Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Internal ribosome entry sites and dicistronic RNAs in mammalian transgenesis

PETER S. MOUNTFORD AND AUSTIN G. SMITH

Modification of the genetic content of cultured cells or of whole animals is now a key strategy in both basic biological research and applied biotechnology. Yet obtaining the desired level and specificity of expression of an introduced gene remains highly problematic. One solution could be to couple expression of a transgene to that of an appropriate intact genomic locus. The identification and functional characterization of RNA sequences known as internal ribosome entry sites now offer the possibility of achieving precise control of transgene expression through the generation of dicistronic fusion mRNAs.

The translation of eukaryotic mRNAs is generally dependent upon 5'-cap-mediated ribosome binding. Ribosomes associate with the 5' cap structure before scanning along the mRNA molecule to the site of translational initiation, in most instances the first AUG codon. While the 'scanning model' of translation initiation[1] suitably accommodates most eukaryotic mRNAs, notable exceptions are provided by the picornavirus family. These viruses produce non-capped transcripts with long (600–1200 nucleotides) 5' untranslated regions (UTRs) containing multiple non-initiating AUG codons. The translational efficacy of these RNAs is dependent on the presence of specific sequences within the UTR called internal ribosome entry sites (IRES)[2,3].

The function of an IRES element was first shown by Pelletier and Sonenberg[2] who constructed artificial dicistronic mRNAs to determine whether internal entry of ribosomes could constitute an alternative pathway to cap-mediated ribosome binding in initiating translation (Fig. 1). Their investigations established that the intercistronic insertion of a defined poliovirus sequence of 490 nucleotides enabled efficient translation of a downstream cistron. Furthermore, cap-mediated translation of the upstream cistron could be blocked without affecting translation of the downstream cistron (Fig. 1c). Thus, the translation of the downstream cistron is not attributable to leaky scanning or reinitiation by ribosomes binding via the cap complex, but is governed independently by direct association of ribosomes with the IRES. In other words, the IRES acts as a ribosome landing pad[2]. It is of interest to note that these elegant experiments used poliovirus infection to block cap-dependent translation. Poliovirus, a member of the picornavirus family, is itself reliant upon IRES-mediated translation for protein synthesis. The normal infective strategy of the virus involves inactivation of the essential cap-binding complex eIF-4F and hence capture of the cellular translational apparatus for its own use. Importantly, however, the function of the IRES is not dependent on viral gene products but relies on interaction with normal cellular proteins[4]. Moreover, members of the cardiovirus subfamily such as encephalomyocarditis virus (EMCV) do not inhibit cellular translation and their IRES sequences function efficiently in competition with cap-mediated initiation (see below).

Internal initiation has been reported for certain cellular transcripts, such as the mRNA encoding the immunoglobulin heavy chain binding protein[5]. Resistance of specific mRNAs to inhibition of cap-mediated translation may be an important antiviral adaptation. It has also been suggested that cellular IRES elements could facilitate selective protein synthesis during mitosis when the activity of eIF-4F is greatly reduced[6]. The discovery of IRES elements in transcripts of the *Antennapedia* gene in *Drosophila*[6] and the mouse *Fgf2* gene[7] has prompted further speculation of a role for cap-independent translation in the developmental regulation of gene expression. It is worth noting that eukaryotic IRES sequences could be more widespread than has hitherto been realized because they cannot be identified by sequence homology; known IRES have been functionally defined and, so far, no conserved features have been found.

Naturally occurring viral or cellular IRES elements have so far been found only in 5' UTR sequences, with the possible exception of a putative IRES in the coding region of mRNA 3 of coronavirus. It is apparent, however, that advantage can be taken of these elements to create artificial dicistronic or multicistronic transcripts. Here, we discuss how IRES elements can be used in transgenic expression constructs to circumvent the constraints of cap-mediated translation and to create polyfunctional RNAs.

Coexpression of transgenes in cell culture

Expression of defined gene products

Many *in vitro* applications for mammalian transgenesis demand the coexpression of heterologous gene products. For example, in order to establish stable cell lines producing a recombinant protein it is generally necessary to introduce vectors for expression both of the protein of interest and of a selectable marker. This is usually achieved either by co-transfecting cells with two independent constructs or by introducing a single vector harbouring two discrete expression cassettes. The first approach is often limited by the inefficiency of co-transfection. The second requires the construction of relatively complex and cumbersome vectors and generally suffers from unreliable and/or low expression of the non-selectable cDNA. The use of an IRES in dicistronic expression vectors can circumvent these problems by enabling a single transcription unit to provide efficient production of both the protein of interest and a selectable marker[8] (Fig. 2). Furthermore, the direct coupling of the selectable marker to the desired cDNA enables effective selective pressure to be maintained for expression of the desired cDNA. Coexpression mediated by an IRES element from EMCV has been demonstrated in single quail QT6 cells *in vitro* following their transfection with RSV-based retroviral vectors[9]. Ghattas and co-workers found that >90% of cells infected

© 1995 Elsevier Science Ltd 0168 - 9525/95/$09.50

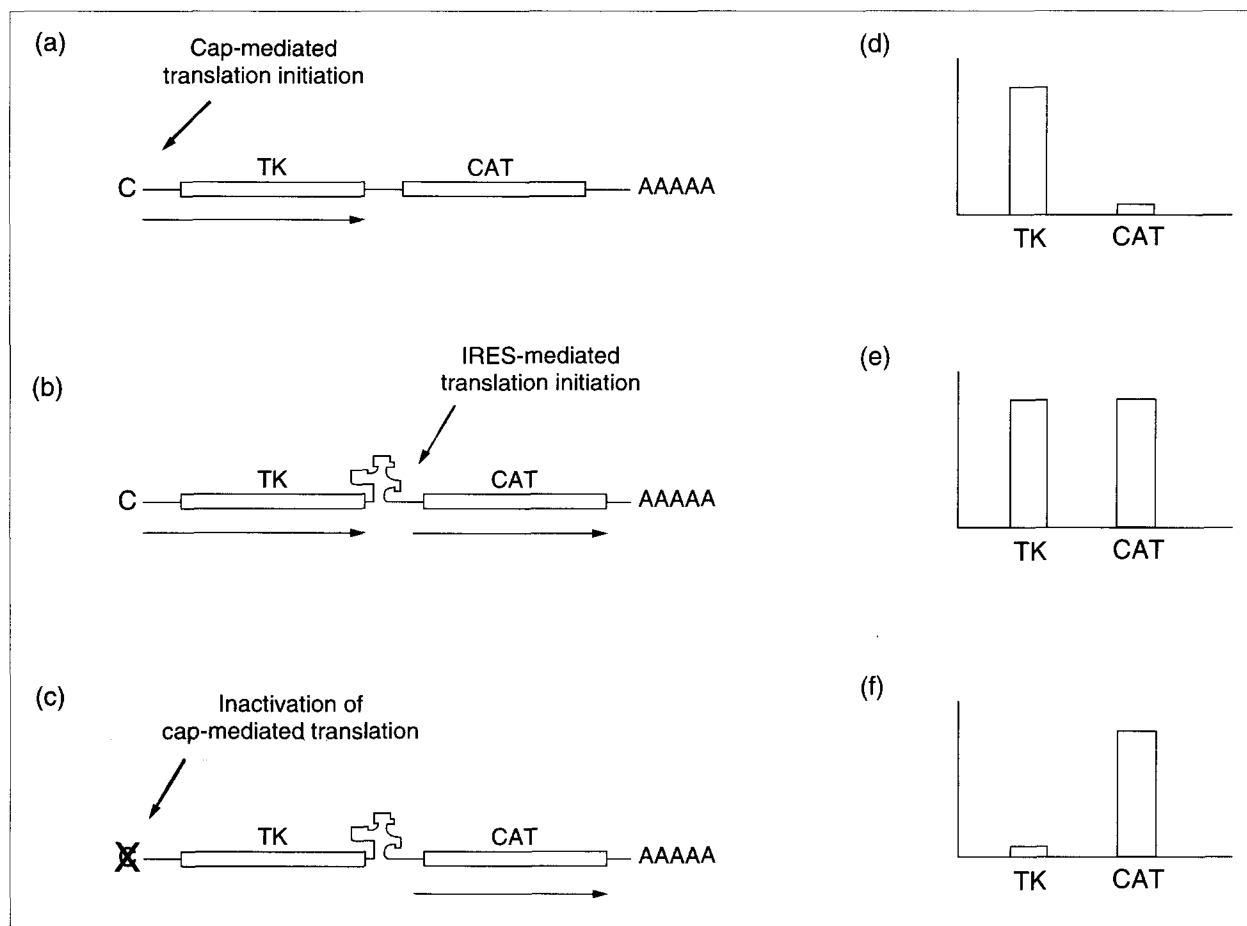

FIGURE 1. The internal initiation of translation mediated by an internal ribosome entry site (IRES). (a–c) Dicistronic constructs comprising the gene encoding the Herpes simplex virus thymidine kinase (TK) as the 5' cistron and the gene encoding bacterial chloramphenicol acetyl transferase (CAT) as the 3' cistron were translated *in vitro* and (d–f) reporter enzyme activity was assayed. (For more details, see Ref. 2.)

with IRES dicistronic constructs expressed both the selectable marker and the linked reporter[9]. By contrast, expression of the non-selectable gene product was rarely observed using viral vectors that rely on twin promoters or alternative splicing.

Most requirements for coexpression can be satisfied by vectors that provide comparable translational efficiencies for both cistrons of a dicistronic mRNA in a given cell type. However, the finding that translational efficiency of the poliovirus IRES varies according to the position of the initiating AUG codon[10] suggests that it may be possible to modulate the expression of the second cistron if desired. This could facilitate accurate correlations between biological response and amount of transgene product.

An IRES element can also be used to coexpress counter-selectable markers. Such cassettes could provide an important safeguard in gene therapy by providing an opportunity for the elimination of genetically manipulated cells in the event of their deregulation. Thus, a retroviral vector has been constructed in which expression of the multidrug resistance gene *MDR1* is linked via an IRES to the gene encoding the Herpes simplex virus thymidine kinase[11]. Cells producing the thymidine kinase could be selected against with the nucleoside analogues gancyclovir or FIAU, if necessary.

Functional expression cloning of novel cDNAs

In addition to facilitating the stable expression of characterized cDNAs, vectors incorporating IRES-mediated coexpression of a selectable marker may also be applied to the isolation of new genes through functional cloning approaches (Fig. 2). For instance, one route to the identification of cDNAs that affect the growth or differentiation of a particular cell type is to screen populations of cells transfected with cDNA expression libraries. Vectors with IRES-linked gene expression of a selectable marker promise significant increases in efficiency by ensuring that the majority of selected transfectants also express cDNA.

A powerful strategy for cloning cDNAs that encode interacting proteins is the two-hybrid system[12]. This is a screen based on the coexpression of a hybrid between a cDNA and an activation domain along with a fusion protein of a DNA binding domain and a target protein. The requirement for production of two proteins suggests that the methodology could be simplified by incorporating an IRES element to produce a single vector for coexpression of both fusion proteins (Fig. 2b). Certain IRES sequences have recently been demonstrated to work in *Saccharomyces cerevisiae*[13], so this approach could be applicable in yeast as well as in analogous mammalian systems[14,15].

IRES function in transgenic mice

Markers of transgene expression

The potential for dicistronic expression of transgenes *in vivo* was first investigated by infecting chick embryos with RSV-based vectors and demonstrating the translation of β-galactosidase mediated by an IRES sequence from EMCV[9]. Similar findings were subsequently described for mouse embryos following transfection of ES cells with plasmids containing IRES–*lacZ* expression cassettes[16,17]. Following on from these observations, plasmid expression vectors containing an IRES element from EMCV have been generated for the co-expression of potentially bioactive gene products and reporter gene products in mice. These vectors have particular application in correlating the ectopic expression of cDNAs with phenotypic effects *in vivo*. For example, ES cells can be transfected with dicistronic 'trifunctional' expression vectors (Fig. 2c) that coexpress cDNA products of interest along with the reporter and selectable marker *βgeo* (which is a fusion of the bacterial *lacZ* reporter gene and the gene for neomycin resistance *neo*[18]). Expression of *βgeo* enables transfected cell lines to be selected and also allows the visualization of sites of transgene expression in subsequent chimaeric analyses *in vivo*. Specific phenotypes, such as developmental perturbations, that may be evident in the chimaeras can then be correlated with cellular expression of the linked marker.

Faithful transgene expression

The ideal transgenic system would provide quantitatively and qualitatively defined expression of the gene product of interest. However, an all too frequent limitation to the successful exploitation of mammalian transgenesis, has been the failure of conventional expression constructs to provide full and controlled or 'faithful' transgene expression.

Most transgenic experimentation has relied upon small expression constructs that incorporate a variety of sequence elements derived, in many instances, from a number of different sources. Promoter and enhancer elements are routinely combined with heterologous introns and polyadenylation signals to generate, in principle, an idealized mini-gene configuration for the expression of an associated cDNA. Although convenient, this type of construct often results in unpredictable and generally low levels of expression. The identification and incorporation of locus control regions (LCRs) has provided a route to full, position-independent transgene expression[19]. However, this approach requires comprehensive knowledge of the specific gene to be mimicked and as such is currently limited to a handful of well-characterized genes.

An alternative and effective strategy that does not require detailed knowledge of contributing regulatory elements is to transfer the gene of interest as an entire genomic locus, for example, via the use of yeast artificial chromosomes[20] or ultimately by using mammalian artificial chromosomes[21]. At its simplest, this approach is

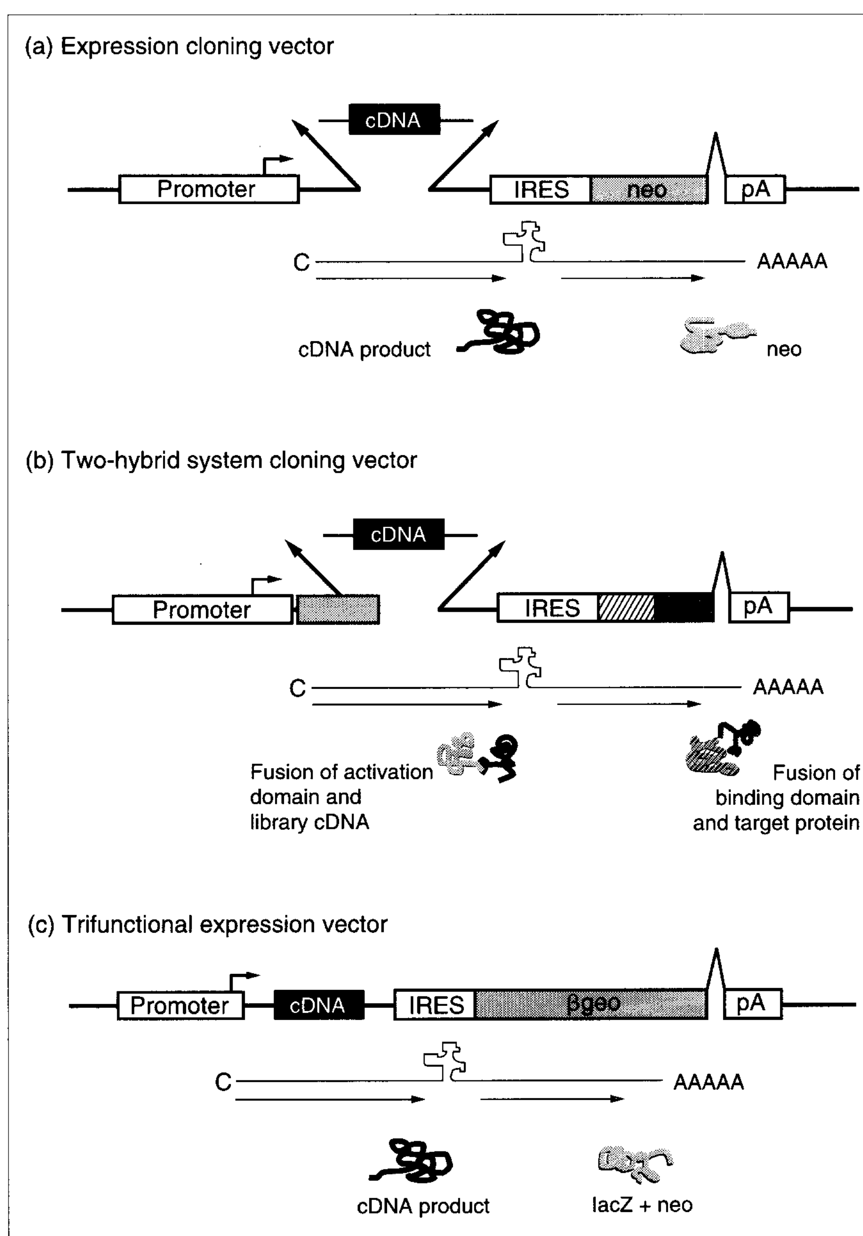

FIGURE 2. Applications of the internal ribosome entry site (IRES) in cloning vectors and transgene expression vectors. (a) Vector for functional screening of cDNA libraries by transfection of mammalian cells. (b) Construct for coexpression of fusion partners in a two-hybrid cloning system. (c) Construct for coexpression of the reporter and selectable marker fusion gene *βgeo*[18], with cDNA in ES cells and mouse embryos.

TABLE 1. Frequency of homologous recombinants isolated using constructs containing an internal ribosome entry site

Gene	Colonies screened	Homologous recombinants	Frequency (%)
Oct3/4 (i)[a]	51	44	86
Oct3/4 (ii)	30	21	70
Dia/Lif	79	21	26
Dia/Lif-R[b]	58	11 (22)	19 (38)
Mecp2[c]	87	68	78

[a]The data for the gene encoding the transcription factor Oct-3/4 were taken from two independent transfections[17].
[b]Gene encoding the receptor for the cytokine known as differentiation inhibiting activity (DIA or LIF); figures in parentheses are the targeted events that deviated in structure from a canonical replacement event (M. Li and A.G. Smith, unpublished).
[c]Methylcytidine binding protein 2 (P. Tate and A. Bird, unpublished).

limited to expressing the endogenous gene product in its native expression pattern. Wider applications can be envisaged if the advantages of genomic regulation can be coupled with expression of heterologous sequences. In previous genomic fusion constructs, cDNA insertion has been restricted to sites upstream of the genomic translation initiation codon. Results have been disappointing due to alterations in RNA splicing and processing and the deletion of internal regulatory sequences[22]. However, the use of IRES elements would bypass the constraints of cap-mediated translation initiation and allow a cDNA sequence to be positioned internally or preferably in the 3' UTR of an intact gene. Insertion of a heterologous reading frame in the last exon of a gene might be particularly effective because this would be compatible with the larger size of final exons in vertebrates[23] and is anticipated to have minimal consequences for RNA processing. Consequently, the desired level and specificity of gene expression should be retained.

One deficiency in our current knowledge is an assessment of whether IRES function is completely independent of surrounding sequences. It is possible that particular mRNA secondary structures could interfere with IRES function or that stretches of homologous sequences could directly disrupt the structure of the IRES. A second area requiring further investigation is that of the relative efficiencies of cap-mediated translation versus IRES-mediated translation. Also, it is not yet certain that all cells can support IRES-mediated translation. IRES elements are used efficiently by various cellular extracts and this is exploited experimentally for enhancing the performance of translation systems *in vitro*. IRES sequences are also effective in expression strategies in cell culture based on bacteriophage RNA polymerases that produce non-capped transcripts[24]. Most compellingly, data generated following the introduction of the IRES–*βgeo* cassette into a variety of chromosomal loci by either gene trapping or gene targeting (see below) have established that the IRES sequence from EMCV is functional *in vivo* in a wide range of both fetal and adult mouse tissues. Whether this is true for other IRES elements and in other species has yet to be determined.

Gene trapping with IRES sequences

Gene trapping is a strategy for identifying developmentally regulated genes based on the random integration of a reporter into chromosomal transcription units[25]. The efficiency of this approach is generally limited by the requirement for the production of in-frame fusion proteins that retain reporter activity. These constraints can be bypassed by using an IRES to allow independent translation of the reporter from fusion transcripts. This is predicted to significantly enhance the frequency of productive integrations because any insertion into an active gene should give functional reporter expression. Indeed, about ten times more G418-resistant ES cell colonies were obtained with the IRES-containing gene trap construct pGT1.8Iresβgeo compared with the parental pGT1.8βgeo vector (P.S. Mountford and W. Skarnes, unpublished). Exploitation of IRES elements could therefore prove invaluable in large-scale gene trap screens.

However, the elimination of fusion proteins between host protein and reporter is not always desirable. Sequences contributed by the host can direct subcellular localization of the fusion product and thereby provide valuable information regarding the nature of the trapped gene product[26]. With this in mind, a gene trap vector called pGTZIN has been designed to express the β-galactosidase reporter as a fusion protein along with IRES-mediated coexpression of the selectable *neo* cassette. Independent translation of the *neo* gene product may prove significant if certain subcellular localizations are incompatible with inactivation of the selective agent G418. Several nuclear-localized fusions have been isolated with this vector. One of these, ZIN40, gives strong and apparently ubiquitous nuclear expression of β-galactosidase in embryonic and adult mouse tissues.

IRES elements and gene targeting

The use of promoterless constructs provides a major enrichment for homologous recombination events in gene targeting by eliminating the selection of random integrations into non-transcribed regions of the genome[27]. However, the constraints on the production of functional selectable markers from fusion transcripts have meant that this approach has had limited application in the past. Exploitation of IRES sequences should alter this situation by simplifying the design of targeting vectors and reducing the need for extensive prior characterization.

The incorporation of an IRES-selectable marker cassette into a gene targeting vector was first described by Wood *et al.*[28]. These workers targeted the gene encoding the immunoglobulin heavy chain in hybridoma cells. They successfully isolated homologous recombinants, but only at a low frequency comparable to that obtained using conventional vectors. However, following

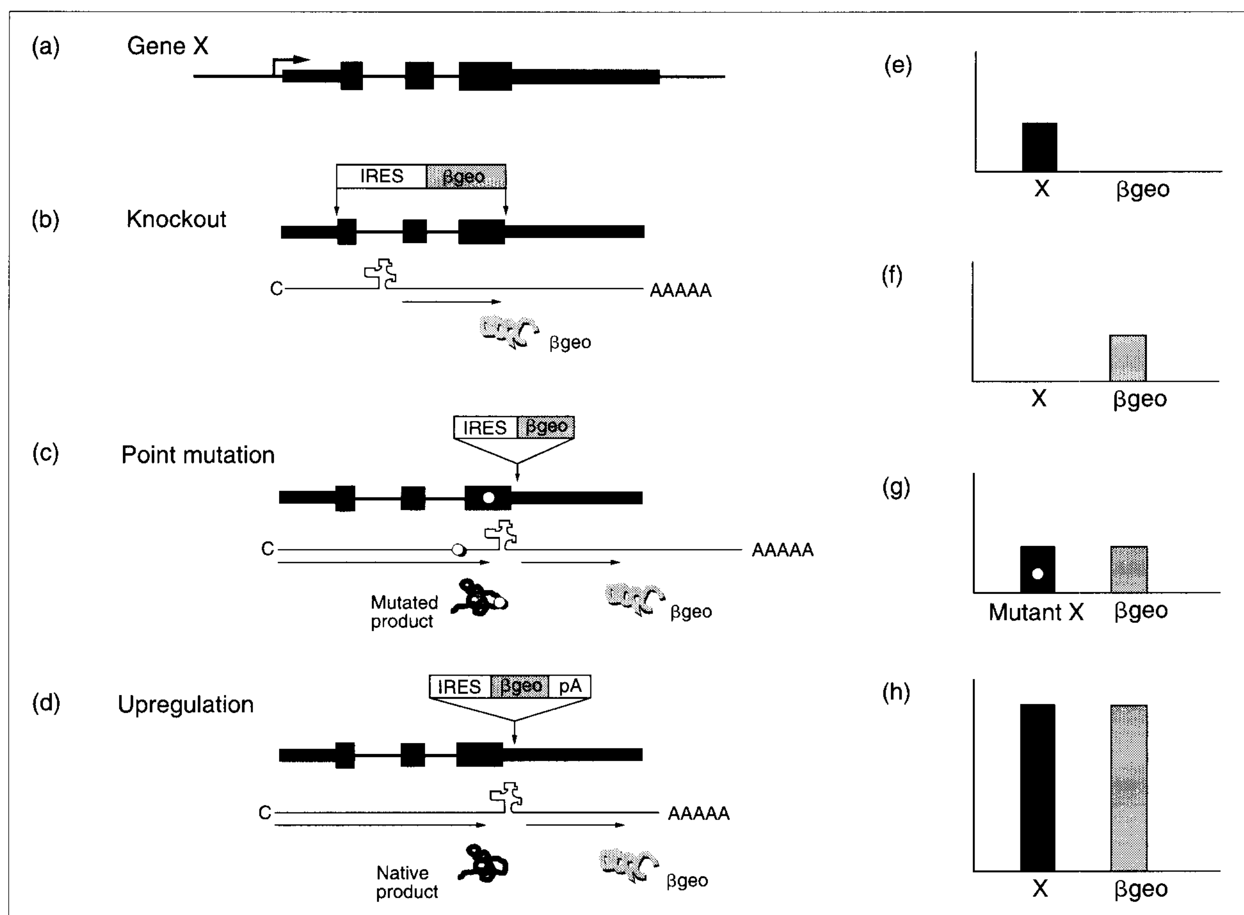

FIGURE 3. Applications of the internal ribosome entry site (IRES) in targeted modification of chromosomal genes. Potential modifications of a target gene (a) through the use of highly efficient IRES-mediated selection. Filled boxes represent coding sequences and thick lines represent untranslated regions. (b) Gene deletion or disruption. (c) Introduction of a subtle mutation in a single step. (Incorporation of the IRES cassette into the 3' UTR is anticipated to have minimal effect on normal expression.) (d) Example of altered expression of a target gene achieved by truncating endogenous destabilizing motifs in the 3' UTR.

the demonstration that the IRES sequence from EMCV is functional in ES cells, the efficacy of this element in gene targeting has been clearly established[17]. The data summarized in Table 1 show that the use of IRES-containing constructs allows homologous recombination events to be isolated at a very high frequency in a variety of loci that exhibit different levels of mRNA expression. Three of these mutations have been transmitted through the mouse germ line, indicating that incorporation of the EMCV IRES does not compromise germ line competence. The results of gene trapping and gene targeting studies in ES cells indicate that a broad range of genes, including many that are developmentally regulated, have sufficient transcriptional activity in ES cells to drive selectable marker expression[17,26,29]. Consequently, the advantages afforded by using an IRES should be widely applicable.

The expression profile of IRES–*lacZ* reporters introduced by homologous recombination reflects that of the endogenous genes, both *in vitro* and *in vivo* (Ref. 17; M. Li and A.G. Smith, unpublished). Exploitation of IRES-linked selectable markers should therefore allow subtle structural or regulatory alterations to be introduced into resident genes in a single step (Fig. 3). Point mutations could be introduced into genes without

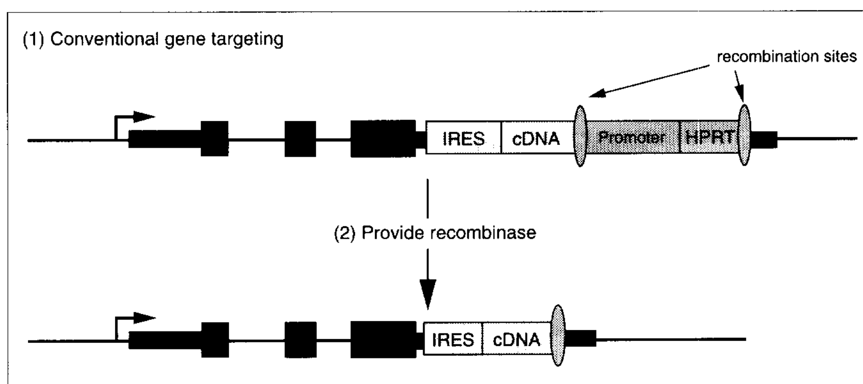

FIGURE 4. Strategy for coupling the expression of a cDNA to a chromosomal gene. Introduction of an IRES–cDNA cassette into the 3' UTR of a target gene, with subsequent excision of the selectable marker by site-specific recombination.

affecting normal expression, or qualitative or quantitative alterations of gene expression could be engendered by interrupting or modifying regulatory sequences. For example, an IRES–*βgeo* cassette with the SV40 polyadenylation signal has been used to truncate the endogenous 3' UTR of the transcript encoding the cytokine known as differentiation inhibiting activity (DIA or LIF)[17]. The effect of this was to delete mRNA-destabilizing sequences and thereby increase the steady-state level of mRNA by several fold. The relative ease with which such modifications can be generated should mean that homologous recombination can be used routinely in both the manipulation of gene expression and the analysis of gene structure and regulation within native chromosomal loci.

A key opportunity afforded by IRES elements is a means of conferring the desired specificity and level of expression on a transgene via coexpression with an appropriate chromosomal gene. As discussed above, an IRES could be incorporated into genomic constructs for conventional additive transgenesis. Another strategy would be to exploit homologous recombination to integrate IRES–cDNA cassettes into 3' UTRs of resident genes. One way of achieving this would be to use polycistronic targeting vectors incorporating both IRES–cDNA cassettes and IRES–selectable-marker elements. In the case of non-expressed genes, the selectable marker could be promoter-driven. The selectable marker cassette could subsequently be excised by site-specific recombination[30] (Fig. 4). Alternatively, sequential targeting[31] could be used: a counter-selectable marker could be incorporated in an initial homologous recombination event, followed by substitution with the IRES–transgene construct in a second step. By either of these routes it should be feasible to place a cDNA under the full regulatory control of an endogenous genomic locus. An interesting application of this approach would be to introduce selectable markers or immortalizing genes into tissue-specific loci in order to facilitate the isolation of particular cell types.

Perspectives

Recent work has established that the IRES element of EMCV is functional in a variety of cultured cell types and in whole animals. Toxic side effects have not been observed at normal expression levels. Importantly, the EMCV IRES does not appear to possess any intrinsic transcriptional regulatory activity. Consequently, it is expected to have a neutral effect on gene expression. However, several important issues still require further investigation. For example, it is not yet clear whether the IRES functions with equal efficacy regardless of sequence context or whether activity is equivalent in all cell types at all stages, nor has the quantitative relationship between translation of upstream and downstream cistrons been determined. Nonetheless, the potential uses of IRES sequences in gene targeting and for the generation of polycistronic transcripts are already apparent. In particular, the possibilities for functional coupling of transgenes to intact chromosomal loci, whether in constructs for pronuclear injection or via homologous recombination, may herald a new era in precision transgenesis.

Acknowledgements

We are indebted to W.C. Skarnes for fruitful discussion and comments on the manuscript. We thank all in A.G. Smith's lab and also W.C. Skarnes, P. Tate and A. Bird for unpublished data. P.S.M. is supported by a Royal Society Endeavour Fellowship.

References

1 Kozak, M. (1989) *J. Cell Biol.* 108, 229–241
2 Pelletier, J. and Sonenberg, N. (1988) *Nature* 334, 320–325
3 Molla, A. *et al.* (1992) *Nature* 356, 255–257
4 Jang, S.K. and Wimmer, E. (1990) *Genes Dev.* 4, 1560–1572
5 Macejak, D.G. and Sarnow, P. (1991) *Nature* 353, 90–94
6 Vagner, S. *et al.* (1995) *Mol. Cell. Biol.* 15, 35–44
7 Oh, S-K. and Sarnow, P. (1993) *Curr. Opin. Genet. Dev.* 3, 295–300
8 Kaufman, R.J., Davies, M.V., Wasley, L.C. and Michnick, D. (1991) *Nucleic Acids Res.* 19, 4485–4490
9 Ghattas, I.R., Sanes, J.R. and Majors, J.E. (1991) *Mol. Cell. Biol.* 11, 5848–5859
10 Pilipenko, E.V. *et al.* (1992) *Cell* 68, 119–131
11 Sugimoto, Y., Aksentijevich, I., Gottesman, M.M. and Pastan, I. (1994) *Biotechnology* 12, 694–698
12 Fields, S. and Song, O-K. (1989) *Nature* 340, 245–246
13 Iizuka, N., Najita, L., Franzusoff, A. and Sarnow, P. (1994) *Mol. Cell. Biol.* 14, 7322–7330
14 Vasavada, H.A. *et al.* (1991) *Proc. Natl Acad. Sci. USA* 88, 10686–10690
15 Fearon, E.R. *et al.* (1992) *Proc. Natl Acad. Sci. USA* 89, 7958–7962
16 Kim, D.G., Kang, H.M., Jang, S.K. and Shin, H-S. (1992) *Mol. Cell. Biol.* 12, 3636–3643
17 Mountford, P. *et al.* (1994) *Proc. Natl Acad. Sci. USA* 91, 4303–4307
18 Friedrich, G. and Soriano, P. (1991) *Genes Dev.* 5, 1513–1523
19 Grosveld, F., Assendeift, G.B.V., Greaves, D.R. and George, K. (1987) *Cell* 51, 975–985
20 Huxley, C. (1994) *Genet. Eng.* 16, 65–91
21 Brown, W.R.A. (1992) *Curr. Opin. Genet. Dev.* 2, 479–486
22 Whitelaw, C.B.A. *et al.* (1991) *Transgenic Res.* 1, 3–14
23 Hawkins, J.D. (1988) *Nucleic Acids Res.* 16, 9893–9908
24 Elroy-Stein, O., Fuerst, T.R. and Moss, B. (1989) *Proc. Natl Acad. Sci. USA* 86, 6126–6130
25 Skarnes, W.C. (1990) *Biotechnology* 8, 827–831
26 Skarnes, W.C. *et al. Proc. Natl Acad. Sci. USA* (in press)
27 Jasin, M. and Berg, P. (1988) *Genes Dev.* 2, 1353–1363
28 Wood, C.R. *et al.* (1991) *Proc. Natl Acad. Sci. USA* 88, 8006–8010
29 Jeannotte, A., Ruiz, J.C. and Robertson, E.J. (1991) *Mol. Cell. Biol.* 11, 5578–5585
30 Jung, S., Rajewsky, K. and Radbruch, A. (1993) *Science* 259, 984–987
31 Stacey, A. *et al.* (1994) *Mol. Cell. Biol.* 14, 1009–1016

P.S. Mountford is in the Centre for Animal Biotechnology, School of Veterinary Science, The University of Melbourne, Parkville, 3052 Victoria, Australia, and A.G. Smith is in the Centre for Genome Research, The University of Edinburgh, King's Buildings, West Mains Road, Edinburgh, UK EH9 3JQ.