

Cryptic transcripts from a ubiquitous plasmid origin of replication confound tests for *cis*-regulatory function

Nathan A. Lemp^{1,2}, Kei Hiraoka², Noriyuki Kasahara^{1,2} and Christopher R. Logg^{2,*}

¹Department of Molecular and Medical Pharmacology and ²Department of Medicine, University of California, Los Angeles, CA 90095, USA

Received February 28, 2012; Revised April 12, 2012; Accepted April 29, 2012

ABSTRACT

A vast amount of research on the regulation of gene expression has relied on plasmid reporter assays. In this study, we show that plasmids widely used for this purpose constitutively produce substantial amounts of RNA from a TATA-containing cryptic promoter within the origin of replication. Readthrough of these RNAs into the intended transcriptional unit potently stimulated reporter activity when the inserted test sequence contained a 3' splice site (ss). We show that two human sequences, originally reported to be internal ribosome entry sites and later to instead be promoters, mimic both types of element in dicistronic reporter assays by causing these cryptic readthrough transcripts to splice in patterns that allow efficient translation of the downstream cistron. Introduction of test sequences containing 3' ss into monocistronic luciferase reporter vectors widely used in the study of transcriptional regulation also created the false appearance of promoter function via the same mechanism. Across a large number of variants of these plasmids, we found a very highly significant correlation between reporter activity and levels of such spliced readthrough transcripts. Computational estimation of the frequency of cryptic 3' ss in genomic sequences suggests that misattribution of *cis*-regulatory function may be a common occurrence.

INTRODUCTION

Plasmid-based reporter assays were first used in animal cells nearly 30 years ago (1) and remain today the most common method of functional testing of candidate *cis*-regulatory sequences. These assays have perhaps found the most widespread application in the study of transcriptional control. Although newer approaches for

identifying regulatory elements, such as those based on chromatin immunoprecipitation, have yielded an abundance of information on protein–DNA interactions, plasmid reporter assays remain the favored method for assessing the functional activity in candidate regulatory sequences (2,3). The study of internal ribosome entry site (IRES) elements has also heavily depended on plasmid reporter assays, and the standard test for these elements is the dicistronic assay (4,5).

Although the integrity and structure of the transcripts generated in reporter assays for *cis*-regulatory function is critical for the validity of these tests, RNA structural analysis is only rarely performed. Mechanisms such as cryptic initiation or aberrant RNA processing may result in unanticipated species that produce anomalous reporter activity and fundamentally alter the outcome of these experiments. As an example, we and others have shown previously that splicing events involving 3' splice sites (ss) in test sequences can cause false positive scores for IRES function in reporter assays (6–8).

An overwhelming majority of reporter plasmids, and molecular cloning vectors in general, have backbone sequences originally derived from the prototype *Escherichia coli* cloning vehicle pBR322 (9) and contain the origin of replication (*ori*) of pMB1 (10) or, more commonly, a point mutant of this *ori* from the pUC series of plasmids (11,12). It has been recognized that plasmid backbone sequences can have some effect on eukaryotic expression cassettes contained within the same plasmid and may consequently influence the results of reporter assays. Accordingly, as a safeguard to prevent aberrant RNAs from reading into the intended transcriptional unit, many reporter plasmids now contain an upstream polyadenylation signal. Such signals are believed to reliably prevent the occurrence of readthrough transcripts that could mediate spurious reporter gene expression (2).

Here, we demonstrate that, despite such safeguards, a cryptic promoter in the pMB1 *ori* gives rise to transcripts that can readily create the artifactual appearance of

*To whom correspondence should be addressed. Tel: +1 310 825 7112; Fax: +1 310 825 5204; Email: logg@ucla.edu

transcriptional or translational regulatory function in reporter assays. Evaluating multiple widely used plasmids, including some that have been employed in thousands of published studies, we show that splicing between sites in the pMB1-driven transcripts and within inserted test sequences can cause robust expression of the downstream reporter gene whether or not the insert possesses genuine transcriptional or translational stimulatory activity. Additionally, we provide new evidence that the *c-myc* 5' untranslated region (UTR) itself harbors a cryptic core promoter responsible for its apparent IRES function in plasmid tests.

MATERIALS AND METHODS

Plasmids

Plasmid pRF and promoterless variants containing the eIF4G, XIAP and *c-myc* inserts have been described previously (13–15). Jian-Ting Zhang (Indiana University) provided the pRF plasmids containing the eIF4G sequence. The pRF plasmids containing the XIAP and *c-myc* sequences were provided by Greg Goodall and Andrew Bert (University of Adelaide). The pβgal/CAT plasmids (16,17) were provided by James Smiley and Holly Saffran (University of Alberta). To construct pRF variants containing a non-specific spacer or fragments of the rabbit β-globin (RBG) or adenovirus E2A genes harboring 3' ss, we annealed synthetic oligonucleotides and inserted them at the SpeI and NcoI sites of pRF containing or lacking the SV40 promoter (SV40P). The oligonucleotides were as follows: spacer, 5'-CTAGTTCGACTGGACACTGGATCTACTACATCGATTGCTGACGGC-3' and 5'-CATGGCCGTTTCAGCAATCGATGTAGTAGATCCAGTGTCCAGTCGAA-3'; β-globin, 5'-CTAGTGCTAACCATGTTTCATGCCTTCTTCTTTTCCTACAGGGCAC-3' and 5'-CATGGTGCCCTGTAGGAAAAGAAGAAGGCATGAACATGGTTAGCA-3'; E2A, 5'-CTAGTACTGACTCCATGATCTTTTCTGCCTATAGGACAC-3' and 5'-CATGGGTGCTCCATAGGCAGAAAAGATCATGGAGTCAGTA-3'. To generate pRF containing no insert, we removed the insert from pR-eIF4G-F with SpeI and NcoI, blunted the ends with T4 DNA polymerase and ligated the resulting fragment. The pRF plasmids containing the eIF4G and XIAP sequences with mutated polypyrimidine tracts (PPTs) were constructed by PCR-based mutagenesis. The five mutations introduced were selected based upon their being the most infrequent bases in the PPTs of naturally occurring 3' ss (18).

pGL3-Control, pGL3-Promoter, pGL3-Enhancer, pGL3-Basic, pGL4.17 and pGL4.10 were from Promega. pGL3-Enhancer-RBG, a variant of pGL3-Enhancer containing the β-globin 3' ss, was derived by cutting promoterless pRF containing the 3' ss with SpeI and NotI, blunting the ends with T4 DNA polymerase and recircularizing the plasmid by ligation. The MluI–NcoI fragment of pGL3-Enhancer-RBG including the β-globin 3' ss was excised and inserted into the same sites in pGL3-Basic to generate pGL3-Basic-RBG. The SacI–NcoI fragment of pGL3-Enhancer containing the 3' ss was

transferred to pGL4.17 and pGL4.10 to generate the β-globin 3' ss-containing variants of these plasmids.

Cell culture and transfection

HeLa cells were used throughout and were grown at 37°C under 5% CO₂ in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 100 units/ml penicillin and 100 μg/ml streptomycin. Transfections were performed with Fugene 6 (Roche) according to the manufacturer's recommendations. For reporter assays, 700 ng of luciferase plasmid was cotransfected with 300 ng of green fluorescent protein (GFP) expression plasmid pGFPemv-cmv[R]-control (Packard Biosciences) and transfection efficiency was assessed by flow cytometric analysis of cells with a Beckman Coulter EPICS XL cytometer to determine the percentage of GFP-positive cells. For RNA isolation, 1 μg of each luciferase plasmid alone was used.

Luciferase assays

For dual measurement of both RLuc and FLuc activities in transfected cells, the Dual-Luciferase Reporter Assay System (Promega) was used and cell lysates were prepared with Passive Lysis Buffer. For measurement of FLuc activities alone, the Luciferase Assay System with Reporter Lysis Buffer was used. Luciferase activities in 20-μl samples of each lysate were determined according to the manufacturer's standard protocol with a Sirius luminometer (Berthold). For relative light unit (RLU) calculations, background autoluminescence values, corresponding to luminescence measured with lysates of mock-transfected cells, were subtracted from each measured value.

RNA isolation and reverse transcription for RT-PCR

The RNeasy Mini kit with optional DNase I treatment (Qiagen) was used for isolation of RNA from transfected cells. QiaShredder columns were used to homogenize the cells prior to RNA purification. Two micrograms of each RNA was reverse transcribed with Superscript III reverse transcriptase and oligo(dT) primer (Invitrogen). The resulting cDNA was treated with RNase H (Invitrogen).

5' rapid amplification of cDNA ends

5' RACE was performed essentially according to the method of Matz *et al.* (19) on RNA isolated as described above. First-strand cDNA was synthesized with Superscript II reverse transcriptase (Invitrogen) and primer 5' TTTTTTT TTTTTTTTTTTTTTTTTT VN-3'. The oligo 5' CAGATGGACGACTTGCGATAGACACGGG-3' was included in the reaction to incorporate a primer-binding site at the 5' end of the cDNA through the template-switching effect. The cDNA was then treated with RNase H and used as template in PCR with Phusion polymerase (Finnzymes) and forward primers 5' ATAGAGCAGTAGTGACTCCGAACAGATGGACGACTTGCGATAGACA-3' (at 0.06 μM) and 5' ATAGAGCATAGTGACTCCGAA-3' (at 0.4 μM). The reverse primer used (at 0.4 μM) depended on the source of the RNA being analysed. For pRF and pGL3 RNA, the *luc*⁺-specific reverse primer 5' GGCCTTCTTTATGTTTTGGCGTCTT-3'

was used. For p β gal/CAT RNA, CAT-specific reverse primer 5' ACGGTCTGGTTATAGGTACATTGAGCAA-3' was used. For pGL4 RNA, *luc2*-specific reverse primer 5' GGTCCCCTCTTCGAGTGGGTAGAAT-3' was used. The resulting products were directly cloned with the Zero Blunt PCR Cloning Kit after cleanup with a PureLink PCR purification kit (both Invitrogen) and sequenced.

Standard RT-PCR

Amplification of reverse transcribed RNA was carried out with Phusion polymerase and primers from Invitrogen. The primer sequences are shown in Supplementary Figure S1. The cycling conditions for amplification of cryptic readthrough transcripts from pGL3 and p β gal/CAT plasmids were 32 cycles of the following: 98°C for 7s, 53°C for 20s and 72°C for 12s, followed by 5 min at 72°C. The conditions for amplification of the intended dicistronic transcripts from p β gal/CAT plasmids were 30 cycles of the following: 98°C for 7s, 56°C for 20s and 72°C for 90s, followed by 5 min at 72°C. The reaction products were run on an agarose gel and the predominant bands were excised, column purified, cloned with the Zero Blunt PCR Cloning Kit and sequenced.

Quantitative RT-PCR

Real-time RT-PCR was performed with primers from Invitrogen, probes and TaqMan Gene Expression Master Mix from Applied Biosystems and a MyiQ2 Real-Time Detection System from Bio-Rad. The primer and probe sequences are shown in Supplementary Figure S1. The cycling conditions were 10 min at 95°C followed by 40 cycles of 15s at 95°C and 1 min at 60°C. β -Actin was used as an endogenous reference to normalize relative levels of the RNA targets (Applied Biosystems human ACTB endogenous control). Standard curves for each primer-probe set were generated for determination of amplification efficiencies. The reaction efficiencies and threshold cycle values obtained were used to calculate relative levels of each target sequence with the Microsoft Excel-based Q-Gene application (20). All values shown were normalized to set the level of spliced readthrough cryptic RNA from pRF containing the globin insert at an arbitrary value of 10 000.

Statistics

Simple linear regression and correlation analysis were performed on log-transformed means of RNA and FLuc expression levels using Prism 5 software (GraphPad).

Determination of incidence of replication origins in GenBank

To determine the number of GenBank sequences containing the ColE1, p15a, pMB1/pUC or pSC101 replication origins, homology searches of the nr/nt nucleotide collection were performed using the Megablast algorithm and the following query sequences: ColE1, CGGATTAGCAGAGCGATGATGGCACAACGGTGCTACAGAGTTCTTGAAGTAGTGGCCCG; p15a, CCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCC

GGTAAAGGCTAAACT; pMB1/pUC, CAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAA; and pSC101, ACAGA GGGTCTAGCAGAATTTACAAGTTTTCCAGCAAA GGTCTAGCAGAATTTACAGATA.

Estimation of incidence of cryptic 3' ss in genomic sequence

To generate sequence to serve as a model of the upstream region of human genes, we used the random-seq program of regulatory sequence analysis tools (RSAT) (21), with the default Markov chain order value of five. 300 kb of random sequence was screened for potential 3' ss with the programs NetGene2 (22,23) (<http://www.cbs.dtu.dk/services/NetGene2>), NNSplice (24) (http://www.fruitfly.org/seq_tools/splice.html), MaxEntScan (25) and Human Splicing Finder (HSF) (26). HSF and MaxEntScan analyses were both performed at <http://www.umd.be/HSF>. The scores for the eIF4G, XIAP, globin and E2A 3' ss were obtained with the pRF plasmids containing the corresponding inserts. The scores for the cryptic 3' ss identified by RT-PCR in the pGL3 plasmids were obtained with the corresponding complete plasmid sequences. The estimated average score for aberrant splice sites in DBASS3 (27) (<http://www.dbass.org.uk/dbass3>) was obtained by randomly selecting one of every five 3' ss in the database, determining their scores in the context of their native pre-mRNA, and averaging.

RESULTS

Test sequences containing 3' ss mimic both IRES elements and promoters in pRF

Sequences of the human eIF4G and XIAP genes have been reported to possess IRES activity (16,28), based on their ability to stimulate expression of a downstream gene in dicistronic reporter assays. Subsequent studies questioned these conclusions and attributed the observed activities to cryptic promoters within the two sequences (13,14), based primarily on the results of reporter assays with promoterless variants of the widely used plasmid pRF (Supplementary Figure S2A) (15), a dicistronic vector derived from the popular pGL3 luciferase reporter plasmids. We noted that, interestingly, a short region within the eIF4G sequence found to be essential for its apparent cryptic promoter function precisely overlaps a 3' ss the sequence contains (Supplementary Figure S2B). Similarly, the putative IRES of the XIAP gene has also been reported to possess 3' ss activity (8), and whether this activity is inherent to the XIAP sequence or artifactual has been a subject of debate (17,29). We subsequently showed that both the eIF4G and XIAP sequences contain 3' ss (6) that are active in a retroviral replication assay and are utilized in the pre-mRNAs of the eIF4G and XIAP genes themselves.

Here, we further examined the eIF4G and XIAP elements in the context of pRF with or without its promoter. As in previous studies with these constructs (13,14), the promoterless variants lacked both the SV40

promoter and the chimeric intron. In the same manner, we also tested the *c-myc* 5' UTR, which similarly was reported to be an IRES (15,30) and later to be a promoter (13). Furthermore, to determine if other sequences containing 3' ss could produce the appearance of IRES and promoter function in assays with these plasmids, we introduced short 3' ss-containing segments from the β -globin and adenovirus E2A genes as test sequences.

As previously reported, the eIF4G, *c-myc* and XIAP inserts all stimulated robust expression of FLuc from pRF (Figure 1). Interestingly, the 30-ss-containing globin and E2A fragments also stimulated FLuc expression with comparable potency. As expected, deletion of the promoter and chimeric intron vastly reduced expression of the upstream RLuc cistron from all constructs. By contrast, this deletion had almost no effect on FLuc expression from any of the plasmids. These findings suggested that 3' ss can mimic both IRES and promoter elements in this reporter system, and in fact, from this type of assay alone, that it is not possible to distinguish whether a given element has splicing, IRES, or promoter activity. Furthermore, the observed persistence of FLuc expression despite deletion of the SV40P indicated that the test inserts that stimulated second-cistron expression either contain cryptic promoter function themselves or were functioning as 3' ss within cryptic transcripts initiating at unknown upstream locations within the plasmid.

To assess the requirement of the 3' ss of the eIF4G and XIAP sequences for their putative IRES and promoter

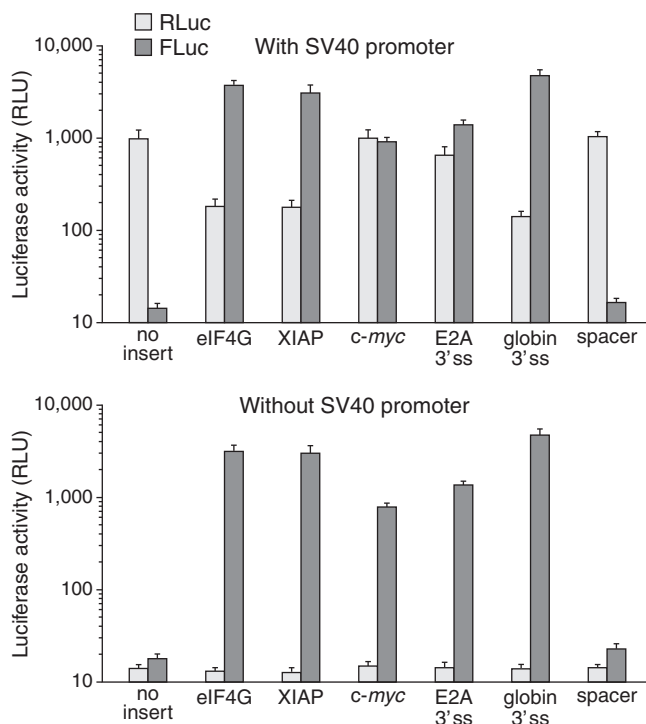


Figure 1. Second-cistron expression from dicistronic reporter plasmid pRF is potentially stimulated by insertion of 3' ss-containing test sequences or putative IRES/promoter elements, but is unaffected by deletion of the SV40 promoter and chimeric intron. RLuc and FLuc activities were measured from cells transfected with the indicated constructs. Shown are mean values \pm SD ($n = 3$).

function, we tested variants in which the PPTs of the 3' ss were mutated to contain several purines (Supplementary Figure S3). The mutant forms stimulated FLuc expression from both promoter-containing and promoterless pRF to levels \sim 2–4% of those of the wild-type sequences, suggesting that splicing at these sites is indeed required for both apparent IRES and promoter activity.

Apparent IRES and promoter function in the eIF4G and XIAP sequences is mediated by cryptic readthrough transcripts from the plasmid *ori*

To identify the initiation sites of the transcripts responsible for FLuc expression from the promoterless eIF4G, XIAP and *c-myc* constructs, we performed 5' RACE, sequencing 22–23 clones per construct. Remarkably, all but one of the transcripts from the eIF4G and XIAP plasmids had initiated from within the plasmid *ori* (Figure 2A and B), and none from within the test insert itself. In each transcript, the previously identified 3' ss within the insert had spliced with a cryptic 5' ss encoded by the plasmid backbone, thereby positioning the FLuc cistron near to the 5' end. These cryptic transcripts had read into the reporter gene unit despite the presence of an upstream transcription termination element. About half were free of AUGs upstream of the FLuc cistron that might impair translation by ribosomes scanning from the m⁷G cap. The plasmid sequence 5' to the major transcription initiation site in the *ori* harbors striking and fortuitous homologies to TATA and GC boxes (Figure 2C) at locations relative to the initiation site that are typical for these elements in eukaryotic promoters. Also surprisingly, one cloned transcript from the eIF4G plasmid had been formed by *trans*-splicing between the eIF4G 3' ss and a 5' ss of an endogenous cellular RNA (Supplementary Figure S4).

All of the cloned transcripts from the *c-myc* plasmid, however, had initiated within the *c-myc* sequence itself, and none had undergone splicing. A number of sites scattered throughout the insert had been used for initiation, but about half were clustered at a location exhibiting sequence hallmarks of a core promoter, including initiator (Inr) and downstream promoter element (DPE) homologies (32) (Figure 2D), indicating that the *c-myc* sequence indeed harbors promoter activity. An earlier study of putative IRES function in the sequence identified a 50-nt segment sufficient to drive second-cistron expression in plasmid assays (33). This short region entirely encompasses the identified cluster of transcription initiation events. Furthermore, detailed dissection of the 50-nt segment in the previous study revealed two 14-nt sub-sequences critical for apparent IRES function, and these closely coincide with the aforementioned Inr and DPE homologies. Considering the additional previous finding that siRNA targeted to the RLuc cistron strongly knocked down RLuc but not FLuc expression from pRF containing the *c-myc* insert (13,34), a compelling case emerges that the mechanism by which this element scores positive for IRES function in tests with this plasmid is primarily, if not exclusively, cryptic transcription.

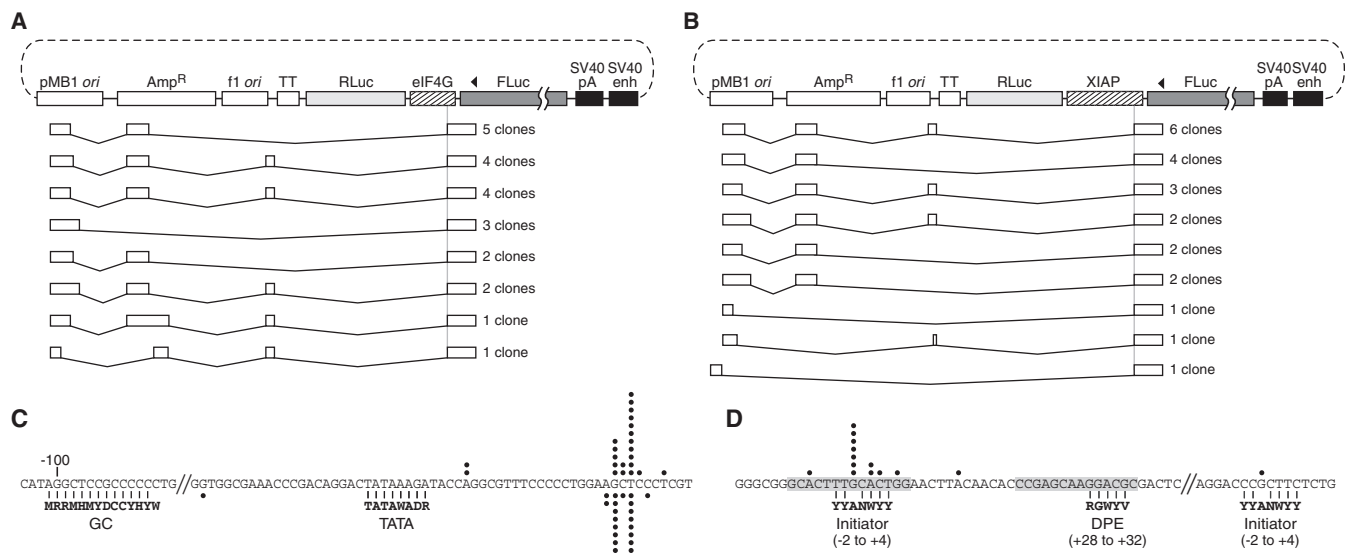


Figure 2. Mapping of the initiation sites of FLuc-encoding RNA produced from promoterless pRF constructs containing the eIF4G, XIAP and *c-myc* sequences. (**A** and **B**) 5' RACE products from promoterless pR-eIF4G-F (**A**) and pR-XIAP-F (**B**). The cloned transcripts are grouped by splicing pattern and aligned to maps of the corresponding plasmids. Amp^R, ampicillin resistance gene; f1 ori, bacteriophage f1 origin; TT, transcriptional terminator. Arrowheads denote the location of the reverse primer used. (**C**) Sequence detail of the proximal cryptic promoter in the pMB1 ori. Black circles above and below the sequence represent the start site of cloned transcripts from promoterless pR-eIF4G-F and pR-XIAP-F, respectively. Homologies to consensus vertebrate GC and TATA boxes (obtained from the Eukaryotic Promoter Database) (31) are indicated. (**D**) Detail of the region of the major and one minor transcription start site identified in the *c-myc* sequence. The region shown is the *c-myc* 5' UTR between -194 and -74 relative to the AUG start codon. Black circles indicate the initiation sites of cloned transcripts. Inr and DPE consensus sequences and their canonical locations relative to transcription start sites (32) are shown below homologous *c-myc* sequence. Sequences highlighted in gray are 14-nt elements previously reported (33) to mediate apparent IRES function in reporter assays.

To determine the relative abundance of the cryptic and expected transcripts produced by pRF constructs, we performed quantitative RT-PCR using splice isoform-specific oligonucleotides. All of the tested plasmids produced high levels of RNA from the cryptic promoter in the *ori*, regardless of the presence or sequence of the intercistronic insert (Figure 3). The presence of a 3' ss-containing insert, however, increased the total levels of cryptic transcript by 2- to 3-fold, and resulted in levels of spliced readthrough RNA that were in most cases higher than that of the expected dicistronic transcript generated from the SV40P. Moreover, the levels of cryptic spliced readthrough transcripts were little affected by the presence or absence of the SV40P. Collectively, our results indicate that these unappreciated readthrough RNAs account for most, if not all, of the FLuc expression stimulated by the eIF4G and XIAP sequences from both promoterless and promoter-containing pRF.

Interestingly, deletion of the SV40P alone from pRF, while leaving the chimeric intron intact, was reported previously to not diminish the expression of RLuc (35). Based on our findings, we speculate that RLuc expression from that variant of the plasmid may result from splicing of the 3' ss of the chimeric intron with 5' ss of the cryptic transcripts from the *ori*.

Cryptic readthrough transcripts explain apparent IRES and promoter function in the XIAP sequence when tested in a second reporter system

As noted, the mechanism by which the XIAP sequence stimulates expression of a downstream reporter gene has

been the subject of ongoing controversy (6,8,17,29,36,37). While the 3' ss it contains has been found to cause aberrant splicing of the intended transcripts of pRF and other multicistronic plasmids (6,8,34), this was reported not to be the case with the RNA produced from pβgal/CAT (Supplementary Figure S5A), the vector originally used to identify this putative IRES (16). A more recent study, however, found that the XIAP element does trigger aberrant splicing of the dicistronic RNA of pβgal/CAT, but its ability to stimulate CAT expression was ultimately attributed to cryptic promoter activity within the element (17). This conclusion was based on the findings that siRNA targeted to β-gal sequence retained in the mis-spliced dicistronic RNAs knocked down β-gal but not CAT expression, and that the XIAP sequence stimulated CAT expression even when the plasmid's native cytomegalovirus (CMV) promoter was removed. Northern blot analyses revealed cryptic transcripts of 1.5–2.0 kb that were produced independently of the CMV promoter and which were concluded to initiate from the XIAP insert and mediate its apparent IRES activity. Consistent with results of that study, our RT-PCR amplification of the entire region of the expected dicistronic pβgal/XIAP/CAT RNA upstream of CAT yielded only spliced forms, none of which appeared to represent templates suitable for efficient CAT translation (Supplementary Figure S5B). However, our 5' RACE analysis of the CAT-encoding transcripts produced by the promoterless form of the plasmid revealed none that had initiated within or near the XIAP insert, but primarily at the same site in the pMB1 *ori* identified in the pRF plasmids (Figure 4A). As in pRF, the XIAP 3' ss had

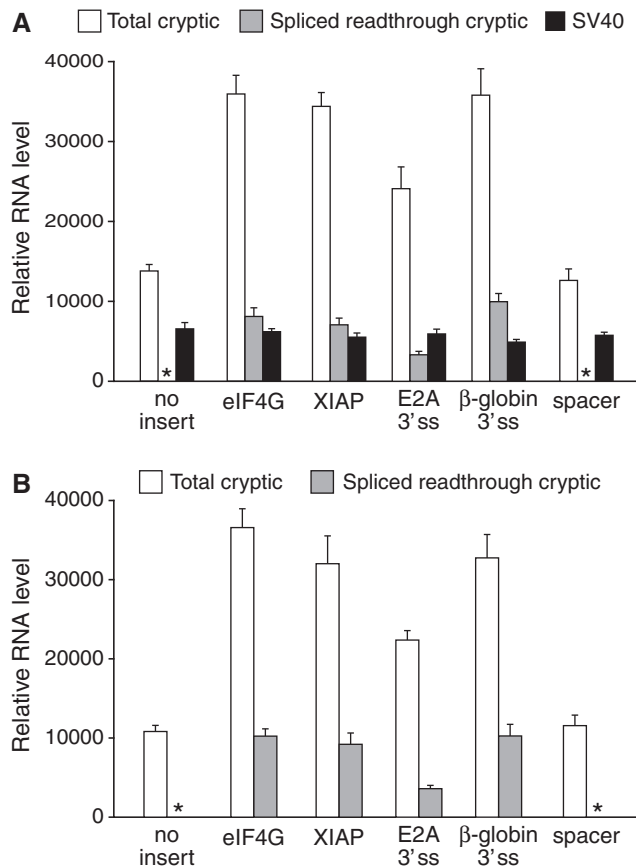


Figure 3. The cryptic promoter in the pMB1 *ori* within pRF is robust and constitutively active. Shown are levels of cryptic and expected RNA produced from pRF constructs containing (A) or lacking (B) the SV40P. RNA from transfected cells was analysed by quantitative RT-PCR to detect total cryptic RNA, spliced readthrough cryptic RNA, or the expected RNA from the SV40P. Asterisks indicate the absence of detectable transcript. Error bars, SD.

triggered splicing events in these transcripts that positioned the downstream reporter cistron close to the 5' cap. Each of these RNAs was completely free of upstream AUGs and were produced whether or not the CMV promoter was present (Figure 4B and C). Additionally, these RNAs are of approximately the same size as the cryptic transcripts identified previously by northern blotting, suggesting that they represent the same species. Together these results strongly indicate that spliced readthrough transcripts account for both apparent IRES and promoter function in the XIAP sequence when tested in the p β gal/CAT system.

Splicing of cryptic readthrough transcripts can mimic promoter function in the pGL3 system

To determine if transcripts from the pMB1 *ori* could influence the outcome of reporter assays with more conventional, monocistronic vectors, we introduced the globin test sequence into the multiple cloning site (MCS) of the promoterless plasmids of the popular pGL3 system (Supplementary Figure S6A). 5' RACE analysis of the FLuc-encoding RNA produced by pGL3-Enhancer containing this insert revealed an array of cryptic readthrough transcripts similar to those produced from promoterless

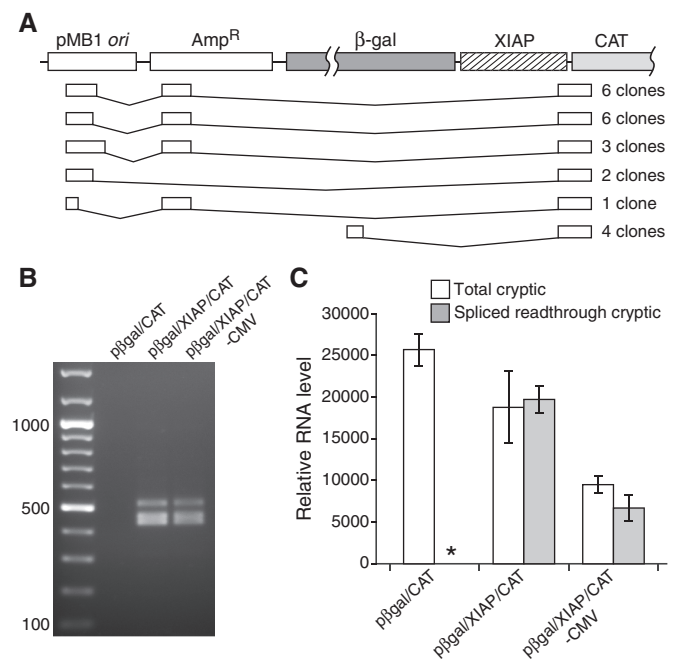


Figure 4. Production of CAT-encoding monocistronic transcripts from the pMB1 cryptic promoter within the p β gal/CAT reporter system. (A) Structure of transcripts identified by 5' RACE analysis of RNA from promoterless p β gal/XIAP/CAT. (B) Amplification by standard RT-PCR of cryptic CAT-encoding transcripts from p β gal/CAT plasmids with or without the XIAP insert or CMV promoter. The forward primer was specific for the first exon and the reverse primer was specific for the CAT gene. Sequencing showed that the three major amplified species correspond to the three most prevalent splice isoforms identified by 5' RACE. (C) Quantitation of cryptic transcripts by real-time RT-PCR. The asterisk indicates the absence of detectable transcript. Error bars, SD.

pRF and p β gal/CAT plasmids harboring the eIF4G and XIAP inserts (Figure 5A).

Insertion of the globin sequence into pGL3-Basic and pGL3-Enhancer increased FLuc activity by 10- and 13-fold, respectively (Figure 5B). While the insert-containing variants of pGL3-Enhancer and promoterless pRF generated roughly similar levels of FLuc activity (12 000 versus 4700 RLU), as might be expected from their structural similarity and the assortment of spliced readthrough transcripts they both produce, the fold-stimulation by the 5' ss was much lower for pGL3-Enhancer. This was due primarily to a nearly 50-fold higher baseline FLuc expression from this plasmid. To determine if this high background might at least in part result from constitutive production of readthrough transcripts spliced at unknown cryptic 3' ss, we performed RT-PCR analysis on cells transfected with the four parental plasmids of the pGL3 system. Such transcripts were indeed produced by all four plasmids, and in each of the major amplified isoforms, a cryptic 3' ss upstream of the FLuc gene had been activated (Figure 5C and D). In pGL3-Basic and pGL3-Enhancer, this cryptic 3' ss was within the MCS and in pGL3-Promoter and pGL3-Control it was within the SV40P. All isoforms lacked out-of-frame upstream AUGs that might impair FLuc translation. Transcription from the pMB1 *ori* in pGL3 plasmids

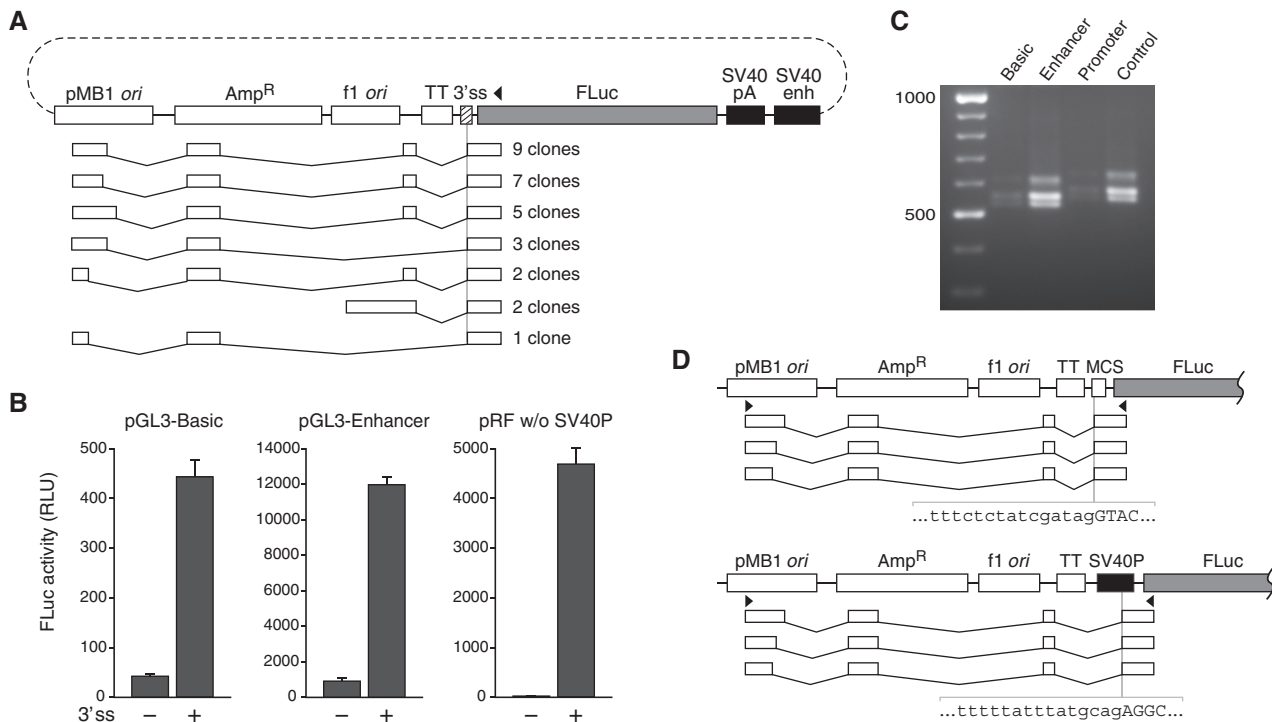


Figure 5. Production of luciferase-encoding cryptic readthrough RNA by the pGL3 system. (A) The presence of a 3' ss in a test sequence introduced into pGL3-Enhancer results in production of spliced readthrough transcripts similar to those observed with pRF. Shown are the 5' RACE products from pGL3-Enhancer containing the globin 30-ss site as the test sequence. (B) Insertion of the globin sequence into pGL3-Basic or pGL3-Enhancer stimulates luciferase activity strongly, despite substantial preexisting background. Results of luciferase assays with promoterless pRF from Figure 1 are shown here in linear scale for comparison. (C and D) Cryptic 3' ss upstream of the FLuc gene in the unmodified parental plasmids of the pGL3 system are activated in readthrough transcripts. (C) Results of RT-PCR reactions to detect RNA from the pMB1 *ori* in which a cryptic 3' ss upstream of the FLuc gene had been utilized. (D) Structure of the RT-PCR products as determined by sequencing. Top, transcripts from pGL3-Basic and pGL3-Enhancer; bottom, transcripts from pGL3-Promoter and pGL3-Control. Arrowheads indicate the location of the primers used. The sequences of the activated cryptic 3' ss are shown.

can thus result in activation of cryptic 3' ss proximal to the FLuc gene to generate FLuc-encoding isoforms.

To more thoroughly assess the relationship between levels of spliced readthrough transcripts and FLuc activity, we performed regression analysis on these two variables across the full set of pRF and pGL3 constructs tested. A very highly significant correlation between these variables (Pearson's correlation coefficient, $r = 0.9760$, $P < 0.0001$) was observed when pGL3-Promoter and pGL3-Control were excluded (Figure 6). Not surprisingly, the two latter plasmids expressed substantially more FLuc than would be predicted from their production of the readthrough RNA alone, as both contain the SV40P directly upstream of the FLuc gene. Among the other plasmids, however, spliced readthrough RNA levels could predict FLuc activity with striking consistency.

Splicing of readthrough transcripts can mimic promoter function in the newer pGL4 system

A reporter plasmid platform designated pGL4 was introduced in 2004 as a replacement for the pGL3 system. This newer plasmid series included modifications such as removal of potential cryptic transcription factor binding sites from within the backbone and reporter gene to reduce aberrant transcription, and codon optimization to increase reporter activity. pGL4 plasmids also contain the pMB1 *ori*,

and this sequence was left unaltered. We therefore sought to determine if similar cryptic transcripts might also generate FLuc activity with this system. For this purpose, we selected the promoterless pGL4.10 and pGL4.17 (Supplementary Figure S6B), due to their being the most structurally comparable to pGL3-Basic and -Enhancer, respectively. Introduction of the globin test sequence into the MCS resulted in a 7-fold increase in FLuc activity from pGL4.10 and a 45-fold increase from pGL4.17 (Figure 7A). 5' RACE analysis of the FLuc-encoding transcripts from pGL4.17 containing the test sequence revealed that, of 41 transcripts cloned, 30 had initiated from the cryptic promoter in the *ori* (Figure 7B), and two had initiated from the SV40P in the plasmid backbone. The initiation site for the remaining nine could not be determined as they, remarkably, had undergone *trans*-splicing with nine different cellular RNAs (Supplementary Figure S7). Thus, pGL4 plasmids appear to be as susceptible as pGL3 plasmids to producing artifactual reporter activities due to transcripts from the pMB1 cryptic promoter.

The pMB1 *ori* and cryptic promoter is ubiquitous among cloning vectors

Because the results of this study were obtained with only a subset of available reporter plasmids, we sought to assess how common the occurrence of cryptic readthrough

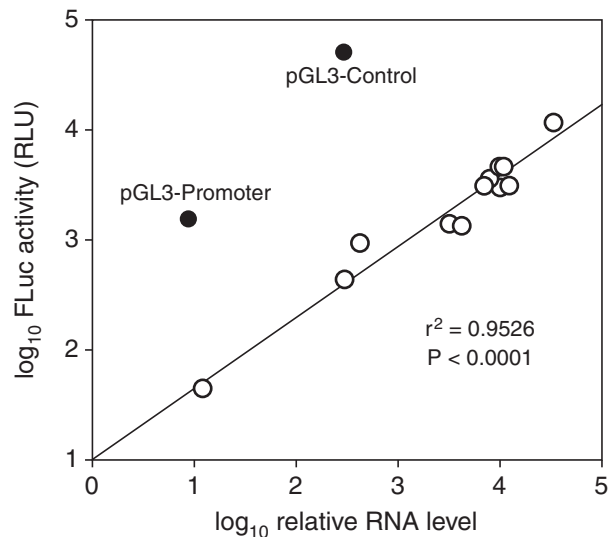


Figure 6. Correlation between levels of spliced readthrough RNA and luciferase expression among the pRF and promoterless pGL3 plasmids tested in this study. Open circles represent the promoterless and promoter-containing pRF plasmids, pGL3-Basic, pGL3-Enhancer and the variants of these plasmids with 3' ss-containing test inserts. The pRF plasmids containing either no insert or the spacer are not shown since they produced no detectable spliced readthrough RNA. The line represents the linear regression for the values obtained with these plasmids. The values for pGL3-Promoter and pGL3-Control, which were excluded from the regression analysis, are shown for comparison. Supplementary Table S1 provides the identities and values for all plasmids represented.

transcripts is likely to be among cloning vehicles generally. We therefore performed homology searches of GenBank to estimate the relative frequency of those with pMB1 or pUC origins versus those with other replicons, including those of ColE1, p15a and pSC101, which are also present in a significant proportion of vectors. Of the GenBank entries matching one of these origins, 95% were to pMB1 or pUC, demonstrating the overwhelming predominance of this element (Supplementary Figure S8A). Additionally, the ColE1 and p15a sequences are both highly homologous to the pMB1/pUC origin, particularly in the region surrounding and upstream of the transcription start site, and thus may also exhibit promoter function in eukaryotic cells (Supplementary Figure S8B). In order for the promoter to stimulate expression of a gene in the same plasmid, it presumably must be in the same orientation. Many commercially available reporter plasmids have this characteristic, and include several in addition to pGL3 and pGL4 that have been in wide use for many years (Supplementary Table S2). Queries of the HighWire Press search engine (<http://highwire.stanford.edu>) using the terms 'luciferase' and either 'pGL3' or 'pGL4' yielded 10872 and 917 studies, respectively, referencing these two systems alone.

Cryptic 3' ss are predicted to occur with high frequency in candidate regulatory sequences

Unlike the transcribed portions of genes, which are subject to selection pressure against cryptic ss that could cause

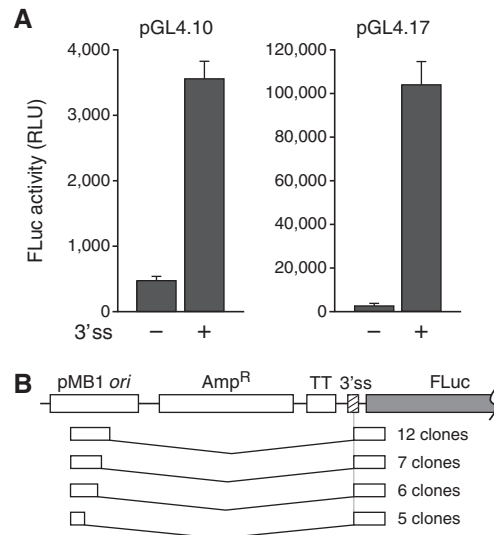


Figure 7. A 3' ss can mimic a promoter in the pGL4 system by inducing production of spliced readthrough transcripts. (A) Expression of luciferase from pGL4.10 and pGL4.17 with and without the globin 3' ss inserted into the MCS. Error bars, SD. (B) 5' RACE products from pGL4.17 containing the globin insert.

aberrant splicing, the upstream region where promoters and other regulatory elements are located are presumably free from such selection. To estimate the chance frequency of cryptic 3' ss in these regions of human genes, we generated random sequences modeled on these areas (21) and screened them with four ss prediction programs for potential 3' ss. Predicted 3' ss scoring as high or higher than the average of the four authentic 3' ss tested in this study (eIF4G, XIAP, E2A and β -globin) occurred 0.25–1.0 times per kb (Figure 8). Predicted 3' ss scoring as high or higher than those we identified in the parental pGL3 plasmids (Figure 5D) occurred approximately 1–10 times per kb, and those scoring as high or higher than that of the estimated average 3' ss in the Database of Aberrant 3' Splice Sites (DBASS3) (27) occurred 3–9 times per kb.

DISCUSSION

Controversy has surrounded the reliability of plasmid tests for IRES function (4,8,14,17,29,38,39). Here we show that assays for *cis*-regulatory function in general are prone to producing artifactual results when plasmids in widespread use for this purpose are employed. We demonstrate that candidate regulatory sequences can strongly stimulate reporter gene expression from both dicistronic and monocistronic vectors by altering the splicing of cryptic RNAs constitutively produced from the plasmid *ori*. For test sequences lacking genuine transcriptional or translational regulatory activity, the level of stimulation primarily reflects the level of altered splicing that is triggered. These splicing events, which can induce reporter expression by up to hundreds-fold, can mimic promoter, enhancer or IRES function in the test sequence, depending on the RNA species the plasmid used is thought to produce.

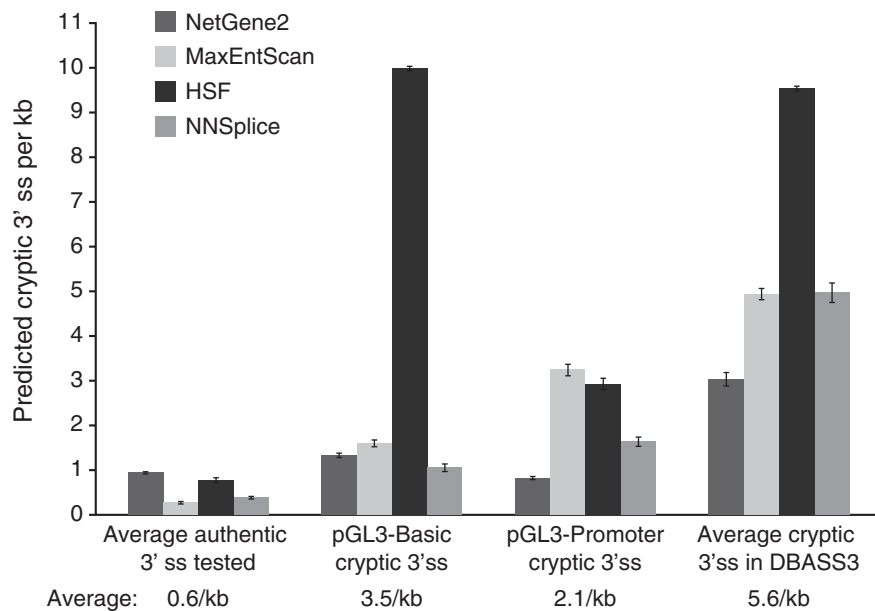


Figure 8. Estimation of the frequency of randomly occurring cryptic 3' ss in the untranscribed upstream region of genes in the human genome. Three 100-kb random sequences, generated as detailed in Materials and Methods, were analysed by the splice site prediction programs NetGene2, MaxEntScan, HSF and NNSplice. Shown are the average number of predicted 3' ss per kilobase having scores at least as high as that of the indicated 3' ss. The 'average authentic 3' ss tested' is the average score of the eIF4G, XIAP, E2A and globin 3' ss. The pGL3-Control and pGL3-Enhancer cryptic 3' ss are those identified in this study and shown in figure 5D. The 'average cryptic 3' ss in DBASS3' is the estimated average score of all 3' ss in the Database of Aberrant 3' ss (27). Error bars, SD.

Although the identified cryptic promoter appears to be transactivatable by enhancer sequences present elsewhere in the plasmid, it remains active even in the absence of such elements. Plasmids pGL3-Basic and pGL4.10, which contain no known eukaryotic enhancer elements, all produced the cryptic RNA, and introduction of 3' ss-containing inserts into both of these vectors potentially stimulated second-cistron expression via these transcripts. In a previous study, we found that introduction of the eIF4G or XIAP test sequences into a pGL4-based dicistronic vector resulted in very little detectable stimulation of second-cistron expression. The likely reason for this is that the second-cistron reporter in these plasmids was GFP, which is a far less sensitive cellular reporter than FLuc.

Earlier studies have noted aberrant reporter gene expression caused by backbone sequences of other plasmids (40–42), although this problem has been viewed as effectively solved by the introduction of termination elements upstream of the intended transcriptional unit (2). Most commercially available reporter plasmids now contain such signals. The upstream terminator in the pGL3, pGL4 and pRF plasmids is a bipartite element comprising a synthetic polyadenylation (SPA) signal and transcriptional pause sequence. The SPA was shown in plasmid assays to terminate transcription with high efficiency when located in an exon (43). When placed within an intron, however, the signal was rendered inactive. Addition of the pause site downstream of the SPA reactivated polyadenylation activity, but only to a fraction of that seen when the SPA was within an exon (44). These findings may thus explain why this otherwise

potent signal cannot prevent the production of spliced readthrough transcripts: the presence of a 3' ss downstream of the termination element creates an intron wherein the element is rendered weak or inactive. A 3' ss in a test insert may therefore stimulate reporter expression both by causing upstream sequences to be spliced out and by facilitating readthrough.

The ability of the eIF4G, XIAP and other putative IRES elements to stimulate expression of the downstream gene in dicistronic assays has been previously attributed, fully or in part, to mis-splicing of the intended dicistronic RNA caused by the 3' ss these sequences harbor (6,7,45). Our results, however, indicate that most or all of this ability is attributable to splicing of the cryptic transcripts from the *ori*, as the presence or absence of the SV40 promoter in pRF had almost no effect on either FLuc expression or levels of spliced readthrough RNA. Moreover, in the case of pβgal/XIAP/CAT, siRNAs targeted to first-cistron sequences retained in the aberrantly spliced dicistronic RNA were found to be incapable of reducing CAT expression (17), suggesting that second-cistron expression from this plasmid as well occurs primarily via the spliced readthrough RNA, which is produced whether or not the CMV promoter is present.

A large number of other sequences originally suspected to be IRES elements have also proven to stimulate second-cistron expression from dicistronic plasmids whether or not the dicistronic unit's promoter was present. In each of the 28 cases we found in a search of the literature, pRF was the plasmid employed and the reporter assay results were interpreted as demonstrating cryptic promoter activity in the test sequence (13,14,46–60). For only two of these

elements (49,56), however, was any attempt made to map the transcription initiation site. In both instances, the method used was a nuclease protection assay with a probe complementary to the insert. This approach would be unable to identify initiation sites outside of the probed region. However, if a 3' ss within the insert had undergone splicing, the analysis would suggest initiation at that location. Perhaps significantly, the putative initiation sites mapped for one of these two sequences, the MYEOV 5' UTR (49), lie adjacent to a predicted 3' ss this sequence contains (Supplementary Figure S9).

The number of inaccurate conclusions that have been drawn regarding *cis*-regulatory function due to unexpected RNA species such as those described here is difficult to estimate but potentially large given the widespread use of vectors harboring the pMB1 *ori*. In light of the frequency at which cryptic 3' ss are predicted to occur in genomic DNA, our findings indicate that the results of reporter tests of candidate regulatory sequences should be regarded as provisional until unexpected and confounding RNA species have been ruled out, or corroborating evidence from other sources is obtained. Going forward, the development of reporter plasmids with reduced cryptic transcription or improved insulation of the reporter gene unit would greatly facilitate the study of *cis*-regulation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2 and Supplementary Figures 1–9.

ACKNOWLEDGEMENTS

We are grateful to Drs Marilyn Kozak and Greg Goodall for critical review of a portion of the data presented. We thank Jian-Ting Zhang, Greg Goodall, Andrew Bert, James Smiley and Holly Saffran for generously providing plasmids.

FUNDING

The National Cancer Institute [CA121258 to N.K.]; National Institute of Diabetes and Digestive and Kidney Diseases [P30 DK 41301 for the CURE: Digestive Diseases Research Center, a Pilot and Feasibility Study award to C.R.L.]; United States Department of Defense [PC074133 to N.K.]. Funding for open access charge: National Cancer Institute [U01 NS059821].

Conflict of interest statement. None declared.

REFERENCES

- Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.*, **2**, 1044–1051.
- Carey, M., Peterson, C.L. and Smale, S.T. (2009) *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Maston, G.A., Evans, S.K. and Green, M.R. (2006) Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.*, **7**, 29–59.
- Kozak, M. (2005) A second look at cellular mRNA sequences said to function as internal ribosome entry sites. *Nucleic Acids Res.*, **33**, 6593–6602.
- Pelletier, J. and Sonenberg, N. (1988) Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature*, **334**, 320–325.
- Baranick, B.T., Lemp, N.A., Nagashima, J., Hiraoka, K., Kasahara, N. and Logg, C.R. (2008) Splicing mediates the activity of four putative cellular internal ribosome entry sites. *Proc. Natl Acad. Sci. USA*, **105**, 4733–4738.
- Sherrill, K.W., Byrd, M.P., Van Eden, M.E. and Lloyd, R.E. (2004) BCL-2 translation is mediated via internal ribosome entry during cell stress. *J. Biol. Chem.*, **279**, 29066–29074.
- Van Eden, M.E., Byrd, M.P., Sherrill, K.W. and Lloyd, R.E. (2004) Demonstrating internal ribosome entry sites in eukaryotic mRNAs using stringent RNA test procedures. *RNA*, **10**, 720–730.
- Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene*, **2**, 95–113.
- Betlach, M., Hershfield, V., Chow, L., Brown, W., Goodman, H. and Boyer, H.W. (1976) A restriction endonuclease analysis of the bacterial plasmid controlling the *ecoRI* restriction and modification of DNA. *Fed. Proc.*, **35**, 2037–2043.
- Lin-Chao, S., Chen, W.T. and Wong, T.T. (1992) High copy number of the pUC plasmid results from a *Rom/Rop*-suppressible point mutation in RNA II. *Mol. Microbiol.*, **6**, 3385–3393.
- Vieira, J. and Messing, J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene*, **19**, 259–268.
- Bert, A.G., Grepin, R., Vadas, M.A. and Goodall, G.J. (2006) Assessing IRES activity in the HIF-1 α and other cellular 5' UTRs. *RNA*, **12**, 1074–1083.
- Han, B. and Zhang, J.T. (2002) Regulation of gene expression by internal ribosome entry sites or cryptic promoters: the eIF4G story. *Mol. Cell. Biol.*, **22**, 7372–7384.
- Stoneley, M., Paulin, F.E., Le Quesne, J.P., Chappell, S.A. and Willis, A.E. (1998) C-Myc 5' untranslated region contains an internal ribosome entry segment. *Oncogene*, **16**, 423–428.
- Holcik, M., Lefebvre, C., Yeh, C., Chow, T. and Korneluk, R.G. (1999) A new internal-ribosome-entry-site motif potentiates XIAP-mediated cytoprotection. *Nat. Cell Biol.*, **1**, 190–192.
- Saffran, H.A. and Smiley, J.R. (2009) The XIAP IRES activates 3' cistron expression by inducing production of monocistronic mRNA in the betagal/CAT bicistronic reporter system. *RNA*, **15**, 1980–1985.
- Zhang, M.Q. (1998) Statistical features of human exons and their flanking regions. *Hum. Mol. Genet.*, **7**, 919–932.
- Matz, M., Shagin, D., Bogdanova, E., Britanova, O., Lukyanov, S., Diatchenko, L. and Chenchik, A. (1999) Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res.*, **27**, 1558–1560.
- Muller, P.Y., Janovjak, H., Miserez, A.R. and Dobbie, Z. (2002) Processing of gene expression data generated by quantitative real-time RT-PCR. *Biotechniques*, **32**, 1372–1374, 1376, 1378–1379.
- Thomas-Chollier, M., Sand, O., Turatsinze, J.V., Janky, R., Defrance, M., Vervisch, E., Brohee, S. and van Helden, J. (2008) RSAT: regulatory sequence analysis tools. *Nucleic Acids Res.*, **36**, W119–W127.
- Brunak, S., Engelbrecht, J. and Knudsen, S. (1991) Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J. Mol. Biol.*, **220**, 49–65.
- Hebsgaard, S.M., Korning, P.G., Tolstrup, N., Engelbrecht, J., Rouze, P. and Brunak, S. (1996) Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. *Nucleic Acids Res.*, **24**, 3439–3452.
- Reese, M.G., Eeckman, F.H., Kulp, D. and Haussler, D. (1997) Improved splice site detection in Genie. *J. Comput. Biol.*, **4**, 311–323.

25. Yeo, G. and Burge, C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.*, **11**, 377–394.
26. Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M. and Beroud, C. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.*, **37**, e67.
27. Buratti, E., Chivers, M., Hwang, G. and Vorechovsky, I. (2011) DBASS3 and DBASS5: databases of aberrant 3' and 5' splice sites. *Nucleic Acids Res.*, **39**, D86–D91.
28. Gan, W., LaCelle, M. and Rhoads, R.E. (1998) Functional characterization of the internal ribosome entry site of eIF4G mRNA. *J. Biol. Chem.*, **273**, 5006–5012.
29. Holcik, M., Graber, T., Lewis, S.M., Lefebvre, C.A., Lacasse, E. and Baird, S. (2005) Spurious splicing within the XIAP 5' UTR occurs in the Rluc/Fluc but not the betagal/CAT bicistronic reporter system. *RNA*, **11**, 1605–1609.
30. Nanbru, C., Lafon, I., Audigier, S., Gensac, M.C., Vagner, S., Huez, G. and Prats, A.C. (1997) Alternative translation of the proto-oncogene c-myc by an internal ribosome entry site. *J. Biol. Chem.*, **272**, 32061–32066.
31. Perier, R.C., Praz, V., Junier, T., Bonnard, C. and Bucher, P. (2000) The eukaryotic promoter database (EPD). *Nucleic Acids Res.*, **28**, 302–303.
32. Smale, S.T. and Kadonaga, J.T. (2003) The RNA polymerase II core promoter. *Annu. Rev. Biochem.*, **72**, 449–479.
33. Cencig, S., Nanbru, C., Le, S.Y., Gueydan, C., Huez, G. and Krays, V. (2004) Mapping and characterization of the minimal internal ribosome entry segment in the human c-myc mRNA 5' untranslated region. *Oncogene*, **23**, 267–277.
34. Andreev, D.E., Dmitriev, S.E., Terenin, I.M., Prassolov, V.S., Merrick, W.C. and Shatsky, I.N. (2009) Differential contribution of the m7G-cap to the 5' end-dependent translation initiation of mammalian mRNAs. *Nucleic Acids Res.*, **37**, 6135–6147.
35. Cobbold, L.C., Spriggs, K.A., Haines, S.J., Dobbins, H.C., Hayes, C., de Moor, C.H., Lilley, K.S., Bushell, M. and Willis, A.E. (2008) Identification of internal ribosome entry segment (IRES)-trans-acting factors for the Myc family of IRESs. *Mol. Cell Biol.*, **28**, 40–49.
36. Kazadi, K., Loeuillet, C., Deutsch, S., Ciuffi, A., Munoz, M., Beckmann, J.S., Moradpour, D., Antonarakis, S.E. and Telenti, A. (2008) Genomic determinants of the efficiency of internal ribosomal entry sites of viral and cellular origin. *Nucleic Acids Res.*, **36**, 6918–6925.
37. Kozak, M. (2008) Faulty old ideas about translational regulation paved the way for current confusion about how microRNAs function. *Gene*, **423**, 108–115.
38. Kozak, M. (2001) New ways of initiating translation in eukaryotes? *Mol. Cell Biol.*, **21**, 1899–1907.
39. Schneider, R. and Kozak, M. (2001) New ways of initiating translation in eukaryotes. *Mol. Cell Biol.*, **21**, 8238–8246.
40. de Wet, J.R., Wood, K.V., DeLuca, M., Helinski, D.R. and Subramani, S. (1987) Firefly luciferase gene: structure and expression in mammalian cells. *Mol. Cell Biol.*, **7**, 725–737.
41. Heard, J.M., Herbomel, P., Ott, M.O., Mottura-Rollier, A., Weiss, M. and Yaniv, M. (1987) Determinants of rat albumin promoter tissue specificity analyzed by an improved transient expression system. *Mol. Cell Biol.*, **7**, 2425–2434.
42. Langner, K.D., Weyer, U. and Doerfler, W. (1986) Trans effect of the E1 region of adenoviruses on the expression of a prokaryotic gene in mammalian cells: resistance to 5' CCGG-3' methylation. *Proc. Natl Acad. Sci. USA*, **83**, 1598–1602.
43. Levitt, N., Briggs, D., Gil, A. and Proudfoot, N.J. (1989) Definition of an efficient synthetic poly(A) site. *Genes Dev.*, **3**, 1019–1025.
44. Enriquez-Harris, P., Levitt, N., Briggs, D. and Proudfoot, N.J. (1991) A pause site for RNA polymerase II is associated with termination of transcription. *EMBO J.*, **10**, 1833–1842.
45. Van Eden, M.E., Byrd, M.P., Sherrill, K.W. and Lloyd, R.E. (2004) Translation of cellular inhibitor of apoptosis protein 1 (c-IAP1) mRNA is IRES mediated and regulated during cell stress. *RNA*, **10**, 469–481.
46. Beltran, M., Puig, I., Pena, C., Garcia, J.M., Alvarez, A.B., Pena, R., Bonilla, F. and de Herreros, A.G. (2008) A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev.*, **22**, 756–769.
47. Bushell, M., Stoneley, M., Kong, Y.W., Hamilton, T.L., Spriggs, K.A., Dobbins, H.C., Qin, X., Sarnow, P. and Willis, A.E. (2006) Polypyrimidine tract binding protein regulates IRES-mediated gene expression during apoptosis. *Mol. Cell*, **23**, 401–412.
48. Cuesta, R., Martinez-Sanchez, A. and Gebauer, F. (2009) miR-181a regulates cap-dependent translation of p27(kip1) mRNA in myeloid cells. *Mol. Cell Biol.*, **29**, 2841–2851.
49. de Almeida, R.A., Heuser, T., Blaschke, R., Bartram, C.R. and Janssen, J.W. (2006) Control of MYEOV protein synthesis by upstream open reading frames. *J. Biol. Chem.*, **281**, 695–704.
50. Dobson, T., Kube, E., Timmerman, S. and Krushel, L.A. (2008) Identifying intrinsic and extrinsic determinants that regulate internal initiation of translation mediated by the FMR1 5' leader. *BMC Mol. Biol.*, **9**, 89.
51. Dumas, E., Staedel, C., Colombat, M., Reigadas, S., Chabas, S., Astier-Gin, T., Cahour, A., Litvak, S. and Ventura, M. (2003) A promoter activity is present in the DNA sequence corresponding to the hepatitis C virus 5' UTR. *Nucleic Acids Res.*, **31**, 1275–1281.
52. Han, B., Dong, Z., Liu, Y., Chen, Q., Hashimoto, K. and Zhang, J.T. (2003) Regulation of constitutive expression of mouse PTEN by the 5' untranslated region. *Oncogene*, **22**, 5325–5337.
53. Han, B., Dong, Z. and Zhang, J.T. (2003) Tight control of platelet-derived growth factor B/c-sis expression by interplay between the 5' untranslated region sequence and the major upstream promoter. *J. Biol. Chem.*, **278**, 46983–46993.
54. Huang, B.L., Dornbach, L.M. and Lyons, K.M. (2007) The 5' untranslated regions (UTRs) of CCN1, CCN2, and CCN4 exhibit cryptic promoter activity. *J. Cell Commun. Signal.*, **1**, 17–32.
55. Li, P.W., Li, J., Timmerman, S.L., Krushel, L.A. and Martin, S.L. (2006) The dicistronic RNA from the mouse LINE-1 retrotransposon contains an internal ribosome entry site upstream of each ORF: implications for retrotransposition. *Nucleic Acids Res.*, **34**, 853–864.
56. Liu, Z., Dong, Z., Han, B., Yang, Y., Liu, Y. and Zhang, J.T. (2005) Regulation of expression by promoters versus internal ribosome entry site in the 5' untranslated sequence of the human cyclin-dependent kinase inhibitor p27kip1. *Nucleic Acids Res.*, **33**, 3763–3771.
57. Mancuso, G. and Rugarli, E.I. (2008) A cryptic promoter in the first exon of the SPG4 gene directs the synthesis of the 60-kDa spastin isoform. *BMC Biol.*, **6**, 31.
58. Timmerman, S.L., Pflingsten, J.S., Kieft, J.S. and Krushel, L.A. (2008) The 5' leader of the mRNA encoding the mouse neurotrophin receptor TrkB contains two internal ribosomal entry sites that are differentially regulated. *PLoS One*, **3**, e3242.
59. Wang, Z., Weaver, M. and Magnuson, N.S. (2005) Cryptic promoter activity in the DNA sequence corresponding to the pim-1 5' UTR. *Nucleic Acids Res.*, **33**, 2248–2258.
60. Young, R.M., Wang, S.J., Gordan, J.D., Ji, X., Liebhaber, S.A. and Simon, M.C. (2008) Hypoxia-mediated selective mRNA translation by an internal ribosome entry site-independent mechanism. *J. Biol. Chem.*, **283**, 16309–16319.