Molecular Brain Research, 19 (1993) 293–302

BRESM 70636

Structure of the human gene for the neural phosphoprotein B-50 (GAP-43)

H.B. Nielander, P.C. De Groen, B.J.L. Eggen, L.H. Schrama, W.H. Gispen and P. Schotman

Division of Molecular Neurobiology, Laboratory for Physiological Chemistry, Rudolf Magnus Institute, Institute of Molecular Biology and Medical Biotechnology, Utrecht University, Utrecht (The Netherlands)

(Accepted 23 February 1993)

Key words: B-50; GAP-43; Neuromodulin; Growth-associated protein; Gene structure; Human gene; Transcription start

The genomic DNA encoding the exons for the human neural phosphoprotein B-50 (GAP-43) was isolated using rat-based cDNA probes and oligonucleotides. Exons 2 and 3 were isolated from a genomic library, exon 1 was amplified by PCR on total genomic DNA. The gene consists of 3 exons and 2 large introns. The first exon encodes the N-terminal 10 amino acids of B-50 involved in membrane association of the protein. Exon 2 encodes the main part of the protein with the sites for protein kinase C-mediated phosphorylation and calmodulin binding, and includes a 10 amino acid residue insert not found in rodents. Exon 3 encodes the last 29 amino acid residues. The reported sequence extends the known cDNA structure to both the 5′ and 3′ ends. The 358 bp region upstream of the translational initiation codon, containing the main transcription starts, is purine-rich and does not include TATA or GC boxes. At the 3′ end potential polyadenylation signals were found 510 bp and 584 bp downstream of the stopcodon in exon 3. The 5′ end of the mRNA is heterogeneous in length, with primer extension products corresponding to a 5′ untranslated region of 159 and 343 bases. Northern hybridizations, however, indicate that the majority of B-50 mRNA has a shorter 5′ untranslated region, as was reported for the rat (Schrama et al., *Soc. Neurosci. Abstr.*, 18 (1992) 333.4). The structural organization of the human gene is similar to that described for the rat (Grabczyk et al., *Eur. J. Neurosci.* 2 (1990) 822–827), and both translated and untranslated regions show a high degree of sequence homology to the rat gene.

INTRODUCTION

B-50 (GAP-43) is a nervous-tissue specific phosphoprotein (M_r 51,000, pI 4.7 in human) implicated in axonal growth, regeneration and neuronal plasticity (see reviews[4,9,19,25,45,46]). The expression of B-50 is high during neuronal development in rat and mouse[6,7,21,29,30] and strongly decreases in most neuronal cells in adulthood, whereas levels of expression remain high in associative regions of the human brain[5,30,31].

Two incomplete cDNA sequences containing the open reading frame for human B-50 have been published[23,31]. These showed a high degree of homology to rat B-50 cDNA[3,32,37]. B-50 is encoded by a single-copy gene[3], localized on chromosome 3 of the human genome[23]. Cloning of the rat B-50 gene has recently been reported[20,28,41].

In 1986, Jacobson et al.[21] suggested transcriptional regulation of B-50 expression during development of the nervous system. The available data, however, are contradictory. Indeed, a close correlation between mRNA and protein levels has been observed in many studies (e.g. refs. 3, 13, 29, 30, 48). Basi et al.[3] showed that changes in B-50 expression are matched by changes in the rate of transcription in isolated rat cortex nuclei, while others using similar nuclear run-on assays did not find indications for an important transcriptional regulation during either cortical development in rat or optic nerve regeneration in goldfish[6,35]. Federoff et al.[16] provided evidence for transcriptional regulation of B-50 expression by glucocorticoids in rat PC12 cells, but Costello et al.[10,11] were unable to confirm this. Increased B-50 levels are associated with enhanced mRNA stability in NGF- or phorbol ester-induced rat

Correspondence: P. Schotman, Division of Molecular Neurobiology, Laboratory for Physiological Chemistry, Rudolf Magnus Institute, Institute of Molecular Biology and Medical Biotechnology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. Fax: (31) 30-537797.
Abbreviations: AMV, avian myeoloblastosis virus; GAP-43, growth associated protein-43; MMLV, moloney murine leukemia virus; NGF, nerve growth factor; PCR, polymerase chain reaction; RT, reverse transcriptase.

PC12 cells[6,10,11,16,35], suggesting that B-50 expression can be regulated at the posttranscriptional level.

To determine the complete mRNA structure and genomic organization of human B-50, we isolated the major part of the gene from a genomic library and obtained the first exon by specific amplification on total human DNA using the polymerase chain reaction (PCR)[38]. We then determined transcription starts using primer extension and northern hybridizations.

Part of this study appeared in abstract form[33,41].

MATERIALS AND METHODS

All radiochemicals were obtained from Amersham. Enzymes were from Pharmacia unless stated otherwise. An amplified human genomic library in λEMBL3[17], prepared from chromosomal DNA of chronic myeloid leukemia cells, was kindly provided by Dr. G. Grosveld (Erasmus University, Rotterdam, The Netherlands). Genomic DNA from human placental tissue and rat liver were obtained from Clontech, USA. Human brain poly(A)$^+$ RNA (15-year-old Caucasian female) was obtained from Clontech.

Phage λ and plasmid isolations, restriction enzyme analysis and subcloning of fragments were performed essentially according to standard protocols[39].

Oligonucleotides

Oligonucleotides were purified either by HPLC, by gel filtration on Sephadex G-25 (Pharmacia LKB), or by using an oligonucleotide purification cartridge (OPC, Applied Biosystems). Sequences and positions are shown in Table I. Oligonucleotides were 5′ end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase.

Blotting and hybridizations

Plaque lifts, Southern and northern blots were prepared on nylon membranes (Hybond-N, Amersham) according to the supplier. RNA was denatured using glyoxal/DMSO[39]. Hybridizations with [α-^{32}P]dCTP-labeled DNA probes (random-primed labeling kit, Boehringer) or [α-^{32}P]CTP-labeled RNA probes (T7 RNA polymerase transcripts, Promega) were performed in 6× SSC, 50% formamide, 1% SDS, 5× Denhardt's solution and 100 μg/ml denatured herring sperm DNA at 42°C (DNA) or 45–48°C (RNA). The final wash was in 0.1× or 0.2× SSC, 0.1% SDS at 65°C. For oligonucleotides, formamide was omitted and hybridization and washing were carried out at various temperatures dependent on the calculated T_m of the oligonucleotide.

Polymerase chain reaction (PCR)

A typical PCR reaction contained 15 pmol of each primer and 1 U Taq DNA polymerase (Perkin-Elmer Cetus) in a volume of 50 μl, and consisted of 30 cycles: denaturing 1 min at 94°C, annealing 1 min at 50–65°C, extension for 1 min (cycles 1–20) or 2 min (cycles 21–30) at 72°C in a thermal cycler (Pharmacia LKB-Gene ATAQ Controller). Amplification was completed by a final incubation at 72°C for 10 min.

PCR fragments were purified on low-melting point agarose gels (Nusieve GTG, FMC bioproducts) in TBE, blunt-ended with T4 DNA polymerase, and ligated into the *Sma*I site of pGEM7 (Promega). Alternatively, subcloning in pCR1000 (TA cloning system, Invitrogen Corporation) was carried out without prior purification of the PCR products.

Primer extension

Primer extension experiments were performed on 0.5 μg of human poly(A)$^+$ RNA essentially according to standard protocols[39], with avian meyeloblastosis virus (AMV) reverse transcriptase (RT) or moloney murine leukemia virus (MMLV)-RT. The oligonucleotide used was located in exon 2 (position +65 to +89 relative to the startcodon, Table I). Following heat denaturation of RNA and oligonucleotide for 5 min at 80°C, other components were added and extension was continued in the presence of [α-^{32}P]dCTP for 2 h at 37°C. Products were analyzed on denaturing 6% polyacrylamide-urea gels.

Sequencing

Subcloned genomic fragments were sequenced by the dideoxynucleotide chain termination method[40], with a T7 DNA polymerase sequencing kit (Pharmacia). Fragments were analyzed on a 6% polyacrylamide 7 M urea sequencing gel. For computer-aided data analyses, Microgenie (Beckman) and Lasergene (DNAstar Ltd., London) were used.

RESULTS

Isolation of genomic exon 2 and 3 clones from a library

Screening of approximately 300,000 plaques of a human genomic library in λEMBL3 with a rat B-50

TABLE I

Oligonucleotides

[1] The codes refer to the species (R = rat, H = human), the DNA strand (S = sense, identical to mRNA; A = antisense), and the number of the 5′ base of the oligonucleotide in the corresponding genomic sequence. [2] Small capitals represent non-homologous nucleotides added to the 5′ end to introduce endonuclease recognition sites (underlined). [3] Nucleotide numbering of the homologous nucleotides in the human cDNA sequence[23] or rat genomic sequence[42]. The first base of the startcodon is numbered +1.

Code [1]	*Sequence 5′ to 3′* [2]	*Position* [3]
5′ RS-791	CTCCACAATTATCAACTGCACACAA	−791 to −767
RS-440	GGAAGCTAGTGAACAATTCTGAGAA	−440 to −416
RS-242	AGAAATGCATATGCGGTGAGCAATA	−242 to −218
RS-214	GATCGCTGTAGACCTTACAGTTGCTGCTAACTGCCCT	−214 to −178
RS-175	TGTGTGTGAGGGAGAGAGAGGGA	−175 to −153
HS-84	AGAAAAGAGGTGGAGAGGGGGGG	−84 to −62
RS-4	CACCATGCTGTCCGGTATGAGAAG	−4 to +20
HS+1089	TGTGTGTGCAATGTTCCGTTC	+1089 to +1109
3′ RA-88	CTTGCTCACTCGCTCTCTCGCGCT	−111 to −88
HA-64	AGC GGATCC CCCTCTCCACCTCTTTTC	−83 to −64
HA+30	TG GAATTC TGTTTGGTTCTTCTCATACAGC	+8 to +30
RA+89	TTATCCTCCGGTTTGACACCATCTT	+65 to +89
RA+1221	AT GAATTC AATATTCAGATTGAAAAAGGAGCCTA	+1221 to +1247

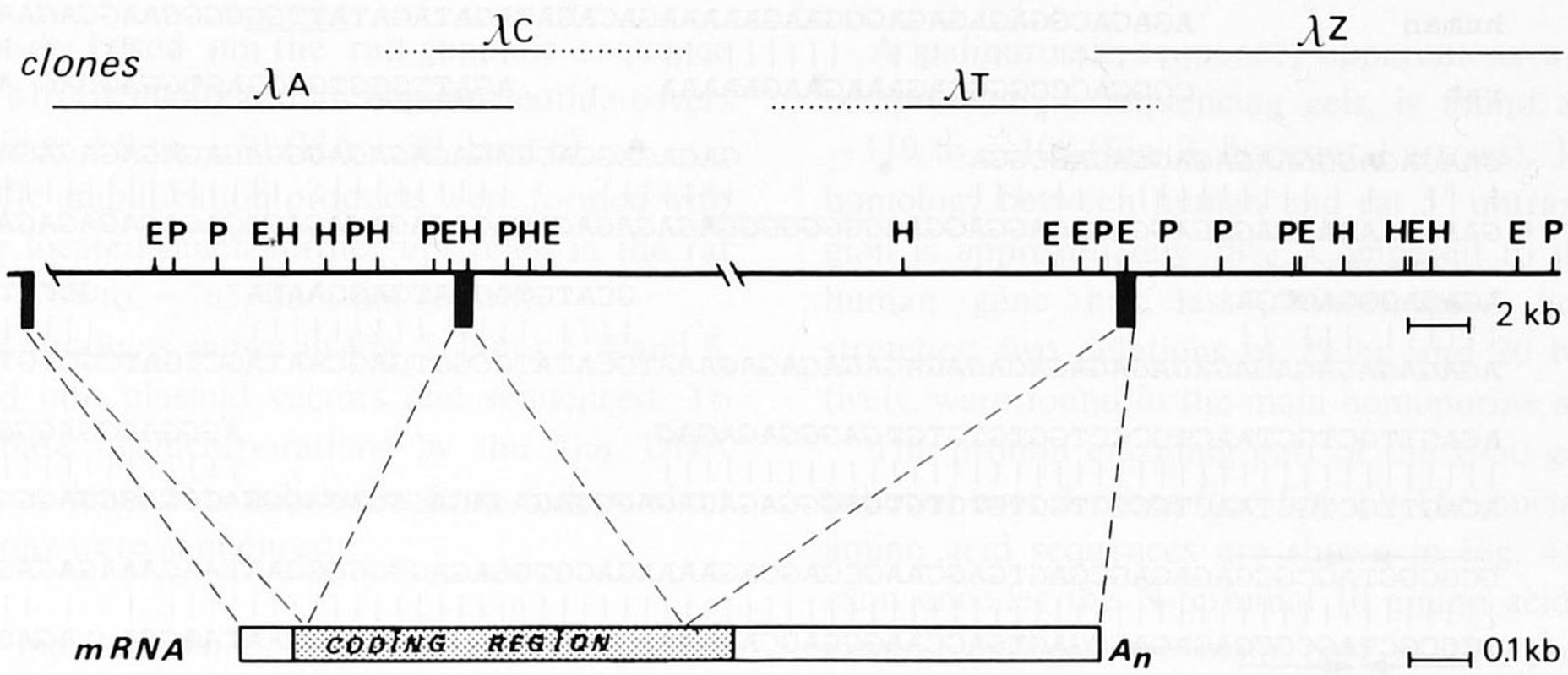

Fig. 1. Structure of the human B-50 gene. The location of genomic clones and a partial restriction map of the human gene for endonucleases *Eco*RI (E), *Hin*dIII (H), and *Pst*I (P) are given. Dotted lines indicate regions in which not all restriction sites were determined. Exons are depicted as dark vertical bars, connected to the corresponding regions in the (proposed) mature mRNA shown directly below. The shaded area is the B-50 coding region: Exon 1 amino acids 1–10, exon 2 amino acids 11–209, exon 3 amino acids 210–238.

cDNA (nucleotides −40 to 1085, numbers relative to the translational start site; see[32]) yielded 4 positive clones. Screening of another 250,000 plaques of the library with a cRNA probe containing the first 222 bases of the rat cDNA (−40 to 182) yielded 3 additional positive phages.

Genomic clones were purified and analyzed by restriction enzyme digestions, Southern blotting and hybridization with cDNA probes. A partial restriction map for the human gene and the position of four of the genomic clones is shown in Fig. 1. The other clones contained either exon 2 or exon 3 and were not further analyzed. No overlap was found between clones containing different exons, indicating a minimal size of 30 kbp for the human B-50 gene (Fig. 1). Restriction fragments containing exon 2 (*Hin*dIII 3.2 kb and *Pst*I 2.4 kb) or exon 3 (*Eco*RI 1.6 kb, *Pst*I 2.3 kb; Fig. 1) were subcloned for sequence analysis.

Hybridization of the genomic clones with an oligonucleotide based on the rat exon 1 sequence −4 to +20 indicated that none of the clones contained the most 5′ sequence of the reported human B-50 cDNAs (data not shown).

Amplification of exon 1 by PCR

Since we did not succeed in isolation of genomic clones containing exon sequences upstream of cDNA base +30 from the human genomic library, we ampli-

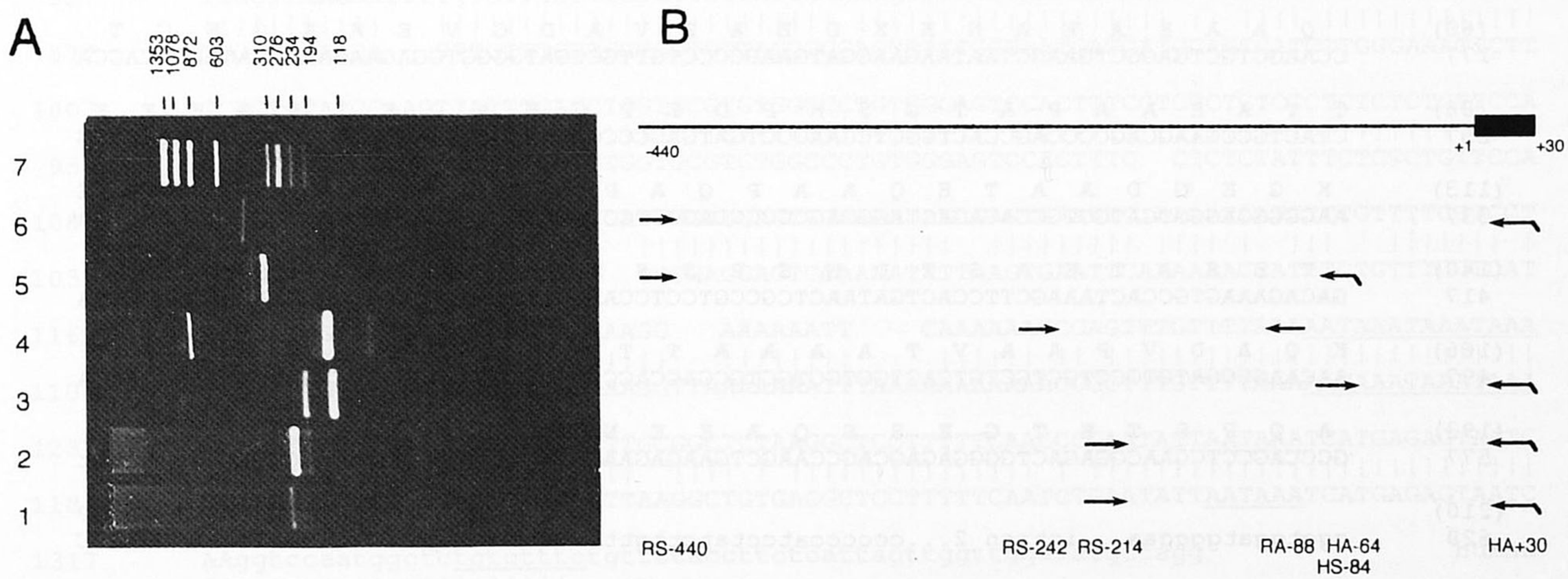

Fig. 2. Amplification of exon 1 by PCR. A: amplification products from human (lanes 2–6) and rat (lane 1) genomic DNA were separated on a 2% agarose gel (ethidium bromide staining). Samples were run from left to right. B: oligonucleotides used for the PCR in the corresponding lanes, and the position of these primers in the rat sequence above. The bold line represents the protein-encoding part of exon 1. Oligonucleotides RS-440, RS-242, RS-214 and RA-88 are based on rat sequences, HS-84, HA-64 and HA+30 are derived from the human cDNA sequence (see Table I). Lane 7 shows the *Hae*III fragments of ΦX-174 DNA.

```
-358   human   AGAGACGGAGAGAGAGGGAAGAAAAAGACAGATAGATAGATATTGGGGGGAAGGAGAAAAAAGGA
               |||| | ||| |||||||||                  |||| |    || | | ||||| | | ||||
-415   rat     CGGGACCCGGGCAGAAAGAAGAAAAA            AGATTGGGTGGGGAGTGGAGGAG AGAGAGGA

-293   GAAGAGAGGGAAGAGAGGACAGCGGA                GAGAGAGCACCAGAGAGAGAGGGAGAGAGAGAGAGAGCGCTAG
       |||| | |||||||||| || |||                 |||||||    |||||||||| |||||||||||||| |  ||
-358   GAAGGAAAAGAAGAGAGGAAAGAGGAGGAAGGGGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG

-224   AGAGAGGGAGCGA                                GCATGTGCGATGAGCAATA       GCTGTGGACCTT
       |||||| ||| ||                               |||| |||| |||||||||       ||||| ||||||
-278   AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAATGCATATGCGGTGAGCAATAGCTGATCGCTGTAGACCTT

-180   ACAGTTGCTGCTAACTGCCCTGGTGTGTGTGAGGGAGAGAG                          AGGGAGGGAGGGAGAGAGA
       |||||||||||||||||||||||||||||||||||||||||                          |||||||||||||| ||||
-198   ACAGTTGCTGCTAACTGCCCTGGTGTGTGTGAGGGAGAGAGAGGGAGAGAGAGGGAGAGGGAGGGAGGGAGGGAGGGAGA

-120   GCGCGCTAGCGCGAGAGAGCGAGTGAGCAAGCGAGCAGAAAAGAGGTGGAGAGGGGGGGAATAAGAAAGAGAGAGAAGGA
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |  | ||||| |||
-118   GCGCGCTAGCGCGAGAGAGCGAGTGAGCAAGCGAGCAGAAAAGAGGTGGAGAGGGGGGGAATAAGGA  AGAGAGGAGGA

-40    AAGGAGAGAAGGCAGGAAGAAGGCAAGGGACGAGACAACCATG                                human
       |||||||||||||||||||||||||| ||| || || ||||||
-40    AAGGAGAGAAGGCAGGAAGAAGGCAGGGGAAGATACCACCATG                                rat
```

Fig. 3. Sequence of the human B-50 gene upstream of the translation start, aligned with the rat genomic sequence. The first base of the startcodon (bold) is numbered +1. A palindromic region is indicated by horizontal arrows. Two uninterrupted homopurine stretches are in italics. The inverse complement of the CCAAT sequence is underlined.

fied the missing sequence directly on total genomic DNA by PCR. The oligonucleotides used for the amplification (Fig. 2B, and Table I) were based on human cDNA[23] or rat genomic sequences[41]. PCR on human genomic DNA with several primer combinations resulted in specific fragments (Fig. 2A, lanes 2–6), as confirmed by hybridization with a rat exon 1 probe (not shown). These fragments were of approximately the size expected for rat genomic DNA (e.g. Fig. 2A, lanes 1 and 2). The largest fragment was obtained with an

```
(1)          M  L  C  C  M  R  R  T  K  Q
1           ATGCTGTGCTGTATGAGAAGAACCAAACAGgtag...intron 1...gtcattgaagccctctctttt

(11)         V  E  K  N  D  D  D  Q  K  I  E  Q  D  G  I  K  P  E  D  K  A  H
31          ttttctcgacaaagGTTGAAAAAAATGATGACGACCAAAAGATTGAACAAGATGGTATCAAACCAGAAGATAAAGCTCAT

(33)         K  A  A  T  K  I  Q  A  S  F  R  G  H  I  T  R  K  K  L  K  G  E  K  K  D  D  V
97          AAGGCCGCAACCAAAATTCAGGCTAGCTTCCGTGGACACATAACAAGGAAAAAGCTCAAAGGAGAGAAGAAGGATGATGT

(60)         Q  A  A  E  A  E  A  N  K  K  D  E  A  P  V  A  D  G  V  E  K  K  G  E  G  T
177         CCAAGCTGCTGAGGCTGAAGCTAATAAGAAGGATGAAGCCCCTGTTGCCGATGGGGTGGAGAAGAAGGGAGAAGGCACCA

(86)        T  T  A  E  A  A  P  A  T  G  S  K  P  D  E  P  G  K  A  G  E  T  P  S  E  E  K
257         CTACTGCCGAAGCAGCCCCAGCCACTGGCTCCAAGCCTGATGAGCCCGGCAAAGCAGGAGAAACTCCTTCCGAGGAGAAG

(113)        K  G  E  G  D  A  A  T  E  Q  A  A  P  Q  A  P  A  S  S  E  E  K  A  G  S  A  E
337         AAGGGGGAGGGTGATGCTGCCACAGAGCAGGCAGCCCCCCAGGCTCCTGCATCCTCAGAGGAGAAGGCCGGCTCAGCTGA

(140)        T  E  S  A  T  K  A  S  T  D  N  S  P  S  S  K  A  E  D  A  P  A  K  E  E  P
417         GACAGAAAGTGCCACTAAAGCTTCCACTGATAACTCGCCGTCCTCCAAGGCTGAAGATGCCCCAGCCAAGGAGGAGCCTA

(166)       K  Q  A  D  V  P  A  A  V  T  A  A  A  A  T  T  P  A  A  E  D  A  A  A  K  A  T
497         AACAAGCCGATGTGCCTGCTGCTGTCACTGCTGCTGCTGCCACCACCCCTGCCGCAGAGGATGCTGCTGCCAAGGCAACA

(193)        A  Q  P  P  T  E  T  G  E  S  S  Q  A  E  E  N  I
577         GCCCAGCCTCCAACGGAGACTGGGGAGAGCAGCCAAGCTGAAGAGAACATAGgtgagcaaccgcgagggtcagatgcaat

(210)                                                                E  A  V  D  E  T  K  P
629         gggtggatggggaa...intron 2...ccccccatcctatcttgttttctttctcagAAGCTGTAGATGAAACCAAACC

(218)        K  E  S  A  R  Q  D  E  G  K  E  E  E  P  E  A  D  Q  E  H  A  *
651         TAAGGAAAGTGCCCGGCAGGACGAGGGTAAAGAAGAGGAACCTGAGGCTGACCAAGAACATGCCTGA
```

Fig. 4. Sequence of the protein encoding part of the human B-50 gene. Start and stop codon are printed bold and intron sequences in lowercase. Exon sequences are numbered starting at the ATG codon. The amino acid sequence is shown above the nucleotide sequence, with numbering in parentheses.

oligonucleotide based on the rat genomic sequence −440 to −416 (RS-440) and an oligonucleotide covering nucleotides +8 to +30 (HA + 30, lane 6).

No specific amplification products were formed with a 5′ primer located much further upstream in the rat sequence (−789 to −765, data not shown).

The PCR products shown in Fig. 2, lanes 2, 4 and 5, were cloned into plasmid vectors and sequenced. To check for base misincorporations by the *Taq* DNA polymerase at least two subclones from independent PCR reactions were sequenced.

Sequence of the B-50 gene

In Fig. 3, the human sequence upstream of the B-50 startcodon is shown, and aligned with the rat genomic sequence[41] (this sequence is identical to that published by Grabczyk et al.[20], except for an additional AG repeat in the purine stretch upstream of position −240[42]). The region upstream of the translation start is purine rich, consisting for 81% of A and G residues. In the 358 bp we sequenced upstream of the translation start no consensus TATA or GC boxes are present. The only well characterized transcription regulatory element[15,22,27,49] is a CCAAT box at position −317 in the non-coding strand.

A palindromic sequence, apparent as a region of compression in sequencing gels, is found at position −119 to −108 (Fig. 3, horizontal arrows). The overall homology between human and rat 5′ untranslated region is approximately 78%. Compared to the rat the human gene has less uninterrupted homopurine stretches; two deletions of 29 bp and 20 bp, respectively, were found in the main homopurine stretches.

The protein encoding part of the B-50 gene is distributed over 3 exons (see Fig. 1). The nucleotide and amino acid sequences are shown in Fig. 4. The first exon encodes the N-terminal 10 amino acids of B-50, exon 2 encodes the main part of the protein (199 amino acids and the first base of codon 210), and exon 3 encodes the remainder of the total of 238 amino acid residues. The sequences at the exon–intron borders conform to primate consensus splice sites[43]. Exon 1 was isolated using PCR, the sequence at the exon 1/intron 1 boundary was obtained from a genomic clone[12].

In Fig. 5 the untranslated part of the human exon 3 sequence is shown in alignment with the rat sequence[20,37]. Polyadenylation signals are found at positions 1223 and 1297 (Fig. 5, underlined). The sequence of 150 bp 3′ to the sequence shown in Fig. 5

```
 715   human                      TGAACTCTAAGAAATGGCTTTCCACATCCCCACCCTCCCCTCTCCTGAGCCTGT
                                  |||||| ||||||| |||||||||| |   ||| | |||     | |||||
 679   rat                        TGAACTTTAAGAAACGGCTTTCCACGT   TGCCCCCACCT    GAACCCTGT

 769   CTCTCCCTACCCTCTTCTCAGCTCCACTCTGAAG TCCCT TCCTGTCCTGCTCACGTCTGTGAGTCTGTCCTTTCCCAC
       |||| ||| |||| ||||||| |||||||||||| | ||| |||||||||||||||| |||||| ||||||| ||| ||
 726   CTCT CCTGCCCT TTCTCAGATCCACTCTGAAGTTTCCTCTCCTGTCCTGCTCACGTGTGTGAGCCTGTCCTCTCCTAC

 847   CCACTAGCCCTCTTTCTCTCTGTGTGGCAAACATTTAAAAAAAAAAAAAAAAAGCAGGAAAGATCCCAAGTCAAACAGTG
       | |  |||||||  |||||||||||||||||||||||||||||||||||     |||||||||||||||||||| |||||||
 804   CTATGAGCCCTC  TCTCTCTGTGTGGCAAACATTTAAAAAAAAAAAA     GCAGGAAAGATCCCAAGTCCAACAGTG

 927   TGGCTTAAACATTTTTTGTTTCTTGGTGTTGTTATGGCAAGTTTTTGGTAATGATGATTCAATCATTTTGGGAAATTCTT
       |||||||||||   |||||||||||||||||||||||| |||||||||||||||||||| || |||| |||||||||||||
 877   TGGCTTAAACA   TTTGTTTCTTGGTGTTGTTATGGCGAGTTTTTGGTAATGATGATGCAGTCATCTTGGGAAATTCTT

1007   GCACTGTATCCAAGTTATTTGATCTGGTGCGTGTGGCCCTGTGGGAGTCCACTTTCCTCTCTCTCTCTCTCTCTGTTCCA
       |||||||| ||  ||| |||||||||||||||||||||||||||||||||||||||||  |||||| | |||||||||||||
 954   GCACTGTACCCCGGTTTTTTGATCTGGTGCGTGTGGCCCTGTGGGAGTCCACTTTC  CTCTCTATTTCTCTCTGTTCCA

1087   A  GTGTGTGTGCAATGTTCCGTTCATCTGAGGAGTCCAAAATATCGAGTGAATTC AAAATCATTTTTGTTTTCCTCCT
       |  |||||||||||||||||||   |||||||||||||||||||  ||||||||| |||| |  |||   ||||||| |
1032   AGTGTGTGTGTGCAATGTTCCGT   TCTGAGGAGTCCAAAATATTAAGTGAATTCAAAAAACCATTTCTGTTTCCTCAT

1164   TTTCAATGTGATGGAATGAACAAAAAGG   AAAAAATT    CAAAAAACCCAGTTTGTTTTAAAAATAAATAAATAAA
       |||||||||||||||||||||||||||   ||||||||     |||||| || |||||||||||||||||||||||||||
1109   TTTCAATGTGATGGAATGAACAAAAAGGTTAAAAAAATTTAAAAAAAAAAACCGGTTTGTTTTAAAAATAAATAAATAAA

1237   GCAAATGTGCCAATTAGCGTAAACTTGCGGCTCTAAGGCTCCTTTTTCAACCCGAATATTAATAAATCATGAGAGTAATC
       |||||||||||||||||||||| |||  |||| | ||||||||||||||| | |||||||||||||||||||||||||||
1189   GCAAATGTGCCAATTAGCGTAA CTTAAGGCTGTGAGGCTCCTTTTTCAATCTGAATATTAATAAATCATGAGAGTAATC

1317   AAggtccaatggctctgtgtttctgttctaccttctcattagttggttggacatgctagg                human
       ||  |    ||  |||||||||  |
1268   AActtt ggtgt tctgtgtttctt                                                     rat
```

Fig. 5. Sequence of the human B-50 gene downstream of the stopcodon (in bold), aligned with the rat genomic sequence. Regions of potential secondary structure are indicated by horizontal arrows. Poly(A) signals and a GT-motif known to follow poly(A) sites are single underlined, a motif implicated in transcript degradation is double underlined[44]. Sequences beyond the proposed polyadenylation sites are printed in lowercase.

revealed no further polyadenylation signals (not shown). The second polyadenylation signal is followed by a sequence motif which is known to occur immediately downstream of polyadenylation sites (Fig. 5, underlined). This sequence is in perfect agreement with both the consensus sequence (YGTGTTYY), and position (24–30 bp downstream of the poly(A) signal) of this motif found in the majority of mammalian genes[26].

The potential stem-loop forming structures in the 3′ untranslated part of exon 3 with the highest free energy content are indicated by horizontal arrows. These are found at comparable positions in human and rat. In the alignment of Fig. 5, 521 out of 604 bases (86%) could be matched.

Primer extension

As the human B-50 sequence up to −358 contains no obvious signals for transcriptional initiation we determined the transcription starts experimentally. Primer extension on total human RNA was performed with a rat B-50-specific oligonucleotide located in exon 2 using two different reverse transcriptases. This resulted in multiple bands, with their intensity varying between experiments. Moreover, different results were obtained with the 2 transcriptases. As shown in Fig. 6, the longest product apparent with both enzymes, corresponding to a 5′ untranslated region of 343 b, was most prominent with MMLV-RT. We are not sure of the nature of the products longer than 343 bases, since these were only found using MMLV, and even were not present in all extension experiments with MMLV. Moreover, we were not able to confirm these products by RT-coupled PCR. The rather strong signal at the top of the figure is close to the origin and might represent material which has not properly entered the gel. Several prominent bands obtained with AMV-RT (e.g. at −273) were not found with MMLV-RT, indicating more premature stops in the reverse transcription with AMV-RT.

No indications for an additional intron were found; the sequence lacks consensus splice signals and PCR on the extension products with different oligonucleotide combinations gave specific fragments of the expected length. Moreover, sequencing of reverse transcriptase-PCR products confirmed their uninterrupted homology with the genomic sequence (data not shown).

Northern blotting

The 5′ heterogeneity of the mRNA was studied by northern hybridization as well. Human brain poly(A)$^+$ RNA was hybridized with two probes containing genomic sequences upstream of the translation start (Fig 7, lanes 3). Probe A (panel A) recognizes mRNA

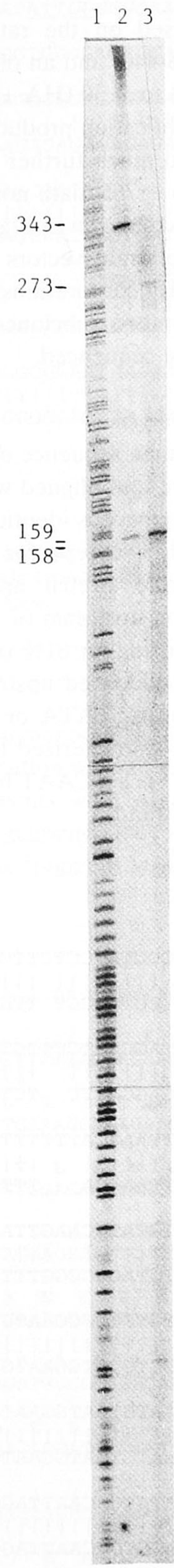

Fig. 6. Primer extension. Primer extension products on human poly(A)$^+$ RNA (0.5 μg) using MMLV-reverse transcriptase (lane 2) or AMV-reverse transcriptase (lane 3). The oligonucleotide used was located in exon 2 (+65 to +89, Table I). Lane 1 contains a length marker, numbers alongside the figure indicate the length upstream of the translation start.

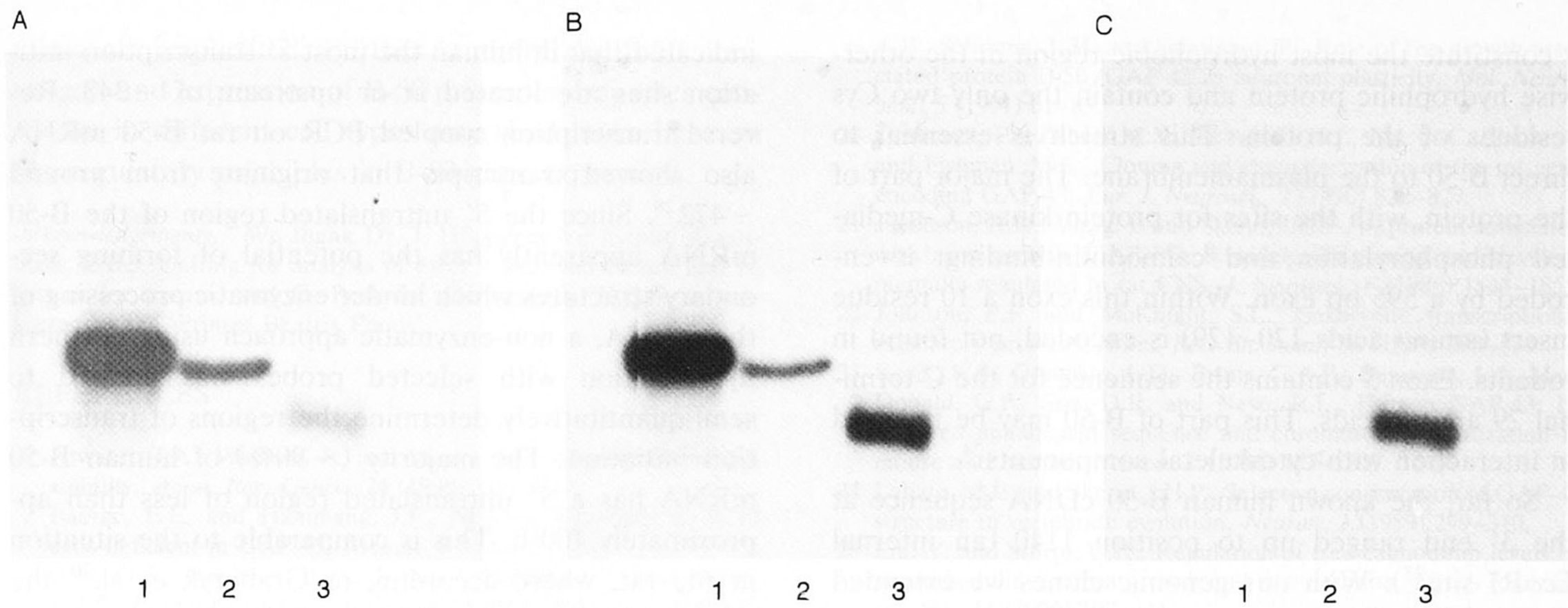

Fig. 7. Northern hybridization. Autoradiogram of a northern blot with human poly(A)$^+$ RNA (2 μg; lane 3), hybridized to three different probes. Lanes 1 and 2 contain 250 and 25 pg, respectively, of a cRNA derived from the human genomic sequence −212 to +30 and approximately 1300 bases of vector-derived sequence. A: hybridization with the sequence −223 to −86, exposure time 72 h. B: hybridization with the sequence from −84 to +30, exposure time 48 h. C: hybridization with the sequence +1089 to +1296, exposure time 18 h. Numbers refer to the human sequence, +1 is the first base of the B-50 start codon.

upstream of −86, probe B (panel B) mRNA downstream of −84. The specific hybridization signal obtained with probe A confirms the presence of sequences upstream of −86 in the poly(A)$^+$ RNA. However, with probe B, of comparable length and specific activity, a much stronger hybridization signal was obtained in the poly(A)$^+$ RNA. A control transcript generated on the genomic exon 1 sequence (lanes 1 and 2) was used as an internal standard to correct for differences in GC content of the probe and hybridization conditions. Both probes have approximately 120 b complementary to the control transcript. In comparison with the signal obtained with 25 pg control transcript (lanes 2), the signals with probe A and probe B on the human mRNA (lanes 3) were estimated by densitometry to be approximately 5-fold less and 2-fold more, respectively. This indicates that 90% of the B-50 messengers contain a 5′ untranslated region of approximately 86 b or less. The probe in panel C is a 207 bp fragment directly upstream of the second polyA signal, a sequence not described for the human cDNA so far. The strong hybridization signal at the same position as found with probe B indicates most or all transcripts contain this sequence. Since this fragment is not represented in the controll transcript quantification of the hybridization signal is not possible.

DISCUSSION

Screening of a human genomic library with rat B-50 cDNA probes resulted in clones for exons 2 and 3. An additional 5′ exon, containing the coding sequence for the first 10 amino acids and the 5′ untranslated region, was missing. This could be due to the rather small size of the homologous region in the rat probes used (70 bp with 94% homology). Alternatively, this region of the human gene might be poorly represented in the amplified library as was found for rat exon 1[28]. Southern hybridization indicated the presence of human sequences highly homologous to a 5′ part of the rat gene (sequence −214 to −178, data not shown). This prompted us to directly amplify sequences by PCR using rat-based oligonucleotides as 5′ primers, and human-based oligonucleotides on the 3′ side (Fig. 2). Multiple partial overlapping independent clones were sequenced, resulting in extension of the 5′ sequence of the gene to 358 bp upstream of the translation start. No deviating sequences were found. We did use the PCR-derived sequences to rescreen an unamplified library, resulting in a number of clones for B-50 exon 1. Partial sequence analysis confirmed our PCR-based sequence analysis.

The genomic sequence presented here confirms the cDNA sequence reported by Kosik et al.[23], except for a C instead of a T at position 1130, as determined from two independent genomic clones. Ng et al.[31] reported a human cDNA sequence deviating upstream of position −44. This might be a cloning artefact, since no consensus 3′ splice site can be recognized at the point of divergence between the two cDNA sequences, and no homologous counterpart of the upstream sequence can be found in the human or the rat B-50 gene.

The intron/exon structure of the human gene is very similar to that of the rat[20], and correlates with proposed functional domains in the protein (refs. see reviews[9,19,25,45,46]). The 10 amino acids encoded by exon

1 constitute the most hydrophobic region in the otherwise hydrophilic protein and contain the only two Cys residues of the protein. This stretch is essential to direct B-50 to the plasmamembrane. The major part of the protein, with the sites for protein kinase C-mediated phosphorylation and calmodulin-binding, is encoded by a 595 bp exon. Within this exon a 10 residue insert (amino acids 120–129) is encoded, not found in rodents. Exon 3 contains the sequence for the C-terminal 29 amino acids. This part of B-50 may be involved in interaction with cytoskeletal components[24].

So far, the known human B-50 cDNA sequence at the 3′ end ranged up to position 1140 (an internal *Eco*RI site[23]). With our genomic clones we extended the sequence further downstream beyond 2 consensus poly(A) signals in exon 3 starting at positions 1223 and 1297. Northern blotting confirmed the presence of sequences 3′ of the internal *Eco*RI site in the mRNA (Fig. 7C). Considering the strict sequence requirements for poly(A) signals[36], the absence of poly(A) signals further downstream, the presence of a consensus GT-motif following the poly(A) signal, and the apparent total length of B-50 messenger RNA (~1.5 kb), it seems likely that one or both of the closely spaced poly(A) signals are used. Up to 16 bp downstream of the second poly(A) signal human and rat sequences are well conserved (Fig. 5). This makes the last conserved residues (positions 1318 to 1320) good candidates for a poly(A) site. For the rat, two B-50 cDNAs have been reported with poly(A) tails; one starting 21 bp downstream of the first poly(A) signal[3], and one starting 15 bp downstream of the second[37].

Untranslated regions in mRNA are often involved in stabilization of transcripts[1]. In rat PC12 cells, an increase in B-50 mRNA stability upon induction with NGF or TPA has been reported[10,11,16,35]. In addition, a B-50-deficient PC12 cell line has been described with a B-50 mRNA truncated at the 3′ untranslated region by 200–300 bases[2]. Comparison of human and rat B-50 exon sequences shows that large parts of the 5′ and 3′ untranslated regions are as highly conserved as the coding regions, with an overall homology of approximately 79% and 87%, respectively (Figs. 3 and 5). Two potential stem loop-forming sequences noted in the 3′ untranslated region of human B-50 cDNA[31] have homologous counterparts in the rat (Fig. 5, see also ref. 35). Both, human and rat contain an internal poly(A) stretch in exon 3, preceded by the sequence ATTTA (Fig. 5, double underlined). This motif has been implicated in selective degradation of messenger RNAs[18,44].

Determination of the B-50 transcription starts with classical methods (RNAse protection, S1 nuclease protection) gave no consistent results. Primer extension indicated that in human the most 5′ transcription initiation sites are located at or upstream of −343. Reverse transcription coupled PCR on rat B-50 mRNA also showed transcripts that originate from around −472[28]. Since the 5′ untranslated region of the B-50 mRNA apparently has the potential of forming secondary structures which hinder enzymatic processing of the mRNA, a non-enzymatic approach using northern hybridization with selected probes was applied to semi-quantitatively determine the regions of transcription initiation. The majority (~90%) of human B-50 mRNA has a 5′ untranslated region of less than approximately 100 b. This is comparable to the situation in the rat, where according to Grabczyk et al.[20] the major start sites are found between −50 and −60 (see also[42]).

In human neuroblastoma cell lines two B-50 messengers of 1.4 and 1.6 kb have been reported [34]. In our experiments the transcripts hybridizing with probe A (−223 to −86) were longer than those hybridizing with probe B (−84 to +30, Fig. 7), however this difference seemed less than 100 b.

The apparent length of human mRNA was previously determined as being 1.4 to 1.5 kb, a similar length is calculated from the northern blot shown in Fig. 7. At present it is difficult to understand the difference between the observed 1.4–1.5 kb and the calculated length of the transcript with a 5′ untranslated region of 343 b (length is 343 bases + 1318 bases + poly(A) tail). Possible explanations are that the longest transcripts are below the detection limit or that the B-50 mRNA is difficult to denature on agarose gels.

Our sequence data indicate that no known cis-acting elements involved in transcriptional regulation are present near the main initiation sites. Promoters of genes with multiple transcription starts often lack TATA and/or CCAAT boxes[14,47]. Like other TATA-less promoters these genes usually contain GC-boxes (Sp1 binding sites) very close to the transcriptional start site(s). These GC boxes promote efficient transcription, and may regulate start site utilization[8]. At present it is not clear how initiation sites in the B-50 gene are selected.

Recently for the rat gene a core promoter has been proposed, located within the sequence −752 to −366[28]. This would implicate a long distance between promoter and main transcription start sites[20,42]. Results from our laboratory indicate promoter activity also in a more downstream rat sequence (Eggen et al., in preparation).

For the specific pattern of expression of B-50 during neuronal development and regeneration various transcriptional and posttranscriptional regulatory mecha-

nisms are implicated. Further study of the B-50 messenger, and expression of B-50 promoter/reporter constructs in different cell types may shed more light on the regulatory principles in B-50 gene expression.

Acknowledgements. We thank Dr. J. Höppener for helpful discussions, E.B.J. Ueffing for analysis of exon 3, A.J. van Rozen and N. Fresen for figures, and G. Peek for photography. H.B.N. is supported by the 'Prinses Beatrix Fonds'.

REFERENCES

1 Atwater, J.A., Wisdom, R. and Verma, I.M., Regulated mRNA stability, *Annu. Rev. Genet.*, 24 (1990) 519–541.

2 Baetge, E.E. and Hammang, J.P., Neurite outgrowth in PC12 cells deficient in GAP-43, *Neuron*, 6 (1991) 21–30.

3 Basi, G.S., Jacobson, R.D., Virág, I., Schilling, J. and Skene, J.H.P., Primary structure and transcriptional regulation of GAP-43, a protein associated with nerve growth, *Cell*, 49 (1987) 785–791.

4 Benowitz, L.I. and Routtenberg, A., A membrane phosphoprotein associated with neural development, axonal regeneration, phospholipid metabolism, and synaptic plasticity, *Trends Neurosci.*, 10 (1987) 527–532.

5 Benowitz, L.I., Perrone-Bizzozero, N.I., Finklestein, S.P. and Bird, E.D., Localization of the growth-associated phosphoprotein GAP-43 (B-50, F1) in the human cerebral cortex, *J. Neurosci.*, 9 (1989) 990–995.

6 Benowitz, L.I. and Perrone-Bizzozero, N.I., The expression of GAP-43 in relation to neuronal growth and plasticity: when, where, how and why?, *Progr. Brain Res.*, 89 (1991) 69–87.

7 Biffo, S., Verhaagen, J., Schrama, L.H., Schotman, P., Danho, W. and Margolis, F.L., B-50/GAP-43 expression correlates with process outgrowth in the embryonic mouse nervous system, *Eur. J. Neurosci.*, 2 (1990) 487–499.

8 Blake, M.C., Jambou, R.C., Swick, A.G., Kahn, J.W. and Azizkhan, A.C., Transcriptional initiation is controlled by upstream GC-box interactions in a TATA-less promoter, *Mol. Biol.*, 10 (1990) 6632–6641.

9 Coggins, P.J. and Zwiers, H., B-50 (GAP-430): biochemistry and functional neurochemistry of a neuron-specific phoshoprotein, *J. Neurochem.*, 56 (1991) 1095–1106.

10 Costello, B., Meymandi, A. and Freeman, J.A., Factors influencing GAP-43 gene expression in PC12 pheochromocytoma cells, *J. Neurosci.*, 10 (190) 1398–1406.

11 Costello, B., Lin, L-H., Meymandi, A., Bock, S., Norden, J.J. and Freeman, J.A., Expression of the growth- and plasticity-associated neuronal protein, GAP-43, in C 12 pheochromocytoma cells, *Progr. Brain Res.*, 89 (1991) 47–67.

12 De Groen, P., Schrama, L.H., Nielander, H.B., Gispen, W.H. and Schotman, P., The complete gene encoding human B-50/GAP-43, *Soc. Neurosci. Abstr.*, 18 (1992) 333.2.

13 De la Monte, S., Federoff, H.J., Ng, S.-C., Grabczyk, E. and Fishman, M.C., GAP-43 gene expression during development: persistence in a distinctive set of neurons in the mature central nervous system, *Dev. Brain Res.*, 46 (1989) 161–168.

14 Dynan, W.S., Promoters for housekeeping genes, *Trends Gen.*, 2 (1986) 196–197.

15 Faisst, S. and Meyer, S., Compilation of vertebrate-encoded transcription factors, *Nucl. Acids Res.*, 20 (1992) 3–26.

16 Federoff, H.J., Grabczyk, E. and Fishman, M.C., Dual regulation of GAP-43 gene expression by nerve growth factor and glucocorticoids, *J. Biol. Chem.*, 263 (1988) 19290–19295.

17 Frischauf, A.-M., Lehrach, H., Poustka, A. and Murray, N., Lambda replacement vectors carrying polylinker sequences, *J. Mol. Biol.*, 170 (1983) 827–842.

18 Gillis, P. and Malter, J.S., The adenosine-uridine binding factor recognizes the AU-rich elements of cytokine, lymphokine, and oncogene mRNAs, *J. Biol. Chem.*, 266 (1991) 3172–3177.

19 Gispen, W.H., Nielander, H.B., De Graan, P.N.E., Oestreicher, A.B., Schrama, L.H. and Schotman, P., Role of the growth-associated protein B-50/GAP-43 in neuronal plasticity, *Mol. Neurobiol.*, 5 (1991) 61–85.

20 Grabczyk, E., Zuber, M.X., Federoff, H.J., Ng, S.-C., Pack, A. and Fishman, M.C., Cloning and characterization of the rat gene encoding GAP-43, *Eur. J. Neurosci.*, 2 (1990) 822–827.

21 Jacobson, R.D., Virág, I. and Skene, J.H.P., A protein associated with axon growth, GAP- 43, is widely distributed and developmentally regulated in rat CNS, *J. Neurosci.*, 6 (1986) 1843–1855.

22 Johnson, P.F. and McKnight, S.L., Eukaryotic transcriptional regulatory proteins, *Annu. Rev. Biochem.*, 58 (1989) 799–839.

23 Kosik, K.S., Orecchio, L.D., Bruns, G.A.P., Benowitz, L.I., MacDonald, G.P., Cox, D.R. and Neve, R.L., Human GAP-43: Its deduced amino acid sequence and chromosomal localization in mouse and human, *Neuron*, 1 (1988) 127–132.

24 Labate, M.E. and Skene, J.H.P., Selective conservation of GAP-43 structure in vertebrate evolution, *Neuron*, 3 (1989) 299–310.

25 Liu, Y. and Storm, D.R., Regulation of free calmodulin levels by neuromodulin: neuron growth and regeneration, *Trends Pharmacol. Sci.*, 11 (1990) 107–111.

26 McLauchlan, J., Gaffney, D., Whitton, J.L. and Clements, J.B., The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini, *Nucl. Acids Res.*, 13 (1985) 1347–13666.

27 Mitchell, P.J. and Tjian, R., Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins, *Science*, 245 (1989) 371–245.

28 Nedivi, E., Basi, G.S., Virag, I and Skene, J.H.P., A neural-specific GAP-43 core promoter located between unusual DNA elements that interact to regulate its activity, *J. Neurosci.*, 12 (1992) 691–704.

29 Neve, R.L., Perrone-Bizzozero, N.I., Finklestein, S., Zwiers, H., Bird, E., Kurnit, D.M. and Benowitz, L.I., The neuronal growth-associated protein GAP-43 (B-50, F1): use of cDNAs to show the neuronal specificity, developmental regulation and regional distribution of the human and rat mRNAs, *Mol. Brain Res.*, 2 (1987) 177–183.

30 Neve, R.L., Finch, E.A., Bird, E.D. and Benowitz, L.I., Growth-associated protein GAP-43 is expressed selectively in associative regions of the adult human brain, *Proc. Natl. Acad. Sci. USA*, 85 (1988) 3638–3642.

31 Ng, S.-C., De la Monte, S.M., Conboy, G.L., Karns, L.R. and Fishman, M.C., Cloning of human GAP-43: growth association and ischemic resurgence, *Neuron*, 1 (1988) 133–139.

32 Nielander, H.B., Schrama, L.H., Van Rozen, A.J., Kasperaitis, M., Oestreicher, A.B., De Graan, P.N.E., Gispen, W.H. and Schotman, P., Primary structure of the neuron-specific phosphoprotein B-50 is identical to growth-associated protein GAP-43, *Neurosci. Res. Commun.*, 1 (1987) 163–172.

33 Nielander, H.B., De Groen, P., Ueffing, E.B.J., Eggen, B.J.L., Schrama, L.H., Schotman, P. and Gispen W.H., Gene and mRNA structure of human B-50 (GAP-43), *Soc. Neurosci. Abstr.*, 17 (1991) 101.4.

34 Örtoft, E., Betsholtz, C., Pahlman, S. and Hammerling, U., Human GAP-43: multiple 5′ untranslated regions and distinct regulation in neuroblastoma cell lines, *Soc. Neurosci. Abstr.*, 18 (1992) 327.5.

35 Perrone-Bizzozero, N.I., Neve, R.L., Irwin, N., Lewis, S.E., Fischer, I. and Benowitz, L.I., Post-transcriptional regulation of GAP-43 mRNA levels during neuronal differentiation and nerve regeneration, *Mol. Cell. Neurosci.*, 2 (1991) 402–409.

36 Proudfoot, N.J., How RNA polymerase II terminates transcription in higher eukaryotes, *Trends Biochem. Sci.*, 14 (1991) 105–110.

37 Rosenthal, A., Chan, S.Y., Henzel, W., Haskell, C., Kuang, W.-J., Chen, E., Wilcox, J.N., Ullrich, A., Goeddel, D.V. and Routtenberg, A., Primary structure and mRNA localization of protein F1, a growth-related protein kinase C substrate associated with synaptic plasticity, *EMBO J.*, 6 (1987) 3641–3646.

38 Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Ehrlich, H.A., Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science*, 239 (1988) 487–495.

39 Sambrook, J., Fritsch, E.F. and Maniatis, T., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Lab., Cold Spring Harbor, New York, 1989.
40 Sanger, F., Nicklen, S. and Coulsen, A.R., DNA sequencing with chain termination inhibitors, *Proc. Natl. Acad. Sci. USA*, 74 (1977) 5463–5467.
41 Schrama, L.H., Nielander, H.B., Van Rozen, A.J., Schotman, P. and Gispen, W.H., Organization of the rat and the human B-50 (GAP-43) gene, *Soc. Neurosci. Abstr.*, 16 (1990) 154.11.
42 Schrama, L.H., Rensen-de Leeuw, M.G.A., Korver, W. and Eggen, B.J.L., Analysis of B-50/GAP-43 transcription starts by northern blotting and genetic variation in H-DNA length of the rat B-50 (GAP-43) gene, *Soc. Neurosci. Abstr.*, 18 (1992) 333.4.
43 Senapathy, P., Shapiro, M.B. and Harris, N.L., Splice junctions, branch point sites, and exons: sequence statistics, identifications and applications to genome project, *Methods Enzymol.*, 183 (1990) 252–280.
44 Shaw, G. and Kamen, R., A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation, *Cell*, 46 (1986) 659–667.
45 Skene, J.H.P., Axonal growth-associated proteins, *Annu. Rev. Neurosci.*, 12 (1989) 127–156.
46 Strittmatter, S.M., Vartanian, T. and Fishman, M.C., GAP-43 as a plasticity protein in neuronal form and repair, *J. Neurobiol.*, 23 (1992) 836–841.
47 Sutcliffe, J.G., mRNA in the mammalian central nervous system, *Annu. Rev. Neurosci.*, 11 (1988) 157–198.
48 Van der Zee, C.E.E.M., Nielander, H.B., Vos, J.P., Lopes da Silva, S., Verhaagen, J., Oestreicher, A.B., Schrama, L.H., Schotman, P. and Gispen, W.H., Expression of growth-associated protein B-50 (GAP43) in dorsal root ganglia and sciatic nerve during regenerative sprouting, *J. Neurosci.*, 9 (1989) 3505–3512.
49 Wingender, E., Compilation of transcription regulating proteins, *Nucl. Acids Res.*, 16 (1988) 1879–1891.