Published online June 24, 2004

Nucleic Acids Research, 2004, Vol. 32, No. 11 ***e90***
DOI: 10.1093/nar/gnh088

Structural analysis of hepatitis C RNA genome using DNA microarrays

María Martell*, Carlos Briones[1], Aránzazu de Vicente[1], María Piron, Juan I. Esteban, Rafael Esteban, Jaime Guardia and Jordi Gómez

Laboratorio Medicina Interna-Hepatología, Hospital Vall d'Hebron, Barcelona, Spain and [1]Laboratorio de Evolución Molecular, Centro de Astrobiología, (CSIC-INTA), Madrid, Spain

Received March 20, 2004; Revised May 20, 2004; Accepted June 3, 2004

ABSTRACT

Many studies have tried to identify specific nucleotide sequences in the quasispecies of hepatitis C virus (HCV) that determine resistance or sensitivity to interferon (IFN) therapy, unfortunately without conclusive results. Although viral proteins represent the most evident phenotype of the virus, genomic RNA sequences determine secondary and tertiary structures which are also part of the viral phenotype and can be involved in important biological roles. In this work, a method of RNA structure analysis has been developed based on the hybridization of labelled HCV transcripts to microarrays of complementary DNA oligonucleotides. Hybridizations were carried out at non-denaturing conditions, using appropriate temperature and buffer composition to allow binding to the immobilized probes of the RNA transcript without disturbing its secondary/tertiary structural motifs. Oligonucleotides printed onto the microarray covered the entire 5′ non-coding region (5′NCR), the first three-quarters of the core region, the E2–NS2 junction and the first 400 nt of the NS3 region. We document the use of this methodology to analyse the structural degree of a large region of HCV genomic RNA in two genotypes associated with different responses to IFN treatment. The results reported here show different structural degree along the genome regions analysed, and differential hybridization patterns for distinct genotypes in NS2 and NS3 HCV regions.

INTRODUCTION

Hepatitis C virus (HCV) is one of the most important causes of chronic liver disease worldwide and 3% of the population is estimated to be infected (1). Molecular studies have shown that the genome of HCV from a single isolate cannot be defined by a single sequence but rather by a population of variant sequences closely related to each other referred to as viral quasispecies (2–5). This enormous diversity has generated genetically distinct groups or genotypes (6). It is well known that HCV genotypes show different response to the currently used antiviral treatment alpha-interferon (IFN-α) in combination with ribavirin, which has a limited long-term efficacy mostly in the HCV genotype 1b (7–9). Consequently, a lot of effort has been put into trying to identify specific sequences in the HCV viral quasispecies that may determine resistance or sensitivity to antiviral therapy. Unfortunately, results have been inconclusive (7–17) probably because IFN-α does not specifically target a particular HCV gene. IFN-α activity enhances host antiviral responses that, in turn, exert selection pressures on various viral genome regions. Thus, the targets of IFN-α action are likely to be numerous and distributed over the entire genome (18–20).

Viral genomic RNA sequences determine secondary and tertiary RNA structures which are part of the viral phenotype. Distinct ordered structures in local regions of single-stranded RNA sequences often correlate with functions such as control of replication, transcription, mRNA processing and translation (21,22). RNA structural motifs of viral genomes have been shown to play important roles, acting as *cis* elements like the internal ribosome entry sites (IRES) (23,24), internal hairpins and pseudoknots directing expression of different open reading frames in retroviruses (25–27), encapsidation signal in human immunodeficiency virus (28–30) and tRNA-like elements described in plant viruses involved in viral replication (31,32). Also, it has been recently demonstrated that the initiation of HCV translation is directed by an RNA structural element (33). Gene expression can also be modulated by RNA structural elements in cellular mRNA in *trans* as in the case of a pseudoknot in IFN-γ mRNA that activates the interferon-inducible protein kinase PKR (34). The RNA structure from the HCV 5′ non-coding region (5′NCR) (35–38) and to a lesser extent those from the core region (39) and the 3′NCR (40–42), have been the subject of many studies. However, the role of variant RNA structures present in the viral quasispecies in relation with IFN sensitivity or resistance is unknown.

In previous studies, structural features of HCV RNA of two different genomic regions, 5′NCR (43,44) and NS2 (45), have been analysed using complementary deoxyoligonucleotide probes and RNase H cleavage. Though this method has proven to be a powerful tool to map RNA structural degree and to determine accessible sites for ribozyme targeting, it is very laborious and time consuming. In the present work, we have

*To whom correspondence should be addressed. Tel: +34 934894034; Fax: +34 934894032; Email: mmartell@vhebron.net

developed a methodology based on hybridizing labelled HCV transcripts to arrays of complementary oligonucleotides, in order to determine the structural degree of large regions in the HCV RNA genome. The hypothesis was that by correlating the intensity of hybridization signals with the degree of genome accessibility, it would be possible to distinguish regions with relatively open RNA structures from those more tightly folded. Here we document that DNA microarray technology can be adapted to map structural degree of the HCV RNA genome. This methodology offers the possibility to evaluate the role of variability in HCV RNA, not at the sequence level but at that of RNA structure, in patients infected with different HCV genotypes and showing different responses to IFN treatment.

MATERIALS AND METHODS

Oligonucleotide design

In a first approach, 20 DNA oligonucleotides were designed for the construction of HCV-specific macroarrays: 6 of them corresponded to the HCV 5′NCR and 14 corresponded to the NS2 coding region. These oligonucleotides were complementary to specific viral regions in which the genomic RNA adopts different degrees of structure, as previously measured by means of RNase H accessibility studies (44,45) (Figure 1A).

In a second step of the analysis, 156 DNA oligonucleotides were designed for the construction of microarrays, corresponding to the entire 5′NCR (nt 1–361), the first three-quarters of the core region (nt 342–751), the E2–NS2 junction (nt 2478–3067) and the first 400 nt of the NS3 region (nt 3419–3819) [numbered as HCV-1b, Gene Bank accession number S62220 (46), Figure 1B]. These oligonucleotides were complementary to genomic regions from HCV-N (S62220) genotype 1b (46) and HCV clone pJ6CF (AF177036) genotype 2a (47). Of them 74 were specific for genotype HCV-1b, 74 were specific for genotype HCV-2a and 8 were common for both genotypes. All the oligonucleotides were 20 nt long and their identification number indicates the nucleotide number at the 5′ position in the original 1b sequence (the sequences of these oligonucleotides are shown in the online supplementary data). They were synthesized (TIB Molbiol) with an amino linker 'C6' [$NH_2(CH_2)_6$] at their 5′end, followed by a spacer of non-specific $(dT)_{15}$ track; purification was performed by high-performance liquid chromatography (HPLC).

Spotting and attachment to glass slides

Oligonucleotides were diluted in spotting solution (Telechem-Arrayit) and spotted onto super-epoxy-coated glass slides (Telechem-Arrayit). Macroarrays were constructed by manually spotting 0.5 µL (dot diameter of approximately 1.5 mm, Figure 2) of oligonucleotide solutions at concentrations of

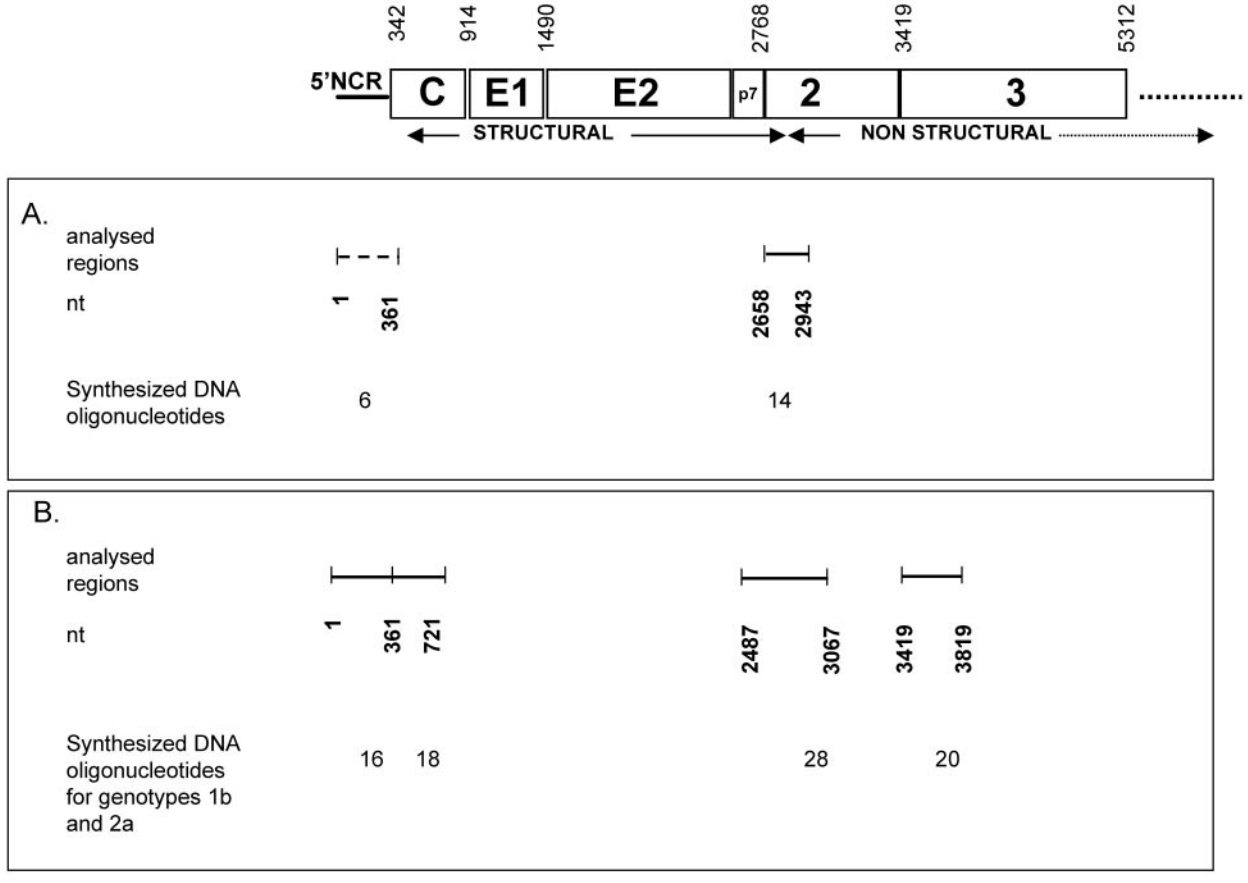


Figure 1. Schematic representation of HCV genome indicating the number of oligonucleotides synthesized in the 5′NCR and NS2 regions for macroarrays (**A**) and in the 5′NCR, core, NS2 and NS3 regions for microarrays (**B**).

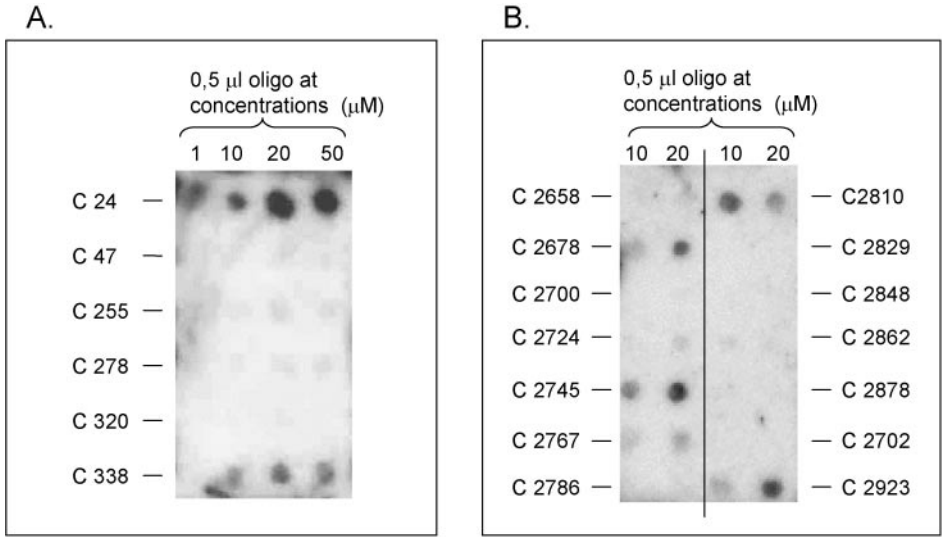


Figure 2. Structural analysis of HCV RNA using macroarrays. Hybridization of HCV RNA transcripts from (**A**) 5′NCR (402 nt long) and (**B**) NS2 region (562 nt long) with oligonucleotides complementary to those regions printed as dots of 1.5 mm in diameter (0.5 µl) and concentrations ranging from 1 to 50 µM. (A) The autoradiogram shows a strong hybridization signal in two accessible regions (C24 and C338). (B) The autoradiogram shows hybridization signals in oligonucleotides corresponding to the most accessible regions of this genomic fragment (C2745, C2923, C2678, C2810).

1, 10, 20 and 50 µM. Microarrays containing 156 oligonucleotides were spotted by means of a GMS 417 arrayer (Affymetrix), defining two contiguous subarrays or grids (genotype 1b at the left side and genotype 2a at the right side) and printing these two-grid configuration twice on the same slide (Figure 3). Each oligonucleotide, at 5 and 25 µM concentration, was spotted in triplicates as overlapping dots of 50 pl and 150 µm in diameter, with a centre-to-centre distance of 150 µm (replicates) or 300 µm (different samples) (table B in the supplementary data). Coupling was achieved 10 min after printing. Printed macro and microarrays were kept, without any further treatment, up to six months at room temperature.

Synthesis and labelling of *in vitro* transcripts

HCV RNA transcripts were derived from plasmid pN(1–4728) that contains a fragment of HCV genotype 1b including nucleotides 1 to 4728, and plasmid pJ6CF, which contains the complete genome of HCV genotype 2a. Templates for *in vitro* transcription were either linearized plasmids (digested by Aat II or Bam HI for 5′NCR or E2/NS2 junction region, respectively) for the macroarrays, or PCR products obtained from the same plasmids and containing a T7 RNA polymerase promoter for the microarrays (Table 1). Standard 25 µL *in vitro* transcription reactions were performed containing 1 µg DNA template, 4 µL [α-^{33}P]UTP (~92 TBq/mmol; 2500 Ci/mmol), 400 µM each NTP and 1.6 U/µL T7 RNA polymerase (Promega). After incubation for 2 h at 37°C, the DNA template was digested with 0.5 U DNase I (Promega) at 37°C for 5 min. Cellulose CF11 chromatography was used to eliminate DNA fragments and non-incorporated nucleotides. When using linearized plasmids, an extra purification step was performed in order to eliminate incomplete or longer transcription products, as described (48). The concentration of radioactive transcripts was determined by calculating the amount of incorporated [α-^{33}P]UTP based on scintillation counting.

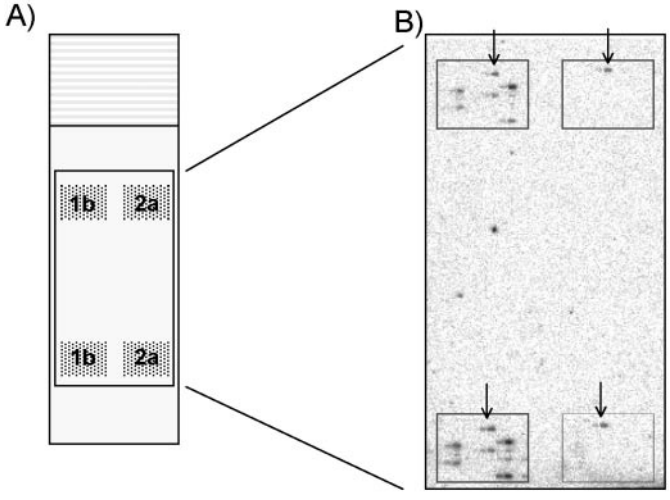


Figure 3. Oligonucleotide arrangement on the microarrays. (**A**) A hundred and fifty six oligonucleotides were spotted on the same slide in two contiguous subarrays or grids: seventy-four oligonucleotides specific for genotype HCV-1b and eight common for both genotypes were printed at the left side grid (1b), and seventy-four specific for genotype HCV-2a were printed at the right side of the slide (2a). This pattern has been printed twice on the microarray (**B**) As an example, hybridization with RNA transcript HCV NS2-1b (plus 5′NCR 1b for location purposes) is shown. Cross hybridization of RNA transcript 1b with genotype 2a-specific oligonucleotides, only occurs in oligo 338 (arrows) sharing all the nucleotides but one between both genotypes.

Hybridization and washing

Slides were processed just before hybridization to remove unbound DNA molecules and printing buffer components from the substrate. Briefly, the slides were washed for 2 min at room temperature in 2× SSC, 0.1% sarcosyl, and for 2 min at room temperature in 2× SSC. Oligonucleotides were denatured for 2 min at 100°C in H_2O, cooled for 10 s at

Table 1. Primers used for PCR amplifications of 5′NCR, Core, NS2 and NS3 regions. (+) Sense primers, (−) antisense primers

Genotype	Primer name	Primer sequence (5′-3′)	Amplified HCV region	RNA transcript length
HCV 1b	1b/2a 5′T7 1(+)	TAATACGACTCACTATAGGACCCGCCCCGATTGGGGGCGA	5′NCR (1–360)	360 nt
	1b 5′ 341 (−)	GGTTTAGGATTCGTGCTCAT		
	1b/2a 5′T7 1(+)	TAATACGACTCACTATAGGACCCGCCCCGATTGGGGGCGA	5′NCR (1–417)	417 nt
	1b C393 (−)	CCGGAATTCCCCGGGAACTTGACGTCCTGTGGGC		
	1b/2a CT7 358(+)	TAATACGACTCACTATAGCCTCAAAGAAAAACCAA	CORE (358–751)	393 nt
	1b/2a 731 (−)	GATACCCCATGAGGTCGGC		
	1b NS2T7 2487(+)	TAATACGACTCACTATAGGGAGTATGTCGTGTTGCTT	E2-NS2 junction (2487–3107)	620 nt
	1b NS2 3087 (−)	AGCGCGTACGAAGTACGGCA		
	1b NS3T7 3419(+)	TAATACGACTCACTATAGGCGCCCATCACGGCCTAC	NS3 (3419–3840)	421 nt
	1b NS3 3820 (−)	CCGAAGAGCCCTTCAAGTAG		
	1b/2a 5′T7 1 (+)	TAATACGACTCACTATAGGACCCGCCCCGATTGGGGGCGA	5′NCR (1–362)	362 nt
	2a 5′ 342 (−)	GAGGTTTAGGATTTGTGCTC		
	1b/2a CT7 358 (+)	TAATACGACTCACTATAGCCTCAAAGAAAAACCAA	CORE (358–731)	393 nt
	1b/2a C 731 (−)	GATACCCCATGAGGTCGGC		
HCV 2a	2a NS2T7 2499(+)	TAATACGACTCACTATAGGGAGTGGGTAATACTCTTA	E2-NS2 junction (2499–3118)	619 nt
	2a NS2 3098 (−)	AGCCCTGACGAAGTACGGCA		
	2a NS3T7 3430(+)	TAATACGACTCACTATAGGCCCCCATCACTGCTTAC	NS3 (3430–3851)	421 nt
	2a NS3 3831 (−)	CTGAGGACCCCTTCAAGGTG		

room temperature, fixed by plunging the slides for 2 min in ice cold 100% ethanol and finally centrifuged for 1 min at 500 *g*.

Hybridization of the labelled transcripts was carried out under non-denaturing conditions, using appropriate temperature and buffer composition to allow binding of the RNA transcript to the immobilized probes, without disturbing its secondary/tertiary structure (44). Hybridization mix consisted of 60 nM labelled RNA transcript, 1× buffer H (20 mM Hepes–HCl pH 7.8, 50 mM KCl, 10 mM $MgCl_2$, 1 mM DTT), 1× BSA, 50 µg/ml tRNA, 50 µg/ml poly A, 0.01% SDS and 40 U RNasin. After 2 h of incubation in the hybridization chamber (Genetix) at 37°C, slides were washed for 15 min in 1× buffer H, 0.2% SDS and for 15 min in 1× buffer H, at room temperature. The slides were finally dried by spinning for 1 min at 500 *g* and immediately exposed either to XAR-5 (Kodak) X-ray films (macroarrays), or to high resolution BAS-SR (Fujifilm) imaging plates (microarrays).

Signals in the autoradiographs of the macroarrays were compared by visual inspection. Hybridization signals in the microarrays were captured by the Radioisotopic Image Analyser FLA-5000 (Fuji), with scanning pixels of 25×25 µm^2. Overlapping dots spotted onto the microarrays allowed hybridization signals to be detected as 'bands' of about 150×450 µm^2. Analysis softwares FLA-Image Reader and Image Gauge (Fuji) were used for reading and quantifying hybridization images. Reproducibility of the method was assessed by comparing the results of three different hybridization experiments for each genotype and region. Error bars (mean and SD) were calculated from the hybridization signals in each position expressed as a percentage of the highest signal in the three experiments (100% of hybridization).

RESULTS AND DISCUSSION

Duplex formation between complementary oligonucleotides and long nucleic acids is not only determined by the sequence complementarity and it is constrained by the secondary/tertiary structure of the target nucleic acid. Experiments to measure the effect of different single-stranded nucleic acid structures, under high salt conditions, on the rate of duplex formation on solid supports, have shown that RNA or cDNA structural motifs may impede hybridization between probe and target and, therefore, bias the results obtained in gene expression arrays (49–52). Here, we have reversed the argument and, by taking advantage of the solid support hybridization techniques, we have aimed at evaluating the structural degree of long RNA molecules with potential application to HCV diagnostics and therapeutics. We have first set up the conditions for probing the accessibility of long HCV RNA transcripts by hybridization to complementary oligonucleotides on a solid support in macroarray format. In a second step, we have constructed DNA microarrays to analyse the structural degree of long HCV fragments in two different HCV genotypes known to be associated to different responses to IFN treatment.

A first approach on macroarrays

Macroarrays were constructed with a set of 20 DNA oligonucleotides, corresponding to HCV regions previously known to display different structural features: the 5′NCR and the E2/NS2 junction region. Macroarrays were hybridized with labelled HCV RNA transcripts under non-denaturing conditions. The autoradiogram in Figure 2A shows the hybridization pattern of a 402 nt long HCV RNA transcript from the 5′NCR. Strong radioactive signal can be visualized in fully accessible regions complementary to oligos C24 and C338, while the other regions of limited or null accessibility cannot be discriminated showing weak or no radioactive signal. Figure 2B shows the autoradiogram of the hybridization signal of a 562 nt long HCV RNA transcript from the E2/NS2 junction region with complementary oligonucleotides. Radioactive signals were obtained in C2745, C2923, C2678 and C2810. Images shown in Figure 2 have been reproducibly observed in two independent experiments obtained by means of different labelled RNA transcript preparations. These results are in agreement with previous ones using RNase H and DNA

oligonucleotides in solution (43–45), and demonstrate the suitability of the proposed method to distinguish highly structured RNA from less structured ones, in the HCV genome. Macroarrays were printed manually and compared by visual inspection and therefore not used for fine quantitation purposes but they suggested that the approach was valid and we could go to the next step using microarrays. In addition, in a control experiment two non-accessible and one accessible sites showed positive signals when the hybridization was performed with labelled unstructured short 21mer RNA molecules (data not shown).

Analysis of the structural profile along the HCV RNA genome using microarrays

Genotype-specific microarrays for HCV 1b and HCV 2a were prepared to cover the entire 5′NCR, the first three-quarters of the core region, the E2–NS2 junction region, and the first 400 nt of the NS3 region. Figure 4 shows the results of hybridization of microarrays of genotype 1b and 2a with HCV RNA transcripts corresponding to these regions. A high signal/background ratio was obtained, together with a remarkable specificity and reproducibility. Cross-hybridization between one particular RNA transcript and non-specific oligonucleotides on the same slide (i.e. RNA 1b transcript with oligonucleotides genotype 2a specific, or vice versa) (see Figure 3) only occurs when the oligonucleotides correspond to accessible regions and differ in 1 or 2 nt between genotypes. This is the case of oligonucleotides 338 in the 5′NCR, and 361 and 381 in the core region. In contrast, oligonucleotides differing in $\geqslant$3 nt between genotype 1b and 2a showed absolute lack of cross-hybridization, thus representing the negative controls of the method. Inter-experiment reproducibility is represented by the error bar graphs showed in Figure 5. They show the mean and standard deviation (SD) of three different experiments for each genotype and region. Hybridization patterns are coincident in the different experiments, as it can be seen in Figure 4, where one single and representative experiment is shown.

As a whole, microarray results confirmed the signals obtained with macroarrays and demonstrated: (i) different structural degree along the genome regions analysed and (ii) different hybridization patterns for different genotypes in certain HCV regions.

Genomic regions of conserved and structured RNA. The 5′NCR and core regions (Figure 4A and B) show an important degree of structural compactness that prevents hybridization with most oligonucleotides: out of 34 oligonucleotides printed (covering the first 750 nt of the HCV genome) only numbers 24, 338, 361 and 381 showed detectable hybridization signals both for genotype 1b and 2a. Interestingly, when comparing the hybridization of a 361 nt long HCV genotype 1b transcript corresponding to the 5′NCR and the first nucleotides of the core region with the hybridization of a 417 nt long HCV transcript from the same region (Figure 6), we observed that the signal corresponding to oligo 338 was not detectable in the hybridization of the longer fragment, and, instead, a hybridization signal appeared in the last oligonucleotide covering this fragment (381). This suggests that hybridization signals with oligonucleotides complementary to the end of the transcript could be in some cases artefactual. Artificial transcript ends could create exposed 5′ and/or 3′ terminal sequences, otherwise compromised in secondary/tertiary interactions in the full length RNA genome, as it seems to be the case for oligonucleotides 338 and 361. It is generally accepted that many secondary structures are local and folding occurs as the RNA molecule is transcribed. This makes some sequences to remain accessible to further long-range interactions, while others show restricted accessibility to distant regions. As elongation of the RNA chain proceeds, more structures are added to the transcript which could mask the accessible sites previously obtained (51). Therefore, one should be cautious when interpreting accessible positions, since long-range interactions may involve not only the ends of an RNA transcript but also its internal regions, otherwise revealed by the loss of hybridization signals that results from extending the transcript length. In contrast, we can be very confident of our results when considering secondary local elements or short-range interactions demonstrated by a reduction or a total lack of heteroduplex formation in microarrays.

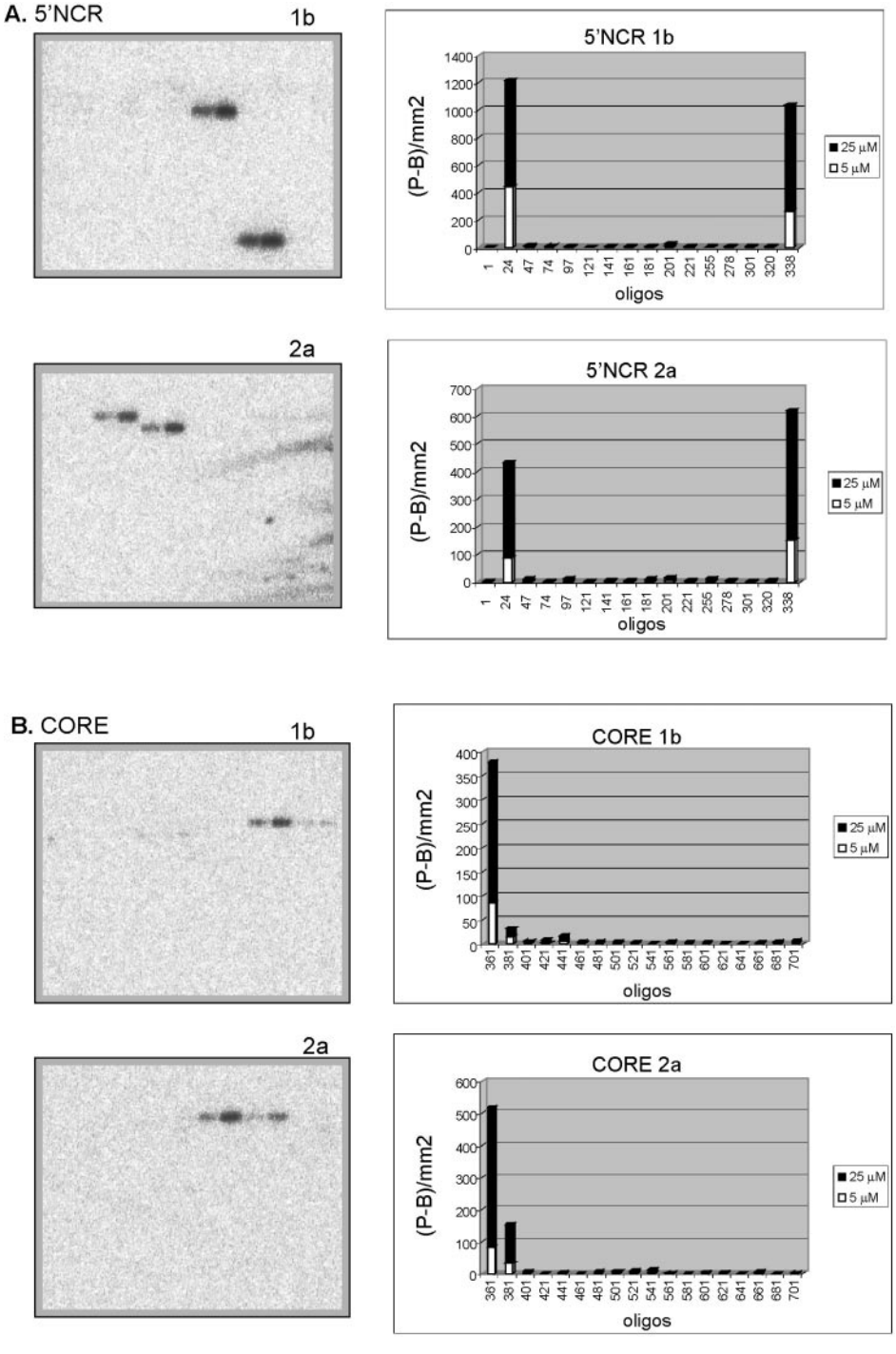
Our results also point out that, except for the position corresponding to oligo 24 (nt 24–44), the structural degree of the 5′NCR and core represents, as a whole, a non-accessible region to probe hybridization. Biochemical and functional studies have established that the HCV 5′NCR folds into a highly ordered complex structure with multiple stem–loops (35,37,53,54). These regions contains two distinct RNA elements: a short 5′ proximal RNA element (nt 1–43) that regulates HCV RNA replication and a longer IRES element (nt 44–370) where protein synthesis starts, eventually leading to production of the polyprotein precursor. The results reported here, both with the 361 nt long and the 417 nt long HCV transcripts (Figure 6), fit quite well with the model proposed for a compact IRES element and an accessible region corresponding to the 5′ proximal RNA element, mainly to nucleotides 24–44 (35,37,53,54).

It is interesting to note that hybridization on solid support is a more restrictive method than mapping the structural degree by RNase H and complementary DNA oligonucleotides in solution. Using our conditions, a relatively reduced signal in the solid support with regard to the experiment in solution would be attributable either to stearic constrains of the folded RNA target to approximate the oligonucleotide or to unspecific interactions of the RNA with the chemically modified solid support. The first hypothesis might be the most appropriate for the case of the 5′ untranslated region of HCV RNA which is an exceptionally highly structured domain and might approach the solid support in a preferred orientation. As an example, RNase H stimulates moderate cleavage in position 255 (44) while it seems non-accessible in the microarray experiment. Also, stem–loop IIIb, which is accessible to single-stranded-specific nucleases and expected to be an accessible region, is not observed in our hybridization on solid support.

Apart from its 5′ end, the lack of accessibility in the entire analysed fragment of the core-encoding region (nt 342–751) was very surprising and suggests a compact folding not expected for a region only involved in protein coding. The earliest evidence that the HCV core RNA sequence harbours additional functions than coding for a single protein was the observation that the rate of synonymous substitution for the core gene was remarkably lower than that for the other HCV genes (55,56). The limitation in the number of synonymous

substitutions in this region has been explained by the possibility of encoding an alternative protein by frame shifting (ARFP) (57,58), and by the existence of internally base-paired stem–loop structures in this region (39). Our results suggest that at least the first three-quarters of the core region contain a number of RNA structural elements that probably lead to structural constraints to nucleotide replacements in this region.

Genomic regions of accessible and variable RNA. A very different situation is observed in the 620 nt long fragment from the E2/NS2 junction region (Figure 4C) and in the 420 nt long fragment from the NS3 region (Figure 4D). Both RNA transcripts show complex hybridization patterns that indicate much less structured regions, where accessible positions alternate with more compact fragments. The hybridization patterns clearly differ between genotype 1b and 2a in both regions. The quantitation results show that in the E2/NS2 junction region (Figure 4C) the fragment corresponding to oligonucleotides 2829–2862 is not accessible at all in genotype 1b, but it is highly accessible in genotype 2a. The opposite situation is found for fragment 2724–2767, where genotype 1b is more accessible than genotype 2a.



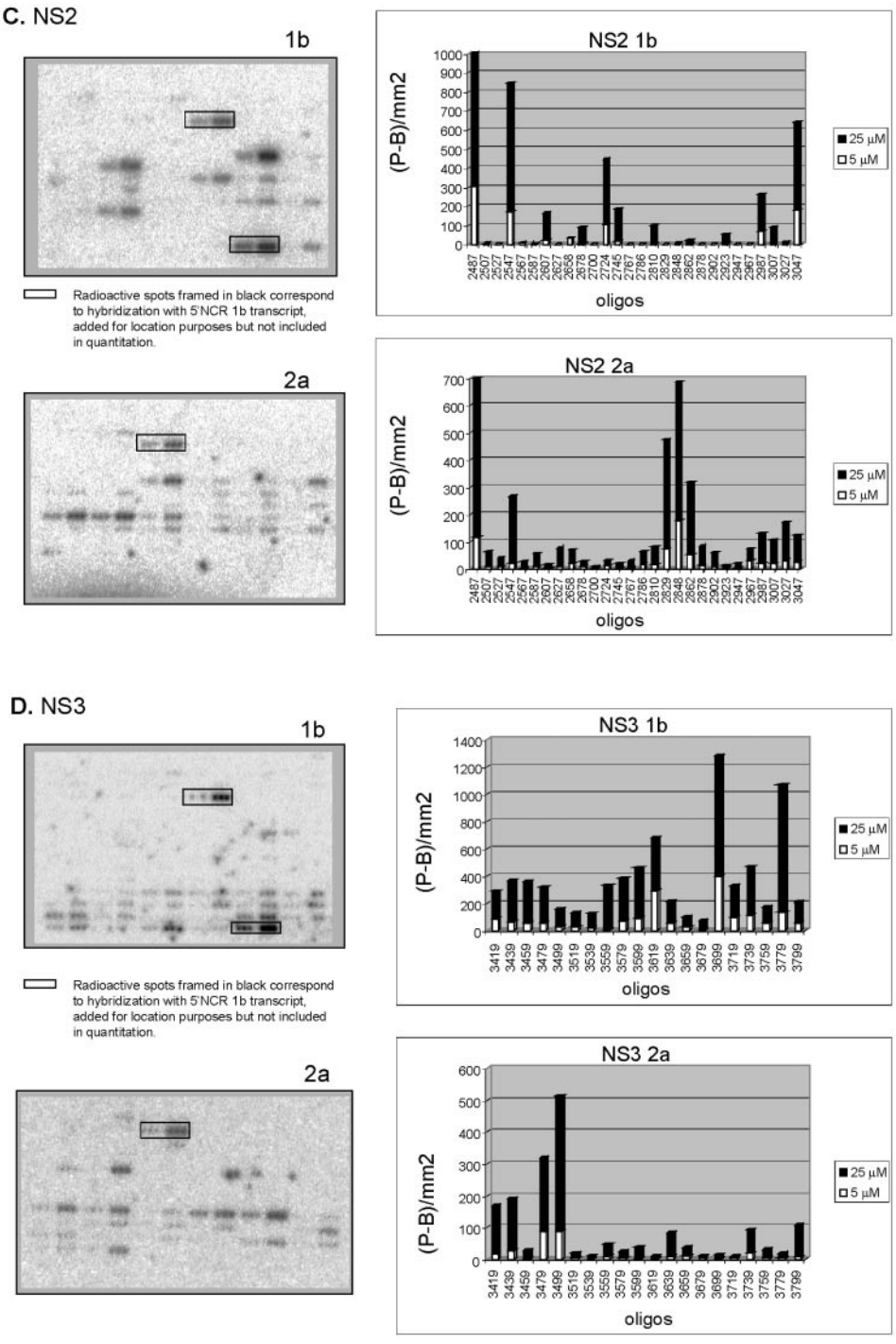
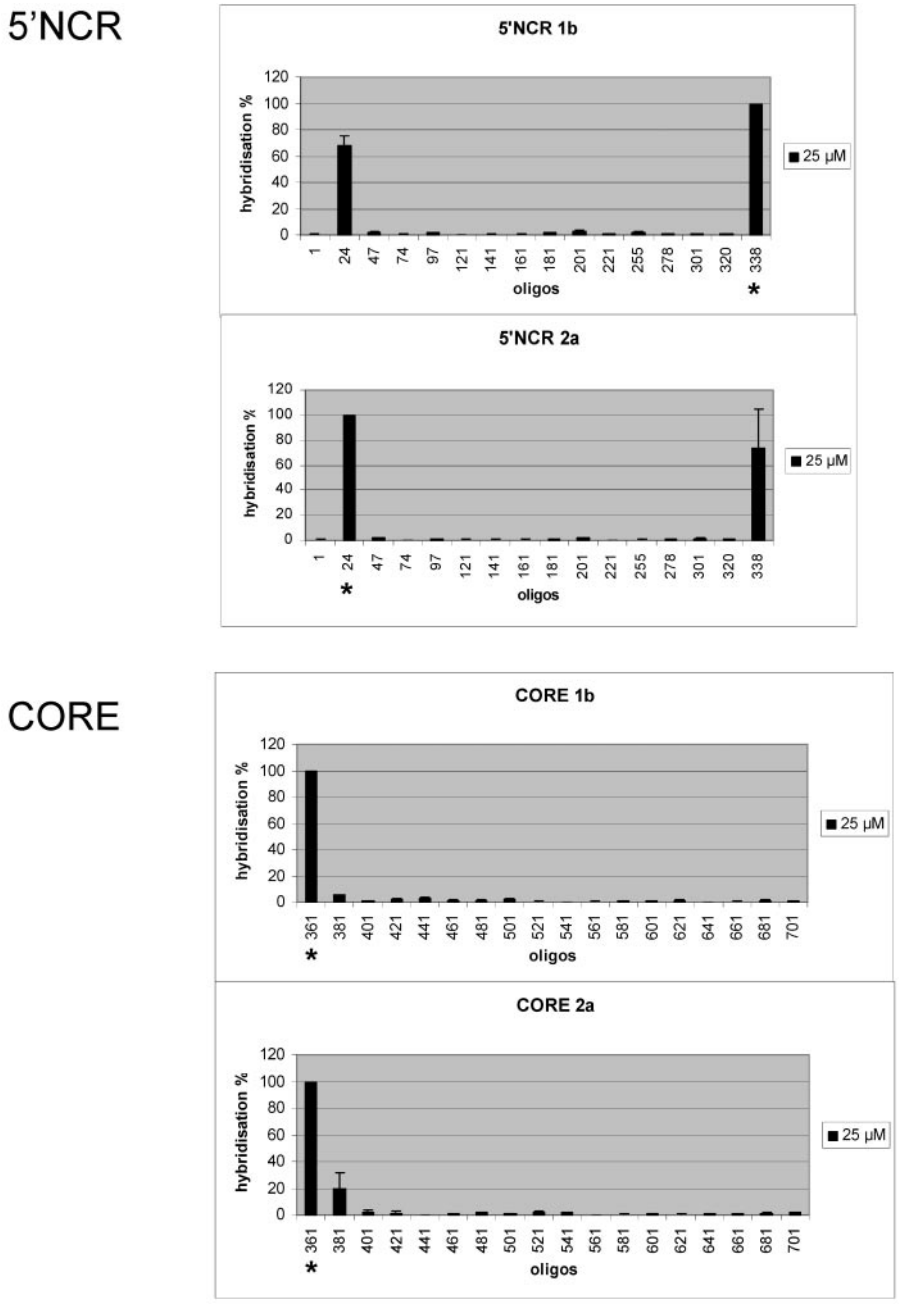


Figure 4. Analysis of the RNA structural degree in different regions of HCV-1b and HCV-2a genomes using DNA microarrays. The oligonucleotides were printed as overlapping triplicate spots of 150 µm in diameter, at two concentrations (5 and 25 µM). The relative positions of the spotted oligonucleotides is different for genotypes 1b and 2a (see details in Table B in the supplementary data). The HCV genome regions analysed were the entire 5′NCR (**A**), the first three quarters of core region (**B**), the E2/NS2 junction region (**C**) and the first 400 nt of NS3 region (**D**). Hybridization signals of RNA transcripts corresponding to HCV genotypes 1b and 2a are shown at the left. 5′NCR-1b RNA transcript was added to the hybridization mix when analysing NS2 and NS3 regions, as positive control in the printed area. Quantitation of hybridization signals is shown at the right part of each panel. $(P-B)/mm^2$ indicates the area density of photo-stimulated luminescence (PSL) of hybridization signal minus background. White bars correspond to the quantitation of hybridization signals with oligonucleotides printed at 5 µM concentration; the sum of white and black bars correspond to the quantitation of hybridization signals with oligos printed at 25 µM.

Other regions, as 2487–2507, 2547–2567 and 2987–3067 are relatively accessible in both genotypes. Interestingly, the presence of a tRNA-like structure, flanking position 2860–2861 in the genome of HCV genotype 1b has been recently demonstrated (48). Also, this structure was consistently maintained in several HCV-1b variants, although they differ in the primary sequence at or near the relevant domain. Our results are consistent with the presence of an



important RNA structural element involving nucleotides 2829–2862 in the RNA transcript corresponding to genotype 1b, while the same fragment seems to be accessible in genotype 2a.

In the NS3 region (Figure 4D), genotype 1b is, in general, much more accessible than genotype 2a, particularly in the fragment corresponding to oligonucleotides 3619–3799.

Perspectives of the DNA microarrays methodology in the study of HCV genome

A specific sequence intrinsically involved in resistance to IFN treatment, i.e. significantly correlated with non-responder patients, has not been described. Nevertheless, an inverse correlation has been demonstrated between quasispecies complexity in different genomic regions and response to IFN treatment (12,59–61). Approaching the extremely high sequence variation of HCV by analysing the limited variability of its RNA structures can simplify the study of quasispecies in relation to IFN resistance or sensitivity in patients infected with different HCV genotypes. Theoretical studies suggest that quasispecies sequence space is much more complex than the corresponding space of variant structures: several variant sequences can adopt the same RNA secondary structure (62) and, additionally, functional secondary/tertiary structures are strongly conserved.

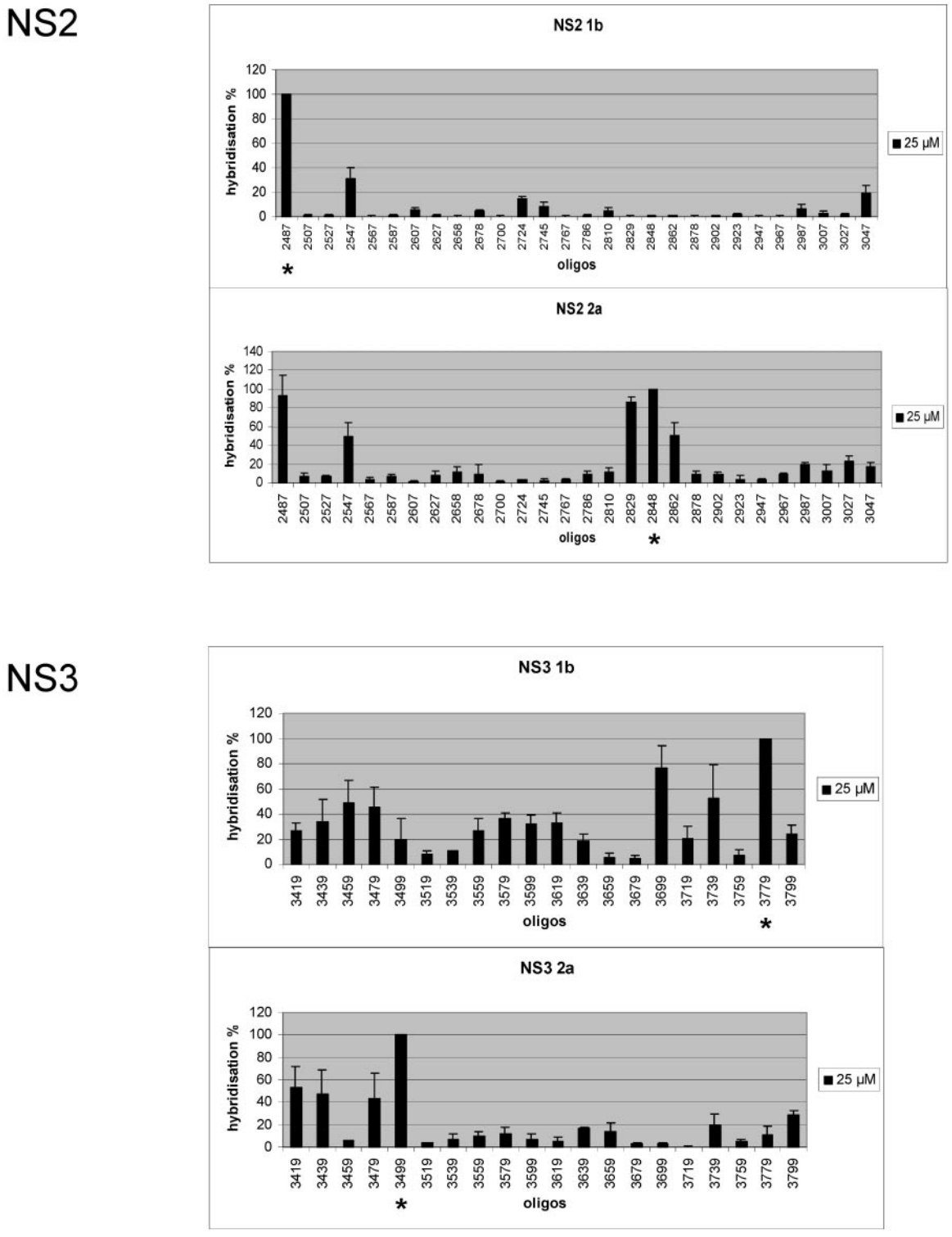


Figure 5. Reproducibility analysis. The results of three different hybridization experiments for each genotype and region were compared. Error bars (mean and SD) were calculated from the hybridization signals in each position expressed as a percentage of the highest signal in the three experiments (100% of hybridization), identified with an asterisk.

The antiviral activity of IFN involves induction of transcription of several antiviral genes, such as the double-stranded (ds) RNA-dependent protein kinase, which is activated in the presence of dsRNA. In this activated form, PKR inhibits the synthesis of (cellular or viral) proteins (63). It has been recently shown that a region of the viral RNA comprising part of the IRES is able to bind to PKR in competition with dsRNA, preventing activation of kinase *in vitro* and diminishing its

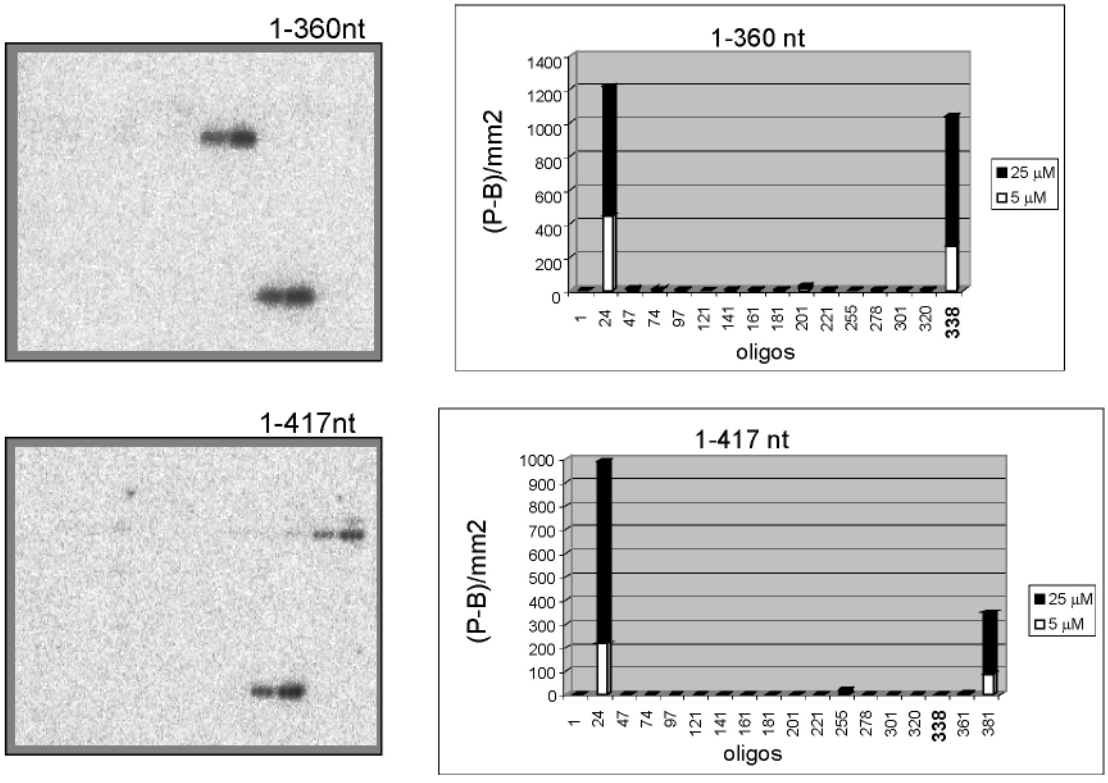


Figure 6. Analysis of artificial end effects. Hybridization of two 5′NCR RNA transcripts of different length, covering positions 1–360 and 1–417 from HCV genotype 1b. Hybridization signals obtained with both fragments are shown at the left: signal corresponding to oligo 338 is not present with the 417 nt long fragment. Quantitation of hybridization signals is represented in the panels at the right.

inhibitory effect (64). However, the HCV IRES, comprising the 5′NCR and the first nucleotides of the HCV coding sequence is highly conserved between isolates. In particular, the nucleotide variations found have little effect on the extensive secondary structure of the IRES, consistently with our results in 5′NCR and core region in two different genotypes. Although current evidences cannot exclude it completely, it seems unlikely that the structural elements present in the HCV IRES could account for the differences in sensitivity or resistance to IFN treatment associated with HCV genotypes.

The results reported here also show a differential structural degree in the internal genomic regions NS2 and NS3 in HCV genotypes 1b and 2a, although, we cannot presume any implication of these structural differences in the modulation of the inhibitory effect of IFN on the synthesis of viral proteins.

In conclusion, we have documented that DNA microarray technology can be used as a high-throughput method to analyse the entire HCV genome for the identification of RNA structural elements. The next step would be to use this technology to analyse complete HCV genomes obtained from infected patients showing different response pattern to antiviral treatment. Classification of hybridization profiles and correlation to IFN resistance or sensitivity could help to identify the genetic determinants of this complex phenomenon.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online: Table A, list of oligonucleotides printed on the microarrays; Table B, location and concentrations of oligonucleotides printed on the microarrays.

ACKNOWLEDGEMENTS

We would like to thank Prof. E. Domingo and Prof. J. Genescà for valuable comments on the manuscript. The authors would also like to thank Dr J. Bukh (National Institute of Health, Bethesda, MD) for kindly providing the hepatitis C virus clone pJ6CF of strain HC-J6 (genotype 2a), Dr V. Parro for helping in the setting up the DNA microarray technology at Centro de Astrobiología (CAB) and Dr L. Sumoy (Centre for Genomic Regulation, Barcelona, Spain) for technical support.

M.M. was supported in part by grants DITTO-HCV QLK2-2000-0836 from the European Commission FP5 and Red Nacional de Gastroenterología y Hepatología (C03/02). The work in Hospital Vall d'Hebron was supported in part by grants SAF2000/0183 and PROFIT (FIT-010000-2003-51) from the Ministerio de Ciencia y Tecnología. The work at CAB was supported by the European Union, Instituto Nacional de Técnica Aerospacial, Ministerio de Ciencia y Tecnología and Comunidad de Madrid.

REFERENCES

1. Choo,Q.L., Kuo,G., Weiner,A.J., Overby,L.R., Bradley,D.W. and Houghton,M. (1989) Isolation of cDNA clone derived from a blood-borne non A non B viral hepatitis genome. *Science*., **244**, 359–362.

2. Abid,K., Quadri,R., Negro,F., Taylor,D.R., Shi,S.T., Romano,P.R., Barber,G.N. and Lai,M.M. (2000) Hepatitis C virus, the E2 envelope protein, and alfa-Interferon resistance. *Science*, **287**, 1555.
3. Gomez,J., Martell,M., Quer,J., Cabot,B. and Esteban,J.I. (1999) Hepatitis C viral quasispecies. *J.Viral Hepat.*, **6**, 3–16.
4. Martell,M., Esteban,J.I., Quer,J., Genesca,J., Weiner,A., Esteban,R., Guardia,J. and Gomez,J. (1992) Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J. Virol.*, **66**, 3225–3229.
5. Domingo,E. (1996) Biological significance of viral quasispecies. *J. Viral Hepat.*, **2**, 247–261.
6. Bukh,J., Miller,R.H. and Purcell,R.H. (1995) Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. *Semin. Liver Dis.*, **15**, 41–63.
7. Lee,W.M. (1997) Therapy of hepatitis C: interferon alfa-2a trials. *Hepatology*, **26**, 89S–95S.
8. Poynard,T., Leroy,V., Cohard,M., Thevenot,T., Mathurin,P., Opolon,P. and Zarski,J.P. (1996) Meta-analysis of interferon randomized trials in the treatment of viral hepatitis C: effects of dose and duration. *Hepatology*, **24**, 778–789.
9. Tong,M., Reddy,R., Lee,W.M., Pockros,P.J., Hoefs,J., Keeffe,E.B., Hollinger,F.B., Hathcote,E.J., White,H., Foust,R.T. *et al.* (1997) Treatment of chronic hepatitis C with consensus interferon: a multicenter, ramdomized, controlled trial. Consensus Interferon study group. *Hepatology*, **26**, 747–754.
10. Enomoto,N., Sakuma,I., Asahina,Y., Kurosaki,M., Murakami,T., Yamamoto,C., Izumi,N., Marumo,F. and 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.
11. Enomoto,N., Sakuma,I., Asahina,Y., Kurosaki,M., Murakami,T., Yamamoto,C., Ogura,Y., Izumi,N., Marumo,F. and Sato,C. (1996) Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N. Engl. J. Med.*, **334**, 77–81.
12. Gonzalez-Peralta,R.P., Liu,W.Z., Davis,G.L., Qian,K.P. and Lau,J.Y. (1997) Modulation of hepatitis C virus quasispecies heterogeneity by interferon-alpha and ribavirin therapy. *J. Viral Hepat.*, **4**, 99–106.
13. Pawlotsky,J.M., Germanidis,G., Neumann,A.U., Pellerin,M., Frainais,P.O. and Dhumeaux,D. (1998) Interferon resistance of hepatitis C virus genotype 1b: relationship to nonstructural 5A gene quasispecies mutations. *J. Virol.*, **72**, 2795–2805.
14. Saiz,J.C., Lopez-Labrador,F.X., Ampurdanes,S., Dopazo,J., Forns,X., Sanchez-Tapias,J.M. and Rodes,J. (1998) The prognostic relevance of the nonstructural 5A gene interferon sensitivity determining region is different in infections with genotype 1b and 3a isolates of hepatitis C virus. *J Infect. Dis.*, **177**, 839–847.
15. Squadrito,G., Leone,F., Sartori,M., Nalpas,B., Berthelot,P., Raimondo,G., Pol,S. and Brechot,C. (1997) Mutations in the nonstructural 5A region of hepatitis C virus and response of chronic hepatitis C to interferon alfa. *Gastroenterology*, **113**, 567–572.
16. Taylor,D.R., Shi,S.T., Romano,P.R., Barber,G.N. and Lai,M.M. (1999) Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. *Science*, **285**, 107–110.
17. Quer,J., Murillo,P., Martell,M., Gomez,J., Esteban,J.I., Esteban,R. and Guardia,J. (2004) Subtype mutations in the envelope 2 region including phosphorylation homology domain of Hepatitis C virus do not predict effectiveness of antiviral therapy. *J. Viral Hepat.*, **11**, 45–54.
18. Goodbourn,S., Didcock,L. and Randall,R.E. (2000) Interferons: cell signaling, immune modulation, antiviral responses and virus countermeasures. *J. Gen. Virol.*, **81**, 2341–2364.
19. Stark,G.R., Kerr,I.M., Williams,B.R., Silverman,R.H. and Schreiber,R.D. (1998) How cells respond to interferons. *Annu. Rev. Biochem.*, **67**, 227–264.
20. Heim,M.H., Moradpour,D. and Blum,H.E. (1999) Expression of hepatitis C virus proteins inhibits signal transduction through the Jak-STAT pathway. *J. Virol.*, **73**, 8469–8475.
21. Bashirullah,A., Cooperstock,R.L. and Lipshitz,I.D. (1998) RNA localization in development. *Annu. Rev. Biochem.*, **67**, 335–394.
22. Rajagopalan,L.E. and Malter,J.S. (1997) Regulation of eukaryotic messenger RNA turnover. *Prog. Nucleic Acid Res. Mol. Biol.*, **56**, 257–286.
23. Martinez-Salas,E., Ramos,R., Lafuente,E. and Lopez,d.Q. (2001) Functional interactions in internal translation initiation directed by viral and cellular IRES elements. *J. Gen. Virol.*, **82**, 973–984.
24. Reynolds,J.E., Kaminski,A., Kettinen,H.J., Grace,K., Clarke,B.E., Carroll,A.R., Rowlands,D.J. and Jackson,R.J. (1995) Unique features of internal initiation of hepatitis C virus RNA translation. *EMBO J.*, **14**, 6010–6020.
25. Chen,X., Kang,H., Shen,L.X., Chamorro,M., Varmus,H.E. and Tinoco,I., Jr (1996) A characteristic bent conformation of RNA pseudoknots promotes -1 frameshifting during translation of retroviral RNA. *J. Mol. Biol*, **260**, 479–483.
26. Jacks,T., Madhani,H.D., Masiarz,F.R. and Varmus,H.E. (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell*, **55**, 447–458.
27. ten Dam,E.B., Pleij,C.W. and Bosch,L. (1990) RNA pseudoknots: translational frameshifting and readthrough on viral RNAs. *Virus Genes*, **4**, 121–136.
28. Darlix,J.L., Gabus,C., Nugeyre,M.T., Clavel,F. and Barre-Sinoussi,F. (1990) Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol.*, **216**, 689–699.
29. Skripkin,E., Paillart,J.C., Marquet,R., Ehresmann,B. and Ehresmann,C. (1994) Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization *in vitro*. *Proc. Natl Acad. Sci. USA*, **91**, 4945–4949.
30. Lanchy,J.M., Ivanovitch,J.D. and Lodmell,J.S. (2003) A structural linkage between the dimerization and encapsidation signals in HIV-2 leader RNA. *RNA*, **9**, 1007–1018.
31. Litvak,S., Tarrago,A., Tarrago-Litvak,L. and Allende,J.E. (1973) Elongation factor-viral genome interaction dependent on the aminoacylation of TYMV and TMV RNAs. *Nat. New Biol.*, **241**, 88–90.
32. Giegé,R., Florentz,C. and Dreher,T.W. (1993) The TYMV tRNA-like structure. *Biochimie*, **75**, 569–582.
33. Wang,T.H., Rijnbrand,R.C.A. and Lemon,S.M. (2000) Core protein-coding sequence, but not core protein, modulates the efficiency of cap-independent translation directed by the internal ribosome entry site of hepatitis C virus. *J. Virol.*, **74**, 11347–11358.
34. Ben Asouli,Y., Banai,Y., Pel-Or,Y., Shir,A. and Kaempfer,R. (2002) Human interferon-gamma mRNA autoregulates its translation through a pseudoknot that activates the interferon-inducible protein kinase PKR. *Cell*, **108**, 221–232.
35. Brown,E.A., Zajac,A.J. and Lemon,S.M. (1994) *In vitro* characterization of an internal ribosomal entry site (IRES) present within the 5′ nontranslated region of hepatitis A virus RNA: comparison with the IRES of encephalomyocarditis virus. *J.Virol.*, **68**, 1066–1074.
36. Smith,D.B., Mellor,J., Jarvis,L.M., Davidson,F., Kolberg,J., Urdea,M., Yap,P.L. and Simmonds,P. (1995) Variation of the hepatitis C virus 5′ non-coding region: implications for secondary structure, virus detection and typing. The International HCV Collaborative Study Group. *J. Gen. Virol.*, **76**, 1749–1761.
37. Wang,C., Sarnow,P. and Siddiqui,A. (1994) A conserved helical element is essential for internal initiation of translation of hepatitis C virus RNA. *J. Virol.*, **68**, 7301–7307.
38. Wang,C., Le,S.Y., Ali,N. and Siddiqui,A. (1995) An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5′ noncoding region. *RNA*, **1**, 526–537.
39. Walewski,J.L., Gutierrez,J.A., Branch-Elliman,W., Stump,D.D., Keller,T.R., Rodriguez,A., Benson,G. and Branch,A.D. (2002) Mutation Master: profiles of substitutions in hepatitis C virus RNA of the core, alternate reading frame, and NS2 coding regions. *RNA*, **8**, 557–571.
40. Blight,K.J. and Rice,C.M. (1997) Secondary structure determination of the conserved 98-base sequence at the 3′ terminus of the hepatitis C virus genome RNA. *J. Virol.*, **71**, 7345–7352.
41. Ito,T. and Lai,M.M. (1997) Determination of the secondary structure of and cellular protein binding to the 3′-untranslated region of the hepatitis C virus RNA genome. *J. Virol.*, **71**, 8698–8706.
42. Yamada,N., Tanihara,K., Takada,A., Yorihuzi,T., Tsutsumi,M., Shimomura,H., Tsuji,T. and Date,T. (1996) Genetic organization and diversity of the 3′ noncoding region of the hepatitis C virus genome. *Virology*, **223**, 255–261.

43. Lima,W.F., Brown-Driver,V., Fox,M., Hanecak,R. and Bruice,T.W. (1997) Combinatorial screening and rational optimization for hybridization to folded hepatitis C virus RNA of oligonucleotides with biological antisense activity. *J. Biol. Chem.*, **272**, 626–638.
44. Lyons,A.J., Lytle,J.R., Gómez,J. and Robertson,H.D. (2001) Hepatitis C virus internal ribosome entry site RNA contains a tertiary structural element in a functional domain of stem–loop II. *Nucleic Acids Res.*, **29**, 2535–2541.
45. Nadal,A., Robertson,H.D., Guardia,J. and Gomez,J. (2003) Characterization of the structure and variability of an internal region of hepatitis C virus RNA for M1 RNA ghide sequence ribozyme targeting. *J. Gen. Virol.*, **84**, 1545–1548.
46. Hayashi,N., Higashi,H., Kaminaka,K., Sugimoto,H., Esumi,M., Komatsu,K., Hayashi,K., Sugitani,M., Suzuki,K. and Tadao,O. (1993) Molecular cloning and heterogeneity of the human hepatitis C virus (HCV) genome. *J. Hepatol.*, **17** (Suppl. 3), S94–S107.
47. Yanagi,M., Purcell,R.H., Emerson,S.U. and Bukh,J. (1999) Hepatitis C virus: an infectious molecular clone of a second major genotype (2a) and lack of viability of intertypic 1a and 2a chimeras. *Virology*, **262**, 250–263.
48. Nadal,A., Martell,M., Lytle,J.R., Lyons,A.J., Robertson,H.D., Cabot,B., Esteban,J.I., Esteban,R., Guardia,J. and Gomez,J. (2002) Specific Cleavage of Hepatitis C Virus RNA Genome by Human RNase P. *J. Biol. Chem.*, **277**, 30606–30613.
49. Southern,E.M., Case-Green,S.C., Elder,J.K., Johnson,M., Mir,K.U., Wang,L. and Williams,J.C. (1994) Arrays of complementary oligonucleotides for analysing the hybridisation behaviour of nucleic acids. *Nucleic Acids Res.*, **22**, 1368–1373.
50. Williams,J.C., Case-Green,S.C., Mir,K.U. and Southern,E.M. (1994) Studies of oligonucleotide interactions by hybridisation to arrays: the influence of dangling ends on duplex yield. *Nucleic Acids Res.*, **22**, 1365–1367.
51. Sohail,M., Akhtar,S. and Southern,E.M. (1999) The folding of large RNAs studied by hybridization to arrays of complementary oligonucleotides. *RNA*, **5**, 646–655.
52. Mir,K.U. and Southern,E.M. (1999) Determining the influence of structure on hybridization using oligonucleotide arrays. *Nature Biotechnol.*, **17**, 788–792.
53. Tsukiyama-Kohara,K., Iizuka,N., Kohara,M. and Nomoto,A. (1992) Internal ribosome entry site within hepatitis C virus RNA. *J. Virol.*, **66**, 1476–1483.
54. Brown,E.A., Zhang,H., Ping,L.H. and Lemon,S.M. (1992) Secondary structure of the 5′ nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res.*, **20**, 5041–5045.
55. Ina,Y., Mizokami,M., Ohba,K. and Gojobori,T. (1994) Reduction of synonymous substitutions in the core protein gene of hepatitis C virus. *J. Mol. Evol.*, **38**, 50–56.
56. Smith,D.B. and Simmonds,P. (1997) Characteristics of nucleotide substitution in the hepatitis C virus genome: constraints on sequence change in coding regions at both ends of the genome. *J. Mol. Evol.*, **45**, 238–246.
57. Walewski,J.L., Keller,T.R., Stump,D.D. and Branch,A.D. (2001) Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame. *RNA*, **7**, 710–721.
58. Roussel,J., Pillez,A., Montpellier,C., Duverlie,G., Cahour,A., Dubuisson,J. and Wychowski,C. (2003) Characterization of the expression of the hepatitis C virus F protein. *J. Gen. Virol.*, **84**, 1751–1759.
59. Kanazawa,Y., Hayashi,N., Mita,E., Li,T., Hagiwara,H., Kasahara,A., Fusamoto,H. and Kamada,T. (1994) Influence of viral quasispecies on effectiveness of interferon therapy in chronic hepatitis C patients. *Hepatology*, **20**, 1121–1130.
60. Okada,S., Akahane,Y., Suzuki,H., Okamoto,H. and Mishiro,S. (1992) The degree of variability in the amino terminal region of the E2/NS1 protein of hepatitis C virus correlates with responsiveness to interferon therapy in viremic patients. *Hepatology*, **16**, 619–624.
61. Pawlotsky,J.M. (2000) Hepatitis C virus resistance to antiviral therapy. *Hepatology*, **32**, 889–896.
62. Schuster,P. (1997) Genotypes with phenotypes: adventures in an RNA toy world. *Biophys. Chem.*, **66**, 75–110.
63. Kaufman,R.J. (2000) The double-stranded RNA-activated protein kinase PKR. In Sonenberg,N., Hershey,J.W.B. and Mathews,M.B. (eds), *Translational Control of Gene Expression*. Cold Spring Harbor laboratory press, Cold Spring Harbor, NY, pp. 503–527.
64. Vyas,J., Elia,A. and Clemens,M.J. (2003) Inhibition of the protein kinase PKR by the internal ribosome entry site of hepatitis C virus genomic RNA. *RNA*, **9**, 858–870.