

# Genetic and phenotypic structure of hepatitis C virus NS3/4 protease quaspecies

Sandra Franco, Mariona Parera, Ester Aparicio Bonaventura Clotet, and Miguel Angel Martínez

Fundació irsiCaixa

Laboratori de Retrovirologia irsiCaixa, Hospital Universitari Germans Trias i Pujol, Universitat Autònoma de Barcelona, 08916 Badalona, Spain

## Introduction

The hepatitis C virus (HCV) NS3/4A serine protease is not only involved in the viral polyprotein processing but also efficiently blocks the retinoic-acid-inducible gen I (RIG-I) and toll-like receptor (TLR)-3 signaling pathways and contributes to virus persistence by enabling HCV to escape the interferon (IFN) antiviral response. The NS3/4 protease performs this function by cleaving Cardif (also termed VISA, MAVS or IPS-1) and TRIF proteins to block RIG-I and TLR-3 signaling, respectively. NS3/4 protease inhibition can prevent Cardif and/or TRIF inactivation during HCV infection, thereby maintaining the innate immune response. Thus, differences in NS3/4 protease catalytic efficiency could be related to both viral pathogenicity and efficacy of IFN-based therapies.

The high mutation rate associated with HCV replication results in the generation of swarms of mutants known as viral quaspecies. Early evolution of the viral quaspecies was shown to predict the clinical outcome of acute hepatitis C. Likewise, the long-term viral evolution was shown to correlate with the severity of liver disease.

## Aim

In the present study, we analyzed, at a high resolution (1%), the genotype and enzymatic activity of 3 HCV NS3/4 protease quaspecies. We determined the catalytic efficiency of each variant present in the quaspecies in order to establish the relationships between genotype, phenotype and fitness.

## Materials and Methods

**Individuals.** Three HCV infected individuals A, B and C were chosen for this study. The HCV genotype for the 3 individuals was 1b. Individuals A and B were coinfecting with human immunodeficiency virus type 1 (HIV-1), while individual C was HCV monoinfected (Table 1). Individual C sample was taken during acute infection, afterward this individual resolved HCV infection.

**Recovery and analysis of HCV sequences.** Individual plasmid clones carrying the HCV NS3 protease coding region were generated from a single time point plasma sample from 3 HCV infected individuals (Table 1). Ninety six, 97 and 103 clones from each individual sample, respectively, were cloned and sequenced.

**Genetic screen for determining the catalytic efficiency of HCV NS3/4 proteases.** The catalytic efficiency of the different HCV NS3/4 protease variants was determined using a bacteriophage lambda ( $\lambda$ ) based genetic screen as previously described (Martinez et al. 2003; Franco et al. 2007).

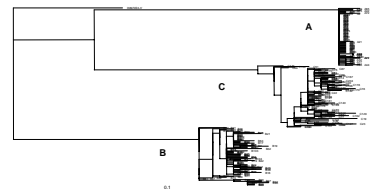


Figure 1. Neighbor-joining phylogram of HCV NS3/4 protease sequences from quaspecies A, B and C. The HCV subgenomic replicon 1389/NS3-3 sequence was used as outgroup. For the phylogenetic analysis PAUP 4.0 software package was used with a GTR + G model of evolution.

**Genetic structure of HCV NS3/4 protease quaspecies.** Neighbor-joining phylogenetic reconstruction of all NS3 protease nucleotide sequences showed that sequences from each individual produced a monophyletic group, which were supported by bootstrap analysis (Fig. 1). Intrasample genetic distances, Shannon entropy (Sn) and nucleotide heterogeneity (Hn) calculations confirmed that quaspecies A was more homogeneous than quaspecies B and C (Table 1). A different picture was observed when calculations were performed at the amino acid level. The lowest values for Da, Ha and Sa could be found within quaspecies C, which was the most heterogeneous quaspecies at the nucleotide level (Table 1 and Fig. 1). This result reflected the different pressure constrains between viral quaspecies. The ds/dn ratios were, in the 3 quaspecies, greater than 1, suggesting a preponderance of genetic drift over selection within the studied coding region. Nevertheless, different values were found for each quaspecies, 1.6, 4.8 and 15.9, respectively, indicating, once again, that different selective constrains could be acting on different quaspecies. Amino acid alignments showed a similar population structure for the 3 studied quaspecies (Fig. 2). One major form with frequencies of 58%, 45% and 66%, respectively, was identified in the 3 quaspecies. A second group of sequences representing between 2 and 6 % of sequences could be also found in every quaspecies. A common characteristic that shared the 3 studied quaspecies was the high proportion of unique variants. Some mutations conferring resistance to NS3/4 protease inhibitors (PIs) were found in 2 quaspecies (Fig. 2). The mutations V36A and R109K (clones B19 and B54) have been involved in the acquisition of resistance to the HCV NS3/4A protease inhibitor VX-950 and SCH 6, respectively. Likewise, substitution V170A (clone C30) has been involved in the acquisition of resistance to the protease inhibitor SCH 503034.

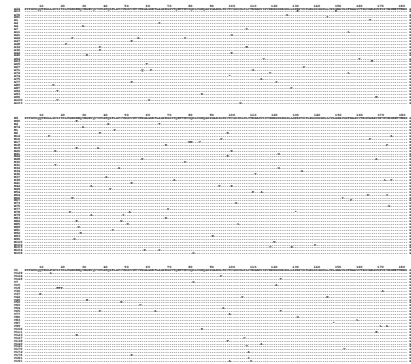


Figure 2. Amino acid sequence alignment of the three HCV NS3/4 protease quaspecies. Amino acid changes are indicated relative to the master (most abundant) sequence. The number (%) of occurrences within each sample of identical amino acid sequences is given on the right at the end of each sequence.

Table 1. Clinical and virological characteristics of the three studied HCV infected individuals.

Individual	HCV viral load (IU/ml)	HCV-1 viral load (RNA copies/ml)	ALT (IU/ml)	AST (IU/ml)	CD4* (cells/ $\mu$ l)	Hb (g/dl)	Hs (%)	Sn (%)	Sa (%)	Dg (%)	Da (%)	Ds (%)	Dn (%)	Ratio (% + SD)	
A	2364	<80	13	20	571	96	48.9	34.3	0.59	0.44	0.3 ± 0.2	0.8 ± 0.8	0.45 ± 0.4	0.28 ± 0.2	1.6
B	109142	<80	37	27	992	97	84.5	48.4	0.92	0.61	1.1 ± 0.5	1 ± 0.7	2.6 ± 1.1	0.54 ± 0.4	4.8
C	1300000	-	862	ND	ND	103	91.3	31.1	0.96	0.38	1.2 ± 0.6	0.5 ± 0.5	3.5 ± 1.2	0.22 ± 0.2	15.9

Hn, nucleotide heterogeneity; Sn, nucleotide normalized Shannon entropy; Hs, amino acid heterogeneity; Sa, amino acid normalized Shannon entropy; Dg, genetic distance; Da, amino acid distance; ds, synonymous substitutions; dn, non-synonymous substitutions.

## Results

**Catalytic efficiency of HCV NS3/4 protease quaspecies variants.** First, we evaluated the enzymatic activities of single variant proteases by engineering in the cl  $\lambda$  repressor the HCV polyprotein NS5A/NS5B cleavage site. The enzymatic activities were related to the activity of the HCV subgenomic replicon 1389/NS3-3 NS3/4 protease (100%). The 3 master protease enzymatic activities were 53.2 ± 10% (A24), 73.9 ± 5% (B9) and 165 ± 10% (C 1) (Fig. 3). The percentage of defective proteases (less than a 1% of the activity) was 39%, 29% and 31% for the A, B and C quaspecies, respectively. Consequently, 67% of all analyzed proteases displayed a detectable enzymatic activity (Fig. 3). Nonfunctional proteases with mutations located at the catalytic triad (H57, D81 and S139) (clones A71, B28, B108 and C7) or at the zinc-binding site (C97, C99, C145 and H149) (clones A22, A42, B6, B20, B42 and C138) (Fig. 2, 3) were found, demonstrating the specificity of the genetic screen employed here to measure the catalytic efficiency of the different protease variants. Remarkably, 42% and 37% of the protease variants from quaspecies B and C, respectively, showed a similar or higher catalytic efficiency than the master protease. Of note, several of the high fitness minority variants, mainly in the C quaspecies, had only one substitution when compared to the master sequence. Finally, the 3 proteases bearing substitutions involved in the acquisition of resistance to current PIs (B19, B54 and C30) showed a catalytic efficiency similar to that displayed by the master protease (Fig. 3).

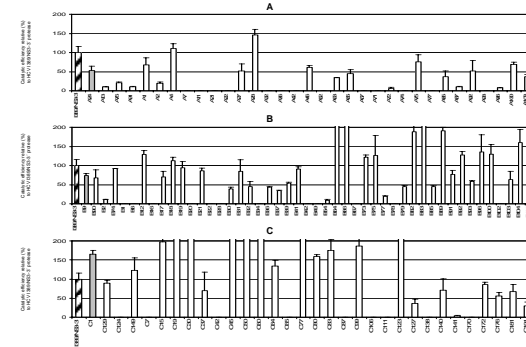


Figure 3. Comparative growth of phages containing different HCV NS3/4 protease single variants targeting the NS5A/NS5B cleavage site. The growth of phages encoding a single protease variant (white bars) was compared to the growth of wild-type 1389/NS3-3 protease (100%) (dashed bar). The gray bar indicates the master protease.

To investigate why high efficiency minority variants were so abundant in quaspecies C, we decided to analyze the catalytic efficiency of the entire C quaspecies with a different target cleavage site. We chose the NS4B/NS5A cleavage site, which is also targeted *in trans* by the NS3/4 protease (Fig. 4).

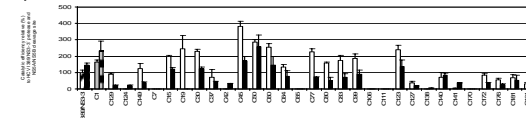


Figure 4. Comparative growth of phages containing quaspecies C HCV NS3/4 protease single variants targeting the NS4B/NS5A cleavage site. The growth of phages encoding a single protease variant (white and black bars) was compared to the growth of wild-type 1389/NS3-3 protease (100%) (dashed bar). The white bars represent phages targeting the NS5A/NS5B cleavage site. The black bars represent phages targeting the NS4B/NS5A cleavage site. The gray and dotted bars indicate the quaspecies C master protease (C1) targeting the NS5A/NS5B and NS4B/NS5A cleavage sites, respectively.

In contrast to the results obtained with the C1 master protease, 85% of quaspecies C minority variants had a lower catalytic efficiency with the NS4B/NS5A cleavage site than with the NS5A/NS5B cleavage site (Fig. 4). Overall, the variants that showed a higher catalytic efficiency than the master protease in front of the NS5A/NS5B cleavage site had a low efficiency when tested with the NS4B/NS5A cleavage site. Remarkably, the C19 variant, that displayed more than 200 % of the activity of the C1 master protease with the NS5A/NS5B cleavage site, showed no activity at all with the NS4B/NS5A cleavage site. Only variant C50 displayed, with both cleavage sites, a better catalytic efficiency than the master protease, suggesting that the viral complexity in this genomic region may not exclusively depend on the enzyme catalytic efficiency. Overall, the results obtained after analyzing this second cleavage site demonstrated that the master proteases could be a quaspecies optimum capable of processing with high efficiency their different target cleavage sites.

## Conclusion

The high resolution of the quaspecies shown here reflects the ability of HCV to explore a huge range of NS3/4 protease genetic configurations. Furthermore, the phenotypic resolution of these quaspecies demonstrates that most of the former genetic configurations had a detectable enzymatic activity, including mutants having substitution involved in the acquisition of resistance to current NS3/4A protease inhibitors. Nevertheless, a different genetic diversification and distribution of catalytic efficiencies was detected in the 3 quaspecies, suggesting that different selective forces are acting in different infected individuals. The genetic variability found here is also remarkable since the effective cleavage of TRIF and Cardif is required for optimal viral replication. The finding that minority variants had a very different catalytic efficiency depending of the target cleavage site tested suggests that some minority variants might be selected on the basis of their better efficiency in cleaving TRIF or Cardif. It remains to be elucidated whether differences in the NS3/4 protease catalytic efficiency can be related to virus fitness, virulence or pathogenicity.

## References

Martinez MA, Clotet B. (2003). A genetic screen to monitor the activity of the Hepatitis C Virus NS3 serine protease. *Antimicrobial Agents and Chemotherapy*, 47, 1760-1765.  
Franco S, Parera M, Aparicio e, Clotet B and Martínez MA. (2007). Genetic and catalytic efficiency structure of HCV NS3/4 protease quaspecies. *Hepatology*, in press.