

Parra *et al.*, 2007

# Intrasplicing Coordinates Alternative First Exons with Alternative Splicing in the Protein 4.1R Gene

Marilyn K. Parra<sup>1\*</sup>, Jeff S. Tan<sup>1\*</sup>, Narla Mohandas<sup>2</sup>, and John G. Conboy<sup>1,†</sup>

<sup>1</sup> Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road,  
Mail stop 74-217, Berkeley, CA 94720

<sup>2</sup> Red Cell Physiology Laboratory, New York Blood Center, New York, NY 10021

† Corresponding author: Building 74, Room 225 (MS 74-217), 1 Cyclotron Road,  
Berkeley, CA 94720, USA

\* The first two authors contributed equally to this work.

Tel: 510-486-6973

Fax: 510-486-6746

Email: [jgconboy@lbl.gov](mailto:jgconboy@lbl.gov)

Running Title: Intrasplicing Regulation of Alternative Splicing

Number of characters: 56,474 (including spaces)

Number of figures: 9

Number of tables: 0

Subject Category: RNA

Supporting grant: NIH grant DK32094 to N.M.

## Abstract

In the protein 4.1R gene, alternative first exons splice differentially to alternative 3' splice sites far downstream in exon 2'/2 (E2'/2). We describe a novel intrasplicing mechanism by which exon 1A (E1A) splices exclusively to the distal E2'/2 acceptor via two nested splicing reactions regulated by novel properties of exon 1B (E1B). E1B behaves as an exon in the first step, using its consensus 5' donor to splice to the proximal E2'/2 acceptor. A long region of downstream intron is excised, juxtaposing E1B with E2'/2 to generate a new composite acceptor containing the E1B branchpoint/pyrimidine tract and E2 distal 3' AG-dinucleotide. Next, the upstream E1A splices over E1B to this distal acceptor, excising the remaining intron plus E1B and E2' to form mature E1A/E2 product. We mapped branchpoints for both intrasplicing reactions and demonstrated that mutation of the E1B 5' splice site or branchpoint abrogates intrasplicing. In the 4.1R gene, intrasplicing ultimately determines N-terminal protein structure and function. More generally, intrasplicing represents a new mechanism whereby alternative promoters can be coordinated with downstream alternative splicing.

Keywords: Alternative first exons/ alternative splicing/ intrasplicing/ Protein 4.1R

## Introduction

Alternative pre-mRNA splicing is a critical regulator of eukaryotic gene expression that enables a relatively small number of genes to encode a much larger proteome. Moreover, many alternatively spliced genes are regulated in a tissue-specific manner, so that the structure and function of the encoded protein isoforms can be tailored to the requirements of various differentiated cell types. Biochemical and computational studies have identified a number of *trans*-acting splicing factors, and *cis*-acting regulatory elements in the pre-mRNA, that work together to carry out specific alternative splicing programs during normal development. Although much remains to be learned about alternative splicing regulation, recent studies document significant progress in deciphering the splicing “code” that regulates the splicing efficiency of alternative exons. Several recent reviews summarize the state of the field in terms of understanding regulation in normal development, and in terms of the impact of aberrant splicing regulation in human disease (Black, 2003; Blencowe, 2006; Faustino and Cooper, 2003; Matlin *et al.*, 2005).

It is increasingly clear that many eukaryotic genes also possess multiple transcriptional promoters associated with alternative first exons (Baek *et al.*, 2007; Cooper *et al.*, 2006; Kim *et al.*, 2005; Tan *et al.*, 2006; Denoeud *et al.*, 2007), a phenomenon that may be as common as alternative splicing of internal exons. Often these alternative first exons are located far upstream of the coding exons (Hong *et al.*, 2006), raising questions as to how very long introns are accurately spliced. Moreover, there is considerable interest in understanding to what extent these 5' transcription decisions influence downstream alternative splicing events. Much has been written about the coupling of transcription and splicing processes (Bird *et al.*, 2004; Das *et al.*,

2006; Dye *et al.*, 2006; Hirose and Manley, 2000; Listerman *et al.*, 2006; Maniatis and Reed, 2002), and a few studies have directly demonstrated that promoter identity can influence downstream alternative splicing (Auboeuf *et al.*, 2004; Cramer *et al.*, 1999; Cramer *et al.*, 1997; Kornblihtt, 2005; Pagani *et al.*, 2003). However, these processes remain poorly understood.

The protein 4.1R gene (EPB41) provides an interesting model system in which to explore some of these issues mechanistically. In the human genome, this complex gene spans ~240kb and is comprised of 22 exons, at least 10 of which can be alternatively spliced (Conboy *et al.*, 1991; Huang *et al.*, 1993; Parra *et al.*, 2004; Schischmanoff *et al.*, 1997; Baklouti *et al.*, 1997). The 5' region of the gene contains three mutually exclusive, alternative first exons designated E1A, E1B, and E1C (Figure 1A). These exons are located far upstream of the coding exons (73-100kb) and possess their own transcriptional promoters (Parra *et al.*, 2003). Most importantly, these first exons exhibit a unique and strictly observed splicing paradigm: transcripts initiated at the 5'-most exon (E1A) always splice to the distal 3' acceptor located within exon 2 (E2), while E1C transcripts always choose the proximal one only 17 nucleotides (nt) upstream (Figure 1B). This splicing decision determines N-terminal protein structure, since translation start site AUG1 is encoded in the alternatively spliced portion of E2. Transcripts including this sequence encode larger isoforms of 4.1R protein containing a long N-terminal headpiece, while those excluding it initiate translation farther downstream at AUG2 located in E4 (Figure 1C). The latter mRNAs encode smaller ~80kDa isoforms of protein that differ significantly in their affinity for binding partners in the red cell membrane (Nunomura *et al.*, 2002). Given this gene structure, cells can vary the ratio of N-terminal isoforms by controlling transcription at the upstream promoters.

We are studying the molecular mechanism that coordinates upstream transcription events with downstream splicing events in the 4.1R gene. Some of the key questions to be addressed

include: How does E1A selectively choose the distal 3' acceptor when there is another one only 17 nt upstream? How is this event regulated across long distances? Since the splicing pattern is conserved across several vertebrate orders and is independent of tissue type (Parra *et al.*, 2003; Tan *et al.*, 2005), it appears not to require any cell type specific splicing factors. In this study, we present data indicating that E1A splicing is coordinated via a two-step mechanism utilizing an exon-like element that has an AG-independent upstream intron and exhibits both exonic and intronic features. Because the pathway involves nested splicing reactions first removing a large internal region of the intron, followed by resplicing over the same region using external splice sites to remove the remaining intron, the model resembles the “intrasplicing” mechanism proposed earlier (Ott *et al.*, 2003). The intrasplicing model is supported by a series of experiments with minigene constructs that recapitulate intrasplicing in a manageable system and allow for analysis of essential regulatory motifs. This analysis reveals that intrasplicing is distinct from recursive splicing, another mechanism for stepwise removal of long introns (Burnette *et al.*, 2005; Hatton *et al.*, 1998).

## Results

### **Accurate E1A splicing to E2'/2 requires a downstream element mapping to the E1B region.**

The proximal 3' splice acceptor site at E2'/2 is preceded by a typical pyrimidine tract and is predicted to function as a strong acceptor (Figure 1B); indeed, the proximal site is selected exclusively in all single-intron constructs examined to date (see below). The distal site, only 17nt downstream, is preceded by purines that apparently weaken its function. To study the molecular mechanism by which E1A selectively splices to the distal 3' splice site ~100 kb downstream in E2'/2, we engineered a series of minigenes that recapitulate this coordinated splicing process in a

cell culture system (Figure 2). Construct pE1A-1 used the natural E1A promoter to drive expression of a small model pre-mRNA containing only E1A, E2'/2, and a few hundred nucleotides of flanking intron sequence. This minimal construct was spliced incorrectly in transfected cells: E1A was joined exclusively to the proximal 3' acceptor of E2 to generate a product that is slightly larger (Figure 3, lane 1) than the expected normal product due to the inclusion of E2'. Multiple attempts at reconstituting proper E1A splicing to the distal 3' splice site were unsuccessful until sequences extending ~7 kb downstream of E1A were included in the minigene (construct pE1A-2L). This longer construct extended past exon 1B (E1B), a previously annotated alternative first exon that is rarely expressed among 4.1R clones in the genetic databases. E1A splicing in this context bypassed the proximal acceptor and instead was completely switched to the appropriate distal acceptor in E2'/2 (Figure 3, lane 2). This result suggested that element(s) several kilobases downstream of E1A, perhaps overlapping with E1B, are powerful determinants of its downstream splicing choice(s).

In order to define the active element that promotes accurate E1A splicing, we systematically assayed a series of constructs spanning various portions of the intronic region between E1A and E1B. The most informative of these are diagrammed in Figure 2 (top panel), and assayed for splicing in Figure 3. Ultimately, construct pE1A-2 revealed that a specific 200 nt fragment was necessary and sufficient for proper E1A splicing (lane 3). This element mapped ~6.5 kb downstream of E1A and encompassed E1B. DNA fragments of similar length, but shifted 100nt upstream or downstream, did not rescue proper E1A splicing (constructs pE1A-2a and pE1A-2b; lanes 4 and 5). These results suggested that multiple elements in the vicinity of E1B, including its 5' splice site and sequences  $\geq 100$ nt upstream, are required for proper splicing of E1A.

**Proper E1A splicing does not require specific transcriptional regulatory elements but does require the E1B 5' splice site.**

Transcriptional promoters can exert a significant influence on downstream splicing events (Kornblihtt, 2005). To determine whether the natural E1A promoter region (Parra *et al.*, 2003) is essential for correct downstream splicing, we tested E1A splicing in constructs expressed from several exogenous promoters. Minigenes that retained the E1B-associated splicing regulator, but were transcribed from a cytomegalovirus (CMV) promoter, an ankyrin promoter (Gallagher *et al.*, 2000), or the natural 4.1R E1C promoter, all mimicked E1A splicing to the distal acceptor site at E2'/2 (Figure 4A, lanes 2 and 3). Because the latter two constructs also substituted a heterologous first exon for 4.1R-E1A, there appears to be no specific primary sequence requirement in the first exon. Together these results suggest that specific sequences in the proximal promoter and /or the first exon are not sufficient to determine downstream splicing preferences at E2'/2. Additional experiments demonstrated further that the E1B-associated regulatory element does not possess transcriptional enhancer activity when tested in a luciferase reporter construct (data not shown). Therefore, neither proximal transcriptional promoter elements nor distal transcriptional enhancer elements can account for E1A splicing specificity. Instead, it appears that the E1B region is the critical determinant that appears capable of mediating appropriate splicing independent of the identity of the upstream promoter / first exon.

We next considered the possibility that the active element encompassing E1B might function via a splicing-based mechanism to promote proper E1A splicing. To test whether the annotated E1B 5' splice site plays a role in this process, we compared E1A splicing patterns in pre-mRNA constructs that contained either a normal (G/gtgagt) or a mutated (G/atgaat) 5' splice site at E1B. Figure 4B shows that abolishing the 5' splice site disrupted the proper splicing of E1A, causing it

to switch predominantly to use of the proximal acceptor site at E2 to generate a PCR product that is larger (lane 3) than the correct product (lane 2). Control experiments revealed that a cryptic 5' splice site (G/gtttga), located 10 nt downstream of the endogenous E1B 5' splice site, could substitute for the natural 5' splice site (data not shown). Together these results indicated that a functional 5' splice site in the E1B region is essential for coupling of E1A to the distal acceptor site at E2, and led us to propose the splicing model presented below.

### **The two-step intrasplicing hypothesis**

Based on similarities in splice site arrangement of the 4.1R gene 5' region to the intrasplicing model proposed earlier on hypothetical grounds (Ott *et al.*, 2003), we proposed a two step splicing mechanism that incorporates the requirement for the E1B-regulatory region. As shown in Figure 5A, this region of the 4.1R pre-mRNA contains two pairs of splice sites: an “internal” set comprising E1B 5' (SD1) and the proximal 3' acceptor site at E2 (SA1), and an “external” set represented by E1A 5' (SD2) and the distal 3' splice site at E2 (SA2). Sequential activation of the internal and then external splice site pairs can remove the entire intron in two steps and insure selection of SA2 in mature mRNA. A unique and essential feature of the model is the regulatory element that drives this process: it has specific features of both exons and introns that allow it to function in both capacities at different steps of intrasplicing. For that reason we designate this element as an “intraexon”; in the 4.1R gene this element has been annotated as a first exon, but its major function may be in intrasplicing.

In the first step of the reaction, we hypothesize that E1B is recognized as an exon and uses its strong 5' splice site to join E1B and E2'2 via splice sites SD1 and SA1. This reaction creates a lariat structure comprising the intronic region between E1B and E2, as well as a splicing intermediate with the overall structure E1A—intron—E1B/E2'2. The partially spliced



intermediate RNA no longer possesses the internal splice site pair (SD1 and SA1) or the natural branchpoint sequence upstream of E2, and the architecture of both E1B and E2'/2 has been completely altered by their juxtaposition to form a composite 3' splice site for the second step of intrasplicing. That is, the branchpoint and pyrimidine region of E1B are now coupled with the distal AG acceptor site of E2'/2, and the intraexon 1B behaves like an intron to promote splicing of E1A (SD2) to the remaining 3' acceptor in E2'/2 (SA2). Accordingly, the intraexon contributes a new branchpoint sequence to which the E1A 5' splice site loops during lariat formation.

To identify candidate regulatory motifs whose function could be experimentally assessed in splicing assays, we examined the primary nucleotide sequences of the E1B region (Figure 5B, upper panel). In addition to its consensus 5' splice site, E1B is preceded by intronic sequence that contains a candidate branchpoint, followed by a modest pyrimidine rich sequence (17/24 pyrimidines) and two nonconsensus 3' splice site dinucleotides TG and GG with no AG-dinucleotides in the vicinity. The observation of two rare EST clones that exhibit utilization of these nonconsensus splice sites (BP364691 and BP315748) supports the notion that E1B can be recognized in an AG-independent manner to promote its splicing to E2'/2. After the first step of intrasplicing, E1B is then juxtaposed with E2' (shaded) to generate the sequence shown in the lower panel of Figure 5B. In this new structure, a novel composite 3' splice site is composed of the branchpoint and pyrimidine tract of E1B and the distal AG-dinucleotide from E2'/2. The predicted branchpoint would now be relatively far upstream of the 3' splice site (~111nt), an unusual but not unprecedented arrangement (Gooding *et al.*, 2006). Indeed, like other introns with distant branchpoints, the E1B region has an AG-exclusion zone that extends from the putative branchpoint to the distal E2 acceptor.

### Experimental analysis of the intrasplicing model

The intrasplicing model proposes that specific motifs associated with E1B, E2'/2, and E1A drive a two step process that proceeds through an obligatory intermediate RNA in which intraexon 1B is joined to E2'/2. This structure can then serve as a direct precursor of the mature product by re-splicing at the external splice site pair SD2 and SA2 so as to join E1A to E2. Furthermore, both steps of intrasplicing should generate specific lariat-branchpoint structures representing the excised intronic sequences. Our experimental analysis of the intrasplicing model focused on testing these predictions by (1) determining whether a splicing intermediate does exist and whether it can function as a direct precursor of mature product, (2) detecting the predicted lariat-branchpoint structures from both steps of intrasplicing, and (3) testing the function of E1B branchpoint mapped in (2).

Initial attempts to detect the splicing intermediate in endogenous cellular RNA utilized mouse erythroblast RNA as a template for RT-PCR, because this cell type is known to be a rich source of 4.1R transcripts. In these experiments we could detect products up to ~3.2kb that include much of the intron between E1A and intraexon 1B, and exhibit E1B already spliced to E2'/2, exactly as expected. However, full length intermediate failed to amplify, most likely due to the large transcript size (~5kb in mouse) and low abundance of this structure. As an alternative approach, we turned to the minigene system in which constructs such as pE1A-2 should yield smaller intermediates that would be easier to detect by RT-PCR. Indeed, amplification of pE1A-2 splicing products with primers F1 and R2 generated two predominant products, the unspliced precursor and the expected intermediate (Figure 6A). DNA sequence analysis confirmed the identity of these bands. It is also worth noting that this PCR assay should also be able to detect aberrant products of direct E1A splicing to the default E2 acceptor; the absence of such products

reinforces the notion that intraexon E1B is a strong regulator and its presence prevents direct splicing of E1A to E2'/2.

To directly test the functionality of this intermediate RNA, we tested whether it could serve as a precursor to mature E1A/E2 product in transfected cells. Figure 6 shows that mouse 3T3 cells, transfected with human construct E1A-2Lint as the sole source of mature product, indeed were able to undertake the hypothesized second step of the reaction to create mature E1A/E2 product (Figure 6B, lane 3) identical to that obtained by processing of model pre-mRNA pE1A-2 (lane 2). No such product was detected in mock transfected cells (lane 4). These findings demonstrate that a pre-spliced model of the predicted endogenous intermediate RNA can splice in a manner consistent with its proposed role in the intrasplicing pathway for production of mature E1A/E2 RNA.

Detection of predicted lariat structures from both stages of intrasplicing was performed using an experimental scheme that facilitates PCR amplification across branchpoint sequences in lariat structures (Figure 7A; Vogel *et al.*, 1997). This approach has been used previously to map branchpoints for intermediate steps of recursive splicing pathways (Burnette *et al.*, 2005; Hatton *et al.*, 1998). The first step of intrasplicing was hypothesized to involve the E1B 5' splice site looping back to a branchpoint upstream of E2, while the second step was predicted to employ the E1A 5' splice site looping to an E1B-associated branchpoint. For both steps, nested PCR reactions successfully amplified candidate branchpoint structures that were subsequently confirmed by DNA sequence analysis (Figure 7B, lanes 2 and 3). As expected, the E1B 5' splice site was joined to a branchpoint located at a fairly typical location (35 bp upstream of the default E2 acceptor site), with the sequence TGCTTA\*T (asterisk denotes the branchpoint A nucleotide). Interestingly, we determined that the lariat for the second step involved the E1A 5'

splice site looping back to a branchpoint sequence (GTCTCA\*C) associated with E1B, 89 nt upstream of its junction with E2', and a total of 106 nt upstream of the distal 3' splice site in E2 that is used in the second intrasplicing step (when the 17 nt alternatively spliced portion of E2 is taken into account).

To rule out the possibility that these branchpoints were artifacts of the human minigene system, and also to confirm the results in second vertebrate species, we mapped branchpoint sequences derived from splicing of endogenous 4.1R pre-mRNAs in mouse erythroblasts. Erythroblast RNA was chosen because 4.1R transcripts initiated at E1A are expressed at high levels in these cells. Use of nested PCR primers allowed us to successfully amplify branchpoint sequences used in both first and second steps of intrasplicing for the mouse 4.1R gene. DNA sequence analysis revealed that the first step branchpoint differs in primary sequence between human and mouse, but the approximate location upstream of E2 is conserved. In contrast, the experimentally determined branchpoint used in the second step (GTCTCA\*C) was identical in the mouse and human genes, both in terms of primary sequence and with respect to its location relative to acceptor site SA2 (see also Figure 8 below).

The functional significance of the conserved intraexon branchpoint was further tested by comparing the splicing phenotype of constructs containing the natural branchpoint sequence GTCTCA\*C, or two mutated branchpoints GTCTCGC and GTCTCGT (Figure 7). Whereas positive control construct pE1A-2 was spliced correctly, with E1A joined to the distal acceptor (lane 2), both of the branchpoint mutants yielded predominantly inappropriate splicing of E1A to the proximal acceptor site (lanes 3 and 4). The presence of incorrectly spliced product indicates that the first step of intrasplicing was substantially blocked by the branchpoint mutation, consistent with a role for the branchpoint sequence in exon definition prior to the first splicing

step. We speculate that the small quantities of properly spliced product result from partial utilization of a nearby cryptic branchpoint sequence; cryptic branchpoints have been observed by others (Gooding *et al.*, 2006).

The substantial abrogation of proper E1A splicing to E2 by mutations in critical E1B motifs- either the 5' splice site or its upstream branchpoint- complement the functional analysis of splicing intermediates, and together these data strongly support intrasplicing as the bona fide major pathway for E1A splicing.

### **Conservation of intrasplicing regulatory motifs**

The results above identify two important sub-components of the intraexon: the conserved branchpoint region, and the consensus 5' splice site (Figure 8). Because functional *cis*-regulatory motifs are frequently conserved through evolution, we examined the conservation of these intraexon E1B features among mammalian genomes. Comparative genomic analysis of several mammals showed that both of these motifs represent islands of highly conserved sequence (Figure 8) embedded in a region that is considerably diverged in the flanking introns and only moderately conserved in the body of E1B region in a wide variety of mammals.

To test whether these sequence properties could facilitate identification of intrasplicing elements in other genes, we examined the paralogous 4.1B gene. This gene resembles 4.1R in its 5' structure, with multiple promoters/first exons upstream and alternative 3' splice sites downstream in exon 2 (Figure 9A). In the 143kb region between exon 1A and exon 2'/2, only a single intrasplicing-like element was identified ~20kb downstream of exon 1A. This element was very similar in structure to its counterpart in 4.1R, with candidate distant branchpoint /pyrimidine tract motifs, a consensus 5' splice site, and AG-deficient sequence. Although this candidate intraexon is not present in any database transcripts and has not been annotated as an

exon, it was able to promote the intrasplicing pathway when positioned downstream of exon 1A in the 4.1R minigene (Figure 9B). Furthermore, mutation of the predicted 5' splice site of this element abrogated its ability to promote intrasplicing (data not shown).

## Discussion

### Proper E1A splicing proceeds via a two-step mechanism

This study reports a novel intrasplicing mechanism that coordinates promoter choice with downstream splicing events, at alternative splice acceptors, over long distances. In the 4.1R gene model, it appears that the *location* of the promoter, rather than any specific feature of promoter sequence, is of paramount importance for intrasplicing: transcripts initiated at promoters downstream of the intrasplicing regulatory element splice in a single step to the dominant proximal acceptor site in exon 2, while transcripts initiated at promoters upstream of a functional intraexon ultimately splice instead to a more distal acceptor site in exon 2. This ability to re-splice via an ordered two-step mechanism provides a satisfying molecular explanation for the otherwise paradoxical long-distance coupling of E1A to the distal acceptor site at E2, a phenomenon that is conserved across several vertebrate orders (Tan *et al.*, 2005). To our knowledge, this is the first experimentally verified example of intrasplicing.

Central to the intrasplicing model is the unique behavior of intraexon E1B as an important mediator of both splicing steps. This element exhibits dual features of both exons and introns embodied among the smaller component motifs that together constitute its structure. In the first step the element behaves like an exon (albeit one whose upstream intron is recognized in an AG-independent manner), providing a strong consensus 5' splice site that exclusively joins to the proximal 3' splice acceptor at E2'/2. Excision of this extensive intervening sequence changes the pre-mRNA architecture substantially by removing not only this internal pair of splice sites, but also the branchpoint region naturally located upstream of E2, and juxtaposing E1B with E2'/2. In the second step, the E1B branchpoint / pyrimidine sequences are complemented by the distal AG-dinucleotide derived from E2'/2 to form a functional composite splice acceptor site. E1B

thus behaves like an intron and is deleted in the second step splicing of E1A to E2. In support of this model, we have used mutant splicing substrates to demonstrate the importance of the 5' splice site and the juxtaposed branchpoint sequence, mapped branchpoint sequences and 5' splice sites, and provided evidence that the predicted intermediate RNA is a bona fide precursor of mature mRNA.

Although the complex behavior proposed for E1B function in intrasplicing is unusual, there is precedence for key aspects of its proposed mechanism of action. In the first step it appears that E1B is recognized as an exon, despite its non-conventional 3' splice site that lacks an AG-dinucleotide. AG-independent introns have been shown to be capable of lariat formation (Reed, 1989); it seems reasonable to propose that exon definition of E1B may occur in a similarly AG-independent manner, likely promoted by branchpoint/pyrimidine tract binding proteins. The absence of an AG is important here because a fully functional 3' splice site would probably be detrimental to the second step of intrasplicing by providing competition with the distal SA2 acceptor, leading to inclusion of E1B as an internal exon. The necessity of avoiding such competition may thus explain the complete absence of AG-dinucleotides in E1B in more than a dozen vertebrate genomes examined to date. Presumably, E1B is recognized through its conserved branchpoint/pyrimidine tract, its strong 5' splice site, and (most likely) exonic splicing enhancer elements. In a preliminary scan of the E1B region for potential splicing enhancer elements, using ESEfinder (Cartegni *et al.*, 2003), we found that E1B itself was relatively enriched in predicted enhancers relative to the flanking introns (results not shown). We speculate that splicing enhancer elements are important to prevent premature skipping of the intraexon, analogous to the observations that exonic enhancers can act as barriers to prevent inappropriate exon skipping (Ibrahim *et al.*, 2005). In the second step of intrasplicing, the conserved



branchpoint sequence is relatively distant (106nt upstream) from the juxtaposed internal E2 acceptor site in the intermediate RNA. How this branchpoint functions at a distance is not clear; however, there is precedence for distal branchpoints and AG exclusion zones in a subclass of mammalian 3' splice sites (Gooding *et al.*, 2006). Consistent with this model, the 4.1R and 4.1B intraexon sequences are similarly devoid of AG dinucleotides in the branchpoint to 5' splice site region.

### **Intrasplicing is distinct from recursive splicing**

Recursive splicing, a mechanism for stepwise removal of long introns in *Drosophila*, is distinct in its molecular details from intrasplicing. In recursive splicing, an exon utilizes its natural 5' splice site in a first step to splice to a downstream 3' splice site AG-dinucleotide in a "zero length" exon- so named because the 3' splice site is immediately adjacent to a 5' splice site with no actual exon sequence between them. Juxtaposition of this 5' element with the original exon thus creates a new composite 5' splice that allows the exon to splice again to a new downstream 3' splice site. In contrast, intrasplicing is essentially a situation of nested introns with internal and external splice site pairs utilized sequentially. Despite these differences, however, we note similarities in the mechanism of recursive splicing and the specific intrasplicing pathway described here. Both require participation of regulatory elements- the zero length exon or the intraexon- that can be buried deep within introns, and are essential for intermediate steps of RNA processing, but are not represented in the mature spliced product. Both processes take advantage of the rearrangement of regulatory elements that results from one splicing step, to promote the next step of splicing in an orderly process that must require dynamic remodeling of the spliceosomal complexes. It will be interesting to explore these mechanisms in more detail in the future.

**Intrasplicing functionally regulates 4.1R protein structure**

The splicing decision at alternative acceptors in E2 directly affects protein 4.1R N-terminal structure, because the short 17 nt sequence flanked by these two sites encodes translation initiation site AUG1. During erythroid differentiation, this splicing decision is tied through intrasplicing to upstream transcription events to insure synthesis of the functional 80 kDa isoforms of the protein. Specifically, 4.1R transcription in late erythroblasts begins almost exclusively at E1A (Parra *et al.*, 2003); intrasplicing directs E1A transcripts to skip translation start site AUG1; and translation begins predominantly downstream at AUG2 in exon 4 to give rise to 80 kDa 4.1R protein. In contrast, other cell types utilize both E1A and E1C to produce a mixture of 80 kDa and higher molecular weight isoforms of the protein.

The preceding analysis suggests also that the original annotation of E1B as an alternative first exon in the 4.1R gene should be re-evaluated. Although standard promoter assays with luciferase reporters indicated that the region immediately upstream of E1B possesses promoter activity, there is little EST evidence for transcripts initiated at E1B. We propose instead that the major function of E1B is in the intrasplicing pathway, and that E1B will thus be represented predominantly in splicing intermediates but not in mature mRNA.

**Other possible roles for intrasplicing**

Intrasplicing could play several distinct roles in processing metazoan pre-mRNAs. Given the abundance of alternative first exons and alternative 3' splice sites in human genes, intrasplicing may offer a new mechanism by which promoter choice can be coupled indirectly to downstream alternative splicing. Recent studies in the ENCODE project have demonstrated that distal 5' transcription start sites are not uncommon in the human genome (Denoeud *et al.*, 2007). It is interesting to speculate that some of these cases may utilize intrasplicing either to progressively

shorten a long first intron, or to couple promoter choice with downstream splicing events as demonstrated here for the 4.1R gene. It is important to distinguish this model from other coupling mechanisms in which promoter identity can influence transcription rate or splicing factor binding activity so as to regulate downstream splicing (Auboeuf *et al.*, 2004; Kornblihtt, 2005). The intrasplicing model is unique in that specific promoter characteristics are not required, as demonstrated by the ability of heterologous promoters to exhibit proper splicing to E2 via this pathway. It appears instead that the physical location of the promoter, upstream of the intraexon regulator, is essential for effects on downstream splicing.

More generally, intrasplicing could be used to regulate alternative 3' splice sites internally within a gene, if a regulatory element similar to the 4.1R intraexon resides upstream of the regulated splice site pair. Regulatory control of intrasplicing at such sites could then be mediated by antagonistic activities of silencer and enhancer motifs, to inhibit or promote the intrasplicing pathway, similar to regulation of tissue-specific alternative exons. Since a substantial fraction of alternative splicing events involve alternative splice donor or splice acceptor sites (Hiller *et al.*, 2004; Sugnet *et al.*, 2004), this mode of regulation could be more widespread than currently appreciated. Such regulation might also explain cases in which use of alternative acceptors does not correlate with splice site strength, or diverges from the normal preference for use of the more proximal site (Zheng *et al.*, 2005). Included among the exons that could be regulated by intrasplicing are those with the so-called NAGNAG acceptors, in which the alternative 3' splice sites are only three nucleotides apart. Such cases are reported to appear in ~5% of human genes, and some may be evolutionarily conserved and tissue specifically regulated (Akerman and Mandel-Gutfreund, 2006; Hiller *et al.*, 2004).

Finally, intrasplicing has been proposed as a mechanism for aiding the removal of long introns (Ott *et al.*, 2003). As many as 10% of human introns are >10kb in length (Sakharkar *et al.*, 2005), and intrasplicing could well be adapted for stepwise removal of long introns in other genes, somewhat analogous to the ability of recursive splicing to do so (Burnette *et al.*, 2005; Hatton *et al.*, 1998).

## Materials and Methods

### General molecular biology and splicing assay

Cloning of the various minigene constructs was performed following standard molecular biology protocols (Sambrook *et al.*, 1989). Restriction enzymes, ligases, and phosphatases were obtained from New England Biolabs, while mutagenesis was performed using Stratagene's QuikChange XL Site-Directed Mutagenesis Kit. Finally, the minigene vector used in this study was pcDNA3, supplied by Invitrogen.

### Minigene construction

Minigene E1A-1 contains the following elements of the human protein 4.1R (EPB41) gene: the E1A promoter region, E1A exon sequences, and a short region of downstream intron sequence joined to E2 with a short upstream intronic region by an overlapping PCR strategy (in the May 2006 version of the human genome assembly, these two regions correspond to coordinates chr1: 29,084,660-29,086,576 and chr1:29,186,261-29,187,004, respectively). This 2.7kb minigene was cloned into the XhoI/XbaI sites of a pcDNA3 vector from which the CMV promoter region had been removed by deleting a 740nt region between NruI and EcoRV sites. The functional pE1A-2L minigene was constructed by extending the E1A fragment described above ~ 7 kb further downstream so as to include E1B and a bit of its downstream intron

sequence (chr1: 29,084,660-29,093,166). This construct was further modified by deleting substantial portions of the intron sequence between E1A and E1B to generate pE1A-2, pE1A-2a, and pE1A-2b. Briefly, pE1A-2 contained only a 200 nt region of E1B (chr1:29092779-29093001; 223nt) inserted between E1A and E2; pE1A-2a contained a region shifted slightly upstream so as to overlap only the 5' half of E1B (chr1:29,092,690-29,092,909; 219nt); and pE1A-2b contained a region shifted slightly downstream so as to overlap only the 3' half of E1B (chr1:29,092,888-29,093,096; 209nt). Minigene pANK-2L is similar to pE1A-2L, except that the promoter and first exon of the erythroid ankyrin gene (chr8:41,773,748-41,774,499 in the UCSC genome browser) were substituted in place of the 4.1R E1A elements. Minigene pE1C-2L likewise substitutes the promoter/first exon of exon 1C from the human 4.1R gene (chr1:29,112,211-29,114,244 in the UCSC genome browser) in place of the 4.1R E1A elements. Construct pE1A-2Lint is identical to pE1A-2L, except the intron between E1B and E2 is deleted. Likewise, pE1A-2int is identical to pE1A-2, except for deletion of the same intron.

### Site-directed mutagenesis

Mutagenesis of candidate splicing elements was performed using the QuikChange XL Site-Directed Mutagenesis Kit (Stratagene). The 5' splice site of E1B was altered from GTGAGT to ATGAAT using the primer 5'-GACGGCTGCTTCTGGATGAATAATGG-TTTGACTTAATG-3' and its reverse complement. The branchpoint upstream of E1B was mutated from GTCTCAC to GTCTCGC using the primer 5'-GGCTGTGCTGTGTGTCTCGCTGCTGTCATTTGCTTCC-3' and its reverse complement. A second mutation to GTCTCGT was performed using primer 5'-GGCTGTGCTGTGTGTCTCGTTGCTGTCATTTGCTTCCA-3' and its reverse complement. Underlined nucleotides in the primers indicate positions of the mutated nucleotides; bold nucleotides indicate normal position of the branchpoint.

## Splicing analysis

Minigenes were transfected into mouse 3T3 or human HEK293T cells using Fugene Transfection Reagent (Roche) as described (Ponthier *et al.*, 2006). Total RNA was extracted from cells with Qiagen's RNeasy Mini Kit, and then reverse-transcribed into cDNA using the Superscript III First Strand Synthesis System (Invitrogen) with a reverse primer targeted specifically against the pcDNA3 vector (5'-TACAAGGCACAGTCGAGG-3'). The following primers were used in subsequent PCR assays designed to distinguish between correct and incorrect splicing at the alternative splice acceptors at E2: F1 (E1A forward primer) 5'-GCAAAGTGGCAGGAACCTCTTAAAG-3'; R1 (E2 reverse primer) 5'-CGAGGAGAATAGTCGTGAAAGTCC-3'; R2 (reverse primer that overlaps E2'/2) 5'-GTCACTAAACTCTTCTCTGTTGTC-3'; ANK1 forward, 5'-CTCCTCCTCAAGCCCCCAAG-3'; E1C forward primer is 5'-CTCCTCCTGCCATTGTTCGT-3'.

## Lariat analyses

Amplification of lariat intermediates was performed as described (Vogel *et al.*, 1997) and diagrammed in Figure 7. To detect the lariat intermediate for the first intrasplicing step joining E1B to E2', the nested PCR reactions were performed with the last step utilizing a forward primer located upstream of exon E2' (5'-GGATGGTATTCGTCCCGGATGGT-3') and a reverse primer downstream of E1B (5'-GCAGTCATTAAGTCAAACCATTA-3'). Similarly, to detect the lariat resulting from second step splicing of splicing of E1A to E2, nested PCR reactions were again performed using forward primers located upstream of intraexon E1B (final primer sequence: 5'-CAGATTATTGCTGACTAGGGCGT-3') and reverse primers downstream of E1A (final primer sequence: 5'-GGACCCTTCGACCCTGGTCTCGC-3'). As described in

the original report, the PCR product spanning the branch point exhibited a “T” in place of the actual branchpoint “A” residue. In multiple experiments to amplify the branchpoints from low abundance pre-mRNAs, we consistently obtained the expected products from the 4.1R gene. However, unrelated products with no sequence homology to the 4.1R gene were also sometimes amplified and we were unable to completely eliminate these additional artifacts.

## **Acknowledgments**

The authors would like to thank lab members Miki Yamamoto and Thomas Gallagher for helpful suggestions and discussions. This work was supported by National Institutes of Health NIH grant DK32094 and by the Director, Office of Biological and Environmental Research, US Department of Energy under contract DE-AC03-76SF00098.

## References

- Akerman, M. and Mandel-Gutfreund, Y. (2006) Alternative splicing regulation at tandem 3' splice sites. *Nucleic Acids Res*, **34**, 23-31.
- Auboeuf, D., Dowhan, D.H., Kang, Y.K., Larkin, K., Lee, J.W., Berget, S.M. and O'Malley, B.W. (2004) Differential recruitment of nuclear receptor coactivators may determine alternative RNA splice site choice in target genes. *Proc Natl Acad Sci U S A*, **101**, 2270-2274.
- Baek, D., Davis, C., Ewing, B., Gordon, D. and Green, P. (2007) Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. *Genome Res*, **17**, 145-155.
- Baklouti, F., Huang, S.C., Vulliamy, T.J., Delaunay, J. and Benz, E.J., Jr. (1997) Organization of the human protein 4.1 genomic locus: new insights into the tissue-specific alternative splicing of the pre-mRNA. *Genomics*, **39**, 289-302.
- Bird, G., Zorio, D.A. and Bentley, D.L. (2004) RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol*, **24**, 8963-8969.
- Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, **72**, 291-336.
- Blencowe, B.J. (2006) Alternative splicing: new insights from global analyses. *Cell*, **126**, 37-47.
- Burnette, J.M., Miyamoto-Sato, E., Schaub, M.A., Conklin, J. and Lopez, A.J. (2005) Subdivision of large introns in *Drosophila* by recursive splicing at nonexonic elements. *Genetics*, **170**, 661-674.



- Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q. and Krainer, A.R. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res*, **31**, 3568-3571.
- Conboy, J.G., Chan, J.Y., Chasis, J.A., Kan, Y.W. and Mohandas, N. (1991) Tissue- and development-specific alternative RNA splicing regulates expression of multiple isoforms of erythroid membrane protein 4.1. *J Biol Chem*, **266**, 8273-8280.
- Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L. and Myers, R.M. (2006) Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res*, **16**, 1-10.
- Cramer, P., Caceres, J.F., Cazalla, D., Kadener, S., Muro, A.F., Baralle, F.E. and Kornblihtt, A.R. (1999) Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol Cell*, **4**, 251-258.
- Cramer, P., Pesce, C.G., Baralle, F.E. and Kornblihtt, A.R. (1997) Functional association between promoter structure and transcript alternative splicing. *Proc Natl Acad Sci U S A*, **94**, 11456-11460.
- Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M. and Reed, R. (2006) Functional coupling of RNAP II transcription to spliceosome assembly. *Genes Dev*, **20**, 1100-1109.
- Denoëud, F., Kapranov, P., Ucla, C., Frankish, A., Castelo, R., Drenkow, J., Lagarde, J., Alioto, T., Manzano, C., Chrast, J., Dike, S., Wyss, C., Henrichsen, C.N., Holroyd, N., Dickson, M.C., Taylor, R., Hance, Z., Foissac, S., Myers, R.M., Rogers, J., Hubbard, T., Harrow, J., Guigó, R., Gingeras, T.R., Antonarakis, S.E. and A., R. (2007) Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res.*, **17**, 746-759.

- Dye, M.J., Gromak, N. and Proudfoot, N.J. (2006) Exon tethering in transcription by RNA polymerase II. *Mol Cell*, **21**, 849-859.
- Faustino, N.A. and Cooper, T.A. (2003) Pre-mRNA splicing and human disease. *Genes Dev.*, **17**, 419-437.
- Gallagher, P.G., Romana, M., Tse, W.T., Lux, S.E. and Forget, B.G. (2000) The human ankyrin-1 gene is selectively transcribed in erythroid cell lines despite the presence of a housekeeping-like promoter. *Blood*, **96**, 1136-1143.
- Gooding, C., Clark, F., Wollerton, M.C., Grellscheid, S.N., Groom, H. and Smith, C.W. (2006) A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. *Genome Biology*, **7**, R1.
- Hatton, A.R., Subramaniam, V. and Lopez, A.J. (1998) Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. *Mol. Cell*, **2**, 787-796.
- Hiller, M., Huse, K., Szafranski, K., Jahn, N., Hampe, J., Schreiber, S., Backofen, R. and Platzer, M. (2004) Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nat Genet*, **36**, 1255-1257.
- Hirose, Y. and Manley, J.L. (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev*, **14**, 1415-1429.
- Hong, X., Scofield, D.G. and Lynch, M. (2006) Intron size, abundance, and distribution within untranslated regions of genes. *Mol Biol Evol*, **23**, 2392-2404.
- Huang, J.P., Tang, C.J., Kou, G.H., Marchesi, V.T., Benz, E.J., Jr. and Tang, T.K. (1993) Genomic structure of the locus encoding protein 4.1. Structural basis for complex

- combinational patterns of tissue-specific alternative RNA splicing. *J Biol Chem*, **268**, 3758-3766.
- Ibrahim el, C., Schaal, T.D., Hertel, K.J., Reed, R. and Maniatis, T. (2005) Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proc Natl Acad Sci U S A*, **102**, 5002-5007.
- Kim, T.H., Barrera, L.O., Qu, C., Van Calcar, S., Trinklein, N.D., Cooper, S.J., Luna, R.M., Glass, C.K., Rosenfeld, M.G., Myers, R.M. and Ren, B. (2005) Direct isolation and identification of promoters in the human genome. *Genome Res*, **15**, 830-839.
- Kornblihtt, A.R. (2005) Promoter usage and alternative splicing. *Curr Opin Cell Biol*, **17**, 262-268.
- Listerman, I., Sapra, A.K. and Neugebauer, K.M. (2006) Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat Struct Mol Biol*, **13**, 815-822.
- Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature*, **416**, 499-506.
- Matlin, A.J., Clark, F. and Smith, C.W. (2005) Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol*, **6**, 386-398.
- Nunomura, W., An, X., Gascard, P., Parra, M., Narla, M. and Takakuwa, Y. (2002) Regulation of 135kDa protein 4.1R interactions with membrane proteins by its head-piece region. *Mol Biol Cell*, **13**, 500a.
- Ott, S., Tamada, Y., Bannai, H., Nakai, K. and Miyano, S. (2003) Intrasplicing--analysis of long intron sequences. *Pac Symp Biocomput*, 339-350.

- Pagani, F., Stuani, C., Zuccato, E., Kornblihtt, A.R. and Baralle, F.E. (2003) Promoter architecture modulates CFTR exon 9 skipping. *J Biol Chem*, **278**, 1511-1517.
- Parra, M., Gee, S., Chan, N., Ryaboy, D., Dubchak, I., Mohandas, N., Gascard, P.D. and Conboy, J.G. (2004) Differential domain evolution and complex RNA processing in a family of paralogous EPB41 (protein 4.1) genes facilitate expression of diverse tissue-specific isoforms. *Genomics*, **84**, 637-646.
- Parra, M.K., Gee, S.L., Koury, M.J., Mohandas, N. and Conboy, J.G. (2003) Alternative 5' exons and differential splicing regulate expression of protein 4.1R isoforms with distinct N-termini. *Blood*, **101**, 4164-4171.
- Ponthier, J.L., Schluepen, C., Chen, W., Lersch, R.A., Gee, S.L., Hou, V.C., Lo, A.J., Short, S.A., Chasis, J.A., Winkelmann, J.C. and Conboy, J.G. (2006) Fox-2 Splicing Factor Binds to a Conserved Intron Motif to Promote Inclusion of Protein 4.1R Alternative Exon 16. *J Biol Chem*, **281**, 12468-12474.
- Reed, R. (1989) The organization of 3'splice-site sequences in mammalian introns. *Genes Dev.*, **3**, 2113-2123.
- Sakharkar, M.K., Perumal, B.S., Sakharkar, K.R. and Kanguane, P. (2005) An analysis on gene architecture in human and mouse genomes. *In Silico Biol*, **5**, 347-365.
- Schischmanoff, P.O., Yaswen, P., Parra, M.K., Lee, G., Chasis, J.A., Mohandas, N. and Conboy, J.G. (1997) Cell shape-dependent regulation of protein 4.1 alternative pre-mRNA splicing in mammary epithelial cells. *J Biol Chem*, **272**, 10254-10259.
- Sugnet, C.W., Kent, W.J., Ares, M., Jr. and Haussler, D. (2004) Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac Symp Biocomput*, 66-77.

- Tan, J.S., Mohandas, N. and Conboy, J.G. (2005) Evolutionarily conserved coupling of transcription and alternative splicing in the EPB41 (protein 4.1R) and EPB41L3 (protein 4.1B) genes. *Genomics*, **86**, 701-707.
- Tan, J.S., Mohandas, N. and Conboy, J.G. (2006) High frequency of alternative first exons in erythroid genes suggests a critical role in regulating gene function. *Blood*, **107**, 2557-2561.
- Vogel, J., Hess, W.R. and Borner, T. (1997) Precise branch point mapping and quantification of splicing intermediates. *Nucleic Acids Res*, **25**, 2030-2031.
- Zheng, C.L., Fu, X.D. and Gribskov, M. (2005) Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. *Rna*, **11**, 1777-1787.

## Figure Legends

**Figure 1. Promoter choice and downstream alternative splicing events in protein 4.1R. A.** Arrangement of the 4.1R gene, 5' region, showing three alternative first exons E1A, E1B, and E1C, and the differential manner in which they splice selectively to alternative acceptors downstream in E2. The proximal site E2 acceptor is pyrimidine rich and is predicted to be a stronger 3' splice site than the distal, more purine rich, site (Tan et al., 2005); the 17 bp region between these acceptors is indicated by E2'. The thin line to E1B indicates that this isoform is rare among database clones. Numbers on the introns indicate length in kb. (B) An important functional consequence of this splicing variation is that E1A transcripts produce a shorter protein isoform compared to that made from E1B/C promoters, with the former missing a critical

‘headpiece’ region (HP) that participates in protein-protein interactions. Other protein 4.1R domains are indicated by the abbreviations FERM (4.1/ezrin/radixin/moesin homology domain), SAB (Spectrin-Actin Binding domain), and CTD (C-Terminal Domain).

**Figure 2. Minigene constructs used to analyze intrasplicing.** Diagrams show the structure of minigenes used in this study to map the intrasplicing regulatory element, to test promoter dependence of intrasplicing, and to investigate motifs essential for the intrasplicing mechanism. Each construct consists of three components derived from the human 4.1R gene: a promoter/first exon, a portion of the 4.1R intron downstream of E1A, and a terminal exon consisting of 4.1R exon 2 with its proximal upstream intron sequence. Gaps represent regions of the natural gene that have been deleted. Promoter identity is indicated in the construct designation; pE1A, 4.1R exon 1A promoter; pCMV, cytomegalovirus promoter; pANK, erythroid ANK1 promoter (Gallagher *et al.*, 2000); and pE1C, 4.1R exon 1C promoter. E1B does not have a consensus 3’ splice site; its approximate 5’ boundary is indicated by the dotted line. Constructs pE1A-2b and pE1A-2a each contains only half of the E1B region. In the bottom panel, sites marked by the “X” indicate position of mutations in the E1B 5’ splice site and E1B-associated branchpoint; 1B\* indicates a candidate regulatory element from the paralogous 4.1B gene.

**Figure 3. A downstream regulatory element is required for proper E1A splicing to exon 2**  
Analysis of spliced products as a function of the presence of downstream regulatory element(s). Shown are the amplified products from cells transfected with the following constructs: lane 1, size standards (indicated in nt); lane 2, pE1A-1; lane 3, pE1A-2L; lane 4, pE1A-2; lane 5, pE1A-2a; lane 6, pE1A-2b. Lane 7 represents a negative PCR control. The larger PCR product (expected size=343 nt) corresponds to inappropriate splicing of E1A to the proximal acceptor in E2, while the smaller product (expected size=326 nt) represents correct splicing to the distal

acceptor. Correct splicing of transcripts initiated at E1A required a regulatory element in the downstream intron that overlaps with E1B.

**Figure 4. Proper E1A splicing is independent of promoter architecture, but requires splicing element(s) associated with E1B.** A. Analysis of spliced products as a function of promoter identity. Shown are the amplified products from cells transfected with the following constructs: lanes 1 and 7, pE1A-2L; lane 2, pCMV-2L; lane 3, pANK-2L; lanes 4 and 6, pE1A-1 negative control, lanes 5 and 9, mock transfection; lane 8, pE1C-2L. Correct splicing of transcripts initiated at promoters upstream of E1B was independent of promoter or first exon identity. B. Dependence of E1A splicing on the E1B 5' splice site. Lanes in panel (B) are as follows: Lane 1, construct E1A-1 lacking the regulatory element; lane 2, construct E1A-2 with an intact E1B 5' splice site; lane 3, construct E1A-2mut5' with a mutant E1B 5' splice site; lane 4, negative PCR control; lane 5, mock transfected cells.

**Figure 5. The two-step intrasplicing hypothesis.** (A). In this model, the splice donor site of E1B (SD1) splices to the proximal 3' acceptor of E2 (SA1) in the first step of the reaction, resulting in an intermediate and an excised lariat structure. This step juxtaposes E1B with E2 and, by eliminating SA1, activates distal acceptor SA2. In the second step, E1B functions as an intron so that the E1A 5' splice site (SD2) splices to SA2 using branchpoint bp2, excising a second lariat and releasing the mature product. (B). Top panel shows the sequence context of E1B prior to the first step of intrasplicing, when it is preceded by a candidate branchpoint / pyrimidine tract and followed by a consensus 5' splice site. Lower panel shows the sequence context after the first step of intrasplicing, when E1B has been juxtaposed to E2'/2 to form a new composite 3' splice site in conjunction with the distal AG-dinucleotide.

**Figure 6. Analysis of the partially spliced RNA intermediate.** (A) Detection of the intermediate RNA produced in the first step of intrasplicing from pE1A-2. RNA from transfected cells was amplified by RT-PCR using the indicated primers. (B) Splicing of intermediate RNA into mature product. Constructs used for splicing assays: lane 1, pE1A-1 negative control; lane 2, pE1A-2L positive control; lane 3, intermediate pE1A-2Lint; lane 4, mock-transfected control. Note that the intermediate functions as a direct precursor of the mature E1A-E2 spliced product. (C). Detection of a partial splicing intermediate from endogenous RNA. Natural RNA purified from mouse splenic erythroblasts was amplified using an antisense primer at the E2'/2 junction, and sense primers in the intron upstream of E1B. In the longest successful amplification (lane 1), ~3.2kb of upstream intron was present upstream of the joined E1B/2'/2. Lane 2 represents size standards in kb.

**Figure 7. Branchpoint analysis.** (A) RT-PCR strategy for detecting lariat intermediates from both steps of intrasplicing (Vogel *et al.*, 1997), showing location of antisense (AS1, AS2) and sense (S1, S2) primers. Branchpoint is indicated by "bp". (B) Gel analysis of the amplified branchpoint structures. Arrows indicate PCR products whose sequences were confirmed to represent lariat structures in which upstream 5' splice sites loop back to the relevant branchpoints for both steps of the proposed mechanism. Lane 1, 100 bp size ladder; lane 2, lariat from second step of intrasplicing; lane 3, lariat from first step of intrasplicing. The additional band in lane 3 is an artifact unrelated to 4.1R transcripts. C. Functional testing of the branchpoint for step 2. Lane 1, incorrectly spliced product from pE1A-1; lane 2, correctly spliced product from pE1A-2; lanes 3-4, mixture of spliced products from two different branchpoint mutants; lane 5, mock transfection; lane 6, negative PCR control. Mutagenesis of the highly conserved branchpoint resulted in reduced production of properly spliced E1A transcript.



**Figure 8. Phylogenetic conservation of the intraexon-associated branchpoint and 5' splice site.**

(A) Organization of the intrasplicing element, showing the conserved branchpoint located 84-92 nt upstream of the conserved, consensus 5' splice site. (B) Comparison of genome sequences shows that these two key motifs were conserved in a wide range of mammalian species. In the branchpoint sequences, nonconserved nucleotides are underlined and the presumed branchpoint A is represented in bold. In the 5' splice site sequences, upper case represents the end of E1B and lower case nucleotides indicate the beginning of the intron.

**Figure 9. Functional analysis of a heterologous intraexon from the paralogous 4.1B gene.**

(A) A candidate intraexon E1B\* identified 20kb downstream of exon 1A in the 4.1B gene was inserted into the 4.1R minigene in place of the natural element, and tested for its ability to promote the intrasplicing pathway. (B). Splicing assay showing that the heterologous element does promote proper intrasplicing.

Figure 1

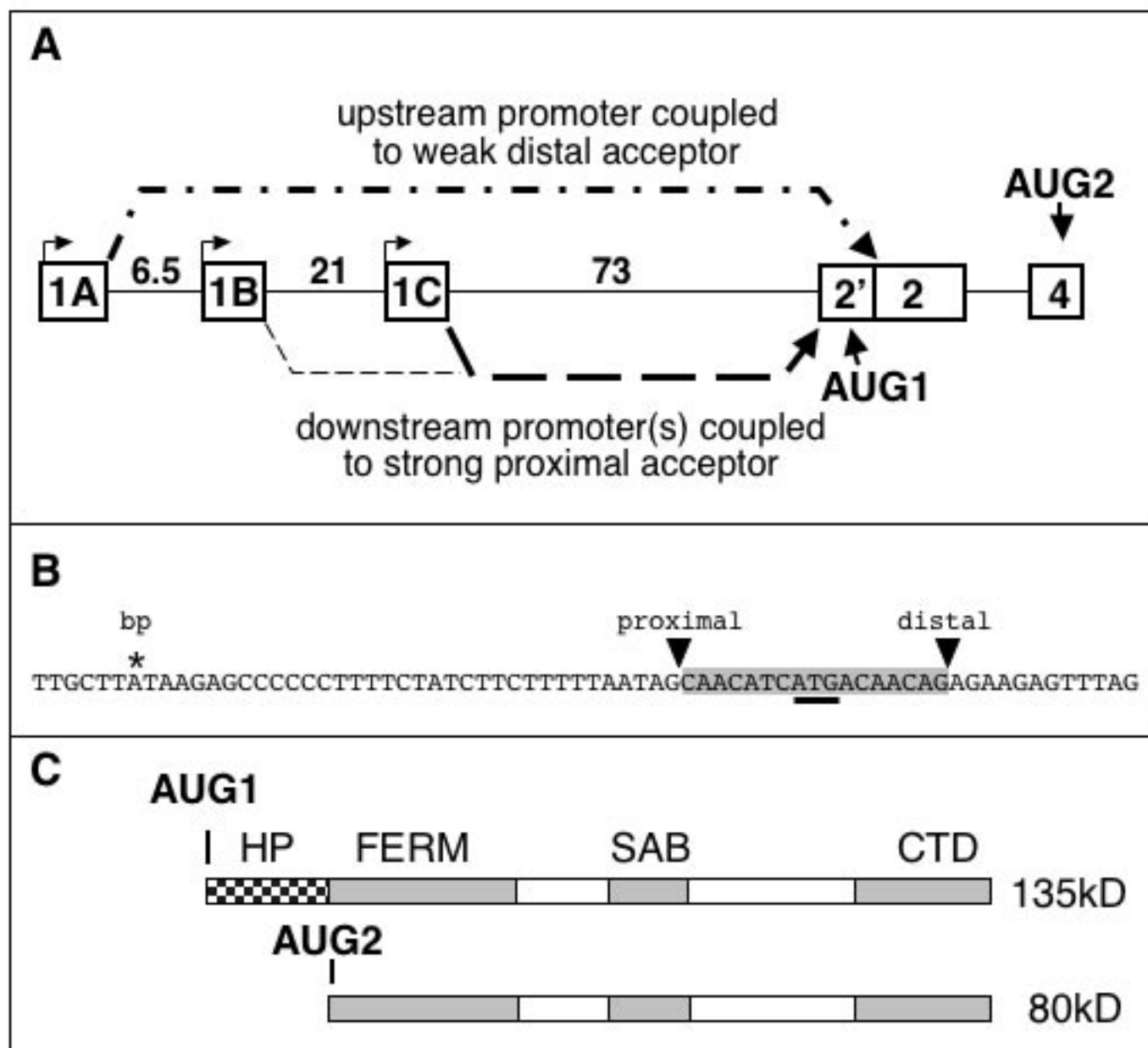


Figure 2

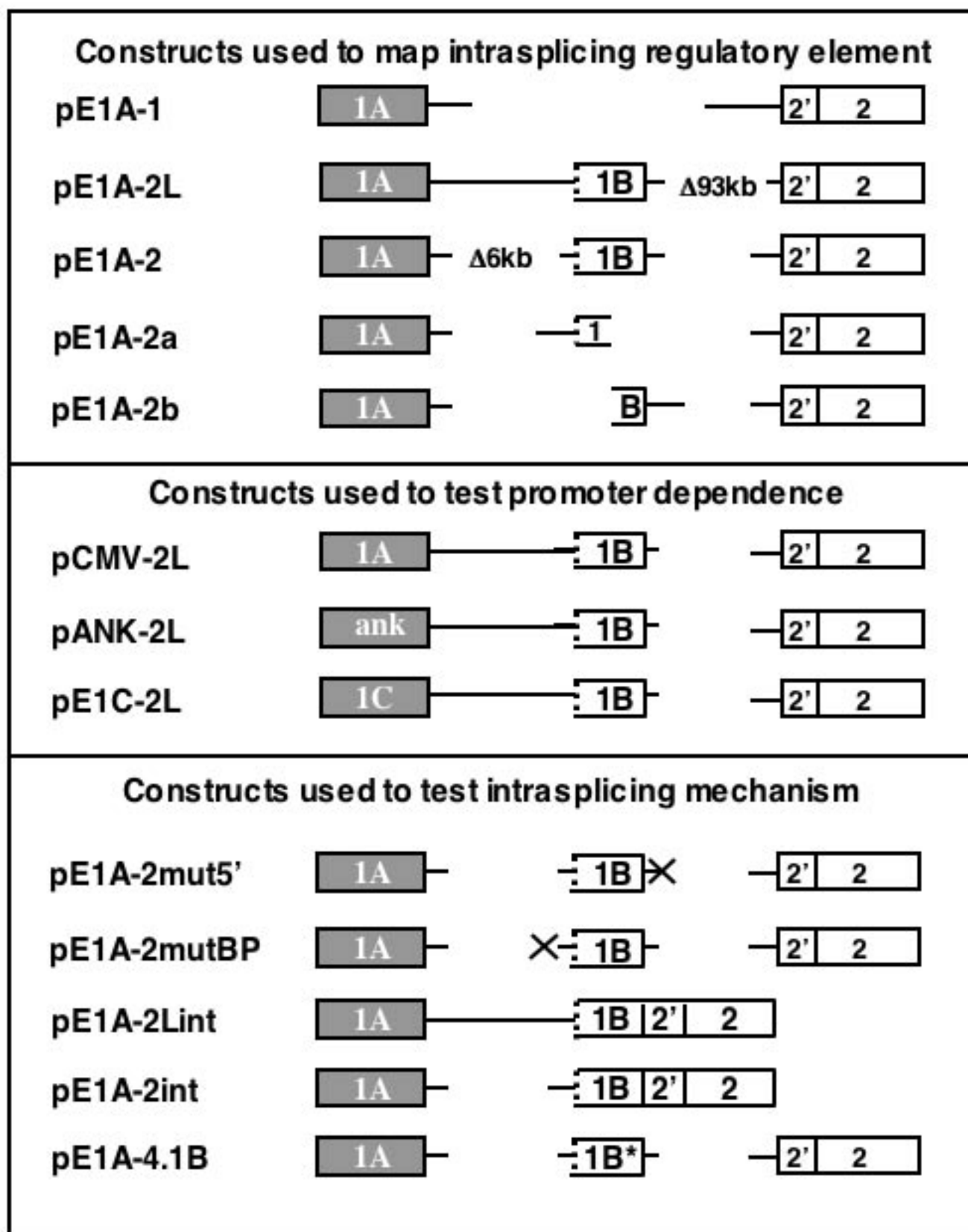


Figure 3

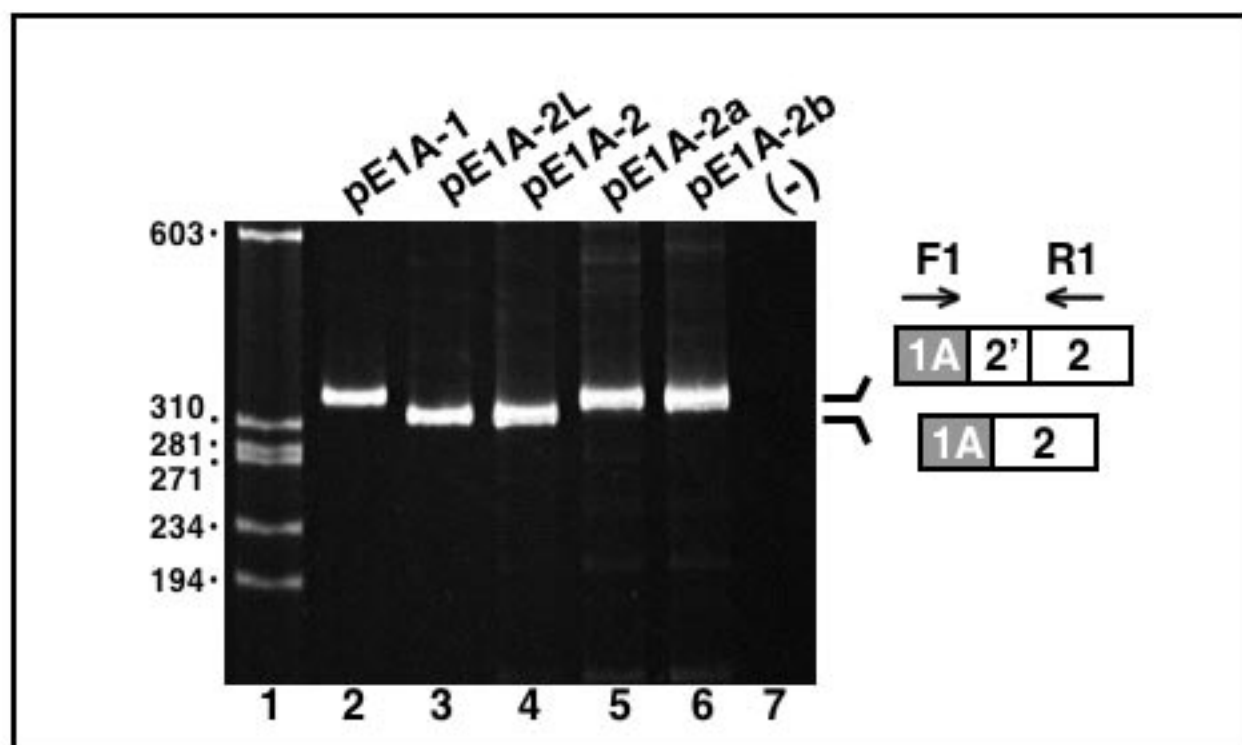


Figure 4

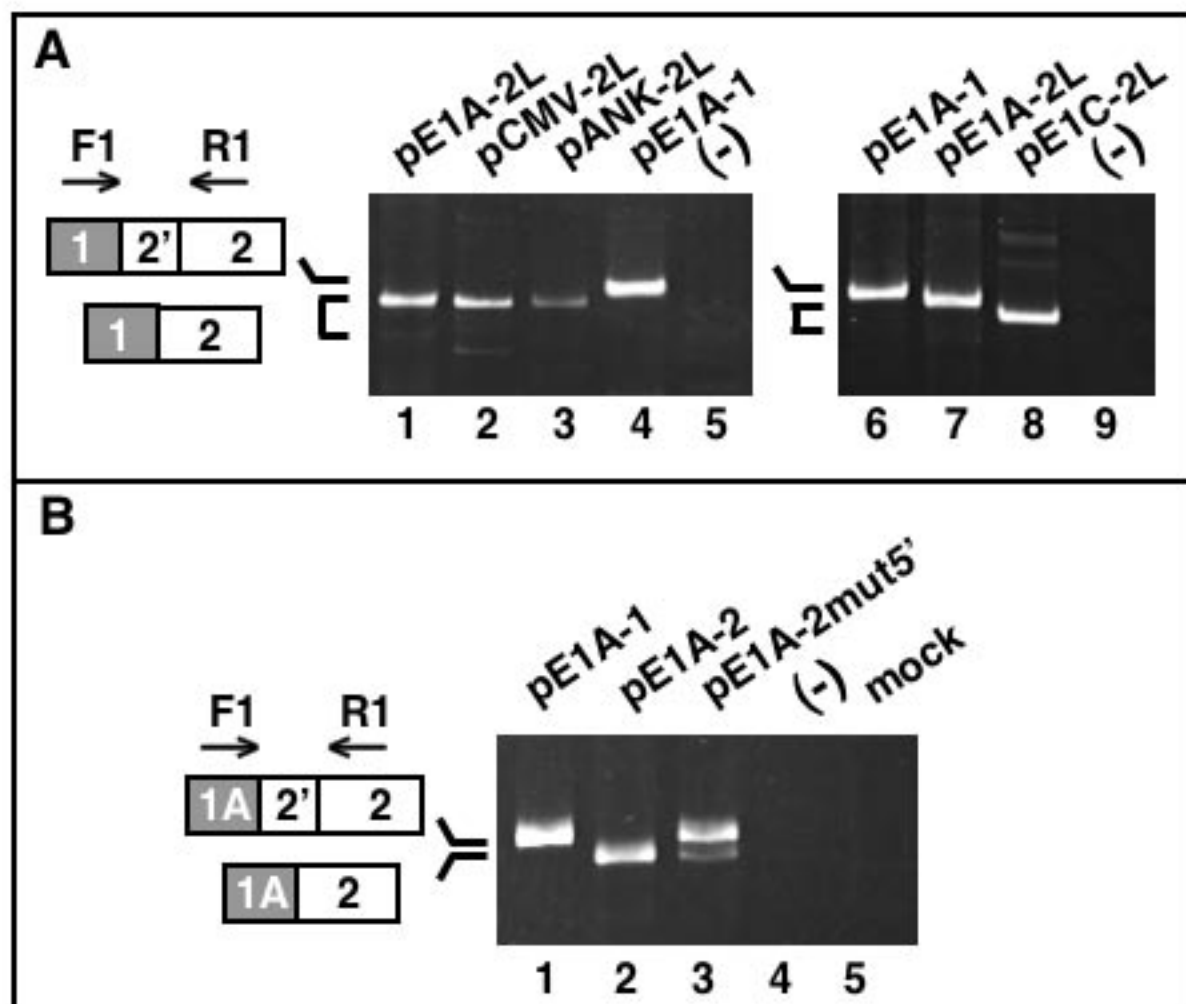


Figure 5A

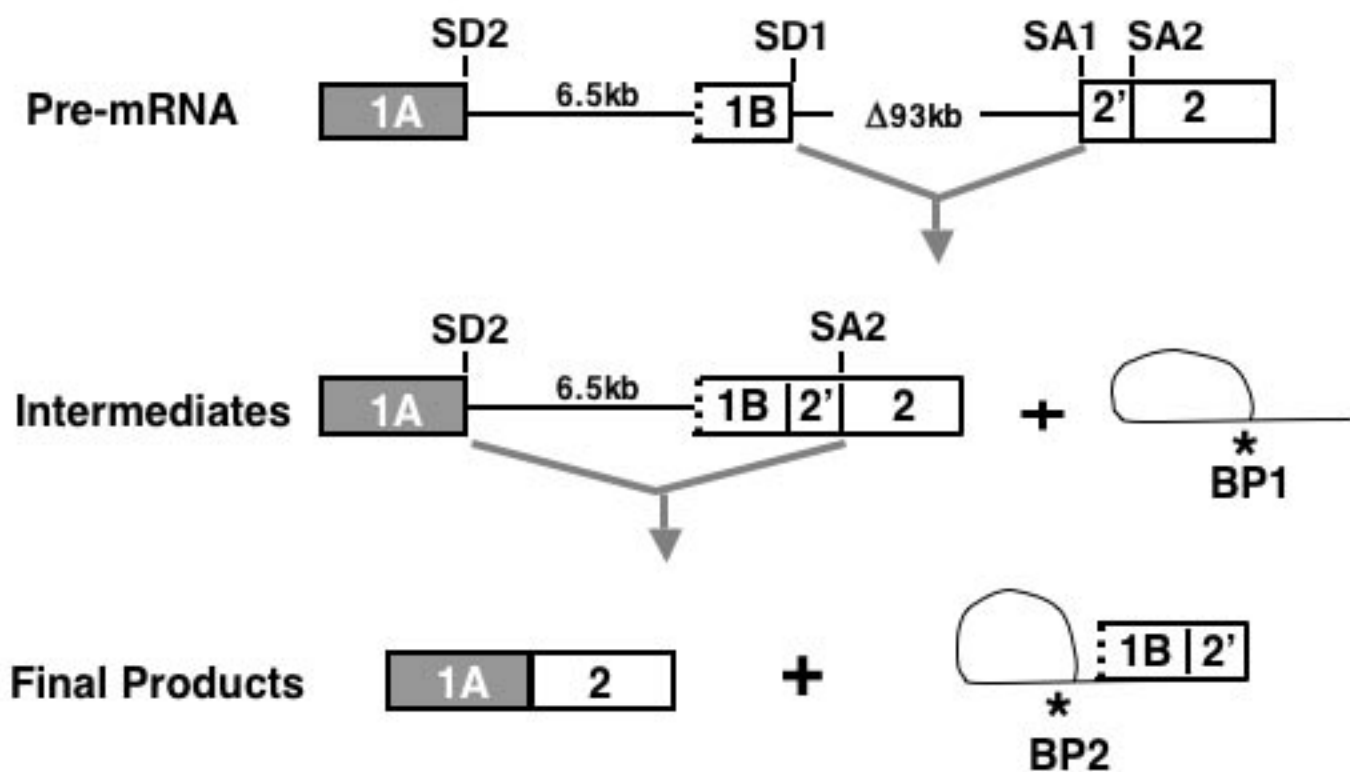


Figure 5B

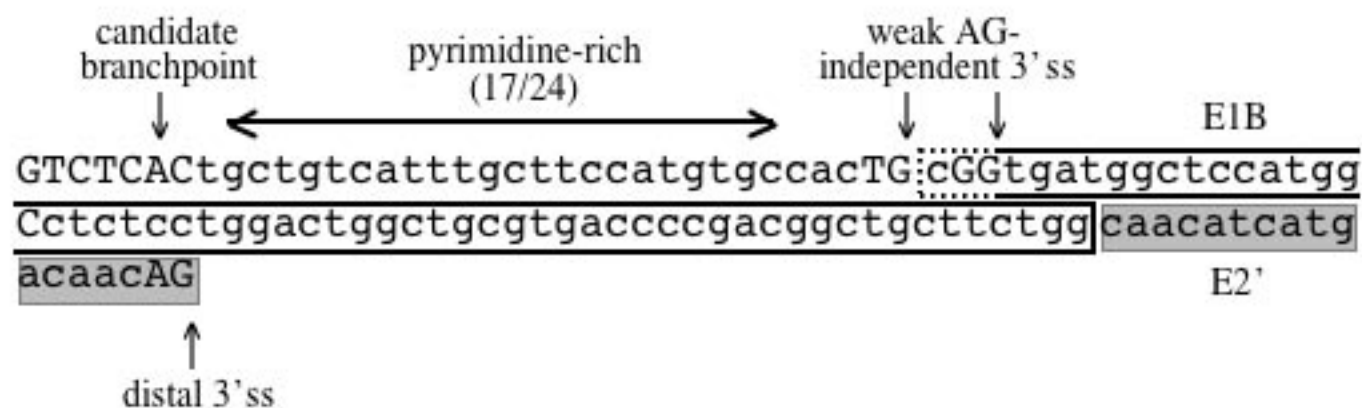
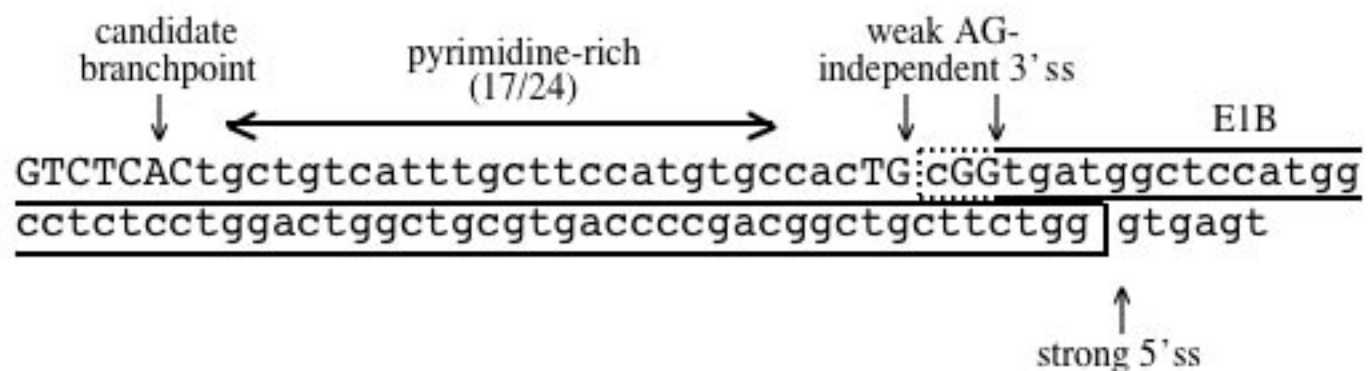


Figure 6

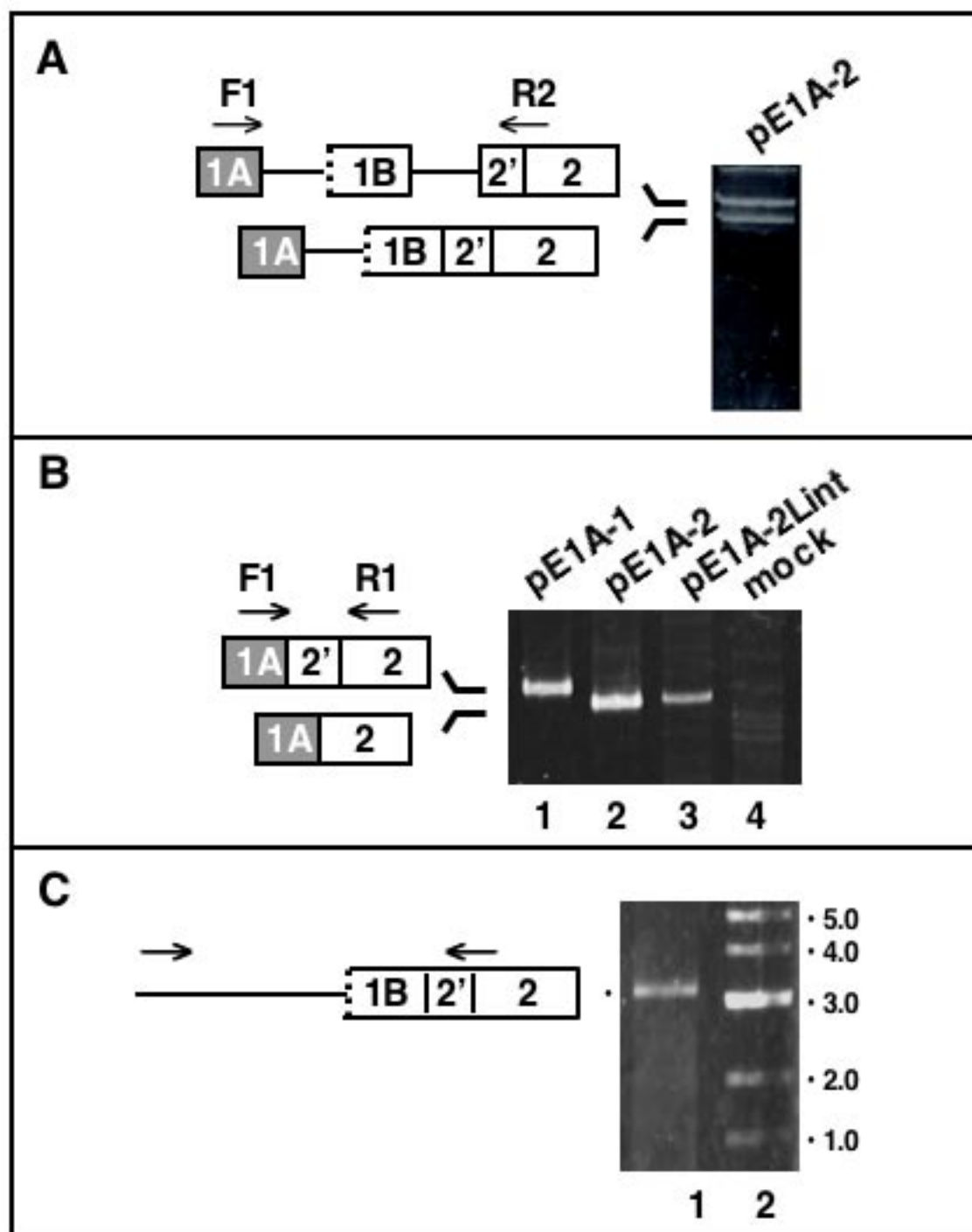




Figure 7

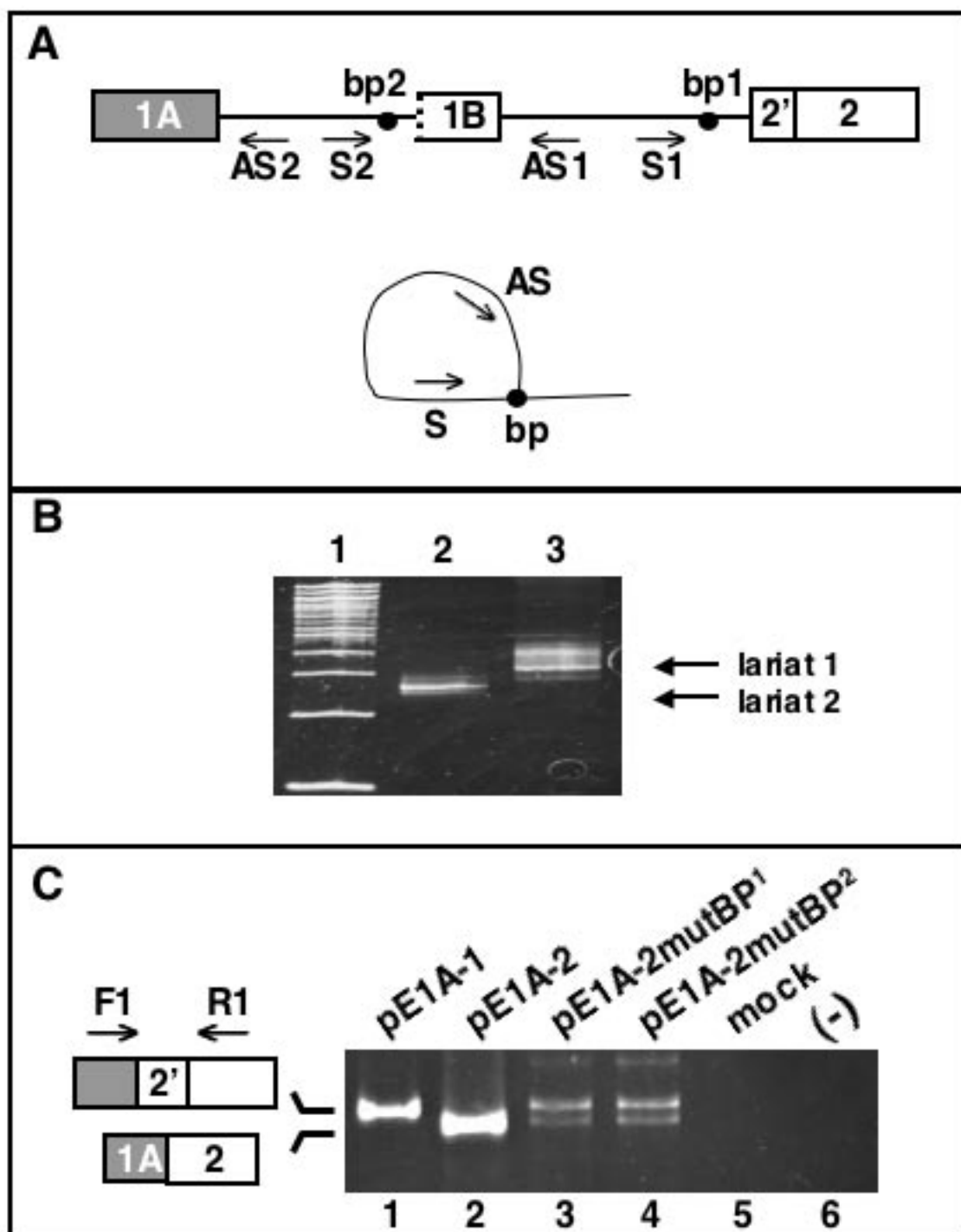
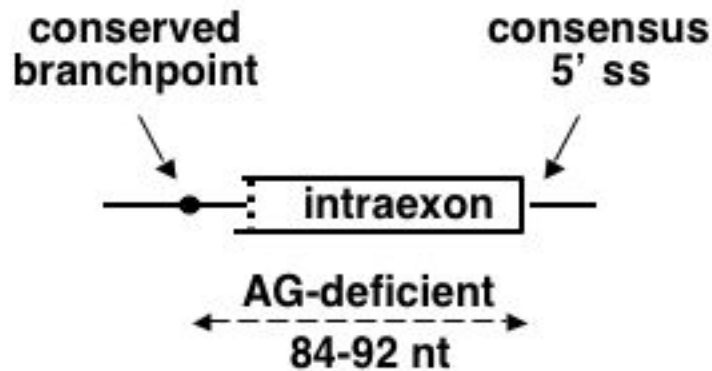


Figure 8



Organism	Branchpoint	distance	5' ss
Human	GTCTCAC	84	TGG/gtgagt
Mouse	GTCTCAC	87	TGG/gtgagt
Cow	GTCTCAC	86	TGG/gtgagt
Brown bat	GTCTCAC	88	TGG/gtgagt
Horse	GTCTCAC	88	TGG/gtgagt
Cat	GTCTC <u>A</u> T	89	TGG/gtgagt
Squirrel	<u>T</u> TCTCAC	92	TGG/gtgagt
Armadillo	GTCTCAC	89	TGG/gtgagt

Figure 9

