Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 664–668, February 1988
Biochemistry

Human 67-kDa calelectrin contains a duplication of four repeats found in 35-kDa lipocortins

(**Ca^{2+}-binding proteins/protein-tyrosine kinase substrates/gene duplication/phospholipase A_2 inhibitor**)

THOMAS C. SÜDHOF*†, CLIVE A. SLAUGHTER*‡, ISABELLA LEZNICKI*, PAUL BARJON*, AND GARY A. REYNOLDS§

*Howard Hughes Medical Institute, Departments of †Molecular Genetics and ‡Biochemistry, and §Division of Cardiology, Department of Medicine, University of Texas Health Science Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75235

Communicated by Michael S. Brown, September 15, 1987

ABSTRACT **The 67-kDa calelectrin is the largest member of a family of Ca^{2+}-binding proteins that associate with membranes and phospholipids in a Ca^{2+}-dependent manner. Oligonucleotide probes based on peptide sequences obtained from purified bovine 67-kDa calelectrin were used to screen a human retina cDNA library, and the complete primary structure of human 67-kDa calelectrin was deduced by DNA sequence analysis. The protein consists of eight 68-amino acid repeats separated by linking sequences of variable lengths. It is highly similar to the human lipocortin I and II sequences, each of which contains four such repeats. The amino termini of the three proteins show no sequence similarity; however, in the repeated regions the proteins are 42–45% identical in sequence. Analysis of the 16 repeats from the three proteins provides insights into the structural basis for Ca^{2+}-dependent phospholipid binding. These data place the calelectrins and the lipocortins into the same gene family and suggest that these proteins have similar functions and have evolved from a common ancestor.**

Ca^{2+} acts as an intracellular second messenger by virtue of its concentration-dependent binding to regulatory proteins. Until recently, all intracellular Ca^{2+}-binding proteins were thought to belong to a single structural group, characterized by the presence of an EF hand motif (looped conformation in Ca^{2+}-binding protein between two helices E and F) (1). We have characterized a distinct family of Ca^{2+}-dependent lipid-binding proteins, the calelectrins, that consist of several immunologically cross-reacting proteins (2–4). Three members of this protein family with apparent molecular masses of 67 kDa, 35 kDa, and 32.5 kDa have been purified to homogeneity, and several additional related proteins have been identified immunologically (2, 3).

The calelectrins are abundant, evolutionarily conserved proteins, the cellular function of which is unknown. They are distinguished by their ability to reversibly bind to phospholipids and cell membranes in physiological Ca^{2+} concentrations. This property has permitted the isolation of these proteins by Ca^{2+}-dependent hydrophobic-interaction chromatography (2). Primary amino acid sequences from some members of this protein family have demonstrated that the calelectrins are structurally related to each other and to proteins that are intracellular protein-tyrosine kinase substrates and are variably called lipocortin I and II, p35 and p36, respectively—or, most recently, "calpactins" (5–9).

Lipocortin I and II are two related proteins each containing four similar repeats. Lipocortin I and II were identified by their ability to inhibit phospholipase A_2 activity and were purified using this inhibition as an assay (6, 7). Unexpectedly, when their amino acid sequences were deduced from the nucleotide sequence of cDNA clones, their sequences were found identical to the amino acid sequences of two intracellular protein-tyrosine kinase substrates of 35 kDa and 36 kDa (named p35 and p36, respectively), the primary structure of which was being determined concomitantly (5–9). p35 and p36 appear to be physiological intracellular substrates for $pp60^{v\text{-}src}$ and the epidermal growth factor receptor protein-tyrosine kinase (8, 9). Although several hypotheses on the function of lipocortin I and II (p35 and p36, respectively) have been advanced, their biological role is still controversial.

We report the cDNA cloning and primary structure of the human 67-kDa calelectrin,¶ the largest member of the calelectrin family. The 67-kDa calelectrin, or a similar protein, has probably also been identified by other scientists, who refer to the protein as a lymphocyte membrane-associated Ca^{2+}-binding protein (10), a chromobindin (11), the 67-kDa calcimedin (12), a 73-kDa protein (13), and protein III (14). The deduced sequence contains eight similar repeats, each consisting of ≈68 amino acids. Analysis of these eight repeats for similarity suggests that the 67-kDa calelectrin has evolved by a series of gene-duplication events. Alignment of the 67-kDa calelectrin sequence with lipocortin I and II protein sequences demonstrates a close relationship of 42–45% identity.

EXPERIMENTAL PROCEDURES

Protein Purification and Sequencing. The 67-kDa calelectrin was purified from bovine liver by Ca^{2+}-dependent hydrophobic-affinity chromatography and anion-exchange chromatography on a fast protein liquid chromatography (FPLC) Mono Q column (Pharmacia) as described (2, 15), except that the Mono Q chromatography was done at pH 7.4. Automated N-terminal amino acid sequence analyses were done with an Applied Biosystems (Foster City, CA) model 470A gas-phase sequencer, and amino acid residues were identified with an Applied Biosystems model 120A on-line phenylthiohydantoin amino acid analyzer, using the manufacturer's standard programs and chemicals. Two hundred picomoles of 67-kDa calelectrin were subjected to N-terminal sequencing and found to be blocked to Edman degradation. Peptides were generated from the intact protein by cleavage with CNBr in 70% (vol/vol) formic acid and were separated by reverse-phase HPLC, using a 2.1 × 100 mm Brownlee (Santa Clara, CA) RP300 column. Elution was accomplished with a 0–50% acetonitrile gradient in 0.1% trifluoroacetic acid. Individual peptide peaks were collected

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

¶The sequence reported in this paper is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03578).

directly onto 1-cm Whatman GF/C glass-filter discs and subjected to automatic sequence analysis. Starting with 1 nmol of purified protein, sequencing yielded 9–30 pmol.

cDNA Cloning and Nucleotide Sequencing. Two peptide sequences obtained from bovine 67-kDa calelectrin were similar to each other and to a repeated sequence in lipocortin I and II (P1 and P4 in Fig. 1). An oligonucleotide family (ATGAAGGG[CAG]CT[GC]GG[CAG]AC[TC]GA[TC]GA-[CAG]GA) complementary to this repeat was synthesized on an Applied Biosystems model 380B DNA synthesizer; nucleotides in brackets designate redundant positions. Oligonucleotides were end-labeled with bacteriophage T4 kinase (United States Biochemical, Cleveland) and [γ-^{32}P]ATP (>5000 Ci/mmol, 1 Ci = 37 GBq; ICN) and used to screen a human retina cDNA library (provided by J. Nathans, Genentech Corp.). Hybridizations were done in 6× SSCP (1× SSCP = 0.15 M NaCl/15 mM $Na_3C_6H_5O_7 \cdot 2H_2O$/1 mM $Na_2H_2P_7O_7$), 10 μg of polyadenylic acid per ml, 10 μg of bovine liver tRNA per ml, and 5× Denhardt's solution (Denhardt's solution is 0.02% polyvinylpyrrolidone/0.2% Ficoll/0.02% bovine serum albumin) for 10–14 hr at 42°C (16). Filters were washed in 6× SSCP/0.1% $NaDodSO_4$ for 20 min at room temperature and twice for 40 min at 55°C. Between 5 and 10 hybridization-positive clones were obtained per 10,000 plaques screened. Selected clones were plaque-purified, and their cDNA inserts were subcloned and sequenced. DNA sequence analysis was done as described (17) using single-stranded bacteriophage M13 DNA tem-

```
GGGCGCAGGGACCCGAGGCCTCCGCTGCGGCTGGATTCTGCTGCGAACCGGAGACCATGGCCAAACCAGCACAGGGTGCCAAGTACCGGGGCTCCATCCATGACTTCCCAGGCTTT 116
                                                    M  A  K  P  A  Q  G  A  K  Y  R  G  S  I  H  D  F  P  G  F  20

GACCCCAACCAGGATGCCGAGGCTCTGTACACTGCCATGAAGGGCTTTGGCAGTGACAAGGAGGCCATACTGGACATAATCACCTCACGGAGCAACAGGCAGAGGCAGGAGGTCTGC 233
 D  P  N  Q  D  A  E  A  L  Y  T  A  M  K  G  F  G  S  D  K  E  A  I  L  D  I  I  T  S  R  S  N  R  Q  R  Q  E  V  C  59

CAGAGCTACAAGTCCCTCTACGGCAAGGACCTCATTGCTGATTTAAAGTATGAATTGACGGGCAAGTTTGAACGGTTGATTGTGGGCCTGATGAGGCCACCTGCCTATTGTGATGCC 350
 Q  S  Y  K  S  L  Y  G  K  D  L  I  A  D  L  K  Y  E  L  T  G  K  F  E  R  L  I  V  G  L  M  R  P  P  A  Y  C  D  A  98

AAAGAAATTAAAGATGCCATCTCGGGCATTGGCACTGATGAGAAGTGCCTCATTGAGATCTTGGCTTCCCGGACCAATGAGCAGATGCACCAGCTGGTGGCAGCATACAAAGATGCC 467
 K  E  I  K  D  A  I  S  G  I  G  T  D  E  K  C  L  I  E  I  L  A  S  R  T  N  E  Q  M  H  Q  L  V  A  A  Y  K  D  A  137

TACGAGCGGGACCTGGAGGCTGACATCATCGGCGACACCTCTGGCCACTTCCAGAAGATGCTTGTGGTCCTGCTCCAGGGAACCAGGGAGGAGGATGACGTAGTGAGCGAGGACCTG 584
 Y  E  R  D  L  E  A  D  I  I  G  D  T  S  G  H  F  Q  K  M  L  V  V  L  L  Q  G  T  R  E  E  D  D  V  V  S  E  D  L  176

GTACAACAGGATGTCCAGGACCTATACGAGGCAGGGGAACTGAAATGGGGAACAGATGAAGCCCAGTTCATTTACATCTTGGGAAATCGCAGCAAGCAGCATCTTCGGTTGGTGTTC 701
 V  Q  Q  D  V  Q  D  L  Y  E  A  G  E  L  K  W  G  T  D  E  A  Q  F  I  Y  I  L  G  N  R  S  K  Q  H  L  R  L  V  F  215

GATGAGTATCTGAAGACCACAGGGAAGCCGATGAAGGCCAGCATCCGAGGGGAGCTGTCTGGGGACTTTGAGAAGCTAATGCTGGCCGTAGTGAAGTGTATCCGGAGCACCCCGGAA 818
 D  E  Y  L  K  T  T  G  K  P  M  K  A  S  I  R  G  E  L  S  G  D  F  E  K  L  M  L  A  V  V  K  C  I  R  S  T  P  E  254

TATTTTGCTGAAAGGCTCTTCAAGGCTATGAAGGGCCTGGGGACTCGGGACAACACCCTGATCCGCATCATGGTCTCCCGTAGTGAGTTGGACATGCTCGACATTCGGGAGATCTTC 935
 Y  F  A  E  R  L  F  K  A  M  K  G  L  G  T  R  D  N  T  L  I  R  I  M  V  S  R  S  E  L  D  M  L  D  I  R  E  I  F  293
                           K  G  L  G  -  R  D  N  T  L  I  R        V  -  -  S  E  L  D
                                                  P1                          P2

CGGACCAAGTATGAGAAGTCCCTCTACAGCATGATCAAGAATGACACCTCTGGCGAGTACAAGAAGACTCTGCTGAAGCTGTCTGGGGGAGATGATGATGCTGCTGGCCAGTTCTTC 1052
 R  T  K  Y  E  K  S  L  Y  S  M  I  K  N  D  T  S  G  E  Y  K  K  T  L  L  K  L  S  G  G  D  D  D  A  A  G  Q  F  F  332

CCGGAGGCAGCGCAGGTGGCCTATCAGATGTGGGAACTTAGTGCAGTGGCCCGAGTAGAGCTGAAGGGAACTGTGCGCCCAGCCAATGACTTCAACCCTGACGCAGATGCCAAAGCG 1169
 P  E  A  A  Q  V  A  Y  Q  M  W  E  L  S  A  V  A  R  V  E  L  K  G  T  V  R  P  A  N  D  F  N  P  D  A  D  A  K  A  371

CTGCGGAAAGCCATGAAGGGACTCGGGACTGACGAAGACACAATCATCGATATCATCACGCACCGCAGCAATGTCCAGCGGCAGCAGATCCGGCAGACCTTCAAGTCTCACTTTGGC 1286
 L  R  K  A  M  K  G  L  G  T  D  E  D  T  I  I  D  I  I  T  H  R  S  N  V  Q  R  Q  Q  I  R  Q  T  F  K  S  H  F  G  410

CGGGACTTAATGACTGACCTGAAGTCTGAGATCTCTGGAGACCTGGCAAGGCTGATTCTGGGGCTCATGATGCCACCGGCCCATTACGATGCCAAGCAGTTGAAGAAGGCCATGGAG 1043
 R  D  L  M  T  D  L  K  S  E  I  S  G  D  L  A  R  L  I  L  G  L  M  M  P  P  A  H  Y  D  A  K  Q  L  K  K  A  M  E  449
                                                                    -  -  A  -  Y  D  A  K  Q  L  K  K  A
                                                                                          P3

GGAGCCGGCACAGATGAAAAGGCTCTTATTGAAATCCTGGCCACTCGGACCAATGCTGAAATCCGGGCCATCAATGAGGCCTATAAGGAGGACTATCACAAGTCCCTGGAGGATGCT 1520
 G  A  G  T  D  E  K  A  L  I  E  I  L  A  T  R  T  N  A  E  I  R  A  I  N  E  A  Y  K  E  D  Y  H  K  S  L  E  D  A  488

CTGAGCTCAGACACATCTGGCCACTTCAGGAGGATCCTCATTTCTCTGGCCACGGGGCATCGTGAGGAGGGAGGAGAAAACCTGGACCAGGCACGGGAAGATGCCCAGGTGGCTGCT 1637
 L  S  S  D  T  S  G  H  F  R  R  I  L  I  S  L  A  T  G  H  R  E  E  G  G  E  N  L  D  Q  A  R  E  D  A  Q  V  A  A  527

GAGATCTTGGAAATAGCAGACACACCTAGTGGAGACAAAACTTCCTTGGAGACACGTTTCATGACGATCCTGTGTACCCGGACGTATCCGCACCTCCGGAGAGTCTTCCAGGAGTTC 1754
 E  I  L  E  I  A  D  T  P  S  G  D  K  T  S  L  E  T  R  F  M  T  I  L  C  T  R  T  Y  P  H  L  R  R  V  F  Q  E  F  566

ATCAAGATGACCAACTATGACGTGGAGCACACCATCAAGAAGGAGATGTCTGGGGATGTCAGGGATGCATTTGTGGCCATTGTTCAAAGTGTCAAGAACAAGCCTCTCTTCTTTGCC 1871
 I  K  M  T  N  Y  D  V  E  H  T  I  K  K  E  M  S  G  D  V  R  D  A  F  V  A  I  V  Q  S  V  K  N  K  P  L  F  F  A  605

GACAAACTTTACAAATCCATGAAGGGTGCTGGCACAGATGAGAAGACTCTGACCAGGATCATGGTATCCCGCAGTGAGATTGACCTGCTCAACATCCGGAGGGAATTCATTGAGAAA 1988
 D  K  L  Y  K  S  M  K  G  A  G  T  D  E  K  T  L  T  R  I  M  V  S  R  S  E  I  D  L  L  N  I  R  R  E  F  I  E  K  644
                K  G  A  G  T  D  E  K  T  L  T  R  I
                                   P4

TATGACAAGTCTCTCCACCAAGCCATTGAGGGTGACACCTCCGGAGACTTCCTGAAGGCCTTGCTGGCTCTCTGTGGTGGTGAGGACTAGGGCCACAGCTTTGGCGGGCACTTCTGC 2105
 Y  D  K  S  L  H  Q  A  I  E  G  D  T  S  G  D  F  L  K  A  L  L  A  L  C  G  G  E  D  *                              673

CAAGAAATGGTTATCAGCACCAGCCGCCATGGCCAAGCCTGATTGTTCCAGCTCCAGAGACTAAGGAAGGGGCAGGGGTGGGGGGAGGGGTTGGGTTGGGCTCTTATCTTCATGGAG 2222

CTTAGGAAACGCTCCCACTCCCACGGGCCATCGAGGGCCAGCACGGCTGAGCGGTGAAAAACCGTAGCCATAGATCCTGTCCACCTCCACTCCCCTCTGACCCTCAGGCTTTCCCAG 2339

CTTCCTCCCCTTGCTACAGCCTCTGCCCTGGTTTGGCTATGTCAGATCCAATAAACATCCTGAACCTCTGTCTGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2447
```

FIG. 1. Nucleotide and deduced amino acid sequence of the 67-kDa calelectrin cDNA. Sequences of peptides from bovine 67-kDa calelectrin are shown below the deduced amino acid sequence and numbered P1–P4. Residue numbers are shown at the right. Nucleotide sequence was determined using M13 subclones of both strands of clones λCE1 and λCE9. The deduced amino acid sequence is given in the single-letter code.

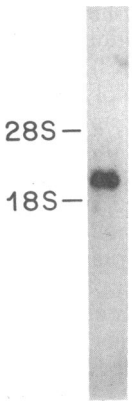

FIG. 2. RNA blot analysis of the 67-kDa calelectrin message using total RNA from transformed human fibroblasts. Positions of 18S and 28S RNA markers are shown at left. The blot was probed with a uniformly ^{32}P-labeled single-stranded probe corresponding to nucleotides 1700–1938 in Fig. 1 and exposed with an intensifying screen for 1 day at −70°C.

plates (18) and specific oligonucleotide primers (19). Nucleotide and protein sequences were analyzed with the Beckman Microgenie program on an IBM PC-AT.

For RNA blot analysis, total RNA from an SV40-transformed human fibroblast cell line (SV589) was electrophoresed and blotted as described (20). The filter was hybridized with a uniformly ^{32}P-labeled single-stranded probe (21) and washed at high stringency using standard procedures.

RESULTS

Two peptide sequences obtained from CNBr fragments of the 67-kDa calelectrin (P1 and P4 in Fig. 1) were similar to each other and to a conserved sequence in human lipocortin I and II (6, 7). An oligonucleotide complementary to this sequence was synthesized to screen a human retina cDNA library. Thirty hybridization-positive plaques were identified, and nine were selected for restriction enzyme analysis and preliminary DNA sequencing. Eight clones encoded the 67-kDa calelectrin, and one encoded lipocortin I. The two largest 67-kDa calelectrin clones, λCE1 and λCE9, were selected for further analysis. These clones had identical restriction enzyme maps and, where analyzed, identical DNA sequences except at their 5′ ends. One clone (λCE1) had an inverted repeat at its 5′ end not present in λCE9, and this was presumed to be a cDNA cloning artifact.

The composite DNA sequence derived from the cDNA inserts of λCE1 and λCE9 and the deduced amino acid sequence are shown in Fig. 1. The nucleic acid sequence translates into a 674-residue protein with a predicted M_r of 75,901. This molecular weight is higher than the value of 67,000 determined by the mobility of the human and bovine proteins on NaDodSO$_4$/polyacrylamide gels (2). Locations

```
                (1)  GS.IHDFPGFDPNQDAEA.LYTA...MKGFGSD.KEAILDIITSRSNRQRQEVCQSYKSLYGKDLIADLKYELTGKFERLIVGLMRPPAY   95
CALELECTRIN
                (5)  GT.VRPANDFNPDADAKA.LRKA...MKGLGTD.EDTIIDIITHRSNVQRQQIRQTFKSHFGRDLMTDLKSEISGDLARLILGLMMPPAH  438

          I     (1)  GSAVSPYPTFNPSSDVAA.LHKA...IMVKGVD.EATIIDILTKRNNAQRQQIKAAYLQETGKPLDETLKKALTGHLEEVVLALLKTPAQ  117
LIPOCORTIN
          II    (1)  GS.VKAYTNFDAERDALN.IETA...IKTKGVD.EVTIVNILTNRSNAQRQDIAFAYQRRTKKELASALKSALSGHLETVILGLLKTPAQ  108

                (2)            CDAKE.IKDA...ISGIGTD.EKCLIEILASRTNEQMHQLVAAYKDAYERDLEADIIGDTSGHFQKMLVVLLQGTRE  167
CALELECTRIN
                (6)            YDAKQ.LKKA...MEGAGTD.EKALIEILATRTNAEIRAINEAYKEDYHKSLEDALSSDTSGHFRRILISLATGHRE  510

          I     (2)            FDADE.LRAA...MKGLGTD.EDTLIEILASRTNKEIRDINRVYREELKRDLAKDITSDTSGDFRNALLSLAKGDRS  189
LIPOCORTIN
          II    (2)            YDASE.LKAS...MKGLGTD.EDSLIEIICSRTNQELQEINRVYKEMYKTDLEKDIISDTSGDFRKLMVALAKGRRA  180

                (3)  EDDVVS.EDLVQQDVQD.LYEAGELKWG..TD.EAQFIYILGNRSKQHLRLVFDEYLKTTGKPMKASIRGELSGDFEKLMLAVVKCIRS  251
CALELECTRIN
                (7)  EGGENLDQAREDAQVAAEILEIADTPSGDKTSLETRFMTILCTRTYPHLRRVFQEFIKMTNYDVEHTIKKEMSGDVRDAFVAIVQSVKN  599

          I     (3)  EDFGV.NEDLADSDARA.LYEAGERRKG..TD.VNVFNTILTTRSYPQLRRVFQKYTKYSKHDMNKVLDLELKGDIEKCLTAIVKCATS  273
LIPOCORTIN
          II    (3)  EDGSVIDYELIDQDARD.LYDAGVKRKG..TD.VPKWISIMTERSVPHLQKVFDRYKSYSPYDMLESIRKEVKGDLENAFLNLVQCIQN  264

                (4)          TPEYFAER.LFKA...MKGLGTR.DNTLIRIMVSRSELDMLDIREIFRTKYEKSLYSMIKNDTSGEYKKTLLKLSGGDDD  326
CALELECTRIN
                (8)          KPLFFADK.LYKS...MKGAGTD.EKTLTRIMVSRSEIDLLNIRREFIEKYDKSLHQAIEGDTSGDFLKALLALCGGED*  673

          I     (4)          KPAFFAEK.LHQA...MKGVGTR.HKALIRIMVSRSEIDMNDIKAFYQKMYGISLCQAILDETKGDYEKILVALCGGN*  346
LIPOCORTIN
          II    (4)          KPLYFADR.LYDS...MKGKGTR.DKVLIRIMVSRSEVDMLKIRSEFKRKYGKSLYYYIQQDTKGDYQKALLYLCGGDD*  339

CONSENSUS                  DAxx.LxxA...MKGxGTD.ExxLIxILxSRSNxxxxxIxxx(Y/F)xxxxxKDLxxxIxx(E/D)TSGDxxKxLLxLxxG
```

FIG. 3. Alignment of the 67-kDa calelectrin and the lipocortin I and II amino acid sequences. Within each of the four sequence blocks, the upper two lines depict the 67-kDa calelectrin sequences derived from the first and second half of the protein molecule. The lower two lines in each block contain lipocortin I and II sequences, respectively. Amino acids shared by two or more sequences within each block are shaded. Numbers of corresponding internal repeats for the sequences are shown in parentheses on the left, and residue numbers are at the right (from Fig. 1 and ref. 7). At bottom, a consensus sequence has been derived for all 16 repeats. This consensus represents amino acids at a given position present in at least half of the sequences. Sequences are shown in the single-letter amino acid code.

and sequences of the four CNBr peptides (P1–P4) derived from the bovine 67-kDa calelectrin are shown below the deduced amino acid sequence of the human protein (Fig. 1). Surprisingly, the sequences of these four peptides are identical in humans and cows. Blot analysis of total RNA from human fibroblasts using a radiolabeled probe derived from λCE1 suggested the presence of a single mRNA of about 2.5 kb (Fig. 2). No cross-hybridizing mRNAs were detected at the high stringency used in these experiments.

Analysis of the predicted amino acid sequence of the 67-kDa calelectrin demonstrates eight imperfect amino acid repeats joined by linker sequences of variable lengths. The linker separating the fourth and fifth repeats is the longest linker, containing 44 amino acids. The sequence is maximally similar when the first four repeats are aligned with the last four repeats, which indicates that the 67-kDa calelectrin consists of two sets of four repeats. Thus, repeat 1 resembles repeat 5 more closely (48%) than repeat 1 resembles the other six repeats; repeat 2 is most similar to repeat 6, etc. (Fig. 3).

Fig. 3 shows that each of the two sets of four repeats can also be aligned with the lipocortin I and lipocortin II sequences, revealing residue identities ranging from 42% to 45%. This alignment excludes the residue at the amino termini of the three proteins and the linker between the fourth and fifth repeats of the 67-kDa calelectrin. These excluded sequences are distinctive for each protein.

Sequences of the two halves of the 67-kDa calelectrin and of lipocortin I and II are aligned in Fig. 3 in a format that also allows comparison of their 16 internal repeats. A consensus repeat sequence of amino acids at a given position in at least one half of the repeats is shown below the alignment. A few residues are conserved in all 16 repeats as, for example, the arginine and isoleucine residues in the center of the consensus sequence. For many positions to which no consensus amino acid can be assigned, residues with similar side-chain hydrophobicities are found. Interestingly, at positions occupied by these hydrophilic residues, little charge conservation of a given side chain exists; one exception is the aforementioned conserved central-arginine residue. Such findings suggest that the repeats fold into a similar tertiary structure but that the surface charge at a given position is not essential for its function, except in the case of the conserved arginine residue.

Mean hydrophobicity of amino acid residues at each position in the 16 calelectrin/lipoctortin repeats is plotted in Fig. 4 over the repeat consensus sequence. Hydrophobic residues are revealed to be regularly spaced at 4- or 5-amino acid intervals over most repeats. Clusters of four to six hydrophobic residues are found at the end of the first third of a repeat and again at the carboxyl terminus. These hydrophobic clusters are indicated by brackets above the hydrophobicity profile in Fig. 4. In each case, these hydrophobic clusters are flanked by very hydrophilic residues. The position of the central conserved arginine is marked by an arrow in Fig. 4. Although secondary structure predictions did not lead to an unequivocal model of the folding of the repeats (data not shown), the repeats clearly must form structures that contain both very hydrophilic and very hydrophobic faces.

DISCUSSION

The sequence of human 67-kDa calelectrin presented here reveals that the protein consists of eight similar repeats separated by linkers of variable length. A 24-amino acid sequence precedes the first repeat, and an unrelated 44-amino acid sequence separates the fourth from the fifth repeat. Although all repeats are similar to each other, maximal similarities are achieved by aligning the first four repeats with the last four repeats, including linking se-

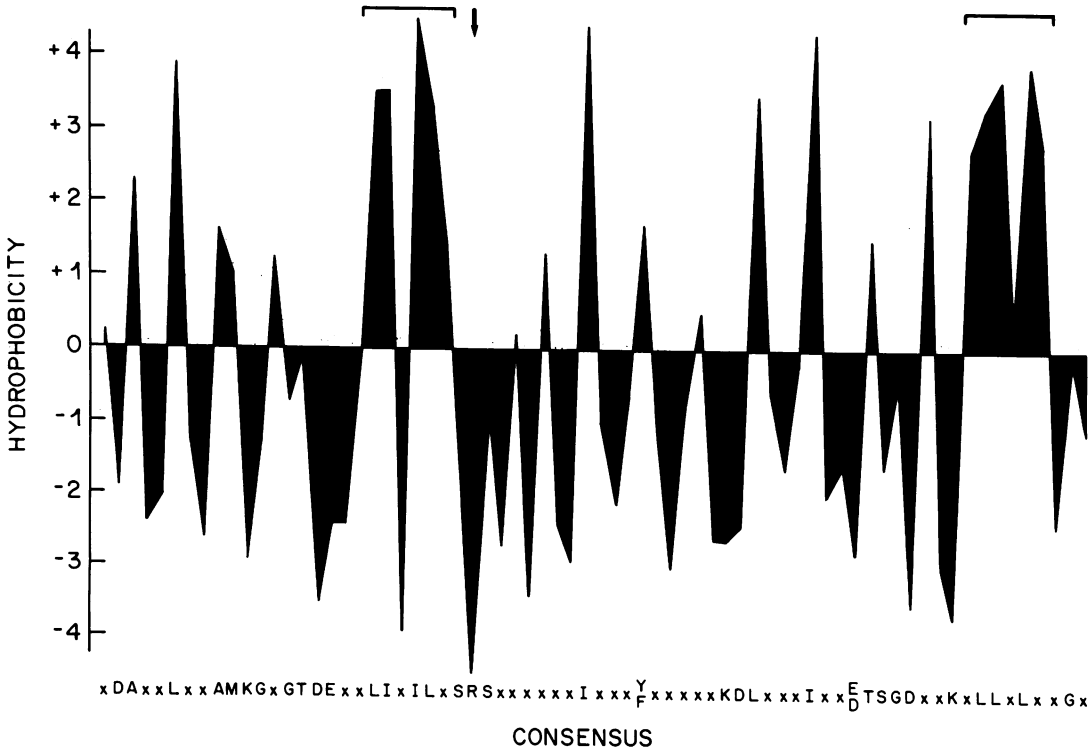

FIG. 4. Distribution of amino acid hydrophobicity across the calelectrin/lipocortin repeats. Calelectrin/lipocortin repeats were aligned as in Fig. 3, and mean hydrophobicity of amino acid residues at a given position was calculated for the 16 repeats, using the hydropathy index of Kyte and Doolittle (22). The ordinate gives mean hydrophobicity, while the consensus amino acid sequence from Fig. 3 is aligned on the abscissa. The two clusters of hydrophobic amino acids are marked by brackets above the profile, and the position of the conserved central arginine residue (R) is marked by an arrow.

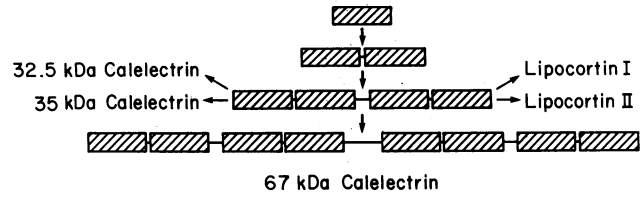

FIG. 5. Proposed evolutionary scheme for the calelectrins/lipocortins. A primordial calelectrin repeat is hypothesized to have undergone two gene duplications to result in a four-repeat precursor. From this precursor all calelectrins in the 32.5- to 35-kDa range and lipocortins I and II have evolved. The precursor duplicated to form 67-kDa calelectrin.

quences. When thus aligned, 48% identity between the two halves of 67-kDa calelectrin is seen.

Although earlier studies had suggested that calelectrins and lipocortins belong to the same protein family, the high degree of similarity between lipocortin I, lipocortin II, and the 67-kDa calelectrin is striking: pairwise comparisons among the three proteins reveal 42–53% identity (Fig. 3). The similarity extends uniformly over the whole protein sequences, except for the amino termini and the middle linker of the 67-kDa calelectrin, which have no common sequences. Again, the similarity between the directly aligned proteins is greater than the similarity between their internal repeats.

This pattern strongly suggests that they evolved from a common evolutionary precursor containing four repeats, which in turn was derived by two gene duplications from a single-repeat protein (Fig. 5). Interestingly, all other calelectrins purified by us have apparent molecular masses on $NaDodSO_4$/polyacrylamide gels of 30–35 kDa (2–4), suggesting that these proteins contain only four repeats. The proposed evolutionary scheme is supported by pairwise comparisons among the repeats, which show that repeats 1 and 3 and repeats 2 and 4 within each set are more similar to each other than the opposite pairings. Duplication of a functional Ca^{2+}-binding module to create proteins with a different number of homologous domains has also been seen for the family of EF-hand-containing Ca^{2+}-binding proteins. In this family there are found two-domain (parvalbumin) and four-domain (calmodulin) members (1).

Although function of the calelectrins/lipocortins is unresolved, there is general agreement that all family members bind Ca^{2+} and bind to phospholipids in a Ca^{2+}-dependent manner (2–4, 13, 23, 24). The localization of this property to the homologous internal repeats of these proteins is a plausible hypothesis. Divergence in the N-terminal regions of the calelectrins/lipocortins may relate to their specific cellular roles. This second hypothesis is suggested by the localization of the phosphorylation sites of the lipocortins and the binding site for the small lipocortin II subunit to these particular sequences (7–9, 25, 26). All calelectrins/lipocortins appear to be abundant proteins differentially expressed in specific cell types (23, 27). Each tissue may have its own pattern of calelectrin/lipocortin expression, with the divergent N-terminal sequences assigning the specific roles of a family member in a given tissue.

Comparison of the 16 human calelectrin/lipocortin repeats reveals a pattern of divergence and structural conservation. No classical EF-hand Ca^{2+}-binding site is present. Furthermore, no acidic-amino acid residue is conserved in all repeats, although the residue hydrophobicity is generally well conserved at a given position. Unlike calmodulin, the calelectrins do not expose a hydrophobic site as a function of Ca^{2+} and appear to preferentially bind to acidic phospholipids (14, 23, 24). The consensus structure of the repeat suggests that the central arginine residue is involved in phospholipid binding, whereas the hydrophobic residues may form a hydrophobic core in the tertiary structure of the repeats.

Future studies on calelectrins must clarify two major questions. (*i*) What is calelectrin function? Studies using the most abundant and widely expressed calelectrin, the 67-kDa calelectrin, may resolve this question. (*ii*) What is the structural basis for the property of calelectrins to bind to lipids in a Ca^{2+}-dependent fashion? This question could be studied on a single-repeat unit and may reveal a new principle in Ca^{2+}-regulation of protein structure (28).

We thank Drs. David W. Russell, Joseph L. Goldstein, and Michael S. Brown for their advice, support, and critical reviews of the manuscript and Dr. Stephen Sprang for secondary structure-prediction analysis. Kim Orth and Carolyn Moomaw provided excellent technical assistance, and we thank Teri Baker for excellent secretarial work.

1. Kretsinger, R. H. (1980) *CRC Crit. Rev. Biochem.* **8**, 119–174.
2. Südhof, T. C., Ebbecke, M., Walker, J. H., Fritsche, U. & Boustead, C. (1984) *Biochemistry* **23**, 1103–1109.
3. Südhof, T. C. (1984) *Biochem. Biophys. Res. Commun.* **123**, 100–107.
4. Südhof, T. C., Walker, J. H. & Fritsche, U. (1985) *J. Neurochem.* **44**, 1302–1307.
5. Geisow, M. J., Fritsche, U., Hexhau, J. M., Dash, B. & Johnson, T. (1986) *Nature (London)* **320**, 636–638.
6. Wallner, B. P., Mattaliano, R. J., Hession, C., Cate, R. L., Tizard, R., Sinclair, L. K., Foeller, C., Chow, E. P., Browning, J. L., Ramachandran, K. L. & Pepinsky, R. B. (1986) *Nature (London)* **320**, 77–80.
7. Huang, K., Wallner, B. P., Mattaliano, R. J., Tizard, R., Burne, C., Frey, A., Hession, C., McGray, P., Sinclair, L. K., Pingchang Chow, E., Browning, J. L., Ramachandran, K. L., Tang, J., Smart, J. E. & Blake Pepinsky, R. (1986) *Cell* **46**, 191–199.
8. De, B. D., Misono, K. S., Lukas, T. J., Mroczkowski, B. & Cohen, S. (1986) *J. Biol. Chem.* **261**, 13784–13792.
9. Saris, C. J. M., Tack, B. F., Kristensen, T., Glenney, J. R., Jr., & Hunter, T. (1986) *Cell* **46**, 201–212.
10. Owens, R. J., Gallagher, C. J. & Crumpton, M. J. (1984) *EMBO J.* **3**, 945–952.
11. Creutz, C. E., Zaks, W. J., Hammar, H. C., Crane, S., Martin, W. H., Gould, K. L., Oddie, K. M. & Parsons, S. J. (1987) *J. Biol. Chem.* **262**, 1860–1868.
12. Moore, P. M. (1986) *Biochem. J.* **238**, 49–54.
13. Davidson, F. F., Dennis, E. A., Powell, M. & Glenney, J. R. (1987) *J. Biol. Chem.* **262**, 1698–1705.
14. Shadle, P. J., Gehrke, V. & Weber, K. (1985) *J. Biol. Chem.* **260**, 16354–16360.
15. Südhof, T. C. & Stone, D. K. (1987) *Methods Enzymol.* **139**, 30–35.
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
18. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
19. Südhof, T. C., Goldstein, J. L., Brown, M. S. & Russell, D. W. (1985) *Science* **228**, 815–822.
20. Lehrman, M. A., Russell, D. W., Goldstein, J. L. & Brown, M. S. (1987) *J. Biol. Chem.* **262**, 3354–3361.
21. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
22. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
23. Südhof, T. C. (1984) *Trans. Biochem. Soc.* **12**, 972–974.
24. Südhof, T. C., Walker, J. H. & Obrocki, J. (1982) *EMBO J.* **1**, 1167–1170.
25. Gerke, V. & Weber, K. (1985) *J. Biol. Chem.* **260**, 1688–1695.
26. Glenney, J. R., Boudreau, M., Galyean, R., Hunter, T. & Tack, B. (1986) *J. Biol. Chem.* **261**, 10485–10488.
27. Silva, F. G., Sherrill, K., Spurgeon, S., Südhof, T. C. & Stone, D. K. (1986) *Differentiation* **33**, 175–183.
28. Kretsinger, R. H. & Creutz, C. E. (1986) *Nature (London)* **320**, 573.