Sequence analysis of the NS5A protein of European hepatitis C virus 1b isolates and relation to interferon sensitivity

Gilles Duverlie,1 Hafida Khorsi,1 Sandrine Castelain,1 Olivier Jaillon,1 Jacques Izopet,2 Françoise Lunel,3 François Eb,1 François Penin4 and Czeslaw Wychowski5

1Virologie, CHU-Hôpital Sud 80054 Amiens cedex 1, France
2Virologie, CHU-Hôpital Purpan, 31059 Toulouse cedex, France
3Virologie, CHU-Hôpital d’Angers, 4 rue Larrey, 49033 Angers cedex, France
4Laboratoire de conformation des protéines, UPR 412-CNRS, Institut de Biologie et Chimie des Protéines, 7 passage du vercors, 69367 Lyon cedex 07, France
5Equipe Hepatite C, CNRS-UMR 319, Institut de Biologie de Lille, BP 447, 59021 Lille cedex, France

Japanese studies have defined the discrete 2209–2248 amino acid region of the non-structural 5A protein (NS5A2209–2248) of hepatitis C virus genotype 1b (HCV 1b) isolates as the interferon sensitivity determining region (ISDR). European studies did not confirm these results since most of the ISDR sequences harboured an intermediate profile. Recently, a direct interaction between the NS5A protein, involving the ISDR, and the interferon-induced protein kinase (PKR) has been reported and presented as a possible explanation of HCV interferon resistance. In the present study, the entire NS5A amino acid sequence from 11 resistant and eight sensitive strains from European HCV 1b isolates was inferred from direct sequencing. The previously described important amino acid stretches and positions in NS5A were compared between the resistant and sensitive groups. Although some variations were observed, no clear differences could be directly correlated with the interferon sensitivity. However, sensitive strains were different, owing to more amino acid changes when compared to a consensus sequence from all strains. The carboxy-terminal region and especially the previously reported NS5A/V3 region showed most of the variations. Moreover, the conformational analysis of NS5A by secondary structure prediction allowed the differentiation of most sensitive strains from resistant ones. It was concluded that other regions different from ISDR were involved in resistance to interferon maybe via the interaction between NS5A and PKR.

Introduction

Hepatitis C virus (HCV) is a major causative agent of chronic hepatitis that can progress to cirrhosis and hepatocellular carcinoma over several decades (Tong et al., 1995). Its single-stranded, positive-sense RNA genome encodes a polyprotein of about 3000 residues, processed into structural and non-structural proteins (Hijikata et al., 1991; Grakoui et al., 1993; reviewed by Major & Feinstone, 1997). So far, interferon administration remains the most successful treatment for HCV infection, although it is of limited long term efficacy.

Virological parameters, such as genotype and virus load are related to the response rate (Martinot-Peignoux et al., 1995; reviewed by Hoofnagle & DiBisceglie, 1997). Enomoto et al. (1995, 1996) have studied the quasispecies selection of HCV during interferon therapy. By direct comparative analysis of full-length sequences of HCV 1b isolates before and during interferon treatment, they identified clusters of amino acid differences in the NS5A gene, especially in a 40 amino acid stretch from codon 2209–2248, which was termed the interferon sensitivity determining region (ISDR). Chronically infected individuals, whose ISDR sequences were identical to that of HCV-J (wild-type), were non-responders, while those harbouring more than three amino acid differences (mutant strains) were complete responders to interferon therapy. Most of the strains with intermediate differences (1–3 amino acids) were found in non-responder patients. Other ISDR sequence comparison studies have been reported by Japanese and

0001-5367 © 1998 SGM
European groups (Chayama et al., 1997; Khorsi et al., 1997; Kurosaki et al., 1997; Squadrito et al., 1997; Zeuzem et al., 1997; summarized by Herion & Hoofnagle, 1997). Japanese groups confirmed the correlation between the number of differences in ISDR and sensitivity to interferon, but European studies did not. In the latter studies, most of the complete responders harboured intermediate strains with one to three changes. A North American study of the ISDR of HCV genotype 1a also found no relation between the amino acid sequences and the response to interferon (Hofgartner et al., 1997). The NS5A protein is variable in length and sequence according to the different genotypes: that of genotype 1b is 447 amino acid long (residues 1973–2419 of the polyprotein), whereas insertions of 1, 18 and 4 residues are present in the NS5A of types 1a, 2 and 3, respectively (reviewed by Maertens & Stuyver, 1997). Numerous differences could be observed between the ISDR sequences of the other genotypes, but correlations with the interferon sensitivity remain to be discovered (Chamberlain et al., 1997).

Recently, Gale et al. (1997) demonstrated that NS5A of genotypes 1a and 1b could be a potent inhibitor of the double-stranded RNA-dependent protein kinase (PKR), one of the most important antiviral proteins induced by alpha-interferon. NS5A–PKR interaction studies using deletion mutants showed that the ISDR was probably required for interaction with, and repression of, PKR. These data provided the first evidence for the molecular mechanisms underlying HCV resistance to interferon therapy, but did not explain why ISDR sequences in European strains did not allow the discrimination between sensitive and resistant strains. Thus it appeared important to compare other regions of NS5A. In this study, we have directly sequenced the entire NS5A gene of European HCV 1b isolates from eight complete responder and 11 non-responder patients. Comparative analyses of their amino acid sequences and conformations suggested that other region(s) of the NS5A protein may also be involved in the sensitivity to interferon, at least in European strains, which frequently had an intermediate profile.

Methods

**Samples.** Sera were collected from 19 chronically HCV 1b-infected patients eligible for interferon treatment, before therapy in eight complete responder and during therapy in 11 non-responder patients. Complete responders were defined as PCR HCV-negative and had normalization of transaminases at least 6 months after the end of treatment. Non-responders showed elevated transaminases and were PCR HCV-positive at the end of the treatment (Khorsi et al., 1997). Genotyping and ISDR sequencing were performed as already described (Castelain et al., 1997; Khorsi et al., 1997). In complete responder patients, we hypothesized that the master sequence, if not all the clones of the HCV quasispecies, corresponded to interferon-sensitive HCV isolates. Conversely, the HCV quasispecies master sequence of non-responder patients after 3–4 months of interferon therapy must correspond to resistant strains.

**NS5A amplification.** The entire NS5A region was amplified by nested RT–PCR. HCV RNA was prepared from 140 µl of each thawed serum using a silica gel extraction method (QiaAmp HCV, Qiagen). RT was combined with the first round of PCR. Ten µl out of 50 µl of isolated RNA was added to 90 µl of master mix containing 320 µM of each dNTP, 50 U of human placental ribonuclease inhibitor (Promega), 0.5 µM of each outer primer, 1 mM dithiothreitol, 1.5 mM MgCl₂, 9 U of avian myeloblastosis virus reverse transcriptase (AMV), 2.5 U Tag DNA polymerase and 1 × Taq buffer (Perkin Elmer Cetus). RT–PCR was performed in one step for 25 min at 42 °C and 15 min at 50 °C, followed by 30 cycles of amplification, each consisting of 90 s at 94 °C, 90 s at 50 °C and 5 min at 72 °C. For the second amplification, 5 µl of the first reaction mixture was removed and further amplified with an inner primer set over 35 cycles after 2 min at 94 °C, each cycle consisting of 1 min at 94 °C, 1 min at 55 °C and 1 min at 72 °C and completed by an extension of 7 min at 72 °C. All primers were designed on the basis of HCV 1b complete genome sequences available in the DDBJ/GenBank/EMBL nucleotide databases. Outer primers were E1 (5′ GAGGGGCTGTG-CAGTGGATG 3′, nucleotides 6057–6077) and E2 (5′ GCTGGCAG-ATGGTGGCACA 3′, nucleotides 7698–7718). Inner primers were I3 (5′ TCCGCTCTGCTGCTAAGGGAA 3′, nucleotides 6246–6265) and I4 (5′ GACAGAGCACATCTCA 3′, nucleotides 7567–7586). Nucleotide positions are numbered according to the HCV-1 sequence and the NS5A gene is within positions 6246–7586. PCR products were then subjected to electrophoresis in a 1% low melting point agarose gel (Nusieve GTG, FMC) and visualized by ethidium bromide staining before purification.

**Direct sequencing of NS5A.** PCR fragments were cut out from the agarose gel and purified using a minicolumn system (Wizard PCR Prep DNA purification system, Promega). Cycle sequencing was performed using the ABI PRISM Dye terminator ready reaction kit following the manufacturer’s instructions (Perkin Elmer Cetus). Inner PCR primers were used as sequencing primers and also sense primers (S277, 5′ CCCATCAACGCCTACACCC 3′; S534, 5′ ATTCAGGTGCGGCCCTCA-AACA 3′; S775, 5′ ATCGAGGCAACCTCCTTG 3′; S1033, 5′ GGTGCCCCDTYTGCCRCT 3′) and antisense primers (AS306, 5′ GACAGGGGCGCCTGGTGTA 3′; AS504, 5′ CGCCGGAGGCTGTTCTG 3′; AS816, 5′ TGTCCGGCCATCTCCT 3′; AS1174, 5′ CGGCHGAYGACYCRAGC 3′) to obtain the entire nucleotide sequences of the two strands. Sequencing reactions were prepared according to the manufacturer’s instructions before automatic electrophoresis of the sequencing products by an ABI PRISM 310 Genetic Analyser (Perkin Elmer Cetus). Extraction, amplification and sequencing were repeated for sera corresponding to sequences S1, S3, S8 and R5 from three complete responders and one non-responder, respectively.

**NS5A sequence analysis.** Electrophoretograms were interpreted using the Sequence Navigator software (Perkin Elmer Cetus). Multiple amino acid sequence alignment, consensus and ambiguity sequences were carried out with CLUSTAL W1.6 software (Thompson et al., 1994). A consensus phylogenetic tree was constructed with the programs Seqboot (Bootstrap), Protdist (Dayhoff PAM matrix) and Neighbour (Neighbour-joining) of the PHYLIP package version 3.5c (Felsenstein, 1993). The robustness of the trees was evaluated by bootstrap resampling (1000 data sets) of the multiple sequence alignment. The topology of the consensus tree and the bootstrap values (percentages) were calculated with the Consense program. Variability of the different sequences was expressed by the number of amino acid changes compared with the consensus sequence to the simple majority for each position of all strains. Numbers of different amino acids at each position were also compared between sensitive and resistant strains. Statistical analyses were performed using Student’s t-test to compare these values. Results were also compared to those published by Enomoto et al. (1995) for the sequences of the NS5A region at position 2141–2419.
Potential phosphorylation sites were localized using the typical sequence signatures reported in the Prosite dictionary (Bairoch, 1993). The secondary structure of proteins was predicted with the double prediction method (Deleage & Roux, 1987), which is a combination of protein class predictions according to Nakashima et al. (1986) and the Chou & Fasman method (1978) which allows the prediction of amino acid states. These methods are implemented in the ANTHEPROT software (Geourjon & Deleage, 1995) and predictions were made using the IBCP web facilities (http://www.ibcp.fr). Multiple alignment of predicted secondary structures was carried out, an identity comparison matrix was calculated, and a tree of NS5A \( 2411-2419 \) conformation similarities drawn with njplot.

Results

The eight interferon-sensitive and the 11 interferon-resistant strains were named S1–S8 and R1–R11, respectively, and their NS5A nucleotide and deduced amino acid sequences have been deposited in the GenBank/DDBJ/EMBL database. An identical amino acid sequence was obtained after another extraction, amplification and sequencing analysis of isolates from sera corresponding to S1, S3, S8 and R5. The consensus sequence of all strains was identical to HCV-J except for 19 positions, i.e. the 11 reported in Fig. 2, and 1978 (arginine), 1989 (threonine), 2006 (valine), 2009 (phenylalanine), 2136 (alanine), 2146 (threonine), 2410 (glutamic acid) and 2413 (serine). The overall comparison of the amino acid variability for all strains and for the eight interferon-sensitive and the 11 interferon-resistant NS5A sequences is presented (Fig. 1A–C) and compared to that observed for the sequences of the NS5A region 2141–2419 published by Enomoto et al. (1995) (Fig. 1D–F). The amino acid changes along each sequence were identified by comparison with the consensus sequence and the total number of changes at each position was reported as a percentage (Fig. 1A, D). The probability of amino acid changes between interferon-sensitive and -resistant strains was calculated by using Student’s \( t \)-test for overlapping windows of 7 amino acids (Fig. 1B, E). Such a window size appeared to be the most relevant among all tested window sizes to identify the most variable regions. Indeed, the use of smaller windows (< 5) did not allow a simple straight identification of variable regions, while probability calculation appeared to be biased for larger windows (> 9) when multiple amino acid changes were observed in the window but involving only one or two sequences over all the sequences. This probability analysis allowed the identification of the most relevant variable regions between sensitive and resistant strains. The differential changes frequency between sensitive and resistant strains was calculated using overlapping windows of 7 amino acids and plotted for each sequence position (Fig. 1C, F). Consequently, a higher sequence variability in sensitive strains than in resistant strains corresponded to a positive peak and vice versa. Finally, the positions of the characteristic regions of the NS5A gene are also shown at the top of each panel in Fig. 1. The combination of all these representations allowed a simple reading of sequence variability between sensitive and resistant strains.

Fig. 1. Amino acid variability in sensitive and resistant strains in the NS5A sequence. (A–C) European strains, this study; (D–F) NS5A \( 2141-2419 \) region of Japanese strains published by Enomoto et al. (1995). NS5A sequences are numbered according to HCV-J. (A, D) Total number of amino acid changes at each position relative to the consensus sequence of all strains of the present study (see text). (B, E) Probability of amino acid changes between interferon-sensitive and -resistant strains calculated using Student’s \( t \)-test for overlapping windows of 7 amino acids. The reported probability at each position is that of the central amino acid of the window. The value \( P = 0.05 \) is represented by a dotted line. (C, F) Differential changes frequency between sensitive and resistant strains (S–R) calculated for each position according to the equation: \( S - R = \Sigma \frac{CS_i}{CR_i/N_i} - \frac{CS_i/N_i}{\Sigma CS_i} \), where \( \Sigma CS_i \) is the number of changes (CS) in the sensitive strains, \( \Sigma CR_i \) is the number of changes (CR) in the resistant strains, \( N_i \) is the number of sensitive strain sequences, and \( N_i \) is the number of resistant strain sequences. Reported S–R values at each position were calculated using an overlapping window of 7 amino acids. The open boxes on the solid lines above the panels indicate potential functional regions of the NS5A gene, i.e. region involved in the association with NS4A (a, 2135–2139), potential cleavage site (c, 2167–2172), hyperphosphorylated region (HP, 2197–2204), interferon sensitivity determining region (ISDR, 2209–2248), nuclear localization signal (NLS, 2326–2334), NS5A/V3 region (V3, 2356–2379).

Variability at different regions and positions

The mean and standard deviation values of different amino acids for each position were also calculated. They were \( 1.273 \pm 0.541 \) for the eight sensitive and \( 1.213 \pm 0.489 \) for the 11 resistant strains \( (P = 0.02) \). The variability compared with the consensus sequence was analysed along the NS5A gene for these two groups. For the sequences described in this study,
sensitive strains were more variable than resistant strains ($P = 0.04$) for the entire carboxy-terminal region of the NS5A sequence, i.e. NS5A_2194–2293 ($P \leq 0.05$). Few stretches showed significant differences between sensitive and resistant strains (Fig. 1). Some amino acids were variable from position 2136 to 2154. At position 2143, glutamic acid was present in all resistant strains instead of aspartic acid in five out of eight sensitive strains. Significant variations were also present around the ISDR since mutant strains S1 and S8 had four and eight changes in this region, respectively. The most important
changes were localized in the variable 2356–2379, previously described as V3 by Inchauspe et al. (1991). For the Japanese isolates, the same observations with higher variations ($P < 0.01$) could be made in the carboxy-terminal region of NS5A. Indeed, as described by Enomoto et al. (1995), ISDR showed most of the variations. Some significant changes were also present around the nuclear localization signal, the V3 region and at the end of NS5A. At position 2413, glycine was present in eight of the nine sensitive strains and in four of the eight sensitive strains in our study. The consensus phylogenetic tree using amino acid sequences presented in this study was constructed for the carboxy-terminal region 2141–2419 (see Fig. 3A). Discrimination of the sensitive isolates was only possible for strains S1, S8, S5 and S3 with a very low bootstrap value, except for the mutant strains S1 and S8 (68%).

### Conformational analyses of NS5A sequences

To highlight the impact of amino acid differences in the local structure along the sequence, a conformational analysis was made by means of the secondary structure prediction for each strain (Fig. 2). The data were obtained using the double prediction method (Deléage & Roux, 1987), which considers four possible conformations for each residue: helix (h), β-sheet (e for ‘extended’), turn (t) and coil (c). It should be noted that the conformation prediction for each amino acid takes into account the nature of amino acids in the local sequence (typically 3 amino acids before and after the considered amino acid). Thus, an isolated amino acid difference may modify the predicted conformation of neighbouring amino acids, but not necessarily its own conformation. The variation of predicted conformation for each residue of each strain was reported relative to the conformation of the consensus sequence. Detailed analysis was thus performed in order to identify differences between sensitive and resistant strains. Inspection of the N-terminal part of NS5A (1973–2196) did not reveal particular features (data not shown), and the main interesting features lie in the carboxy-terminal part of NS5A sequences that are reported (Fig. 2). Among the 122 amino acid substitutions observed for resistant strains in the 2141–2419 region, 51% were conservative. In contrast, sensitive strains appeared more variable with only 41% conservation among 143 substitutions. It should be noted that most of the conservative substitutions concerned hydrophobic–hydrophobic amino acid substitution, and that almost no important changes involving hydrophobic amino acids were observed. As these amino acids generally constitute the hydrophobic core of the protein, it can be concluded that both sensitive and resistant NS5A proteins likely present the same 3D folding. Conversely, both important point changes and clusters of changes exhibited relatively large modifications of predicted conformation. These differences were mostly observed for polar and charged amino acids that are mainly located at the surface of proteins, indicating that these changes may perturbate the interaction of NS5A with other compounds. The comparison of the NS5A predicted conformation seemed to allow a better differentiation of sensitive and resistant strains than the consensus phylogenetic tree, as shown in Fig. 3 (B) (note that conformational comparisons were done using an identity matrix; for convenience, the results are presented as a tree). The mutant strains S1 and S8 belonged to the same family of NS5A conformations as intermediate strains S2, S5, S6 and the sensitive strain S3 with a wild-type ISDR. S4 and S7 NS5A conformation appeared to be rather different from the other sensitive strains. These results suggested that variations in the NS5A carboxy-terminal region might be related to the sensitivity of HCV 1b strains to interferon. Comparison of resistant and sensitive NS5A sequences and conformations reported in Fig. 2 allowed the identification of a marked difference in the V3 region for the GSSESSAV sequence (2353–2360), which appeared well conserved in the resistant strains, while being highly variable in the sensitive strains. This was essentially due to several changes in the sensitive strains, and in particular of glutamic acid (2356), glycine (S2, S3, S6, S8) or aspartic acid (S4, S5) or lysine (S1). Similar analysis of Enomoto’s published sequences...
(1995, 1996) showed important predicted conformational variations for sensitive strains in the ISDR, in the 2255–2258 segment, and in the 2412–2418 segment while resistant strains exhibited poor variations (data not shown). Conversely, these three regions appeared almost invariant in the strains described here. Moreover, residues 2353–2360 that were variable in our sensitive strains appeared variable in both resistant and sensitive Japanese strains. These results clearly indicate large differences between Japanese and French HCV-1b strains which are difficult to reconcile when considering the ISDR as the single region involved in PKR interaction. The variability of our strains in relation to the known features of the NS5A protein was also analysed and the results are presented below.

Phosphorylation sites

The NS5A protein can be phosphorylated and, at least for genotype 1b, hyperphosphorylated in the central region near the ISDR, i.e. serine residues 2197, 2201 and 2204 (Kaneko et al., 1994; Tanji et al., 1995). Between 40 and 44 serine residues could be present in the different sequences at 52 positions, 35 of which were always present, and eight were highly conserved. Serine was only present once in the nine remaining positions, in place of threonine for four cases and proline for three cases. Threonine replaced serine in 55% (11/20) of amino acid changes found in all sequences, and always at positions 1997, 2073 and 2079. Positions described as important for the basal- and hyperphosphorylation were conserved. Two positions seemed less conserved, i.e. position 2278 (Fig. 2) and 2413 (already mentioned). Analysis of potential phosphorylation sites presenting typical sequence signature did not show important or typical variation between resistant and sensitive strains. The putative phosphorylation sites for the CMGC kinase group recently shown to be likely involved in NS5A phosphorylation (Reed et al., 1997) were well conserved. In conclusion, interferon sensitivity did not seem to be related to the presence of identified or potential phosphorylation sites.

Amino acid residues 2135–2139

It was previously shown that two phosphoproteins, a faster-migrating 56 kDa protein and a slower-migrating 58 kDa protein were produced from the NS5A region (Kaneko et al., 1994; Tanji et al., 1995). The region NS5A<sub>2135-2139</sub> (P-A-C-R-P) seemed to be important for NS4A-dependent phosphorylation and production of the 58 kDa protein. The stretch P-A-C-K-P was highly conserved in the sequences presented in this study, except at position 2136 where valine was present three times and glutamic acid once. At position 2138, arginine was present once in a complete responder. As already mentioned, a variable region in sensitive strains could be considered from position 2136 to 2154 (Fig. 1), but its importance remains to be determined.

Cleavage site after serine/threonine at position 2172

A potential cleavage site in NS5A protein has been described after threonine at position 2172 which contains the consensus motif D-X-X-X-T/S (Markland et al., 1997). It was conserved in all sequences, but a single residue at position 2169 inside the motif was somewhat different. Sensitive strains harboured alanine instead of threonine in seven out of eight cases and resistant strains harboured threonine in six out of eleven cases and valine once. Modification of the threonine residue could interfere with the potential cleavage site.

ISDR 2209–2248

Briefly, following the Enomoto criteria, the NS5A<sub>2209-2248</sub> region was identical to HCV-J in five cases (wild-type), of which one was sensitive (S3), intermediate in eight cases with arginine at position 2218 instead of histidine, of which three were sensitive (S5, S6, S7), intermediate with glutamine at position 2218 in one sensitive strain (S2), and intermediate with two changes in two resistant strains and with three changes in one sensitive strain (Fig. 2). Two mutant strains with more than three changes were sensitive (S1, S8). Arginine at position 2218 was the most frequent in resistant strains as well as in the sensitive one, and a sensitive strain could have the HCV-J<sub>2209-2248</sub> sequence. Thus, interferon sensitivity of only two strains (S1 and S8) could be related to differences in the ISDR.

Nuclear localization signal at position 2326–2334

Recently, Ide et al. (1996) demonstrated that the amino acid sequence from position 2326 to 2334 of HCV-H (P-P-R-K-K-R-T-V), also present in HCV 1b (Tanji et al., 1995), can function as a nuclear localization signal (NLS) when fused to the amino terminus of the E. coli β-galactosidase protein. As this sequence may influence the location of the fusion protein, we analysed the different amino acids present in the NLS signal. Our results indicated that the consensus sequence was modified to P-P-R-K-R-T-V-V and some variations were possible at different positions, i.e. P-P-R-R(K)<sub>β</sub>-R(K)<sub>β</sub>-R(K)<sub>β</sub>-T-V(A<sub>j</sub>,V<sub>j</sub>)<sub>j</sub>-V (i<sub>j</sub>), the subscripts indicating the frequency of the amino acid changes (Fig. 2). Only one change was present in six modified sequences, and two changes for one resistant strain with R-K at positions 2330–2331 and for one sensitive strain with K-V at positions 2331–2332. These amino acid changes were not related to interferon sensitivity in our strains. Some changes were also present in Japanese strains (Fig. 1), but they seemed to be conservative (data not shown).

Discussion

Resistance to interferon is one major characteristic of HCV and understanding the molecular mechanisms which are involved in this process is important from a fundamental and therapeutic point of view. The role of the interferon system in
the natural evolution of HCV infection is not clear. However, it seems that at least some quasispecies could already be selected before interferon therapy is begun. HCV has been described as a poor, but prolonged, interferon inducer (Jakschies et al., 1994). This could explain the fact that no change in virus load could be seen in a significant proportion of non-responder patients (reviewed by Hoofnagle & DiBisceglie, 1997). The recent finding of a physical and functional interaction between NS5A and PKR proteins has highlighted a possible mechanism for virus escape from the interferon response (Gale et al., 1997). Enomoto et al. (1995, 1996) has previously shown, by sequence comparison, important changes in NS5A before and after interferon therapy, involving residues 2209–2248, which were defined as the ISDR. Gale et al. (1997) have also shown in a deletion study that ISDR could be important for interaction with PKR. However, recent European studies could not confirm these results since many strains harbouring an intermediate ISDR profile were found in sensitive or resistant strains. In a previous study, we sequenced the ISDR from 43 HCV 1b isolates and only three out of 17 sensitive strains were classified as mutant strains following the Enomoto criteria (Khorsi et al., 1997). These mutant strains have not been reported frequently in Europe (Khorsi et al., 1997; Squadrito et al., 1997; Zeuzem et al., 1997).

To explain the differences observed between the Japanese and European studies, we analysed different amino acid positions or regions along the entire NS5A protein which have been reported as being potentially important. Our observations indicated that there were more amino acid changes in sensitive than in resistant strains when compared to the consensus sequence of all strains. Moreover, independently from the reference sequence, more differences in amino acids could be present along the NS5A sequence in sensitive strains than in resistant ones. This may suggest that the NS5A interaction with PKR is only possible with limited variations in the sequence and structure of NS5A. As in the Japanese strains reported by Enomoto et al. (1995), the carboxy-terminal region of NS5A and the NS5A/V3 region was more variable in sensitive than in resistant strains, even though the ISDR was not included. Recent findings revealed that NS5A was phosphorylated at serine residues in its carboxy-terminal region or additionally phosphorylated at serine residues in the central region of amino acids 2197–2204 (Tanji et al., 1995). The analyses of potential sites of phosphorylation indicated good conservation of the serine residues in resistant and sensitive strains which were perfectly conserved at the positions important for basal phosphorylation or hyperphosphorylation of NS5A. In most of the changes that we observed, the serines were replaced by threonine residues, which could also be phosphorylated. Recently, Reed et al. (1997) reported that a minor fraction of threonine residues were phosphorylated in the NS5A protein of a 1a strain, which could also be phosphorylated in NS5A of a 1b strain according to our sequence data. However, it seems unlikely that differences in serine residues could play a direct role in interferon resistance. Recent studies reported by Shimotohno’s group have shown that amino acids 2135–2139 of NS5A type 1b were important for the interaction with the NS4A protein (Asabe et al., 1997). We found that the amino acids 2135–2139 were changed at some positions in a few sequences, maybe suggesting co-variation between NS5A and NS4A to preserve their interaction. The NS5A protein possesses a nuclear localization-like signal sequence in which some modifications of amino acids have been mentioned. Their importance remains to be determined, since HCV NS5A has never been detected in the nucleus. The cleavage site observed at position 2172 of NS5A in some expression systems (Markland et al., 1997) seemed to be always conserved. This cleavage site has not been reported to be functional in mammalian cells, but it may be of importance as two recent articles have reported that amino-terminus-deleted NS5A 1a or 1b proteins can function as potent transactivators in yeast and human hepatoma cells (Tanimoto et al., 1997; Kato et al., 1997). Overall, no differences concerning the known features of the NS5A protein could be clearly related to interferon sensitivity.

In the absence of a 3D structure for NS5A, it is difficult to evaluate the impact of amino acid changes on its structure and/or reactivity. Indeed, although each amino acid has a propensity to adopt a preferred conformation, almost all of them can adopt any of the four conformations (helix, sheet, turn, coil), and their final conformation in the protein is highly dependent on their local and global environment (i.e. their neighbouring amino acid in the sequence and in the 3D structure, respectively). Thus, the conformational analysis of NS5A made by means of the secondary structure prediction for every strain helped us to reveal local variations in the sequence that are not obvious from the simple observation of amino acid changes. This strategy allowed the differentiation of the variations that presented almost no measurable effect on the predicted conformation from the important differences (at least for local conformation). All strains exhibit a high level of heterogeneity. However, most changes concerning hydrophobic amino acids were conservative, and most changes did not modify the main secondary structure elements, i.e. helices and β-sheets. This indicated that the hydrophobic backbone of NS5A was always conserved and that the overall 3D folding of NS5A was essentially identical for all strains. In contrast, the non-conservative changes concerned essentially polar and charged amino acids that are generally localized at the surface of proteins. Such changes may yield mutant NS5A proteins unable to interact with other compounds such as PKR. The ISDR can play a critical role in mediating the interaction with PKR to inhibit its biological activity (Gale et al., 1997), and differences in the ISDR of S1 and S8 strains can explain the sensitivity of these strains to interferon. In intermediate strains, depending on the nature and localization of mutation in the interaction region, a graduation of binding affinity may be
expected that represents gradual change in sensitivity to interferon. However, the low level of changes observed in the ISDR of the six other strains cannot explain their sensitivity. The sequence and conformation variability of these sensitive strains suggested that other regions of NS5A may be involved in PKR interaction. Our prediction of NS5A conformation seemed to better differentiate between sensitive and resistant strains than the amino acid sequence comparison (Fig. 3). This approach allowed us to identify a marked difference at the beginning of the V3 region previously described by Inchauspe et al. (1991). Particularly, the GSSESSAV motif (2353–2360) appeared very well conserved in the resistant strains but variable in the sensitive strains.

In conclusion, as in the Japanese strains, the carboxy-terminal region of the NS5A 1b protein of European isolates was more variable in sensitive than in resistant strains. Nevertheless, the variability of NS5A 1b was lower in European isolates than in the Japanese isolates. These differences between NS5A proteins of HCV 1b isolates from different countries further complicated interpretation of the published studies (Enomoto et al., 1995; Squadrito et al., 1997). Thus, it was necessary to sequence all of the carboxy-terminal region to differentiate most of the sensitive strains from the resistant strains by conformational analyses. Variations around the V3 region seemed particularly interesting and need further study. Direct sequencing was helpful in defining the master sequence of NS5A of the HCV isolates, since cloning after PCR could introduce artefactual mutations, misleading the conclusions obtained (Smith et al., 1997). However, functional assays studying the NS5A–PKR interactions with the different cloned and expressed NS5A proteins with a very close ISDR are necessary to conclude to the importance of these amino acid changes, even if they could express faint or gradual differences. It remains to determine the importance of the NS5A–PKR interaction in the global resistance of HCV to interferon. The NS5A protein (as other HCV proteins) may also interact with other molecular compounds involved in the complex interferon signal transduction pathway.

The authors thank Dr Jean-Michel Pawlotsky for providing some samples, all the participants of the ‘Réseau Hépatite C de Picardie’ and Catherine Duggan for correction of the manuscript. This work was supported in part by the contract ARC no. 5068, by the Programme Hospitalier de Recherche Clinique (1995) and by the Biobanque de Picardie.

References


signal and subcellular distribution of hepatitis C virus nonstructural protein NS5A. Gene 182, 203–211.


Received 1 December 1997; Accepted 18 February 1998