Cloning and Sequence Analysis of cDNA Encoding p38, a Major Synaptic Vesicle Protein

Kathleen M. Buckley, Eric Floor,* and Regis B. Kelly

Department of Biochemistry and Biophysics, University of California at San Francisco School of Medicine, San Francisco, California 94143-0448; and *Department of Anatomy, University of Wisconsin Medical School, Madison, Wisconsin 53706

**Abstract.** We have isolated from a lambda gt11 rat brain cDNA library cDNA clones encoding >95% of the open reading frame and untranslated regions of the mRNA for p38, the most abundant of the integral membrane proteins of the synaptic vesicle. Phage containing cDNA that encoded vesicle proteins were identified by screening fusion proteins with a polyclonal serum to rat brain synaptic vesicles. To identify phage carrying p38 sequences, fusion proteins were used to affinity purify monospecific antibodies from the original heterogeneous serum; antibodies to a 38,000-D protein were then identified by Western blotting. Inserts carrying DNA-encoding p38 sequences were subcloned into plasmid vectors and used to generate cDNA probes for Northern blot analysis. A major transcript of 2.4 kb was expressed specifically in brain and endocrine tissue but not in liver, consistent with the tissue-specific expression of the protein detected by antibody techniques. Using three overlapping clones that encoded fusion proteins, we identified and sequenced ~85% of the cDNA. Two additional Eco RI fragments at the 5' end of the mRNA were obtained from a fourth clone identified by screening a second lambda gt11 library with a 5' cDNA probe. The cDNA encoded an open reading frame of 298 amino acids with a 3' untranslated region of 1.4 kb. The protein shares no sequence homology with other Ca2+-binding proteins. The availability of a cDNA clone for an integral synaptic vesicle protein should facilitate studies of its function in transmitter release, its intracellular targeting, and regulation of synaptic vesicle biogenesis during development and regeneration of nerve terminals.

SYNAPTIC vesicles in all parts of the nervous system have in common three integral membrane proteins: p65 (Matthew et al., 1981), SV2 (Buckley and Kelly, 1985), and p38 (Wiedenmann and Franke, 1985; Jahn et al., 1985; Navone et al., 1986). In addition to being universally present in synaptic endings of a variety of transmitter types, these proteins are also highly conserved in vertebrate nervous systems, from cartilaginous fish to humans. Such widespread distribution and evolutionary conservation suggest the vesicle proteins play a critical but as yet unknown role in the biogenesis, packaging, transport, or fusion of synaptic vesicles. Despite the lack of functional information, the antibodies that have been generated against synaptic vesicle proteins have proved powerful tools in the study of secretory vesicle biogenesis (e.g., Navone et al., 1986; Lowe et al., 1988) and the development and assembly of synaptic connections (Greif and Reichardt, 1982; Knaus et al., 1986; Sarthy and Bacon, 1985). If cDNA probes were available for the synaptic-specific membrane proteins, some of the steps in neuronal development and differentiation could be investigated at the transcriptional level. Furthermore, the restriction of the membrane proteins to secretory vesicles implies a specific targeting mechanism. The availability of cDNA probes for the vesicle proteins will make possible an in vitro mutagenesis analysis, which is currently the favored strategy for exploring targeting domains. Finally, the amino acid sequence deduced from DNA sequence sometimes yields insight about function. Towards these ends, we have identified and sequenced 95% of the cDNA encoding the synaptic vesicle-specific membrane protein p38, a transmembrane glycoprotein that has been reported to be the major Ca2+-binding protein of the synaptic vesicle membrane (Rehm et al., 1986).

**Materials and Methods**

Identification of cDNA Clones with Antibody Probes

A rat brain cDNA library constructed in the expression vector lambda gt11 (kindly donated by Dr. Andrew Dowsett and Dr. Norman Davidson, California Institute of Technology) was screened according to the method of Huynh et al. (1985) with R10, a rabbit antiserum generated against rat brain synaptic vesicles. Positive clones were identified with horseradish peroxidase-conjugated goat anti-rabbit IgG (BioRad Laboratories, Richmond, CA).

To identify clones coding for p38, R10-positive phage were used to affinity-purify monospecific antibodies to p38 by the following procedure. Plaque-purified phage were plated on *Escherichia coli* Y1090, grown for 4 h at 42°C and induced with isopropyl β-D-thiogalactopyranoside—impregnated nitrocellulose filters at 37°C for 3 h. The filters with adsorbed fusion protein were then used to affinity-purify monospecific antibodies from the
heterologous serum by the technique of Smith and Fisher (1984). Eluted antibo-
dies were incubated overnight at room temperature with Western blots of
rat brain homogenates, separated by SDS-PAGE in 10% gels, and binding sites
were visualized with horseradish peroxidase–conjugated goat anti-
rabbit IgG. Affinity-purified antibodies that reacted with p38 in Western
blots were concentrated 10- to 20-fold by centrifugation in Centricon spin-
dialysis chambers (Amicon Corp., Danvers, MA) and tested for binding to
syncytia at sections of rat spinal cord by immunofluorescence (Buckley and
Kelly, 1985). Three overlapping clones coding for 85% of the
cDNA were obtained from this library.

The p38-specific monoclonal antibody, SY38 (Boehringer Mannheim Biochemica,
Biberach, Germany), was also used to screen fusion proteins as described.
This antibody immunoprecipitates intact synaptic vesicles and so
must recognize a cytoplasmic domain of p38 (Weidemann and Franke,
1985).

### Screening Libraries with cDNA Probes

A rat brain lambda gt11 cDNA library (Clontech, Palo Alto, CA) generated with random primers and constructed by using RNAs H for second-strand
synthesis was screened by plaque hybridization as described by Maniatis et
al. (1982). Hybridization was performed overnight at 42°C with ∼10 e cpm
of cDNA probe, generated by the random-primer extension method of Fein-
described. This antibody immunoprecipitates intact synaptic vesicles and so
is specific for synaptic vesicles in vivo.

### Subcloning and Sequencing

DNA from lambda gt11 clones was prepared either by the plate lystate method of Maniatis et al. (1982) or, in some cases, purified with an antibody
to the lambda phage head (LambdaSorb; Promega Biotec, Madison, WI),
digested with Eco RI, and the inserts were subcloned into either Bluescribe
(Strategene, San Diego, CA) or pUC8 plasmid vectors in the e. coli JM101.
For sequencing by the dideoxynucleotide termination method of Sanger et
al. (1977), DNA was subcloned into M13mp18 RF. Progressive deletions of the
cDNA were generated by the method of Henikoff (1984), except that mung
bean nuclease was used to remove single-stranded DNA. Certain regions of the
sequence were determined using synthetic oligonucleotides as primers. Sequences were also determined in the lamba phage directly
using commercial primers (Clontech and New England Biolabs, Beverly,
MA) and synthetic oligonucleotides by the method of Chen and Seeburg
(1985). DNA from four overlapping lambda clones was sequenced in both
directions to give the final sequence. The orientation of the four Eco RI frag-
ments in the final sequence was determined by sequence analysis of overlap-
ping clones and by sequencing across the three 5' Eco RI sites in the recom-
binant lambda phage B35 (Fig. 3). In addition, we attempted to determine
the NH2-terminal sequence by NH2-terminal protein sequencing. Edman
degradation of purified p38 was not successful due to apparent blockage of the
amino terminus (Buckley, K. M., L. Clift-O'Grady, and R. B. Kelly,
unpublished observations). We did not attempt primer extension sequencing
because of the low abundance of p38 mRNA in brain tissue.

### Northern Analysis

RNA was prepared from a variety of tissues by homogenization in 5 M
guanidinium isothiocyanate–5% 2-mercaptoethanol–10 mM EDTA–50 mM
Na Hepe, pH 7.4, followed by centrifugation through 5.7 M CsCl and the
pol/ctohroform extraction (Maniatis et al., 1982). RNA from cell cultures
was isolated by the method of Chelley and Anderson (1984). Pol yA mRNA was
selected by oligo d(T) cellulose chromatography. RNA was subject-
ed to electrophoresis in 1% agarose–2.2 M formaldehyde gels, transferred to
nitrocellulose or Nytran (Schleicher & Schuell, Keene, NH), and immobi-
]ized to nitrocellulose by baking or to Nytran by exposure to UV (Church
and Gilbert, 1984). Northern blots were probed with hybridized overnight with 10 cpm of cDNA probe (sp act of 106 cpm/μg) generated by the random-primer method of Feinberg and Vogelstein
(1983), and washed at a final stringency of 0.1 X SSC/0.1% SDS at 60°
Molecular weights of RNA transcripts were determined by comparison with RNA standards ranging in size from 0.3 to 9.5 kb (RNA ladder; Bethesda
Research Laboratories, Bethesda, MD).

### In Vitro Translation and Immunoprecipitation

RNA synthesized from cDNA inserts in a plasmid vector which contains
binding sites for both the T3 and T7 polymerase (Bluescribe, Stratagene),
and poly(A) mRNA isolated from rat brain were translated in a rabbit
reticulocyte lystate system in the presence of [35]methionine as suggested
by the manufacturer (BRL). Aliquots of the translation products were ana-
lyzed directly on 12% acrylamide–SDS gels or first immunoprecipitated
with either one of two polyclonal sera to rat brain synaptic vesicles (Joe B.,
generated by Dr. Daniel Cutler and Dr. Anson Lowe, and R10, generated
by Dr. Erik Floor) or a rabbit antiserum to rat brain–coated vesicles (Miro-
slaw, generated by Dr. Anson Lowe and Dr. Daniel Cutler), all of which re-
act predominantly with p38 in Western blots of rat brain membranes (unpub-
lished observations). For immunoprecipitation, samples were diluted 10-fold in
1% NP-40, 0.4% deoxycholate, 66 mM EDTA, 10 mM Tris-HCl, pH 7.4,
0.1% SDS, and incubated overnight at 4°C with 1-2 μl of antiserum. Im-
mune complexes were precipitated with the addition of 300 μl pansorbin
(Sapryptococcus aureus cells, Calbiochem-Behring Corp., La Jolla, CA)
prewashed with the detergent buffer, and washed three times by centrifuga-
tion through 15% sucrose in the detergent buffer. Gels were impregnated
with 1 M Na salicylate, dried, and exposed at −70°C with an intensifying
screen. The relative mobilities of proteins in SDS-PAGE were determined
by comparison with the following protein standards: bovine albumin
(66,000), ovalbumin (45,000), glyceraldehyde-3-phosphate dehydrogenase
(36,000), carbonic anhydrase (29,000), bovine pancreas trypsinogen
(24,000), trypsin inhibitor (20,100), and α-lactalbumin (14,200) (Sigma
Chemical Co., St. Louis, MO).

### Sequence Analysis

Analysis of the nucleic acid sequence was performed using the Bionet pro-
gram (IntelliGenetics, Inc., Palo Alto, CA). Membrane propensity predic-
tion (Kuhn and Leigh, 1985), hydropathy analysis using the values of
Eisenberg et al. (1982), and immunogenicity analysis of the secondary
structure were obtained with the generous assistance of Robert Stroud,
University of California, San Francisco. The secondary structure plots
shown in Fig. 9 were generated with MacGene Plus software (Applied
Genetics Technology, Fairview Park, OH), using the method of Garnier et
al. (1978) and hydropathy analysis of Kyte and Doolittle (1982). The hy-
drophobic domains determined with the membrane propensity values of
Kuhn and Leigh (1985) or with hydropathy analysis using the values of
Eisenberg et al. (1982) or Kyte and Doolittle were in close agreement. The
regions of the sequence underlined in Fig. 8 are an average of all three deter-
minations; the hydropathy plot displayed in Fig. 9 uses only the Kyte and
Doolittle values. Immunogenic regions of the amino acid sequence were
predicted using a window of +7 to −7 residues by the method of Finer-
Moore et al. (1988).

### Results

#### Identification of cDNA Clones Encoding p38

To identify cDNA clones encoding p38, a rat brain cDNA library constructed in the expression vector lambda gt11 was screened with R10, a polyclonal serum to rat brain synaptic vesicles. Because the R10 serum recognizes several polypep-
tides in addition to p38 (Fig. 1), the positive clones identified with the whole serum were separated into those cDNAs related to p38 and those that coded for other proteins recognized by the polyclonal serum. To determine which clones expressed p38 fusion protein, plaque-purified phage were plated at high density on 90-mm plates and fusion proteins induced by isopropyl β-d-thiogalactopyranoside were adsorbed to nitrocellulose filters. Antibodies specific for each fusion protein were isolated by incubating the filters overnight with the polyclonal serum. The filters were washed to remove unbound antibody, and the bound antibody was eluted with low-pH buffer (Smith and Fisher, 1984). To identify the vesicle protein identified by each monoclonal antibody, proteins in rat brain homogenates were separated by gel electrophoresis, transferred to nitrocellulose filters, and each filter was incubated overnight with the eluted antibody. The proteins in rat brains recognized by the affinity-purified antibodies were visualized with a horseradish peroxidase–
conjugated second antibody. 14 clones identified as positives
Immunoreactivity of antibodies affinity purified with lambda gt11 fusion proteins. Rat brain proteins were subjected to electrophoresis in preparative SDS gels and transferred to nitrocellulose. The filter was cut into strips, and each strip incubated with a different affinity-purified antibody. Binding sites were then visualized with horseradish peroxidase-conjugated goat anti-rabbit IgG. First six lanes show antibodies, affinity purified with lambda gt11 recombinant fusion proteins, that bind to p38. Lanes marked R/O were incubated with the heterologous polyclonal serum without affinity purification, and the lane marked 1-1 shows antibody that recognizes a 70-kD protein in rat brain (arrowhead).

To verify that the 38-kD protein was a synaptic vesicle protein, the same antibody solutions used in Western blots were used for immunofluorescence on cryostat sections of fixed, frozen rat spinal cord. As reported for the monoclonal antibody to p38 (Wiedenmann and Franke, 1985; Navone et al., 1986) and as expected for an antibody to a synaptic vesicle protein (Matthew et al., 1981; Buckley and Kelly, 1985) the immunofluorescence was confined to small varicosities that were distributed throughout the spinal cord and could be seen to outline cell bodies and dendrites (Fig. 2). This characteristically synaptic pattern of staining was seen with antibodies eluted from only one of the lambda phage fusion proteins, lambda 1-5. The remaining antibodies showed no binding above background, possibly because the supernatants were depleted after binding in the Western blots.

Figure 1. Immunoreactivity of antibodies affinity purified with lambda gt11 fusion proteins. Rat brain proteins were subjected to electrophoresis in preparative SDS gels and transferred to nitrocellulose. The filter was cut into strips, and each strip incubated with a different affinity-purified antibody. Binding sites were then visualized with horseradish peroxidase-conjugated goat anti-rabbit IgG. First six lanes show antibodies, affinity purified with lambda gt11 recombinant fusion proteins, that bind to p38. Lanes marked R/O were incubated with the heterologous polyclonal serum without affinity purification, and the lane marked 1-1 shows antibody that recognizes a 70-kD protein in rat brain (arrowhead).

Figure 2. Immunofluorescence localization of affinity-purified antibodies bound to sections of rat spinal cord. Antibodies eluted from the lambda phage 1-5 fusion proteins were incubated with 10-12 µm cryostat sections of rat spinal cord and localized with fluorescein-conjugated goat anti-rabbit IgG. (a) Phase micrograph of motoneuron cell body at the edge of the ventral horn. (b) Immunofluorescence micrograph of the same section as in a incubated with affinity-purified antibodies to p38. The cell body is surrounded with immunofluorescent varicosities. The white matter is unstained. (c) Immunofluorescence micrograph of dorsal horn of the spinal cord, showing that the staining is localized to synaptic structures in this region as well. Bars: (a and b) 15 µm; (c) 22 µm.
To identify overlapping clones, a cDNA probe was prepared from the largest Eco RI fragment, lambda 2-4B (see Fig. 3), and used to probe the remaining clones by Southern blotting. Lambda 2-4B and two of the largest inserts containing this sequence, lambda 1-5 and lambda 2-2, were subcloned into plasmid vectors and into M13 RF for sequence analysis by the dideoxy chain termination method (Sanger, 1977). As expected, the three clones contained overlapping sequence and together coded for an open reading frame of ∼200 amino acids followed by a 3' untranslated region of 1416 bp and a poly A+ tail (Fig. 3). The sequence of the Eco RI site at the 5' end of clone 2-4B was not that of the linkers used in the construction of the library, implying that this site was part of the p38 sequence and that an additional Eco RI fragment existed at the 5' end. To identify the remaining cDNA at the 5' end, the first 200 bp of lambda 2-4B was used to prepare a cDNA probe for screening a second lambda gtll library. The second library was made by random priming and without S1 nuclease, which should maximize the chances of cloning the 5' end. Out of 20 positive clones, five containing additional sequence to the 5' end of the probe were identified. In all five clones an Eco RI fragment of 110 bp was located proximal (5') to the sequence obtained from lambda 2-4B (Fig. 3, Eco RI site 2-site 3). Three of these clones, lambda B35, lambda B51, and lambda B7, contained another Eco RI fragment at the 5' end, but although the sequence of this fragment was homologous in all three clones, only the 160-bp Eco RI fragment in lambda B35 coded for an open reading frame (Fig. 3, Eco RI site 1-2). The terminal RI fragment in B51 was homologous to B35 except for two deletions: between positions 2 and 7 and between positions 50 and 92. The 5' sequence of B7 was also homologous to B35 and B51 but diverged between positions 1 and 119. Because the sequence diverged in approximately the same region in these clones, it would seem likely that the differences represent artifacts generated during reconstruction of the library due to secondary structure in the mRNA. Because it coded for an open reading frame, only sequence for B35 is given. Because the Eco RI sequence at the 5' end of B35 differed from the linker sequence used to construct the library, we surmise that an additional Eco RI fragment exists at the 5' end of the p38 cDNA.

**Northern Analysis of mRNA Transcripts**

To determine the tissue specificity of the mRNA transcripts recognized by the cDNAs, we generated probes from lambda 2-4B, the nonoverlapping Eco RI fragment of lambda 1-5 coding for the 3' untranslated region, and the 5' Eco RI fragment of 160 bp from lambda B35. The 2-4B cDNA probe hybridized to a 2.4-kb transcript that was specific for rat brain and was not present in rat liver (Fig. 4a). The 2.4-kb transcript was also present in cells of the rat pheochromocytoma cell line, PC12, the mouse pituitary cell line, AtT-20,
and in bovine brain (Fig. 4, c and d), as well as in the rat superior cervical ganglion (data not shown), and all of which are known to express p38 protein (Navone et al., 1986; Buckley, K. M., and R. B. Kelly, unpublished observations). Qualitatively similar results were obtained when either the 5' Eco RI fragment from lambda B35 or the 3' untranslated region of lambda 1-5 was used as probe (data not shown).

A minor transcript of \(~1.2\) kb was detected in AtT20 cells (Fig. 4 b). Differential hybridization with cDNA probes from the 5' open reading frame and the 3' untranslated sequence revealed that the smaller transcript was missing much of the 3' untranslated region found in the 2.4-kb transcript (not shown).

**Identity of p38 as Synaptophysin**

The protein encoded by our cDNA clones shares at least one antigenic determinant with a protein that has an apparent molecular weight of \(38,000\), is expressed in neuronal and endocrine tissue, and has a distribution that, by immunofluorescence, is identical to that of other known vesicle proteins. To confirm that the R10 serum recognized p38 (or synaptophysin) and not another protein with a similar molecular weight in rat brain, we detergent-solubilized a synaptic vesicle-enriched preparation from rat brain and applied it to a DEAE column. Elution of p38 was monitored by silver stain analysis and immunoblotting of column fractions with R10. p38 has been shown to bind to DEAE columns until eluted with \(90-110\) mM salt (Rehm et al., 1986). After the procedure for purification of p38 described by Rehm et al. (1986), we analyzed the material that did not bind to the column as well as the fractions eluted with a linear gradient of \(0-250\) mM NaCl. Silver stain analysis of the column fractions revealed the presence of a protein with an \(M_r\) of \(38,000\) present in the flow-through fractions as well as a \(38,000\)-D protein eluted from the column as a broad peak around \(100-150\) mM NaCl (Fig. 5 a). The eluted protein appeared as a characteristic doublet, and when aliquots of all fractions, including the flow-through material, were transferred to nitrocellulose and probed with the R10 serum, only the \(M_r = 38,000\) doublet that was bound to and subsequently eluted from the DEAE column by the gradient reacted with the antiserum, whereas the \(38,000\)-mol wt protein in the flow-through did not (Fig. 5 b). The protein characterized here is, therefore, almost certainly p38 (synaptophysin).

**In Vitro Translation of p38 cRNA**

In some cases the electrophoretic mobility of a protein in SDS-PAGE is a poor measure of its true molecular weight. The nonglycosylated form of p38 migrates as a protein of \(M_r = 34,000\) (Rehm et al., 1986; Navone et al., 1986). To ask if one of the methionines in the open reading frame could be an initiator methionine, a partial cDNA containing the first methionine in the p38 sequence, up to and including the third to fourth Eco RI site; see Fig. 3) was transcribed into RNA using T3 and T7 polymerase. In vitro translation of these RNAs resulted in detectable protein synthesis only with the T3 RNA transcript and not with the T7 transcript. The total translation products from the T3 cRNA are shown in lane 1. The two peptides specifically immunoprecipitated from the total translation products by a rabbit polyclonal serum that contains antibodies to p38, one of which (Joe B.) is shown in Fig. 6, lane 2. To compare the \(M_r\) of these proteins with the \(M_r\) of p38 made in vitro from native rat brain RNA, we translated rat brain poly A + RNA and immunoprecipitated with the same polyclonal serum. The two major proteins synthesized from the RNA transcribed by the T3 polymerase migrated with an \(M_r\) of 24,000 and 26,000 (the upper band at 45,000 is a background band seen in the absence of added mRNA). These two proteins were specifically immunoprecipitated by three different polyclonal sera that contain antibodies to p38, one of which (Joe B.) is shown in Fig. 6, lane 2. The band seen at \(M_r = 45,000\) in lane 1 is the nonglycosylated form of p38.
positions of molecular mass standards in kilodaltons.

Third and major protein immunoprecipitated by antibodies to p38 RNA also migrate at these molecular weights from the rat brain RNA translation products by the polyclonal antiserum. The two major translation products from the I3 RNA transcript. Lane 3, peptides immunoprecipitated has an Mr of 36,000 by the polyclonal serum R10 (data not shown). Consistent products, suggesting that two internal initiation sites present mobility (Mr = 26,000 and 24,000) as the T3 translation of 36,000. Two additional proteins migrated with the same general labeled proteins were seen in these immunoprecipitates (Fig. 7, lane 3). The primary translation product had an M, of 36,000. Two additional proteins migrated with the same mobility (M, = 26,000 and 24,000) as the T3 translation products, suggesting that two internal initiation sites present in the native RNA are represented in the cDNA we have sequenced. These same three peptides were also precipitated by the polyclonal serum R10 (data not shown). Consistent with an internal initiation hypothesis, the methionines at positions 105, 126, and 127 all fulfill the Kozak criteria for initiation sites (A/CXXAUGG) (Kozak, 1981) (Fig. 8), and should code for two polypeptides differing in size by ~2,000. We conclude that the primary translation initiation site is missing from our cDNAs. Initiation at the true methionine would add ~10,000 to the apparent molecular weight.

Identification of the cDNA Encoding the Cytoplasmic Domain of p38

The monoclonal antibody against p38 generated by Weidenmann and Franke (1985), SY38, is known to precipitate intact synaptic vesicles. Furthermore, protease treatment of intact vesicles destroys the ability of the monoclonal to bind to p38 in Western blots (Rehm et al., 1986). To determine the region of cDNA that contains the coding region for the cytoplasmic domain of this integral membrane protein, a lambda gt11 phage, lambda 1-5, containing only the sequence from 760 bp to the poly A' tail (Fig. 3) was induced to produce fusion protein in plaques plated on E. coli Y1090, and the fusion protein was transferred to nitrocellulose and probed with the mAb SY38. The monoclonal reacted very strongly with this fusion protein and not with control phage (data not shown), suggesting that the 3' end of the cDNA open reading frame encodes the cytoplasmic tail of the protein.

Sequence Analysis of p38 cDNA

The cDNA clones isolated by a combination of antibody and cDNA screening coded for 894 bp of open reading frame and 1,416 bp of 3' untranslated sequence. The orientation of the sequences and the various Eco RI fragments was determined as diagrammed in Fig. 3.

The consensus sequence for polyadenylation is located 14 nucleotides from a poly A tail in a 3' untranslated sequence of 1,416 bp (Fig. 8). The open reading frame codes for a protein of 298 amino acids with a predicted pl of 4.6 and contains two potential N-glycosylation sites at codons 44 and 188, in agreement with an observed pl of 4.6-4.8 and experimental evidence for glycosylation of p38 from several studies (Jahn et al., 1986; Wiedenmann and Franke, 1985; Navone et al., 1986; Rehm et al., 1986). The predicted molecular weight for this amino acid sequence is 32,425; if both N-glycosylation sites are used, the predicted molecular weight increases to ~38,500. In addition the amino acid sequence also contains three phosphorylation sites within the predicted cytoplasmic tail: two tyrosines (positions 264 and 280) and one threonine (position 212). However, no data has been published concerning the phosphorylation of p38.

Hydrophobic and potential transmembrane domains in the amino acid sequence were predicted using three different sets of values: membrane propensity analysis by the method of Kuhn and Leigh (1985), hydropathicity analysis using the values of Eisenberg et al. (1982), and hydropathicity analysis using the values of Kyte and Doolittle (1982). Each of the three methods predicted four hydrophobic domains occurring at similar positions in the amino acid sequence (Figs. 8 and 9).

The last 90 amino acids of the open reading frame that follow the fourth hydrophobic domain have a distinct amino acid composition and secondary structure compared with the rest of the protein. The last eight amino acids in this region, which we have defined as the terminal cytoplasmic domain by the mAb-binding experiments, is predicted to be the main immunogenic region of the protein by the method of Piner-Moore et al. (1988), consistent with published observations that most polyclonal and monoclonal antibodies to p38 bind to the cytoplasmic tail (Jahn et al., 1985; Wiedenmann and Franke, 1985; Rehm et al., 1986). This region also contains a glycine-rich domain between residues 230-290 (Fig. 8); the overall percentage of glycine in the protein is 12%, but in this region the percentage increases to 39%. This stretch of amino acids is also enriched for glutamine. From residues 1-240, glutamine accounts for only 2% of the amino acids, whereas from 240 to 290 the percentage increases to 20%. Also contained in this region are a series of six Gly-X-Gly repeats, three Gly-Pro-Glu repeats, and two repeats of seven to eight amino acids Gly-Gly-Tyr-Gly-Pro-Gln-(Gly)-Asp. Because of the unusual amino acid composition, the predicted secondary structure of the terminal domain has a large number of turns, compared with the rest of the sequence (Fig. 9).

No significant sequence homologies were obtained by searching through protein data banks. However, a number of proteins with glycine-rich domains, such as types I and II cytoskeletal keratin and bovine cartilage alpha 1 (II) chain, show homology to the glycine-rich domain of p38.

Discussion

The cDNA we have cloned and sequenced encodes p38, a transmembrane glycoprotein of synaptic vesicles (Wieden-
mann and Franke, 1985; Jahn et al., 1986) and of endocrine secretory granules (Lowe et al., 1988). We believe this cDNA represents the coding sequence for p38 based on the following evidence. Fusion proteins expressed by recombinant lambda gt11 phage share an antigenic determinant with p38, and because p38 is the only protein in brain with that determinant, both by size (Fig. 1) and by DEAE chromatography (Fig. 7), the two proteins must be the same. Consistent with this prediction, the protein sequence deduced from the cDNA sequence has several of the properties predicted for p38, including hydrophobic, potential transmembrane domains, two N-glycosylation sites, an immunodominant region in the cytoplasmic tail, and approximately the same molecular weight.

The open reading frame derived from the cDNA clones described in this paper does not appear to be complete. The sequence flanking Eco RI site 1, at the 5' end of the cDNA, is underlined, as well as the consensus sequence for polyadenylation (AATAAA).
at position 105 in the amino acid sequence, and a protein beginning at this methionine would have a predicted molecular weight of ~21,000. However, the cDNA fragment that is missing is likely to be small. We have accounted for over 2.3 kb out of a predicted 2.4 kb mRNA, so the remaining open reading frame and 5' untranslated region together are probably <100 bp. Furthermore, the difference between the *M*ₐ of unglycosylated p38 (34,000; Rehm et al., 1986; Navone et al., 1986) and the predicted molecular weight for the amino acid sequence deduced from our cDNA clones (32,425) is ~1,600, which could be encoded by 15 amino acids. Another possibility suggested by our data is that the electrophoretic mobility overestimates the molecular weight, and that the entire amino acid sequence of the 34,000-D peptide is encoded by the cDNA we have sequenced. The *M*ₐ of p38 synthesized in vitro from rat brain mRNA was 36,000 (Fig. 6). The largest peptide produced by in vitro transcription/translation of p38 cDNA, which is presumably initiated from the first methionine at position 105, had an *M*ₐ of 26,000 (Figs. 6 and 7). Note that this is more than the size predicted by the amino acid sequence, which is 20,725, implying that the carboxyl end of the protein has properties that confer on it an unexpectedly low electrophoretic mobility. Translation of the preceding 105 amino acids would add ~II,500. Simple addition would give a primary translation product that migrated as a protein of *M*ₐ 37,500, compared with the observed 36,000. If the amino-terminal hydrophobic domain in the predicted amino acid sequence acted as a signal peptide and was cleaved, removing the first 30 amino acids, the electrophoretic mobility of the mature unglycosylated p38, estimated by simple subtraction, would correspond to a protein with an *M*ₐ of 34,000 or a predicted molecular weight (from

---

**Figure 9.** Secondary structure predictions from the p38 amino acid sequence (hydropathicity plot using the values of Kyte and Doolittle). Positive values are indicated by vertical bars extending above the horizontal line; negative values fall below the line. The ruler at the top indicates 25 amino acids per division.
Figure 10. Models for the disposition of \( \text{p38} \) protein in the synaptic vesicle membrane. (1) All four predicted hydrophobic regions are presented as transmembrane domains. (2) If the first hydrophobic acids of mature \( \text{p38} \) presented as transmembrane domains. (2) If the first hydrophobic acids of mature \( \text{p38} \). (Rehm et al., 1986; Jahn et al., 1985). The predicted structure of three hydrophobic domains (Fig. 10, model 2) has two N-glycosylation sites on the lumenal side, as required, and a cytoplasmic immunogenic domain recognized by the mAb SY38. However, protease digestion of intact synaptic vesicles appears to remove a single large fragment from \( \text{p38} \) of \( \sim 13 \text{ kD} \) (Rehm et al., 1986; Jahn et al., 1985), suggesting a single cytoplasmic domain and one transmembrane region. If three or four of the hydrophobic domains are indeed transmembrane (Fig. 10), as suggested by the hydropathicity analysis, the cytoplasmic loop would have to be protease insensitive, perhaps because of its small size. In such a case, \( \text{p38} \) could have several transmembrane domains yet only a single large fragment containing the carboxyl-terminal 90 amino acids would be removed by protease treatment.

\( \text{p38} \) is the major integral membrane protein of the synaptic vesicle, and it has been reported to be the major \( \text{Ca}^{2+} \)-binding protein in the vesicle membrane. Because this activity is lost with protease digestion of intact synaptic vesicles, it has been proposed that this activity is localized to the cytoplasmic tail (Rehm et al., 1986). However, no homology was found between the amino acid sequence of \( \text{p38} \) and the \( \text{Ca}^{2+} \)-binding proteins characterized by an "EF hand" (Kretsinger, 1980) or a number of membrane-associated \( \text{Ca}^{2+} \)-binding proteins that contain a \( \text{Ca}^{2+} \) consensus sequence of 17 amino acids (Geisow et al., 1986). A third class of \( \text{Ca}^{2+} \)-dependent integral membrane proteins called cadherins have been described that are involved in intercellular adhesion but show no apparent homology to the other two classes of \( \text{Ca}^{2+} \)-binding proteins. Thus, a variety of sequences or secondary structures exist that confer \( \text{Ca}^{2+} \)-binding activity, and \( \text{p38} \) may contain a previously unrecognized \( \text{Ca}^{2+} \)-binding sequence.

Although the role played by \( \text{p38} \) in neurons is not known, there are several possible functions for a transmembrane synaptic vesicle protein. \( \text{p38} \) could link the vesicle membrane to the cytoskeleton or to the plasma membrane. It could participate in packaging molecules into the synaptic vesicle, even \( \text{Ca}^{2+} \), because \( \text{p38} \) is reported to bind \( \text{Ca}^{2+} \) (Rehm et al., 1986). Alternatively, \( \text{p38} \) may play a role in the \( \text{Ca}^{2+} \)-dependent exocytosis of synaptic vesicles. The \( \text{Ca}^{2+} \)-binding activity of \( \text{p38} \) is thought to be localized to the cytoplasmic tail of the protein. As this region is encoded by the cDNA clones we have sequenced, it may now be possible to explore the consequences of hypopexpression or deletion of the cytoplasmic tail.

The cDNA clones encoding \( \text{p38} \) are also immediately valuable for the study of membrane protein targeting. Sorting of soluble proteins and membrane proteins from the Golgi complex into lysosomes, regulated secretory granules, or into vesicles that transport them directly to the plasma membrane appears to result from the presence of sorting domains. Because \( \text{p38} \) is known to be a membrane component of dense-core secretory granules (Lowe et al., 1988), it may now be possible to examine membrane protein sorting in the regulated secretory pathway, asking, for example, if the cytoplasmic domain specifies destination, either during biogenesis or during recycling of secretory vesicle membrane components from the plasma membrane to the Golgi region (Patka and Winkler, 1986). The regulated secretory pathway presents a unique challenge to models of membrane sorting because, unlike other intracellular organelles, secretory granules and synaptic vesicles appear de novo during development. The availability of a \( \text{p38} \)-specific cDNA probe provides a unique opportunity to examine how the regulated pathway is induced, during synaptogenesis and in the development of endocrine cells.

We are extremely grateful to Lois Clift-O'Grady for her assistance with many aspects of this project, especially the sequencing and protein purification; Dr. Jeff Edman and Dr. Linda Matsutchi for advice on library screening and cloning; Dr. Robert Stroud and Dr. Janet Finer-Moore for assistance with the secondary structure analysis; Dr. Jeffrey Boone Miller for his assistance in preparation of the manuscript; and the members of the Kelly lab for helpful comments.

This work was supported by National Institutes of Health grant NS-09878-16 to Dr. Kelly.

Received for publication 12 August 1987.

References

Buckley, K. M., and R. B. Kelly. 1985. Identification of a transmembrane glycoprotein specific for secretory vesicles of neural and endocrine cells. J. Cell Biol. 100:1284–1294.

Cheley, S., and R. Anderson. 1984. A reproducible microanalytical method for the detection of specific RNA sequences by dot-blot hybridization. Anal. Biochem. 137:15–19.

Chen, E. Y., and P. Seeburg. 1985. Supercill sequencing: a fast and simple method for sequencing plasmid DNA. DNA (NY). 4:165–170.

Church, G. M., and W. Gilbert. 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA. 81:1991–1995.

Eisenberg, D., P. Weiss, T. Terwilliger, and W. Wilcox. 1982. Hydrophobic moments and protein structure. Faraday Symp. Chem. Soc. 17:109–120.

Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6–13.

Finer-Moore, J., J. F. Bazan, J. Rubin, and R. M. Stroud. 1988. Prediction of Protein Structure and the Principles of Protein Conformation. D. Fasman, editor. Plenum Press, New York. In press.
Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120:97–120.

Geisow, M. J., U. Fritsche, J. M. Hexham, B. Dash, and T. Johnson. 1986. A consensus amino acid sequence repeat in Torpedo and mammalian Ca++. dependant membrane-binding proteins. Nature (Lond.) 320:636–638.

Greif, K., and L. F. Reichardt. 1982. Appearance and distribution of neuronal cell surface and synaptic vesicle antigens in the developing rat superior cervical ganglion. J. Neurosci. 2:843–852.

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

Huynh, T. V., R. A. Young, and R. W. Davis. 1985. Construction and screening cDNA libraries lambda gtl0 and lambda gtl1. In DNA Cloning: A Practical Approach. Vol. I. D. M. Glover, editor. IRL Press, Washington, DC. 49–78.

Jahn, R., W. Scheibler, C. Ouimet, and P. Greengard. 1985. A 38,000 dalton membrane protein (p38) present in synaptic vesicles. Proc. Natl. Acad. Sci. USA. 82:4137–4141.

Knaus, P., H. Betz, and H. Rehm. 1986. Expression of synaptophysin during postnatal development of mousebrain. J. Neurochem. 47:1302–1304.

Kozak, M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. 9:5233–5236.

Kretsinger, R. H. 1980. Structure and evolution of calcium-modulated proteins. CRC Crit. Rev. Biochem. 8:119–174.

Kuhn, L. A., and J. S. Leigh, Jr. 1985. A statistical technique for predicting membrane structure. Biochim. Biophys. Acta. 828:351–361.

Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydrophobic character of a protein. J. Mol. Biol. 157:105–132.

Lowe, A. W., L. Mededdu, and R. B. Kelly. 1988. Endocrine secretory granules and neuronal synaptic vesicles have three integral membrane proteins in common. J. Cell Biol. In press.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Matthew, W., R. Jahn, G. DiGioia, H. Sukenbrok, P. Greengard, and P. DeCamilli. 1986. Protein p38: an integral membrane protein specific for small vesicles of neurons and neuroendocrine cells. J. Cell Biol. 103:2511–2527.

Patzak, A., and H. Winkler. 1986. Exocytotic exposure and recycling of membrane antigens of chromaffin-granules: ultrastructural evaluation after immunolabeling. J. Cell Biol. 102:510–515.

Rehm, H., B. Wiedenmann, and H. Betz. 1986. Molecular characterization of synaptophysin, a major calcium-binding protein of the synaptic vesicle membrane. EMBO (Eur. Mol. Biol. Organ.) J. 5:535–541.

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

Sarthy, P. V., and W. Bacon. 1985. Developmental expression of a synaptic vesicle-specific protein in the rat retina. Dev. Biol. 112:284–291.

Smith, D., and P. Fisher. 1984. Identification, developmental regulation, and response to heat shock of two antigenically related forms of a major nuclear envelope protein in Drosophila embryos: application of an improved method for affinity purification of antibodies using polypeptides immobilized on nitrocellulose blots. J. Cell Biol. 99:20–28.

Wiedenmann, B., and W. Franke. 1985. Identification and localization of synaptophysin, an integral membrane glycoprotein of M, 38,000 characteristic of presynaptic vesicles. Cell. 41:1017–1028.