Conserved Tertiary Structure Elements in the 5' Untranslated Region of Human Enteroviruses and Rhinoviruses

SHU-YUN LE,[1] JIH-H. CHEN,[2] NAHUM SONENBERG,[3] AND JACOB V. MAIZEL

Laboratory of Mathematical Biology, Division of Cancer Biology, Diagnosis and Centers,
National Cancer Institute, NIH, Building 469, Room 151, Frederick, Maryland 21702

A combination of comparative sequence analysis and thermodynamic methods reveals the conservation of tertiary structure elements in the 5' untranslated region (UTR) of human enteroviruses and rhinoviruses. The predicted common structural elements occur in the 3' end of a segment that is critical for internal ribosome binding, termed "ribosome landing pad" (RLP), of polioviruses. Base pairings between highly conserved 17-nucleotide (nt) and 21-nt sequences in the 5' UTR of human enteroviruses and rhinoviruses constitute a predicted pseudoknot that is significantly more stable than those that can be formed from a large set of randomly shuffled sequences. A conserved single-stranded polypyrimidine tract is located between two conserved tertiary elements. R. Nicholson, J. Pelletier, S.-Y. Le, and N. Sonenberg (1991, *J. Virol.* 65, 5886–5894) demonstrated that the point mutations of 3-nt UUU out of an essential 4-nt pyrimidine stretch sequence UUUC abolished translation. Structural analysis of the mutant sequence indicates that small point mutations within the short polypyrimidine sequence would destroy the tertiary interaction in the predicted, highly ordered structure. The proposed common tertiary structure can offer experimentalists a model upon which to extend the interpretations for currently available data. Based on these structural features possible base-pairing models between human enteroviruses and 18 S rRNA and between human rhinoviruses and 18 S rRNA are proposed. The proposed common structure implicates a biological function for these sequences in translational initiation. © 1992 Academic Press, Inc.

INTRODUCTION

Poliovirus, coxsackievirus, and human rhinovirus are members of the *Picornaviridae* family. They contain an unusually long, conserved 5' UTR. All picornavirus mRNAs, unlike most eukaryotic mRNAs, lack a 5' cap structure (Hewlett *et al.*, 1976; Nomoto *et al.*, 1976). *In vitro* and *in vivo* experiments have shown that the internal initiation of cap-independent translation requires a large segment (nt 140 to 630) of the 5' UTR of poliovirus (Pelletier and Sonenberg, 1988; Trono *et al.*, 1988a,b). The importance of several stem–loop structures in the 5'-end of the segment termed "ribosome landing pad" (RLP) were demonstrated in poliovirus type 2 (Lansing strain) *in vivo* (Nicholson *et al.*, 1991) and *in vitro* (Meerovitch *et al.*, 1991). The exact ribosome binding sequence, however, is not known.

Genetic analysis (Trono *et al.*, 1988b; Kuge and Nomoto, 1987) of poliovirus mutants suggests that the RLP assumes a distinct tertiary structure that can be recognized by trans-acting factor(s) and/or 40 S ribo-

somes. Although some common RNA secondary structures in the 5' UTR of the enterovirus and rhinovirus had been derived by phylogenetic comparative sequence analysis (Pilipenko *et al.*, 1989; Rivera *et al.*, 1988; Skinner *et al.*, 1989) or by combining comparative sequence analysis and a thermodynamic suboptimal folding method (Le and Zuker, 1990), the existence of a three-dimensional structure was not proposed directly.

A pseudoknot is an important structural element that can significantly contribute to tertiary interactions in RNA folding (Puglisi *et al.*, 1988). Experimental evidence has implicated an important role for pseudoknots in protein recognition, translational mediation and the replication of the plant viruses (Tang and Draper, 1989; McPheeters *et al.*, 1988; Brierley *et al.*, 1989, 1991; Pleij *et al.*, 1985; Portier *et al.*, 1990; Bredenbeek *et al.*, 1990). Mutational analyses in the α operon mRNA (Tang and Draper, 1989) and rpsO mRNA (Portier *et al.*, 1990) of *Escherichia coli*, as well as in the autogenous regulation of gene 32 protein expression in bacteriophage T4 (McPheeters *et al.*, 1988) have also suggested that pseudoknot structures are specifically recognized by components of the translation apparatus. To understand the mechanism of internal initiation of cap-independent translation in the poliovirus, coxsackievirus, and human rhinovirus, the determina-

---

[1] To whom reprint requests should be addressed.
[2] Biomedical Supercomputing Center, Program Resources, Inc., NCI/FCRDC, DynCorp., Frederick, MD 21702.
[3] Department of Biochemistry and McGill Cancer Center, McGill University, Montreal, Quebec, Canada H3G 1Y6.

tion of tertiary structure of the RLP in their 5' UTRs is necessary.

Presently available programs of energy minimization for predicting secondary structure exclude the pseudoknot motif and other possible tertiary interactions because thermodynamic parameters for the structural elements of tertiary interactions are unknown.

In this paper, we use a new method (Chen et al., 1992) to predict possible pseudoknots or tertiary structures formed in the 5' UTR. The method is a combination of the prediction of significantly stable RNA secondary structures and evaluation of statistical significance for tertiary interactions. The validity of the predicted tertiary interactions in the sequence is then examined by conservation in all homologous sequence RNAs in the 5' UTRs of the poliovirus, coxsackievirus, and human rhinovirus.

Based on the theoretical structure model, potential base pairings between RLP and human 18 S rRNA are determined. A functional role of the proposed common tertiary structure of RLP is suggested in the recognition of trans-acting factors and/or 40 S ribosomes to the RLP. It is proposed that the pyrimidine stretch downstream from the conserved 21-nt sequence contributes to an efficient translation of viral mRNA in a manner analogous to that of the Shine-Dalgarno sequence in prokaryotic systems (Pilipenko et al., 1992).

MATERIALS AND METHODS

The sequence data of 11 polioviruses, 6 coxsackieviruses, and 5 human rhinoviruses were obtained from the GenBank and EMBL databases. Except for PV2W (poliovirus type 2, strain W-2) (Pevear et al., 1990), CA21 (coxsackievirus A21) (Hughes et al., 1989), CA9 (coxsackievirus A9) (Chang et al., 1989), and CVB3 (coxsackievirus B3) (Klump et al., 1990), the other 18 sequences are the same as those reported in a previous paper (Le and Zuker, 1990).

The multiple alignment of 22 sequences was performed using Zuker's multiple alignment program, MAL (Le and Zuker, 1990). MAL was written in Fortran 77 and could run under IRIS/UNIX or VAX/VMS.

The pseudoknot structures in these sequences were predicted using the RNAKNOT program (Chen et al., 1992). RNAKNOT was written in Fortran 77 on a CRAY-YMP/8128 (UNICOS) system. The method for predicting pseudoknots or tertiary structures in a given RNA sequence can be summarized as follows:

(1) Construct a list (List A) of nonoverlapped regions that are thermodynamically and statistically significant in the given sequence (Le et al., 1990).

(2) Create a list (List B) of nonoverlapped regions that are thermodynamically significant by comparing

only their stability scores (Le et al., 1990) for each segment in the sequence. Each item in List B is nonoverlapped with the regions in List A.

(3) Search and build a list of all possible pseudoknots that satisfy steric constraints and would not overlap with regions in List A and List B from both the 5' and 3' sides of the hairpin loop structure under consideration for each item in List A.

(4) Eliminate less stable pseudoknots from the list mentioned in the step 3.

(5) Evaluate the statistical significance of these computed potential pseudoknots using three scores z, n_1, and n_2.

In the evaluation, the z score is defined as $z = (n\text{obs} - r\text{mean})/\text{SD}$ for each tertiary interaction pattern in the sequence, where $n\text{obs}$ is the number of times the pattern occurs in the real sequence, $r\text{mean}$ is the average number of times the occurrence of the pattern in a set of scrambled sequences whose composition and length are identical to the actual sequence, and SD is the standard deviation. The scores n_1 and n_2 are defined as the numbers of randomized sequences that have a pseudoknot thermodynamically more stable than the real sequence. n_1 and n_2 differ in that the free energy contributed by a pseudoknot is calculated using different simulation rules. Thus, a large value of z and/or a small value of n_1 or n_2 might be an indication of nonrandomness of the occurrence of the pseudoknot folded in the sequence. It provides supporting evidence for the existence of the pseudoknot in RNA molecules.

RESULTS

Possible tertiary interactions in the 5' UTR of polioviruses

The possible pseudoknot interactions in the 5' UTR of PV2L (poliovirus type 2, Lansing), HRV89 (human rhinovirus type 89) and CB4 (coxsackievirus B4) were searched throughout the sequence by the program RNAKNOT. By extensive simulations five potential pseudoknots in PV2L were selected. Two were located at the 5'-end of the UTR and three other tertiary interactions were found in the RLP of polioviruses and its downstream region. The two potential pseudoknots predicted in the 5'-end of PV2L UTR are highly conserved in 11 polioviruses. The occurrence of the two pseudoknots, however, will destroy a significantly stable RNA secondary structure formed in the folding region 2–88 (Le and Zuker, 1990). Furthermore, the two pseudoknots are not conserved in human rhinoviruses and coxsackieviruses. Therefore, the two tertiary

structures cannot be considered significant in enteroviruses and rhinoviruses.

The predicted pseudoknots, K2 (tertiary interactions: between the region 497–501 and 550–554) and K3 (579–581 and 600–602) in the RLP of PV2L were found to be totally conserved in all 22 human enteroviruses and rhinoviruses. The hairpins involved in the two common pseudoknots K2 and K3 correspond to two significant structure motifs designated as SM2 (484–514) and SM3 (585–620) in a previous paper (Le and Zuker, 1990). The 5-nt UGGGU in totally conserved 21-nt sequence of the 22 picornaviruses participates in the tertiary interaction of pseudoknot K2. Also, compensatory base changes were detected in the computed pseudoknot K3. That is, two or more independent covariations occur in the complementary sequences such that the predicted pseudoknot K3 is preserved in all sequences. For instance, 5'-GUG/CAC-3' base pairings in poliovirus type 1 vary to 5'-AUG/CAU-3' in PV2L and poliovirus type 3, to 5'-AUU/AGU-3' in HRV89, and to 5'-AUAU/AUAU-3 in HRVPI and HRV14 (see Fig. 1).

Based on comparative sequence analysis, another well-conserved tertiary structural element, K1, was detected (459–464 and 559–564 in PV2L) in all 22 picornaviruses. The involved hairpin in the conserved tertiary structure element is a small hairpin of 4 bp that is thermodynamically significant. This small hairpin is a structural element in the common RNA secondary structure for these viruses (Le and Zuker, 1990). Interestingly, the perfectly conserved 6-nt sequence GUG-UUU (559 to 564) is included in the distinct tertiary interaction. The conservation of these three conserved tertiary interactions (K1, K2, and K3) in 22 sequences is shown in Fig. 1.

Evaluation of predicted tertiary interactions in the 5' UTR of human enteroviruses and rhinoviruses

The significance of the predicted conserved tertiary structure elements in 22 sequences of human enteroviruses and rhinoviruses was assessed individually. The three significance scores, z, n_1, and n_2, of these pseudoknots or tertiary structures (see Materials and Methods) were calculated from 1000 randomly shuffled sequences. The results are listed in Table 1. For most of the pseudoknots K1 and K2 the values for n_2 are not large (i.e., their ratios are less than 0.1). Both K1 and K2 have small n_1 values. For example, the three significance scores of z, n_1 and n_2 of the pseudoknot K1 in PV2L are 8.36, 31, and 84. It means that we can expect the average 31 (or 84) observations in 1000 random samples would form more stable pseudoknots than that of the predicted pseudoknot K1 in PV2L. It

indicates that the pseudoknot K1 predicted in PV2L is highly stable and statistically significant. Among these three significance scores, the parameter z may not be a sensitive measure from our experience (Chen et al., 1992). The large values in z scores (say $z > 4.5$) for these pseudoknots shows that the occurrences of the specific base-pairing patterns in their tertiary interactions are not random.

Conserved tertiary structure elements in the 5' UTR of human enteroviruses and rhinoviruses

Figures 2–4 show the common tertiary structures in the 5' UTR of PV2L, HRV89, and CB4, respectively. In these common structural models, an 11-nt sequence of a 21-nt conserved sequence (545 to 565) is involved in two conserved tertiary structure elements, K1 and K2. The AUG-7 (nt 588 to 590 in poliovirus type 2, whose position relative to the pyrimidine stretch is relatively constant in all picornaviruses; the number changes depend on the viral serotype) downstream from the 21-nt sequence is embraced in the common tertiary structure element K3.

Based on the computed common tertiary structure, a new base-pairing model between PV2L and human 18 S rRNA is proposed (Fig. 2). In the model, the single-stranded purine rich region at the 3'-end of 18 S rRNA (1823 to 1837) that is followed by a significantly stable stem–loop structure (Le and Maizel, 1989) can be complementary to the pyrimidine stretch between the conserved 21-nt sequence and AUG-7. The single-stranded pyrimidine stretch is positioned between the two common tertiary structure elements. The base-pairing models between HRV89 and human 18 S rRNA and between CB4 and 18 S rRNA are the same as that of PV2L (Figs. 3 and 4). The results show that the possible base pairings between the viral mRNAs and human 18 S rRNAs are well conserved in the human enteroviruses and rhinoviruses (Fig. 5). The possible interactions between these mRNAs of picornaviruses and 18 S rRNAs include most nt of the polypyrimidine tract downstream from the conserved 21-nt sequence. The sequence complementary to the viral pyrimidine stretch is also evolutionary conserved among 18 S rRNAs of eukaryotes. The chemical and RNase probing results also suggest that this complementary sequence is not base paired in the 18 S rRNAs (Rairkar et al., 1988).

DISCUSSION

Results from a number of laboratories (Jang et al., 1988, 1989; Belsham and Brangwyn, 1990; Kuhn et al., 1990; Pelletier and Sonenberg, 1988; Bienkowska-Szewczyk and Ehrenfeld, 1988; Jackson et al., 1990)

```
                              10        20        30        40        50        60        70        80        90
#1  PV1A  UTR(444-619)
#2  PV1B  UTR(447-621)
#3  PV1S  UTR(447-621)
#5  PV2S  UTR(447-622)
#10 PV3XX UTR(450-625)
#4  PV2L  UTR(448-623)
#11 PV2W  UTR(451-626)
#6  PV3L  UTR(450-624)
#7  PV3P3 UTR(450-624)
#8  PV3S  UTR(452-626)
#9  PV3X3 UTR(450-624)
#12 CB1   UTR(450-626)
#13 CB3   UTR(449-625)
#17 CVB3  UTR(450-626)
#16 CA9   UTR(452-628)
#14 CB4   UTR(452-628)
#15 CA21  UTR(445-620)
#18 HRV14 UTR(453-624)
#22 HRVPI UTR(457-628)
#19 HRV1B UTR(448-622)
#20 HRV2  UTR(438-610)
#21 HRV89 UTR(446-618)

          A    K1   a    B    C    D    E    K2   e    d    F    f    c    b
```

```
                              100       110       120       130       140       150       160       170
#1  PV1A  UTR(444-619)
#2  PV1B  UTR(447-621)
#3  PV1S  UTR(447-621)
#5  PV2S  UTR(447-622)
#10 PV3XX UTR(450-625)
#4  PV2L  UTR(448-623)
#11 PV2W  UTR(451-626)
#6  PV3L  UTR(450-624)
#7  PV3P3 UTR(450-624)
#8  PV3S  UTR(452-626)
#9  PV3X3 UTR(450-624)
#12 CB1   UTR(450-626)
#13 CB3   UTR(449-625)
#17 CVB3  UTR(450-626)
#16 CA9   UTR(452-628)
#14 CB4   UTR(452-628)
#15 CA21  UTR(445-621)
#18 HRV14 UTR(453-624)
#22 HRVPI UTR(457-628)
#19 HRV1B UTR(448-622)
#20 HRV2  UTR(438-610)
#21 HRV89 UTR(446-618)

          k2        k1                   K3        H    k3        h
```
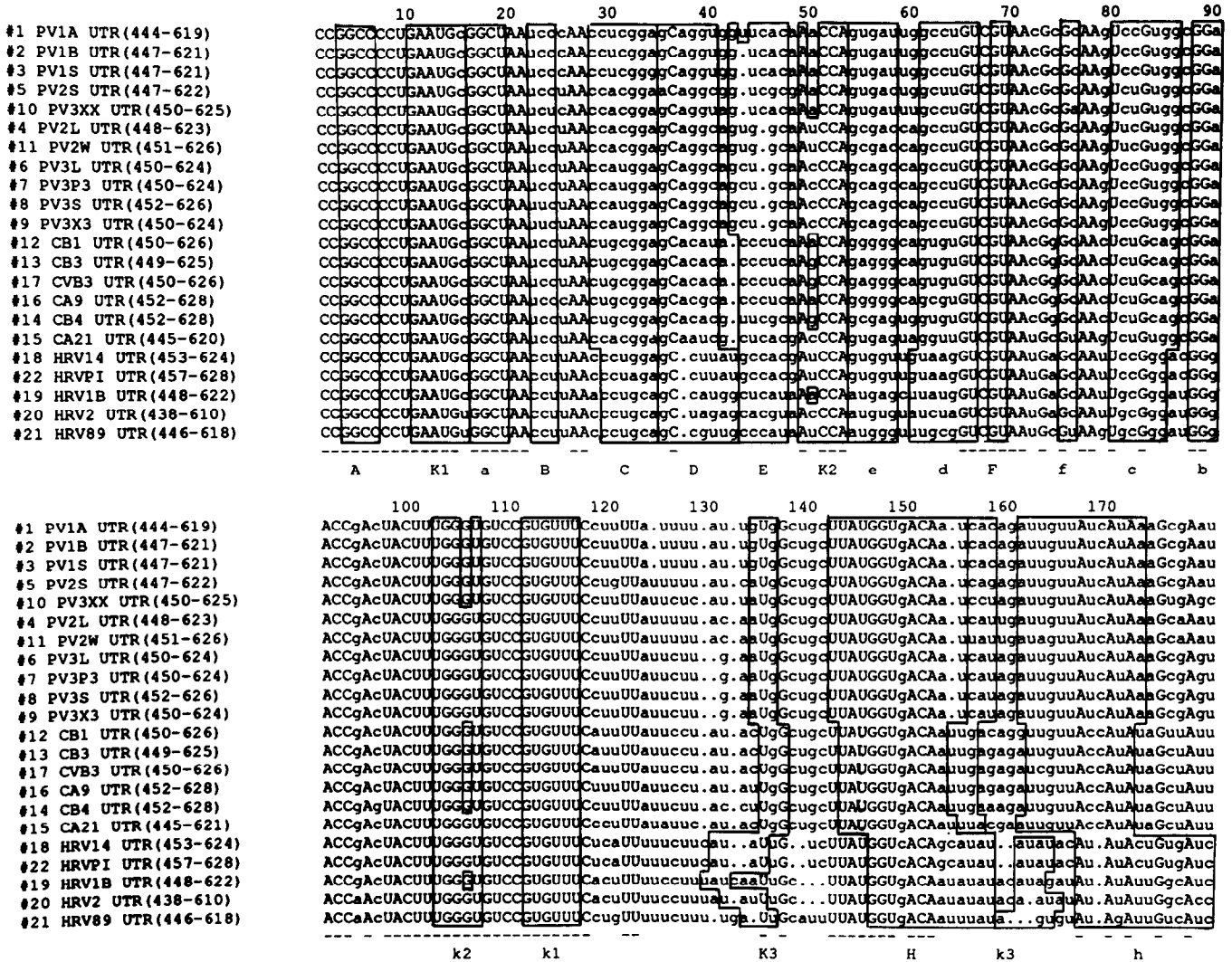
FIG. 1. Alignment of the sequences of 5' UTRs from 22 picornaviruses. Deletions are denoted by dots. The conserved base-pairing regions are marked by boxes and labeled by letter A–H, K1–K3, a–h, and k1–k3. Among them, the boxes of K1–K3 and k1–k3 correspond to three tertiary structure elements. The sequences used in the alignment are PV1A (poliovirus type 1, Mahoney), PV1B (poliovirus type 1, Mahoney, Baltimore), PV1S (poliovirus type 1, Sabin 1), PV2L (poliovirus type 2, Lansing), PV2S (poliovirus type 2, Sabin 2), PV2W (poliovirus type 2, strain W-2), PV3L (poliovirus type 3, P3/Leon/37), PV3P3 (poliovirus type 3, P3/119), PV3S (poliovirus type 3, Sabin 3, Vaccine), PV3X3 (poliovirus type 3, P3/Leon 12 a1b), PV3XX (poliovirus type 3, 23127), CB1 (coxsackievirus B1), CB3 (coxsackievirus B3), CB4 (coxsackievirus B4), CA21 (coxsackievirus A21), CA9 (coxsackievirus A9), CVB3 (coxsackievirus B3), HRV14 (human rhinovirus type 14, Stanway), HRV1B (human rhinovirus type 1B), HRV2 (human rhinovirus type 2), HRV89 (human rhinovirus type 89), and HRVPI (human rhinovirus type 14).

indicate that the translational initiation of picornaviruses involves a mechanism of internal ribosome entry. Nicholson et al. (1991) recently reported that the 5' border of the RLP core element of poliovirus type 2 was located between nt 134 and 155. Deletions of 240–302, 350–380, and 450–480 in PV2L cause the complete abolition of internal initiation of translation. It demonstrates that a highly conserved and extensive stem–loop structure (237–441 in PV2L) is essential for internal initiation of translation for polioviruses. Small mutations within a 9-nt polypyrimidine sequence 5'-UUUCCUUUU-3' (560 to 568) of PV2L showed that the

4-nt pyrimidine stretch sequence UUUC (560 to 563) was also required for the internal initiation (Nicholson et al., 1991). Systematic mutation of this stretch of sequence revealed that the sequence UUUCCUU constituted an important cis-acting RNA element for translation of poliovirus in vitro (Pestova et al., 1991). Recently, Nomoto and co-workers (Iizuka et al., 1991) also demonstrated that the conserved 21-nt sequence (546–566) was essential for in vitro translation of coxsackievirus B1 (CB1). The experimental data strongly suggest that the 3'-end of RLP constitutes a highly ordered tertiary structure, which can be specifically

LE ET AL.

TABLE 1

STATISTICS OF PREDICTED TERTIARY INTERACTIONS IN THE 5' UTR OF HUMAN ENTEROVIRUSES AND RHINOVIRUSES

| Virus | Conserved tertiary structure elements | | | | | | | | |
| | K1 | | | K2 | | | K3 | | |
	n_1	$n_2{}^a$	z	n_1	$n_2{}^a$	z	n_1	$n_2{}^a$	z
PV1A	40	83	4.66	80	141	4.96	35	54	1.34
PV1B	40	83	4.66	80	141	4.96	35	54	1.34
PV1S	40	83	4.66	80	141	4.96	35	54	1.34
PV2L	31	84	8.36	21	64	5.78	49	83	1.56
PV2S	47	120	4.27	77	139	4.27			
PV2W	31	84	8.09	44	99	4.45	120	170	1.15
PV3L	56	128	4.55	24	27	3.32	36	89	1.54
PV3P3	42	99	4.55	24	27	3.32	22	85	1.56
PV3S	42	99	4.55	24	27	3.32	22	85	1.56
PV3X3	42	99	4.55	24	27	3.32	22	85	1.56
PV3XX	35	92	4.96	53	102	6.12			
CA21	37	73	9.90	9	12	3.99	182	251	2.72
CA9	36	93	7.38	56	95	5.78	180	392	3.54
CB1	36	99	9.48	37	74	6.66	241	471	2.80
CB3	68	171	5.89	92	250	0.95	180	497	4.36
CB4	63	141	5.68	133	358	0.51	161	425	1.36
CBV3	68	171	5.89	92	250	0.95	294	643	2.66
HRV14	25	64	6.66	25	60	6.00	60	235	2.16
HRV1B	25	76	7.18	49	88	6.00	464	702	1.82
HRV2	57	136	7.60	8	9	6.37	255	396	0.24
							(60	73	1.34)[b]
HRV89	39	94	2.59	32	68	4.10	218	574	1.04
HRVPI	25	55	6.66	25	60	5.25	38	191	1.85

[a] n_2 was computed by using parameter $c = 2.0$ (Chen et al., 1992).
[b] If the stem 2 of the pseudoknot K3 in HRV2 is UUG and AAC instead of UAUU and AUAA.

recognized by 40 S ribosomes and/or trans-acting factor(s). However, the details of the tertiary structure of internal ribosomal entry site have not been determined yet.

In this paper we propose a common RNA tertiary structural model for the 3' part of the 5'-UTR of 22 picornaviruses. The conserved tertiary structure is situated just 3' to the functional stem–loop structure (237 to 441 in PV2L) for internal initiation of translation. The conserved 17-nt (445 to 461 in PV2L) and 21-nt (543 to 563 in PV2L) sequences are involved in the superstructure. Base pairings between the two highly conserved regions contribute two significant tertiary structure elements K1 and K2 in the predicted structure model. The point mutations (Nicholson et al., 1991) of 3 nt UUU out of an essential 4-nt pyrimidine stretch sequence UUUC would destroy the predicted tertiary interactions in the pseudoknot K1. The proposed common tertiary structure can offer experimentalists a rationale upon which to extend interpretations of current data (Nicholson et al., 1991; Pestova et al., 1991). It is conceivable that the pseudoknot is specifically recog-

nized by components of the translation apparatus as in the case of the translational control of the ribosomal S4 (Tang and Draper, 1989) and ribosomal protein S15 (Portier et al., 1990) of the E. coli. Similarly, the conserved tertiary structure element K2 is suggested as a potential binding site for the interaction between by 40 S ribosomes and/or trans-acting factor(s) and viral mRNAs.

Base complementarity between the pyrimidine stretch of the picornavirus mRNA and 18 S rRNA may participate in the interaction between viral mRNA and the small ribosomal subunit (Kuhn et al., 1990; Nicholson et al., 1991). The pyrimidine stretch UUUCC has been suggested to be an analogue of the prokaryotic Shine-Dalgarno sequence by Agol and co-workers (Pilipenko et al., 1992). Based on the common RNA secondary structures (Le and Zuker, 1990) of the 5' UTR in 18 human enteroviruses and rhinoviruses, two sequences complementary to the highly conserved polypyrimidine sequence in all picornaviruses were identified in human 18 S rRNA (Nicholson et al., 1991). They are located at nt 1194 to 1207 (3'-GGCAGUUAAGG-

ciency of poliovirus (Pelletier *et al.*, 1988a,b; Pilipenko *et al.*, 1992). Structural analyses for these variants described by Pilipenko *et al.* (1992) and Pestova *et al.* (1991) reveal that the base-pairing interaction between viral mRNA and 18 S rRNA would be destabilized in the mutated transcripts. The template activity of the mutated transcripts showed markedly lower than that of the wild type (Pilipenko *et al.*, 1992; Pestova *et al.*, 1991).

The pseudoknot K3, which contains an AUG-7 and AUG-6 (in PV2L) is conserved in all 22 tested picornavirus sequences. This tertiary interaction in coxsackieviruses is extremely weak (at least two G–U base pairings in the structure) and is not supported by the statistical analysis (large n_1 and n_2 values) in our procedure. However, it can be supported by compensatory base
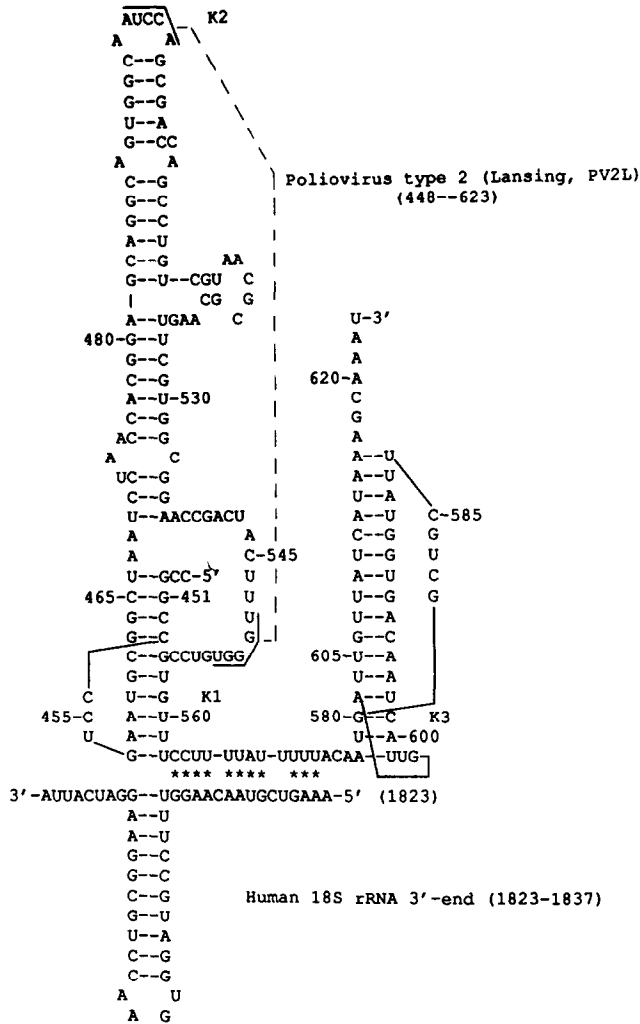
FIG. 2. Predicted base-pairing model between poliovirus type 2 (Lansing, referred to as PV2L) and human 18 S rRNA. The conserved tertiary structure folded in the RLP of PV2L is highly stable and statistically significant. The three conserved tertiary structure elements in 22 picornaviruses are denoted by K1, K2, and K3. The RNA secondary structure of human 18 S rRNA 3'-end is based on the published model (Huysmans and Wachter, 1986; Chan *et al.*, 1984). An asterisk between two sequences denotes the base pairing between PV2L and human 18 S rRNA.

AAA-5') and nt 1822 to 1837 (3'-GGAACAAUGCUGA-AAA-5'). The two regions are not base-paired in the common RNA secondary structural model of 18 S rRNAs and are highly conserved throughout evolution. According to the new common tertiary structure model, the sequence 3'-GGAACAAUGCUGAAAA-5' spanning nt 1822 to 1837 in human 18 S rRNA is a better complementary sequence to the polypyrimidine stretch of 22 picornaviruses. For polioviruses, at least 11 of 14 nt can be base pairing with the complementary sequence. It is known that the sequence upstream from the AUG-7 (in PV2L) affects translational effi-
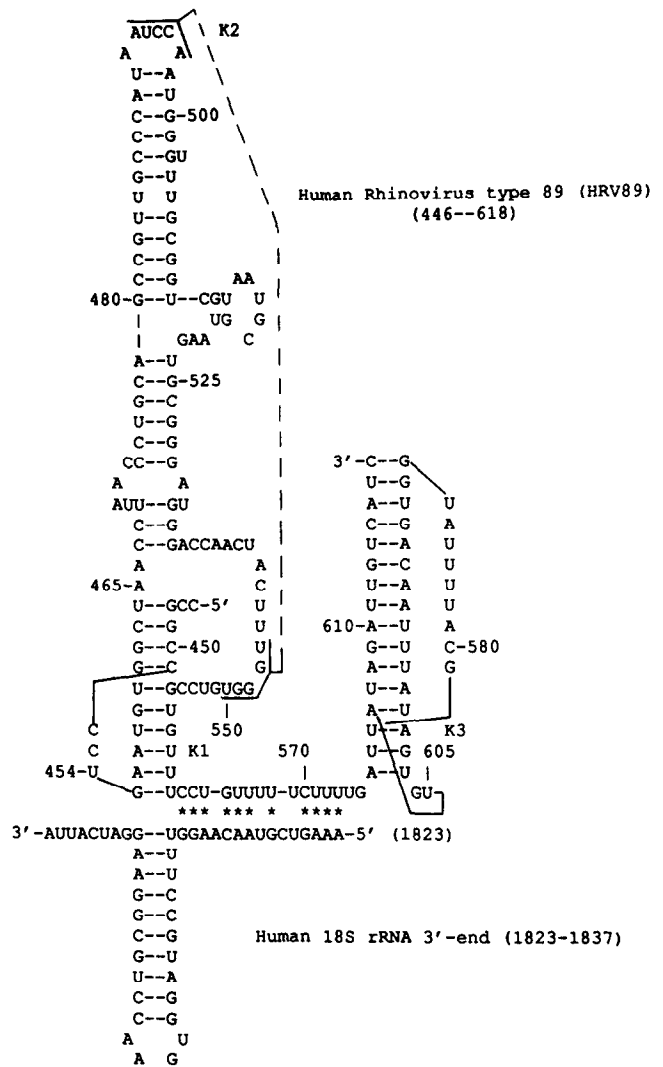
FIG. 3. Predicted base-pairing model between human rhinovirus type 89 (HRV89) and human 18 S rRNA. For further details see the caption to Fig. 2.

```
        AGCC\\ K2
       A    A\  \
         C--G       \
         G--C        \
         C--G         \
         U--A          \
         U--GU          \
       G   G            \
         C--G            |  Coxsackievirus B4 (CB4)
 490-A--U                |     (452--628)
         C--G    520     |
         A--U     |      |
         C--G    AA      |
       G--U--CGU  C      |
       |     CG   G      |
       A--UCAA    G      |  U-3'
         G--C            |  U
         G--U            |  A
         C--G            |  U
         G--C            |  C
         U--A-535        |  G
       AC--G             |  A
      A    C             |  U
       UC--G             |  620-A--U
         C--G            |  U--A
         U--AACCGAGU     |  A--U   U-590
         A        A      |  C--G   C
         A        C  |   |  C--G   G
 470-U--GCC-5'   U   |   |  A--U   U
       C--G-455  U   |   |  U--G   C
       G--C      U   |   |  U--A
       G--C      G   |   |  G--C
       C--GCCUGUGG/      |  U--A
       G--U     |        |  610-U--A
     C U--G    555       |  G--U
     C A--U K1           |  580 G--U  K3
 460-U A--U              |     | U--G
       G--UCCUU-UUAUU-CUUACCU-\AAAGA-
          ****  ****  ***
 3'-AUUACUAGG--UGGAACAAUGCUGAAA-5' (1823)
          A--U
          A--U
          G--C
          G--C
          C--G        Human 18S rRNA 3'-end (1823-1837)
          G--U
          U--A
          C--G
          C--G
        A     U
        A  G
```
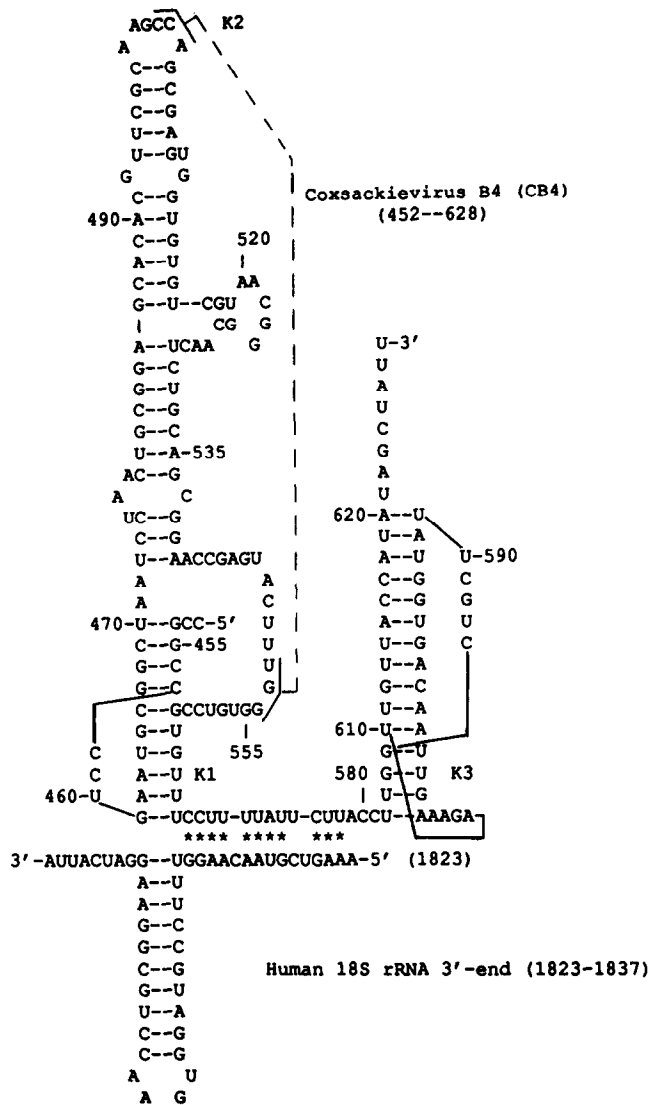
FIG. 4. Predicted base-pairing model between coxsackievirus B4 (CB4) and human 18 S rRNA. For further details see the caption to Fig. 2.

changing found in these predicted structural models. It is worth noting that the single-stranded sequence spanning between pseudoknots K1 and K3 is partially complementary to the 3'-end of human 18 S rRNA. For human rhinovirus RNA, Borman and Jackson (1992) recently demonstrated that the ribosome entry site included the single-stranded sequence.

Several lines of evidence indicated that pseudoknot motifs might be involved in protein binding to a mRNA (McPheeter et al., 1988; Tang and Draper, 1989; Portier et al., 1990). Recently, Brown et al. (1991) proposed a putative pseudoknot formed in the 3'-end of the 5' UTR of hepatitis A virus. Although our theoretical model awaits verification by experiments, our model does provide a reasonable explanation for current mu-

tagenesis results in the translational initiation of human enteroviruses (Nicholson et al., 1991; Pilipenko et al., 1992; Pestova et al., 1991). Brierley et al. (1991) recently demonstrated that the RNA pseudoknot in coronavirus mRNA could not be functionally replaced by a hairpin structure that was composed of an equivalent set of base pairs. Also, the removal of a single base-pair contact in either stem of the pseudoknot reduced or abolished the ribosomal frameshifting. Thus, it is conceivable that two helical stems have to be in proximity to generate a specific functional structure, perhaps stacked coaxially. It is such a structural feature that the RNA pseudoknot can provide as a recognition site for protein binding.

Sonenberg and co-workers (Meerovitch et al., 1989) reported that a novel protein (p52) could specifically bind to nt 559 to 624 in 5' UTR. Wimmer and co-workers (Pestova et al., 1991) also indicated that a ribosome-bound protein p57 was an important cofactor for the translation of poliovirus RNA in vitro and it could bind upstream of the pyrimidine stretch. Thus, we suggest that the common superstructure formed in the sequence (nt 450 to 616 in PV2L) is essential for the translational initiation of human enteroviruses and rhinoviruses. We conjecture that the 40 S ribosome subunit can recognize the pseudoknot K1, and probably K2, in translation initiation of poliovirus mRNAs. The

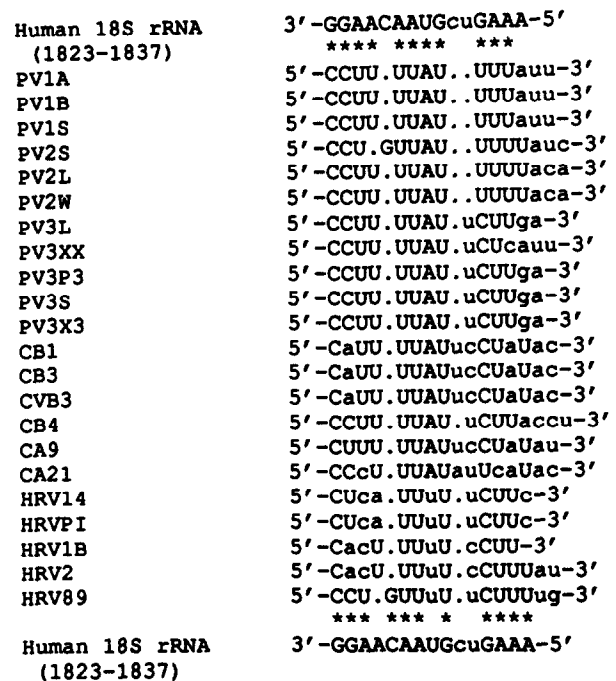|	
Human 18S rRNA (1823-1837)	3'-GGAACAAUGcuGAAA-5'
	**** **** ***
PV1A	5'-CCUU.UUAU..UUUauu-3'
PV1B	5'-CCUU.UUAU..UUUauu-3'
PV1S	5'-CCUU.UUAU..UUUauu-3'
PV2S	5'-CCU.GUUAU..UUUUauc-3'
PV2L	5'-CCUU.UUAU..UUUUaca-3'
PV2W	5'-CCUU.UUAU..UUUUaca-3'
PV3L	5'-CCUU.UUAU.uCUUga-3'
PV3XX	5'-CCUU.UUAU.uCUcauu-3'
PV3P3	5'-CCUU.UUAU.uCUUga-3'
PV3S	5'-CCUU.UUAU.uCUUga-3'
PV3X3	5'-CCUU.UUAU.uCUUga-3'
CB1	5'-CaUU.UUAUucCUaUac-3'
CB3	5'-CaUU.UUAUucCUaUac-3'
CVB3	5'-CaUU.UUAUucCUaUac-3'
CB4	5'-CCUU.UUAU.uCUUaccu-3'
CA9	5'-CUUU.UUAUucCUaUau-3'
CA21	5'-CCcU.UUAUauUcaUac-3'
HRV14	5'-CUca.UUuU.uCUUc-3'
HRVPI	5'-CUca.UUuU.uCUUc-3'
HRV1B	5'-CacU.UUuU.cCUU-3'
HRV2	5'-CacU.UUuU.cCUUUau-3'
HRV89	5'-CCU.GUUuU.uCUUUug-3'
	*** *** * ****
Human 18S rRNA (1823-1837)	3'-GGAACAAUGcuGAAA-5'

FIG. 5. Base pairings between human 18 S rRNA 3'-end and polypyrimidine stretches in the 5' UTR of human enteroviruses and rhinoviruses. An asterisk between two sequences denotes the base pairing between these virus mRNAs and human 18 S rRNA.

binding interaction of pseudoknot K1 (and/or K2) and 40 S ribosome subunit and/or trans-acting factors can be strengthened by the strong base pairing between the sequences in the polypyrimidine stretch and 3'-end of 18 S rRNA. These plausible structural predictions add to the known intriguing properties of the long untranslated 5' regions of picornavirus genome/messages.

ACKNOWLEDGMENTS

REFERENCES

BELSHAM, G. J., and BRANGWYN, J. K. (1990). A region of the 5' noncoding region of foot-and-mouth disease virus directs efficient internal initiation of protein synthesis within cells: Involvement with the role of L protease in translational control. *J. Virol.* **64**, 5389–5395.

BIENKOWSKA-SZEWCZYK, K., and EHRENFELD, E. (1988). A internal 5'-noncoding region required for translation of poliovirus RNA *in vitro*. *J. Virol.* **62**, 3068–3072.

BORMAN, A., and JACKSON, R. J. (1992). Initiation of translation of human rhinovirus RNA: Mapping the internal ribosome entry site. *Virology* **188**, 685–696.

BREDENBEEK, P. J., PACHUK, C. J., NOTEN, A. F. H., CHARITE, J., LUYTJES, W., WEISS, S. R., and SPAAN, W. J. M. (1990). The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-59: A highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res.* **18**, 1825–1832.

BRIERLEY, I., DIGARD, P., and INGLIS, S. C. (1989). Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an RNA pseudoknot. *Cell* **57**, 537–547.

BRIERLEY, I., ROLLEY, N. J., JENNER, A. J., and INGLIS, S. C. (1991). Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* **220**, 889–902.

BROWN, E. A., DAY, S. P., JANSEN, R. W., and LEMON, S. M. (1991). The 5' non-translated region of hepatitis A virus RNA: Secondary structure and elements required for translation *in vitro*. *J. Virol.* **65**, 5828–5838.

CHAN, Y.-L., GUTELL, R., NOLLER, H. F., and WOOL, I. G. (1984). The nucleotide sequence of a rat 18 S ribosomal ribonucleic acid gene and a proposal for the secondary structure of 18 S ribosomal ribonucleic acid. *J. Biol. Chem.* **259**, 224–230.

CHANG, K., AUVINEN, P., HYYPIA, T., and STANWAY, G. (1989). The nucleotide sequence of coxsackievirus A9: Implications for receptor binding and enterovirus classification. *J. Gen. Virol.* **70**, 3269–3280.

CHEN, J.-H., LE, S.-Y., and MAIZEL, J. V., JR. (1992). A procedure for RNA pseudoknot prediction. *Comp. Appl. Biosci.* **8**, 243–248.

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.

HUGHES, P. J., NORTH, C., MINOR, P. D., and STANWAY, G. (1989). The complete nucleotide sequence of coxsackievirus A21. *J. Gen. Virol.* **70**, 2943–2952.

HUYSMANS, E., and WACHTER, R. (1986). Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res.* **14**, r73–r118.

IIZUKA, M., YONEKAWA, H., and NOMOTO, A. (1991). Nucleotide sequences important for translation initiation of enterovirus RNA. *J. Virol.* **65**, 4867–4873.

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.

JANG, S. K., KRAUSSLICH, H.-G., NICKLIN, M. J. H., DUKE, G. M., PALMENBERG, A. C., and WIMMER, E. (1988). A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* **62**, 2636–2643.

JANG, S. K., DAVIES, M. V., KAUFMAN, R. J., and WIMMER, E. (1989). Initiation of protein synthesis by internal entry of ribosomes into the 5' nontranslated region of encephalomyocarditis virus RNA *in vivo*. *J. Virol.* **63**, 1651–1660.

KLUMP, W. M., BERGMANN, I., MUELLER, B. C., AMEIS, D., and KANDOLF, R. (1990). Complete nucleotide sequence of infectious coxsackievirus B3 cDNA: Two initial 5' uridine residues are regained during plus-strand RNA synthesis. *J. Virol.* **64**, 1573–1583.

KUGE, S., and NOMOTO, A. (1987). Construction of viable deletion and insertion mutants of the sabin strain of type 1 poliovirus: Function of the 5' noncoding sequence in viral replication. *J. Virol.* **61**, 1478–1487.

KUHN, R., LUZ, N., and BECK, E. (1990). Functional analysis of the internal translation initiation site of foot-and-mouth disease virus. *J. Virol.* **64**, 4625–4631.

LE, S.-Y., and MAIZEL, J. V., JR. (1989). A method for assessing the statistical significance of RNA folding. *J. Theoret. Biol.* **138**, 495–510.

LE, S.-Y., CHEN, J.-H., and MAIZEL, J. V., JR. (1990). Efficient searches for unusual folding regions in RNA sequences. *In* "Structure & Methods: Human Genome Initiative and DNA Recombination" (Sarma, R. H., and Sarma, M. H., Eds.), Vol. 1, pp. 127–136. Adenine Press.

LE, S.-Y., and ZUKER, M. (1990). Common structures of the 5' noncoding RNA in enteroviruses and rhinoviruses: Thermodynamical stability and statistical significance. *J. Mol. Biol.* **216**, 729–741.

McPHEETERS, D. S., STORMO, G. D., and GOLD, L. (1988). Autogenous regulatory site on the bacteriophage T4 gene 32 messenger RNA. *J. Mol. Biol.* **201**, 517–535.

MEEROVITCH, K., PELLETIER, J., and SONENBERG, N. (1989). A cellular protein that binds to the 5'-noncoding region of poliovirus RNA: Implications of internal translation initiation. *Genes Dev.* **3**, 1026–1034.

MEEROVITCH, K., NICHOLSON, R., and SONENBERG, N. (1991). *In vitro* mutational analysis of cis-acting RNA translational elements within the poliovirus type 2 5' untranslated region. *J. Virol.* **65**, 5895–5901.

NICHOLSON, R., PELLETIER, J., LE, S.-Y., and SONENBERG, N. (1991). Structural and functional analysis of the ribosome landing pad of poliovirus: *In vivo* translation studies. *J. Virol.* **65**, 5886–5894.

NOMOTO, A., LEE, Y. F., and WIMMER, E. (1976). The 5' end of poliovirus mRNA is not capped with $m^7G(5')ppp(5')Np$. *Proc. Natl. Acad. Sci. USA* **73**, 375–380.

PELLETIER, J., and SONENBERG, N. (1988). Internal initiation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320–325.

PELLETIER, J., KAPLAN, G., RACANIELLO, V. R., and SONENBERG, N. (1988a). Cap-independent translation of poliovirus mRNA is conferred by sequence elements within the 5' noncoding region. *Mol. Cell. Biol.* **8**, 1103–1112.

PELLETIER, J., FLYNN, M. E., KAPLAN, G., RACACIELLO, V. R., and SONENBERG, N. (1988b). Mutational analysis of upstream AUG codons of poliovirus RNA. *J. Virol.* **62**, 4486–4492.

PESTOVA, T. V., HELLEN, C. U. T., and WIMMER, E. (1991). Translation of poliovirus RNA: Role of an essential cis-acting oligopyrimidine element of a cellular 57-kilodalton protein. *J. Virol.* **65**, 6194–6204.

PEVEAR, D. C., OH, C. K., CUNNINGHAM, L. L., CALENOFF, M., and JUBELT, B. (1990). Localization of genomic regions specific for the attenuated, mouse adapted poliovirus type 2 strain W-2. *J. Gen. Virol.* **71**, 43–52.

PILIPENKO, E. V., BLINOV, M., ROMANOVA, L. I., SINYAKOV, 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 controlling translation and neurovirulence. *Virology* **168**, 201–209.

PILIPENKO, E. V., GMYL, A. P., MASLOVA, S. V., SVITKIN, Y. V., SINYAKOV, A. N., and AGOL, V. I. (1992). Prokaryotic-like cis elements in the cap-independent internal initiation of translation on picornavirus RNA. *Cell* **68**, 119–131.

PLEIJ, C. W. A., RIETVELD, K., and BOSCH, L. (1985). A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.* **13**, 1717–1731.

PORTIER, C., PHILIPPE, C., DONDON, L., GRUNBERG-MANAGO, M., EBEL, J. P., EHRESMANN, B., and EHRESMANN, C. (1990). Translational control of ribosomal protein S15. *Biochim. Biophys. Acta* **1050**, 328–336.

PUGLISI, J. D., WYATT, J. R., and TINOCO, I., JR. (1988). A pseudoknotted RNA oligonucleotide. *Nature* **331**, 283–286.

RAIRKAR, A., RUBINO, H. M., and LOCKARD, R. E. (1988). Chemical probing of adenine residues within the secondary structure of rabbit 18 S ribosomal RNA. *Biochemistry* **27**, 582–592.

RIVERA, V. M., WELSH, J. D., and MAIZEL, J. V., JR. (1988). Comparative sequence analysis of the 5' noncoding region of enteroviruses and rhinoviruses. *Virology* **165**, 42–50.

SKINNER, M. A., RACANIELLO, V. R., DUNN, G., COOPER, J., MINOR, P. D., and ALMOND, J. W. (1989). New model for the secondary structure of the 5' non-coding RNA of poliovirus is supported by biochemical and genetic data that also show that RNA secondary structure is important in neurovirulence. *J. Mol. Biol.* **207**, 379–392.

TANG, C. K., and DRAPER, D. E. (1989). Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. *Cell* **57**, 531–536.

TRONO, D., PELLETIER, J., SONENBERG, N., and BALTIMORE, D. (1988a). Translation in mammalian cells of a gene linked to the poliovirus 5' noncoding region. *Science* **241**, 445–448.

TRONO, D., ANDINO, R., and BALTIMORE, D. (1988b). An RNA sequence of hundreds of nucleotides at the 5' end of poliovirus RNA is involved in allowing viral protein synthesis. *J. Virol.* **62**, 2291–2299.