Major Epiplasmic Proteins of Ciliates Are Articulins: Cloning, Recombinant Expression, and Structural Characterization

Irm Huttenlauch,* Norbert Geisler,* Uwe Plessmann,* Robert K. Peck,‡ Klaus Weber,* and Reimer Stick§

*Max Planck Institute for Biophysical Chemistry, Department of Biochemistry, D-37018 Göttingen, Federal Republic of Germany; ‡Department of Zoology and Animal Biology, University of Geneva, CH-1211 Geneva, Switzerland; and §Institut für Biochemie und Molekulare Zellbiologie, Abteilung Entwicklungsbiochemie, Universität Göttingen, D-37073 Göttingen, Federal Republic of Germany

Abstract. The cytoskeleton of certain protists comprises an extensive membrane skeleton, the epiplasm, which contributes to the cell shape and patterning of the species-specific cortical architecture. The isolated epiplasm of the ciliated protist *Pseudomicrothorax dubius* consists of two major groups of proteins with molecular masses of 78–80 kD and 11–13 kD, respectively. To characterize the structure of these proteins, peptide sequences of two major polypeptides (78–80 kD) as well as a cDNA representing the entire coding sequence of a minor and hitherto unidentified component (60 kD; p60) of the epiplasm have been determined. All three polypeptides share sequence similarities. They contain repeated valine- and proline-rich motifs of 12 residues with the consensus VPVP--V-V-V-. In p60 the central core domain consists of 24 tandemly repeated VPV motifs. Within the repeat motifs positively and negatively charged residues, when present, show an alternating pattern in register with the V and P positions. Recombinant p60 was purified in 8 M urea and dialyzed against buffer. Infrared spectroscopic measurements indicate 30% β-sheet. Electron microscopy reveals short filamentous polymers with a rather homogenous diameter (~15–20 nm), but variable lengths. The small polymers form thicker filaments, ribbons, and larger sheets or tubes. A core domain similar to that of *P. dubius* p60 is also found in the recently described epiplasmic proteins of the flagellate *Euglena*, the so-called articulins. Our results show that the members of this protein family are not restricted to flagellates, but are also present in the distantly related ciliates where they are major constituents of the epiplasm. Comparison of flagellate and ciliate articulins highlights common features of this novel family of cytoskeletal proteins.

THE membrane skeleton of eukaryotic cells is part of the cytoskeleton. It is involved in the organization and maintenance of cell shape and functions in the generation and stabilization of specialized membrane domains. This submembraneous organization is usually dictated by microfilamentous structures containing actin and a variety of actin-binding proteins. Particularly well-analyzed examples are, for instance, the membrane skeletons of vertebrate red blood cells (for review see Bennett, 1990) and the intestinal brush border cells of various vertebrates (for review see Louvard, 1989). An actin-based subplasma membrane organization occurs not only in metazoan cells, but is also typical for some protists such as *Dictyostelium* and *Physarum* (for reviews see Schleicher and Noegel, 1992; Stockem and Brix, 1994).

In contrast, other protists possess a unique cortical cytoplasm: the epiplasm. The epiplasm is always in close contact with a membrane (Peck, 1977; Grain, 1986). In flagellates like *Euglena* it is the plasma membrane, in ciliates it is the inner membrane of the alveolus, a membrane system that closely apposes the plasma membrane (Peck, 1977; Bricheux and Brugerolle, 1986, 1987) and is thought to be involved in regulation of the intracellular calcium level (Stelly et al., 1991). It is generally believed that the epiplasm functions in maintaining cell shape. Moreover, there is experimental evidence that it is involved in pattern-forming processes like cortical organelle positioning (Aufderheide, 1983; Peck, 1986). In ciliates, cilia or ciliary complexes with their basal bodies that are homologous to centrioles in metazoans, are among these cortical organelles. They are polar structures, and their precise and asymmetrical arrangement defines the anterioposterior polarity (corresponding to the normal swimming direction) and the left–right or circumferential asymmetry of these unicellular organisms. The cortical patterns are species specific. Moreover, they have to be precisely reproduced during cell division: during the equatorial mode of cell division the anterior division product has to reconsti-

Address all correspondence to Irm Huttenlauch, Max Planck Institute for Biophysical Chemistry, Department of Biochemistry, P.O. Box 2841, D-37018 Göttingen, FRG. Tel.: 49 (551) 201-1486. Fax: 49 (551) 201-1578.

tute posterior structures, and the posterior part anterior structures, respectively (for reviews see Aufderheide et al., 1980; Cohen and Beisson, 1988; Frankel, 1989).

The epiplasm of the ciliate *Pseudomicrothorax dubius* is a thick, continuous, proteinaceous layer (Peck, 1977). It includes distinct regions such as the terminal plates and the docking sites, where basal bodies and trichocysts are attached. These regions, as well as the ribs extending between adjacent rows of terminal plates, are precisely positioned relative to one another (for a detailed ultrastructural description of the epiplasm see Peck, 1977; Peck et al., 1991). The epiplasm of *P. dubius* can be isolated in milligram quantity in a highly purified form, free of cortical fibers (microtubules, microfilaments). This has allowed a detailed biochemical analysis (Huttenlauch and Peck, 1991; Peck et al., 1991). The epiplasm of *P. dubius* consists of two major groups of polypeptides, 78–80 and 11–13 kD, as well as a series of minor components of 18–62 kD. The group of 78–80-kD proteins consists of at least three different polypeptides and numerous isoelectric variants. Immunelectron microscopy revealed that at least one of these proteins is located throughout the entire epiplasm, while many of the minor components, which are glycosylated, are located at the outer surface of the epiplasm, or might even be membrane proteins that are tightly bound to the epiplasm (Curtenaz and Peck, 1992). Ultrastructural analysis of some protists such as dinoflagellates and entodiniomorphid ciliates indicates a filamentous substructure of the epiplasm with filament diameters of 2–3 nm (Viguès et al., 1984; Cachon et al., 1987; Viguès and David, 1989). In other protists, including *P. dubius*, the epiplasm appears homogenous with no obvious indication for the existence of filaments. Moreover, in no case have filaments been observed by electron microscopy in material obtained by in vitro reconstitution experiments of *P. dubius* epiplasm (Peck et al., 1991).

To unravel the nature of the epiplasmic proteins of ciliates we have carried out primary sequence analysis of several epiplasmic polypeptides of *P. dubius*. We present here partial peptide sequences of two of the major constituents as well as the full-length cDNA sequence of one minor species. All three polypeptides show sequence similarities and contain repeated valine- and proline-rich, 12-residue-long motifs with the consensus sequence VPVP--V-V-V-. Similar amino acid repeats have recently been described for two major epiplasmic proteins of the flagellate *Euglena gracilis* (Marrs and Bouck, 1992), which have been named articulins.

Articulins have no significant sequence similarity to other known proteins. The presently known flagellate articulins and the ciliate epiplasmic proteins described here show overall sequence similarity, both are characterized by repeated VPV motifs with characteristic charge distribution in their central domains, and thus belong to the same class of cytoskeletal proteins. Members of this novel protein family comprise the major protein constituents of the epiplasm of ciliates as well as flagellates.

Materials and Methods

Cultures and Epiplasm Preparation

P. dubius strain N5b was cultivated, harvested, and the epiplasm was prepared as described (Peck et al., 1991).

Electrophoretic Procedures

SDS-PAGE using 7.5 or 10% (wt/vol) polyacrylamide slab gels was performed as described (Huttenlauch and Peck, 1991). The M_r values of the polypeptides were determined from mobilities on SDS-PAGE compared with the following protein standards of $M_r \times 10^{-3}$: phosphorylase b (94), BSA (67), and ovalbumin (43). Two-dimensional gel electrophoresis and electrophoretic transfer of proteins to nitrocellulose sheets was as described (Huttenlauch and Peck, 1991) except that transfer was carried out at 240 mA for 8 h.

Immunolabeling of Gel Blots

Immunolabeling of gel blots was carried out as described (Stick, 1988). Antiserum 018 and 015 (Peck et al., 1991) were diluted 1:1,250. The secondary antibody was alkaline phosphatase–conjugated goat anti–rabbit IgG (Dianova, Hamburg, FRG) diluted 1:2,000.

Isolation of Epiplasmic Proteins, Preparation of Fragments and Peptide Sequences

Epiplasmic polypeptides separated by NEPHGE followed by SDS-PAGE were visualized by staining with 0.1% Coomassie brilliant blue R in 45% (vol/vol) methanol, 10% (vol/vol) acetic acid. Individual spots were excised from 15 gels and stored at −70°C until use. Electrophoretic protein concentration, transfer to poly(vinylene difluoride) membrane, digestion with endoproteinase Asp-N or trypsin, HPLC of the resulting peptides, and microsequencing were essentially as described by Ungewickell et al. (1994).

RNA Techniques and Northern Analysis

RNA from *P. dubius* was isolated according to Chomczynski and Sacchi (1987). Poly(A)$^+$ RNA was isolated by selection on oligo(dT)-cellulose (Boehringer Mannheim Biochemicals, Mannheim, FRG) using standard procedures (Maniatis et al., 1982). Northern blots of RNA separated on denaturing formaldehyde agarose gels were hybridized in 5 × SSC at 65°C overnight with a 500-bp PCR fragment (generated with primers P4 and P5) radiolabeled with [^{32}P]dCTP (3,000 Ci/mmol, Amersham Buchler GmbH, Braunschweig, FRG) to high specific activity using a random prime kit (Pharmacia Diagnostics AB, Uppsala, Sweden). Filters were washed two times 30 min each in 1 × SSC, 1% SDS at 65°C, and once for 30 min in 0.5 M Na-phosphate, pH 7.0, at 65°C. Filters were exposed to x-ray films (XAR-5; Eastman Kodak Co., Rochester, NY) at −70°C using intensifying screens for 2–3 d.

Oligonucleotides

Peptide amino acid sequences were used to synthesize degenerated oligonucleotides (I, inosine) of the coding as well as the noncoding strand: spot 4 peptide *e* (DRQIEKP) sense primer P1 5′GAT/CAGAT/CAIATIGAIAAICC3′ and (PVYIEVE) antisense primer P2 5′TCIACT/CTCIATITAIACIGG3′; peptide *g*: (ERPVPVA) antisense primer P3 5′GCIACIGGIACIGGTCTTTC3′. Oligonucleotides deduced from the nucleotide sequence of the PCR products obtained in the first round of RT-PCR: spot 4 peptide *e* (VEVPVER) sense primer P4 5′GTCGAGGTCCCAGTTGAGCG3′, antisense primer P5 5′CACTGGCACTTCAACTGGAG3′. Oligonucleotides deduced from the partial cDNA 3-1 for the amplification of the 5′ end of the cDNA: antisense primer P6 5′CCAACTGGCGTCCCTGCTGCCTGC3′; antisense nested primer P7 5′GCCGCTTGAAGTTGTCCTTGTAGAGC3′. Oligonucleotide designed to amplify the open reading frame of p60 flanked by restriction sites BamHI and HindIII: sense primer with BamHI site P8 5′CGGGATCCATGATTCCCAACACTGCACTAC3′.

Polymerase Chain Reaction

First-strand cDNA was synthesized from total RNA using the Gene Amp reverse transcriptase-PCR kit from Perkin-Elmer Cetus (Norwalk, CT). 1 μg total RNA was primed per reaction with a mixture of random hexanucleotides following the manufacturer's protocol. PCR was carried out with combinations of degenerated primers (0.4 μm each) deduced from the peptide sequences. Primer combinations that gave specific products were P1/P2 and P1/P3. The cycling parameters were: 4 cycles (1 min, 94°C; 1 min, 42°C; 2.5 min, 72°C), and 30 cycles (1 min, 94°C; 1 min, 50°C; 2.5 min, 72°C), and a final polymerization step (12 min, 72°C). To appropriate aliquots of the reaction mixture 0.1 vol of all four dNTPs (10 mM) and 1

μl Klenow polymerase (5 U/μl) (Boehringer Mannheim Biochemicals) was added. After incubation at 37°C for 15 min individual fragments were separated on agarose gels (1–2.2%) and electroeluted. The fragments were treated with polynucleotide kinase (Boehringer Mannheim Biochemicals) under standard conditions for 1 h, extracted with phenol/chloroform and chloroform, and precipitated with ethanol. The fragments were then cloned into the EcoRV restriction site of the Bluescript KS vector (Stratagene Corp., La Jolla, CA) and subjected to double-stranded sequencing by the dideoxy chain termination method with either T7 and T3, or KS and SK primers using the Sequenase kit (United States Biochemical Corp., Cleveland, OH).

cDNA Library Construction and Screening

An oligo (dT)-primed cDNA library was synthesized from *P. dubius* poly $(A)^+$ RNA and ligated unidirectionally into λZAP vector arms using the λZAP–cDNA synthesis kit (Stratagene Corp.). The library was packaged in Gigapack II Gold packaging extract (Stratagene Corp.) and plated on *E. coli* SURE cells (Stratagene Corp.). The library had a complexity of $>2 \times 10^6$ plaque-forming units with average insert sizes between 0.7 and 2 kbp. The library was amplified once on *E. coli* SURE. About 1×10^6 pfu of the amplified library were screened with a ~500-bp PCR fragment (obtained in an RT-PCR with primers P4/P5) radiolabeled with [^{32}P]dCTP (3,000 Ci/mmol) (Amersham Corp.) to high specific activity using a random prime kit (Pharmacia Diagnostics AB). Library screening, plaque purification, conversion of λZAP into Bluescript plasmid, DNA isolation, subcloning, and sequencing were by standard methods (Maniatis et al., 1982) and according to manufacturer's instructions. Four independent phage clones were isolated that represented partial cDNA clones encoding p60. The cDNA insert of the largest clone (termed pBst SK 3-1) was sequenced on both strands.

Amplification of the 5′ end of the cDNA encoding p60 was done with the gene-specific primers P6 and P7 using the 5′-Amplifinder RACE kit (Clontech, Palo Alto, CA) following the protocol of the manufacturer. Cycling parameters of the PCR step were: 40 cycles (45 s, 94°C; 45 s, 60°C; 2 min, 72°C), and a final polymerization step (7 min, 72°C). The resulting PCR fragment was cloned into the EcoRV restriction site of the Bluescript KS vector (Stratagene Corp.) (pBst KS 5′ p60).

Plasmid Constructions and Protein Expression

To obtain a clone containing the entire coding sequence of p60 the StyI/StyI restriction fragment of plasmid pBst SK 3-1 containing the bases +71 to +1714 of the cDNA (numbering with respect to the cDNA sequence given in Fig. 4) was cloned into pBst KS 5′ p60 that had been linearized with StyI. Clones containing the StyI/StyI fragment in sense orientation were identified by PCR and restriction fragment analysis. p60 was translated in vitro using the TNT-coupled transcription translation system (Promega Corp., Madison, WI) and [^{35}S]methionine (Amersham Buchler GmbH) following the instructions of the manufacturer.

To express p60 in *E. coli*, the entire open reading frame was amplified from the plasmid containing the full-length cDNA by PCR using a sense primer that started at position +1 and that contained a BamHI recognition site at its 5′ end (P8) and, as antisense primer, the vector-derived KS primer that primed downstream of the cDNA insert in the multiple cloning site. Cycling parameters were: initial denaturation for 2 min, 94°C followed by 35 cycles (1 min, 94°C; 1 min, 55°C; 2 min, 72°C) and a final polymerization step (7 min, 72°C). The PCR product was purified using a spin column (QIAGEN Inc., Chatsworth, CA), double digested with BamHI and HindIII, and gel purified by agarose gel electrophoresis. The fragment was cloned into BamHI/HindIII double-digested pINDU vector (Bujard et al., 1987) and used to transform *E. coli* BL 21. Transformed bacteria were grown overnight in rich medium (QIAGEN Inc.) in the presence of 100 μg/ml ampicillin. 200 μl culture was harvested, centrifuged for 5 min at 6,000 rpm in a tabletop centrifuge, resuspended in 100 μl SDS sample buffer, and boiled for 15 min. 5–20 μl of this lysate was loaded per lane. As a control, *E. coli* BL 21 cells were transformed with pINDU encoding an irrelevant polypeptide of ~45 kD.

For purification of recombinant p60, inclusion bodies were isolated by standard methods and dissolved in 8 M urea buffer. Purification of p60 was achieved by anion exchange chromatography in 8 M urea buffer on Mono Q (Pharmacia Diagnostics AB). Alternatively, p60 was purified in 8 M urea buffer on Mono S followed by Mono Q. In addition, such material was also subjected to gel filtration on Superose 6 (Pharmacia Diagnostics AB) in the presence of 8 M urea. The purity of the recombinant protein was monitored by SDS-PAGE and amino-terminal sequencing. All three preparations gave similar images in electron microscopy.

Deglycosylation of Epiplasmic Proteins

Deglycosylation of *P. dubius* epiplasmic proteins with *N*-glycosidase F (Boehringer Mannheim Biochemicals) was essentially as described by Huttenlauch and Peck (1991).

Spectroscopy

For Fourier-transform-infra-red (FTIR)[1] spectroscopy, the urea solution was concentrated to 8 mg/ml protein and dialyzed for 48 h at 4°C against D_2O containing 10 mM Tris-HCl at pD 8.0 and 1 mM 2-mercaptoethanol. The spectrum was recorded on a spectrometer (IFS25; Bruker Instruments, Karlsruhe, FRG) using a CaF_2 cell with a 50-μm Teflon spacer. 20 interferograms were taken in the range 400–4,000 cm^{-1} with a resolution of 2 cm^{-1}. Fourier self-deconvolution was in part performed with software provided by the laboratory of H. H. Mantsch (Institute for Biodiagnostics, Winnipeg, Canada) (Kauppinen et al., 1981). Line narrowing of the slightly smoothed spectrum was performed using a triangular apodization function, assuming an initial line-width of half-width at half-height = 17 cm^{-1} with a line-narrowing factor of k = 2.0. The components in the amide I region of the spectrum were fitted with Gaussian bands by using a deconvoluted spectrum over the range from 1,598 to 1,695 cm^{-1} (Byler and Susi, 1986). A constraint on the fitting was that all lines but the less well-defined random coil band at 1,641 cm^{-1} should have comparable widths that are consistent with the deconvolution parameters.

Electron Microscopy

Purified, recombinant p60 in 8 M urea buffer was dialyzed against 10 mM Tris-HCl, pH 8.0, 1 mM $MgCl_2$ or 1 mM $CaCl_2$, and 1 mM 2-mercaptoethanol. Dialysis was performed at room temperature using 0.025 μm filters (Millipore Corp., Milford, MA). Protein concentration was 0.5 mg/ml. Specimens were negatively stained using either 1% uranyl acetate or 2% phosphotungstic acid adjusted to pH 7.5 with NaOH.

Results

The protein pattern of the isolated epiplasm of *P. dubius* consists of two major groups of polypeptides with molecular masses of 78,000–80,000 and 11,000–13,000, respectively, as well as several minor components of 18,000–62,000. The 78,000–80,000-mol wt group of polypeptides is resolved into more than 10 spots by two-dimensional-gel electrophoresis (Peck et al., 1991; see Fig. 1). Peptide map analysis of six major spots (1–6; Fig. 1) indicated that some

1. *Abbreviation used in this paper*: FTIR, Fourier-transform-infra-red.

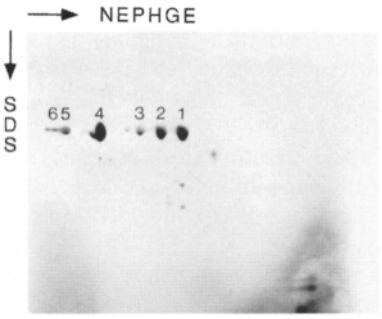

Figure 1. Two-dimensional gel electrophoretic analysis of epiplasm from the ciliate *P. dubius*. Epiplasmic polypeptides were separated by NEPHGE in the first dimension, and by SDS-PAGE in the second dimension. Upper portion of a Coomassie brilliant blue–stained gel is shown. Major spots of the 78,000–80,000-mol wt group of polypeptides are numbered 1–6. Spots 1 and 4 have been subjected to amino acid sequence analysis in this study.

of these spots represent isoelectric variants of the same polypeptides (1, 2, 3/5, 6, respectively) (Peck et al., 1991). In this report we have focused our sequence analysis on spots 1 and 4. They differ in their peptide maps and represent the two major components of the 78,000–80,000-mol wt group of polypeptides of *P. dubius* epiplasm (Peck et al., 1991).

To isolate polypeptides for protein sequencing, purified epiplasm was subjected to NEPHGE/SDS-PAGE. Individual spots were cut from the gels. Polypeptides from several gels were electrophoretically concentrated into a single gel slot, transferred electrophoretically to a polyvinylene difluoride membrane, and treated with endoproteinase Asp-N or with trypsin. Peptides released from the membrane were subjected to HPLC, and the resulting elution profile was screened by microsequencing. Sequence information of some 300 residues of spot 1 and 200 residues of spot 4 were obtained in total. Assuming a molecular mass of 80,000 for each of the polypeptides this would account for about 40 and 30%, respectively, of the complete sequences. About two-thirds of the peptides analyzed from both spots are rich in valine and proline, and more strikingly, these fragments show an alternating pattern of these two amino acids (VPV repeats). Alignments of the sequences are shown in Fig. 2 together with the consensus sequence of the VPV repeats of the epiplasmic protein p60 for which the complete amino acid sequence has been deduced by cDNA sequencing (see below). Although peptides from both spots display the same pattern of VPV-repetitive sequences, no identical peptide was obtained from both components. This holds also for the nonrepetitive peptide sequences obtained (results not shown). Thus it seems that spots 1 and 4 are indeed the products of two distinct genes as previously deduced from comparative peptide map analysis (Peck et al., 1991).

To generate hybridization probes for cDNA library screening we used reverse transcriptase-PCR. Degenerated oligonucleotide primers were synthesized based on selected peptide sequences (see Materials and Methods). The peptide sequence *e* of spot 4 (Fig. 2) was of sufficient length to allow the design of a primer pair from the same fragment. Two alternative primers with opposite orientation were synthesized for the other peptide sequences selected. Random primed cDNA was used as template in PCR amplifications with primer pairs in all possible combinations. With the two primers deduced from sequence *e* of spot 4, a single product of the expected size was obtained. This product, 79 bp in length, encoded the amino acid sequence situated between the two primers (underlined in Fig. 2). Since no size prediction could be made for the products of the other PCR reactions, all products were subcloned into plasmid vectors and sequenced from both ends. One product, 300-bp in length, encoded VPV-repetitive sequences. All other fragments were products of false priming; they either encoded completely unrelated sequences or had no open reading frame. The 300 bp product had been generated in a PCR with the downstream primer of peptide *e* (P1), and the upstream primer of peptide *g* (P3) of spot 4 (Fig. 2; see Materials and Methods). Unexpectedly, this fragment did not encode the amino acids downstream of primer P1, but a rather similar sequence. We obtained a similar result in a PCR that was

```
seq. consensus p60    V P V P - - V - V - V p
charge consensus        - + - + - + - + - + -

peptide sequences spot 1

a)          V Q E P V A  V P N P V A V P Q P Y A  V P Q P Y A V Q
b)          D I I X V E  V P V Y V E K R V E V P
c)            D V V V E  K P V I I E R K V E
d)                       D  V P I V V P R Y N
e)                       D  V E R P V P V P F
f)                       D  V P V E R P R Y
g)                       D  V P V Q I P V Q
h)                       D  V P V Q I P I Q
i)                       D  V P I Q R P I
j)                       D  K P I Y V Q I
k)                       D  V P Y V V T R
l)                  D R V  V P V P H
m)                       D  Q P V N R P V T V P K
n)                D D R F  V P R A V
o)              D T P F E  V P V N N P

seq. consensus          V P V - v P
charge consensus        + - + - + - + +

peptide sequences spot 4

a)                  D V A  V P V P V E R Q I P I E  R P V E
b)                     E   I P V P V E R Q V R I E  Q P V Y
c)                     D   V P V P I G R P V P Q P  V
d)              D R P V E  R P V P V E Q I I Q K E  L L I E R P A F
e)              D R Q I E  K P V V I E Q I V E V P  V E
                           R P V Y I E V E N P I
f)              E E V I E  R P V P R P V Y V E Q
g)                     E   R P V P V A V P R E V
h)                     E   R P V P R P V Y V E Q

seq. consensus          R P V P v E v - V E -
charge consensus        +       + - + - + - + -
```

Figure 2. Peptide sequence alignment of *P. dubius* epiplasmic proteins. Epiplasmic polypeptides 1 and 4, as depicted in Fig. 1, were characterized by peptide sequencing (see Materials and Methods). Only peptides from the repeat regions are compiled. Sequences are aligned according to the consensus of the VPV-repeat motif. Peptide *a* of spot 1 was obtained after trypsin digestion; all other peptides were obtained after digestion with endoproteinase Asp-N. The amino acid sequence underlined in peptide *e* of spot 4 has been independently confirmed by nucleotide sequence analysis of the corresponding PCR fragment. The sequence consensus motif and the ordered charge consensus motif of spot 1 and 4 peptides are given below the peptide sequences. The corresponding consensus motifs of p60 are given in the two top lines. Uppercase letters, residue in ≥50% of the repeats; lowercase letters, residue in ≥33% of the repeats; +, majority of charged residues positively charged; −, majority of charged residues negatively charged. Note that the consensus motifs of spot 1 and 4 proteins might be biased since only incomplete sequence information is currently available.

carried out to amplify the cDNA fragment spanning the gap between the two previously isolated PCR products. In this PCR, perfect matching primers (P4, P5) deduced from the nucleotide sequence of these fragments were used. A 500-bp cDNA fragment was amplified that represented a 5′ extension of the 300-bp PCR product. While the 3′ end of this fragment perfectly matched the 5′ end of the 300-bp product, the sequence at its 5′ end was not identical to the sequence of the fragment from which the 5′ primer had been deduced. The whole region encoded VPV repeats, but none of these were identical with peptide sequences of either spot 4 or 1. We therefore assumed that, while the 79-bp PCR product was derived from the cDNA encoding spot 4 polypeptide, the other two PCR products represented cDNA fragments coding for an additional epiplasmic protein. From the analysis of the complete cDNA sequence of this polypeptide we could infer that the cross-hybridization of the PCR primers was due to the high degree of amino acid similarity between these two polypeptides.

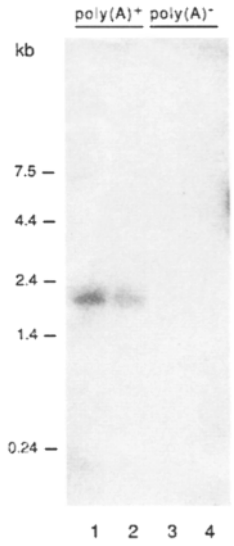

Figure 3. Northern blot analysis of *P. dubius* p60 mRNA. 5 μg (lane *1*) and 2.5 μg (lane *2*) of poly(A)$^+$ RNA, and 10 μg (lane *3*) and 20 μg (lane *4*) of poly(A)$^-$ RNA were electrophoresed through an agarose–formamide gel and transferred to a nitrocellulose membrane as described in Materials and Methods. A single transcript of 2.1 kb was detected with a ^{32}P-labeled 500-bp probe. Positions of RNA standards are given in kilobases at left.

When Northern blots of poly(A$^+$)-enriched RNA were probed with the ^{32}P-labeled 500-bp PCR fragment a single RNA of about 2,100 bp was detected. No hybridization was found with up to eight times the amount of poly(A$^-$) RNA, indicating that the corresponding mRNA is in the polyadenylated fraction (Fig. 3). The 500-bp PCR fragment was used in subsequent cDNA library screening. 1×10^6 plaques of a λZAP cDNA library of oligo dT primed *P. dubius* mRNA were screened under high stringency. Four cross-hybridizing clones were isolated. The largest clone (3-1), contained a cDNA insert of 1832 bp. It has an open reading frame of 1699 bp (566 amino acids) that codes for a VPV-repetitive protein. Its sequence perfectly matched that of the 300- and 500-bp cDNA fragments obtained by PCR. However, the insert of clone 3-1 was shorter than expected from the Northern blot analysis, and more significantly, it did not contain a methionine start codon. The first methionine of clone 3-1 is situated COOH-terminal of four VPV repeats, ~200 codons downstream of the 5′ end. To obtain a full-length clone, we carried out a primer extension reaction with poly(A$^+$) RNA, ligation of an anchor oligonucleotide to the 3′ end of the single stranded cDNA, followed by PCR with a cDNA-specific nested primer and a primer complementary to the anchor sequence (see Materials and Methods). We isolated a 397-bp cDNA fragment that extended the existing cDNA by 261 bp. The complete cDNA is 2093 bp long (Fig. 4). This size is in good agreement with the results of Northern blot analysis, indicating that a full-length cDNA clone had been generated. The cDNA contains an open reading frame encoding a polypeptide of 567 amino acid residues. The methionine initiation codon is embedded in a sequence favorable for translational initiation (Kozak, 1989). It is preceded by four in-frame stop codons.

Codon usage in some ciliates like *Paramecium*, *Tetrahymena*, *Stylonychia*, and *Oxytricha*, differs from the universal genetic code by translating either UAA or UAG as glutamine and by the use of UGA as the sole termination codon (Caron and Meyer, 1985; Helftenbein, 1985; Kuchino et al., 1985; Preer et al., 1985; Hanyu et al., 1986). In other ciliates, like *Euplotes*, UGA is translated as cysteine, and UAA functions as termination codon (Cohen et al., 1990; Kaufmann et al., 1992). Our analysis of the cDNA encoding *P. dubius* p60 shows no direct evidence that an unusual genetic code is used. If the UAG codon at the end of the open reading frame would code for glutamine, as in other ciliates, the polypeptide would be extended by 16 residues. UAA would then function as termination codon, since it is the only in-frame stop codon present in the 3′ untranslated region of the p60 mRNA.

The epiplasm is part of the membrane skeleton of protists. In ciliates like *P. dubius*, it is in close contact with the inner membrane of the alveolus, while in flagellates like *Euglena*, which lack alveoli, it underlies directly the plasma membrane. In *Euglena*, articulins interact with the membrane by noncovalent binding to a major integral plasma membrane protein (Rosiere et al., 1990). For ciliates, it is not known how the corresponding interaction with membranes occurs. We therefore searched for potential membrane-spanning domains in the sequence of p60. The NH_2-terminal domain of p60 contains a segment of predominantly hydrophobic residues (positions 37–74) with hydropathic indices ≥1.5. However, this region does not display the typical features of a transmembrane domain. It is nearly twice as long as a typical membrane-spanning domain, and it is not flanked by charged amino acids. One might therefore assume that p60 does not directly interact with membranes.

When the full-length synthetic RNA is translated in vitro in a rabbit reticulocyte lysate in the presence of [^{35}S]methionine, a single major polypeptide with an apparent molecular mass of ~60,000 is detected by fluorography (Fig. 5, lane *1*). This is in good agreement with the molecular mass of 60,998, calculated from the predicted protein sequence. The molecular mass of the polypeptide (p60) encoded by this cDNA is, however, much lower than the 78–80 kD determined for the major epiplasmic polypeptides of spots 1 and 4 by SDS-PAGE (Peck et al., 1991). To decide whether this discrepancy was due to differences in the gel systems used in the two investigations, epiplasmic polypeptides and in vitro synthesized p60 were separated on the same gel. The epiplasmic 78–80-kD group is separated into two closely spaced bands on 7.5% SDS-PAGE with apparent masses of 76,500 and 80,000 (Fig. 5, lane *2*) well separated from the in vitro synthesized p60 protein (Fig. 5, lane *1*).

We have also expressed p60 in *E. coli.* For this purpose, the complete coding region was cloned into a bacterial expression vector (pINDU; Bujard et al., 1987). The resulting fusion protein contained the 567 amino acids of p60, as well as four vector derived amino acids (MRGS) at the NH_2 terminus. The NH_2-terminal sequence of the fusion protein was confirmed by protein sequencing (results not shown). A total-cell lysate of bacteria transformed with pINDU-p60 (Fig. 5, lane *3*) and, as a control, a lysate of bacteria transformed with pINDU carrying an unrelated and smaller cDNA insert (Fig. 5, lane *4*) were separated on the same gel as the epiplasmic polypeptides and the in vitro synthesized p60. Lysates of the pINDU–p60-transformed bacteria show a prominent band at 60 kd that is absent in the control lysate. This band migrates at the same position as the protein synthesized in the reticulocyte lysate (Fig. 5, compare lanes *3* and *1*).

A rabbit polyclonal antiserum against total epiplasm

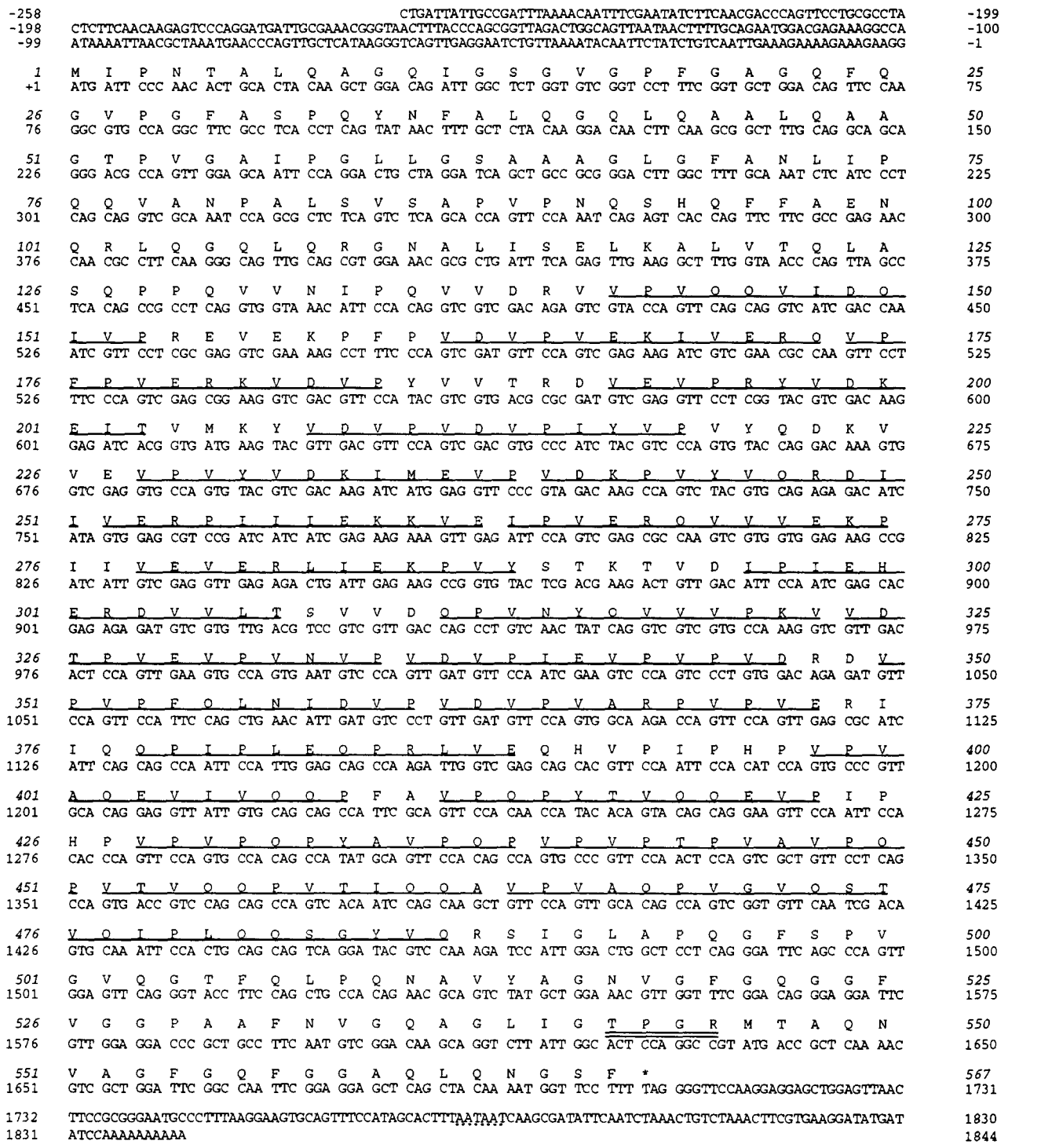

Figure 4. cDNA sequence and deduced amino acid sequence of *P. dubius* p60. Amino acid residues are represented by single letters above the first base of each codon. Numbering of amino acid residues is in italics. Numbering of nucleotides starts with +1 at the ATG initiation codon. The stop codon TAG is designated by an asterisk. The 24 VPV-repeated motifs of the core domain are underlined. A putative $p34^{cdc2}$ protein kinase consensus site (*TPGR*) in the COOH-terminal domain is double underlined. A putative polyadenylation signal (*AATAAT*) is designated by a dotted line. These sequence data are available from GenBank under accession number L41557.

(antiserum 018; Peck et al., 1991) was used to characterize further the bacterially expressed p60. This antiserum had previously been shown to react with epiplasmic proteins, but not with other cytoskeletal elements. On Western blots it labels the major components at 78–80 kD, as well as some quantitatively minor bands (Fig. 5, lane *7*; see also Peck et al., 1991). The antiserum recognizes p60 in total bacterial cell lysates (Fig. 5, lane *6*). The specificity of the reaction is demonstrated by the complete absence of staining of the control lysate (Fig. 5, lane *5*). Taken together,

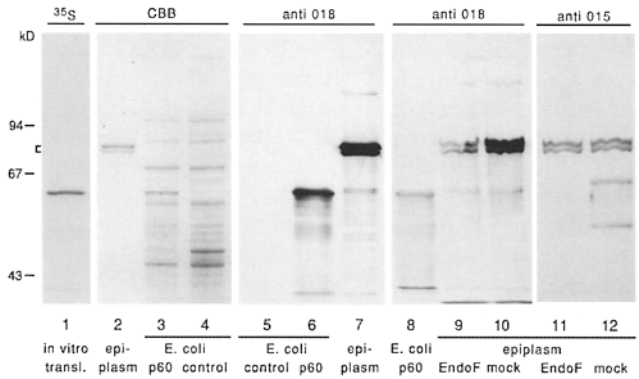

Figure 5. Characterization of in vitro translated and bacterially expressed *P. dubius* articulin p60. p60 translated in a coupled in vitro transcription–translation system in the presence of [^{35}S]methionine (lane *1*), polypeptides of an epiplasmic isolate of *P. dubius* (lanes *2, 7, 9–12*), total cell lysate of *E. coli* BL 21 expressing *P. dubius* p60 (lanes *3, 6,* and *8*), total cell lysate of *E. coli* BL 21 expressing a control plasmid (see Materials and Methods) (lanes *4* and *5*) were separated on SDS-PAGE (7.5%). In vitro synthesized polypeptides were detected by fluorography (lane *1*). Polypeptides in lanes *2–4* were stained with Coomassie brilliant blue, and polypeptides in lanes *5–7* were electrophoretically transferred to nitrocellulose membrane and probed and detected with antiserum 018 as described in Materials and Methods. Samples shown in lanes *1–7* were separated on the same gel. Note that the recombinant p60 is detected by serum 018 (lane *6*). p60 is a minor component in isolated epiplasm (lane *7*). Lanes *8–12*: gel electrophoretic comparison of bacterially expressed p60 (lane *8*) with epiplasmic polypeptides of *P. dubius* deglycosylated with *N*-glycosidase F (lanes *9* and *11*), and mock-incubated epiplasmic polypeptides (lanes *10* and *12*). Polypepides were separated on SDS-PAGE (7.5%), electrophoretically transferred to nitrocellulose membrane, and probed and detected with either antiserum 018 (lanes *8–10*) or antiserum 015 (lanes *11* and *12*). Serum 018 recognizes protein epitopes of epiplasmic polypeptides, while serum 015 recognizes protein, as well as carbohydrate epitopes. The absence of immunoreactivity with serum 015 of the lower mol wt bands in *N*-glycosidase F–treated samples (compare lanes *11* and *12*) demonstrates the efficiency of the enzymatic deglycosylation. Note that the electrophoretic mobility of p60 from isolated epiplasm remains unaltered upon *N*-glycosidase F treatment. Positions of molecular size standards in kilodaltons are given at the left.

we have isolated a cDNA that encodes an epiplasmic protein of 60 kD. p60 shows sequence similarity to and immunological cross-reactivity with the major epiplasmic proteins of 78–80 kD. Antiserum 018 is specific for epiplasmic polypeptides. It does not react with cytosolic components (Peck et al., 1991). Since no prominent bands are found in the molecular weight range of 60,000 in Coomassie-stained gels of epiplasmic, as well as total cytoskeletal preparations, p60 must represent a minor component of the epiplasm of *P. dubius* relative to the 78- and 80-kD polypeptides. A possible candidate polypeptide for p60 is detected on immunoblots of epiplasmic proteins probed with antiserum 018. It has a similar electrophoretic mobility as bacterially expressed p60 (Fig. 5, lanes *6* and *7*). Its electrophoretic mobility remains unaltered after treatment with *N*-glycosidase F (Fig. 5, lanes *9* and *10*). Therefore, in contrast to several other minor epiplasmic polypeptides in this size range it is not glycosylated. *N*-glycosidase F, which cleaves a broad range of asparagine-bound *N*-glycans, has been shown to effectively deglycosylate several of the minor epiplasmic polypeptides (Huttenlauch and Peck, 1991). The effectiveness of the enzyme has been monitored by the loss of reactivity with antiserum 015, which recognizes the glycomoieties of these proteins (Fig. 5, compare lanes *11* and *12*).

P. dubius p60 Is an Articulin

The protein encoded by the cDNA described here has a molecular mass of 60,998 and a calculated isoelectric point of ~4.8. It contains an unusually high number of valine and proline residues (valine ~20%, proline ~14%; Table I). In the central core domain, which comprises about two-thirds of the total length of the polypeptide, valine plus proline accounts for more than 45% of the residues (Table I). The core domain consists of 24 tandem repeats of 12 residues that are occasionally separated by two to eight residue long linkers. The repeats show an alternating pattern of valine and proline residues (VPV repeats) (Fig. 6). Strikingly, these repeats resemble those of *Euglena* articulins p80 and p86 (Marrs and Bouck, 1992) and thus *P. dubius* p60 represents a ciliate homologue of the flagellate articulins (see below).

The consensus sequences for the 12 residue repeats of the core domain are VPVP--V-V-Vp (*P. dubius* p60), VPVPVeviV-v- (*E. gracilis* p80), and VPVPyp/eV-v/k-Ve (*E. gracilis* p86; Marrs and Bouck, 1992), respectively. Uppercase letters represent occupancy in >50% of the repeats, while lowercase letters indicate >33% (see Fig. 6 for details). The sequences are compatible with a series of alternating V and P residues, which is relatively strict for p60, and somewhat more relaxed for few positions of the *E. gracilis* proteins. V positions never show a proline, while a few P positions can display a valine. The high num-

Table I. Amino Acid Composition of p60 Domains

	Total	NH_2-terminal head	Core	COOH-terminal tail
Val	20.28	8.51	27.75	8.75
Pro	13.93	9.22	17.63	6.25
Gln	11.29	14.89	9.54	12.50
Gly	6.88	12.06	0.58	25.00
Ala	6.53	14.89	2.02	11.25
Ile	6.17	4.26	7.80	2.50
Glu	4.94	1.42	7.51	0.00
Leu	4.41	10.64	1.73	5.00
Asp	3.70	0.71	5.78	0.00
Arg	3.35	2.13	4.05	2.50
Phe	3.35	4.96	1.16	10.00
Thr	3.00	2.13	3.18	3.75
Asn	2.82	5.67	0.87	6.25
Lys	2.65	0.71	4.05	0.00
Ser	2.65	5.67	1.16	3.75
Tyr	2.47	0.71	3.47	1.25
His	0.88	0.71	1.16	0.00
Met	0.71	0.71	0.58	1.25
*				

NH_2-terminal domain, residues 1–141; core domain, residues 142–487; COOH-terminal domain, residues 488–567. The frequency of amino acids is given in percent.
*p60 does not contain cysteine and tryptophan residues.

***Pseudomicrothorax dubius* articulin p60** **24 repeats**

1	V	P	V	Q	Q	V	I	D	Q	I	V	P	R	E	V	E	K	P	F	F
2	V	D	V	P	V	E	K	I	V	E	R	Q								
3	V	P	F	P	V	E	R	K	V	D	V	P	Y	V	V	T	R	D		
4	V	E	V	P	R	Y	V	D	K	E	I	T	V	M	K	Y				
5	V	D	V	P	V	D	V	P	I	Y	V	P	V	Y	Q	D	K	V	V	E
6	V	P	V	Y	V	D	K	I	M	E	V	P								
7	V	D	K	P	V	Y	V	Q	R	D	I	I								
8	V	E	R	P	I	I	I	E	K	K	V	E								
9	I	P	V	E	R	Q	V	V	V	E	K	P	I	I						
10	V	E	V	E	R	L	I	E	K	P	V	Y	S	T	K	T	V	D		
11	I	P	I	E	H	E	R	D	V	V	L	T	S	V	V	D				
12	Q	P	V	N	Y	Q	V	V	V	P	K	V								
13	V	D	T	P	V	E	V	P	V	N	V	P								
14	V	D	V	P	I	E	V	P	V	P	V	D	R	D						
15	V	P	V	P	F	Q	L	N	I	D	V	P								
16	V	D	V	P	V	A	R	P	V	P	V	E	R	I	I	Q				
17	Q	P	I	P	L	E	Q	P	R	L	V	E	Q	H	V	P	I	P	H	P
18	V	P	V	A	Q	E	V	I	V	Q	Q	P	F	A						
19	V	P	Q	P	Y	T	V	Q	Q	E	V	P	I	P	H	P				
20	V	P	V	P	Q	P	Y	A	V	P	Q	P								
21	V	P	V	P	T	P	V	A	V	P	Q	P								
22	V	T	V	Q	Q	P	V	T	I	Q	Q	A								
23	V	P	V	A	Q	P	V	G	V	Q	S	T								
24	V	Q	I	P	L	Q	Q	S	G	Y	V	Q								
	1		3		5		7		9		11									
sequence consensus	V	P	V	P	-	-	V	-	V	-	V	p								
V	20	0	16	0	7	1	12	2	12	1	13	1								
P	0	13	0	15	0	4	0	5	0	6	0	11								
-	0	9	0	3	0	9	0	5	0	8	0	4								
+	0	0	2	0	3	0	5	1	5	1	3	0								
charge consensus		-	+	-	+	-	+	-	+	-	+	-								

***Euglena gracilis* articulin p80** **33 repeats**

	1		3		5		7		9		11	
sequence consensus	V	P	V	P	V	e	v	i	V	-	v	-
V	27	0	25	0	20	0	13	3	23	3	11	2
P	0	24	0	30	0	6	0	2	0	7	0	1
-	1	1	0	2	0	18	0	5	1	8	2	7
+	1	0	0	0	2	0	11	2	6	0	16	4
charge consensus				-	+	-	+	-	+	-	+	-

***Euglena gracilis* articulin p86** **33 repeats**

	1		3		5		7		9		11	
sequence consensus	V	P	V	P	y	p/e	V	-	v/k	-	V	e
V	25	0	26	1	9	1	22	3	15	8	19	5
P	0	18	0	30	0	12	0	5	0	3	0	0
-	0	4	2	0	0	13	0	6	0	6	3	20
+	5	4	2	0	2	0	5	3	14	4	5	3
charge consensus	+				+	-	+	-	+	-	+	-

articulin repeat motif

sequence	V	P	V	P	-	-	V	-	V	-	V	-
charge				-	+	-	+	-	+	-	+	-

Figure 6. Sequence alignment of VPV-repeat motifs of *P. dubius* articulin p60 and *E. gracilis* p80 and p86. Amino acid sequences were deduced from the nucleotide sequence of the corresponding cDNA (see Fig. 4). Residues between two adjacent repeats, when present, are given at the right. Data for *E. gracilis* articulins were taken from Marrs and Bouck (1992). The sequence consensus is given below the alignment. Uppercase letters, residue in ≥50% of the repeats; lowercase letters, residue in ≥33% of the repeats. The number of valine (*V*) and proline (*P*), as well as positively (+) and negatively (−) charged residues for each column is recorded. A charge consensus is given in the last row. +, majority of charged residues positively charged; −, majority of charged residues negatively charged. A consensus for all three articulins is given at the bottom.

ber of tyrosines in position 5 and isoleucines in position 8 are a peculiarity of *E. gracilis* p86 and p80, respectively. Protein-specific individualities are also seen in the periodicity of the charge distributions. In the *P. dubius* p60 protein, there is a striking segregation of charged residues. Except for position 1, which has no charged residue, the P positions show practically only negative charges, while the V positions harbor the positive charges. This segregation holds in *E. gracilis* p80 only for positions 4 to 12, and in *E. gracilis* p86 for positions 1, 6, 7, 9, and 12 (Fig. 6).

The NH_2- and COOH-terminal domains flanking the central core show quite different amino acid sequences. Both domains are rich in glycine and alanine, as well as phenylalanine. These three amino acids together comprise >45% of all residues of the COOH-terminal domain (Table I). More generally, with the exception of glutamine, that is found with similar frequency in all three domains, as well as those amino acids that occur with low overall frequency, the abundance of most amino acids shows a characteristic segregation between the central core domain and the two terminal domains (Table I). The terminal domains contain few charged residues (NH_2-terminal domain, 7 out of 141 residues; COOH-terminal domain, 2 out of 80 residues). Moreover, the first 100 residues of the NH_2-terminal domain are completely devoid of charged residues (Fig. 4). Also, in the central repeat domain, a segregation of charged residues is found: the last six repeats lack charged amino acids, while several repeats in the first two-thirds of the central domain are particularly rich in acidic and basic amino acids (Fig. 6).

The COOH-terminal domain contains the motif Thr-Pro-Gly-Arg at positions 542 to 545. This sequence conforms to the proposed consensus site for phosphorylation by $p34^{cdc2}$ protein kinase (Moreno and Nurse, 1990). Phosphorylation of *P. dubius* epiplasmic proteins has not yet been analyzed. However, other cytoskeletal proteins of the ciliate cortex, i.e., the polypeptides that constitute the ciliary rootlets (kinetodesmal fibres) show hyperphosphorylation during their disassembly in the course of cell division (Sperling et al., 1991). The role of phosphorylation in mitotic disassembly and reassembly of nuclear lamins has been analyzed in detail in vertebrate cells (Heald and McKeon, 1990; Peter et al., 1990, 1991). Here, $p34^{cdc2}$ is involved directly in the mitotic disassembly of the nuclear lamina.

Spectroscopic and Electron Microscopic Analysis of Recombinant p60

Secondary structure analysis of the core domain using the method of Chou and Fasman (1974) predicts many short stretches of β-sheets for this region. The FTIR spectrum of p60 has a pronounced maximum at 1,641 cm^{-1}, indicative of random coil structure. A second maximum is situated at 1,616 cm^{-1}, a region that is typical for β-sheet or extended chain bands (Fig. 7 *A*). A detailed analysis yields ~50% random coil, and ~28% β-/extended-chain contributions (Table II). The overall spectral shape can be compared with random-coil polypeptides such as polylysine at neutral pH. Low wavenumber β-strand bands of high intensity have predominantly been observed in denatured proteins which aggregate and have been attributed to intermolecular interactions.

Electron microscopic analysis of purified recombinant p60 reveals rod-shaped, short filamentous structures, with diameters of ~15–20 nm (Fig. 8 *A*). Their length distribution is heterogeneous, ranging from 20 nm for single round "subunits" to about 120 nm (Fig. 8 *A*, *inset*). The longer filaments seem to be formed by the smaller subunits in longitudinal array. In addition, a tendency for the formation of larger structures is noticed. These are about 25–30-nm-thick filaments with lengths in the μm range (Fig. 8 *A*),

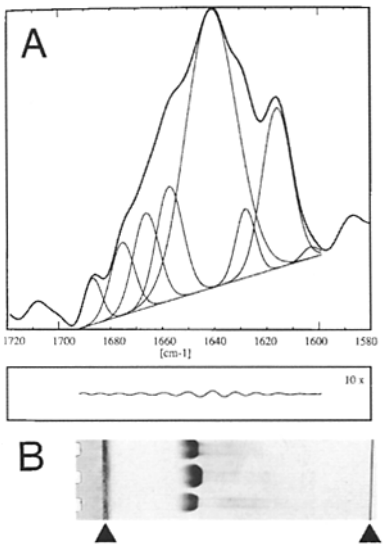

Figure 7. FTIR spectrum of recombinant *P. dubius* articulin p60. *A* shows the deconvoluted spectrum in the amide I region. It also includes the fitted bands. Abscissa is given as wavenumber. The quality of the fitting is shown in the box below *A*, which gives the difference between the deconvoluted curve and the sum of the fitted bands in a 10-fold amplification. For assignments of the fitted bands see Table II. *B*: SDS-PAGE (12%) of recombinant *P. dubius* articulin p60 after purification. The gel was run from left to right; arrowheads mark top and bottom of the separation gel. The protein fraction shown here was used to record the FTIR spectrum (this figure), as well as the electron micrographs shown in Fig. 8.

and broader, large aggregates which often appear to be tubes. The latter often branch, forming interconnected, wavy networks (Fig. 8 *B*). The presence of divalent cations does not seem to change significantly their appearance, but the structures look smoother and straighter. To exclude the possibility that the larger structures formed on the grid during the staining procedure, samples were also centrifuged for 15 min at 20,000 *g* before staining. About 25% of the protein was pelleted. Amino-terminal sequencing of pellet and supernatant provided the same sequence. The supernatant showed images like those given in Fig. 8 *A* inset.

Table II. Secondary Structure Assignment of Fitted FTIR Bands of Purified, Recombinant P. dubius Articulin p60

Band	Position	Percent area	Secondary structure feature
	cm^{-1}		
1	1,641.2	49.4	Random coil
2	1,616.0	16.7	β
3	1,628.1	4.7	β
4	1,675.6	6.1	β
5	1,666.7	8.6	β-turn*
6	1,687.5	2.9	β-turn
7	1,657.6	10.9	α

The β-sheet bands (2–4) amount to ~30%.
*The band at 1666.7 cm^{-1} could not unambiguously be assigned.

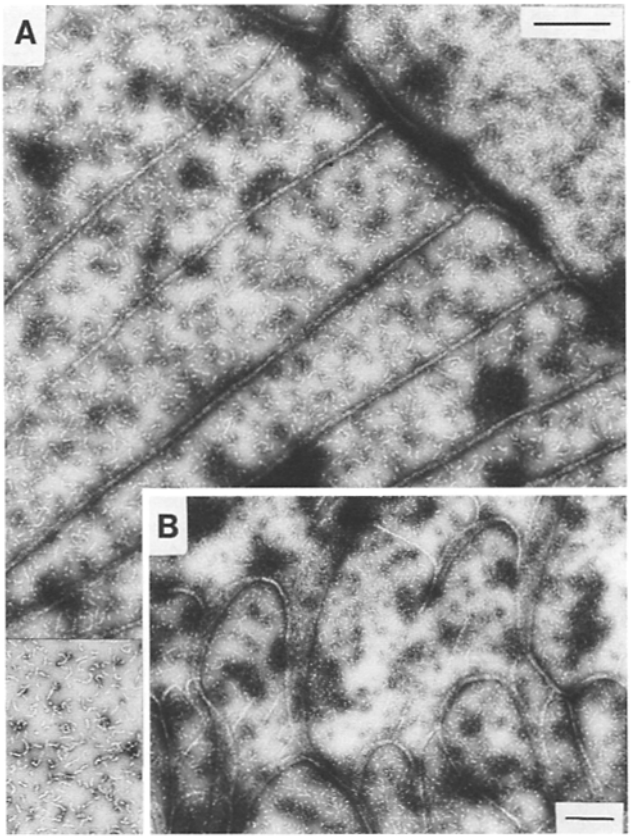

Figure 8. Electron microscopy of recombinant *P. dubius* p60. Structures obtained after dialysis of purified, recombinant articulin p60 against renaturation buffer in the presence of 1 mM $MgCl_2$ are shown. The smaller types of structures are densely and uniformly scattered over the whole area in *A* and *B*. They are thin filaments of variable length (see inset; *A*). The larger structures shown in *A* are thicker and rather straight filaments. *B* presents another type of larger structures: they consist of broader forms of aggregations, often, probably, tubes, which interconnect to arch-shaped superstructures. Bars, 0.4 μm.

Discussion

Here, we report sequence results on epiplasmic proteins of the ciliated protist *Pseudomicrothorax dubius*. The sequence of a complete cDNA encoding a polypeptide of 60 kD (p60), that represents a quantitatively minor component of the epiplasm, as well as partial peptide sequences of two distinct major components with molecular masses of 78,000–80,000 are presented. Previous studies (Peck et al., 1991) demonstrated that the 78–80-kD group of polypeptides can be separated into several components by two-dimensional-gel electrophoresis. Based on peptide map analysis, several of these components represent isoelectric variants of the same polypeptides (spots 1, 2, and 3; see Fig. 1), while others seem to reflect products of separate genes (spots 1 and 4; see Fig. 1). Our sequence data directly confirm this view. From the apparent molecular masses of the two major polypeptides, we calculate that about one-third of their sequences are covered by our peptide sequence analysis. Although the proteins are clearly related, none of the peptide sequences obtained from the

two components are identical. This strongly suggests that the two components (spot 1 and 4) are indeed the products of two distinct genes.

Epiplasmic Proteins of P. dubius Are Characterized by VPV-repetitive Motifs

The most striking finding of this study concerns the identification of a 12-amino acid-long repeat sequence in all three epiplasmic proteins. This motif resembles that of *Euglena* articulins (Marrs and Bouck, 1992). It is tandemly repeated 24 times in p60. Comparison of the peptide sequences obtained for the two major components with the complete amino acid sequence of p60 indicates that all three *P. dubius* proteins have a similar molecular architecture.

Secondary structural prediction rules of the sequence in the central domain with its 24 VPV repeats indicates a high content of β-sheet, and FTIR spectroscopical measurements of the bacterially expressed protein show that p60 contains ~30% β-sheet. Comparison of the repeats of all three *P. dubius* polypeptides reveals a remarkable distribution of charged amino acids. Positively and negatively charged residues show an alternating distribution in register with the pattern of valine and proline residues. In a β-sheet structure, this distribution would result in the separation of negatively and positively charged residues into two layers above and below the polypeptide backbone, which could have an important function in the structural organization of the epiplasmic proteins. Although the VPV-repeat consensus of spots 1 and 4 polypeptides is still based on incomplete sequence information, and therefore might be biased, the 24 repeats of p60 are fully characterized by the complete cDNA.

Electron Microscopical Observations

The epiplasm of *P. dubius* appears homogeneous in electron microscopical ultrathin sections. Reconstitution experiments with total epiplasmic preparations have failed to detect filamentous structures (Peck, 1977). The electron microscopical appearance of recombinant p60 shows that the majority of the molecules are found in elongated or short filamentous forms. These have a uniform diameter (~15–20 nm), but variable length distribution. They seem to be constructed of smaller roundish units with a diameter similar to the thickness of the short filaments. The small polymers have a tendency to form thicker and longer filaments, ribbons or tubes, and large sheets. p60, however, is a minor component of the epiplasm. It remains to be seen whether the major articulins form similar structures in vitro, and how these structures are related to the epiplasmic organization in vivo.

Epiplasmic Proteins of Flagellates and Ciliates are Members of the Same Multigene Family, the Articulins

The sequences of two epiplasmic proteins of the flagellate *Euglena gracilis* have recently been described. These proteins were named articulins (Marrs and Bouck, 1992). Articulins and the ciliate polypeptide p60 described in this study show the same domain organization. All three ciliate proteins contain VPV repeats that match the consensus sequence of the articulin repeats. The complete sequence of an articulin from the ciliate *P. dubius* extends the results obtained from the two articulin sequences of the flagellate *E. gracilis*. Articulins have a tripartite sequence organization. A central core domain of repetitive 12 residue motifs starting with the consensus sequence VPVP . . . is the hallmark of this novel protein family (Marrs and Bouck, 1992). So far the number of repeats is either 33 (*E. gracilis* p80 and p86) or 24 (*P. dubius* p60). When present, the linkers between the repeats are maximally six to eight residues long. The terminal domains flanking the core differ in sequence and length. The NH_2-terminal head domain harbors 76 (*E. gracilis* p80), 112 (*E. gracilis* p86), or 141 residues (*P. dubius* p60). The COOH-terminal tail domains are 67 (*P. dubius* p60), 110 (*E. gracilis* p86), or 149 residues (*E. gracilis* p80) long. Interestingly, the two terminal domains of each articulin show some similarity. They all contain a long segment free of charged residues. In the head domain, these segments cover 39 (*E. gracilis* p80), 62 (*E. gracilis* p86), or even 93 residues (*P. dubius* p60), while the tail domains display such segments with 33 (*E. gracilis* p80) or 55 continuous residues (*P. dubius* p60). An additional argument for a structural homology between heads and tails is, so far, restricted to the *Euglena* articulins. A heptapeptide around the sequence APVTYGA is found four times either in the head (*E. gracilis* p86) or in the tail domain (*E. gracilis* p80). The length variability of both terminal domains and the core domain opens the possibility that articulins can strongly vary in polypeptide molecular weight (see also below). Since the articulins preserve sequence principles rather than actual sequences the overall sequence identities are relatively low. The two *E. gracilis* articulins share only 37% identity (Marrs and Bouck, 1992), and the *P. dubius* articulin is even more remote with <20% identity. Despite these low identity values, the conservation of the sequence principles, as well as the overall structure of the polypeptides from *P. dubius* and *E. gracilis* is in favor of common ancestry, rather than convergent evolution. In addition, protein data bank search did not reveal any significant similarity to other proteins.

In conclusion, the epiplasmic proteins of two distantly related groups of protists, the flagellates and the ciliates, which are viewed as representatives of different phyla, are members of the same protein family (for discussion of protist phylogeny see Baroin et al., 1988; Baroin-Tourancheau et al., 1992; Lee and Kugrens, 1992; Cavalier-Smith, 1993).

Articulins Might Be the Major Constituents of the Epiplasm of Various Protists

An epiplasmic layer has been documented for many ciliate genera and several other protists including dinoflagellates and flagellates (for review see Grain, 1986). In all these organisms, the epiplasm shows a strikingly similar appearance and is always in close contact with cortical membranes, either the plasma membrane or the inner alveolar membrane. This apparent similarity has raised the question of whether the epiplasmic proteins in different protists are homologous. Biochemical analysis of a variety of species revealed that the epiplasm is composed of a multitude of polypeptides with a great diversity in molecular masses (Dubreuil and Bouck, 1985, 1988; Williams, 1986;

Viguès et al., 1987; Peck et al., 1991; Nahon et al., 1993). Interestingly, heterogeneity in apparent molecular mass is even seen among closely related species (Williams, 1986). However, immunological studies with polyclonal as well as monoclonal antibodies indicate the presence of common epitopes in epiplasmic polypeptides of different molecular masses within a particular species, as well as between different protists (Viguès et al., 1987; Nahon et al., 1993; Curtenaz et al., 1994). Polyclonal antisera raised against epiplasmic proteins of distantly related protists, the ciliate *P. dubius*, the flagellate *Euglena acutus*, and the dinoflagellate *Noctiluca scintillans* (Viguès et al., 1987), showed cross-reaction of the major epiplasmic polypeptides of all three species. While the epiplasmic proteins of *P. dubius* and *Euglena* have similar ranges of apparent molecular masses (11–86 kD), polypeptides of much higher apparent molecular masses are found in *Noctiluca* (45–220 kD; Viguès et al., 1987). With respect to the cDNA characterization of *P. dubius* articulin p60, it is interesting to note that sera raised against the euglenoid and the dinoflagellate proteins both recognize a quantitatively minor component of ~60 kD in epiplasmic preparations of *Pseudomicrothorax*. Interspecies cross-reactivity has also been observed with some mAbs (Nahon et al., 1993; Curtenaz et al., 1994). However, as often observed, such immunological cross-reactions can display unpredictable patterns with respect to the phylogenetic distances separating the organisms. Thus an mAb against *Paramecium* epiplasmic proteins cross-reacts with major epiplasmic polypeptides of *Pseudomicrothorax* (articulins p78 and p80), *Euplotes*, and *Euglena acus*, but fails to recognize the corresponding polypeptides in much more closely related species such as *Tetrahymena* and *Colpidium*.

The finding that ciliate and flagellate epiplasmic proteins are members of the same protein family suggests that the repetitive VPV motifs of articulins might represent the common epitopes that are recognized by several of the antibodies directed against epiplasmic proteins, a hypothesis that now can be tested by using peptide sequences or recombinant articulins to produce bona fide articulin antibodies. It further could imply that the broad interspecies cross-reactivity of these antibodies indeed reflects common ancestry of epiplasmic proteins. We therefore assume that VPV proteins (articulins) are major epiplasmic constituents of most, if not all, protists that contain this structure. Given the length variability of both terminal domains and the core domain of flagellate and ciliate articulins, it seems likely that the large epiplasmic proteins of the dinoflagellate *Noctiluca scintillans*, which show immunological cross-reactivity with the epiplasmic proteins of *Euglena acus* and *P. dubius* (Viguès et al., 1987) are most likely also articulins. Thus, the modular structure of the articulins could explain the diversity of molecular masses of epiplasmic proteins observed in different protists. This type of diversification is not without precedence. Intermediate filament proteins, one of the major classes of cytoskeletal proteins of metazoans, are only one example of how a modular architecture has been used to generate protein diversity while maintaining structural similarities (for review see Fuchs and Weber, 1994). The epiplasm is a very elaborate structure with numerous ridges and depressions in some protists, but it may be much simpler in others. Its structural organization is thus characteristic for each species examined so far. On the other hand, it is always associated with a cortical membrane along its external face, and cortical fiber systems and other organelles, e.g., mitochondria and secretory granules, are anchored to its internal face. The articulins, with their conserved core domain and more variable terminal domains, appear to be well adapted to a role of assuring functions common to all epiplasms, yet permitting the epiplasmic structural diversity particular to each species. The demonstration that epiplasmic proteins of distantly related protists share common molecular principles will help to clarify the roles of core and terminal domains, as well as to define the limits of diversity within this novel family of skeletal proteins.

We thank Thomas Heimburg (Max Planck Institute [MPI] for Biophysical Chemistry, Göttingen) for the FTIR analysis. The excellent technical assistance of Jürgen Schünemann is greatly acknowledged. I. Huttenlauch would like to thank Herbert Jäckle (MPI for Biophysical Chemistry, Göttingen) for hospitality during the initial phase of this work.

I. Huttenlauch acknowledges the support of the Deutsche Forschungsgemeinschaft (Hu 549/1-1 and 549/1-2).

Received for publication 9 March 1995 and in revised form 18 April 1995.

References

Aufderheide, K. 1983. Mitochondrial associations with specific components of the cortex of *Tetrahymena thermophila*: microtubules and the epiplasm. *J. Protozool.* 30:457–460.

Aufderheide, K., J. Frankel, and N. E. Williams. 1980. Formation and positioning of surface-related structures in protozoa. *Microbiol. Rev.* 44:252–302.

Baroin, A., R. Perasso, L. Q. Qu, J. P. Bachellerie, and A. Adoutte. 1988. Partial phylogeny of the unicellular eukaryotes based on rapid sequencing of 28 S ribosomal RNA. *Proc. Natl. Acad. Sci. USA.* 85:3474–3478.

Baroin-Tourancheau, A., P. Delgado, R. Perasso, and A. Adoutte. 1992. A broad molecular phylogeny of ciliates: identification of major evolutionary trends and radiations within the phylum. *Proc. Natl. Acad. Sci. USA.* 89: 9764–9768.

Bennet, V. 1990. Spectrin-based membrane skeleton: a multipotential adaptor between plasma membrane and cytoplasm. *Physiol. Rev.* 70:1029–1065.

Bricheux, G., and G. Brugerolle. 1986. The membrane cytoskeleton complex of euglenoids. I. Biochemical and immunological characterization of the epiplasmic proteins of *Euglena acus*. *Eur. J. Cell Biol.* 40:150–159.

Bricheux, G., and G. Brugerolle. 1987. The pellicular complex of euglenoids. II. A biochemical and immunological comparative study of major epiplasmic proteins. *Protoplasma.* 140:43–54.

Bujard, H., R. Gentz, M. Lanzer, D. Stueber, M. Mueller, I. Ibrahimi, M.-T. Haeuptle, and B. Dobberstein. 1987. A T5 promoter-based transcription–translation system for the analysis of proteins in vitro and in vivo. *Methods Enzymol.* 155:416–433.

Byler, D. M., and H. Susi. 1986. Examination of the secondary structure of proteins by deconvoluted FTIR spectra. *Biopolymers.* 25:469–487.

Cachon, J., M. Cachon, A. Boillot, and D. L. Brown. 1987. Cytoskeletal and myonemal structures of Dinoflagellates are made of 2–3-nm filaments. *Cell Motil.* 7:325–336.

Caron, F., and E. Mayer. 1985. Does *Paramecium primaurelia* use a different genetic code in its macronucleus? *Nature (Lond.).* 314:185–188.

Cavalier-Smith, T. 1993. Kingdom protozoa and its 18 phyla. *Microbiol. Rev.* 57:953–994.

Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162:156–159.

Chou, P. Y., and G. D. Fasman. 1974. Prediction of protein conformation. *Biochemistry.* 13:222–245.

Cohen, J., and J. Beisson. 1988. The cytoskeleton. *In* Paramecium. H.-D. Görtz, editor. Springer-Verlag New York Inc., New York. 363–392.

Cohen, J., P. Dupuis, and B. Viguès. 1990. Expression of a ciliate gene in *Escherichia coli* using a suppressor tRNA to read the UAA and UAG glutamine codons. *J. Mol. Biol.* 216:189–194.

Curtenaz, S., and R. K. Peck. 1992. A monoclonal antibody study of protein distribution in the membrane skeleton of the ciliate *Pseudomicrothorax*. *J. Cell Sci.* 103:1117–1125.

Curtenaz, S., P. Nahon, F. Iftode, and A. Fleury. 1994. Interspecific immunological cross-reactions among cortical proteins of four ciliates. *Eur. J. Protistol.* 30:440–450.

Dubreuil, R. R., and G. B. Bouck. 1985. The membrane skeleton of a unicellular organism consists of bridged, articulating stripes. *J. Cell Biol.* 101:1884–

1896.
Dubreuil, R. R., and G. B. Bouck. 1988. Interrelationships among the plasma membrane, the membrane skeleton and surface form in a unicellular flagellate. *Protoplasma.* 143:150–164.
Frankel, J. 1989. Pattern Formation. Ciliate Studies and Models. Oxford University Press, New York. 19–161.
Fuchs, E., and K. Weber. 1994. Intermediate filaments: structure, dynamics, function and disease. *Annu. Rev. Biochem.* 63:345–382.
Grain, J. 1986. The cytoskeleton in protists: nature, structure and functions. *Int. Rev. Cytol.* 104:153–249.
Hanyu, N., Y. Kuchino, S. Nishimura, and H. Beier. 1986. Dramatic events in ciliate evolution: alteration of UAA and UAG termination codons to glutamine codons due to anticodon mutations in two *Tetrahymena* $tRNAs^{Gln}$. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:1307–1311.
Heald, R., and F. McKeon. 1990. Mutations of phosphorylation sites in lamin A that prevent nuclear lamina disassembly in mitosis. *Cell.* 61:579–589.
Helftenbein, E. 1985. Nucleotide sequence of a macronuclear DNA molecule coding for α-tubulin from the ciliate *Stylonychia lemnae.* Special codon usage: TAA is not a translation termination codon. *Nucleic Acids Res.* 13:415–431.
Huttenlauch, I., and R. K. Peck. 1991.The membrane skeleton of *Pseudomicrothorax.* II. Biochemical and immunological characterization of the glycosylated cytoskeletal proteins. *J. Cell Sci.* 100:707–715.
Kaufmann, J., V. Florian, and A. Klein. 1992. TGA cysteine codons and intron sequences in conserved and nonconserved positions are found in macronuclear RNA polymerase genes of *Euplotes octocarinatus. Nucleic Acids Res.* 20:5985–5989.
Kauppinen, J. K., D. J. Moffat, H. H. Mantsch, and D. G. Cameron. 1981. Fourier self-deconvolution: a method for resolving intrinsically overlapped bands. *Appl. Spectrosc.* 35:271–276.
Kozak, M. 1989. The scanning model for translation: an update. *J. Cell Biol.* 108:229–241.
Kuchino, Y., N. Hanyu, F. Tashiro, and S. Nishimura. 1985. *Tetrahymena thermophila* glutamine tRNA and its gene that corresponds to UAA termination codon. *Proc. Natl. Acad. Sci. USA.* 82:4758–4762.
Lee, R. E., and P. Kugrens. 1992. Relationship between the flagellates and the ciliates. *Microbiol. Rev.* 56:529–542.
Louvard, D. 1989. The function of the major cytoskeletal components of the brush border. *Curr. Opin. Cell Biol.* 1:51–57.
Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.
Marrs, J. A., and G. B. Bouck. 1992. The two major membrane skeletal proteins (articulins) of *Euglena gracilis* define a novel class of cytoskeletal proteins. *J. Cell Biol.* 118:1465–1475.
Meyer, F., H. J. Schmidt, E. Plümper, A. Hasilik, G. Mersmann, H. E. Meyer, A. Engström, and K. Heckmann. 1991. UGA is translated as cysteine in pheromone 3 of *Euplotes octocarinatus. Proc. Natl. Acad. Sci. USA.* 88:3758–3761.
Moreno, S., and P. Nurse. 1990. Substrates for $p34^{cdc2}$: in vivo veritas? *Cell.* 61: 549–551.
Nahon, P., G. Coffe, H. Le Guyader, J. Darmanaden-Delorme, R. Jeanmaire-Wolf, J.-C. Clérot, and A. Adoutte. 1993. Identification of the epiplasmins, a new set of cortical proteins of the membrane skeleton in *Paramecium. J. Cell Sci.* 104:975–990.
Peck, R. K. 1977. The cortical ultrastructure of the somatic cortex of *Pseudomicrothorax dubius*: structure and function of the epiplasm in ciliated protozoa. *J. Cell Sci.* 25:367–385.
Peck, R. K. 1986. The trace element selenium induces trichocyst formation in the ciliated protozoan *Pseudomicrothorax dubius. Eur. J. Cell Biol.* 41:174–181.
Peck, R. K., F. Duborgel, I. Huttenlauch, and G. de Haller. 1991. The membrane skeleton of *Pseudomicrothorax.* I. Isolation, structure, and composition. *J. Cell Sci.* 100:693–706.
Peter, M., J. Nakagawa, M. Dorée, J. C. Labbé, and E. A. Nigg. 1990. In vitro disassembly of the nuclear lamina and M phase–specific phosphorylation of lamins by cdc2 kinase. *Cell.* 61:591–602.
Peter, M., E. Heitlinger, M. Häner, U. Aebi, and E. A. Nigg. 1991. Disassembly of in vitro formed lamin head-to-tail polymers by cdc2 kinase. *EMBO (Eur. Mol. Biol. Organ.) J.* 10:1535–1544.
Preer, J. R., Jr., L. B. Preer, B. M. Rudman, and A. J. Barnett. 1985. Deviation from the universal code shown by the gene for surface protein 51A in *Paramecium. Nature (Lond.).* 314:188–190.
Rosiere, T. K., J. A. Marrs, and G. B. Bouck. 1990. A 39-kD plasma membrane protein (IP39) is an anchor for the unusual membrane skeleton of *Euglena gracilis. J. Cell Biol.* 110:1077–1088.
Schleicher, M., and A. A. Noegel. 1992. Dynamics of the *Dictyostelium* cytoskeleton during chemotaxis. *New Biol.* 4:461–472.
Sperling, L., G. Keryer, F. Ruiz, and J. Beisson. 1991. Cortical morphogenesis in *Paramecium*: a transcellular wave of protein phosphorylation involved in ciliary rootlet disassembly. *Dev. Biol.* 148:205–218.
Stelly, N., J.-P. Mauger, M. Claret, and A. Adoutte. 1991. Cortical alveoli of *Paramecium*: a vast submembranous calcium storage compartment. *J. Cell Biol.* 113:103–112.
Stick, R. 1988. cDNA cloning of the developmentally regulated lamin L_{III} of *Xenopus laevis. EMBO (Eur. Mol. Biol. Organ.) J.* 7:3189–3197.
Stockem, W., and K. Brix. 1994. Analysis of microfilament organization and contractile activities in *Physarum. Int. Rev. Cytol.* 149:145–215.
Ungewickell, E., U. Plessmann, and K. Weber. 1994. Purification of golgi adaptor protein 1 from bovine adrenal gland and characterization of its β_1 (β') subunit by microsequencing. *Eur. J. Biochem.* 222:33–40.
Viguès, B. D., and D. David. 1989. Demonstration of a 58 kd protein as a major component of the plasma membrane skeleton in the ciliate *Entodinium bursa. Protoplasma.* 149:11–19.
Viguès, B., G. Metenier, and J. Sénaud. 1984. The sub-surface cytoskeleton of the ciliate *Polyplastron multivesiculatum*: isolation and major protein components. *Eur. J. Cell Biol.* 35:336–342.
Viguès, B., G. Brichieux, C. Metivier, G. Brugerolle, and R. K. Peck. 1987. Evidence for common epitopes among proteins of the membrane skeleton of a ciliate, an euglenoid and a dinoflagellate. *Eur. J. Protistol.* 23:101–110.
Williams, N. E. 1986. Evolutionary change in cytoskeletal proteins and cell architecture in lower eukaryotes. *Prog. Protistol.* 1:1309–1324.