Polypurine (A)-rich sequences promote cross-kingdom conservation of internal ribosome entry

Yuri L. Dorokhov*, Maxim V. Skulachev†‡, Peter A. Ivanov†‡, Svetlana D. Zvereva†, Lydia G. Tjulkina†, Andres Merits§, Yuri Y. Gleba‡, Thomas Hohn¶, and Joseph G. Atabekov†‖

*A. N. Belozersky Institute of Physico-Chemical Biology, and †Department of Virology, Moscow State University, 119899 Moscow, Russia; §Institute of Biotechnology, Viikki Biocenter, P.O. Box 56 (Viikinkaari 9), Helsinki University, Fin-00014 Helsinki, Finland; ‡Icon Genetics, Biozentrum Halle, Weinbergweg 22, D-06120 Halle (Saale), Germany; and ¶Friedrich Miescher Institute, Maulbeerstrasse 66, 4058 Basel, Switzerland

Communicated by Meinhart H. Zenk, Universität Halle, Halle/Salle, Germany, February 22, 2002 (received for review September 15, 2001)

The internal ribosome entry sites (IRES), $IRES^{CR}_{CP,148}$ and $IRES^{CR}_{MP,75}$, precede the coat protein (CP) and movement protein (MP) genes of crucifer-infecting tobamovirus (crTMV), respectively. In the present work, we analyzed the activity of these elements in transgenic plants and other organisms. Comparison of the relative activities of the crTMV IRES elements and the IRES from an animal virus—encephalomyocarditis virus—in plant, yeast, and HeLa cells identified the 148-nt $IRES^{CR}_{CP,148}$ as the strongest element that also displayed IRES activity across all kingdoms. Deletion analysis suggested that the polypurine (A)-rich sequences (PARSs) contained in $IRES^{CR}_{CP,148}$ are responsible for these features. On the basis of those findings, we designed artificial PARS-containing elements and showed that they, too, promote internal translation from dicistronic transcripts *in vitro*, in tobacco protoplasts and in HeLa cells. The maximum IRES activity was obtained from multiple copies of either $(A)_4G(A)_2(G)_2$ or $G(A)_{2–5}$ as contained in $IRES^{CR}_{CP,148}$. Remarkably, even homopolymeric poly(A) was moderately active, whereas a poly(G) homopolymer was not active. Furthermore, a database search for existing PARS sequences in 5′-untranslated regions (5′UTR) of genes in tobacco genome allowed the easy identification of a number of IRES candidates, in particular in the 5′UTR of the gene encoding *Nicotiana tabacum* heat-shock factor 1 (NtHSF1). Consistent with our prediction, the 5′UTR of NtHSF1 turned out to be an IRES element active *in vitro*, in plant protoplasts and HeLa cells. We predict that PARS elements, when found in other mRNAs, will show a similar activity.

Translation of most eukaryotic mRNAs occurs by traditional cap-dependent ribosome scanning (1–5). However, the initiation of translation of a variety of viral and cellular mRNAs takes place by an alternative mechanism of internal ribosome entry mediated by internal ribosome entry sites (IRESs). IRESs of about 350–450 nt have been identified and most extensively characterized in the 5′-untranslated regions (5′UTRs) of RNA of viruses belonging to the *Picornaviridae* and *Flaviviridae* families (6–8), whereas IRESs of about 200 nt were found on the RNAs of insect RNA viruses (9–10). The IRES elements of different origin differ largely in structural organization, sequence, length and functional requirements. It is generally believed that there are kingdom-specific limitations of viral IRES activity; thus none of the animal virus IRES elements seem to be active in yeast cells (7). Contrary to this concept, Urwin *et al.* (11) reported that the encephalomyocarditis virus (EMCV) IRES ($IRES^{EMCV}$) was also moderately active in plant cells.

IRES elements have also been found in the 5′UTRs of several animal mRNAs. Importantly, IRES-dependent translation has been reported for cellular mRNAs when their cap-dependent translation is impaired (e.g., under conditions of viral infection, heat shock, apoptosis, and at the G_2/M phase of the cell cycle) (12–14).

It is obvious that mRNAs of those plant viruses that are naturally uncapped (e.g., members of the *Potyviridae*, *Comoviridae*, and *Luteoviridae* families) must be translated by a cap-independent process (15–18). Indeed, two distinct regulatory elements revealed within the 5′UTR of tobacco etch potyvirus were capable of mediating internal translation from dicistronic constructs (19).

In accordance with the ribosome-scanning mechanism, only the 5′-proximal gene of tobamovirus genomic RNA can be directly translated by ribosomes, whereas the other genes are expressed from two separate 3′-coterminal subgenomic RNAs (sgRNAs). The dicistronic I_2 sgRNA is translated to produce the movement protein (MP), whereas the 3′-proximal coat protein (CP) gene is silent. The CP gene is expressed from a small monocistronic sgRNA (for review, see ref. 20). Recently, a new tobamovirus [crucifer-infecting tobamovirus (crTMV)] capable of systemically infecting members of the *Brassicaceae* family has been isolated and characterized (21). We reported that the 148-nt region upstream of the CP gene of crTMV RNA contains an IRES ($IRES^{CR}_{CP,148}$), promoting cap-independent and internal translation of the CP gene and different reporter genes from dicistronic constructs (22, 23). Recently, the ability of $IRES^{CR}_{CP,148}$ to promote internal translation was confirmed in a potato virus X vector-based system (24). The capacity of crTMV $IRES^{CR}_{CP,148}$ to mediate internal translation distinguishes this tobamovirus from the well-known type member of the genus, TMV U1; the equivalent 148-nt sequence from TMV U1 RNA ($U1^{CR}_{CP,148}$) was incapable of mediating internal translation (22). Recently, it has been shown that the 228- and 75-nt regions upstream of the MP gene of crTMV RNA, $IRES^{CR}_{MP,228}$ and its 5′-truncated variant $IRES^{CR}_{MP,75}$, are also active (23).

In this study, the activities of $IRES^{CR}_{CP,148}$, $IRES^{CR}_{MP,75}$, and the well-characterized mammalian $IRES^{EMCV}$ were compared in a dicistronic translation assay in plant, yeast, and HeLa cells. It was found that $IRES^{CR}_{CP,148}$ exhibited a high capacity to mediate translation of the 3′-proximal β-glucuronidase (GUS) gene located on a dicistronic transcript in all of the types of cells tested. The sequence elements responsible for this cross-kingdom activity were identified. The results allowed us to artificially design novel IRES elements and to identify plant-derived IRES elements in plant genes that all demonstrate cross-kingdom activity.

Materials and Methods

Plasmid Constructs. Dicistronic plasmids contained crTMV CP or green fluorescent protein (GFP) gene as the first cistron and the GUS gene as the second, separated by various intercistronic sequences (ICS). T7 and 35S promoter-based constructs of CP-ICS-GUS and hairpin (H)-CP-ICS-GUS series were described previously (22, 23). The analogue procedure (precisely described in supporting information on the PNAS web site, www.pnas.org) was used to construct T7-H-GFP-ICS-GUS and 35S-GFP-ICS-GUS plasmids. Artificial ICS were obtained from pairs of complementary oligonucleotides. *Nicotiana tabacum* heat-shock factor 1

BIOCHEMISTRY

Abbreviations: IRES, internal ribosome entry site; crTMV, crucifer-infecting tobamovirus; MP, movement protein; CP, coat protein; EMCV, encephalomyocarditis virus; ICS, intercistronic sequences; RRL, rabbit reticulocyte lysates; WGE, wheat germ extracts; PPT, polypurine tract; PARS, polypurine A-rich sequences; GFP, green fluorescent protein; GUS, β-glucuronidase; H, hairpin; NtHSF, *N. tabacum* heat-shock factor.

‖To whom reprint requests should be addressed. E-mail: atabekov@genebee.msu.su.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

(NtHSF-1) mRNA (European Molecular Biology Laboratory accession no. AB014483) untranslated leader was obtained by PCR from *N. tabacum* total genomic DNA. Corresponding 35S-based cassettes were transferred into pBIN19 vector for plant transformation. To perform experiments in yeast the CP-ICS-GUS, fragments of T7-based plasmids were inserted into pYeDP1/8–2 yeast expression vector.

In vitro transcription and translation were performed according to manufacturer protocols for the RiboMax kit, wheat germ extract, and rabbit reticulocyte lysate (all from Promega). Transcripts were purified by 2M LiCl precipitation. The mRNA concentration in translation reactions was 0.5 μM in all cases.

Generation and Characterization of Transgenic Tobacco Plants. Plasmid constructs were transformed into *Agrobacterium tumefaciens* strain AGL1 (25) by using standard procedures (26). Transgenic R_0 plants were obtained from discs of *N. tabacum* var. Samsun and characterized by Northern and Western analyses, which were performed according to membrane manufacturer protocols (Amersham Pharmacia)—Hybond N+ and polyvinylidene difluoride membrane, correspondingly.

Protoplast Preparation and Transfection. Protoplasts were isolated from *N. tabacum* (cv. W38) leaves and electroporated with pFF19-based dicistronic DNA constructs "CP-ICS-GUS" and "GFP-ICS-GUS" as described earlier (23).

Determination of GUS Activity. GUS activity was determined according to ref. 27 and measured in relative light units. GUS activity was normalized with the protein concentration estimated by using a Bio-Rad protein assay kit. For each experiment, background GUS activity associated with nontransfected protoplasts was subtracted. The mean values (with SE bars) for three to ten independent experiments are given.

Yeast Cell Transformation and Analysis. The yeast strain 2805 was transformed according to ref. 28. Transformants were selected on minus-histidine medium. Bicistronic mRNA transcription was induced by galactose. Total protein was extracted from yeast spheroplasts by three cycles of freezing in liquid nitrogen and rapid warming up to 42°C in the presence of 0.1% sarkosyl/0.1% Triton X-100. The supernatant, clarified by centrifugation, was collected, and the total protein content was determined (29).

Transfection of HeLa Cells by Modified Vaccinia Virus Encoding T7 RNA Polymerase and T7 Promoter-Based GUS-Expressing Plasmids "H-GFP-ICS-GUS." HeLa cell monolayers were grown on 3.5-cm-diameter Petri dishes in Dulbecco's modified MEM supplemented with 10% heat-inactivated FCS and 100 units/ml of streptomycin and penicillin. Virus stocks of modified vaccinia virus Ankara, expressing bacteriophage T7 RNA polymerase, were prepared as described (30). HeLa cell dishes that were 80–90% confluent were infected with virus at 30–40 plaque-forming units/cell. After 45-min absorption, the cells were washed and transfected by using Opti-MEM (Life Technologies, Gaithersburg, MD), plasmid DNA, and Lipofectin (GIBCO/BRL). A transfection mixture of 2 μg of DNA in 5 μl of Lipofectin was used for each 3.5-cm plate; six plates were used in each experiment for each construct. Cells were incubated at 37°C for 6 h. After incubation, the medium was removed, cells were washed twice with PBS and lysed directly on the plate in 250 μl of lysis buffer (100 mM potassium phosphate, pH 7.8/0.2% Triton X-100/0.5 mM DTT) for 10 min. The lysate was collected, clarified by centrifugation at 2,000 $\times$ *g* for 10 min, and stored at −70°C.

Results

crTMV IRES-Mediated Expression of the 3′-Proximal GUS Gene in Transgenic Plants. crTMV RNA contains two IRES elements capable of promoting internal translation of the 3′-proximal genes from dicistronic constructs even when translation of the first gene was blocked by a 5′-terminal H structure (22, 23). These results were obtained in cell-free translation systems [rabbit reticulocyte lysates (RRL) wheat germ extracts (WGE)] and in electroporated protoplasts. To exclude possible discrepancies between the functional activities of crTMV-derived IRESs *in vitro* and *in planta*, we compared the relative efficiencies of different crTMV IRESs in transgenic tobacco plants. To this end, a series of R_0 tobacco plants were generated, transgenic for dicistronic constructs containing a 5′-proximal crTMV CP gene separated from the second gene (GUS) by one of the IRES elements. Fig. 1*A* shows that GUS activity could be readily detected by histochemical methods in plants transgenic for dicistronic $IRES^{CR}$-carrying constructs. The integrity of dicistronic and monocistronic transcripts produced in these transgenic plants was proven by Northern blotting. Average RNA samples extracted from the leaves of plants transformed with monocistronic (II in Fig. 1*B*) or dicistronic (III–V in Fig. 1*B*) transgenes were used for analysis. It can be seen from Fig. 1*B* that dicistronic and monocistronic constructs yielded transcripts of the predicted size; no visible bands corresponding to monocistronic products of dicistronic transcript degradation could be detected. In addition, these results provided evidence that the $IRES^{CR}_{CP,148}$ sequence functions as an IRES and not as a transcriptional promoter. The results of Western blot analysis with antibodies to crTMV CP indicated that the 5′-proximal CP gene was expressed in transgenic plants, although the expression level varied from plant to plant (Fig. 1*C*). In Fig. 1*D*, the relative GUS activity mediated by crTMV IRESs was normalized relative to the amount of CP produced by the 5′-proximal gene of dicistronic CP-IRES-GUS mRNAs in individual transgenic lines (Fig. 1 *C* and *D*). In other words, GUS activity was normalized with respect to the amount of dicistronic transcript produced by individual lines.

The relative efficiency of GUS gene expression by monocistronic transgene and by the 3′-proximal GUS gene of dicistronic transgene was also examined. Comparison of GUS activities in samples taken from plants transgenic for dicistronic (CP-IRES-GUS) and monocistronic (GUS) constructs (Fig. 1, Roman numerals III–V and II, respectively) showed that the average levels of IRES-mediated GUS expression (in relative light units) reached 21% ($IRES^{CR}_{MP,75}$ and $IRES^{CR}_{MP,228}$) and 31% ($IRES^{CR}_{CP,148}$) of monocistronic GUS expression.

Cross-Kingdom Conservation of $IRES^{CR}_{CP,148}$ Activity: Comparative Efficiencies of $IRES^{EMCV}$ and crTMV IRESs in Tobacco Protoplasts, HeLa, and Yeast Cells. Our earlier data showed that the crTMV IRES elements were active both in plant cell-derived (WGE) and animal cell-derived (RRL) cell-free systems (22, 23). In a series of experiments, the relative efficiencies of GUS gene expression mediated by $IRES^{EMCV}$, $IRES^{CR}_{MP,75}$ and $IRES^{CR}_{CP,148}$ were compared. The activity of $IRES^{EMCV}$ was negligible in WGE, whereas in RRL, the level of $IRES^{EMCV}$-directed GUS gene activity was more than two times higher than $IRES^{CR}_{CP,148}$, the most active crTMV IRES element (Fig. 2 *D* and *E*).

We then compared the relative activities of the EMCV and two crTMV IRESs ($IRES^{CR}_{CP,148}$ and $IRES^{CR}_{MP,75}$) in different types of cell cultures, including tobacco protoplasts, HeLa cells, and yeast cells. Appropriate promoter-based plasmids were constructed to transcribe a dicistronic RNA including the IRES sequences in question intercistronically upstream of the GUS gene. In all cases tested, GUS expression was negligible from negative control constructs (Fig. 2 *A–C*). In tobacco protoplasts (Fig. 2*A*), the relative activity of the crTMV RNA-derived IRESs was markedly higher than that of $IRES^{EMCV}$, which is consistent with the negligible activity of $IRES^{EMCV}$ in WGE (Fig. 2*D*). Furthermore, the relative efficiency of $IRES^{EMCV}$ was considerably higher in HeLa cells than in nonanimal (tobacco protoplasts and yeast) cells, and the activity of $IRES^{CR}_{MP,75}$ was high in plant protoplasts but dramatically reduced in the nonplant cells. The most unexpected result was that the

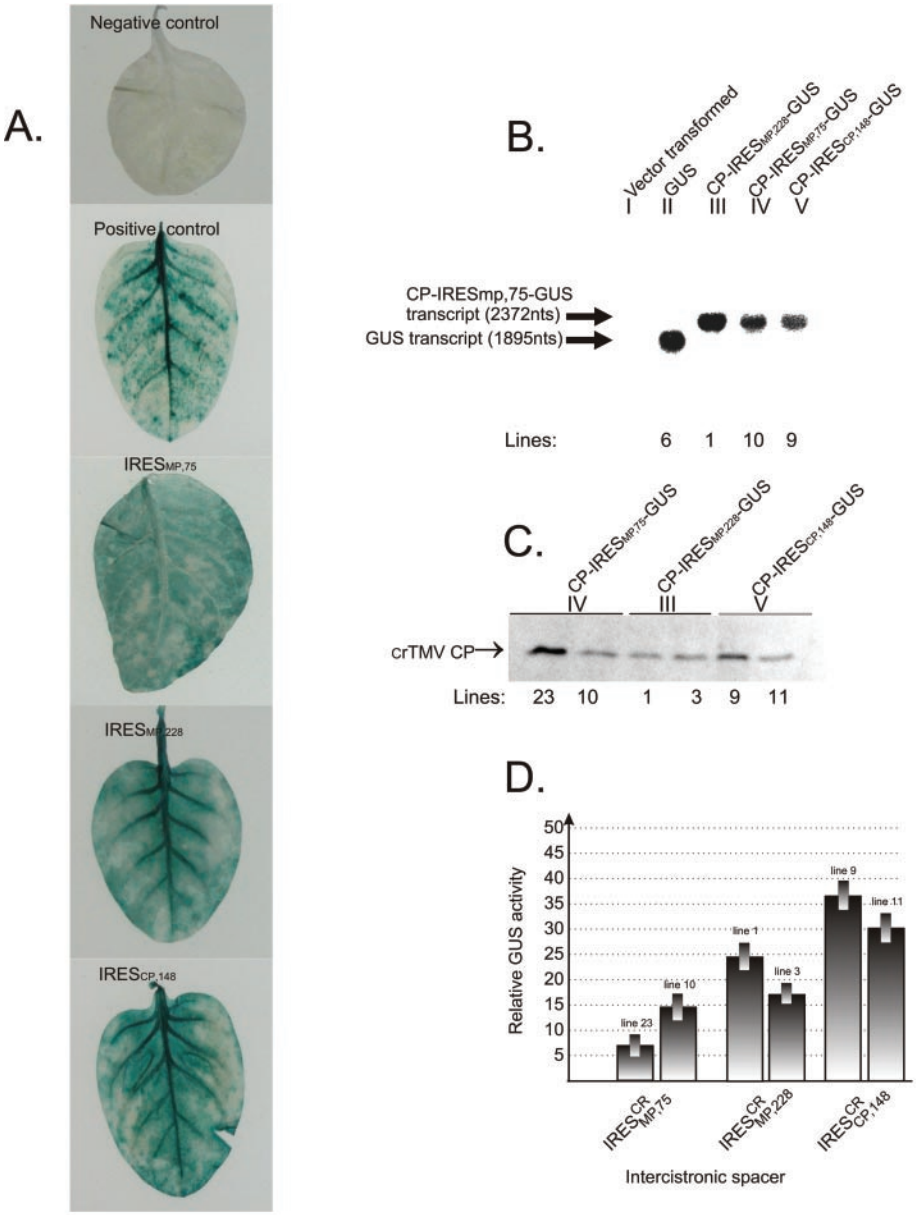

Fig. 1. IRES-mediated GUS gene expression in tobacco plants transgenic for dicistronic CP-IRES-GUS constructs. Five series of transgenic plants differing in IRES sequences were generated: (I) Negative control: vector-transformed plants; (II) positive control: plants transgenic for monocistronic GUS gene; (III–V) $IRES^{CR}_{MP,228}$, $IRES^{CR}_{MP,75}$, and $IRES^{CR}_{CP,148}$ elements, respectively, were used as intercistronic spacers. (*A*) Histochemical detection of GUS activity. (*B*) Northern blot of total RNA isolated from transgenic tobacco leaves probed with a GUS gene DNA probe. Positions of synthetic monocistronic (GUS) and dicistronic (CP-$IRES^{CR}_{MP,75}$-GUS) RNA transcripts are marked by arrows. (*C*) Western blot analyses of the crTMV CP in transgenic tobacco leaves. The arrow indicates the position of crTMV CP. Arabic numerals (*B*, *C*) denote the number of the transgenic plant line used. Roman numerals denote transgenic plants transformed with different constructs indicated above. (*D*) $IRES^{CR}$-mediated GUS activity expressed in two different transgenic lines (denoted by Arabic numerals). The relative GUS activity was normalized to the CP content measured by densitometry of the CP bands presented in *C*.

activity of the 148-nt $IRES^{CR}_{CP,148}$ was invariably the highest in all of the cells tested. Remarkably, it was even higher than that of $IRES^{EMCV}$ in HeLa cells (Fig. 2*B*). This observation provides additional evidence that the requirements for IRES-mediated translation may differ *in vitro* and *in vivo*. These results demonstrate an unusual cross-kingdom conservation of the crTMV $IRES^{CR}_{CP,148}$ activity. It was therefore reasonable to assume that some specific features of this sequence are responsible for its functional universality.

Deletion Analysis. The structural organization of $IRES^{CR}_{CP,148}$ is relatively simple (see figure 6 in ref. 22). It can be folded into a secondary structure containing the 32-nt polypurine tract (PPT), PPT_{32}, upstream of a potentially stable hairpin, and the 11-nt PPT_{11}, just downstream of this stem–loop structure (22). In addition, the short 5′-5562-AGAAGUA-5568–3′ motif (PPT_7) is located downstream of PPT_{32} (22). Earlier studies had shown that neither the 5′ nor the 3′ half of $IRES^{CR}_{CP,148}$, comprising nucleotides 5456–5568 and 5569–5603, was active as single sequences (22). Here we analyzed whether less drastic deletions of $IRES^{CR}_{CP,148}$ retain activity. In fact, sequence 5501–5603 (PPT^{+}_{32}), which includes all three of the PPTs, retained about 70% of intact $IRES^{CR}_{CP,148}$ activity, whereas sequence 5533–5603 (PPT^{-}_{32}), lacking PPT_{32}, was inactive in WGE and extremely low in plant protoplasts (Fig. 3 *B* and *C*, respectively). Sequence 5501–5592, lacking PPT_{11}, lost about 30% of activity in WGE (data not shown). In summary, these results indicate that PPT_{32} and PPT_{11} are the essential elements of $IRES^{CR}_{CP,148}$.

Artificial Polypurine (A)-Rich Sequence (PARS) Elements Exhibit IRES Activity *in Vitro* and in Plant and Animal Cells. PPT_{32} contains the 19-nt sequence AAAAGAAGGAAAAAGAAGG (PPT_{19}) representing a direct tandem repeat of the AAAAGAAGG [$(A_4)G(A_2)G_2$] element in combination with the 11-nt sequence GAAGAAAAGGG. A similar motif (GAAAGAAGAAA) is present in PPT_{11} (Fig. 3*A*). Therefore, all three PARSs can be seen as multiple copies of a $G(A)_{2–5}$ module. To test whether these modules are in fact the important elements of $IRES^{CR}_{CP,148}$, artificial PARSs were created and used as intercistronic spacers in dicistronic H-GFP-ICS-GUS constructs. The IRES activity of

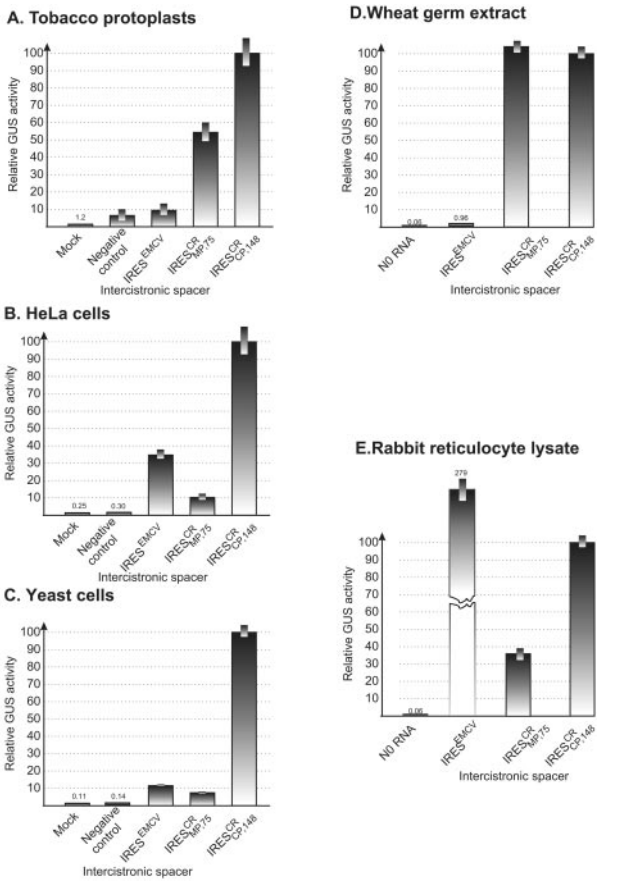

Fig. 2. Cross-kingdom conservation of $IRES^{CR}_{CP,148}$ activity. Expression of the 3′-proximal GUS gene from dicistronic CP-IRES-GUS constructs in tobacco protoplasts (*A*), HeLa (*B*), yeast (*C*) cells and cell-free translation systems WGE (*D*), and RRL (*E*). The 72-nt synthetic GC-rich polylinker-derived (PL_{72}) spacer (23) (*A*, *B*) and the 148-nt region upstream from start codon of the CP gene of TMV U1 ($U1^{SP}_{CP,148}$, ref. 22) (*C*) were used as negative controls. GUS gene expression in HeLa (*B*) and yeast (*C*) cells transfected with animal cell or yeast cell promoter-based dicistronic constructs H-GFP-IRES-GUS and CP-IRES-GUS, respectively.

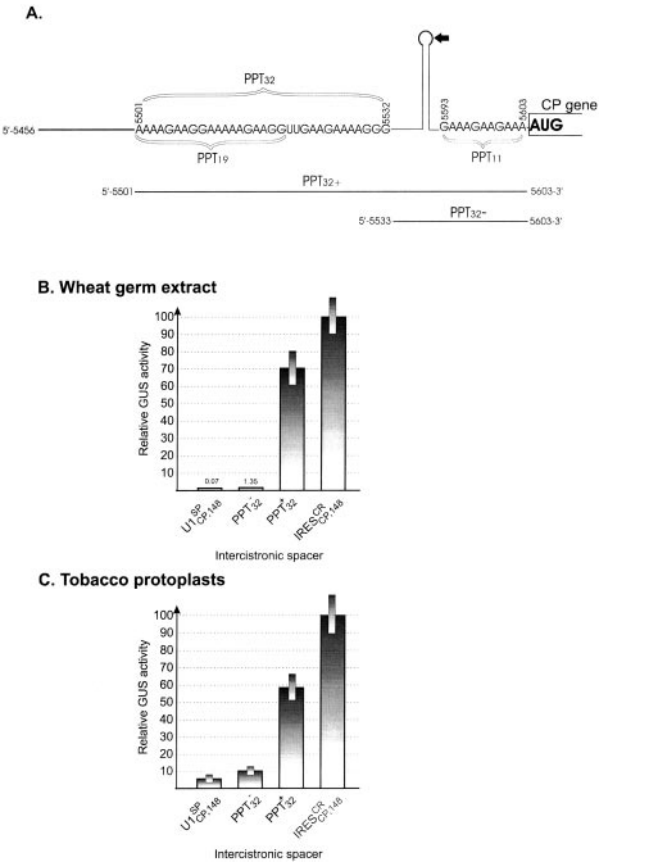

Fig. 3. $IRES^{CR}_{CP,148}$ deletion analysis. (*A*) Simplified schematic representation of the $IRES^{CR}_{CP,148}$ structure (see ref. 22 for details) and its deletion mutants. Letters indicate the sequences of the 32-nt polypurine tract PPT_{32} (which includes the 19-nt element PPT_{19}) located upstream of the hairpin–loop structure and the 11-nt tract (PPT_{11}) just upstream of the CP gene, respectively. Arabic numerals indicate the nucleotide positions in full-length crTMV genomic RNA (21). The arrow points to the position resulting in formation of two deletion mutants (Δ5′IREScp and Δ3′IREScp) described in ref. 22. The lines indicated by PPT^{+}_{32} and PPT^{-}_{32} correspond to the respective $IRES^{CR}_{CP,148}$ deletion mutants used in the present study. GUS gene expression by internal translation from dicistronic constructs in WGE (*B*) and tobacco protoplasts (*C*) under control of the intact $IRES^{CR}_{CP,148}$ and its deletion mutants (PPT^{+}_{32} and PPT^{-}_{32}). $UI^{SP}_{CP,148}$ sequence (22) was taken as a negative control.

these artificial sequences was examined in WGE, tobacco protoplasts, and HeLa cells (Fig. 4). $(PPT_{19})_4$ and $(PPT_{19})_8$ were in fact able to promote internal translation in WGE (Fig. 4*A*). It is noteworthy that $(GAAA)_{16}$ also directed translation of the downstream GUS gene efficiently, despite the fact that translation of the first gene (GFP) was invariably blocked by the hairpin structure H. The results of *in vitro* translation are generally consistent with those experiments made in tobacco protoplasts (Fig. 4*B*) and HeLa cells (Fig. 4*C*). The IRES activity of the sequence $(GAAA)_{16}$ was notably high in both tobacco protoplasts and human HeLa cells. Comparable levels of activity in HeLa cells were exhibited by the spacers $(GAAA)_{16}$ and $(PPT_{19})_4$ (Fig. 4*C*). Significantly, these levels approached (or even exceeded) the level of GUS gene expression promoted by $IRES^{EMCV}$ in HeLa cells (Fig. 4*C*). On the other hand, the levels of GUS gene expression mediated by $(GUUU)_{16}$ were very low *in vitro* (Fig. 4*A*) and in tobacco protoplasts (Fig. 4*B*), and the 68-nt GCU-rich (GCU-R) sequence did not exhibit IRES activity in HeLa cells (Fig. 4*C*).

To estimate the impact of each type of purine on the IRES activity of PARSs, the homopolymers $poly(A)_{60}$ and $poly(G)_{60}$ were inserted into the bicistronic transcripts H-GFP-GUS and tested (Table 1). $Poly(G)_{60}$ exhibited no IRES activity *in vitro*, whereas $poly(A)_{60}$ promoted GUS gene expression even more efficiently than $IRES^{CR}_{CP,148}$. To show that the GUS gene is not being translated from degraded H-GFP-$poly(A)_{60}$-GUS RNA, the ^{32}P-labeled dicistronic transcripts were incubated in RRL. No significant changes in electrophoretic mobility or integrity of transcripts were observed after 60-min incubation (data not shown). Remarkably, the IRES activity of the $poly(A)_{60}$ sequence was drastically reduced in HeLa cells, suggesting that a combination of A and G nucleotides is required for IRES activity *in vivo* (Table 1). Taken together, these data suggest that multiple PARS modules are responsible for conservation of cross-kingdom activity of $IRES^{CR}_{CP,148}$.

IRES Activity of a PARS-Containing 5′UTR Derived from Tobacco Heat-Shock Factor mRNA. It is reasonable to expect that IRES-mediated translation is typical for mRNAs with long and highly structured 5′UTRs. The 5′UTR sequences of a number of plant genes encoding heat-shock protein (HSP) mRNAs were examined for the presence of purine-rich tracts by using the European Molecular Biology Laboratory cDNA nucleotide database. Some of the sequences analyzed contained PARSs of different sizes (accession nos. AB014483, AB017273, AF005993, and AF035460). For example, two long polypurine tracks were revealed in the 453-nt 5′UTR of NtHSF-1 mRNA (accession no. AB014483), 5′-74-

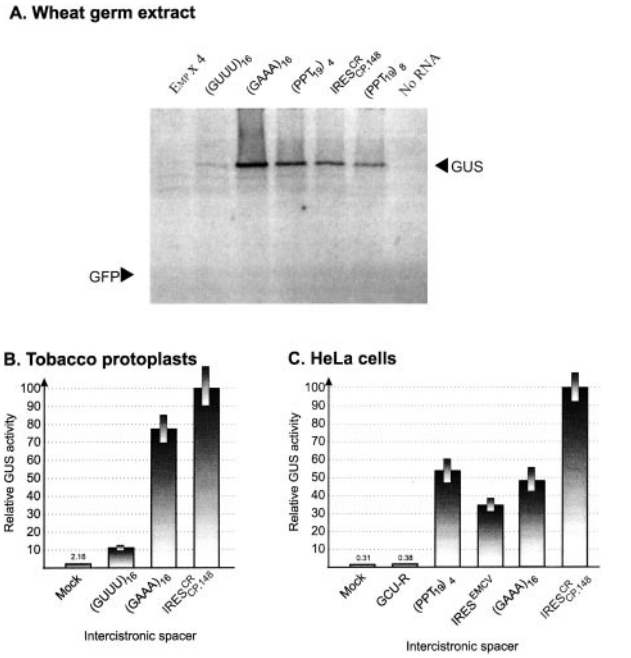

Fig. 4. Comparative dicistronic analysis of IRES activities of multiple G(A)$_3$ modules and natural IRESs (IRES$^{CR}_{CP,148}$ and IRES^{EMCV}) in WGE (*A*), tobacco protoplasts (*B*), and HeLa cells (*C*). Artificial sequences tested: (*i*) $(PPT_{19})_4$ and $(PPT_{19})_8$ representing the tandem repeats of four (76-nt) and eight (152-nt) copies of the 19-nt AAAAGAAGGAAAAAGAAGG sequence derived from PPT_{32} (see Fig. 3), respectively; (*ii*) the 64-nt $(GAAA)_{16}$ sequence consisting of 16 G(A)$_3$ elements; (*iii*) control U-rich sequence $(GUUU)_{16}$; (*iv*) the control Emp × 4 sequence consisting of four copies of the U-rich CGUUUGCUUUUUGUAGUA element derived from another crTMV IRES (IRES$^{CR}_{MP,75}$) and (*v*) the GCU-rich sequence (GCU-R) containing four copies of CGCGGGCG blocks linked via the 7-nt sequence UUUGUUU used as an additional negative control. (*A*) Analysis of proteins directed in WGE by dicistronic H-GFP-ICS-GUS T7 transcripts containing artificial sequences as the intercistronic spacer. Arrows indicate the position of GUS and GFP. (*B* and *C*) GUS gene expression in tobacco protoplasts (*B*) and HeLa (*C*) cells transfected with dicistronic GFP-IRES-GUS constructs containing different IRES sequences. "Mock" indicates that DNA-free solution was used for transfection.

AAAGAAGAGAGAAAACUGAAAAGGCAGAAAA-105–3′ and 5′-420-AGAGAAACAGAGAAAUACAGGGGAAAAA-CAAGGG**AUG**-456–3′), suggesting that the 5′-leader of NtHSF-1 mRNA exhibits IRES activity. To test this hypothesis, the 453-nt 5′UTR of NtHSF-1 (5′UTR NtHSF) was isolated from tobacco genomic DNA and used as an intercistronic spacer in dicistronic analysis of IRES activity. GFP and GUS expression was analyzed from an H-GFP-5′UTR-NtHSF-GUS construct in RRL (Fig. 5*A*), as well as in transfected tobacco protoplasts (Fig. 5*B*) and HeLa cells (Fig. 5*C*). The presence of H at the 5′-terminal position abolished GFP gene expression initiated by ribosome scanning. However, GUS was expressed by internal initiation in all three systems (Fig. 5) and also in WGE (not shown). The activities were

Table 1. Dicistronic analysis of IRES activity of poly(A) and poly(G) sequences used as an intercistronic spacers

Spacer	Activity in RRL assay (%)	Activity in HeLa cells (%)
Poly(A)$_{60}$	150	10
Poly(G)$_{60}$	1	1
IRES$^{CR}_{CP,148}$	100	100
No RNA (mock)	0.3–0.8	0.5–1.0

The H-GFP-spacer-GUS constructs were used in dicistronic assays. Relative GUS activity was expressed as in ref. 23; IRES$^{CR}_{CP,148}$ activity was taken as 100%. The mean values of three independent experiments are given.

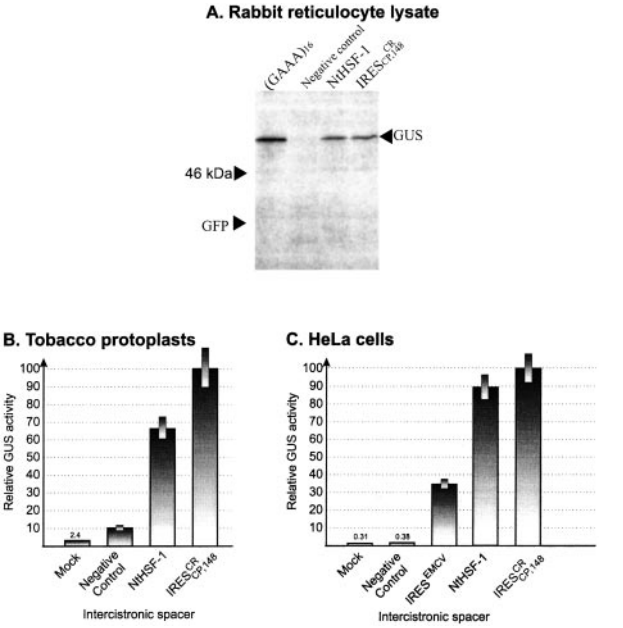

Fig. 5. Dicistronic analysis of IRES activity of the 5′-UTR of NtHSF-1 mRNA (5′UTR NtHSF) in RRL (*A*), tobacco protoplasts (*B*), and HeLa cells (*C*). Tested H-GFP-ICS-GUS RNA transcripts contained as intercistronic spacers the 453-nt 5′UTR of NtHSF-1 mRNA (5′UTR NtHSF) and other synthetic sequences indicated in the legend to Fig. 4.

comparable to those mediated by IRES$^{CR}_{CP,148}$, which was used as a positive control. The 68-nt GCU-rich (GCU-R) sequence and $(GUUU)_{16}$ were used as negative controls. Consequently, the 5′UTR of NtHSF-1 mRNA is an IRES and it exhibits cross-kingdom conservation of internal ribosome entry activity. In both of these otherwise unrelated elements, IRES$^{CR}_{CP,148}$ and IRES$^{NtHSF-1}$, multiple PARS elements are apparently responsible for the activity across the kingdoms analyzed.

Our preliminary results indicate that 5′UTR regions from two other mRNAs of this type, i.e., those encoding the tobacco poly(A)-binding protein and 48-kDa MAP kinase, also promote internal translation in a similar way (see below).

Discussion

IRESs of different origins differ greatly in sequence, length, secondary structure organization, and functional requirements (4). Significant variability was revealed in sets of translation initiation factors and/or noncanonical transacting factors required for the activity of different IRES elements (8, 31–34). It was reasonable to presume that the activity of IRESs in heterologous cell types will be limited because of kingdom-specific differences in cap-independent translation mechanisms. Therefore, it was not unexpected that animal virus (picornaviruses, hepatitis C virus) IRESs were inactive in yeast cells (35–37), despite the fact that IRES-mediated translation of cellular mRNAs has been reported in yeast (38). On the other hand, Urwin *et al.* (11) found that IRES^{EMCV} was active both in animal and, moderately, in plant cells. Taken together, the problem of kingdom-specific differences in IRES activity have so far remained ill-defined.

In the first series of experiments presented here, we showed that IRES$^{CR}_{CP,148}$ and IRES$^{CR}_{MP,75}$ are functionally active in plants transgenic for dicistronic constructs. In addition, these results provided evidence that the IRES$^{CR}_{CP,148}$ sequence functions *in vivo* as an IRES, rather than a transcriptional promoter. Next, the activity of crTMV IRESs (IRES$^{CR}_{CP,148}$ and IRES$^{CR}_{MP,75}$) was compared with that of IRES^{EMCV} in plant, animal, and yeast cells. Surprisingly, comparison of the relative activities of these IRESs showed that IRES$^{CR}_{CP,148}$

invariably exhibited a unique cross-kingdom conservation of maximum IRES activity in organisms as diverse as plants, animal cells and yeast (Fig. 2). The relative efficiency of $IRES^{EMCV}$ was considerably higher in HeLa cells than in nonanimal (tobacco and yeast) cells; however, the relative activity of $IRES^{CR}_{CP,148}$ was invariably the greatest in all cell types tested (Fig. 2 *A–C*).

The $IRES^{CR}_{CP,148}$ sequence can be folded into a simple secondary structure containing two PARSs: the 32-nt PARS (PPT_{32}) upstream of a stable hairpin loop and the 11-nt PARS (PPT_{11}) upstream of the CP gene AUG codon. In addition the short $(PPT)_7$ also contains a PARS motif. Both the PPT_{32} and PPT_{11} sequences could be represented as multiple copies of a $G(A)_{2–5}$ module. Deletion analysis suggested that PARSs might be responsible for the activity of $IRES^{CR}_{CP,148}$. The capacity of artificial sequences to promote internal translation from dicistronic H-GFP-IRES-GUS constructs was tested *in vitro* (WGE, Fig. 4*A*) and *in vivo* (tobacco protoplasts, Fig. 4*B*; HeLa cells, Fig. 4*C*). In agreement with our proposal, the maximum IRES activity was exhibited by PARS elements, in particular by $(GAAA)_{16}$. Remarkably, the activities of $IRES^{CR}_{CP,148}$-derived $(PPT_{19})_4$ and of $(GAAA)_{16}$ appeared to be even somewhat higher than that of $IRES^{EMCV}$. In contrast, the GCU-rich and $(GUUU)_{16}$ spacers had almost no effect on the *in vivo* expression of the second gene. Taken together, our results suggest that PARS elements are archetypal IRES elements responsible for cross-kingdom conservation of IRES activity. To further explore the impact of each type of purine in IRES activity of PARSs, the homopolymers poly(A) and poly(G) were used in dicistronic analysis. Table 1 shows that in RRL, the IRES activity of $poly(A)_{60}$ was very high, whereas $poly(G)_{60}$ was not active. However, in HeLa cells, the IRES activity of $poly(A)_{60}$ was considerably lower than that of $IRES^{CR}_{CP,148}$ (Table 1) or the heteropolymeric artificial IRESs listed above (data not shown). These data indicate that a certain optimal ratio and sequence arrangement of A and G residues is required in order for a PARS to exhibit IRES activity.

As outlined above, IRESs of different origins differ significantly in their translational requirements. In particular, the IRES of hepatitis virus C (HCV) is distinct from the EMCV- and poliovirus-like groups of IRESs. In particular, IRESs of the HCV type bind 40S ribosomal subunits in the absence of initiation factors (8, 33). This phenomenon could be because of: (*i*) specific interaction of the IRES with ribosomal protein(s) (33, 39) and/or (*ii*) base pairing between the IRES and the 18S rRNA. The possibility that complementarity between short modules in eukaryotic mRNAs to 18S rRNA might play a role in IRES–ribosome interaction has been discussed (e.g., see refs. 38, 40–42).

The mechanism of possible interaction of the 40S ribosome with PARS elements is obscure. Apparently, the requirement for internal initiation of translation in a plant cell may differ from requirements in animal and yeast cells. Presumably, such IRES elements can overcome kingdom-specific barriers to translation of the second gene because of their unique capability to exploit only those translation initiation factors and noncanonical transacting proteins that are able to express their function universally in different types of cell. It is possible that the ribosome *per se*, as the most conserved element of the eukaryotic translation apparatus, is responsible for cross-kingdom IRES activity.

It is believed that IRES-mediated translation of cellular mRNAs is activated by physiological stimuli, which play a regulatory role in switching from traditional cap-dependent to IRES-dependent mechanisms (reviewed in refs. 7, 12, and 14). In particular, HSP mRNAs could be regarded as possible candidates for dual cap-dependent and IRES-mediated translation. We suggested that PARSs naturally occurring in long 5′UTRs of plant mRNAs (*i*) confer IRES activity and (*ii*) confer this activity across kingdoms. Two long (32- and 34-nt) and several short PARS elements with multiple $(G)_{1–4}(A)_{2–5}$ modules were identified in the 453-nt 5′UTR of *N. tabacum* heat-shock factor 1 mRNA. When tested in dicistronic constructs (GFP-5′UTR-NtHSF1-GUS), this sequence functioned as an IRES in WGE, RRL, and in tobacco protoplasts and human HeLa cells. These data further support the idea that PARSs are involved in cross-kingdom conservation of IRES activity.

To the best of our knowledge, no IRES elements of plant origin have been described to date.

Analysis of European Molecular Biology Laboratory databases showed that the 5′UTRs of numerous cellular mRNAs contain PARSs that could be regarded as putative plant IRESs. Our preliminary results indicate that two additional mRNAs of this type, i.e., those encoding the tobacco poly(A)-binding protein (43) and 48-kDa mitogen-activated protein kinase (44), also promote internal translation. The approach could thus be used to identify IRES elements in eukaryotic genomes.

We thank Dr. Helen Rothnie for correcting this paper and R. Karapetian for technical assistance in yeast work. This work was funded in part by the Russian Foundation for Basic Research (RFBR) Grant 9904-49119, International Association for the Promotion of Cooperation with Scientists from the New Independent States of the Former Soviet Union (INTAS) Grant 99-00720, RFBR-Deutsche Forschungsgemeinschaft Grant 9904-04000, and Federation of European Biochemical Societies short-term fellowship grant (M.S.).

1. Pain, V. M. (1996) *Eur. J. Biochem.* **236,** 747–771.
2. Kozak, M. (1999) *Gene* **234,** 187–208.
3. Dever, T. E. (1999) *Trends Biol. Sci.* **24,** 398–403.
4. Jackson, R. J. (2000) in *Translation Control of Gene Expression*, eds. Sonenberg, N., Hershey, J. W. B. & Mathews, W. B. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 127–184.
5. Gale, M., Tan, S.-L. & Katze, M. G. (2000) *Microbiol. Mol. Biol. Rev.* **64,** 239–280.
6. Belsham, G. J. & Sonenberg, N. (2000) *Trends Microbiol.* **8,** 330–335.
7. Martinez-Salas, E., Ramos, R., Lefuente, E. & Lopez de Quinto, S. (2001) *J. Gen. Virol.* **82,** 973–984.
8. Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I. & Hellen, C. U. T. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 7029–7036.
9. Sasaki, J. & Nakashima, N. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 1512–1515.
10. Wilson, J. E., Powell, M., Hoover, S. E. & Sarnow, P. (2000) *Mol. Cell. Biol.* **20,** 4990–4999.
11. Urwin, P., Yi, L., Martin, H., Atkinson, H. & Gilmartin, P. M. (2000) *Plant J.* **24,** 583–589.
12. Holcik, M., Lefebvre, C., Yeh, C., Chow, T. & Korneluk, R. G. (1999) *Nat. Cell. Biol.* **1,** 190–192.
13. Werner, R. (2000) *Int. Union Biochem. Mol. Biol. (IUBMB) Life* **50,** 173–176.
14. Sachs, A. B. (2000) *Cell* **101,** 243–245.
15. Carrington, J. C. & Freed, D. D. (1990) *J. Virol.* **64,** 1590–1597.
16. Tacke, E. D., Prufer, F., Salamini, F. & Rohde, W. (1990) *J. Gen. Virol.* **71,** 2265–2272.
17. Thomas, A. A. M., Ter Haar, E., Wellink, J. & Voorma, H. O. (1991) *J. Virol.* **65,** 2953–2959.
18. Gallie, D. R., Tanguay, R. & Leathers, V. (1995) *Gene* **165,** 233–238.
19. Niepel, M. & Gallie, D. R. (1999) *J. Virol.* **73,** 9080–9088.
20. Palukaitis, P. & Zaitlin, M. (1986) in *Plant Virus*, eds. Van Regenmortel, M. H. V. & Fraenkel-Conrat, M. (Plenum, New York) Vol. 2, pp. 105–131.
21. Dorokhov, Yu. L., Ivanov, P. A., Novikov, V. K., Agranovsky, A. A., Morozov, S. Yu., Efimov, V. A., Casper, R. & Atabekov, J. G. (1994) *FEBS Lett.* **350,** 5–8.
22. Ivanov, P. A., Karpova, O. V., Skulachev, M. V., Tomashevskaya, O. L., Rodionova, N. P., Dorokhov, Yu. L. & Atabekov, J. G. (1997) *Virology* **232,** 32–43.
23. Skulachev, M. V., Ivanov, P. A., Karpova, O. V., Korpela, T., Rodionova, N. P., Dorokhov, Yu. L. & Atabekov, J. G. (1999) *Virology* **263,** 139–154.
24. Toth, R. L., Chapman, S. Carr, F. & Santa Cruz, S. (2001) *FEBS Lett.* **489,** 215–219.
25. Lazo, G. R., Stein, P. A. & Ludwig, R. A. (1991) *Biotechnology* **9,** 963–967.
26. Horsh, R. B., Fry, J. E., Hoffman, N. L., Eichholtz, D., Rogers, S. G. & Fraley, R. T. (1985) *Science* **227,** 1229–1231.
27. Jefferson, R. A. (1987) *Plant Mol. Biol. Rep.* **5,** 387–405.
28. Hill, J., Donald, K. A., Griffiths, D. E. & Donald, G. (1991) *Nucleic Acids Res.* **19,** 5791.
29. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248–254.
30. Wyatt, L. S., Moss, B. & Rozenblatt, S. (1995) *Virology* **210,** 202–205.
31. Kolupaeva, V. G., Hellen, C. U. T. & Shatsky, I. N. (1996) *RNA* **2,** 1199–1212.
32. Kaminski, A. & Jackson, R. J. (1998) *RNA* **4,** 626–638.
33. Pestova, T. V., Shatsky, I. N., Fletcher, S. P., Jackson, R. J. & Hellen, C. U. T. (1998) *Genes Dev.* **12,** 67–83.
34. Pilipenko, E. V., Pestova, T. V., Kolupaeva, V. G., Khitrina, E. V., Poperechnaya, A. N., Agol, V. I. & Hellen, C. U. T. (2000) *Genes Dev.* **14,** 2028–2045.
35. Coward, P. & Dasgupta, A. (1992) *J. Virol.* **66,** 286–295.
36. Evstafieva, A. G., Beletsky, A. V., Borovjagin, A. V. & Bogdanov, A. A. (1993) *FEBS Lett.* **335,** 273–276.
37. Das, S., Ott, M., Yamane, A., Venkatesan, A., Gupta, S. & Dasgupta, A. (1998) *Front. Biosci.* **3,** D1241–D1252.
38. Zhou, W., Edelman, G. M. & Mauro, W. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 1531–1536.
39. Fukushi, S., Okada, M., Stahl, J., Kageyama, T., Hoshino, F. B. & Katayama, K. (2001) *J. Biol. Chem.* **276,** 20824–20826.
40. Pestova, T. V., Hellen, C. U. T. & Wimmer, E. (1991) *J. Virol.* **65,** 6194–6204.
41. Mauro, V. P. & Edelman, G. M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 422–497.
42. Hu, M. C.-Y., Tranque, P., Edelman, G. M. & Mauro, V. P. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 1339–1344.
43. Le, H. & Gallie, D. R. (2000) *Plant Sci.* **152,** 101–114.
44. Zhang, S. & Klessig, D. F. (1997) *Plant Cell* **9,** 809–824.