Structure and Novel Exons of the Human τ Gene[†,‡]

Athena Andreadis,* William M. Brown, 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, 221 Longwood Avenue, Boston, Massachusetts 02115

Received May 6, 1992; Revised Manuscript Received July 23, 1992

ABSTRACT: The microtubule-binding protein τ is important in establishing and maintaining neuronal morphology and is a major component of the neurofibrillary tangles (NFTs) characteristic of Alzheimer's brain. The neuron-specific τ transcript undergoes complex alternative splicing. The human τ gene has been cloned and mapped. The restriction analysis and partial sequencing of the gene shows that it contains (1) four alternatively spliced exons previously described in rodent and bovine but not in human τ cDNAs and (2) two CpG islands, one associated with the promoter region, the other with exon 9. Examination of human τ mRNA indicates that the human cerebrocortical splicing pattern differs from that previously reported for the murine and bovine τ mRNAs, despite conserved exon organization in all three genes.

Microtubule-associated proteins (MAPs) are a disparate group of proteins capable of regulating the microtubule polymer state as well as interacting with other cytoskeletal and subcellular components [reviewed in Olmsted (1986)]. τ is a MAP localized to the axonal segment of peripheral and central neurons (Binder et al., 1986; Kowall & Kosik, 1987; Peng et al., 1986) which can organize microtubules into bundles when introduced into nonneural cells (Drubin et al., 1986; Kanai et al., 1989; Lewis et al., 1989). Aberrant forms of τ protein are also found within neurofibrillary tangles (NFTs), characteristic pathological structures in Alzheimer's brain (Delacourte & D'efossez, 1986; Grundke-Iqbal et al., 1986; Kosik et al., 1986; Nukina et al., 1986; Wood et al., 1986).

The τ gene has been reported to be present as a single-copy gene in the human (Neve et al., 1986), rat (Drubin et al., 1984), and cow (Himmler, 1989) genome. The human gene has been localized to chromosome 17 by in situ hybridization (Neve et al., 1986). The τ primary transcript undergoes complex, regulated alternative splicing, giving rise to several mRNA species that migrate on Northern blots at approximately 6 and 9 kb (Drubin et al., 1986, 1988; Himmler et al., 1989). The resulting protein isoforms are a series of closely spaced bands from 58 to 66 kDa and in some neural tissues a 110-kDa isoform (Cleveland et al., 1977; Drubin et al., 1986, 1988; Oblinger et al., 1991). Most of the alternatively spliced τ exons are cassettes (Figure 1), but in the 3′ end of the gene, the length of the τ reading frame depends on differential processing of the last intron (Goedert et al., 1989a,b, 1992; Himmler, 1989; Himmler et al., 1989; Kosik et al., 1989; Lee et al., 1988).

A set of overlapping genomic clones that cover the entire human τ gene and span over 100 kbp have been characterized in this study. Restriction analysis of the clones indicates that the gene contains two CpG islands, one associated with the 5′ end of the gene, the other with exon 9. The gene also contains homologs to four exons (4A, 6, 8, and 14) previously reported in bovine and rodent τ cDNAs (Couchie et al., 1992; Goedert et al., 1992; Himmler, 1989; Lee et al., 1988), but not yet described in human. No mRNAs containing these exons can be detected in human τ cerebrocortical mRNA, although these exons are flanked by splice sites resembling the consensus sequences.

MATERIALS AND METHODS

Libraries. The human genomic library in the cosmid vector pWE15 was a gift from Dr. R. Tanzi (Massachusetts General Hospital, Boston, MA). Two human genomic libraries in the λ vector EMBL3 were gifts from Dr. S. Karathanasis (Lederle Laboratories, New York, NY) and Dr. J. Ladias (Deaconess Hospital, Boston, MA). The human fetal brain cDNA library was a gift from Dr. R. Neve (McLean Hospital, Boston, MA), and the adult human retina library was purchased from Clontech (Palo Alto, CA).

Probe Synthesis and Purification. Probes for screening the libraries and the Southern blots were prepared from template DNAs by the random-primer technique (Feinberg & Vogelstein, 1983). A commercial random-primer kit (Boehringer Mannheim) was used with [^{32}P]-α-dCTP (Amersham PB.10205, specific activity >3000 Ci/mmol). Probes were purified from unincorporated nucleotides using NucTrap columns (Stratagene) according to the manufacturer's protocol. The purified probe typically had a specific activity of $\sim 7 \times 10^8$ cpm/μg. Oligonucleotides for PCR and sequencing were synthesized on an Applied Systems 380B oligonucleotide synthesizer located in the Center for Neurologic Diseases.

Probes Used. The probe initially used was human fetal τ cDNA clone p19. This clone contains exons 1–14 up to an *Eco*RI site in exon 14, but lacks exons 2, 3, and 10 (Lee et al., 1989). The exons are numbered according to the bovine τ gene notation (Himmler et al., 1989). Subsequently, restriction fragments derived from the 5′ and 3′ ends of the genomic clones obtained were used to walk along the gene in both directions. These are indicated in Figure 1.

Isolation of Genomic Clones. λ and cosmid libraries were plated according to standard techniques (Sambrook et al., 1989). Duplicate lifts were taken from the plated libraries on nitrocellulose membranes (Schleicher & Schuell, BA85), which were treated according to Sambrook et al. (1989) before being fixed by baking for 2 h at 80 °C in a vacuum oven. The membranes were prewetted in 2 × SSC (1 × SSC is 150 mM NaCl/15 mM sodium citrate), 0.1% (w/v) sodium dodecyl sulfate (SDS) before being prehybridized for at least 3 h at 65 °C in 6 × SSC/5 × Denhardt's solution [1 × Denhardt's

† This work was supported by Alzheimer's Faculty Scholar Award FSA-90-005 to A.A. and NIH Grant AG06601 to K.S.K.

‡ The novel nucleic acid sequences in this paper have been submitted to GenBank under Accession Nos. M93652, X61371, X61375, and X61373.

* Corresponding author.

0006-2960/92/0431-10626$03.00/0

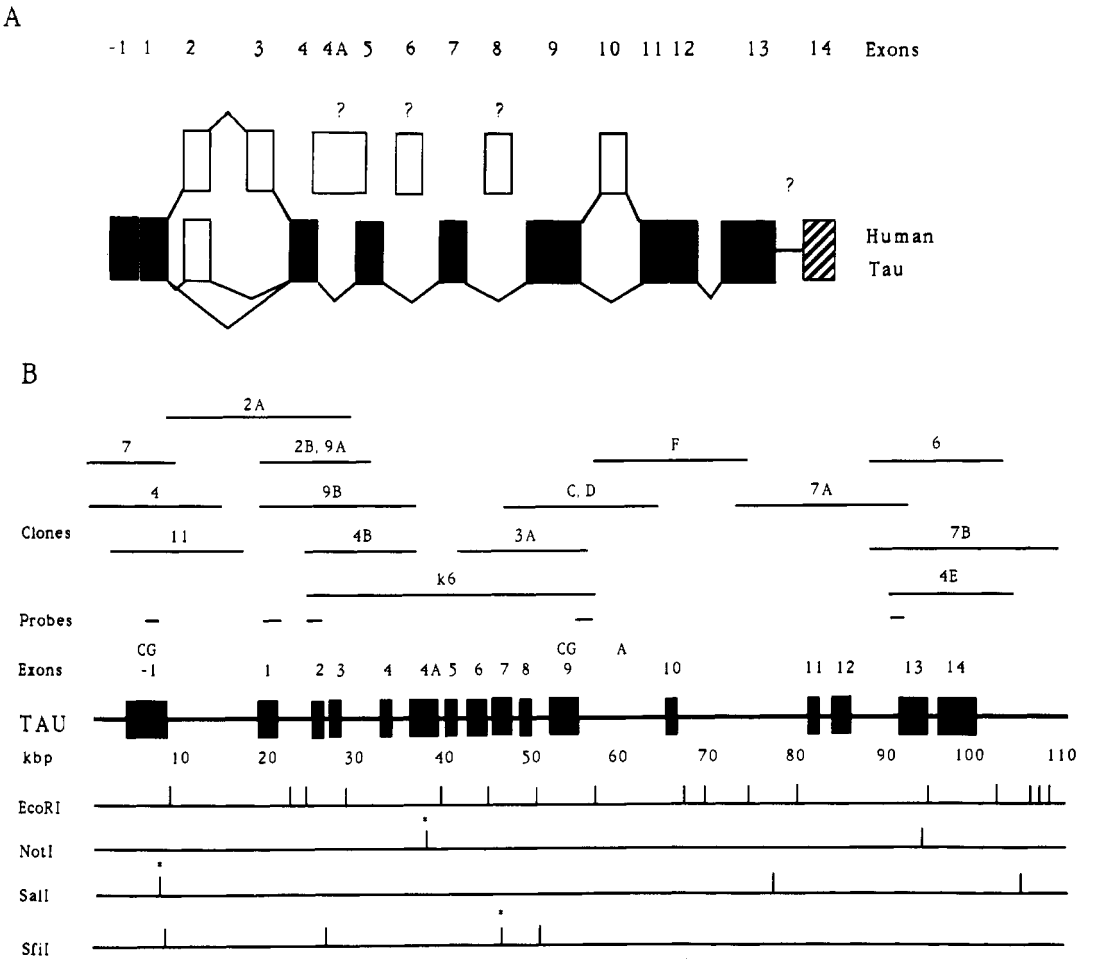

FIGURE 1: (A) Organization and splicing of τ exons in the human gene. Constitutive exons are black, cassette exons are white, and the alternative reading frame in the carboxy-terminus is striped. (B) Genomic organization of the human τ gene. The phage and cosmid clones are shown above the gene map, sites of four restriction enzymes below. The restriction sites with asterisks also appear in the cDNA. CG indicates the position of CpG islands (shown in detail in Figure 2). A indicates the position of an Alu element (discussed in the text). The relative positions of the exons are correct, but exon size is greatly exaggerated with respect to intron size.

solution is 0.02% (w/v) bovine serum albumin, 0.02% (w/v) poly(vinylpyrrolidone), 0.02% (w/v) Ficoll-400)], 0.1% (w/v) SDS, and 100 μg/mL heat-denatured herring sperm DNA. Hybridization was carried out overnight (>17 h) at 65 °C in the same solution to which the heat-denatured probes (~5 × 10^6 cpm/mL) were added. Following overnight hybridization, the membranes were washed for 45 min at 65 °C in 0.1 × SSC/0.1% (w/v) SDS with three changes of buffer. Wet filters were wrapped in Saran Wrap and exposed against Kodak X-Omat film at −70 °C in cassettes with intensifying screens. Positive clones were picked, replated at reduced densities, and rescreened using the same conditions until single isolates were recovered. The cDNA libraries were screened in a similar manner.

DNA Purification, Restriction Mapping, and Southern Blot Analysis. λ DNA was purified from positive phage clones using the small-scale liquid lysate technique (Sambrook et al., 1989) and Qiagen Tip-20 columns according to the manufacturer's protocol (Diagen Inc.). Cosmid DNA was purified by the alkaline lysis method (Sambrook et al., 1989). Restriction endonucleases were purchased from New England Biolabs. All unique λ clones were mapped with *Sal*I and *Eco*RI, and the unique cosmid clone was mapped with *Not*I and *Eco*RI; in both cases the restriction enzymes used are absent from the respective vectors and define the vector/insert boundaries. In addition, to locate any CpG islands associated with the human τ gene, the genomic clones and various subclones derived from them were digested with the restriction endonucleases *Not*I, *Sfi*I, *Bss*HII, *Eag*I, *Sac*II, and *Sma*I, all of which are "rare cutters" that recognize sites comprising entirely G/C palindromes. Selected digests were separated by agarose gel electrophoresis and capillary-transferred to Hybond-N nylon membranes (Amersham). The Southern blots were prehybridized, hybridized, and washed as above, using random-primed probes derived from the p19 fetal cDNA, genomic sequences, or specific τ fragments generated by PCR.

DNA Sequencing and Analysis. All restriction fragments that gave positive signals on Southern blot analysis were subcloned into the cloning vector pKS$^+$ Bluescribe (Stratagene). To locate exons quickly within subcloned genomic fragments, where possible, the inserts were digested with restriction enzymes known to cut the exon and the pKS polylinker and intramolecular ligations were then performed. This procedure placed the standard sequencing primers adjacent to exon sequence and greatly facilitated identification. Plasmid DNA was purified using Qiagen Tip-20's or by cesium chloride banding, according to Sambrook et al. (1989). The purified plasmid DNA was denatured according to Lim and Péne (1988) and ethanol-precipitated. Alternatively, single-stranded phagemid was produced by the modified Stratagene technique (Katayama, 1990). Single-stranded or denatured double-stranded templates were sequenced by the dideoxynucleotide method of Sanger et al. (1977) using a modified T7 DNA polymerase (Tabor & Richardson, 1989) and [^{35}S]-α-dATP (Amersham SJ.1304, specific activity >1000 Ci/

Table I

A. Primers Used in PCR

name	length	orientation	location
HTS4	14	sense	beginning of exon 4
HT2	30	antisense	within exon 7
HT3	24	sense	within exon 7
648H	24	antisense	within exon 9
HT841[a]	25	sense	beginning of exon 13
HT14[a]	30	antisense	region homologous to bovine exon 14

B. Expected PCR Products

exon	primer pair exon location	RNA species	size
6	HTS4/HT2	6^-	154
	4/7	6^+	352
8	HT3/648H	8^-	328
	7/9	8^+	382
14	HT841/HT14	14^-	1454
	13/14	14^+ murine	500

[a] HT841 and HT14 are indicated in Figure 3D.

mmol) according to the modifications of Biggin et al. (1983). Alternatively the "fmol" kit from Promega was used to directly sequence purified λ DNA with sequence-specific oligonucleotide primers end-labeled with [^{32}P]-γ-ATP (Amersham PB.10218, specific activity >6000 Ci/mmol). The products of the reactions were separated on a denaturing (8.3 M urea) 6% (w/v) polyacrylamide gel. Sequences so obtained were compiled and analyzed using the University of Wisconsin software package (Devereux et al., 1984).

PCR Analysis of τ Transcripts. Total RNA from human fetal brain, adult cerebrum, and LA-N-5 human neuroblastoma cells was isolated by the RNAzol method [Biotecx; Chornczynski and Sacchi (1987)]. The primers used for PCR are listed in Table IA, and two are indicated on the sequence in Figure 3D. For PCR analysis of cDNA libraries (Innis et al., 1990), 1 μL of the human fetal brain and adult retina cDNA libraries (both in λgt11, with titers of 10^{11} pfu/mL) was amplified for 25 cycles (denaturation at 94 °C for 1 min, annealing at 55 °C for 2 min, extension at 72 °C for 10 min). For PCR analysis of RNA, 5 μg of total RNA from LA-N-5 cells or human cerebrum (fetal or adult) was reverse-transcribed using random hexamer primers and 200 units of reverse transcriptase (RNAase H$^-$ Superscript, BRL), in a total volume of 20 μL for 2 h at 42 °C. Part of this reaction mix (7 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 35 cycles (denaturation at 94 °C for 1 min, annealing at 55 °C for 2 min, extension at 72 °C for 5 or 10 min). The fetal human τ cDNA p19 mentioned above was used as a positive control (Figure 5).

RESULTS

Organization of the Human τ Gene. Isolation and characterization of the human τ genomic clones indicated that the gene extends for over 100 kbp and that it includes at least 15 exons which range in size from 54 to 753 nucleotides (Figure 1). The exon termed −1 is transcribed but not translated. All the human τ exons and their flanking regions were sequenced. The 5′ and 3′ splice site sequences for all the human τ exons, as well as their lengths, are listed in Table II. All intron/exon boundaries conform to the consensus (Mount, 1982). Examination of the clones for G/C-rich regions by employment of restriction enzymes that recognize sites composed entirely of G and C (*Bss*HII, *Eag*I, *Not*I, *Sac*II, *Sfi*I, *Sma*I) showed that two CpG islands are present in the gene, associated with exons −1 and 9 (Figure 2). Limited sequencing of these regions indicates that their observed/expected ratio of the CpG dinucleotide is 77% and 42%, respectively, much higher than the average of 20–25% (Gardiner-Garden & Frommer, 1987).

Exons 4A, 6, 8, and 14. Analyses of the nucleotide sequence employing the University of Wisconsin software package (Devereux et al., 1984) showed that the human τ gene contains sequences highly homologous to bovine exons 6 and 8 and to murine exons 4A, 6, and 14 [(Figures 3 and 4C) Couchie et al., 1992; Goedert et al., 1992; Himmler, 1989; Lee et al., 1988]. The human exons 6 and 8 are flanked by consensus splice site sequences (Mount, 1982), have strong branch point/polypyrimidine tract regions (Reed & Maniatis, 1988), and retain the appropriate reading frame upon readthrough (Figure 3B,C). The human exon 4A is also flanked by consensus splice site sequences, but there is no obvious branch point or polypyrimidine tract upstream (Figure 3A). Since the two rodent sequences of exon 4A are of different length, the 3′ splice site for the human exon 4A was chosen to retain the reading frame and maximize amino acid homology with its rodent counterparts (Couchie et al., 1992; Goedert et al., 1992).

However, there are differences between species; the human exon 4A is longer than its N115 counterpart by 42 nucleotides and shorter than the PC12 exon 4A by 9 nucleotides (Figure 4A). The insertions/deletions that give rise to the length variation are interspersed throughout the exon, with the least degree of conservation at the 5′ end of the exon. The human exon 6 has a C inserted at nucleotide 125 (reading ACTG; Figure 3B) where its bovine equivalent has an ATG codon which was proposed to serve as an alternative translation start site (Himmler et al., 1989). There are two other methionine codons within the upstream intron in frame with exon 6, until a stop codon is reached 81 nucleotides from the exon 6 3′ splice site. The homology between bovine and human exons 6 extends into the upstream intron of the gene, but stops 177 nucleotides from the 3′ splice site (Figure 4C).

The region in the human sequence homologous to the murine extended carboxy-terminus contains a consensus 3′ splice site at exactly the same position as its murine equivalent; however, if paired to the human 5′ splice site within exon 13, this would result in a short extension (three additional amino acids) before a stop codon is reached. There is a −1 shift from the reading frame that would result in a peptide sequence similar to the murine extended carboxy-terminus and ending at the same codon (Lee et al., 1988) due to a G missing before nucleotide 1190 of the human exon 14 (Figure 3D). However, if a different donor is used (for example, the GT at position 261), that would generate the murine extended reading frame. The region homologous to the bovine exon 14 has what could be considered a 3′ splice site, but displaced 9 nucleotides upstream with respect to its bovine equivalent (nucleotide 1408 in human exon 14; Figure 3D). Proceeding up to 40 nucleotides upstream from either of the putative 3′ splice sites, there is no contiguous stretch of polypyrimidines longer than 8 nucleotides, nor is there a sequence strongly similar to the branch point consensus YNYTRAY [N = any nucleotide, R = purine, Y = pyrimidine; Reed and Maniatis (1988)].

Two human cDNA libraries, fetal brain and adult retina, were screened with genomic probes from the regions corresponding to exons 4A, 6, and 8. Neither library gave any positive clones, although retina contains a high molecular weight τ protein (Kosik and Nixon, unpublished observations). PCR amplification of human fetal or adult mRNA (Innis et al., 1990) from cerebrum or LA-N-5 cells (following first-strand DNA synthesis by reverse transcriptase) gave no visible

Table II: Intron/Exon Junctions of the Human τ Gene

exon	a	3′ ss cagG	exon (bp)	5′ ss CAG gtaagt	a	intron (kbp)
-1		N/A	(>130)	ACTATCAG gtaagcgccg Untranslated	(8/9)	6.5
1	(4/4)	ctttccccag GTGAACTT Untranslated	(150)	CTG AAA G gttagtggac Leu Lys G	(7/9)	5
2	(3/4)	tgtgttccag AA TCT CCC lu Ser Pro	(87)	GCG GAA G gtgggccccc Ala Glu A	(5/9)	2.6
3	(2/4)	tggtttctag AT GTG ACA sp Val Thr	(87)	ACC ACA G gtgagggtaa Thr Thr A	(7/9)	4.2
4	(3/4)	catacaccag CT GAA GAA la Glu Glu	(66)	ACC CAA G gtcagtgaac Thr Gln G	(7/9)	10
4A	(3/4)	ctccacacag AG CCT GAA lu Pro Glu	(753)	CTC AAA G gtctgtgtct Leu Lys A	(6/9)	2
5	(3/4)	attttatcag CT CGC ATG la Arg Met	(56)	GCC AAG gtaagctgac Ala Lys	(7/9)	1
6	(2/4)	tatgtttaag ACA TCC Thr Ser	(198)	CTC AAG gtaaggaaac Leu Lys	(7/9)	1
7	(4/4)	tcatttacag GGG GCT Gly Ala	(127)	AGC TCT G gtaagaagaa Ser Ser A	(7/9)	3
8	(2/4)	tctctttaag CG ACT AAG la Thr Lys	(54)	GAG AGA G gtactcggaa Glu Lys G	(5/9)	3.5
9	(4/4)	tccttcccag GT GAA CCT ly Glu Pro	(266)	GGG AAG gtgagagtgg Gly Lys	(6/9)	13
10	(3/4)	gctaccaaag GTG CAG Val Gln	(93)	GGC AGT gtgagtacct Gly Ser	(5/9)	>13
11	(4/4)	tcatctccag GTG CAA Val Gln	(82)	AAA CCA G gtagccctgt Lys Pro G	(6/9)	1.5
12	(3/4)	tgtgttctag GA GGT GGC ly Gly Gly	(113)	AAA AAG gtaaaggggg Lys Lys	(6/9)	10
13	(3/4)	cttcttgcag ATT GAA Ile Glu	(208)	AAG CAG G gtttgtgatc Lys Gln G	(5/9)	1.3
14	(3/4)	tgctccacag AA ACC CTG lu Thr Leu	(>310)	N/A		

[a] The numbers indicate the agreement between the τ splice site and the consensus sequence (Mount, 1982).

long products which would arise from the inclusion of exon 6 or 8. In contrast, the human mRNA did give the shorter products expected from a 6^- and 8^- τ species (Figure 5A; Table IB). Furthermore, probing of a Southern blot containing the above PCR products with probes specific for exons 6 and 8 gave no signal (data not shown).

PCR amplification of both fetal and adult human cDNA libraries using primers within exons 13 (sense) and 14 (antisense) gave only the long product (1.4 kbp) that would result from retention of the intron (Figure 5B; Table IB). This suggested that no splice site pair is utilized within the region shown in Figure 3D to produce either a bovine- or a murine-type extended carboxy-terminus (Himmler, 1989; Lee et al., 1988).

Thus, the results from sequencing, cDNA library screening, and RNA analysis suggest that, although the equivalents of exons 4A, 6, and 8 are present in the human genome, flanked by splice sites that fit the consensus, they are excluded in human cerebral cortex. Likewise, although the equivalent of τ exon 14 is present in the human genome, preceded by a canonical splice site, the intron between exons 13 and 14 is not excised in human cerebral cortex.

DISCUSSION

CpG Islands of the Human τ Gene. When CpG islands are found in regulated genes, they occur at both the 5′ end and internally, are associated invariably with an exon and occasionally with a repetitive element, and may extend up to 2 kbp (Gardiner-Garden & Frommer, 1986; Lindsay & Bird, 1987).

τ is a neuron-specific gene, and its two CpG islands are typical. They are more than 1 kb in length and are each associated with an exon (–1 and 9, respectively, Figure 2). Furthermore, the 5′ CpG island conforms to being well upstream of the translation start and extending into the transcribed region. The internal CpG island is associated with an AluI repetitive element (Jelinek & Schmid, 1982)

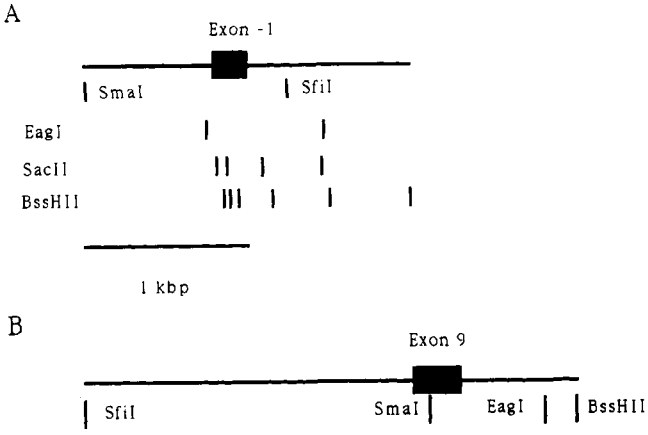

FIGURE 2: Two CpG islands associated with the human τ gene: (A) 5′, located around exon −1; (B) internal, located around exon 9. The sites for the restriction enzymes recognizing G/C palindromes are indicated.

which is located downstream of exon 9 (sequence data not shown; position indicated in Figure 1). The DNA sequence downstream of the CpG island is conserved between MAP2 and τ, whereas upstream of the CpG island the sequences diverge widely (Lewis et al., 1988). It is possible that this CpG island represents a site of gene duplication in an ancestral MAP.

τ Alternative Splicing. The neuron-specific τ transcript (Binder et al., 1986) contains at least 16 exons, of which 8 (exons 2, 3, 4A, 6, 8, 10, 13, and 14) are alternatively spliced [(Figure 1) Couchie et al., 1992; Goedert et al., 1989a,b, 1992; Himmler, 1989; Himmler et al., 1989; Kosik et al., 1989; Lee et al., 1988].

Human exons 4A, 6, and 8 are cassettes (Couchie et al., 1992; Goedert et al., 1992; Himmler et al., 1989), whereas in the bovine gene it has been reported that exon 6 either is internal or gives rise to a novel amino-terminus (Himmler 1989). In this connection, it is interesting that the homology between the human and bovine exon 6 extends into the region of the upstream intron (Figure 4C).

Differential inclusion of the two cassette exons 2 and 3 near the amino-terminus of τ produces three out of possible four combinations: 2^-3^-, 2^+3^-, and 2^+3^+ (Goedert et al., 1989a; Himmler 1989; Kosik et al., 1989). The asymmetry of the 2/3 expression patterns indicates that splicing in that region probably proceeds in a defined temporal order. Near the carboxy-terminus, alternative processing can introduce a fourth

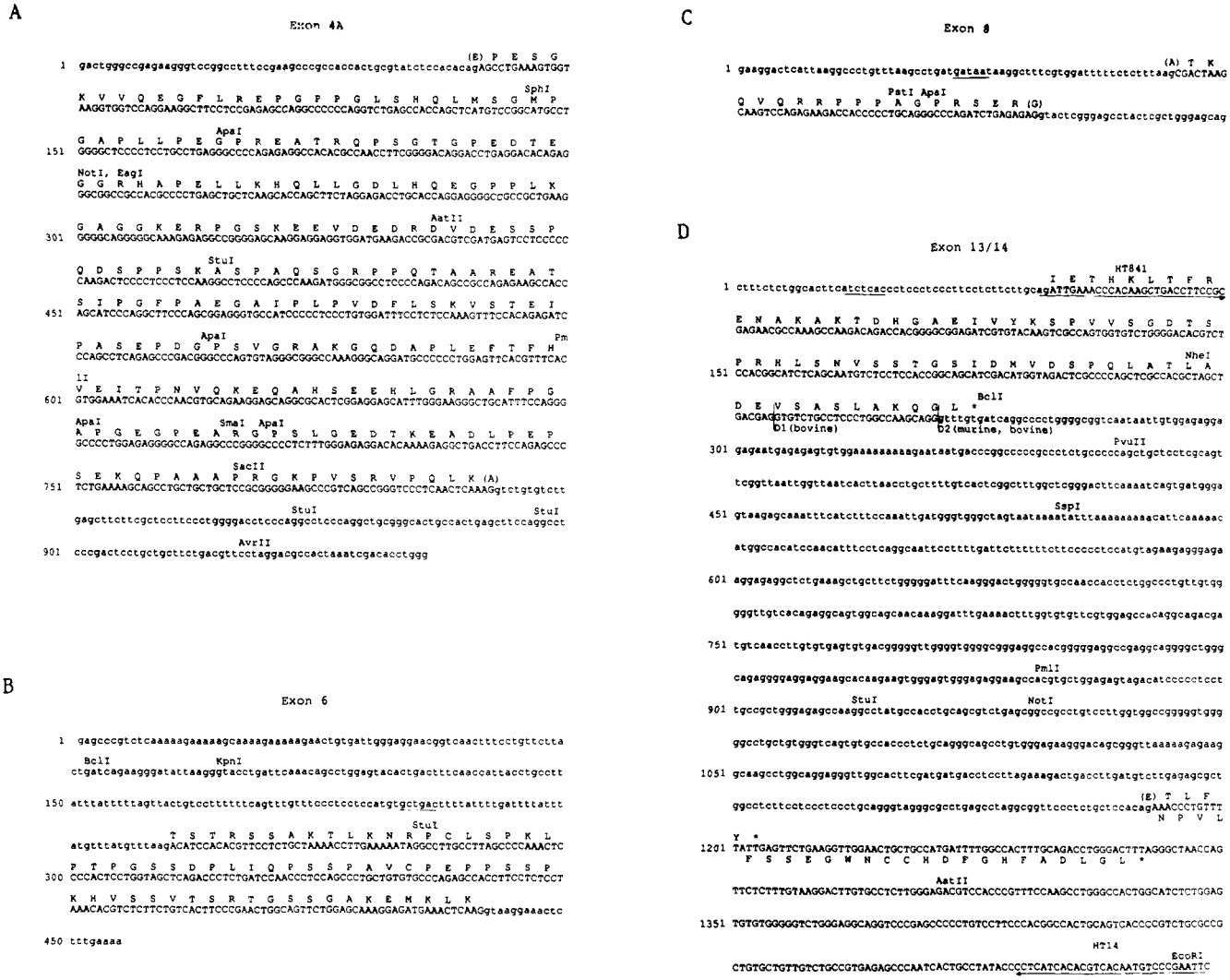

FIGURE 3: (A–D) Genomic sequence in the region of human exons 4A, 6, 8, and 13/14. The coding sequence (capitalized) and reading frame for all exons are indicated. Selected restriction sites are marked above the sequence. Protein sequence is in single letter code; amino acids in parentheses indicate split codons. Numbers on the left refer to the nucleotide sequence. Respective accession numbers for the four exons are M93652, X61371, X61375, and X61373. (B–D) Putative branch points are underlined. (D) Also shown are the bovine and murine alternative 5′ splice sites, the resulting reading frame if the murine (or distal bovine) 5′ splice site is utilized (above the sequence), the resulting reading frame if the 5′ signal at position 261 is utilized (below the sequence), and the oligomers used for PCR analysis.

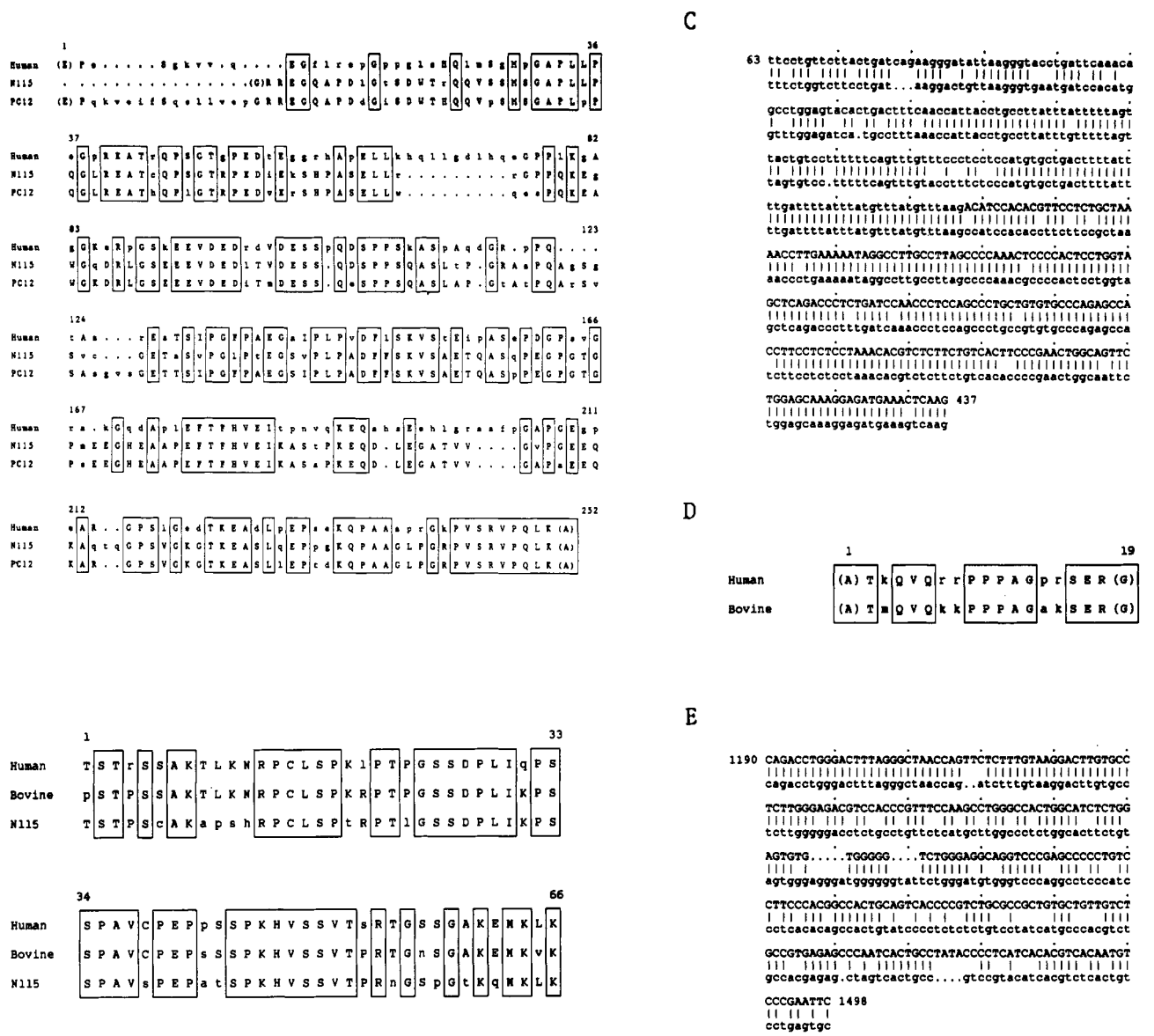

FIGURE 4: Amino acid and nucleotide homologies between the novel human exons and their bovine and murine counterparts. N115 and PC12 indicate τ sequences from mouse N115 and rat PC12 cells, respectively. Protein sequence is in single letter code; amino acids in parentheses indicate split codons. Alignments were made using the BESTFIT and ALIGN programs of the University of Wisconsin software package (Devereux et al., 1984) and further optimized manually. A minimum number of gaps, indicated by "...", were introduced to maximize the alignment. Identical residues are boxed. Numbers refer to the human sequence. Similarity was calculated according to Dayhoff et al. (1972). (A) Amino acid homology between human (top), N115 (middle), and PC12 (bottom) exon 4A. Identity is 48%, similarity 58%. (B) Amino acid homology between human (top), bovine (middle), and N115 (bottom) exon 6. Identity is 68% overall, 88% between human and bovine; similarity is 76%. (C) Nucleotide homology between human (top) and bovine (bottom) exon 6, extending beyond the exon. Identity is 88%. (D) Amino acid homology between human (top) and bovine (bottom) exon 8. Identity is 74%, similarity 90%. (E) Nucleotide homology between human (top) and murine (bottom) exon 14. Identity is 75%.

homologously repeated sequence of 31 amino acids (exon 10, also a cassette) (Goedert et al., 1989a,b; Himmler 1989; Kosik et al., 1989) into the microtubule-binding domain of τ (Lewis et al., 1989; Himmler et al., 1989; Lee et al., 1989). Exons 2, 3, and 10 are adult-specific in rat and human brain; all six possible product combinations of the latter two splicing events have been observed (Goedert et al., 1989a,b; Kosik et al., 1989).

In the bovine and murine gene, multiple 5′ splice sites in exon 13 give rise to different carboxy-termini of varying composition and length (Himmler 1989; Lee et al., 1988). In the murine gene, the choice is between one 5′ splice site or intron retention. In the bovine gene there are two potential 5′ splice sites (the downstream one coincides with the single 5′ splice site in the murine gene), which result in two different reading frames in exon 14. If exon 13 does not splice to exon 14, the translational stop codon within exon 13 is read, and the rest of the sequence acts as a 3′ untranslated region (3′-UTR).

It is possible that splicing of exon 14 is under developmental regulation. The murine cDNA clone which excises the 13/14 intron lacks the adult-specific cassette exons 2, 3, and 10 (Lee et al., 1988). When rat τ switches from the fetal to the adult form, its mRNA increases in size from 6 to 6.5 kb (Neve et al., 1986), possibly reflecting the shift from excision to retention of the 13/14 intron.

In rodents and cows, a ~9-kb τ mRNA species appears in certain tissues in addition to the more common 6 kb one (Drubin et al., 1988; Himmler et al., 1989; Oblinger et al., 1991). Whereas bovine τ cDNA primarily hybridizes to the 6-kb mRNA, sequences specific to exon 6 preferentially hybridize to the 9-kb species (Himmler et al., 1989). In mouse N115 cells, which only express the 9-kb τ mRNA, it is clear that exons 4A and 6 are associated with this high molecular

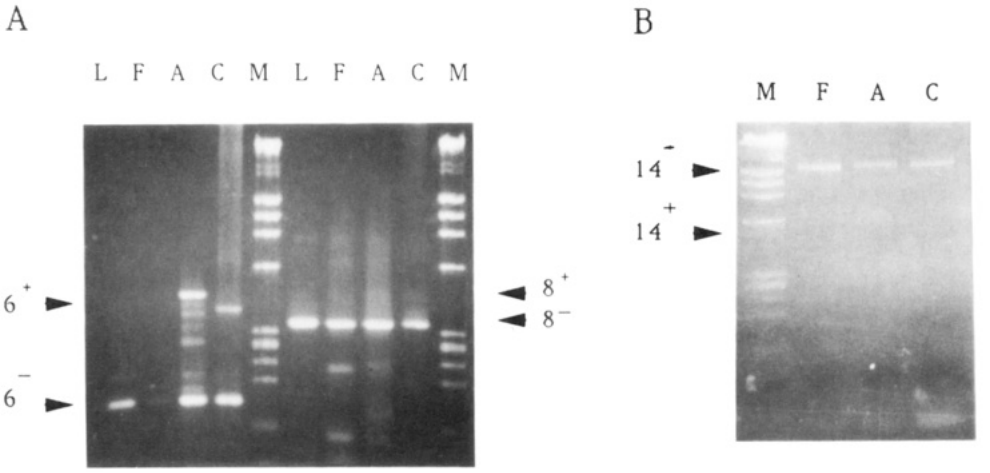

FIGURE 5: (A) Reverse transcriptase PCR analysis of the splicing behavior of human exons 6 and 8: L, LA-N-5 RNA; F, fetal human brain RNA; A, adult human brain RNA; C, fetal human τ cDNA; M, ϕX *Hae*III plus λ *Hind*III. Primer pairs: for exon 6, HTS4 (exon 4)/HT2 (exon 7); for exon 8, HT3 (exon 7)/648H (exon 9). The longer bands seen in the adult human and the positive control for PCR of exon 6 are variable and nonreproducible and were negative on Southern blot analysis. (B) Human library PCR analysis of the splicing behavior of human exons 13/14: F, fetal human brain cDNA; A, adult human brain cDNA; C, fetal human τ cDNA; M, ϕX *Hae*III plus λ *Hind*III. Primer pair: HT841 (exon 13)/HT14 (exon 14).

weight species (Couchie et al., 1992). In contrast to RNA from bovine brain and N115 cells, only a 6-kb species hybridized to a τ probe in RNA from human cerebral cortex (data not shown), which correlates with the absence of exons 4A and 6.

The relative positions of human exons 6, 8, and 14 correspond very closely to the arrangement and distances found in the bovine genome (Himmler, 1989). However, the detailed genomic structure of human τ differs from its bovine and murine homologs in several respects. Whereas exons 6 and 8 are absolutely conserved in length and highly conserved in sequence (Figure 4B,D), exon 4A is of variable length (Figure 4A). Human exon 4A falls between its rat PC12 and mouse N115 counterparts in number of nucleotides, possibly indicating that evolutionary and/or functional constraints might be relaxed on this particular exon. In human exon 6, the ATG codon that might serve as the novel amino-terminus is not conserved (Figure 4C, nucleotide 125), despite the very high homology (88%) of this region between bovine and human τ; in exon 14 a reading frame shift has occurred, so that the protein resulting from excision of the intron would contain only 3 extra amino acids, whereas the extended reading frames contain 23 additional amino acids in murine τ and either 29 or 69 amino acids in bovine τ, depending on which 5′ splice site is utilized (Himmler, 1989; Lee et al., 1988). The human τ gene also differs from the other two in having a slightly modified exon 1—it contains 10 additional intercalated amino acids after residue 15 (Goedert et al., 1989a; Lee 1990). Furthermore, despite the sequence conservation and availability of consensus splice sites, the human τ gene does not splice to include exons 6 and 8 in cerebral cortex, nor does it yield an alternative reading frame in exon 13. The intron between exons 13 and 14 is retained and the entire region exists as a very long 3′-UTR. All these observations are consistent with the fact that no human τ cDNA clones from the central nervous system containing exons 4A, 6, and 8 or missing the 13/14 intron have been isolated [this work, Goedert et al. (1989a, 1989b), and Lee et al. (1989)].

Regulation and Role of the Additional Coding Sequences in τ. Alternative splicing by definition occurs when splice sites that adhere to the consensus sequences are sometimes bypassed when the primary transcript is processed [reviewed in Andreadis et al. (1987) and Smith et al. (1989)]. In the case of cassette exons, the sequences responsible for inclusion or exclusion seem to reside within or very close to the exons themselves [reviewed by Smith et al. (1989)]. If their 5′ or 3′ splice sites are mutated to conform closer to consensus, cassette exons often become constitutive (Black, 1991; Grabowski et al., 1991; Tacke & Goridis, 1991). In this connection, it is interesting that the 5′ splice sites of exons 2, 4A, 8, and 10 and the 3′ splice sites of exons 3, 6, and 8 conform markedly less to the consensus than those of the constitutive τ exons (Table II).

In the case of retained introns, the best studied example is the *Drosophila melanogaster* P transposon [reviewed by Siebel and Rio (1990)]. As in τ, retention of the last intron in the P element results in early termination, whereas its excision gives rise to a longer reading frame which codes for the transposase enzyme. Excision of the intron is made inefficient by the presence of imperfect 5′ splice site sequences around the bona-fide one (Siebel & Rio, 1990). There are imperfect 5′ splice site sequences around both potential 5′ splice sites of the human exon 13 (Figure 3D). However, as such short sequences are rather common throughout the genome, it is impossible to postulate a model for τ similar to that of the P element on sequence comparisons alone.

The only τ cassette exon which has an assigned function is exon 10. Its presence increases the affinity of τ protein for microtubules (Goedert et al., 1989b; Kosik et al., 1989). The specific function of the other cassette exons (2, 3, 4A, 6, 8) is, as yet, unknown. Exons 2 and 3, like exon 10, are developmentally regulated. On the other hand, it is becoming increasingly evident that exons 4A and 6 are tissue-specific and apparently associated with neurons of the peripheral nervous system (Couchie et al., 1992; Goedert et al., 1992). Exon 4A, an unusually long internal exon, would very likely give rise to an extended random coil configuration, since it is rich in prolines and glycines (Chou & Fasman, 1978). If the peptide encoded by exon 4A is needed only to add length to the τ molecule, this may explain its relatively large degree of sequence drift. Conversely, the peptides encoded by exons 6 and 8 are highly conserved, suggesting a specific possible function for these two exons. The role of the alternative carboxy-termini in the bovine and murine genes has not been explored either. An obvious possibility is that they might modulate microtubule affinity.

Finally, although 3′ terminal exons tend to be very long, the case of τ is extreme; retention of the intron between exons 13 and 14, which produces the canonical τ isoform, results in a 3′-UTR of at least 2 kb, one-third of the total mRNA length and more than the size (1.4 kb) of the protein-encoding region. The ratio between coding region and total mRNA length is even more skewed in the case of the high molecular weight τ (2.4 versus 9 kb). The common supposition is that such features regulate mRNA stability and localization [reviewed in Andreadis et al. (1987) and Smith et al. (1989)]; nevertheless, the role of the τ 3′-UTR is also unknown.

ACKNOWLEDGMENT

We thank Drs. Sotiris Karathanasis, John Ladias, and Rudy Tanzi for the human genomic libraries; Dr. Rachel Neve for the human fetal cDNA library; Margaret Erhardt for synthesizing the oligonucleotides; and Paul Watkins, Catherine Linder, Lisa Davis Orecchio, Renee von Toerne, and Audrey Loftis for help with the sequencing.

REFERENCES

Andreadis, A., Gallego, M. E., & Nadal-Ginard, B. (1987) *Annu. Rev. Cell. Biol. 3*, 207–242.
Berk, A. J., & Sharp, P. (1978) *Cell 14*, 695–711.
Biggin, M. D., Gibson, T. J., & Hong, C. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.
Binder, L. I., Frankfurter, A., & Rebhun, L. I. (1986) *Ann. N.Y. Acad. Sci. 466*, 145–166.
Black, D. (1991) *Genes Dev. 5*, 389–402.
Chornczynski, P., & Sacchi, N. (1987) *Anal. Biochem. 162*, 156–159.
Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251–276.
Cleveland, D. W., Hwo, S., & Kirschner, M. W. (1977) *J. Mol. Biol. 116*, 207–225.
Couchie, D., Mavilia, C., Georgieff, I. S., Liem, R. K. H., Shelanski, M. L., & Nuñez, J. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 4378–4381.
Dayhoff, M. O., Eck, R. V., & Park, C. M. (1972). In *Atlas of Protein Sequence and Structure* Dayhoff, M./O., Ed. pp 89–99, National Biomedical Research Foundation, Washington, DC.
Delacourte, A., & Défossez, A. (1986) *J. Neurol. Sci. 76*, 173–186.
Devereux, J. R., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.
Drubin, D. G., Caput, D., & Kirschner, M. W. (1984) *J. Cell Biol. 98*, 1090–1097.
Drubin, D. G., Kobayashi, S., & Kirschner, M. W. (1986) *Ann. N.Y. Acad. Sci. 466*, 257–268.
Drubin, D. G., Kobayashi, S., Kellogg, D., & Kirschner, M. (1988) *J. Cell Biol. 106*, 1583–1591.
Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6–13.
Gardiner-Garden, M., & Frommer, M. (1987) *J. Mol. Biol. 196*, 261–282.
Goedert, M., Spillantini, M. G., Jakes, R., Rutherford D., & Crowther, R. A. (1989a) *Neuron 3*, 519–526.
Goedert, M., Spillantini, M. G., Potier, M. C., Ulrich, J., & Crowther, R. A. (1989b) *EMBO J. 8*, 393–399.
Goedert, M., Spillantini, M. G., & Crowther, R. A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 1983–1987.
Grabowski, P. J., Nasim, F. H., Kuo, H.-C., & Burch, R. (1991) *Mol. Cell Biol. 11* 5919–5928.
Grundke-Iqbal, I., Iqbal, K., Tung, Y.-C., Quinlan, M., Wisniewski, H., & Binder, L. I. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4913–4917.
Himmler, A. (1989) *Mol. Cell. Biol. 9*, 1389–1396.
Himmler, A., Drechsel, D., Kirschner, M. W., & Martin, D. W. (1989) *Mol. Cell. Biol. 9*, 1381–1388.
Innis, M. A., Gelfand, F. H., Sninsky, J. J., & White, T. J. (1990) *PCR Protocols: A Guide to Methods and Applications*, Academic Press, New York.
Jelinek, W. R., & Schmid, C. W. (1982) *Annu. Rev. Biochem. 51*, 813–844.
Kanai, Y., Takemura, R., Oshima, T., Mori, H., Ihara, Y., Yanagisawa, M., Masaki, T., & Hirokawa, N. (1989) *J. Cell. Biol. 109*, 1173–1184.
Katayama, C. (1990) *Strategies 4*, 56–58.
Kosik, K. S., Joachim, C., & Selkoe, D. J. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4044–4048.
Kosik, K. S., Orecchio, L. D., Binder, L. I., Trojanowski, J., Lee, V., & Lee, G. (1988) *Neuron 1*, 817–825.
Kosik, K. S., Orecchio, L. D., Bakalis, S., & Neve, R. L. (1989) *Neuron 2*, 1389–1397.
Kowall, N. W., & Kosik, K. S. (1987) *Ann. Neurol. 22*, 639–643.
Lee, G. (1990) in *Cell Motility and the Cytoskeleton*, Vol. *15*, pp 199–203. Wiley-Liss Press, New York.
Lee, G., Cowan, N., & Kirschner, M. (1988) *Science 239*, 285–288.
Lee, G., Neve, R. L., & Kosik, K. S. (1989) *Neuron 2*, 1615–1624.
Lewis, S. A., Wang, D., & Cowan, N. J. (1988) *Science 242*, 936–939.
Lewis, S. A., Ivanoy, I. E., Lee, G. H., & Cowan, N. J. (1989) *Nature (London) 342*, 498–505.
Lim, H. M., & Péne, J. J. (1988) *Gene Anal. Tech. 5*, 32–39.
Lindsay, S., & Bird, A. P. (1987) *Nature (London) 327*, 336–338.
Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.
Neve, R. L., Harris, R., Kosik, K. S., Kurnit, D. M., & Donlon, T. A. (1986) *J. Mol. Brain Res. 1*, 271–280.
Nukina, N., & Ihara, Y. (1986) *J. Biochem. 99*, 1541–1544.
Oblinger, M., Argasinski, A., Wong, J., & Kosik, K. S. (1991) *J. Neurosci. 11*, 2453–2459.
Olmsted, J. B. (1986) *Annu. Rev. Cell Biol. 2*, 421–457.
Peng, A., Binder, L. I., & Black, M. M. (1986) *J. Cell Biol. 102*, 252–262.
Reed, R., & Maniatis, T. (1988) *Genes Dev. 2*, 1268–1276.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. U.S.A. 74*, 5463–5467.
Siebel, C. W., & Rio, D. C. (1990) *Science 248*, 1200–1208.
Smith, C. W. J., Patton, J. G., & Nadal-Ginard, B. (1989) *Annu. Rev. Genet. 23*, 527–577.
Tabor, S., & Richardson, C. C. (1989) *J. Biol. Chem. 264*, 6447–6458.
Tacke, R., & Goridis, C. (1991) *Genes Dev. 5*, 1416–1429.
Wood, J. G., Mirra, S., Pollock, N. J., & Binder, L. I. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4040–4043.