed: 9380524]
- Higgins DG, Sharp PM. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 1988;73:237–244. [PubMed: 3243435]
- Kolykhalov AA, Agapov EV, Blight KJ, Mihalik K, Feinstone SM, Rice CM. Transmission of hepatitis C by intrahepatic inoculation with transcribed RNA. *Science* 1997;277:570–574. [PubMed: 9228008]
- Kolykhalov AA, Feinstone SM, Rice CM. Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA. *J Virol* 1996;70:3363–3371. [PubMed: 8648666]

- Kuiken C, Yusim K, Boykin L, Richardson R. The Los Alamos HCV Sequence Database. *Bioinformatics* 2005;21:379–384. [PubMed: 15377502]
- Kumar S, Tamura K, Nei M. MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment, Brief. *Bioinform* 2004;5:150–163.
- Lareu RR, Harve KS, Raghunath M. Emulating a crowded intracellular environment in vitro dramatically improves RT-PCR performance. *Biochem Biophys Res Commun* 2007;363:171–177. [PubMed: 17854768]
- Matz M, Shagin D, Bogdanova E, Britanova O, Lukyanov S, Diatchenko L, Chenchik A. Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res* 1999;27:1558–1560. [PubMed: 10037822]
- Simmonds P, Tuplin A, Evans DJ. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. *RNA* 2004;10:1337–1351. [PubMed: 15273323]
- Spiess AN, Ivell R. A highly efficient method for long-chain cDNA synthesis using trehalose and betaine. *Anal Biochem* 2002;301:168–174. [PubMed: 11814287]
- Tanaka T, Kato N, Cho MJ, Sugiyama K, Shimotohno K. Structure of the 3' terminus of the hepatitis C virus genome. *J Virol* 1996;70:3307–3312. [PubMed: 8627816]
- Tuplin A, Evans DJ, Simmonds P. Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. *J Gen Virol* 2004;85:3037–3047. [PubMed: 15448367]
- Tuplin A, Wood J, Evans DJ, Patel AH, Simmonds P. Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. *RNA* 2002;8:824–841. [PubMed: 12088154]
- van Leeuwen HC, Liefhebber JM, Spaan WJ. Repair and polyadenylation of a naturally occurring hepatitis C virus 3' nontranslated region-shorter variant in selectable replicon cell lines. *J Virol* 2006;80:4336–4343. [PubMed: 16611892]
- Viguera E, Canceill D, Ehrlich SD. In vitro replication slippage by DNA polymerases from thermophilic organisms. *J Mol Biol* 2001;312:323–333. [PubMed: 11554789]
- Wakita T. Isolation of JFH-1 strain and development of an HCV infection system. *Method Mol Biol* 2009;510:305–327.
- Wei S, To SS. Influence of RNA secondary structure on HEV gene amplification using reverse-transcription and nested polymerase chain reaction. *J Clin Virol* 2003;27:152–161. [PubMed: 12829037]
- Zhang YJ, Pan HY, Gao SJ. Reverse transcription slippage over the mRNA secondary structure of the LIP1 gene. *Biotechniques* 2001;31:1286–1290. [PubMed: 11768657]

Figure 1. Schematic representation of the HCV 3'UTR. Following NS5B are a variable domain (VD) of about 40 bp, a poly(U/UC) tract and a 98-bp 3' X tail. In the optimized protocol, the complete HCV 3'UTR has been determined through two overlapped domains, fragment 1 and fragment 2.

Figure 2.

Estimation of processing activity and fidelity of 5 DNA polymerases in the amplification of ~130 bp poly(T/TG) domain. The plasmid p90H/FL pol- consisting of the full-length HCV genome was used as the PCR template. A series of 1:10 dilution of the plasmid template, starting at 4 ng, was applied and matched to lanes 1, 2, 3 and 4, respectively. Lane 5, negative control. M, 123 DNA ladder (Invitrogen).

Figure 3.

Amplification of the HCV poly(U/UC) tract and 3' X tail. Three reverse primers, located at different sites of 3' X tail, were used for RT and the first round PCR. While the RT-PCR with either 3HBVR1UTR2 or 3HBVR1UTR3 generated DNA bands with expected sizes, ~300 and 324 bp, respectively, the primer 3HBVR1UTR1, located on the uttermost part of 3' X tail, gave a much smaller product (~123 bp), indicating the existence of potential replication slippage. Lanes 1, 2, 3, 4 represent samples #069, #0273, #1099 and #1564, respectively; Lane 5, negative control. M, 123 bp DNA ladder (Invitrogen).

Figure 4.

Mapping of the priming sites showing replication slippage in the amplification of the HCV poly(U/UC) tract and 3' X tail. Lanes 1 and 2 were the amplicons by using primers 3HBVR1UTR1 and 3HBVR1UTR2, respectively. Lanes 3 through 15 indicated 13 overlapped 20-bp primers with each 3-bp extension toward HCV 3' terminus. The 5' end of the primer in lane 3 was located in position 9582 according to the numbering of HCV strain H77 (GenBank accession number: AF009606). Lanes 14 and 15 represented a negative amplification and a shortened product, respectively. Lane 16, negative control. M, 123 bp DNA ladder (Invitrogen).

Figure 5.

Amplification of the HCV 3'UTR by ligation-mediated RT-PCR. Three linkers were estimated for their efficiency in the ligation and subsequent RT-PCR. Lanes 1, 2, 3, 4 represent samples #069, #0273, #1099 and #1564, respectively; Lane 5, negative control. M1, 100 bp DNA ladder (Invitrogen). M2, 123 bp DNA ladder (Invitrogen).

Figure 6. Successful amplification of the complete HCV 3'UTR using optimized protocol. The HCV 3'UTR was divided into two overlapped domains, fragment 1 and fragment 2. Lanes 1 through 15 represented 15 serum samples with various HCV RNA titers. Lane 16, negative control. M, 123 bp DNA ladder (Invitrogen).

Figure 7.

Alignment of the complete HCV 3'UTR consensus sequences derived from 15 patient samples. Each consensus sequence was determined through the cloning and amplification of two overlapped fragments using the optimized protocol. Partial NS5B and linker sequences were removed. HCV strain H77 (AF009606) was used as the reference on the top line. Dot indicates identity and dash shows deletions.

Table 1

List of primers and adaptors used in this study. Primer numbering is according to the HCV H77 strain (GenBank accession no. AF009606). All primers and adaptors were designed using the software program Eugene (version 1.01). Star indicates the inclusion of HBV sequences in either primers or adaptors. HBV, hepatitis B virus.

[illegible]

Primer	Polarity	Sequence (5'-3')	Position	Applications
HBVRIlinker3	Anti-sense	tctgtagagaaaggcacagacggggacgggtacgggcagacggagaa	HBV*	
3HBVRI3x20	Anti-sense	tctgtagagaaaggcacagacggggtcagcactctctgcagtcatgc	9601–9622*	
3UTRF1	Sense	ctccatcttagccctatgtcacgg	9554–9576	
3UTRF2EcoRI	Sense	gtactgaattctctggaagggtccgtgagcc	9581–9600	Optimized protocol
CircleF269.5EcoRI	Sense	gtactgaattcgcctggacttgcctcggctcggttc	9233–9254	
3HBVRIHindIII	Anti-sense	actcgaagctttgtagagaaaggcacagacgggg	HBV*	