Nucleic Acids Research, 2001, Vol. 29, No. 24 **5079–5089**

Conserved RNA secondary structures in *Picornaviridae* genomes

Christina Witwer[1], Susanne Rauscher[1], Ivo L. Hofacker[1] and Peter F. Stadler[1,2,*]

[1]Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Währingerstraße 17,
A-1090 Wien, Austria and [2]The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Received August 6, 2001; Revised and Accepted October 10, 2001

ABSTRACT

The family *Picornaviridae* contains important pathogens including, for example, hepatitis A virus and foot-and-mouth disease virus. The genome of these viruses is a single messenger-active (+)-RNA of 7200–8500 nt. Besides coding for the viral proteins, it also contains functionally important RNA secondary structures, among them an internal ribosomal entry site (IRES) region towards the 5′-end. This contribution provides a comprehensive computational survey of the complete genomic RNAs and a detailed comparative analysis of the conserved structural elements in seven of the currently nine genera in the family *Picornaviridae*. Compared with previous studies we find: (i) that only smaller sections of the IRES region than previously reported are conserved at single base-pair resolution and (ii) that there is a number of significant structural elements in the coding region. Furthermore, we identify potential *cis*-acting replication elements in four genera where this feature has not been reported so far.

INTRODUCTION

The genomes of RNA viruses not only code for proteins, but often contain functionally active RNA structures that play a role during the various stages of the viral life cycle. Determining which parts of the huge RNA structure, formed by a viral genome, are functionally relevant is a difficult task. In general, the secondary structures of such regions do not look significantly different from structures formed by random sequences. RNA secondary structures are, however, quite susceptible to point mutations: computer simulations (1,2) showed that a small number of point mutations are very likely to cause large changes in the secondary structures: mutations in 10% of the sequence positions already lead 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 conservation.

If selection acts to preserve a structural element then it must have some function. This observation led to the design of algorithms, briefly outlined in the Materials and Methods section, that reliably detect conserved RNA secondary structure elements in a small sample of related RNA sequences (3,4). Of course, it is not possible to determine the function of the conserved structure elements. Nevertheless, knowledge about their location can be used to guide, for instance, deletion studies (5).

The family *Picornaviridae* is currently divided into nine genera: aphthovirus, cardiovirus, enterovirus, hepatovirus, rhinovirus, parechovirus, erbovirus, kobuvirus and teschovirus (6). These viruses are among the smallest ribonucleic acid-containing mammalian viruses known.

Although the members of the different genera exhibit only a small degree of sequence similarity at the nucleic acid level, they are homologs with a similar genomic structure and gene organization (see Fig. 1 and Table 1). The genome consists of a single strand messenger-active (+)-RNA of 7200–8500 nt that is polyadenylated at the 3′-terminus and carries a small protein (virion protein, genome; VPg) covalently attached to its 5′-end. The major part of the genomic RNA consists of a single large open reading frame (ORF) coding a polyprotein. The 5′-non-translated region (5′-NTR) is unusually long and contains multiple AUG triplets prior to the initiator of the viral translation. All *Picornaviridae* have an internal ribosomal entry site (IRES) instead of a 5′-cap structure (7,8).

A number of conserved RNA secondary structure elements have been described at least in some genera. Here we provide a comprehensive survey.

MATERIALS AND METHODS

The algorithms alidot and pfrali for searching conserved secondary structure patterns in large RNAs have been described in detail previously (3,4). An ANSI C implementation is available from http://www.tbi.univie.ac.at/RNA/. The 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. In this respect, alidot and pfrali are similar to programs such as construct (9,10) and x2s (11; see also 12). In contrast to efforts to simultaneously compute alignment and secondary structures (13,14) the present approach emphasizes that the sequences may have common

*To whom correspondence should be addressed at: Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Währingerstraβe 17, A-1090 Wien, Austria. Tel: +43 1 4277 52737; Fax: +43 1 4277 52793; Email: studla@tbi.univie.ac.at

structural motifs but no single common structure. In this sense alidot/pfrali combines structure prediction and motif search (15).

The algorithms implement a combination of thermodynamic structure prediction and phylogenetic comparison. In the first step, a set of thermodynamically plausible candidate base pairs is obtained by computing the matrix of base pairing probabilities using McCaskill's partition function algorithm (16) for each sequence and retaining all pairs with a thermodynamic equilibrium probability greater than 3×10^{-3}. The computations were performed using the Vienna RNA Package (17), based on the energy parameters published previously (18). In the second step, the candidate base pairs are ranked based on the sequence variation across the sample.

A sequence is *compatible* with base pair (i, j) if the two nucleotides at positions i and j of the multiple alignment can form either a Watson–Crick (GC, CG, AU or UA) pair or a wobble (GU, UG) pair. When different pairing combinations are found for a particular base pair (i, j) we speak of *consistent* mutations. If we find combinations such as GC and CG or GU and UA, where both positions are mutated at once we have compensatory mutations. The occurrence of consistent and, in particular, compensatory mutations strongly supports a predicted base pair, whereas non-consistent mutations contradict the conservation of a structural feature.

The ranking is hierarchical, first by the number of sequences that are inconsistent with a base pair and then by the number of pair types times the average equilibrium probability. In the final step, high-ranking pairs are combined into structural elements.

The multiple sequence alignments are calculated using CLUSTAL W (19) and Ralign (20). The latter program produces an amino acid sequence alignment for the coding parts of viral genomes, which is translated back to the underlying RNA sequence and combined with RNA alignments of the non-coding regions. The alignments are used without further modifications (except where stated explicitly). The quality of the alignment has a strong effect on the results, as small errors in the alignment can easily hide a conserved feature. While false positives remain rare, the number of conserved structures that are found decreases with the diversity of the sequences analyzed. The best results are obtained when the sequence diversity is large enough to provide many compensatory mutations, but low enough to allow accurate alignments, typically at pairwise identity of, say, 80%.

Results are presented as conventional secondary structure drawings, as mountain plots (e.g. Fig. 4, top) or dot plots (Fig. 8).

Mountain plots. A base pair (i, j) is represented by a slab ranging from i to j. The 5′ and 3′ sides of stems thus appear as uphill and downhill slopes, respectively, whereas plateaus indicated unpaired regions. Mountain plots are equivalent to the conventional drawing but have the advantage that: (i) they can be compared more easily and (ii) it is easier to display additional information.

Dot plots. Dot plots are useful when structural alternatives have to be displayed. Each pair is shown as a small square at positions (i, j) and (j, i). The upper right and the lower left triangle can be used to compare structures obtained by different methods.

Table 1. Complete genomic RNA of *Picornaviridae*

Genus	N	l_A	η	Coding region	η (5′)
Aphthovirus	9	8231	0.791	1088–8124	0.543[a]
Cardiovirus	6	8233	0.665	1088–8103	0.614
Enterovirus	29	7664	0.651	777–7548	0.765
Hepatovirus	10	7526	0.911	759–7449	0.901
Parechovirus	3	7391	0.774	716–7283	0.828
Rhinovirus	7	7296	0.687	652–7138	0.771
Teschovirus	25	7135	0.893	433–7063	0.945

We list the number N of available sequences, the length l_A of their alignment, their average pairwise sequence identity η, the location of the coding region in the alignment, and the mean pairwise sequence identity in the 5′-NTR. Only one complete sequence is known for both erbovirus and kobuvirus, which is not sufficient for the analysis presented here. Teschovirus sequences do not include the S-fragment.

[a]Seven of the nine sequences from the 5′-terminus are incomplete in the GenBank entries.

Mountain plots and dot plots contain information about both sequence variation (color code) and thermodynamic likeliness of a base pair (indicated by the height of the slab and the size of the dot, respectively). Colors in the order red, ocher, green, cyan, blue and violet indicate one to six different types of base pairs. Pairs with one or two inconsistent mutations are shown in (two types of) pale colors.

In the conventional graphs, paired positions with consistent mutations are indicated by circles around the varying position. Compensatory mutations are thus shown by circles around both pairing partners. Inconsistent mutants are indicated by gray instead of black lettering.

Supplemental material

The complete data, consisting of the multiple alignments for each genus, the thermodynamic structure predictions for each sequence, the output of the pfrali program, and the secondary structure elements listed in Figure 1, are accessible through a web interface at http://rna.tbi.univie.ac.at/virus/.

RESULTS

The putative conserved structural features that have been identified for each genus are summarized in Figure 1. The largest pieces of conserved structure are located in the 5′-UTR. All groups, with the exception of rhinovirus, show a hairpin motif close to the 3′-end. In addition, there are a substantial number of possibly conserved RNA structures within the ORF.

5′-Non-translated region

The most prominent feature in the 5′-NTR is the IRES. The literature distinguished two or three types of IRES structures. Most common in the literature is the distinction between the RE type found in rhinovirus and enterovirus, and the ACH type structure of aphthovirus, cardiovirus and hepatovirus (21–24). Some authors distinguish three groups: (i) aphthovirus and cardiovirus, (ii) enterovirus and rhinovirus and (iii) hepatovirus (25,26). Figure 2 summarizes the results from our

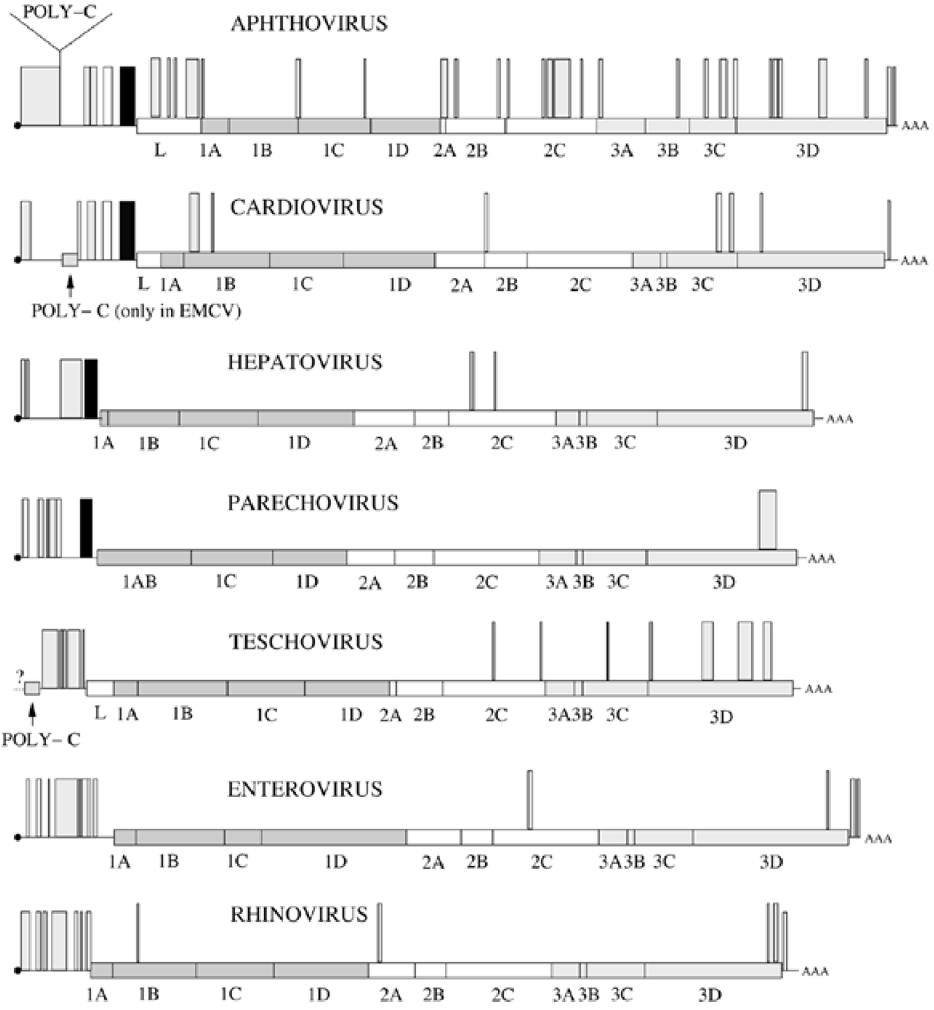

POLY–C
APHTHOVIRUS
L 1A 1B 1C 1D 2A 2B 2C 3A 3B 3C 3D
AAA
CARDIOVIRUS
L 1A 1B 1C 1D 2A 2B 2C 3A 3B 3C 3D
POLY– C (only in EMCV)
HEPATOVIRUS
1A 1B 1C 1D 2A 2B 2C 3A 3B 3C 3D
PARECHOVIRUS
1AB 1C 1D 2A 2B 2C 3A 3B 3C 3D
TESCHOVIRUS
?
POLY– C
L 1A 1B 1C 1D 2A 2B 2C 3A 3B 3C 3D
ENTEROVIRUS
1A 1B 1C 1D 2A 2B 2C 3A 3B 3C 3D
RHINOVIRUS
1A 1B 1C 1D 2A 2B 2C 3A 3B 3C 3D

Figure 1. Overview of Picorna genomes. Putative conserved RNA elements are indicated above the diagrams of the reading frames. The black boxes indicate the J,K-element, the white box is the Ib-element (for details see text). Proteins: leader protein L (only present in aphthovirus, cardiovirus and teschovirus), capsid proteins 1A–1D, viral protease 2A, proteins involved in RNA synthesis 2B, 2C, unknown function 3A, VPg 3B, major viral protease 3C, RNA-dependent RNA-polymerase 3D (24,50,51).

analysis, which includes teschovirus, parechovirus and erbovirus for the first time.

Overall, we find that the IRES structure is less conserved even within the genera than expected. Figure 2 indicates in color those features that are conserved within a group at the level of individual base pairs. Non-shaded parts of the structure are taken from folding for each genus (the reference sequence listed in Appendix A) using the conserved base pairs as constraints.

We recover the close structural similarities of rhinovirus and enterovirus. A comparison of aphthovirus, cardiovirus, hepatovirus and parechovirus sets hepatovirus apart from the other two groups (see also the details of the J- and K-elements in Fig. 3). The one available complete erbovirus genome also shares the elements J and K with aphthovirus and cardiovirus (26). Teschovirus forms a distinct fourth group of IRES structures that appears only vaguely related to the other groups.

Cardiovirus and aphthovirus. The secondary structure of the 5′-NTR of cardiovirus has been discussed previously (20–23,27). Our results are very similar to previous work (23) on the EMC virus. The main difference is that elements Ia and Ic, which flank element Ib, are not present in the TME virus and therefore not conserved in the genus cardiovirus. In addition, the H-element is longer in our data. In comparison with the earlier studies, both our results and the structures in a previous study (23) have shorter conserved stem–loop regions.

The 5′-NTR of the aphthovirus FMDV has been discussed previously (21,22,28). There is only a single sequence for ERV-1 (equine rhinovirus I), which has only marginal sequence similarity with FMDV and hence was considered separately. Our data for FMDV are similar to the earlier studies. However, we find that the conserved parts of the stem–loop regions are significantly shorter than the ones reported (21).

A comparison of cardiovirus and aphthovirus structures shows the following main differences: (i) the stem–loop structure A1 at the 5′-end is much longer in FMDV compared with cardiovirus and (ii) the D-element in FMDV is enlarged at the expense of F.

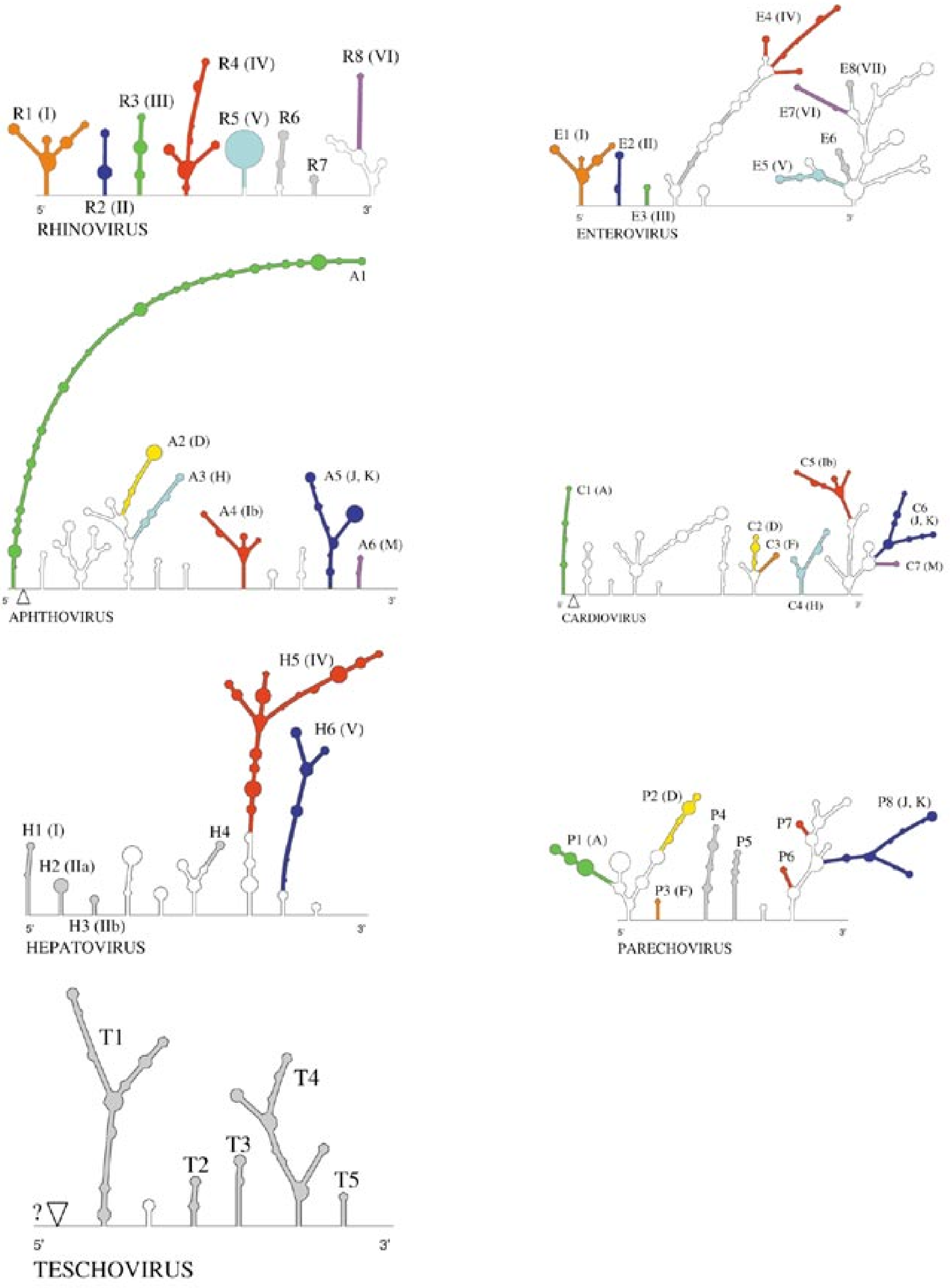


Figure 2. Schematic illustration of the 5′-NTRs of *Picornaviridae*. The minimum energy structure of one sequence of the respective virus genus is represented. Colored backgrounds mark regions that are conserved within all investigated sequences of that genus. The labels in parentheses correspond to the notation of the ‘classic’ model of the IRES (24). The sequence positions of the structure elements in a reference structure are listed in Appendix A. Δ denotes a poly(C) region, ? indicates the missing data for the 5′-end of the teschovirus sequences.

The stem–loop structure H is very similar in aphthovirus and cardiovirus, the loop sequence UCUUU is strongly conserved in both genera. The stem contains many compensatory mutations that CLUSTAL W failed to correctly align in this region. Manual improvement of the alignment shows that the Ib-elements as well as the M-elements of both genera can be superimposed and hence are structurally almost identical (Fig. 4, left and right). In contrast, only the J-stem of the J,K feature is structurally (almost) identical in the two genera (Fig. 4, middle) despite the fact that the topology of the J,K-elements is conserved (Fig. 3).

Parechovirus. Until recently parechoviruses echovirus 22 and echovirus 23 were classified as members of the genus enterovirus.

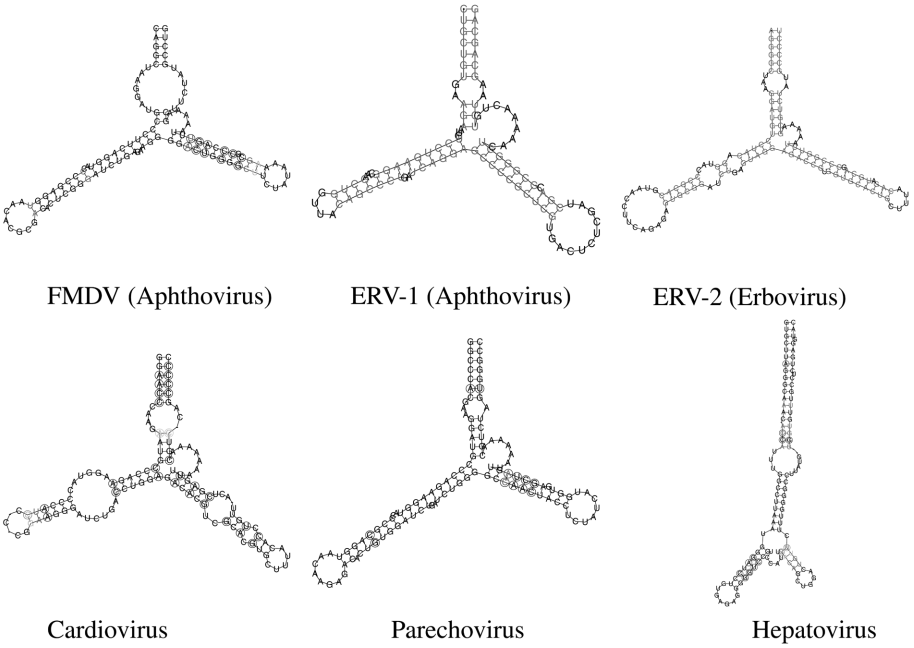


Figure 3. Consensus structures of the J- and K-elements. There appears to be some variation within aphthovirus. The sizes of the loops vary between different strains, FMDV and ERV-1, in this example. Hepatovirus shows an analogous structure.

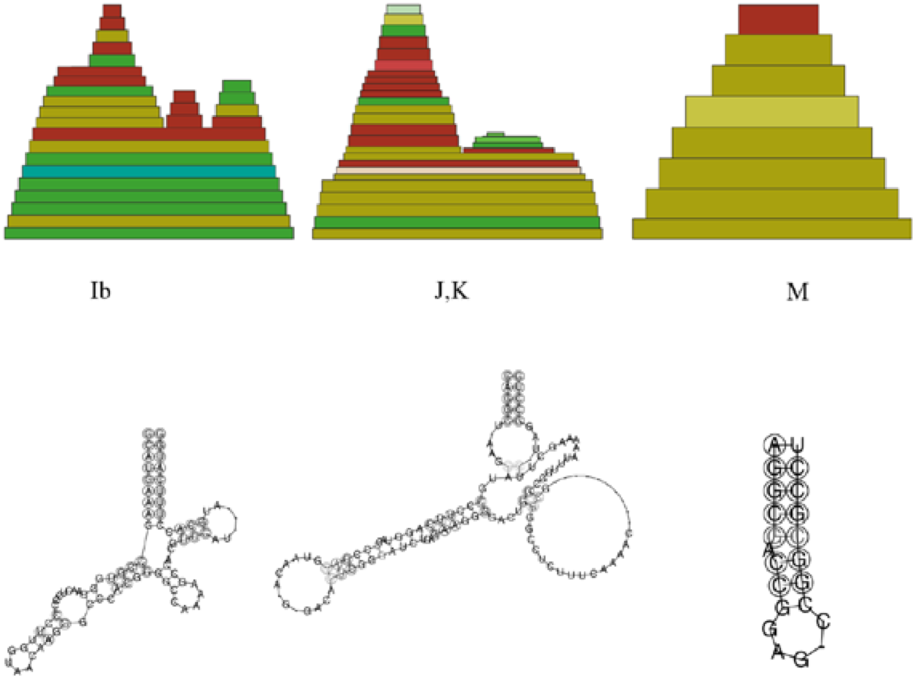


Figure 4. Common features in the IRES of aphthovirus and cardiovirus. Alignment manually improved.

The secondary structure of their 5′-NTR has been described previously (29,30). As our analysis is based on only three sequences, we might still overestimate the conserved parts, in particular in regions P1, P4 and P6 where the sequence is highly conserved. Ghazi *et al*. (30) finds the same structure for parechovirus and cardiovirus. Our results agree in part with this analysis. In particular, our element P1 corresponds to A, but is shorter. P2 corresponds to D and P3 corresponds to F. Our element P4 is located in the region of Ghazi’s H. The sequence is rather conserved in this region and both Ghazi’s H and our P4 have comparable thermodynamic stability, with a pairing probability of approximately $P = 1/3$ for each of the

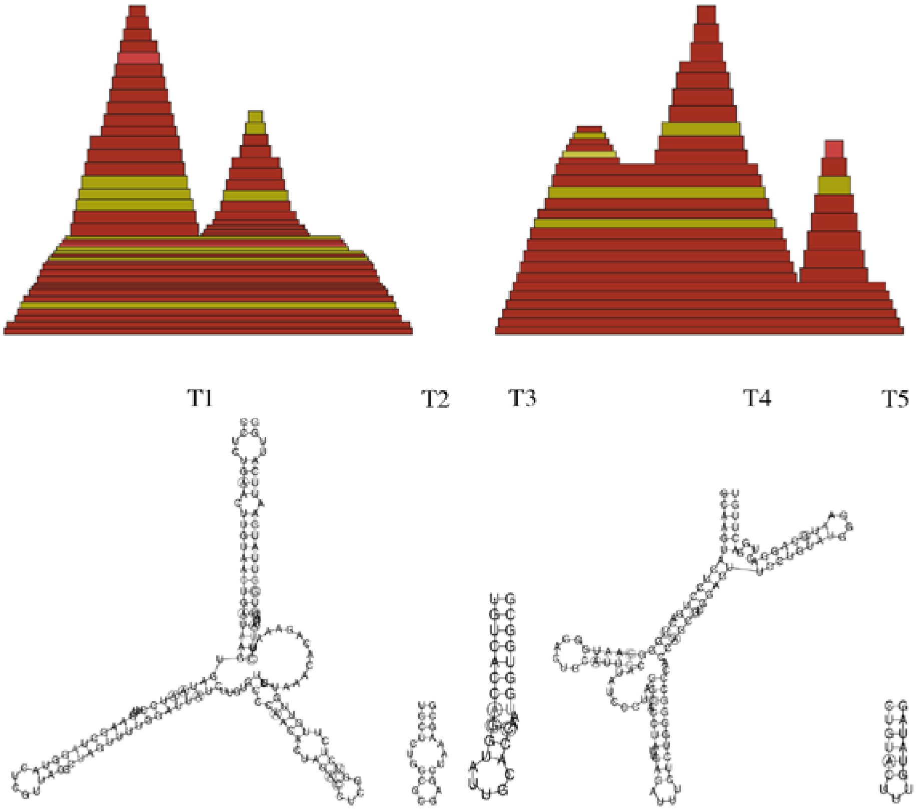


Figure 5. Conserved structures in the 5′-NTR of teschovirus. Mountain plots are given only for T1 and T4.

two alternatives. Both variants have little similarity with the H-element in cardiovirus and aphthovirus. The element P5 has been reported before. P8 contains the J- and K-elements, where J is identical to Ghazi's structures, our prediction for the K-element shows minor differences with previous studies. Our elements P6 and P7 are part of Ghazi's clover leaf-like motif I. The clover leaf-structure is thermodynamically feasible ($P = 1/4$), but appears to have significant structural variability in this genus so that it is not detected as a conserved feature. The analysis in Oberste *et al.* (29), which used only two sequences and Zuker's mfold program, agrees in part with our consensus structure.

Hepatovirus. The secondary structure of hepatovirus RNA was considered by Brown *et al.* (31). The sequences in our data set have ~90% pairwise identity, hence we have only a small number of compensatory mutations to verify structural features predicted based on the thermodynamic rules.

Elements I and II are identical to Brown's structure; the sequences are completely conserved in these regions, i.e. there are no co-variations to verify the thermodynamic prediction in the region. Domain III is not present in our data. We find high structural flexibility here. It was noted (31) that 'the structure of domain III was poorly defined'. The only possibly significant structure in this region is a stem–loop with a completely conserved sequence around position 300, which does not appear in Brown's prediction.

Stem IV is significantly shorter in our analysis and the multiloop of the clover leaf structure is slightly different. The deletion studies reported previously (31) indicate that domain IV is critically involved in formation of an HAV IRES element. Compared with the earlier study we find a larger element V at the expense of part of IV.

The conserved secondary structure elements of hepatovirus cannot be compared directly with that of aphthovirus and cardiovirus. But there is a structural analogy of the clover leaf structure (Ib in cardiovirus/aphthovirus and IV in hepatovirus) and the branching stem–loop (J,K in cardiovirus/aphthovirus and V in hepatovirus).

Teschovirus. The sequences of the teschovirus 5′-NTR are not known completely. The presence of an oligo-C stretch was demonstrated for F65 (32). The nucleotide sequence of the 5′-NTR up to this C tract could be determined successfully only in three of the 25 assayed strains (Talfan, Bozen and Vir-1626/89) (33). Hence we report no structure before the oligo-C region.

The secondary structure of the 5′-NTR of teschovirus has not been studied previously. Only element T4 shows some similarities with element V in hepatovirus. The other conserved structures do not have obvious similarities with conserved elements in other picorna genera. The conserved elements T1–T5 are shown in Figure 5.

Enterovirus and rhinovirus. The secondary structure of IRES of enterovirus and rhinovirus has been the subject of a larger number of studies (25,34,35). We recover elements I–VII in enterovirus and I–VI in rhinovirus, some of them with a slightly shorter stem. In addition, we found the stem–loop structures E6 and R6, and R7, respectively (see Fig. 2). E6 and R7 are homologous structures; the sequence is absolutely conserved here. Elements R5 (=V) and R7 can be detected unambiguously only in HRV-A.

An attempt to extract the common structures of enterovirus and rhinovirus for a common multiple alignment yielded only a fraction of the structures found in each genus separately. In part, this is due to small differences in the structural elements and, in part, the lack of structures can be attributed to the poor quality of the alignment.

Other Picornaviridae. According to Wutz *et al.* (26) the IRES structure of ERV-1 (=ERAV, aphthovirus) and ERV-2 (=ERBV, erbovirus) is similar to that of aphthovirus and cardiovirus. The similarity to FMDV is insufficient for a good alignment of ERV-2 with the FMDV sequences. The computed minimum energy structure shows an identifiable J,K-element. The one complete sequence of kobuvirus does not exhibit any features that can be matched unambiguously with the conserved structural elements of the other genera.

Coding region

Cis-acting replication element. A *cis*-acting replication element (CRE) within the coding region of several picornaviruses has been described in a number of different picornaviruses. The function of the CRE probably involves the initiation of the synthesis of the negative-sense strand template RNA during virus replication (36). The CRE has been identified in HRV14 in region 1B of the genome (37), in cardiovirus in region 1B (38) and in poliovirus in region 2C (36).

Although located within a protein-coding segment of the genome, the CRE function is independent of its translation. Thus, this segment of the viral RNA has dual functions, both encoding the VP1 capsid protein and participating directly in the replication of the viral genome. The existence of the computer-predicted structure was confirmed by mutational analysis (37,38). Furthermore, the activity of the CRE is not position dependent (36).

In cardiovirus we recover the CRE in the 1B region, which encodes the capsid protein VP2. For EMCV (excluding Mengo-virus) and theilovirus, our structure agrees with the one reported previously (38). In the study by Palmenberg and Sgro (23) a different structure is given for Mengo-virus. We find that the Mengo-virus CRE-structures agree with the consensus of the other species. In enterovirus we recover the CRE in 2C (36), in HRV-B the element is found in 1B (37).

We find putative CRE elements in the 2C region of aphthovirus and teschovirus, and in the 2A region of HRV-A. There are three conserved elements in the coding region of hepatovirus. The most likely candidate for a CRE is located in region 2C. For parechovirus we were not able to identify a putative CRE. The locations of the (putative) CREs are summarized in Table 2.

The loop of the CRE is relatively large in all genera and contains predominantly A and C (Fig. 6). We note that an alignment of region 2C of all aphthovirus and teschovirus sequences shows that the CRE element is conserved between the two genera.

Table 2. Position of (putative) CRE

Genus	Accession nos	Gene	Position	Remark
Aphthovirus	AJ007347	2C	4834–4859	Putative
Enterovirus	V01150	2C	4456–4494	As in ref. (36)
Cardiovirus	M81861	1B	1308–1340	As in ref. (38)
Hepatovirus	K02990	2C	4187–4245	Possible
Teschovirus	AF231769	2C	4228–4249	Putative
Rhinovirus-A	M16248	2A	3325–3357	Putative
Rhinovirus-B	K02121	1B	1727–1764	As in ref. (37)
Parechovirus	?			

Other conserved elements. There appears to be no structural feature in the coding region that is shared among all picorna genera besides the CRE. On the other hand, we find a number of structures that are conserved within a genus (Fig. 1). There are five such structures in cardiovirus, a single feature in the three-dimensional region of parechovirus, six in yeschovirus, three in hepatovirus, two in enterovirus, 25 (!) in aphthovirus and one in rhinovirus. A complete list is provided in electronic form at http://rna.tbi.univie.ac.at/virus/. We do not have an explanation for the large number of conserved elements in genus aphthovirus in comparison with all other *Picornaviridae*.

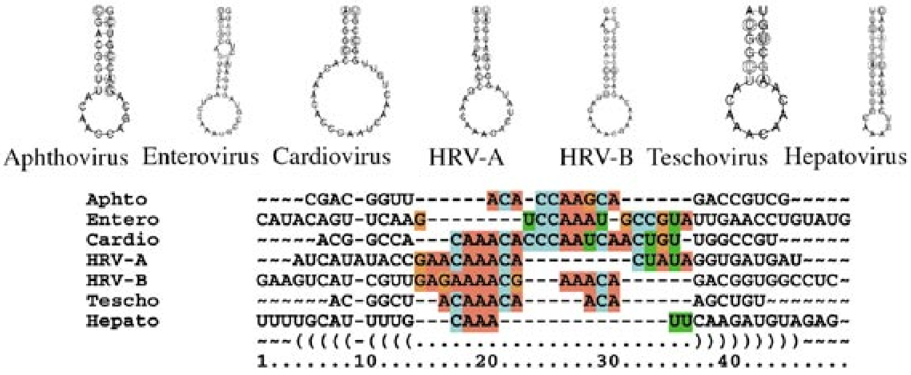
The two species HRV-A and HRV-B in the genus rhinovirus show significant differences at the level of their secondary structures. In HRV-A we find four conserved elements inside the coding region, only one of which is also conserved between HRV-A + HRV-B. These differences are emphasized by the fact that the CRE is located in different parts of the genome in these two species (Table 2).

3′-Non-translated region

Recent deletion studies (39,40) show that the 3′-NTR, which ends in a poly(A) region of variables length in all genera, is important for RNA synthesis.

Cardiovirus. Duque and Palmenberg (40) report three conserved stem–loop motifs (I, II, III) in the 3′-NTR. Deletion studies by the authors indicated that deletion of III is lethal, deletion of II resulted in marginal RNA synthesis activity but failure of transfection with genomic RNA, while stem I was found to be dispensable for viral growth. Surprisingly, we find stem I has large thermodynamic stability and a significant number of compensatory mutations (Fig. 7), whereas regions II and III do not form a conserved structure in our data set. The structures reported (40) are consistent with all sequences of our data set but are not thermodynamically favorable in any one of them (Fig. 8). Neither II nor III can be recovered by considering the species theilovirus and EMCV separately.

The secondary structures reported by Cui and Porter (39) are completely different from both our results and the structures reported by Duque and Palmenberg (40). Cui and Porter (39) suggest that a U-rich stretch, essentially region III (40), interacts with the poly(A) tail.



```
Aphto    ----CGAC-GGUU------ACA-CCAAGCA------GACCGUCG-----
Entero   CAUACAGU-UCAAG--------UCCAAAU-GCCGUAUUGAACCUGUAUG
Cardio   -----ACG-GCCA---CAAACACCCAAUCAACUGU-UGGCCGU------
HRV-A    ---AUCAUAUACCGAACAAACA---------CUAUAGGUGAUGAU----
HRV-B    GAAGUCAU-CGUUGAGAAAACG---AAACA------GACGGUGGCCUC-
Tescho   ------AC-GGCU--ACAAACA-----ACA------AGCUGU-------
Hepato   UUUUGCAU-UUUG---CAAA--------------UUCAAGAUGUAGAG-
         ---((((-((((.........................))))))))----
         1.......10........20........30........40.........
```

Figure 6. Known and putative CREs in *Picornaviridae*: secondary structures (top) and sequences (below). Nucleotides in the loops are highlighted to emphasize the AC-rich composition.

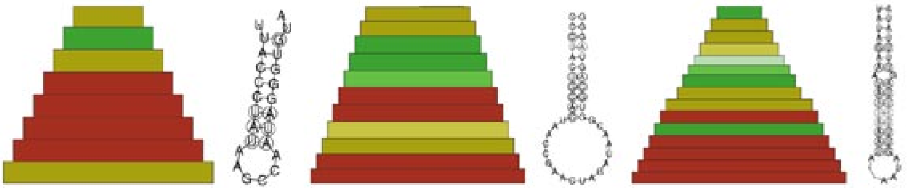


Figure 7. The most prominent features in the 3′-NTR. Cardiovirus region I (left); region II of enterovirus cluster 3 (middle); hairpin structure of rhinovirus (right).

Enterovirus. There is ample literature on the structure of the 3′-NTR of enterovirus (35,41–45). None of the reported structures are conserved within the entire genus. Following the previous studies, we have therefore split the 29 available genomes into three clusters because there are not enough sequences available for the fourth cluster described by Zell and Stelzner (35).

Cluster 1 consists of poliovirus and human enterovirus C, cluster 2 contains human enterovirus A (35,46). The 3′-NTR sequences are highly conserved within each of these two clusters. While we find the published structures in our data, their equilibrium probabilities are small and they appear as one of a number of thermodynamically feasible alternatives.

Cluster 3 contains human enterovirus B. We find domains I and II as reported previously (35,41). It is interesting to note that the structure of domain II is supported by a substantial number of compensatory mutations.

Other genera. The 3′-NTR of aphthovirus apparently has not been considered before. We found two hairpin structures, one of which has an almost conserved GAAA sequence motif in the loop. In one of the nine sequences we found GCAAA instead.

A hairpin motif, which was already reported (41), is detected unambiguously in rhinovirus (Fig. 7). The structure is conserved between HRV-A and HRV-B.

In parechovirus, hepatovirus and teschovirus there are no significant conserved secondary structures in the 3′-NTR. In particular, we could not confirm the published minimum energy structures for individual teschovirus (32,33) and hepatovirus (42) sequences as conserved features.

DISCUSSION

Structural genomics, the systematic determination of all macro-molecular structures represented in a genome, is at present focused almost exclusively on proteins. Over the past two decades it has become clear, however, that a variety of RNA molecules have important, and sometimes essential, biological functions beyond their roles as rRNAs, tRNAs or mRNAs. To comprehensively understand the biology of a cell, it will ultimately be necessary to know the identity of all encoded RNAs, the molecules with which they interact and the molecular structures of these complexes (47). Viral RNA genomes, because of their small size and the strong selection that acts upon them, are an ideal proving ground for techniques that aim at identifying functional RNA structures.

A combination of structure prediction based on the thermodynamic rules of the ‘standard energy model’ for nucleic acid secondary structures and the evaluation of consistent and compensatory mutations can be employed for scanning complete viral genomes for functional RNA structure motifs. This contribution reports a detailed, comprehensive survey of such structural features for those six (out of nine) genera of the family of *Picornaviridae* for which sufficient sequence information is currently available: aphthovirus, cardiovirus, enterovirus, hepatovirus, parechovirus, rhinovirus and teschovirus.

The 5′-region of a number of these viruses has been studied previously because of the particular interest in the IRES region. Our automatic approach confirms many of the patterns identified previously based on smaller data sets. However, we find that in many cases the parts of these features that are conserved base pair by base pair are significantly smaller. This

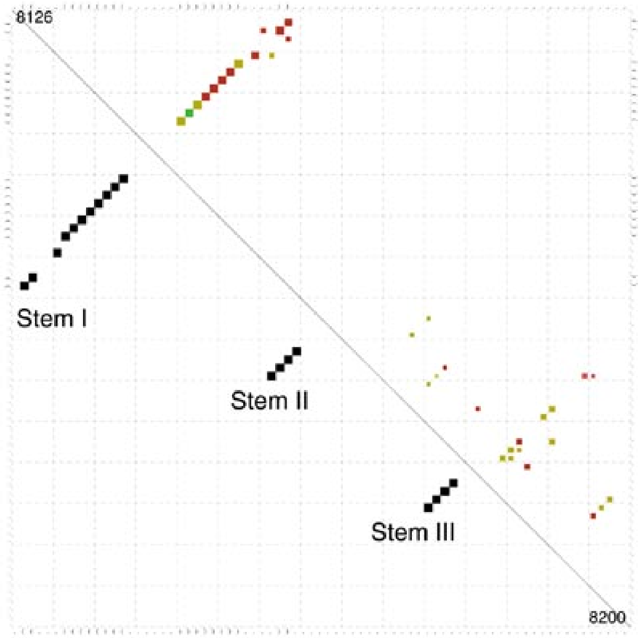


Figure 8. Dot plot of the 3′-NTR of cardiovirus. (Upper triangle) Output of pfrali showing only stem I. Its prediction is well supported by a number of consistent mutations and the absence of inconsistent mutations. The size of the squares is proportional to log$P$ here. The signals in the area of stems II and III have probabilities of only a few percent and hence are not significant. The lower triangle shows the structures reported in (40) for comparison.

conclusion is based mainly on the fact that some sequences that are now contained in the database simply cannot form parts of the structures that have been reported previously as conserved. The same conclusion can be drawn for the 3′-NTR.

On the other hand, there are a large number of secondary structure elements that have not been described before, most importantly within the coding region. Most notably, we have been able to identify likely or at least possible candidates for the CRE region in aphthovirus, hepatovirus, rhinovirus-A and teschovirus, apart from recovering the known locations of the CRE in enterovirus, cardiovirus and rhinovirus-B. Only for parechovirus we did not find a significant signal.

The approach used here goes beyond search software such as RNAMOT (48) in that it does not require any a priori knowledge of the functional structure motifs and it goes beyond searches for regions that are thermodynamically especially stable or well-defined (49) in that it returns a specific prediction for a structure if and only if there is sufficient evidence for structural conservation. The results collected here (and in the supplemental material available at http://rna.tbi.univie.ac.at/virus/) could be used to refine descriptors, e.g. for the CRE that can then be used for structure-specific scans in other RNAs.

ACKNOWLEDGEMENTS

This work is supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung, Project Nos P-12591-INF (S.R.) and P-13545-MAT (C.W.).

REFERENCES

1. Fontana,W., Konings,D.A.M., Stadler,P.F. and Schuster,P. (1993) Statistics of RNA secondary structures. *Biopolymers*, **33**, 1389–1404.
2. 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. Roy. Soc. London B*, **255**, 279–284.
3. 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. *Nucleic Acids Res.*, **26**, 3825–3836.
4. Hofacker,I.L. and Stadler,P.F. (1999) Automatic detection of conserved base pairing patterns in RNA virus genomes. *Comp. Chem.*, **23**, 401–414.
5. 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 flavivirus. *J. Virol.*, **72**, 2132–2140.
6. Pringle,C. (1999) Virus taxonomy at the XIth international congress of virology, Sydney, Australia. *Arch. Virol.*, **144**, 2065–2070.
7. Hewlett,M.J., Rose,J.K. and Baltimore,D. (1976) 5′ Terminal structure of poliovirus polyribosomal RNA is pUp. *Proc. Natl Acad. Sci. USA*, **73**, 327–330.
8. Nomoto,A., Lee,Y.F. and Wimmer,E. (1976) The 5′-end of poliovirus mRNA is not capped with m7g(5′)pppg(5′)np. *Proc. Natl Acad. Sci. USA*, **73**, 375–380.
9. 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.
10. Lück,R., Graf,S. and Steger,G. (1999) ConStruct: a tool for thermodynamic controlled prediction of conserved secondary structure. *Nucleic Acids Res.*, **27**, 4208–4217.
11. Juan,V. and Wilson,C. (1999) RNA secondary structure prediction based on free energy and phylogenetic analysis. *J. Mol. Biol.*, **289**, 935–947.
12. Le,S.Y. and Zuker,M. (1991) Predicting common foldings of homologous rnas. *J. Biomol. Struct. Dyn.*, **8**, 1027–1044.
13. Sankoff,D. (1985) Simultaneous solution of the RNA folding, alignment, and proto-sequence problems. *SIAM J. Appl. Math.*, **45**, 810–825.
14. Corodkin,J., Heyer,L.J. and Stormo,G.D. (1997) Finding common sequences and structure motifs in a set of RNA molecules. In Gaasterland,T., Karp,P., Karplus,K., Ouzounis,C., Sander,C. and Valencia,A. (eds), *Proceedings of the ISMB-97*. AAAI Press, Menlo Park, CA, pp. 120–123.
15. Dandekar,T. and Hentze,M.W. (1995) Finding the hairpin in the haystack: searching for RNA motifs. *Trends Genet.*, **11**, 45–50.
16. McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
17. 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.
18. Mathews,D., 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.
19. Thompson,J.D., Higgs,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties, and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
20. Stocsits,R., Hofacker,I.L. and Stadler,P.F. (1999) Conserved secondary structures in hepatitis B virus RNA. In Giegerich,R. (ed.), *Computer Science and Biology, Proceedings of the German Conference on Bioinformatics GCB '99*. University of Bielefeld, Bielefeld, Germany.
21. Pilipenko,E.V., Blinov,V.M., Chernov,B.K., Dmitrieva,T.M. and Agol,V.I. (1989) Conservation of the secondary structure elements of the 5′-untranslated region of cardio- and aphthovirus RNAs. *Nucleic Acids Res.*, **17**, 5701–5711.
22. Le,S.Y., Chen,J.H., Sonenberg,N. and Maizel,J.V. (1993) Conserved tertiary structural elements in the 5′ nontranslated region of cardiovirus, aphthovirus and hepatitis A virus RNAs. *Nucleic Acids Res.*, **21**, 2445–2451.
23. Palmenberg,A.C. and Sgro,J. (1997) Topological organization of picornaviral genomes: statistical prediction of RNA structural signals. *Semin. Virol.*, **8**, 231–241.

24. Rueckert,R.R. (1996) *Picornaviridae*: the viruses and their replication. In Fields,N., Knipe,D. and Howley,P. (eds), *Virology,* 3rd Edn. Lippincott-Raven Publishers, Philadelphia, New York, Vol. 1, pp. 609–654.
25. Jackson,R.J., Howell,M.T. and Kaminski,A. (1990) The novel mechanism of initiation of picornavirus RNA translation. *Trends Biochem. Sci.*, **15**, 477–483.
26. Wutz,G., Nowotny,N., Grosse,B., Skern,T. and Kuechler,E. (1996) Equine rhinovirus serotypes 1 and 2: relationship to each other and to aphthoviruses and cardioviruses. *J. Gen. Virol.*, **77**, 1719–1730.
27. Duke,G.M., Hoffman,M.A. and Palmenberg,A.C. (1992) Sequence and structural elements that contribute to efficient encephalomyocarditis virus RNA translation. *J. Virol.*, **66**, 1602–1609.
28. Clarke,B.E., Brown,A.L., Currey,K.M., Newton,S.E., Rowlands,D.J. and Carroll,A.R. (1987) Potential secondary and tertiary structure in the genomic RNA of foot and mouth disease virus. *Nucleic Acids Res.*, **15**, 7067–7079.
29. Oberste,M.S., Maher,K. and Pallansch,M.A. (1998) Complete sequence of echovirus 23 and its relationship to echovirus 22 and other human enteroviruses. *Virus Res.*, **56**, 217–223.
30. Ghazi,F., Hughes,P.J., Hyypiae,T. and Stanway,G. (1998) Molecular analysis of human parechovirus type 2 (formerly echovirus 23). *J. Gen. Virol.*, **79**, 2642–2650.
31. Brown,E.A., Day,S.P., Jansen,R. and Lemon,S.M. (1991) The 5′ nontranslated region of hepatitis A virus RNA: secondary structure and elements required for translation *in vitro. J. Virol.*, **65**, 5828–5838.
32. Doherty,M., Todd,D., McFerran,N., and Hoey,E.M. (1999) Sequence analysis of a porcine enterovirus serotype 1 isolate: relationships with other picornaviruses. *J. Gen. Virol.*, **80**, 1929–1941.
33. Zell,R., Dauber,M., Krumbholz,A., Henke,A., Birch-Hirschfeld,E., Stelzner,A., Prager,D. and Wurm,R. (2001) Porcine teschoviruses comprise at least eleven distinct serotypes: molecular and evolutionary aspects. *J. Virol.*, **75**, 1620–1631.
34. Pilipenko,E.V., Blinov,V.M., Romanova,L.I., Sinyakow,A.N., Maslova,S.V. and Agol,V.I. (1989) Conserved structural domains in the 5′-untranslated region of picornaviral genomes: an analysis of the segment contolling translation and neurovirulence. *Virology*, **168**, 201–209.
35. Zell,R. and Stelzner,A. (1997) Application of genome sequence information to the classification of bovine enteroviruses: the importance of 5′- and 3′-nontranslated regions. *Virus Res.*, **51**, 213–229.
36. Goodfellow,I., Chaudhry,Y., Richardson,A., Meredith,J., Almond,J.W., Barclay,W., and Evans,D.J. (2000) Identification of a *cis*-acting replication element within the poliovirus coding region. *J. Virol.*, **74**, 4590–4600.
37. McKnight,K. and Lemon,S.M. (1998) The rhinovirus type 14 genome contains an internally located RNA structure that is required for viral replication. *RNA*, **4**, 1569–1584.
38. Lobert,P.-E., Escriou,N., Ruelle,J. and Michiels,T. (1999) A coding RNA sequence acts as a replication signal in cardioviruses. *Proc. Natl Acad. Sci. USA*, **96**, 11560–11565.
39. Cui,T. and Porter,A.G. (1995) Localization of the binding site for encephalomyocarditis virus RNA polymerase in the 3′-noncoding region of the viral RNA. *Nucleic Acids Res.*, **23**, 377–382.
40. Duque,H. and Palmenberg,A.C. (2001) Phenotypic characterization of three phylogenetically conserved stem-loop motifs in the mengovirus 3′ untranslated region. *J. Virol.*, **75**, 3111–3120.
41. Pilipenko,E.V., Maslova,S.V., Sinyakov,A. and Agol,V.I. (1992) Towards identifaction of *cis*-acting elements involved in the replication of enterovirus RNAs: a proposal for the existence of tRNA-like terminal structures. *Nucleic Acids Res.*, **20**, 1739–1745.
42. Rohll,J.B., Moon,D.H., Evans,D.J. and Almond,J.W. (1995) The 3′ untranslated region of picornavirus RNA: Features required for efficient genome replication. *J. Virol.*, **69**, 7835–7844.
43. Mirmomeni,M.H., Hughes,P.J. and Stanway,G. (1997) A tertiary structure in the 3′ untranslated region of enteroviruses is necessary for efficient replication. *J. Virol.*, **71**, 2363–2370.
44. Wang,J., Bakkers,J.M., Galama,J.M., Bruins Slot,H.J., Pilipenko,E.V., Agol,V.I. and Melchers,W.J. (1999) Structural requirements of the higher order RNA kissing element in the enteroviral 3′ UTR. *Nucleic Acids Res.*, **27**, 485–490.
45. Meredith,J.M., Rohll,J .B., Almond,J.W. and Evans,D.J. (1999) Similar interactions of the poliovirus and rhinovirus 3d polymerases with the 3′ untranslated region of rhinovirus 14. *J. Virol.*, **73**, 9952–9958.
46. King,A.M.Q., Brown,F., Christian,P., Hovi,T., Hyypiä,T., Knowles,N.J., Lemon,S.M., Minor,P.D., Palmenberg,A.C., Skern,T. and Stanway,G. (2000) Picornaviridae. In Van Regenmortel,M.H.V., Fauquet,C., Bishop,D.H.L., Calisher,C.H., Carsten,E.B., Estes,M.K., Lemon,S.M., Maniloff,J., Mayo,M.A., McGeoch,D.J., Pringle,C.R. and Wickner,R.B. (eds), *Virus Taxonomy. Seventh Report of the International Committee for the Taxonomy of Viruses.* Academic Press, New York, San Diego, pp. 657–673.
47. Doudna,J.A. (2000) Structural genomics of RNA. *Nature Struct. Biol.*, **7**, 954–956.
48. Gautheret,D., Major,F. and Cedergren,R. (1990) Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA. *Comput. Appl. Biosci.*, **6**, 325–331.
49. Jacobson,A.B. and Zuker,M. (1993) Structural analysis by energy dot plot of large mRNA. *J. Mol. Biol.*, **233**, 261–269.
50. Vance,L.M., Moscufo,N., Chow,M. and Heinz,B.A. (1997) Poliovirus 2C region functions during encapsidation of viral RNA. *J. Virol.*, **71**, 8759–8765.
51. Rowlands ,D.J. (1999) Foot-and-mouth disease viruses (picornaviridae). In Webster,R. and Granoff,A. (eds), *Encyclopedia of Virology,* 2nd Edn. Academic Press, pp. 586–575.
52. Le,S.Y. and Zuker,M. (1990) Common structures of the 5′ non-coding RNA in enteroviruses and rhinoviruses. *J. Mol. Biol.*, **216**, 729–741.
53. Leckie,G. (1998) Cloning and sequencing of the genome of human rhinovirus 9. PhD Thesis, University of Leicester, UK.

APPENDICES

Appendix A

The schematic drawings in Figure 2 are obtained from a typical strain with the strictly conserved structural features indicated by shadings. Here we give the reference sequences, alternative nomenclature where available, and the exact sequence positions of the outermost base pair of each of the indicated elements.

Aphthovirus: FMDV, strain C3Arg85 (accession no. AJ007347) (24):

A1 2–367, A2 (D) 587–640, A3 (H) 648–703, A4 (Ib) 769–846, A5 (J,K) 924–1033, A6 (N) 1037–1058.

Cardiovirus: TMEV, strain DA (accession no. M20301) (23,24,27):

C1 (A) 1–86, C2 (D) 524–554, C3 (F) 580–602, C4 (H) 610–680, C5 (Ib) 749–831, C6 (J,K) 909–1020, C7 (M) 1023–1042.

Parechovirus: HPeV-1, strain Harris (accession no. S45208 L00675) (30):

P1 (A) 14–67, P2 (D) 157–205, P3 (F) 239–253, P4 261–325, P5 327–373, P6 416–431, P7 452–464, P8 (J,K) 550–661. P6 and P7 are part of Ib.

Teschovirus: PTV-11, strain Dresden (accession no. AF296096), no S-fragment:

T1 19–166, T2 187–208, T3 212–242, T4 257–372, T5 401–415.

Hepatovirus: HAV, strain MBB (accession no. M20273) (31):

H1 (I) 5–37, H2 (II) 49–72, H3 (IV) 349–545, H4 (V) 577–688.

Enterovirus: Coxackievirus B, strain 1 Japan (accession no. M16560), (24,35,52):

E1 (I) 2–86, E2 (II) 127–165, E3 (III) 200–215, E4 (IV) 240–443, E5 (V) 477–534, E6 535–559, E7 (VI) 583–622, E8 (VII) 625–641.

Rhinovirus: strain HRV89 (accession no. M16248, A10937), (24,35,52):

R1 (I) 3–85, R2 (II) 128–166, R3 (III) 183–229, R4 (IV) 272–405, R5 (V) 422 462, R6 479–511, R7 536–548, R8 (VI) 582–624.

Appendix B: access codes

Aphthovirus/FMDV: AF18915, AJ133359, AF154271, AJ007347, X00871, X00429, X74812, M10975, AJ251473, M14409, M14408, L11360, Y18531, X74811, X83209;

Aphthovirus/Equine rhinitis A virus: X96870;

Cardiovirus: L22089, M81861, M22457, K01410, M16020, M20562, M20301, M80887;

Hepatovirus: X75214, AB020569, AB020567, AB020565, AB020564, D00924, X83302, M20273, K02990, M59808;

Parechovirus: AJ005695, AF055846, L02971;

Teschovirus: AJ011380, AF23176, AF231768, AF296104, AF296100, AF296102, AF296087, AF296107, AF296108, AF296109, AF296088, AF296089, AF296111, AF296112, AF296113, AF296090, AF296091, AF296115, AF296117,

Kobuvirus: AB010145.

AF296092, AF296093, AF296118, AF296094, AF296119, AF296096;

Rhinovirus: M16248, D00239, L24917, X02316, K02121, L05355, U60874, RV-0007 (Stanway et al. unpublished, accessible at the The Picornavirus Home Page: http://www.iah.bbsrc.ac.uk/virus/Picornaviridae/picornavirus.htm (Institute for Animal Health, UK), RV-0002 (53);

Enterovirus: V01150, X00595, D00625, K01392, X04468, U05876, AF177911, AF176044,

U22521, U22522, D00627, M16560, AF085363, U57056, M16572, X05690, S76772, X67706, AF083069, U16283, X92886, X84981, X80059, X79047, AF162711, D00435, D00538, D90457;

Erbovirus: X96871;