Identification, characterization, and molecular cloning of a novel transporter-like protein localized to the central nervous system

Jay A. Gingrich[a], Peter H. Andersen[a,*], Mario Tiberi[a], Salah El Mestikawy[a], Per N. Jorgensen[b], Robert T. Fremeau Jr.[a] and Marc G. Caron[a]

[a]Howard Hughes Medical Institute Research Laboratories, Depts. of Cell Biology and Medicine, Duke University Medical Center, Durham, NC 27710, USA and [b]Dept. of Monoclonal Antibody, Biozyme Novo Nordisk A/S, Bagsvaerd DK-2880, Denmark

During the course of large scale purification of the D_1 dopamine receptor from rat brain, a protein of ~87,000 daltons (p87) was observed to copurify with the D_1 receptor through four chromatographic steps. To characterize the nature of this protein, bovine and rat cDNA clones were isolated and sequenced. The bovine and rat clones were highly conserved (98.5% identity). Each clone possessed an open reading frame of 2226 base pairs encoding a protein of 742 amino acids (calculated MW of 82,500), containing three stretches of peptide sequence obtained from p87 sequence analysis. Comparison of the deduced peptide sequence of this protein with those found in available databanks revealed that it was a novel protein related to the family of nutrient transport proteins from eukaryotes and bacteria, including, the mammalian facilitated glucose transporters, the yeast transporters for maltose, lactose, and glucose, and the proton-driven bacterial transporters for arabinose, xylose, and citrate. In addition p87 also shares with these transporters a similar hydropathicity profile that suggests the presence of 12 transmembrane segments. The mRNA for p87 appears to be localized primarily, if not exclusively, to the central nervous system. Northern blot analysis reveals a message of ~4.8 kb in cortex, hippocampus, brain stem, and cerebellum, but no detectable signal in peripheral tissues such as spleen, liver, kidney, lung, heart, or skeletal muscle. Evidence form Western blot analysis and immunohistochemistry suggests that this protein may be expressed in intracellular organelles or the membrane of synaptosomes rather than plasma membrane. Based on its structure and properties, p87 appears to define a new class of transporter-like proteins.

Transporter-like protein; Central nervous system; Molecular cloning; Protein characterization

1. INTRODUCTION

Given the complexity of the CNS, numerous methods have been developed to study its function. One avenue of research has been the identification and study of proteins which appear to be unique to the central nervous system: proteins which have been estimated to comprise perhaps as many as two-thirds of the products encoded by the genome. This approach does not depend on a prior understanding of function, has proven to be enormously successful. For example, several new members of the family of G-protein coupled receptors [24] and transport proteins were cloned before their endogenous ligand or substrate was known. However, many of these 'orphan' receptors and transporters have now been identified including receptors for 5-HT (5-HT$_{1A}$, 5-HT$_{1D}$), cannabinol, and adenosine [12,18,25,27], and the Na^+ and Cl^--dependent transporters for proline, serotonin, to list only a few exam-

ples [4,13]. Likewise, understanding of the mechanisms of synaptic vesicle function has advanced through the purification and characterization of their protein components without prior knowledge of their function [2,20,28,29,31]. The power of this approach continues to grow as the structure of more functional families of proteins are defined.

It was the recent success of this reverse approach to understanding biochemical function which led us to characterize, p87, a protein of molecular weight 87,000 Da which persistently copurified with the D_1 receptor through four chromatographic steps including a ligand-based affinity chromatographic step based on the selective antagonist SCH23,390. During this procedure, the D_1 receptor was enriched more than 3,000-fold and p87 was estimated to represent approximately 50% of the total protein. We reasoned that p87 could be of potential interest first because of its potential ability to associate with the D_1 dopamine receptor or second because of the possibility that this protein might interact with benzazepine compounds such as SCH 23,390.

In this report we describe the purification, cloning, and characterization of p87. The rat and bovine cDNAs for this protein revealed significant homology of its deduced sequence with a family of nutrient transporter proteins which includes the mammalian glucose trans-

Correspondence address: J.A. Gingrich, Howard Hughes Medical Institute Research Laboratories, Depts. of Cell Biology and Medicine, Duke University Medical Center, Durham, NC 27710, USA.

*On sabbatical leave from: The Department of Molecular Pharmacology and Bioscience, Novo Nordisk A/S Bagsvaerd, DK-2880, Denmark.

porter, and bacterial transporters for xylose, arabinose, citrate, lactose and tetracycline. Based on Northern and Western blot analysis, this protein was found to be localized predominantly to the CNS. The predicted membrane topography of this transporter-like protein suggests that it represents the first member of a novel transporter-like family. Preliminary accounts of this work have previously been published in abstract form [15,16].

2. MATERIALS AND METHODS

2.1. *Purification of the D_1 dopamine receptor*

The D_1 dopamine receptor was purified as described elsewhere (Gingrich et al., in preparation) with some modifications to accommodate the increased scale. Briefly, approximately 2,000 rat brains were dissected to remove their striata in three batches. Each batch was solubilized and affinity-purified independently [14]. These eluates were pooled and concentrated on Q-Sepharose, and then chromatographed on two Superose 6 columns in tandem. The peak fractions were pooled and applied to a WGA agarose column and eluted with N-acetylglucosamine. From this procedure 417 pmol of [³H]SCH23390 binding activity was recovered with a specific activity estimated at 6 nmol [³H]SCH23390 bound/mg protein, or about 50% pure. For analysis, a 1% aliquot of the purified receptor preparation was removed and radioiodinated with 0.2 mCi [¹²⁵I]Bolton-Hunter reagent for 4 h at 4°C (Langstone, 1980). The samples were elecrophoresed on preparative SDS-PAGE, the bands localized by autoradiography, cut and electro-eluted. The samples were then lyophilized (Savant Instruments) and resuspended in 70% formic acid containing 0.5 M CNBr in a polypropylene Eppendorf tube. The reaction was allowed to proceed 24 h in the dark at room temperature and then lyophilized.

The samples were then resuspended in 0.175 ml of 0.1% tri-fluoroacetic acid (TFA) in water (solution A), loaded on a 2.1 mm × 100 mm C-8 column, and eluted with a gradient of solution B (70% acetonitrile (Jackson/Burdick), 29.9% water, 0.1% TFA) at 0.5% acetonitrile/min (flow rate = 0.1 ml/min, 1 fraction/ min). Optical density (O.D.) at 214 nm was monitored throughout the run. Fractions were spotted onto polybrene coated glass fiber filters and subjected to automated Edman degradation sequencing.

2.2. *Library screening*

Oligonucleotide probes were synthesized based on peptide sequence information obtained from p87. The longest of these probes (based on the codon bias information in [23]) was radiolabelled by addition of [³²P]phosphate using T4 polynucleotide kinase, and used to screen a bovine brain cDNA library constructed in lambda ZAP II (Stratagene). Positive clones were purified by successive rounds of plaque purification, and cDNA inserts recovered by helper phage mediated excision. Both strands of each clone were then sequenced using the Sanger dideoxy chain termination method (Sequenase, USB).

2.3. *Northern blot analysis*

Total RNA was isolated from frozen tissues using guanidinium isothiocyanate and centrifugation through a cushion of cesium chloride. Total RNA (30 mg) was electrophoresed on a 1% agarose/formaldehyde gel, transferred to nitrocellulose, hybridized with a *Nco*I fragment of the p87 cDNA (1 × 10⁶ cpm/ml) for 18 h, washed (1 × SSC 55°C), and apposed to X-ray film with intensifying screens for 12 h at -70°C.

2.4. *Expression of p87 in mammalian cell systems*

To express p87 in mammalian cells, the entire cDNA of 27B (see Fig. 2) was subcloned into pCMV4 (pCMV4-27B) and used to transiently transfect COS-7 cells (15 mg of plasmid per 15 cm flask) using the DEAE-dextran method [9]. Transfection efficiency was controlled by simultaneous transfection of COS-7 cells with a plasmid containing the D_2 dopamine receptor [6]. Permanent transfection in L(tk⁻) cells

was achieved by cotransfecting these cells with a plasmid encoding for resistance to the antibiotic G418 (pNEO) and pCMV4-27B (ratio 1:20) using the calcium phosphate precipitation method, and then selecting for cells with G418 resistance (400 mg/ml). Cells demonstrating resistance to G418 were then selected clonally and tested for the expression of pCMV4-27B by Northern and Western blot analysis.

2.5. *Cell culture and uptake assays*

Ltk⁻ cells were maintained in Dulbecco minimal essential medium supplemented with 10% heat inactivated fetal bovine serum and 0.1 mg/ml gentamycin at 37°C in 5% CO_2. Positive expressing cells or controls were washed twice with phosphate buffered saline (PBS), incubated for 5–20 minutes at 37°C with PBS containing 0.1–1 mM of the radiolabelled substrate, washed 3 times with ice-cold PBS and then lysed with a solution of 0.1 N NaOH, 1% SDS. An aliquot of this solution was then counted by liquid scintillation spectophotometry.

Fig. 1. Determination of peptide sequences for purified p87. (Inset) Autoradiogram of the purified material. A 1% aliquot of the preparative purification material was removed and iodinated with [¹²⁵I]Bolton-Hunter reagent, desalted on Sephadex G-50 to remove excess labelling reagent and then added back to the remainder of the unlabeled sample, concentrated on an Amicon YM-30 Centricon unit and then separated by 10% SDS-PAGE. The wet gel was then covered with plastic wrap and apposed to film at 4°C with two intensifying screens for 2 h to produce the image shown. The autoradiogram was then used as a guide to excise the appropriate bands. (Main figure) Chromatogram of the peptide fragments of p87 following CNBr cleavage. The samples were separated by reverse-phase HPLC as described in Methods. The lower tracing (blank) was observed with solution alone whereas the middle tracing (control) represents when the non-protein containing blank sample was injected (processed identically to the protein containing sample). The upper tracing (sample) was that obtained from the protein containing sample derived from the higher molecular weight band on the preparative SDS-polyacrylamide gel (inset).

3. RESULTS

3.1. Purification and protein microsequencing of the p87 protein

The p87 protein was found to copurify with the D_1 dopamine receptor during its preparative-scale purification. Fig. 1 (inset) shows the autoradiogram obtained following radioiodination and SDS-PAGE electrophoresis of the purified material. Two prominent proteins which have relative molecular weights of 68 and 87 kDa were observed. Based on the relative size of the protein, the 68 kDa protein (p68) was thought to be the D_1 dopamine receptor. The identity of the 87 kDa protein (p87) was unknown. The p87 band was processed for amino acid sequence analysis by electroelution, dialysis, CNBr cleavage, and separation of peptides by microbore reverse phase HPLC (see Methods). The sequence of three of the peaks of p87 was obtained, and they neither corresponded to any sequence in available protein databases nor bore homology with G-protein coupled receptors or guanine nucleotide binding proteins.

3.2. Cloning of bovine and rat cDNAs for p87 from brain libraries

A non-degenerate 54-base probe designed using the sequence of the longest peptide was radiolabelled, and used to detect positive clones in a bovine brain cDNA library by plaque hybridization. Screening of 1×10^6 recombinants yielded 54 positive clones. One of these clones, BT27B, was found to contain a long open reading frame of 2226 bases coding for a protein of 742 amino acids (calcd. MW = 82.5 kDa). Importantly, the long open reading frame contained the sequence of three of the p87 peptides, thus confirming the identity of this clone. Moreover, the calculated molecular weight of the BT27B clone (82,500 daltons) was in close agreement with the molecular weight of the p87 protein (~87,000 daltons) which was detected on SDS-PAGE by autoradiography. Using a probe from the bovine p87 cDNA, the rat homologue of BT27B was isolated (RT10A) from a rat brain cDNA library (96 positive clones out of 1×10^6 recombinants) and sequenced. The rat and bovine p87 cDNAs were found to be highly conserved having 98.5% identity at the amino acid level. The nucleotide sequence for clone BT27B and the deduced amino acid sequence of the rat and bovine open reading frames are shown in Fig. 2. In each case the open reading frame is chosen to begin at the first Kozak consensus sequence which is in frame with the rest of the open reading frame and which is preceded by a termination codon.

3.3. Structural features of p87 protein

The hydropathicity profile of p87 determined using the algorithm of [21] suggests that p87 possesses 12 hydrophobic domains which are of sufficient length to span a lipid bilayer. The location of the putative trans-

membrane domains in the protein suggests that p87 possesses a large hydrophilic amino-terminal domain, a short carboxyl-terminal domain, and two large hydrophilic loops between putative transmembrane (TM) segments 6–7 and TM segments 7–8. The p87 sequence contains three possible sites for N-linked glycosylation (N-X-S/T) within the loop defined by the putative 7th and 8th TM segments (Asn^{498}, Asn^{548}, Asn^{573}). The sequence of p87 also contained several potential sites of regulatory phosphorylation by protein kinase A (Ser^{47}, Thr^{447}, and Thr^{684}) and protein kinase C (Thr^{27}, Ser^{42}, Ser^{357}, Ser^{679}). Interestingly, these consensus sites for regulatory phosphorylation are all found in domains which are predicted to be on the same side of the membrane bilayer and which, moreover, are on the opposite side of the membrane where the consensus sites for glycosylation are found. Interestingly, two consensus sites for ATP binding motifs (GXGXXG(7–18aa)K) are found in the sequence at aa129–143 lying in the amino-terminus near the putative TM1 and aa266–288 located in the region of putative TM4 and the loop between TM4 and TM5 [19]. A model for the possible transmembrane organization and the relative position of the various sites of post-translational modifications are shown in Fig. 3a.

3.4. Homologies of p87 sequence with known transport proteins

Comparison of the p87 protein sequence with a database of protein sequence motifs reveals that p87 contains two amino acid 'signature' sequences which are conserved among proteins known to transport sugars and other nutrients across the plasma membrane of eukaryotic cells and bacteria [1,17] (see Fig. 3b legend). Moreover, when protein databanks were searched for homologies to p87, additional similarities were found with several members of the transporter protein family. Specifically, the best matches were found to be with the glucose transporters of rat, man and yeast (~26% identity over an ~180 amino acid (AA) overlap), and the bacterial transporters for xylose (22% over 148 AA), arabinose (24% over 213 AA), citrate (24% over 176 AA) and tetracycline (21% over 134 AA). In addition, significant homology was also found with the rat sodium channel (21% over 201 AA). This region of homology resides in the C-terminal half of the molecule.

Several regions of striking homology between selected members of the transport family are shown in Fig. 3b. It is apparent that the more highly conserved regions between the different members of this class of protein are clustered to the same side of the membrane bilayer and occur primarily in the short hydrophilic loops connecting TM segments (i.e. A, B, A', and B') and the hydrophilic regions adjoining TM segments (i.e. C and C'). In addition, the alignment demonstrates that the amino-terminal half of these proteins is better conserved among different members than the carboxyl-ter-

minal half. This is evidenced by the large regions of conservation which extend well into the TM segments of regions A, B, and C. In contrast, the regions of conservation in the areas A′, B′, and C′ are mostly limited to the short, putative extramembraneous regions. Fig. 3b also illustrates the presence of the transport protein signature sequences mentioned earlier. The first such motif, ([LIVMF]X[GLIVMFA]X[G]X$_8$[LY]X$_2$[EQ]X$_6$[RK]), is found in region B and is perfectly conserved in p87 (LsGvGiGgsipivfsYfsEflaqekR). The second motif, ([LIVMST][DE]X [LIVMFA]GR[RK]X$_{4,5}$G), is represented in region A′ and again this sequence is well conserved (although not perfectly) in p87 (LDkiGRlRmlaG) since the presence of a Leu residue between the two Arg residues interrupts the motif.

Another feature that p87 shares with these other transport proteins is a significant degree of internal homology. This is evidenced by repetition of certain motifs in the regions A/A′, B/B′ and C/C′. In the regions defined by A and A′ there is conservation of a sequence which connects TM segments 2 and 3 and 8 and 9 (DRLGRR in A and DKIGRLR in A′). In regions B/B′ the internal repeat follows the sequence EFLAQEKR in B and ELYPSDKR in B′. Finally, another internal repeat can be found in regions C/C′ (PES in region C and PET in region C′).

3.5. mRNA for p87 are found predominantly in the central nervous system

An mRNA (4–4.8 kb) could be detected for p87 in several species including rat, bovine, and man (data not shown). Fig. 4 shows the results of Northern blot analysis for the expression of p87 message in several regions of the rat CNS and periphery. In the CNS, a prominent message is detected in the cortex, brain stem, hippocampus, and cerebellum. In contrast, no detectable message could be found in any of the peripheral tissues examined such as heart, liver, lung, kidney, spleen, or skeletal muscle. This analysis suggests that p87 is localized predominantly to the CNS. Moreover, the message appears to be widespread in its distribution within the CNS since it is found in roughly the same abundance in the four regions examined.

3.6. Subcellular distribution of p87

In order to determine the subcellular distribution of the p87 protein, a differential centrifugation technique [22] was used to generate a four fraction separation. In this fractionation method the 'N' fraction is the nuclear fraction and a high concentration of lysozomes. The 'M' and 'L' fractions contain the heavy and light mitochondria respectively. These fractions also contain a high concentration of synaptosomes, and to a lesser degree plasma membrane, and lysozomes. The 'P' fraction contains the majority of the plasma membranes. Detection of the protein p87 was determined using an antipeptide

antibody (MIL9) targeted towards a sequence in the large hydrophilic amino-terminal domain (RRSYSRFEEEDDDDFPAPA) of p87. Using MIL 9 an immunoreactive band was detected at 85 kDa, and this labelling was blocked by simultaneous incubation of MIL 9 with an excess of the antigenic peptide (– and + respectively). As can be seen in Fig. 5, The 'M' and 'L' fractions contain the majority of the p87 immunoreactivity with far less being detected in the nuclear and plasma membrane fractions. This immunoreactivity in the subcellular fractions corresponds in size to the immunoreactive band observed in crude rat striatum homogenates. Interestingly, by immunohistochemistry the p87 signal is detected exclusively in neurons and localizes essentially with intracellular organelles (B. Guldhammer et al., data not shown). Thus, the distribution of p87 subcellularly is most consistent with a localization in synaptosomes/synaptic vesicles. Interestingly, no detectable immunreactive signal is seen in purified bovine adrenal chromaffin vesicles, supporting the CNS specific nature of p87.

3.7. Functional expression of the p87 protein

To determine the possible function of p87 in the CNS, p87 cDNA was expressed in both transient and stable fashion in COS-7 and L(tk⁻) cells respectively. In these cells transport of several radiolabelled compounds including: dopamine, serotonin, norepinephrine, glutamate, GABA, choline, 3-O-methyl glucose, 2-deoxyglucose, methylglucopyranoside, and mixtures of amino acids was not detected. In addition, no detectable increases in [³H]SCH23390 binding were observed.

4. DISCUSSION

4.1 P87 is homologous to the family of nutrient transporters

p87 is related to a class of proteins known to transport nutrients across the plasma membrane of mammalian, primitive eukaryotic (yeast, fungus, algae), and bacterial cells (reviewed in [1]). This family of transport proteins share several common features between themselves and with p87. First, these proteins generally possess 12 stretches of hydrophobic amino acids which very likely are membrane spanning segments [26]. The

→

Fig. 2. Nucleotide and peptide sequence of p87. The nucleotide sequence of the bovine p87 (clone BT27B) is shown on the top line. The amino acid sequence is shown below. The deduced sequence of the rat p87 clone (RT10A) was found to be identical to the bovine amino acid sequence except where noted (line below the bovine aa sequence). The underlined amino acid sequences correspond to putative transmembrane domains predicted from hydropathicity analysis. The regions highlighted by asterisks indicate sequences which were found during peptide microsequence analysis of the CNBr fragments of rat p87 protein.

```
                                   -433 AGAGTAGACGCCGCTGCCGCCTGCCCCACTGCCGGGACCCCCGGCCTCACCCAGACTCCCT -372
TATCTGGGGGACTCAGCCCAGATCTCCAGAGACTCCTGCAAGAGAAGTAGCAGCCTTTTCTGAGTTATCCTGGCTCCCCAGCTTAGCCTCCCCAACCTA -273
GCTCCCCCTCCCGTTCCAGTACCCTCTCTCCCCCTCTTCCTTCCCAATGCATCCAACTGAACTGGGTGTAGGCCAAAGCTCTTCCCTCCCTCCCCTCC -174
TAAACCCTCACTGGCCAGATCCTATATTCTATTCTGTGCCCAGAGTTACTCAAACACACTTCATCTAAGTAACTCGGGGAAGTGGTCTTTCTGACTCCA -75

CGAGCCCTTGTGCAAAATCAGAATAAGGAAGGGGGGAAGATTCGGTGAAGTAGTCCCAAAGCCCCATCATG GAA GAG GGC TTC AGA GAC CGG   24
                                                                        M   E   E   G   F   R   D   R    8

GCA GCT TTC ATC CGT GGG GCC AAA GAC ATT GCC AAG GAA GTC AAG AAG CAT GCA ACC AAG AAG GTG GTG AAG GGC    99
 A   A   F   I   R   G   A   K   D   I   A   K   E   V   K   K   H   A   T   K   K   V   V   K   G    33
                                                                       A
CTG GAC AGA GTC CAG GAT GAA TAT TCC CGG AGA TCC TAC TCC CGC TTT GAG GAG GAG GAC GAT GAT GAT GAC TTC   174
 L   D   R   V   Q   D   E   Y   S   R   R   S   Y   S   R   F   E   E   E   D   D   D   D   D   F    58
                                                                     E
CCC GCC CCT GCT GAT GGC TAT TAC CGC GGG GAA GGG GCC CAG GAT GAG GAG GAA GGC GGC GCA TCT AGT GAT GCC   249
 P   A   P   A   D   G   Y   Y   R   G   E   G   A   Q   D   E   E   E   G   G   A   S   S   D   A    83

ACC GAG GGC CAC GAC GAG GAT GAT GAG ATC TAC GAG GGG GAA TAT CAG CGC ATC CCC CGG GCA GAG TCT CGG GGC   324
 T   E   G   H   D   E   D   D   E   I   Y   E   G   E   Y   Q   R   I   P   R   A   E   S   R   G   108

AAA GGC GAG CGG ATG GCA GAT GGG GCA CCC CTG GCT GGA GTG ACG CGG GGC TTG GGT GAT GGG GAG GGT CCC CCG   399
 K   G   E   R   M   A   D   G   A   P   L   A   G   V   R   G   G   L   G   D   G   E   G   P   P   133
                                                             S
GGG GGG CGG GGA GAA GCG CAG CGG CGG AAA GAA CGG GAA GAA CTA GCC CAG CAG TAT GAA GCC ATC CTA CGG GAG   474
 G   G   R   G   E   A   Q   R   R   K   E   R   E   E   L   A   Q   Q   Y   E   A   I   L   R   E   158
                                      D
TGC GGC CAT GGC CGC TTC CAG TGG ACA CTC TAT TTC GTG CTT GGT CTG GCA CTG ATG GCC GAT GGT GTT GAG GTC   549
 C   G   H   G   R   F   Q   H   T   L   Y   F   V   L   G   L   A   L   M   A   D   G   V   E   V   183
                                          **************************
TTC GTG GTG GGC TTC GTG CTG CCC AGT GCT GAG AAA GAC ATG TGC CTG TCT GAC TCC AAC AAA GGC ATG CTG GGC   624
 F   V   V   G   F   V   L   P   S   A   E   K   D   M   C   L   S   D   S   N   K   G   M   L   G   208
**********************************
CTC ATT GTC TAC CTG GGC ATG ATG GTG GGA GCC TTC CTC TGG GGA GGG CTG GCT GAT CGG CTG CGT CGA AGA CAG   699
 L   I   V   Y   L   G   M   M   V   G   A   F   L   W   G   G   L   A   D   R   L   G   R   R   Q   233

TGT GTG CTC ATC TCA CTC TCA GTC AAC AGT GTC TTC GCC TTT TTC TCA TCT TTC GTC CAG GGT TAT GGC ACT TTC   774
 C   V   L   I   S   L   S   V   N   S   V   F   A   F   F   S   S   F   V   Q   G   Y   G   T   F   258

CTT TTC TGC CGT CTC CTT TCT GGG GTT GGG ATC GGA GGG TCC ATC CCC ATC GTC TTC TCC TAT TTC TCG GAG TTT   849
 L   F   C   R   L   L   S   G   V   G   I   G   G   S   I   P   I   V   F   S   Y   F   S   E   F   283
                                                      H
CTG GCA CAG CAG AAA CGT GGG GAG CAT TTG AGC TGG CTC TGC ATG TTT TGG ATG ATT GGT GGA GTG TAT GCA GCT   924
 L   A   Q   Q   K   R   G   E   H   L   S   H   L   C   M   F   W   M   I   G   G   V   Y   A   A   308

GCT ATG GCC TGG GCC ATC ATC CCC CAC TAT GGG TGG AGC TTT CAG ATG GGG TCT GCT TAC CAG TTC CAC AGC TGG   999
 A   M   A   H   A   I   I   P   H   Y   G   W   S   F   Q   M   G   S   A   Y   Q   F   H   S   W   333

AGG GTC TTT GTC CTC GTC TGC GCT TTC CCT TCT GTG TTT GCC ATT GGG GCT CTG ACC ACA CAG CCT GAA AGC CCC  1074
 R   V   F   V   L   V   C   A   F   P   S   V   F   A   I   G   A   L   T   T   Q   P   E   S   P   358

CGT TTC TTC CTG GAG AAT GGG AAG CAT GAT GAG GCC TGG ATG GTA CTG AAG CAG GTC CAT GAC ACC AAC ATG CGG  1149
 R   F   F   L   E   N   G   K   H   D   E   A   W   M   V   L   K   Q   V   H   D   T   N   M   R   383
                                                   ***********************************
GCC AAG GGG CAT CCT GAG CGA GTC TTC TCG GTA ACC CAC ATT AAG ACA ATT CAT CAG GAG GAT GAG TTC ATT GAG  1224
 A   K   G   H   P   E   R   V   F   S   V   T   H   I   K   T   I   H   Q   E   D   E   L   I   E   408

ATC CAG TCT GAC ACA CGG GCC TGG TAC CAG CGC TGG GGG GTC CGG GCC TTG AGC CTG GGA GGG CAG GTC TGG GGG  1299
 I   Q   S   D   T   G   A   H   Y   Q   R   W   G   V   R   A   L   S   L   G   G   Q   V   W   G   433
                  T
AAT TTC CTC TCT TGT TTT GGT CCA GAA TAC CGC CGC ATC ACT CTG ATG ATG ATG GGT GTG TGG TTC ACC ATG TCA  1374
 N   F   L   S   C   F   G   P   E   Y   R   R   I   T   L   M   M   M   G   V   W   F   T   M   S   458
                      S
TTC AGC TAC TAT GGC CTG ACT GTC TGG TTT CCC GAC ATG ATC CGC CAT CTC CAA GCG GTG GAC TAT GCA GCC CGC  1449
 F   S   Y   Y   G   L   T   V   W   F   P   D   M   I   R   H   L   Q   A   V   D   Y   A   A   R   483

ACC AAA GTG TTC CCT GGG GAA CGT GTG GAG CAC GTG ACT TTT AAC TTC ACC TTG GAG AAT CAG ATC CAC CGA GGG  1524
 T   K   V   F   P   G   E   R   V   E   H   V   T   F   N   F   T   L   E   N   Q   I   H   R   G   508

GGA CAG TAC TTC AAT GAC AAG TTC ATT GGG CTA CGT CTG AAG TCA GTA TCC TTT GAG GAC TCC CTG TTT GAG GAA  1599
 G   Q   Y   F   N   D   K   F   I   G   L   R   L   K   S   V   S   F   E   D   S   L   F   E   E   533

TGT TAT TTC GAG GAT GTC ACA TCC AGC AAC ACA TTT TTC CGC AAC TGC ACG TTC ATC AAC ACC GTG TTC TAT AAC  1674
 C   Y   F   E   D   V   T   S   S   N   T   F   F   R   N   C   T   F   I   N   T   V   F   Y   N   558

ACT GAC CTG TTT GAG TAC AAG TTT GTG AAC AGC CGT CTG GTG AAC AGC ACA TTC CTG CAC AAC AAG GAG GGC TGC  1749
 T   D   L   F   E   Y   K   F   V   N   S   R   L   V   N   S   T   F   L   H   N   K   E   G   C   583

CCC CTG GAC GTG ACG GGG GGT GAA GGC GCC TAC ATG GTG TAT TTT GTC AGC TTC TTG GGG ACG CTG GCT GTG  1824
 P   L   D   V   T   G   T   G   E   G   A   Y   M   V   Y   F   V   S   F   L   G   T   L   A   V   608
                                           *****************************
CTT CCT GGG AAC ATT GTG TCT GCT CTG CTC ATG GAC AAG ATT GGC AGG CTC CGA ATG CTT GCT GGC TCC AGC GTG  1899
 L   P   G   N   I   V   S   A   L   L   M   D   K   I   G   R   L   R   M   L   A   G   S   S   V   633

ATG TCC TGT GTC TCC TGC TTC TTC CTG TCT TTC GGG AAC AGT GAG TCC GCC ATC ATC GCT CTG CTC TGC CTT TTC  1974
 M   S   C   V   S   C   F   F   L   S   F   G   N   S   E   S   A   M   I   A   L   L   C   L   F   658
 L
GGG GGG GTC AGC ATC GCA TCC TGG AAC GCG CTG GAC GTG TTG ACT GTT GAA CTC TAC CCC TCG GAC AAG AGA ACC  2049
 G   G   V   S   I   A   S   W   N   A   L   D   V   L   T   V   E   L   Y   P   S   D   K   R   T   683

ACA GCC TTC CGC TTC CTG AAT GCC CTG TGT AAG CTG GCA GCT GTG CTG GGG ATC AGC ATC TTC ACG TCC TTT GTG  2124
 T   A   F   G   F   L   N   A   L   C   K   L   A   A   V   L   G   I   S   I   F   T   S   F   V   708
     P
GGA ATC ACC AAG GCT GCC CCC ATC CTC TTT GCC TCC GCT GCC CTT GCC CTC GGG AGT TCT CTG GCC CTG AAG CTG  2199
 G   I   T   K   A   A   P   I   L   F   A   S   A   A   L   A   L   G   S   S   L   A   L   K   L   733

CCC GAG ACC CGG GGG CAG GTG CTG CAG TGA GGGAATCTCGGGCTTTGCGGGTGGCAGGCACACTGTGAGACCAACAATTCCCTCTTCCC  2274
 P   E   T   R   G   Q   V   L   Q   .                                                                 742

GTTCCCTGCCCTGCCATCCTGGCTGCCAGAGCCCTCCTTCCCCGCCTCCTGTTTAGTCTTAGCCATGTATGCGTGAGTGTGTGTCTGTATGTATGTGTG 2373
TGCATGTGTGAGACCCCCGTGGGCAGGGACTACAGCGAAGGCTCCTCCATCCTAGAATTTGGGGTGTAGCTCTGTTTCCCACCTCCTACCACCCCTGACTT 2472
TGCACAAGAGAAGGCTTGGCCTCACCCTTCTCCCCCAGTGTTAGTGAGAGGCCCTTGCTCTTCCCAGCCTCCTGGGGCTTTTCCCCCTTATTCCTTCTGC 2571
CTAGGCTCTGGCGAAACCACAAGCAGGTGGTTATGCCGGGGGCTCAAGGCTTGATATAGACCATGGACCAAAAGAACTTCTTAGAGTTGGAAGGGCCTAG 2670
AGTGCCCTCTCATGCCTCCGTGTTGGATGCTGGGAAAGTAGCAATAAACCTCAGCCCCCTGGCCTCCAGTTCCCCCTCAATTTGGGCTACAAGCCTGAAT 2769
TTTATCAAATTACCTTTCTGATTCTTATTTATTTATATATTAAGTTCTGAGGCAGCTCAGCAGCACTGTGTGTGTGTGAATGCTTGTATACACTTATGT 2878
ATGTGTGTGTGTGTGTGTATGTGTGCTGTGGGGTGGGGGTGTCATTATACTGTCCTAAAATACAAGCCAAGGGTAAATTTCAGTGGACACACACACAGGT 2977
ACCTGCTCCCCACAGTACCTCCCCTAATCTTGTTTCTGAGCCTGGCCTGGGTGGCAGGAGCCTCGGCAGGCTGGGAGAAGAGTCTCACAGTCTAGGAAC 3076
AGCCAAGGGCACCTGACAAAGTTCATCTCCTCCCCATCTCTCAAGAAGCACAGGCCTCCTCTGGAGTGAGAGGCTCTCCCTACACAGCACAGACGGCAA 3175
CAGCACAGCTGCCCTCCCACCTCGACTGTCTCCTCACAGGCAGAGGAGCAGGGGAGGGGAAGAAACCAGTATGTGGTCAAACCAGCAGATCAAAAGCAC 3274
AAGGAGCTGGGGCAGGGCAGGAACCGGGCTTTCCTAGCAACTCCTGCACAGTGGGAAGAGTTTCAGCAGTAAGCCGGGGACCCCTAATGGCAGGACCAG 3373
AAGCCTCAGTTTCCTCATCTCACCCTCCCCACAGGCATAGCCAGTGTAGGTTAAGCTGCCCCTGCC                                  3438
```
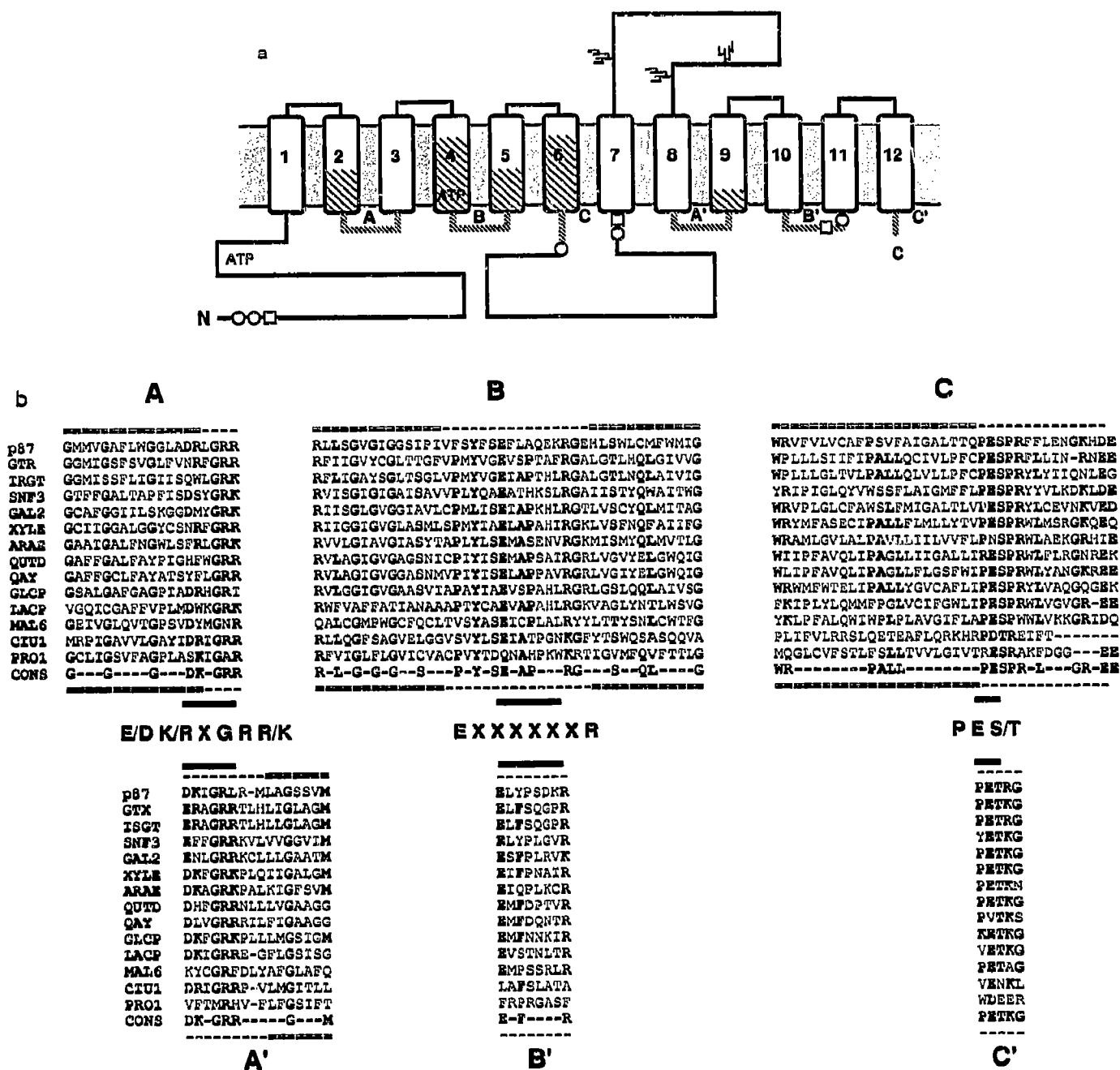
Fig. 3. Predicted transmembrane topography, post translational modifications and limited sequence alignment of p87. (A) Predicted post translational modification sites and topological membrane organization of p87. Based on the hydropathicity analysis of p87, the protein is expected to possess 12 TM segments which are indicated by the white rectangles. The membrane is indicated by the grey rectangle behind the TM segments. The polypeptide backbone which connects these TM segments is indicated by the black lines. The squares and circles imposed upon these lines indicate the presence of a consensus sites for protein kinase A and C respectively (Ser[47], Thr[147], Thr[089] and Thr[27], Ser[42], Ser[357], Ser[679]). The branch-like structures indicate the presence of consensus sites for N-linked glycosylation (Asn[479], Asn[548], Asn[573]). (B) Alignment of selected regions of various transport proteins. The regions shown in these alignments correspond roughly to the hatched areas of the p87 model shown in Fig. 3a. These regions are labelled A, B, C, A', B', and C'. Residues that are conserved in the majority of the transporter proteins listed are printed in bold letters. Areas thought to reside in transmembrane domains are bracketed by double broken lines. The transport proteins for which sequences are listed are: (1) p87, (2) the facilitated glucose transporter of rat (GluGTR); (3) the insulin responsive human glucose transporter (IRGT); (4) the high affinity glucose transporter of S. cerevisiae (SNF3), (5) the galactose transporter of S. cerevisiae (GAL2), (6) the xylose transport protein of E. coli (XYLE), (7) the arabinose transport protein of E. coli (ARAE), (8) the hexose-proton cotransporter of green algae (QUTD), (9) the quinate transporter of fungus (QAY), (10) the glucose transport protein of blue-green algae (GLCP), (11) the lactose permease of S. cerevisiae (LACP), (12) maltose permease gene from S. cerevisiae (MAL6), (13) the citrate-proton symport of E. coli (CIU1), (14) the probable transport protein from leishmania (PRO1). A consensus sequence is shown on the bottom line of each cluster. Sequences for these transporter proteins were obtained from the Swiss protein databank.

amino- and carboxyl-termini are thought to lie on the same face of the membrane bilayer and in the case of the mammalian glucose transporter, this model of transmembrane topology has been tested and found to be consistent with experimental results [5,7,10,11]. Thus, based on the similarity of the hydropathicity of p87 with these other proteins, we have proposed a model of transmembrane organization (Fig. 3a) which places the amino and carboxy-termini of p87 on the 'cytoplasmic' face of the membrane. This orientation places the large loop between putative TM 7 and 8 on the 'extracellular' face of the membrane which is consistent with the presence of consensus sites for N-linked glycosylation in this domain. Similarly, this orientation also places all of the consensus substrate sites for regulatory protein kinases on the putative cytoplasmic domains of the protein. The relative position of these potential sites of post-translational modification in p87 are summarized in Fig. 3a.

The second feature which most of these proteins share is the presence of one or both of the so-called nutrient transporter 'signature' sequences which are well conserved in p87 and located in the expected position (between TMs 4–5 and TMs 8–9). These signature sequences are highly specific and unlikely to be found by chance. The p87 protein also contains two versions of the well conserved 'RXGRR' motif which is found in many members of this family and which is thought to form a b-turn connecting TMs 2 and 3 and TMs 8 and 9 [1].

4.2. Does p87 function as a transporter?

The homology which p87 shares with the family of nutrient transporters suggests that p87 likely functions

Fig. 5. Subcellular localization of p87 immunoreactivity. Membrane fractions were prepared as described [22], electrophoresed, blotted onto nitrocellulose and incubated with MIL9 antibody in the presence (+) and absence (−) of the antigenic peptide. The primary antibody was detected by a secondary anti-rabbit antibody conjugated with horseradish peroxidase and diaminobenzidine. The arrow on the right-hand side indicates the immunoreactive band which is blocked by peptide. See text for details of the composition of each fraction.

as a transport protein in the CNS. Thus, the obvious question which remains to be answered is the nature of the substrate for p87. It seemed likely that given the rather specific localization of p87 to the brain, the substrate would likely be a compound which is found predominantly in the CNS. However, it is now known that the Na^+ and Cl^- dependent plasma membrane transport proteins for compounds such as the neurotransmitters (e.g. GABA, NE, DA and 5-HT) bear little or no sequence homology to p87 (reviewed in [32]). Thus, not unexpectedly, expression studies in mammalian cells of p87 failed to show any accumulation across the plasma membrane of several major neurotransmitter compounds (data not shown). The strong homology which p87 shows to the proton-driven bacterial transporter family suggests that the substrate would likely be a 'nutrient.' However, the only sugar utilized by the brain for metabolism is glucose, and glucose transport proteins have already been identified in the brain [3,17]. Moreover, we have failed to detect the plasma membrane transport of a variety of carbohydrates in cells transfected (transiently and permanently) with p87. In addition, p87 does not appear to transport amino acid substrates such as glutamate, aspartate, glycine, GABA (data not shown).

Another possibility for the apparent lack of detectable transport activity for p87 might be that the protein is directed to another cellular organelle than the plasma membrane. Several properties of p87 are consistent with this hypothesis. First, given the neuronal specificity of p87, it might be expected to subserve a function common to neurons but not all cells. One of the intracellular specializations unique to neurons is the synaptic vesicle. Although not conclusive in this respect, the subcellular distribution of p87 presented here (Fig. 5) is consistent with this possibility. Second, a vesicular location of p87 would fit with its structural homology with the bacterial transporters which like vesicular transporters are often

Fig. 4. Northern blot analysis of p87 mRNA. In each of the lanes 30 ng of total RNA from each of the indicated tissues was electrophoresed on a 1% agarose/formaldehyde gel, transferred to nitrocellulose, hybridized to a radiolabelled probe derived from the coding region of p87 cDNA, washed, and exposed to X-ray film for 24 h to obtain the autoradiogram shown. The arrows to the right-hand side of the autoradiogram indicate the position of mRNA standards.

proton-driven (Stern-Bach et al., 1990). In addition, the widespread distribution of p87 is not consistent with any one single neurotransmitter system in the brain and would suggest the p87 substrate could well be a common constituent of several types of synaptic vesicles. It is tempting to speculate, given the presence of two consensus ATP binding sites in this protein, that p87 could play a role in the transport of ATP or protons which are found in many types of synaptic vesicles. A transporter for ATP might be expected to display some specificity for sugar residues (e.g. ribose), thus the similarity of p87 with sugar transporters. Obviously, this possibility remains to be thoroughly tested. The difficulty in testing such a hypothesis has been the development of suitable assay systems.

Thus, in this report we describe a new CNS specific protein which appears to define a new transporter family. Thusfar the basis for the copurification of p87 with the D_1 dopamine receptor is still not understood. However, the further characterization of this novel protein using the tools (cDNA probes, cell lines, antibodies, etc) which have been developed should eventually lead to a more complete understanding of the functional role of p87 in the brain. Further work may likely reveal other members of this family of proteins and provide new insights into brain function.

REFERENCES

[1] Baldwin, S.A. and Henderson, P.J.F. (1989) Annu. Rev. Physiol. 51, 459–471.

[2] Bennett, M.K., Calakos, N. and Scheller, R.H. (1992) Science 257, 255–259.

[3] Birnbaum, M.J., Haspel, H.C. and Rosen, O.M. (1986) Proc. Natl. Acad. Sci. USA 83, 5784–5788.

[4] Blakeley, R.D., Berson, H.E., Fremeau Jr., R.J., Caron, M.G., Peek, M.M., Prince, H.K. and Bradley, C.C. (1991) Nature 354, 66–70.

[5] Bolfield, M.C. and Wilson, T.H. (1989) J. Biol. Chem. 264, 11649–11652.

[6] Bunzow, J.R., Van Tol, H.H.M., Grandy, D.K., Albert, P., Salon, J., Christie, M., Machida, C.A., Neve, K.A. and Civelli, O. (1988) . Nature 336, 783–787.

[7] Cairns, M.T., Alvarex, J., Panico, M., Gibbs, A.F., Morris, H.R. et al. (1987) Biochim. Biophys. Acta 905, 295–310.

[8] Catterall, W.A. (1988) Science 242, 50–61.

[9] Cullen, B.R. (1987) Methods Enzymol. 152, 684–704.

[10] Davies, A., Meeran, K., Cairns, M.T. and Baldwin, S.A. (1987) J. Biol. Chem. 262, 9347–9352.

[11] Eckert, B. and Beck, C.F. (1989) J. Biol. Chem. 264, 11663–11670.

[12] Fargin, A., Raymond, J.R., Lohse, M.J., Kobilka, B.K., Caron, M.G. and Lefkowitz, R.J. (1989) Nature 335, 358–360.

[13] Fremeau Jr., R.T., Caron, M.G. and Blakely, R.D. (1992) Neuron 8, 915–926.

[14] Gingrich, J.A., Amlaiky, N., Senogles, S.E., Chang, W.K., McQuade, R.M., Berger, J.G. and Caron, M.G. (1988) Biochemistry 27,3907–3912.

[15] Gingrich, J.A., Andersen, P.H. and Caron, M.G. (1990) Clin. Res. 38, 290A.

[16] Gingrich, J.A., Andersen, P.H., Joergensen, T.N., Guldhammer, B.S., El Mestikawy, S., Fremeau Jr., R.T., Tiberi, M. and Caron, M.G. (1990) Soc. Neurosci. Abstr. 287.7.

[17] Gould, G.W and Bell, G.I. (1990) Trends Biochem. Sci. 15, 18–23.

[18] Hamblin, M.W. and Metcalf, M.A. (1991) Mol. Pharmacol. 40, 143–148.

[19] Hunter, T. (1991) Methods Enzymol. 200, 3–37.

[20] Jahn, R., Schiebler, W., Ouimet, C. and Greengard, P. (1985) . Proc. Natl. Acad. Sci. USA 82, 4137–4141.

[21] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.

[22] Laduron, P. (1977) Int. Rev. Neurobiol. 20, 251–281.

[23] Lathe, R. (1985) J. Mol. Biol. 183, 1–12.

[24] Libert, F., Parmentier, M., Lefort, A., Dinsart, C., Van Sande, J., Maenhaut, C., Simons, M.J., Dumont, J.E. and Vassart, G. (1989) Science 244, 569–572.

[25] Libert, F., Schiffmann, S.N., Lefort, A., Parmentier, M., Gérard, C., Dumont, J.E., Vanderhaeghen, J.J. and Vassart, G. (1991) EMBO J. 10, 1677–1682.

[26] Maiden, M.C.J., Davis, E.O., Baldwin, S.A., Moore, D.C.M. and Henderson, P.J.F. (1987) Nature 325, 641–643.

[27] Matsuda, L.A., Lolait, S.J., Brownstein, M.J., Young, A.C. and Bonner, T.I. (1990) Nature 346, 561–564.

[28] Matthew, W.D., Tsavaler, L. and Reichardt, L.F. (1981) J. Cell Biol. 91, 257–269.

[29] Perin, M.S., Fried, V.A., Mignery, G.A., Jahn, R. and Südhof, T.C. (1990) Nature 345, 260–263.

[30] Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski J, Lok, S., Plavsic, N., Chou, J., Drumm, M.L., Iannuzzi, M.C., Collins, F.S. and Tsui, L.-C. (1989) Science 245, 1066–1073.

[31] Südhof, T.C., Baumert, M., Perin, M.S. and Jahn, R. (1989) Neuron 2, 1475–1481.

[32] Uhl, G.R. (1992) Trends Neurosci. 15, 265–268.

Note added in proof

The cDNA and protein sequences for rat p87 reported in this paper are virtually identical to the sequences reported recently for rat synaptic vesicle protein 2 (SV2), a membrane glycoprotein localized to secretory vesicles (Bajjalieh, S.M., Peterson, K., Shinghal, R. and Scheller, R.H., Science 257, 1271–1273, 1992; Feany, M.B., Lee, S., Edwards, R.H. and Buckley, K.M., Cell 70, 861–867, 1992).