Genetic variability in the 5′ UTR and NS5A regions of hepatitis C virus RNA isolated from non-responding and responding patients with chronic HCV genotype 1 infection

Flávio MG Araújo/+, Ivan V Sonoda, Nilton B Rodrigues, Rosangela Teixeira¹, Rodrigo AF Redondo², Guilherme C Oliveira/³

Instituto de Pesquisa René Rachou-Fiocruz, Av. Augusto de Lima 1715, 30190-002 Belo Horizonte, MG, Brasil ¹Instituto Alfa de Gastroenterologia, Hospital das Clínicas ²Laboratório de Biodiversidade e Evolução Molecular, Departamento de Biologia Geral, ICB, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil ³Programa de Pós-Graduação e Pesquisa, Santa Casa de Belo Horizonte, Belo Horizonte, MG, Brasil

Sequence variation among different hepatitis C virus (HCV) isolates has adaptive significance and reflects the modes and intensities of selection mechanisms operating on the virus. In this work, we sought to investigate using classical population genetics parameters, the genetic variability of HCV genotype 1 using the 5′ UTR and NS5A regions from treatment non-responding and responding groups of patients. Both regions showed low genetic variability and the 5′ UTR showed neutral deviation. No differences were observed in the nonsynonymous/synonymous nucleotide substitution ratio among groups for NS5A. The analysis of molecular variance test of the 5′ UTR region showed an 11.94% variation among groups. Phylogenetic analysis showed no correlation between sequence variations and therapeutic responses.

Key words: hepatitis C virus - 5′ UTR - NS5A - therapeutic

Hepatitis C virus (HCV) is the only species in the genus Hepacivirus within the Flaviviridae family. HCV genetic variability results from a typical Darwinian evolutionary process that, by continuous diversification of viral populations, leads to the environment permanently selecting the fittest variants (Pawlotsky 2006). This variability is favored by rapid virus replication, large viral population sizes and high mutation rates, a consequence of the low fidelity of HCV RNA-dependent RNA polymerase (Duarte et al. 1994). Hence, HCV circulates within an infected host as a heterogeneous viral population containing closely related but genetically distinct variants, known as quasispecies (Bukh et al. 1995).

Phylogenetic analyses of partial or full-length HCV sequences from strains isolated in different regions of the world have led to the identification of six major genotypes, containing closely related subtypes (Simmonds et al. 1993). Genotype 1 is the most prevalent in Brazil, followed by genotypes 3 and 2 (Campiotto et al. 2005, Martins et al. 2006, da Silva et al. 2007, Freitas et al. 2008). The genomes of these genotypes differ from each other by approximately 30-35% in their nucleotide compositions, whereas different subtypes typically differ by about 20-25% (Simmonds 2004). HCV genomes that belong to a single genotype have shown 5-8% diversity in nucleotide sequences and 4-5% diversity in amino acid sequences (Kato 2001).

The HCV genome contains a highly conserved region, the 5′ untranslated region (5′ UTR) and moderately variable regions such as NS5A, which codes for a non-structural protein. Those regions are commonly used in HCV genotyping and evolutionary studies (Marggraf et al. 2004, Fan et al. 2005). The 5′ UTR is a region where mutations are rare and sometimes compensatory, preserving the base-pairing pattern in order to conserve structural properties related to translation efficiency (Soler et al. 2002). NS5A has limited natural amino acid variability, which conserves its functional properties in vivo (Pawlotsky 2006).

In this study, we sought to investigate, using classical population genetics parameters, the genetic variability in the viral 5′ UTR and NS5A regions in not responding to treatment and responding patients with chronic hepatitis C from state of Minas Gerais, Brazilian Southeast region.

According to the exclusion criteria of association, individuals with metabolic, alcoholic or autoimmune liver diseases, or who were co-infected with HBV/HIV, were excluded from the study. A total of 33 serum samples from individuals with confirmed HCV genotype 1 chronic hepatitis were obtained during the pre-treatment clinical evaluation.

All patients were treated with pegylated interferon and ribavirin for 48 weeks. At the end of the treatment, serum samples were collected again and re-evaluated by RT-PCR (HCV RNA qualitative, Roche, lower limit of detection of 50 UI/mL). Patients were then classified into two groups according to their virological responses: non-responder (N) (n = 20 with positive PCR) or responder (R) (n = 13 with negative PCR). The study was approved by the Research Ethics Committee of the Universidade Federal de Minas Gerais (Belo Horizonte,
MG, Brazil). A written informed consent was obtained from each participant prior to the proceedings.

HCV RNA was isolated from 140 µL serum samples using the QIAamp viral RNA kit (Qiagen, USA). 5' UTR and NS5A cDNA synthesis was performed using SuperScript™ II Reverse Transcriptase (Invitrogen, USA) and specific antisense primers: Quasi 3' for the 5' UTR (Laporte et al. 2000), 1aNS5AR1 for the NS5A genotype 1a and 1bNS5AR1 for the NS5A genotype 1b. These primers amplified the PKR-binding domain (PKRBD) and the interferon sensitivity-determining region (ISDR) (Macquillan et al. 2004). The 5' UTR and NS5A cDNA fragments were amplified by nested PCR. For the 5' UTR, the Quasi 3' and Quasi 3' primers were used for first-round amplification and the IRES 5' and IRES 3' primers were used for the second round, producing a 350 bp amplicon. Primer IRES 5' has a restriction site for BamHI and primer IRES 3' has a restriction site for PstI, to facilitate cloning into the pIRF vector (Laporte et al. 2000). Plasmid DNA was purified using the QIAprep Spin Miniprep Kit (Qiagen, USA). For NS5A genotype 1a, the 1aNS5AF1 and 1aNS5AR1 primers were used for first-round amplification and the IRES 5' and IRES 3' primers were used for the second round (nt 6633-7379), producing a 396 bp amplicon (Macquillan et al. 2004). PCR conditions were the same for both regions: 35 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min, followed by a final elongation step of 72°C for 7 min. PCR fragments were isolated by electrophoresis in agarose gels and purified using Qiagen DNA purification kits.

The 5' UTR plasmid DNA and the NS5A PCR amplicons were sequenced using the ThermoSequence II dye terminator cycle sequencing kit (GE Healthcare, USA) in a MegaBace 1000 sequencer (GE Healthcare, USA). Quality attributes of the nucleotide sequences were obtained with Phred (Ewing et al. 1998). Six recombinant clones of the 5' UTR and six amplicons of each NS5A genotype were sequenced. The nucleotide sequences were deposited in GenBank under the accession numbers: EU360315 to EU360324 for the 5' UTR and EU360325 to EU360353 for NS5A.

Sequences of the HCV 5' UTR plasmids, NS5A amplicons and 5' UTR-NS5A concatamers from the 33 serum samples were aligned using Clustal X v1.81 (Thompson et al. 1997) and used for population-genetic and phylogenetic analyses. Different measures of polymorphism were used to estimate genetic variability in the HCV sequences. DnaSP 4.5 (Rozas et al. 2003) was used to estimate the number of singletons (\(\eta_s\)), number of haplotypes (\(n\)), nucleotide diversity (\(\pi\)), diversity of haplotypes (\(h\)), average number of nucleotide differences (\(K\)), Tajima’s D (Tajima 1989), Fu and Li’s D* and F* (Fu and Li 1993) tests of evidence for non-neutral evolution and the mean \(\omega\) ratio (non-synonymous - Ka and synonymous - Ks). The mean \(\omega\) ratio of Ka and Ks is an important indicator of selective pressure at the protein level: \(\omega = Ka/Ks = 1\) is interpreted as neutral evolution, \(\omega < 1\) as purifying selection and \(\omega > 1\) as diversifying selection.

AMOVA (Excoffier et al. 1992) was applied to evaluate genetic differentiation among NR and R and among samples within the groups. These analyses were conducted using Arlequin v3.1 (Excoffier et al. 2005).

In order to summarize the results, we constructed a neighbor-joining dendrogram (Saitou & Nei 1987) with the Kimura two-parameter substitution model distance, using the 5' UTR-NS5A concatamers and MEGA v3.1 software (Kumar et al. 2004).

The NR group showed the highest \(n\) the highest values of \(\eta_s\) (\(\Phi_s\)) and, like the R group, low values for \(\pi\) and \(K\) for both regions. The NR group also showed slightly higher \(h\) in the NS5A region (Table) compared with R, which may be explained by the occurrence of specific

### Table

<table>
<thead>
<tr>
<th>Genetic variability</th>
<th>5' UTR</th>
<th>NS5A</th>
<th>Concatamers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>NR</td>
<td>R</td>
</tr>
<tr>
<td>(\eta_s)</td>
<td>14</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>(n)</td>
<td>5</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>(h)</td>
<td>0.731</td>
<td>0.584</td>
<td>0.987</td>
</tr>
<tr>
<td>(\pi)</td>
<td>0.007</td>
<td>0.005</td>
<td>0.153</td>
</tr>
<tr>
<td>(K)</td>
<td>2.692</td>
<td>1.837</td>
<td>28.167</td>
</tr>
<tr>
<td>Neutrality tests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tajima’s D</td>
<td>-1.850</td>
<td>-2.203</td>
<td>-0.131</td>
</tr>
<tr>
<td>Fu and Li D*</td>
<td>-2.522</td>
<td>-3.312</td>
<td>0.503</td>
</tr>
<tr>
<td>Fu and Li F*</td>
<td>-2.676</td>
<td>-3.470</td>
<td>0.380</td>
</tr>
<tr>
<td>Nucleotides substitution</td>
<td>0.141</td>
<td>0.122</td>
<td></td>
</tr>
</tbody>
</table>

\(a: p < 0.05\); \(h\): haplotypes diversity; \(K\): average number of nucleotide differences; \(Ka\): average number of non-synonymous substitutions; \(Ks\): average number of synonymous substitutions; \(n\): number of haplotypes; NR: non-responders patients; R: responders patients; \(\eta_s\): number of singletons; \(\pi\): nucleotide diversity.
selection mechanisms correlated to immune evasion operating on the NS5A region. However, this possibility requires further analysis as it may be random. NS5A protein is not only related to the complex virus replication mechanism, but also has activities such as binding to and inactivating PKR (Gale et al. 1997), blocking apoptotic pathways through sequestration of p53, modulating intracellular calcium levels and inducing anti-inflammatory interleukin 8 secretion (Polyak et al. 2001).

Both groups (NR and R) showed deviations from neutrality in the 5' UTR region. In contrast, no significant deviation values were observed for NS5A and the 5' UTR-NS5A concatemers (Table). Negative values for Tajima’s $D$ can arise in cases of population expansion (for instance, after a bottleneck), selective sweeps and weak purifying selection (Hedrick 2005). At the beginning of HCV infection, there is a reduction in the virus population, a bottleneck effect. It has been demonstrated that the bottleneck effect reduces infection progress in acute infection (Crandall et al. 1999). One general interpretation of our results is that either the 5' UTR region is under negative selective pressure or our samples have experienced a recent bottleneck, or perhaps both. On the other hand, negative values for Fu and Li’s $D^*$ and $F^*$ tests are frequently attributed to an excess of recently arrived $n$, suggesting that either population expansion or background selection has occurred (Fu & Li 1993). However, it is impossible to distinguish between selection and expansion as the cause of negative $D^*$ and $F^*$ values.

Similar values for $K_a$ and $K_s$ were observed in both groups (Table). Neither $K_s$ nor $K_a$ substitution rates, calculated for the NS5A region, showed differences among groups and, in both cases, $K_s/K_a$ substitution ratio was smaller than 1, indicating purifying selection. In this context, the evolutionary features of HCV show no indication of positive diversifying selection in both NR and R groups.

AMOVA tests showed 11.94% variation among groups and 88.06% variation within groups, with a $\Phi_{st} = 0.12$ for the 5' UTR region and no significant values for NS5A. Although we observed changes in the 5' UTR and NS5A regions and high percentages of variation among groups, we observed no significant differentiation values for any region when analyzing both $\Phi_{st}$ and AMOVA values. The 5' UTR-NS5A dendogram summarizes the phylogenetic analysis in which no correlation was observed between sequence variations and therapeutic responses (Figure).

HCV genotypes and subtypes exhibit complex epidemiological patterns with respect to geographical distribution, prevalence, response to treatment and transmission mode. The recognition of factors responsible for this complex epidemiology is difficult, but will undoubtedly contribute to a better comprehension of HCV genetic dynamics, which is required to establish a preventive strategy of disease control (Jimenez-Hernandez et al. 2007).

To our knowledge, this is the first study performed in Brazil aiming to investigate HCV genetic variability in serum samples from patients with chronic hepatitis C using molecular-genetic population tools.

In this work, we found evidence for greater positive selection of the viral population via the 5' UTR region for both NR and R groups.

In conclusion, virological response to HCV is a complex issue (Enomoto et al. 1995, Yamamoto et al. 1997, Hofgartner et al. 1997) and is believed to be better correlated to host factors than to genetic variability of the virus. However, this issue deserves further investigation, including tests of other viral markers and whole-genome screening.

ACKNOWLEDGEMENTS
To Dr. Annie Cahour, for the pIRF plasmids.

REFERENCES
Campiotto S, Pinho JR, Carrilho FJ, Da Silva LC, Souto FJ, Spinelli


