BAYLOR
COLLEGE OF
MEDICINE

One Baylor Plaza
Houston, Texas 77030-3498
Department of Pathology
Section of Molecular
Pathobiology
TEL: (713) 798-4661
FAX: (713) 798-5838

November 14, 1996

Philip Migliore, M.D.
Research Director
Moran Foundation

Dear Phil:

Enclosed please find my progress report for Moran Foundation Project number 1-95-0079. I would like to express my appreciation for the support provided by the Moran Foundation.

Enclosed is a copy of a manuscript to be submitted for publication that acknowledges support from the Moran Foundation

If you have any questions, please feel free to contact me.

Sincerely,

Thomas A. Cooper, M.D.
798-3141 office
798-5838 fax
tcooper@bcm.tmc.edu

Thomas A. Cooper, M.D. **PROGRESS REPORT** **11/14/96**

MORAN FOUNDATION PROJECT 1-95-0079: Cardiac troponin T isoform transitions during cardiac failure: a molecular basis for decreased contractility

The goals of this project were to identify and molecularly clone cDNAs for human cardiac troponin T isoforms and examine the distribution of these mRNA isoforms in normal and diseased heart. This analysis was carried out entirely by Dr. Rose Anton, a Pathology resident.

Our first step was to analyze heart RNAs from tissue from 12 diseased hearts and one normal heart. Heart RNA was analyzed by primer extension using an oligonucleotide complementary to a region of human cTNT mRNA that is downstream of the variable region. This analysis demonstrated one predominant isoform in all samples and several different isoforms that differed in different samples. These differences were analyzed further by RT-PCR using two sets of primers. The two primer sets allowed analysis of different regions of the mRNA that were shown by other laboratories to contain variability. As in the primer extension, different isoforms were detected in different diseased hearts and the patterns in the diseased hearts were different from that in the normal heart. Studies in other laboratories demonstrated that there is little variation in the pattern of isoform expression in normal hearts. Therefore we conclude that different isoforms generated by alternative splicing are generated in response to heart disease.

Soon after receiving funding for this project, two reports described the expression of cTNT alternatively spliced mRNAs in normal and diseased heart. The two labs differed in their findings: one found different mRNA isoforms expressed in heart disease and the other found no difference between normal and diseased heart. Our results support the conclusions of the former study. Both published studies involved a small number of patients. Our plan was to screen a large number of heart samples available at the Ben Taub and Methodist hospitals. Unfortunately, the demands on Dr. Anton's time did not allow her to complete all of the scientific goals of this project. I was impressed with Dr. Anton's progress in her molecular evaluation of cTNT expression in heart disease and I believe that this project achieved a second goal of providing valuable molecular biology training to Dr. Anton.

The following are updates with regard to previous Moran Foundation Awards:

Preliminary results funded by Moran Foundation Award 1-94-0074, entitled "A cell-free assay for regulated pre-mRNA alternative splicing", contributed to obtaining a four year NIH competitive renewal (NIH (R01-HL45565-06A1), "Troponin T Alternative Splicing in Embryonic Heart", 7/1/96-6/30/00. $120,318 direct annual costs; $639,248 total direct costs. Thomas A. Cooper, M.D., Principal investigator.). These results were presented at a national meeting [Ryan, K.J and Cooper, T.A. (May 1996) Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle. RNA Processing Meeting, Madison, Wisconsin)] and have been submitted to a second meeting [Ryan, K.J., Philips, A., and Cooper, T.A. (April, 1997) Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal and cardiac muscle. Keystone Symposium, Molecular Biology of Muscle Development. Snowmass, CO.]

Results funded in part by the Moran Foundation (1-93-0066; "Selection of exon sequences that facilitate pre-mRNA splicing.") will be submitted for publication in *Molecular and Cellular Biology.* A copy of the manuscript is enclosed. Support from the Moran Foundation is noted in the acknowledgments.

Identification of a new class of exonic splicing enhancers by *in vivo* selection

Lydia R. Coulter, Mark A. Landree, and Thomas A. Cooper*

Department of Pathology and Program in Cell and Molecular Biology, Baylor College of Medicine, Houston, TX 77030

* corresponding author (713) 798-3141 telephone
(713) 798-5838 fax
tcooper@bcm.tmc.edu

ABSTRACT

Splicing of complex metazoan pre-mRNAs occurs with remarkable fidelity. For some vertebrate exons, splice site selection is aided by purine-rich splicing enhancers located within the exon. To identify other classes of splicing enhancers, we developed an iterative scheme to select *in vivo* for exon sequences that enhance exon inclusion. This approach, modeled on the *in vitro* SELEX procedure, is the first *in vivo* application of an iterative procedure to enrich RNA processing signals. Two predominant sequence motifs were enriched after three rounds: a purine-rich motif that resembles previously identified splicing enhancers and a novel class of A/C-rich splicing enhancers (ACE). Individual selected ACEs enhanced splicing *in vivo* and *in vitro*. ACE splicing activity was competed by RNAs containing the purine-rich splicing enhancer from cardiac troponin T exon 5. Thus, ACE activity is likely to require a subset of the SR splicing factors previously shown to mediate activity of this purine-rich enhancer. ACE motifs are found in two vertebrate exons previously demonstrated to contain splicing enhancer activity as well as in the well-characterized *Drosophila doublesex* (*dsx*) splicing enhancer. We demonstrate that one copy of the *dsx* repeat enhances splicing of a vertebrate exon in vertebrate cells and that this enhancer activity requires the ACE motif. Our results suggest the possibility that ACEs are evolutionarily related to the *dsx* repeat element.

INTRODUCTION

Most metazoan pre-mRNAs contain multiple exons that are precisely joined during splicing. Remarkable progress has been made toward understanding the biochemistry of intron removal (30), however, many questions remain regarding the initial events of spliceosome assembly. Conserved sequences at the intron/exon borders (splice sites) are required for splicing but these sequences do not contain sufficient information to distinguish bona fide splice sites from cryptic splice sites of similar sequence found in introns and exons. How the authentic splice sites are initially recognized, cryptic splice sites are ignored, and exons are contiguously joined without exon skipping remain unknown (reviewed in 3).

Recent evidence suggests that exons are defined early in the splicing reaction by factors that bind to the splice sites and communicate across the exon (2). Cooperative interactions of a large

number of factors that bind to multiple *cis* elements may compensate for the lack of information content in individual splicing elements (33). Some exons are defined by auxiliary elements in addition to the splice sites. For example, intronic elements have been found associated with exons that are unusually small and/or alternatively spliced (4, 6, 9, 19, 25, 35, 37). Some of these elements are associated with cell specific splicing while others appear to be ubiquitously recognized. Auxiliary splicing elements have also been identified within exons (1, 5, 11, 17, 20, 22, 36, 38, 40, 47, 49, 51, 52). Most of these exonic splicing enhancers are purine-rich, however, non-purine rich enhancers have also been described (15, 39, 47, 48).

Activation of purine-rich exonic splicing enhancers requires binding of SR proteins, a conserved family of essential splicing factors (22, 32, 39, 40). Differences in substrate specificity among SR protein family members have been revealed by SELEX analysis and have been demonstrated directly by RNA binding and *in vitro* splicing assays (reviewed in 16, 29). For example, we demonstrated that a subset of SR proteins bound and activated the purine-rich splicing enhancer in cardiac troponin T (cTNT) exon 5 (32). Protein-protein interactions between SR proteins and other splicing factors are proposed to contribute to the cooperative assembly of early splicing complexes (2, 29, 33).

One of the best characterized exonic splicing enhancers is in the fourth exon of the *Drosophila doublesex* gene (reviewed in 34). This enhancer contains six copies of a 13 base pair sequence (the *dsx* repeat) and a purine-rich element (PRE) located between the fifth and sixth repeats. In contrast to the splicing enhancers identified in vertebrate exons which appear to be ubiquitously recognized, the *dsx* enhancer promotes female specific splicing of exon 4 (34). RNA:protein and protein:protein interactions between the *dsx* repeats and two *Drosophila* SR-like proteins, Tra2 and Tra, recruit essential SR proteins into a complex that promotes *dsx* enhancer activity in HeLa nuclear extracts (18, 26, 27, 44, 45).

The prevalence of exonic splicing enhancers is unknown. Most enhancers have been found in exons that are associated with an alternative splicing event, either within an alternative exon or an adjacent constitutive exon. The association of enhancers with alternative splicing may reflect a mechanistic link: alternative splicing generally requires weak splice sites to allow regulated use and enhancers would ensure that these signals are not completely ignored. Alternatively, this

association may reflect the bias towards investigations of alternatively spliced exons. It is possible that splicing enhancers are common among both constitutive and alternative exons.

Here we describe an iterative procedure to select for exon sequences that enhance inclusion of an alternative exon during splicing *in vivo*. We used this approach to identify a novel class of exonic splicing enhancers. A thirteen nucleotide randomized cassette was inserted into an alternative exon that is predominantly skipped in the absence of a splicing enhancer. Exons that were spliced into the mRNA were selectively amplified by RT-PCR; repeated cycles of ligation, transfection and amplification were used to enrich for sequences that enhanced exon inclusion. After three rounds of selection, two sequence motifs predominated: a purine-rich motif resembling the previously defined enhancers and an A/C-rich motif. Exons that were neither purine-rich nor A/C-rich were also isolated indicating that a variety of sequences can enhance splicing.

METHODS

Selection procedure: 100 ng of oligonucleotide containing 13 randomized positions (Figure 1A, top) was made double stranded and amplified by PCR using Vent DNA polymerase and oligonucleotides that flanked the randomized region (GACGTACGGATCCATTC and TAATACGACTCACTATA). The PCR product was phenol/chloroform extracted and digested with SalI and BamHI. The 24 nucleotide SalI/BamHI fragment was isolated from a 6% nondenaturing polyacrylamide gel and quantitated. 100 µl ligation reactions contained 1 µg gel-isolated minigene with an equal molar amount of insert. Ligation reactions were transfected directly into QT35 quail fibroblast cultures (two 60 mm plates of 10^6 cells) using a calcium phosphate protocol as described previously (51). QT35 cells were chosen for their high transfection efficiency. Total RNA was harvested 40-48 hours following transfection using guanidine thiocyanate (50) and treated with DNase. RNAs containing the alternative exon were selectively amplified by RT-PCR using oligos A (GTGGTGAGGCCCTGGGCAGGTC) and B (AAGGGTAGACCACCAGCAGCCTGGA) (Figure 1A) which prime within the alternative exon (nucleotides complementary to the alternative exon are underlined). The PCR product was phenol/chloroform extracted and processed through two more rounds. Conditions for ligation and preparation of vector and insert were optimized using a bacterial transformation assay.

Minigene constructs, transient transfection, and RNA analysis: The SXN13 selection minigene contains a 34 nucleotide alternative exon flanked by duplicated intron 1 from human ß-globin such that the first and third exons of the minigene are globin exons 1 and 2. The fragment containing exons 1-3 of this minigene was constructed using a PCR approach based on a design that is modified from Dominski and Kole (13), see also (51). Minigene intron 3 and exon 4 are derived from cTNT intron 6 and exon 18 (details are available upon request). The cTNT minigene (Figure 4) was derived from ΔPB (51) by replacing natural exon 4 sequence (to remove a previously introduced BamHI site). Exon 5 was modified as shown in Figure 4 to accommodate the SalI/BamHI cassette containing the randomized region. Transient transfections of cloned minigenes contained 5 µg plasmid DNA per 60 mm plate of 10^6 cells. Total cellular RNA from one half plate of cells was assayed by primer extension using 2 ng of 32P-end labeled oligonucleotides (51) (AGAACCTCTGGGTCCAAGGGTAG is complementary to globin exon 2 and GTCTCCTCTTCCTCCTCGTCTACCTGATCC is complementary to cTNT exon 6). The products of primer extension were quantitated directly from the gel using a Betagen Betascope 603 analyzer. The percent exon inclusion was calculated as [cpm exon inclusion / (cpm exon inclusion + cpm exon skipping)] X 100. Except for some of the nonselected clones in Figure 2, results from all minigenes were confirmed by at least two transfections.

***In vitro* RNA assays**

The *in vitro* splicing substrates in Figure 5 were constructed by an EcoRI/NdeI collapse of the transfection plasmids in Figure 4 that brought exon 4 just downstream of the T3 polymerase promoter in the Bluescript plasmid. These RNAs contain (5' to 3') 28 nucleotides of polylinker, 8 nucleotides of cTNT intron 3, the 18 nucleotide cTNT exon 4, cTNT intron 4, and the selection exon shown in Figure 4. Uniformly labeled splicing substrates were synthesized in 20 µl reactions containing 500 ng DNA (linearized with BamHI), 40 mM Tris pH 7.5, 6 mM $MgCl_2$, 2 mM spermidine, 50 mM NaCl, 10 mM DTT, 0.5 mM each NTP, 1.25 mM GpppG, 20 units RNAsin, 25 µCi ^{32}P-αGTP, and 20 units T3 RNA polymerase. All RNAs were gel isolated. Splicing reactions (10 µl) contained 40,000 cpm substrate (approximately 0.01 pmole), 0.8% PEG, 1 mM $MgCl_2$, 0.625 mM ATP, 25 mM creatine phosphate, and 40% HeLa nuclear extract. Reactions were stopped by addition of 90 µl of

urea solution [7M urea, 0.35M NaCl, 0.01M Tris (pH 7.4), 0.01M EDTA, 0.05g/ml SDS] and 90 µl of water, phenol/chloroform extracted, then run on a 4% denaturing polyacrylamide gel.

Competitor RNAs in Figure 6 contain (5'->3'): 20 nucleotides of vector, the last 50 nucleotides of cTNT intron 4, and the selected exon to the *BamHI site (Figure 4). UP, WT, and* DOWN competitor RNAs containing cTNT exon 5 enhancer mutants were described previously (32). Competitor RNAs were synthesized in 100 µl reactions containing 6 µg linearized DNA, 40 mM Tris (pH 7.5), 20 mM $MgCl_2$, 2 mM spermidine, 50 mM NaCl, 10 mM DTT, 4 mM each NTP, 40 units RNAsin, 2 µCi ^{32}P-αGTP (for quantitation), and 60 units T7 RNA polymerase. Competitor RNAs were gel isolated and quantitated by either scintillation counting or UV absorbance.

RESULTS

The selection scheme is diagrammed in Figure 1A. A synthetic DNA cassette containing 13 contiguous randomized positions (approximately 6.7×10^7 sequences) was directionally ligated into the middle exon of a three exon minigene derived from human ß-globin. Ligation reactions were transfected directly into QT35 quail fibroblast cultures. In the absence of a splicing enhancer, the middle exon is predominantly skipped (51) due to its small size (34 nucleotides) and a non-canonical 5' splice site (14). To ensure that splicing enhancers would be detected, the minigene was designed such that the alternative exon is included at a low level in the absence of a splicing enhancer. We have shown previously that inclusion of this exon is induced by a purine-rich splicing enhancer (51). Following transfection, RNAs that contained the middle exon were selectively amplified by RT-PCR using oligonucleotides complementary to the junctions of exons 1 and 2 (oligo A, Figure 1A) and exons 2 and 3 (oligo B). To enrich for the strongest splicing enhancers, the selected pool of randomized cassettes was excised from the RT-PCR product and cycled though two additional rounds of ligation/transfection/amplification.

Two predominant sequence motifs were enriched by the selection (Figure 1B and Table 1). One motif resembles the previously characterized purine-rich splicing enhancers, which validates the *in vivo* selection approach. The second motif is a novel A/C-rich element we refer to as an ACE (defined in Table 1 legend). After three rounds of selection, 16/30 (53%) of selected exons contained an ACE (Figure 1B) compared to 5/31 (16%) of randomly picked non-selected exons

(data not shown). Strikingly, 22 of the 30 (73%) third round exons that were sequenced contained either an ACE, a purine-rich motif, or both motifs.

To test directly whether the selected sequences functioned as splicing enhancers, individual clones containing selected or non-selected exons were picked at random and transiently transfected into QT35 cultures, and the level of exon inclusion was assayed by primer extension. The selected exons showed a significantly higher level of exon inclusion than non-selected exons (representative results are in Figure 2). The median level of exon inclusion for selected exons was 39% (n=25) compared to 9% (n=20) for exons containing non-selected sequences. We conclude that the selection procedure enriched for sequences with splicing enhancer activity.

Enhancer activities of three individual selected sequences are shown in Figure 3. These clones, which contain an ACE (4.11.12, lane 3), a purine-rich motif (4.24.4, lane 5) or both motifs (4.11.1, lane 1) show higher levels of exon inclusion than individual non-selected exons (Figure 2). In addition, exon inclusion is enhanced in these clones above a baseline level of 22% determined by transfecting a plasmid pool from >400 bacterial colonies containing non-selected exons (pool, lane 7). To directly determine whether the A/C-rich motif was required for enhanced exon inclusion, we introduced point mutations into two selected clones. Mutations which reduced the A/C content of 4.11.1 (lane 2) and 4.11.12 (lane 4) each decreased the level of exon inclusion compared to the selected sequence. A synthetic A/C-rich sequence also enhanced exon inclusion, further supporting a role for this motif as a splicing enhancer (SYN, lane 6).

To test the enhancer activity in a different minigene, selected purine-rich and ACE motifs were inserted in place of the purine-rich enhancer in exon 5 of the cardiac troponin T (cTNT) gene. Splicing of cTNT exon 5 requires an exonic enhancer, making this a good context in which to test selected sequences (7, 51). All three of the selected exons shown in Figure 3 as well as the synthetic ACE enhanced splicing of the cTNT exon compared to pooled non-selected exons (Figure 4). Three independent mutations in the ACE decreased splicing of 4.11.12, demonstrating a sequence-specific requirement for the A/C-rich motif for enhancer activity. We conclude that splicing enhancer activity of the selected motifs is independent of the minigene used for selection.

We next tested whether enhancer activities of selected exons were reproduced *in vitro*. A single intron splicing substrate was derived from the cTNT minigene (shown in Figure 4) such that

the downstream exon contained the selection cassette (Figure 5). The results presented in Figure 5 demonstrate that the selected motifs enhanced removal of the upstream intron. Mutations that disrupted ACE activity *in vivo* decreased the level of splicing to that observed for the non-selected sequence. As described previously for the cTNT purine-rich enhancer (51), mutations in the ACE did not affect splicing of the downstream intron (data not shown). Our data suggest that the ACE promotes exon inclusion by enhanced splicing of the upstream intron and not the downstream intron. Although a role for differences in stability in determining mRNA ratios *in vivo* cannot be ruled out, these *in vitro* results directly demonstrate that the selected enhancers affect splicing.

To determine whether the ACE enhanced splicing via direct interactions with titratable *trans*-acting factors, *in vitro* splicing of the 4.11.12 ACE substrate (shown in Figure 5) was challenged with cold competitor RNAs containing the 4.11.12 downstream exon linked to the branch site and 3' splice site of intron 4. RNAs containing either the 4.11.12 or the SYN ACE exons efficiently competed splicing (Figure 6 and data not shown). Competitors containing 4.11.12 mutations that inactivated splicing *in vivo* did not compete splicing (12 mu1 and 12 mu2) demonstrating that competition is ACE-dependent. The 4.11.12 competitor RNA did not compete *in vitro* splicing of an enhancer-independent splicing substrate (cTNT exons 8 and 9) and therefore is not a general inhibitor of splicing (data not shown).

To determine whether the ACE and purine-rich enhancers require the same or different enhancer-binding factors for their activities, splicing of 4.11.12 was challenged with RNAs containing the cTNT exon 5 enhancer and two mutants that increase or decrease exon 5 enhancer activity *in vivo*. These RNAs were used previously to demonstrate that the strength of the purine-rich enhancer directly correlates with binding affinity for four members of the SR protein family of essential splicing factors, SRp30a, SRp40, SRp55, and SRp75 (32). The results presented in Figure 7 demonstrate that the UP and WT cTNT purine-rich enhancers titrate factors required for ACE-dependent splicing. Furthermore, the levels of competition by the three enhancers directly correlate with their binding affinities for SR proteins (32). Competition is not due to general titration of SR proteins or other general splicing factors since cTNT exon 5 does not compete splicing of three precursor RNAs that lack purine-rich or A/C-rich enhancers (data not shown). Therefore, we

conclude that at least one of the factors that binds to the cTNT purine-rich enhancer is required for splicing activity of the ACE-containing exon.

RNA elements rich in adenosines and cytosines have recently been shown to affect RNA processing in several experimental systems (Table 2). *All but two of the elements shown in Table 2* are naturally occurring splicing enhancers (*Drosophila dsx*, cTNT exon 16, and calcitonin/CGRP exon 4), while human ß-globin splicing was stimulated by a mutation of 6 Us to 6 Cs in a heterologous exon sequence (ß-globin/URA3 chimera). The HSV-TK fragment allows cytoplasmic accumulation of an otherwise unstable intronless human ß-globin mRNA by an unknown mechanism (24). Identification of the ACE-like motif in the *Drosophila dsx* enhancer is particularly intriguing. To determine whether a single copy of the 13 nucleotide *dsx* element functions in a vertebrate exon, a sequence that is found in three of the six repeats (31) was inserted into the selectable exon cassette and was tested in both the globin and cTNT minigenes in QT35 cells (Figure 8). In both minigenes, the *dsx* element is one of the strongest enhancers tested (Figure 8, lanes 1 and 4). To determine the sequence requirements for *dsx* enhancer activity in vertebrate cells, we introduced separate mutations into the T/C-rich 5' region and the A/C-rich 3' region. A three nucleotide substitution in the A/C-rich segment of the motif significantly decreased enhancer activity (dsx mu2, lanes 3 and 6). Mutation of the first two nucleotides from TC -> GA (dsx mu1, lanes 2 and 5) also decreased enhancer activity. This mutation has been shown to prevent Tra2 binding and to inactivate Tra/Tra2-dependent splicing in transfected *Drosophila* Kc cells (21). Our results demonstrate that a single copy of the *Drosophila dsx* repeat element strongly enhances splicing of vertebrate splice sites in vertebrate cells. Furthermore, the sequence requirements for enhancer activity in vertebrate cells include but are not limited to the ACE-like region.

DISCUSSION

We describe a novel approach to enrich for an RNA processing signal in vertebrate cells. To our knowledge, this is the first application of an iterative procedure *in vivo* to identify RNA processing signals. *In vivo* randomization-selection approaches have been used in *Saccharomyces cerevisiae* to demonstrate RNA-RNA interactions required for splicing. Madhani and Guthrie (28) identified functional intermolecular base paring of U2 and U6 snRNAs and Libri et al. (23) demonstrated a role for pre-mRNA intramolecular base-pairing in splicing. In these procedures,

selection was based on cell survival and occurred in a single step. Our approach is more like a SELEX procedure (46) that has been performed in living cells rather than in microfuge tubes. A pool of pre-mRNAs containing randomized sequence is expressed in cells, the desired product is selected by the choice of PCR priming sites, amplified, and cycled through multiple rounds. The advantage of this approach is the ability to enrich for active sequences. Splicing signals in vertebrate genes rarely fit a strict consensus making it difficult to identify them on the basis of sequence alone. Using an iterative procedure allowed us to identify an enriched motif even though no clear consensus could be discerned. It should be possible to use this approach to identify cell-specific splicing enhancers by performing the selection procedure in differentiated cell types. It may also be useful to perform directed "*in vivo* SELEX" in cells that overexpress individual splicing factors.

Interestingly, the ACE motifs described in this report were not identified in a similar *in vitro* splicing selection scheme that used HeLa nuclear extracts (42). This may reflect differences in the splicing enhancers preferred in HeLa compared to QT35 cells or differences between *in vivo* and *in vitro* selection approaches. As in our study, these authors identified a number of enhancer sequences that were represented only once in the set of sequenced clones. We could not find common motifs among this group of non-ACE, non-purine-rich enhancers isolated in the two studies. In transfection studies, we find that enhancer activity of non-ACE, non-purine-rich selected exons are as strong or stronger than the purine-rich or ACE motifs (data not shown). Therefore, our results are consistent with those of Tian and Kole (42) in suggesting that additional classes of splicing enhancers remain to be identified.

A/C-rich motifs can be found in elements shown to affect RNA processing in several systems (Table 2). Our results identify this motif as a class of splicing enhancers. Exon 16 of cTNT was found to contain an element that enhanced splicing of heterologous transcripts *in vitro* and promoted binding of U2AF65 to the upstream 3' splice site (48). U2AF binding was competed by the 3' one-third of cTNT exon 16 which contains A/C-rich and purine-rich motifs. Point mutants in the A/C-rich but not purine-rich motif prevented competition of U2AF binding. Therefore, an ACE in cTNT exon 16 promotes binding of U2AF to the upstream 3' splice site. Exon-facilitated binding of U2AF to the

3' splice site required nuclear components in addition to U2AF. Presumably these factors include those that make direct contact with the splicing enhancer.

The specific factors that interact with the ACE and mediate enhancer activity remain to be determined. We show that the cTNT purine-rich enhancer titrates factors required for ACE activity (Figure 7) suggesting that at least one of the SR proteins required for cTNT enhancer activity is also required for ACE activity. We previously demonstrated that the cTNT enhancer binds four of the essential SR proteins and that the binding affinity of the SR proteins directly correlated with enhancer strength *in vivo* and *in vitro* (32). The ACE identified in cTNT exon 16 was shown to immunoprecipitate with anti-SC35 antibodies (48). However, since SC35 neither binds nor activates the cTNT enhancer (32), this is not likely to be the factor titrated by an excess of cTNT exon 5 RNA.

Our results suggest the possibility that the *dsx* enhancer is a member of a previously unrecognized class of conserved ACE splicing enhancers. By sequence, the *dsx* repeat appears to contain two motifs: a T/C-rich motif at the 5' end and an A/C rich motif at the 3' end. Our results indicate that both motifs are required for its function (Figure 8). Consistent with a bipartite structure, these two motifs have recently been shown to bind different sets of splicing factors by site-specific UV crosslinking (27). The T/C-rich 5' half binds to an SR protein (RBP1 in *Drosophila* and 9G8 in mammals) and the A/C-rich component binds *Drosophila* Tra2. In these experiments, Tra2 was added exogenously to HeLa nuclear extracts. However, previous results suggested that HeLa extracts contain Tra2-like activity, as addition of Tra alone to HeLa nuclear extracts activated the *dsx* female specific 3' splice site (43). Interestingly, two human homologues of Tra2 have recently been identified (8). Taken together, these results suggest the interesting possibility that the ACE motifs selected in this procedure are targets for human Tra2.

In summary, our results demonstrate that it is feasible to perform iterative selection for RNA processing signals *in vivo*. Active motifs were enriched within three rounds suggesting that *in vivo* selection conditions are inherently stringent. This is a powerful approach to identify nucleic acid sequences that are directly relevant to RNA processing in living cells.

ACKNOWLEDGMENTS

We would like to thank Sue Berget and Bill Mattox for advice during the course of the project and Gil Cote, Bill Mattox, Miles Wilkinson, and members of the Cooper lab for comments on the manuscript. This project was supported by the American Cancer Society and the Moran Foundation. T.A.C is an Established Investigator of the American Heart Association.

REFERENCES

1. **Amendt, B. A., D. Hesslein, L. J. Chang and C. M. Stoltzfus.** 1994. Presence of negative and positive cis-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1. Mol Cell Biol **14:**3960-3970.
2. **Berget, S. M.** 1995. Exon recognition in vertebrate splicing. J Biol Chem **270:**2411-2414.
3. **Black, D.** 1995. Finding splice sites within a wilderness of RNA. RNA **1:**763-771.
4. **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.
5. **Caputi, M., G. Casari, S. Guenzi, R. Tagliabue, A. Sidoli, C. A. Melo and F. E. Baralle.** 1994. A novel bipartite splicing enhancer modulates the differential processing of the human fibronectin EDA exon. Nucl. Acids Res. **22:**1018-1022.
6. **Carlo, T., D. A. Sterner and S. M. Berget.** 1996. An intron splicing enhancer containing a G-rich repeat facilitates inclusion of a vertebrate micro-exon. RNA **2:**342-353.
7. **Cooper, T. A.** 1992. In vitro splicing of cardiac troponin-T precursors - exon mutations disrupt splicing of the upstream intron. J Biol Chem **267:**5330-5338.
8. **Dauwalder, B., F. Amaya-Manzanare and W. Mattox.** 1996. A human homologue of the Drosophila sex determination factor transformer-2 has conserved splicing regulatory functions. Proc Natl Acad Sci, USA **(In press):**.
9. **Del Gatto, F. and R. Breathnach.** 1995. Exon and intron sequences, respectively, repress and activate splicing of a fibroblast growth factor receptor 2 alternative exon. Mol Cell Bio **15:**4825-4834.

10. **Dietz, H. C., D. Valle, C. A. Francomano, R. J. Kendzior, R. E. Pyeritz and G. R. Cutting.** 1993. The Skipping of Constitutive Exons In Vivo Induced by Nonsense Mutations. *Science* **259:***680-683.*

11. **Dirksen, W. P., R. K. Hampson, Q. Sun and F. M. Rottman.** *1994. A purine-rich exon* sequence enhances alternative splicing of bovine growth hormone pre-mRNA. J Biol Chem **269:**6431-6436.

12. **Dirksen, W. P., Q. A. Sun and F. M. Rottman.** 1995. Multiple splicing signals control alternative intron retention of bovine growth hormone pre-mRNA. J Biol Chem **270:**5346-5352.

13. **Dominski, Z. and R. Kole.** 1991. Selection of splice sites in pre-mRNAs with short internal exons. Mol Cell Biol **11:**6075-83.

14. **Dominski, Z. and R. Kole.** 1992. Cooperation of pre-RNA sequence elements in splice site selection. Mol Cell Biol **12:**2108-2114.

15. **Dominski, Z. and R. Kole.** 1994. Identification of exon sequences involved in splice site selection. J Bio Chem **269:**23590-23596.

16. **Fu, X. D.** 1995. The superfamily of arginine/serine-rich splicing factors. RNA **1:**663-680.

17. **Gontarek, R. R. and D. Derse.** 1996. Interactions among SR proteins, an exonic splicing enhancer, and a lentivirus Rev protein regulate alternative splicing. Mol Cell Biol **16:**2325-2331.

18. **Heinrichs, V. and B. S. Baker.** 1995. The Drosophila SR protein RBP1 contributes to the regulation of doublesex alternative splicing by recognizing RBP1 RNA target sequences. EMBO J **14:**3987-4000.

19. **Huh, G. S. and R. O. Hynes.** 1994. Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element. Genes Dev. **8:**1561-1574.

20. **Humphrey, M. B., J. Bryan, T. A. Cooper and S. M. Berget.** 1995. A 32-nucleotide exon-splicing enhancer regulates usage of competing 5' splice sites in a differential internal exon. Molecular and Cellular Biology **15:**3979-3988.

21. **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 U S A **89:**8092-8096.

22. **Lavigueur, A., H. La Branche, A. R. Kornblihtt and B. Chabot.** 1993. A splicing enhancer in the human fibronectin alternate ED1-exon interacts with SR-proteins and *stimulates U2-snRNP binding. Genes Dev.* **7:**2405-2417.

23. **Libri, D., F. Stutz, T. Mccarthy and M. Rosbash.** 1995. RNA structural patterns and splicing: Molecular basis for an RNA-based enhancer. RNA **1:**425-436.

24. **Liu, X. D. and J. E. Mertz.** 1995. HnRNP L binds a cis-acting RNA sequence element that enables intron-independent gene expression. Genes Dev **9:**1766-1780.

25. **Lou, H., Y. Yang, G. J. Cote, S. M. Berget and R. F. Gagel.** 1995. An intron enhancer containing a 5' splice site sequence in the human calcitonin/calcitonin gene-related peptide gene. Mol Cell Biol **15:**7135-7142.

26. **Lynch, K. W. and T. Maniatis.** 1995. Synergistic interactions between two distinct elements of a regulated splicing enhancer. Genes Dev **9:**284-293.

27. **Lynch, K. W. and T. Maniatis.** 1996. Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. Genes Dev **10:**2089-2101.

28. **Madhani, H. D. and C. Guthrie.** 1994. Randomization-selection analysis of snRNAs in vivo: evidence for a tertiary interaction in the spliceosome. Genes Dev. **8:**1071-1086.

29. **Manley, J. L. and R. Tacke.** 1996. SR proteins and splicing control. Genes Dev **10:**1569-1579.

30. **Moore, M. J., C. C. Query and P. A. Sharp.** 1993. Splicing of precursors to mRNAs by the spliceosome, p. 305-358. *In* R. F. Gesteland and J. F. Atkins (ed.), RNA World. Cold Spring Harbor Press, .

31. **Nagoshi, R. N. and B. S. Baker.** 1990. Regulation of sex-specific RNA splicing at the Drosophila doublesex gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. Genes Dev **4:**89-97.

32. **Ramchatesingh, J., A. M. Zahler, K. M. Neugebauer, M. B. Roth and T. A. Cooper.** 1995. A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. Mol Cell Biol **15:**4898-4907.

33. **Reed, R.** 1996. Initial splice-site recognition and pairing during pre-mRNA splicing. Curr Opin Gen Dev **6:**215-220.

34. **Rio, D.** 1993. Splicing of pre-mRNA: mechanism, regulation and role in development. Curr Opin Genet Dev **3:**574-584.

35. **Ryan, K. R. and T. A. Cooper.** 1996. Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle. *Mol. Cell. Biol.* **16:**4014-4023.

36. **Santisteban, I., F. X. Arredondovega, S. Kelly, M. Loubser, N. Meydan, C. Roifman, P. L. Howell, T. Bowen, K. I. Weinberg, M. L. Schroeder and M. S. Hershfield.** 1995. Three new adenosine deaminase mutations that define a splicing enhancer and cause severe and partial phenotypes: Implications for evolution of a CpG hotspot and expression of a transduced ADA cDNA. Hum Mol Gen **4:**2081-2087.

37. **Sirand-Pugnet, P., P. Durosay, E. Brody and J. Marie.** 1995. An intronic (A/U)GGG repeat enhances the splicing of an alternative intron of the chicken ß-tropomyosin pre-mRNA. Nuc Acids Res **23:**3501-3507.

38. **Staffa, A. and A. Cochrane.** 1995. Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. Mol Cell Biol **15:**4597-4605.

39. **Staknis, D. and R. Reed.** 1994. SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol Cell Biol **14:**7670-7682.

40. **Sun, Q., A. Mayeda, R. K. Hampson, A. R. Krainer and F. M. Rottman.** 1993. General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. Genes Dev. **7:**2598-2608.

41. **Tanaka, K., A. Watakabe and Y. Shimura.** 1994. Polypurine sequences within a downstream exon function as a splicing enhancer. Mol Cell Biol **14:**1347-1354.

42. **Tian, H. C. and R. Kole.** 1995. Selection of novel exon recognition elements from a pool of random sequences. Mol Cell Biol **15:**6291-6298.

43. **Tian, M. and T. Maniatis.** 1992. Positive control of pre-messenger-RNA splicing in vitro. Science **256:**237-240.

44. **Tian, M. and T. Maniatis.** 1993. A splicing enhancer complex controls alternative splicing of doublesex pre-messenger RNA. Cell **74:**105-114.

45. **Tian, M. and T. Maniatis.** 1994. A splicing enhancer exhibits both constitutive and regulated activities. Genes Dev. **8:**1703-1712.

46. **Tuerk, C. and L. Gold.** 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science **249:**505-510.

47. **van Oers, C. M., G. J. Adema, H. Zandberg, T. C. Moen and P. D. Baas.** 1994. Two different sequence elements within exon 4 are necessary for calcitonin-specific splicing of human calcitonin/calcitonin gene-related peptide I pre-mRNA. Mol Cell Biol **14:**951-960.

48. **Wang, Z., H. M. Hoffmann and P. J. Grabowski.** 1995. Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. RNA **1:**21-35.

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

50. **Xie, W. and L. I. Rothblum.** 1991. Rapid, small-scale RNA isolation from tissue culture cells. BioTechniques **11:**325-327.

51. **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.

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

	ACE	ACE and/ or PRE	n
NON-SELECTED	16%	23%	31
FIRST RND	25%	39%	28
SECOND RND	34%	47%	32
THIRD RND	53%	73%	30

Table 1: ACE and purine-rich sequences (PRE) were enriched during *in vivo* selection for exonic enhancers. Percent of sequenced clones that contain an ACE or PRE. Sequences rich in A/C and G/A were clearly enriched, however, no consensus for either motif was discernible. Therefore, ACE and purine rich motifs are defined as at least five consecutive A+C or G+A motifs, respectively. To exclude homopolymers, which are unlikely to have enhancer activity (41), the motif must contain both nucleotides at a ratio of ≥1:5. This criteria was based on the length and sequence requirements for activity of natural and synthetic purine rich motifs (12, 32, 41). It should be noted that a five nucleotide purine-rich sequence contributes significantly to the enhancer in bovine growth hormone exon 5 (12). To obtain non-selected exons, ligations containing the SalI/BamHI randomized cassette and SXN13 minigene vector were transformed into bacteria and colonies were picked at random for sequencing.

Gene	Sequence	Ref.
CT/CGRP exon 4	**ACUUCAACAA**GUU	(47)
cTNT exon 16	**CCACCA**GAAGGU	(48)
ß-globin/URA3 chimera	**CCAAGUACAA**CCCCCC**A**CU	(15)
dsx (1)	UCUUCAAUCAACAAG	(31)
dsx (2)	UCUUCAAUCAACAUU	"
dsx (3)	UCUACAAUCAACAUU	"
dsx (4)	UCUUCAAUCAACAAU	"
dsx (5)	UCAACAAUCAACAUA	"
HSV thymidine kinase	UCGCG**AACA**UCU**ACACCACACAACACC**GCCUCGA	(24)

Table 2: RNA processing elements containing A/C-rich motifs.

ACE-like elements are in bold font. The underlined nucleotides in the ß-globin/URA3 chimera sequence is a (U)6 -> (C)6 substitution that activated splicing of ß-globin splice sites. Five of the six repeats within the *Drosophila dsx* exon 4 enhancer are shown with the 13 nucleotide consensus sequences underlined.

FIGURE LEGENDS

Figure 1: *In vivo* selection for splicing enhancers. (A) See the Methods section for details of the selection procedure. (B) The sequence of the SXN13 middle exon is shown at the top of the column. Intron sequence is in lower case, exon sequence is in upper case; randomized positions are indicated as "N" and the SalI and BamHI restriction sites are underlined. All sequenced third round exons are shown below. In the selected sequences, purine-rich motifs are underlined and ACE motifs are in bold font (these motifs are defined in the legend to Table 1). To avoid selecting against premature stop codons which can result in exon skipping (10), the natural globin translation start codon was removed from the SXN13 minigene. Selected sequences from all three rounds contain equivalent numbers of stop codons in all three reading frames, indicating that there was no selection against stop codons.

Figure 2: Exons selected in the SXN13 minigene enhanced exon inclusion compared to non-selected exons. Individual clones were transfected into QT35 cultures and assayed by primer extension. The selected clones are, from left to right: 5.25.10, 5.25.13, 5.25.17, 6.5.19, 5.25.5, 5.25.8. All are from the third round and their sequences are shown in Figure 1B. Nonselected clones were obtained as described in the legend to Table 1. The primer extension products are 175 and 209 nucleotides for exon skipping and inclusion, respectively.

Figure 3: *In vivo* analysis of individual selected exons. Top: sequences of selected exons and mutants thereof used in Figures 3-6 are shown below the SalI/BamHI selection cassette. Nucleotide substitutions are highlighted in black. SYN is a synthetic ACE. To test pooled clones in panels Figures 3 and 4, a ligation containing the SXN13 or cTNT minigene and the SalI/BamHI randomized cassette was transformed into bacteria. More than 400 colonies were pooled, inoculated into 100 ml of broth, and grown overnight for preparation of plasmid DNA for transfection. Lane to lane differences in the level of expression are primarily due to differences in transfection efficiency. We determined that the level of exon inclusion is not affected by the total level of minigene RNA (data not shown).

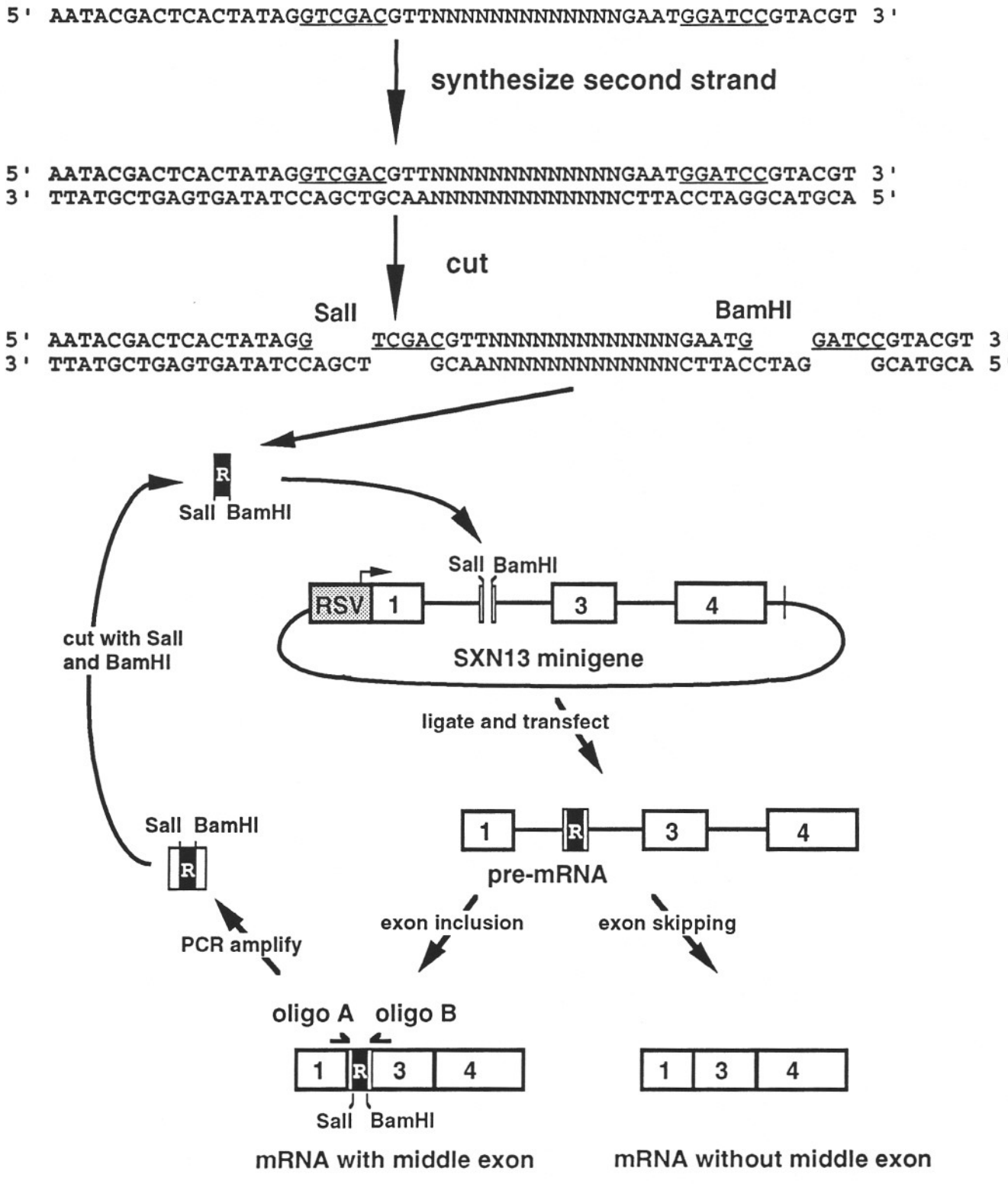
Figure 4: Selected exons enhance splicing in a different minigene. The cTNT minigene is shown at the top. Below the minigene is the sequence of the modified exon 5 in which the natural purine-rich enhancer was replaced by the selectable SalI/BamHI cassettes. Intron sequence is in lower case. SalI and BamHI sites are underlined. The primer extension products are 169 and 210 nucleotides.

Figure 5: *In vitro* analysis of selected and non-selected exons. The splicing substrate was derived from intron 4 of the cTNT minigene shown in Figure 4 (see Methods). All splicing substrates are identical except for the sequences shown in Figure 3. Reactions were performed in HeLa nuclear extract for 0, 45, and 90 minutes. Splice products and intermediates are identified to the left of the gel. The nonselected exon contains the sequence CACTCCTTACCGA and expresses 7% exon inclusion in the SXN13 minigene *in vivo*.

Figure 6: Splicing of the 4.11.12 substrate shown in Figure 5 was challenged using 4, 10, and 20 pmoles of competitor RNA. Competitor RNAs contain the last 50 nucleotides of cTNT intron 4 linked with the exon shown in Figure 4.

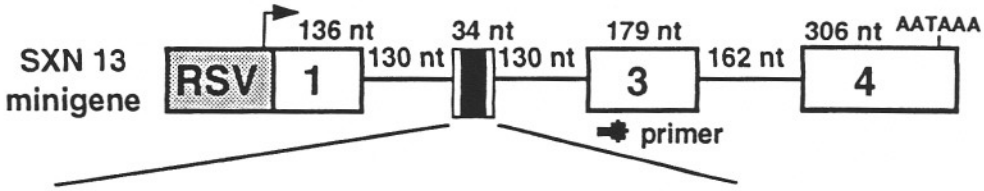
Figure 7: Splicing of 4.11.12 was challenged with 10, 30, and 50 pmoles of RNAs containing the cTNT exon 5 purine-rich splicing enhancer. The UP, WT and DOWN RNAs contain, respectively, a mutation that increases enhancer activity *in vivo*, the unmodified exon, and a mutation that nearly eliminates activity (32).

Figure 8: A single copy of the *Drosophila dsx* repeat enhances splicing of vertebrate splice sites in vertebrate cells. Sequences of the unmodified and mutant *dsx* elements within the 24 nucleotide SalI/BamHI selection cassette are shown at the top of the figure. Nucleotide substitutions are highlighted in black. The cassette was tested in the SXN13 and cTNT minigenes shown in Figures 2 and 4, respectively. The results from transient transfection and primer extension analysis are shown.


5' AATACGACTCACTATAGGTCGACGTTNNNNNNNNNNNNNNGAATGGATCCGTACGT 3'
synthesize second strand
5' AATACGACTCACTATAGGTCGACGTTNNNNNNNNNNNNNNGAATGGATCCGTACGT 3'
3' TTATGCTGAGTGATATCCAGCTGCAANNNNNNNNNNNNNNCTTACCTAGGCATGCA 5'
cut
SalI
BamHI
5' AATACGACTCACTATAGG TCGACGTTNNNNNNNNNNNNNNGAATG GATCCGTACGT 3'
3' TTATGCTGAGTGATATCCAGCT GCAANNNNNNNNNNNNNNCTTACCTAG GCATGCA 5'
R
SalI BamHI
SalI BamHI
RSV
1
3
4
SXN13 minigene
cut with SalI
and BamHI
ligate and transfect
SalI BamHI
R
1
R
3
4
pre-mRNA
PCR amplify
exon inclusion
exon skipping
oligo A
oligo B
1
R
3
4
SalI BamHI
1
3
4
mRNA with middle exon
mRNA without middle exon

ag<u>GTCGAC</u>GTT**NNNNNNNNNNNNN**GAAT<u>GGATCC</u>AGgt

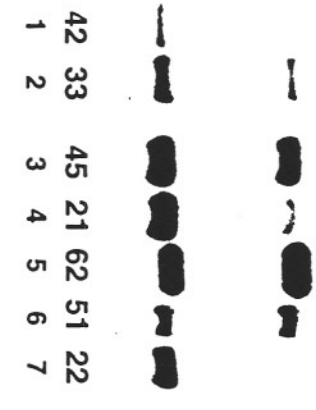
4.11.1	ACC<u>GAGAA</u>**CCCAC**
4.11.2	GGGAC<u>AGAGG</u>CGT
4.11.3	<u>GGAAAAAG</u>CGGCG
4.11.4	**CACCC**TCGGACGA
4.11.6	ATG**CAACCC**GGCT
4.11.7	CTGACCG**CCCCAA**
4.11.8	GGACGAGCCTGTG
4.11.9	CG**AACAC**<u>GAAGG</u>C
4.11.10	GAA**CCCACC**TGCC
4.11.11	GATGCC<u>GAGAG</u>CC
4.11.12	**CACCA**GT**CACC**GC
4.11.13	GATGTTAACCGCG
4.11.14	ATGGCAGCGACGA
4.24.1	CAGACTATGCCGC
4.24.4	<u>GGAAGAAGG</u>CGTG
4.24.5	GGG**C<u>AACA</u>**<u>GAAG</u>C
4.24.6	GCCGCATG**ACCAC**
4.24.7	**CCACCA**GAATGGC
5.25.1	**CACCC**CCGGCACT
5.25.3	TACG**ACAACCACC**
5.25.10	T**CCACA**GGGCCGC
5.25.13	TCGGCAGCTCGCT
5.25.17	CCATG**AACCAC**GC
5.25.2	**CCCACCAA**GCGCA
5.25.5	GTGGTGCAGCAGT
5.25.6	AC<u>GAGAG</u>CGCGTG
5.25.8	GATCGTGGCTCGC
6.5.19	<u>AGAAGAG</u>CGGCCC
6.5.25	TGCTGCGAGACGT
6.5.28	TTGACCG**ACACCC**



agGTCGACGTTNNNNNNNNNNNNNNGAATGGATCCAGgt

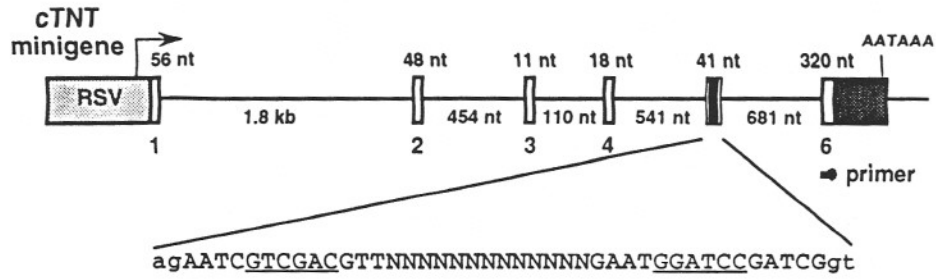
GTCGACGTTNNNNNNNNNNNNNNGAATGGATCC

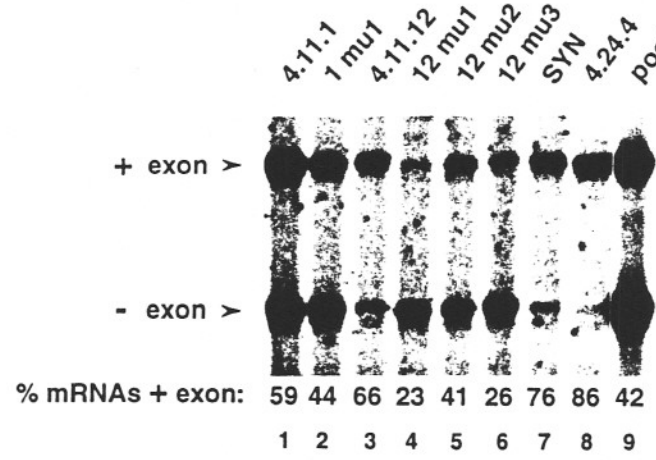
4.11.1	ACCGAGAAACCCAC
1 mu1	CTCGAGTTACCTGC
4.11.12	CACCAGTCACCGC
12 mu3	TATCAGTTTACCGC
12 mu1	CTGGCGTCGTCGC
12 mu2	CTGCTGCTTACGGC
4.24.4	GGAAGAAGGCGTG
SYN	GACCACAACAGAC

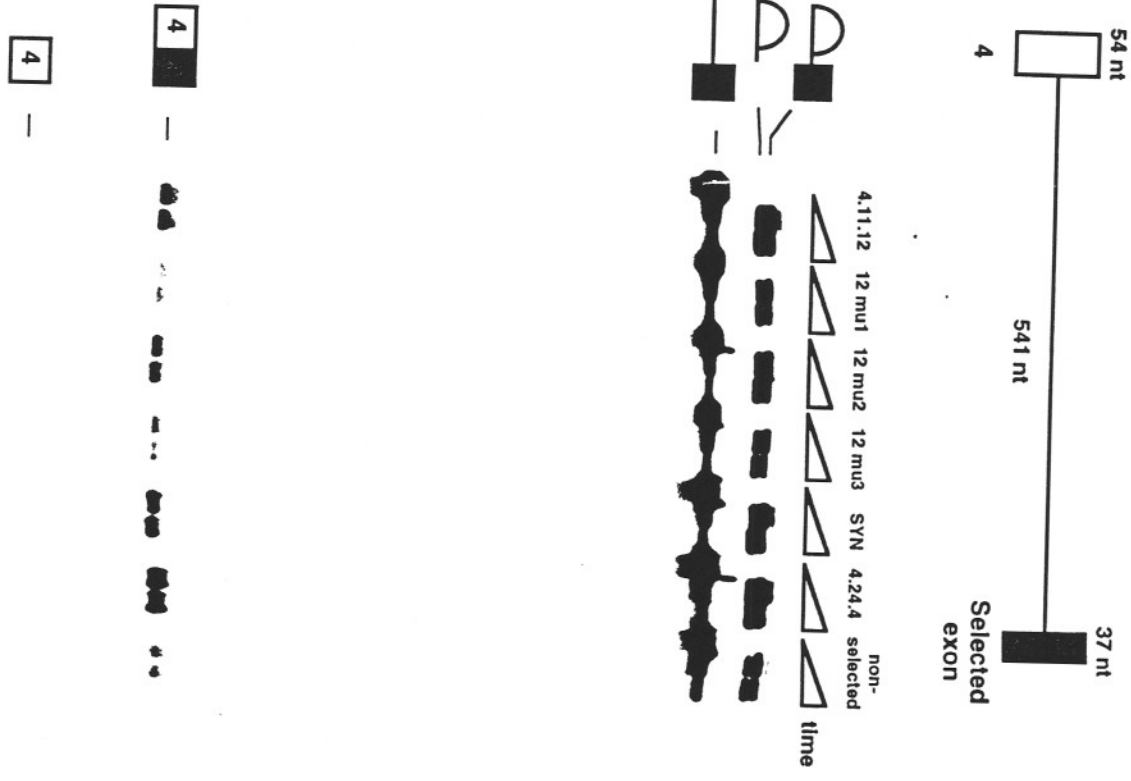


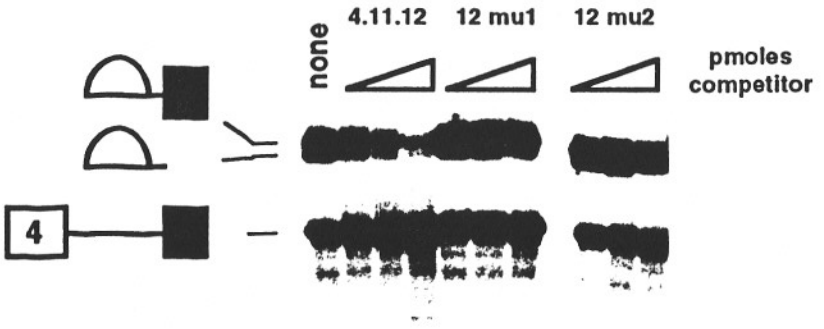
% mRNA + exon: 42 33 45 21 62 51 22

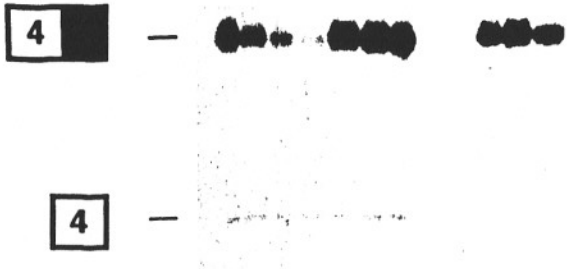
1 2 3 4 5 6 7

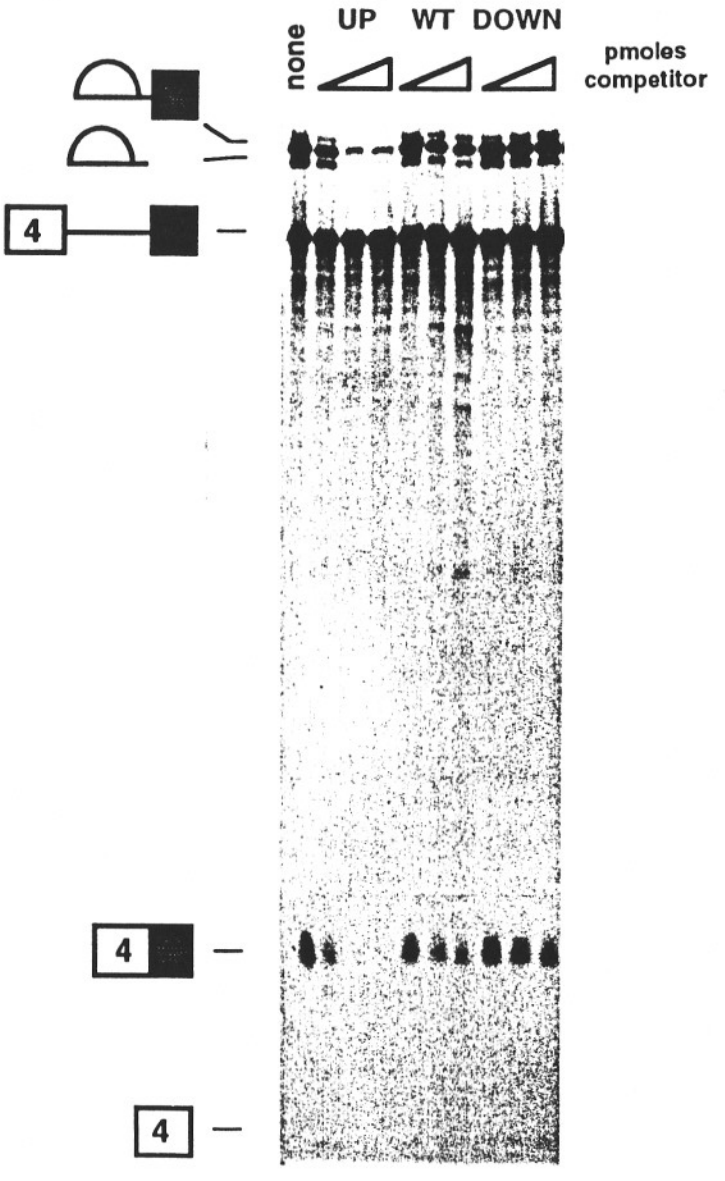

cTNT
minigene
56 nt
RSV
1.8 kb
48 nt
11 nt
18 nt
41 nt
320 nt
AATAAA
454 nt
110 nt
541 nt
681 nt
1
2
3
4
6
primer
agAATCGTCGACGTTNNNNNNNNNNNNNNGAATGGATCCGATCGgt

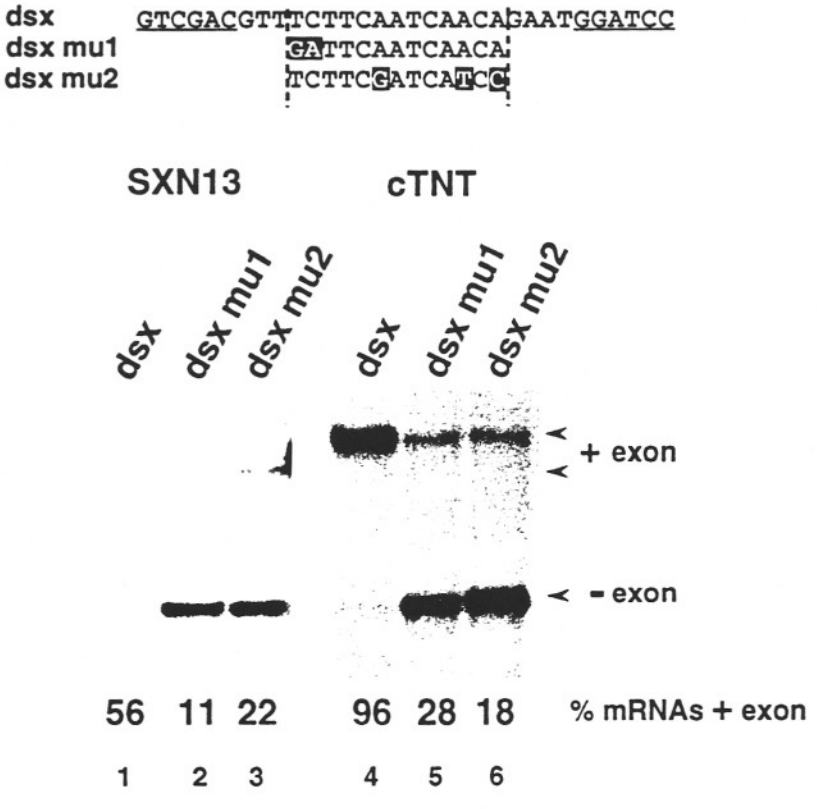

4.11.1
1 mu1
4.11.12
12 mu1
12 mu2
12 mu3
SYN
4.24.4
pool
+ exon
- exon
% mRNAs + exon:
59 44 66 23 41 26 76 86 42
1 2 3 4 5 6 7 8 9


4
54 nt
541 nt
37 nt
Selected exon
4.11.12
12 mu1
12 mu2
12 mu3
SYN
4.24.4
non-selected
time
4
4


none
4.11.12
12 mu1
12 mu2
pmoles competitor
4


4
4


none
UP
WT
DOWN
pmoles competitor
4
4
4


dsx GTCGACGTTTCTTCAATCAACAGAATGGATCC
dsx mu1 GATTCAATCAACA
dsx mu2 TCTTCGATCATCG
SXN13
cTNT
dsx
dsx mu1
dsx mu2
dsx
dsx mu1
dsx mu2
+ exon
- exon
56 11 22 96 28 18 % mRNAs + exon
1 2 3 4 5 6