Available online at www.sciencedirect.com
SCIENCE DIRECT®
Methods 37 (2005) 331–340
METHODS
www.elsevier.com/locate/ymeth

Use of minigene systems to dissect alternative splicing elements

Thomas A. Cooper *

Departments of Pathology and Molecular and Cellular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Accepted 5 July 2005

Abstract

Pre-mRNA splicing is an essential step for gene expression in higher eukaryotes. The splicing efficiency of individual exons is determined by multiple features involving gene architecture, a variety of *cis*-acting elements within the exons and flanking introns, and interactions with components of the basal splicing machinery (called the spliceosome) and auxiliary regulatory factors which transiently co-assemble with the spliceosome. Both alternative and constitutive exons are recognized by multiple weak protein:RNA interactions and different exons differ in the interactions which are determinative for exon usage. Alternative exons are often regulated according to cell-specific patterns and regulation is mediated by specific sets of *cis*-acting elements and *trans*-acting factors. Transient expression of minigenes is a commonly used in vivo assay to identify the intrinsic features of a gene that control exon usage, identify specific *cis*-acting elements that control usage of constitutive and alternative exons, identify *cis*-acting elements that control cell-specific usage of alternative exons, and once regulatory elements have been identified, to identify the *trans*-acting factors that bind to these elements and modulate splicing. This chapter describes approaches and strategies for using minigenes to define the *cis*-acting elements that determine splice site usage and to identify and characterize the *trans*-acting factors that bind to these elements and regulate alternative splicing.

Keywords: Minigene; Deletion analysis; RNA binding; Exon splicing enhancer; Exon splicing silencer; Intronic splicing enhancer; Intronic splicing silencer

1. Introduction

The majority of human genes express multiple mRNAs via alternative splicing. These mRNAs can encode functionally diverse protein isoforms or can promote the on:off control of gene expression by introducing premature translation stop codons [1]. A large fraction of these alternative splicing events are regulated according to cell-specific patterns or in response to acute stimulation [2]. Whether or not a splice site is used is determined by a large number of features. Some features establish a basal level of exon recognition and others are determinative for the modulation of alternative splicing in response to specific cues. The features that establish a basal level of exon recognition include splice site strength, the size of the exon, the sizes of flanking introns, and the presence of secondary structures. In addition, multiple negative and positive *cis*-acting elements contribute to both the basal level of exon recognition and the modulation of exon use in response to specific signals (Fig. 1). Alternative exons are typically recognized less efficiently than constitutive exons allowing for their modulated use. The consensus 5′ and 3′ splice sites flanking alternative exons are typically non-canonical and are less efficient binding sites for the components of the basal splicing machinery (the spliceosome). Exons contain exonic splicing enhancers (ESEs) and exonic splicing silencers (ESSs) that enhance or repress exon usage, respectively (Fig. 1). The first ESEs to be identified in vivo based on functional analyses using minigenes were in alternative exons [3–5]. The best characterized ESEs are purine-rich and their enhancer activities are mediated though binding to members of the SR protein family [6]. A lesser characterized A/C-rich family of ESEs mediate their effects by binding to the cold box protein, YB-1 [7]. ESSs bind to hnRNP proteins and functional analysis using minigenes have demonstrated that this binding of hnRNP proteins

* Fax: +1 713 798 5838.
E-mail address: tcooper@bcm.edu.

1046-2023/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2005.07.015

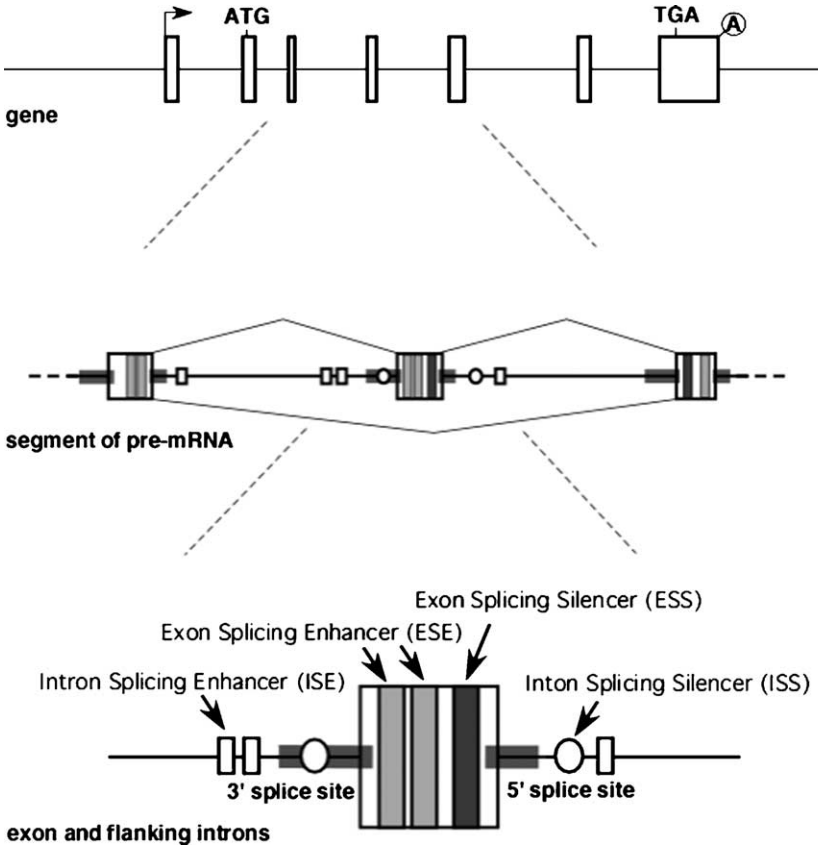


Fig. 1. Exonic and intronic elements.

mediates repression [8]. Recent analyses that combine biocomputational identification and functional analysis in minigenes have determined that ESEs and ESSs are common to both alternative and constitutive exons [8–11]. While some of the motifs identified by biocomputational analysis match previously identified ESE and ESS motifs, the majority do not, indicating that a great deal of information remains to be discovered regarding how exons are recognized. Intronic splicing enhancers (ISEs) and intronic splicing silencers (ISSs) within the upstream and downstream introns (Fig. 1) also serve to enhance or repress, respectively, use of associated splice sites [12].

Minigene constructs are an important tool for the identification and in vivo analysis of the *cis*-acting regulatory elements and *trans*-acting factors that establish splicing efficiency and that regulate alternative splicing. Expression of minigene pre-mRNAs by transient transfection provides a rapid assay for loss-of-function and gain-or-function analyses for *cis*-elements and *trans*-acting factors that affect splicing regulation. This chapter will present approaches to define the intronic and exonic *cis*-acting elements that establish basal splicing efficiency, identify elements responsible for cell-specific regulation, and identify specific sequence motifs that are required for regulation by individual *trans*-acting factors. Some considerations for the identification of elements that regulate cell-specific alternative splicing have been previously described [13].

2. Initial considerations

Minigenes can be used to address a number of different questions with regard to alternative splicing regulation. The first step of an investigation to dissect the features that regulate alternative splicing is to establish the specific issue(s) to be addressed. The common uses of minigenes include: (i) determine the role of the splice sites in establishing a basal level of exon recognition, (ii) identify exonic and intronic elements that enhance or repress splicing, (iii) determine whether an allelic variant has an effect on splicing efficiency, (iv) identify elements required for cell-specific regulation, and (v) identify elements required for regulation by a specific *trans*-acting factor. The approaches for these goals differ and will be addressed separately in the section entitled, "Analysis of results."

3. Description of the method

3.1. Minigene design

Minigenes contain a genomic segment from the gene of interest that includes the alternatively spliced region and flanking genomic regions. The genomic segment is generated by PCR amplification either directly from genomic DNA or a cloned genomic DNA fragment as template. The lengths of the flanking genomic segments depend on the goals of the

analysis and can range from tens to thousands of nucleotides. To determine whether an allelic variant is responsible for altered splicing or to test whether a specific motif is the response element for a putative regulatory factor, the minigene will include the *cis*-element(s) of interest. To identify the *cis*-acting elements required for cell-specific regulation of an alternative exon or in response to overexpression and knockdown studies of specific proteins, one must choose a fragment that is likely to contain the primary regulatory elements. Based on current knowledge, the primary elements regulating alternative splicing are typically within 200–300 nucleotides upstream and/or downstream of the regulated exon. While there are cases in which cell-specific regulatory elements have been found more than one kilobase from the regulated splice site [14–16], a genomic fragment containing up to 300 nucleotides of flanking intronic sequences upstream and downstream from the exon is a reasonable starting point (Fig. 2). Once a genomic segment has been shown to contain elements that are sufficient for regulated splicing, sequential deletions from the 3′ and 5′ ends can be used to define the smallest genomic fragment that is necessary and sufficient for regulation. It is advantageous to define a minimal responsive genomic segment to facilitate identification of the relevant *cis* elements by eliminating functionally redundant elements. In cases in which the initial genomic fragment fails to show regulation, we have been successful using larger genomic segments to demonstrate regulation. For example, a genomic fragment spanning from the exon upstream of the alternative exon to the exon downstream of the alternative exon is very likely to include all relevant regulatory elements necessary and sufficient for the regulation since regulatory elements are rarely identified upstream or downstream of the exons that flank the regulated exon.

3.2. Amplification of the genomic fragment

The genomic fragment to be cloned into the minigene is amplified from genomic DNA using oligonucleotides containing restriction enzyme sites at their 5′ ends that match restriction sites in the recipient plasmid. These sites will depend on the available cloning sites within the minigene (RS#1 and RS#2, Fig. 2). It is most convenient to select two restriction enzymes that cut efficiently in the same buffer. The cloning sites within the priming oligonucleotides need to be unique to the PCR product. Usually the available minigene plasmid has only one or two restriction enzyme sites that can be used for insertion of the genomic segment. If the cloning site within the vector is also found in the genomic segment to be cloned, different restriction enzyme sites with identical overhangs can be incorporated into the PCR primers to incorporate a site unique to the PCR fragment that can be used for cloning. For example, if the cloning sites within the minigene are *Bam*HI (5′ site) and *Xho*I (3′ site) but there are *Bam*HI and *Xho*I sites within the genomic segment to be amplified, *Bgl*II and *Sal*I restriction sites can be incorporated into the upstream and downstream PCR primers, respectively. The *Bam*HI/*Bgl*II and *Sal*I/*Xho*I pairs have identical overhangs and can be ligated allowing for use of the limited number of sites within the minigene for cloning. Note that ligation of different sites with the same "sticky ends" results in a loss of both sites. A full list of restriction sites with compatible overhangs can be found at (http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/compatible_cohesive_overhangs.asp).

Because many restriction enzymes do not cut efficiently at sites that are at the end of the molecule, six nucleotides (often ATATAT) are included as the first nucleotides of the oligonucleotide. These nucleotides are then followed by the restriction site and then 18–21 nucleotides that are complementary to the gene of interest and serve to prime the PCR. The digestion efficiency of several restriction enzymes at the end of molecules has been tested and can be found at http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/cleavage_linearized_vector.asp.

PCRs are performed using a high fidelity polymerase so as not to introduce nucleotide substitutions. We use either Pfu and Taq Plus Precise (both from Stratagene) according to the manufacturer's protocols. PCRs using Taq Plus Precise are performed in 50 μl reactions and contain 50–500 ng

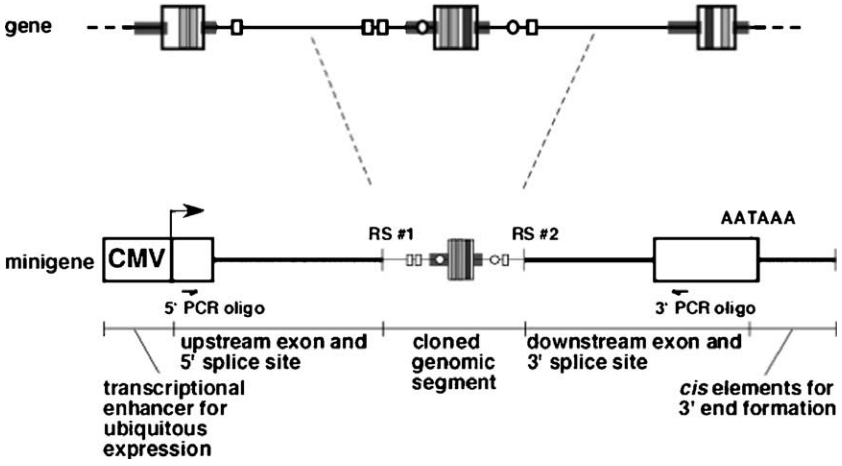


Fig. 2. Minigene vector. CMV, cytomegalovirus transcriptional enhancer/promote.

genomic DNA (purchased from Clontech), 1× Taq Plus Precise buffer (supplied with the enzyme), 0.25 mM dNTPs, 2.5 μl DMSO, 150 ng of each oligo, and 2.5 U of Taq Plus Precise. A master mix is made for four reactions including all reagents except the enzyme and genomic DNA. The master mix is aliquoted into the PCR tubes which contain 50, 100, 500 ng and a no DNA control. A ''hot-start'' reduces background amplification. There are several hot-start polymerases available. The low tech but effective approach for hot-start when the number of tubes is small is to add the polymerase after the temperature of the first cycle reaches 65 °C and mix by pipetting up and down.

3.3. Transfection and RNA extraction

The choice of cells to use for transfection depends upon the goals being pursued. Cells with the highest transfection efficiency such as HeLa, COS, and HEK293 are generally used to define the basis for ''non-constitutive'' splicing (that is, to identify the features that result in inefficient splicing), to identify general regulatory elements, to test the effects of allelic variants, and to test whether splicing is regulated by co-expression of putative regulators. To identify *cis*-acting elements required for cell-specific regulation, the ideal approach is to use one cell culture which can be induced to undergo a physiological transition that is associated with the splicing transition under investigation (e.g., differentiation or stimulation of a cellular response by external stimuli). Ideally, the minigene can be demonstrated to undergo the same regulated splicing transition as the endogenous gene.

In lieu of a single-culture system, splicing effects are often compared in two different cell lines, one expressing the ''regulated'' splicing pattern reflecting the cell-specific pattern of interest and the other expressing a different splicing pattern. The caveat with this approach is that different cell lines can exhibit different ratios of alternative splicing patterns for reasons that are unrelated to the natural regulation that occurs in the tissue of interest. Indeed, we have found that expression of the same minigene in several different and commonly used cell lines exhibit a variety of splicing efficiencies (data not shown). Since none of these cell lines reflect a cell-specific differentiated state, these splicing differences are unlikely to reflect regulation that is physiologically relevant to tissues. Also note that the ''same'' cell line will vary between labs due to differences in propagation which affects features such as transfection efficiency and splicing patterns. In our lab, a single operator has observed changes in splicing patterns within the same cell line over time. It is wise to begin by purchasing a fresh cell stock (from ATCC, for example) and to expand and freeze down a large number of stock vials and monitor cultures for changes that warrant use of a fresh culture. This will help reduce variability that can complicate comparisons from transfections performed months or years apart.

When using cells with high transfection efficiency (>30% of cells transfected), transfections can be performed in six well plates using 200 ng to 1 μg of total plasmid DNA (minigene plus protein expression vector). COSM6 cells ($1.5–2 \times 10^5$ cells) are plated such that they reach 100% confluency no sooner than the day after transfection (day two after plating). Alternatively, we have found that our COSM6 cells can be transfected 6 h after plating. Typically we transfect a total of 1 μg total plasmid DNA including 500 ng minigene, 500 ng of protein expression vector plasmid. When variable amounts of plasmid are used, we add a carrier plasmid lacking an RNA polymerase II promoter such that all transfections contain the same amount of DNA. The constant amount of DNA reduces variable transfection efficiency. A large number of transfection reagents are available and can be used according to the manufacturer's instructions. We prefer Fugene6 (Roche) due to its high transfection efficiency in multiple cell lines, ease of use, and low toxicity. RNA is typically extracted on the second day (40–48 h after transfection), however, RNA as well as coexpressed protein can be readily detected by RT-PCR on the day after transfection in cells that transfect with high efficiency. Total cellular RNA is extracted using the Trizol reagent (Invitrogen) according to the manufacturer's protocols. Following extraction, RNA is stored in RNase free water at −20 or −80 °C.

3.4. Quantification of results

Alternative splicing patterns can be analyzed using S1 nuclease protection, RNase protection, RT-PCR analyses or the RNA invasive cleavage assay [35]. RT-PCR is the most commonly used assay because of its sensitivity and straightforward protocol. We commonly use the ''two-tube'' cDNA and PCRs in which cDNA synthesis is performed separately from PCR amplification. In addition, one oligonucleotide is used to prime cDNA synthesis in the reverse transcription reaction and as the reverse primer for PCR. First-strand synthesis of cDNA is performed in 20 μl using 1–2 μg of total RNA from transfected cells, 10 ng of the downstream (reverse) primer, 0.5 mM dNTPs and 4 U AMV reverse transcriptase (Life Sciences) in First-Strand Synthesis Buffer provided by the manufacturer at 42 °C for 1 h. PCR amplification is performed using half of the cDNA synthesis reaction (added directly to the PCR mixture), 200 ng of upstream and downstream oligonucleotides, and includes 1.5 ng ^{32}P 5′-end-labeled forward primer (see below), 1× *Taq* $MgCl_2$-free buffer (Promega), 1.75 mM $MgCl_2$, 0.2 mM dNTPs, and 2.5 U *Taq* polymerase (Promega) [14]. The labeling reaction for the forward primer is made up as a master mix to label 1.5 ng of the forward primer per reaction and contains 70 mM Tris (pH 7.5), 1 mM $MgCl_2$, 0.5 mM DTT, 0.7 U of polynucleotide kinase (USB), an equal molar amount of [γ^{32}P]ATP in a 10 μl volume and is incubated at 37 °C for 30 min. The kinasing reaction is then added directly to the PCR master mix. As DTT can negatively affect the PCR, it is important to keep the volume of the kinase reaction (as well as the reverse transcription reaction) as small as possible. The

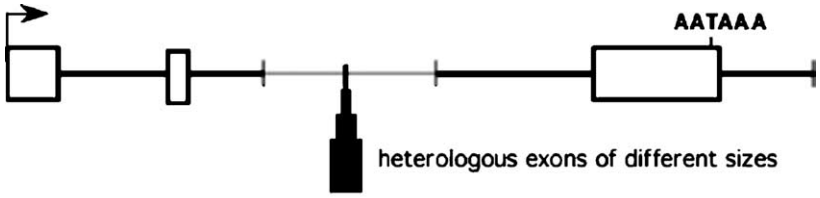
annealing temperatures and number of cycles in the PCR is determined empirically, however, we attempt to keep the cycle number below 21 cycles. Given that minigene mRNA is expressed at relatively high levels in cells that transfect efficiently, it should not be necessary to use high cycle numbers to detect minigene mRNA. The PCR products are separated on 5% non-denaturing polyacrylamide gels, which are dried down, and the bands are quantified using a PhosphorImager (Molecular Dynamics).

The ratios of the PCR products (amplicons) must be representative of the ratios of the different mRNA splice variants. Small differences in the efficiency of reverse transcription of the alternatively spliced mRNAs or differences in the efficiency of amplification of the different amplicons can result in mis-representation of the ratio of alternatively spliced mRNAs. These disparities between mRNA and PCR product ratios can be reduced using conditions within the linear range of amplification, limiting the analysis to no more than two splice variants per oligo pair, and since PCR amplification favors that smallest amplicons, by choosing amplicons that different in size by no more than 40%. A straightforward test is to remove aliquots of the PCR following a range of cycle numbers and then quantify the bands. The ratios of different bands remain constant within the linear range. RT-PCR can be optimized on in vitro transcribed RNAs derived from the different amplicons. For example, PCR products can be readily cloned into a plasmid containing the T7 RNA polymerase promoter using the Topo cloning kits (Invitrogen). Alternatively, the PCR product can be re-amplified using a forward primer that includes the short T7 promoter [17]. The RNAs are in vitro transcribed [18], quantified by UV absorbance, and then mixed at different ratios. These mixtures are then used to confirm that the ratio of the PCR products match the ratios of the input RNA.

Transient transfection assays are inherently variable due to differences in minigene and protein expression and the state of the cells. It is important to confirm the results of transient transfection assays in at least three independent transfections. The variability of the results is also determined and expressed as the standard deviation. In cases in which putative splicing regulators are co-expressed with the minigene, the level of protein expression should be confirmed by western blot analysis for each transfection. Transfecting increasing amounts of expression plasmid does not necessarily result in the expected range of protein expression levels and the levels of protein expression should be confirmed. In addition, transient transfection does not always recapitulate the regulation observed for the endogenous pre-mRNA [36]. Stable cell lines usually work well in these cases.

Results from minigene transfections can be expressed in different ways such as the percent spliced mRNA containing or lacking the variable region or as a ratio of mRNAs including and excluding the variably spliced region. We feel that the results are best expressed as the percent of mRNA including (or excluding) the variable region rather than as a ratio. Similarly, when comparing the splicing pattern in different RNA samples, we calculate the percent of mRNAs that include the variable region in different conditions (different cell types, with and without putative regulatory proteins) and use the arithmetic difference of the percent mRNAs that include the variable region to determine the extent of a splicing change. The reason for using arithmetic difference is that comparisons of inclusion:skipping ratios or of the fold change is strongly influenced by the basal level of exon inclusion which is usually more dependent on the intrinsic (and artificial) nature of the minigene construct than on a regulatory event. For example, if the basal level of exon inclusion is 50% (50% of mRNAs include the variable region), there can be no more than a 2-fold increase in exon inclusion. In contrast, if the basal level of exon inclusion is 5%, a 3-fold change would mean that 15% of the mRNAs contain the variable region. The splice variant is still only a minority of the total mRNAs expressed from the gene and a change of only ten percentage points is not a robust regulatory change and may or may not be biologically significant. A 3-fold change in which from 20 to 60% of the mRNAs include the variable region reflects a change of 40% points and a transition from being the minority splice variant to the majority variant. It is important to note that relatively small changes in the relative abundance of different splice variants and the proteins they encode can have significant biological impact [19], however, the point here is to emphasize the importance of expressing the results such that the composition of the mRNA population is apparent. When a splicing transition is expressed only as the fold change, it is not possible to evaluate the change in the relative abundance of different mRNA isoforms in the mRNA population.

Another factor leading us to use the arithmetic difference rather than a fold change is an experiment in which we modified a muscle-specific exon to have a range of exon sizes and therefore different levels of basal inclusion in fibroblasts and muscle cultures (Fig. 3). The difference in the levels of exon inclusion was then compared in fibroblasts and muscle cultures. Note that the *cis*-acting elements that regulate muscle-specific inclusion are located within the introns so that changes in the exon are not expected to have strong effects on the enhanced exon inclusion observed in muscle cultures [20]. This experiment reveals the responsiveness of the exon to fibroblast versus muscle at different basal levels of inclusion. Except for exons with relatively high or low basal levels of exon inclusion, the responsiveness of most exons was similar based on arithmetic change (Fig. 3). In contrast, the fold change varied and did so according to the basal level of exon inclusion rather than the actual change in the abundance of the mRNAs. For example, while R35(32) and R35(30) showed similar changes in mRNA abundance of 48 and 52% points, the fold change was higher for the construct with lower basal exon inclusion (4.0 versus 2.5). A construct showing only a medium change in mRNA abundance of 34 points [R35(28)], showed the highest fold change



clone name	exon size	F	M	Diff.	Fold
R35(46)	46	87	97	10	1.1
R35(37)	37	64	94	30	1.5
R35(33)	33	35	83	48	2.4
R35(32)	32	31	79	48	2.5
R35(30)	30	18	72	54	4.0
R35(28)	28	6	40	34	6.7

Fig. 3. Effects of exon size on the regulation of exon inclusion. Artificial exons of different sizes were inserted in place of chicken cardiac troponin T exon 5 within the context of the regulatory elements in the flanking introns. The percent of mRNAs containing the exon in fibroblasts (F) or muscle (M) cultures calculated as (mRNA + exon)/[(mRNA − exon) + (mRNA − exon)] × 100. Comparison in F and M are expressed as arithmetic difference (Diff.) and fold change (Fold) based on the percent exon inclusion (Fold). The results are taken from [21].

(6.7-fold). The results strongly suggested that the responsiveness of all but the largest exon in muscle cultures was similar despite different basal levels of inclusion in the fibroblasts and that expressing the arithmetic difference between different regulatory states is a better representation of the level of regulation than the fold difference.

A second observation that came from this experiment was that regulation appears to be less robust if the basal level of exon inclusion is high or low. When inclusion in fibroblasts was >60%, the change in muscle was smaller than for constructs with lower basal levels due to the restricted amount of change possible [R35(46) and R35(37), Fig. 3]. It is also appears that the exon with low basal inclusion had a blunted response as if in these transfection assays, it is difficult for exons with a poor basal recognition to be fully regulated, possibly due to a "weakened" state of overall recognition [R35(28), Fig. 3]. While not tested here, it should be considered that exons that are regulated primarily by repression might show a blunted response to repressors in minigenes with a high basal level of exon recognition.

A third point from this experiment is that the features of an alternative exon that establish its basal level of recognition can be "adjusted" to enable the maximal level of response to mutations, cell-specific regulation, or co-expressed regulatory factors. We prefer a basal level of 30–40% exon inclusion in the absence of co-expressed regulatory protein or in the non-specific cell type to observe activation of exon inclusion. To modulate basal levels of exon inclusion, we typically modify one of two features: the 5′ splice site or the size of the alternative exon. Since there are rarely regulatory elements within 5′ splice sites, we most often modify positions 3 and 4 of the intron to weaken or strengthen the 5′ splice site. To do this we use "doped" oligos in which more than one nucleotide is present in each position to be modified so that several mutants can be generated from a single cloning step. Individual mutants are then tested by transfection to identify those with the preferred level of exon inclusion. In cases in which intronic regulatory elements are being tested independent of exonic elements, exon size can be modified to increase or decrease basal levels of exon inclusion as in Fig. 3.

3.5. Analysis of results

The five primary uses of minigenes listed at the beginning of this chapter differ with regard to the question to be addressed, experimental setup, and analysis of results. Parameters for four of these approaches are considered separately in this section. Strategies for the fifth use of minigenes (to identify cell-specific regulatory elements) has been described previously [13].

3.6. Determine the role of the splice sites in establishing a basal level of exon recognition

The question typically being addressed in these types of experiments is: "why is this exon not constitutively spliced?" All exons, constitutive and alternative, have multiple elements that contribute to splicing efficiency. In early investigations of alternative splicing, minigenes were used to demonstrate that the sequence of the splice sites as well as the size of the exon have a profound effect on exon recognition [5]. In the case of the chicken cardiac troponin T alternative exon 5, the fifth nucleotide of the intron is a non-consensus "T" which is found in less than 5% of 5′ splice sites. Mutation of this one nucleotide to the consensus "G" resulted in constitutive splicing demonstrating that this nucleotide was a critical feature for establishing exon skipping [5]. In the case of exons that are regulated in a cell-specific manner, such analyses identify the basis for exon skipping but do not identify the basis for cell-specific modulation of exon use. That is, a mutation in the 5′ splice site that improves the match to the consensus and induces constitutive inclusion does not indicate that the 5′ splice

site is the target for cell-specific regulation of the alternative exon. It is quite possible, for example, that while the weak 5′ splice site contributes to inefficient exon recognition, a regulatory factor that binds to an exonic or intronic element enhances inclusion by promoting stronger recognition of the 3′ splice site.

As noted above in Fig. 3, exon size has also been shown to have a strong effect on the efficiency of exon recognition. Several studies demonstrated that increasing the size of an exon, independent of exonic sequence, enhance exon recognition [22–24]. In one study, the exon was replaced by a restriction enzyme site and artificial exons with a range of sizes were generated by inserting fragments from plasmid DNA. Each fragment was tested in both orientations. While exon sequence from the different orientations had some effects, the general trend was that the larger fragments resulted in increased exon inclusion [22].

3.7. Identify exonic and intronic elements that enhance or repress splicing

There is a growing appreciation of the abundance and diversity of ESEs and ESSs within the human genome [8,9,11,25]. Minigenes are useful tools to identify and demonstrate the activities of exonic and intronic elements in vivo. An analysis of ESEs should begin with the use of biocomputational tools such as ESEFinder [26] and RESCUE-ESE to identify putative elements [27]. Similarly, resources for ESSs are growing [8]. Biocomputational resources for intronic elements has not progressed as far as for exonic elements, however, there are several resources [28].

To test whether a motif within an exon functions to enhance or suppress exon inclusion, results from both loss of function and gain of function mutations analyses provide the strongest support. Loss of function analyses are performed by introducing mutations into the putative element. Since substitutions introduced into an element can have effects on splicing independent of the sequence being mutated, it is best to test more than one substitution within a putative element to confirm that the nucleotides being tested are required for splicing activity. A gain of function experiment provides particularly strong evidence for a splicing enhancer or repressor. Typically, the segment containing the putative enhancer or repressor is placed within the context of a heterologous exon. Because a change in exon size can have a strong impact on the efficiency of exon inclusion, it is important that when exonic elements are being tested, exons with and without the putative exonic element are identical in size. Fragments containing the unmodified and mutated elements are compared. We use a modified version of a minigene designed by Dominski and Kole [23] in which the middle exon of a three exon construct contains unique restriction sites for the insertion of heterologous sequences. This minigene has been used to test ESEs identified by loss of function studies [22], for in vivo selection of sequence motifs that enhance exon inclusion [29], as well as to test ESEs identified biocomputationally [9].

The factors that bind to the exonic elements can be identified based on previously determined binding preferences of proteins known to bind to exonic elements such as SR proteins [6]. When it is of interest to identify the factors that bind to the *cis*-acting element, the functional analyses should be performed such that in vivo activity can be correlated with in vitro analysis of RNA binding proteins that are assayed using gel shift analysis or UV crosslinking analyses [30]. A useful strategy is to generate a series of mutations with a range of in vivo activities and then test for correlations between in vivo activity and binding affinity of a specific protein. This approach was used to demonstrate binding of SR proteins to the ESE within cTNT exon 5 [30] and YB-1 binding to the A/C-rich ESE identified by in vivo SELEX [7]. While inactivation of a splicing element is relatively straightforward, it is more challenging to increase the activity of an element. One approach is to identify to key components of the active sequence motif and increase the representation of that sequence motif by multimerization.

3.8. Determine whether an allelic variant has an effect on splicing efficiency

Prior to the identification of ESEs and ESSs, when mutations within splice sites were the only known disease-causing mutations that affected splicing, it was predicted that 15% of point mutations that caused disease did so by disrupting splicing [31]. Now that the roles of intronic and exonic elements have been better defined, current estimates for some genes are that up to 50% of disease-causing mutations disrupt normal splicing [6]. In the search for disease-causing mutations, nucleotide differences are identified in candidate genes of affected individuals compared to unaffected family members. It is often unclear whether a nucleotide substitution is a disease-causing mutation or an allelic variant not associated with disease. In addition, once disease-causing mutations are identified, unless the substitutions clearly disrupt the consensus splice sites, it is also often unclear whether substitutions within or near exons could be affecting splicing efficiency. For example, missense mutations that are linked with disease are often assumed to disrupt the function of the protein and the altered amino acids are assumed to be key to protein function. It is now clear that many of the exonic mutations once thought to disrupt protein function actually disrupt splicing of the exon that contains the substitution [6,32]. The best approach to determine whether a nucleotide substitution or allelic variant affects splicing is to assay splicing of the endogenous RNA from the relevant tissue of affected individuals. Because tissue samples from affected individuals are often not available, the alternative approach is to compare splicing of the variant and wild-type alleles in side-by-side analyses using minigenes. In designing minigenes to test for sequence variants within exons, it is best

to include the exon and its flanking splice sites including at least 70 nucleotides of the upstream and downstream flanking introns. Similarly, variants within introns should be tested in the context of the nearest exon(s). Mutations within exons or splice sites most often result in exon skipping, however, use of cryptic splice sites within the exon or flanking introns are also common [6]. While rare, retention of introns is also observed. To enhance the likelihood that an aberrant splicing pattern is detected, large genomic segments can be tested containing the exon and several hundred nucleotides of the upstream and downstream flanking introns.

The identical genomic segments containing the wild-type and variant alleles are amplified from genomic DNA by PCR and cloned into the minigene. The PCR-amplified genomic segments need to be sequenced to confirm that they are identical except for the variations to be tested.

3.9. Identify elements required for regulation by specific trans-acting factors

Minigenes are particularly useful for demonstrating that individual splicing events are affected by the loss and gain of activities of specific regulators. Furthermore, once a protein has been shown to regulate a splicing event, a combination of in vivo analyses and in vitro RNA binding assays can be used to identify the specific element(s) required for responsiveness and to correlate responsiveness to binding of the protein to the pre-mRNA. Typically, minigenes are first tested by transient expression with epitope-tagged versions of the protein of interest to determine whether there is an effect on splicing. Expression of the protein is confirmed by western blot analysis. A caveat with this approach is the possibility that overexpressing an RNA binding protein to unnaturally high intracellular concentrations will have general effects on splicing. One can imagine that at sufficiently high concentrations, an RNA binding protein will bind to low affinity non-natural binding sites within the minigene pre-mRNA and modulate splicing. Therefore, it is not desirable to express large amounts of exogenous protein to monitor splicing effects. Another caveat is that in the cells used for transient transfection, the endogenous activity of the protein could already be at maximum levels so that adding additional protein will have no significant effect. One solution is to perform transient transfection experiments in at least two different cell lines to confirm either that there are or are not effects on splicing.

In cases of cell-specific regulation in which different splice variants are expressed in different cell cultures, different states of the same cell culture, and in response to changes in protein steady state levels (overexpression or depletion), it is possible that a change of mRNA stability of one variant rather than splice site selection is determinative for the change in mRNA ratios. To address this possibility, each of the cDNA PCR products for the splice variants derived from RT-PCR are cloned into the minigene expression plasmid. The plasmids expressing the different spliced mRNA variants are mixed at different ratios for transient transfection (for example, for two splice variants, typically at 1:3, 1:1, and 3:1). RT-PCR analysis of the RNA expressed from these transfections should reflect the ratios of the input plasmid DNAs. If splice site choice and not differential mRNA stability is determinative for the change in the ratio of mRNAs, the ratio of PCR products should be unchanged under the different conditions being tested.

RNAi-mediate depletion experiments are an important complement to overexpression studies. There are several considerations for RNAi depletion experiments. First, the cell line must express the protein of interest and ideally not express proteins with potentially redundant activity. Second, the efficiency of depletion must be assayed, preferably by western blot and/or immunostaining of the endogenous protein. If an antibody is not available, mRNA expression can be monitored by northern blot or quantitative RT-PCR. We have had success in HeLa cell cultures using 21 nucleotide duplexes from Dharmacon designed using the Dharmacon siDESIGN program (www.dharmacon.com). For siRNA-mediated depletion of muscleblind-like 1 (MBNL1) in HeLa cell cultures, cells were plated at 300,000 cells per well in a six well plate in DMEM plus 10% fetal bovine serum. Twelve hours after plating, the media are replaced with serum-free media (DMEM alone) and 2.66 μg siRNA duplex is transfected using Oligofectamine (Invitrogen, Carlsbad, CA). One milliliter of serum-containing media (DMEM supplemented with 30% FBS) is added after 4 h. After another 12 h, the media is replaced with antibiotic-free media and the cells are transfected with 0.5 μg of minigene plasmid DNA and 2.66 μg of siRNA duplex using Lipofectamine 2000 (Invitrogen, Carlsbad, CA). The media were exchanged with antibiotic-free media containing 10% fetal bovine serum 6 h later. RNA and protein are harvested 48 h after the minigene is transfected. SiRNA duplexes complementary to non-endogenous proteins such as luciferase or GFP are used to control for the general effects of siRNA transfection. At a minimum, decreased protein or RNA expression from the siRNA target gene and the effects of depletion on splicing need to be observed with at least two independent siRNA duplexes which target different regions of the mRNA. Another often-cited criterion is that splicing in siRNA-depleted cells should be restored by expression of the protein or a paralogue not affected by the siRNA treatment. It should be noted, however, that if overexpression of the protein is known to affect splicing in the absence of siRNA depletion, this experiment is likely to affect splicing of the minigene in the expected manner even if siRNA treatment affects splicing due to effects unrelated to the targeted protein. This is why a true knockdown approach should attempt to restore the level of regulatory protein to pre-knockdown levels rather than overexpression.

For both overexpression and depletion studies, it is important to examine splicing of the endogenous gene.

Response of the endogenous alternative splicing event to depletion of the protein of interest is particularly strong evidence for regulation by the protein. On the other hand, it has been difficult to demonstrate regulation of endogenous splicing events by overexpression of particular proteins. One major obstacle in these studies is transfection efficiency since non-transfected cells contribute a "background" signal of the splicing pattern in the absence of the overexpressed protein that reduces the overall measurable response. In contrast to plasmid transfection protocols, siRNA transfection and depletions can be highly efficient, depending on the cell cultures used. The non-regulated background patterns also results from pre-exiting pools of mRNAs synthesized and processed before the expression of transfected factors or siRNAs. The level of this background depends upon the half-lives of the processed RNAs and the time interval between treatment of the cultures and harvest of the RNA. Demonstrating responsiveness of the endogenous alternative splicing event to modulation of protein levels establishes strong support for regulation. Minigenes can be used to confirm this regulation, map the element(s) required for responsiveness to protein levels, and demonstrate a correlation between responsiveness and protein binding to the pre-mRNA.

Once a minigene has been confirmed as responsive to overexpression and depletion of a particular protein, the same transient transfection assay can be used to identify the region of the minigene required for this response. While overexpression or depletion experiments demonstrate responsiveness, they do not demonstrate that the protein is mediating the response by binding directly to the pre-mRNA. To establish direct regulation, it is necessary to identify the binding site(s) within the pre-mRNA and to demonstrate that binding correlates with the ability to affect splicing. The goal is to first identify a minimal element required for a splicing response and then to determine whether binding of the protein to this site in vitro correlates with activation of splicing. If putative binding sites for the protein of interest are known, a reasonable starting point is to introduce nucleotide substitutions to disrupt binding and compare the responsiveness of wild-type and mutant minigenes to changes in protein expression. If binding sites are not known, a systematic deletion analysis of upstream and downstream introns is performed. Deletions are generated simply by placement of oligonucleotides for PCR amplification of a sequentially smaller genomic segment using the cloned genomic segment as the template. Once regions required for responsiveness are defined by deletion analysis, point mutations can be introduced to identify the specific regions required for responsiveness to the overexpression and siRNA-mediated depletion of the protein. Experiments to map response elements to specific proteins are best performed in parallel with in vitro RNA binding assays to map the preferred binding site. UV crosslinking assays are the most straightforward and are typically performed using recombinant protein expressed in bacteria or baculovirus [30]. However, we have performed UV crosslinking experiments using protein transiently expressed in eukaryotic cells [33]. This is particularly useful for proteins that are insoluble in bacterial or baculovirus expression systems.

4. Closing comments

The main strengths of the minigene approach are the abilities to demonstrate that specific nucleotide changes affect splicing efficiency and to define elements required for responsiveness to cell type and specific splicing regulators. It is important, however, to be cognizant of the caveats of this approach when designing experiments. First, the context of the exon within the minigene influences basal recognition in a way that cannot be predicted. The flanking exons and their adjacent intronic segments can influence the basal recognition of the alternative exon under study. A related caveat is that mutations that affect splicing of an exon in one context might not affect splicing in another context, and in particular, not in the context of the full-length pre-mRNA. Therefore, it is possible that splicing "elements" defined by mapping within a minigene context might have no consequence in the context of the full-length pre-mRNA expressed under its own promoter. In addition, splicing efficiency is promoter dependent, due primarily due to the processivity of elongation by RNA polymerase II [34]. It is now clear that splicing (as well as other RNA processing events) is linked with transcription. In particular, different promoters have different effects on basal exon recognition. On the whole, however, minigenes can be effective tools to demonstrate that particular nucleotide changes result in altered splicing patterns, identifying cell-specific regulatory elements [6].

References

[1] L.F. Lareau, R.E. Green, R.S. Bhatnagar, S.E. Brenner, Curr. Opin. Struct. Biol. 14 (2004) 273–282.
[2] D.L. Black, Annu. Rev. Biochem. 27 (2003) 27–48.
[3] H.J. Mardon, G. Sebastio, F.E. Baralle, Nucleic Acids Res. 15 (1987) 7725–7733.
[4] R.K. Hampson, L.L. Follette, F.M. Rottman, Mol. Cell. Biol. 9 (1989) 1604–1610.
[5] T.A. Cooper, C.P. Ordahl, Nuc. Acids Res. 17 (1989) 7905–7921.
[6] L. Cartegni, S.L. Chew, A.R. Krainer, Nat. Rev. Genet. 3 (2002) 285–298.
[7] E. Stickeler, S.D. Fraser, A. Honig, A.L. Chen, S.M. Berget, T.A. Cooper, EMBO J. 20 (2001) 3821–3830.
[8] Z. Wang, M.E. Rolish, G. Yeo, V. Tung, M. Mawson, C.B. Burge, Cell 119 (2004) 831–845.
[9] W.G. Fairbrother, R.F. Yeh, P.A. Sharp, C.B. Burge, Science 297 (2002) 1007–1013.
[10] M. Sironi, G. Menozzi, L. Riva, R. Cagliani, G.P. Comi, N. Bresolin, R. Giorda, U. Pozzoli, Nucleic Acids Res. 32 (2004) 1783–1791 (Print 2004).
[11] X.H. Zhang, L.A. Chasin, Genes Dev. 18 (2004) 1241–1250.
[12] E.J. Wagner, M.A. Garcia-Blanco, Mol. Cell. Biol. 21 (2001) 3281–3288.
[13] T.A. Cooper, in: S.R. Haynes (Ed.), Methods in Molecular Biology, Vol. 118, Humana Press, Totowa, NJ, 1999, pp. 391–403.
[14] R.S. Savkur, A.V. Philips, T.A. Cooper, Nat. Gen. 29 (2001) 40–47.
[15] S. Kawamoto, J. Biol. Chem. 271 (1996) 17613–17616.

[16] B.K. Dredge, R.B. Darnell, Mol. Cell. Biol. 23 (2003) 4687–4700.
[17] R. David, D. Wedlich, Biotechniques 30 (2001) 769–774.
[18] J.K. Yisraeli, D.A. Melton, Methods Enzymol 180 (1989) 42–50.
[19] D.A. Buchner, M. Trudeau, M.H. Meisler, Science 301 (2003) 967–969.
[20] K.J. Ryan, T.A. Cooper, Mol. Cell. Biol. 16 (1996) 4014–4023.
[21] T.A. Cooper, Mol. Cell. Biol. 18 (1998) 4519–4525.
[22] R. Xu, J. Teng, T.A. Cooper, Mol. Cell. Biol. 13 (1993) 3660–3674.
[23] Z. Dominski, R. Kole, Mol. Cell. Biol. 11 (1991) 6075–6083.
[24] D.L. Black, Genes Dev. 5 (1991) 389–402.
[25] W.G. Fairbrother, L.A. Chasin, Mol. Cell. Biol. 20 (2000) 6816–6825.
[26] L. Cartegni, J. Wang, Z. Zhu, M.Q. Zhang, A.R. Krainer, Nucleic Acids Res. 31 (2003) 3568–3571.
[27] W.G. Fairbrother, G.W. Yeo, R. Yeh, P. Goldstein, M. Mawson, P.A. Sharp, C.B. Burge, Nucleic Acids Res. 32 (2004) W187–W190.
[28] A.N. Ladd, T.A. Cooper, Genome Biol. 3 (2002) 8.1–8.16.
[29] L.R. Coulter, M.A. Landree, T.A. Cooper, Mol. Cell. Biol. 17 (1997) 2143–2150.
[30] J. Ramchatesingh, A.M. Zahler, K.M. Neugebauer, M.B. Roth, T.A. Cooper, Mol. Cell. Biol. 15 (1995) 4898–4907.
[31] M. Krawczak, J. Reiss, D.N. Cooper, Hum. Genet. 90 (1992) 41–54.
[32] N.A. Faustino, T.A. Cooper, Genes Dev. 17 (2003) 419–437.
[33] N. Charlet-B, P. Logan, G. Singh, T.A. Cooper, Mol. Cell. 9 (2002) 649–658.
[34] A.R. Kornblihtt, Curr. Opin. Cell Biol. 17 (2005) 262–268.
[35] E.J. Wagner, M.L. Curtis, N.D. Robson, A.P. Baraniak, P.S. Eis, M.A. Gracia-Blanco, RNA 9 (2003) 1552–1561.
[36] D. Libri, J. Marie, E. Brody, M.Y. Fiszman, Nucleic Acids Res. 17 (1989) 6449–6462.