

# 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<sup>1</sup>\*, Ivan V Sonoda, Nilton B Rodrigues, Rosângela Teixeira<sup>1</sup>,  
Rodrigo AF Redondo<sup>2</sup>, Guilherme C Oliveira<sup>3</sup>

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

Financial report: CNPq (472862/2004-9), FAPEMIG (EDT-80/05)

\* Corresponding author: araujo@cpqrr.fiocruz.br

Received 8 May 2008

Accepted 19 August 2008

MG, Brazil). A written informed consent was obtained from each participant prior to the proceedings.

HCV RNA was isolated from 140  $\mu$ L 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 genotype 1a/1b cDNA fragments were amplified by nested PCR. For the 5' UTR, the Quasi 5' 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 *Bam*HI and primer IRES 3' has a restriction site for *Pst*I, 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 (nt 6633-7379) and the 1aNS5AF2 and 1aNS5AR2 primers were used for the second round (nt 6904-7193), producing a 290 bp amplicon. For NS5A genotype 1b, the 1bNS5AF1 and 1bNS5AR1 primers were used for first-round amplification (nt 6638-7278) and primers 1bNS5AF2 and 1bNS5AR2 were used for the second round (nt 6834-7229), 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 concatemers from the 33 serum samples were aligned using Clustal X v1.81 (Thompson et al. 1997) and used for population-genetic and phylogenetics 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 -  $K_a$  and synonymous -  $K_s$ ). The mean  $\omega$  ratio of  $K_a$  and  $K_s$  is an important indicator of selective pressure at the protein level:  $\omega = K_a/K_s = 1$  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 concatemers 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

HCV genetic variability, neutrality tests and nucleotides substitution rates values for 5' UTR, NS5A and 5' UTR-NS5A concatemers

Genetic variability	5' UTR		NS5A		Concatemers	
	R	NR	R	NR	R	NR
$\eta_s$	14	15	16	35	30	50
$n$	5	7	12	19	13	20
$h$	0.731	0.584	0.987	0.995	1.000	1.000
$\pi$	0.007	0.005	0.153	0.143	0.053	0.050
$K$	2.692	1.837	28.167	26.521	30.859	28.358
Neutrality tests						
Tajima's $D$	-1.850 <sup>a</sup>	-2.203 <sup>a</sup>	-0.131	-1.208	-0.401	-1.344
Fu and Li $D^*$	-2.522 <sup>a</sup>	-3.312 <sup>a</sup>	0.503	-0.797	0.030	-1.130
Fu and Li $F^*$	-2.676 <sup>a</sup>	-3.470 <sup>a</sup>	0.380	-1.076	0.030	-1.395
Nucleotides substitution						
$K_a/K_s$	-	-	0.141	0.122	-	-

<sup>a</sup>:  $p < 0.05$ ;  $h$ : haplotypes diversity;  $K$ : average number of nucleotide differences;  $K_a$ : average number of non-synonymous substitutions;  $K_s$ : 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).

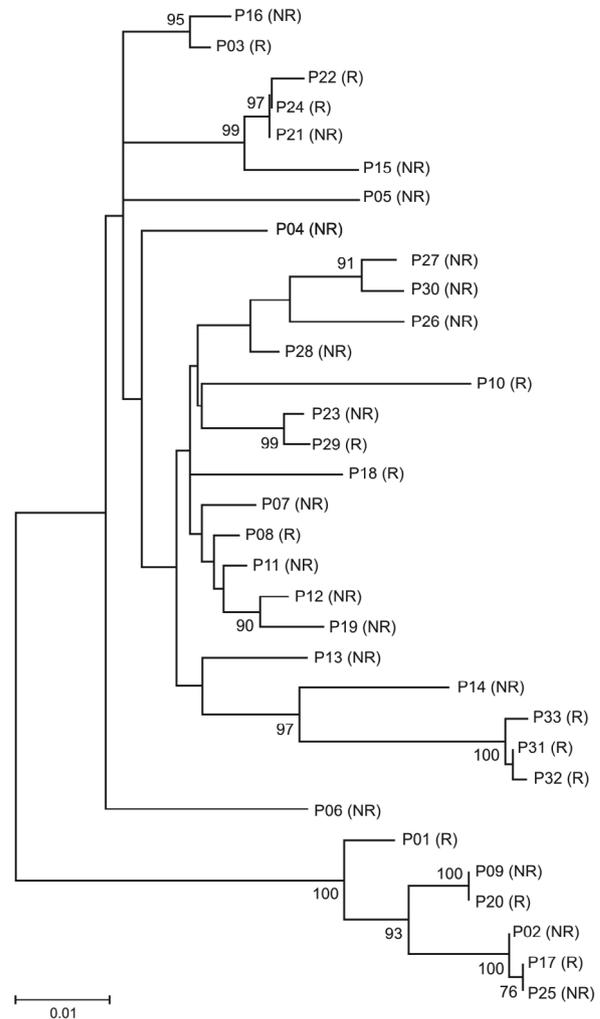
Both groups (NR and R) showed deviations from neutrality in the 5' UTR region. In contrast, no significant deviation values were observed for NS5A and the 5' UTR-NS5A concatemers (Table). Negative values for Tajima's *D* can arise in cases of population expansion (for instance, after a bottleneck), selective 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 negatives *D\** and *F\** values.

Similar values for *K<sub>a</sub>* and *K<sub>s</sub>* were observed in both groups (Table). Neither *K<sub>s</sub>* nor *K<sub>a</sub>* substitution rates, calculated for the NS5A region, showed differences among groups and, in both cases, *K<sub>s</sub>/K<sub>a</sub>* 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 dendrogram 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.



Neighbor-joining dendrogram showing phylogenetic relationships among HCV samples. Branch lengths are proportional to the expected number of mutations following the Kimura two parameter model of evolution. The dendrogram was constructed using the 5' UTR-NS5A concatemers. Bootstrap values, obtained after 1,000 sampling replications, are shown only when higher than 70%. NR: non-responder patients; R: responder patients.

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**

Bukh J, Miller RH, Purcell RH 1995. Genetic heterogeneity of the hepatitis C virus. *Princess Takamatsu Symp* 25: 75-91.  
 Campiotto S, Pinho JR, Carrilho FJ, Da Silva LC, Souto FJ, Spinelli

- V, Pereira LM, Coelho HS, Silva AO, Fonseca JC, Rosa H, Lacet CM, Bernardini AP 2005. Geographic distribution of hepatitis C virus genotypes in Brazil. *Braz J Med Biol Res* 38: 41-49.
- Crandall KA, Vasco DA, Posada D, Imamichi H 1999. Advances in understanding the evolution of HIV. *AIDS* 13 (Suppl A): S39-S47.
- da Silva CM, Costi C, Krug LP, Ramos AB, Grandi T, Gandolfi VL, Menezes ME, Ocampos M, Niel C, Rossetti ML 2007. High proportion of hepatitis C virus genotypes 1 and 3 in a large cohort of patients from Southern Brazil. *Mem Inst Oswaldo Cruz* 102: 867-870.
- Duarte EA, Novella IS, Weaver SC, Domingo E, Wain-Hobson S, Clarke DK, Moya A, Elena SF, de la Torre JC, Holland JJ 1994. RNA virus quasispecies: significance for viral disease and epidemiology. *Infect Agents Dis* 3: 201-214.
- Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Izumi N, Marumo F, Sato C 1995. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *J Clin Invest* 96: 224-230.
- Ewing B, Hillier L, Wendl MC, Green P 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175-185.
- Excoffier L, Lawlor S, Schneider S 2005. Arlequin ver 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1: 47-50.
- Excoffier L, Smouse PE, Quattro JM 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.
- Fan W, Zhu W, Wei L, Wang Q, Yin L, Du S, Zhuang H 2005. Non-structural 5A gene variability of hepatitis C virus (HCV) during a 10-year follow up. *J Gastroenterol* 40: 43-51.
- Freitas SZ, Cunha RV, Martins RMB, Teles SA, Ibanhes ML, Motta-Castro ARC 2008. Prevalence, genotypes and risk factors associated with hepatitis C virus infection in hemodialysis patients in Campo Grande, MS, Brazil. *Mem Inst Oswaldo Cruz* 103: 405-408.
- Fu YX, Li WH 1993. Statistical tests of neutrality of mutations. *Genetics* 133: 693-709.
- Gale MJ Jr, Korth MJ, Tang NM, Tan SL, Hopkins DA, Dever TE, Polyak SJ, Gretch DR, Katze MG 1997. Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. *Virology* 230: 217-227.
- Hedrick PW 2005. *Genetics of Populations*, 3rd ed., Jones and Bartlett, Boston, 737 pp.
- Hofgartner WT, Polyak SJ, Sullivan DG, Carithers RL Jr, Gretch DR 1997. Mutations in the NS5A gene of hepatitis C virus in North American patients infected with HCV genotype 1a or 1b. *J Med Virol* 53: 118-126.
- Jimenez-Hernandez N, Torres-Puente M, Bracho MA, Garcia-Robles I, Ortega E, del Olmo J, Carnicer F, Gonzalez-Candelas F, Moya A 2007. Epidemic dynamics of two coexisting hepatitis C virus subtypes. *J Gen Virol* 88: 123-133.
- Kato N 2001. Molecular virology of hepatitis C virus. *Acta Med Okayama* 55: 133-159.
- Kumar S, Tamura K, Nei M 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5: 150-163.
- Laporte J, Malet I, Andrieu T, Thibault V, Toulme JJ, Wychowski C, Pawlowsky JM, Huraux JM, Agut H, Cahour A 2000. Comparative analysis of translation efficiencies of hepatitis C virus 5' untranslated regions among intraindividual quasispecies present in chronic infection: opposite behaviors depending on cell type. *J Virol* 74: 10827-10833.
- Macquillan GC, Niu X, Speers D, English S, Garas G, Harnett GB, Reed WD, Allan JE, Jeffrey GP 2004. Does sequencing the PKRBD of hepatitis C virus NS5A predict therapeutic response to combination therapy in an Australian population? *J Gastroenterol Hepatol* 19: 551-557.
- Margraf RL, Erali M, Liew M, Wittwer CT 2004. Genotyping hepatitis C virus by heteroduplex mobility analysis using temperature gradient capillary electrophoresis. *J Clin Microbiol* 42: 4545-4551.
- Martins RM, Teles SA, Freitas NR, Motta-Castro AR, Souto FJ, Musisi A, Amorim RM, Martins CR 2006. Distribution of hepatitis C virus genotypes among blood donors from mid-west region of Brazil. *Rev Inst Med Trop São Paulo* 48: 53-55.
- Pawlowsky JM 2006. Hepatitis C virus population dynamics during infection. *Curr Top Microbiol Immunol* 299: 261-284.
- Polyak SJ, Khabar KS, Rezeiq M, Gretch DR 2001. Elevated levels of interleukin-8 in serum are associated with hepatitis C virus infection and resistance to interferon therapy. *J Virol* 75: 6209-6211.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.
- Saitou N, Nei M 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406-425.
- Simmonds P 2004. Genetic diversity and evolution of hepatitis C virus - 15 years on. *J Gen Virol* 85: 3173-3188.
- Simmonds P, Holmes EC, Cha TA, Chan SW, Mcomish F, Irvine B, Beall E, Yap PL, Kolberg J, Urdea MS 1993. Classification of hepatitis-C virus into 6 major genotypes and a series of subtypes by phylogenetic analysis of the Ns-5 region. *J Gen Virol* 74: 2391-2399.
- Soler M, Pellerin M, Malnou CE, Dhumeaux D, Kean KM, Pawlowsky JM 2002. Quasispecies heterogeneity and constraints on the evolution of the 5' noncoding region of hepatitis C virus (HCV): Relationship with HCV resistance to interferon-alpha therapy. *Virology* 298: 160-173.
- Tajima F 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876-4882.
- Yamamoto C, Enomoto N, Kurosaki M, Tazawa J, Izumi N, Marumo F, Sato C 1997. Nucleotide sequence variations in the 5' untranslated region of hepatitis C virus-1B: No association with efficacy of interferon therapy or serum HCV-RNA levels. *Hepatology* 26: 327-328.