Transcriptional slippage prompts recoding in alternate reading frames in the hepatitis C virus (HCV) core sequence from strain HCV-1

Maxime Ratinier,1 Steeve Boulant,2 Christophe Combet,1 Paul Targett-Adams,2 John McLauchlan2 and Jean-Pierre Lavergne1

1IBCP (Institut de Biologie et Chimie des Protéines), CNRS, UMR 5086, Université de Lyon, IFR 128, 7 passage du Vercors, F-69367 Lyon, France
2MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, UK

Since the first report of frameshifting in HCV-1, its sequence has been the paradigm for examining the mechanism that directs alternative translation of the hepatitis C virus (HCV) genome. The region encoding the core protein from this strain contains a cluster of 10 adenines at codons 8–11, which is thought to direct programmed ribosomal frameshifting (PRF), but formal evidence for this process has not been established unequivocally. To identify the mechanisms of frameshifting, this study used a bicistronic dual luciferase reporter system in a coupled transcription/translation in vitro assay. This approach revealed +1 as well as –1 frameshifting, whereas point mutations, selectively introduced between codons 8 and 11, demonstrated that PRF did not readily account for frameshifting in strain HCV-1. Sequence analysis of cDNAs derived from RNA transcribed by T7 RNA polymerase in the dual luciferase reporter system, as well as in both a subgenomic replicon and an infectious clone derived from strain JFH1, identified additions and deletions of adenines between codons 8 and 11 due to transcriptional slippage (TS). Moreover, RNA isolated from cells infected with virus generated by JFH1 containing the A-rich tract also contained heterogeneity in the adenine sequence, strongly suggesting TS by the NS5B viral polymerase. These findings have important implications for insight into frameshifting events in HCV-1 and demonstrate for the first time the involvement of transcriptional slippage in this recoding event.

INTRODUCTION

Hepatitis C virus (HCV) infection is a major cause of chronic hepatitis, liver cirrhosis and hepatocellular carcinoma worldwide (Giannini & Brechot, 2003). HCV belongs to the family Flaviviridae and is a positive-sense, single-stranded RNA virus. Its genome encodes a polyprotein precursor of about 3000 aa, which is further processed by cellular and viral proteases to yield mature structural and non-structural proteins (for reviews, see Reed & Rice, 2000; Penin, 2003; Moradpour et al., 2007). Evidence from computer predictions has identified a cluster of unusually conserved synonymous codons in the core protein coding region, suggesting that an overlapping open reading frame (ORF) could exist in this part of the polyprotein. In support of protein translation in an alternative reading frame, specific IgGs for three of four peptides, whose sequences were derived from the alternate ORF, have been detected in sera from patients with chronic HCV infection (Walewski et al., 2001).

In 2001, Xu et al. (2001) obtained the first experimental evidence for translation of an alternative core ORF product in the HCV-1 strain. Their results suggested that −2/+1 programmed ribosomal frameshifting (PRF), promoted by a cluster of 10 consecutive adenines, occurred at or near codon 9. Direct sequence analysis of frameshift products gave ambiguous results, and therefore the precise mechanism of frameshifting and the exact amino acid sequence at the shift junction remain unknown (Xu et al., 2001). Additional studies have revealed that this A-rich sequence is also able to induce –1 frameshifting (Choi et al., 2003). More recently, it has been shown that internal initiation near codons 85/87 (Vassilaki & Mavromara, 2003) or at codon 26 (Baril & Brakier-Gingras, 2005) in the core coding sequence may generate products in the +1 reading frame. Previously, we demonstrated another pathway for expression of alternative reading frame products from a genotype 1b sequence (GenBank accession no. D89872) that does not involve the A-rich motif (Boulant et al., 2003). In this case, a +1 PRF at codon 42 could be followed by rephasing through a –1 frameshift at the stop codon at position 144 in the +1 reading frame to give an alternative form of core, named DFC HCV. These proteins, collectively named ARFPs for alternative reading frame proteins, have never been directly observed in native tissue and the main evidence for their expression has been

Received 22 November 2007
Accepted 26 February 2008
specific immune humoral and cellular responses mounted against them in patients with chronic infection (Bain et al., 2004; Komurian-Pradel et al., 2004; Troesch et al., 2005; Fiorucci et al., 2007).

In this report, we examined the sequence requirements within the A-rich tract between codons 8 and 11 of strain HCV-1 that are necessary for efficient frameshifting. To investigate whether frameshifting could arise from transcriptional slippage (TS) rather than ribosomal frameshifting, the nucleotide sequences of clones derived from RNA transcribed by T7 RNA polymerase were determined. We also modified infectious strain JFH1 (Lindenbach et al., 2005; Wakita et al., 2005; Zhong et al., 2005) to incorporate an A-rich tract at codons 8–11 to determine whether transcription by the viral NS5B RNA polymerase could generate genomes with insertion or deletion of adenines leading to decoding in alternative reading frames.

**METHODS**

**Plasmid DNA construction and site-directed mutagenesis.** Plasmids were constructed by inserting a PCR-generated fragment containing a PstI site at both ends into the unique PstI site of pIRF, a dual luciferase reporter plasmid (Laporte et al., 2000). Briefly, pIRF is a bicistronic reporter vector under the control of both T7 and cytomegalovirus (CMV) promoters, composed of the firefly luciferase (FLuc) gene followed by the HCV 5′ untranslated region (UTR) sequence up to the codon 10 of the core and a PstI site at the end of the ORF. The pIRF vector contains the first 10 codons for the core protein (Fig. 1a). The upstream (control) FLuc reporter is translated in a cap-dependent manner, whereas the downstream (assay) RLuc reporter is under the control of the HCV internal ribosome entry site (IRES). The ATG codon of RLuc was converted to GGC encoding glycine by site-directed mutagenesis using a Quickchange kit (Stratagene) (Fig. 1a). Additional mutagenesis converted the PstI site (CTGCAG) to ACCAAA, which is identical to codon sequences 11 and 12 in HCV strain H77. For plasmids to detect +1 frameshifting, a nucleotide was added after codon 12 (AAAc; Fig. 1a). For plasmids to detect −1 frameshifting, 2 nt were introduced after codon 12 (AAAcg; Fig. 1a). This strategy gave three constructs containing the 5′-terminal 12 codons for the core from HCV strain H77, including the start codon ATG (boxed) up to a PstI cloning site (italics), followed by the ATG of the RLuc gene (underlined). This sequence was converted to encode the first 12 codons of HCV-1 core protein as described in Methods. The nucleotides added to place the RLuc gene in the −1 and +1 ORFs relative to the H77 0 ORF sequence are denoted in lower case. (b) pSGR-Luc JFH1 subgenomic replicon constructs. The nucleotide sequences for codons 8–11 are shown for each construct. ECMV IRES, encephalomyocarditis virus IRES.

**Table 1.** Sequences cloned in the bicistronic dual luciferase reporter vector

Nucleotides varying from the H77 sequence are denoted in lower case and X YYYY ZZZY slippery sequences are underlined. The codons are numbered from the ATG initiation codon of the core protein. The nucleotide number of the wobble position is indicated for each codon in the H77 sequence. Each sequence is shown in the 0 frame but constructs were also made in the −1 and in the +1 frame relative to the RLuc gene.

In vitro frameshifting efficiency measurements. PRF efficiency from the RLuc gene was measured in a rabbit reticulocyte lysate system, using a TnT Quick Coupled Transcription/Translation Systems kit (Promega) that employs T7 RNA polymerase for transcription. Transcription/translation reactions were carried out with 200 ng plasmid DNA in a total volume of 10 μl at 30 °C for 1 h. Three aliquots of 2.5 μl were assayed for chemiluminescence using an MLX Microtiter plate luminometer (Dynex Technologies) and a Dual-Glo Luciferase Assay System kit (Promega) according to the manufacturer’s instructions. The mean RLuc:FLuc ratio of the three assays was determined for each reading frame of a given construct. The ratios for individual constructs were expressed relative to the sum of RLuc:FLuc ratios (normalized to 100%) for each plasmid series (e.g. Mut1 0, −1 and +1 plasmids) and corrected for background.
values obtained with the control plasmid. Experiments were performed in triplicate.

Introduction of a poly(A) track in the JFH1 subgenomic replicon and infectious clone. Construction of the bicistronic JFH1 subgenomic replicon (pSGR-Luc-JFH1) has been described previously (Targett-Adams & McLauchlan, 2005). The first cistron consisted of the 5’ UTR of JFH1 and the 5’-terminal 57 nt of the core ORF fused to the gene encoding FLuc (Fig. 1b). Site-directed mutagenesis was carried out using a Quickchange II XL kit (Stratagene) to convert the JFH1 sequences to HCV-1 [pSGR-Luc-JFH1poly(A)] nucleotide sequences between codons 8 and 11 (Fig. 1b). Transcripts were synthesized in vitro from linearized plasmids by T7 RNA polymerase and electroporated into Huh7 cells as described previously (Targett-Adams & McLauchlan, 2005). Luciferase activity within cell extracts was assayed at 4 h after electroporation using a Luciferase Assay System (Promega) and a Turner 20/20 luminometer (Promega).

To create the JFH1poly(A) clone, site-directed mutagenesis was carried out on pJFH1 (a gift from Takaji Wakita, National Institute of Infectious Diseases, Tokyo, Japan) using a Quickchange II XL kit and mutagenesis oligonucleotides, designed to convert the JFH1 sequence to HCV-1 nucleotide sequences between codons 8 and 11. Huh7 cells were electroporated with either JFH1wt or JFH1poly(A) for 3 days. Virus released into growth medium at this time point was used to infect monolayers of naïve cells on coverslips. Prior to infection, medium containing supernatant virus was filtered using a 0.2 μm Minisart filter (Sartorius) and diluted with fresh Dulbecco’s minimal essential medium. The tissue culture 50% infective dose (TCID50) for virus released from cells was calculated based on the method described by Lindenbach et al. (2005). To prepare RNA from infectious virus, medium containing supernatant virus was filtered and loaded on to a cushion of 2 ml 20% sucrose, 50 mM Tris/HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA. Virus was pelleted by centrifugation at 100 000 g for 4 h at 4 °C. RNA in the pellet material was isolated using Trizol reagent (Invitrogen) according to the manufacturer’s instructions.

Isolation of RNA, and cDNA sequence analysis. To isolate RNA, 1 ml Trizol reagent was added to in vitro transcription reactions and RNA extraction was performed according to the manufacturer’s instructions. The same method was used to extract RNA from Huh7 cells. To eliminate DNA from samples, RNA was digested with amplification-grade DNase I (Invitrogen). DNase-treated RNA (50 ng) was used to perform RT-PCR using an Access RT-PCR System (Promega). The RT-PCR protocol was performed according to the manufacturer’s instructions as follows: reverse transcription was carried out using avian myeloblastosis virus reverse transcriptase at 45 °C for 45 min followed by denaturation at 94 °C for 2 min. cDNA amplification using Tfl DNA polymerase was carried out for 40 cycles of 30 s at 94 °C, 1 min at 52 °C and 1 min at 68 °C, followed by a final step at 68 °C for 4 min. Reverse and forward primers were located in the RLuc coding sequence and the HCV IRES, respectively, to allow amplification of a 540 bp cDNA fragment from the pLRF-generated RNA. For cDNA amplification of RNA obtained from either the subgenomic replicon or infectious JFH1 RNA, the reverse primer was specific for FLuc and E1 coding sequences, respectively.

To control for insertions or deletions of adenine residues during RT-PCR, RNA of 50 nt was synthesized chemically (Invitrogen). The RNA sequence comprised a stretch of 10 consecutive adenines in the same nucleotide context as in the pLRF vector used for in vitro experiments (5’-GAGCAGGAGAATCATCAAAAATAC-3’). Primers used for the RT-PCR control reaction were: 5’-GAGCAGGATCTAAACG-3’ (sense) and 5’-CTTTGGAGTGGG-3’ (antisense). RT-PCR was performed as described above with slight modifications to hybridization (45 s at 48 °C) and elongation (30 s at 68 °C) conditions for the PCR.

All cDNA fragments were cloned into pGEM-T Easy vectors (Promega) for sequencing. Nucleotide sequences were determined by MWG Biotech AG (Germany) and at in-house facilities in the MRC Virology Unit. Sequence comparisons were performed using CLUSTAL_W multiple-alignment software (http://npsa-pbil.ibcp.fr; Combet et al., 2000). Statistical analysis was performed using a χ² test.

RESULTS

Both +1 and −1 frameshifting occur at the adenine cluster in the HCV-1 core sequence

Initially, we compared frameshifting for strains H77 and HCV-1 by cloning their core coding sequences in a bicistronic luciferase reporter vector (Fig. 1a). The results revealed that the H77 core sequence gave low RLuc values in an in vitro translation system (Fig. 2), confirming previous reports that this virus strain does not generate frameshifted protein efficiently. By contrast, the HCV-1 core sequence produced readily detected frameshifting (Fig. 2). We found that rephasing in the +1 ORF (requiring −1 frameshifting) was 20% (± 3.1), whilst rephasing in the −1 ORF (requiring +1 frameshifting) was 11% (± 1.6).

The HCV-1 sequence contains two possible motifs, A24AAA27AAA30 and A27AAA30AAC33 (underlined in Table 1), that match the consensus −1 PRF slippery sequence X XXY YYZ (where X and Y = adenine and Z = adenosine or cytosine) (reviewed by Farabaugh, 1996; Gesteland & Atkins, 1996; Dinman et al., 1998). To identify the segment likely to be involved in frameshifting, nucleotides A26 and A32 in the HCV-1 sequence were measured. Filled bars, rephasing in the −1 ORF as a result of +1 PRF; shaded bars, rephasing in the +1 ORF as a result of −1 PRF.
converted separately to G$_{26}$ and C$_{32}$, generating two constructs named Mut2 and Mut1, respectively. Analysis with these constructs showed that converting A$_{26}$ to G$_{26}$ (Mut2) almost abolished frameshifting, as the Rluc:Fluc ratio was only slightly higher than that obtained with the strain H77 sequence (Fig. 2). By contrast, changing A$_{32}$ to C$_{32}$ (Mut1) had a less pronounced effect, lowering $-$1 frameshifting to 12 % ($\pm$1%) and $+$1 frameshifting to 7 % ($\pm$1%). These experiments demonstrated that adenine-rich sequences between nt 24 and 30 of strain HCV-1 promoted synthesis of frameshifted proteins and that the introduction of guanine at nt 26 blocked their production. Moreover, the frequency of rephasing in the $+$1 ORF (which requires $-$1 frameshifting) was consistently higher than in the $-$1 ORF (which requires $+$1 frameshifting).

**Translational rephasing does not readily account for frameshifting in the HCV-1 core sequence**

The above results suggested that a consensus slippery sequence X XXY YYZ (where X, Y and Z are adenines) between nt 24 and 30 in strain HCV-1 was able to drive $-$1 frameshifting and that this process could be blocked by mutating A$_{26}$$\rightarrow$G within the slippery sequence. To examine the additional sequence requirements for $-$1 rephasing, we created A$\rightarrow$C mutations at nt 24 (Mut3) and 30 (Mut4). These substitutions were introduced into Mut1, as the slippery sequence between nt 24 and 30 had greater frameshifting activity than the second element between nt 27 and 33. Adenine was mutated to cytosine, as slippage in the 5' direction would generate an unfavourable C–U codon/anticodon interaction at the third base in codons 8 and 10 (Lim & Curran, 2001; Agris, 2004). We predicted that the A$_{24}$$\rightarrow$C substitution in Mut3 should prevent frameshifting, whereas the A$_{30}$$\rightarrow$C alteration in Mut4 should have no appreciable effect. We found that $-$1 frameshifting was blocked as predicted for Mut3 (Fig. 2). Surprisingly, $-$1 frameshifting for Mut4 was reduced to 4.1 % ($\pm$0.5%) compared with values obtained for Mut1 (Fig. 2). These data for Mut4 suggested that the A-rich sequence between nt 24 and 30 either did not behave as a typical motif for $-$1 frameshifting or that another mechanism generated the frameshifted products.

Our initial results indicated that the segment incorporating the slippery sequence in HCV-1 and Mut1 also directed $+$1 frameshifting, which could arise from either a $+$1 PRF or a $-$2 PRF event. A $+$1 PRF event is generally induced by the presence of a ‘hungry’ codon or a stop codon at the ribosomal A site in the 0 ORF together with an abundant tRNA (reviewed by Farabaugh, 2000; Baranov et al., 2002; Namy et al., 2004). This situation does not arise in the HCV-1 sequence where codons 9 and 10 at the A site in the 0 ORF are identical in the $-$1 ORF, i.e. AAA. Our results with Mut2 demonstrated that converting codon 9 from AAA to AGA abolished $+$1 frameshifting and hence this triplet must play a central role in any such event. However, frameshifting would also rely on contributions from either codon 8 or 10. We tested the contributions of both codons based on two scenarios in which codon 9 would occupy either the A site (Fig. 3a) or the P site (Fig. 3b) on the ribosome. If it was positioned in the A site, codon 8 would be located in the P site and a mutation at the wobble position in codon 10 (Mut4, Table 1) should exert no effect on $+$1 PRF. Similarly, positioning of codon 9 in the P site would place codon 10 in the A site and a mutation at the wobble position in codon 8 (Mut3, Table 1) should not influence $+$1 PRF events. Surprisingly, no $+$1 PRF products were detected for either Mut3 or Mut4 (Fig. 2). As $+$1 PRF should require the third base in codon 8 or 10, but not both nucleotides, we concluded that tRNA slippage or overlap of a fourth base by the P-site tRNA is unlikely to explain the observed $+$1 frameshifting by the HCV-1 sequence.

As an alternative to $+$1 PRF, it was possible that tRNAs decoding codons 9 (AAA$_{24}$) and 10 (AAA$_{30}$), respectively, in the P and A sites could slip by two bases in the 5' direction, generating a $-$2 PRF event (Fig. 3c). This process is possible with HCV-1 and Mut1, as they contain the sequence CAA$_{24}$AAA$_{27}$AAA$_{30}$ in the 0 frame, which could be decoded as C$_{22}$AAA$_{25}$AAA$_{28}$AA$_{30}$ in the $-$2 ORF. However, mutations at the second or third bases in codon 8 would destroy $-$2 PRF. Although mutation at codon 8 from CAA$_{24}$ to CAC$_{24}$ (Mut3) completely abolished $+$1 PRF, altering CAA$_{24}$ to CCA$_{24}$ (Mut5) still permitted 2.8 % ($\pm$0.3%) apparent $+$1 frameshifting (Fig. 2). This failure...
to eliminate frameshifting by Mut5 indicated that a –2 PRF process was unlikely to account for rephasing in the +1 ORF.

Taken together, our results question whether the production of frameshifted products by rephasing in either the –1 or +1 ORFs arises solely through translational processes in HCV-1.

**TS occurs at the adenine cluster in vitro**

As an alternative to translational rephasing by ribosomes, frameshifted proteins could also arise from TS, which introduces or deletes nucleotides at runs of adenines or thymines in mRNA during transcription by RNA polymerases (Wagner et al., 1990; Volchkov et al., 1995). We generated cDNA from RNA that had been produced in rabbit reticulocyte lysate by T7 RNA polymerase with HCV-1 and Mut1 plasmids. cDNA fragments were cloned and the nucleotide sequences were determined for the HCV core region flanking the A-rich sequences between codons 8 and 11. To control for errors in transcribed RNA that could be derived from alterations in the length of the poly(A) tract in the plasmids added to *in vitro* translation reactions, 20 clones of each of the wt and HCV-1 constructs were sequenced. No insertion or deletion of adenine was detected in any of the clones examined (data not shown). To control for insertion or deletion of adenines during reverse transcription and PCR amplification, cDNA was also produced from a synthetic RNA of 50 nt, which contained 10 consecutive adenines.

For Mut1, we obtained 6.6 ± 0.9 % +1 frameshifting in the luciferase assay and two clones out of 47 (4.3 %) with a deletion of one adenine (Table 2 and Fig. 4a). For the same construct, we obtained 12.0 ± 0.9 % –1 frameshifting in the luciferase assay and four clones out of 47 (8.5 %) with the addition of one adenine (Table 2 and Fig. 4a). With HCV-1, +1 frameshifting was detected at a level of 11.0 ± 1.6 % in the luciferase assay, and three clones out of 27 contained a deletion of one adenine, plus one clone had the addition of two adenines (14.8 %) (Table 2 and Fig. 4b). For the same construct, –1 frameshifting was higher at 20.0 ± 3.1 % in the luciferase assay and a correspondingly greater number of six clones out of 27 (22.2 %) had the addition of one adenine (Table 2 and Fig. 4b); we also obtained one sequence with deletion of three adenines and one sequence with addition of three adenines, both leading to the decoding of luciferase in the 0 ORF (Fig. 4b). By contrast, no insertions or deletions were found in 30 clones derived from synthetic RNA containing 10 adenines, indicating that any heterogeneity in the cDNAs did not result from the RT-PCR conditions (data not shown).

Our data indicate that the extent of frameshifting measured by luciferase activity closely matched the proportion of

<table>
<thead>
<tr>
<th>Mut1</th>
<th>HCV-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLuc/FLuc (%)</td>
<td>RLuc/FLuc (%)</td>
</tr>
<tr>
<td>+1</td>
<td>6.6 ± 0.9</td>
</tr>
<tr>
<td>–1</td>
<td>12.0 ± 0.9</td>
</tr>
<tr>
<td>Sequences (%)</td>
<td>Sequences (%)</td>
</tr>
<tr>
<td>(frequency*)</td>
<td>(frequency*)</td>
</tr>
<tr>
<td>4.3 (2/47)</td>
<td>14.8 (4/27)</td>
</tr>
<tr>
<td>8.5 (4/47)</td>
<td>22.2 (6/27)</td>
</tr>
</tbody>
</table>

*Number of clones leading to translation of the corresponding alternative reading frame/total number of sequenced clones.

**Table 2.** Comparison of the percentage of *in vitro* recoding as determined by luciferase activity and sequence analysis of clones from cDNA.
adenines either inserted or deleted in the A-rich tract. Hence, we conclude that TS by T7 RNA polymerase in the A-rich tract between codons 8 and 11 is the major contributory factor accounting for detection of frameshifted products in the in vitro luciferase assay system.

**Frameshifting in subgenomic and full-length JFH1 RNAs can arise from TS in transcripts synthesized by T7 and viral NS5B RNA polymerases**

Due to the difficulties with propagation of HCV, studies of its replication have relied on in vitro transcription of RNA by T7 RNA polymerase. To determine whether frameshifted products arising from TS could be detected in a replicon system, we introduced the A-rich tract from HCV-1 into a subgenomic replicon, pSGR-Luc-JFH1, that gives high replication levels and produces luciferase activity from a chimeric protein consisting of the N-terminal 19 aa of the core protein linked to the enzyme (Targett-Adams & McLauchlan, 2005); this generated a construct termed pSGR-Luc-JFH1poly(A) (Fig. 1b). To control for errors in transcribed RNA that could be derived from alterations in the length of the poly(A) tract in the plasmids added to in vitro translation reactions, 20 clones of each of the wt and HCV-1 constructs were sequenced. No insertion or deletion of adenine was detected in any of the clones examined (data not shown). Following in vitro transcription by T7 RNA polymerase, cDNA was generated from the synthesized RNA and cloned into vector pGEM-T Easy.

Sequence analysis of 44 clones containing cDNAs from SGR-Luc-JFH1poly(A) RNA revealed that seven clones had an adenine insertion in the A-rich tract (15.9 % of clones) and the same number of clones had an adenine deletion (Table 3). Thus, almost 32 % of clones containing an A-rich tract in the JFH1 subgenomic replicon had alterations introduced by TS by T7 RNA polymerase. By contrast, cDNA inserts from wt SGR-Luc-JFH1 RNA contained no additions or deletions in the equivalent region between codons 8 and 11 (Table 3). To support evidence for enhanced production of frameshifted products by SGR-Luc-JFH1poly(A) arising by TS, we constructed replicons SGR-Luc-JFH1poly(A)/+1 and SGR-Luc-JFH1poly(A)/−1 with

<table>
<thead>
<tr>
<th>No. sequencesa</th>
<th>SGR-Luc-JFH1</th>
<th>SGR-Luc-JFH1poly(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Adenine deletion</td>
<td>0</td>
<td>7 (15.9 %)</td>
</tr>
<tr>
<td>1 Adenine insertion</td>
<td>0</td>
<td>7 (15.9 %)</td>
</tr>
<tr>
<td>Sum</td>
<td>0</td>
<td>14 (31.8 %)</td>
</tr>
</tbody>
</table>

*Total number of analysed sequences per construct was 44.

1 and 2 nt, respectively, removed from the BglII site at the junction between the coding regions for the core protein and luciferase. We predicted that replicon SGR-Luc-JFH1poly(A)/+1 would give luciferase activity from in vitro-transcribed RNAs with a nucleotide insertion, whilst SGR-Luc-JFH1poly(A)/−1 would generate activity from RNAs with a single nucleotide deletion. Corresponding constructs (SGR-Luc-JFH1wt/+1 and SGR-Luc-JFH1wt/−1) were also constructed for wt SGR-Luc-JFH1. Luciferase activity measured at 4 h after electroporation of RNA into cells, corresponding to a period before the onset of replication and indicative of translation from input RNA (Jones et al., 2007), gave enzyme levels for SGR-Luc-JFH1poly(A)/+1 and SGR-Luc-JFH1poly(A)/−1 that were 23 and 9.7 % of those obtained for SGR-Luc-JFH1poly(A) (Fig. 5a). By contrast, the luciferase activities measured for SGR-Luc-JFH1wt/+1 and SGR-Luc-JFH1wt/−1 were only 1 % of those obtained for SGR-Luc-JFH1. The difference in values between the extent of frameshifted products produced by SGR-Luc-JFH1poly(A)/+1 (23 %) and SGR-Luc-JFH1poly(A)/−1 (9.7 %) compared with the numbers of clones with insertions (15.9 %) or deletions (15.9 %) produced by in vitro transcription of SGR-Luc-JFH1poly(A) could arise from the sample size of clones whose sequences were determined (n=44). Nonetheless, these results support our evidence that TS by T7 RNA polymerase in the poly(A) tract in SGR-Luc-JFH1poly(A) gives higher levels of frameshifted products.

To extend analysis of TS in the context of the viral genome, the sequence of codons 8–11 was modified to contain an A-rich tract in the cDNA for infectious strain JFH1, to give pJFH1poly(A). Following transcription of pJFH1 poly(A) by T7 RNA polymerase, sequence analysis from a total of 76 cDNA clones gave eight clones (10 %) with deletion of an adenine residue, whilst 12 clones (16 %) had an adenine insertion (Fig. 5b). Similar to results obtained with the SGR-Luc-JFH1 subgenomic replicon, 20 cDNA clones of wt JFH1 RNA contained no insertions or deletions of adenine residues in the region encoding codons 8–11 (data not shown). Following electroporation of JFH1poly(A) RNA into cells, the titre of infectious virus released into the medium was determined to establish whether the heterogeneity in the RNA population affected virus production. By 72 h post-electroporation, the TCID50 value for JFH1poly(A) reached 6×104, compared with 1.5×105 for wt JFH1, and thus there was a 60 % reduction in virus titre.

The ability to generate infectious JFH1poly(A) enabled monitoring of any changes in the RNA population not only after electroporation but also following infection of naive cells. Hence, RNA was extracted from cells at 72 h after electroporation with JFH1poly(A) RNA and infection with JFH1poly(A) cell-released virus at an m.o.i. of less than 1. Sequence analysis of 76 cDNA clones from RNA isolated at 72 h post-electroporation revealed a higher proportion of clones with an adenine deletion (25 %) compared with those with an adenine insertion (4 %). From RNA isolated at 72 h post-infection, a similar pattern was evident (18 % of cDNAs had an adenine deletion, whereas 4 % contained

---

*Table 3. Rates of TS on T7 RNA polymerase in vitro-transcribed wt JFH1 and HCV-1 sequences in the subgenomic replicon pSGR-Luc-JFH1*
an adenine insertion; Fig. 5b). We also prepared RNA from JFH1_poly(A) cell-released virus particles isolated by centrifugation through a sucrose density gradient. From analysis of cDNA amplified from this material, loss of an adenine residue was present in 15% of clones, whilst an additional adenine residue was found in 9% of clones (Fig. 5b). As viral genomes with loss and gain of adenine residues in the A-rich tract were detected during infection, our results indicate that TS at this sequence occurs not only with T7 RNA polymerase but also with the HCV NS5B RNA polymerase. In control experiments, analysis of 20 cloned cDNA inserts generated from RNA isolated from wt JFH1 RNA-electroporated and wt JFH1-infected cells revealed no addition or removal of sequences in the core coding region (data not shown).

**DISCUSSION**

**TS causes recoding in alternative reading frames at the adenine cluster in strain HCV-1**

Since the first report of frameshifting in strain HCV-1, its sequence has been the paradigm to examine mechanisms that direct alternative translation of the HCV genome. Several studies have concluded that frameshifted products are synthesized by a process requiring tRNA slippage on the mRNA template, either in the +1 or –1 direction, during translation (Xu et al., 2001; Choi et al., 2003). Frameshifting is directed by a stretch of 10 consecutive adenine residues between codons 8 and 11 containing two potential slippery signals. We have shown that interrupting this poly(A) segment by substitution of guanine at position 26 is sufficient almost to abolish frameshifting. Analysis with additional mutants did not support a model that frameshifting in the –1 and +1 ORFs occurred solely through a translational mechanism. Rather, sequence analysis of cDNAs prepared from RNA containing the A-rich segment revealed addition and deletion of adenines between codons 8 and 11 that could account for the synthesis of frameshifted products. Therefore, slippage by RNA polymerases during transcription offers an alternative explanation for frameshifting in the strain HCV-1 genome. Nevertheless, we cannot completely rule out the possibility that –1 PRF also occurs in strain HCV-1 due to the presence of heptameric slippery sequences.

We examined cloned cDNA fragments produced from two in vitro-transcribed RNA templates (HCV-1 and Mut1) and found that an appreciable number of clones contained variable numbers of adenine residues in the A-rich tract region. There was a close correlation in the frequency of insertions/deletions in the HCV-1 and Mut1 sequences and the extent of frameshifting detected by the luciferase assay. Lower overall frequency of insertions/deletions in Mut1 was also reflected in reduced frameshifting efficiency with this mutant compared with HCV-1. We conclude that the tract of 10 adenine residues in HCV-1 promotes heterogeneity at this sequence in RNA transcribed by T7 RNA polymerase. Shortening the length of consecutive adenines by one residue (Mut1) reduced heterogeneity and insertion of non-adenines abolished TS. This correlation between the frequency of adenine insertions/deletions and the decoding in the –1/+1 ORFs indicates for the first time that TS rather than PRF at the A-rich tract is a key mechanism for frameshifting in vitro in strain HCV-1.
In vitro transcription of DNA by bacterial RNA polymerases introduces a bias in frameshifting studies on HCV-1

This is the first study on frameshifting in the HCV genome in the context of viral RNA replication and demonstrates that insertion and deletion of residues during in vitro transcription by T7 RNA polymerase plays a critical role in the synthesis of frameshifted products. Our results may explain discrepancies observed in previous experiments on frameshifting in HCV-1 using different RNA polymerases in vitro. T7 RNA polymerase has shown varying levels of frameshifting [e.g. 12–15 % +1 frameshifting and 30–45 % −1 frameshifting in a study by Choi et al. (2003) and 30 % +1 frameshifting described by Roussel et al. (2003)]. By measuring luciferase activity, we found 11 % +1 frameshifting and 20 % −1 frameshifting with the same polymerase. However, experiments with SP6 RNA polymerase generated very high levels (54 %) of +1 frameshifting (Varakioti et al., 2002). As part of the study presented here, we also determined the extent of insertion and deletion in the HCV-1 poly(A) tract sequence using a plasmid in which mRNA including the core sequences was transcribed by RNA polymerase II in HepG2 cells from a CMV promoter. Sequence data obtained from 85 clones revealed that 4.7 % of clones had an adenine insertion, whilst 5.9 % contained an adenine deletion. Overall, the percentage slippage from RNA polymerase II was 10.6 % compared with 29.2 % for the combined cDNA sequence data for transcripts generated by T7 RNA polymerase. These values were significant (P=0.0038, χ² test). Hence, different RNA polymerases give varying levels of TS efficiency, which would influence the extent of detected +1 frameshifting. If frameshifting had occurred solely through a translational mechanism, the use of different polymerases to transcribe RNA would be predicted to have no effect. Moreover, the different extent of TS by T7 RNA polymerase and RNA polymerase II provides further evidence that insertion and deletion in the poly(A) tract in the HCV-1 sequence is not an artefact of cDNA synthesis by reverse transcriptase. As HCV propagation directly from infectious sera in tissue culture cells is not possible, RNA replication and virus production systems have relied on generating transcripts by prokaryotic polymerases (Lohmann et al., 1999; Lindenbach et al., 2005; Wakita et al., 2005; Zhong et al., 2005). In vitro-synthesized transcripts direct production of HCV not only in tissue culture cells but also in animal models through intrahepatic injection of RNA. For studies investigating the significance of frameshifted proteins in chronic infection, our analysis highlights the potential for difficulties with data interpretation, as the frequency of frameshifting in sequences containing the poly(A) tract are most probably enhanced in systems utilizing RNA produced in vitro.

Involvement of NS5B-dependent TS in recoding

Despite the heterogeneity in the A-rich tract introduced by T7 RNA polymerase in JFH1_poly(A) RNA, it was possible to examine TS, which was dependent on the viral NS5B RNA polymerase, by inoculating cells with virus progeny produced by this construct. We predict that RNAs generated by T7 RNA polymerase, with either an insertion or deletion of an adenine residue, would not produce the viral polypeptide and hence would be incapable of replication. Given the amount of RNA used for electroporation, it is likely that cells would contain both replication-competent and -incompetent genomes. Replication of incompetent genomes would be possible through complementation with replicase complexes generated by RNAs with 10 adenines in the A-rich tract. Moreover, incompetent genomes, which are replicated by this mechanism, would be available for packaging. Our detection of a heterogeneous RNA population in infectious virus particles confirms this notion. However, propagation of replication-incompetent genomes is highly unlikely for two reasons. Firstly, the m.o.i. was less than 1 and, secondly, multiple infection of tissue culture cells by HCV is highly inefficient (Schaller et al., 2007). Therefore, it is probable that cells are inoculated with a single infectious unit. Under such circumstances, it would be predicted that any RNAs with heterogeneity in the A-rich tract would reduce in abundance. However, the proportion of RNAs with heterogeneous sequences remained similar at the same time period after electroporation and infection. From these results, we conclude that TS in infected cells arises from transcription and replication of viral genomes by the NS5B RNA polymerase.

In addition to propagation of a heterogeneous RNA population, we observed a change in bias towards loss of adenine residues after electroporation and infection, which was highly significant (P=0.00037). By contrast, RNA transcribed by T7 RNA polymerase showed no significant difference in the probability for inserting or deleting adenines. The source of this shift is not obvious, but could arise from different properties of T7 and NS5B RNA polymerases such that removal of adenine in a poly(A) tract is favoured by the HCV enzyme. Alternatively, decoding of genomes with adenine insertions would terminate translation at codon 26. Such RNAs with abortive translation could be rapidly degraded by the non-sense-mediated decay pathway (Conti & Izaurralde, 2005; Weischenfeldt et al., 2005). By contrast, genomes with an adenine deletion would give a translated product of 125 aa. Hence, RNA giving rise to this longer polypeptide may not be targeted to the same decay pathway.

Relevance of TS-prone sequences to production of frameshifted proteins in vivo

Analysis of 3442 HCV sequences deposited in the European HCV database (http://euhcvdb.ibcp.fr) revealed nine other entries with sequences similar to that of HCV-1 that could promote TS (data not shown). This suggests that insertion or deletion of adenines between codons 8 and 11 is likely to be extremely rare, and no such heterogeneity has been
reported. Nevertheless, HCV exists in infected patients as a complex mixture of closely related variants referred to as quasispecies (Weiner et al., 1991; Martell et al., 1992). These variants are generated continuously during virus replication, as a result of mutations introduced by the viral RNA polymerase, which lacks proofreading ability. Hence, minor quasispecies may arise that harbour TS-prone sequences, which could facilitate insertion or deletion of nucleotides. This mechanism would allow a frameshift protein production, thereby providing an explanation for the prevalence of antibodies against such products (10–62% depending on the study group; Komurian-Pradel et al., 2004; Troesch et al., 2005). Sequence analysis of the minor quasispecies from infected patients with high levels of antibodies against frameshifted proteins would help to address this question. An alternative explanation for the high prevalence of anti-ARFP antibodies would be expression of the core protein +1 ORF by alternative mechanisms such as internal initiation, as described previously (Vassilaki & Mavromara, 2003; Baril & Brakier-Gingras, 2005).

To conclude, our results reveal that TS is an alternative mechanism to account for frameshifting in the HCV-1 sequence and is dependent on a noninterrupted, contiguous stretch of adenine residues. Hence, transcriptional as well as translational processes may contribute to the detection of frameshifted products generated by HCV.

ACKNOWLEDGEMENTS

We thank Dr A. Cahour for the gift of the pIRF vector, Dr C. Wychowski for the gift of the H77 clone and Dr T. Wakita for the gift of the JFH1 clone. We are grateful to Clare Addison for determining nucleotide sequences and Derek Gatherer for statistical analysis. J.-P.L. and M.R. were supported by grants from the Centre National de la Recherche Scientifique; Université Claude Bernard; Agence Nationale de Recherche sur le Sida et les hépatites virales (60/0620 to J.-P.L., fellowship to M.R.) and the European Community (HepCVax Cluster EU FP5 grant QLK2-2002-01329, VIRGIL Network of Excellence EU FP6 grant LSHM-CT-2004-503359 to the European HCV database). S.B. was supported by a Marie Curie Intra-European Fellowship (contract number 025198).

REFERENCES


