Conserved RNA Secondary Structures in Flaviviridae Genomes

Caroline Thurner[†], Christina Witwer[†], Ivo L. Hofacker[†,*], and Peter F. Stadler[†,‡,§]

[†]Institut für Theoretische Chemie und Molekulare Strukturbiologie Universität Wien, Währingerstraße 17, A-1090 Wien, Austria

[‡]Bioinformatik, Institut für Informatik, Universität Leipzig, Kreuzstraße 7b, D-04103 Leipzig, Germany

[§]The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

[*]Address for correspondence:

Ivo L. Hofacker, Institut für Theoretische Chemie und Molekulare Strukturbiologie Universität Wien, Währingerstraße 17, A-1090 Wien, Austria
Tel: ++43 1 4277 52738, Fax: ++43 1 4277 52793
Email: {caro,ivo}@tbi.univie.ac.at, peter.stadler@bioinf.uni-leipzig.de

Abstract.

We report here on a comprehensive computational survey of evolutionarily conserved secondary structure motifs in the genomic RNAs of the family *Flaviviridae*. This virus family consists of the three genera *Flavivirus*, *Pestivirus*, *Hepacivirus* and the group of *Hepatitis G Viruses* with a currently uncertain taxonomic classification. Based on the control of replication and translation, two subgroups were considered separately: The genus Flavivirus with its type I cap structure at the 5' untranslated region (UTR) and a highly structured 3'UTR, and the remaining three groups that exhibit translation control by means of a Internal Ribosomal Entry Site (IRES) in the 5'UTR and a much shorter less structured 3'UTR. The main findings of our survey are: strong evidence for the possibility of genome cyclization in Hepatitis C and G viruses in addition to the Flaviviruses; a surprisingly large number of conserved RNA motifs in the coding regions; and a lower level of detailed structural conservation in the IRES and 3'UTR motifs than reported in the literature. An electronic atlas organizes the information on the more than 150 conserved, and therefore putatively functional, RNA secondary structure elements.

Keywords. Flaviviridae, RNA secondary structure, Internal Ribosomal Entry Site (IRES), Genome Cyclization.

1. Introduction

Viral RNA genomes not only code for proteins but in many instances carry RNA motifs that play a crucial role in the viral life cycle. Well known examples are the internal ribosomal entry sites (IRES), the RRE motif in HIV, or the CRE hairpin in Picornaviridae. The detection of such functional motifs in viral genome is a difficult task because almost all RNA molecules form secondary structures and functional structures are not significantly different from the structures formed by random sequences (Fontana *et al.*, 1993; Rivas and Eddy, 2000).

RNA secondary structures have been shown to be very sensitive to mutations (Fontana *et al.*, 1993; Schuster *et al.*, 1994): mutations in about 10% of the sequence positions already leads almost surely to unrelated structures if the mutated positions are chosen randomly. Secondary structure elements that are consistently present in a group of sequences with less than, say 95%, average pairwise identity are therefore most likely the result of stabilizing selection, not a consequence of the high degree of sequence homology. This fact can be exploited to design algorithms, briefly outlined in section 2 that reliably detect conserved RNA secondary structure elements in a small sample of related RNA sequences (Hofacker *et al.*, 1998; Hofacker and Stadler, 1999). This method was recently applied quite successfully to a survey of the genomes of Picornaviridae (Witwer *et al.*, 2001) and the RNA pre-genome of Hepadnaviridae (Stocsits *et al.*, 1999).

In this contribution we report a comprehensive survey of the family *Flaviviridae* which possess a single stranded positive sense RNA (ss+RNA) genome. The family is subdivided into the three genera *Flavivirus*, *Pestivirus*, *Hepacivirus* and the group of *Hepatitis G Viruses* with a currently uncertain taxonomic classification (van Regenmortel *et al.*, 2000). The RNA genome, which has a size of $9.6 - 12.3$kb is characterized by a similar organization (Fig. 1) in all genera and acts as the only mRNA found in infected cells. It contains one single long open reading frame flanked by a 5' and 3' untranslated region (UTR). These are known to form into specific secondary structures required for genome replication and translation. Viral proteins are synthesized as one single polyprotein, which is co- and posttranslationally cleaved by viral and cellular proteinases.

Based on the control of replication and translation it is useful to consider two subgroups of Flaviviridae: The first group is formed by the genus Flavivirus and is characterized by a type I cap structure at the 5'UTR and a highly structured 3'UTR. In this group there is evidence that the 5' and 3' ends stack together to cause a cyclization of the genome (sometimes referred to as "panhandle structure") which might be an important feature for RNA-replication (Hahn *et al.*, 1987; Khromykh *et al.*, 2001).

The second group consisting of Hepatitis C Virus (HCV), Pestivirus (PESTI), and Hepatitis G Virus (HGV), controls translation by means of an *Internal Ribosomal Entry Site* (IRES) in the 5'UTR and has a short, less structured 3'UTR. Pestivirus and Hepacivirus have very similar IRES regions (Pestova *et al.*, 1998); the IRES of hepatitis G virus is 50% longer and structurally quite different (Simons *et al.*, 1996). In this contribution we treat these two groups separately.

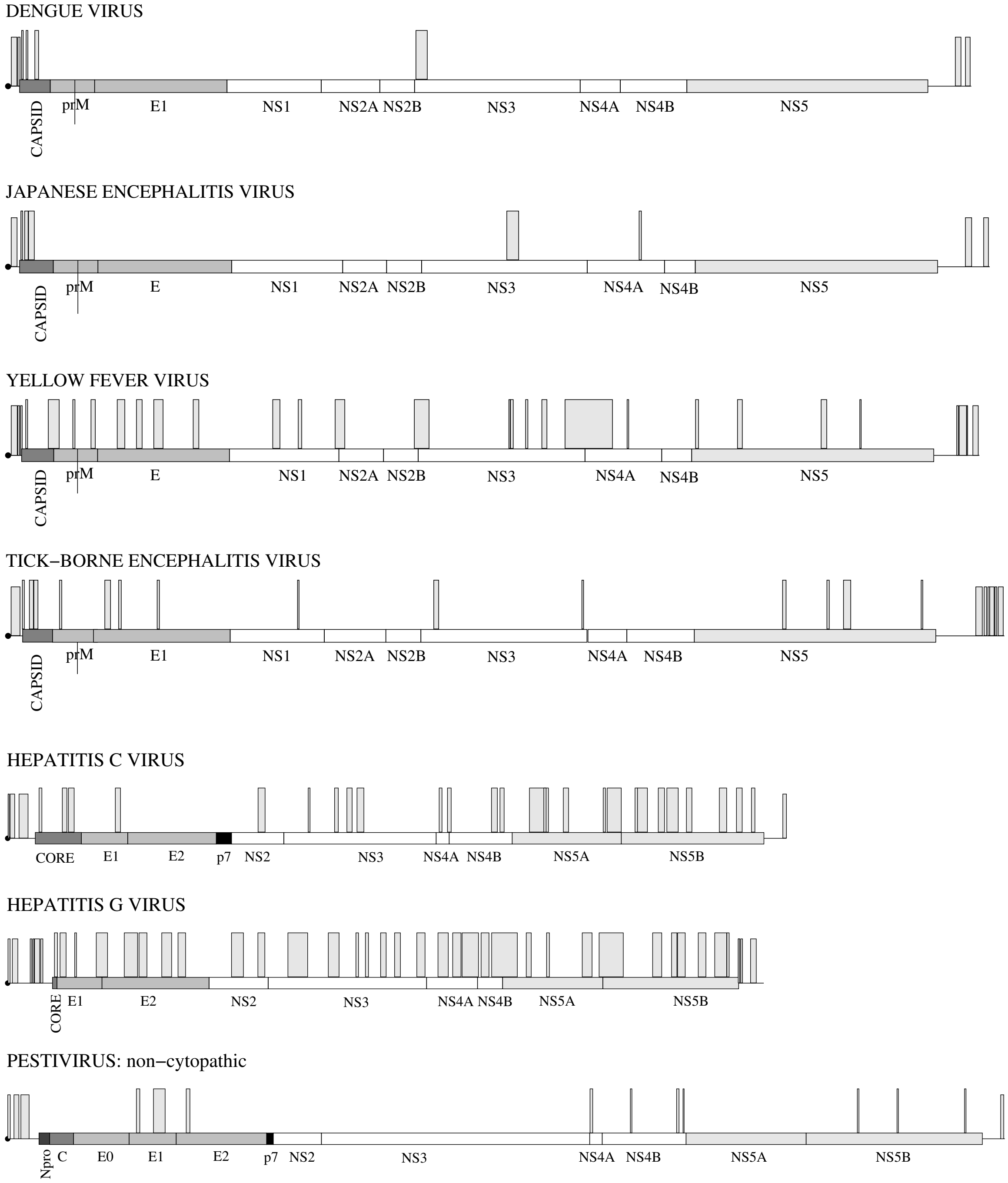


Figure 1. Genome map for the Flavivirus species Dengue Virus, Japanese Encephalitis Virus, Yellow Fever Virus and Tick-borne Encephalitis Virus and the genera Hepatitis C Virus, Hepatitis G Virus and Pestivirus. Putative conserved secondary structures are indicated by the boxes above the RNA sequence.

In general, the 5' and 3'UTRs of Flaviviridae genomic RNA are fairly well understood. Very little is known, however, about the secondary structures of the coding regions despite some evidence that the coding region might also contain functionally RNA motifs (Simmonds and Smith, 1999; Tuplin *et al.*, 2002).

In the survey reported here we found many putative structural elements, as indicated in Fig. 1. A complete description of each one of them cannot be displayed in print because of space constraints, but see Fig. 6 for selected examples. The complete material including positional information, sequences, accession numbers, multiple sequence alignments, structure predictions, structure drawings, and information on the sequence covariation are available as supplemental material in electronic form in our Viral RNA Structure Database at `http://rna.tbi.univie.ac.at/virus/`. This website can also be used to retrieve the computational results for regions that we have not identified as structurally conserved. Unless noted otherwise, the names used to denote individual conserved helices follow the scheme used on the website.

2. Materials and Methods

Sequence data were obtained from the NCBI genome database. The phylogenetic distribution and the pairwise sequence similarity of the available data was such that the following groups of Flaviviridae could be investigated in details: the genera Pestivirus, Hepacivirus, the unclassified group Hepatitis G Virus, and some species of the genus Flavivirus. These are Dengue Virus (DEN), Japanese Encephalitis Virus (JEV), Yellow Fever Viruses (YFV) and Tick-borne Encephalitis Virus (TBE). Statistical information on these sequence data is compiled in Table 1.

The genus Pestivirus can be subdivided into two groups, the cytopathic Pestiviruses, which cause cell shrinkage, membrane blabbing and cell death, and the non-cytopathic ones. Cytopathic viruses develop from non-cytopathic viruses by RNA-recombination, resulting in genome duplicates, rearrangements, deletions and insertions (Myers and Thiel, 1996). Characteristically they have at least one additional copy of the NS3 protein, isolated from NS2/NS3 through Ubiquitin (Ub) or cIns insertions (Myers and Thiel, 1996; Tautz *et al.*, 1999). In this study we only use full-length genomes of non-cytopathic Pestiviruses, because insertions of Ub and cIns cause extended gaps in the multiple alignments that interfere with the analysis. On the other hand, a detailed study of cytopathic viruses and particularly the effects of the extended insertions on the secondary structure of Pestiviruses, was not possible because there were too few sequences available in public databases.

Multiple sequence alignments are calculated using `CLUSTAL W` (Thompson *et al.*, 1994). All sequence positions reported in the following refer to the multiple sequence alignments that are available as part of the supplemental material.

RNA genomes were folded in their entirety using McCaskill's partition function algorithm (McCaskill, 1990) as implemented in the `Vienna RNA Package` (Hofacker *et al.*, 1994), based on the energy parameters published in (Mathews *et al.*, 1999). The result of this computation is a matrix of base pairing probabilities P_{ij} for each potential base pair (i, j) of the genomic RNA.

The `alidot` algorithm (Hofacker *et al.*, 1998; Hofacker and Stadler, 1999) was used to search the base pairing probability matrix for conserved secondary structure patterns. This method requires an independent prediction of the secondary structure for each of the sequences and a multiple sequence alignment that is obtained without any reference to the predicted secondary structures. The algorithm ranks base pairs

Table 1. Number of analyzed sequences N, length of our alignments, and of 5' and 3'UTR, and mean pairwise sequence identities σ.

Group	N	length	σ	5'UTR	σ	IRES	coding region	3'UTR	σ
Group 1: Flavivirus									
DEN	16	10775	80.4	1..98	87.4		96..10288	10289..10775	85.4
JEV	17	10979	95.5	1..95	98.9		96..10394	10395..10979	95.6
YFV	7	10863	96.4	1..121	99.8		121..10351	10352..10863	91.7
TBE	6	11143	86.5	1..132	91.2		133..10374	10375..11143	69.8
Group 2: Pestivirus, Hepacivirus, and Hepatitis G Virus									
HGV	10	9397	89.8	1..556	94.2	45-556	557..9085	9086..9397	96.7
HCV	9	9679	87.1	1..342	98.3	43-354	343..9399	9400..9679	85.1
PESTI	11	12393	74.9	1..388	80.7	65-388	389..12114	12115..12393	62.0

using both the thermodynamic information contained in the base pairing probability matrix and the information on compensatory, consistent (e.g. GC $\rightarrow$ GU), and inconsistent mutations contained in the multiple sequence alignment. The approach is different from efforts to simultaneously compute alignment and secondary structures (Corodkin *et al.*, 1997; Sankoff, 1985) and from programs such as `construct` (Lück *et al.*, 1996, 1999) and `alifold` (Hofacker *et al.*, 2002) because it does not assume that the sequences have a single common structure. An implementation of this algorithm is available from `http://www.tbi.univie.ac.at/RNA/`. The `alifold` algorithm (Hofacker *et al.*, 2002) is used to obtain consensus structures of regions with significant structural conservation.

Computational results are shown as Hogeweg-style mountain plots (Hogeweg and Hesper, 1984) with color codes indicating sequence covariations, see the caption of Fig. 3 for details.

3. Results I: Genus Flavivirus

The genus Flavivirus is a wide spread genus containing very diverse species. We focused on the species of Dengue Virus (DEN), Japanese Encephalitis Virus (JEV), Yellow Fever Virus (YFV) and Tick-borne Encephalitis Virus (TBE) since at least 6 sufficiently diverged genomic sequences are necessary for our analysis method. The genus Flavivirus differs from other members of the family Flaviviridae in that its 5'UTR is obviously shorter and carries no IRES function but a type I cap-structure (Brinton and Dispoto, 1988). On the other hand the 3'UTR is much longer and seems to be involved in replication processes of the viral genome (Proutski *et al.*, 1999). In Fig. 2 we show an overview of the conserved secondary structure elements of the 5' and 3' ends and the cyclization domains called P1', P1, P2, and CS or CS"A", respectively, we found by applying our algorithms.

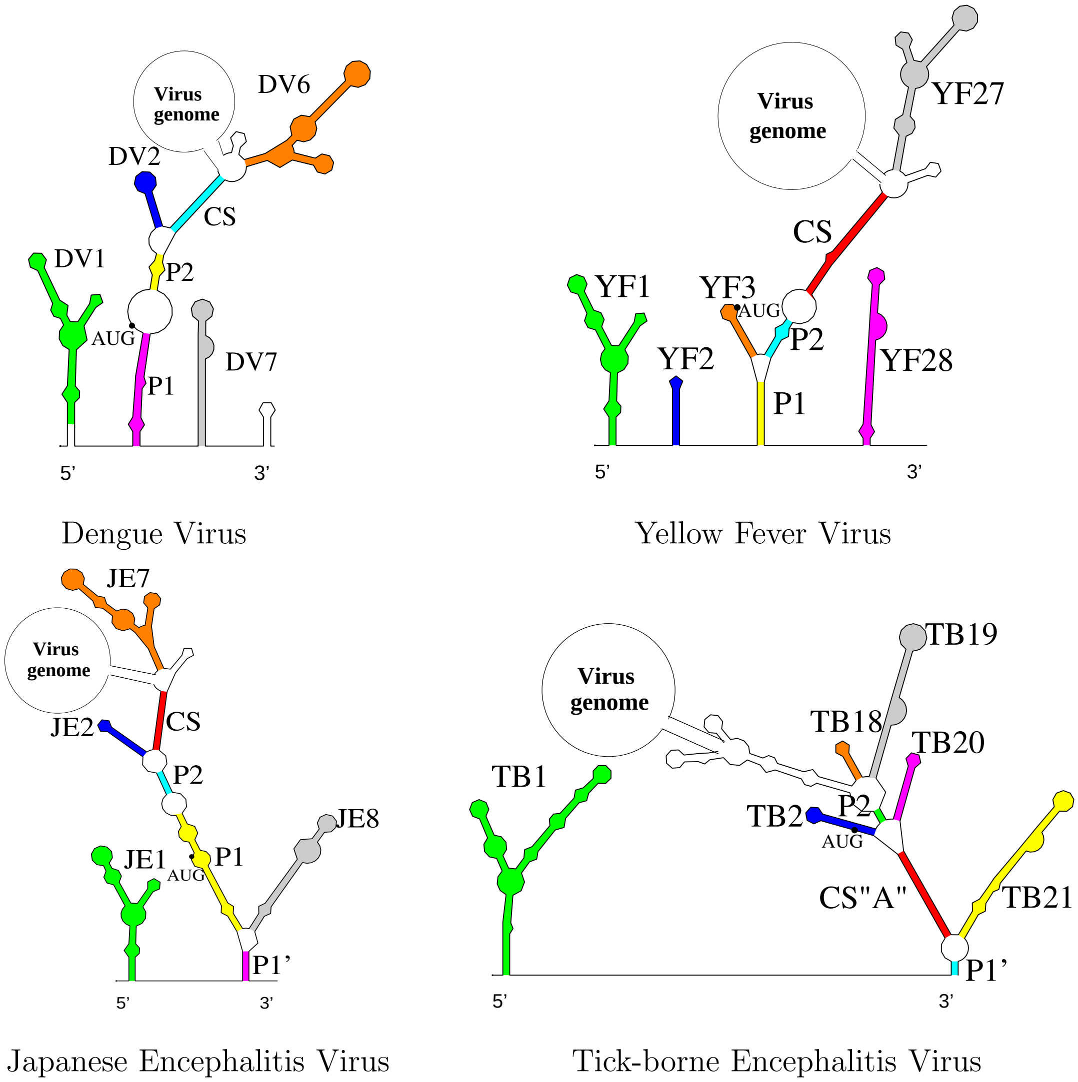


Figure 2. The minimum free energy structure of one sequence of the respective virus species is represented. Colored backgrounds mark regions which our folding algorithm and selection criteria allowed for all sequences. The nomenclature of the structures correspond to the website atlas of structures, see `rna.tbi.univie.ac.at`. Conserved secondary structures where 5' and 3'UTRs are involved in genome cyclization are called P1', P1, and P2; CS and CS"A" are taken from Hahn *et al.* (1987) and Khromykh *et al.* (2001) respectively. The exact positions of these structure elements are given in Supplemental Material B.

3.1. Genome Cyclization

Hahn found complementary sequences (cyclization sequences "CS") close to the 5' end and the 3' end of the genome and concluded that the two ends of the genome of Flavivirus stick together in a panhandle-like structure (Hahn *et al.*, 1987). Recently, it has been shown that RNA synthesis *in vitro* requires both 5' and 3' ends present, either connected in the same RNA sequence, or added in trans (You and Padmanabhan, 1999). Another piece of evidence for the cyclization of the genomic RNA is the finding that the 1st stack in 5'UTR and the last stack in 3'UTR together with the cyclization sequences (CS) are necessary and sufficient for virus translation and replication (Khromykh *et al.*, 2001).

The mean pairwise sequence identity of all four species of the genus Flavivirus (less than 50%) was too small to yield good alignments. The species Tick-borne Encephalitis Virus differs most from the other species in both, sequence and structure. From an alignment of the remaining species, DEN, JEV and YFV, we obtained a common structure for the CS, see Fig. 2, which confirms the prediction of Hahn *et al.* (1987). We then compared only DEN and JEV. These two species shared an extended cyclization sequence as shown already by Khromykh *et al.* (2001). In our data the CS contained no sequence variation but was predicted with pair probabilities close to 1. Adjacent to the CS we found a further stack which participates in genome cyclization and which contains several sites of sequence variation (see P2 in Supplemental Material A). Between CS and P2 there is a well conserved hairpin structure supported by numerous compensatory mutations (see DV2/JE2 in Supplemental Material A).

Tick-borne Encephalitis Virus. The conserved cyclization motifs first reported by Hahn *et al.* (1987) for mosquito-borne viruses are absent in other strains. Putative CSs were proposed for Powassan virus RNA (Mandl *et al.*, 1993) and for TBE virus and Cell fusing Agent (CFA) (Khromykh *et al.*, 2001).

In all proposed motifs for genome cyclization again we did not find any mutations in sequences, thus we could not use our method to confirm the predicted structure by means of sequence covariation. Thermodynamic folding, however, provided strong evidence for the CS“A”-motif in Fig. 2 (Khromykh *et al.*, 2001; Mandl *et al.*, 1993) because these base pairs appeared with probabilities close to 1 in the folds of the complete genome. Khromykh’s region CS“B” was folded only by one single sequence (TEU27491) and thus could not be considered as a common motif for all members of the species TBE.

3.2. 5’UTR

Flaviviruses carry a common type I cap structure and thus exhibit no IRES region. Nevertheless we found well-conserved secondary structures in the 5’UTRs. The 5’UTRs of DEN, JEV and YFV formed into a very similar secondary structure, while the structure for TBE turned out significantly different, see Fig. 3, DV1, JE1, YF1 and TB1 respectively. For DEN we found structural conservation, while the sequences of JEV and YFV were very conserved. A manually improved alignment of JEV and YFV to DEN of this region showed that there was a significant structure conservation among all three genera. Furthermore all structures contained an internal loop of one single U at the 5’ and two Us at the 3’ strand of the stem (data not shown).

For Dengue Virus a structure similar to DV1 is proposed by Leitmeyer *et al.* (1999) and Khromykh *et al.* (2001). A stem-loop structure from positions 80-105 reported in Leitmeyer *et al.* (1999) has an absolutely conserved sequence but is predicted by `RNAfold`.

A stem carrying the initiator `AUG` proposed by Hahn *et al.* (1987) for Yellow Fever Virus was not conserved among the available sequences for this species.

The 5'UTR structure proposed by Khromykh *et al.* (2001) for TBE was inconsistent with the available sequence data. We found a quite different structure that was confirmed by several mutations, both consistent and compensatory.

3.3. Coding Region

Several conserved secondary structures were found in the coding regions of DEN, JEV, YFV and TBE. The structures are available on the web site. So far, no functions have been proposed for these regions. Stem-loop DV2 was proposed already by Hahn *et al.* (1987) for Den-2 Virus.

3.4. 3'UTR

Similar structures were found in all four species, Fig. 3. Sequence variation in the stem DV6a was high in DEN, and present in JEV. For YFV we did not find a structure comparable to the whole of DV6 or JE7, but only a single stem corresponding to the the stem loop DV6a and JE7a.

Structures similar to DV6, JE7, YF27 or TB19 were also proposed by Hahn *et al.* (1987) for DEN 2 and YFV, by Khromykh *et al.* (2001) for DEN, YFV, JEV, and TBE by Rauscher *et al.* (1997) (B for DEN, YFV, and JEV and I, II and III for TBE), by Proutski *et al.* (1999) (TL1/RCS2 or TL2/CS2 for DEN and JEV, and "stem loop 1 in subregion I" for YFV), and by Leitmeyer *et al.* (1999) for DEN.

Dengue virus. For the DEN 3'UTR we found the same structures as by Rauscher *et al.* (1997) where the analysis was restricted to the isolated 3'UTR. None of the long-range interactions interfered with any of these structural motifs. Leitmeyer *et al.* (1999) proposes additional base pairings that we could not find, because they conflicted with the Cyclization Domains.

We found only parts of the secondary structures proposed by Proutski *et al.* (1997) for DEN2 as conserved for all DEN species. In particular we did not find his structures I2 and I3, II1 and III except region 3'LSH (our DV7). DV6 and DV7 are also discussed by Proutski *et al.* (1999) for DEN4. All other structures reported in that study are disrupted by the cyclization of the viral genome. Assuming that cyclization of the genome is vital we can re-interpret the deletion studies reported by Proutski *et al.* (1999) in the following way: Deletion of DV6a (TL2) yields a delayed and reduced growth in simian and mosquito cells. When the deletions extended more to the 3' end of the sequence the CS region is destroyed [Proutski's mutant 3'172-83], hence no viable viruses are found. A non-viable mutant 3'172-107 may be explained by the importance of the sequence motif CAAAAA for virus propagation (Men *et al.*, 1996). Our data indicate that in this case the sequence motif rather than any structure associated with it is important. In the mutants 3'd333-183 and 3'384-183 Proutski measures a greatly delayed and reduced growth in living cells. We would argue that these deletions destroy a possible prolongation of the cyclization region that we found for Dengue viruses (data not shown). Our data indicate that each sequence allows additional stems for cyclization in this region even though their exact positions vary

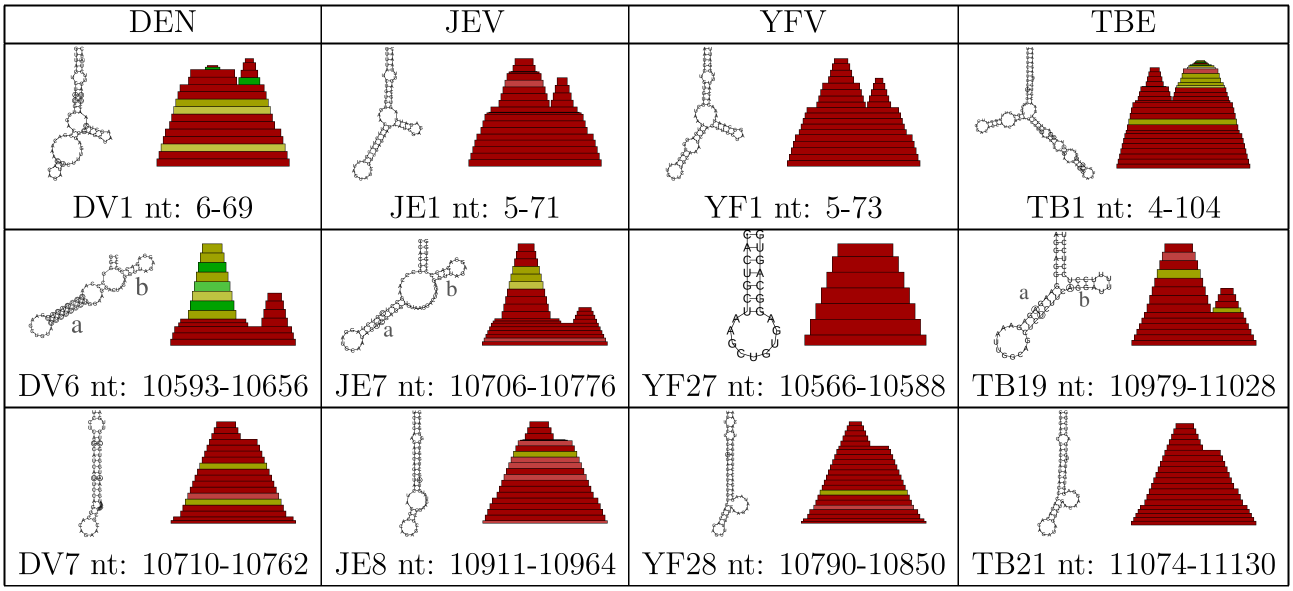


Figure 3. Conserved secondary structures of the flavivirus species DEN, JEV, YFV and TBE in the 5'UTR (first raw) and 3'UTR (second and third raw).
Mountain plots (Hogeweg and Hesper, 1984) faithfully represent secondary structures: each base pair (i, j) is represented by a slab ranging from position i to j; its height is proportional to the base pairing probability in thermodynamic equilibrium, computed with McCaskill's algorithm. Colors indicate the number of different types of base pairs (red 1, ocher 2, green 3, turquoise 4, blue 5, violet 6). Saturated color indicates that all sequences can form the base pair, while two levels of pale color mean that 1 or 2 input sequences have non pairing bases at positions i and j. If there are more than 2 non-compatible sequences the pair is not displayed.
In the conventional drawings, consistent and compensatory mutations are indicated by circles around bases that have mutations. Gray letters indicate inconsistent mutations.

slightly. It is plausible that such an extended cyclization region adds to the efficiency of viral replication but is not necessarily essential for its viability.

Yellow Fever Virus and **Japanese Encephalitis Virus.** The sequences in our data set had about $\sigma = 91.7\%$ pairwise homology in the 3' non translated region. We observed only a small number of compensatory mutations to verify structural features predicted based on our thermodynamic algorithm. We found essentially the same structures as Rauscher *et al.* (1997); again none of the structures reported by Rauscher *et al.* (1997) conflicted with CS regions. YF28 was shorter by 9 base pairs than reported by Hahn *et al.* (1987). YF28 and YF27 corresponded to 3'LSH and I1 respectively, JE7 and JE8 to 3'LSH and II2 respectively as proposed by Proutski *et al.* (1997). More structures could not be found for similar reasons as explained for DEN above.

Tick borne encephalitis virus. We recovered the structures III, VI, VII, VIII, and IX as reported by Mandl *et al.* (1998) and Rauscher *et al.* (1997). Structure A1 was shorter because of conflicts with cyclization sequences P4 and P3. Structure A2 did not seem to be conserved. For structure MS we had evidence from thermodynamic folding. However, this structure would form a pseudo-knot with P2. Structure V

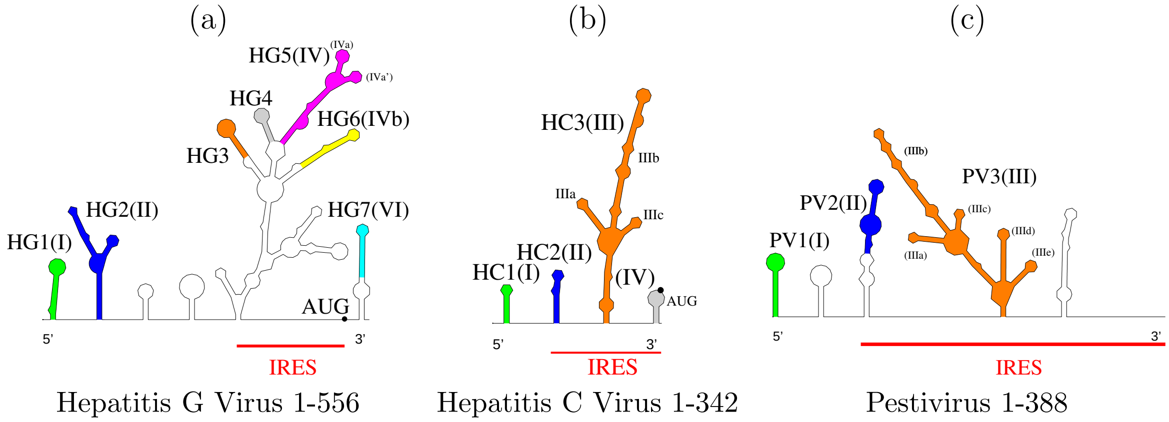


Figure 4. Schematic illustration of 5'UTRs of HGV, HCV and Pestivirus. Conserved structures are discussed in the text. Notations in brackets correspond in (a) to (Simons *et al.*, 1996), (b) to (Honda *et al.*, 1996a) and (c) to (Brown *et al.*, 1992).

conflicted with P2 and was therefore not predicted from the complete genome folds. The structure we found here differed slightly to what was proposed by Proutski *et al.* (1997).

4. Results II: Pestivirus, Hepacivirus, and Hepatitis G Virus

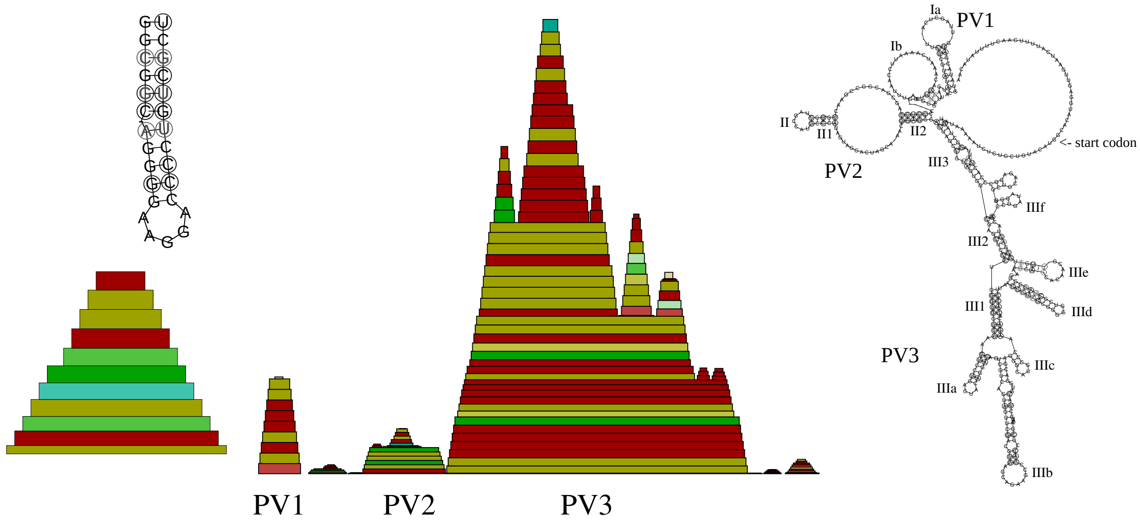
4.1. The 5'UTR

The 5'UTRs of these virus groups contain an IRES. We found that the RNA structure was much less conserved than what one would expect from the published literature; an overview is given in Fig. 4. This was consistent with the data reported by Witwer *et al.* (2001) for Picornaviridae. Surprisingly, there was a large amount of variation in the IRES structures of HGV and HCV, while the IRES structure of Pestivirus turned out to be highly conserved.

The IRES structures of HCV and Pestivirus shared a common overall structure despite the fact that they were not comparable at the sequence level. Nevertheless, they shared a few significant details: the IIIa stem carried a completely conserved loop sequence and stem IIIc was conserved in its sequence.

Hepatitis G Virus. The 5'UTR sequences of HGV were much higher conserved than the rest of the genome (Tab. 1). The mutations in this part of the region were concentrated to nucleotide (nt) positions 410 to 437, which comprised the structure element HG6 (IVb), Fig. 5(a). This motif was predicted also in previous studies (Simons *et al.*, 1996; Smith *et al.*, 1997).

We found a stack HG2 (FIG. 4(a)) which was shorter and more shifted to the 5' end of the IRES than stack II reported by Simons *et al.* (1996). Our prediction was supported by compensatory mutations (data not shown). The reason for the discrepancy was the formation of a panhandle-like structure by means of a base pairing interaction from nt 163-175 with nt 9213-9201 (see section 4.4., “Genome Cyclization”).



(a) (b)

Figure 5. (a) HGV: 5'UTR nt: 410-437, IRES conserved element HG6(IVb) (b) Pestivirus 5'UTR nt:1-420; the IRES is supposed to begin with stack PV2(II).

The sequences were too conserved in the remainder of the 5'UTR for to support predicted structures by means of sequence variation. The thermodynamic folding algorithm, however, found partly structures similar to those proposed previously (Katayama *et al.*, 1998; Simons *et al.*, 1996; Smith *et al.*, 1997).

Hepatitis C Virus. The 5'UTR of HCV comprises 324-341 nucleotides. The `RNAfold` algorithm recovered structures similar to those reported in previous studies (Collier *et al.*, 2002; Honda *et al.*, 1996b; Kalliampakou *et al.*, 2002; Kieft *et al.*, 2001; Kolupaeva *et al.*, 2000a; Odreman-Macchioli *et al.*, 2000; Psaridi *et al.*, 1999; Spahn *et al.*, 2001; Tang *et al.*, 1999), see Fig. 4(b).

Due to high sequence conservation (Tab. 1) we found only two sites with compensatory mutations in HC3 (called IIIa, b, and c by Honda *et al.* (1996b)) in our data set of 9 complete genomic sequences. If additional sequences of the IRES region were included in the analysis the structure was well supported by compensatory mutations. This structure HC3 has received considerable attention since it appears to act as binding site for the eIF3-40S complex. It has an internal loop which is twisted in itself (Collier *et al.*, 2002). Even though we had a mean homology of 98.3% in this region, there were two compensatory mutations just before and after this highly structured part of the HCV IRES. This confirms Collier's interpretation that here the shape of the backbone rather than the sequence composition is important for translation initiation.

We found a stem HC2 which corresponds to IIa proposed by Honda *et al.* (1999). The remaining nucleotides of stem II were more likely involved in long range interactions with nucleotides 8571 to 8552 (NS5B). When the isolated IRES region (i.e. nt 44-357) was folded separately, stem II was recovered.

Pestivirus. As with HCV and HGV the 5'UTR region was more conserved than the rest of the genome, Tab. 1.

In stack PV1 (Ia in (Brown *et al.*, 1992), domain A in (Deng and Brock, 1993)) we found two consistent and two compensatory mutations, Fig. 5(b). The inconsistent mark in the first base pair of PV1 was due to a gap in the first position of the alignment in sequence AY072924.

Fletcher and Jackson (2002) observed that a deletion of nucleotides comprising stack PV2 (II in (Brown *et al.*, 1992), domain C in (Deng and Brock, 1993)) produces a decrease of IRES activity to 19%. Though the pair probabilities in stack PV2 were small, see the mountain-plot of Fig. 5(b), we found no inconsistencies and a considerable amount of compensatory mutations. This might point out the importance of the structure rather than the sequence to IRES function in this region.

As in previous studies (Deng and Brock, 1993; Fletcher and Jackson, 2002; Kolupaeva *et al.*, 2000b; Moser *et al.*, 2001) our method detected stem PV3 as an important feature of Pestivirus IRES structure.

4.2. Coding region

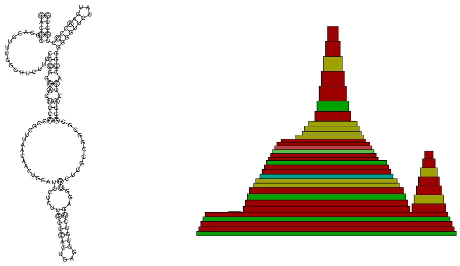
Hepatitis G Virus. We found two significant conserved stacks (HG9 and HG10) in the E1 region, which were proposed previously by Simmonds and Smith (1999) based on a different algorithm (data presented in the supplemental material).

Conserved secondary structures seemed to be concentrated in the NS5A and NS5B region of the Hepatitis G virus genome, Fig. 6. Some of these were proposed already by Cuceanu *et al.* (2001), Fig. 6(c) and (d). Furthermore we found HG38 as Cuceanu's $SL_{NS5B}V$ and HG39 as $SL_{NS5B}IV$. In our data Cuceanu's $SL_{NS5B}I$ motif was completely conserved in the sequence, in $SL_{NS5B}VI$ we found more inconsistent than compensatory mutations (data not shown), and the proposed $SL_{NS5B}VII$ structure could not be found with our method.

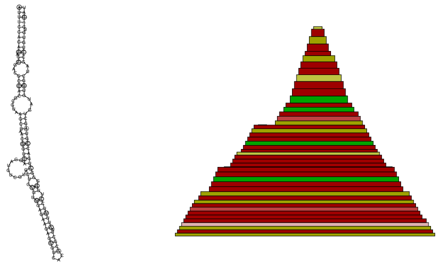
Hepatitis C Virus. Again we found most of the conserved structures in NS5A and NS5B regions. Some of these have been reported previously as important for the efficiency of the IRES function (Tuplin *et al.*, 2002; Zhao and Wimmer, 2001). One of the motifs detected in this study is HC4, shown in Fig. 6(e). Tuplin further finds HC6 as SL443, HC27 as SL8828 and HC28 as SL9011. According to our data there was no evidence for the existence of SL7730 and SL9118. Tuplin's SL8926 showed too many inconsistencies in our data, and SL8376 was not folded because of interactions of this region with the 3'UTR (see section 4.4., "Genome Cyclization").

Ray *et al.* (1999) argues that the HCV persistence is associated with sequence variability in putative envelope genes E1 and E2. We found a conserved RNA structure, HC7 in Fig. 6(f), in this region.

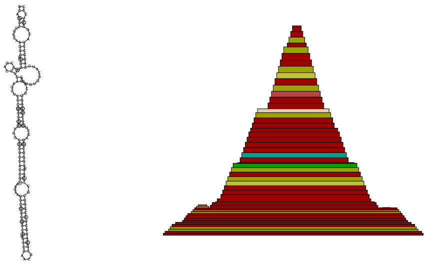
Pestivirus All putative conserved secondary structure elements in the coding region of Pestivirus were very short. A stem loop downstream of the initiator AUG that is important for translation initiation has been reported by Myers *et al.* (2001). In our data this structure had too many inconsistencies.



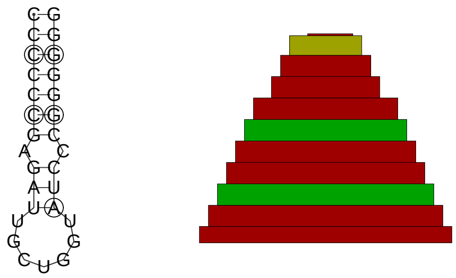
(a) HGV HG13; E2 nt: 1915 to 2038



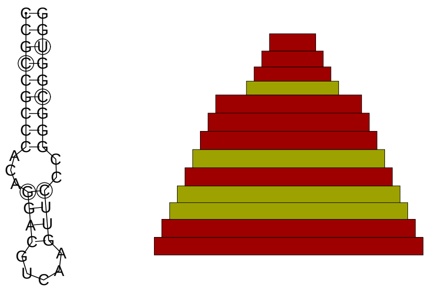
(b) HGV HG31; NS5A nt: 7141 to 7263



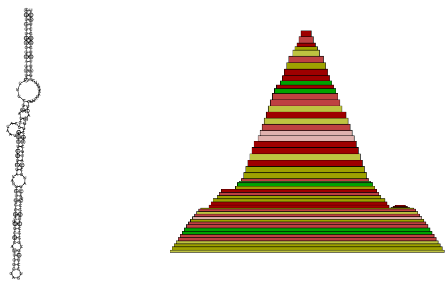
(c) HGV HG40; NS5B nt: 8788 to 8937



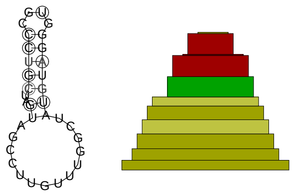
(d) HGV HG41; NS5B nt: 8939 to 8967



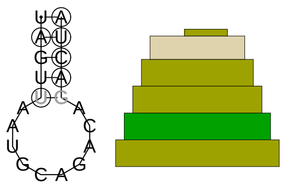
(e) HCV HC4; core nt: 390 425



(f) HCV HC7; E1 nt: 1338 1472



(g) Pestivirus PV8; NS4A nt: 7238 7271



(h) Pestivirus PV14; NS5B nt: 11892 11911

Figure 6. Conserved secondary structures in the coding region of HGV, HCV and Pestivirus. (a) and (b) new structures we propose among others for HGV. Structures proposed by Cuceanu *et al.* (2001) as: (c) HG40 (SL_{NS5B}III) and (d) HG41 (SL_{NS5B}II). (e) Structure proposed by Tuplin *et al.* (2002) for HCV HC4 as SL47. (f) Structure we propose in HCV's coding region. In (g) and (h) we propose Conserved secondary structures for Pestivirus coding region.

4.3. 3'UTR

Hepatitis G Virus. Similar to the 5'UTR, the 3'UTR of Hepatitis G Virus is highly conserved ($\sigma = 96.7\%$). Not surprisingly, we found structures partly similar to those reported priviously (Katayama *et al.*, 1998; Okamoto *et al.*, 1997; Xiang *et al.*, 2000) but we could not confirm all of them based on sequence co-variation (data not shown). Some of the structures proposed therein were not found because of long range interactions with the 5'UTR (see section 4.4. "Genome Cyclization"). HG43 was proposed also by Cuceanu *et al.* (2001) and Xiang *et al.* (2000) to be conserved and turned out to be well supported also by our data.

Hepatitis C Virus. The 3'UTR consists of a short sequence of variable length and composition (variable region), an U rich stretch (poly-U-UC region) variable in its length and a highly conserved sequence of approximately 100 nucleotides at the 3' end (conserved region, X-tail) (Kolykhalov *et al.*, 1996; Tanaka *et al.*, 1996; Yamada *et al.*, 1996). Within this X-tail we found only a single mutation (which is compatible with the predicted structure). Our stem HC29 corresponded to SL1 as reported previously (Blight and Rice, 1997; Ito and Lai, 1997; Yamada *et al.*, 1996). Ito's and Blight's stems SL2 and SL3 competed in our data with the formation of two long range interactions LR1 and LR2, see section 4.4. "Genome Cyclization". The probability of base pairs in LR1 was around $p = 0.54$, which was significantly higher than HC29. The elements SL2 and SL3 were thermodynamically unfavorable in the genomic context and could only be detected when a sequence window was used that was too small to contain the long-range interactions.

Pestivirus. Pestiviruses are very inhomogeneous in their 3'UTR region, due to extended AU rich insertions in some strains. The only RNA feature that was shared among all available sequences is the terminal stem PV15 that was originally described by Deng and Brock (1993), see also (Becher *et al.*, 1998; Yu *et al.*, 1999).

4.4. Genome Cyclization

Surprisingly, we discovered strong evidence for genome cyclization not only in the genus of Flavivirus, where this effect has already been described in literature, but also within the Hepatitis C and G viruses. The most prominent of them are shown in Fig. 7.

In HGV genome cyclization is localized to base pairings between nt 33-48 with nt 9367-9353 (pair probabilities ≈ 0.6), nt 128-140 with nt 9224-9214 and nt 163-175 with nt 9213-9201 see Fig. 7 (pair probabilities ≈ 0.7). These domains are very conserved in sequence. We found only one consistent mutation in the base pair (42,9357). On the other hand there was one sequence carrying an inconsistent mutation at base pair (130,9222).

In HCV putative cyclization domains comprised base pairs of nt 1-3 with nt 8627-8625, 88-92 with 8602-8606 and 95-110 with 8556-8571. The last one had two sites of compensatory mutations, see Fig. 7. Here nucleotides from the IRES region paired with nucleotides within the coding region for the protein NS5B. At the same time we observed two regions of the 3'UTR to fold forward to the NS5B region as well:

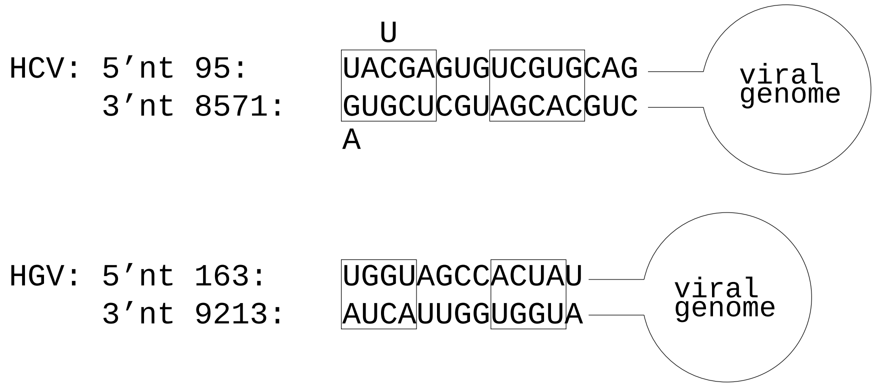


Figure 7. Putative cyclization regions in HCV and HGV genomes. The boxed areas point out sequences that might be read as palindrome sequences and maybe play a functional role in replication processes.

(i) LR1: nt 8628-8661 (NS5B) paired with nt 9623-9599 (3’UTR) and (ii) LR2: nt 8989-9007 (NS5B) with nt 9583-9598 (3’UTR). This brought 5’ and 3’ regions into very close proximity.

5. Discussion

We have employed a combination of structure prediction based on thermodynamic rules and the evaluation of consistent and compensatory mutations to search Flaviviridae genomes for functional RNA structure motifs. While the UTRs of some of these viruses have been studied previously, this contribution reports a comprehensive survey of structural features across the full genomes of the whole family Flaviviridae.

Furthermore, instead of using a “sliding window” technique, all predictions were carried out for the complete genomic RNA sequences. This enables our algorithm to find long range interactions, in particular we found significant probability for cyclization in all genera except Pestivirus.

In the genus Flavivirus a cyclization of the genome was already described in the literature and localized to very conserved cyclization sequences. Apart from recovering these known cyclization sequences, we detected further sequences which took part in cyclization for all species in this study. These sequences varied considerably in sequence, length and position. (Proutski *et al.*, 1999) showed, that deletion of these sequences lead to a greatly delayed and reduced growth in simian and mosquito cells. It is possible that these additional cyclization domains are not strictly necessary for virus viability, but only support and stabilize viral genome cyclization.

Most surprisingly we found viral genome cyclization also in HGV and HCV, which had not been reported before. Our algorithm made out base pair probabilities for both, previously reported secondary structures in 5’ and 3’UTRs as well as for genome cyclization. For both cases, our data revealed no inconsistencies. Thus known structures compete with genome cyclization. Our evaluation conditions favored genome cyclization based on both, thermodynamic prediction and, in the case of HCV, even sequence variation. This result can be interpreted either as a relict of ancient ancestors between these genera and the genus Flavivirus or, more speculatively, as a switch providing different functions in different states of the viral life cycle (e.g. a switch between replication and translation states of the virus).

While in Flavivirus and HGV the 5' and 3' end pair within the untranslated regions, we found base pairing in HCV between the 5' end and the 3' end to a region some 1000 nt upstream of the 3' end (i.e. a region within the NS5B protein). Most interestingly we observed that in this way 5' and 3' ends were brought closely together. This could be a reason for the particular importance of the NS5B region as assumed in the literature.

Furthermore in this contribution (and in the supplementary material available online) we present a large number of secondary structure elements that have not been described before, most importantly within the coding region. This information could be used to identify additional regions which might be important for virus viability and propagation, and thus to gain more insight into the life-cycle of the members of the family Flaviviridae.

Acknowledgments

C.T. and C.W. are supported by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung*, Project No. P-13545-MAT. The work of PFS is supported in part by the DFG Bioinformatics Initiative, BIZ-6/1-2.

References

Becher P., Orlich M., and Thiel H.J., 1998. Complete genomic sequence of border disease virus, a pestivirus from sheep. *J. Virol.* **72**:5165–5173.

Blight K.J. and Rice C.M., 1997. Secondary structure determination of the conserved 98-base sequence at the 3' terminus of hepatitis C virus genome RNA. *J. Virol.* **71**:7345–7352.

Brinton M.A. and Dispoto J.H., 1988. Sequence and secondary structure analysis of the 5'-terminal region of flavivirus genome RNA. *Virology* **162**:290–299.

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. *Nucl. Acids Res.* **20**:5041–5.

Collier A.J., Gallego J., Klinck R., Cole P.T., Harris S.J., Harrison G.P., Aboul-Ela F., Varani G., and Walker S., 2002. A conserved RNA structure within the HCV IRES eIF3-binding site. *Nat. Struct. Biol.* **9**:375–380.

Corodkin J., Heyer L.J., and Stormo G.D., 1997. Finding common sequences and structure motifs in a set of RNA molecules. In *Proceedings of the ISMB-97*, editors Gaasterland T., Karp P., Karplus K., Ouzounis C., Sander C., and Valencia A., pages 120–123. AAAI Press, Menlo Park, CA.

Cuceanu N.M., Tuplin A., and Simmonds P., 2001. Evolutionarily conserved RNA secondary structures in coding and non-coding sequences at the 3' end of the hepatitis G virus/GB-virus C genome. *J. Gen. Virol.* **82**:713–722.

Deng R. and Brock K.V., 1993. 5' and 3' untranslated regions of pestivirus genome: primary and secondary structure analyses. *Nuc. Acids Res.* **21**:1949–1957.

Fletcher S.P. and Jackson R.J., 2002. Pestivirus internal ribosome entry site (IRES) structure and function: elements in the 5' untranslated region important for IRES function. *J. Virol.* **76**:5024–5033.

Fontana W., Konings D.A.M., Stadler P.F., and Schuster P., 1993. Statistics of RNA secondary structures. *Biopolymers* **33**:1389–1404.

Hahn C.S., Hahn Y.S., Rice C.M., Lee E., Dalgarno L., Strauss E.G., and Strauss J.H., 1987. Conserved elements in the 3'untranslated region of flavivirus RNAs and potential cyclization sequences. *J. Mol. Biol.* **198**:33–41.

Hofacker I.L., Fekete M., Flamm C., Huynen M.A., Rauscher S., Stolorz P.E., and Stadler P.F., 1998. Automatic detection of conserved RNA structure elements in complete RNA virus genomes. *Nucl. Acids Res.* **26**:3825–3836.

Hofacker I.L., Fekete M., and Stadler P.F., 2002. Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* **319**:1059–1066. SFI Preprint 01-11-067.

Hofacker I.L., Fontana W., Stadler P.F., Bonhoeffer S., Tacker M., and Schuster P., 1994. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* **125**:167–188.

Hofacker I.L. and Stadler P.F., 1999. Automatic detection of conserved base pairing patterns in RNA virus genomes. *Comp. & Chem.* **23**:401–414.

Hogeweg P. and Hesper B., 1984. Energy directed folding of RNA sequences. *Nucl. Acids Res.* **12**:67–74.

Honda M., Beard M.R., Ping L.H., and Lemon S.M., 1999. A phylogenetically conserved stem-loop structure at the 5' border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation. *J. Virol.* **73**:1165–1174.

Honda M., Brown E.A., and Lemon S.M., 1996a. Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *RNA* **2**:955–68.

Honda M., Ping L.H., Rijnbrand R.C., Amphlett E., Clarke B., Rowlands D., and Lemon S.M., 1996b. Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. *Virology* **222**:31–42.

Ito T. and Lai M.M.C., 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.

Kalliampakou K.I., Psaridi-Linardaki L., and Mavromara P., 2002. Mutational analysis of the apical region of domain II of the HCV IRES. *FEBS Lett.* **511**:79–84.

Katayama K., Fukushi S., Kurihara C., Ishiyama N., Okamura H., and Oya A., 1998. Full-length GBV-C/HGV genomes from nine japanese isolates: characterization by comparative analysis. *Arch. Virol.* **143**:1–13.

Khromykh A.A., Meka H., Guyatt K.J., and Westaway E.G., 2001. Essential role of cyclization sequences in flavivirus RNA replication. *J. Virol.* **75**:6719–6728.

Kieft J.S., Zhou K., Jubin R., and Doudna J.A., 2001. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* **7**:194–206.

Kolupaeva V.G., Pestova T.V., and Hellen C.U., 2000a. An enzymatic footprinting analyses of the interaction of 40S ribosomal subunits with the internal ribosomal entry site of hepatitis C virus. *J. Virol.* **74**:6242–6250.

Kolupaeva V.G., Pestova T.V., and Hellen C.U., 2000b. Ribosomal binding to the internal ribosomal entry site of classical swine fever virus. *RNA* **6**:1791–807.

Kolykhalov A.A., Feinstone S., and Rice C.M., 1996. Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA. *J. Virol.* **70**:3363–71.

Leitmeyer K.C., Vaughn D.W., Watts D.M., Salas R., de Chacon I.V., Ramos C., and Rico-Hesse R., 1999. Dengue virus structural differences that correlate with pathogenesis. *J. Virol.* **73**:4738–4747.

Lück R., Gräf S., and Steger G., 1999. ConStruct: A tool for thermodynamic controlled prediction of conserved secondary structure. *Nucl. Acids Res.* **27**:4208–4217.

Lück R., Steger G., and Riesner D., 1996. Thermodynamic prediction of conserved secondary structure: Application to the RRE element of HIV, the tRNA-like element of CMV, and the mRNA of prion protein. *J. Mol. Biol.* **258**:813–826.

Mandl C.W., Holzmann H., Kunz C., and Heinz F.X., 1993. Complete genomic sequence of Powassan virus: evaluation of genetic elements in tick-borne versus mosquito-borne Flaviviruses. *Virol.* **194**:173–184.

Mandl C.W., Holzmann H., Meixner T., Rauscher S., Stadler P.F., Allison S.L., and Heinz F.X., 1998. Spontaneous and engineered deletions in the 3' noncoding region of tick-borne encephalitis virus: construction of highly attenuated mutants of a flavivirus. *J. Virol.* **72**:2132–2140.

Mathews D.H., Sabina J., Zucker M., and Turner H., 1999. Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. *J. Mol. Biol.* **288**:911–940.

McCaskill J.S., 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**:1105–1119.

Men R., Bray M., Clark D., Chanock R.M., and Lai C.J., 1996. Dengue type 4 virus mutants containing deletions in the 3' noncoding region of the RNA genome: analysis of growth restriction in cell culture and altered viremia pattern and immunogenicity in rhesus monkeys. *J. Virol.* **70**:3930–3937.

Moser C., Bosshart A., Tratschin J.D., and Hofmann M.A., 2001. A recombinant classical swine fever virus with a marker insertion in the internal ribosome entry site. *Virus Genes* **23**:63–68.

Myers G. and Thiel H.J., 1996. Molecular characterization of pestiviruses. *Adv. Vir. Res.* **47**:53–118.

Myers T.M., Kolupaeva V.G., Mendez E., Baginski S.G., Frolov I., Hellen C.U., and Rice C.M., 2001. Efficient translation initiation is required for replication of bovine viral diarrhea virus subgenomic replicons. *J. Virol.* **75**:4226–4238.

Odreman-Macchioli F.E., Tisminetzky S.G., Zotti M., Baralle F.E., and Buratti E., 2000. Influence of correct secondary and tertiary RNA folding on the binding of cellular factors to the HCV IRES. *Nucl. Acids Res.* **28**:875–885.

Okamoto H., Nakao H., Inoue T., Fukuda M., Kishimoto J., Iizuka H., Tsuda F., Miyakawa Y., and Mayumi M., 1997. The entire nucleotide sequences of two GB virus C/hepatitis G virus isolates of distinct genotypes from japan. *J. Gen. Virol.* **78**:737–745.

Pestova T.V., Shatsky I.N., Fletcher S.P., Jackson R.J., and Hellen C.U., 1998. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev.* **12**:67–83.

Proutski V., Gould E.A., and Holmes E.C., 1997. Secondary structure of the 3' untranslated region of flaviviruses: similarities and differences. *Nucl. Acids Res.* **25**:1194–1202.

Proutski V., Gritsun T.S., Gould E.A., and Holmes E.C., 1999. Biological consequences of delitions within the 3'-untranslated region of flaviviruses may be due to rearrangements of RNA secondary structure. *Virus Research* **64**:107–123.

Psaridi L., Georgopoulou U., Varaklioti A., and Mavromara P., 1999. Mutational analysis of a conserved tetraloop in the 5' untranslated region of hepatitis C virus identifies a novel RNA element essential for the internal ribosome entry site function. *FEBS Lett.* **453**:49–53.

Rauscher S., Flamm C., Mandl C.W., Heinz F.X., and Stadler P.F., 1997. Secondary structure of the 3'-noncoding region of flavivrus genomes: Comparative analysis of base pairing probabilities. *RNA* **3**:779–791.

Ray S.C., Wang Y.M., Laeyendecker O., Ticehurst J.R., Villano S.A., and Thomas D.L., 1999. Acute hepatitis C virus structural gene sequences as predictors of persistent viremia: hypervariable region 1 as a decoy. *J. Virol.* **73**:2938–2946.

Rivas E. and Eddy S.R., 2000. Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics* **16**:573–585.

Sankoff D., 1985. Simultaneous solution of the RNA folding, alignment, and protosequence problems. *SIAM J. Appl. Math.* **45**:810–825.

Schuster P., Fontana W., Stadler P.F., and Hofacker I.L., 1994. From sequences to shapes and back: A case study in RNA secondary structures. *Proc. Royal Society London B* **255**:279–284.

Simmonds P. and Smith D.B., 1999. Structural constraints on RNA virus evolution. *J. Virol.* **73**:5787–5794.

Simons J.N., Desai S.M., Schultz D.E., Lemon S.M., and Mushahwar I.K., 1996. Translation initiation in GB viruses A and C: evidence for internal ribosome entry and implication for genome organization. *J. Virol.* **70**:6126–6135.

Smith D.B., Cuceanu N., Davidson F., Jarvis L.M., Mokili J.L., Hamid S., Ludlam C.A., and Simmonds P., 1997. Discrimination of hepatitis G virus/GBV-C geographical variants by analysis of the 5' non-coding region. *J. Gen. Virol.* **78**:1533–1542.

Spahn C.M., Kieft J.S., Grassucci R.A., Penczek P.A., Zhou K., Doudna J.A., and Frank J., 2001. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science* **291**:1959–1962.

Stocsits R., Hofacker I.L., and Stadler P.F., 1999. Conserved secondary structures in hepatitis B virus RNA. In *Computer Science in Biology*, pages 73–79. Univ. Bielefeld, Bielefeld, D. Proceedings of the GCB'99, Hannover, D.

Tanaka T., Kato N., Cho M.J., Sugiyama K., and Shimotohno K., 1996. Structure of the 3' terminus of the hepatitis c virus genome. *J. Virol.* **70**:3307–12.

Tang S., Collier A.J., and Elliott R.M., 1999. Alterations to both the primary and predicted secondary structure of stem-loop IIIc of the hepatitis C virus 1b 5' untranslated region (5'UTR) lead to mutants severely defective in translation which cannot be complemented in trans by the wild-type 5'UTR sequence. *J. Virol.* **73**:2359–2364.

Tautz N., Harada T., Kaiser A., Rinck G., Behrens S., and Thiel H.J., 1999. Establishment and characterization of cytopathogenic and noncytopathogenic pestivirus replicons. *J. Virol.* **73**:9422–9432.

Thompson J.D., Higgs D.G., and Gibson T.J., 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties, and weight matrix choice. *Nucl. Acids Res.* **22**:4673–4680.

Tuplin A., Wood J., Evans D.J., Patel A.H., and Simmonds P., 2002. Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. *RNA* **8**:824–841.

van Regenmortel M.H.V., Fauquet C., Bishop D., Carstens E., Estes M., Lemon S., Maniloff J., Mayo M., McGeoch D., Pringle C., *et al.*, 2000. *Virus Taxonomy: The Classification and Nomenclature of Viruses. The Seventh Report of the International Committee on Taxonomy of Viruses.* Academic Press, SanDiego. `http://www.ncbi.nlm.nih.gov/ICTVdb/`.

Witwer C., Rauscher S., Hofacker I.L., and Stadler P.F., 2001. Conserved RNA secondary structures in picornaviridae genomes. *Nucl. Acids Res.* **29**:5079–5089.

Xiang J., Wunschmann S., Schmidt W., Shao J., and Stapleton J.T., 2000. Full-length GB virus C (hepatitis G virus) RNA transcripts are infectious in primary CD4-positive T cells. *J. Virol.* **74**:9125–9133.

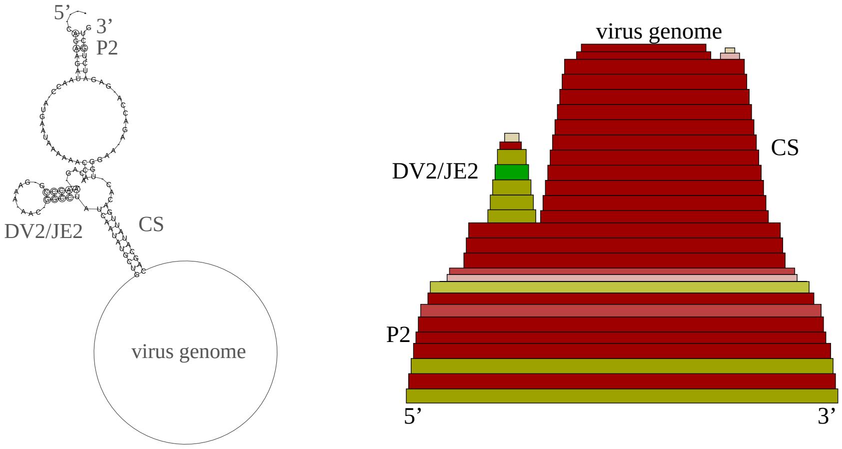
Yamada N., Tanihara K., Yorihuzi T.T., Tsutsumi M., Shimomura H., Tsuji T., and Date T., 1996. Genetic organization and diversity of the 3' noncoding region of the hepatitis C virus genome. *Virology* **223**:255–261.

You S. and Padmanabhan R., 1999. A novel in vitro replication system for dengue virus. initiation of RNA synthesis at the 3'-end of exogenous viral RNA templates requires 5'- and 3'-terminal complementary sequence motifs of the viral RNA. *J. Biol. Chem.* **274**:3714–3722.

Yu H., Grassmann C.W., and Behrens S.E., 1999. Sequence and structural elements at the 3' terminus of bovine viral diarrhea virus genomic RNA: functional role during RNA replication. *J. Virol.* **73**:3638–3648.

Zhao W.D. and Wimmer E., 2001. Genetic analysis of a poliovirus/hepatitis C virus chimera: new structure for domain II of the internal ribosomal entry site of hepatitis C virus. *J. Virol.* **75**:3719–3730.

Supplemental Material A



Conserved structures close to the CS of DEN and JEV.

Supplemental Material B

	DEN	JEV	YFV	TBE
P1'		73-10970		109-11143
		78-10965		111-11141
P1	83-10710	108-10883	134-10779	
	98-10694	112-10879	140-10773	
P2	112-10686	79-10910	100-10789	166-10958
	117-10681	102-10887	109-10780	168-10956
CS/CS"A"	141-10680	136-10876	147-10767	115-11073
	151-10670	146-10866	165-10750	129-11059

Sequence positions of putative motives which take part in genome cyclization for the Flavivirus species DEN, JEV, YFV and TBE. P1', P1, P2 and CS/CS"A" are shown in Fig. 2.