Nucleic Acids Research, 1993, Vol. 21, No. 9 2217–2221

The exon trapping assay partly discriminates against alternatively spliced exons

Athena Andreadis, Paul E.Nisson[1], Kenneth S.Kosik and Paul C.Watkins[1]
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 and [1]GIBCO-BRL/Life Technologies Inc., Gaithersburg, MD 20898, USA

Received November 13, 1992; Revised and Accepted March 22, 1993

ABSTRACT

A cosmid containing eight exons of the gene coding for the microtubule-associated tau protein was subjected to the exon trapping assay. All the constitutive exons contained in the cosmid (4, 5, 7 and 9) were efficiently captured regardless of size. Of the four alternatively spliced exons, three (3, 4A and 8) were not isolated by the assay, but the behavior of exon 6 depended on the identity of its flanking exons.

INTRODUCTION

The human gene which codes for the neuron-specific microtubule-associated protein tau contains at least sixteen exons, of which six are cassettes (Figure 1A). Exons 4A, 6 and 8 have recently been found in the human genome (1), but not yet in human cDNAs; in rodents the former two are found in tau mRNAs of the peripheral nervous system (2, 3). Exons 2, 3 and 10 are adult-specific in the central nervous system (4–8). Furthermore, the expression patterns of exons 2 and 3 (3 is never found without 2) indicate that these two exons are not equivalent with regard to splicing.

In the present study, we have used a cosmid containing eight exons (3 to 9, Figure 1B) of the human tau gene as a model substrate for exon trapping (9). All the constitutive exons were efficiently captured, but the behavior of the alternatively spliced ones varied. The results suggest that this assay system, although entirely reliable with constitutive exons, may both under- and overestimate the potential cDNA species arising from a gene.

MATERIALS AND METHODS

Cell culture

COS stock cell cultures were maintained in Dulbecco's modified Eagle medium (GIBCO-BRL) supplemented with 10% fetal bovine serum (GIBCO-BRL, heat-inactivated at 56°C for 1 hour).

Plasmid construction

All human tau genomic fragments originated from cosmid χ6, which contains a 35 kbp insert cloned into the BamHI site of pWE15 (Figure 1B; 1). Restriction mapping and partial sequencing showed that the χ6 insert includes tau exons 3 to 9 and the 3′ end of exon 2, truncated at an internal Sau3AI site (the cosmid library was generated by partially digesting human genomic DNA with Sau3AI, pooling the larger fragments and inserting them into the BamHI site of pWE15; R. Tanzi, personal communication). The tau fragments were cloned into eukaryotic expression vectors pSPL1 (9) and pSVIRB (Figure 2A). The latter is a derivative of pSVI2 (10) in which the BamHI site in the second exon and the EcoRI site in the second intron of the rat insulin gene are unique. Chimeric constructs which included exons 2 to 9 (Figure 3A) were produced by standard cloning methodology (11), using the unique BamHI and EcoRI restriction sites in pSPL1 and pSVIRB introns, respectively, as convenient insertion points for the relevant tau genomic fragments.

Transfections and RNA preparation

Plasmid DNA was purified using Qiagen Tip-20's according to the manufacturer's protocol (Diagen Inc.) or by cesium chloride banding, according to Sambrook et al. (11). Plasmids were introduced into COS cells by either the calcium phosphate (11; Promega) or the lipofectase (12; GIBCO-BRL) method. Cells were plated on 100 mm plates and fed two hours prior to transfection. Upon reaching ~40% confluence, each plate was transfected with 10μg of construct DNA. The medium was changed 16 hours after addition of the calcium phosphate/DNA precipitate, without glycerol shock. Total RNA was isolated 48 hours post-transfection by the RNAzol method (13) according to the manufacturer's instructions (Cinna/Biotecx Laboratories, Houston, TX).

Reverse transcription and polymerase chain reactions

For PCR analysis of RNA, 3μg of total RNA from transiently transfected COS cells were reverse transcribed using either random hexamers or oligomer SA4 (Table I), which is specific for pSPL1, and 200 units of reverse transcriptase (RNAase H$^-$ Superscript, GIBCO-BRL), in a total volume of 20μl for either 30 min or 1 hour at 42°C. The reaction was then treated with RNAase H (GIBCO-BRL) for 10 min at 55°C. Part of this reaction mix (5μl) was then diluted to a final volume of 50μl, the concentrations of the buffer and dNTPs were adjusted for

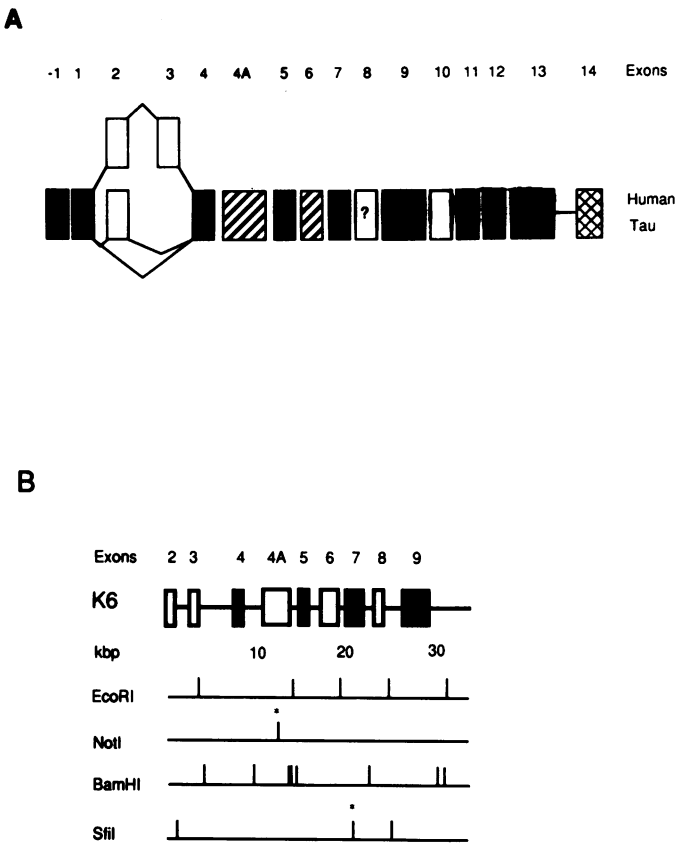

Figure 1.(A) Organization and splicing of exons in the human tau gene. Constitutive exons are black, developmentally regulated cassette exons white, tissue-specific cassette exons striped, the alternative C-terminus crosshatched. The three possible species in the 2/3 region are diagrammed. The question mark in exon 8 indicates that its regulation type is unknown. **(B)** Organization of cosmid clone κ6. Sites of four restriction enzymes are shown below the gene map. The restriction sites with asterisks also appear in the cDNA. The relative positions of the exons are correct, but exon size is greatly exaggerated with respect to intron size.

PCR and the mixture was amplified (30 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, extension at 72°C for 2 min for pSVIRB constructs; 35 cycles of denaturation at 94°C for 45 sec, annealing at 58°C for 30 sec, extension at 72°C for 1 sec for pSPL1 constructs). The primers used are listed in Table I and shown in Figure 2A. The pSPL1 and pSVIRB DNAs were used as controls (Figure 3B). The reverse transcribed, PCR-amplified (RT-PCR) product of any construct which was larger in size from that of spliced pSPL1 or pSVIRB was cloned and sequenced.

Cloning of RT-PCR products

The fragments generated by RT-PCR were cloned by two methods: 1) The PCR products from pSVIRB were denatured for 2 min at 90°C in 'forward heat' buffer (11), phosphorylated in the presence of kinase 'forward' buffer, 100mM ATP and 10 units of T4 polynucleotide kinase (New England Biolabs), run on a low melt agarose gel, extracted from the agarose by treatment with β-agarase (New England Biolabs) and blunt-end ligated into SmaI-digested dephosphorylated pKS$^+$-Bluescribe (Stratagene); 2) The RT-PCR products from pSPL1 were subjected to another PCR amplification with nested primers (SD1 and SA1, Figure 2A). These secondary PCR primers contained

Table I. Primers used in PCR (Indicated in Figure 2A)

Name	Length	Orientation	Location
SVP	30	Sense	SV40 early promoter
INS3	30	Antisense	Rat insulin exon 3
SD2	25	Sense	Upstream β-globin exon
SD1	22+12*	Sense	Upstream HIV tat exon
SA1	21+12*	Antisense	Downstream HIV tat exon
SA4	22	Antisense	Downstream β-globin exon

*The additional 12 nucleotides are the 4 CUA repeats used for UDG cloning.

4 repeats of the trimer CUA, U being deoxyuracil (14; Figure 2B). This base is cleaved off by uracil DNA glycosylase, thus producing 3' overhangs that can anneal to vector pUC19 which has been engineered to contain complementary overhangs (14; GIBCO-BRL).

Sequencing

Plasmid DNA was prepared by either alkaline lysis (11) or the rapid boiling method of Holmes and Quigley (15), denatured according to Lim & Péne (16) and ethanol-precipitated, then sequenced by the dideoxynucleotide method of Sanger et al. (17) using a modified T7 DNA polymerase (18) and [^{35}S]-α-dATP (Amersham SJ.1304) according to the modifications of Biggin et al. (19). Alternatively, the PCR products were sequenced by the dsDNA cycle sequencing system (20; GIBCO-BRL). The sequences were resolved on a denaturing (8.3 M urea) 6% (w/v) polyacrylamide gel and analyzed using the University of Wisconsin software package (21).

RESULTS

The constitutive tau exons are efficiently captured but three of the cassette tau exons are not amplified

Human tau exons 3 to 9 were inserted into expression vectors and transiently transfected into COS cells. The sizes of the inserts, exons and flanking introns are shown in Figure 3A.

The RT-PCR results show that all the constitutive tau exons (4, 5, 7, 9) contained in the cosmid were present in the mRNA of the expression constructs (Figure 3, constructs 2, 4, and 6). Furthermore, sequencing of the RT-PCR products shows that these exons were spliced correctly; no cryptic 5' or 3' splice sites were utilized.

The insert containing tau exons 8 and 9 was placed in vector pSPL1 without checking for orientation. In this case, there were two PCR products (Figure 3B). The major one contained exon 9; sequencing of the minor one (Figure 3B, band X) showed that it contained a 162 nucleotide insert from the nonsense strand downstream of exon 9, fortuitously flanked by sequences in good agreement with the splicing consensus cag/G and CAG/gtaagt (22; Figure 4B).

In contrast to the constitutive exons, the cassette exons 3, 4A and 8 were absent from the mRNA of the expression constructs (Figure 3, constructs 1, 3 and 6). In the case of exon 3, the genomic insert in the expression construct also carried the 3' half of exon 2; the exon 2 fragment did not inhibit splicing nor produce aberrant products. In the case of exon 4A, insertion of the relevant fragment into pSPL1 in either orientation gave no spliced product besides the vector (data not shown). In the case of exon 8, the genomic insert used for the expression construct also carried exon 9, which was captured.

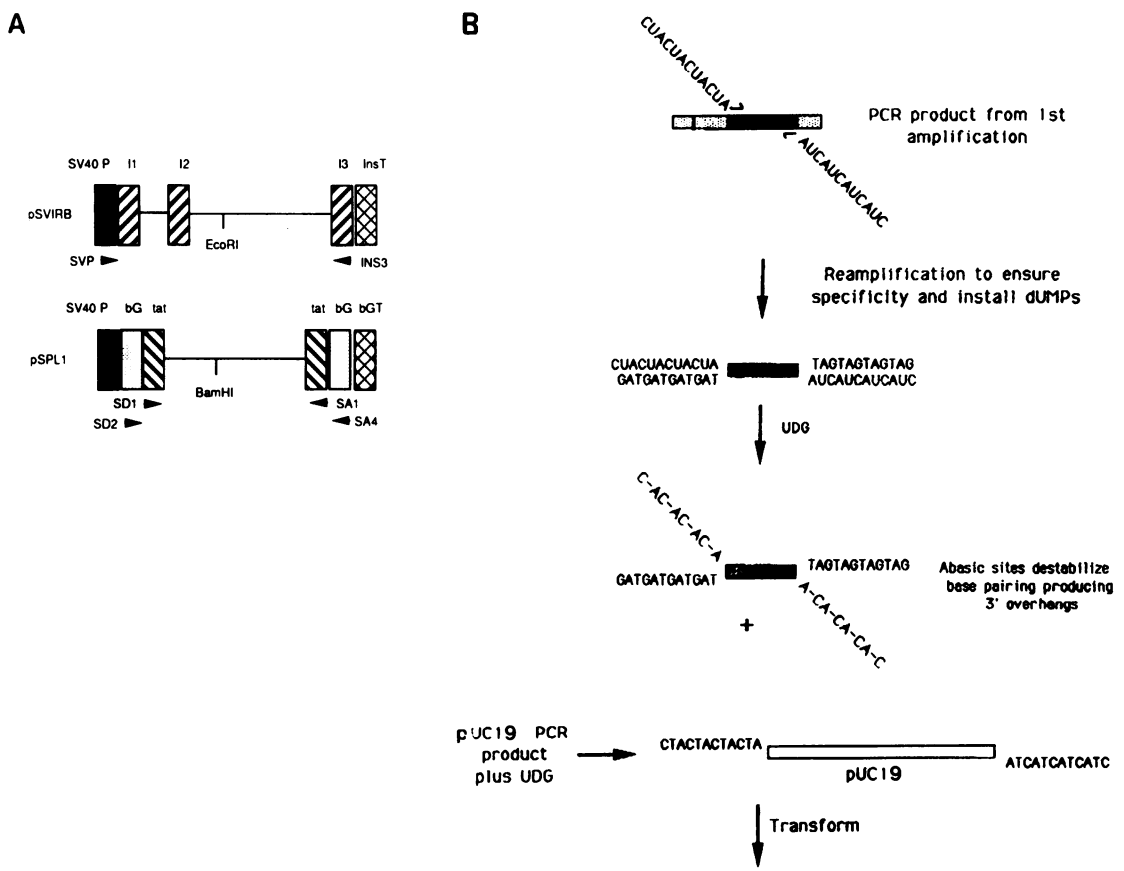

Figure 2. (A) Schematic diagram of expression vectors pSPL1 and pSVIRB. The SV40 promoter/enhancer element is black, β-globin exons are dotted, insulin and HIV tat exons striped, insulin and globin terminators crosshatched. The two unique restriction sites utilized and the primers used for RT-PCR are indicated. (B) Outline of the UDG method for cloning PCR fragments.

The behavior of exon 6 depends on the identity of the flanking exons

Expression constructs containing cassette exon 6 were made with either homologous or heterologous (insulin) flanking exons. Interestingly enough, exon 6 was excluded from the construct mRNA from COS cells if flanked by exons 5 and 7, but was included if flanked by insulin exons 2 and 3 (Figure 3, constructs 4 and 5, respectively). Furthermore, when exon 6 was flanked by heterologous exons (construct 5), it gave rise to three different splicing products, arising from uniform utilization of the correct 5′ splice site coupled to alternative 3′ splice sites (Figure 3B and 4A, bands A, B and C). In fact, the major PCR product corresponded not to the authentic splice site of exon 6 (Figure 3B and 4B, band A) but to a site within the exon (Figure 3B and 4A, band B) slightly closer to the 3′ splice consensus of cag/G (22). However, other sites within exon 6 equally similar to the consensus were ignored.

There is clearly a dichotomy between exons 4A and 6, because in the 4A construct the tau exon was flanked by heterologous exons and was completely excluded (data not shown). On the other hand, constitutive exon 4 was captured even though it was flanked by heterologous exons (Figure 3, construct 2).

DISCUSSION

In this study, a cosmid containing both constitutive and alternative (cassette) exons of the human tau gene was used as a model substrate for exon trapping (9). The results from the expression constructs indicate that all the constitutive tau exons on the cosmid are amplified by this assay regardless of size: the constitutive exons range from 56 (exon 5) to 266 nucleotides (exon 9). Also, it does not seem to matter whether one or more constitutive exons are within a genomic fragment (for example, exons 5 and 7 are trapped as a unit) nor whether the constitutive exons are flanked by heterologous ones (for example, exon 4).

The situation is very different for alternative exons. In the case of tau, all the alternatively spliced exons on χ6 are cassettes and under different phenotypic regulation: 2 and 3 are small exons with large introns (1, 6) which are developmentally regulated in the central nervous system (4–8). Furthermore, exon 3 is never incorporated without exon 2, making it an ambiguous entity between a true cassette and a member of a coordinated exon pair. Conversely, 4A and 6 are large exons with small introns (1, 6) which until now have only been found in the peripheral nervous system or cell lines derived therefrom (2, 3). The phenotypic regulation of exon 8 is unclear; it has been found in bovine brain, but not yet in tau cDNAs of either the human central (1) or rodent peripheral (2, 3) nervous system.

Three of the four alternative tau exons (3, 4A and 8) are not captured by the trapping assay, regardless of size, length of flanking introns or phenotypic regulation. Their exclusion may be due to different reasons. Exon 3, given its behavior in the normal context of the tau gene (it never appears without 2; 4–8), may require pre-splicing to exon 2 for activation. The extremely large exon 4A (753 nucleotides) exceeds the length postulated by the tenets of exon definition (23). The exclusion of exon 8

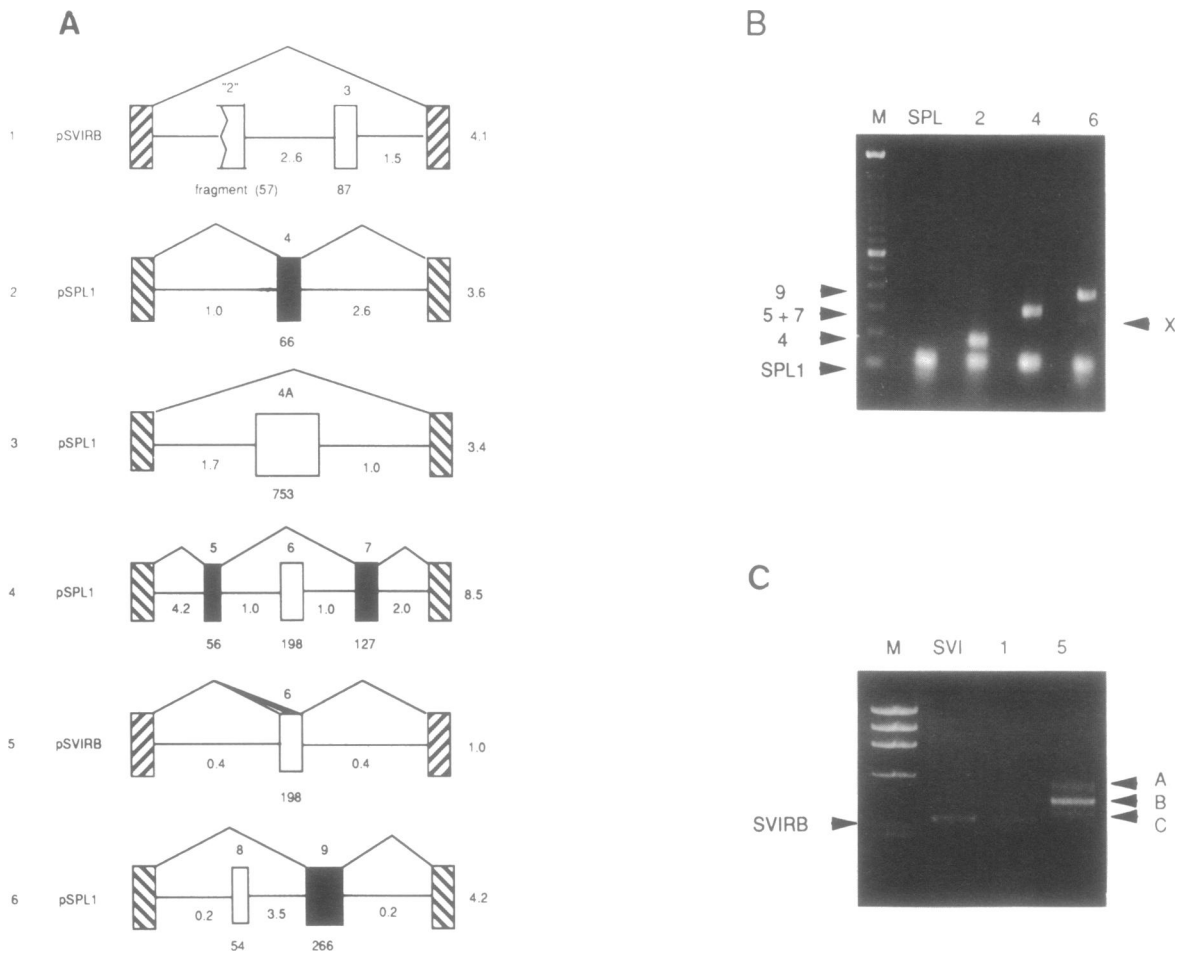

Figure 3. **(A)** Summary of expression constructs and their behavior in COS cells. Constitutive tau exons are black, tau cassette exons white, vector exons striped. For the sake of clarity, only the exons immediately flanking the tau insert are shown. The expression vector used is indicated on the left and the size of the tau genomic insert (in kbp) on the right side of each construct; the size of the tau exons (in nucleotides) and introns (in kbp) is shown below each. **(B)** and **(C)** RT-PCR of the expression constructs. pSPL1 and pSVIRB are included as controls. Markers are the 100 bp ladder (panel B, GIBCO-BRL) and ϕX HaeIII (panel C, New England Biolabs). The exact sequence of bands A, B and C from construct 5 and X from construct 6 are shown in Figure 4.

A

Exon 6

```
                2/4
                  T  S  T  R  S  S  A  K  T  L  K  N  R  P  C  L  S  P  K  L
1 atgtttatgtttaagACATCCACACGTTCCTCTGCTAAAACCTTGAAAAATAGGCCTTGCCTTAGCCCCAAACTC
                  A
                                          3/4
  P  T  P  G  S  S  D  P  L  I  Q  P  S  S  P  A  V  C  P  E  P  P  S  S  P
  CCCACTCCTGGTAGCTCAGACCCTCTGATCCAACCCTCCAGCCCTGCTGTGTGCCCAGAGCCACCTTCCTCTCCT
                                             B
                                3/4
  K  H  V  S  S  V  T  S  R  T  G  S  S  G  A  K  E  M  K  L  K
151 AAACACGTCTCTTCTGTCACTTCCCGAACTGGCAGTTCTGGAGCAAAGGAGATGAAACTCAAGgtaaggaaactc
                                   C
                                                                7/9
```

B

Captured sequence from the nonsense strand

```
                                3/4
1 aggccaggcccccgaagtctgtgcttctccacctttagGAAGCTGGAGAGCTCGCGAGCACCCACTTGCTCCCAG

  AAGAATCAGCTGATCAAAACAGAACCCCCAATCTCCCCAGAATCCAGAGGAACCCAACCTAAGCAGATGGCCCAC

151 GAGTGGAGATCTGACCTATGGGTGGCAGTGTATTCTGCCCCAGAAGGCAGgtggtggggaaccgctgtgggtgcc

                                                                5/9
```

Figure 4. Sequence of variant splicing products. The exon and pseudoexon are designated in capital letters. The numbers indicate agreement with the 3′ and 5′ splice site consensus sequences (22). **(A)** Human tau exon 6, bands A, B, C. **(B)** Nonsense strand downstream of human tau exon 9, band X.

may be dictated by the extremely poor agreement of its splice sites with the consensus (1, 22).

For exon 6, the identity of the flanking exons appears crucial. The homologous tau exons suppress it, whereas heterologous exons allow its inclusion. The involvement of the flanking exons in splicing selection seen in the case of exon 6 has been documented for the cardiac troponin T cassette miniexon (24), as well as for the mutually exclusive exons of myosin light chain 1/3 (25). Regardless of the mechanisms operating on selection of exon 6, what is important for the validity of the exon trapping assay is the fact that, when exon 6 is captured, it gives multiple products.

The behavior of exon 6 and the capture of a random sequence from the nonsense strand near exon 9 indicate that context is as important in the trapping assay as in splicing itself. The splice sites of the alternative tau exons conform less well to the consensus (22) than those of the constitutive ones (1) and almost certainly are involved in regulation. It is also interesting that the sequences which give rise to products B, C and X have neither a long polypyrimidine tract nor a likely branch point (26, 27).

The sensitivity of the trapping assay (it contains two amplification steps, the intracellular increase in plasmid copy number caused by the T antigen and the PCR reaction) can clearly lead to artifacts. Insertion of certain genomic fragments may give false positives: Pseudoexons such as that within the intron between exons 8 and 9 seem to conform to the size and boundary requirements for real exons (Figure 4B). One reading frame of this captured sequence even gives a continuous peptide, though not one corresponding to any known protein within the non-redundant combined peptide database (28). Thus, one of the advantages ascribed to the assay—namely, the retrieval of exons without need of cDNA isolation (9)—needs to be qualified. Furthermore, use of this assay may lead to both under- and overestimation of possible mRNA species: retrieval of differing products need not indicate alternative splicing *per se*; exon 6 has always been isolated as an entire unit (2, 3, 6), so the band B and C products (Figure 3B) would lead to the conclusion that it actually has multiple 3′ splice sites. On the other hand, isolation of variant products may yield some information about the regulation of alternative exons.

Taken together, the results from this study suggest that this particular method of exon identification will generally not unmask all alternatively spliced exons. It will give the skeleton outline of a gene, i.e. the constitutive exons; it will also capture certain subclasses of alternative exons.

ACKNOWLEDGEMENTS

We wish to thank Rudi Tanzi for providing the human cosmid library, Alice Young and Robert Blakesley for help with the cycle sequencing, Catherine Linder for contributing to the restriction mapping, Albarosa Alicea for help with the minigene constructs and Margaret Erhardt for synthesizing oligonucleotides SVP and INS3. This work was supported by Alzheimer's Faculty Scholar Award FSA-90-005 to A.A. and NIH grant AG06601 to K.S.K.

REFERENCES

1. Andreadis, A., Brown, W. M. and Kosik, K. S. (1992) Biochemistry **31**, 10626–10633.
2. Couchie, D., Mavilia, C., Georgieff, I. S., Liem, R. K. H., Shelanski, M. L. and Nuñez, J. (1992) Proc. Natl. Acad. Sci. USA **89**, 4378–4381.
3. Goedert, M., Spillantini, M. G. and Crowther, R. A. (1992) Proc. Natl. Acad. Sci. USA **89**, 1983–1987.
4. Goedert, M., Spillantini, M. G., Jakes, R., Rutherford, D. and Crowther, R. A. (1989) Neuron **3**, 519–526.
5. Goedert, M., Spillantini, M. G., Potier, M. C., Ulrich, J. and Crowther, R. A. (1989) EMBO J. **8**, 393–399.
6. Himmler, A., Drechsel, D., Kirschner, M. W. and Martin, D. W. (1989) Mol. Cell. Biol. **9**, 1381–1388.
7. Himmler, A. (1989) Mol. Cell. Biol. **9**, 1389–1396.
8. Kosik, K. S., Orecchio, L. D., Bakalis, S. and Neve, R. L. (1989) Neuron **2**, 1389–1397.
9. Buckler, A. J., Chang, D. D., Graw, S. L., Brook D. A., Haber, D. A., Sharp, P. A. and Housman, D. E. (1991) Proc. Natl. Acad. Sci. USA **88**, 4005–4009.
10. LoMedico, P. T. (1982) Proc. Natl. Acad. Sci. USA **79**, 5798–5802.
11. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Laboratory Press.
12. Whitt, M. A., Buonocore, L., Rose, J. K., Ciccarone, V., Chytil, A. and Gebeyehn, G. (1991) Bethesda Res. Lab. Focus **13**, 8–12.
13. Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**, 156–159.
14. Nisson, P. E., Rashtchian, A. and Watkins, P. C. (1991) PCR Methods and Applications **1**, 120–123.
15. Holmes, D. S. and Quigley, M. (1981) Anal. Biochem. **114**, 193–197.
16. Lim, H. M. and Péne, J. J. (1988) Gene Anal. Techn. **5**, 32–39.
17. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
18. Tabor, S. and Richardson, C. C. (1989) J. Biol. Chem. **264**, 6447–6458.
19. Biggin, M. D., Gibson, T. J. and Hong, C. F. (1983) Proc. Natl. Acad. Sci. USA **80**, 3963–3965.
20. Craxton, M. (1991) Methods: A Companion to Methods in Enzymology **3(1)**, 20–26.
21. Devereux, J. R., Haeberli, P. and Smithies, O. (1984) Nucl. Acids Res. **12**, 387–395.
22. Mount, S. M. (1982) Nucl. Acids Res. **10**, 459–472.
23. Robberson, B. L., Cote, G. J. and Berget, S. M. (1990) Mol. Cell Biol. **10**, 84–94.
24. Cooper, T. A. and Ordahl, C. P. (1989) Nucl. Acids Res. **17**, 7905–7921.
25. Gallego, M. E. and Nadal-Ginard, B. (1990) Mol. Cell Biol. **10**, 2133–2144.
26. Reed, R. (1989) Genes Dev. **3**, 2113–2123.
27. Mullen, M. P., Smith C. W. J., Patton J. G. and Nadal-Ginard, B. (1991) Genes Dev. **5**, 642–655.
28. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). J. Mol. Biol. **215**, 403–410.