Structure, Expression, and Some Regulatory Mechanisms of the Rat Preprotachykinin Gene Encoding Substance P, Neurokinin A, Neuropeptide K, and Neuropeptide γ

Mark S. Carter and James E. Krause

Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri 63110

The rat preprotachykinin (PPT) gene encoding the neuropeptides substance P (SP), neurokinin A (NKA), neuropeptide K (NPK), and neuropeptide γ was isolated from a λ Charon 4A genomic library. Two overlapping clones contained all of the exons present in β-PPT, including some 7 and 9 kb 5′ and 3′ flanking sequence, respectively. The presence of 1 major and 2 minor transcription initiation sites was determined from primer extension and nuclease protection experiments. Analysis of the nucleotide sequence homology between the rat and bovine revealed the presence of highly conserved regions throughout the entire coding region and within the 5′ flanking sequences. Primer extension and nuclease protection experiments demonstrated that the primary transcript is differentially spliced primarily into γ- and β-PPT mRNA in all tissues examined in the adult rat where the gene is expressed. β-PPT mRNA contains all of the exons, whereas γ-PPT mRNA lacks exon 4, which encodes part of the N-terminus of NPK. The α-PPT mRNA, which lacks exon 6 (the sequence encoding NKA and processing sites), comprises about 1% of the total PPT mRNA. An RNA secondary structure model is proposed to account for these specific exon exclusion events in the RNA splicing process. These results are discussed with regard to the mechanisms regulating SP gene expression and the functional significance of differential RNA splicing in the rat.

The tachykinins are a family of peptides so named due to their ability to rapidly induce the contraction of gut tissue, and they are neurotransmitter and/or neuromodulator substances in both the CNS and PNS (Pernow, 1983; Maggio, 1988; Krause et al., 1989b). Peptides in this family include substance P (SP), neurokinin A (NKA), neuropeptide K (NPK), neuropeptide γ (NPγ), and neurokinin B (NKB). These peptides are derived by post-translational processing of polyprotein preprotachykinin (PPT) precursors. The mRNAs that encode SP, NKA, NPK, and NPγ are derived from a single gene (Nawa et al., 1983, 1984; Krause et al., 1987), whereas NKB is encoded by a distinct gene (Kotani et al., 1986; Bonner et al., 1987). Alternative RNA splicing of

the SP gene primary transcript results in the generation of 3 mRNAs called α-, β-, and γ-PPT mRNA, with each mRNA differing in exon usage corresponding to the protein coding region. SP precursor sequences are encoded by all 3 PPT mRNAs, whereas NKA precursor sequences are present in β- and γ-PPT mRNAs. NPK and NPγ are N-terminally extended derivatives of NKA that appear to be final products in the posttranslational processing of β- and γ-PPT in some tissues (Tatemoto et al., 1985; Arai and Emson, 1986; Kage et al., 1988; Krause et al., 1989b; Takeda et al., 1989). Thus, multiple tachykinin peptides with related biological activities can be derived from the SP gene.

Since primarily the rat has been used in studies directed toward understanding aspects of biosynthesis, posttranslational processing, axonal transport, secretion, function, and regulation of protachykinin-derived peptides, it is important to define the structure of the rat SP gene and to determine the mechanisms that regulate its expression. Moreover, since tissue-specific alternative RNA splicing of the bovine SP gene primary transcript has been reported (Nawa et al., 1983, 1984), it is essential to determine how general the pattern of splicing is in different species. Important differences have already been reported (Krause et al., 1987, 1989a) in terms of the types of SP-encoding RNAs detected in the rat CNS and PNS compared to those of the bovine.

In the experiments reported here, we detail the isolation and characterization of the rat SP gene, as well as some of its 5′ flanking sequence. We document the site(s) of transcription initiation, provide further evidence for the expression of multiple mRNAs from this gene, and demonstrate that the 3 PPT mRNAs are cytoplasmic. A comparison of the nucleic acid and deduced amino acid sequences of the coding regions of the bovine, rat, and human genes reveals a striking degree of homology of the tachykinin peptide coding sequences including NPK and NPγ, as well as in other regions of the precursors that have been generally referred to as "spacer" regions. This work provides a basis for further analysis of the regulation of the rat PPT gene encoding, SP, NKA, NPK, and NPγ.

Materials and Methods

Materials. Restriction endonucleases, avian myeloblastosis reverse transcriptase, RQ1 DNase, DNA polymerase I, and Klenow fragment of DNA polymerase I, T4 polynucleotide kinase, T4 DNA ligase, SP6 polymerase, T7 polymerase, calf intestinal phosphatase, universal oligonucleotides for DNA sequence analysis, and placental ribonuclease inhibitor were purchased from Bethesda Research Laboratories, Boehringer Mannheim, New England Biolabs, or Promega Biotec. Proteinase K, nuclease S1, and ribonucleases A and T1 were obtained from Sigma

Chemical Company. dNTPs and ddNTPs were from Boehringer Mannheim. PPT gene–related oligonucleotides were synthesized by the Washington University Protein Chemistry Facility. Other chemicals were the highest purity available. [γ-^{32}P]ATP($>$6000Ci/mmol), [α-^{32}P]dCTP (3000Ci/mmol), [α-^{32}P]CTP(800Ci/mmol), and [α-^{32}P]UTP(800Ci/mmol) were from Amersham or ICN, and [α-^{35}S]-dATP($>$1300Ci/mmol) was from New England Nuclear.

Tissue isolation and RNA and DNA preparation. Adult male rats of the Sprague-Dawley strain (Holtzman Company, Madison, WI) were used. Tissues were dissected or isolated, rinsed with 10 mM PBS, and frozen on dry ice. For dissection of the CNS, the landmarks and method of Glowinski and Iversen (1966) were used. The entire spinal cord was also isolated. Gastrointestinal tissues were flushed with PBS prior to freezing. Nuclear and cytoplasmic subcellular fractions were prepared as described previously (Krause and Karavolas, 1980). RNA was prepared by the guanidinium–cesium chloride centrifugation method (Chirgwin et al., 1979). High-molecular-weight DNA was prepared as described (Maniatis et al., 1982) using pulverized frozen tissue.

Isolation and characterization of genomic clones. A rat genomic library was provided by T. Sargent (Sargent et al., 1979) and was screened using a nick-translated cDNA insert derived from pSP27-4 (Krause et al., 1987), which is near full length β-PPT cDNA. Of the approximately 5 genome equivalents of DNA screened, 2 genomic clones ($\lambda\phi$11-1 and $\lambda\phi$1-1) were isolated and characterized by restriction endonuclease mapping. Three PstI hybridizing fragments (see Fig. 1) were further analyzed. From $\lambda\phi$11-1, 1 and 5.8 kb fragments were isolated and subcloned into the plasmid pUC19. Similarly, from $\lambda\phi$1-1, a 1.4 kb PstI fragment was isolated and subcloned into pUC19. Southern blot hybridizations (Maniatis et al., 1982) were performed with the endonuclease-digested inserts of the above plasmids to further map the locus.

Nucleotide sequence strategy and analysis. The genomic fragments inserted into pUC19 were subcloned into M13mp18 and mp19 for sequence analysis by the dideoxynucleotide chain termination method of Sanger et al. (1977). Various PPT-based oligonucleotide primers as well as M13 universal primers were used for sequence analysis as shown in Figure 1. Deleted subclones were created using the Exo III digestion method of Henikoff (1984). Nucleotide sequences were analyzed with an IBM AT computer and Microgenie (Queen and Korn, 1984) software. The method of Tinoco et al. (1973) was used for predicting RNA secondary structure. GenBank was used for nucleotide sequence homology searches. The rat SP gene sequence has been submitted to GenBank, accession numbers m 34159–m 34162.

Plasmid constructions for cRNA transcriptions and nuclease protection experiments. A 1.0 kb PstI fragment from $\lambda\phi$11-1 harboring exon 1 was subcloned into pGem1 and the orientation of the insert was verified by restriction mapping. The resulting plasmid was called pG1PP1.

The construction of pG1α-PPT, pG1β-PPT, and pG1γ-PPT has been described previously (MacDonald et al., 1988; Krause et al., 1989a,b).

Primer extension experiments. The general procedures of Ghosh et al. (1980) and Gardner et al. (1987) were used as a guide. Specifically, total RNA (25 μg) was coprecipitated with approximately 1 ng end-labeled oligonucleotide and was resuspended in 10 mM Tris-HCl, pH 7.5, containing 1 mM EDTA (TE). Annealing of probe and total RNA in TE was achieved by sequential 1 hr incubations at 10° below and 15° below the calculated T_m for the duplex [T_m = 4(G+C) + 2)(A + T)]. Reverse transcription was performed by incubating the oligonucleotide-RNA solution at 42°C for 30 min in 10 mM Tris-HCl, pH 8.0, containing 10 mM MgCl$_2$, 5 mM DTT, 0.25 mM each dNTP, 5 units RNasin, and 15 units reverse transcriptase. Nucleic acids were ethanol precipitated and resuspended in a 99% formamide solution containing dye and were electrophoresed on a 6% polyacrylamide gel containing 7 M urea at 25 V/cm.

cRNA transcription reactions and nuclease protection experiments. An RNA probe complementary to exon 1 and upstream sequences was prepared using Hind III linearized pG1PP1 and T7 polymerase and was used for transcription initiation site analysis. An RNA probe complementary to the coding region of β-PPT mRNA was prepared using Hind III linearized pG1β-PPT and T7 polymerase. Message sense RNAs corresponding to α-, β-, and γ-PPT mRNAs were prepared using EcoR1 linearized pG1α-PPT, pG1β-PPT and pG1γ-PPT and SP6 polymerase (MacDonald et al., 1988; Krause et al., 1989a,b). Transcription reactions were carried out using the procedures suggested by Promega. Template DNA was subsequently digested with RQ1 DNase. The specific activity of the antisense RNA probe derived from pG1β-PPT was 4.5 $\times$ 10^8 cpm/μg and that derived from pG1PP1 was 5.6 $\times$ 10^8 cpm/μg.

The general procedures for nuclease protection experiments have been discussed (Krause et al., 1987, 1989a). In general, antisense ^{32}P-probe (2 $\times$ 10^5 dpm) was used for each protection assay. Probe was coprecipitated with total RNA (generally 25 μg), and the precipitate was resuspended in 10 μl of 40 mM PIPES, pH 6.4, containing 400 mM NaCl, 1 mM EDTA, and 80% (vol/vol) formamide. Annealing was performed for $\geq$12 hr at 45°C. Different procedures were used for digestion of nonannealed RNAs in each nuclease protection assay. For the antisense RNA probe derived from pG1PP1 (transcription initiation site analysis), nonannealed RNAs were digested with RNases A (40 μg/ml) and T$_1$ (2 μg/ml) in 300 μl as described (Krause et al., 1989a,b). After digestion for 30 min at 37°C, proteinase K was added and the nucleic acids were precipitated from 2-propanol prior to electrophoresis. For the antisense RNA probe derived from pG1β-PPT (mRNA splicing form analysis), nonannealed RNAs were digested with S$_1$ nuclease (100–1000 U/ml, depending on batch) in 300 μl as described by Krause et al. (1989a). After addition of 75 μl termination mix (2.5 M NH$_4$OAc containing 50 mM EDTA), carrier RNA was added and nucleic acids were precipitated from 2-propanol prior to electrophoresis. For both primer extension and nuclease protection analyses, electrophoresis was carried out at 25 V/cm on 6% polyacrylamide gels containing 7 M urea.

To quantitate the amount of specific PPT mRNAs, the optical density of the autoradiographic signals corresponding to the protected α-, β-, and γ-PPT mRNA regions was determined as described previously (Krause et al., 1989a,b). ^{32}P-dCTP was used for the preparation of the ^{32}P-β-PPT antisense cRNA probe; consequently, the total radioactive bases corresponding to protected PPT mRNA species were as follows:

$$\alpha\text{-PPT mRNA} = \frac{96\text{ C's}}{364\text{ bases}} + \frac{18\text{ C's}}{90\text{ bases}} = \frac{114\text{ C's}}{454\text{ bases}},$$

$$\beta\text{-PPT mRNA} = \frac{126\text{ C's}}{508\text{ bases}},$$

$$\gamma\text{-PPT mRNA} = \frac{83\text{ C's}}{295\text{ bases}} + \frac{35\text{ C's}}{168\text{ bases}} = \frac{118\text{ C's}}{463\text{ bases}}.$$

Since the total number of radioactive bases protected for α-, β-, and γ-PPT mRNAs was similar (9.5% or less difference from the β-PPT mRNA species), the optical density of the sum of protected species (for α-PPT mRNA, and/or γ-PPT mRNA was used to measure mRNA abundance.

To determine the number of transcripts initiated from different transcription initiation sites by nuclease protection procedures, the specific activity of the hybridizing probe region and density of the autoradiographic signals derived by protection were used. ^{32}P-ATP was used for pG1PP1 antisense cRNA probe preparation, and the total radioactive bases protected corresponding to transcription initiation at different sites were as follows:

$$\text{TATA I} = \frac{7\text{ A's}}{100\text{ bases}},$$

$$\text{TATA II} = \frac{9\text{ A's}}{125\text{ bases}},$$

$$\text{TATA III} = \frac{42\text{ A's}}{278\text{ bases}}.$$

Thus, the protected fragments from TATA II and TATA III sites had specific radioactivities 1.3 and 6.0 times greater than the protected fragment from the TATA I site.

Southern and Northern blot hybridizations. Southern blotting of genomic DNA was performed as described (Maniatis et al., 1982) using nick-translated ^{32}P-cDNA insert from pSP 27-4 (Krause et al., 1987) or with a β-actin cDNA probe (Nadel et al., 1983). Northern blots were performed as described (Krause et al., 1987) using a ^{32}P-RNA probe transcribed from pG1β-PPT. RNA blots were washed at 68°C in 0.1 $\times$ SSPE containing 0.1% sarcosyl.
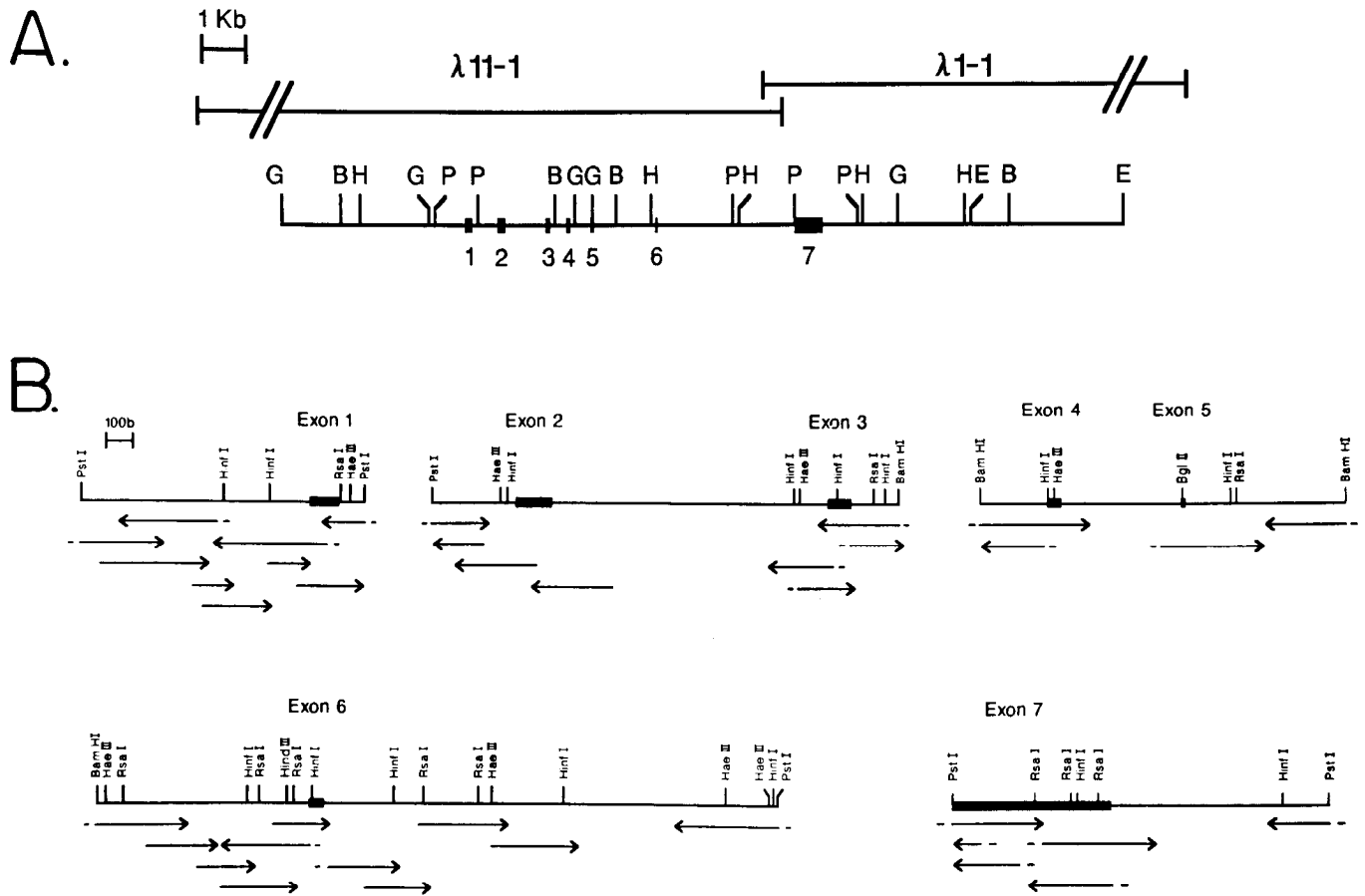
Figure 1. Isolation and characterization of the rat PPT gene encoding substance P neurokinin A. *A*, Restriction map of the transcription unit as determined by blot and nucleotide sequence analysis of Charon 4A genomic clones λ11-1 and λ1-1. The *numbers* below the map represent exons 1–7, and the *letters* above represent restriction enzyme sites. Abbreviations: *G*, BglII; *B*, BamHI; *H*, Hind III; *P*, PstI; *E*, EcoRI. *B*, Strategy used for determining the nucleotide sequence of various M13 genomic subclones. The direction and extent of sequence obtained is displayed by the *arrows*, and relevant restriction sites are indicated. Universal or other synthetic oligonucleotides, as described in Materials and Methods, were used to prime synthesis and are denoted by the *bars*.

Results

Isolation and characterization of the rat SP gene. We previously isolated a near full length 1035 base pair rat β-PPT cDNA clone, pSP27-4 (Krause et al., 1987). The nick-translated insert derived from this clone was used to screen a rat Charon 4A genomic library. Two overlapping clones were isolated that contained the entire rat SP gene, based upon nucleic acid sequence analysis. Figure 1A displays the restriction map of the rat PPTI gene as determined from the analysis of genomic clones λφ1-1 and λφ11-1, which consisted of approximately 18 and 20.5 kb, respectively. Relevant hybridizing regions of the genomic clones were subcloned into pUC 19 and into M13 for further characterization by restriction and nucleotide sequence analysis as described in Materials and Methods. Figure 1B displays the strategy for nucleotide sequence analysis of various regions of the SP gene locus.

Structure of the rat PPT I gene. Comparison of the structure of β-PPT cDNA with the genomic DNA sequence allowed for the determination of the structural organization of the rat gene (Figs. 1, 2). The PPT I gene is about 8 kb in length and contains 7 exons, similar to the bovine gene (Nawa et al., 1984).

Exons 1–7 consist of 100, 132, 97, 45, 24, 54, and 593 bases, respectively. The site(s) of transcription initiation was identified by primer extension and nuclease protection experiments described below. Exon 1 encodes most of the 5' untranslated sequence, and the remaining 9 bases of 5' untranslated sequence are encoded on exon 2. PPT precursor sequence is encoded by 123 nucleotides of exon 2, exons 3–6, and 47 bases on the 5' end of exon 7. Exon 6 is absent in α-PPT mRNA, whereas exon 4 is absent in γ-PPT mRNA. Part of exon 3 encodes the SP sequence and its processing sites, and exon 6 encodes the NKA sequence and its processing sites. Intron A (425 bp) interrupts the 5' untranslated sequence, and introns B (~1035 bp), C (~440 bp), D (~440 bp), E (~1.47 kbp), and F (~3.2 kbp) interrupt the protein coding sequence.

Potential cis regulatory sequences in the 5' flanking region of the SP gene. A variety of nucleotide consensus sequences have been identified in many genes that are essential for transcriptional regulation. A comparison of the 5' flanking sequence of the rat SP gene with these consensus sequences has allowed for the identification of some putative regulatory regions. A sequence with 85% homology to an estrogen response element (Maurer and Notides, 1987) is located at −137 to −149, and sequences with 67 and 75% homology to a cAMP response element (Roesler et al., 1988) are found at −30 to −41 and at −398 to −409. A sequence similar to the so-called dyad symmetry or serum response element (Triesman, 1986), a *cis* se-

ctgcagagctccaaaggtaagcatccagcctttctagtccccaacaaggctaaaggggagagaggcacaattatcctcttcccacccccttctgccttcagggtgtgcctgggaagaagctgtaggggaacaaaagatgccttagaatggctgatgggta   -706

agttctacatgagaaaggaggtttaaattcctctttcccctaaatgtaaaacaaacctgccttcatcctctgaagcgggagaccggaaacacttttgcagtgctagagaaatgagaatattctgactgatttggtggggaggggggttgggggggtgtgt   -546

tccagccctagatataacacctcataaaccttaagacacataaagtagaaatgaaaugaaaacccgcttgcttcatccctctgaagtgcttgctggtgtcttagtattattcacaaggttttgctgctcaagttatttggctgtcctcaaagcgcaata   -386

ttccctgatgcctcttgagagaaaagttccctaagtccgaagcatgagtcacttcgctcagttttgatgagtaatctcaggtgtcactgaaccttgttcqgaagaagaggggaggggggcgtcagatttgcagacggaagaaaacaggtctctctggatt   -226

TATA III   **TATA II**
ggatggcgagacctcgacttccctaaaattcgtcatttcgaacccaattggtccagatgttatggactccgacgggttaccgtctcggaaactctatcacgcaagcaaaggcgaggggcggctaattaaatattgagcagaaagtcgcgtgggqag   -66

                                    TATA I   **Exon 1**
agtgtcacgtggctctccaggctcatcacgcctgagataagtaaggcgaagcaggagcagggactAGAGCGCACTCGGACCAGCTCCACTCCAGCACCGCGGCGGAGGAGAGCGAGGAGCGCCCAGCAAGTGCGCACCTGCGGAGCATCACCGGGTCCGA   +95

CCGCAgtgagtaccactcccgcctgccatctgccgctgcctggggccgtcgtgccttggtcagaggtggcgtccgcccgggggtttcacctgcagcagccacagacccggagagaaactgtgcaaataccaacacctctctttgtctgccttcaggcttc   255

ggagtgtgggtcagtgggtaggggggctgggacgttgagaggcaaagagaggaggacttgaggctctttggcacgtcagtagccttcttaaagggtttggagaaatgttcgacctggtgggtttttgtttttgtttaatcttggtttagcgagacctctt   415

                                                                   Exon 2
ctttcttccttgtggcattgaaattcagctgctctaacgtcttgggtggccccccgcctacatatgtataacgattctttctctccttctctgtctcctactctctccaAAATCCAACATGAAAATCCTCGTGGCGGTGGCGGTCTTTTTTCTCGTTTC   575
                                                                            MetLysIleLeuValAlaValAlaValPhePheLeuValSer
                                                                            1   10

CACTCAACTGTTTGCAGAGGAAATCGGTGCCAACGATGATCTAAATTATTGGTCCGACTGGTCCGACAGTGACCAAATCAAGgtgagaccaaccgggcgtgggagctgttgccgtccccttgtggtagcaccctgactagcctagctggagcctgagagg   735
rThrGlnLeuPheAlaGluGlyIleGlyAlaAsnAspAspLeuAsnTyrTrpSerAspTrpSerAspSetAspGlnIleLys
         20   30   40

ctcattggattgtaagggcgtcataccggacaaagcacaaggcttattgcccagaactaagtttctccctgagggaggaaaatgtcctgggtcagtgtctccccaacagaaaaggagttcttctttccttcctccctccctccctccctcc---   895
-IntronB, 0.58 kb----ccctgatagactttgggggagaaggaaatctcaggcgggtccgggttggggaaagtgctgcatagattctagtatggccggtctcatgctagcggctgcttctgtggctctgaggaaagctcaggcacaccgtgaat   1613

                                             Exon 3
gtggaccttgttcacttaatttgtctactcgctggtttgtccctccagGAGGCAATGCCGGAGCCCTTTGAGCATCTTCTTCAGAGAATCGCCCGAAGACCCAAGCCTCAGCAGTTCTTTGGATTAATGGGCAAACGGGATGCTGgtgagatgagcaata   1773
                                               GluAlaMetProGluProPheGluHisLeuLeuGlnArgIleAlaArgArgProLysProGlnGlnPhePheGlyLeuMetGlyLysArgAspAlaA
                                               50   60   70

gtcactaggtctatcaggctacccggtctgcatgcctgctccttcctggagcacccaagtgtctcttgtaccgcgagtataattttacatgcctaaagacttttgattccctgaagtcttcgatccactaacacttgcactagctcaaagtggagggggag   1933
gggggatccattttcccttgcttcacagcaccaaggtgacccaaattggaaactaaccttagctaaacactcccctgactaaagatcccaattacaagggtcctcaggtatgctaatggtccaggctcaggtgaagatgttttcttggatLgggttagat   2093

                                             Exon 4
ttggctataccctcgctacagtgataggaattttcaattgccttgagttgatggcacagtcagtggagcttgtgaagagttaatgtcaattcatctcttgtcagATTCCTCAATTGAAAAACAAGTGGCCCTGTTAAAGGCTCTTTATGgtaaacatccct   2253
                                                                                                          spSerSerIleGluLysGlnValAlaLeuLeuLysAlaLeuTyrG
                                                                                                          80

ataatcttatttcctatatgatgatcatgtaaaagtatttgatggtataatcagacttgcagtgtgcctatagagagtgtgtgtgtgtgtgtgtgtg----IntronD, 0.21 kb----catctactttagtcctcagaacagcacataaataagcg   2598

                                             Exon 5
ttaaatatttcgaaatcaactccttaagttctctggaaaagataaaatagctttaaaaaagaaacccatgtatctgctaactggctctcttttcttctagGGCATGGTCAGATCTCTCACAAAGtaagttcacaatcatttgacatttatcaaatgct   2758
                                                                                                     lyHisGlyGlnIleSerHisLysA
                                                                                                     90

aatgttaatgatagtgaatattgcatcctgtgttaccttttgcagcacaacatagagcccaatggtttgattgagataatgtagggaaagaaagacagaatgtgatttgctgaatgatcactgactagactctttatgtagacagattacgtaccaaat   2918
gcacattacattgagagaaaagatacactatagttttttagtttatattctctctagggtaatcacatgagcactttccagcctgtccagaagacgagagagagagagagagagagagagagagagagagagaacacttatatacacccttctaat   3078
tcatattcatgaactacctttttataatggcaggaagggattaagtagactgggtgaatatttgagtagaaaatgcactcgaaatgtgatagagatacaggtttattttccccttgatcccataattgaatgggaacctgggagagaagaggtgcagat   3238
tgcagaactggagggtccttagagctatttcctagttaattggatccagaggagacaggaagtcccttgggcccctggaaatgttgcctgcttctatgagggtcatgaattggactaaccctggcagggtacaccccaaaccattaacactgtggtctca   3398
actaacacaatggaaactaagaaaaatcctccccgttgcagtatgataataacttccccaaagccttcatctttggggaaaaaaaatgacatttactaagggtttaaatttaaagaaatcttatacagcaaaaaaagtaataaaatgaatgtatctta   3558
ttaatgttaaaaaacaaattgctttcaggaacagtattttaagtctaaatctactaaatgtgacccctttagaatattccaatactaaggatctgacttaaagttcccttgttgagaaatgcaacatatttggtaataaagacataaacgtgaatgtctt   3718
ttcccacacaaaaacatataattaaagtccccaaaattcaaagctactccagtggcaaccatatttaaaacttcataatagataaaacataatacagataaaagattctgctaatgtctagataaagaactgagaaaagtactcacaatatgttgttgc   3878
agtgtggtttgcaaggcttatataagtctttgtgggtcaccagtgggtgactagaaacattcatttattttattctacaggcatatattcagaagcttgttaactttgtactagttatgagtttatttcttcaaaaacatacatactaaaatacccta   4038

                  Exon 6
aatgtattttccagGGCATAAAAACAGATTCCTTTGTTGGACTAATGGGCAAAAGAGCTTTAAATTCTGgtatgtataaaatcatgtctgaaaatatgtaaatcaacttctgtaattttttacggaaagtaacaatttttaacatactaaaaaaaaaag   4198
             rgHisLysThrAspSerPheValGlyLeuMetGlyLysArgAlaLeuAsnSerV
             100   110

gagccaatagacttaacatgagccatttctgtagacaaatttttcataggagcgttttgtagccaagagtgggcagaaagtcgattaagctgataccaccacaatgtaagtatatttctctcaaagtgccattctgtccagcatctgccccacttagtcact   4358
acatatcctgaatctttggtctagtctgacactctttttcacagatattcttaataaaatgtcctggaccccttagagagagatgagagcttttttacttacaaactgaagtgatgtggtacttttcatcagactaaccatttagtgatttttctactagaaa   4518
aataaatagataatgacaagcaaaatgtttctaaaattactgtctctctctgaagttatggaagctcaaaatccctgtcttaatgggtggcaaagaggctgagtgtagtggtgtccattctaacgccttcaaaccattcttcctgttccagatttagacc   4678
tacatttaaaagtaccacacgccccaaacccatttaaacctaagcgggattttccttggcccagctgcataatctctcatttatcctgcttgaacttctcatataaatgccacagcatatctctaaagccccctcactgttcacctcatttaatcaatta   4838
tattcaatgttaaataatcccattattgttttcatgaaaattggctatcagtgaaaatgggttaaaatgataaacacctgattagaaattaaccagcatctgtttatgatatttagttcacagaaaatttataaagatgtgctggcttaattttaaatat   4998

                                             Exon 7
ctgactcatcccaaaataggtcactctataggtatcttaatcaatgtagacatatttatatttt----IntronF, 2.04 kb----ctgcagTGGCTTATGAAAGAAGCGCAATGCAGAACTACGAAAGAAGGCGTAAATAAACCCTGTAACGCAC   7173
                                                                                          alAlaTyrGluArgSerAlaMetGlnAsnTyrGluArgArgArgLysEnd
                                                                                          120   130

TATCTATTCATCTCCATCTGTGTCCGCGAGCAGTGAGCGGTAAAATAAAAATGTGCGCTATGAGGAATGATTATTTATTTAATATCAAATGTTGTTATGAGTGAAAAACTCAAAAAAGTGTTTATTTTTTCATATTGTGCCAATAAGCATTGTAATTCTA   7333

ATGTGGTGACCTCCTCAGACAGAAGTAGAAATTAGTTGTAACTTCAGCAAAGCACAGTGTTGATGGAGTTGTACAAGTTTGCCAGCGATGCAAGTCTCCAAAGACAGAAAGGCTGCTGTGAGGCAGTGCAGGCGGCTGCTGCTGGAGGCAGAGAAACTCC   7493

TGTGTGTCTTGCGCTTCCCTTGGTTGCTTTTATCCTAATGATGTACTGAGAGTTTGGTATCTGACTCTATTTGTATCCTAGCAGCATGTTTCCTGTGTTGTGACTATATAGAGATGTTTTTAAAAGTTTCAATGTACTTCTCTGGTCTTCAGTCATTGTA   7653

TGATGTGTTGTGATAGCTACCATTTTAAATAAAAGAATGTATCTTCAGGaattttgtcacattctttcattaatatttatgctaaccacaccatctctcattaggatgagtgtgatcaagctgtttgggtaaacacttctatctctctcaggcacagtgca   7813

aattgagtgtgcgcgcaacacacacacacacacacacacacacacacacatacctccacacacattttt   7882

Figure 2. Nucleotide sequence of the rat PPT gene encoding SP and NKA. The *upper case letters* correspond to the sequence of exons 1–7, and the *lower case letters* correspond to the sequence of 5' or 3' flanking, or intervening sequences. The nucleotide sequence is numbered on the right panel with the first nucleotide (+1) of the major transcription initiation site as an adenine base. Other transcription initiation sites are indicated by the *arrowheads,* and appear downstream of the TATA-like regions indicated. The *underlined* sequences in the 5' flanking region correspond to putative regulatory regions, some of which are discussed in the text. The deduced amino acid sequence of β-PPT is shown beneath the appropriate exons. Intron A is designated as that between exon 1 and exon 2. The indicated introns were not completely sequenced, and the approximate number of nucleotides where sequence was not obtained is indicated adjacent to the intron letter designation. The total lengths of the introns are as follows: intron A, 425 bp; intron B ~1010 bp, intron C, 438 bp; intron D, ~460 bp; intron E, 1331 bp; intron F, ~3000 bp.

Figure 3. Determination of the SP gene transcription initiation sites by primer extension and solution hybridization–nuclease protection experiments. *A,* Schematic illustration of the plasmid, pG1PP1 constructed for the protection assays as described in Materials and Methods. The *arrow* indicates the direction of cRNA synthesis using T₇ polymerase and its promoter, which yields a 1.05 kb transcript. The region corresponding to exon 1 is illustrated as the enlarged region upstream from the T₇ promoter. Abbreviations: *H,* Hind III; *P,* PstI; *AS,* antisense; *Ampʳ,* ampicillin-resistance gene; *ori,* origin of replication. *B,* The antisense cRNA produced, and the potential annealing of mRNAs that are produced as a result of transcription from TATA I, TATA II, or TATA III regions. The size of the primary transcript and the protected fragments is indicated above each. *C,* Results of nuclease protection and

quence that mediates some growth factor effects on transcription (Greenberg et al., 1987), is found at nucleotides -206 to -224 at 75% homology. In addition, GC-containing sequences called the GC box are generally adjacent to the TATA-like sequences upstream from the site of transcription initiation (Dynan and Tjian, 1985; Briggs et al., 1986). Thus, GC-rich motifs are located at -100 to -108 bases and at -261 to -270 bases. Also, a CCAAT sequence is present at -177 to -181, which may interact with one of the family of CCAAT binding proteins (Santoro et al., 1988). Potential TATA-like sequences are located at -22 to -28, -88 to -98 and at -196 to -202, which may be sites of RNA polymerase II interaction (see below).

Genomic blot hybridization analysis of the rat SP gene. Genomic hybridization experiments were performed to document the number of SP gene(s). Blot hybridization analysis of rat liver genomic DNA digested with BamHI, EcoRI, and BglII was performed with the full-length β-PPT cDNA and with the 5′ and 3′ β-PPT cDNA fragments as described in Materials and Methods. With the full-length probe, a single hybridizing species was observed with EcoRI-digested liver genomic DNA, whereas 2 hybridizing species were observed with BamHI or BglII digested genomic DNA (data not shown). Using either the 5′ β-PPT cDNA fragment or the 3′ β-PPT cDNA fragment, a single hybridizing species was observed which corresponded to those seen with the full-length β-PPT cDNA, indicating a single gene. Further genomic blot experiments were performed with DNA isolated from a variety of tissues and CNS regions after digestion with EcoRI and BamHI to assess for gene duplication and were compared with results obtained with a single copy gene, β-actin (Nadel et al., 1983). Similar size and intensity autoradiographic signals were observed using the β-PPT cDNA probe with liver, spleen, whole brain, striatum, hypothalamus, trigeminal ganglia, and cerebellum genomic DNA. Based upon these results and the frequency of genomic clones in the library, it seems that a single copy SP gene is present in the rat haploid genome.

Analysis of PPT I gene transcription initiation sites. Primer extension and solution hybridization–nuclease protection analyses were used to determine the transcription initiation site(s) of the rat SP gene. A synthetic oligonucleotide of 20 bases corresponding to the 3′ end of exon 1 (primer A) was used for these experiments. Primer extension was performed with ³²P-primer A, and total RNA was isolated from rat liver, striatum, and hypothalamus. Gel analysis of reaction products (Fig. 3C, right) showed the presence of bands at 100, 125, and 278 nucleotides with striatal and hypothalamic RNA, respectively, but not with liver RNA. The intensity of the 100 base signal was much greater than those observed at 125 and 278 bases, and longer exposure times were required for unequivocal detection of the larger species. A solution hybridization nuclease protection assay was used to further analyze these putative transcription start site(s). A 1050 base antisense cRNA was transcribed from pG1PP1 using T₇ polymerase as shown in Figure 3A. A schematic illustration of pG1PP1 transcript annealing to potential PPT mRNAs initiated is displayed in Figure 3B. The solution hybridization–ribonuclease protection experiment displayed in Figure 3C (left) shows protection by striatal and hypothalamic RNA of a major species at 100 bases and minor species at 125 and 278 bases.

←

primer extension analyses. The sizes of the protected or extended fragments specific to striatum or hypothalamus are indicated on the right. ³²P-labeled MspI fragments of pBR322 were used as size standards.
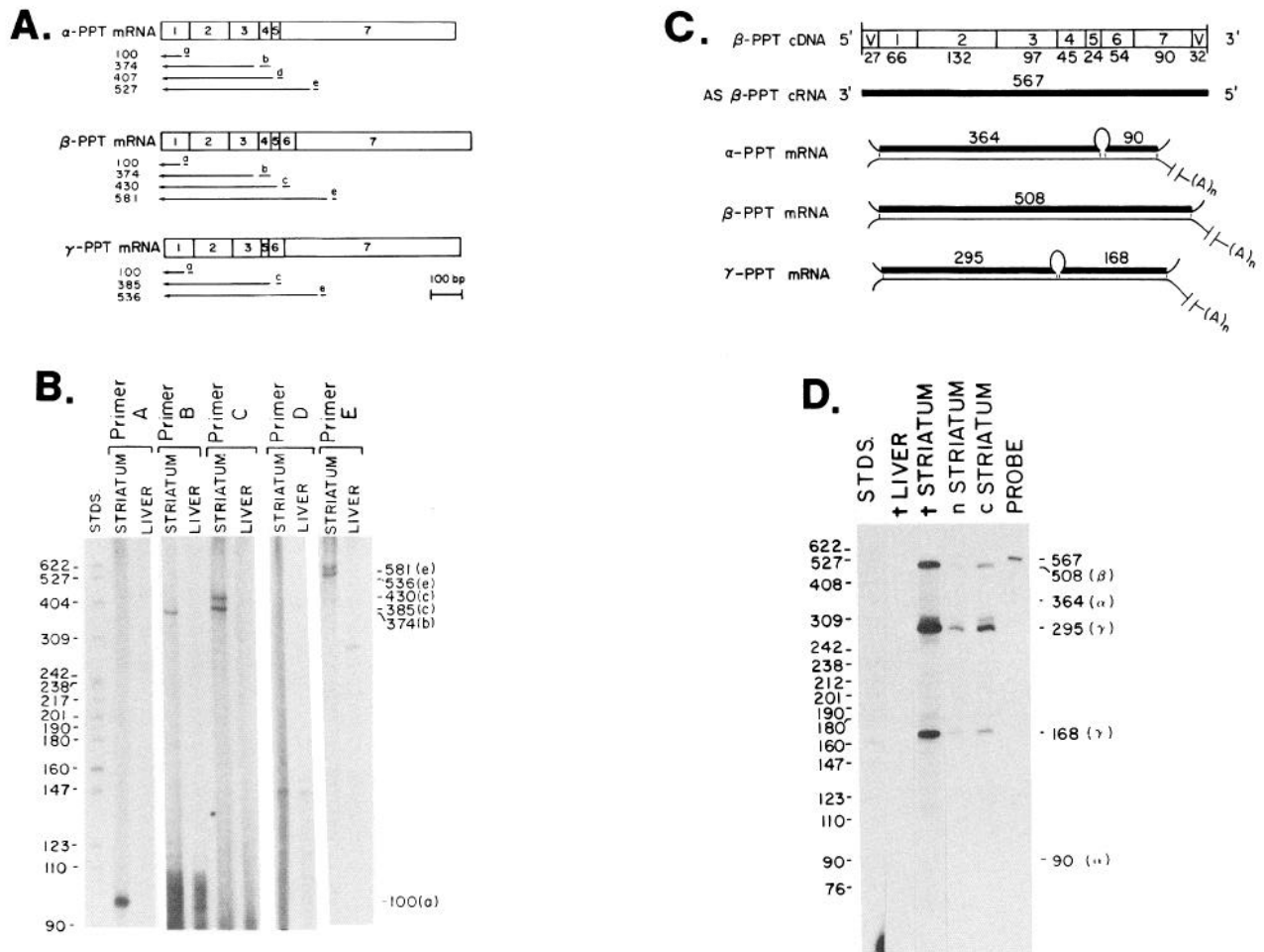
Figure 4. Analysis of striatal SP gene expression by primer extension and nuclease protection experiments. *A,* Schematic of the 3 mRNAs, with exon-corresponding regions (1–7) indicated for α-, β-, and γ-PPT mRNA. Below each mRNA is indicated the site for primer annealing to the mRNAs, and the sizes of the primer elongated transcripts are indicated on the left when the major transcription initiation site is used. The primers correspond to regions of the gene sequence displayed in Figure 2 as follows: *a,* 83–100; *b,* 2197–2241; *c,* 4068–4087; *d,* 2713–2723 colinear with 7110–7118; and *e,* 7224–7238. *B,* Gel analyses of the primer extension experiments performed with total striatal or liver RNA. The sizes of the various primer extended fragments for primers A–E are indicated on the right by extrapolation to the sizes of end-labeled MspI-digested pBR322. *C,* Probe used for nuclease protection experiments, and the annealing of the antisense β-PPT cRNA with α-, β-, and γ-PPT mRNAs. The *numbers within* the cDNA correspond to exons, and the *numbers below* the exons correspond to their length in base pairs. The size of the homologous segments protected after annealing and nuclease digestion is shown above the annealed structures. *V,* vector. *D,* Results from a nuclease protection assay in which RNAs isolated from liver (25 μg), striatum (25 μg), a nuclear fraction of striatum (10 μg), or a cytoplasmic fraction of striatum (10 μg) were used as described in Materials and Methods. An autoradiogram of a 7 M urea–polyacrylamide gel is shown; size standards were end-labeled MspI-digested pBR322. In the cases where 10 μg RNA was used, 15 μg liver RNA was added to achieve a total of 25 μg. An aliquot of probe alone (2.5% of that in the assay) was electrophoresed adjacent to the protected species. *t,* total RNA; *n,* nuclear RNA; and *c,* cytoplasmic RNA.

These species were not observed with liver RNA, whereas a nonspecific band of about 190 bases was present with all 3 RNAs examined. Examination of the gene sequence 5' to exon 1 reveals the presence of TATA-like sequences (TATA I, TATA II, and TATA III; see Fig. 2) at nucleotides −28, −98, and −202. Quantitation of the nuclease protection assay results (see Materials and Methods and Fig. 3C) indicated that >90% of the SP-encoding mRNA transcripts initiated was a result of the TATA I site. A variety of CNS and peripheral tissue RNAs were examined, and no obvious brain region or tissue-specific choices in transcription initiation sites were observed.

Differential splicing of the PPT I gene primary transcript: primer extension and nuclease protection experiments. Figure 4A schematically depicts α-, β-, and γ-PPT mRNAs, with the regions encoded by individual exons displayed numerically (1–7). For

primer extension experiments, various synthetic oligonucleotides (primers A–E) complementary to regions of α-, β-, or γ-PPT mRNA were used, and the sizes of the potential reverse transcription products are displayed below each mRNA illustrated. ^{32}P-primer A resulted in a major species of 100 bases with striatal but not liver RNA (Fig. 4B). Primer B yielded a major species of 374 bases with striatal but not liver RNA. Primer C yielded 2 major products (385 and 430 bases) with striatal RNA, as did primer E (536 and 581 bases), while no products with either primer were observed with liver RNA. Primer D is an oligonucleotide that is complementary with 10 bases on the 3' end of exon 5 and 10 bases on the 5' end of exon 7, i.e., the sequence present in α-PPT mRNA. No detectable specific product was observed with primer D in striatum compared to liver, though 2 nonspecific products were observed with striatal and
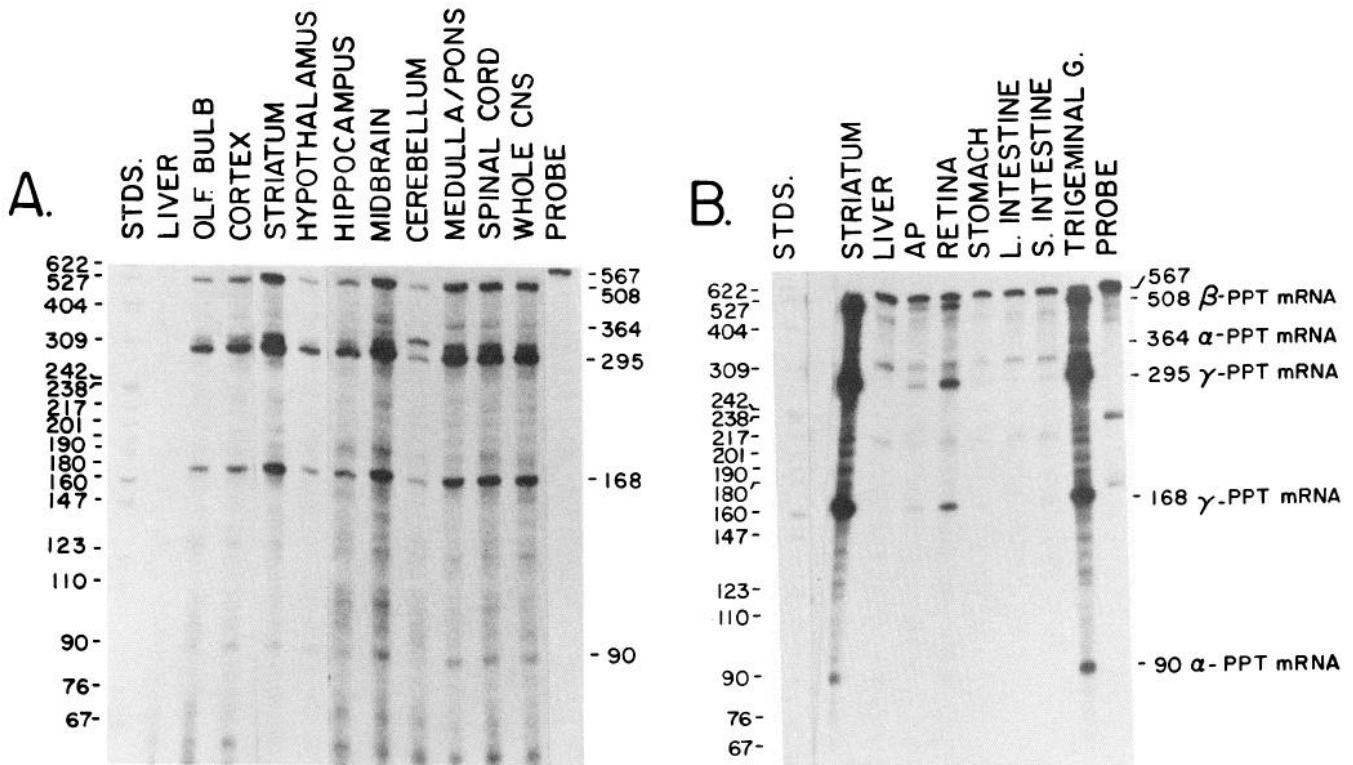
Figure 5. Predominant expression of β- and γ-PPT mRNAs in rat CNS and in peripheral tissues. Nuclease protection assays were performed as described in Figure 5 and in the text. *A,* Results with CNS RNAs. *B,* Results with RNAs isolated from various peripheral tissues. In *A* and *B,* the autoradiographic exposure time for striatal and hypothalamic RNA is 20% of that shown in the other lanes. The probe lane used 2.5% of the total amount of pG1β-PPT cRNA used in the assay. The standards are end-labeled Msp-I-digested pBR322 DNA.

liver RNAs. Taken together, these primer extension experiments indicate the presence of 2 major PPT mRNAs in rat striatum, one lacking and one containing exon 4 sequences. Consequently, we conclude that γ- and β-PPT mRNAs are the most abundant SP-encoding mRNAs in rat striatum.

We previously devised a solution hybridization–nuclease protection assay to quantitate the rat PPT mRNAs (Krause et al., 1989a). This assay consists of the use of a uniformly labeled ³²P-β-PPT cRNA probe, and subsequent digestion of nonannealed nucleic acids with nuclease S₁ or ribonucleases A and T₁. Figure 4C shows the annealing of α-, β-, and γ-PPT mRNAs to the β-PPT antisense cRNA. The hybridization probe is 567 bases, and α-PPT mRNA results in the protection of specific autoradiographic signals at 364 and 90 bases, γ-PPT mRNA protects species of 295 and 168 bases, and β-PPT mRNA protects a species of 508 bases. RNA from striatal nuclear and cytoplasmic subcellular fractions was examined. Figure 4D shows that striatal RNA (unfractionated) results in 3 major protected bands (508, 295, and 168 bases) corresponding to β- and γ-PPT mRNAs, respectively. Bands corresponding to α-PPT mRNAs (364 and 90 bases) are barely detectable at the exposure shown. The same species are also observed with both nuclear and cytoplasmic striatal RNA preparations (Fig. 4D). Based upon densitometric scanning of the autoradiograms, all 3 mRNA species are present at the same relative abundance in both nuclear and cytoplasmic fractions, with γ-PPT mRNA comprising 72% of the total SP-encoding mRNAs, β-PPT mRNA comprising 27%, and α-PPT mRNA comprising 1%.

Tissue distribution of the multiple SP-encoding mRNAs. Nuclease protection experiments were performed with the pG1β-

PPT cRNA probe and various tissue RNAs obtained from various CNS regions, PNS structures, and other tissues. Figure 5A shows the results of a typical experiment performed with RNA isolated from various CNS regions. In all CNS regions where the SP gene is expressed, major autoradiographic signals were observed at 295 and 168 bases, respectively, corresponding to the γ-PPT mRNA. The autoradiographic signal at 508 bases is also evident in all tissues where the gene is expressed and corresponds to the β-PPT mRNA species and is much stronger than the signals at 364 and 90 bases, respectively, which correspond to α-PPT mRNA. Figure 5B displays the results of a similar experiment, where RNA isolated from several gastrointestinal tissues, anterior pituitary, retina, and trigeminal ganglia was used. The same patterns of SP-encoding RNAs are observed in these tissues, though the abundance level of the mRNAs in the gastrointestinal tissues and anterior pituitary is much lower than in the various CNS tissues. It should be pointed out that the α-PPT mRNA species is not detectable in RNA isolated from the anterior pituitary, even upon long exposure times. The results from multiple experiments performed with at least 2 independent RNA preparations from each tissue are summarized in Table 1. From this analysis, it is clear that the γ-PPT mRNA comprises from 65 to 90% of the total SP-encoding mRNA in all tissues, and the β-PPT comprises from 9 to 34% of the total SP-encoding mRNA. The α-PPT mRNA species is the least abundant, comprising ≤1% of the total SP-encoding mRNA.

Potential mechanism(s) of PPT mRNA differential splicing. Because of the major differences observed in SP precursor RNA splicing between rat and bovine species (Nawa et al., 1984;

Table 1. Expression of α-, β-, and γ-PPT mRNAs in rat CNS and peripheral tissues

Tissue	Relative abundance	mRNA type (%)		
		α	β	γ
CNS				
Olfactory bulb	0.03	1	21	78
Cortex	0.05	<1	27	73
Striatum	1.00	1	27	72
Hypothalamus	0.29	<1	25	75
Hippocampus	0.03	<1	20	80
Midbrain	0.13	<1	24	76
Cerebellum	0.01	<1	20	79
Medulla/pons	0.09	<1	25	75
Spinal cord	0.09	<1	22	78
Other tissue				
Kidney	ND	—	—	—
Liver	ND	—	—	—
Anterior pituitary	0.01	ND	26	74
Retina	0.05	<1	20	79
Stomach	0.01	<1	34	65
Large intestine	0.01	<1	10	89
Small intestine	0.01	<1	9	90
Trigeminal ganglia	0.71	<1	27	72
Thyroid	ND	—	—	—

Results represent the mean values compiled from at least 3 nuclease protection experiments using RNA from 2 different preparations. The relative abundance of SP-encoding mRNAs was determined from the sum of the autoradiographic signals corresponding to α-, β-, and γ-PPT mRNAs compared to the levels present in rat striatum (see Fig. 6). In all cases, the variation observed for the relative abundance values and the percent mRNA type was less than 20% of the mean value reported. ND, not detectable.

Krause et al., 1987; Figs. 4 and 5 above), we compared the SP gene donor and acceptor splice sites from these species (Table 2). Rat and bovine SP gene donor splice sites are identical except for that region adjacent to the 1st exon, which displays a 50% identity between the 2 species but is not alternatively utilized in precursor RNA splicing. The acceptor splice sites show more variability between rat and bovine species (from 56 to 88% homology where the sequence has been fully determined), yet those 3' splice sites that are adjacent to the differentially spliced exons (i.e., exons 4 and 6) show 88 and 94% homology, respectively. Both donor and acceptor splice sites conform to the consensus sequences compiled by Padgett et al. (1986). Thus, it appears that differences in donor and acceptor splice sites cannot account for the differences in PPT precursor RNA splicing observed between rat and bovine.

Regions throughout the SP gene were examined for stem and loop structures according to the rules of Tinoco et al. (1973) in order to identify potential regions of intrastrand base pairing that may affect splicing. Among the various regions of the SP gene primary transcript, 2 regions adjacent to the differentially excluded exons (i.e., exons 4 and 6) showed the potential of stem and loop structures in which the regions were A/T-rich and at times displayed negative ΔG values. Other regions of the SP gene primary transcript displayed the potential for stem and loop formation, but in these few cases they were short sequences with a relatively high G-C content (see below and Discussion). Figure 6A displays hypothetical stem and loop structures that form with the SP primary transcript adjacent to exon 4 and thus may result in the exclusion of exon 4 in the splicing process, resulting in the formation of γ-PPT mRNA. Two intronic regions could potentially anneal with regions of exon 4: a region

Table 2. Comparison of rat and bovine SP precursor RNA donor and acceptor splice sites

Species	Exon	Donor splice site		Intron	Acceptor splice site		Exon
		5' Splice site	% Homology		3' Splice site	% Homology	
		NAG/GUAAGUNN[a]			(Py)₁₁NCAG/GN[a]		
Rat	1	GCA/GUGAGUAC	50	A	CUCCUACUCUCUCCAG/AA	75	2
		‖ ‖ ‖			‖‖‖‖‖‖‖‖ ‖‖‖‖		
Bovine		GCG/GUAAUGGC			UCCCUACUCUUCCAG/AA		
Rat	2	AAG/GUGAGACC	100	B	UGGUUUGUCCCUCCAG/GA	56	3
		‖‖‖‖‖‖‖‖‖			‖‖ ‖‖ ‖‖‖‖ ‖‖		
Bovine		AAG/GUGAGACC			UGUFUUUGUCCUCUAG/GA		
Rat	3	CUG/GUGAGAUG	100	C	AUUCAUCUCUUGUCAG/AU	88	4
		‖‖‖‖‖‖‖‖‖			‖‖‖ ‖‖‖‖‖‖‖‖‖‖		
Bovine		CUG/GUGAGAUG			AUUUGUCUCUUGUCAG/AU		
Rat	4	AUG/GUAAACAU	100	D	CUCUCUUUUCUUCUAG/GG	63	5
		‖‖‖‖‖‖‖‖‖			‖‖‖ ‖ ‖‖ ‖ ‖‖‖		
Bovine		AUG/GUAAACAU			AUCUUUCUUUCUAUAG/GA		
Rat	5	AAA/GUAAGUUC	100	E	UAAAUGUAUUUUCCAG/GG	63 (94)	6
		‖‖‖‖‖‖‖‖‖			‖‖ ‖‖‖‖‖‖‖‖		
Bovine		AAA/GUAAGUUC			AAAUGUAUUUUCCAG/GG		
Rat	6	CUG/GUAUGUAU	100	F	CUGCAG/UG		7
		‖‖‖‖‖‖‖‖‖			‖‖‖		
Bovine		CUG/GUAUGUAU			UCUCAG/UG		

[a] The consensus donor and acceptor splice sites from from Padgett et al. (1986) and the sequences of the bovine splice sites are from Nawa et al. (1984). The vertical lines between the rat and bovine sequences indicate nucleotide identity. The acceptor splice site adjacent to exon 6 displays a 63% homology between rat and bovine; however, the absence of an A nucleotide in the bovine at the −9 position would result in a homology of 94%.

in intron C could anneal near the 5' end of exon 4 [10/10 match; A-U-rich (7/12); $\Delta G = -8.4$ kcal/mol], and a region in intron D could anneal with the 3' end of exon 4 [9/12 match; A-U-rich (8/10); $\Delta G = 1.6$ kcal/mol]. Figure 6*B* displays hypothetical stem and loop structures that may form with the SP primary transcript adjacent to exon 6 and thus may result in the exclusion of exon 6 in the splicing process, which would result in the formation of α-PPT mRNA. Two exonic regions could potentially anneal with sequences present in exon 6: a region in intron E that includes part of the 5' (donor) splice site could anneal near the 5' end of exon 6 [8/10 match; A-U rich (8/10); $\Delta G = +4.4$], and another region of intron E could anneal with the 3' end of exon 6 and part of the 5' end of intron F [12/14 match; A-U-rich (10/14); $\Delta G = -6.8$].

Discussion

In the current studies we have screened a rat λ Charon 4A genomic library using a rat β-PPT cDNA as a hybridization probe and have isolated the rat gene that encodes the tachykinin peptides SP, NKA, NPK, and NPγ. The rat SP gene is comprised of 7 exons and is approximately 8 kb in length. SP is encoded on part of exon 3, and NKA is encoded by exon 6. Exons 4 and 6 of this PPT gene are separately excluded in the splicing process, thus resulting in the formation of γ-PPT mRNA and α-PPT mRNA, respectively. The complete sequence of the 7 exons, 3 introns (A, C, and E), and 866 bases of 5' flanking sequence have been determined, as well as the partial sequences of the remaining introns (B, D, and F) and 180 bases of 3' flanking sequence. Both genomic clone isolation and analysis, and Southern blot analysis indicate that the gene responsible for SP and NKA production is single copy in the rat haploid genome. The structural organization of the rat gene is similar to that of the bovine (Nawa et al., 1984); however, many features of expression, including the pattern of precursor RNA splicing, are distinctly different.

The control of gene expression can occur via interaction of regulatory protein(s) with specific DNA sequences existent upstream from the site of transcription initiation. Based upon primer extension analyses and nuclease protection experiments with striatal RNA, we conclude that 3 transcription initiation sites are utilized for transcription of the rat SP/NKA gene. The major transcription initiation site is 21 bases downstream from the sequence TAAATAA; thus, this site appears to be a site of interaction of RNA polymerase II for transcription initiation (Nevins, 1983). Nuclease protection and primer extension studies revealed 2 additional minor sites of transcription initiation in the striatum and hypothalamus; these are located 25 and 178 bases upstream from the major transcription initiation site. The nuclease protection assay results showed these minor sites more clearly than the primer extension experiments, presumably due to the greater annealing efficiency of the former technique (Krause et al., 1989a) and the higher specific activity of the uniformly labeled cRNA probe compared to the end-labeled oligonucleotide. The first nucleotide of all transcripts is A, which is generally the preferred nucleotide for capping in eukaryotes (Breathnach and Chambon, 1981). If 5' flanking regions of the SP gene are involved in transcriptional regulatory mechanisms, they may likely be defined by a determination of sequence conservation between the rat and cow. Such a comparison of the SP gene indicates an overall homology of 82% from the rat major transcription initiation site extending 346 bases upstream. Aside
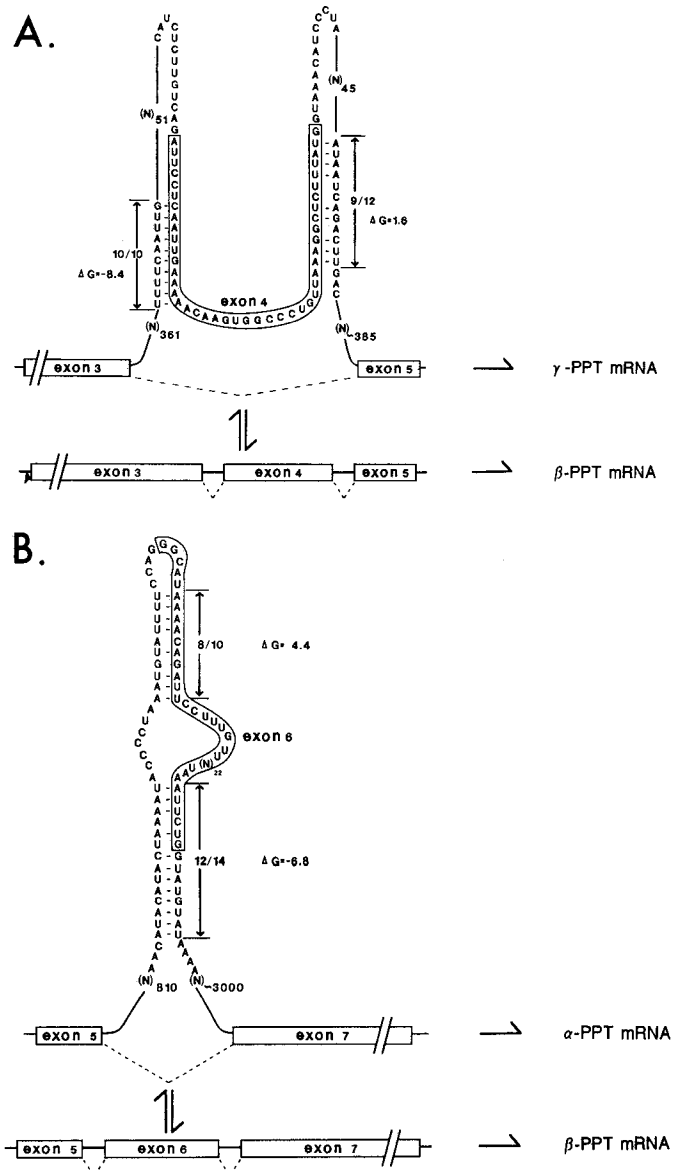
Figure 6. Schematic illustration of potential RNA secondary structures formed by intrastrand annealing of the primary PPT transcript. *A*, A potential structure in which exon 4 may undergo intrastrand annealing which could prevent inclusion of this exon, thus resulting in the formation of γ-PPT mRNA. The ΔG values were calculated using the method of Tinoco et al. (1973). *B*, A potential structure in which exon 6 may undergo intrastrand annealing which could prevent inclusion of this exon, thus resulting in the formation of α-PPT in RNA. The exon regions are enclosed in a *box*, and the sites of potential annealing are displayed.

from the rat, no further 5' genomic sequences have been reported for the bovine (Nawa et al., 1984) or any other species. Therefore, these highly conserved sequences identified here may represent regulatory elements involved in the regulation of SP gene expression. Many additional regions in the 5' flanking area display homology to established consensus sequences of eukaryotic transcription factor binding sites. These putative regulatory sites are considered to be tentative and require functional testing. Some of these sequences in the rat SP gene include those homologous to DNA response elements for estrogen (Maurer and Notides, 1987) and serum factor(s) (Triesman, 1986; Greenberg
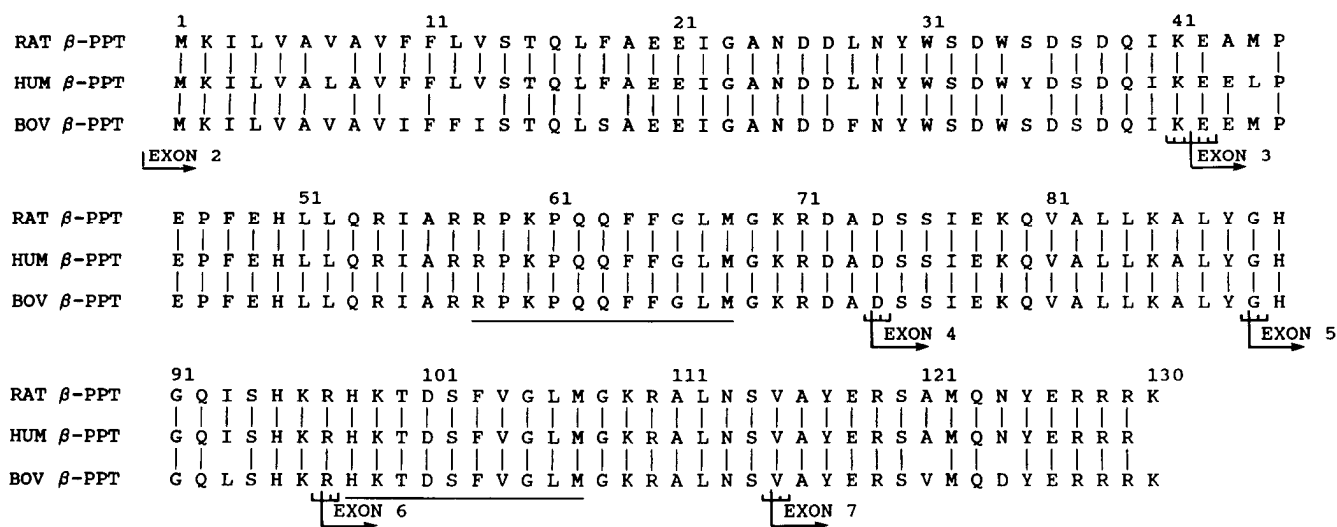
```
             1                 11                21                31                41
RAT β-PPT    M K I L V A V A V F F L V S T Q L F A E E I G A N D D L N Y W S D W S D S D Q I K E A M P
HUM β-PPT    M K I L V A L A V F F L V S T Q L F A E E I G A N D D L N Y W S D W Y D S D Q I K E E L P
BOV β-PPT    M K I L V A V A V I F F I S T Q L S A E E I G A N D D F N Y W S D W S D S D Q I K E E M P
             |EXON 2                                                                       |EXON 3
```

```
                   51                61                71                81
RAT β-PPT    E P F E H L L Q R I A R R P K P Q Q F F G L M G K R D A D S S I E K Q V A L L K A L Y G H
HUM β-PPT    E P F E H L L Q R I A R R P K P Q Q F F G L M G K R D A D S S I E K Q V A L L K A L Y G H
BOV β-PPT    E P F E H L L Q R I A R R P K P Q Q F F G L M G K R D A D S S I E K Q V A L L K A L Y G H
                                                              |EXON 4                     |EXON 5
```

```
             91                101               111               121               130
RAT β-PPT    G Q I S H K R H K T D S F V G L M G K R A L N S V A Y E R S A M Q N Y E R R R K
HUM β-PPT    G Q I S H K R H K T D S F V G L M G K R A L N S V A Y E R S A M Q N Y E R R R
BOV β-PPT    G Q L S H K R H K T D S F V G L M G K R A L N S V A Y E R S V M Q D Y E R R R K
                   |EXON 6                               |EXON 7
```

Figure 7. Structural comparison of β-PPT from rat, bovine, and human species. The primary structures of the 3 β-PPTs are from the results of Nawa et al. (1983), Krause et al. (1987), and Harmar et al. (1986) and are displayed with the regions encoded by the various exons indicated below the sequence. The residues are numbered above the sequence and the regions corresponding to SP and NKA are *underlined*. Note that the 3′ end of exons 3–6 split codons after the first base. Residues displaying identity are indicated by the *vertical line*.

et al., 1987); thus, these hormonal systems may be involved in regulating the expression of the SP gene. In fact, evidence has been presented for a role of nerve growth factor (Kessler and Black, 1981; Lindsay and Harmar, 1989) and estrogen (Aronin et al., 1986; Brown et al., 1990) in the regulation of SP gene expression. A second class of control nucleotide sequences, including the so-called CAAT box and GC box, the binding sites for the CAAT-binding transcription factor(s) and Sp1 transcription factor are also located in the 5′ flanking region of the SP gene. The GC box sequence has been shown to be responsible for maintenance of transcriptional efficiency (Lee et al., 1987), whereas the CAAT sequence may interact with multiple transcription factors (Briggs et al., 1986).

Previously we had demonstrated by cDNA cloning and nuclease protection experiments using end-labeled cDNA probes that 3 SP-encoding mRNAs are derived from the rat SP gene by differential RNA splicing. We have confirmed and extended these results by further nuclease protection studies using a uniformly labeled cRNA probe and by primer extension experiments using oligonucleotides that anneal with RNA sequences corresponding to individual exon-encoded regions of SP-encoding mRNAs. These and previous results (Krause et al., 1987, 1989a) demonstrate the existence of α-, β-, and γ-PPT mRNAs in rat striatum, and a quantitative analysis of these results indicates that γ-PPT mRNA > β-PPT mRNA ≫ α-PPT mRNA at a ratio of approximately 80:20:<1 (see Table 1). Bannon and coworkers (Haverstick et al., 1989) have also provided evidence recently that γ- and β-PPT mRNAs are the most abundant SP-encoding mRNAs in rat striatum. Since this ratio of mRNAs exists in all rat tissues where the SP gene is expressed, tissue-specific splicing of the SP gene primary transcript does not occur in the adult rat. The bovine SP gene primary transcript has previously been shown to be differentially spliced in a tissue-specific fashion (Nawa et al., 1983, 1984). Thus, bovine α-PPT mRNA was more abundant in CNS tissue compared to β-PPT mRNA (approximately 1 mol β-PPT mRNA for 3 mol α-PPT mRNA), whereas β-PPT mRNA was more abundant than α-PPT

mRNA in gut and thyroid tissues. An mRNA species encoding γ-PPT was not identified in the bovine and consequently was not quantitated. Therefore, the pattern of splicing is substantially different in rat compared to bovine (Nawa et al., 1983, 1984), and consequently the mechanism(s), function(s), and/or consequence(s) of differential splicing of the SP gene primary transcript may be species-specific to a certain extent.

The biochemical mechanism(s) of differential precursor RNA splicing in the production of the multiple SP-encoding mRNAs is unknown. The α-PPT mRNA transcript lacks the exon 6-encoded sequence, whereas the γ-PPT mRNA transcript lacks the exon 4-encoded sequence. This process of exon exclusion may arise as a result of base-pairing interactions between sequences of introns adjacent to the exons that are to be excluded. Moreover, specific protein factors may promote or hinder these potential base-pairing interactions. It has been shown that RNA secondary structure can play a role in splice site selection. For example, splice sites sequestered in genetically engineered RNA hairpins can be omitted in the splicing process (Solnick, 1985; Eperon et al., 1986), and it appears that loop length is an important factor in splice site selection (Solnick and Lee, 1987; Eperon et al., 1988). With the SP gene primary transcript (e.g., as shown in Fig. 6), sequences in intron C and D could potentially interact with exon 4 sequences, resulting in the exclusion of exon 4 in the mRNA formed. Similarly, sequences in intron E could potentially interact with exon 6 sequences, resulting in the exclusion of exon 6 in the mRNA formed. These potential base-pairing regions are short, are A-U-rich (58–80%), and generally have negative ΔG values. No other base-pairing interactions with the same characteristics were found in the SP gene sequence analyzed. Most biologically important RNA base-pairing interactions are short, such as tRNA (Abelson, 1979), snRNA (Guthrie and Patterson, 1988), RNase P RNA (James et al., 1988), and Tetrahymena IVS RNA (Cech et al., 1983). Also, there is precedent for short A-U-rich helices as well (see, for example, the 9R and 9R′ interaction in the Tetrahymena IVS; Cech et al., 1983). Thus, the proposed base-pairing interactions

of the SP gene primary transcript provides a testable hypothesis by mutational analysis for the role of these sequences in differential RNA splicing.

The α- and β-PPT mRNAs are present in the cytoplasm of neurons as determined by subcellular fractionation with subsequent isolation and analysis of cytoplasmic and nuclear RNAs. Consequently, all 3 PPT mRNAs should be available to the cellular translation machinery, and it has been demonstrated previously that α-, β-, and γ-PPT mRNA are efficiently translated *in vitro* (MacDonald et al., 1988). α-, β-, and γ-PPT enter the secretory pathway as a result of their signal peptide-induced import into the lumen of the endoplasmic reticulum (MacDonald et al., 1988).

If differential splicing of the SP/NKA gene in rat tissues is a stochastic process with γ- and β-PPT mRNAs as the most abundant species, what is the functional significance of the differential RNA splicing? As shown in Figure 7, the entire β-PPT precursor between cow (Nawa et al., 1983), rat (Krause et al., 1987), and human (Harmar et al., 1986) is highly conserved (from 91 to 96% residue identity), and even the precursor regions excluding the SP and NKA portions are substantially conserved (from 89 to 96% residue identity). The exon 4-encoded sequence is identical in the 3 species, whereas the exon 5-encoded sequence is identical except for a conservative isoleucine-to-leucine change in the bovine species. This high degree of homology may indicate that the peptides encoded therein may be functionally significant. Other neuropeptide polyprotein precursors do not necessarily exhibit such a high degree of conservation throughout their sequence (e.g., see the interspecies comparison of pre-proenkephalin sequences by Yoshikawa et al., 1984, and Rosen et al., 1984). The polyprotein precursors encoding γ- and β-PPT can be differentially processed posttranslationally such that NKA and/or NPγ can be produced from γ-PPT (Kage et al., 1988; MacDonald et al., 1989) and that NKA and/or NPK can be produced from β-PPT (Tatemoto et al., 1985; MacDonald et al., 1989). Also, both NPγ and NPK can be secreted from cells by secretogogue stimulation (MacDonald et al., 1989). The biological activities of NPγ and NPK are not entirely worked out, but these peptides are potent on certain tachykinin-sensitive end points and in some cases are more potent than NKA. For example, NPγ is potent in rat salivary secretion (Takeda and Krause, 1989a), and NPK is potent on plasma extravasation (Tatemoto et al., 1985) and on salivary secretion (Takeda and Krause, 1989b), and both are much more potent than NKA. Recent studies have established that NPK (Beaujouan et al., 1988) and NPγ (Dam et al., 1989; Takeda and Krause, 1989c) can be more potent ligands at certain tachykinin NK-2 receptor sites compared to NKA, and it may be that multiple types of tachykinin NK-2 receptors exist with differential affinities for NKA, NPγ, and NPK. We suggest that the function of differential splicing of the rat SP gene primary transcript is for production and secretion of SP and multiple NKA-related peptides.

In summary, we have characterized the rat PPT gene encoding SP, NKA, NPγ, and NPK by restriction mapping and nucleotide sequence analysis, and we have analyzed various aspects of gene expression. This report forms the basis for detailed studies on the mechanisms underlying expression of the rat gene. We have also provided detailed evidence that the pattern of differential RNA splicing of this SP gene primary transcript is regulated quite differently from that of the bovine (Nawa et al., 1983, 1984). Since the γ- and β-PPT mRNA comprises essentially all of the SP-encoding mRNA in all tissues of the adult rat where the gene is expressed, it appears likely that a single cell expresses both γ- and β-PPT mRNA.

References

Abelson J (1979) RNA processing and the intervening sequence problem. Annu Rev Biochem 48:1035–1069.

Arai H, Emson PC (1986) Regional distribution of neuropeptide K and other tachykinins (neurokinin A, neurokinin B and substance P) in rat central nervous system. Brain Res 399:240–249.

Aronin N, Coslovsky R, Leeman SE (1986) Substance P and neurotensin: their roles in the regulation of anterior pituitary function. Annu Rev Physiol 48:537–549.

Beaujouan J-C, Saffroy M, Petitet F, Torrens Y, Glowinski J (1988) Neuropeptide K, scyliorhinin I and II: new tools in the tachykinin receptor field. Eur J Pharmacol 151:353–354.

Bonner T, Affolter HU, Young AC, Young WS (1987) A cDNA encoding the precursor of the rat neuropeptide neurokinin B. Mol Brain Res 2:243–249.

Breathnach R, Chambon P (1981) Organization and expression of eucaryotic split genes coding proteins. Annu Rev Biochem 50:349–383.

Briggs MR, Kadonga JT, Bell SP, Tjian R (1986) Purification and biochemical characterization of the promoter-specific transcription factor, SP1. Science 234:47–52.

Brown ER, Harlan RE, Krause JE (1990) Gonadal steroid regulation of substance P and substance P-encoding messenger ribonucleic acids in the rat anterior pituitary and hypothalamus. Endocrinology 126:330–340.

Cech TR, Tanner NK, Tinoco I, Weir BR, Zuken M, Peckman PS (1983) Secondary structure of the Tetrahymena ribosomal RNA intervening sequence: structural homology with fungal mitochondrial intervening sequences. Proc Natl Acad Sci USA 80:3903–3909.

Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294–5299.

Dam T-V, Takeda Y, Krause JE, Escher E, Quirion R (1989) γ-Preprotachykinin-(72–92)-peptide amide: an endogenous preprotachykinin I gene derived peptide which preferentially binds to neurokinin-2 receptors. Proc Natl Acad Sci USA 87:246–250.

Dynan WS, Tjian R (1985) Control of eukaryotic messenger RNA synthesis by sequence specific DNA-binding proteins. Nature 316:774–778.

Eperon LP, Estibeiro JP, Eperon IC (1986) The role of nucleotide sequences in splice site selection in eucaryotic pre-messenger RNA. Nature 324:280–282.

Eperon L, Graham IR, Griffiths AD, Eperon IC (1988) Effect of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase. Cell 54:393–401.

Gardner DG, Vlasuk GP, Baxter JD, Fiddes JC, Lewicki JA (1987) Identification of atrial natriuretic factor gene transcripts in the central nervous system of the rat. Proc Natl Acad Sci USA 84:2175–2179.

Ghosh PK, Reddy VB, Piatak M, Lebowitz P, Weissman SM (1980) Determination of RNA sequences by primer directed synthesis and sequencing of their cDNA transcripts. Methods Enzymol 65:580–595.

Glowinski J, Iversen LL (1966) Regional studies of catecholamines in the rat brain: the disposition of [³H] norepinephrine, [³H] dopamine and [³H]DOPA in various regions of the brain. J Neurochem 13:655–669.

Greenberg ME, Siegfried Z, Ziff EB (1987) Mutation of the c-fos gene dyad symmetry element inhibits serum inducibility of transcription *in vivo* and the nuclear regulatory factor binding in vitro. Mol Cell Biol 7:1217–1225.

Guthrie C, Patterson B (1988) Spliceosomal snRNAs. Annu Rev Genet 22:387–419.

Harmar AJ, Armstrong A, Pascall JC, Chapman K, Rosie R, Curtis A, Going J, Edwards CRW, Fink G (1986) cDNA sequence of human β-preprotachykinin, the common precursor to substance P and neurokinin A. FEBS Lett 208:67–72.

Haverstick DM, Rubenstein A, Bannon MJ (1989) Striatal tachykinin gene expression regulated by interaction of D-1 and D-2 dopamine receptors. J Pharmacol Exp Ther 248:858–862.

Henikoff S (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351–359.

James BD, Olsen GJ, Liu J, Pace NR (1988) The secondary structure of ribonuclease P RNA, the catalytic element of ribonucleoprotein enzyme. Cell 52:19–26.

Kage R, McGregor GP, Thim L, Conlon JM (1988) Neuropeptide γ: a peptide isolated from rabbit intestine that is derived from γ-pre-protachykinin. J Neurochem 50:1412–1417.

Kessler JA, Black IB (1981) Nerve growth factor stimulates development of substance P in the embryonic spinal cord. Brain Res 208: 135–145.

Kotani H, Hoshimaru M, Nawa H, Nakanishi S (1986) Structure and gene organization of bovine neuromedin K precursor. Proc Natl Acad Sci USA 83:7074–7078.

Krause JE, Karavolas HJ (1980) Pituitary 5α-dihydroprogesterone 3α-hydroxysteroid oxidoreductases. Subcellular location and properties of NADH- and NADPH-linked activities. J Biol Chem 255:11807–11814.

Krause JE, Chirgwin JM, Carter MS, Xu ZS, Hershey AD (1987) Three rat preprotachykinin mRNAs encode the neuropeptides substance P and neurokinin A. Proc Natl Acad Sci USA 84:881–885.

Krause JE, Cremins JD, Carter MS, Brown ER, MacDonald MR (1989a) Solution hybridization-nuclease protection assays for sensitive detection of differentially spliced substance P- and neurokinin A-encoding messenger ribonucleic acids. Methods Enzymol 168:634–652.

Krause JE, MacDonald MR, Takeda Y (1989b) The polyprotein nature of substance P precursors. BioEssays 10:62–69.

Lee W, Haslinger A, Karin M, Tjian R (1987) Two factors that bind and activate the human metallothionein II_A gene in vitro also interact with the SV40 promoter and enhancer regions. Nature 325:368–372.

Lindsay RM, Harmar AJ (1989) Nerve growth factor regulates expression of neuropeptide genes in adult sensory neurons. Nature 337:362–364.

MacDonald MR, McCourt DW, Krause JE (1988) Posttranslational processing of α-, β- and γ-preprotachykinins. Cell-free translation and early posttranslational processing events. J Biol Chem 263:15176–15183.

MacDonald MR, Takeda J, Rice CM, Krause JE (1989) Multiple tachykinins are produced and secreted upon post-translational processing of the three substance P precursor proteins, α-, β- and γ-preprotachykinin. Expression of the preprotachykinins in AtT-20 cells infected with vaccinia virus recombinants. J Biol Chem 264:15578–15592.

Maggio JE (1988) Tachykinins. Annu Rev Neurosci 11:13–28.

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual, pp 88–94. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Maurer RA, Notides AC (1987) Identification of an estrogen-responsive element from the 5'-flanking region of the rat prolactin gene. Mol Cell Biol 7:4247–4254.

Nadel U, Zakut R, Shami M, Neuman S, Levy Z, Yaffe D (1983) The nucleotide sequence of the rat cytoplasmic β-actin gene. Nucleic Acids Res 11:1759–1791.

Nawa H, Hirose T, Takashima H, Inayama S, Nakanishi S (1983) Nucleotide sequences of cloned cDNAs for two types of bovine substance P precursor. Nature 306:32–36.

Nawa H, Kotani H, Nakanishi S (1984) Tissue specific generation of two preprotachykinin mRNAs from one gene by alternative RNA splicing. Nature 312:729–734.

Nevins JR (1983) The pathway of eukaryotic mRNA formation. Annu Rev Biochem 52:441–466.

Padgett RA, Grabowski PJ, Konarska MM, Seiler S, Sharp PA (1986) Splicing of messenger RNA precursors. Annu Rev Biochem 55:1119–1150.

Pernow B (1983) Substance P. Pharmacol Rev 35:85–141.

Queen C, Korn LJ (1984) A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res 12:581–599.

Roesler WJ, Vandenback GR, Hanson RW (1988) Cyclic AMP and the induction of eucaryotic gene transcription. J Biol Chem 263:9063–9066.

Rosen H, Douglass J, Herbert E (1984) Isolation and characterization of the rat preproenkephalin gene. J Biol Chem 259:14309–14313.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467.

Santoro C, Mermod N, Andrews PC, Tjian R (1988) A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. Nature 334: 218–224.

Sargent TD, Wu J-R, Sala-Trepat JM, Wallace RB, Reyes AA, Bonner J (1979) The rat serum albumin gene: analysis of clones sequences. Proc Natl Acad Sci USA 76:3256–3260.

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

Solnick D, Lee SI (1987) Amount of RNA secondary structure required to induce a secondary splice. Mol Cell Biol 7:3194–3198.

Takeda Y, Krause JE (1989a) γ-Preprotachykinin-(72–92)-peptide amide potentiates substance P-induced salivation. Eur J Pharmacol 161:267–271.

Takeda Y, Krause JE (1989b) Neuropeptide K potently stimulates salivary gland secretion and potentiates substance P-induced salivation. Proc Natl Acad Sci USA 86:392–396.

Takeda Y, Krause JE (1989c) Characterization of a high affinity neuropeptide γ binding site in rat duodenum, urinary bladder and vas deferens. Soc Neurosci Abstr 15:346.

Takeda Y, Takeda J, Smart BM, Krause JE (1990) Regional distribution of neuropeptide γ and other tachykinin peptides derived from the substance P gene in the rat. Regul. Peptides (in press).

Tatemoto K, Lundberg JM, Jörnvall H, Mutt V (1985) Neuropeptide K: isolation, structure and biological activities of a novel brain tachykinin. Biochem Biophys Res Commun 128:947–953.

Tinoco U, Borer PN, Dengler B, Levine MD, Uhlenbeck OC, Crothers DM, Gralla J (1973) Improved estimation of secondary structure in nucleic acids. Nature [New Biol] 246:40–41.

Triesman R (1986) Identification of a protein binding site that mediates transcriptional response of the c-fos gene to serum factors. Cell 46:567–574.

Yoshikawa K, Williams C, Sabol SL (1984) Rat brain preproenkephalin mRNA, cDNA cloning, primary structure and distribution in the central nervous system. J Biol Chem 259:14301–14308.