Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element

Gene S. Huh and Richard O. Hynes

Howard Hughes Medical Institute and Center for Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307 USA

The alternatively spliced exon EIIIB is regulated in a cell type-specific manner in the rat fibronectin gene. Splicing of EIIIB into fibronectin mRNA is dependent on sequences in the intron immediately downstream of EIIIB. We show that a short, highly repeated TGCATG motif in this intron is important for cell type-specific recognition of EIIIB as an exon. This motif enhances usage of the EIIIB 5' splice site; furthermore, this repeated TGCATG sequence can activate an alternatively spliced exon in the unrelated rat preprotachykinin pre-mRNA. Interestingly, this sequence can also be found within cis-acting elements identified previously in other alternatively spliced genes. This short repeated TGCATG motif is therefore a cell type-specific element that, in addition to controlling fibronectin alternative splicing, may participate in the regulation of other alternative RNA processing events.

[Key Words: Alternative splicing; fibronectin; EIIIB; hexanucleotide element]

Received February 22, 1994; revised version accepted May 16, 1994.

Alternative pre-mRNA splicing represents a fundamental mode of post-transcriptional metazoan gene regulation. This phenomenon generates distinct mRNAs from a single RNA transcript, often generating multiple proteins from a single gene in the process, and can be regulated in a developmental or cell type-specific fashion (for review, see Smith et al. 1989; McKeown 1992).

Splicing is a multistep process involving numerous proteins and small nuclear ribonucleoprotein particles (snRNPs) U1, U2, U4, U5, and U6; much has been learned about this reaction from biochemical studies utilizing simple model pre-mRNAs. However, considerably less is known about the faithful recognition or regulation of splice sites in vivo. Naturally occurring splice sites often vary from known consensus sequences; in addition, long pre-mRNA transcripts can harbor numerous splice site-like sequences, only a fraction of which are authentic signals. Therefore additional mechanisms must exist that specify introns and exons; these mechanisms must also exhibit flexibility for regulated splicing to occur (for review, see Green 1991; Moore et al. 1993).

In addition to its resemblance to a consensus, splice site usage can be affected by many diverse parameters. For example, internal exon identification in vertebrates may be aided by coordinate recognition of an exon's splice sites, a process termed exon definition [Robberson et al. 1990; Talerico and Berget 1990; Hoffman and Grabowski 1992]. This process may be compromised if the exon is of an inappropriate length [Robberson et al. 1990; Black 1991; Dominski and Kole 1991]. Splice site usage can also be affected by adjacent intron sequences [Helfman et al. 1990; Black 1991; Gallego et al. 1992], exon sequences [Somasekhar and Mertz 1985; Reed and Maniatis 1986; Mardon et al. 1987; Cooper and Ordahl 1989; Streuli and Saito 1989; Libri et al. 1990], or RNA secondary structure [Solnick 1985; Eperon et al. 1988; D'Orval et al. 1991; Libri et al. 1991]. Many parameters can therefore affect splicing in ways that are complex and not readily predictable.

In Drosophila melanogaster, sophisticated genetic analyses have identified genes that can encode negative or positive splicing regulators [for review, see Baker 1989; Mattox et al. 1992, McKeown 1992]. However, the identification of vertebrate splicing regulators has been more difficult. Because multiple pre-mRNA sequences affect splicing, it can often be difficult to distinguish between constitutive and regulatory elements by mutational analysis. Furthermore, the genetic methods that have identified Drosophila regulators cannot yet be applied easily to the identification of splicing regulators in vertebrates. Among the proteins implicated in the biochemistry of splicing, a number can affect splice site selection in vitro [Ge and Manley 1990; Krainer et al. 1990; Fu et al. 1992; Mayeda et al. 1993; Zahler et al. 1993], suggesting that differential concentrations of general splicing factors might regulate some alternative splicing events in vivo [Maniatis 1991].

1Corresponding author.
To address some of these questions, we have been studying the regulation of alternative splicing in the rat fibronectin (FN) gene. The EIIIB exon (also called ED2, EDB, or EDIIIB in this gene) is differentially spliced into FN mRNA to varying degrees in different cell types and developmental stages; for example, EIIIB is absent from adult liver FN mRNA but is present in nearly all early embryonic FN messages (for review, see Hynes 1990). Previous transfection studies have established that EIIIB is an inefficiently recognized exon and that its cell type-dependent inclusion into mRNA requires (1) a balance between competing splice sites in the EIIIB region, and (2) sequences located in the intron downstream of EIIIB (Huh and Hynes 1993).

The present study extends our previous work on EIIIB-activating intron elements. Surprisingly, we have found that EIIIB inclusion can be activated by repeated copies of a hexanucleotide sequence (TGCA TG) that normally occurs several times in the EIIIB-activating intron element. These TGCA TG repeats constitute a major requirement for cell type-specific EIIIB recognition and represent a novel example of mammalian splicing regulation. TGCA TG sequences can also direct cell type-specific regulation of splicing in the rat preprotachykinin (PPTK) gene transcript and, furthermore, can be found within elements identified in other alternatively spliced genes. Our findings regarding this hexameric control element have implications that may bear on the regulation of alternative splicing in general.

Results

A repeated sequence coincides with intron elements required for EIIIB usage

To study the regulation of EIIIB splicing, we used a three-exon minigene that contained the EIIIB region of the rat FN gene. Previous studies showed that this minigene, called 7iBi89, reproduced cell type-specific EIIIB regulation when expressed by a number of cell lines. In addition, these studies established that EIIIB inclusion requires cis-acting sequences that lie at least 445 bases away from EIIIB, spanning >500 bases in the 3’ region of the intron immediately downstream of EIIIB (called IVS2 in 7iBi89; Fig. 1). Deletions in this region caused near-complete EIIIB skipping in cell lines that normally splice EIIIB into mRNA; deletion of a particular section of this region (termed the intronic control region, or ICR) attenuated the cell type dependency of exon inclusion, indicating that cell type-specific elements existed within IVS2 (Huh and Hynes 1993). Inspection of the 7iBi89 sequence revealed nine GCATG repeats within the 1071-base IVS2 (Fig. 1). Eight of these have the sequence TGCA TG; five have the sequence TGCA TG. This clustering of short sequences was striking, because one would expect to find a given pentanucleotide only once every 1024 nucleotides by random chance. The distribution of these repeats, in addition, correlates well with the EIIIB-activating region and the ICR within IVS2. A comparison between rat and human FN genes revealed significant homology in the EIIIB-activating region of IVS2 relative to other regions in this intron (Fig. 1, bottom). In particular, seven of nine GCATG repeats are conserved. In previous studies, IVS2 deletions that reduced EIIIB inclusion removed at least one GCATG repeat; removal of more repeats had greater effects. Given (1) the remarkably high incidence of this short repeat in IVS2 and (2) the striking correlation between these repeats and conserved EIIIB-activating cis-acting sequences, we hypothesized that GCATG or TGCATG repeats mediate EIIIB inclusion.

TGCA TG repeats activate EIIIB exon recognition

To test whether TGCA TG sequences could activate EIIIB inclusion, we synthesized a DNA oligonucleotide that contained two TGCA TG hexamers (oligonucleotide T in Fig. 2A). Because the repeats in IVS2 were often followed by adenosine or cytidine nucleotides, oligonucleotide T contained one each of the sequences TGCA TG and TGCA TG, separated by pyrimidine-rich spacers. Up to three copies of oligonucleotide T were inserted in tandem into 7iBi89 ΔGA, a version of 7iBi89 from which 495 bases of IVS2 were removed; deletion ΔGA, which abolishes EIIIB inclusion (Huh and Hynes 1993), removed four endogenous GCATG repeats. As controls, we tested oligonucleotides that contained either mutant hexamers (TGCA TG; oligonucleotide M) or scrambled hexamers (AGTCGT; oligonucleotide S). Inserts were designated by name and copy number (e.g., insert T3 contains three copies of oligonucleotide T).

Minigenes were transfected stably into F9 teratocarcinoma cells, which normally exhibit high levels of EIIIB inclusion. S1 nuclease analysis of transfected RNA was carried out using a minigene-specific cDNA probe that contains the third 7iBi89 exon and overlaps EIIIB. The 7iBi89 minigene exhibited significant EIIIB inclusion, as assessed by the ratio of EIIIB + to EIIIB- mRNA (Fig. 3A, lane 1). EIIIB inclusion was reduced dramatically by the ΔGA deletion (lane 2). However, the insertion of one to three copies of oligonucleotide T into 7iBi89 ΔGA (two to six TGCA TG repeats) restored inclusion in an additive manner (lanes 3–5). With six TGCA TG repeats, EIIIB inclusion was indistinguishable from that exhibited by the undeleted minigene (cf. lanes 5 and 1). In contrast, insertion of repeated mutant or scrambled hexamers (lanes 6–11) did not restore EIIIB inclusion. Therefore, in F9 cells, repeated TGCA TG sequences can substitute for certain EIIIB-activating sequences in IVS2.

We then tested whether TGCA TG repeats could also substitute for other repeat-containing sequences in IVS2. The IVS2 deletion ΔAR2a removes sequences not deleted by ΔGA (Fig. 2B). ΔAR2a removes three TGCA TG repeats and also reduced EIIIB inclusion (Fig. 3A, lane 12). Insertion of four TGCA TG repeats in ΔAR2a (insert T2, lane 13) enhanced EIIIB inclusion to levels slightly higher than those of 7iBi89, an effect not observed with mutant hexamers (insert M2, lane 14). Transient transfection of ΔGA or ΔAR2a derivatives in F9 or 293 cells.
yielded similar results; comparison with cotransfected internal controls in 293 cells (which also normally include EIIIB) showed no gross differences in overall expression among minigenes [data not shown]. Thus, it is unlikely that these effects are attributable to differential expression levels. We conclude that repeated TGCATG sequences can substitute for either of two distinct EIIIB-activating, GCATG-containing regions in IVS2.

We then examined whether TGCATG repeats also had effects in EIIIB-skipping cell lines by transfecting these constructs into HeLa cells [which normally include EIIIB at low levels] and into COS cells [which completely skip EIIIB in 7iBi89]. In HeLa cells, TGCATG repeats in 7iBi89AGA stimulated EIIIB inclusion detectably (Fig. 3A, lanes 15–28). Six synthetic TGCATG repeats (T3) had weaker effects in HeLa than in F9 cells, although HeLa EIIIB inclusion was greater in 7iBi89AGA + T3 than in 7iBi89 [cf. lanes 15 and 19]. Otherwise, the effects of test and control repeats in HeLa cells paralleled those in F9 cells. In contrast, complete EIIIB skipping in COS cells was unaffected by deletions or repeat insertions in IVS2 [Fig. 3B, lanes 29–36]. Thus, TGCATG repeats activate the splicing of EIIIB only in cell types that normally exhibit EIIIB inclusion.

Activity of TGCATG repeats is affected by bases flanking the hexamer

We then tested whether TGCATG activity was affected by bases immediately flanking the repeat. The original test oligonucleotide (T) contained one each of the sequences TGCATGA and TGCATGC; as five repeats in the EIIIB-III-8a intron were TGCATGA sequences, we tested an oligonucleotide that contained two TGCATGA heptamers (TA; Fig. 2A). Another oligonucleotide (T) contained one each of the sequences TGCATGA and TGCATG_A; as five repeats in the EIIIB alternative splice patterns, the 3' endpoint of this region is undefined [Hub and Hynes 1993]. The solid area within this shaded bar represents a segment [the ICR] that is particularly important for EIIIB inclusion (see text). GCATG repeats are shown as vertical lines beneath 7iBi89. The nucleotide sequence below 7iBi89 is of IVS2 (1071 nucleotides), flanked by 10 bases each of exons EIIIB and III-8a. The rat sequence is shown [uppercase letters] bases conserved between rat and human FN genes; [lowercase letters], bases absent or not conserved in human, [dashes] gaps generated in comparing the two sequences. Nucleotide +1 is the first base in IVS2. Sequences important for EIIIB inclusion are enclosed within dashed lines; the ICR is shaded. Heptamers that contain GCATG repeats are enclosed in boxes. Letters in boldface type indicate stretches of rat–human homology [13 or more identical bases of 15]. Certain regions [bases +1 to +50 and +500 to +1071] contain several relatively uninterrupted stretches of rat–human sequence identity.
respect (Fig. 3B, lanes 3, 6, 7). Therefore, six alternating fine spacers (MP3) had no such effect (lane 9).

TGCATG and TGCATGC repeats stimulated EIIIB inclusion in the cell type-specific way. This suggests that TGCATG sequences (Huh and Hynes 1993) can substitute for ICR sequences. Neither mutant nor scrambled TGCATG repeats can restore cell type-specific exon usage.

In HeLa cells, inserts TA3 and TC3 had little effect on EIIIB splicing (Fig. 3B, lanes 15,16). As in F9 cells, insert TP3 stimulated HeLa EIIIB inclusion more effectively than did insert T3 [lane 17]. For any given minigene, HeLa inclusion was always lower than F9 inclusion. None of these inserts affected the complete EIIIB-skipping pattern exhibited by transiently transfected COS cells [lanes 21–24]. We conclude that although sequence context is not absolutely critical for hexamer function, repeat-dependent EIIIB inclusion is influenced to some extent by sequences flanking the TGCATG motif.

TGCATG repeats can restore cell type-specific exon usage

The above results suggested that the TGCATG sequence represents a cell type-specific element. However, it was possible that these repeats enhanced EIIIB inclusion non-specifically while other sequences provided cell type specificity. In past studies, the role of IVS2 sequences in cell type-specific EIIIB regulation was explored by using a chimeric FN-related minigene called M2. The M2 construct contains 7iBi89 sequences that spanned IVS2 [solid rectangles in Fig. 4A]. The upstream portion of M2 [open and hatched rectangles] is derived from a constitutively spliced region of the rat FN gene (see legend to Fig. 4). Regulated inclusion of the hybrid III-9/EIIIB central exon of M2 was detectable in all cell lines and also correlated with EIIIB regulation; exon inclusion was low in COS cells, intermediate in HeLa cells, and high in F9 cells (Fig. 4b). Certain IVS2 deletions in M2 reduced cell type-specific regulation by reducing inclusion in HeLa and F9 relative to COS cells. This effect was only observed when at least three GCATG repeats were removed, particularly those within the empirically determined ICR (Fig. 1). These data established that cell type-specific elements within IVS2 contained GCATG sequences (Huh and Hynes 1993). To address directly whether TGCATG repeats acted in a cell type-specific fashion, synthetic TGCATG-containing oligonucleotides were tested for their ability to substitute for ICR function in the M2 minigene.

The deletion of four GCATG repeats from M2 [using the IVS2 deletion ΔGA, which also removes the ICR; minigene M2ΔGA in Fig. 4A] significantly reduced inclusion in F9 and HeLa cells while leaving COS cell inclusion relatively unaffected, reducing the differential between F9 and COS cell inclusion from 11- to 3-fold (Fig. 4B). The insertion of two synthetic TGCATG hexamers (M2ΔGA + T1) modestly increased the F9/COS differential [from three- to fivefold]. Six repeats in the form of insert T3 [M2ΔGA + T3] enhanced M2 exon inclusion significantly in all three cell lines, even in COS cells. Because COS cell inclusion was also enhanced, insert T3 [in contrast to T1] did not increase the F9/COS differential significantly. Insert T3 also enhanced EIIIB inclusion HeLa cells to levels similar to F9. Therefore, a moderate number of TGCATG repeats can partially substitute for ICR sequences. Neither mutant nor scrambled repeats enhanced M2 exon usage (M2ΔGA + M1, M3,

---

### Figure 2

TGCATG repeats can substitute for the EIIIB-activating region. (A) (Left) The names of repeat-containing oligonucleotides used in this study. Monomeric forms of the inserts are shown at right; differences between each sequence and that of oligonucleotide T is shaded. (B) The 7iBi89 minigene and its deletion derivatives with single or multiple oligonucleotide inserts. Inserts were placed at the ΔGA or ΔAR2a deletion junctions.

| Oligo | Sequence |
|-------|----------|
| T     | GATCC TGCATGA CTACTGCT TGCATGC TCGTCATA |
| M     | GATCC TGCATGA CTACTGCT TGCATGC TCGTCATA |
| S     | GATCC TGCATGA CTACTGCT TGCATGC TCGTCATA |
| TA    | GATCC TGCATGA CTACTGCT TGCATGC TCGTCATA |
| TC    | GATCC TGCATGA CTACTGCT TGCATGC TCGTCATA |
| TP    | GATCC TGCATGA CTACTGCT TGCATGC TCGTCATA |
| MP    | GATCC TGCATGA CTACTGCT TGCATGC TCGTCATA |

| Minigene | EIIIB inclusion |
|----------|-----------------|
| 7iBi89   |                  |
| AGA     |                  |
| ΔGA+T1  |                  |
| ΔGA+T2  |                  |
| ΔGA+T3  |                  |
| ΔGA+M1  |                  |
| ΔGA+M2  |                  |
| ΔGA+M3  |                  |
| ΔGA+S1  |                  |
| ΔGA+S2  |                  |
| ΔGA+S3  |                  |
| ΔGA+TA3 |                  |
| ΔGA+TC3 |                  |
| ΔGA+TP3 |                  |
| ΔGA+MP3 |                  |
| ΔAR2a   |                  |
| ΔAR2a+T2|                  |
| ΔAR2a+M2|                  |

Table: TGCATG repeats can substitute for the EIIIB-activating region.

- **A**: (Left) The names of repeat-containing oligonucleotides used in this study. Monomeric forms of the inserts are shown at right; differences between each sequence and that of oligonucleotide T are shaded.
- **B**: The 7iBi89 minigene and its deletion derivatives with single or multiple oligonucleotide inserts. Inserts were placed at the ΔGA or ΔAR2a deletion junctions. GCATG repeats are shown as vertical lines. GCATG repeats are shown as vertical lines.

Each of the inserts TA3, TC3, and TP3 significantly enhanced EIIIB inclusion in F9 cells [Fig. 3B]. Surprisingly, neither TA3 nor TC3 worked as well as T3 in this respect [Fig. 3B, lanes 3, 6, 7]. Therefore, six alternating TGCATGA and TGCATGC repeats stimulated EIIIB inclusion to a greater extent than did six repeats of either sequence alone. Interestingly, insert TP3 [with purine spacers] enhanced EIIIB inclusion much more effectively than did T3 [cf. lanes 3 and 8]. Mutant repeats with purine spacers [MP3] had no such effect [lane 9].
Hexamer repeats regulate alternative splicing

Figure 3. S1 analyses of transfected 7iBi89 minigenes. (A) The EIIIB + 7iBi89 mRNA is shown at top left (conventions as in Fig. 1), along with the cDNA-derived 5’-end-labeled S1 probe, the electrophoretic positions of undigested probe [S1 probe] and of fragments generated from protection by EIIIB + or EIIIB - mRNAs are indicated at left. The ratio of these fragments (EIIIB + / EIIIB - ) indicates the level of EIIIB inclusion. Cell lines are indicated at the top. Cells were transfected with each of the indicated constructs [7iBi89AGA and 7iBi89AAR2a were abbreviated to AGA and AAR2a, respectively; (Ctrl) untransfected cells]. Shown are the effects of inserts derived from oligonucleotides T, M, and S (see Fig. 2A for sequences). For analysis, cytoplasmic RNA (from stably transfected F9, HeLa, and COS cells) or total RNA (from transiently transfected COS cells) was used. [Lane M] End-labeled molecular weight markers (pBR322/MspI fragments). [B] S1 analyses of transfections using minigenes containing inserts TA3, TC3, TP3 and MP3 (see Fig. 2A for sequences). Labeling conventions are as in A.

The effects of variant inserts upon M2AGA splicing were found to differ from each other, both quantitatively and qualitatively. Insert TC3 (with six TGCATGC repeats) had effects similar to those of insert T3. In contrast, insert TA3 (with six TGCATGA repeats), in addition to enhancing the F9/COS differential (from three- to fivefold), exhibited a pattern of cell type specificity most closely resembling that of intact M2, with inclusion levels in HeLa being intermediate between those in F9 and COS. Insert TP3 (with purine spacers) enhanced the F9/COS differential to the greatest extent (from 3- to 10-fold). Insert MP3 did not exhibit this effect. We conclude

+ S1, and + S3); rather, inclusion was nonspecifically reduced by these control inserts.
Figure 4. TGCATG repeats regulate cell type-specific exon recognition. (A) The M2 and M2ΔGA minigenes. The portion of M2 derived from 7iBi89 is indicated by solid rectangles (exons) and solid bars (introns). The upstream portion of M2 (hatched and open rectangles and bars indicate exons and introns, respectively) is derived from a constitutively spliced region of the rat FN gene and contains exon III-8b, the III-8b/III-9 intron, and part of exon III-9. GCATG repeats are indicated as vertical lines within IVS2. Oligonucleotides were inserted at the AGA deletion junction. (B) Histograms indicating the amount of central exon inclusion exhibited by each minigene (exon+/exon− ratios); error bars represent standard deviations from transient transfections that were carried out at least twice for COS and HeLa cells and at least three times for F9 cells. (Open bars) COS cells; (shaded bars) HeLa cells; (solid bars) F9 cells. The F9/COS differential for each construct (calculated as a ratio between F9 and COS values) is shown below each triplet histogram set.

that TGCATG repeats are capable of cell type-specific action and that cell type specificity is influenced by hexamer context.

TGCATG repeats promote EIIIB 5′ splice site usage

In general, IVS2 deletions shown previously to reduce exon inclusion removed, at most, four GCATG repeats [Huh and Hynes 1993]. When we deleted seven IVS2 repeats from the M2 minigene, an unusual splicing phenotype resulted. This phenotype was detected by using an RNase protection probe that spanned the central exon of M2 (Fig. 5A). Differently sized probe products were obtained, depending on which of the central exon splice sites [3′ or 5′] were used. Analysis of M2 expression in COS or HeLa cells yielded primarily a product corresponding to usage of both splice sites of the M2 central exon (Fig. 5B, lanes 1,10). In contrast, when a large deletion removing seven repeats was tested (M2AL2R2a; Fig. 5A), this fragment was not detected, instead, an IVS2-containing RNA accumulated (Fig. 5B, lanes 5,14). This RNA, which lacks IVS1, did not accumulate if three- or four-repeat sections of the AL2R2a region were deleted (M2AR2a, lanes 4 and 13; M2AL2, lanes 3 and 12), nor did it accumulate if a large deletion that only removed four repeats was tested (M2GA, lanes 2 and 11). Other RNase protection assays indicated that the splice sites of both the upstream and downstream flanking exons were utilized in all of these constructs, although usage of the downstream exon 3′ splice site was only partial for M2AL2R2a [data not shown]. Taken together, these results suggested that when seven repeats were deleted, IVS1 splicing and exon skipping still occurred but that the EIIIB-derived 5′ splice site of the M2 central exon was not used, resulting in production of a partially spliced RNA (summarized in Fig. 5A).

As these results suggested that GCATG repeats were required for EIIIB 5′ splice site usage, we analyzed the expression of M2AL2R2a derivatives that contained synthetic repeat inserts T3, M3, TP3, or MP3. Both minigenes M2AL2R2a+T3 and M2AL2R2a+TP3 expressed detectable amounts of completely spliced M2 exon, although partially spliced RNA was still present (Fig. 5B, lanes 6,8,15,17). Mutant repeats did not exhibit this effect [M2AL2R2a+M3 or M2AL2R2a+MP3; lanes 7,9,16,18]. Splice site usage of the flanking exons in
M2AL2R2a remained relatively constant regardless of insert identity (data not shown). Therefore, TGCATG repeats can partially rescue EIIIB 5' splice site usage. Rescue by insert TP3 (as assessed by ratios of partially to completely spliced RNA) was slightly greater than that by insert T3, particularly in HeLa cells, thus correlating with the effects of these inserts in 7iBi89AGA [Fig. 2]. Because only partial rescue was effected by six TGCATG repeats, it is possible that additional sequences in IVS2 are required for proper 5' splice site usage; alternatively, partial effects may have been a consequence of inappropriate synthetic repeat positioning within this intron. Nonetheless, these results suggest that one function of TGCATG sequences is to promote more efficient usage of the EIIIB 5' splice site.

**TGCATG repeats can activate a heterologous alternatively spliced exon**

We then tested whether these repeats could activate exons unrelated to EIIIB by testing TGCATG inserts in the rat PPTK gene transcript. The PPTK gene contains two alternative cassette-type exons, E4 and E6. We focused on exon E4, as E4 is only spliced into PPTK mRNA ~25% of the time in a wide range of cell types [Krause et al. 1987; Nasim et al. 1990]. Therefore, any positive effects of TGCATG repeats on E4 selection should be detectable. If so, it should also be possible to assess whether TGCATG-dependent exon activation is cell type dependent. We transferred a number of oligonucleotide inserts into a minigene expressing exons E2–E7 of the rat PPTK gene transcription unit; inserts were positioned in the intron between E4 and E5 [Fig. 6A]. RNA from transiently transfected COS, HeLa, NIH-3T3, and F9 cells was analyzed for E4 splicing by RNase protection, using a probe derived from a PPTK cDNA containing exons E1–E5.

As expected, E4 inclusion in the unmodified PPTK minigene was relatively low in these four cell types (Fig. 6B, lanes 1, 7, 13, 19). Although some quantitative differ-
Figure 6. TGCATG repeats function in a heterologous gene. (A) Diagram of the rat PPTK minigene, with the locations of alternatively spliced exons E4 (shaded) and E6. This minigene contains the SV40 early promoter and 3’ poly[A] signals [Nasim et al. 1990]. Other labeling conventions are as in Fig. 1. Synthetic inserts were placed at a BglII site (asterisk) within the E4-E5 intron, located 136 nucleotides downstream of E4 and 328 nucleotides upstream of E5. (B) RNase protection analysis of the transfected PPTK minigene and its derivatives. A cDNA containing PPTK exons E1–E5 was used as a probe. Analyses were carried out on total RNA from untransfected cells (Ctrl) or from transiently transfected COS, HeLa, 3T3, or F9 cells (cell lines are indicated above each panel; transfected minigenes are indicated above each lane). The protected fragments corresponding to E4+ and E4− mRNAs are indicated at left. The full-length riboprobe (Probe) is large and migrates much more slowly than the E4+/E4− fragments and is not shown here.

ences were apparent, E4 inclusion did not differ drastically between cell lines. Insert T3 was found to have a significant positive effect on E4 inclusion [PPTK + T3; lanes 3, 9, 15, 21] relative to the control inserts M3 (lanes 5, 11, 17, 23) or S3 (lanes 6, 12, 18, 24). The effects of two-repeat inserts were also tested (T1 vs. M1). Enhancement of E4 usage by insert T1 was slight but significant compared with insert M1. The degree of exon activation ranged from low (COS) to very high (HeLa, F9). The degree of TGCATG-dependent E4 activation among COS, NIH-3T3, and F9 cells [COS < T3 < F9; Fig. 6B] correlated with EIIIB regulation in 7bBi89 (Huh and Hynes 1993). We note that although EIIIB inclusion is normally low in HeLa cells, TGCATG-dependent E4 activation in HeLa cells was unexpectedly high. We conclude that TGCATG repeats are capable of activating alternatively spliced exons unrelated to EIIIB. Additionally, these results establish that TGCATG repeats provide at least part of the cell type specificity of EIIIB regulation in the FN gene.

Discussion

We have identified a short hexanucleotide element (TGCATG) that activates the splicing of the EIIIB exon in the rat FN gene. This element, as a repeated array, can also promote the cell type-specific inclusion of a heterologous alternatively spliced exon. This novel short motif therefore regulates splice site selection when present in alternatively spliced pre-mRNAs.

TGCATG sequences in IVS2 activate splicing of EIIIB

The EIIIB minigene 7bBi89 exhibited high levels of EIIIB inclusion in F9 cells, low inclusion in HeLa cells, and complete EIIIB skipping in COS cells. Previous studies identified a region in the intron downstream of EIIIB [IVS2] that was important for the recognition of EIIIB for splicing (Huh and Hynes 1993). Several TGCATG repeats are present in this relatively large (>500 nucleotide) region (Fig. 1). Deletions of separate repeat-containing regions in IVS2 abolished EIIIB inclusion in F9 cells. Replacement of these deleted regions by synthetic TGCATG repeats restored EIIIB inclusion in F9 cells (Figs. 2 and 3). Repeats had reduced effects in HeLa cells (which normally express low EIIIB inclusion) and no effect in COS cells (which skip EIIIB completely); repeat activity therefore appeared to vary according to cell type. In a related FN minigene, M2 (Fig. 4), deletion of certain TGCATG-containing IVS2 sequences reduced F9 and HeLa exon inclusion significantly while having little effect on the low level of inclusion in COS cells. The effects of this deletion could be reversed by reinserion of an appropriate number of synthetic TGCATG repeats, confirming that these repeated sequences are capable of cell type-dependent action.

TGCATG sequences also activated PPTK exon E4 inclusion when placed downstream of E4 (Fig. 6). Therefore, exons EIIIB and E4 may both be intrinsically poorly recognized for similar reasons. In vitro studies of PPTK alternative splicing have indicated that the 5′ splice site of E4 binds U1 snRNP poorly, as a consequence, recruitment of the splicing factor U2AF to the E4 3′ splice site

1568 GENES & DEVELOPMENT
is also compromised [Kuo et al. 1991; Hoffman and Grabowski 1992]. In addition, mutations in the E4 or EIIIB 5' splice sites that improve donor complementarity to U1 snRNA can also increase inclusion significantly [Nasim et al. 1990; Huh and Hynes 1993]. Because TGCATG repeats can activate the EIIIB 5' splice site [Fig. 5], TGCATG sequences may act by facilitating U1--5' splice site interactions and the exon selection events that follow.

Sequence specificity of the element

Besides the TGCATG consensus, no other obvious inter-repeat homologies are apparent in IVS2 [Fig. 1]. Because hexamer elements can act from within different sequence contexts [Figs. 2, 3B, and 6], specific RNA secondary structures are probably not critical for hexamer function. Therefore, it is likely that the TGCATG motif comprises (or contains) the essential cis-active element.

Both the degree and cell type specificity of repeat-dependent exon inclusion can be affected by sequences flanking the synthetic hexamers. For example, TGCATG sequences were most active in a purine-rich context. Interestingly, three TGCATG/TGCATGC repeat pairs stimulated EIIIB inclusion more than either six TGCATGAs or six TGCATGCs [Figs. 2 and 3]. This finding could reflect either subtle context effects or a synergistic interaction between distinct repeat-binding factors with overlapping sequence specificities. These context-dependent differences could also reflect differential hexamer accessibility or factor affinity. Cell type specificity was mimicked most closely by TGCATG repeats (Fig. 4), which could explain why a majority of IVS2 repeats have this sequence. Many of the naturally occurring GCATG repeats in IVS2 are adjacent to evolutionarily conserved sequences [Fig. 1], possibly reflecting a requirement for appropriate context.

Taken together, these data indicate that repeated TGCATG hexamers can account for the activity of most, if not all, of the cis-acting elements identified previously within IVS2. Although the presence of other EIIIB-regulating elements cannot be ruled out, we have shown that TGCATG sequences can regulate EIIIB if present in enough copies and in the proper sequence context. Therefore, TGCATG repeats play a major essential role in the splicing of EIIIB into FN mRNA.

Is the TGCATG element a cell type-specific splicing signal?

Whereas a constitutive signal might be expected to exert equal effects in all cell types, a genuine regulatory element should exert cell type-dependent effects. The repeated TGCATG element meets this latter criterion in three separate minigenes. First, TGCATG repeats activated EIIIB usage in a minigene in which inclusion was compromised by an IVS2 deletion [7Bi89ΔGA]. TGCATG-dependent EIIIB inclusion was highest in F9, lower in HeLa and undetectable in COS cells, correlating with the ability of each cell line to utilize EIIIB. Furthermore, in two other minigenes that exhibited detectable inclusion in all cell lines [M2 and PPTK], TGCATG repeats again had cell type-dependent effects [Figs. 4 and 6]. Therefore these repeats appear to comprise cell type-specific control signals rather than constitutively recognized elements.

We note that although HeLa cells normally exhibit low EIIIB inclusion, the response of HeLa cells to TGCATG repeats in many cases was disproportionately high. For example, PPTK E4 activation was unexpectedly high in HeLa cells, despite apparently EIIIB-like regulation of E4 in the other cell lines tested [Fig. 6]. Hence, it is possible that other FN gene sequences contribute to EIIIB regulation, either by acting independently or perhaps by affecting hexamer context.

A number of studies in other vertebrate genes have identified sequences that control splicing. Secondary structure and intron sequences near exon 7 in the rat β-tropomyosin gene (exon 6B in chicken) can repress usage of this exon [Goux-Pelletan 1990; Helfman et al. 1990; D’Orval et al. 1991; Guo et al. 1991; Libri et al. 1991; Gallego et al. 1992]. An element that inhibits splicing in Rous sarcoma virus can also function in a heterologous intron [Arrigo and Beemon 1988; McNally et al. 1991]. Elements that activate splicing in the c-src gene and others that inhibit splicing in the calcitonin/CGRP systems have also been described [Emeson et al. 1989; Black 1992]. For many of these cases, minimal essential elements that affect splicing have yet to be defined; however, parallels may exist between regulation of the c-src and calcitonin/CGRP systems and the regulation of EIIIB, as will be commented on below.

In comparison, the regulatory element we have identified is novel in a number of respects. Repeated TGCATG sequences are capable of functioning in a cell type-dependent fashion in a heterologous context, a finding that has not been established conclusively for many of the other potential vertebrate elements reported to date. In addition, the remarkably diminutive character of this repeat may have widespread implications for regulated alternative splicing.

Short repeats in RNA biology

Given the complexity of heterogeneous nuclear RNA [hnRNA], a single UGCAUG sequence would have little informational capacity to recruit trans-acting factors unless additional signals were present. Therefore, it is not surprising that this motif is repeated extensively in IVS2. The role of repeated sequences in RNA splicing has been clearly documented in one other system, the regulation of Drosophila doublesex [dsx] pre-mRNA splicing by the transformer [tra] and transformer-2 [tra-2] genes. The tra and tra-2 gene products [Tra and Tra-2] positively regulate female-specific dsx splice acceptor usage; Tra and Tra-2 interact with the dsx pre-mRNA via six 13-nucleotide repeats located downstream of the female-specific acceptor [Hedley and Maniatis 1991; Hoshijima et al. 1991; Ryner and Baker 1991; Inoue et al. 1992; Tian and Maniatis 1992, 1993]. In contrast, the repeats that we
Document are shorter and can promote 5′ splice site usage. Interestingly, each dsrc 13-mer repeat can be depicted as two hexamers separated by an adenosine residue, the first hexamer being a variable copy of the second [UCaCA A UCAACA; Burtis and Baker 1989; Inoue et al. 1992].

Other examples of reiterated RNA elements include the AUUUA destabilizing sequence found in the 3′ untranslated regions of granulocyte–macrophage colony-stimulating factor (GM-CSF) and other short-lived mRNAs [Shaw and Kamen 1986]. The human immunodeficiency virus Rev protein may interact with multiple sites within its RNA target element [Kjems et al. 1991]. In Drosophila, the Sex-let al [Sxal] gene product blocks male-specific splicing of the tra gene by binding to a single critical U-rich element located at the tra male-specific 3′ splice site [Sosnowski et al. 1989; Inoue et al. 1990; Valcarcel et al. 1993], autoregulation by the Sxal protein of its own pre-mRNA, on the other hand, may involve interactions with multiple poly(U) intron elements in the Sxal transcript [Sakamoto et al. 1992; Horabin and Schedl 1993]. Recently, an exonic purine-rich element has been identified in a number of genes that promotes recognition of both constitutive and alternative exons, apparently by aiding 3′ splice site usage. This element, although degenerate and somewhat variable in length, is repeated in some exons and can also promote heterologous exon recognition [Lavigueur et al. 1993; Sun et al. 1993; Watakabe et al. 1993; Xu et al. 1993; Yeakley et al. 1993]. Regulation of RNA metabolism by short repeated elements is therefore an increasingly recognized phenomenon.

**How might TGCATG elements work?**

The removal of seven GCATG repeats from IVS2 resulted in inactivation of the EIIIB splice donor, a phenotype that was partially rescued by synthetic TGCATG repeats (Fig. 5). Therefore these elements may recruit factors that directly activate nearby 5′ splice sites [Fig. 7A]. As discussed previously, activation of EIIIB [or PPTK exon E4] could occur by repeat-facilitated U1 snRNP binding. Alternatively, these repeats may render IVS2 more accessible for splicing, perhaps by recruiting factors that configure this intron appropriately. A number of hnRNP proteins exhibit sequence or transcript specificity [Swanson and Dreyfuss 1988; Bennett et al. 1992; Matunis et al. 1993]; some of these may prefer to bind UGCAUG. Other possible UGCAUG-binding candidates include members of the SR protein family, which are splicing factors that can promote proximal 5′ splice site selection in cell-free extracts [Fu et al. 1992; Zahler et al. 1992]. SR proteins have been implicated in Drosophila dsrc regulation in vitro [Tian and Maniatis 1993] and may also interact with the purine-rich exon recognition element [Lavigueur et al. 1993]. Different SR proteins have also been shown to exhibit substrate specificity in vitro, either for commitment to splicing [Fu 1993] or for proximal 5′ splice site selection [Zahler et al. 1993]; these observations may reflect distinct sequence specificities among individual SR family members.

Alternatively or additionally, these repeats may inhibit or delay the usage of the IVS2 splice acceptor (which is used regardless of splice choice), allowing more opportunity for EIIIB recognition [Fig. 7A], this might explain why several repeats are located in the 3′ portion of IVS2. Regulation of this type is not without precedent,
Hexamer repeats regulate alternative splicing

because 5' splice site selection in SV40 and adenovirus pre-mRNAs can also be affected by alterations at a common 3' splice site (Fu and Manley 1988; Ulfendahl et al. 1989). In principle, it is also possible that these repeats represent transcriptional pause sites, in which case a delay in IVS2 acceptor commitment might result from pausing of RNA polymerase II during transcription of the repeat-rich region within IVS2.

In any event, it is likely that differentially expressed cellular factors mediate these repeat-dependent effects, either by interacting with hexamer targets directly or with ubiquitous hexamer-binding proteins. To distinguish among these models, experiments are currently in progress in an attempt to reproduce hexamer-dependent regulation in vitro. These experiments, in addition to those characterizing hexamer-specific cellular factors, should offer further insights regarding mechanisms of alternative splicing.

Do TGCATG sequences regulate other alternatively spliced genes?

Given our observations, short sequences could, in principle, govern regulation in other alternatively spliced genes. The TGCATG element occurs in at least two other instances of splice regulation. Studies of the c-src gene have identified at least two intron elements, located downstream of the alternatively spliced N1 exon, that are important for neural-specific N1 exon inclusion (Black 1992). It is interesting to note that a TGCATG sequence occurs in one of these elements and that a mutation that reduced N1 inclusion also disrupted this hexamer [Fig. 7B]. Moreover, extensive deletions that removed both intron elements resulted in production of unspliced and partially spliced RNA [Black 1992], a result similar to that obtained upon removal of seven repeats from FN minigenes (our results). These parallels suggest that TGCATG-related sequences may participate in c-src splicing regulation, probably in conjunction with additional neural-specific elements.

In the calcitonin/α-CGRP gene, HeLa and thyroid C cells produce calcitonin by using exon 4, whereas neurons and F9 cells produce calcitonin gene-related peptide (CGRP) by using exons 5 and 6 instead [Leff et al. 1987; Emeson et al. 1989; Fig. 7B]. Regulation is evident at the calcitonin-specific exon 4 splice acceptor, which is used in HeLa cells but is inactive in F9 cells [Emeson et al. 1989]. We note that eight TGCATG repeats are clustered within 700 bases surrounding this splice acceptor in the human gene [Fig. 7C]. Five of these are TGCATG. In studies using the rat gene, derepression of the exon 4 acceptor in F9 cells could be induced by nearby upstream mutations, a number of which removed one or more TGCATG repeats [Emeson et al. 1989]. Although a more recent and comprehensive analysis of the rat gene failed to identify any single element that was consistently critical for regulation [Yeakley et al. 1993], the influence of multiple dispersed elements could explain these recent findings, particularly if such elements were functionally redundant. CGRP-specific splicing in F9 cells might then occur by repeat-dependent down-regulation of the exon 4 splice acceptor, similar to the way in which these repeats may attenuate the 3' splice site downstream of EIIIB in the same cells [Fig. 7A]. Although there is as yet no conclusive evidence that these repeats regulate calcitonin/CGRP processing, it is nevertheless intriguing that multiple copies of a splicing signal for EIIIB should occur (at a much higher density than predicted by chance) in a gene characterized by a different mode of regulated alternative splicing. It seems unlikely that TGCATG-dependent splicing regulation is unique to EIIIB, given these correlations among FN, c-src and calcitonin/CGRP, it is tempting to speculate that TGCATG-related motifs may participate in a range of embryo- or neural-specific RNA processing events.

Our findings regarding EIIIB have allowed us to achieve a more refined understanding of splicing regulation for one vertebrate alternative exon. Perhaps more important, however, are the novel insights that these findings also provide with regard to splice site selection mechanisms in general. Extensive research has revealed a plethora of alternatively spliced genes that are regulated in many different ways. The regulation of many of these could conceivably be a functional consequence of the collective action of multiple short sequences that are combinatorially unique to each system. Such a fascinating possibility would pose many interesting questions for future consideration.

Materials and Methods

Oligonucleotides and computer analyses

Each double-stranded DNA oligonucleotide consisted of an annealed sense and antisense oligonucleotides. Sense strand sequences are shown in Fig. 2A; antisense oligonucleotides consisted of the sequence 5'-GATC-3' followed by the reverse complement to nucleotides 5-35 of each sense sequence.

Comparisons of rat FN [Schwarzbauer et al. 1987, R. Patel-King, J.E. Schwarzbauer, P.A. Norton, and R.O. Hynes, unpubl.; GenBank accession number L20801] and human FN genomic sequences [Paolella et al. 1988; GenBank accession number X07717] were carried out using version 7.0 of the UWCGG software package [Devereux et al. 1984]; the comparison shown in Figure 1 was adapted from the output from a BESTFIT comparison (using default parameters). Analysis of repeats in the rat FN gene and in the human calcitonin/α-CGRP gene [Broad et al. 1989, GenBank accession number X15943] was carried out using the nucleotide interpretation program of the STADEN software package [Staden 1990].

DNA constructs and probes

DNA manipulation procedures were as described previously [Ausubel et al. 1987; Sambrook et al. 1989]. Restriction enzymes were obtained from New England Biolabs. Minigenes 7ibi89 and M2 in expression vectors pBAGH or pBAGH5v, along with their deletion derivatives and the S1 analysis probe, have been described previously (Huh and Hynes 1993). The deletions in this study (in 7ibi89 or in M2) removed the following nucleotides from the 1071 base IVS2 (numbering as in Fig. 1): ΔGA, bases 235-729, ΔAR2a, bases 734-1039, ΔAL2, bases 417-729, ΔAL2R2a, bases 417-1039.
The following procedure was carried out with each of the DNA oligonucleotide pairs to create single or multiple inserts in 7IBi89AGA. After gel purification, 800 pmoles of each of the sense and antisense oligonucleotides were annealed [in a total volume of 40 µl TE (10 mM Tris-HCl, 1 mM EDTA at pH 8.0)] by cooling slowly from 90°C to 30°C. Twenty-microliter aliquots of the annealed mixture were then phosphorylated in a total volume of 40 µl using 40 units of 3'-phosphatase-free polynucleotide kinase (Boehringer Mannheim). After phenol/chloroform [PC] extraction and ethanol precipitation, the mixture was precipitated, and digested with BamHI and BglII to eliminate all but head-to-tail ligation joins [for inserts TP and MP, BglII was used instead of BamHI]. This ligation-digestion cycle was repeated two more times to ensure an adequate production of appropriately ligated multimers. The final digest was then blunt-ended with Klengon fragment and dNTPs; this mixture was then ligated to AffIII-cut, blunt-ended 7IBi89AGA/pBAGH. DNAs from transformed colonies were sequenced to ensure appropriate insert size, identity, and orientation. 7IBi89AGA/pBAGH.Sv derivatives were made by transferring a XhoI-EcoRI fragment [containing the entire transcription unit] from 7IBi89AGA/pBAGH into pBAGH.Sv.

Inserts were excised from 7IBi89AGA/pBAGH via BglII excision; these were ligated to BglII-cut M2 GA/pBAGH.Sv or M2 AL2R2a/pBAGH.Sv to make the insert-containing M2AGA and M2 AL2R2a derivatives. The RNase protection probe used for assessing M2 central exon splice site usage (Fig. 5) was derived from a HindIII–PstI fragment of the M2 minigene that contains the central exon [in pBluescript SK(–); Stratagene]. KpnI-linearized plasmid was the template for T3 RNA polymerase [Stratagene].

PPTK minigene constructs and probes were derived from pPSVPαA-2’-7’ [a plasmid containing exons E2–E7 of the rat PPTK gene inserted between SV40 promoter and poly(A) signals] and pRPC1, a β-PPTK cDNA clone in pBS (Stratagene). Oligonucleotide inserts were inserted into a BglII site in pPSVPαA-2’-7’, in the intron between E4 and E5 [Fig. 6A]. The RNase probe used for assessing E4 inclusion was made by cutting pRPC1 with BglII and BamHI and religating the vector, generating a subclone containing exons E1–E4 and part of E5. This was linearized with EcoRI and transcribed with T3 RNA polymerase.

Cell culture and RNA analysis

Cell culture methods, transfection protocols, and RNA analysis by S1 nuclease or RNase protection assays were as described previously (Huh and Hynes 1993). S1 analysis was carried out using 5′ end-labeled probes, whereas RNase protection analyses were carried out with uniformly labeled riboprobes. Quantitation of bands was carried out using a Molecular Dynamics PhosphorImager.

Acknowledgments

We are grateful to Paula Grabowski for providing us with the rat preprotachykinin plasmids used in some of these studies. We also wish to thank Ben Blencowe, Melissa Moore, Charles Query, Phillip Sharp, Peggy Kolm, Mark Borowsky, Grant Wheeler, and Pamela Norton for helpful discussions and criticisms of the manuscript. This work was supported by a grant from the National Cancer Institute (U.S. Public Health Service grant RO1 CA 17007) and by the Howard Hughes Medical Institute. G.S.H. was a Howard Hughes Medical Institute predoctoral fellow, and R.O.H. is a Howard Hughes Medical Institute Investigator.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Arrigo, S. and K. Beemon. 1988. Regulation of Rous sarcoma virus RNA splicing and stability. Mol. Cell. Biol. 8: 4858-4867.

Auszubel, F.M., R. Brent, R.E. Kingston, J.G. Seidman, J.A. Smith, and K. Struhl. 1987. Current protocols in molecular biology John Wiley/Greene, New York.

Baker, B.S. 1989. Sex in flies: The splice of life. Nature 340: 521-524.

Bennett, M., S. Pinol-Roma, D. Staknis, G. Dreyfuss, and R. Reed. 1992. Transcript-dependent packaging of pre-mRNA in hnRNP complexes prior to spliceosome assembly in vitro. Mol. Cell. Biol. 12: 3165–3175.

Black, D.L. 1991. Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells? Genes & Dev. 5: 389–402.

Black, D.L. 1992. Activation of c-src neuron-specific splicing by an unusual RNA element in vivo and in vitro. Cell 69: 795–807.

Broad, P.M., A.J. Symes, R.V. Thakker, and R.K. Craig. 1989. Structure and methylation of the human calcitonin/α-CGRP gene. Nucleic Acids Res. 17: 6999–7011.

Burtis, K.C. and B.S. Baker. 1989. Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell 56: 997–1010.

Cooper, T.A. and C.P. Ordahl. 1989. Nucleotide substitutions within the cardiac troponin T alternative exon disrupt pre-mRNA alternative splicing. Nucleic Acids Res. 17: 7905–7921.

Devereux, J., P. Haebeler, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12: 387–395.

Domingik, Z. and R. Kole. 1991. Selection of splice sites in pre-mRNAs with short internal exons. Mol. Cell. Biol. 11: 6075–6083.

D’Orval, B.C., Y.d’A. Carafa, P. Sirard-Pugnet, M. Gallego, E. Brody, and J. Marie. 1991. RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts. Science 252: 1823–1828.

Emeson, R.B., F. Hedjran, J.M. Yeakley, J.W. Guise, and M.G. Rosenfeld. 1989. Alternative production of calcitonin and CGRP mRNA is regulated at the calcitonin-specific splice acceptor. Nature 341: 76–80.

Eperon, I.C., I.R. Graham, A.D. Griffiths, and I.C. Eperon. 1988. Effects of RNA secondary structure on alternative splicing of pre-mRNA: Is folding limited to a region behind the transcribing RNA polymerase? Cell 54: 393–401.

Fu, X.D. 1993. Specific commitment of different pre-mRNAs to splicing by single SR proteins. Nature 365: 82–85.

Fu, X.D., A. Mayeda, T. Maniatis, and A.R. Krainer. 1992. General splicing factors SF2 and SC35 have equivalent activities in vitro, and both affect alternative 5′ and 3′ splice site selection. Proc. Natl. Acad. Sci. 89: 11224–11228.

Fu, X.Y., H. Ge, and J.L. Manley. 1988. The role of the polyppy-
rimidine stretch at the SV40 early pre-mRNA 3' splice site in alternative splicing. *EMBO J.* 7: 809–817.

Gallego, M.E., L. Balvay, and E. Brody 1992. Cis-acting sequences involved in exon selection in the chicken β-tropomyosin gene. *Mol. Cell. Biol.* 12: 5415–5425.

Ge, H. and J.L. Manley. 1990. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell* 62: 25–34.

Goux-Pelletan, M., D. Libri, Y.d'A Carafa, M. Fiszman, E. Brody, and J. Marie. 1990. In vitro splicing of mutually exclusive exons from the chicken β-tropomyosin gene: Role of the branch point location and very long pyrimidine stretch. *EMBO J.* 9: 241–249.

Green, M.R. 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu. Rev. Cell Biol.* 7: 559–599.

Guo, W., G.J. Mulligan, S. Wormsley, and D.M. Helfman. 1991. Alternative splicing of β-tropomyosin pre-mRNA: Cis-acting elements and cellular factors that block the use of a skeletal muscle exon in nonmuscle cells. *Genes & Dev.* 5: 2096–2107.

Hedley, M.L. and T. Maniatis. 1991. Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that binds specifically to tra-2 protein in vitro. *Cell* 65: 579–586.

Helfman, D.M., R.F. Rosciglio, G.J. Mulligan, L.A. Finn, and K.S. Weber. 1990. Identification of two distinct intron elements involved in alternative splicing of β-tropomyosin pre-mRNA. *Genes & Dev.* 4: 98–110.

Hoffman, B.E. and P.J. Grabowski. 1992. U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes & Dev.* 6: 2554–2568.

Horabin, J.I. and P. Schedl. 1993. Sex-lethal autoregulation requires multiple cis-acting elements upstream and downstream of the male exon and appears to depend largely on controlling use of the male exon 5' splice site. *Mol. Cell. Biol.* 13: 7734–7746.

Hoshijima, K., K. Inoue, I. Higuchi, H. Sakamoto, and Y. Shimura. 1991. Control of doublesex alternative splicing by transformer and transformer-2 in *Drosophila*. *Science* 252: 833–836.

Huh, G.S. and R.O. Hynes. 1993. Elements regulating an alternatively spliced exon of the rat fibronectin gene. *Mol. Cell. Biol.* 13: 5301–5314.

Hynes, R.O. 1990. *Fibronectins*. Springer-Verlag, New York.

Inoue, K., K. Hoshijima, H. Sakamoto, and Y. Shimura. 1990. Binding of the *Drosophila* Sex lethal gene product to the alternative splice site of transformer primary transcript. *Nature* 344: 461–463.

Inoue, K., K. Hoshijima, I. Higuchi, H. Sakamoto, and Y. Shimura. 1992. Binding of the *Drosophila* transformer and transformer-2 proteins to the regulatory elements of doublesex primary transcript for sex-specific RNA processing. *Proc. Natl. Acad. Sci.* 89: 8092–8096.

Kjemps, J., M. Brown, D.D. Chang, and P.A. Sharp. 1991. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc. Natl. Acad. Sci.* 88: 683–687.

Krause, J.E., J.M. Chirgwin, M.S. Carter, Z.S. Xu, and A.D. Hersh. 1987. Three rat preproteocthyrinin mRNAs encode the neuropeptides substance P and neurokinin A. *Proc. Natl. Acad. Sci.* 84: 881–885.

Kuo, H.C., F.H. Nasim, and P.J. Grabowski. 1991. Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science* 251: 1045–1050.

Lavigneur, A., H. La Branche, A.R. Kombliht, and B. Chabot. 1993. A splicing enhancer in the human fibronectin alternative ED1 exon interacts with SR proteins and stimulates U2 snRNPs binding. *Genes & Dev.* 7: 2405–2417.

Leff, S.E., R.M. Evans, and M.G. Rosenfeld. 1987. Splice commitment dictates alternative RNA processing in calcitonin/CGRP gene expression. *Cell* 48: 517–524.

Libri, D., M. Goux-Pelletan, E. Brody, and M.Y. Fiszman. 1990. Exon as well as intron sequences are cis-acting regulating elements for the mutually exclusive alternative splicing of the β-tropomyosin gene. *Mol. Cell. Biol.* 10: 5036–5046.

Libri, D., A. Piseri, and M.Y. Fiszman. 1991. Tissue-specific splicing in vivo of the β-tropomyosin gene: Dependence on an RNA secondary structure. *Science* 252: 1842–1845.

Maniatis, T. 1991. Mechanisms of alternative pre-mRNA splicing. *Science* 251: 33–34.

Mardon, H.J., G. Sebastio, and F.E. Baralle. 1991. A role for exon sequences in alternative splicing of the human fibronectin gene. *Nucleic Acids Res.* 15: 7725–7733.

Mattos, W., L. Ryner, and B.S. Baker. 1992. Autoregulation and multifunctionality among trans-acting factors that regulate alternative pre-mRNA splicing. *J. Biol. Chem.* 267: 919023–19026.

Matunis, E.L., M.J. Matunis, and G. Dreyfuss. 1993. Association of individual hnRNP proteins and snRNPs with nascent transcripts. *J. Cell Biol.* 121: 219–228.

Mayeda, A., D.M. Helfman, and A.R. Kainer. 1993. Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Mol. Cell. Biol.* 13: 2993–3001.

McKeown, M. 1992. Alternative mRNA splicing. *Annu. Rev. Cell Biol.* 8: 133–155.

McNally, M.T., R.R. Gotarek, and K. Beemon. 1991. Characterization of Roux sarcoma virus intronic sequences that negatively regulate splicing. *Virology* 185: 99–108.

Moore, M.J., C.C. Query, and P.A. Sharp. 1993. Splicing of precursors to mRNA by the spliceosome. In *The RNA world* (ed. R. Gesteland and J. Atkins), pp. 303–357. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Nakanishi, H., P.A. Sears, H.M. Hoffman, H.C. Kuo, and P.J. Grabowski. 1990. A sequential splicing mechanism promotes selection of an optional exon by repositioning a downstream 5' splice site in preproteocthyrinin pre-mRNA. *Genes & Dev.* 4: 1172–1184.

Paoletti, G., C. Henschcliffe, G. Sebastio, and F.E. Baralle. 1988. Sequence analysis and in vivo expression show that alternative splicing of ED-B and ED-A regions of the human fibronectin gene are independent events. *Nucleic Acids Res.* 16: 3545–3557.

Reed, R. and T. Maniatis. 1986. A role for exon sequences and splice-site proximity in splice-site selection. *Cell* 46: 681–690.

Roberson, B.L., G.J. Cote, and S.M. Berget. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* 10: 84–94.

Ryner, L.C. and B.S. Baker. 1991. Regulation of doublesex pre-mRNA processing occurs by 3'-splice site activation. *Genes & Dev.* 5: 2071–2085.

Sakamoto, H., K. Inoue, I. Higuchi, Y. Ono, and Y. Shimura. 1992. Control of *Drosophila* Sex-lethal splicing by its own female-specific product. *Nucleic Acids Res.* 20: 5533–5540.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
Huh and Hynes

Schwarzbauer, J.E., R.S. Patel, D. Fonda, and R.O. Hynes. 1987. Multiple sites of alternative splicing of the rat fibronectin gene transcript. EMBO J. 6: 2573–2580.

Shaw, G. and R. Kamen. 1986. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell 46: 659–667.

Smith, C.W.J., J.G. Patton, and B. Nadal-Ginard. 1989. Alternative splicing in the control of gene expression. Annu. Rev. Genet. 23: 537–577.

Solnick, D. 1985. Alternative splicing caused by RNA secondary structure. Cell 43: 667–676.

Somasekhar, M.B. and J.E. Mertz. 1985. Exon mutations that affect the choice of splice sites used in processing the SV40 late transcripts. Nucleic Acids Res. 13: 5591–5609.

Sosnowski, B.A., J.M. Belote, and M. McKeown. 1989. Sex-specific alternative splicing of RNA from the transformer gene results from sequence-dependent splice site blockage. Cell 58: 449–459.

Staden, R. 1990. Searching for patterns in protein and nucleic acid sequences. Methods Enzymol. 183: 193–211.

Streuli, M. and H. Saito. 1989. Regulation of tissue-specific alternative splicing: Exon-specific cis elements govern the splicing of leukocyte common antigen pre-mRNA. EMBO J. 8: 787–796.

Sun, Q., R. Hampson, and F.M. Rottman. 1993. In vitro analysis of bovine growth hormone pre-mRNA alternative splicing. J. Biol. Chem. 268: 15659–15666.

Swanson, M.S. and G. Dreyfuss. 1988. Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. Mol. Cell. Biol. 8: 2237–2241.

Talerico, M. and S.M. Berget. 1990. Effect of 5' splice site mutations on splicing of the preceding intron. Mol. Cell. Biol. 10: 6299–6305.

Tian, M. and T. Maniatis. 1992. Positive control of pre-mRNA splicing in vitro. Science 256: 237–240.

———. 1993. A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. Cell 74: 105–114.

Ulfendahl, P.J., J.P. Kreivi, and J. Akusjärvi. 1989. Role of the branch site/3' splice site region in adenovirus-2 E1A pre-mRNA alternative splicing: Evidence for 5' and 3' splice site cooperation. Nucleic Acids Res. 17: 925–937.

Valcarcel, J., R. Singh, P.D. Zamore, and M.R. Green. 1993. The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. Nature 362: 171–175.

Watakabe, A., K. Tanaka, and Y. Shimura. 1993. The role of exon sequences in splice site selection. Genes & Dev. 7: 407–418.

Xu, R., J. Teng, and T.A. Cooper. 1993. The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. Mol. Cell. Biol. 13: 3660–3674.

Yeakley, J.M., F. Hedjran, J.P. Morfin, N. Merillat, M.G. Rosefeld, and R.B. Emeson. 1993. Control of calcitonin/calcitonin gene-related peptide pre-mRNA processing by constitutive intron and exon elements. Mol. Cell. Biol. 13: 5999–6011.

Zahler, A.M., W.S. Lane, J.A. Stolik, and M.B. Roth. 1992. SR proteins: A conserved family of pre-mRNA splicing factors. Genes & Dev. 6: 837–847.

Zahler, A.M., K.M. Neugebauer, W.S. Lane, and M.B. Roth. 1993. Distinct functions of SR proteins in alternative pre-mRNA splicing. Science 260: 219–222.
Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element.

G S Huh and R O Hynes

Genes Dev. 1994, 8:
Access the most recent version at doi:10.1101/gad.8.13.1561