Relative exon affinities and suboptimal splice site signals lead to non-equivalence of two cassette exons

Athena Andreadis^{*,+,§}, Jennifer A. Broderick⁺ and Kenneth S. Kosik

Department of Neurology (Neuroscience), Harvard Medical School and Center for Neurologic Diseases,
Department of Medicine (Division of Neurology), Brigham and Women's Hospital, Boston, MA 02115, USA

Received March 13, 1995; Revised and Accepted June 30, 1995

GenBank accession nos L35768 and L35769

ABSTRACT

Tau is a microtubule-associated protein whose transcript undergoes complex regulated splicing in the mammalian nervous system. Exons 2 and 3 of the gene are alternatively spliced cassettes in which exon 3 never appears independently of exon 2. Expression of tau minigene constructs in cells indicate that exon 2 resembles a constitutive exon, while a suboptimal branch point connected to exon 3 inhibits inclusion of exon 3 in the mRNA. Splicing of the two tau exons is controlled by their relative affinities for each other versus the affinities of their flanking exons for them.

INTRODUCTION

Alternative splicing is a versatile and widespread mechanism for generating multiple mRNAs from a single transcript (1). Splicing choices are regulated by both tissue/cell type and developmental stage; the mRNAs arising from such processing produce functionally diverse protein isoforms.

Splicing occurs in two transesterification reactions, with the participation of the spliceosome, a large complex of proteins and snRNAs (1,2). *In vitro* studies (3,4) and splicing mutant studies (listed in 4) have shown that the snRNAs play an important role in exon definition and splice site hierarchies. In addition to the invariant 5' and 3' splice sites, *in vitro* splicing studies led to the discovery of a less well-conserved splicing signal, the branch point (consensus sequence YRYURAY, R = purine, Y = pyrimidine). The branch point is normally located 30–50 nt upstream of the 3' splice site and is followed by a polypyrimidine [poly(Y)] tract (5). Within a first approximation, the 3' splice site is defined as the first YAG downstream of the branch point (6,7).

Whereas much knowledge has accumulated about both the *cis* and *trans* requirements of constitutive splicing (1,2), a major unanswered question is what distinguishes a cryptic splicing site from an authentic one. Such a distinction is important, because mammalian splice sites are loosely defined with respect to sequence and thus redundant in the genome. The exon definition model (in which snRNAs are postulated to recognize exon

boundaries by attachment at the 3' splice site of an exon and scanning downstream for a 5' splice site within a certain distance) has provided a partial explanation for splice site authentication (3). However, in alternative splicing, splice sites which adhere to the consensus sequences are either utilized or by-passed when the primary transcript is processed (1,8).

When alternatively spliced genes are expressed in inappropriate contexts they are not spliced constitutively, but produce one of the potential correct mRNAs, called the default mode (1,8). This finding implies that even higher order splicing decisions are made partly or wholly *in cis*.

Studies of *cis* determinants of alternative splicing have shown that exon behavior can be dictated by exon length (9–11), agreement of the splice sites with the consensus (9,10,12–14), location of the branch point (6,7,15–17) and length and composition of the associated poly(Y) tract (7,13,15–20). Suboptimal splice sites and/or displaced branch points lead to regulation of splice site selection, apparently via intrinsic hierarchies defined by complementarity of splice site signals to specific snRNAs (20,21). In several other systems exon inclusion is promoted by purine-rich sequences within the regulated exon (22–24).

The tau protein is enriched in neurons (25) and is the major component of the Alzheimer neurofibrillary tangles (26). The tau transcript contains 16 exons, of which six are cassettes (27). Two tau exons, 2 and 3, fall in the rare 'incremental combinatorial' category, in which the downstream exon of an alternatively spliced pair never appears by itself, but all other combinations are allowed (Fig. 1A). Exons 2 and 3 are adult-specific in the central nervous system (28–30), although they seem to be constitutive in the peripheral nervous system (31,32).

The present study has focused on the behavior of constructs containing these tau cassette exons in non-neural (COS) and neuroblastoma (SK-N-SH) cells. The relevant tau genomic region and the expression vector utilized are diagrammed in Figure 1B and C respectively. The expression constructs showed that exon 2 was almost always present in the construct mRNAs in both neuroblastoma and non-neural cells, regardless of the identity of the flanking exons. On the other hand, exon 3 was incorporated inefficiently in the construct mRNAs unless either

* To whom correspondence should be addressed at: Department of Biomedical Sciences, E.K. Shriver Center

Present addresses: ⁺Department of Biomedical Sciences, E.K. Shriver Center for Mental Retardation, Waltham, MA 02254, USA and [§]Department of Neurology, Harvard Medical School and Massachusetts General Hospital, Boston, MA 02116, USA

Figure 1. (A) Splicing patterns of the human tau gene in the region of exons 1–4. (B) Detailed restriction map of the region around tau exons 2–4. The relative positions of the exons are correct, but exon size is exaggerated. Sites marked with an asterisk occur within the exons. (C) Diagrammatic representation of expression vector pSVIRB and the largest genomic fragment used from the tau 2/3 region. SV40 promoter/enhancer regions are black, insulin exons and terminators diagonally striped, tau exons 2 and 3 white. For the sake of clarity, in subsequent figures insulin exon 1 and the small intron between insulin exon 1 and 2 are not shown in any of the constructs and the truncated insulin exons 2 and 3 are not shown in constructs containing tau exons 1 and/or 4. Primers used for PCR are indicated.

the region around its putative branch point/polypyrimidine tract was modified or exon 2 was pre-spliced to the upstream exon.

MATERIALS AND METHODS

Cell culture

Cells were cultured on 100 mm plates. Media and supplements came from Gibco-BRL, except for fetal calf serum (FCS; Hyclone). COS (monkey kidney) cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% FCS. SK-N-SH (human neuroblastoma) cells were maintained in Eagle's minimal essential medium supplemented with non-essential amino acids and 10% FCS.

Plasmid construction

The parental vector used for cloning was pSVIRB (33; Fig. 1C). pSVIRB contains an additional intron between fused SV40–insulin exon 1 and insulin exon 2 that is not shown in Figures 3A–7A, but which served as an internal splicing control. The tau genomic fragments originated from cosmid or λ clones (27).

Tau chimeric constructs were produced by standard cloning methodology or by PCR, insertion into pKS⁺ Bluescribe (Stratagene) and subsequent cloning by directional ligation (34). The most frequently used primers for PCR are shown in Figure 1C. All inserts generated by PCR were sequenced, to ensure absence of mutations. The following minigene constructs were generated:

SV2/3 contains tau exons 2 and 3 and their flanking introns (1.0 kbp upstream of exon 2, 1.4 kbp downstream of exon 3 and 2.6 kbp between them). SV2/ Δ 3 contains tau exon 2 and its flanking introns (1.0 kbp upstream and 1.0 kbp downstream). SV Δ 2/3

contains tau exon 3 and its flanking introns (1.1 upstream and 1.4 kbp downstream).

The SV1/N/4 constructs contain tau exon 1 plus 73 nt of its downstream intron fused to insulin exon 2 and tau exon 4 plus ~500 nt of its upstream intron fused to insulin exon 3. With respect to tau exons 2 and 3, SV1/2/3/4, SV1/2/4 and SV1/3/4 contain the same inserts as SV2/3, SV2/ Δ 3 and SV Δ 2/3 respectively.

SVf2/3 contains the 3' two thirds of tau exon 2 fused to insulin exon 2, the entire 2.6 kbp intron between tau exons 2 and 3, tau exon 3 and 1.4 kbp of its downstream intron. SV Δ 2/3ST and SVf2/3ST contain tau exon 3, 84 nt of its upstream and 125 nt of its downstream intron respectively embedded into SV Δ 2/3 and SVf2/3 from which the original long tau exon 3 insert had been excised (this leaves 1.2 kbp of intron downstream of exon 2).

SVf2/3KT, 3MT and 3LT differ from SVf2/3ST in the length of the tau exon 3 upstream intron: 278, 187 and 142 nt respectively. SVf2/3BPT and BPiT are identical to SVf2/3LT, except they contain non-tau inserts of 20 and 140 nt respectively at a distance of 84 nt upstream from the start of tau exon 3. The starting points for all 3NT constructs are shown in Figure 2 (N = K, M, L or S). In all the 3NT constructs the length of the intron between tau exons 2 and 3 is at least 1.6 kbp: the 5' 1.5 kbp of the intron is invariant and is the native sequence downstream of exon 2.

SV Δ 2/3UST contains tau exon 3 up to position 142 in its upstream intron and 125 nt of its downstream intron embedded into SV Δ 2/3 from which the original long tau exon 3 insert had been excised. Within the upstream intron, 83 nt have been deleted. The deletion ends exactly at the start of tau exon 3.

SV7 contains tau exon 7 and its flanking introns (1.0 kbp upstream and 2.0 kbp downstream). SV3/7 contains the 5' one third of tau exon 3 and 1.6 kbp of its upstream intron fused to exon 7 in

```

-158  ttggtccctttgtgggtttgttcagggcggtgtccagctgtttccacagggagcgattttcagctccacaggac
-83   actgctccccagttcctcctgagaacaaaagggggcgctggggagagggccaccgttctgagggctcactgtatg
      (3/4)      (5/9)
-8    tgttccagAAT(exon 2)AAGgtgggc - 1.5 kbp -- SpeI -- 0.8 kbp --
      ┌ 3K
-280  tgcagcagctctgtttcactgcagcgtttacacagggctgccgggctttcctgggtgatgagctggcgcttcatga
      ┌ 3M
-205  gccagaaccactcagcagcatgtcagtgcttctctggggagactggtagcaggggctccgggcctacttcaggg
      BPd(5/6) polyY      ┌ 3S
-130  ctgctttctggcatatggctgatccctcctcactcctcctcctcctgcatgctcctgcgcaagaagcaagggtga
      BPp(3/6)      (2/4)      (7/9)
-55   ggggctgggtatggctcgtcctggccctctaaggtggatctcggtggtttctagATG(exon 3)CAGgtgagg

```

Figure 2. The sequence around human tau exons 2 and 3 (partial, both exons are 87 nt long). Exon sequences are in upper case, intron sequences in lower case. The two sets of numbers show the distance from the beginning of each exon. The two putative branch points of exon 3 are underlined. PolyY indicates the polypyrimidine tract associated with the putative distal element. Numbers in parentheses show agreement of splice sites and branch points with the consensus sequences. 3K, M, L and S are the starting points for constructs discussed in the text and in Figures 5–7. The *SpeI* site demarcates the extent of the left intron side for all the f2/3 constructs.

SV7. The SV3N/7 series were generated by substituting the corresponding fragments from the SVf2/3NT constructs into SV3/7 (regenerating the fused 3/7 exon and variable lengths of upstream intron). SV7/3 contains the 3' two thirds of tau exon 3 and 1.4 kbp of its downstream intron fused to exon 7 in SV7.

Sequencing

Plasmid DNA was prepared by the Qiagen miniprep method, then sequenced by the modified dideoxynucleotide method (35,36) with Sequenase 2.0 (US Biochemicals). The sequencing ladders were resolved on denaturing (8.3 M urea) 6% (w/v) polyacrylamide gels and analyzed using GCG software (37).

Transfections and RNA preparation

Plasmid DNA was purified using Qiagen Tip-100s or by cesium chloride banding (34). Plasmids were introduced into COS cells by the calcium phosphate method and into SK-N-SH cells by the lipofection method (DOTAP; Boehringer-Mannheim; 38).

Cells that had reached ~30% confluence were fed 2 h prior to transfection. Each plate was transfected with 10 µg construct DNA. The medium was changed 16 h after transfection, without glycerol shock. Total RNA was isolated 48 h post-transfection by the RNazol method (Cinna/Biotech).

Reverse transcription and polymerase chain reactions

For PCR analysis of RNA 5 µg total RNA from transiently transfected cells were reverse transcribed using random hexamer primers and 200 U reverse transcriptase (RNAase H⁻ Superscript; BRL) in a total volume of 20 µl for 1 h at 42°C. Part of this reaction mix (5 µl) was then diluted to a final volume of 50 µl, the concentrations of the buffer and dNTPs were adjusted for PCR and the mixture was amplified for 25 cycles (denaturation at 94°C for 1 min, annealing at 65°C for 1 min, extension at 72°C for 2 min; 39). The primers were chosen to only amplify products arising from the constructs. The primer pair most commonly used was INS1 with INS3 (shown in Fig. 1C). INS1 was also used with HTE3 to detect the presence or absence of exon 3 (Figs 3C and 4C).

The RT-PCR product of any construct which was different in size from that of spliced vector pSVIRB was sequenced, to ensure that splicing had occurred at the correct junctions. The RT-PCR experiments were done three times with RNAs from three independent transfections, to ensure that the results were reproducible.

The sizes of the RT-PCR products were calculated as follows: The relevant fragment of insulin exon 1 is 30 nt, the tau 1–insulin 2 fused exon is 170 nt, intact insulin exon 2 is 200 nt, the relevant fragment of insulin exon 3 is 60 nt and the tau 4–insulin 3 fused exon is 90 nt. Thus RT-PCR products arising from insulin-specific primers (Figs 3B and 4B) produce bands of the same size (290 nt if both tau exons were excluded, 380 or 470 nt if one or both tau exons were included) because the two 'halves' of the constructs compensate for each other in length (30 + 170 + 90 versus 30 + 200 + 60). However, RT-PCR products arising from use of INS1 and HTE3 (Figs 3C and 4C) differ depending on the exon upstream of the 2/3 insert; those constructs containing tau exon 1 give rise to a product which is 30 nt shorter than that from constructs containing intact insulin exon 2.

RNAase protection assays

The probe used (SV3p) was an RT-PCR product of a construct in which tau exon 3 was fused to insulin exon 2 at the *Bam*HI site (INS1/INS3 primer pair). An antisense riboprobe uniformly labeled with [α -³²P]CTP (PB20382; Amersham) was generated according to the Promega *in vitro* transcription protocol.

Full-length riboprobe was isolated by running the reaction mixture on a 5% acrylamide–8.3 M urea gel, excising the corresponding slice and eluting it in 0.5 M NH₄OAc at 37°C overnight. Equivalent amounts of the probe were dissolved in 22 µl Ambion hybridization buffer, then added to 10 µg total RNA from COS cells transfected with the tau constructs, heated at 94°C for 3 min, then allowed to hybridize at 42°C for 16 h. A 1:100 dilution of RNase A/RNase T1 mixture and digestion buffer (Ambion) were added to each tube and allowed to react for 30 min at 37°C. The reaction was precipitated, dissolved in formamide loading buffer and run on a 6% acrylamide–8.3 M urea gel to resolve the protected fragments. For each probe, two

Figure 3. In non-neuronal (COS) cells, the default pattern for human tau exons 2 and 3 is inclusion and exclusion respectively. (A) Schematic representation of expression constructs containing intact exons 2 and/or 3 with homologous or heterologous flanking exons. SV40 promoter/enhancer regions are black, insulin exons and terminators diagonally striped, tau exons 1 and 4 vertically striped, tau exons 2 and 3 white. Major splicing pathways are shown by solid, minor by dashed lines. RT-PCR of the expression constructs using (B) INS1/INS3 and (C) INS1/HTE3 primer pairs. The identities of the spliced species are indicated.

tubes containing yeast RNA, one with and one without RNase, served as negative and positive controls respectively.

RESULTS

The default pattern for exon 2 is inclusion, although it lacks a polypyrimidine tract

The intron upstream of human tau exon 2 is very purine-rich and contains no polypyrimidine tract longer than 8 nt up to 158 nt upstream of exon 2 (Fig. 2). Shortening of the polypyrimidine tract to <9 nt results in inhibition of splicing (6). Inhibition of splicing also results from placing purine-rich tracts at or near the branch point (10).

Nevertheless, this intrinsic characteristic of exon 2 does not result in inhibition of its splicing. The RT-PCR results showed that in COS cells exon 2 was present in the mRNA of the construct. The presence or absence of exon 3 made no difference (Fig. 3A and B; SV2/3 or SV2/Δ3). Furthermore, exon 2 was always included, whether the flanking exons were heterologous (Fig. 3A and B; SV2/3 and SV2/Δ3) or homologous (Fig. 3A and B; SV1/2/3/4 and SV1/2/4).

Human neuroblastoma SK-N-SH (SKN) cells express tau at low levels. The endogenous tau of SKN expresses all three 2/3 isoforms (2⁻3⁻, 2⁺3⁻, 2⁺3⁺). In contrast to COS cells, in SKN cells exon 2 was alternatively spliced: species with and without exon 2 were

detected whether this exon was flanked by homologous or heterologous exons and regardless of the presence or absence of exon 3 (Fig. 4A and B; SV2/3, SV2/Δ3, SV1/2/3/4, SV1/2/4).

The products visible in Figures 3B and C and 4B and C were the correct size for the indicated species. Cloning and sequencing of the RT-PCR products showed that the tau exons were precisely spliced; no cryptic 5' or 3' splice sites were utilized. The fact that the SV1/2/3/4 minigene reproduces the *in vivo* tau splicing patterns indicates that the SKN cells are a valid system for studying tau splicing regulation.

Exon 3 is inefficiently spliced, but not due to competition from exon 2

In both COS and SKN cells, exon 3 was inefficiently incorporated into the construct mRNA (Figs 3A and B and 4A and B; SV2/3 and SV1/2/3/4). When exon 2 was deleted the predominant species in both cell types was 3⁻ (Figs 3A and B and 4A and B; SVΔ2/3 and SV1/3/4). The identity of the flanking exons appeared irrelevant (Fig. 3A and B; SV2/3, SVΔ2/3, SV1/2/3/4 and SV1/3/4). Therefore, exons 2 and 3 are not in competition; exon 3 is excluded due to an intrinsically suboptimal splicing signal.

In SKN cells, where exon 2 became optional (previous section), exon 3 did not become incorporated into the construct mRNA by itself when exon 2 was present: use of an antisense primer specific

Figure 4. In neuroblastoma (SKN) cells, tau exon 2 is alternatively spliced, but tau exon 3 is still largely excluded. (A) Schematic representation of expression constructs containing intact exons 2 and/or 3 with homologous or heterologous flanking exons. Color/pattern conventions for the constructs and their splicing pathways are as in Figure 3A. RT-PCR of the expression constructs using (B) INS1/INS3 and (C) INS1/HTE3 primer pairs. The identities of the spliced species are indicated.

for exon 3 showed only a band corresponding to a 2^+3^+ species, regardless of the identity of the flanking exons (Fig. 4A and C; SV2/3 and SV1/2/3/4). Thus in the presence of exon 2 the splicing pattern duplicates that seen *in vivo* by giving no 2^-3^+ product. However, in both cell types, when exon 2 was physically absent exon 3 was incorporated by itself, although inefficiently: the PCR product including exon 3 became visible only when a primer specific for exon 3 was utilized, so that there was no competition for the probe among different species in the PCR reaction (compare SVΔ2/3 and SV1/3/4 products in Figs 3B versus C and 4B versus C).

In COS cells, whereas exon 3 was almost undetectable in the absence of exon 2 (Figs 3A and B and 4A and B; SVΔ2/3 and SV1/3/4), a readily visible species that included exon 3 was detectable in the construct mRNA when exon 2 was pre-spliced to rat insulin exon 2 (Fig. 5A and B; compare SVΔ2/3 with SVf2/3). The same result was obtained by RT-PCR (data not shown). This result indicates that exon 2, once its own 3' splice site is removed, has higher intrinsic affinity for exon 3 than either homologous exons 1 and/or 4 or the heterologous insulin exons.

The behavior of exon 3 arises from a region near its branch point

The sequence around the 3' splice site of exon 3 contains two potential branch points, one very close in sequence to the consensus and followed by a long polypyrimidine tract but displaced upstream (−90), the other in poor agreement with the

consensus and with a very short polypyrimidine stretch but located at −50, within the usual location range for branch points (Fig. 2).

To test the possibility of competition between branch points, a deletion was made that contained exon 3 with minimal flanking introns (Fig. 5A and B; SVf2/3ST and SVΔ2/3ST). Whether exon 2 was deleted or pre-spliced upstream, exon 3 inclusion increased dramatically over the equivalent constructs which contained exon 3 with large segments of its flanking introns. The tendency towards including more of exon 3 if exon 2 was upstream persisted in both constructs compared with the original analogous constructs with longer introns (SVf2/3 and SVΔ2/3 respectively). The RT-PCR results concurred with those from the RNase protection experiments; sequencing of the products again showed that the utilized spliced junctions were correct in all cases.

Subsequently several constructs were made which progressively reconstructed the upstream intron flanking exon 3 (Fig. 6). Starting points were chosen to bracket a sequence stretch that could potentially base pair with the putative proximal branch point (K and M or L), while others framed the putative distal branch point (M or L and S). All these constructs precisely spliced exon 2 to exon 3 in COS cells (Fig. 6A and B; SVf2/3NT, where N = K, L or S). The two tau exons also spliced to each other in two constructs in which the distance between the upstream and downstream elements had been increased by insertion of 20 and 150 heterologous nucleotides (Fig. 6A and B; SVf2/3BPT; the results from SVf2/3BPiT, which contains the longer insertion, are not shown, but are identical to those of SVf2/BPT).

Figure 5. The identity of the exon upstream of tau exon 3 is important. (A) Schematic representation of expression constructs containing exon 3 and (a) either long or minimal flanking introns with (b) exon 2 either deleted or pre-spliced to insulin exon 2. Color/pattern conventions for the constructs and their splicing pathways are as in Figure 3A. The numbers under the intron lines represent total intron length (kbp). (B) RNase protection of the expression constructs in COS cells. – and +, lanes with probe in the absence of RNA or RNase respectively. The identities of the spliced species are indicated. The probe utilized is diagrammatically depicted below the autoradiogram, with the sizes of all potential spliced species shown.

Finally, construct SVΔ2/3UST was made in order to determine whether a 3' splice site can be recruited if the branch point region proximal to exon 3 has been deleted. In this construct exon 3 is primarily excluded. However, the minor product that includes it uses the correct 3' splice site for exon 3 (Fig. 5A and B and results from cloning and sequencing the RT-PCR product).

The behavior of all the constructs in this section in SKN cells was identical to that seen in COS cells (data not shown). Thus the behavior of exon 3 is not due to competing branch point elements; rather, the branch point of exon 3 is intrinsically suboptimal.

Exon 3 can confer its splicing behavior to a heterologous exon

To examine if the behavior of exon 3 could be transferred, constructs were made in which either the 5' or 3' side of exon 3 was fused to a constitutive exon (Fig. 7).

Exon 7 is constitutively included when by itself (Fig. 7A and B; SV7). When the 3' side of exon 3 was fused to exon 7 the resulting 7/3 fusion exon was alternatively spliced in COS cells (Fig. 7A and B; SV7/3). On the other hand, when the 5' side of exon 3 was fused to exon 7 the resulting construct by-passed the 3/7 fusion exon in both COS (Fig. 7A and B; SV3/7) and SKN cells (data not shown). The low efficiency of exon 3 inclusion

persisted regardless of the intron length (Fig. 7A and B; SV3N/7, where N = M, BP or S). These results are consistent with those from the previous section in suggesting that the splicing defect of exon 3 resides primarily in its branch point/3' splice site region.

DISCUSSION

Comparison of tau exons 2 and 3 with other 'incremental combinatorial' exons

The splicing pattern of tau exons 2 and 3 has been documented in two other systems: exons 7 and 8 of the amyloid precursor protein gene (40) and neuron-specific exons N1 and N2 of the *c-src* gene (9,41). The default pattern of exon N1 is exclusion, dictated partly by its extreme shortness (9); however, full activation requires sequences of the downstream intron (42). Alignments of sequences from equivalent regions of the three genes by the GCG programs (37) showed no obvious homologies.

The role of exon length and sequences

Short cassette exons show a default phenotype of exclusion and become constitutive when lengthened (9,10). The strong implication is that their extreme shortness prevents spliceosomes from assembling at both ends simultaneously. However, tau exons 2 and

Figure 6. The behavior of exon 3 is not due to competition between branch point elements. (A) Schematic representation of expression constructs containing exon 3 with various lengths of its upstream intron and exon 2 pre-spliced to insulin exon 2. Color/pattern conventions for the constructs and their splicing pathways are as in Figure 3A. The numbers under the intron lines represent total intron length (kbp). (B) RNase protection of the expression constructs in COS cells. Lane marking and probe depiction conventions are as in Figure 5B.

3 are long enough (87 nt) that steric hindrance between their 5' and 3' splice sites would not be the operative regulatory mechanism in their case.

It is unlikely that sequences within tau exon 3 itself modulate splicing: the 3N/7 and f2/3NT series behave very similarly, with the former containing only the 5' most 30 nt of exon 3 and the latter the entire exon.

The role of flanking sequences

Flanking exons do not play a role in most alternative splicing systems. The exceptions so far are tau exon 6 (33), fibronectin exon EIIIB (43), the cardiac troponin T mini-exon (12) and the mutually exclusive exons in myosin light chain 1/3 (44).

Flanking exons do not play a role in splicing of exon 2. The implications from this and from its constitutive inclusion in COS cells are that: (i) the determinants of exon 2 selection must lie within the exon itself; (ii) its absence from fetal neurons is most likely due to an inhibitory *trans* factor(s). For exon 3 the identity of the downstream exon also appears irrelevant; however, it is strongly activated by the presence of pre-spliced exon 2 upstream.

For cassette exons, intronic regulatory sequences tend to be near the exon and redundant. In the *c-src* N1 exon, the putative positive region ends 142 nt downstream of the exon (42). The

exception to this is fibronectin exon EIIIB, which is regulated by elements located at least 519 nt from the exon (43,45).

In all tau constructs intron lengths exceed 140 nt. In particular, the intron between tau exons 2 and 3 is always at least 1.6 kbp long. Thus the behavior of exon 3 in the f2/3NT constructs (Figs 5 and 6) cannot be attributed to intron length, but must arise from the intronic region upstream of and proximal to the exon.

Exon 3 splice site choice and the modified scanning hypothesis

Exon 3 seems to have suboptimal splice sites, but its 3' splice site is particularly implicated in its suppression. Deletions or insertions in the 3' proximal region of the upstream intron of exon 3 do not alter the behavior of exon 3. The suboptimality of the exon 3 branch point is apparently due to its poor agreement with the consensus sequence and its very short polypyrimidine tract. Such a sequence would be by-passed by the splicing apparatus unless an activating factor attracted it there.

The hierarchy of 3' splice site strengths is $CAG \approx TAG > AAG \gg GAG$ *in vivo* and *in vitro*. It was originally thought that the 3' splice site was passively defined as the first YAG after the branch point (6). However, recent experiments (46) indicate that scanning is combined with weighting of splice site strength.

Figure 7. Tau exon 3 can confer its behavior to a heterologous exon. (A) Schematic representation of expression constructs containing fusions of tau exons 3 and 7. Color/pattern conventions for the constructs and their splicing pathways are as in Figure 3A; tau exon 7 is vertically striped. (B) RT-PCR of the expression constructs in COS cells using the INS1/INS3 primer pair. The identities of the spliced species are indicated.

The 3' splice site of exon 3 is TAG/A, which deviates from the consensus CAG/G; this has been seen in other neuron-specific exons (47). Upstream of exon 3 is an AAG located between the proximal putative branch point and the authentic 3' splice site TAG of exon 3. Whenever incorporated, exon 3 never utilized this cryptic site. Interestingly, if used it would result in a tau protein containing seven additional amino acids.

Splicing regulation in the tau 2/3 region

The specific functions of the protein domains encoded by tau exons 2 and 3 are unknown. They are not directly involved in microtubule binding, which is localized in the C-terminal domain of tau (48,49). The default inclusion of exon 2 correlates with its constitutive presence in the 9 kb tau mRNA of peripheral neurons (31,32). Conversely, the constitutive presence of exon 3 in that mRNA (31,32) implies the existence of a specific activator.

The results from this study show that the splicing behavior of tau exons 2 and 3 can be largely explained by *cis* characteristics of the exons themselves: the splice sites of exon 3 are suboptimal and the 5' splice site of exon 2 has the highest affinity for the 3' splice site of exon 3, once exon 2 has been incorporated.

Such behavior has been documented for the mutually exclusive exons of β -tropomyosin (16,50,51). Thus, like tropomyosin, the

two tau exons are weakly coupled. Moreover, evidence from many systems indicates that the position and sequence of the branch point play a crucial role for regulation of alternative splicing (5–7,15–20), as is the case for tau exon 3. The characteristics of the tau system ensure that a 2⁺3⁺ species will not be produced and that mRNAs which include exon 3 will be minor in the absence of specific activating factors.

The regulation of these splicing events may be involved in the Alzheimer's disease process. There is recent evidence that the 2⁺3⁺ tau isoform has increased phosphorylation of Thr39 in the domain encoded by exon 1 (52) and that the normal ratio of tau isoforms is disturbed in Alzheimer brain, with the 3⁺ form increased (53).

ACKNOWLEDGEMENTS

We wish to thank Drs Norma Gerard for the try-out batch of DOTAP reagent, Chris Smith for the COS cells and Doug Black for the gift of the mouse genomic clone containing exon N2, Albarosa Alicea (NHLBI grant HL 07718), Soojin Ryu, Ho Leung Ng, Paul Wu and Bridget Wagner for help with the minigene constructs and Lisa Davis Orecchio for additional technical assistance. This work was supported by Alzheimer's Award FSA-90-005 and NSF grant IBN9222061/9496276 to AA and NIH grant AG06601 to KSK.

REFERENCES

- Sharp, P.A. (1994) *Cell*, **77**, 805–815.
- Green, M.R. (1991) *Annu. Rev. Cell Biol.*, **7**, 559–599.
- Robberson, B.L., Cote, G.J. and Berget, S.M. (1990) *Mol. Cell. Biol.*, **10**, 84–94.
- Talerico, M. and Berget, S.M. (1990) *Mol. Cell. Biol.*, **10**, 6299–6305.
- Reed, R. and Maniatis, T. (1988) *Genes Dev.*, **2**, 1268–1276.
- Smith, C.W.J. and Nadal-Ginard, B. (1989) *Cell*, **56**, 749–758.
- Smith, C.W.J., Porro, E.B., Patton, J.G. and Nadal-Ginard, B. (1989) *Nature*, **342**, 243–247.
- Andreadis, A., Gallego, M.E. and Nadal-Ginard, B. (1987) *Annu. Rev. Cell Biol.*, **3**, 207–242.
- Black, D. (1991) *Genes Dev.*, **5**, 389–402.
- Dominski, Z. and Kole, R. (1991) *Mol. Cell. Biol.*, **11**, 6075–6083.
- Stern, D.A. and Berget, S.M. (1993) *Mol. Cell. Biol.*, **13**, 2677–2687.
- Cooper, T.A. and Ordahl, C.P. (1989) *Nucleic Acids Res.*, **17**, 7905–7921.
- Dominski, Z. and Kole, R. (1992) *Mol. Cell. Biol.*, **12**, 2108–2114.
- Tacke, R. and Goridis, C. (1991) *Genes Dev.*, **5**, 1416–1429.
- Chebli, K., Gattoni, R., Schmitt, P., Hildwein, G. and Stevenin, J. (1989) *Mol. Cell. Biol.*, **9**, 4852–4861.
- Helfman, D.M., Ricci, W.M. and Finn, L.A. (1988) *Genes Dev.*, **2**, 1627–1638.
- Libri, D., Balvay, L. and Fiszman, M.Y. (1991) *Mol. Cell. Biol.*, **11**, 3204–3215.
- Mullen, M.P., Smith, C.W.J., Patton, J.G. and Nadal-Ginard, B. (1991) *Genes Dev.*, **5**, 642–655.
- Noble, J.C.S., Pan, Z.Q., Prives, C. and Manley, J.L. (1987) *Cell*, **50**, 227–236.
- Noble, J.C.S., Prives, C. and Manley, J.L. (1988) *Genes Dev.*, **2**, 1460–1475.
- Grabowski, P.J., Nasim, F.H., Kuo, H.-C. and Burch, R. (1991) *Mol. Cell Biol.*, **11**, 5919–5928.
- Watakabe, A., Tanaka, K. and Shimura, Y. (1993) *Genes Dev.*, **7**, 407–418.
- Xu, R., Teng, J. and Cooper, T.A. (1993) *Mol. Cell. Biol.*, **13**, 3660–3674.
- Yeakley, J.M., Hedjran, F., Morfin, J.-P., Merillat, N., Rosenfeld, M.G. and Emeson, R.B. (1993) *Mol. Cell. Biol.*, **13**, 5699–6011.
- Matus, A. (1988) *Annu. Rev. Neurol.*, **11**, 29–44.
- Kosik, K.S. and Greenberg, S.M. (1994) In Terry, R., Katzman, R. and Bick, K. (eds), *Alzheimer's Disease*. Raven Press, p. 335–344.
- Andreadis, A., Brown, W.M. and Kosik, K.S. (1992) *Biochemistry*, **31**, 10626–10633.

- 28 Goedert,M., Spillantini,M.G., Jakes,R., Rutherford,D. and Crowther,R.A. (1989) *Neuron*, **3**, 519–526.
- 29 Himmler,A., Drechsel,D., Kirschner,M.W. and Martin,D.W. (1989) *Mol. Cell. Biol.*, **9**, 1381–1388.
- 30 Kosik,K.S., Orecchio,L.D., Bakalis,S. and Neve,R.L. (1989) *Neuron*, **2**, 1389–1397.
- 31 Goedert,M., Spillantini,M.G. and Crowther,R.A. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 1983–1987.
- 32 Couchie,D., Mavilia,C.I., Georgieff,S., Liem,R.K.H., Shelanski,M.L. and Nuñez,J. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 4378–4381.
- 33 Andreadis,A., Nisson,P., Kosik,K.S. and Watkins,P. (1993) *Nucleic Acids Res.*, **21**, 2217–2221.
- 34 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 35 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
- 36 Biggin,M.D., Gibson,T.J. and Hong,C.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.
- 37 Devereux,J.R., Haerberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
- 38 Rose,J.K., Buonocore,L. and Whitt,M.A. (1991) *Biotechniques*, **10**, 520–525.
- 39 Innis,M.A., Gelfand,F.H., Sninsky,J.J. and White,T.J. (1990) *PCR Protocols: A Guide to Methods and Applications*. Academic Press.
- 40 Golde,T.E., Estus,S., Usiak,M., Younkin,L.H. and Younkin,S.G. (1990) *Neuron*, **4**, 253–267.
- 41 Pyper,J.M. and Bolen,J. (1990) *Mol. Cell. Biol.*, **10**, 2035–2040.
- 42 Black,D. (1992) *Cell*, **69**, 795–807.
- 43 Huh,G.S. and Hynes,R.O. (1993) *Mol. Cell. Biol.*, **13**, 5301–5314.
- 44 Gallego,M.E. and Nadal-Ginard,B. (1990) *Mol. Cell Biol.*, **10**, 2133–2144.
- 45 Huh,G.S. and Hynes,R.O. (1994) *Genes Dev.*, **8**, 1561–1574.
- 46 Smith,C.W.J., Chu,T.T. and Nadal-Ginard,B. (1993) *Mol. Cell. Biol.*, **13**, 4939–4952.
- 47 Stamm,S., Zhang,M.Q., Marr,T.G. and Helfman,D.M. (1994) *Nucleic Acids Res.*, **22**, 1515–1526.
- 48 Lee,G. (1990) *Cell Motil. Cytoskeleton*, **15**, 199–203.
- 49 Lee,G. and Brandt,R. (1992) *Trends Cell Biol.*, **2**, 286–289.
- 50 Helfman,D.M. and Ricci,W.M. (1989) *Nucleic Acids Res.*, **17**, 5633–5650.
- 51 Guo,W., Mulligan,G.J., Wormsley,S. and Helfman,D.M. (1991) *Genes Dev.*, **5**, 2096–2107.
- 52 Greenwood,J.A., Scott,C.W., Spreen,R.C., Caputo,C.B. and Johnson,G.V.W. (1994) *J. Biol. Chem.*, **269**, 4373–4380.
- 53 Liu,W.K., Dickson,D.W. and Yen,S.H. (1994) *Am. J. Pathol.*, **142**, 387–394.