Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 9832–9836, October 1994
Cell Biology

Cloning and sequencing of parafusin, a calcium-dependent exocytosis-related phosphoglycoprotein

(*Paramecium*/sequence homology/Southern blot analysis/PCR)

SUKANYA V. SUBRAMANIAN, ELZBIETA WYROBA, ANDERS P. ANDERSEN, AND BIRGIT H. SATIR*

Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461

Communicated by Berta V. Scharrer, May 27, 1994

ABSTRACT **A cDNA for parafusin, an evolutionarily conserved phosphoglycoprotein involved in exocytosis, has been cloned and sequenced from a unicellular eukaryote, *Paramecium tetraurelia*. A *Paramecium* cDNA library was screened with an oligonucleotide probe synthesized to an internal amino acid sequence of isolated parafusin. The insert was 3 kb long with an open reading frame of 1.75 kb. Data base searches of the deduced amino acid sequence showed that *Paramecium* parafusin had a 50.7% sequence identity to rabbit muscle phosphoglucomutase, although no detectable phosphoglucomutase activity has been detected in isolated parafusin. The deduced parafusin amino acid sequence had four inserts and two deletions, which might confer on the protein specific functions in signal transduction events related to exocytosis. Furthermore, searches for potential phosphorylation sites showed the presence of a protein kinase C site (KDFSFR) specific to parafusin. Southern blot analysis with probes specific for parafusin and phosphoglucomutase suggested that these proteins were products of different genes. We propose that parafusin and phosphoglucomutase are members of a superfamily that conserve homologies important for the tertiary structure of the molecules.**

Previously we discovered a cytosolic phosphoprotein, parafusin, that plays a role in regulated exocytosis in the unicellular eukaryote *Paramecium* (1, 2) and that is evolutionarily conserved (3). Parafusin has been shown to be phosphorylated via a Ca^{2+}-dependent protein kinase (4). Surprisingly, parafusin is also a phosphoglycoprotein in which a short chain of mannose residues is O-linked to serine. This chain is phosphoglucosylated by a glucose-1-phosphate phosphotransferase that uses UDP glucose (5). We have recently demonstrated that dephosphoglucosylation is catalyzed by a Ca^{2+}-activated phosphodiesterase. Cells in which parafusin is normal but that are unable to release the content of their dense core secretory vesicles upon stimulation show inactive phosphodiesterase, suggesting that dephosphoglucosylation is a critical event in the pathway to exocytosis (4).

Tryptic digests of parafusin purified as described earlier (6) have been obtained and oligonucleotide probes were made to relevant polypeptide sequences. We now report the cloning and sequencing of the cDNA for *Paramecium* parafusin.[†] Although the deduced amino acid sequence had 50.7% identity to rabbit muscle phosphoglucomutase (PGM), Southern hybridization analysis suggests that parafusin and PGM in *Paramecium* are products of different genes.

MATERIALS AND METHODS

Screening of *Paramecium* cDNA Library. A degenerate 63-mer oligonucleotide antisense probe was synthesized using *Paramecium* codon usage (7) according to the amino acid sequence NTDHFVTVEEIVTQYWQQFGR, obtained after tryptic digestion of purified parafusin. The oligonucleotide probe was end labeled with [τ-^{32}P]ATP using T4 polynucleotide kinase (Promega), and a λgt10 *Paramecium* cDNA library (courtesy of Helmut Schmidt, Westfälische Wilhelms Universität, Münster) was screened (≈10^5 plaques). Agarose gel electrophoresis of the *Eco*RI digest of the purified λ DNA from two positive clones identified two fragments of 1.6 and 1.4 kb. The 1.4- and 1.6-kb inserts were purified and subcloned into pGEM plasmids.

PCR. PCRs were performed with 0.25 μg of total *Paramecium* DNA for 35 cycles with 50–100 ng of primers (listed below), which were internal sequences of parafusin cDNA. Primers: 17, 5′-ACCAGGTATGGTCGTCTAA-3′; 21, 5′-CTAAGTTGCCTCACTCACC-3′; 36, 5′-TTGTTC-CTCCTCCCACTACGATTAGGCCACAACCTATGGC-GAATAGAGGCACCAAGAGTTTAAGTA-3′; 14, 5′-AGT-TATGACAGTTGGTTCA-3′; 15, 5′-GGAGCTGCATGTGA-TGGAG-3′; 12, 5′-CATGTAATAAATCACCAAC-3′. The PCR products were separated by 1.2% agarose gel electrophoresis.

Isolation of Total DNA and Southern Hybridization. The *P. tetraurelia* cell pellet was lysed by heating at 65°C in 1% SDS/0.5 M EDTA/1 M Tris·HCl, pH 9.5. Proteinase K (Boehringer Mannheim) was added to a final concentration of 100 μg/ml for 30 min and the DNA was isolated by a procedure slightly modified from ref. 8. The isolated DNA was digested with restriction endonucleases *Bam*HI, *Eco*RI, or *Hin*dIII (Promega) (1.8–2.9 units per μg of DNA) and electrophoresed (9 μg per lane) on 1% agarose, denatured, blotted to Zeta-Probe genomic tested (GT) blotting membranes (Bio-Rad).

(*i*) Southern hybridization was performed with the τ-^{32}P-end-labeled 63-mer oligonucleotide probe at 55°C to 5–10 μg of *Eco*RI-digested λ or plasmid DNA, run on 1% agarose gel, and blotted onto GeneScreen membranes under denaturing conditions. (*ii*) Southern hybridization to total *Paramecium* DNA was performed with oligonucleotides 3′-end-labeled with digoxygenin-ddUTP (Boehringer Mannheim) using terminal deoxynucleotidyltransferase (Promega) and purified on Bio-Spin 6 columns (Bio-Rad) using a nonradioactive chemiluminescent technique (9). Blots were hybridized overnight at 45°C with the appropriate probe. Signal detection was achieved using Lumi-Phos 540 (Boehringer Mannheim).

Sequencing. The plasmid DNA was purified with the Magic Plasmid Miniprep (Promega) to serve as template for sequencing with the Sequenase kit (United States Biochemical). Dideoxynucleotide chain termination (10) was used for sequencing in both directions with multiple primers.

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.

Abbreviations: PGM, phosphoglucomutase; pPFUS, *Paramecium* parafusin.
*To whom reprint requests should be addressed.
†The sequence reported in this paper has been deposited in the GenBank data base (accession no. L12471).

Computer Analysis. Analysis of Wisconsin Package 7.3.1 on UNIX, the nucleotide and amino acid sequence was done using the Genetics Computer Group, Inc. in Scientific Computing Facility at Einstein College of Medicine. The FASTA program was used in searches for homology in the GenBank/EMBL (for nucleotides) and Swiss-Prot (for proteins) data banks. The GAP program was used for alignment of homologous sequences.

RESULTS AND DISCUSSION

Parafusin cDNA Isolation. A λgt10 *Paramecium* cDNA library was screened with an antisense oligonucleotide (63-mer) probe. This probe was synthesized complementary to a polypeptide obtained from purified parafusin that had no significant homology to any other protein as revealed by searches through the Protein Data Bank. Upon *Eco*RI digestion of the phage DNA, all positive clones exhibited two fragments of 1.6 and 1.4 kb (Fig. 1*A*). Both the 1.6- and 1.4-kb inserts were subcloned into the plasmid pGEM7Z (Fig. 1*C*). Southern blots probed with the ^{32}P-labeled 63-mer oligonucleotide showed hybridization with the 1.6-kb fragment and the uncut 3-kb insert (Fig. 1 *B* and *D*).

Sequence of Parafusin cDNA. The cDNA and the deduced amino acid sequences of parafusin are shown in Fig. 2*A*. The entire cDNA insert consists of 3088 bp, 62% A + T residues, with an initiation codon at nucleotide 913. The initiation codon is GTA, which normally codes for valine, but at the initiating position it codes for methionine. Although not common in higher eukaryotes this occurs frequently in prokaryotes and has been reported earlier for *Paramecium* mitochondrial genes (11, 12). Just preceding the initiation codon is an in-frame termination codon. Confirmation of this initiation codon comes from the result obtained after tryptic digestion of isolated parafusin. One of the peptides obtained starts at the 22nd amino acid in the predicted protein sequence and there is no methionine between this and the initiation codon.

The open reading frame is 1749 bp long and stops at nucleotide 2661 with TGA as the stop codon. The translated sequence from the initiation codon predicts a polypeptide of 583 amino acids with M_r 65,000, in good agreement with the purified protein (M_r 63,000). Two-dimensional gel electrophoresis has demonstrated the presence of three isoforms of parafusin with pI values between 5.8 and 6.3 (6). In contrast, the value from the cDNA deduced amino acid sequence indicates a pI of 6.5. This discrepancy could be due to different states of phosphorylation and/or oxidation (13) of the protein. The cDNA deduced amino acid sequence of parafusin predicts it to be a highly charged molecule. A hydropathy plot of parafusin shows the protein to be largely hydrophilic, with a prominent hydrophobic segment located at the N terminus (residues 1–10) that may act as a membrane association sequence.

The 5′ untranslated region consists of 900 bp, the first 600 of which show 75% homology with elongation factor α corresponding to bases 1350–1940, which is thought to be involved in autoregulation (14). The significance of this homology is yet to be determined. The 3′ untranslated region consists of 427 bp and most likely has a polyadenylylation signal (AATAAA) at position 2679. Fig. 2*B* shows a restriction map of the parafusin cDNA. Restriction enzymes *Bam*HI, *Eco*RI, and *Hin*dIII cut the cDNA insert at 704, 1459, and 2593 nucleotides, respectively.

To establish that the cDNA clone contained the coding sequences specific for parafusin, we have compared the deduced cDNA amino acid sequence to nine different tryptic peptides obtained from digestion of the purified parafusin. All sequences matched the deduced amino acid sequences (Fig. 2*A*, single underlined letters) and are preceded by a basic amino acid residue consistent with the specificity of trypsin digestion. PCR analysis revealed that amplification of total *Paramecium* DNA with internal primers (Fig. 3) gave amplified products of ≈1740 kb (lane 2), ≈1060 kb (lane 3), and ≈960 kb (lane 4), the same size as expected from the cDNA sequence (Fig. 2*B*). This may indicate that there are no introns in the parafusin gene.

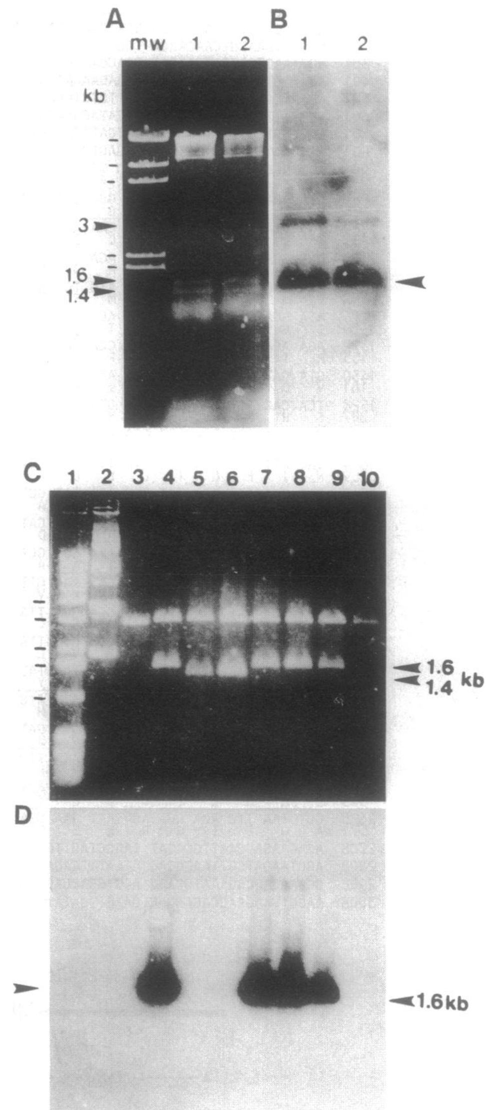

FIG. 1. Isolation and subcloning of the parafusin cDNA. (*A*) Agarose gel of the *Eco*RI digest of purified phage DNA from two positive clones showing two fragments of 1.6 and 1.4 kb and a faint 3-kb fragment (lanes 1 and 2, arrowheads). Lane mw, molecular size marker λ *Hin*dIII (top to bottom: 23.1, 9.4, 6.5, 2.3, and 2 kb, respectively). (*B*) Southern hybridization of the *Eco*RI digest (*A*) with the ^{32}P-end-labeled 63-mer oligonucleotide probe. The probe hybridizes with the 1.6-kb insert as well as with the 3-kb fragment (lanes 1 and 2). (*C*) Agarose gel of plasmids with parafusin cDNA inserts. The 1.6- and 1.4-kb inserts were purified and subcloned into pGEM7Z plasmids. Lane 1, molecular size markers 4, 3, 2, 1.6, and 1 kb from the top, respectively; lane 2, uncut control pGEM plasmid; lane 3, *Eco*RI-digested control pGEM plasmid; lanes 4–9, *Eco*RI digests of plasmid DNA from positive colonies; lane 10, *Eco*RI digest of plasmid DNA from negative control. Arrowheads show the 1.6- and 1.4-kb inserts. (*D*) Southern blot of the gel in *C* with the ^{32}P-end-labeled 63-mer oligonucleotide probe. The probe hybridized only to the 1.6-kb insert (arrowhead).

Parafusin Amino Acid Sequence Analysis. Searches for homology to parafusin revealed 50.7% identity to rabbit muscle PGM in a 578-amino acid overlap scattered through-

```
A    1 ATTCCGAAACCACTCAGATT ACCATTACAAGACGTCTACA AGATCGGAGGTATTGGTACA GTCCCAGTCGGAAGAGTTGA AACTGGTGTCTTGAAACCAG  100
   101 GTATGGTCGTCTAATTCGCC CCATCAGCCATCACAACTGA AGTCAAGTCAGTTGAAATGC ATCATGAAGCCCTCCCAGAA GCCGTCCCAGGAGATAACGT  200
   201 CGGATTCAACGTCAAGAACA TCGCCGTCAAAGATTTGAAG AGAGGATTCGTCTGCTCAGA CTCAAAGAGTGACCCAGCAA GAGAATGCCAATCATTCAAC  300
   301 GCTTAAGTCATCATCATCAT CAACCATCCAGGATAAATCC AAAACGGATATTGCCCAGTA TTGGATTGCCACACAGCCCA CATTGCTTGCAAATTCTAAG  400
   401 AAATTCTTTCAAAGAATGAC AGAAGAACAGGCAAAGTCAT TGAAGAAGAACCAAAATTCG TCAAATCAGGAGATGCTGCT ATGGTCAAATTAATCCCAAC  500
   501 CAAACCAATGTGCGTTGAAA TCTTCTCTGAATATCCCCCA CTCGGTAGATTCGCCGTCAG AGATATGAAATAAACCGTTG CCGTCGGTGTCATCAAGGTT  600
   601 GTTTTTTTATTTAATTTGTA CACATAGGAATATTAAATTA TTGTGGATATAATCATTTCA GTAACAACTCTTTGGCTTCA TTGATTTTAGTGGCCACATA  700
   701 TGTGGATCCTCCTTGATCTG GATGATTGAGCATCAACAAT GTGCGATGTCGAGTCTTGAT TTGGTCTAGCGGGGTGCCTT CTCTGACTCCGAGAATCAAT  800
   801 TGGGCTTCTCTCCTGGTGAG TTGAGTGGAGAAGGTTCCCT TGTAGAAGGCCCTTGAGGTT TCCACAGTCTTGAAAAACTC CTTGCTCTTGATTTGTCTGT  900
   901 AGGTGCGAATGA GTA GTT TTG TTC CTC CTC CCA CTA CGA TTA GGC CAC AAC CTA TGG CGA ATA GAG GCA CCA AGA GTT TAA   981
     1           *  V   V   L   F   L   L   P   L   R   L   G   H   N   L   W   R   I   E   A   P   R   V   Q     24
   982 GTA ACA TAA CCA TAT GCA GGG CAA AAG CCA GGT ACT TCA GGG TTA CGT AAA AAG GTG AGT GAG GCA ACT TAG CCT AAT  1059
    25 V   T   Q   P   Y   A   G   Q   K   P   G   T   S   G   L   R   K   K   V   S   E   A   T   Q   P   N     50
  1060 TAT TTA GAG AAC TTT GTG CAA AGC ATT TTC AAC ACA TTA CGC AAG GAT GAA TTG AAG AAT GTT CTC TTT GTT GGT GGA  1037
    51 Y   L   E   N   F   V   Q   S   I   F   N   T   L   R   K   D   E   L   K   N   V   L   F   V   G   G     76
  1138 GAT GGA AGG TAC TTT AAT AGA TAG GCT ATT TTC TCA ATA ATT CGT TTG GCT TAT GCA AAT GAT ATA AGC GAG GTA CAT  1215
    77 D   G   R   Y   F   N   R   Q   A   I   F   S   I   I   R   L   A   Y   A   N   D   I   S   E   V   H    102
  1216 GTT GGC TAA GCA GGT TTA ATG TCT ACT CCT GCC TCT TCA CAC TAT ATT CGT AAG GTT AAT GAA GAA GTT GGA AAT TGC  1293
   103 V   G   Q   A   G   L   M   S   T   P   A   S   S   H   Y   I   R   K   V   N   E   E   V   G   N   C    128
  1294 ATA GGA GGA ATA ATC CTT ACT GCT TCT CAT AAT CCA GGT GGC AAA GAA CAT GGA GAT TTT GGA ATC AAG TTC AAT GTA  1371
   129 I   G   G   I   I   L   T   A   S   H   N   P   G   G   K   E   H   G   D   F   G   I   K   F   N   V    154
  1372 AGA ACA GGT GCT CCT GCT CCA GAA GAC TTC ACT GAT TAA ATC TAC ACA CAT ACG ACT AAA ATT AAG GAA TAT TTA ACA  1449
   155 R   T   G   A   P   A   P   E   D   F   T   D   Q   I   Y   T   H   T   T   K   I   K   E   Y   L   T    180
  1450 GTA GAT TAT GAA TTC GAG AAA CAC ATA AAT TTA GAC TAA ATA GGA GTA TAC AAA TTC GAA GGC ACT AGA TTA GAA AAA  1527
   181 V   D   Y   E   F   E   K   H   I   N   L   D   Q   I   G   V   Y   K   F   E   G   T   R   L   E   K    206
  1528 TCA CAC TTT GAA GTG AAG GTA GTA GAC ACA GTT TAG GAC TAT ACT CAA TTG ATG TAG AAA CTT TTT GAT TTC GAT TTG  1605
   207 S   H   F   E   V   K   V   V   D   T   V   Q   D   Y   T   Q   L   M   Q   K   L   F   D   F   D   L    232
  1606 CTT AAA GGC TTA TTC TCA AAT AAA GAT TTT TCA TTT AGA TTT GAT GGT ATG CAT GGT GTT GCA GGA CCT TAT GCT AAA  1683
   233 L   K   G   L   F   S   N   K   D   F   S   F   R   F   D   G   M   H   G   V   A   G   P   Y   A   K    258
  1684 CAC ATT TTT GGC ACC TTA TTA GGA TGT TCA AAG GAG TCT CTA TTG AAT TGT GAT CCA TCT GAA GAT TTT GGA GGC GGT  1761
   259 H   I   F   G   T   L   L   G   C   S   K   E   S   L   L   N   C   D   P   S   E   D   F   G   G   G    284
  1762 CAT CCA GAT CCA AAT TTG ACT TAT GCT CAT GAT CTA GTT GAA TTA TTG GAT ATT CAT AAG AAA AAG GAT GTA GGT ACA  1839
   285 H   P   D   P   N   L   T   Y   A   H   D   L   V   E   L   L   D   I   H   K   K   K   D   V   G   T    310
  1840 GTG CCA TAA TTT GGA GCT GCA TGT GAT GGA GAT GCC GAT AGA AAT ATG ATC TTA GGT AGA TAA TTC TTT GTA ACA CCT  1917
   311 V   P   Q   F   G   A   A   C   D   G   D   A   D   R   N   M   I   L   G   R   Q   F   F   V   T   P    336
  1918 TCT GAT TCC TTG GCA GTT ATA GCT GCT AAT GCT AAT TTA ATA TTT AAA AAT GGC TTA TTA GGT GCA GCT AGA TCT ATG  1995
   337 S   D   S   L   A   V   I   A   A   N   A   N   L   I   F   K   N   G   L   L   G   A   A   R   S   M    362
  1996 CCA ACT TCA GGA GCT TTG GAT AAG GTT GCA GCC AAG AAT GGA ATA AAA CTA TTT GAA ACA CCA ACA GGT TGG AAA TTC  2073
   363 P   T   S   G   A   L   D   K   V   A   A   K   N   G   I   K   L   F   E   T   P   T   G   W   K   F    388
  2074 TTT GGA AAT TTA ATG GAT GCA GGA TTG ATC AAT CTA TGT GGA GAA GAA AGT TTT GGA ACT GGG TCA AAT CAT ATC AGA  2151
   389 F   G   N   L   M   D   A   G   L   I   N   L   C   G   E   E   S   F   G   T   G   S   N   H   I   R    414
  2152 GAA AAA GAT GGA ATT TGG GCA GTG TTA GCA TGG TTG ACT ATT TTG GCA CAT AAA AAT AAG AAT ACT GAT CAT TTT GTC  2229
   415 E   K   D   G   I   W   A   V   L   A   W   L   T   I   L   A   H   K   N   K   N   T   D   H   F   V    440
  2230 ACT GTT GAA GAA ATA GTC ACT TAA TAT TGG TAA TAA TTT GGT AGA AAC TAT TAC TCA AGA TAT GAT TAT GAA TAA GTT  2307
   441 T   V   E   E   I   V   T   Q   Y   W   Q   Q   F   G   R   N   Y   Y   S   R   Y   D   Y   E   Q   V    466
  2308 GAT TCT GCT GGT GCT AAT AAA ATG ATG GAG CAT TTA AAA ACT AAA TTC TAA TAT TTT GAA TAA TTG AAA CAA GGA AAT  2385
   467 D   S   A   G   A   N   K   M   M   E   H   L   K   T   K   F   Q   Y   F   E   Q   L   K   Q   G   N    492
  2386 AAA GCC GAT ATT TAT GAT TAT GTT GAT CCT GTT GAT CAA TCA GTG TCA AAA AAC CAA GGA GTT AGA TTC GTT TTT GGA  2463
   493 K   A   D   I   Y   D   Y   V   D   P   V   D   Q   S   V   S   K   N   Q   G   V   R   F   V   F   G    518
  2464 GAT GGA TCA AGA ATT ATA TTT AGA TTG AGT GGT ACT GGA TCA GTA GGT GCT ACT ATT AGA ATC TAC TTT GAA TAA TTT  2541
   519 D   G   S   R   I   I   F   R   L   S   G   T   G   S   V   G   A   T   I   R   I   Y   F   E   Q   F    544
  2542 GAA TAA TAG TAG ATT CAA CAT GAA ACA GCT ACT GCT TTA GCC AAT ATT ATT AAG CTT GGA TTG GAA ATC AGT GAT ATT  2619
   545 E   Q   Q   Q   I   Q   H   E   T   A   T   A   L   A   N   I   I   K   L   G   L   E   I   S   D   I    570
  2620 GCT TAA TTC ACA GGA AGA AAT GAA CCA ACT GTC ATA ACT TGA ACAATC AATCATATTAATAAATCAAT ATACACAAAAGCAGCAAAAG 2707
   571 A   Q   F   T   G   R   N   E   P   T   V   I   T   *                                                         583
  2708 ACACAAGACCAATTGCAGAT TAAGCTACTTCAGAGGCTGC ATGCACATATGCCGCTCTCA TCTTGTATGAAGATAATCAA GATATCGACGCTACCAAGTT 2807
  2808 AGCTAAGATCGTTAAGGCAG CTAATTTGAGAGTTGAACCA ATTTGGACAAAAGTTTTCGA AAAGGCTTTGAAAGGAAAAA AAGTTGGTGATTTATTACAT 2907
  2908 GGAAACAGCGGAAATGCAGG AGGTGCACCAGCAGCACAAG CTACAGCCACACCAGCTGCA GCAGAAGCAAAGAAGCCTGA GCCTGTTAAGGAGGTTAAGA 3007
  3008 AAGCTGAAGAACCAGAAGAA GATGTTGATATGGGTGGTTT ATTTGATTGAAAATGAATAC AAAATCGTACATTAATATTC G  3088
```

B

5' BamH I 1 EcoR I 2 3 Hind III 3'

913 2661

958 1740 1057

Translated region ▲— n —▲ Expected PCR product

FIG. 2. (*A*) Nucleotide and deduced amino acid sequence of parafusin. Parafusin cDNA was isolated, purified, and sequenced as described. Nucleotide sequence is in the 5′ to 3′ direction with the initiating codon GTA at position 913. In-frame stop codons are marked by *. Underlined amino acid sequences are the sequences that match peptide fragments obtained from tryptic digests of purified parafusin. Poly(A) signal is located at position 2679. (*B*) Restriction map and schematic diagram of parafusin cDNA showing the expected size (in bp) of the PCR products amplified with the specific primers (listed in *Materials and Methods*). The enzyme *Bam*HI cuts in the untranslated region at position 704, while *Eco*RI and *Hin*dIII cut the insert in the translated region at positions 1459 and 2593, respectively. Probes used for Southern blot analysis (see Fig. 5) are represented as open bars over the translated region (1, N terminus; 2, pPFUS specific; 3, 63-mer). Upper line marks every 100 nucleotides.

out the molecule. PGM, a well known and extensively studied glycolytic enzyme, catalyzes an intramolecular transfer of phosphate (bound to Ser^{116}) between C-1 and C-6 of glucose. Fig. 4 shows the alignment of the cDNA deduced *Paramecium* parafusin (pPFUS) sequence to the known PGM sequences: rabbit muscle PGM (15), rabbit muscle PGM type 1 (16), and human PGM (17), with the asterisk representing residues identical to rabbit muscle PGM. pPFUS exhibits a unique N-terminal domain, four regions of specific insertions (I-1–I-4), two major (D-1 and D-2) and several minor deletions in the amino acid sequence as compared to PGM (Fig. 4). One of the insertions (I-1) is located near the active site region of rabbit muscle PGM (Ser^{116}, shaded). pPFUS also has a potential N-glycosylation site at position 289 (Fig. 4, boldface underlined). A computer search for motifs revealed a consensus sequence for a prokaryotic membrane lipoprotein lipid attachment site (18) on a cysteine residue (Fig. 4, arrowhead). In addition, searches for phosphorylation sites (19), showed the presence of a protein kinase C site (KXXSXR) KDFSFR (Fig. 4, double underlined) unique to pPFUS. Moreover, one Ca^{2+}/calmodulin-dependent protein kinase site (KXXT) was identified (KPGT, underlined), which was present in all four proteins. An additional four Ca^{2+}/calmodulin-dependent protein kinase sites (residues underlined) were present in the other three PGM proteins (16). Two of these four sites are not present in pPFUS since they are interrupted by regions of insertions (I-2 and I-4). The remaining two sites are located close to the major deletion (D-2) in pPFUS in a region where there is little homology to rabbit muscle PGM sequences. The pPFUS N terminus contains 21 unique amino acid residues. The N-terminal residues of rabbit muscle PGM type 1 also differ from rabbit muscle PGM (16) but are not homologous to the pPFUS sequence.

It cannot be ruled out that parafusin may be an isoform of PGM, with completely different properties, that may have resulted from alternative mRNA splicing and/or gene duplication and subsequent independent mutational remolding of

FIG. 3. PCR amplification of total *Paramecium* DNA using internal primers of the parafusin cDNA sequence. Lane 1, molecular weight markers (123-bp ladder); lane 2, primers 36 (bp 919–985) and 14 (bp 2640–2659); lane 3, primers 15 (bp 1854–1873) and 12 (bp 2892–2911); lane 4, primers 17 (bp 96–115) and 21 (bp 1035–1054). Arrowheads show amplification products.

the respective gene products and/or posttranslational modifications. In a range of organisms, multiple isozymes of PGM have been identified, thought to be due to multiple PGM loci (20). cDNAs encoding two isoforms of PGM have been isolated from rabbit skeletal muscle (16). One isoform is rabbit muscle PGM type 1 (Fig. 4). The second, rabbit muscle PGM type 2, is identical to a M_r 60,000 Ca^{2+}/calmodulin-dependent phosphoprotein in the sarcoplasmic reticulum that is believed to potentially regulate Ca^{2+} release via its phosphorylation/dephosphorylation. The presence of another PGM-related protein has been reported from smooth muscle cells and fibroblasts; similarly to parafusin (21), this protein has no PGM activity (22). Several other examples of homologous proteins with varied functions are currently being reported (23).

Southern Blot Analysis. To explore whether there are different genes for parafusin and PGM in *Paramecium*, Southern blot analysis was performed on *Paramecium* total DNA cut with *Bam*HI, *Eco*RI, and *Hin*dIII restriction enzymes. Specific oligonucleotide probes have been synthesized to the N terminus (Fig. 2*B*, 1) and to the I-3 of pPFUS (Fig. 2*B*, 2) and to a specific region (using *Paramecium* codon usage) common to the three mammalian PGM molecules corresponding to a deletion region (D-2) in pPFUS. We have used these probes as well as with the 63-mer probe (Fig. 2*B*, 3) on identical blots to test whether *Paramecium* possesses a putative PGM isoform that is distinct from parafusin. In the *Hin*dIII digests (Fig. 5) the 63-mer oligonucleotide recognized three major bands of 8.2, 5.0, and 4.0 kb (lane 1, arrowheads). The 4.0- and 5.0-kb DNA species are both seen with the pPFUS-specific probe (lane 2, arrowheads), and the 4.0-kb band is detectable with the N-terminal probe (lane 4, arrowhead). In contrast, no prominent species were seen hybridizing with the PGM-specific probe (lane 3). Since there are no introns in the parafusin gene (Fig. 3), these results indicate that there are separate genes in *Paramecium* for parafusin and PGM. In the *Eco*RI digests (data not shown),

FIG. 5. Southern blot analysis of total *Paramecium* DNA cut with *Hin*dIII (lanes 1–4) or *Bam*HI (lanes 5 and 6) using different probes specific for parafusin and PGM sequentially after stripping the same blot. Molecular size markers are indicated on the left in kb; arrowheads indicate prominent hybridization reactions. Lane 1, 63-mer, corresponding to parafusin amino acids 435–455; lane 2, 29-mer oligonucleotide corresponding to parafusin residues 299–308 (see Fig. 4, I-3); lanes 3 and 5, 39-mer, corresponding to PGM residues 445–457, which are deleted (see Fig. 4, D-2) in parafusin; lanes 4 and 6, 66-mer, specific for the N-terminal region of parafusin residues 3–24. Location of probes is shown in diagram in Fig. 2*B*.

```
pPFUS    1 MVLFLLPLRLGHNLWRIEAPR*Q*-*QP*AG**********K*SE-ATQP**L***V***FN*LR-KDELKNV*F******YFNRQ**FS*I*L*Y**D*SE
rPGM     1                   VKI-VTVKTKAYPDQKPGTSGLRKRVKVFQSSTNYAENFIQSIISTVEPAQRQEATLVVGGDGRFYMKEAIQLIVRIAAANGIGR
rPGM1    1          MEEGPLPLLTIR---TAPYH-***********KTYYFEDKPC*L*******FFSIDLKDRQGSS********YFN*S**ET*LQM********
hPGM     1                      M***-*****Q**Q*******************A**********************************************A**********

                       I-1                                                        D-1
pPFUS  100 VHV**A*LM****S*HY***VNEEVGNC**************KEH********VRT*A****DF**Q*YTHTTK*K**-----*T**YEFEKHINL*QIGVY
rPGM    85 LVIGQNGILSTPAVSCIIRKI-----KAIGGIILTASHNPGG-PNGDFGIKFNISNGGPAPEAITDKIFQISKTIEEYAICPDLKVDLGVLGKQQFDLENKF
rPGM1   90 *********************----*****************-*************************************************************
hPGM    86 *********************----*****************-*******************************************V*****************

        I-2
pPFUS  197 *FEGTRLEKSHFE*KV**T*QD*TQLMQKL***DL**G*F*NKD-FSF*F*G****A***A*H*FGTL**CSKESELL**D*S*****G*********H***
rPGM   181 K---------PFTVEIVDSVEAYATMLRNIFDFNALKELLSGPNRLKIRIDAMHGVVGPYVKKILCEELGAPANSA-VNCVPLEDFGGHHPDPNLTYAADLV
rPGM1  186 *---------***********************************************************************-*************************
hPGM   182 *---------*****************S****S**************C*********************************-*************************

          I-3          ▼
pPFUS  298 *LLDIHKKKDVGTVPQ****C***A*******R-Q***T****L******---ANLIFKN*LL*A***********K**AKNG*K*F***************
rPGM   273 ETMK-------SGEHDFGAAFDGDGDRNMILGKHGFFVNPSDSVAVIAANIFSIPYFQQTGVRGFARSMPTSGALDRVANATKIALYETPTGWKFFGNLMDA
rPGM1  278 *****------*********************************************************************************************
hPGM   274 *****------**************************************************************************S*******************

                                                       I-4                                           D-2
pPFUS  396 GLIN***********N*******I*******T***HKNKNTDHFVT**E*VTQYWQQ****YYS*****Q*DSA**N***EH*KT----------K*QYFEQ
rPGM   368 SKLSLCGEESFGTGSDHIREKDGLWAVLAWLSILATRKQ------SVEDILKDHWHKFGRNFFTRYDYEEVEAEGATKMMKDLEALMFDRSFVGKQFSANDK
rPGM1  373 ***************************************------*************************************************************
hPGM   369 ***************************************------**********Q*H********************N****************************

pPFUS  488 LKQGN***IYD*V****Q*******V*FV*G*************V*****I*FEQF*QQ--Q*QHETATA**NI*KLG*EI*DIAQF***NE*****
rPGM   464 VYTVEKADNFEYHDPVDGSVSKNQGLRLIFADGSRIIFRLSGTGSAGATIRLYIDSYEKDNAKINQDPQVMLAPLISIALKVSQLQERTGRTAPTVIT
rPGM1  469 **************************************************************************************************
hPGM   465 ***********S******I*R********T*****V**************************V***********************************
```

FIG. 4. Computerized alignment of the deduced pPFUS amino acid sequence to the known mammalian PGM sequences (rPGM, rabbit muscle PGM; rPGM1, rabbit muscle PGM type 1; hPGM, human PGM). *, Amino acids identical to rabbit muscle PGM. Potential Ca^{2+}/calmodulin phosphorylation sites are underlined. Double-underlined residues represent a potential protein kinase C phosphorylation site present only in parafusin. Active site Ser^{116} in PGM is shaded and prokaryotic membrane lipoprotein lipid attachment site is marked with an arrowhead. Potential N-glycosylation site common to all sequences is in boldface (underlined).

the 63-mer hybridized prominently at 4.8 and 2.4 kb. These two major bands were also seen with the pPFUS-specific probe, whereas the PGM-specific probe hybridized to a single DNA species of 3.7 kb. However, this 3.7-kb band was the main band hybridizing with the N-terminal probe, indicating that parafusin and *Paramecium* PGM may have similar N termini. In *Bam*HI digests (Fig. 5), a band at 6.5 kb hybridizes with both the parafusin N-terminal (lane 6) and 63-mer (data not shown) probes and a band at 4.6 kb also hybridizes with both the PGM-specific (lane 5) and the pPFUS N-terminal-specific (lane 6) probes. This may suggest that the N terminus of *Paramecium* PGM is homologous to the hydrophobic putative membrane association sequence in parafusin or that this sequence is lost in subsequent processing to the mature PGM in all species. Southern blot analysis of rat, mice, and yeast DNA using the same parafusin-specific probe suggests the presence of parafusin counterpart(s) in these species (24).

The sequence and structural information presented here suggest that the region near the active site of PGM is altered in the pPFUS molecule, which is consistent with the observation that isolated parafusin does not have detectable PGM activity (21). We do not know what role the other structural alterations in the PGM motifs play, but the specific disruption of four Ca^{2+}/calmodulin kinase sites is quite striking and probably highly significant. Anti-peptide antibodies made against the derived amino acid sequence of the pPFUS I-4 (see Fig. 4) not present in mammalian PGMs recognize a protein band on immunoblots corresponding to parafusin in *Paramecium* and in other organisms but do not crossreact either with purified rabbit muscle PGM or with *Paramecium* PGM-enriched fractions recently isolated by us (21). Furthermore, we have previously shown that parafusin incorporates uridine 5′-[β-[^{35}S]thio]diphosphate glucose and this label can be chased off not with unlabeled glucose 1-phosphate but only with unlabeled UDP glucose (4, 5). This is in contrast to PGM, which incorporates glucose 1-^{32}P and does not incorporate UDP glucose (21). Therefore, we hypothesize that parafusin and PGM are related molecules but with different functions in the cell. The protein kinase C site that is present in pPFUS but not in PGM could be related to the serine phosphorylation of pPFUS as part of the exocytic cycle (4). It will be important to locate the O-glucosylation site in pPFUS and to determine whether a comparable site is present in PGM.

In all probability, parafusin and PGM belong to a superfamily of proteins. Parafusin and PGM join the list of examples of proteins that have conserved critical regions of primary structure to maintain overall tertiary structure during evolution, while diverging in function.

The authors thank Dr. A. Eisen and Dr. M. Rapport for valuable technical advice and comments on this manuscript. E.W. is a Fulbright Fellow from the Department of Cell Biology, Nencki Institute of Experimental Biology, Warsaw, Poland. This work was supported in part by National Institutes of Health Grant DK 41918.

1. Gilligan, D. M. & Satir, B. H. (1982) *J. Biol. Chem.* **257**, 13903–13906.
2. Satir, B. H., Busch, G., Vuoso, A. & Murtaugh, T. J. (1988) *J. Cell. Biochem.* **24**, 429–443.
3. Satir, B. H., Hamasaki, T., Reichman, M. & Murtaugh, T. J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 930–932.
4. Subramanian, S. V. & Satir, B. H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 11297–11301.
5. Satir, B. H., Reichman, M., Srisomsap, C. & Marchase, R. B. (1990) *J. Cell Biol.* **111**, 901–907.
6. Murtaugh, T. J., Gilligan, D. M. & Satir, B. H. (1987) *J. Biol. Chem.* **262**, 15734–15739.
7. Martindale, D. W. (1989) *J. Protozool.* **36**, 29–34.
8. Godiska, R., Aufderheide, K. J., Gilley, D., Hendrie, P., Fitzwater, T., Preer, L. B., Polisky, B. & Preer, J. R., Jr. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7590–7594.
9. Martin, R., Hoover, C., Grimme, S., Grogan, C., Holtke, J. & Kessler, C. (1990) *BioTechniques* **9**, 762–768.
10. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
11. Pritchard, A. E., Seilhamer, J. J. & Cummings, D. J. (1986) *Gene* **44**, 243–253.
12. Mahalingam, R., Seilhamer, J. J., Pritchard, A. E. & Cummings, D. J. (1986) *Gene* **49**, 129–138.
13. Dawson, D. M. & Greene, J. M. (1975) in *Isozymes, Third International Conference*, ed. Market, C. (Academic, New York), Vol. 1, pp. 381–388.
14. Linz, J. E., Lira, L. M. & Sypherd, P. S. (1986) *J. Biol. Chem.* **261**, 15022–15029.
15. Ray, W. J., Jr., Hermodson, M. A., Puvathingal, J. M. & Mahoney, W. C. (1983) *J. Biol. Chem.* **258**, 9166–9174.
16. Lee, Y. S., Marks, A. R., Gureckas, N., Lacro, R., Nadal-Ginard, B. & Kim, D. H. (1992) *J. Biol. Chem.* **267**, 21080–21088.
17. Whitehouse, D. B., Putt, W., Lovegrove, J. U., Morrison, K., Hollyoake, M., Fox, M. F., Hopkinson, D. A. & Edwards, Y. H. (1992) *Genetics* **89**, 411–415.
18. Hayashi, S. & Wu, H. C. (1990) *J. Bioenerg. Biomembr.* **22**, 451–471.
19. Pearson, R. & Kemp, B. E. (1991) *Methods Enzymol.* **200**, 62–81.
20. Quick, C. B., Fisher, R. A. & Harris, H. (1974) *Eur. J. Biochem.* **42**, 511–517.
21. Andersen, A. P., Wyroba, E., Zhao, H., Reichman, M. & Satir, B. H. (1994) *Biochem. Biophys. Res. Commun.* **200**, 1353–1358.
22. Belkin, A. M., Klimanskaya, I. V., Lukashev, M. E., Lilley, K., Critchley, D. R. & Koteliansky, V. E. (1994) *J. Cell Sci.* **107**, 159–173.
23. Singh, R. & Green, M. R. (1993) *Science* **259**, 365–368.
24. Wyroba, E. & Satir, B. H. (1994) *Mol. Biol. Cell* **4**, 430a, 2495.