Amino Acid Sequence of P-57, a Neurospecific Calmodulin-Binding Protein†

Bassam T. Wakim,‡ Kenneth A. Alexander,§ H. Robert Masure,§ Berta M. Cimler,§ Daniel R. Storm,§ and Kenneth A. Walsh*,‡

Departments of Biochemistry and Pharmacology, University of Washington, Seattle, Washington 98195

Received May 5, 1987; Revised Manuscript Received June 15, 1987

ABSTRACT: The amino acid sequence was determined for bovine brain P-57, a neurospecific, membrane-associated, calmodulin-binding protein. It consists of a single 239-residue polypeptide chain blocked at its amino terminus and containing an unusually hydrophilic amino acid composition. Seventy percent of the molecule is composed of Glu/Gln, Ala, Lys, Asp/Asn, and Pro; there is only one aromatic residue. A lack of favorable cleavage sites required that a particularly wide variety of digests and subdigests be performed to obtain appropriate sets of overlapping peptides. This protein is clearly homologous with the cDNA-derived sequence of mouse brain P-57, although the bovine protein is 12 amino acid residues longer; the homology is less obvious in the middle sections of the two sequences. Bovine brain P-57 lacks homology with any other protein in an updated sequence database. A segment reported to interact with calmodulin (Arg-Gly-His-Ile-Thr-Arg-Lys-Lys-Leu) is placed at residues 43–51 within the only extended segment of P-57 that carries the net positive charge that would favor that interaction. There is no hydrophobic segment characteristic of many proteins that interact with membranes.

P-57 is a neurospecific protein that binds calmodulin (CaM).[1] It was discovered in bovine brain membrane fractions by virtue of its unusual interaction with CaM–Sepharose (Andreasen et al., 1983; Cimler et al., 1985), where it displayed greater affinity for CaM in the absence than in the presence of Ca^{2+}. In contrast, CaM interacts with a majority of target proteins with greater affinity in the presence of Ca^{2+}. For example, its affinity for three CaM-regulated enzymes, myosin light chain kinase, cyclic nucleotide phosphodiesterase, and phosphorylase kinase, is greatly enhanced by Ca^{2+} (Blumenthal & Stull, 1980; Crouch et al., 1981; Burger et al., 1983; Olwin et al., 1984; Olwin & Storm, 1985). There are few other examples of proteins that bind CaM with equal or greater affinity in the absence of Ca^{2+} (Van Eldik & Burgess, 1983; Burgess et al., 1984).

P-57 migrates on SDS–PAGE gels with an apparent molecular weight of 57 000. However, hydrodynamic studies demonstrated that its molecular weight is closer to 25 700, with a Stokes radius of 4.58 nm (Masure et al., 1986). Sucrose density gradient sedimentation and fluorescence polarization measurements indicated that P-57 is an unusually elongated molecule with an axial ratio of 16:1 (Masure et al., 1986). Its amino acid composition is distinctive, containing a single phenylalanine residue, no other aromatic amino acid, and unusually high percentages of alanine, glutamic acid, lysine, and proline (Masure et al., 1986).

In the course of exploring the interaction of P-57 and CaM, we sought details of the structure of P-57. The amino acid sequence of the bovine brain protein was determined, as described herein.

Materials and Methods

Bovine brain P-57 was prepared as previously described (Masure et al., 1986) and then reduced and *S*-carboxymethylated as described by Takio et al. (1983). S-CM-P-57 was subjected to a final purification step on a tandem pair of TSK SW3000 columns in the presence of 6 M guanidine hydrochloride and 10 mM phosphate, pH 6.0.

Cleavage at methionine residues was performed by incubating the protein (20 nmol) with 10 mg of cyanogen bromide in 72% v/v formic acid for 15 h at room temperature in the dark. Cleavage at aspartic acid was achieved by incubation in 2 M formic acid under vacuum at 100 °C for 4 h.

Digestion of S-CM-P-57 with TPCK–trypsin (Worthington) took place in 0.1 M NH_4HCO_3, pH 8.0, at 37 °C using a protease:substrate mole ratio of 1:100 for 2 h. For cleavage at arginine residues, CM-P-57 was citraconylated according to the method of Atassi and Habeeb (1972), dialyzed at pH 8.0 to remove excess reagents, and then digested with TPCK–trypsin as above.

Fragment M1 was subdigested with chymotrypsin in 0.1 M NH_4HCO_3, pH 8.0, for 6 h at 37 °C using a protease:substrate weight ratio of 1:100 and *p*-aminobenzamidine (10 mM) as a trypsin inhibitor. Fragment M2 was cleaved at lysine residues with endopeptidase Lys-C (1:100 w/w) in 50 mM Tris-HCl, pH 9.0, for 2 h at 37 °C. C2 was subdigested with thermolysin (1:100 w/w) in 0.1 M NH_4HCO_3 and 1 mM $CaCl_2$, pH 7.8, for 6 h at 37 °C. Fragments R2 and C2 were separately subdigested with *Staphylococcus aureus* V8 protease (1:100 w/w) in 0.1 M NH_4HCO_3, pH 8.0, for 20 h at 37 °C.

Mixtures of peptides were first fractionated by exclusion chromatography on tandem TSK columns according to the procedure of Titani et al. (1986). Pooled fractions were desalted and subfractionated by reversed-phase HPLC using acetonitrile gradients and Altex Ultrapore RPSC, SynChropak RP-P or RP-8, or Hypersil ODS columns. In some cases more

† This work was supported by National Institutes of Health Grants GM-15731 (to K.A.W.) and GM-33708 (to D.R.S.), National Research Service Awards NS-07985 (H.R.M.) and GM-07270 (B.M.C.), the Life and Health Insurance Medical Research Fund (K.A.A.), and Postdoctoral Training Grant AM-07441 (B.T.W.).

* Address correspondence to this author.

‡ Department of Biochemistry.

§ Department of Pharmacology.

[1] Abbreviations: CaM, calmodulin; CMC, *S*-carboxymethyl-Cys; EGTA, ethylene glycol bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid; HPLC, high-performance liquid chromatography; P-57, a bovine brain protein of M_r 24 721 with anomalous SDS–PAGE mobility of M_r 57 000; S-CM, *S*-carboxymethyl; SDS–PAGE, polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate; TPCK, L-1-(tosylamido)-2-phenylethyl chloromethyl ketone; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride.

0006-2960/87/0426-7466$01.50/0

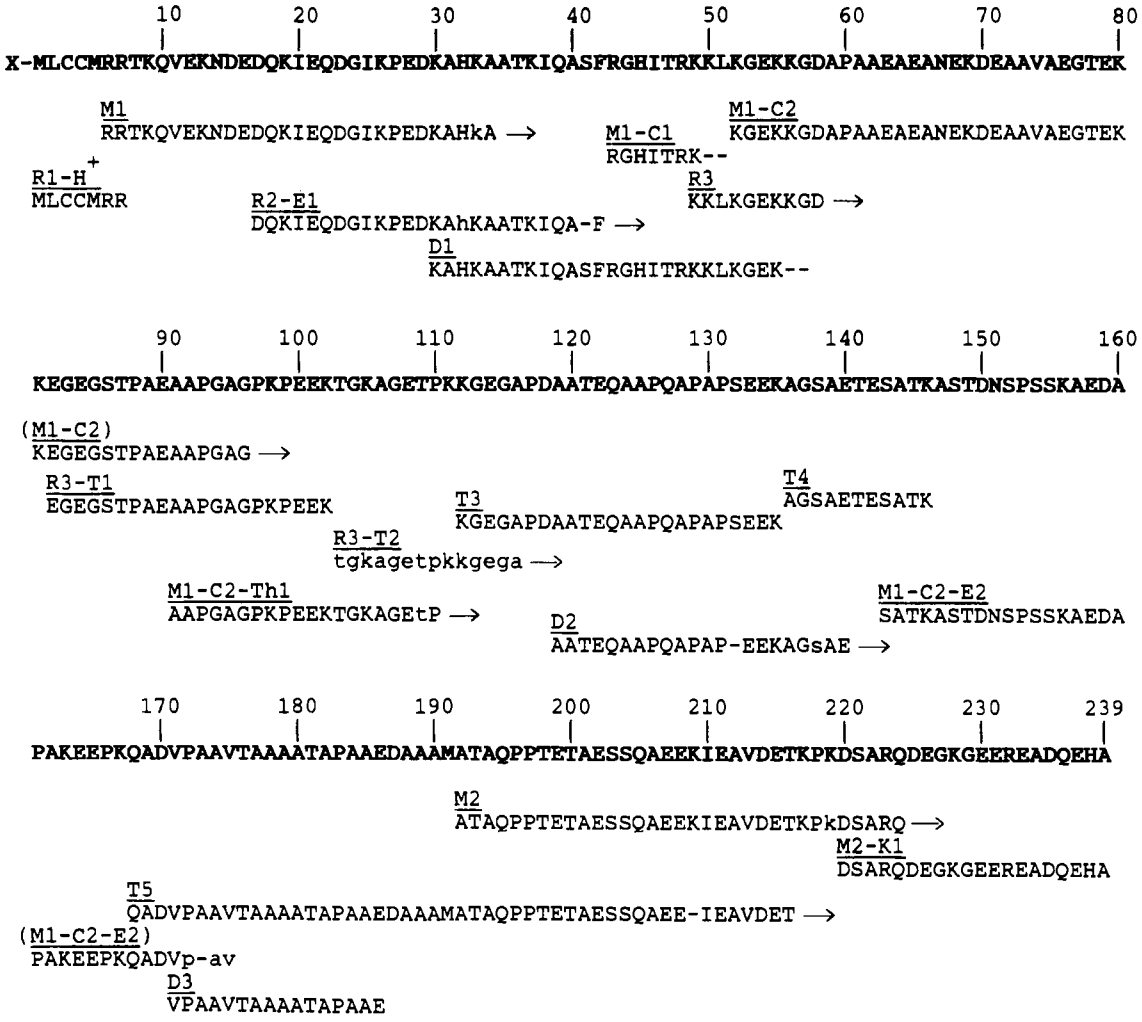

FIGURE 1: Detailed summary of the proof of sequence of bovine brain P-57. The sequences derived by Edman degradation of selected peptides (names are underlined) are displayed below the summary sequence (bold type). Hyphenated names of peptides indicate products of subdigestion. Prefixes M, R, E, K, and D are peptides derived by cleavage at methionyl, arginyl, glutamyl, lysyl, and aspartyl residues, respectively. Prefixes T, C, and Th refer to peptides derived from tryptic, chymotryptic, and thermolytic digests, respectively. R1-H^+ denotes the product of mild acid treatment of R1 (see text). Unidentified residues are indicated by hyphens or by arrows for longer unidentified sequences. Lower case letters indicate tentative identifications of phenylthiohydantoins toward the end of an analysis (e.g., R2-E1) or by resolution of a mixture (R3-T2).

than one column separation was needed. Fractions R2-E1, R3-T1, R3-T2, M1-C2-E2, and M2-K1 were purified directly on a SynChropak RP-P column without prior size-exclusion chromatography.

Amino acid compositions were determined in 20-h hydrolysates by reversed-phase HPLC with a Waters Picotag system (Bidlingmeyer et al., 1984).

Amino acid sequences were determined with either a Beckman 890C sequencer or an Applied Biosystems 470A sequencer with a Model 120A on-line HPLC system. The phenylthiohydantoins from the Beckman sequencer were identified in complementary HPLC systems (Glajch et al., 1985; Ericsson et al., 1977).

Searches for homologous sequences and alignments of related sequences were done on a VAX/VMS computer using the SEARCH, ALIGN, and RELATE programs of Dayhoff et al. (1983) and the National Biomedical Research Foundation data base of December 4, 1986.

RESULTS

The amino acid composition of P-57 (Masure et al., 1986) indicated that there would be several problems in developing a successful strategy for the analysis of this 239-residue protein. Specifically, there appeared to be too many tryptic cleavage sites (30 Lys and 6 Arg) and too many potential glutamyl cleavage sites (50 Glx). There was also a distinct lack of chymotryptic cleavage sites (two Leu, two Met, and one Phe), although in the end the amino acid sequence (Figure 1) revealed a third methionyl residue. Apparently significant oxidation of methionine had occurred during hydrolysis. In addition, it was observed that the protein was blocked at its amino terminus, and the high content of proline (7.5%) was expected both to interfere with enzymatic cleavages and to decrease the efficiency of Edman degradations.

In preliminary experiments, a strategy was based on cleavage at three methionyl residues with cyanogen bromide and at six arginine residues of the citraconylated protein with trypsin. However, the distribution of these residues was quite uneven. Two of the three methionines are within five residues of the amino terminus, and the six arginine residues are so located that there are 175 residues between the fourth and the fifth arginines. We turned next to a chymotryptic digest, but both leucines and the single phenylalanine were located in the amino-terminal 51 residues, leaving a very large fragment from Lys-52 to Met-191. Thus, after Edman degradation of the fragments available from the above digests, it was clear that it would be difficult to obtain sequence information from the portion of the molecule corresponding to residues 97–191 (Figure 1). Information was obtained from tryptic digests wherein several lysine residues resisted tryptic attack, yielding

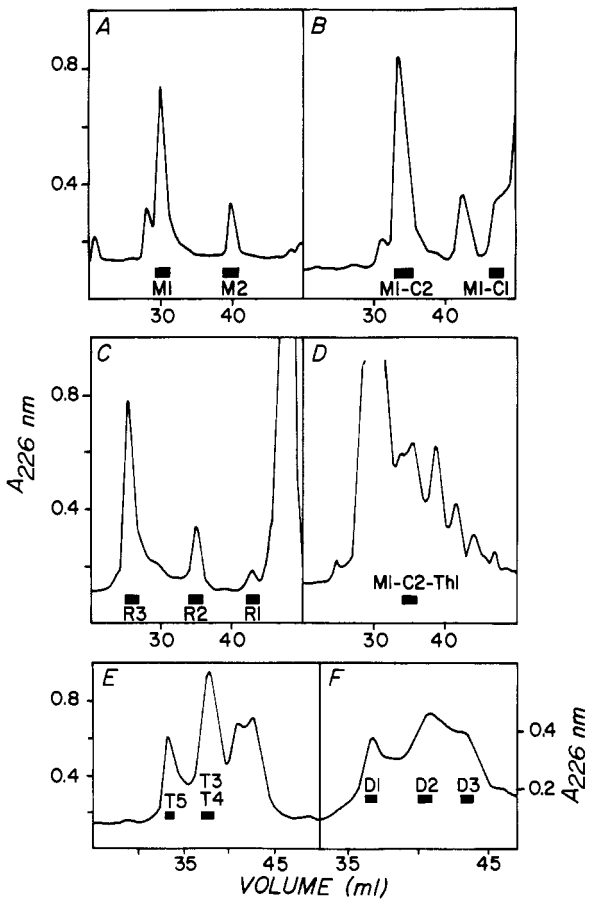

FIGURE 2: Primary separations of peptide mixtures resulting from (A) Met cleavage in S-CM-P-57, (B) a chymotryptic digest of M1, (C) arginine cleavage in S-CM-P-57, (D) a thermolytic digest of M1-C2, (E) a tryptic digest of S-CM-P-57, and (F) aspartic acid cleavage in S-CM-P-57. In each case, peptides were separated by size in 6 M guanidine hydrochloride and 10 mM phosphate, pH 6.0, at a flow rate of 1 mL/min. The cyanogen bromide (A) and chymotryptic (B) digests were separated on a tandem pair of TSK SW3000 columns. The other digests were separated on a tandem pair of TSK SW2000 columns. Each pooled fraction was desalted and further purified with an acetonitrile gradient (in dilute aqueous trifluoroacetic acid) and reversed-phase HPLC. Peptides from cyanogen bromide cleavage and from the chymotryptic digest were purified on an Ultrapore RPSC column. Tryptic fragments were purified on an RP-8 column. The thermolytic peptide, M1-C2-Th1, was desalted on an RP-8 column and purified on a Hypersil ODS column. Peptides from aspartic acid cleavage were purified on an RP-P column.

peptides of satisfactory size from this central region. It was still necessary to employ a thermolytic digest and cleavage at aspartic acid to complete the sequence analysis.

Another unexpected difficulty, also related to the unusual composition of this protein, was recognized during the purification of peptides by reversed-phase HPLC. The paucity of hydrophobic residues and the ubiquity of Ala, Pro, Glu, and Lys yielded many peptides of similar mobility on reversed-phase HPLC. Most peptides eluted from reversed-phase columns in 15–20% acetonitrile. In addition, amino acid composition data were less helpful than usual in seeking critical peptides because the compositions tended to be so similar.

The primary separations of the various digests were made according to size on tandem TSK columns, as illustrated in Figure 2. Peptides in pooled fractions were then purified by reversed-phase HPLC using acetonitrile gradients. Peptides in subdigests of R2, R3, and M2 were directly purified on a SynChropak RP-P column. Each of the purified peptides was subjected to Edman degradation, and the overlapping sets of information are summarized in Figure 1.

P-57 is blocked at the amino terminus. The peptide R1, obtained by cleavage at Arg-6, was also blocked. R1 eluted from a SynChropak RP-P column in 19% acetonitrile as might be expected with an *N*-acetyl blocking group (Walsh & Sasagawa, 1984). Mild treatment with acid (2 M formic acid, 4 h, 110 °C, under vacuum) unblocked about 20% of this peptide, and Edman degradation yielded the sequence MLCCMRR, in accord with its amino acid composition. This overlapped M1, which in turn overlapped other segments of sequence information to yield a definitive sequence to Pro-110, although Thr-109 was only tentatively identified. Similarly, a segment from Lys-112 to Ala-239 was deduced from the overlapping segments illustrated in Figure 1. Evidence of Ala-239 as the C-terminal residue is based solely on its observation as the last phenylthiohydantoin from the Edman degradation of M2-K1.

To complete the analysis, it remained only to link Pro-110 to Lys-112. This was accomplished with the sequence labeled R3-T2 in Figure 1. This sequence was deduced by analysis of a mixture of two major peptides, with N-termini at residues 103 and 112, respectively. By subtracting the known sequence of residues 112–125, we derived the sequence illustrated for R3-T2. This deduced sequence also confirmed the presence of Thr at residue 109.

The amino acid sequence of P-57 was examined by the RELATE program of Dayhoff et al. (1983) for evidence of internal homology, but none was found. Nor was any homologous protein or protein segment identified by the SEARCH program and the protein data base of the National Biomedical Research Foundation. In a separate paper (Cimler et al., 1987), the cDNA-derived sequence of *mouse* brain P-57 is presented. Its similarity to bovine brain P-57 is illustrated in Figure 3. Bovine P-57 is 12 amino acid residues longer and, as displayed in Figure 3, the homology between the two proteins is evident throughout, although less striking in the central region.

DISCUSSION

The amino acid composition of P-57 is distinctive; 50% is composed of Ala, Glu, and Lys residues. Of the 239-residue total, there are 90 charged groups and only a single aromatic amino acid. Moreover, this unusual composition is distributed in an uneven fashion. For example, four of the five sulfur-containing residues are found in the amino-terminal five residues, and 50% of the residues between 171 and 198 are alanine.

The most difficult problems in the analysis of the sequence of P-57 were related to this composition and to the lack of ideal cleavage points between residues 51 and 191. That region of the molecule appeared to be too rich in lysine and glutamic acid residues and virtually lacking in other scissile bonds. Fortunately, nearby Pro and Glu residues restricted tryptic digestion, facilitating the generation of peptides that provided necessary overlaps in this region.

The weakest part of the proof of sequence is in the area of Thr-109 to Lys-111, including the overlap between R3-T2 and T3 in Figure 1. The sequence of R3-T2 was deduced from a mixture of two major peptides, the second of which was placed elsewhere in the structure. In three separate attempts we were unable to determine the sequence of M1-C2-Th1 for more than 20 residues in spite of its larger size (5 kDa). This was in part due to the three prolines early in the sequence that tend to give incomplete Edman degradation. The high proline content of P-57 decreased the efficiency of sequencing several other peptides. The amino acid composition based on the

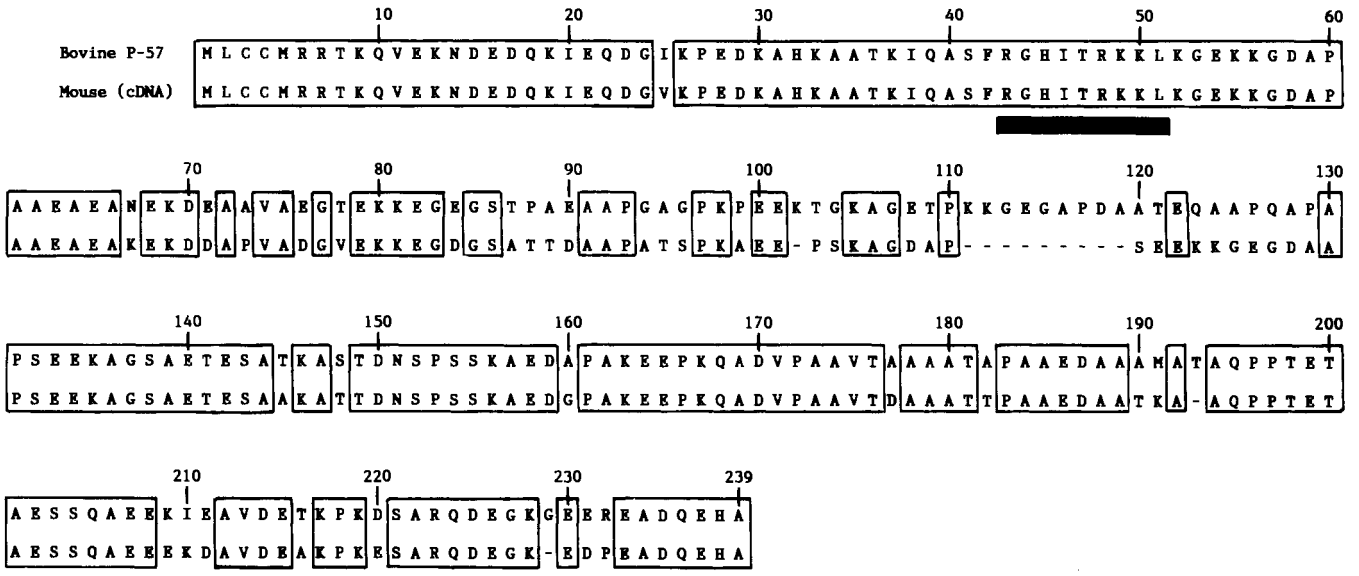

FIGURE 3: Sequence homology between bovine brain P-57 (this study) and mouse brain P-57 (Cimler et al., 1987) as optimized by the ALIGN program of Dayhoff et al. (1983). Identical residues are enclosed in boxes. The CaM-binding segment in the bovine protein, reported by Wakim et al. (1986), is underlined with a dark bar.

Table I: Amino Acid Composition of S-CM-P-57 and Selected Peptides[a]

	fragment (residues)					
	P-57 (1–239)	M1 (6–191)	M2 (192–239)	M1-C1 (43–51)	M1-C2 (52–191)	R1 (1–7)
Asx (D/N)	16.4 (18)	10.9 (14)	2.7 (4)		4.6 (9)	
Glx (E/Q)	49.2 (49)	39.3 (34)	14.0 (15)		27.3 (26)	
CMC (C)	0.5 (2)	0.2 (0)				1.6 (2)
Ser (S)	11.4 (12)	9.3 (9)	2.9 (3)		7.7 (8)	
Gly (G)	15.3 (16)	15.9 (14)	3.0 (2)	0.7 (1)	13.1 (12)	
His (H)	3.1 (3)	1.8 (2)	1.0 (1)	0.4 (1)		
Arg (R)	7.1 (6)	3.8 (4)	1.9 (2)	2.0 (2)	1.4 (0)	1.9 (2)
Thr (T)	16.7 (17)	9.4 (13)	3.1 (4)	0.9 (1)	9.2 (10)	
Ala (A)	41.4 (52)	36.8 (44)	8.8 (8)	1.3 (0)	43.2 (40)	
Pro (P)	20.4 (19)	17.5 (16)	3.4 (3)		16.7 (15)	
Val (V)	5.0 (5)	4.1 (4)	1.3 (1)		3.4 (3)	
Met (M)[b]	0.5 (3)	0 (1)			0.3 (1)	1.6 (2)
Ile (I)	4.9 (5)	4.0 (4)	1.3 (1)	1.1 (1)		
Leu (L)	2.1 (2)	1.7 (1)		1.2 (1)		1.1 (1)
Phe (F)	1.1 (1)	1.0 (1)				
Lys (K)	27.6 (29)	23.6 (25)	4.3 (4)	1.7 (2)	19.1 (16)	
total residues	239	186	48	9	140	7
yield (%)		39	51	23	44	25[c]

[a] Residues per molecule by Picotag amino acid analysis; numbers in parentheses are derived from the sequence in Figure 1. [b] Determined as homoserine in M1, M2, M1-C1, and M1-C2. [c] Approximate total of several chromatographic fractions, presumably the result of partial oxidation of Met and CMC.

sequence is in reasonable agreement with that determined in hydrolysates (Table I) except for alanine, which was lower in hydrolysates. Although our evidence for the identity of the carboxyl terminus is not ideal, it is consistent with that of the *mouse* brain protein, where the nucleotide sequence displays a stop codon after the Glu-His-Ala sequence (Cimler et al., 1987).

The high content of glutamic acid (15.9%) and alanine (21.8%), both of which are strong α-helix formers, is indicative of possible α-helices in P-57. However, analysis of the sequence for secondary structure by the method of Chou and Fasman (1978) predicted no α-helix formation between residues 1 and 172. By that same method, three short sequences between residues 173 and 217 are compatible with α-helix. Direct measurements, however, by circular dichroism spectroscopy indicate less than 1% α-helix in P-57 and about 78% random coil (Masure et al., 1986).

Considering the axial ratio of the protein, its lack of secondary structure, and its lack of components for a hydrophobic core, it is possible that its interaction with the apocalmodulin

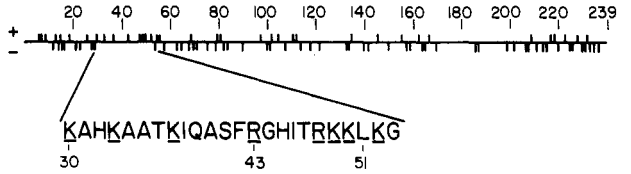

FIGURE 4: Distribution of charged residues in bovine P-57. Attention is drawn to residues 30–53, which lack negative charges and include a cluster of basic residues. As the only extended sequence with a net positive charge, this region has the potential of interacting with the highly negatively charged apocalmodulin molecule.

molecule is due solely to a segment of the primary structure of P-57. Since calmodulin is a very acidic protein, a putative binding segment in P-57 would be expected to carry a net positive charge in spite of the overall negative charge (−20) of the whole molecule. Examination of the charge distribution within P-57 reveals that the longest segment lacking Asp or Glu (residues 30–53) contains a cluster of six Lys, two Arg, and two His within the only region carrying a predominantly positive charge (Figure 4). It has already been reported that

a peptide from this region, M1-C1, resembled the whole molecule in having greater affinity for a CaM–Sepharose column in EGTA than in Ca^{2+} (Wakim et al., 1986). The sequence of that peptide is RGHITRKKL (residues 43–51). Experiments are being designed to examine its binding in greater detail and to confirm or identify the CaM-binding segment within P-57.

The cDNA-derived sequence of mouse brain P-57 has recently been determined by Cimler et al. (1987). It displays 77% identity with the bovine brain protein and is clearly homologous, although four gaps are necessary to optimize their alignment (Figure 3). The putative CaM-binding sequence of bovine P-57 lies within the longest segment of identity (residues 26–70). Half of the differences between the proteins from the two species lie between residues 87 and 129, including a nine-residue gap in the mouse protein. Since calmodulin is a very highly conserved protein, it is unlikely that residues 87–129 in P-57 contribute to the binding of calmodulin in either species.

If it is correct that an amino-terminal segment of P-57 interacts with apoCaM and a middle segment is unimportant, one is left with a question concerning the function of the carboxyl-terminal ~100 residues. Since P-57 binds membranes but lacks a hydrophobic anchor, one can speculate that the observed conservation in this segment is related to an interaction with a membrane component.

Added in Proof

After submission of the manuscript, the cDNA sequence of a rat neuronal, growth-related protein "GAP-43" was reported by Karns et al. (1987). That sequence is so closely related to those illustrated in Figure 3 that rat GAP-43 must correspond to bovine and mouse P-57.

Acknowledgments

We thank Roger D. Wade, Maria Harrylock, and Gerard Doyle for technical assistance, Lowell Ericsson and Drs. Koji Takio, Koiti Titani, and Harry Charbonneau for many valuable discussions, and Mary K. Woods for assistance with the manuscript.

References

Andreasen, T. J., Luetje, C. W., Heideman, W., & Storm, D. R. (1983) *Biochemistry 22*, 4615.

Atassi, M. Z., & Habeeb, A. F. S. A. (1972) *Methods Enzymol. 25*, 546.

Bidlingmeyer, B. A., Cohen, S. A., & Tarvin, T. L. (1984) *J. Chromatogr. 336*, 93.

Blumenthal, D. K., & Stull, J. T. (1980) *Biochemistry 19*, 5608.

Burger, D., Stein, E. A., & Cox, J. A. (1983) *J. Biol. Chem. 258*, 14733.

Burgess, W. H., Watterson, D. M., & Van Eldik, L. J. (1984) *J. Cell Biol. 99*, 550.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45.

Cimler, B. M., Andreasen, T. J., Andreasen, K. I., & Storm, D. R. (1985) *J. Biol. Chem. 260*, 10784.

Cimler, B. M., Giebelhaus, D. H., Wakim, B. T., Storm, D. R., & Moon, R. T. (1987) *J. Biol. Chem.* (in press).

Crouch, T. H., Holroyde, M. J., Collins, J. G., Solaro, R. J., & Potter, J. D. (1981) *Biochemistry 20*, 6318.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524.

Ericsson, L. H., Wade, R. D., Gagnon, J., MacDonald, R. M., Granberg, R. R., & Walsh, K. A. (1977) in *Solid Phase Methods in Protein Sequence Analysis* (Previero, A., & Coletti-Previero, M.-A., Eds.) p 137, Elsevier/North-Holland, New York.

Glajch, J. L., Gluckman, J. C., Charikofsky, J. G., Minor, J. M., & Kirkland, J. J. (1985) *J. Chromatogr. 318*, 23.

Karns, L. R., Ng, S., Freeman, J. A., & Fishman, M. C. (1987) *Science (Washington, D.C.) 236*, 597.

Masure, H. R., Alexander, K. A., Wakim, B. T., & Storm, D. R. (1986) *Biochemistry 25*, 7553.

Olwin, B. B., & Storm, D. R. (1985) *Biochemistry 24*, 8081.

Olwin, B. B., Edelman, A. M., Krebs, E. G., & Storm, D. R. (1984) *J. Biol. Chem. 259*, 10949.

Takio, K., Smith, S. B., Walsh, K. A., Krebs, E. G., & Titani, K. (1983) *J. Biol. Chem. 258*, 5531.

Titani, K., Kumar, S., Takio, K., Ericsson, L. H., Wade, R. D., Ashida, K., Walsh, K. A., Chopek, M. W., Sadler, J. E., & Fujikawa, K. (1986) *Biochemistry 25*, 3171.

Van Eldik, L. J., & Burgess, W. H. (1983) *J. Biol. Chem. 258*, 4539.

Wakim, B. T., Alexander, K. A., Masure, H. R., Walsh, K. A., & Storm, D. R. (1986) *Abstracts of Papers, Fifth International Symposium on Calcium Binding Proteins in Health and Disease*, Asilomar, p 132, Plenum, New York.

Walsh, K. A., & Sasagawa, T. (1984) *Methods Enzymol. 106*, 22.