A Complete Mutational Fitness Map of the Hepatitis C Virus Nonstructural 3 Protease: Relation to Recognition by Cytotoxic T Lymphocytes

Jonas Söderholm and Matti Sällberg
Division of Clinical Virology, Karolinska Institutet at Karolinska University Hospital Huddinge, Stockholm, Sweden

The hepatitis C virus nonstructural (NS) 3/4A protease sequence is highly conserved for reasons not fully understood. We determined the protease activity in 181 NS3/4A gene products in which each protease residue was replaced by alanine or glycine. Unexpectedly, most (87%) protease residues could be replaced and protease activity would be retained. Using these data, we were able to identify a human leukocyte antigen A2–restricted epitope in which substitutions at 5 of 9 residues destroyed the protease. The NS3 protease shows an unexpectedly high plasticity, and it is therefore important to identify target sequences in which the appearance of mutations is restricted by viral fitness.

A characteristic of the hepatitis C virus (HCV) is its high degree of genetic variability. This has been proposed as one factor contributing to the high rate of chronic HCV infections, because of the introduction of escape mutations into immunological epitopes [1, 2]. However, the genetic variability is not evenly distributed within the viral genome. An additional factor that may contribute to viral persistence is that the nonstructural (NS) 3 protein may directly inhibit the antiviral response of the host cell in vitro [3] and in vivo [4]. Some regions are quite well conserved, such as the protease domain of the NS3 protein. One likely explanation for this sequence conservation is that many mutations may affect the fitness of the viral protease and, hence, the rate of viral replication. We have recently shown that this is partly true by identifying residues within a highly conserved human leukocyte antigen (HLA) A2 epitope that were essential for both protease fitness as well as for binding to the HLA-A2 molecule [5]. Thus, immune escape at these positions was effectively prevented by impairing viral replication [5]. However, this only revealed a small part of the complete picture. The true plasticity of the NS3 protease is not known, and, on the basis of the high sequence conservation of this region, one would predict that possibly a majority of mutations negatively affect the protease activity. We, therefore, performed a complete mutational fitness analysis of the protease domain of the HCV NS3 protein and found an unexpectedly high plasticity of the NS3 protease. Defining the residues in which mutations reduce viral fitness, in particular within epitopes recognized by cytotoxic T lymphocytes (CTLs), is of key importance in both drug and vaccine development.

Materials and methods. The full-length NS3/4A gene encoding a functional protease has been described elsewhere [6]. A total of 181 mutants of the protease domain covering aa 1027–1207 of the original NS3/4A sequence were sequentially replaced by alanine (Ala) or glycine (Gly) by the use of in vitro mutagenesis, in accordance with the manufacturer’s protocol (GeneTailor Site-Directed Mutagenesis; Invitrogen). The mutant sequences were all confirmed by sequencing.

Peptides were produced corresponding to HLA-A2 epitopes of the cytomegalovirus pp65 protein [5] (sequence NLVPMVATV) or the HCV NS3 HLA-A2 epitopes GLLGCIITSL, YLVRHADV, and LLCPAGHAV [7, 8], by use of an automated synthesizer [9]. The NS3/4A protease activity was determined by in vitro translation (TNT; Promega), as described elsewhere [5, 6]. A functional in vitro–translated NS3 protease generates 2 protein bands on an SDS-PAGE gel, a shorter liberated NS3 protein, and a longer NS3-NS4A fusion protein [5, 6]. Because a nonfunctional protease generates only the NS3-NS4A fusion protein band, the assay can be effectively used to screen mutations that affect protease activity. Importantly, most mutations identified by this approach also affect the replication of viral RNA [5].

HLA-A2–restricted RMAS cells, expressing the human β2-microglobulin and HLA-A2 chains combined with the α3 transmembrane cytoplasmic domain from the mouse (RMAS-HHD; gift of F. Lemonnier, Institut Necker, Paris, France), were maintained as described elsewhere [5]. The ability of NS3/4A-derived HLA-A2 peptides to bind to HHD molecules was determined by a flow cytometry–based RMAS-HHD stabilization assay [5,
Results. We sequentially mutated the 180 protease-domain residues to Ala or Gly in a functional full-length NS3/4A gene, and all gene products were then tested for protease activity by the presence of cleavage at the NS3-NS4A junction by use of an in vitro translation [5]. We have previously shown that the protease activity as determined by the in vitro translation assay correlates well (albeit not to 100%) with the ability of the NS3 protease to cleave the precursor polyprotein and to support the replication of viral RNA in a replicon-based system [5].

Ala or Gly substitutions at 23 (13%) of 181 tested protease domain residues resulted in a nonfunctional protease as determined by the absence of a free NS3 band (figure 1, red boxes). Ala substitutions at 5 residues resulted in enhanced protease activity, as evidenced by a complete, or almost complete, liberation of NS3 (figure 1, blue boxes). Thus, most (87%) residues of the NS3/4A protease accept Ala/Gly substitutions while retaining fitness of the NS3 protease, suggesting that the protease has an unexpectedly high plasticity. Moreover, as would be predicted, we found that the residues previously identified as being involved in the zinc finger or the catalytic triad could not be substituted by Ala, confirming that the technique used correctly identified these residues as essential (figures 1 and 2). Also, one of the previously identified residues responsible for the development of resistance against protease inhibitors was found not to accept an Ala substitution (figure 2).

We compared the location of the residues essential for protease fitness in vitro with the natural variation of the NS3 protease domain in 148 HCV genotype 1a strains. In these sequences, no polymorphisms were observed at 111 (61%) of the 181 protease-domain residues, whereas 70 (39%) were polymorphic. Seventeen residues (74%) essential for protease fitness were located in nonpolymorphic sites (figure 2), and 6 (26%) were polymorphic, showing no significant association between conserved residues and protease fitness. Thus, other factors, such as interactions with the host cell and other viral proteins, are likely contributors to the conservation of the NS3 protease sequence.

We have recently shown that immunological escape within an HLA-A2–restricted epitope was prevented by the fitness of the NS3 protease [5]. We, therefore, characterized 3 additional HLA-A2–restricted protease CTL epitopes, GLLGCIITSL, YLV-TRHADV, and LLCPAGHAV. No residues essential for protease activity were located in the first epitope at residues 1038–1047. Adding to a previous report [8], we found that 3 natural variants of the epitope bound to HLA-A2 and were recognized by CTLs primed by genetic immunization of HLA-A2 Tg mice (figure 2). Thus, the protease does not seem to limit the variability of this epitope. As reported previously, 2 residues essential for protease activity were located in the immunodominant HLA-A2–restricted epitope at residues 1073–1081 [5] (figure 2). Epitope positions 2, 7, and 9 were essential for binding to the HLA-A2 molecules in the 1131–1139 and 1169–1177 epitopes (figure 2), which is consistent with findings regarding the 1073–1081 epitope [5]. Interestingly, these 2 epitopes differed with respect to the presence of residues essential for protease activity (figure 2). In the seemingly low-avidity HLA-A2–binding 1131–1139 epitope, no residues were essential for protease activity, and 4 positions were polymorphic (figure 2). Thus, immune escape may also occur within this epitope without affecting protease fitness. In the 1169–1177 epitope, 5 residues at epitope positions 1, 2, 3, 6, and 7 were essential for the fitness of the viral protease (figure 2). Of these, positions 2, 6, and 7 also interacted with the HLA-A2 molecule (figure 2). Finally, 3 residues were polymorphic, of which 1 was essential for protease activity (figure 2).

Discussion. The sequence of the HCV NS3 protein is highly conserved, and the reason for this is not fully understood. The host immune response to NS3 is considered to be pivotal for controlling the infection. It has been assumed that a key factor restricting the variability of the NS3 protease is that many mutations may impair the viral protease and, hence, also viral fitness. It was recently shown that this may be true for 2 types of viral escape, immune escape within a CTL epitope [5] and escape from a protease inhibitor [12]. With respect to escape from the CTL response, we have recently shown that residues highly conserved in wild HCV strains were often those that could not be mutated with a retained protease activity [5]. With respect to escape from viral protease inhibitors, it was found that such resistant viruses may have reduced viral fitness [12]. The most important resistance mutations reported to date for genotype 1b replicons seem to be positions 1135 (or 109), 1179 (or 153), 1182 (or 156), and 1194 (or 168) (figure 2) [12]. Taken together, the cost of mutations to viral fitness would help to explain the conserved nature of the NS3 protease. However, because the complete plasticity of the NS3 protease has not been determined, it is impossible to estimate the contribution of protease fitness in maintaining the conservation of the NS3 protease domain sequence. We have now determined the plasticity of the NS3 protease domain by a rapid in vitro assay. The effects of 181 Ala or Gly substitutions in a functional genotype 1a NS3/4A protease on cleavage at the NS3 and NS4A junction were tested by an in vitro translation assay [5, 6]. With this approach, we found that Ala or Gly could substitute 87% of the 181 protease residues. This suggests that the plasticity of the NS3 protease domain may have been underestimated and implies that additional factors are needed to explain the con-
Figure 1. Fitness of the nonstructural (NS) 3 protease determined, by use of an in vitro transcription and translation assay with a wild-type (wt) NS3/4A gene or with 181 NS3/4A genes in which alanine or glycine was substituted for each protease-domain residue. The mutant genes were tested for protease activity after translation by use of an in vitro transcription and translation assay (TNT; Invitrogen). The upper band corresponds to the noncleaved NS3/4A fusion protein, and the lower band corresponds to the free NS3 protein. A single NS3/4A-fusion band represents a nonfunctional protease (red boxes), and a single NS3 band represents a protease sequence with enhanced protease activity (blue boxes). Developed films were photographed using a GeneGenius gel documentation camera (SynGene) and were analyzed using GeneTools (SynGene). The intensity plot of each lane is shown below each lane. A single or clearly dominant peak indicates destroyed or enhanced protease activity, compared with the dual peak appearance of the wt NS3/4A gene.

served nature of the NS3 protease domain, such as interactions with host cell factors or with other HCV proteins. Also, it is important to note that our current approach has its limitations in that it determines only NS3-NS4A cleavage and not the cleavage of other viral targets or host cell sequences [13]. Ala or Gly substitutions may not reflect events occurring in nature, and the substitutions may be too conservative or too radical. Regardless, our data provide a road map identifying residues that are of particular interest, such as those in CTL epitopes or those with a yet-unknown function. Such residues need to be further analyzed with respect to better-targeted substitutions.

The complete fitness map of the NS3 protease domain allowed us to relate the residues essential for protease activity to key elements involved in the protease as well as to immune recognition. The in vitro translation assay correctly pinpointed the residues involved in the zinc-finger domain as well as in

\[ \text{Figure 1. Fitness of the nonstructural (NS) 3 protease determined, by use of an in vitro transcription and translation assay with a wild-type (wt) NS3/4A gene or with 181 NS3/4A genes in which alanine or glycine was substituted for each protease-domain residue. The mutant genes were tested for protease activity after translation by use of an in vitro transcription and translation assay (TNT; Invitrogen). The upper band corresponds to the noncleaved NS3/4A fusion protein, and the lower band corresponds to the free NS3 protein. A single NS3/4A-fusion band represents a nonfunctional protease (red boxes), and a single NS3 band represents a protease sequence with enhanced protease activity (blue boxes). Developed films were photographed using a GeneGenius gel documentation camera (SynGene) and were analyzed using GeneTools (SynGene). The intensity plot of each lane is shown below each lane. A single or clearly dominant peak indicates destroyed or enhanced protease activity, compared with the dual peak appearance of the wt NS3/4A gene.} \]
Figure 2. Presence of synonymous (syn) and nonsynonymous (nonsyn) substitutions after alignment of 148 genotype 1a nonstructural (NS) 3 protease sequences. Analysis was done by use of SNAP software (available at: http://hiv-web.lanl.gov/content/hiv-db/SNAP/WEBSNAP/SNAP.html; version 9/15/98). A total of 124 genes were from GenBank (accession nos. AF009606, AF011751–AF011753, AF271632, AF290978, AF369214, AF369217–AF369222, AF369224–AF369226, AF369228–AF369235, AF369237, AF369239–AF369240, AF369242–AF369245, AF511950, AJ278830, AY588701–AY588718, AY588736–AY588753, AY588814–AY588828, AY588849–AY588881, AY615798, AY695436–AY695437, D10749, M32084, M62321, and M67463), and 24 genes had been cloned and sequenced in-house [5]. Where polymorphisms exist, the most commonly found amino acid is the upper amino acid on the X-axis label. Red boxes mark mutations that destroy protease activity, and blue boxes mark mutations that enhance protease activity. The catalytic triad, zinc finger [11], and mutations mediating resistance to protease inhibitors [12] are also indicated in the graph. Also shown is the immunological characterization of 3 human leukocyte antigen (HLA) A2–restricted cytotoxic T lymphocyte (CTL) epitopes located in the protease domain (B–E). The binding of the 1038–1047 NS3 peptide (at 50 μmol/L) to HHD-RMAS cells is shown (B), and values are the median fluorescence. The recognition of the same epitope by CTLs from HLA-A2 Tg mice genetically immunized by NS3/4A are shown (C), and data are the percentage of specific lysis. The legend and the peptide sequences used for the symbols in panel C can be found in the Y-axis labeling of panel B. The analysis of the binding of naturally occurring and Ala/Gly-substituted NS3-derived HLA-A2–restricted peptides 1131–1139 (D) and 1169–1177 (E) to HLA-A2–transfected RMAS cells is also shown. Values are the lowest peptide dilution inducing positive fluorescence. Low peptide concentrations indicate a high HLA-A2–binding avidity, and vice versa. CMV, cytomegalovirus; HCV, hepatitis C virus.
the catalytic triad as essential for the protease, supporting the accuracy of the screening method. With respect to the host response, we know that these NS3-specific CTLs can enter the liver and eliminate hepatocytes expressing HCV antigens [14]. We have now characterized 3 HLA-A2–restricted CTL epitopes and determined whether potential escape mutations within these epitopes could be introduced while retaining protease activity. Two epitopes contained several polymorphic residues found in wild-type NS3 sequences and did not contain any residue essential for protease activity. However, in the 1169–1177 epitope, Ala/Gly could not be substituted for a majority of the residues without destroying the protease. Thus, HLA-A2–restricted epitopes differ greatly in how mutations affect both immune recognition and the fitness of the viral protease.

Our present findings have several important implications. We provide a complete mutational map of the in vitro fitness of the HCV NS3 protease domain, which will aid in the understanding of the plasticity of the HCV genome. Our newly identified residues essential for protease activity should be tested in the context of the newly developed HCV clones that can productively infect HuH-7 cells [15]. Following the mapping of protease fitness, we could identify CTL epitopes in which immune escape are unlikely to occur, because many mutations have negative effects on the fitness of the viral protease. This will certainly be of help in designing epitope-based vaccines, where the included epitopes preferably should be those in which the virus is unable to mutate due to viral fitness.

References