

A Neural-Specific GAP-43 Core Promoter Located between Unusual DNA Elements that Interact to Regulate Its Activity

Elly Nedivi,^a Guriqbal S. Basi,^b Ildikó Virág Akey,^c and J. H. Pate Skene

Department of Neurobiology, Stanford University, Stanford, California 94305

In an effort to identify *cis*-acting elements that respond to signals controlling different stages of neural differentiation, we have analyzed the promoter and surrounding regulatory sequences of the rat GAP-43 gene. Expression of this gene is both neural specific and, within neurons, strongly modulated by signals related to axon integrity. Expression analysis in cell lines and primary rat cortical cultures demonstrates that neural-selective gene expression can be directed by a 386 base pair GAP-43 promoter fragment that contains canonical TATA and CCAAT box consensus sequences. A short region of homology with other neural-specific genes, identified upstream of the core promoter, is not essential for selective expression in neuronal cells. Within cortical cell cultures, expression is strongly modulated by two interacting elements on either side of the promoter, each of which contains a sequence with the potential to adopt an unusual DNA conformation. While each of these flanking elements reduces expression when added alone to the core promoter, each counteracts the negative influence of the other when both elements are present.

Neural differentiation is marked by a series of steps leading from the initial induction of stem cells in the neural epithelium to the emergence of mature neurons, glia, or neuroendocrine cells (Anderson, 1990; McKay, 1989). Genes whose expression is restricted to neural cells may be induced at different steps in these differentiation pathways, presumably in response to distinct environmental or intracellular signals. One strategy for unraveling the control of these distinct differentiation steps is to identify *cis*-acting elements that link expression of particular genes to specific stages in neural differentiation.

Of the few genes that have been dissected to identify *cis*-acting elements mediating neural-specific expression, most are char-

acteristically expressed at highest levels in the mature nervous system (Sehgal et al., 1988; Maue et al., 1990; Sauerwald et al., 1990). In contrast, several genes encoding protein components of growth cones and immature neurites are also highly specific but are expressed at highest levels in immature rather than mature neurons (Miller et al., 1987; Hoffman and Cleveland, 1988; Stein et al., 1988; Skene, 1989). Such genes must contain regulatory elements that allow them to be induced selectively at an appropriate stage of neural differentiation and then to respond to signals that strongly modulate their expression within differentiated neurons.

A well-characterized example of a gene expressed in almost all neurons during a discrete phase of development is the single copy gene for an abundant growth cone protein, GAP-43. Expression of GAP-43 occurs exclusively in neural cells, and most prominently in immature neurons (Basi et al., 1987; Karns et al., 1987; Neve et al., 1987; Rosenthal et al., 1987). It may also be expressed in glial cells under certain circumstances (Vitkovic et al., 1988; Deloulme et al., 1990). Within committed, post-mitotic neurons, expression of GAP-43 is tightly correlated with a reversible transition between a "growth state" typical of immature neurons during axon extension and synaptogenesis, and the chronic "maintenance" state of most mature neurons, characterized by maintenance of a mature axon and synaptic terminals (Skene and Willard, 1981a,b). The reversibility of this transition is illustrated by the ability of many mature neurons to reenter a functional growth state after certain kinds of axon injury, accompanied by reexpression of GAP-43 and other genes characteristic of immature neurons (Hoffman, 1989; Miller et al., 1989; Skene, 1989; Doster et al., 1991; Tetzlaff et al., 1991). The pattern of GAP-43 expression indicates that this gene is strongly induced in the course of neuronal differentiation and, within differentiated neurons, is chronically modulated by additional signals closely linked to the control of axon growth.

In this report, we describe the isolation and characterization of a functional promoter and surrounding regulatory elements from the rat GAP-43 gene. In contrast to several other neural-specific genes, in which neural specificity is conferred by distant silencer elements acting on a relatively promiscuous promoter, we found that 386 base pairs (bp) of DNA surrounding the GAP-43 promoter is sufficient to confer strongly neural-specific gene expression of a marker gene in transient transfection assays. Unlike another neural-specific core promoter, from the synapsin I gene (Sauerwald et al., 1990), the GAP-43 promoter fragment contains canonical TATA and CCAAT consensus sequences. Within neural cultures, expression directed by the GAP-43 promoter is strongly modulated by two interacting regulatory elements that contain sequences with the potential to adopt unusual DNA conformations.

Received Aug. 13, 1991; revised Sept. 25, 1991; accepted Oct. 1, 1991.

We are grateful to Drs. Laurel Bolin, Anson Lowe, and Dennis Clegg for providing cell lines, and to Dr. Susan Meakin for providing 7SNGF and the CAT expression vectors. We also thank Dr. Cornelia Gorman, who suggested the experiments using PCR to amplify primer extension products and provided a detailed protocol, and Dr. Tom Schwartz for helpful suggestions regarding these experiments. We thank Dr. Richard Scheller for providing cDNA libraries and for helpful advice, and Suzanne Muller for sharing transfection protocols prior to publication. We are particularly grateful to Dr. Stuart Feinstein for excellent advice, help with transfection experiments, and critical reading of the manuscript. This work was supported by NIH Grant EY07397.

Correspondence should be addressed to Pate Skene, Department of Neurobiology, Box 3209, Duke University Medical Center, Durham, NC 27710.

^a Present address: Department of Hormone Research, Weizmann Institute of Science, P.O. Box 26, Rehovot 76100, Israel.

^b Present address: Protein Design Laboratories, Mountain View, CA 94043.

^c Present address: Department of Biophysics, Boston University Medical School, Boston, MA 02118.

Copyright © 1992 Society for Neuroscience 0270-6474/92/120691-14\$05.00/0

Materials and Methods

Construction of a rat genomic library and isolation of GAP-43 genomic clones. High molecular weight rat brain genomic DNA was partially digested with *Sau3A* and size fractionated on a 0.6% agarose Tris-acetate gel (Maniatis et al., 1982), and DNA of 15–45 kilobase pairs (kbp) was recovered by electroelution. This size-fractionated DNA was ligated into *Bam*HI/*Eco*RI-digested EMBL3 DNA (Frischauf et al., 1983; Karn et al., 1983) and packaged into λ -phage (Maniatis et al., 1982). The resulting primary library (approximately 5×10^6 independent clones), assuming an average insert size of 17.5 kbp, thus represented over 26 rat genome equivalents.

Both this library and a Charon 4A rat liver genomic library were screened according to the method of Benton and Davis (1977) using cDNA probes that were 32 P labeled by random hexamer priming (Feinberg and Vogelstein, 1983). The Charon 4A library was screened using a full-length GAP-43 cDNA clone pG43-1.2 (Basi et al., 1987) as a probe. The EMBL3 library was screened with a 100 bp *Eco*RI/*Ava*II restriction fragment from the GAP-43 cDNA 5' end. Hybridizations were performed in $5 \times$ saline-sodium citrate (SSC), $1 \times$ Denhardt's solution, 0.1% SDS, 100 μ g/ml sonicated salmon sperm DNA, and 50% formamide at 42°C for 48 hr, followed by multiple washes in $0.1 \times$ SSC, 0.1% SDS at 65°C.

Analysis and sequencing of genomic clones. Restriction maps of the genomic DNA inserts were made by analysis of double and sequential digestions with pairs of restriction enzymes. Southern blots of these digests hybridized with cDNA probes were used to identify the smallest digest fragments containing coding segments of the GAP-43 gene. These fragments were subsequently subcloned in both orientations into the HindIII site of the Bluescript KS+ phagemid vector (Stratagene). *Eco*RI and *Sac*I were used to generate 5' and 3' protruding ends and progressive unidirectional deletions were generated from both orientations using the Erase-a-Base System (Promega) based on a previously developed procedure (Henikoff, 1984). The sequential deletion products were then sequenced off single-stranded templates by the dideoxynucleotide chain termination method (Sanger et al., 1977) with Sequenase (U.S. Biochemicals). Nucleotide sequence alignments and analyses were performed with the software package of the University of Wisconsin Genetics Computer Group (Devereux et al., 1984).

Primer extension and PCR amplification of extension products. Two synthetic oligonucleotide primers were prepared to examine endogenous RNA. Both these nonoverlapping primers were specific for 5' coding sequence of the GAP-43 gene. The first primer (PEM) was complementary to coding nucleotides 67–94 and contained the sequence 5' GAGCCTTATCCTCCGGTTTGACACCAT 3'. The second primer (PE0) was complementary to coding nucleotides 2–29 and contained the sequence 5' CTGTTTGGTTCTTCTCATAACAGCACAGC 3'. PolyA⁺ mRNA from tissue and cortical cultures was isolated using Fast Track (Invitrogen) and annealed for 3 hr at 60°C to primers 32 P-labeled with polynucleotide kinase. Extensions were carried out for 1 hr at 42°C with M-MuLV reverse transcriptase (Bethesda Research Laboratories). Products were then resolved on a 6% acrylamide-urea gel.

Primer extensions for amplification by PCR were done as described above using unlabeled primer PE0. After extension, samples were treated with RNase A for 15 min at 37°C, extracted with phenol/chloroform, precipitated, resuspended in water, and then used for PCR with the Perkin-Elmer Gene Amp kit. The 100 μ l reaction containing 50–100 pmol of each upstream and downstream primer was overlaid with mineral oil and cycled for 30 rounds of amplification in a Perkin-Elmer Cetus thermal cycler at 95°C for 1 min, 65°C for 3.5 min, and 72°C for 1.5 min, ending with 72°C for 10 min. 2.5 U of Ampliqaq (Perkin-Elmer Cetus) was added to start and after 15 cycles. The samples were extracted with chloroform and analyzed on a 1.6% agarose gel in Tris borate buffer. PE0 was the downstream primer. Upstream primers were nucleotides +206 to +235 (PCR2.5) containing the sequence 5' GCA-TATGCGGTGAGCAATAGCTGATCGCTG 3', nucleotides +71 to +86 (PCR2.7) containing the sequence 5' CCCGGGCAGAAAGAA-GAAAAAGATTGG 3', and nucleotides –4 to +23 (PCR3) containing the sequence 5' GTCTAGAAATGGGGGTAGGGGCTATGG 3'. Samples containing the upstream primer PCR3 were also amplified under lower-stringency conditions with annealing done at 50°C.

cDNA clone isolation. Two oligonucleotides complementary to the 5' untranslated region of the published rat cDNA sequence (Rosenthal et al., 1987) were 32 P labeled with polynucleotide kinase and used as probes for screening a rat E13 spinal cord cDNA library in λ gt11 (Furley et al.,

1990). One oligonucleotide (PE1) was complementary to nucleotides +427 to +462 and contained the sequence 5' TCTTCCCCTGCCTC-TTCTGCTTCTCTCCTTTC 3'; the second oligonucleotide (PE2) was complementary to nucleotides +348 to +375 and contained the sequence 5' ACTCGTCTCTCGCGCTAGCGGCTCTC 3'. PE1 was hybridized as described for the genomic probes, but at 37°C; PE2 likewise with 35% formamide 37°C. This screen resulted in the isolation of three independent cDNA clones. Inserts from all three purified phage clones were subcloned into Bluescript. Double-stranded *CsCl* purified plasmid was then used as template for sequencing in from both ends of the insert with Sequenase.

Northern blot analysis. PolyA⁺ mRNA from rat liver, brain, primary cortical cultures, or tissue culture cell lines was isolated using Fast Track (Invitrogen). Quantities (1.5 μ g) of each RNA were resolved by electrophoresis through a 0.66 M formaldehyde, 1.2% agarose gel and capillary transferred with $20 \times$ SSC to nitrocellulose membrane. Blots were hybridized under same conditions as genomic screens, with a GAP-43 cDNA probe 32 P labeled by random hexamer priming (Feinberg and Vogelstein, 1983).

Construction of promoter-chloramphenicol acetyltransferase gene hybrids. Because of the lack of a convenient restriction site in the 5' untranslated region, close to the translation start, the first promoter construct (1 kb) containing the entire upstream region from +466 to the HindIII site at –523 was made by deleting in from the 3' end of the 2 kbp HindIII fragment subcloned in Bluescript KS-. *Cla*I and *Apa*I were used to generate 5' and 3' protruding ends, and the *Exo*III deletion was done using the Erase-a-Base system. The deleted insert was then cut out using HindIII and *Kpn*I and force cloned into the unique HindIII promoter cloning site of pSVoATCAT situated 61 bp upstream of the chloramphenicol acetyltransferase (CAT) translation start. When conveniently situated restriction sites were not present, subsequent 3' deletions for promoter constructs were generated in the same way. All deletion endpoints were accurately mapped by sequencing. Except for the 1 kb construct, all fragments generated by deletions or restriction digests were blunt ended for subcloning either by treatment with T4 DNA polymerase for chewback of protruding 3' ends or by Klenow fragment of *Escherichia coli* DNA polymerase for filling in 5' overhangs. These fragments were then cloned into the HindIII site of pSVoATCAT after addition of HindIII linkers (Pharmacia). The expression vector pSVoATCAT (Meakin et al., 1989), which contains the promoterless bacterial CAT gene, is a derivative of pSV2ACAT (Kadesch and Berg, 1984) without the SV40 promoter. It also contains two polyadenylation signals derived from the early region of SV40 immediately upstream of the promoter cloning site and a translation stop region between the promoter cloning site and the CAT initiation codon. Consequently, CAT activity in the absence of an inserted promoter fragment is practically nonexistent.

Cell culture and DNA transfections. Rat cortical neurons were grown in dissociated cultures by a method previously described (Dichter, 1978) and modified as follows. Embryonic day 16–18 (E16–E18) embryos were removed using sterile technique. Cortices were removed, peeled of meninges, and collected in cold Hank's buffered salt solution (HBSS). The tissue was gently spun 5 min at $500 \times g$ and then resuspended in HBSS without Ca/Mg. After 10 min at 37°C, tissue was spun as above and resuspended in HBSS without Ca/Mg containing 0.01% trypsin. Subsequent to digestion 10 min at 37°C with trypsin, tissue was spun again, resuspended in HBSS with 100 μ g/ml of DNase, and incubated 10 min at 37°C. DNase was removed by rinsing with fresh Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% supplemented calf serum. The tissue was then triturated with fire-polished Pasteur pipettes with lumens of decreasing dimensions; cells were then counted and plated at a density of $2-4 \times 10^6$.

PC12 cells (Greene and Tischler, 1976) were grown in DMEM supplemented with 5% supplemented calf serum and 5% horse serum; RN22 cells (Pfeiffer and Wechsler, 1972), in DMEM containing 10% heat inactivated horse serum; HTC cells (Thompson et al., 1966), in DMEM containing 10% calf serum; and AtT-20 (Buonassisi et al., 1962), in DMEM containing 10% fetal calf serum. All cells were maintained in a humidified 37°C incubator—PC12 cells with a 12% CO₂, 88% air atmosphere; RN22, HTC, and AtT-20 with an 8% CO₂, 92% air atmosphere; and cortical cultures with a 5% CO₂, 95% air atmosphere. All tissue culture medium was obtained from GIBCO and contained $1 \times$ penicillin/streptomycin/fungisome (PSF), also from GIBCO. Serums were from Hyclone.

For transfections, PC12 cells were plated on polylysine-coated 60 mm tissue culture dishes (Muller et al., 1990) at a density of 2×10^6 cells/

dish and grown for 2–3 d with 7S NGF added to 100 ng/ml. Cortical cultures were plated on polylysine-coated 35 mm tissue culture dishes at a density of $2\text{--}4 \times 10^6$ cells/dish and grown for 5 d before transfection. Fresh medium was added every 2 d.

Transfections were carried out as described (Muller et al., 1990) with 40 μ l of Lipofectin (1 mg/ml; Bethesda Research Laboratories) and 20–40 μ g of DNA (normalized by molar ratio to 20 μ g of the pSVoATCAT) per dish. Five hours later, the transfection medium was removed and replaced with fresh serum containing medium. Cells were harvested for CAT assays or polyA⁺ selection 2 d following transfection.

CAT assays and quantitation. For CAT assays, cells were harvested as described previously (Muller et al., 1990) and then assayed by the modified ³H-sodium acetate method (Nordeen et al., 1987). Two time points were taken from each plate to ensure that measurements were within the linear range with respect to reaction time.

Except in experiments shown in Figure 5B, promoter activity for each construct was calculated as the average CAT activity from a minimum of three independent experiments, each experiment containing duplicate transfections for each construct. Within each experiment, accurate cell plating resulted in excellent reproducibility between duplicate plates transfected with the same construct. All experiments contained RSVCAT-transfected cultures as a positive control, and both PSVoCAT-transfected and untransfected cultures as negative controls. Promoter activity was calculated after subtraction of background, represented by the PSVoCAT-transfected samples.

Results

Relation to genomic exon/intron structure to protein functional domains

To isolate genomic DNA clones encoding the rat GAP-43 gene, we used a full-length GAP-43 cDNA clone (Basi et al., 1987) to screen an amplified library of genomic DNA from rat liver (Sargent et al., 1979). Screening of approximately 35–40 genome equivalents yielded two sets of nonoverlapping clones accounting for the coding parts of the GAP-43 gene downstream of codon 10. Since an additional 50 genome equivalents screened from this library failed to produce clones accounting for the first 10 codons and the mRNA 5' untranslated region, we constructed a primary unamplified rat brain genomic library and screened it with a restriction fragment from the GAP-43 cDNA 5' end. A screen of 13 genome equivalents yielded three overlapping clones, λ 2A-21, λ 4A-21, and λ 9B, which were found to contain a sequence corresponding to codons 1–10 of the cDNA, followed on the 3' end by intron sequence. At their 5' ends, these clones encoded all of the 5' untranslated sequences contained in previously characterized cDNAs, plus at least 5 kbp of additional DNA upstream of the known cDNA sequences. The coding portion of this sequence was designated exon 1 (Fig. 1). A second set of clones, selected from the amplified library, comprised four independent isolates containing a second exon encoding amino acids 11–199, flanked by intronic sequences at either end. This set of clones is represented by λ 6A in Figure 1. A third set of genomic clones, comprising λ 5A-1 and λ 5B-1A, contained exon 3, beginning at the second nucleotide in codon 199 and extending through the termination codon to the polyadenylation site(s). Restriction mapping of clones λ 5A-1 and λ 5B-1A indicated differences in their 3' ends that were not investigated further.

The organization of the GAP-43 gene deduced from these genomic clones is illustrated in Figure 1. The sequences at each of the exon/intron borders in the GAP-43 gene are in agreement with the consensus sequences for eukaryotic donor and receptor splice junctions, and splicing obeys the GT/AG rule (Mount, 1982). The genomic clones isolated to date contain a total of 28 kbp, and as they are nonoverlapping, this is a minimal estimate for the size of the GAP-43 gene. The organization of the

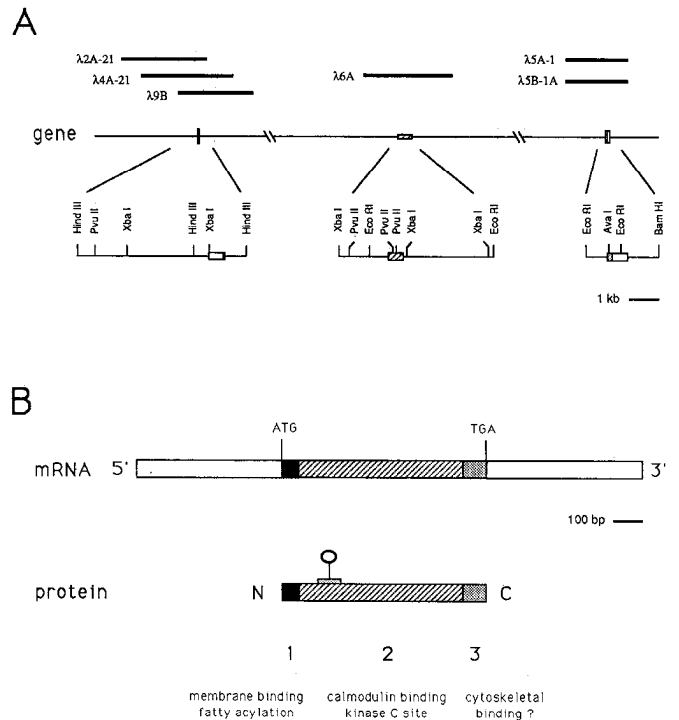


Figure 1. Restriction map and exon/intron organization of the rat GAP-43 gene. *A*, Genomic clones containing sequences corresponding to GAP-43 mRNA (*thick lines*) are aligned above a low-resolution schematic map of the rat GAP-43 gene. Exon segments are marked by *boxes*. Below the low-resolution map is an expanded view showing restriction maps of the genomic DNA surrounding each exon. From these mapped regions, the smallest restriction fragments containing coding parts of the gene were identified on Southern blots probed with GAP-43 cDNA, subcloned, and sequenced for accurate mapping of exon/intron boundaries. *B*, Schematic maps of GAP-43 mRNA and protein showing the relationship of exon/intron boundaries to functional sites in the protein. *Solid boxes* indicate translated portions of the mRNA. Exon 2 (*hatched box*) codes for a large central protein domain containing several functional and structural units. Two of these are indicated by the *bracket* and *circle* above the protein map, marking a proposed calmodulin-binding domain centered on the principal serine residue phosphorylated by protein kinase C (Alexander et al., 1988; LaBate and Skene, 1989; Liu et al., 1991). Exon 1 (*solid box*) codes for a small amino-terminal domain involved in GAP-43 binding to membranes, and exon 3 (*shaded box*) encodes a portion of the carboxy-terminal protein domain that may interact with elements of the submembranous cytoskeleton.

rat GAP-43 gene reported here agrees with that described previously by Grabczyk et al. (1990).

The genomic organization of the three GAP-43 exons conforms to some degree with the generally advanced hypothesis that exons of eukaryotic genes often encode discrete functional units (Blake, 1978; Gilbert, 1978; Go, 1983). The majority of GAP-43 is encoded by a single exon, which is flanked on either end by exons encoding small amino-terminal and carboxy-terminal segments. The amino-terminal 10 amino acids, encoded by exon 1, correspond to the proposed site for membrane binding via fatty acylation (Skene and Virag, 1989; Zuber et al., 1989; Liu et al., 1991). On the 3' end, exon 3 encodes a 27 amino acid, carboxy-terminal domain that encompasses a phylogenetically conserved sequence motif (F motif of LaBate and Skene, 1989) and part, but not all, of a region with limited sequence similarity to neurofilament proteins. The conserved sequences within this region have been proposed as possible

content has been previously noted to be a conserved feature of the 5' untranslated sequences of GAP-43 cDNAs (Kosik et al., 1988; LaBate and Skene, 1989). In two portions of the purine-rich genomic sequence, the A+G content increases to 100%. One stretch of 50 bp (+302 to +352) consists of interspersed and overlapping repetitions of the sequences AGGGA and GA. A longer sequence spans 125 contiguous purines (+107 to +232) on the coding strand, including 35 pairs of directly alternating guanosine and adenine residues, (GA)₃₅. Homopurine sequences of this type, with their complementary homopyrimidine strands, have been shown to be capable of forming an H-DNA structure, in which partial melting of the DNA allows one strand to fold back to form an intramolecular triple helix, leaving the other strand unpaired (Johnston, 1988; Wells et al., 1988; Htun and Dahlberg, 1989). A different kind of repeating sequence occurs approximately 380 bases upstream of the proposed TATA box. This element consists of 17 pairs of alternating purine and pyrimidine residues, (GT)₁₇, on the coding strand, which under appropriate conditions can adopt the left-handed Z-DNA conformation (Rich et al., 1984).

Mapping of the GAP-43 transcription start site

As a first step toward locating the GAP-43 transcription start site within the 5' genomic sequence, we carried out primer extension analysis on mRNA from neonatal brain. Extensions from two nonoverlapping primers complementary to the GAP-43 coding region (PE0 and PEM) yielded three discrete extension products mapping to positions +349, +388, and +418, in addition to multiple early terminations (Fig. 4A). Primers complementary to sequences upstream of the translation start site either failed to generate extension products or produced multiple nonspecific products (not shown). As in any primer extension analysis, the discrete extension products can represent either legitimate mRNA termini or artifactual premature terminations of the extension reaction. In the case of GAP-43 mRNA, several considerations make the second possibility of special concern. The sequence upstream of the translation start site indicates a significant potential for secondary structure that could interfere with primer extensions. Indeed, during the sequencing of the genomic DNA, we noted a strong tendency for polymerase reactions to terminate at discrete sites, including nucleotide +349. Furthermore, the three principal extension products (Fig. 4) map to positions in the genomic sequence far from any known consensus sequences for transcription initiation. Finally, the previous isolation (Rosenthal et al., 1987) of a GAP-43 cDNA starting farther upstream than any of the start sites marked by primer extension suggests the presence of an upstream transcription start site not accounted for among the major primer extension products.

Because of the unusual sequence within this region, RNase protection analysis was not able to resolve the issue of the transcription start site(s). RNase protection analysis yielded multiple bands even with an RNA transcript of known size, generated from the cloned genomic DNA by *in vitro* transcription initiated 1 kbp upstream of the translation start codon (not shown).

We therefore used two alternative approaches to test for the existence of such an upstream transcript, by amplifying low-frequency primer extension products extending beyond regions of strong secondary structure. We first screened a cDNA library from embryonic rat spinal cord with two oligonucleotide probes complementary to the 5' end of the longest published GAP-43 cDNA (Rosenthal et al., 1987). Of 300,000 clones screened,

Rat GAP-43	-308 to -312			TGGAG
Rat GAP-43	-271 to -286	CAGCTGCCAGGAGAT		
Rat type II Na ⁺ channel	-58 to -34	TTGT	CCAGGAGATGGAGCTGTCC	
Rat peripherin	-168 to -155	CAGC	CCAGGAGA	
Rat peripherin	-72 to -62			GGAGCTGTCC
Rat SCG10	-320 to -332	CTGT	CCAGGAG	
Mouse neurofilament	-78 to -62	AGC	CCAGG	TGGAGC

Figure 3. Sequence motif common to the 5' flanking regions of several neural-selective genes. The sequence shown in Figure 2 was examined for the occurrence of a sequence motif previously noted to be common to several neural-specific genes (Thompson and Ziff, 1989; Maue et al., 1990). This region of homology was also compared with the recently published sequence from the 5' end of the rat SCG10 gene (Mori et al., 1990). The numbers refer to the nucleotide positions within the 5' flanking region of each gene. The rat GAP-43 and SCG10 sequences occur on the noncoding strand and are therefore inverted compared to the other three genes. Nonconsecutive sequences from the same gene are set on different lines. *Spaces* represent unmatched bases.

three hybridized with both probes. Sequence analysis showed that all three were GAP-43 clones colinear with the genomic sequence, indicating that there is no additional intron upstream of the translation start site. Two clones (A3 and B1, Fig. 2) extend upstream of the longest published cDNA sequence, and both of these included the (GA)₃₅ repeating sequence. The existence of these low-frequency cDNAs indicated that there is a GAP-43 transcript starting at least 400 bp upstream of the translation start.

In a complementary approach, we used a PCR to amplify low-frequency products of a traditional primer extension reaction. Each PCR reaction contained the same downstream primer as in the initial primer extension (PE0), together with one of three upstream primers representing sequences at increasing distances upstream (Fig. 4B). Two of the upstream primers, PCR2.5 and PCR2.7, were located on either end of the GA repeat, which is fully contained within the longest cDNA clone, A3. The third upstream primer, PCR3, was complementary to a genomic sequence upstream of the longest cDNA and downstream of the potential TATA box. The rationale for these experiments was that the PCR reaction would generate a product only if the upstream primer was included within the original mRNA transcript serving as a template for primer extension. As it turned out, PCR reactions with all three primers on primer extensions of brain-derived mRNA resulted in products of a size consistent with transcripts colinear with the genomic DNA (Fig. 4B). Upstream primer PCR3, however, yielded a product only at a hybridization stringency lower than that used for the upstream primers. At the higher stringency, PCR3 failed to amplify the primer extension products from mRNA, even though it was able to amplify plasmid DNA containing the corresponding genomic sequence (Fig. 4B). This suggests that the hybrid between PCR3 and the longest primer extension products contains several unpaired bases. One interpretation of this finding is that PCR3 extends several bases beyond the 5' end of the longest GAP-43 mRNA transcripts.

On the basis of the amplified primer extension results, we propose that GAP-43 transcription can initiate far enough upstream to include the PCR3 sequence. The differential stringency requirements for hybridization of PCR3 to primer extension products suggest, in fact, that transcription may initiate within the PCR sequence. Consistent with this possibility, the genomic DNA beginning 33 bases upstream of the PCR3 sequence contains a TATA-like consensus sequence and two associated CCAAT boxes (Fig. 2). To provide a single reference point for

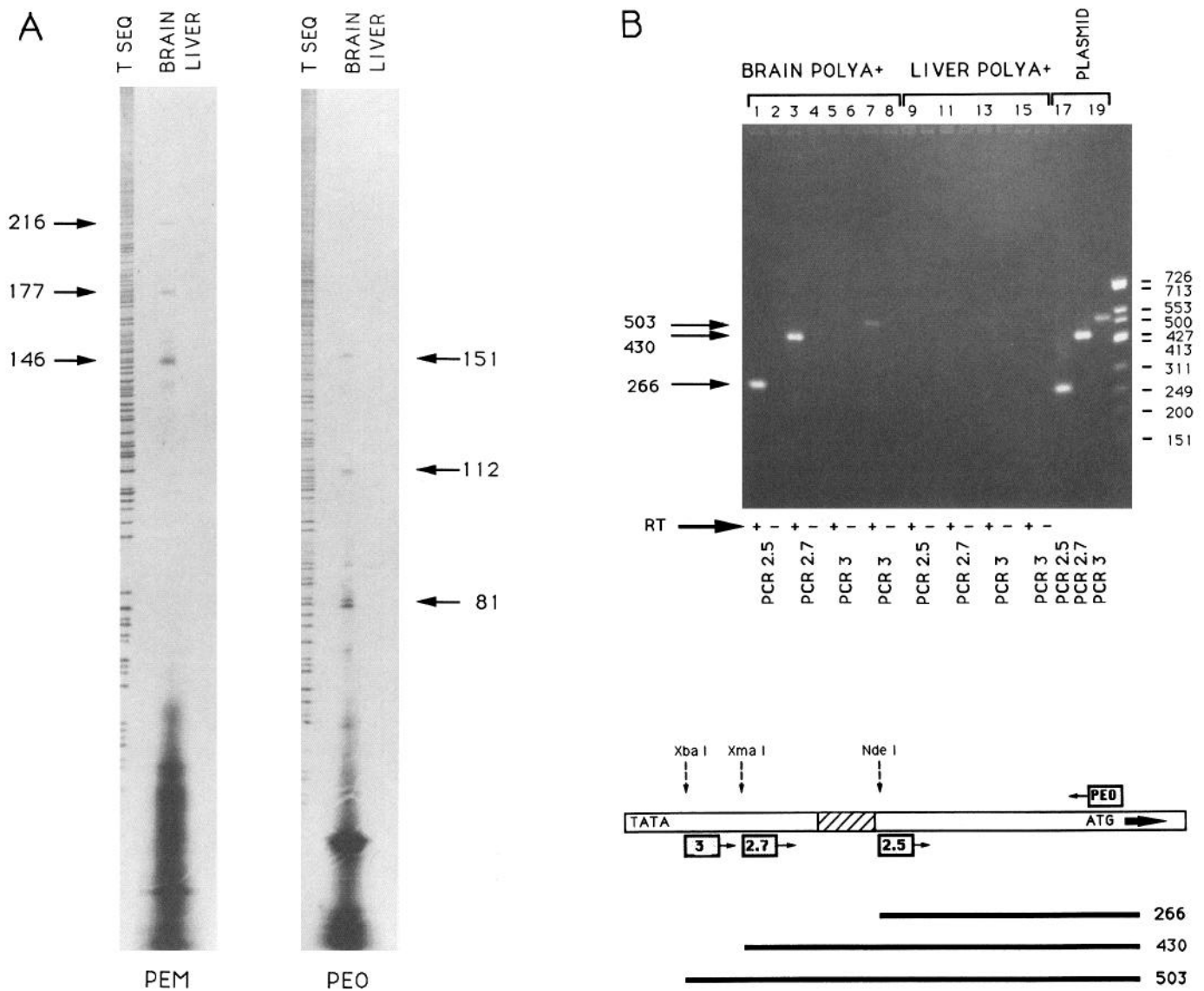


Figure 4. Localization of potential transcription start site(s) by standard and PCR-amplified primer extension analysis. *A*, Standard primer extension analysis. Radiolabeled primers complementary to GAP-43 coding nucleotides 67–94 (*PEM*) or 2–29 (*PEO*) were hybridized with 8 μ g of polyA⁺ mRNA from 8 d rat brain or rat liver and then extended using reverse transcriptase as described in Materials and Methods. Shown are radiolabeled extension products resolved on a denaturing polyacrylamide gel, adjacent to a dideoxy sequencing T ladder (*T SEQ*) of the 5' genomic clone primed with *PEM* and *PEO*, respectively. Three major extension products generated with each primer are indicated with arrows; the numbers indicate the actual length in nucleotides for each extension product. Measured from the appropriate primer sequence, the 5' ends of the three major products for either reaction correspond to nucleotides +418, +339, and +388 in Figure 2. *B*, PCR-amplified primer extension. For each lane, 8 μ g of polyA⁺ mRNA from 8 d rat brain (lanes 1–8) or rat liver (lanes 9–16) were hybridized with the primer *PEO* as in *A* and added to standard primer extension reactions with reverse transcriptase (odd lanes, marked +) or without reverse transcriptase (even lanes, marked –). Each pair of samples was then taken for PCR amplification with the same downstream primer *PEO* and one of three possible upstream primers, PCR2.5, PCR2.7, or PCR3. The annealing step of the PCR reaction was done at 65°C for all lanes except 7–8 and 15–16, which were annealed at 50°C. Products were separated on an agarose gel and stained with ethidium bromide. All three primer pairs were able to amplify brain-derived primer extension products to produce PCR fragments (lanes 1, 3, and 7) of a size corresponding to the genomic distance between their primer sequences (control lanes 17–19). Amplification with primer PCR3 generated a PCR product of appropriate size at reduced hybridization stringency (lane 7), but not at the higher stringency used for the other reactions (lane 5). No PCR products were generated from mRNA from liver (lanes 9–16), which does not express GAP-43. The failure to generate PCR products from brain mRNA in the absence of reverse transcriptase confirms that the PCR products are generated from primer extension products and not from contaminating DNA. Standard size markers (lane 20) are ϕ x174 DNA digested with *Hinf*I. The diagram at the bottom shows the PCR primers (bold boxes) aligned with corresponding sequences in a schematic map of GAP-43 genomic DNA. Approximate positions of the *TATA* consensus, the GA repeating sequence (hatched box), and the translation start site (*ATG*) are indicated, along with three restriction sites whose precise positions are indicated in Figure 2. Bold lines below the map show the position and the length in base pairs of the expected PCR product for each pair of primers, if the upstream primer sequence is included in GAP-43 mRNA.

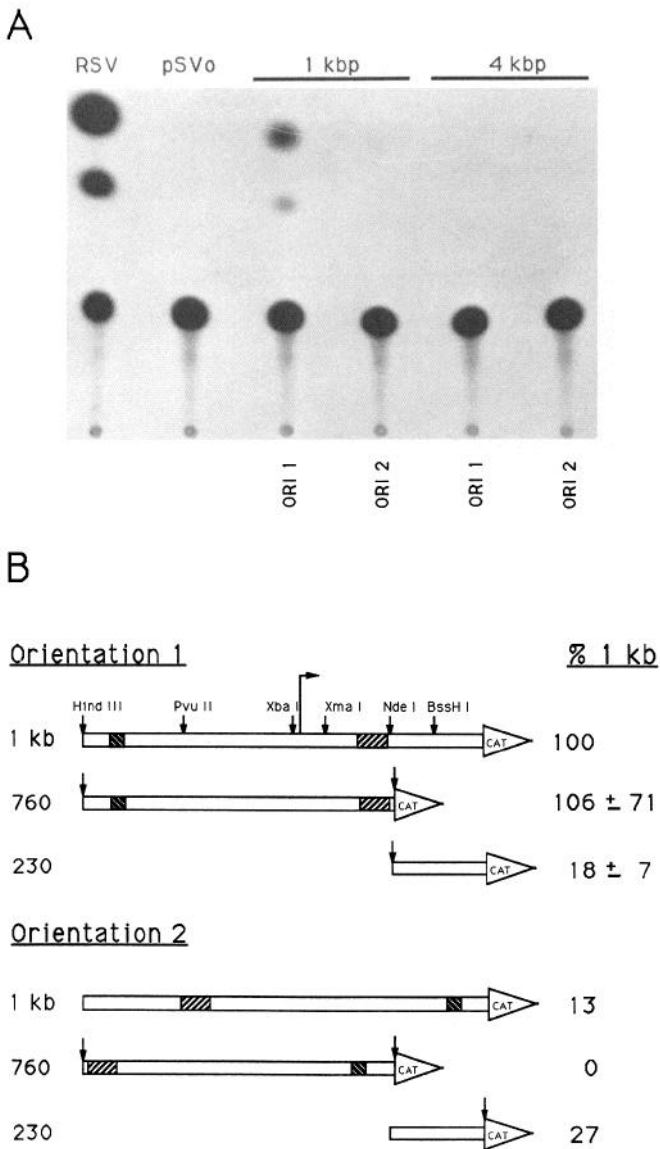


Figure 5. Functional identification of a GAP-43 promoter. DNA fragments from the 5' end of the GAP-43 gene were inserted into plasmid pSVoCAT, upstream of a promoterless gene for the bacterial enzyme CAT, and assayed for their ability to direct expression of CAT after transfection into primary cultures of cells from embryonic rat cortex. CAT activity was assayed as the ability to convert radioactive chloramphenicol to acetylated products resolved on thin-layer chromatography (*A*), or to transfer radioactive acetate to chloramphenicol (*B* and subsequent figures). *A*, Separate fragments containing either the first 1 kbp of DNA upstream of the GAP-43 coding sequence (*1 kbp*), or the next 4 kbp of more upstream DNA (*4 kbp*), were inserted into the pSVoCAT vector either in their normal orientation with regard to the direction of transcription (*ORI 1*) or in the opposite orientation (*ORI 2*). The inserts comprised 1 kbp downstream, or 4 kbp upstream, of the HindIII site indicated in Figure 2. Primary cortical cultures were transfected with equimolar amounts of each construct and assayed for CAT activity 48 hr after transfection as described by Muller et al. (1990). The acetylated products produced by CAT activity are visualized as two upper spots in the thin-layer chromatogram illustrated. Parallel cultures were transfected with the same CAT vector bearing an RSV promoter (*RSV*) as a positive control, or with the pSVoCAT vector with no promoter insert (*pSVo*) as a negative control. *B*, The 1 kbp fragment assayed in *A* was subdivided into a 760 bp fragment extending upstream of the GA repeating sequence, and a 230 bp fragment extending downstream of the repeating sequence to 4 bp before the translation start

the sequence numbering scheme used here, the first A within the sequence spanned by PCR3 is designated nucleotide +1. Designation of this particular base as +1 is solely for numbering purposes and should not be taken to indicate that transcription initiates precisely at this nucleotide. Mapping of a start site in this general region suggests that the GAP-43 promoter is located nearby, perhaps surrounding the promoter consensus sequences situated closely upstream.

Functional identification of the GAP-43 promoter

To identify the GAP-43 promoter functionally, we generated a series of constructs containing various segments of GAP-43 genomic DNA fused to the promoterless bacterial CAT gene. These constructs were tested for promoter activity by transient transfection into primary cortical cultures. The first construct tested, containing the entire 1 kbp fragment from 4 bp upstream of the translation start back to the first HindIII site, yielded activity that was 60–70-fold lower than that exhibited by the Rous sarcoma virus (RSV) promoter fused to CAT (RSVCAT; Gorman et al., 1982) introduced into parallel cultures. This activity was exhibited only in the correct orientation with regard to transcription of the GAP-43 gene, confirming its identification as a functional promoter element (Fig. 5*A*). The next 4 kbp of more upstream DNA failed to show any detectable promoter activity (Fig. 5*A*), indicating that the only GAP-43 promoter(s) is contained within the 1 kbp fragment.

Within this 1 kbp fragment, PCR on primer extension products, cDNA information, and sequence consensus are all consistent with a promoter located upstream of the polypurine-polyrimidine repeating sequence (GA repeat). However, because additional primer extension products terminate downstream of this repeating sequence (Fig. 4; Grabczyk et al., 1990), the presence of an additional downstream promoter could not be ruled out. To localize a possible GAP-43 promoter(s) functionally, we divided the 1 kbp fragment into two subfragments, a 760 bp fragment containing the proposed TATA and CCAAT boxes, and a 230 bp fragment encompassing the sites corresponding to the 5' ends of the major primer extension products. Each of these fragments was then tested separately for promoter activity (Fig. 5*B*). The 760 bp fragment (plasmid 760) exhibited orientation-specific activity that was more variable in magnitude but was comparable on average to the full 1 kbp fragment (Fig. 5*B*). Similarly strong expression from the 760 bp fragment was observed in additional experiments in which CAT expression was assayed by thin-layer chromatography but not quantified (data not shown).

←

site. The positions of the GT and GA repeating sequences (*hatched boxes*) are indicated for reference, and the length of each promoter fragment (in base pairs) is indicated at left. *Downward arrows* indicate restriction sites, and the *tall arrow* preceding the Xba I site indicates the position of a proposed transcription start region, corresponding to the sequence spanned by primer PCR3 in Figure 2. Expression from fragments inserted in the correct orientation with regard to transcription (*Orientation 1*) was quantitated from three independent experiments, each containing duplicate transfections of each construct. Activity is expressed as a percentage of the activity driven from the 1 kbp plasmid and is shown as the mean ± SEM. Expressions from fragments inserted in the opposite orientation (*Orientation 2*) were quantitated from a single experiment, containing duplicate transfections of each construct, and were confirmed qualitatively in two additional experiments.

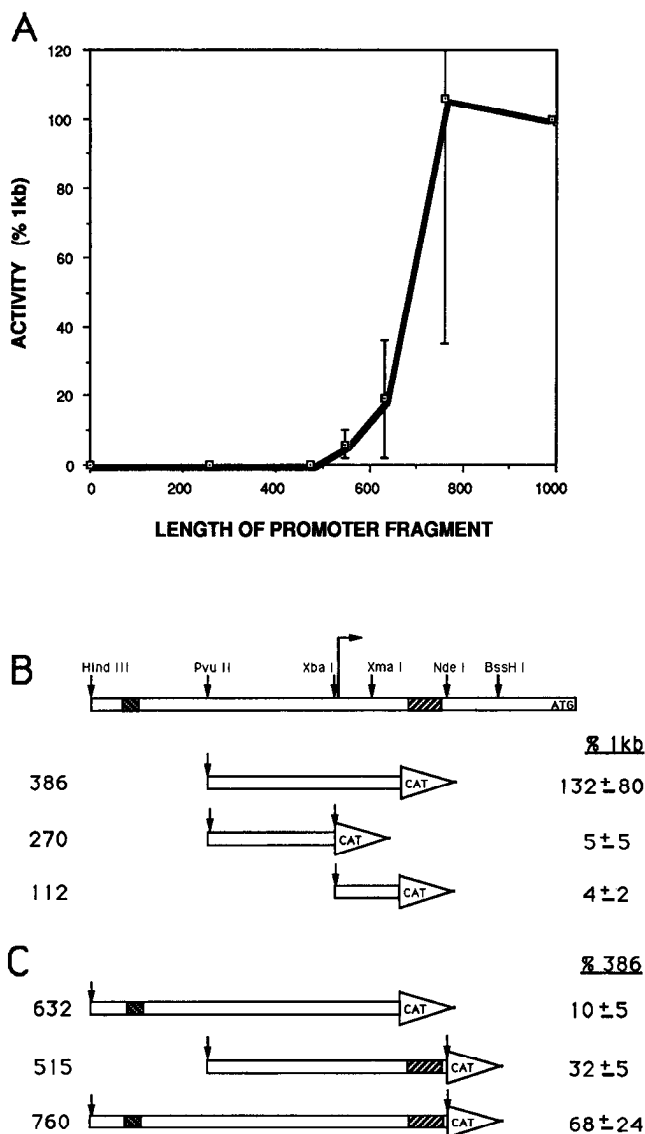


Figure 6. Resolution of a GAP-43 promoter, and its regulation by two flanking elements. *A*, Effect of progressive 3' deletions on promoter activity of the 1 kbp genomic fragment immediately upstream of the GAP-43 translation start site. Due to the lack of convenient restriction sites, the 3' deletions for fragments 632, 548, and 474 were generated by *ExoIII* digestions as described in Materials and Methods. *B*, Structure and transcriptional activity of GAP-43 promoter constructs with double deletions. Diagrams of construct structure are aligned with a map of the 1 kbp fragment, situated above them. The alignment and scale of the diagrams correspond to the x-axis in *A*. *C*, Expression from the GAP-43 386 bp fragment containing the core promoter is strongly influenced by two interacting elements on either side of it. Promoter activity in these experiments is expressed relative to the activity of the 386 bp core strongly inhibits expression (plasmid 632). The strongly inhibitory upstream fragment contains the GT repeating sequence (*darker hatched box*). A downstream region comprising the GA repeating sequence (*lighter hatched box*) and additional polypurine sequence moderately inhibits expression when added alone to the 386 bp core (plasmid 515). Addition of the upstream and downstream elements together (plasmid 760), however, does not produce an additive negative effect on expression. Instead, each element largely counteracts with negative influence of the other. The schematic diagrams are to the same scale and alignment as in *A* and *B*.

The 230 bp fragment (plasmid 230) also exhibited measurable activity, but this activity was consistently low relative to plasmid 760 and was not orientation specific (Fig. 5*B*). Low-level, orientation-independent transcription has been observed previously with plasmids containing regulatory elements introduced in the absence of a promoter (Moreau et al., 1981; Waslyk et al., 1983). The downstream 230 bp fragment thus exhibits neither conventional promoter activity nor consensus sequence elements associated with other eukaryotic promoters. We have therefore focused our further analysis on the 760 bp fragment containing a GAP-43 promoter that accounts functionally for the majority of the promoter activity represented in the larger 1 kbp fragment and that complies with conventional functional and structural criteria.

To map this promoter further, we constructed a series of plasmids bearing inserts with successive deletions from the 3' end of the 760 bp fragment. Promoter activity decreased progressively with deletions from base +237 (760 bp fragment) to base -50 (474 bp fragment; Fig. 6*A*). These results define the region between these bases as important for transcription.

Although a large initial drop in expression was observed upon deletion of the GA repeat, two fragments that retained the putative TATA and CCAAT boxes exhibited low but detectable levels of activity even in the absence of the GA repeat (632 bp and 548 bp fragments; Fig. 6*A*). Deletion of 250 bp from the 5' end, including the GT repeating sequence, restored full activity to a promoter construct without the downstream GA repeat (plasmid 386; Fig. 6*B*). The resulting 386 bp fragment was the smallest fragment to exhibit robust promoter activity. This activity was orientation dependent. A GAP-43 core promoter is thus functionally mapped within the 386 fragment, clearly resolved from the GA repeat.

While the 386 bp fragment consistently directed substantial CAT expression in the embryonic cortical cultures, the magnitude of this expression varied more from one transfection experiment to another than did expression from the larger 1 kbp fragment (Figs. 6, 7). Similar variations in the magnitude of expression were observed with other promoter fragments derived by 3' deletions from the larger 1 kbp fragment, most notably the 760 bp fragment (Fig. 6*A*). The increased variability is apparent whether the expression from the 3' deleted constructs is normalized to the 1 kbp fragment (Fig. 6), or whether expression from each GAP-43 promoter construct is normalized separately to an RSV control promoter (Fig. 7). This suggests that elements contained in the 1 kbp fragment 3' to the core promoter may help stabilize expression in these transient transfection assays.

To complete the functional mapping of the GAP-43 promoter, we took advantage of a conveniently located *XbaI* restriction site to divide the 386 bp fragment into two subfragments, separating the putative TATA box from a proposed transcription start site (plasmids 270 and 112; Fig. 6*A*). Neither of these constructs exhibited promoter activity on its own (Fig. 6*B*). The inability of the upstream 270 bp fragment to initiate transcription indicates that the transcription start site(s) used by the GAP-43 promoter lies no farther upstream than the *XbaI* site. On the other hand, the PCR-primer extension data described previously showed that the longest GAP-43 transcripts contain sequences corresponding to the genomic DNA immediately downstream of the *XbaI* site (Fig. 2). The functional and structural data therefore agree on a transcriptional start within a few bases downstream of this restriction site.

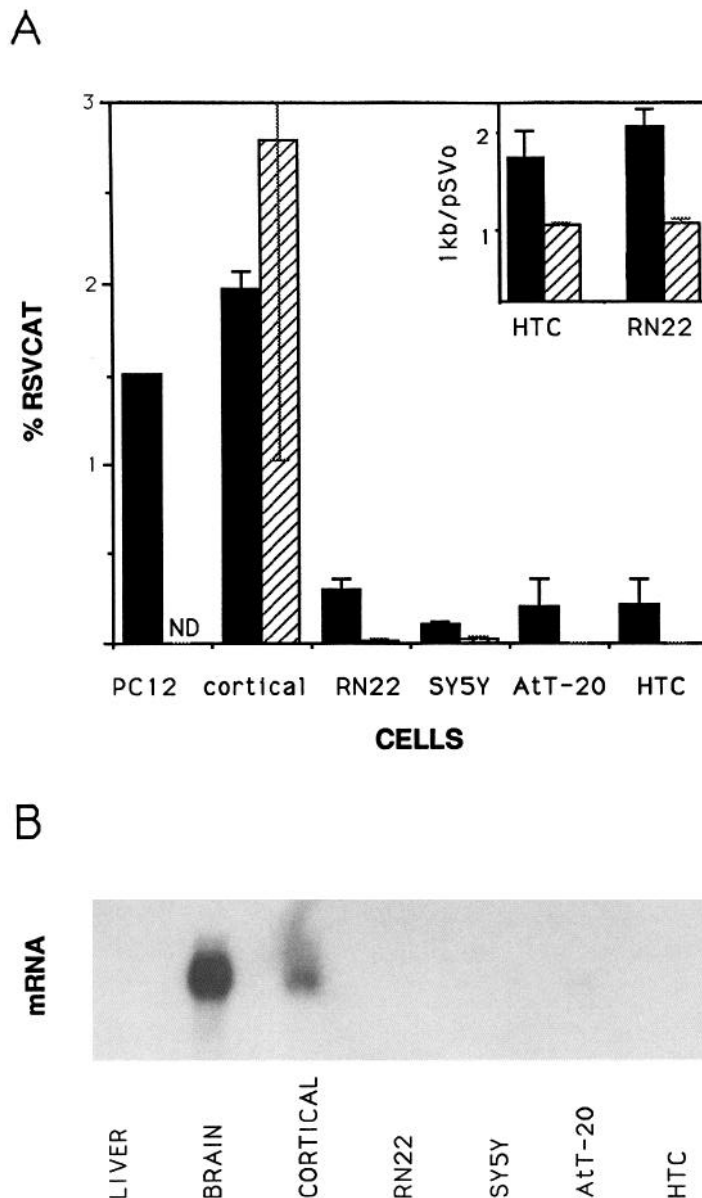


Figure 7. Neuron-selective expression directed by a small GAP-43 promoter-containing fragment. **A**, Plasmids bearing the largest (1 kbp) and smallest (386 bp) fragments that exhibited GAP-43 promoter activity in primary cortical cultures were transfected into a variety of neural and non-neural cells. NGF-treated PC12 cells (*PC12*) and primary cultures of cells from embryonic rat cortex (*cortical*) contain differentiated neurons or neuron-like cells. A Schwannoma cell line (*RN22*), an undifferentiated neuroblastoma cell line (*SY5Y*), and a pituitary-derived neuroendocrine cell line (*AtT-20*) are all non-neuronal cells derived from neural precursors. Non-neuronal cells are represented here by a hepatoma tissue culture cell line (*HTC*). *Solid bars* represent activity of the 1 kbp plasmid, and *hatched bars* represent activity of the 386 bp core promoter in each cell type, expressed as a percentage of RSV CAT activity in the same cell type. The values for each cell type show the mean \pm SEM from three or four independent experiments, each experiment containing duplicate transfections of each construct. Activity of plasmid 386 was not determined in the NGF-treated PC12 cells. A closer examination of low-level activity of plasmids 1 kbp and 386 in *HTC* and *RN22* cells is displayed in the *inset*. Promoter activity in the *inset* is expressed as a fraction of background activity represented by cultures transfected with the promoterless plasmid vector, pSVoCAT. **B**, Northern analysis of endogenous GAP-43 mRNA expression in rat tissues and in cultured cells used for transfection experiments. A 1.5 μ g quantity of polyA⁺ mRNA from each source was separated on a formaldehyde-agarose gel and transferred to nitrocellulose. Nitrocellulose membranes were probed with a ³²P-labeled GAP-43 cDNA insert.

The GAP-43 promoter lies between two elements that interact to affect its activity

Deletion experiments described above showed that serial removal of a polypurine-polypyrimidine tract containing the GA repeat, and an upstream fragment containing the GT repeat, strongly influence expression from a core promoter lying between them. Figure 6B shows that addition of the upstream GT-containing fragment alone to the 386 bp core promoter fragment strongly depresses promoter activity (plasmid 632). Similarly, addition of the downstream GA repeat alone (plasmid 515) decreases expression of the marker enzyme, but to a lesser extent. However, the negative effects of these flanking elements are not additive when both elements are present. Instead, repression by either element separately can be counteracted by the presence of the other on the opposite side of the promoter (plasmid 760; Fig. 6B), indicating that their effects on promoter activity are dependent on interaction between them.

A core promoter fragment is sufficient to direct neural-selective expression in transient expression assays

A prominent regulatory feature of GAP-43 is that its expression is confined to neural cells, predominantly neurons. To determine the degree to which the core GAP-43 promoter and its surroundings can account for this neural-specific expression, we tested our smallest (plasmid 386) and largest (1 kbp) promoter constructs for expression in a variety of neural and non-neural cells (Fig. 7A).

Activity of the 1 kbp construct was highest in cultures containing differentiated neurons (embryonic rat cortex) or neuron-like cells (NGF-treated PC12 cells). Expression of this construct was 10–20-fold lower not only in the liver-derived HTC cell line, but also in AtT-20, a pituitary-derived cell line that behaves as a neuroendocrine precursor; in SY5Y, a committed neural progenitor neuroblastoma line that was undifferentiated; and in RN22, a neural crest-derived Schwannoma cell line. Cells that

direct only low levels of CAT expression from the GAP-43 promoter express low to undetectable levels of endogenous GAP-43 mRNA (Fig. 7B), suggesting they lack some component(s) necessary for activating this promoter. Similarly high expression of the 1 kbp in NGF-treated PC12 cells and primary cortical cultures (Fig. 7A), and comparably low expression in non-neuronal cell lines and primary astrocyte cultures (not shown) argued that preferential expression in cortical cultures was not an artifact of comparing primary cells and cell lines.

The 386 bp core promoter fragment not only retained neural selectivity in its expression but in fact displayed stricter selectivity than the intact 1 kbp. Although clearly selective, the 1 kbp construct in some cell lines exhibited activity up to two times the background CAT activity of a promoterless construct (Fig. 7A, inset). The smaller 386 bp fragment never exhibited activity detectable above background in these cultures.

Discussion

We have identified a functional promoter located on the GAP-43 gene approximately 500 nucleotides upstream of the protein coding region. A short fragment of DNA encompassing this promoter contains elements sufficient to direct neural-selective transcription. Two regions of DNA flanking this promoter, each containing sequences with the potential to adopt unusual DNA conformations, interact to modulate expression from the promoter lying between them.

Characterization of the promoter

The GAP-43 promoter defined here appears to be structurally and functionally similar to other eukaryotic promoters. Functionally, the 386 bp fragment containing the core promoter directs transcription of a marker gene in a strictly orientation-dependent fashion. Structurally, both the mapping of transcription start sites by PCR-amplified primer extension (Fig. 4B), and the functional elimination of promoter activity after digestion at a defined XbaI site (Fig. 6B), indicate that the promoter is located within the 3' half of the 386 bp fragment. This region contains a TATA consensus sequence and two CCAAT sequences at distances from each other and from a proposed transcriptional start site similar to those observed in other eukaryotic promoters (Breathnach and Chambon, 1981).

The present evidence suggests that the 386 bp fragment contains the principal, and probably the only, promoter(s) used *in vivo* to direct GAP-43 expression. This fragment accounts for the majority of the promoter activity contained in 1 kbp of DNA upstream of the protein coding sequence, and a functional survey of more than 4 kbp of DNA upstream of the 386 bp fragment produced no evidence for additional promoter activity.

Sequences downstream of the 386 bp fragment do exhibit a limited ability to drive expression of a marker gene in transient transfection assays, but this activity is small compared to the upstream promoter and, more important, is orientation independent (230 bp fragment in Fig. 5B). No *in vivo* promoter yet described has shown such orientation-independent expression. In the context of a promoterless plasmid construct, however, some types of nonpromoter regulatory sequences have been found to direct low-level, orientation-independent expression of marker genes (Moreau et al., 1981; Wasylyk et al., 1983). Our results suggest that the 230 bp downstream fragment may contain modulatory elements of this type but is unlikely to contain an additional GAP-43 promoter. The increased vari-

ability in promoter activity apparent after deletion of the 230 bp fragment from the 3' end of the larger 1 kbp fragment (Figs. 6, 7) is consistent with some modulatory or stabilizing influence of the 230 bp fragment, at least in transient transfection conditions.

Assignment of transcription initiation sites

Expression data that localize the GAP-43 promoter do not identify the site(s) at which this promoter initiates transcription. PCR-primer extension data strongly suggest that transcription can be initiated within the sequence spanned by the primer PCR3, but it is not clear whether additional downstream sites can also be used for transcription initiation. The predicted length of mature transcripts produced from the most upstream initiation site is approximately 1.7 kb, significantly larger than the 1.4 kb previously estimated for *in vivo* GAP-43 transcripts on the basis of agarose gel electrophoresis (Basi et al., 1987; Karns et al., 1987; Rosenthal et al., 1987). This apparent discrepancy may result from underestimation of the size of GAP-43 mRNA, or GAP-43 transcription could be initiated most frequently from additional downstream sites. If transcription is initiated at sites downstream of the PCR3 sequence, these sites may or may not include those mapped by standard primer extension and RNase protection analysis (Fig. 4A; Grabczyk et al., 1990). Extensive secondary structure within the purine-rich sequence of the longer transcript and in DNA:RNA hybrids, including cruciforms and single-stranded loops, could produce both strong stops in primer extension assays and artifactual RNase cleavage sites. In addition, probes complementary to any part of this region can hybridize at multiple sites. These sorts of artifacts can produce artifactual bands that do not correspond to actual transcription start sites, confounding both primer extension and RNase protection analysis. It is therefore difficult to conclude with confidence whether or how frequently GAP-43 transcription may be initiated at sites downstream of the GA repeat. Despite this uncertainty over additional downstream start sites, both PCR-amplified primer extension analysis and isolation of long cDNA clones show that transcription can be initiated at a site well upstream of the GA repeat, close to the functionally mapped GAP-43 promoter.

5' Sequence motifs common to other neural-specific genes

As more neural-specific genes are isolated and analyzed, there has been increasing interest in identifying sequence motifs common to this category of genes that might mediate their shared expression characteristics. Recently, a shared consensus motif was suggested for the rat and human synapsin I genes (Sauerwald et al., 1990), a mouse neurofilament gene (Lewis and Cowan, 1986), and the human low-affinity NGF receptor gene (Sehgal et al., 1988). This sequence was not found in the GAP-43 gene 5' flanking region. However, the GAP-43 gene did contain a sequence element previously noted (Maue et al., 1990) to be common to the rat type II sodium channel gene, the rat peripheral gene (Thompson and Ziff, 1989), and a mouse neurofilament gene (Lewis and Cowan, 1986). This sequence homology, also found in the more recently reported upstream sequence of the SCG10 (Mori et al., 1990), is centered around a conserved core of seven nucleotides (CCAGGAG). Although it is shared by several neural-specific genes, this sequence motif does not seem to be either necessary or sufficient to restrict gene expression to specific neural cell types. In the GAP-43 gene a 386 bp fragment that does not contain this motif is able to direct neural-specific

transcription. Conversely, in the type II sodium channel gene, a 134 bp promoter fragment containing this sequence is not sufficient to confer cell type specificity (Maue et al., 1990).

It is intriguing that in the growing inventory of neural specific genes there are sequence elements common to only a subclass of these genes, and that these are not associated in a straightforward manner with the functional restriction of expression to cells in the neural lineage. Differentiation of neural cells is not a single event but a series of transitions, and induction of different "neural-specific" genes can occur at different stages in this series and in different branches of the neural lineage. There is no reason to assume that a single signal acting on one common sequence element would be used for gene activation at all steps in such a differentiation pathway. Different *cis*-acting elements or combinations of elements therefore may be required to permit activation of subsets of neural-specific genes at appropriate stages in neural differentiation.

Neural selectivity of the GAP-43 promoter

Functional studies of the neural specificity of the GAP-43 promoter and other neural promoters support the notion that there is more than one mode of conferring selective expression in neural cells. Previous investigations of two other neural-specific genes indicated that an upstream "silencer" element imposes neural-selective expression on a relatively "neutral" core promoter (Maue et al., 1990; Mori et al., 1990). In contrast, we found that a small, 386 bp fragment containing the core GAP-43 promoter exhibits substantial tissue specificity. The difference between these two modes of generating neural specificity corresponds to two general modes of regulating tissue-specific gene activity. Enhancers or silencers located substantially upstream and clearly separated from their promoters can act in an orientation-independent fashion to confer specificity on neutral promoters (Ornitz et al., 1987; Fischer and Maniatis, 1988). In other cases, tissue specificity seems to reside primarily in the promoter itself (Mar et al., 1988; Sauerwald et al., 1990). These two modes of operation are not mutually exclusive, and indeed tissue or cell-type distribution often results from the combined actions of cell type-selective upstream elements and relatively specific promoters (Edlund et al., 1985; Foster et al., 1985; Hammer et al., 1987; Maniatis et al., 1987). Our results therefore leave open the possibility of further spatial or temporal modulation of the neural-specific GAP-43 core promoter by upstream or downstream elements not yet defined.

It is possible that the mechanism used to regulate neural-specific expression of a particular gene is related to the evolutionary history of its tissue specificity requirements. Members of multigene families, such as the SCG10 gene (Mori et al., 1990) and the type II sodium channel gene (Maue et al., 1990), presumably arose through duplication of an ancestral gene that was not neural specific, so that neural specificity would have to be imposed by a later addition, such as an upstream silencer. The GAP-43 gene, in contrast, has been a single copy gene for as much of its evolutionary history as we have traced (LaBate and Skene, 1989). The GAP-43 promoter, therefore, may have evolved as a neural-specific promoter from the origins of the structural gene.

In our experiments, both the core GAP-43 promoter and the larger 1 kbp fragment containing it were selectively expressed in cultures containing differentiated neurons, even in preference to non-neuronal cells such as glial and neuroendocrine cell lines. All of the latter cells were assayed under conditions in which

expression of the endogenous GAP-43 gene was very low or undetectable (Fig. 7B). Previous reports show that some non-neuronal cells can express more substantial levels of GAP-43 under some conditions (Vitkovic et al., 1988; Deloulme et al., 1990). We do not know whether either the core promoter or the 1 kbp fragment would drive expression under those conditions. More extensive analysis of expression from these two fragments in neurons and non-neuronal cells under various conditions and at various stages of differentiation should clarify whether regions surrounding the core may further modulate the cell type specificity of the GAP-43 gene.

Counter modulation by elements on opposite sides of the promoter

One region that clearly modulates expression from the GAP-43 promoter is a long polypurine-polypyrimidine tract that includes the repeating sequence (GA)₃₅. This region moderately depresses expression when added alone to the 386 bp core promoter fragment, but is able to counteract the strong repressive effect of an upstream element when both flanking regions are present. Thus, two separate DNA elements located on opposite sides of the core GAP-43 promoter each can act as a negative regulatory element, but each can also counteract the repressive influence of the other.

Because the polypurine-polypyrimidine sequence is transcribed, some of its effects on CAT expression could be exerted at the posttranscriptional levels of RNA processing, stability, or translation (Pelletier and Sonnenberg, 1985). Constructs in which this sequence is added alone to the core promoter fragment (plasmid 515; Fig. 6C) show that any such posttranscriptional effects result in a net reduction in expression of the marker enzyme. However, the presence of this same polypurine-polypyrimidine sequence downstream of the promoter can counteract the negative influence of a nontranscribed element located upstream of the promoter. Thus, the negative influence of the downstream polypurine-polypyrimidine sequence is not additive with the negative influence of the upstream element. The ability of the downstream polypurine-polypyrimidine sequence to counteract the negative influence of a nontranscribed element located upstream of the promoter is difficult to explain except as an influence exerted at the level of transcription. We therefore suggest that the downstream polypurine-polypyrimidine sequence, containing the GA repeat, interacts with the upstream fragment containing the GT repeating sequence to modulate transcription from the promoter located between them. This does not rule out an additional, independent influence of the GA region on RNA processing, stability, or translation when incorporated into the mRNA.

Previous investigation of polypurine-polypyrimidine sequences similar to the GA repeat offer clues to the mechanisms by which it might regulate transcription. Simple purine pyrimidine repeating sequences can, under appropriate conditions of high superhelicity and/or low pH, undergo a structural transition to form an intramolecular triplex. The resulting structure, designated H-DNA (Johnston, 1988; Wells et al., 1988; Htun and Dahlberg, 1989), consists of a normal double helix to which a third strand is bound within the major groove via Hoogsteen base pairing, leaving the complement of the Hoogsteen base-paired strand single stranded and sensitive to the single-strand-specific nuclease S1. Homopurine-homopyrimidine sequences of this type occur frequently in eukaryotic genomes (Birboim et al., 1979; Behe, 1987), often in the vicinity of gene promoters,

and are often correlated with the formation of S1-sensitive sites indicative of a non-B DNA structure (McKeon et al., 1984; Siegfried et al., 1986; Boles and Hogan, 1987).

It has recently been shown for the *c-myc* gene that a potential H-DNA-forming region that behaves as a positive *cis*-acting transcription element and adopts an S1-sensitive conformation *in vivo* also binds a factor with the properties of a ribonucleoprotein (Davis et al., 1989). In an immunoglobulin switch region, transcription through a potential H-DNA-forming sequence has been shown to generate a complementary RNA transcript that can stabilize the H-DNA triplex by base pairing with the region of single-stranded DNA generated by its formation, resulting in formation of stable, less supercoiled DNA conformations (Reaban and Griffin, 1990). The presence of the GA repeat within a transcribed region of the GAP-43 gene raises the possibility that a similar mechanism could operate to stabilize H-DNA formation in this gene.

Polypurine-polypyrimidine tracts may also influence transcription by other mechanisms. In the *Drosophila hsp 26* gene, for example, a polypurine-polypyrimidine sequence is required *in vivo* for optimal transcription (Glaser et al., 1990), but further experiments suggested that the effect of the CT/GA repeat on expression may not involve formation of H-DNA structure *in vivo*, but that the primary sequence per se is critical for function.

It remains to be determined whether the GA repeating sequence in the GAP-43 promoter region can adopt the H-DNA conformation *in vivo* and, if so, whether this conformational switch is responsible for the transcriptional effects of the polypurine-polypyrimidine element. A novel feature of the GAP-43 gene, however, is that its promoter lies between such a homopurine-homopyrimidine sequence and a second sequence element with the potential to undergo a transition to an alternative non-B DNA structure, the left-handed Z-DNA conformation (Rich et al., 1984). Z-DNA-forming sequences also occur frequently in eukaryotic genomes and also have been proposed to play a role as *cis*-acting transcriptional regulators. Both H-DNA and Z-DNA are induced or stabilized by negative supercoiling, and each relaxes negative supercoiling in adjacent regions of DNA (Rich et al., 1984; Wells et al., 1988). Structural transitions to non-B DNA conformations, and the resulting torsional effects on adjacent DNA, therefore offer one possible mechanism for the functional influences of these elements on each other and on the promoter lying between them.

Although the downstream polypurine-polypyrimidine region and the upstream fragment containing the GT repeat can exert an influence on the promoter lying between them, the regulatory role of these interacting influences is not yet apparent. A DNA fragment containing the core promoter without these flanking elements can exhibit at least some forms of neural-selective expression. Interactions of the upstream and downstream flanking elements, however, might further modulate the cell type specificity of GAP-43 expression. Alternatively, the interaction of these regulatory elements could be involved in modulating expression of the GAP-43 gene within a defined cell type, such as its axon-dependent decline within differentiated neurons late in development. Further dissection of regulatory features mediated by the core promoter and by specific *cis*-acting elements flanking it should provide access to signaling pathways mediating induction of this gene during neural differentiation and its responses to axon disruption, which appears to be closely linked to signals controlling neuron competence for axon growth.

References

- Alexander KA, Wakim BT, Doyle GS, Walsh KA, Storm DR (1988) Identification and characterization of the calmodulin binding domain of neuromodulin, a neurospecific calmodulin-binding protein. *J Biol Chem* 263:7544–7549.
- Anderson DJ (1990) The neural crest cell lineage problem: neurogenesis? *Neuron* 3:1–12.
- Apel ED, Byford MF, Au D, Walsh KA, Storm DR (1990) Identification of the protein kinase C phosphorylation site in neuromodulin. *Biochemistry* 29:2330–2335.
- Basi GS, Jacobson RD, Virag I, Schilling J, Skene JHP (1987) Primary structure and transcriptional regulation of GAP-43, a protein associated with nerve growth. *Cell* 49:785–791.
- Behe MJ (1987) The DNA sequence of the human β -globin region is strongly biased in favor of long strings of contiguous purine or pyrimidine residues. *Biochemistry* 26:7870–7875.
- Benton W, Davis R (1977) Screening λ gt recombinant clones by hybridization to single plaques *in situ*. *Science* 196:180–182.
- Birnboim HC, Sederoff RR, Paterson MC (1979) Distribution of polypyrimidine-polypurine segments in DNA from diverse organisms. *Eur J Biochem* 98:301–307.
- Bjelfman C, Meyerson G, Cartwright CA, Mellstrom K, Hammerling U, Pahlman S (1990) Early activation of endogenous pp60^{src} kinase activity during neuronal differentiation of cultured human neuroblastoma cells. *Mol Cell Biol* 10:361–370.
- Blake CCF (1978) Do genes-in-pieces imply proteins-in-pieces? *Nature* 273:267.
- Boles TC, Hogan ME (1987) DNA structure equilibria in the human *c-myc* gene. *Biochemistry* 26:367–376.
- Breathnach R, Chambon P (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu Rev Biochem* 50:349–383.
- Buonassisi V, Sato G, Cohen AI (1962) Hormone-producing cultures of adrenal and pituitary tumor origin. *Proc Natl Acad Sci USA* 48:1184–1190.
- Davis TL, Firulli AB, Kinniburgh AJ (1989) Ribonucleoprotein and protein factors bind to an H-DNA-forming *c-myc* DNA element: possible regulators of the *c-myc* gene. *Proc Natl Acad Sci USA* 86:9682–9686.
- Deloulme J-C, Janet T, Au D, Storm DR, Sensenbrenner M, Baudier J (1990) Neuromodulin (GAP-43): a neural protein kinase C substrate is also present in O-2A glial cell lineage. Characterization of neuromodulin in secondary cultures of oligodendrocytes and comparison with the neuronal antigen. *J Cell Biol* 111:1559–1570.
- Devereux J, Haerberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12:387–395.
- Dichter MA (1978) Rat cortical neurons in cell culture: culture methods, cell morphology, electrophysiology, and synapse formation. *Brain Res* 149:279–293.
- Doster SK, Lozano AM, Aguayo AJ, Willard MB (1991) Expression of the growth-associated protein GAP-43 in adult rat retinal ganglion cells following axon injury. *Neuron* 6:1–13.
- Edlund T, Walker MD, Barr PJ, Rutter WJ (1985) Cell-specific expression of the rat insulin gene: evidence for role of two distinct 5' flanking elements. *Science* 230:912–916.
- Federoff HJ, Grabczyk E, Fishman MC (1988) Dual regulation of GAP-43 gene expression by nerve growth factor and glucocorticoids. *J Biol Chem* 263:19290–19295.
- Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 132:6–13.
- Fischer JA, Maniatis T (1988) *Drosophila* Adh: a promoter element expands the tissue specificity of an enhancer. *Cell* 53:451–461.
- Foster J, Stafford J, Queen C (1985) An immunoglobulin promoter displays cell-type specificity independently of the enhancer. *Nature* 315:423–425.
- Frischauf A-M, Lehrach H, Poustka A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. *J Mol Biol* 170:827–842.
- Furley AJ, Morton SB, Manalo D, Karagogeos D, Dodd J, Jessell TM (1990) The axonal glycoprotein TAG-1 is an immunoglobulin superfamily member with neurite outgrowth-promoting activity. *Cell* 61:157–170.

- Gilbert W (1978) Why genes in pieces? *Nature* 271:501.
- Glaser RL, Thomas GH, Siegfried E, Elgin SCR, Lis JT (1990) Optimal heat-induced expression of the *Drosophila* hsp26 gene requires a promoter sequence containing (CT)_n·(GA)_n repeats. *J Mol Biol* 211:751–761.
- Go M (1983) Modular structural units, exons and function in chicken lysozyme. *Proc Natl Acad Sci USA* 80:1964–1968.
- Gorman CM, Merlino GT, Willingham MC, Pastan I, Howard BH (1982) The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA mediated transfection. *Proc Natl Acad Sci USA* 79:6777–6781.
- Grabczyk E, Zuber MX, Federoff HJ, Ng S-C, Pack A, Fishman MC (1990) Cloning and characterization of the rat gene encoding GAP-43. *Eur J Neurosci* 2:822–827.
- Greene LA, Tischler A (1976) Establishment of a nonadrenergic clonal line of rat pheochromocytoma cells which respond to nerve growth factor. *Proc Natl Acad Sci USA* 73:2424–2428.
- Hammer RE, Krumlauf R, Camper SA, Brinster RL, Tilghman SM (1987) Diversity of alpha-fetoprotein gene expression in mice is generated by a combination of separate enhancer elements. *Science* 235:53–58.
- Henikoff S (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene (Amst)* 28:351–359.
- Hoffman PN (1989) Expression of GAP-43, rapidly transported growth-associated protein, and class II beta tubulin, a slowly transported cytoskeletal protein, are coordinated in regenerating neurons. *J Neurosci* 9:893–897.
- Hoffman PN, Cleveland DW (1988) Neurofilament and tubulin expression recapitulates the developmental pattern during axonal regeneration: induction of a specific β -tubulin isotype. *Proc Natl Acad Sci USA* 84:4530–4533.
- Htun H, Dahlberg JE (1989) Topology and formation of triple-stranded H-DNA. *Science* 243:1571–1576.
- Johnston BH (1988) The S1-sensitive form of d(C-T)_n·d(A-G)_n: chemical evidence for a three-stranded structure in plasmids. *Science* 241:1800–1804.
- Kadesch T, Berg P (1986) Effects of position of the simian virus 40 enhancer on expression of multiple transcription units in a single plasmid. *Mol Cell Biol* 6:2593–2601.
- Kam J, Brenner S, Barnett L (1983) New bacteriophage lambda vectors with positive selection for cloned inserts. *Methods Enzymol* 101:3–21.
- Karns LR, Ng S-C, Freeman JA, Fishman MC (1987) Cloning of complementary DNA for GAP-43, a neuronal growth-related protein. *Science* 236:597–600.
- Kosik KS, Orecchio LD, Bruns GAP, MacDonald GP, Cox DR, Neve RL (1988) Human GAP-43: its deduced amino acid sequence and chromosomal localization in mouse and human. *Neuron* 1:127–132.
- LaBate ME, Skene JHP (1989) Selective conservation of GAP-43 structure in vertebrate evolution. *Neuron* 3:299–310.
- Lewis SA, Cowan NJ (1986) Anomalous placement of introns in a member of the intermediate filament multigene family: an evolutionary conundrum. *Mol Cell Biol* 6:1529–1534.
- Liu Y, Chapman ER, Storm DR (1991) Targeting of neuromodulin (GAP-43) fusion proteins to growth cones in cultured rat embryonic neurons. *Neuron* 6:411–420.
- Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular cloning: a laboratory manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Maniatis T, Goodbourn S, Fischer JA (1987) Regulation of inducible and tissue-specific gene expression. *Science* 236:1237–1245.
- Mar JH, Antin PB, Cooper TA, Ordahl CP (1988) Analysis of the upstream regions governing expression of the chicken cardiac troponin T gene in embryonic cardiac and skeletal muscle cells. *J Cell Biol* 107:573–585.
- Maue RA, Kraner SD, Goodman RH, Mandel G (1990) Neuron-specific expression of the rat brain type II sodium channel gene is directed by upstream regulatory elements. *Neuron* 4:223–231.
- McKay RDG (1989) The origins of cellular diversity in the mammalian nervous system. *Cell* 58:815–821.
- McKeon C, Schmidt A, de Crombrughe B (1984) A sequence conserved in both the chicken and mouse $\alpha 2(I)$ collagen promoter contains sites sensitive to S1 nuclease. *J Biol Chem* 259:6636–6640.
- Meakin SO, Reddan JR, Tsui L, Breitman M (1989) A rabbit lens epithelial cell line supports expression of an exogenous crystallin gene characteristic of lens fiber cell differentiation. *Exp Eye Res* 48:131–137.
- Miller FD, Naus CCG, Durand M, Bloom FE, Milner RJ (1987) Iso-types of α -tubulin are differentially regulated during neuronal maturation. *J Cell Biol* 105:3065–3073.
- Miller FD, Tetzlaff W, Bisby MA, Fawcett JW, Milner RJ (1989) Rapid induction of the major embryonic α -tubulin mRNA in adults following neuronal injury. *J Neurosci* 9:1452–1463.
- Mitchell PJ, Tjian R (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245:371–378.
- Moreau P, Hen R, Wasyluk B, Everett R, Gaub MP, Chambon P (1981) The SV40 72bp repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* 9:6047–6068.
- Mori N, Stein R, Sigmund O, Anderson DJ (1990) A cell type-preferred silencer element that controls the neural-specific expression of the SCG10 gene. *Neuron* 4:583–594.
- Mount SM (1982) A catalogue of splice junction sequences. *Nucleic Acids Res* 10:459–472.
- Muller SR, Sullivan PD, Clegg DO, Feinstein SC (1990) Efficient transfection and expression of heterologous genes in PC12 cells. *DNA Cell Biol* 9:221–229.
- Neve RL, Perrone-Bizzozero NL, Finklestein S, Zwiers H, Bird E, Kurnit DM, Benowitz LI (1987) The neuronal growth-associated protein GAP-43 (B-50, F1): neuronal specificity, developmental regulation, and regional distribution of the human and rat mRNAs. *Mol Brain Res* 2:177–183.
- Nordeen SK, Green PP III, Fowlkes DM (1987) A rapid, sensitive, and inexpensive assay for chloramphenicol acetyltransferase. *DNA* 6:173–178.
- Ornitz DM, Hammer RE, Davidson BL, Brinster RL, Palmiter RD (1987) Promoter and enhancer elements from the rat elastase I gene function independently of each other and of heterologous enhancers. *Mol Cell Biol* 7:3466–3472.
- Pelletier J, Sonnenberg N (1985) Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell* 40:515–526.
- Pfeiffer SE, Wechsler W (1972) Biochemically differentiated neoplastic clone of Schwann cells. *Proc Natl Acad Sci USA* 69:2885–2889.
- Reaban ME, Griffin JA (1990) Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature* 348:342–344.
- Rich A, Nordheim A, Wang AHJ (1984) The chemistry and biology of left handed Z-DNA. *Annu Rev Biochem* 53:791–846.
- Rosenthal A, Chan SY, Henzel W, Haskell C, Kuang W-J, Chen E, Wilcox JN, Ullrich A, Goeddel DV, Routtenberg A (1987) Primary structure and mRNA localization of protein F1, a growth related protein kinase C substrate associated with synaptic plasticity. *EMBO J* 6:3641–3646.
- Sanger F, Nickelen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467.
- Sargent TD, Wu J, Sala-Trepat JM, Wallace RB, Reyes AA, Bonner J (1979) The rat serum albumin gene: analysis of cloned sequences. *Proc Natl Acad Sci USA* 76:3256–3260.
- Sauerwald A, Hoesche C, Oschwald R, Kilimann MW (1990) The 5' flanking region of the synapsin I gene. *J Biol Chem* 265:14932–14937.
- Sehgal A, Patil N, Chao M (1988) A constitutive promoter directs expression of the nerve growth factor receptor gene. *Mol Cell Biol* 8:3160–3167.
- Siegfried E, Thomas GH, Bond UM, Elgin SCR (1986) Characterization of a supercoil-dependent S1 sensitive site 5' to the *Drosophila melanogaster* hsp 26 gene. *Nucleic Acids Res* 14:9425–9445.
- Skene JHP (1989) Axonal growth-associated proteins. *Annu Rev Neurosci* 12:127–156.
- Skene JHP, Virag I (1989) Posttranslational membrane attachment and dynamic fatty acylation of a neuronal growth cone protein, GAP-43. *J Cell Biol* 108:613–624.
- Skene JHP, Willard M (1981a) Changes in axonally transported proteins during axon regeneration in toad retinal ganglion cells. *J Cell Biol* 89:86–95.
- Skene JHP, Willard M (1981b) Axonally transported proteins associated with axon growth in rabbit central and peripheral nervous systems. *J Cell Biol* 89:96–103.
- Stein R, Mori N, Matthews K, Lo L-C, Anderson DJ (1988) The

- NGF-inducible SCG10 mRNA encodes a novel membrane-bound protein present in growth cones and abundant in developing neurons. *Neuron* 1:463–476.
- Tetzlaff W, Alexander SW, Miller FD, Bisby MA (1991) Response of facial and rubrospinal neurons to axotomy: changes in mRNA expression for cytoskeletal proteins and GAP-43. *J Neurosci* 11:2528–2544.
- Thompson EB, Tomkins GM, Curran JF (1966) Induction of tyrosine α -ketoglutarate transaminase by steroid hormones in a newly established tissue culture cell line. *Proc Natl Acad Sci USA* 56:296–303.
- Thompson M, Ziff EB (1989) Structure of the gene encoding peripherin, an NGF-regulated neuronal-specific type III intermediate filament protein. *Neuron* 2:1043–1053.
- Vitkovic L, Steisslinger HW, Aloyo VJ, Mersel M (1988) The 43-kDa neuronal growth-associated protein (GAP-43) is present in plasma membranes of rat astrocytes. *Proc Natl Acad Sci USA* 85:8296–8300.
- Wasylyk B, Wasylyk C, Augereau P, Chambon P (1983) The SV40 72bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. *Cell* 32:503–514.
- Wells RD, Collier DA, Hanvey JC, Shimizu M, Wohlrab F (1988) The chemistry and biology of unusual DNA structures adopted by oligopurine·oligopyrimidine sequences. *FASEB J* 2:2939–2949.
- Zuber MX, Strittmatter SM, Fishman MC (1989) A membrane-targeting signal in the amino terminus of the neuronal protein GAP-43. *Nature* 341:345–348.