Identification of two short internal ribosome entry sites selected from libraries of random oligonucleotides

Geoffrey C. Owens*†, Stephen A. Chappell‡, Vincent P. Mauro‡, and Gerald M. Edelman*‡

*The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121; and ‡Department of Neurobiology, The Scripps Research Institute, and The Skaggs Institute for Chemical Biology, 10550 North Torrey Pines Road, La Jolla, CA 92037

Contributed by Gerald M. Edelman, December 20, 2000

Sequences that control translation of mRNA may play critical roles in regulating protein levels. One such element is the internal ribosome entry site (IRES). We previously showed that a 9-nt segment in the 5′ leader sequence of the mRNA encoding Gtx homeodomain protein could function as an IRES. To identify other short sequences with similar properties, we designed a selection procedure that uses a retroviral vector to express dicistronic mRNAs encoding enhanced green and cyan fluorescent proteins as the first and second cistrons, respectively. Expression of the second cistron was dependent upon the intercistronic sequences and was indicative of IRES activity. B104 cells were infected with two retroviral libraries that contained random sequences of 9 or 18 nt in the intercistronic region. Cells expressing both cistrons were sorted, and sequences recovered from selected cells were reassayed for IRES activity in a dual luciferase dicistronic mRNA. Two novel IRESes were identified by this procedure, and both contained segments with complementarity to 18S rRNA. When multiple copies of either segment were linked together, IRES activities were dramatically enhanced. Moreover, these synthetic IRESes were differentially active in various cell types. These properties are similar to those of the previously identified 9-nt IRES module from Gtx mRNA. These results provide further evidence that short nucleotide sequences can function as IRESes and support the idea that some cellular IRESes may be composed of shorter functional modules. The ability to identify IRES modules with specific expression properties may be useful in the design of vectors for biotechnology and gene therapy.

Translation in eukaryotes is thought to occur predominantly by a ribosome scanning mechanism in which a preinitiation complex comprising a 40S ribosome subunit, the initiator tRNA, and other factors is recruited by the cap structure at the 5′ end of mRNA (1–3). The translation of some mRNAs, however, occurs by a cap-independent mechanism. For instance, the genomes of some RNA viruses, in particular the *Picornaviridae*, although uncapped, are translated efficiently because of the recruitment of ribosomes at sequences termed internal ribosome entry sites (IRESes) (3). IRESes have also been identified in cellular mRNAs (4), and they appear to make it possible for the translation of some mRNAs to continue under conditions where cap-dependent translation is reduced, for example, during viral infection, cell cycle progression, stress, and apoptosis (5–10). The sequences of the cellular IRESes characterized to date are dissimilar (3), and although a Y-shaped RNA secondary structure has been proposed to occur in some of these cellular IRESes (11), there is no convincing evidence supporting the idea that such a motif is essential for internal initiation at these IRESes.

In contrast to the notion that IRESes need to possess particular secondary structures, our recent identification and characterization of an IRES within the mRNA encoding the Gtx homeodomain protein points to the importance of short nucleotide sequences. The Gtx IRES contains several nonoverlapping segments that were shown to mediate internal initiation. Within one of these segments, a 9-nt sequence was identified that functioned independently as an IRES (12). When synthetic IRESes composed of multiple linked copies of this 9-nt IRES module were examined for activity, internal initiation increased greatly, which was reminiscent of the effects on transcription of multimerizing enhancers (13–15). The 9 nt of the Gtx IRES module are complementary to a sequence contained within 18S rRNA, and it has been suggested that the 18S rRNA within 40S ribosomal subunits may directly base pair to this sequence (16–18). The notion that some cellular IRESes are modular is consistent with the observations that nonoverlapping fragments of some IRESes were able to independently mediate internal initiation (12, 19–23).

Further analysis of other cellular IRESes is likely to identify other short nucleotide sequences that function as IRES modules. Inasmuch as sequence comparisons of different cellular IRESes have failed to identify any motifs that are common to all cellular IRESes, it is possible that internal initiation would be mediated by a wide variety of sequences. We therefore designed a procedure to select functional IRES modules from a population of random oligonucleotides. Sequences were assayed for IRES activity by inserting them between two open reading frames in a dicistronic mRNA and showing that they mediate efficient translation of the second cistron independently of the translation of the first cistron (24, 25).

In our IRES selection method, the dicistronic mRNA is expressed by a recombinant retroviral vector, with the first and second cistrons encoding enhanced green fluorescent protein (EGFP) and enhanced cyan fluorescent protein (ECFP), respectively. Translation of ECFP in this dicistronic arrangement is therefore dependent upon sequences inserted into the intercistronic region. In the present experiments, cells expressing both cistrons were selected by fluorescence-activated cell sorting (FACS). The intercistronic sequences were isolated and retested in a dicistronic mRNA, with the first and second cistrons encoding luciferase proteins from *Renilla* (sea pansy) and *Photinus* (firefly), respectively.

Two of the short sequences identified by this selection procedure exhibited enhanced IRES activity when expressed as multiple linked copies. These synthetic IRES constructs were also differentially active in various cell types and contained segments that are complementary to 18S rRNA. This latter finding is consistent with the idea that some mRNAs may recruit ribosomes by base pairing directly with rRNA.

Materials and Methods

Construction of Retroviral Libraries. The retroviral vector used in this study, designated MESV/EGFP/ECFP/RSVpro, is shown

Abbreviations: IRES, internal ribosome entry site; EGFP, enhanced green fluorescent protein; ECFP, enhanced cyan fluorescent protein; FACS, fluorescence-activated cell sorting; EMCV, encephalomyocarditis virus; ICS, intercistronic sequence.

†To whom reprint requests should be addressed. E-mail: gowens@nsi.edu.

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.

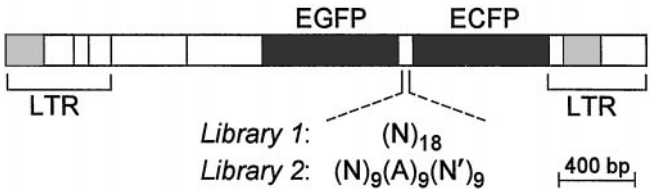

Fig. 1. The IRES selection vector: schematic outline of the retroviral vector used in these studies. This vector expresses a dicistronic mRNA with EGFP as the first cistron and ECFP as the second cistron. This dicistronic mRNA is under the transcriptional control of the upstream long terminal repeat (LTR). Both upstream and downstream LTRs contain the U3 region from Rous sarcoma virus (lightly shaded boxes). The two libraries of random nucleotide sequences, where N represents A, C, G, or U, were inserted as indicated (see *Materials and Methods*). In the second library, N and N′ represent different random nucleotide sequences.

in Fig. 1. It is based on a previously described vector used in a method to select DNA sequences capable of promoter activity (26). This modified vector expresses a dicistronic mRNA that encodes EGFP (CLONTECH) as the first cistron and ECFP (CLONTECH) as the second cistron. Transcription of the dicistronic mRNA is driven by the upstream long terminal repeat. The Rous sarcoma virus U3 enhancer region is located in both the upstream and downstream long terminal repeats. The intercistronic region of the dicistronic mRNA contains a polylinker with unique *Nhe*I, *Pac*I, and *Mlu*I restriction sites. The *Mlu*I site is located 10 nt upstream of the ECFP initiation codon. Two retroviral libraries were generated. In the first library, an oligonucleotide containing 18 random nucleotides, $(N)_{18}$, was cloned into the *Mlu*I site of the polylinker. The sequence of this oligonucleotide is acgcgtgatcca$(N)_{18}$cgagcgacgcgt (26). In the second library, an oligonucleotide containing two segments of nine random nucleotides, $(N)_9$, was cloned into the *Pac*I and *Mlu*I sites of the polylinker. The sequence of this oligonucleotide is ttaattaagaattcttctgacat$(a)_9$ttctgacat$(a)_9$ttctgacat$(a)_9(N)_9(N)_9(a)_9(N')_9(a)_9$gactcacaaccccagaaacagacatacgcgt, where N and N′ are different random nucleotide sequences. The design of this oligonucleotide was based on another previously described oligonucleotide $(S_{III}/S_{II})_5\beta$ (12). This oligonucleotide did not have IRES activity and was used as a spacer control. The first library consisted of $\approx 2.5 \times 10^5$ bacterial clones, and the second consisted of $\approx 1.5 \times 10^5$ bacterial clones; thus both libraries used in these experiments represented only a small fraction of the potential sequence complexity of the random oligonucleotides ($\approx 6.9 \times 10^{10}$).

Transduction and Selection of Cells. The retroviral libraries were packaged in Cos1 cells. Subconfluent cells were triply transfected with the use of the FuGENE 6 reagent (Roche Molecular Chemicals) with plasmids encoding (*i*) the retroviral library; (*ii*) Moloney murine leukemia virus *gag* and *pol* genes (pCMV-$GP_{(Sal)}$), kindly provided by the University of California at San Diego Gene Therapy Program, Vector Core Lab; and (*iii*) the vesicular stomatitis virus G glycoprotein (27). After 48 h, retroviral particles were recovered from culture supernatant, filtered through a 0.45-μm membrane, and then used to infect B104 rat neural tumor cells (28). Approximately 2×10^6 Cos1 cells were transfected, and approximately the same number of B104 cells were subsequently infected. After 72 h, cells were harvested and sorted by FACS on a FACSVantage SE (Becton Dickinson). EGFP was excited with an argon laser tuned to 488 nm, and fluorescence was recorded through a 530-nm bandpass filter. ECFP was excited with a krypton/argon laser tuned to 457 nm, and fluorescence was measured through a 495-nm bandpass filter. As controls for the FACS, B104 cells were infected with the following reference viruses: the parent vector (MESV/EGFP/ECFP/RSVpro), a virus encoding EGFP, a virus encoding ECFP, and a virus that contains the IRES from the encephalomyocarditis virus (EMCV) in the intercistronic region of the parent vector. Cells coexpressing both EGFP and ECFP were isolated and returned to culture for 14 days. These cells were then resorted, and high coexpressors were isolated and further expanded in culture for 5–7 days.

Recovery and Analysis of Selected Sequences. Genomic DNA was prepared with a QIA amp DNA miniprep kit (Qiagen, Chatsworth, CA). Intercistronic sequences were PCR amplified with flanking primers and cloned into the intercistronic region of *RP*h (12, 19). *RP*h is a dicistronic vector that encodes *Renilla* luciferase protein as the first cistron and *Photinus* luciferase protein as the second cistron. B104 cells were transiently cotransfected with the dual luciferase vector and with a vector expressing β-galactosidase, and luciferase and β-galactosidase assays were performed as previously described (12). *Photinus* luciferase activity values were normalized for transfection efficiency by means of β-galactosidase activity and were then normalized to the activity of the *RP*h parent vector (first library) or of *RP*h containing the $(S_{III}/S_{II})_5\beta$ oligonucleotide as a spacer control (second library). Sequences were determined with the use of an ABI system sequencer (PE Biosystems, Foster City, CA) and were compared with the use of the CLUSTAL X multiple sequence alignment program (29). Sequence comparisons were also carried out with the use of the BESTFIT program from the Genetics Computer Group software package (30). Sequence matches were evaluated by comparing BESTFIT quality scores with those obtained when the selected sequences were randomly shuffled 10 times and compared with 18S rRNA. Secondary structure predictions were made with the use of MFOLD, version 3.0 (31, 32). Northern blots were performed as previously described (12, 16), with a riboprobe encompassing the entire coding region of the *Photinus* luciferase gene.

Results

Selection of Sequences with IRES Activity. In a previous study, we developed a retrovirus-based strategy to select for promoter elements from a library of random nucleotide sequences (26). In the present study, a similar approach was used to select short sequences that function in the internal initiation of translation. A retroviral vector was constructed that expressed a dicistronic mRNA with EGFP as the first cistron and ECFP as the second cistron (Fig. 1). EGFP and ECFP are spectral variants of green fluorescent protein that have distinct emission maxima (33), and the simultaneous expression of green fluorescent protein variants can be readily analyzed in single cells by FACS (34).

We first constructed a retroviral library in which the intercistronic region contained a random 18-nt sequence. This library, derived from 2.5×10^5 retroviral plasmids, was used to infect approximately 2×10^6 rat B104 neural tumor cells. After 72 h, cells that coexpressed both EGFP and ECFP, corresponding to approximately 0.5% of the cells, were isolated by FACS. These cells were cultured for 14 days and resorted, and high coexpressors, corresponding to approximately 4% of cells, were collected and regrown. Fig. 2*A* shows the profile of EGFP and ECFP expression in these twice-sorted cells. For comparison, Fig. 2*B* shows cells that had been infected with the virus that contained the EMCV IRES between the EGFP and ECFP genes. Both cell populations showed variable expression, suggesting that IRES activity may vary among individual cells, perhaps reflecting cell cycle differences in the population.

Intercistronic sequences contained within the population of twice-sorted cells were isolated by genomic PCR and cloned into the intercistronic polylinker of the *RP*h vector in which the simian virus 40 promoter drives the expression of a dicistronic mRNA encoding two different luciferase genes (12, 19). This dual luciferase vector has a stable hairpin-forming sequence in

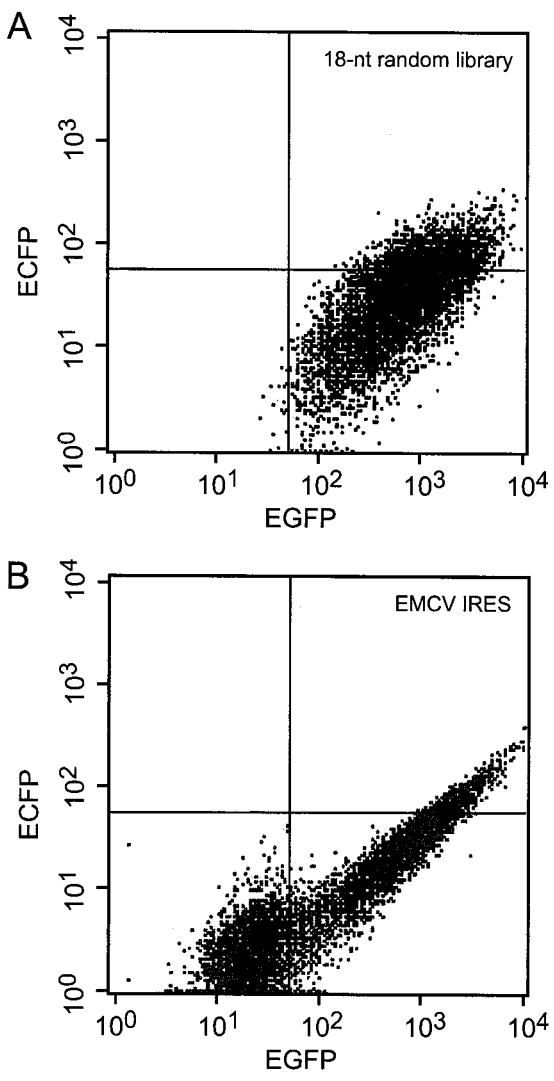

Fig. 2. Profiles of EGFP and ECFP expression by infected cells. Two-dimensional frequency histogram plots show the log of fluorescence emission attributable to EGFP (*x* axis) and ECFP (*y* axis). The quadrants were determined in control experiments with B104 cells that expressed EGFP (*Lower right quadrant*), cells that expressed ECFP (*Upper left quadrant*), and nonexpressing cells (*Lower left quadrant*). Cells in the upper right quadrant expressed both EGFP and ECFP. (*A*) B104 cells infected with the 18-nt random library and sorted twice for coexpression of EGFP and ECFP. (*B*) B104 cells infected with a dicistronic viral vector containing the EMCV IRES in the intercistronic region between EGFP and ECFP.

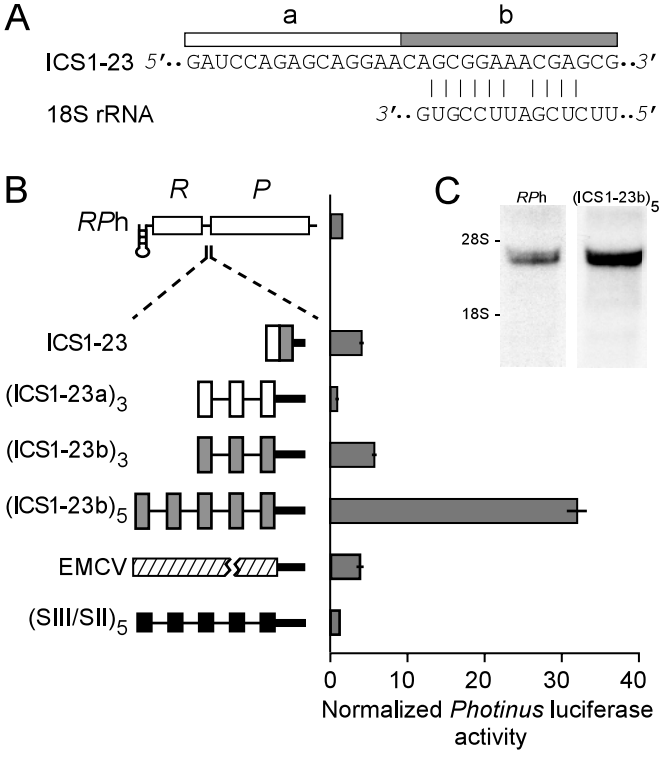

Fig. 3. Sequence and dicistronic analysis of selected sequence ICS1–23. (*A*) Complementarity between a segment of ICS1–23 and 18S rRNA. The upper sequence is that of ICS1–23, and the lower sequence is that of the rat 18S rRNA at nucleotides 1311–1324; complementary nucleotides are indicated by vertical lines. The open and gray bars indicate the nucleotide sequences that comprise sequences designated as ICS1–23a and ICS1–23b, respectively. (*B*) Dicistronic analysis of ICS1–23 and of multiple copies of ICS1–23a and ICS1–23b. A schematic representation of the dual luciferase construct used in this analysis is shown. *Insets* include one copy of selected sequence ICS1–23, three copies of ICS1–23a (open bars), and three or five copies of ICS1–23b (gray bars). Multiple copies of ICS1–23a and ICS1–23b are spaced nine nucleotides apart with poly(A) (thin lines) and separated from the *Photinus* initiation codon by 25 nt of the mouse β-globin 5′ untranslated region (thick black line). Other *Insets* include the EMCV IRES (hatched bar) and the (SIII/SII)β_5 spacer control (12). This spacer control consists of five copies of a 9-nt segment from the mouse β-globin 5′ untranslated region (black bars) spaced 9 nt apart with poly(A). The histogram shows normalized *Photinus* luciferase activities (± SEM) (see *Materials and Methods*) in transfected B104 neuroblastoma cells. (*C*) Northern blot of total RNA purified from B104 cells transfected with either the *RP*h or (ICS1–23b)$_5$/*RP*h dicistronic construct and probed with a *Photinus* luciferase riboprobe. The positions of the 28S and 18S rRNAs are indicated.

the transcribed leader region upstream of the *Renilla* ORF. The hairpin structure blocks scanning ribosomes and therefore suppresses translation of the first cistron. Indeed, inclusion of this structure has been shown to minimize expression of the second cistron by translation reinitiation or leaky scanning (19). Fifty clones were picked at random, and plasmid DNA was prepared, sequenced, and transiently transfected into B104 cells. Of the 45 clones that were successfully sequenced, 39 contained unique 18-nt inserts. The sequences of the other six clones were each represented more than once, which may reflect the relatively low complexity of selected sequences in these twice-sorted cells. All of the sequenced clones were tested in transfected cells, and most activities were weak or at a background level. Only one sequence, designated intercistronic sequence 1–23 (ICS1–23), enhanced *Photinus* luciferase activity by approximately 8-fold relative to the control constructs. This level of activity was similar to that observed for one copy of the previously studied 9-nt Gtx IRES-module (12).

A Segment of ICS1–23 with Complementarity to 18S rRNA Functions as an IRES. A sequence comparison between ICS1–23 and 18S rRNA revealed a complementary match between the 3′ end of the IRES and 18S rRNA at nucleotides 1311–1324 (Fig. 3*A*). This match has a BESTFIT quality score that is significantly greater than that obtained with 10 randomized variations of this sequence (see *Materials and Methods*). The results of other studies from our laboratory have shown that complementarity to 18S rRNA might be the basis for base-pairing interactions with 40S ribosomal subunits (12, 16–18). To address whether the region of complementarity within ICS1–23 had IRES activity, the 30-nt ICS1–23 sequence, which includes the 18-nt random sequence together with 12 nt of flanking sequence, was divided into two segments of 15 nt each. The first 15-nt segment lacked any complementarity to 18S rRNA (ICS1–23a), whereas the

second segment contained the complementary match to 18S rRNA (ICS1–23b). We previously observed that multiple linked copies of the 9-nt Gtx IRES module were more active than the corresponding monomer (12), and we therefore synthesized multimers of each segment of ICS1–23. Each repeated segment was separated by 9 nt of poly(A). As shown in Fig. 3*B*, although three-linked copies of ICS1–23a did not enhance *Photinus* luciferase expression, three- and five-linked copies of ICS1–23b enhanced *Photinus* luciferase activity as compared with ICS1–23, indicating that within ICS1–23, the nucleotides with complementarity to 18S rRNA might function as an IRES. A Northern analysis of RNA from cells expressing the five-linked copies of ICS1–23b identified a single hybridizing band corresponding in size to the full-length dicistronic mRNA (Fig. 3*C*). This finding confirmed that ICS-23b did not enhance *Photinus* luciferase activity by other mechanisms such as alternative splicing or by functioning as a promoter.

Multiple Linked Copies Greatly Enhance Activity of a Second IRES Identified by Selection. To search for even shorter IRES modules, a second retroviral library was generated with two random 9-nt segments separated by a 9-nt poly(A) spacer in the intercistronic region of the encoded dicistronic mRNA. This design was prompted by the results of an earlier study in which we showed that two copies of the 9-nt Gtx IRES module separated by a 9-nt poly(A) spacer exhibited greater IRES activity than a single copy of the module (12). We reasoned that stronger IRESes might be isolated as a consequence of additive or synergistic interactions between individual short IRES modules. B104 cells ($\approx 2 \times 10^6$) were transduced with the second retroviral library, which was derived from 1.5×10^5 retroviral plasmids. Approximately 0.3% of the cells were selected by FACS. As with the first library, cells were cultured and resorted, and approximately 3% of these cells were high coexpressors. Sequences were recovered by genomic PCR and shotgun cloned into the intercistronic region of the *RP*h. One hundred clones were picked at random, and 84 were successfully sequenced, yielding 37 different sequences. Fifteen of these sequences were represented two or more times, indicating that the complexity of the sequences represented in these twice-sorted cells was somewhat lower than that of the first library. When tested by transient transfection in B104 cells, most sequences enhanced *Photinus* luciferase activity weakly (≈2-fold or less above background), and none were as active as ICS1–23. Inasmuch as six of these sequences were isolated four or more times from the twice-sorted cells, these sequences were chosen for further analysis. Each of these sequences contained two 9-nt segments, which were tested individually as five-linked copies. One of these constructs, containing a 9-nt segment designated ICS2–17.2, showed enhanced *Photinus* luciferase activity. In contrast, five-linked copies of ICS2–17.1, the other 9-nt segment contained within selected sequence ICS2–17, did not have IRES activity (Fig. 4*A*). RNA analysis confirmed that a single transcript was produced from this construct (Fig. 4*B*) and that the increase in *Photinus* luciferase activity was derived from an intact dicistronic mRNA, supporting the notion that ICS2–17.2 functions as an IRES. In addition, five-linked copies of both ICS1–23b and ICS2–17.2 were tested in the 5′ untranslated region of a monocistronic reporter mRNA. In seven cell lines tested, $(ICS1–23b)_5$ blocked translation by approximately 70%, and $(ICS2–17.2)_5$ enhanced translation slightly (data not shown). In both cases, mRNA levels appeared to be unaffected. This observation provides further evidence that ICS1–23b and ICS2–17.2 functioned as IRESs in the dicistronic mRNAs and not as transcriptional promoters or enhancers. As with ICS1–23b, sequence comparisons identified a complementary match between ICS2–17.2 and 18S rRNA (Fig. 4*C*) with a BESTFIT quality score that is significantly greater than that obtained with 10 randomized variations of this sequence (see *Materials and Methods*).

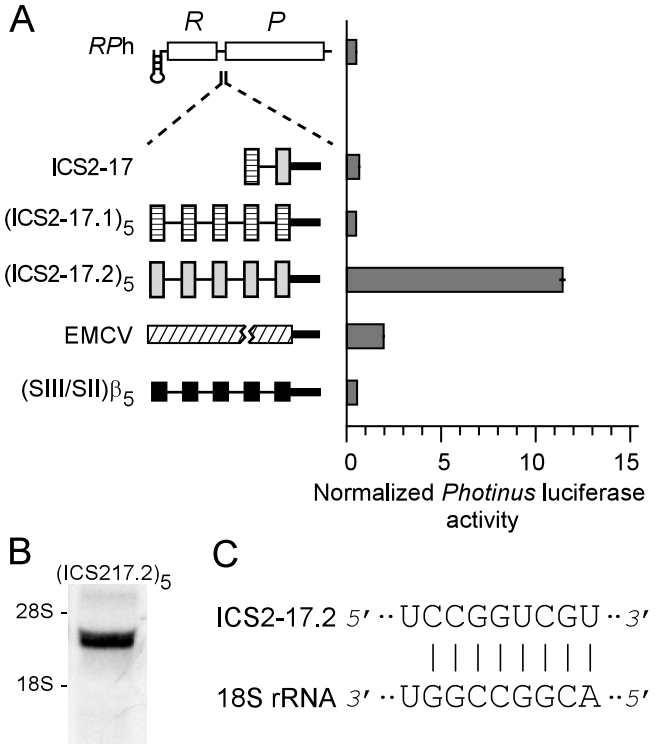

Fig. 4. Dicistronic analysis of ICS2–17 and segments of this sequence. (*A*) *Insets* include one copy of selected sequence ICS2–17 (hatched bar, thin line, gray bar) and five copies of either ICS2–17.1 (hatched bars) or ICS2–17.2 (gray bars). Other *Insets* are as described in Fig. 3. The histogram shows normalized *Photinus* luciferase activities (± SEM) in transfected B104 cells. (*B*) Northern blot of total RNA purified from B104 cells transfected with the $(ICS-2–17.2)_5$/*RP*h dicistronic construct and probed with a *Photinus* luciferase riboprobe. The positions of the 28S and 18S rRNAs are indicated. (*C*) Complementarity between ICS2–17.2 and 18S rRNA. The upper sequence is that of the selected sequence ICS2–17.2, the lower sequence is that of rat 18S rRNA at nucleotides 68–76, with complementary nucleotides indicated by vertical lines.

The Activity of the Selected IRES Modules Varies in Different Cell Lines. ICS1–23b and ICS2–17.2 were selected with the use of the neuroblastoma cell line B104. To determine whether these two IRES modules were also active in other cell types, we assayed five-linked copies of each module in several other cell lines. Although both modules were active in every cell line tested, the activities of these synthetic IRESes varied by as much as 10-fold between cell lines and varied with respect to each other (Fig. 5). It was notable that the pattern of activity of the ICS-23b module in the different cell lines tested was similar to that observed for 10-linked copies of the 9-nt Gtx IRES module (12).

Discussion

To complement our systematic analysis of individual naturally occurring cellular IRESes, we have developed a high throughput *in vivo* selection method to identify IRES modules. In screens of two retroviral libraries containing random sequences of either 18 or 9 nt inserted into the intercistronic region of the expressed dicistronic mRNAs, two different sequences with IRES activity were obtained. A segment from each of these sequences, ICS1–23b and ICS2–17.2, showed enhanced IRES activity when tested as multiple linked copies. Both sequences were validated as IRES modules by ruling out the possibilities that these sequences functioned as splice acceptor sites or as transcriptional promoters.

In this study, two IRES modules were selected from a minute sampling of the total complexity of the random oligonucleotides.

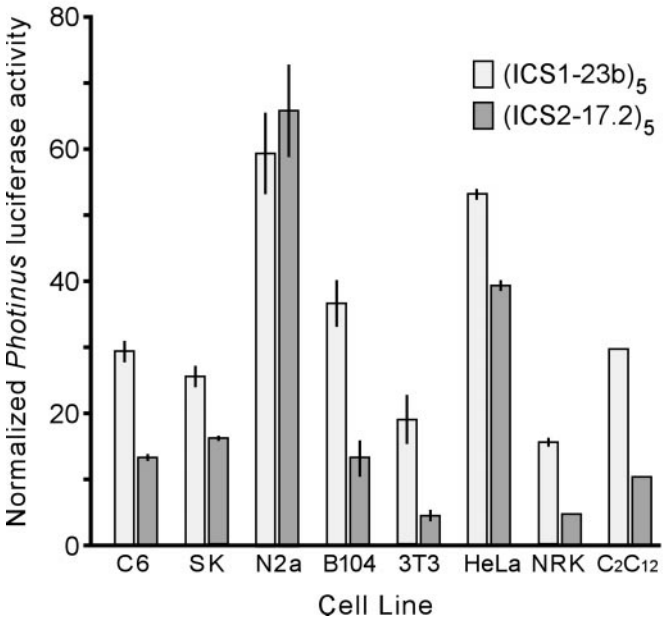

Fig. 5. IRES activities of synthetic IRESs in various cell lines. The multimeric dual luciferase constructs (ICS1–23b)$_5$/*RP*h (see Fig. 3) and (ICS2–17.2)$_5$/*RP*h (see Fig. 4) were transfected into eight different cell lines. Vertical bars represent mean normalized *Photinus* activities with corresponding SEMs. The following cell lines were used in this experiment: rat glioma C6, human neuroblastoma SK, mouse neuroblastoma N2a, rat neuroblastoma B104, mouse fibroblast NIH 3T3 (3T3), human cervical carcinoma HeLa, normal rat kidney NRK, and mouse muscle myoblast line C_2C_{12}.

It is therefore likely that screening of more complex libraries containing the random oligonucleotides used here will identify additional short nucleotide sequences with IRES activity. Two important variables that could usefully be explored further in retroviral libraries are the length of the random oligonucleotide used and the spacing between two random oligonucleotides. In the case of the Gtx IRES module, for example, the length and sequence of the spacer were found to dramatically affect the activity of linked copies (12).

The question arises as to how internal initiation is promoted by short IRESes such as ICS1–23b, ICS2–17.2, and the previously identified Gtx IRES module. All three IRES modules contain a complementary match to different segments of 18S rRNA. A tenable hypothesis is that a direct interaction occurs between the IRES module and the 40S ribosomal subunit via base pairing to 18S rRNA (discussed in ref. 12). Additional studies may determine whether ICS1–23b and ICS2–17.2 recruit ribosomes by this mechanism. An alternative model is that one or more of the IRES modules recruit 40S ribosomal subunits by interacting with a protein component of the translational machinery. Such a protein might include a ribosomal protein, an initiation factor, or a novel bridging protein. The ability to initiate translation internally by binding to an initiation factor has been illustrated in a recent study in which an iron response element and the bacteriophage λ transcriptional antiterminator box B element were both demonstrated to function as IRESes in the presence of fusion proteins between the appropriate binding protein for these RNA elements and eIF4G (35). The lack of appreciable sequence similarities between the selected IRES modules, and between cellular IRESes in general (3), suggests that a wide variety of nucleotide sequences are likely to function in internal translation initiation and that different sequences may recruit preinitiation complexes by different mechanisms. It should be noted that the short sizes of the two IRES modules selected in this study argue against a role for secondary structure in recruitment of the translation machinery at these sequences. This conclusion is further supported by analyses of these two sequences with the MFOLD program, which did not reveal any stable secondary structures (data not shown).

The observation that synthetic IRESes comprising multimers of ICS1–23b, ICS2–17.2, or the 9-nt Gtx IRES module (12) show enhanced IRES activity as compared with the corresponding monomers suggests that multiple copies of the IRES module may increase the probability of recruiting 40S ribosomal subunits. A similar observation has been made for eIF4G tethered to the iron response element-binding protein, for which there was an approximately linear increase in translation when the number of iron response element binding sites was increased from one site to three (35).

An arresting feature of cellular IRESes, as well as of the two IRES modules identified in the present study, is their variable potency in different cell types (12, 36–39). Selection for IRESes in a variety of cell types is likely to identify additional elements with cell- and tissue-specific activities. If ribosomal recruitment requires direct interaction of IRESes with 18S rRNA, variations in efficiency may reflect differences in the accessibility of particular segments of 18S rRNA in different cell types. Alternatively, some IRES modules may require (or be blocked by) binding proteins that are differentially expressed in various cell types. To discriminate among these possibilities will require the determination of the proteins or components of the translation machinery that bind to particular IRES sequences in various differentiated cells.

Given the apparent modular nature of cellular IRESes, it will be of interest to investigate the properties of synthetic IRESes containing combinations of more than one type of module with different expression properties. It may be possible by such a combinatorial approach to design synthetic IRESes having either highly restricted or widespread translational activity. Most of the currently available multicistronic vectors use viral IRESes such as EMCV (e.g., refs. 40 and 41), and some of these do not function optimally in particular target tissues (39). Further applications of the selection method described here may lead to the identification of sequences with specific expression properties that may be useful in the design of vectors for research, biotechnology, and gene therapy.

We thank Megan Weber and Erin Powrie for excellent technical assistance and Drs. Bruce Cunningham, Joseph Gally, and Frederick Jones for critical reading of the manuscript. FACS was carried out at The Scripps Research Institute's flow cytometry core facility. This work was funded by the Neurosciences Support Corporation and by grants from the National Sciences Foundation (MCB9982574) to V.P.M., from the U.S. Public Health Service (NS39837) to G.M.E., and from the G. Harold and Leila Y. Mathers Charitable Foundation (to G.M.E.).

1. Pain, V. M. (1996) *Eur. J. Biochem.* **236,** 747–771.
2. Kozak, M. (1999) *Gene* **234,** 187–208.
3. Jackson, R. J. (2000) in *Translational Control of Gene Expression*, ed. Sonenberg, N., Hershey, J. W. B. & Mathews, M. B. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 127–183.
4. van der Velden, A. W. & Thomas, A. A. M. (1999) *Int. J. Biochem. Cell Biol.* **31,** 87–106.
5. Johannes, G., Carter, M. S., Eisen, M. B., Brown, P. O. & Sarnow, P. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 13118–13123.
6. Cornelis, S., Bruynooghe, Y., Denecker, G., Van Huffel, S., Tinton, S. & Beyaert, R. (2000) *Mol. Cell* **5,** 597–605.
7. Pyronnet, S., Pradayrol, L. & Sonenberg, N. (2000) *Mol. Cell* **5,** 607–616.
8. Stein, I., Itin, A., Einat, P., Skaliter, R., Grossman, Z. & Keshet, E. (1998) *Mol. Cell. Biol.* **18,** 3112–3119.

9. Holcik, M., Lefebvre, C., Yeh, C., Chow, T. & Korneluk, R. G. (1999) *Nat. Cell. Biol.* **1,** 190–192.
10. Stoneley, M., Chappell, S. A., Jopling, C. L., Dickens, M., MacFarlane, M. & Willis, A. E. (2000) *Mol. Cell. Biol.* **20,** 1162–1169.
11. Le, S.-Y. & Maizel, J. V., Jr. (1997) *Nucleic Acids Res.* **25,** 362–369.
12. Chappell, S. A., Edelman, G. M. & Mauro, V. P. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 1536–1541.
13. Ptashne, M. & Gann, A. (1997) *Nature (London)* **386,** 569–577.
14. Zenke, M., Grundstrom, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. & Chambon, P. (1986) *EMBO J.* **5,** 387–397.
15. Ondek, B., Shepard, A. & Herr, W. (1987) *EMBO J.* **6,** 1017–1025.
16. Mauro, V. P. & Edelman, G. M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 422–427.
17. Tranque, P., Hu, M. C.-Y., Edelman, G. M. & Mauro, V. P. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 12238–12243.
18. Hu, M. C.-Y., Tranque, P., Edelman, G. M. & Mauro, V. P. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 1339–1344.
19. Stoneley, M., Paulin, F. E. M., Le Quesne, J. P. C., Chappell, S. A. & Willis, A. E. (1998) *Oncogene* **16,** 423–428.
20. Yang, Q. & Sarnow, P. (1997) *Nucleic Acids Res.* **25,** 2800–2807.
21. Sella, O., Gerlitz, G., Le, S.-Y. & Elroy-Stein, O. (1999) *Mol. Cell. Biol.* **19,** 5429–5440.
22. Lauring, A. & Overbaugh, J. (2000) *Mol. Cell* **6,** 939–945.
23. Huez, I., Creancier, L., Audigier, S., Gensac, M.-C., Prats, A.-C. & Prats, H. (1998) *Mol. Cell. Biol.* **18,** 6178–6190.
24. Pelletier, J. & Sonenberg, N. (1988) *Nature (London)* **334,** 320–325.
25. Jang, S. K., Krausslich, H. G., Nicklin, M. J., Duke, G. M., Palmenberg, A. C. & Wimmer, E. (1988) *J. Virol.* **62,** 2636–2643.
26. Edelman, G. M., Meech, R., Owens, G. C. & Jones, F. S. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 3038–3043. (First Published March 21, 2000; 10.1073/pnas.040569897)
27. Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F. H., Verma, I. M. & Trono, D. (1996) *Science* **272** (5259), 263–267.
28. Bottenstein, J. E. & Sato, G. H. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 514–517.
29. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882.
30. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
31. Zuker, M., Mathews, D. H. & Turner, D. H. (1999) in *RNA Biochemistry and Biotechnology*, eds. Barciszewski, J. & Clark, B. F. C. (Kluwer, Boston), pp. 11–43.
32. Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999) *J. Mol. Biol.* **288,** 911–940.
33. Beavis, A. J. & Kalejta, R. F. (1999) *Cytometry* **37,** 68–73.
34. Anderson, M. T., Tjioe, I. M., Lorincz, M. C., Parks, D. R., Herzenberg, L. A. & Nolan, G. P. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 8508–8511.
35. De Gregorio, E., Preiss, T. & Hentze, M. W. (1999) *EMBO J.* **18,** 4865–4874.
36. Ye, X., Fong, P., Iizuka, N., Choate, D. & Cavener, D. R. (1997) *Mol. Cell. Biol.* **17,** 1714–1721.
37. Stoneley, M., Subkhankulova, T., Le Quesne, J. P., Coldwell, M. J., Jopling, C. L., Belsham, G. J. & Willis, A. E. (2000) *Nucleic Acids Res.* **28,** 687–694.
38. Oumard, A., Hennecke, M., Hauser, H. & Nourbakhsh, M. (2000) *Mol. Cell. Biol.* **20,** 2755–2759.
39. Creancier, L., Morello, D., Mercier, P. & Prats, A.-C. (2000) *J. Cell Biol.* **150,** 275–281.
40. Fan, L., Owen, J. S. & Dickson, G. (1999) *Atherosclerosis* **147,** 139–145.
41. Laufs, S., Kim, S. H., Kim, S., Blau, N. & Thony, B. (2000) *J. Gene Med.* **2,** 22–31.