Nucleic Acids Research, 1993, Vol. 21, No. 10 2445–2451

Conserved tertiary structural elements in the 5′ nontranslated region of cardiovirus, aphthovirus and hepatitis A virus RNAs

Shu-Yun Le, Jih-H.Chen[1], Nahum Sonenberg[2] and Jacob V.Maizel Jr
Laboratory of Mathematical Biology, Division of Cancer Biology, Diagnosis and Centers, National Cancer Institute, NIH, Building 469, Room 151, Frederick, MD 21702, [1]Frederick Biomedical Supercomputing Center, Program Resources, Inc., NCI/FCRDC, DynCorp., Frederick, MD 21702, USA and [2]Department of Biochemistry and McGill Cancer Center, McGill University, Montreal, Quebec H3G 1Y6, Canada

Received January 28, 1993; Revised and Accepted April 6, 1993

ABSTRACT

Statistical analyses of RNA folding in 5′ nontranslated regions (5′NTR) of encephalomyocarditis virus, Theiler's murine encephalomyelitis virus, foot-and-mouth disease virus, and hepatitis A virus indicate that two highly significant folding regions occur in the 5′ and 3′ portions of the 5′NTR. The conserved tertiary structural elements are predicted in the unusual folding regions (UFR) for these viral RNAs. The theoretical, common structural elements predicted in the 3′ parts of the 5′NTR occur in a *cis*-acting element that is critical for internal ribosome binding. These structural motifs are expected to be highly significant from extensive Monte Carlo simulations. Nucleotides (nt) in the conserved single-stranded polypyrimidine tract for these RNAs are involved in a distinctively tertiary interaction that is located at about 15 nt prior to the initiator AUG. Intriguingly, the proposed common tertiary structure in this study shares a similar structural feature to that proposed in human enteroviruses and rhinoviruses. Based on these common structural features, plausible base pairing models between these viral RNAs and 18 S rRNA are suggested, which are consistent with a general mechanism for regulation of internal initiation of cap-independent translation.

INTRODUCTION

Although the nucleotide sequences of encephalomyocarditis virus (EMCV), Theiler's murine encephalomyelitis virus (TMEV), foot-and-mouth disease virus (FMDV), and hepatitis A virus (HAV) share little sequence homology with human enterovirus and rhinovirus, they all have similar genomic structure and gene organization (38). Unlike most eukaryotic mRNAs, they lack a 5′ cap structure (11, 29) and have an unusually long 5′ nontranslated region (5′NTR) that contains multiple AUG triplets prior to the initiator of the viral translation. It is already known that poliovirus RNA utilizes a mechanism of translation initiation which differs from the ribosomal scanning process employed in the translation of most eukaryotic mRNAs (2, 30). Combined with the experimental evidence in FMDV (1, 17), HAV (6), TMEV (48) and EMCV (10, 15, 16, 18) these data suggest that the 5′NTR of these viruses also contains an internal ribosome entry site (IRES), and has an essential functional organization similar to that of human enterovirus and rhinovirus. Although details of the structure of IRES in FMDV, HAV, TMEV and EMCV are unknown, a large segment preceding the open reading frame of the viral polyprotein has been demonstrated to be involved in the regulation of translation initiation. The sizes of the RNA stem-loops in the region of IRES are approximately 350 nucleotides (nt) in HAV and EMCV and 450 nt in FMDV and TMEV, and are similar to known structural motifs that are important components of cap-independent translation (6, 10, 17, 18, 48).

Mutagenesis of the sequences at the 5′NTR of FMDV, HAV, TMEV and EMCV suggests that IRES would essentially consist of a central RNA core structure to which ribosomes have a strong affinity (6, 10, 17, 18). Although some conserved RNA secondary structures in IRES of EMCV, TMEV, and FMDV (32, 10, 9), as well as HAV (6) have been proposed, it is not clear how the structural motifs direct the binding of the 40 S ribosomal subunit to the viral RNA in a cap-independent translation. Recently, we proposed a common higher-order RNA structural model consisting of 3 conserved tertiary structural elements for IRES of 22 human enterovirus and rhinovirus RNAs (22). The pseudoknot conformation in the common superstructure was assumed to be specifically recognized by components of the translation apparatus, 40 S ribosomal subunit and/or *trans*-acting factor(s) in a cap-independent translation of enteroviruses and rhinoviruses. According to the suggested common structural motif, a conserved base pairing model between enterovirus (or

* To whom correspondence should be addressed

rhinovirus) and 18 S rRNA can provide a rational interpretation for currently available mutagenesis data of 5'NTR of human enteroviruses.

To refine our understanding of the mechanism of internal initiation of cap-independent translation in the mRNAs of FMDV, EMCV, TMEV, and HAV the plausible tertiary structural motifs in their 5'NTR are proposed in this paper. The theoretical model is computed by a new procedure in which all thermodynamically favored helices are generated from a computational experiment wherein a large number of energy rules for RNA folding are simulated by fluctuating energy parameters. The validity of predicted helices is then examined by their conservation in all homologous sequences. Significant tertiary interactions are evaluated by extensive Monte Carlo simulations. The proposed base pairing models between these viral RNA and human 18 S rRNA share a similar structural feature to that formed between 18 S rRNA and the 3' portion of IRES of human enteroviruses and rhinoviruses. The feature is that a statistically significant and highly stable pseudoknot contains partial nt of the conserved polypyrimidine stretch. Moreover, the conserved pseudoknot occurs at a region about 15 nt prior to the open reading frame of polyproteins for EMCV, TMEV, FMDV and HAV. This signifies that the common structural features in these diverse RNAs function in the regulation of internal initiation of cap-independent translation.

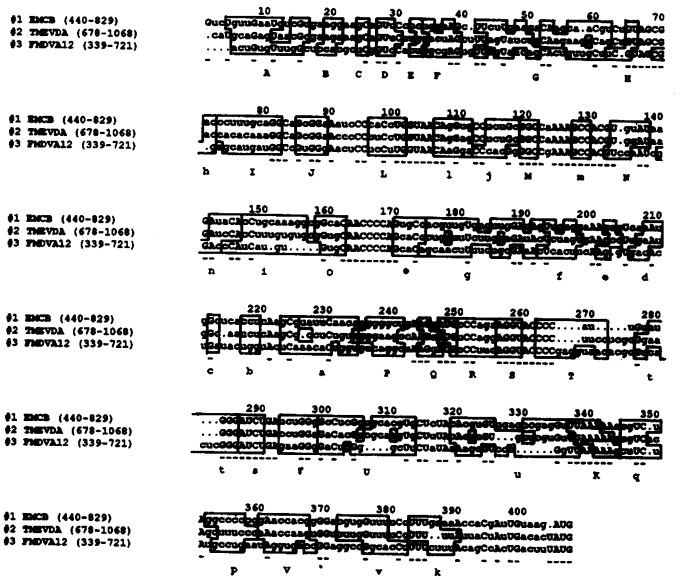

Figure 1. Alignment of UFR in the 3' portion of 5'NTR from EMCV (strain B), TMEV (strain DA), and FMDV (strain A12-119). They are referred to EMCB, TMEVDA, and FMDVA12 in the alignment. Deletions are denoted by dots. The conserved base pairing regions are marked by boxes and labeled by letter A–V and a–v. Among them, the boxes of K and k correspond to the tertiary structural element.

MATERIALS AND METHODS

In this study, 3 EMCV (10, 43), 3 TMEV (44–46), 9 FMDV (39), and 9 HAV sequences (6) were used. The unusual folding regions (UFR) in an RNA sequence are detected by a statistical method, SEGFOLD (19, 20, 41). The significance of an RNA folding is evaluated by comparing the predicted thermodynamic stability of the actual segment with those of its random permutations. A region that is both highly stable (lower stability score as compared to the other possible foldings) and more significant (lower significance score related to the random shuffling segments) is referred to as an UFR and is assumed to have a significant folding form, hinting a structural role for the sequence information.

The RNA higher-order structure is predicted by a new procedure. This procedure includes (a) an improved prediction of alternative RNA secondary structures (referred to as EFFOLD, ref. 23) based on fluctuating thermodynamic parameters, (b) the development of a list of all phylogenetic helices with thermodynamic stability from the prediction of all alternative structures and (c) a 'tree search' for generating the combination of compatible helices from the list. In the prediction of alternative structures, uncertainties of energy parameters for RNA folding are assumed to follow a normal distribution. The free energy parameters from Turner's energy rule (42) are perturbed about their values according to a normal distribution within the ranges of the experimental errors. In general, 50 'simulated energy rules' may be a suitable sample size for the simulation of an RNA folding (23). The lowest free energy structure is searched for each 'simulated energy rule.' In the simulation, the computed lowest free energies (say 50 observations) follow approximately a normal distribution (23) that can be used in the evaluation for the computed structures and helical stems in the next stage. Thus, all thermodynamically favored helices that occurred in the simulation are compiled.

In combination with the multiple sequence alignment (21) the phylogenetic conserved helices are selected from the compiled list mentioned above by the COMFOLD program (24). In the examination, the sequence position of each helix is allowed to shift a short range so that the bias in the sequence alignment can be partially eliminated. The information of compensatory base changes in these phylogenetic helices is saved in a stem list and can be used in the construction of an RNA structure at the next step. A 'tree search' is used to generate RNA higher-order structures by the combination of compatible helices from the list. In practice, the predicted common structures are constructed mainly by two ways. One is to generate a structure with the maximum number of base pairs. The second is based on the phylogenetic conserved information such as numbers of compensatory base changes in a substructure found in the previous examination. The detailed algorithm of the COMFOLD program will be described in another paper (24).

If a potential pseudoknotting or tertiary interaction is detected in the procedure, the pseudoknot needs to be assessed by using three scores, n_1, n_2 and z (8, 22). n_1 and n_2 are defined as the numbers of randomized sequences that have a pseudoknot (or tertiary interaction) thermodynamically more stable than those formed in the real biological sequence in the Monte Carlo simulation. The smaller n_1 and n_2, the more significant the tertiary interaction. The n_1 and n_2 scores differ in that the free energy contributed by tertiary interactions is calculated using different simulated rules (8). The score z is defined as $z = (nobs - rmean)/sd$ for the tertiary interaction pattern in the sequence, where *nobs* is the number of times the pattern occurs in the real sequence, *rmean* is the average number of the occurrence of the pattern in a set of random permutations, and *sd* is the standard deviation. Using the RNAKNOT program (8) potential pseudoknot structures in the sequence can also be searched thoroughly from the compiling list of the thermodynamically favored phylogenetic helices.

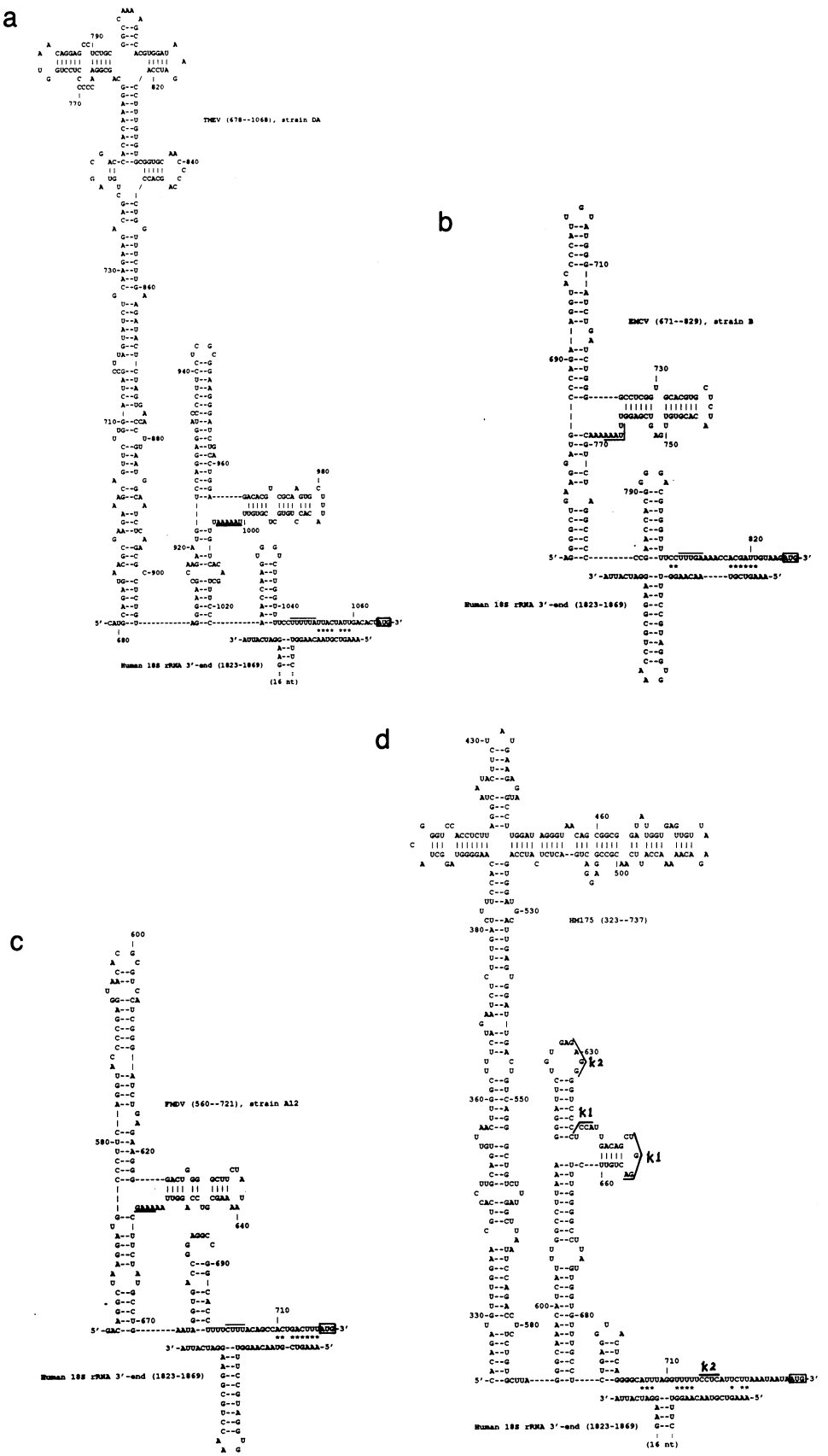

Figure 2. Predicted base pairing models between the 3′ portion of 5′NTR of viral RNAs and human 18 S rRNA. (**a**) TMEV (strain DA), (**b**) EMCV (strain DA), (**c**) FMDV (strain A12-119) and (**d**) HAV (wt HM175). In (b) and (c) the structural models have been simplified that 3′-end parts containing the distinct pseudoknot are only displayed. The tertiary interactions in TMEV, EMCV, FMDV and HAV are denoted by underlines. The RNA secondary structure of the 3′-end of human 18 S rRNA is based on the published model (12, 7). An asterisk between two sequences denotes the base pairing between viral RNAs and human 18 S rRNA.

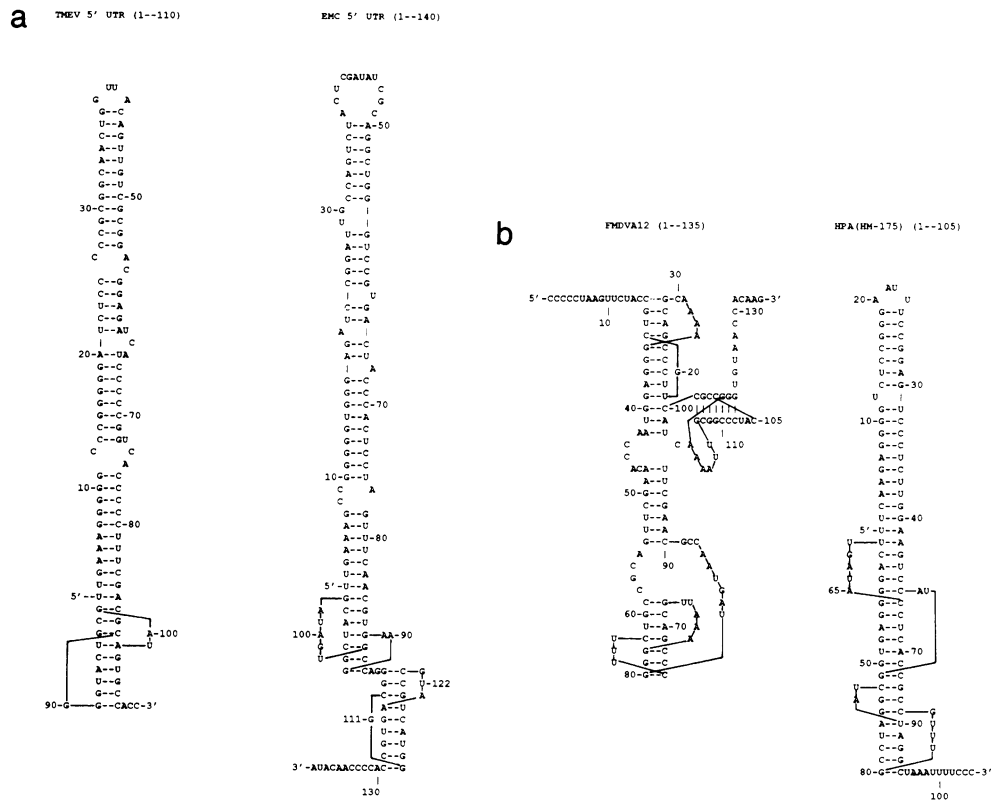

Figure 3. Predicted tertiary structures folded in the 5'-end of 5'NTR of viral RNAs. (**a**) TMEV (strain DA) and EMCV (strain DA), (**b**) FMDV (strain A12-119) and HAV (wt HM175). They are referred to TMEV, EMC, FMDVA12 and HPA in the figure. The tertiary structural elements are highly stable and statistically significant.

RESULTS

Possible higher-order structures in UFR

UFR in the 5'NTR of TMEV, EMCV, FMDV and HAV were searched by the program SEGFOLD(20, 41). The results show that UFR in the 5'NTR of TMEV, EMCV, FMDV and HAV occur mainly in both the 5' and 3' portions of their 5'NTR. Extensive simulations revealed that the segment of 366-nt (678−1043) in TMEV (strain DA) was a distinct UFR and its significance score (Sigscr) was −3.61 SD by using the Turner energy rule. Also UFR of EMCV (strain B), FMDV (strain A12-119), and HAV (wt HM175) were identified in the 362-nt fragment (442−803), 360-nt (338−697) and 390-nt fragment (312−701), respectively. Their corresponding significance scores were −4.56, −3.01, and −2.47 SD and were also calculated by the Turner energy rule. These localized UFR coincide with the suggested structural motifs in the 3' portion of these 5'NTR (6, 10, 17, 18) that were supposed to be important components of cap-independent translation for EMCV, FMDV and HAV.

Three detected UFR sequences of EMCV (strain B), TMEV (strain DA), and FMDV (strain A12-119) were compared by multiple sequence alignment (21). The sequence alignment of these three segments (extended to their initiators of the viral translation) is shown in Fig. 1. From Fig. 1, it is noteworthy that the three UFR shared the stronger sequence conservation than other parts in these 5'NTR. Among them, the TMEV sequence was both close to the EMCV and FMDV sequence. Thus, the TMEV sequence was taken as a master sequence for the prediction of their common structure for these three segments.

The 5'NTR (1065-nt) of TMEV (strain DA) was folded by EFFOLD(23). In the calculation, 200 simulated energy rules for RNA folding were generated. Forty alternative structures were chosen from 200 computed lowest free energy structures. Among 40 predicted structures, 490 thermodynamically favored helical stems were compiled. Based on the predicted stem list and the multiple sequence alignment of three UFR identified from TMEV, EMCV, and FMDV, the common structure for these segments were derived (Fig. 1−2). This computed-aided structural model is similar to that proposed by Pilipenko et al. (32) based on experimental and phylogenetic analyses.

Using the same procedure, a common RNA structure for UFR of HAV mRNAs was computed by means of the stem list derived from the HAV (wt HM175) and the multiple sequence alignment of 8 HAV sequences. The predicted structure is similar to that suggested by Brown et al. (6). However, the predicted folding form in the 3'-end (Fig. 2d) of 5'NTR is different from Brown's published model in the region from nt 590 to 700.

Prediction and evaluation for tertiary interactions in the 5'NTR of EMCV, TMEV, FMDV, and HAV

The possible tertiary interactions were found in the 5'- and 3'-end of the 5'NTR of these viruses. For the 3' portion of the 5'NTR, a plausible pseudoknot was predicted in EMCV, TMEV, and FMDV. The predicted pseudoknots share a conservative structural feature in these three RNAs. The tertiary structural elements are localized at 14−16 nt prior to the initiator of these RNAs. The distinctively tertiary interactions include nt that are in the regions 1001−1006 and 1046−1051 for the TMEV (strain

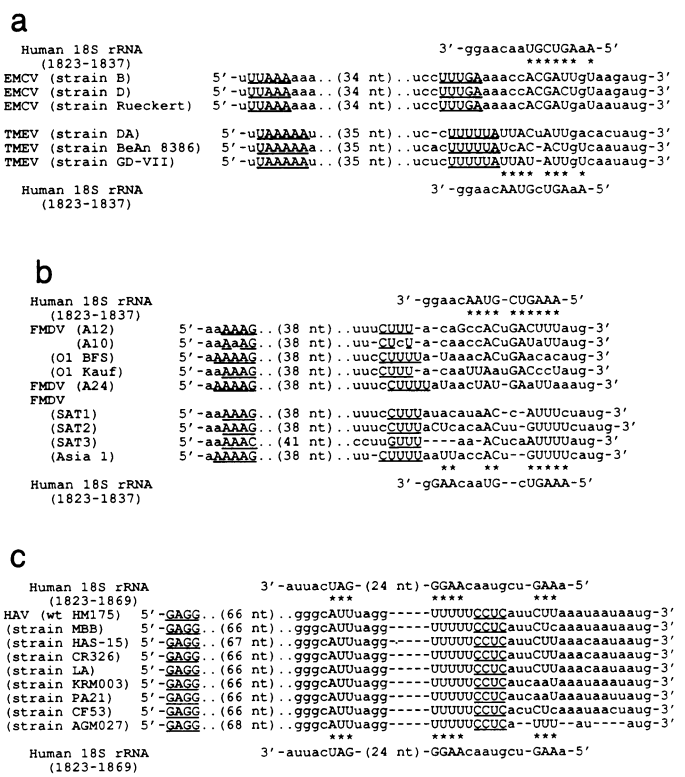

Figure 4. Base pairings between the 3'-end of human 18 S rRNA and single-stranded regions in the 5'NTR of (**a**) TMEV and EMCV (**b**) FMDV and (**c**) HAV. An asterisk between two sequences denotes the base pairing between these viral RNAs and human 18 S rRNA. The predicted tertiary interactions are denoted by underlines.

DA, see Fig. 2a), in the regions 760−765 and 806−810 for the EMCV (strain B, see Fig. 2b), and the regions 656−659 and 700−703 for the FMDV (strain A12-119, Fig. 2c). The predicted tertiary structural elements are supported by compensatory base changes observed in their multiple sequence alignments (Fig. 4a−b). Also, two pseudoknots were computed in the 3'-end of 5'NTR of HAV (Fig. 2d). One of them has been suggested by Brown et al. (6). The new found one occurred in the regions 629−632 and 717−720. The nt in the region 717−720 of HAV (wt, HM175) have been shown a strong tendency (6) of base pairing by the indication of the cleavage site with double-stranded-specific RNase (V_1). This predicted tertiary structure is perfectly conserved in nine HAV sequences (Fig. 4c).

The significance of these computed tertiary interactions was evaluated by Monte-Carlo simulations. The three significance scores, z, n_1 and n_2 of these tertiary structures were calculated from 1000 randomly shuffled sequences. The computed n_1 was 8 for the tertiary structural element in TMEV, 27 in EMCV, 47 in FMDV and 73 in HAV. The n_2 was 22 in TMEV, 68 in EMCV, 134 in FMDV and 131 in HAV. Also, these tertiary interactions had a large z value (e.g. 5.68 in HAV and 5.42 in EMCV). This means that we can expect on average of 27 (or 68) observations in 1000 random samples would form more stable pseudoknots than that of the predicted pseudoknot in EMCV. These measures indicate that the pseudoknot predicted in the 3'-end of the 5'NTR of EMCV (also in TMEV, FMDV, and HAV) is highly stable and statistically significant.

Using the same procedure, highly stable and statistically significant tertiary structures were also computed in the 5'-end of the 5'NTR of these viruses (Fig. 3). For EMCV, two successive pseudoknots occurred in the region 85−129 (Fig. 3a). The three scores of n_1, n_2 and z of the pseudoknot folded in the region 85−105 were 8, 8, and 7.0, respectively. The clustered pseudoknots were also predicted in the 5'-end of the 5'NTR of FMDV and HAV(HPA, Fig. 3b). All computed pseudoknots show unusual statistical significance and highly thermodynamic stability above the random background.

Base pairing model between human 18 S rRNA and the 5'NTR of EMCV, TMEV, FMDV, and HAV

Based on the computed common structure, a base pairing model between EMCV and human 18S rRNA is proposed (Fig. 2b). In our model, the single-stranded region at the 3'-end of 18 S rRNA (1825 to 1830) that is followed by a significantly stable stem-loop structure (19) can be complementary to the single-stranded region prior to the initiator AUG of EMCV. The partial nt within the single-stranded pyrimidine stretch was found to be entangled in a distinctively conserved tertiary structural element. The complementary sequence with 18 S rRNA just follows to the distinct pseudoknot. The base pairing models between TMEV and 18 S rRNA (Fig. 2a), and between FMDV and 18 S rRNA (Fig. 2c) show the same feature with EMCV. The base pairing model between HAV and 18 S rRNA is little different from EMCV, TMEV and FMDV. In the HAV model, most of base pairings between HAV mRNA and 18 S rRNA are prior to the pseudoknot (Fig. 2d). The complementary sequence in HAV also includes partial nt in the single-stranded polypyrimidine stretch. This feature is similar to that found in the base pairing model between human enteroviruses and 18 S rRNA (22).

Our results show that the plausible base pairings between these viral RNAs and 18 S rRNA are conserved in all 24 tested sequences (Fig. 4a−c). The proposed pseudoknots for EMCV, TMEV, and FMDV are supported by the covariant mutation found in 15 sequences. The sequence complementary to these RNAs presented here is also evolutionally conserved among all 18 S rRNAs of eukaryotes (12). The chemical and RNase probing results also suggest that this complementary sequence is not base paired in the 18 S rRNAs (37).

DISCUSSION

The tertiary structures predicted in the 5' portion of the 5'NTR of EMCV, FMDV and HAV are clustered. In this study, statistical test shows that these predicted higher-order structures are highly significant above random background. Our entirely computed structural models are similar to those suggested by Duke et al. (10), Brown et al. (6) and Clarke et al. (9).

The theoretical higher-order structure proposed in the 3' portion of 5'NTR for EMCV, TMEV and FMDV is highly conserved. This higher-order structure includes 24 helical stems labelled in letters A−J/a−j and L−V/l−v, and a tertiary structural element labelled in letters K/k (Fig. 1). It can be divided into three parts, a large sized stem-loop structure consisting of helical stems A−O (except K), a middle sized stem-loop (stems P−U) and a hairpin structure (stem V) in the 3'-end (Fig. 1 and 2). The tertiary structural element K connects parts 2 and 3 and therefore contributes to the formation of a compact superstructure. Each of these structural elements is supported by compensatory base changes except for small helical stems M and S (see Fig. 1).

There is no direct homology in the primary and secondary RNA structures between sequences of the HAV family and sequences of EMCV, TMEV and FMDV. However, the consensus RNA structures predicted in the 3′ portion of 5′NTR for eight HAV mRNAs are similar in principle to that folded in EMCV, TMEV, and FMDV. The predicted higher-order structures in IRES of these diverse viral RNAs consist of three parts that have similar shape and size. Moreover, the predicted tertiary structural elements in these models have also been conserved in their positions relative to the initiator AUG (14−16 nt prior to the initiator).

It should be mentioned that some sequences (e.g. GGAA) corresponding to tetraloops were predicted to engage in base pairing in the proposed model, though an assumed extra −2 kval/mol (50) for each possible tetraloop have been incorporated in the program EFFOLD. It is clear that the more accurate prediction requires knowledge of the thermodynamics of the tertiary interactions and unusual loop structures such as tetraloops.

This paper is another case of successfully detecting interesting structural motifs relative to functional properties using the UFR approach. UFR are sequence segments that are statistically more significant than those randomly shuffled sequences. Numerous results from computer simulations associated with extensive mutagenesis experiments (34, 36, 47) have shown that the distinct UFR identified in retroviral mRNAs by our procedures correlates with various biological functions in mRNA processing and translation. Our results strongly suggest that the significant folding forms in UFR of EMCV, TMEV, FMDV and HAV play an important structural role for their sequence information. The high structural similarity between these viral RNAs of differing sequence implies that the proposed common RNA higher-order structures in this paper have a general function in the internal initiation of cap-independent translation.

Several lines of evidence indicate that 5′NTR of poliovirus, human rhinovirus, EMCV, TMEV, FMDV, and HAV promote cap-independent translation initiation in which the ribosome is directed to bind with IRES without scanning from the 5′ terminus of 5′NTR (1−3, 6, 14, 16−18, 30, 48). The RNA secondary structures folded in IRES of these RNAs were assumed to play critical roles in translation initiation (6, 10, 17, 27−28, 33). The consensus superstructures presented in this study include those functional structure motifs suggested in the 5′NTR of EMCV, FMDV and HAV (6, 10, 17). Intriguingly, the proposed superstructures include a consensus pseudoknot structure that contains partial nt within a conserved polypyrimidine subsequence. The effects of a single nt mutation in the pyrimidine stretch sequence of FMDV (17) have demonstrated that the nt of the first, third and forth position in segment CUUUU are absolutely required for translational initiation. The single pyrimidine-purine exchanges of the third base reduced translational activities of the template up to 15% *in vitro*. The pyrimidine-purine exchanges of the first, third and forth bases would destroy the tertiary interactions in our theoretical superstructural model of FMDV. Similar cases have also been observed in polio, coxsackie, and human rhinoviruses (22).

Experimental results from numbers of laboratories have also demonstrated that pseudoknot structures function as a target for proteins binding to mRNAs (4, 5, 25, 35, 40). The structural feature in a pseudoknot cannot be replaced by a hairpin structure that was composed of an equivalent set of base-pairs. If the distinct configuration of a pseudoknot is destroyed, its biological function will be reduced or abolished. Although the proposed tertiary structure in this study needs to be verified through experiments, our model provides a rational interpretation for available mutagenesis data in the translation initiation of these viral RNAs (17). Kaminski et al. (18) have suggested that a large superstructure directed a precise binding of ribosomes to the 3′-end of IRES for the translational initiation of the EMCV RNA. The single-stranded 5-nt CUUUA of the hairpin-loop structure in strain Rueckert of EMCV (in strain B it is CUCUA, 738−742, and hairpin was labelled by U/u in Fig. 2) has been reported as an efficient binding site for eukaryotic initiation factor eIF-2/2B (13). This binding site is located nearby the predicted pseudoknot in EMCV. Also, Meerovitch et al. (26) indicated that a p52 protein could specifically bind the 3′ portion (559−624) of the 5′NTR of poliovirus. Taken together, the proposed significant pseudoknot structures formed in EMCV, TMEV, FMDV and HAV are assumed to function as an efficient binding site for ribosome and/or other initiation factors in the translational initiation. This is similar to that suggested for the conserved pseudoknots of polio, coxsackie, and human rhinoviruses (22). However, the predicted tertiary structural elements in EMCV, TMEV, FMDV, as well as HAV, have some different features with that identified in polio, coxsackie, and rhinoviruses. The major difference is that the pseudoknots in IRES of EMCV, TMEV, FMDV and HAV are immediately followed by the initiator of viral translation. In the case of human enterovirus and rhinovirus, the predicted pseudoknots are at least 140-nt prior to the initiator. It can be expected that enterovirus and rhinovirus follow somewhat different translation strategies than EMCV, TMEV, FMDV and HAV.

The theoretical base pairing models between human 18 S rRNA and these RNAs of EMCV, TMEV, FMDV and HAV also share similar structural features to that formed between 18 S rRNA and human enterovirus and rhinovirus RNAs. The partial sequences involved in the two types of base pairing model are all located at the same region (1823−1837) in the 3′-end of human 18 S rRNA (22) that is not base-paired in its established RNA secondary structure model (7, 12, 37). Although the theoretical base pairing model awaits verification through experiments, the sequence and structural analyses for those variants described by Pestova et al. (31) and Pilipenko et al. (33) show that the base pairing interaction between poliovirus mRNA and human 18 S rRNA would be destabilized in the mutated transcripts that template activity was markedly lower than that of wt. We suggest that the complementary sequence to the 18 s rRNA in IRES of EMCV, TMEV, FMDV and HAV, similar to that in IRES of the enterovirus and rhinovirus, can play an functional role in a manner analogous to that of the Shine-Dalgarno sequence documented for the translational initiation of prokaryotic mRNAs.

Another example of internal ribosome binding has also been demonstrated in the translation of a cellular mRNA, human immunoglobulin heavy-chain binding protein (BiP) mRNA (49). A conserved higher-order structure that contains a tertiary structural element prior to the initiator has also been computed in the 5′NTR of BiP family sequences (results are not shown in this paper). These plausible common structure models add to the known intriguing properties of 5′NTR of all picornaviral mRNAs. We consider that it is the structural role of IRES detected in picornaviruses is crucial in the internal initiation of cap-independent translation.

ACKNOWLEDGMENTS

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

REFERENCES

1. Belsham, G.J., and Brangwyn, J.K. (1990) *J. Virol.* **64,** 5389–5395.
2. Bienkowska-Szewczyk, K. and Ehrenfeld, E. (1988) *J. Virol.* **62,** 3068–3072.
3. Borman, A., and Jackson, R. J. (1992) *Virology* **188,** 685–696.
4. Brierley, I., Digard, P., and Inglis, S.C. (1989) *Cell* **57,** 537–547.
5. Brierley, I., Rolley, N.J., Jenner, A.J., and Inglis, S.C. (1991) *J. Mol. Biol.* **220,** 889–902.
6. Brown, E.A., Day, S.P., Jansen, R.W., and Lemon, S.M. (1991) *J. Virol.* **65,** 5828–5838.
7. Chan, Y.-L., Gutell, R., Noller, H.F., and Wool, I.G. (1984) *J. Biol. Chem.* **259,** 224–230. 8. Chen, J.-H., Le, S.-Y., and Maizel, J.V.,Jr. (1992) *Comp. Appl. Biosci.* **8,** 243–248.
9. Clarke, B.E., Brown, A.L., Currey, K.M., Newton, S.E., Rowlands, D.J., and Carroll, A.R. (1987) *Nucl. Acids Res.* **15,** 7067–7079.
10. Duke, G.M., Hoffman, M.A., and Palmenberg, A.C. (1992) *J. Virol.* **66,** 1602–1609.
11. Hewlett, M.J., Rose, J.K., and Baltimore, D. (1976). *Proc. Natl. Acad. Sci. USA* **73,** 327–330.
12. Huysmans, E., and Wachter, R. (1986) *Nucleic Acids Res.* **14,** r73–r118.
13. Scheper, G.C., Thomas, A., and Voorma, H.O. (1991) *Biochim. Biophys. Acta* **1089,** 220–226.
14. Trono, D., Andino, R., and Baltimore, D. (1988) *J. Virol.* **62,** 2291–2299.
15. Jang, S.K., Krausslich, H.-G., Nicklin, M.J.H., Duke, G.M., Palmenberg, A.C., and Wimmer, E. (1988) *J. Virol.* **62,** 2636–2643.
16. Jang, S.K., Davies, M.V., Kaufman, R.J., and Wimmer, E. (1989) *J. Virol.* **63,** 1651–1660.
17. Kuhn, R., Luz, N., and Beck, E. (1990). *J. Virol.* **64,** 4625–4631.
18. Kaminski, A., Howell, M., and Jackson, R.J. (1990) *EMBO J.* **9,** 3753–3759.
19. Le, S.-Y., and Maizel, J.V.,Jr. (1989) *J. Theoret. Biol.* **138,** 495–510.
20. Le, S.-Y., Chen, J.-H., and Maizel, J.V.,Jr. (1990) *In* 'Structure \ & Methods: Human Genome Initiative and DNA Recombination', (Sarma, R.H. and Sarma, M.H., Eds), Vol. 1, pp. 127–136. Adenine Press.
21. Le, S.-Y., and Zuker, M. (1990) *J. Mol. Biol.* **216,** 729–741.
22. Le, S.-Y., Chen, J.-H., Sonenberg, N., and Maizel, J.V.,Jr. (1992) *Virology* **191,** 858–866.
23. Le, S.-Y., Chen, J.-H., and Maizel, J.V.,Jr. (1993) *Nucl. Acids Res.* **21,** in press.
24. Le, S.-Y. and Maizel, J.V.,Jr. to be submitted.
25. McPheeters, D.S., Stormo, G.D., and Gold, L. (1988) *J. Mol. Biol.* **201,** 517–535.
26. Meerovitch, K., Pelletier, J., and Sonenberg, N. (1989) *Genes Dev.* **3,** 1026–1034.
27. Meerovitch, K., Nicholson, R., and Sonenberg, N. (1991) *J. Virol.* **65,** 5895–5901.
28. Nicholson, R., Pelletier, J., Le, S.-Y., and Sonenberg, N. (1991) *J. Virol.* **65,** 5886–5894.
29. Nomoto, A., Lee, Y.F., and Wimmer, E. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 375–380.
30. Pelletier, J. and Sonenberg, N. (1988) *Nature* **334,** 320–325.
31. Pestova, T.V., Hellen, C.U.T., and Wimmer, E. (1991) *J. Virol.* **65,** 6194–6204.
32. Pilipenko, E.V., Blinov, M., Chernov, B.K., Dmitrieva, T.M., and Agol, V.I. (1989) *Nucl. Acids Res.* **14,** 5701–5710.
33. Pilipenko, E.V., Gmyl, A.P., Maslova, S.V., Svitkin, Y.V., Sinyakov, A.N., and Agol, V.I. (1992) *Cell* **68,** 119–131.
34. Malim, M, Hauber, J., Le, S.-Y., Maizel, J.V., and Cullen, B.R. (1989) *Nature* **338,** 254–257.
35. Portier, C., Philippe, C., Dondon, L., Grunberg-Manago, M., Ebel, J.P., Ehresmann, B., and Ehresmann, C. (1990) *Biochim. Biophys. Acta* **1050,** 328–336.
36. Fenrick, R., Malim, M., Hauber, J., Le, S.-Y., Maizel, J.V., and Cullen, B.R. (1989) *J. Virol.* **63,** 5006–5012.
37. Rairkar, A., Rubino, H.M., and Lockard, R.E. (1988) *Biochemistry* **27,** 582–592.
38. Rueckert, R.R. (1985) *In* 'Virology', (B.N. Fields, D.M. knipe, R.M. Chanock, J.L. Melnick, and B. Roizman eds.) pp. 705–738. Raven Press, New York.
39. Sangar, D.V., Newton, S.E., Rowlands, D.J., and Clarke, B.E.(1987) *Nucl. Acids Res.* **15,** 3305–3315.
40. Tang, C.K. and Draper, D.E. (1989) *Cell* **57,** 531–536.
41. Le, S.-Y., Malim, M.H., Cullen, B.R., and Maizel, J.V.,Jr. (1990) *Nucl. Acids Res.* **18,** 1613–1623.
42. Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N. Caruthers, M.H., Neilson, T., and Turner, D.H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9373–9377.
43. Cohen, S.H., Naviaux, R.K., Vanden Brink, K.M., and Jordan, G.W. (1988) *Virology* **166,** 603–607.
44. Pevear, D.C., Calenoff, M., Rozhon, E., and Lipton, H.L. (1987) *J. Virol.* **61,** 1507–1516.
45. Ohara, Y., Stein, S., Fu, J., Stillman, L. Klaman, L., and Roos, R.P. (1988) *Virology.* **164,** 245–255.
46. Pevear, D.C., Borkowski, J., Calenoff, M., Oh, C.K., Ostrowaki, B. and Lipton, H.L. (1988) *Virology* **61,** 1–12.
47. Tiley, L.S., Brown, P.H., Le, S.-Y., Maizel, J.V., Clements, J.E. and Cullen, B.R. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 7497–7501.
48. Randyopadhyay, P.K., Wang, C. and Lipton, H.L. (1992) *J. Virol.* **66,** 6249–6256.
49. Macejak, D.G. and Sarnow, P. (1991) *Nature* **353,** 90–94.
50. Jaeger, J.A., Turner, D.H. and Zuker, M. (1990) *Proc. Natl. Sci. USA* **86,** 7706–7710.