362–369 *Nucleic Acids Research, 1997, Vol. 25, No. 2*

A common RNA structural motif involved in the internal initiation of translation of cellular mRNAs

Shu-Yun Le* and Jacob V. Maizel Jr

Laboratory of Mathematical Biology, Division of Basic Sciences, National Cancer Institute, NIH, Building 469, Room 151, Frederick, MD 21702, USA

Received September 11, 1996; Revised and Accepted November 11, 1996

ABSTRACT

The 5′-non-translated regions (5′NTR) of human immunoglobulin heavy chain binding protein (BiP), Antennapedia (Antp) of *Drosophila* and human fibroblast growth factor 2 (FGF-2) mRNAs are reported to mediate translation initiation by an internal ribosome binding mechanism. In this study, we investigate predicted features of the higher order structures folded in these 5′NTR sequences. Statistical analyses of RNA folding detected a 92 nt unusual folding region (UFR) from 129 to 220, close to the initiator AUG in the BiP mRNA. Details of the structural analyses show that the UFR forms a Y-type stem–loop structure with an additional stem–loop in the 3′-end resembling the common structure core found in the internal ribosome entry site (IRES) elements of picornavirus. The Y-type structural motif is also conserved among a number of divergent BiP mRNAs. We also find two RNA elements in the 5′-leader sequence of human FGF-2. The first RNA element (96 nt) is 2 nt upstream of the first CUG start codon located in the reported IRES element of human FGF-2. The second (107 nt) is immediately upstream of the authentic initiator AUG of the main open reading frame. Intriguingly, the folded RNA structural motif in the two RNA elements is conserved in other members of FGF family and shares the same structural features as that found in the 5′NTR of divergent BiP mRNAs. We suggest that the common RNA structural motif conserved in the diverse BiP and FGF-2 mRNAs has a general function in the internal ribosome binding mechanism of cellular mRNAs.

INTRODUCTION

Human fibroblast growth factor 2 (FGF-2) belongs to a family of potent mitogens that are implicated in many aspects of cell growth and differentiation, both in normal and neoplastic settings. The intracellular distribution of FGF-2 plays an important role in cell behavior. Four forms of human FGF-2, a small protein of 18 kDa and three larger proteins of 21, 21.5 and 22.5 kDa, are synthesized from the same mRNA (1). These different proteins result from alternative initiation of translation by using a canonical initiator AUG and three in-frame, upstream, non-canonical CUG initiators (2).

Most mRNAs encoding oncoproteins and cell factors related to cell proliferation possess a long, GC-rich and structured 5′-leader sequence with one or more AUG triplets. The leader sequence of human FGF-2 contains a 301 nt long 5′-non-translated region (5′NTR) with two AUG triplets that is located upstream of the first start codon (CUG) and a 165 nt alternatively translated region (ATR) from the first start codon CUG to the major translational initiator AUG of the FGF-2 mRNA. The base composition of the ATR sequence is up to 90% GC. Kozak (3) suggested that the expression of this type of gene, with a GC-rich and highly structured leader sequence, is translationally regulated. Recently, Vagner *et al*. (4) demonstrated that the alternative translation of human FGF-2 mRNA was mediated by an internal ribosome binding mechanism. Furthermore, a 165 nt RNA element just upstream of the first start codon CUG is proposed as an internal ribosome entry site (IRES). These results indicate that the regulation of translational initiation plays a crucial role in the control of cell proliferation.

The internal ribosome binding mechanism, independent of the 5′-cap structure and of the cap binding protein complex eIF-4E, was first observed in poliovirus (5) and subsequently in cardiovirus (6), aphthovirus (7–9), rhinovirus (10) and hepatitis A virus (11,12). The phenomenon is of general importance as it is not restricted to picornaviruses. It is also found in other types of viruses, including hepatitis C virus (HCV) (13,14) and bovine viral diarrhea virus (15), as well as some cellular mRNAs, such as human immunoglobulin heavy chain binding protein (BiP) mRNA (16), Antennapedia (Antp) mRNA of *Drosophila* (17) and FGF-2 mRNA (4). Capped eukaryotic mRNAs can be translated without significant amounts of the intact protein complex eIF-4F. Initiation and translation by an internal ribosome binding mechanism independent of an intact eIF-4F holoenzyme complex may be advantageous for those mRNAs whose translation is important at mitosis in the cell cycle, because of the presence of underphosphorylated eIF-4F.

Internal ribosome binding is dependent on interactions between a *cis*-acting IRES and a *trans*-acting, cell type-specific factor, such as p57/PTB (polypyrimidine tract binding protein) required in the internal initiation of enteroviruses (18). The RNA–protein interaction involves the specific recognition of sequences and/or structural elements of the IRES element within the 5′NTR by the cellular factors. Thus, it is important to determine the common structural motif folded within these IRES elements in BiP, Antp and FGF-2 mRNAs. Recently, we proposed a common RNA structural motif (19) that is conserved in all IRES elements

* To whom correspondence should be addressed. Tel: +1 301 846 5576; Fax: +1 301 846 5598; Email: shuyun@fcrfv1.ncifcrf.gov

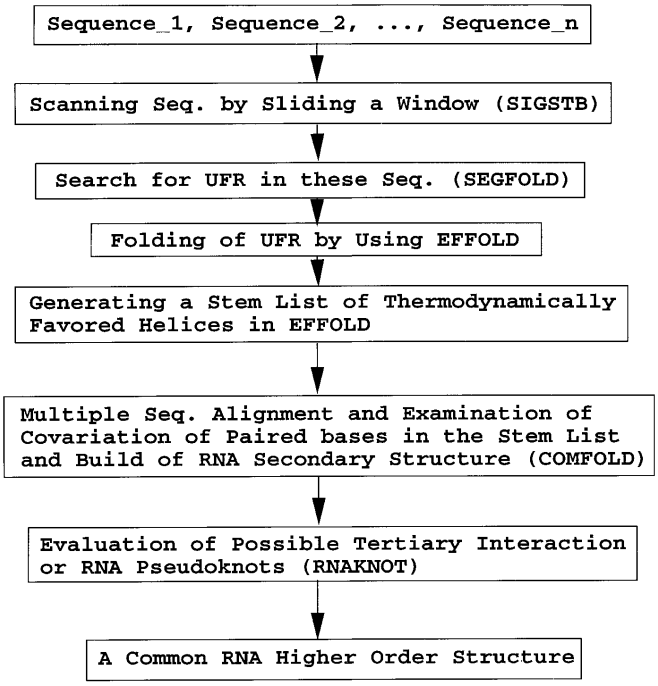

Figure 1. Flowchart of the procedure for RNA folding and structural analysis. For details of the programs SIGSTB, SEGFOLD, EFFOLD, COMFOLD and RNAKNOT see previous publications.

examined to date involved in the internal ribosome binding mechanism, from picornavirus to pestivirus and HCV. The common structural motif shares a structural feature similar to that observed in the catalytic core of group I introns.

Sequence comparison of the 221 nt 5′NTR of BiP mRNA with the IRES of picornaviruses did not reveal any notable homology. Up to now, we do not know whether there is any significant relationship between the viral and cellular IRES. Does the functional IRES in the 5′NTR of cellular mRNAs comprise an undiscovered, conserved structural motif that could be correlated with the mediation of internal ribosome entry in BiP and FGF-2 mRNAs? We address this question in this study.

MATERIALS AND METHODS

In this paper, RNA sequences and RNA folding were analyzed by the procedure developed in our laboratory as shown in Figure 1. For implementation, we set up three groups of mRNA sequences. Group 1 included the *Saccharomyces cerevisiae* karyogamy gene (20; accession no. M25064) and six BiP mRNA sequences from human (21; M19645), rat (22; M14866), Chinese hamster (23; M17169), *Giardia lamblia* (24; U04875), *Trypanosoma brucei* (25; L14477) and *Caenorhabditis elegans* (26; M26604). Group 2 contained FGF-2 mRNAs from human (1; J04513), bovine (27; M13440) and rat (28; M22427) and other FGF members including human FGF-5 (29; M37825) and FGF-9 (30; D14838), human Kaposi's sarcoma oncogene FGF (31; M17446), human keratinocyte growth factor (32; M60828), mouse FGF-6 (33; X51552) and FGF-8 (34; U18673), mouse hst/KFGF (35; X14849) and rat FGF-9 (30; D14839). Group 3 had three sequences of Antp mRNAs (36) from *Drosophila melanogaster* (M20704), *D.virilis* (M95825) and *D.subobscura* (X60995). Sequence comparison was performed using the NUCALN program (37). Each group of sequences was aligned by Zuker's MAL program (38). The alignment data were then edited, combined and refined in terms of the common structural information.

Among these sequences, we searched for UFRs in the human BiP mRNA sequence using the programs SIGSTB and SEGFOLD (39). We folded the detected UFRs using the EFFOLD program (40). These predicted, thermodynamically favored, helical stems in human BiP mRNA were computed in the RNA folding simulation by fluctuating free energy parameters in the Turner energy rules (41,42) within the range of experimental error. Based on these thermodynamically favored stems, we constructed a stem list using COMFOLD (43) in which each stem was supported by phylogenetic comparative analysis (44,45) among the BiP mRNA sequences from human, rat and Chinese hamster. As a result, we built a theoretical common RNA structural motif for the three sequences. Based on this common structural motif, we determined the common RNA folding in the other four related BiP mRNAs. In practice, we first chose the conserved stems from the thermodynamically favored stem list, then added stems from the stem pool of all possible base pairing regions. Some manual inspection and selection for equivalent base pairing are required.

Using the detected IRES element that forms the common structure within the BiP mRNAs, we searched for a similar structure in the 5′NTR sequence of human FGF-2 mRNA by means of the same procedure. Combined with RNA folding simulations and statistical tests, we also determined the structural motif of IRES elements in FGF-2. Similarly, we computed the common structural motif in the 5′NTR of other members of the FGF family and the Antp mRNAs based on the common structural information found in the BiP and FGF-2 sequences.

RESULTS

The UFR and common RNA structural motif of BiP IRES elements

A UFR from nt 129 to 220 is detected in the segment including the 5′NTR and 50 nt downstream of the AUG start codon. This is one of the most stable regions in the 5′ portion of BiP mRNA. Although the segment 129–220 has the lowest stability scores computed by both the Tinoco (59) and Turner (41) energy rules, its scores related to the random shuffling sequences are not statistically significant (Fig. 2). This means that thermodynamic stability alone may not be a good predictor for representing the functional form of this RNA segment.

The proposed RNA structural models (Fig. 3) of the UFR in the three BiP mRNAs from human, rat and Chinese hamster were computed by a combination of thermodynamic, phylogenetic and statistical methods. Among these structures, the junction domain of three stems, A–C, occur with high frequency in the RNA folding simulation, where 50 simulated energy rules were generated. For human BiP mRNA, the frequencies of stems B, C and A (138–142/183–187) are 76, 58 and 58% for folding the complete 5′UTR and 92% for folding the UFR respectively. Although only 43% sequence similarity between human and rat and 52% between human and Chinese hamster are detected among these RNA fragments, the folded structures shown in Figure 3 are quite conserved. Iizuka *et al.* (46) indicated that the RNA fragment spanning nt 128–220 could direct internal initiation when placed in the intercistronic spacer region of a

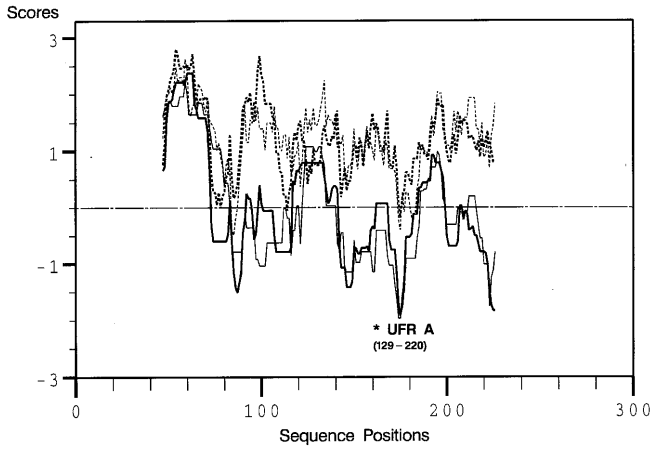

Figure 2. Distributions of the significance score (Sigscr, broken curve) and stability score (Stbscr, continuous curve) in the human BiP mRNA sequence. The two scores were computed by both Turner (thick line) and Tinoco (thin line) energy rules (59). In the plot, the window size was set at 92 nt. The searched sequence consists of the 5′NTR segment and the 50 nt downstream sequence of the start codon AUG. The detected UFR is located from nt 129 to 220 (marked by an asterisk) and is consistent with the BiP IRES element of 128–220 (46). The profile was obtained by plotting the Sigscr and Stbscr of each segment window against the position of the middle nucleotide in the segment by sliding the window throughout the sequence. The two scores are defined as follows: Sigscr = $(E - E_r)/SD_r$ and Stbscr = $(E - E_w)/SD_w$, where E is the lowest free energy of formation for RNA folding of a specific RNA fragment, E_r and SD_r are the mean and standard deviation of the lowest free energies from a great number of random shuffling RNA foldings for the fragment and E_w and SD_w are the mean and standard deviation of the lowest free energies resulting from sliding a window of the same size as the fragment throughout the sequence from the 5′- to 3′-end. In practice, E_r and SD_r were computed by a set of coefficients that were derived from a least squares fit to the 500 random shuffling sequences (60).

dicistronic mRNA. The common RNA structure folded in the region 129–220 of human BiP mRNA is a candidate structural motif of the RNA functional element.

The proposed Y-type structure of the BiP IRES elements of human, rat and Chinese hamster is also observed in the 5′NTR of other divergent BiP mRNAs, including *G.lamblia* (GL), *T.brucei* (TB), *C.elegans* (CE) and the *S.cerevisiae* karyogamy gene (KAR2, a homolog of mammalian BiP). The sequence similarity among these BiP RNA structural models is only ~41% or less (Table 1). The RNA structures folded in these protozoan and metazoan BiP mRNAs are not as stable as those folded in human, rat and Chinese hamster BiP sequences (Fig. 3), because they are not GC-rich sequences.

By comparing the common structural motif (19) found in the IRES elements of picornavirus (47), HCV and pestivirus with the proposed common structural model of the human BiP IRES, we find similarities to the IRES elements of groups B and C of picornavirus. This common feature is a Y-type stem–loop structure immediately followed by a stem–loop structure just upstream of the authentic initiator AUG. Schematic diagrams of the common structural motif are shown in Figure 4.

The common RNA structural motif of FGF-2 IRES elements

The sequence comparison between the UFR sequence of human BiP and the 5′NTR sequence of human FGF-2 mRNA did not indicate any notable sequence similarity. The closest RNA sequence related to the human BiP IRES element was found in an ~100 nt RNA fragment (204–299) located just upstream of the alternative initiator CUG (302–304). The RNA fragment is located within the suggested IRES element (154–318) of human FGF-2 mRNA (4). RNA folding analysis indicated that the higher order structure formed bears a striking resemblance to the conserved structural motif observed among the IRES elements of BiP mRNAs. The RNA segment (204–299) within the proposed FGF-2 IRES element is a good candidate for a role in the IRES-dependent translation of human FGF-2. Similarly, we also detected a corresponding IRES element (331–426) in the rat FGF-2 5′NTR. The IRES of rat FGF-2 is just 4 nt upstream of the corresponding codon CUG, which is similar to that observed in the human FGF-2 IRES. The conserved RNA structural models of the two IRES elements of human and rat FGF-2 are shown in Figure 5. The RNA functional elements in human and rat FGF-2 are referred to as IRES core E1.

Table 1a. Sequence similarity among the common structural motifs of the 5′NTR of Bip mRNAs

	Hamster	Rat	Yeast	GL	TB	CE
Human	0.52	0.43	0.30	0.28	0.27	0.22
Hamster		0.83	0.33	0.32	0.16	0.31
Rat			0.35	0.38	0.32	0.33
Yeast				0.38	0.41	0.32
GL					0.20	0.38
TB						0.29

Table 1b. Sequence similarity among the common structural motifs of the 5′NTR of Bip and FGF-2 mRNAs

	Human FGF-2		Rat FGF-2		Bovine	Human Bip	Rat Bip
	E1	E2	E1	E2	FGF-2 E2		
Human E1	1.00	0.47	0.39	0.29	0.49	0.35	0.29
FGF-2 E2	0.47	1.00	0.39	0.70	0.88	0.15	0.32

RNA folding simulations indicated that the segment 360–467 is one of the most stable folding regions in the first 600 nt domain of the 5′-portion of human FGF-2 mRNA (5′NTR plus the following 134 nt in the coding region). This UFR is immediately upstream of the authentic initiator AUG of the major translation product of human FGF-2 mRNA. This RNA fragment has 47% sequence similarity to the IRES core E1 (204–299) of human FGF-2 (Table 1). Similarly, we also detected an analogous segment in the 5′-leader sequence of rat FGF-2 mRNA (438–532) and bovine FGF-2 mRNA (16–104). Intriguingly, the three RNA fragments share a striking resemblance to the structural feature observed in the IRES elements of BiP and FGF-2 mRNAs (Fig. 6 and Table 2). We suggest that the RNA fragment is another IRES element related to internal translational initiation of the major protein of FGF-2 (termed E2). These IRES elements, E1 and E2, are very close to the authentic initiation codons of FGF-2 mRNAs. This is the same feature that we observed in BiP mRNAs (Fig. 6 and Table 2).

Based on the conserved structural feature observed among these cellular IRES elements, we performed RNA structural analyses of other mRNA sequences of the FGF family and their related sequences. The common structural motif predicted in BiP and FGF-2 is also formed in human FGF-5 and FGF-9, human Kaposi's sarcoma oncogene FGF (KSFGF), human keratinocyte growth factor (KGF), mouse FGF-6 and FGF-8, mouse hst/KFGF and rat FGF-9 (Fig. 6 and Table 2).

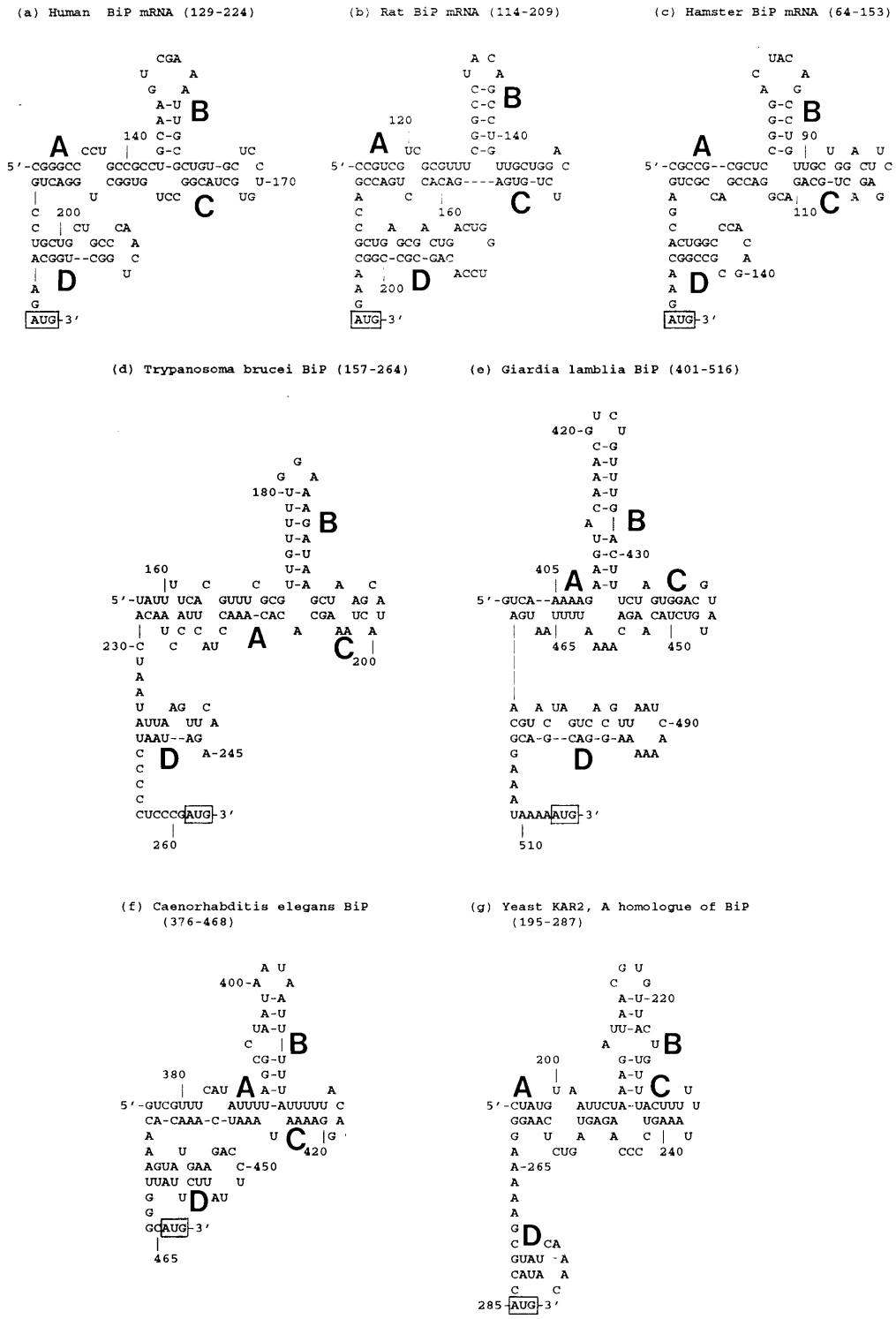

Figure 3. Conventional representations for the common structural motif predicted in the IRES element detected in the 5′NTR of human BiP mRNA (**a**). The common structures predicted in the 5′NTR upstream of the start codon AUG of other BiP mRNAs, rat, Chinese hamster, *Trypanosoma brucei*, *Giardia lamblia*, *Caenorhabditis elegans* and yeast KAR2 are shown in (**b**–**g** respectively). The start codon AUG of the open reading frame is boxed. In the model, the Y-type stem–loop is labeled as stems A–C. The following stem–loop is labeled by the letter D. Stem A is often interrupted by an internal loop and stems B–D may include a small bulge or internal loop.

Structural features of IRES elements in BiP and FGF-2

The sequence alignment of BiP and FGF-2 IRES elements and the 5′UTR of other BiP and FGF family members is shown in Figure 6. The alignment was gradually refined by means of their common structural information. In the common Y-type structure, stem A is usually interrupted by an internal loop and its size ranges from 8 to 13 bp. There is an intervening unpaired base

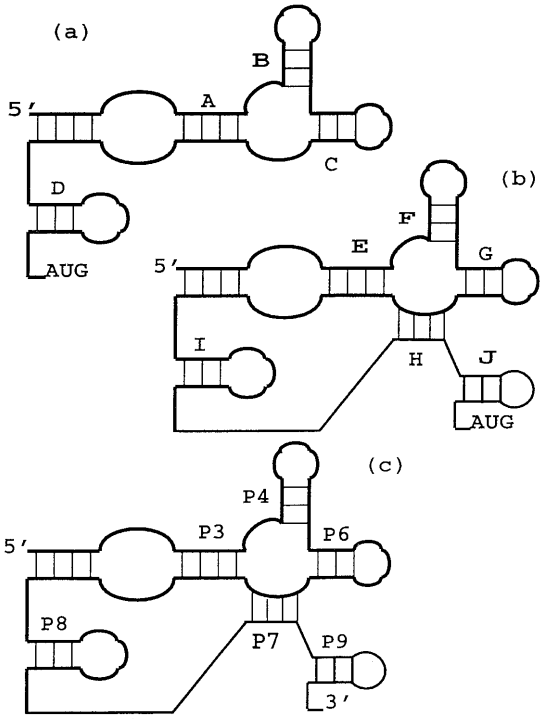

Figure 4. Schematic diagram of the common structural core in (**a**) IRES elements of cellular mRNAs, (**b**) IRES elements of picornavirus, HCV and pestivirus and (**c**) group I introns (48). Stems A, E and P3 are represented as including an internal loop. The possible unpaired nucleotides are not shown in stems B–D in (a), stems F–J in (b) and stems P4 and P6–P9 (c). In the structural core of the cellular IRES (a), stems A–D correspond to stems E–I/J in (b) and to stems P3, P4, P6 and P8 in (c) respectively. The coaxial stem stacking between stems B and C suggested in (a) is consistent with the stem stacking between stems F and G in (b) and between stems P4 and P6 in (c). In the structural core of the viral IRES (b), stems E–J correspond to stems P3, P4 and P6–P9 of group I introns respectively. Among these, stems E and H in (b) and stems P3 and P7 in (c) form a similar RNA pseudoknot. In (b) stem J is deleted in group B of picornavirus, HCV and pestivirus; stem I is deleted in group A of picornavirus, HCV and pestivirus.

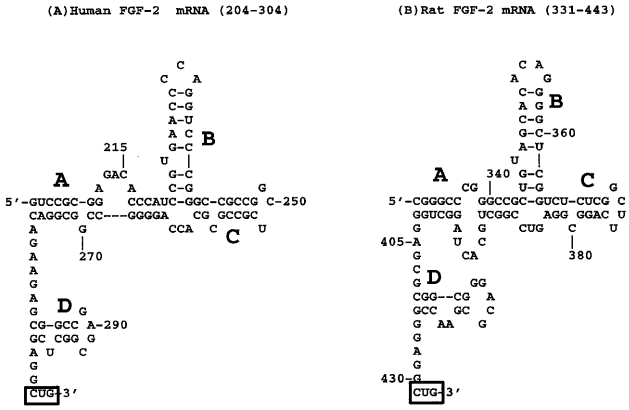

Figure 5. Conventional representations for the common structural motif predicted in the IRES elements detected in the 5′NTR sequence upstream of the start codon CUG of human (**a**) and rat (**b**) FGF-2 mRNAs. The IRES element shown in human FGF-2 was supported by experimental data (4). The start codon CUG is marked by a box. For further details see the caption to Figure 3.

Table 2. Conservation and co-variation of the common structural motif

Species	A:	B:	C:	D:
H. Bip	CGGGCC..GCCGC. GUCAGG..CGGUG.	UGCAA GCGUU	CUGU. GGCA.	.UGCUG. .ACGGU.
R. Bip	CCGUCG..GCGUU. GCCAGU..CACAG.	CGGGC GUCCG	UUGC. AGUG.	.GCUG. .CGGC.
Hams. Bip	CGCCG--CGCUC .GUCGC..GCCAG.	CGGG GUCC	UUGC. GACG.	.ACUG. .CGGC.
Y. KAR2	CUAUG..AUUCU .GGAAC..UGAGA.	AAAG. UUUU.	ACUUU UGAAA	.GUAU. .CAUA.
CE Bip	..GUUU..AUUU.. ..CAAA..UAAA..	UAGG. AUUU.	UUUUU AAAAG	.AGUA.. .UUAU..
GL Bip	UCA--AAAA.. ...AGU..UUUU..	AAGU. UUCA.	UCUaGU. AGAaCA.	.CGUaC. .GCA-G.
TB Bip	UUuUCA..UUcGCG AA.AUU..AA-CAC	UUGUUU. AAUAGA.	GCUaAG CGA.UC	.AUUA. .UAAU.
H. FGF-5	UAUGUC..GGC-GA ACGUAG..CCGaCU	GGCGG. ACGCC.	CAGGG. GACCC.	.UUAG. .GGUC.
R. FGF-2 E1	CGGGCC..GGCCG. GGUCGG..UCGGC.	CUGuAG. GGC-UC.	UCU-CU. AGGcGG.	.CGG.. .GCC..
H. FGF-2 E1	CGC-GG..CCCAU GCGgCC..GGGGA.	CCGuGA. GGC-CC.	GC-CG. CGcGC.	CG-GCC GCuCGG
H. FGF-2 E2	GGGGCC--GUGC CCUCGG--CGCG.	CCCGG. GGGCC.	GCCG.. CGGC..	.CGGGG. .GCCCC.
B. FGF-2 E2	GGAG----CGCG CCUC--..GCGC.	UCGG GGCC	GGCCG UCGGC	.CGGGG. .GUUCC.
R. FGF-2 E2	CCCGC..GGCGA GGGCG..CCGUC.	GCCGCG. CGGAGC.	CGGGG. GCCCC.	GUCCC CGGGG
H. FGF-9	CAGUAA..GAGGG GUUAUU..CUACC.	GAGUUG CUCCAU	GCCUA UGGGU	.CUUG. .GAGC.
R. FGF-9	CAGCAA..GAGGG GUUAUU..CUACC.	GAGUUG CUCCAU	GCCUAG UUGGUC	.CUUG. .GAGC.
M. FGF-6	UUUUAG--GGCCA AGGGUC--CCGGU.	UUA AGU	CCACGU. GGUGGA.	.GAG .CUU
M. FGF-8	GCAC..UUCGG. .CGUG..GGGCU.	UUG. GAC.	UGGG. GCCC.	.GGC .CCG
H. KGF	.AAUGA..CCUAG .UUAUU..GUAUU.	GAGU CUCA	AAGA UUUU	.GGAG .UCAC
H. KSFGF	GGUCG..GCGCUC. .CCAGU..CGCGGG.	GCG CGC	AGCU UCGA	.CCUG. .GGGC.
M. KFGF	UUUG..GCUAC. AAGC..CGACG.	UAGGU. ACUCA.	CCG GGC	CUCACG GGGCGC
D. mel	GUUU..AUCA. CAAA..UAGG.	AAUUG. UCAAC.	AGAGG UCAUC	
D. vir	GUU-UC..AAAUC. CAAcGG..UUUAG.	AAUUG. UCAAC.	AGAGGU UCAUCG	
D. sub	GUUU..AUCA. CAAA..UAGG.	AAUUG. UUAAC.	AGAGG UCAUC	

The stems A–D are four structural units found in the common structural motif of cellular IRES elements of human BiP and FGF-2 mRNAs, as shown in Figures 4 and 6. A dash between two nucleotides indicates that no nucleotide exists between them and a dotted sign represents the nucleotides that are omitted in the table for simplicity. The IRES elements of the homeotic gene Antp mRNAs of *Drosophila melanogaster*, *D.virilis* and *D.subobscura* are denoted by D.mel, D.vir and D.sub respectively (see also Fig. 7).

between stems A and B in some cases. Stem B and stem C contain 4–8 bp and a hairpin loop of 3–8 unpaired bases. In some cases, stems B and C are broken by a small bulge or internal loop. Stem C is frequently connected to stem A by a loop of 3–5 nt. The Y-type structure is followed by an additional stem–loop D, in which there are differences in sequence length. One remarkable property is that stem–loop D is just a few nucleotides upstream from the authentic initiator.

The proposed common structural motif of these IRES elements is confirmed by the intra- and/or interphylum covariance seen in Figure 6 and Table 2 that maintains the base pairing potential in stems A–D. The structural conservation of these divergent

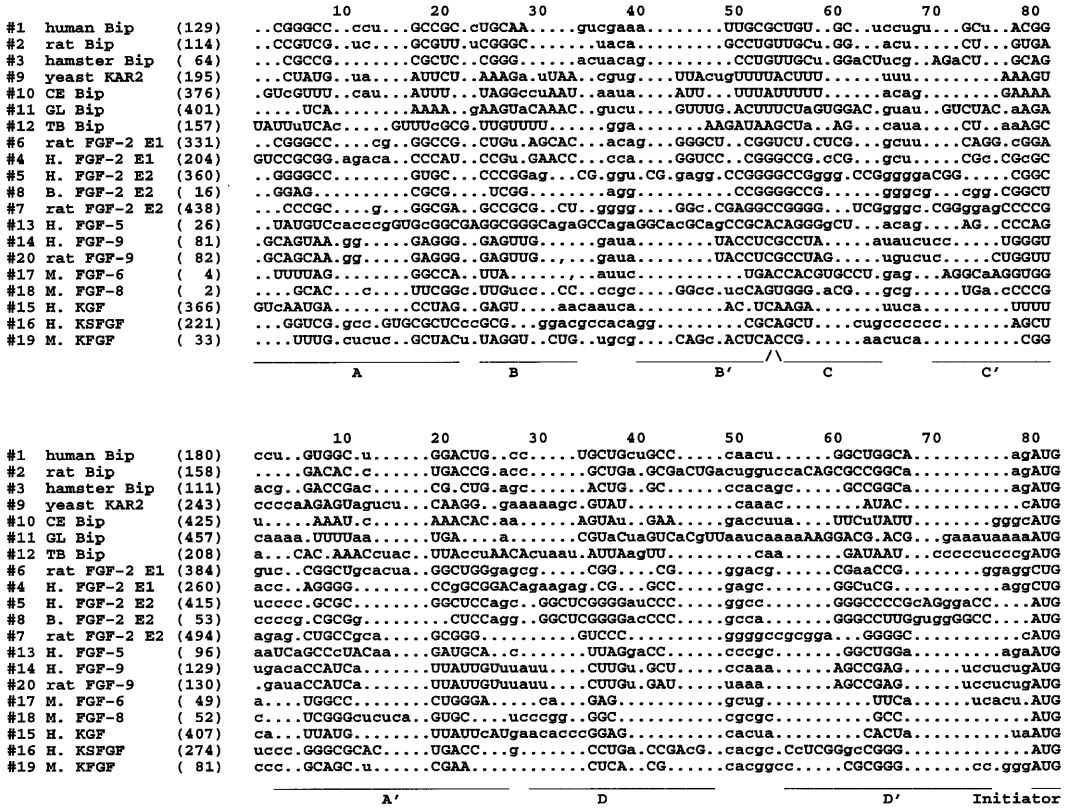

Figure 6. The structural conservation found among the IRES elements of human BiP and FGF-2 and among the 5′UTR of the BiP and FGF families. In the alignment, deletions are denoted by dots. The upper case letter indicates the nucleotide folded in the conserved base pairing regions. The base pairing region is also underlined and labeled by the letters A–D and A′–D′. In the plot, the FGF-2 IRES elements just prior to the first start codon CUG and major initiator AUG are referred to as E1 and E2 respectively.

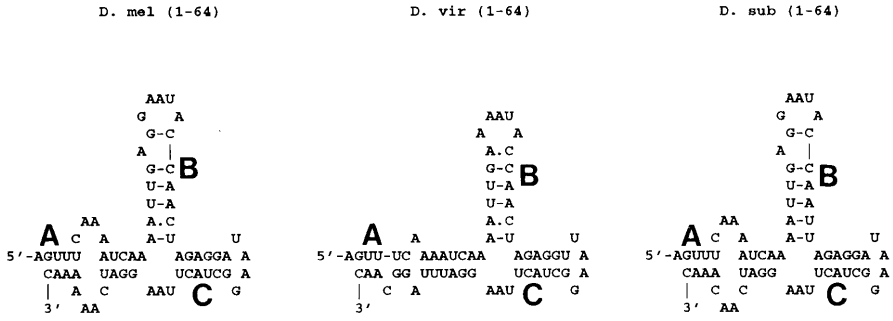

Figure 7. Conventional representations for the common structural motif predicted in the IRES elements detected in the 5′NTR of the homeotic gene Antp mRNAs of *D.melanogaster*, *D.virilis* and *D.subobscura*. In the figure, exon D starts with the nucleotide numbered 1 and the Y-type stem–loop is labeled as stems A–C.

sequences indicates that the predicted common RNA structural motif involved in internal initiation of the translation of cellular BiP and FGF-2 mRNAs could play an important role in internal ribosome binding control of cellular mRNAs.

DISCUSSION

A common RNA structural motif, including a Y-type stem–loop structure followed by a small stem–loop structure just upstream of the translational initiator, is a conserved property found in IRES elements identified in the cellular BiP and FGF-2 mRNAs. Although the Antp gene of *D.melanogaster* contains an exceptionally long 5′NTR (>1500 nt), the 5′-border of exon D functions as the IRES element (17). In the 5′-border of exon D there is a 55 nt RNA sequence highly conserved among *D.melanogaster*, *D.virilis* and *D.subobscura*. Preliminary experiments from Sarnow's laboratory (17,46) indicated that the highly conserved 55 nt sequence (nt 1–55 in Fig. 7) was required for Antp IRES function. The conserved 55 nt sequence is one of the smallest IRES elements reported, which is located over 350 nt upstream of the AUG translational start codon. A Y-type stem–loop structure similar to that observed in other IRES elements of cellular mRNAs can be formed in the RNA fragment that includes the conserved 55 nt and the following 8 nt. However, these folded RNA structures are

not highly stable. One possibility is that IRES-dependent translation of Antp requires additional protein factors to stabilize the folded RNA common structure. Ribosome movement of the 350 nt from the IRES to the initiating AUG may occur by conventional scanning.

Among the common RNA structural motifs of the IRES elements of BiP, FGF-2 and Antp mRNAs, the most striking structural property may be represented by the conserved, conceivably coaxial stacking between stem B and stem C within the common Y-type stem–loop structure. Sequence comparison (Table 1) indicated that while these suggested RNA functional elements diverge in sequence similarity, the folded structures have conserved folding shapes and positions, just a few nucleotides (except for Antp mRNAs) upstream of the authentic initiation codon. The conservation of these structural motifs among these divergent BiP sequences, from protozoa to human, strongly supports their importance. Moreover, the predicted structural motifs of BiP IRES elements are also conserved in FGF-2 IRES elements, although their sequence similarities are ≤35%. The evolutionary stability of this structural motif is consistent with a crucial role in the IRES-dependent translation of cellular mRNAs.

Previously, we proposed a structural core of secondary and tertiary structures (consisting of stems E–H) that is common to all the IRES elements of picornaviruses, HCV and pestiviruses (19). This conserved structural motif (Fig. 4), found among a large number of dissimilar sequences, is likewise a candidate structural core essential to the function of these viral IRES elements. Interestingly, the proposed conserved superstructures for these viral IRES elements share a striking structural resemblance to the higher ordered structure of group I introns. The common structural core (Fig. 4) of these viral IRES, composed of stems E–H, shares tertiary structural features analogous to the core structure centered on helical regions P3, P4, P6 and P7 in group I introns (48). The stacking between stems E and H corresponds to the stacking of helices P3 and P7 and the possible coaxial stacking of stems F and G corresponds to the stacking between P4 and P6. By means of such coaxial stem stacking and other tertiary interactions, the group I introns form a compact three-dimensional structure.

A comparison of the two structural motifs of the viral and cellular IRES elements indicated that they share some resemblance in the folding shape, stem stacking and sequence location of the 5′NTR region (Fig. 4). For the two common structural motifs, stems A–D in cellular IRES correspond to stems E, F, G and I/J of viral IRES elements respectively. Possible coaxial stacking between stems B and C in the cellular IRES corresponds to the stacking of viral stems F and G. The coaxial stacking between two adjacent stems may be one of the most important properties observed in cellular and viral IRES elements. Stem stacking is an important tertiary structural interaction, as shown in group I introns (49,50), rRNAs (51,52) and tRNAs (53). The newly developed energy parameters for coaxial stacking between two adjacent stems (54) indicate that such a distinct structural element is quite favorable to the stabilization of RNA folding. Similar to group I introns, the common core structures of the viral and cellular IRES can form a compact tertiary structure by means of stem stacking. However, the new energy parameters contributed by the stem stacking have not been included in our current Monte Carlo simulation procedure.

Comparatively little is known about IRES elements in cellular mRNAs relative to the viral IRES. Limited information on the boundaries of cellular IRES elements has been determined by deletion analysis. However, the exact positions of these IRES elements have not been determined by extensive experimental testing. We suggest that the function of both viral and cellular IRES elements is correlated with the distinct, conserved RNA higher order structures folded in these elements.

The difference between the two structural motifs of the viral and cellular IRES is also clear. The conserved RNA pseudoknot composed of stems E and H in all viral IRES elements was not observed in the cellular IRES. The RNA pseudoknot has been demonstrated to be essential for HCV IRES function (55). If the pseudoknot is an important structural property of the viral IRES, then there is a difference in the IRES-dependent mechanism that occurs in cellular and viral internal initiation. Currently, two cellular *trans*-acting factors, the La antigen (56) and PTB (57,58), have been found to bind to picornavirus IRES elements and to be essential for their internal initiation of translation. However, PTB binds extremely poorly to the BiP IRES and two other proteins, ~60 and 95 kDa in size, can bind specifically to the BiP IRES (46). It is possible that different *trans*-acting factors are required which are dependent on the specific IRES.

Few IRES elements have been found in cellular mRNAs. It is not clear what advantage the IRES-dependent mechanism offers relative to the conventional ribosome scanning mechanism in translational initiation of cellular mRNAs. Although the proposed common structural motif of the cellular IRES requires confirmation by experimental data, our model is supported by phylogenetic analysis of divergent sequences from BiP, FGF-2 and the FGF family. The structural conservation is consistent with the fact that these IRES elements are functionally related to each other. The common structural motif of cellular IRES elements may provide useful information on the relationship between viral and cellular IRES elements. It is also helpful for searching more cellular mRNAs with an IRES. Knowledge of the distribution of these structural motifs should further the discovery of the underlying principles of cap-independent translation and internal initiation of translation in eukaryotic cellular mRNAs.

ACKNOWLEDGEMENTS

The contents of this publication do 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 US Government.

REFERENCES

1 Prats,H., Kaghad,M., Prats,A.C., Klagsbrun,M., Lelias,J.M., Liauzun,P., Chalon,P., Tauber,J.P., Amalric,F., Smith,J.A. and Caput,D. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1836–1840.
2 Prats,A.C., Vagner,S., Prats,H. and Amalric,F. (1992) *Mol. Cell. Biol.*, **12**, 4796–4805.
3 Kozak,M. (1991) *J. Cell Biol.*, **115**, 887–903.
4 Vagner,S., Gensac,M.C., Maret,A., Bayard,F., Amalric,F., Prats,H. and Prats,A.C. (1995) *Mol. Cell. Biol.*, **15**, 35–44.
5 Pelletier,J. and Sonenberg,N. (1988) *Nature*, **334**, 320–325.
6 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.
7 Kuhn,R., Luz,N. and Beck,E. (1990) *J. Virol.*, **64**, 4625–4631.
8 Belsham,G.J. (1992) *EMBO J.*, **11**, 1106–1110.
9 Bandyopadhyay,P.K., Wang,C. and Lipton,H.L. (1992) *J. Virol.*, **66**, 6249–6256.
10 Borman,A. and Jackson,R.J. (1992) *Virology*, **188**, 685–696.
11 Brown,E.A., Day,S.P., Jansen,R.W. and Lemon,S.M. (1991) *J. Virol.*, **65**, 5828–5838.
12 Glass,M.J. and Summers,D.F. (1993) *Virology*, **193**, 1047–1050.

13 Tsukiyama-Kohara,K., Iizuka,N., Kohara,M. and Nomoto,A. (1992) *J. Virol.*, **66**, 1476–1483.
14 Wang,C., Sarnow,P. and Siddiqui,A. (1994) *J. Virol.*, **68**, 7301–7307.
15 Poole,T.L., Wang,C., Popp,R.A., Potgieter,L.N.D., Siddiqui,A. and Collett,M.S. (1995) *Virology*, **206**, 750–754.
16 Macejak,D.G. and Sarnow,P. (1991) *Nature*, **353**, 90–94.
17 Oh,S.K., Scott,M.P. and Sarnow,P. (1992) *Genes Dev.*, **6**, 1643–1653.
18 Borman,A., Howell,M.T., Patton,J.G. and Jackson,R.J. (1993) *J. Gen. Virol.*, **74**, 1775–1788.
19 Le,S.-Y., Siddiqui,A. and Maizel,J.V.,Jr (1996) *Virus Gene*, **12**, 135–147.
20 Rose,M.D., Misra,L.M. and Vogel,J.P. (1989) *Cell*, **57**, 1211–1221.
21 Ting,J. and Lee,A.S. (1988) *DNA*, **7**, 275–286.
22 Chang,S.C., Wooden,S.K., Nakaki,T., Kim,Y.K., Lin,A.Y., Kung,L., Attenello,J.W. and Lee,A.S. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 680–684.
23 Ting,J., Wooden,S.K., Kriz,R., Kelleher,K., Kaufman,R.J. and Lee,A.S. (1987) *Gene*, **55**, 147–152.
24 Gupta,R.S., Aitken,K., Falah,M. and Singh,B. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 2895–2899.
25 Bangs,J.D., Uyetake,L., Brickman,M.J., Balber,A.E. and Boothroyd,J.C. (1993) *J. Cell Sci.*, **105**, 1101–113.
26 Heschl,M.F.P. and Baillie,D.L. (1989) *DNA*, **8**, 233–243.
27 Abraham,J.A. *et al*. (1986) *Science*, **233**, 545–548.
28 Shimasaki,S., Emoto,N., Koba,A., Mercado,M., Shibata,F., Cooksey,K., Baird,A. and Ling,M. (1988) *Biochem. Biophys. Res. Commun.*, **157**, 256–263.
29 Haub,O., Drucker,B. and Goldfarb,M. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8022–8026.
30 Miyamoto,M., Naruo,K., Seko,C., Matsumoto,S., Kondo,T. and Kurokawa,T. (1993) *Mol. Cell. Biol.*, **13**, 4251–4259.
31 Delli,B.P., Curatola,A.M., Kern,F.G., Greco,A., Ittmann,M. and Basilico,C. (1987) *Cell*, **50**, 729–737.
32 Finch,P.W., Rubin,J.S., Miki,T., Ron,T. and Aaronson,S.A. (1989) *Science*, **245**, 752–755.
33 de Lapeyriere,O. *et al*. (1990) *Oncogene*, **5**, 823–831.
34 MacArthur,C.A., Shankar,D.B. and Shackleford,G.M. (1995) *J. Virol.*, **69**, 2501–2507.
35 Brookes,S., Smith,R., Thurlow,J., Dickson,C. and Peters,G. (1989) *Nucleic Acids Res.*, **17**, 4037–4045.
36 Hooper,J.E., Perez-Alonso,M., Bermingham,J.R., Prout,M., Rocklein,B.A., Wagenbach,M., Edstrom,J.E., de Frutos,R. and Scott,M.P. (1992) *Genetics*, **132**, 453–469.
37 Wilbur,W.J. and Lipman,D.J. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 726–730.
38 Le,S.-Y. and Zuker,M. (1990) *J. Mol. Biol.*, **216**, 729–741.
39 Le,S.-Y., Chen,J.-H. and Maizel,J.V.,Jr (1990) In Sarma,R.H. and Sarma,M.H. (eds), *Structure & Methods: Human Genome Initiative and DNA Recombination.* Adenine Press, Schenectady, NY, Vol. 1, pp. 127–136.
40 Le,S.-Y., Chen,J.-H. and Maizel,J.V.,Jr (1993) *Nucleic Acids Res.*, **21**, 2173–2178.
41 Freier,S.M., Kierzek,R., Jaeger,J., Sugimoto,N., Caruthers,M.H., Neilson,T. and Turner,D.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9373–9377.
42 Jaeger,J.A., Turner,D.H. and Zuker,M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7706–7710.
43 Le,S.-Y., Zhang,K. and Maizel,J.V.,Jr (1995) *Comp. Biomed. Res.*, **28**, 53–66.
44 Gutell,R.R., Weiser,B., Woese,C.R. and Noller,H.F. (1985) *Prog. Nucleic Acid Res. Mol. Biol.*, **23**, 155–195.
45 James,B.D., Olsen,G.J., Liu,J. and Pace,N.R. (1988) *Cell*, **52**, 19–26.
46 Iizuka,N., Chen,C., Yang,Q., Johannes,G. and Sarnow,P. (1995) *Curr. Topics Microbiol. Immunol.*, **203**, 155–177.
47 Pilipenko,E.V., Blinov,M., Chernov,B.K., Dmitrieva,T.M. and Agol,V.I. (1989) *Nucleic Acids Res.*, **17**, 5701–5710.
48 Cech,T.R. (1988) *Gene*, **73**, 259–271.
49 Murphy,F.L., Wang,Y.H., Griffith,J.D. and Cech,T.D. (1994) *Science*, **265**, 1709–1712.
50 Michel,F. and Westhof,E. (1990) *J. Mol. Biol.*, **216**, 585–610.
51 Gutell,R.R., Larsen,N. and Woese,C.R. (1994) *Microbiol. Rev.*, **58**, 10–26.
52 Laing,L.G. and Draper,D.E. (1994) *J. Mol. Biol.*, **237**, 560–576.
53 Kim,S.H. (1979) In Schimmel,P.R. *et al*. (eds), *Transfer RNA: Structure, Properties, and Recognition.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 83–100.
54 Walter,A.E. and Turner,D.H. (1994) *Biochemistry*, **33**, 12715–12719.
55 Wang,C., Le,S.-Y., Ali,N. and Siddiqui,A. (1995) *RNA*, **1**, 526–537.
56 Meerovitch,K., Svitkin,Y.V., Lee,H.S., Lejbkowicz,F., Kenan,D.J., Chan,E.K.L., Agol,V.I., Keene,J.D. and Sonenberg,N. (1993) *J. Virol.*, **67**, 3798–3807.
57 Hellen,C.U.T., Witherell,G.W., Schmid,M., Shin,S.H., Pestova,T.V., Gil,A. and Wimmer,E. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7642–7646.
58 Kaminski,A., Hunt,S.L., Patton,J.G. and Jackson,R.J. (1995) *RNA*, **1**, 924–938.
59 Cech,T.R., Tanner,N.K., Tinoco,I.,Jr, Weir,B.R., Zuker,M. and Perlman,P.S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3903–3907.
60 Chen,J.-H., Le,S.-Y., Shapiro,B., Currey,K.M. and Maizel,J.V.,Jr (1990) *Curr. Adv. Biol. Sci.*, **6**, 7–18.