

This is the peer reviewed version of the following article: Kloetzel, John A.; Aubusson-Fleury, Anne; Butler, Maurice D.; Banerjee, Deben; Mozzicafreddo, Matteo; Sequence and Properties of Cagein, a Coiled-Coil Scaffold Protein Linking Basal Bodies in the Polykinetids of the Ciliate *Euplotes aediculatus*; Eukaryotic Microbiology (2021); <https://onlinelibrary.wiley.com/doi/10.1111/jeu.12850> , which has been published in final form at <https://doi.org/10.1111/jeu.12850>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

DR. JOHN A. KLOETZEL (Orcid ID : 0000-0003-3062-8619)

Article type : Original Article

Kloetzel et al. – Cagein Gene Sequence in *Euplotes*

Sequence and Properties of Cagein, a Coiled-Coil Scaffold Protein Linking Basal Bodies in the Polykinetids of the Ciliate *Euplotes aediculatus*

John A. Kloetzel^{a,1}, Anne Aubusson-Fleury^{b,2}, Maurice D. Butler^{a,3},
Deben Banerjee^{c,4}, and Matteo Mozzicafreddo^d

a Department of Biological Sciences
University of Maryland Baltimore County
Baltimore, Maryland 21250, USA

b Biogenese et fonction des structures centriolaires,
I2BC, Université Paris Saclay,
91190 Gif sur Yvette, France

c New York Blood Center
New York, New York 10021, USA

d Scuola di Bioscienze e Medicina Veterinaria
Università di Camerino

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/JEU.12850](#)

This article is protected by copyright. All rights reserved

¹ Present address: 351 Wiley St., Ashland, Oregon 97520, USA.

² Present address: 1 rue des buis, Venant, 91870 Boissy le Sec, France

³ Present address: 225 E 34th St, New York, New York 10016, USA

⁴ Present address: 915 B Halcyon Ave., Nashville, Tennessee 37204, USA

Correspondence

J. Kloetzel, 351 Wiley St., Ashland, Oregon 97520, USA

Telephone number: +1 541-488-9818; email: kloetzel@umbc.edu

ABSTRACT

In the hypotrich ciliate *Euplotes*, many individual basal bodies are grouped together in tightly packed clusters, forming ventral polykinetids. These groups of basal bodies (which produce compound ciliary organelles such as cirri and oral membranelles) are cross-linked into ordered arrays by scaffold structures known as ‘basal body cages’. The major protein comprising *Euplotes* cages has been previously identified and termed ‘cagein’. Screening a *E. aediculatus* cDNA expression library with anti-cagein antisera identified a DNA insert containing most of a putative cagein gene; standard PCR techniques were used to complete the sequence. Probes designed from this gene identified a macronuclear ‘nano-chromosome’ of ca. 1.5 kb in Southern blots against whole-cell DNA. The protein derived from this sequence (463 residues) is predicted to be hydrophilic and highly charged; however, the native cage structures are highly resistant to salt/detergent extraction. This insolubility could be explained by the coiled-coil regions predicted to extend over much of the length of the derived cagein polypeptide. One frameshift sequence is found within the gene, as well as a short intron. BLAST searches find many ciliates with evident homologues to cagein within their derived genomic sequences.

Key Words

Cilia; ciliate introns; cytoskeleton; cortical morphogenesis; euplotid; frameshift; hypotrich; intermediate filaments; kinetosome; protist.

FUNDAMENTAL structures in the cytoskeleton of many ciliated protists (and potentially other ciliated cell types as well) are frameworks, often referred to as ‘baskets’ or ‘cages’, within which are assembled numerous basal bodies. From the clustered basal bodies inside these frameworks emerge closely-packed cilia that function together in coordinated patterns of beating, enabling the cells to gather food particles, as well as carry out motile functions such as swimming or creeping on substrates. Cirri and so-called oral membranelles are well described ciliary organelles of this type (Adoutte and Fleury 1996; Lynn and Corliss 1991; Soares et al. 2019). A previous study has described in some detail the structure of the ‘basal-body cage’ scaffolds in the ciliate *Euplotes aediculatus* (Kloetzel and Brann 2012). That work also identified one of the major proteins from which this cell’s cages are constructed, a protein we termed ‘cagein’ (with an apparent M_r of 45kD by SDS-PAGE). The work described here has used an antibody generated against cagein to screen a *Euplotes* expression library. The candidate gene fragment identified by this screen was extended and amplified by conventional molecular biology techniques, yielding a full-length gene.

The cagein protein predicted from this gene exhibits a number of characteristics (notably extensive coiled-coil domains) that comport well with the cages’ structural, supportive roles *in vivo*. Database searches reveal that a wide variety of other ciliated protists contain proteins (currently identified only as “unknown” or “hypothetical” predicted proteins, derived from genomic databases) that have a high degree of homology with the sequence of cagein. These so-far unidentified polypeptides would seem likely candidates to be serving structural roles in basal-body supportive frameworks in these cells.

MATERIALS AND METHODS

Euplotes cell culture, cortical preparations, and antibody production

Previous publications have described in detail the methods used to raise *Euplotes aediculatus*, prepare purified cortical extracts from them, and raise antibodies against select cortical proteins (Kloetzel

1991; Kloetzel et al. 2003a; Kloetzel and Brann 2012). In brief, *E. aediculatus* (ATCC 30859) were cultivated in modified Pringsheim's medium, or in commercial Volvic water, and fed either bacteria or *Tetrahymena*. Extraction of cortical fractions from large quantities of washed *Euplotes* cells utilized a Triton-high salt (THS) protocol modified slightly from that published by Williams et al. (1989). Protein bands from THS-extracted cortical preparations were separated by conventional SDS-PAGE and electroblotted onto nitrocellulose membranes; polyclonal antibodies were then produced in rabbits immunized with individual bands cut from the nitrocellulose, as described (Kloetzel and Brann 2012).

***Euplotes* expression library screening**

A cDNA expression library was constructed from a pool of mRNA extracted from a growing population of *E. aediculatus* at the Université Paris-Sud, Orsay, France. Details of its construction in λ gt11, and its screening with antibodies against another *Euplotes* protein, platein, have been described (Kloetzel et al. 2003a). In the present study, the library was screened with a rabbit antibody generated against the most abundant protein (termed 'cagein') in cortical preparations that had been enriched in basal body cages. This antibody, when affinity-purified ("AP-6"), has shown strong, highly specific reactivity with *Euplotes* basal-body cage structures (cirrus and membranelle bases) by immunofluorescence and immuno-electron microscopy (Kloetzel and Brann 2012). Nitrocellulose filters were layered onto agar plates containing lawns of *E. coli* Y1090 infected with the λ gt11/*Euplotes* cDNA library at dilutions of 10^{-4} - 10^{-5} . LacZ/cDNA fusion protein expression was induced with IPTG over 3-5 hr of growth at 37 °C. Filters were collected, rinsed, and exposed to AP-6 anti-cagein antibody (pre-adsorbed with Y1090 lysate) at a dilution of 1:500 overnight at 25 °C. Secondary antibody exposure (goat anti-rabbit IgG, coupled to alkaline phosphatase) at a dilution of 1:7500 was for 45-60 min at 25 °C. Development with alkaline phosphatase reaction mixture revealed several plaques expressing presumptive cagein DNA inserts. Some of these were picked from the agar plates; subclones were rescreened for positive expression at least twice prior to insert characterization. Inserts were amplified by standard polymerase chain reaction (PCR) methods, utilizing λ gt11 forward and reverse primers. Successful reactions were gel-purified and cloned into a double-stranded plasmid vector (pCR®2.1-TOPO, version L; Invitrogen) following manufacturer's instructions. Plasmids with inserts were purified and sequenced in both directions on an ABI 373

DNA sequencer with ABI Prism kits (Applied Biosystems Division of Perkin-Elmer Inc., Foster City, CA, USA).

Completion of the cagein gene sequence

The largest partial cagein gene insert cloned from the *Euplotes* cDNA library (C4.2) was 1216 nucleotides (nt) in length. It included sufficient coding information for a long open reading frame (ORF) concluding with a TAA stop codon; we presumed from its lack of a start codon in this frame that it was missing its upstream sequences. To complete the 5' end of the putative cagein gene, as well as the short 3' gene terminus, a PCR approach was applied (Methods S1), using purified *Euplotes* whole-cell DNA as the template.

The nucleotide sequence of the cagein gene reported here has been submitted to GenBank™ under accession number MN659341.

Southern blotting

Agarose-separated nucleic acids from whole cells (thus yielding primarily natural 'nanochromosome'-size DNA from macronuclei) were screened with cagein-specific probes (labeled using random hexanucleotide primers). The labeling and hybridization procedures (using Hybond-N membranes; Amersham) closely followed those employed previously (Kloetzel et al. 2003a) for screening with platein probes. Hybridizations were carried out overnight at 65 °C, followed by washes and exposure to X-ray film using standard procedures (Sambrook et al. 1989).

RESULTS AND DISCUSSION

Sequence properties of the cagein gene

The predicted nucleotide sequence of the cagein gene is presented in Fig. 1A, including the telomeric sequences characteristic of all *Euplotes* macronuclear genes (Hoffman et al. 1995; Prescott 1994).

Pentameric nucleotide sequences thought to signal excision from the micronuclear genome in *Euplotes* (E-Cbs; Klobutcher et al. 1998) are found at the consensus 17 nucleotides interior to the telomere addition sites at either end of the cagein gene. Ghosh et al. (1994) analyzed a variety of *Euplotes* genes, and describe the most likely signal sequence for poly(A) addition being

5'[A/T]TAAAA3'. Such a candidate poly(A)-addition signal is found near the 3' end of the sequence shown; it is highlighted in gray, as is a presumptive poly(A) addition site. The presence of these features indicates that the gene sequence as shown represents an entire cagein-encoding nanochromosome.

In Fig. 1A the start and termination codons are indicated in red; the total open reading frame (ORF) shown encompasses 1417 nucleotides (nt). Initially we focused on the long ORF ending with a TAA at nt 1125. This sequence would encode a polypeptide of 354 amino acids, with a predicted M_r of 41.4 kDa -- not dissimilar to the 45 kDa mass shown for cagein by SDS-PAGE (Kloetzel and Brann, 2012). This version of the cagein gene and its derived protein were presented at several conferences (Kloetzel et al. 2002; Kloetzel and Butler 2003). More recently, several papers have presented strong evidence that programmed ribosomal frameshifting is very common in *Euplotes* genes (Lobanov et al. 2017; Wang et al. 2016). Upon examining the cagein gene sequence more closely, it became evident that the TAA we had assumed terminated the cagein ORF actually was part of a clear candidate “slippery” frame-shift hexanucleotide signal sequence. A +1 frameshift turned out to link to a second long (351 nt) downstream ORF, adding an additional 117 amino acids to the upstream ORF. The resulting fused cagein protein would then contain 463 amino acids, with an M_r of 54 kDa. This frameshifted sequence is that depicted in Fig. 1A; the location of the frameshift is indicated. The predicted frameshift region is illustrated in detail in Fig. 1B, with both nucleotide and peptide sequences.

Several lines of evidence support the likelihood of the frameshift proposed here. One analysis focused on the relative GC content of coding regions versus 5' and 3' untranslated (UTR) gene segments. Both the pre- and post-frameshift ORFs reveal much higher GC content (34% and 39% respectively) than the UTRs (12-13% GC). This is in agreement with findings for other ciliate genes (Hoffman et al. 1995; Prescott 1994). Furthermore, *Euplotes* 3'-UTRs are typically very short, often less than 50 nt (Ghosh et al. 1994; Lobanov et al. 2017). The cagein trailer sequence shown is only 31 nt. Without a frameshift, the 3'-UTR would extend over an uncharacteristic 383 nt. Additional evidence for the proposed frameshift is presented in the following section, as the properties of the encoded cagein protein are examined and discussed.

Within the frame presented for the upstream ORF, a TAG codon is found at nt 264-266. This seemed problematic, although Lobanov et al. (2017) found that this assumed stop codon, like TAA, is frequently part of a frameshift sequence in the large number of *Euplotes* genes they analyzed. Further evidence to be considered next (in sections on both the derived protein and on sequence comparisons with other ciliates) argued against a frameshift at this location, and stimulated a closer look at this region of the gene. It was found that a 24 nt sequence concluding with the TAG contained the canonical dinucleotides indicative of an intron (GT at the 5'-junction, AG at the 3'-end). Realizing that this part of the gene sequence was determined by PCR, using whole-cell DNA as the template, we assume that this intron segment is indeed part of the natural gene that is spliced out before translation, and we do not include it in our considerations of the cagein protein, to follow next.

As presented in Fig. 1A, the cagein gene, including telomeres, is 1,538 nucleotides in length. When total cellular DNA from *Euplotes* was reacted with a cagein-specific probe by Southern blotting, a band of very close to this size was revealed (Fig. 2). This further supports the idea that the cagein gene presented in Fig. 1A represents the full size of a native macronuclear gene that exists *in situ*. DNA from *Tetrahymena* was nonreactive with this cagein probe.

Sequence features and properties of the derived cagein polypeptide

Analysis of the amino acid sequence derived from the cagein gene reveals features that seem to explain a great deal about how the cagein protein might function in the cytoskeletal cortex of *Euplotes*. Fig. 3 shows the predicted peptide sequence for cagein (463 residues), using the Euplotid code (Code 10: NCBI). Residues 69-463 derive from the gene segment identified in the original library screen. The portion highlighted in yellow indicates the N-terminal 68 amino acids derived from the sequence completed by PCR, following intron removal.

Examining the amino acid composition of the cagein protein, of note is the very high number of negatively and positively charged residues (97 Asp+Glu; 99 Lys+Arg, respectively). Together these comprise 42.3% of the polypeptide's total residues; the theoretical pI of the protein is 7.98.

Hydrophobicity plots (Manavalan; Kite & Dolittle, not shown) confirm cagein to be a highly hydrophilic polypeptide, as expected by an analysis of its component amino acids. The predicted highly charged and hydrophilic nature of the derived cagein polypeptide were unexpected, since the

basal-body cages in which it seems to play an important structural role are highly insoluble. The Triton/High Salt (THS) extraction procedure is expected to solubilize all membranous organelles, nuclear material, and almost all cytoplasmic proteins. These resulting ‘cortical preparations’, representing the insoluble residue, certainly would not be expected to contain hydrophilic proteins -- and yet basal body cages are a prominent feature of such preparations.

In silico analyses of the derived cagein protein help explain this apparent discrepancy. Secondary structure predictions utilizing the JPred4 program (Drozdetskiy et al. 2015) suggest that cagein exists in a helical configuration over a majority of its length (78.6%). Application of the Coils program (v.2.2; ExPASy) reveals (Fig. S1) that many of these helical regions are predicted to exist in a coiled-coil conformation (approximately 52%, scattered throughout the protein’s total length; these regions are also indicated on the amino acid sequence shown in Fig. 3). Since the amino-acid heptads involved in coiled-coil formation are rich in charged residues (Mason and Arndt 2004), this may help explain the high percentage of charged amino acids cited above in cagein’s composition. Cytoskeletal proteins often possess significant domains of coiled-coil structure (see discussion in Preisner et al. 2018); thus it seems reasonable to interpret cagein as an intermediate-filament (IF) type protein that assembles into ‘cage’ scaffolds, grouping basal body arrays in the *Euplotes* cortex (Kloetzel and Brann 2012).

The potential for serine, threonine, and tyrosine (STY) residues in the cagein protein to serve as substrates for protein phosphorylation was explored using the NetPhos3.1a program (ExPASy/CBS). The output of this analysis shown in Fig. S2; of 48 total STYs, 37 meet the 0.5 probability threshold. Choosing more selectively, the 21 individual STY residues that are predicted with high probability (>0.85) of serving as phosphoacceptors are highlighted in Fig. 3. Note that 11 of these high-likelihood phosphoacceptors lie within coiled-coil regions. It therefore seems reasonable to propose that reversible phosphorylation of cagein plays a role in controlling the assembly/disassembly dynamics of basal-body cages during remodeling of the cortex in *Euplotes*, as discussed in the section on cortical morphogenesis below.

An issue that initially seemed hard to reconcile was the predicted M_r for the cagein polypeptide (~54 kDa) compared with its migration on SDS-PAGE gels (~45 kDa). When the cagein protein sequence was submitted to SwissModel (ExPASy), the best model that resulted showed cagein’s

tertiary structure in the form of a helix-loop-helix hairpin. This analysis was refined, using the I-TASSER server to predict the three-dimensional structure of cagein (Methods-S2). This model is shown in Fig. 4. Rath et al. (2009) demonstrated experimentally that such coiled-coil structures yield anomalous migrations (“gel shifting”) by SDS-PAGE, migrating up to 30% faster than their formula weights would suggest, due to altered detergent binding. Such a gel-shift may be expected as well for cagein.

Even though several papers have reported that TAG codons in diverse *Euplotes* genes may sometimes be used to code for amino acids (Miceli et al. 1994; Ricci et al. 2018; Wang et al. 2020), identification of an in-frame TAG codon for residue 68 of the cagein upstream ORF was unexpected. This finding helped lead to the identification of the presumptive intron in the cagein gene described above. As noted there, Euplotids have a very high incidence of +1 programmed ribosomal frameshift transcripts. An in-frame TA[A/G] is frequently a prelude to a frameshift, especially when occurring toward the 5’ end of a suspected ORF. However, the TAG sequence found at nucleotides 264-266 in the cagein gene does not appear to fit into this category for several reasons. First, the codon upstream from the TAG (= CTT in cagein) is not among the prominent frameshift-associated sequences identified to date (Lobanov et al. 2017). Second, any change in reading frame from this TAG codon downstream produces sequences that have an abundance of standard *Euplotes* stop codons (particularly TAAs). Additionally, the sequences both upstream and downstream of the given TAG nucleotide sequence, when translated in the same frame, yield polypeptide sequences that predict a high percentage of coiled-coil content (Fig. 3, S1). No coiled-coil sequences are found in the predicted product of alternative downstream frame translations. These findings give us confidence in our assumption that the TAG codon is used neither in translation nor involved in a frameshift, but rather is removed as part of an intron splicing event.

It would be welcome to have additional experimental evidence demonstrating that the gene we have isolated does indeed encode the major protein (‘cagein’) of the polykinetid basal body scaffold structures. Clearly this gene is expressed in *E. aediculatus*, since it was isolated from a cDNA library. At present, three diverse lines of evidence support our identification. First, the same affinity-purified antiserum used in screening the *Euplotes* expression library exhibits impressive specificity for both the cage scaffold structures in *Euplotes* polykinetids (by immunofluorescence and immuno-EM

staining, *in situ* and in THS cytoskeletal preparations) as well as strong reactivity with the major protein of isolated cage preparations by Western blotting (Kloetzel & Brann 2012). Secondly, the polypeptide encoded by the isolated gene demonstrates the properties found in many (other) cytoskeletal proteins, namely abundant coiled-coils within its predicted conformation. Finally, as discussed next, by far the strongest homologs identified in BLAST searches, using our putative cagein protein as the query, are proteins (so far unidentified) of other ciliated protists that organize their cilia into polykinetid groupings, as in *Euplotes*.

Cagein-type proteins likely exist in other ciliated protists

The genome of the fresh-water *Euplotes* species *E. octocarinatus* has been sequenced and made available by Wang et al. (2018). A search of this database revealed a cagein homologue (contig 6760) with a very strong similarity to *E. aediculatus* cagein (78% identity, 87% similarity; Fig. S3).

Interestingly, the *E. octocarinatus* gene has an intron in the same position as the *E. aediculatus* intron described above. However, in this related species the cognate intron is not 3n (divisible by 3), and thus must be spliced out to preserve the ORF. Bondarenko and Gelfand (2016) point out that 3n introns in *Paramecium* have a high frequency of in-frame stop codons, as is the case for the TAG terminating the 24 nt *E. aediculatus* sequence. These factors provide additional evidence for the removal of this short segment from the pre-mRNA prior to translation of the cagein gene.

Comparisons of the two derived *Euplotes* proteins strongly suggest that the *E. octocarinatus* homologue has a shift in frame about 230 nt downstream from the intron. This is reinforced by the presence of the common frameshift hexanucleotide AAATAA (Lobanov et al. 2017) in the corresponding portion of the gene. Sequence uncertainty in the region of the *E. octocarinatus* gene encoding this reported contig (related to a potential intron) precludes precise determination of the frameshift mechanism utilized. In the alignment shown in Fig. S3, and in the multiple sequence analysis described next, two “X” residues were inserted to accommodate this uncertainty.

The marine *Euplotes* species *E. crassus* and *E. focardii* also have had their genomes sequenced (Lobanov et al. 2017), and evident homologues of the cagein gene have been identified in both. A candidate *E. focardii* cagein mRNA has been found as well in the transcriptome deposited in the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP0205; iMicrobe.us).

Keeling et al. 2014; Youens-Clark et al. 2019). The similarity scores for the derived proteins were robust: *E. crassus* residues showed 51% identity, 69% similarity to *E. aediculatus* cagein (Fig. S4); for *E. focardii* the scores were 52% identity, 69% similarity (Fig. S5). All four *Euplotes* species, analyzed by multiple sequence alignment (Clustal Omega 1.2.4; EMBL), are shown in Fig. 5. From an evolutionary perspective, it is of interest that three of these species display a frameshift, but each in a different location; *E. crassus* has no evident frameshift. We observed that if the 8-residue segment we assume now to represent an intron in *E. aediculatus* was included in the MSA shown in Fig. 5, it created a noticeable gap at this region in the other three species. This was most obvious in comparisons with the *Euplotes* sequences, the closest relatives, but was consistently evident as well in the alignments with more diverse ciliates, discussed next. This further supports our assumption that this region of the *E. aediculatus* gene is spliced out before translation.

When the derived cagein polypeptide sequence was submitted to BLAST (Fasta; NCBI) or HMMER (2.41.1; EMBL) in a search for additional related sequences, the most significant matches were with other ciliated protists: *Stentor coeruleus*, *Stylonychia lemnae*, *Oxytricha trifallax*, and *Paramecium tetraurelia*. All these proteins were derived from partial or full genome database entries, and described by their annotators as “unknown” or “hypothetical protein”. When submitted to the Coils program, each of these cagein-related sequences demonstrated significant blocks of coiled-coiled structure, ranging from 43% to 63% of the respective proteins’ full lengths. Much lower-stringency matches to cagein were recorded for several proteins exhibiting significant amounts of coiled-coil structure, such as myosin heavy-chain tails.

The regions of the ciliate cagein homologues showing the closest matches with *E. aediculatus* cagein were most concentrated at the N-termini of the proteins. When only the 130 N-terminal cagein residues were used as the search query, three additional ciliate sequences emerged -- “uncharacterized proteins” from the genomes of *Tetrahymena thermophila*, *Ichthyophthirius multifiliis*, and *Pseudocohnilembus persalinus*. An example of alignments covering this portion of the cagein gene is shown in Fig. 6 for two ciliates, *P. tetraurelia* and *T. pyriformis*, quite distantly related to *Euplotes*. One particular area of the N-terminal cagein sequence, centered around residues 91 - 112, showed a surprising degree of conservation among all the candidate ciliate homologues. Fig. 7 shows a multiple sequence alignment for this short region. In the five species shown, the percentage of

identical matches for these cells ranged from 67-70%, and positives from 86-90%. Given the quite wide evolutionary separation of these species (Gao et al. 2016; Lynn 2008), it seems plausible that this short region might serve as a hallmark of cagein-related proteins in ciliates generally. In the *E. aediculatus* 21-mer sequence, serine residues at positions 92 and 102 are predicted with greater than 96% likelihood to serve as substrates for phosphorylation (NetPhos 3.1a). Serine 102 is the first residue in what is predicted (Coils v2.2) to be the second major coiled-coil region of the protein (Fig. 3, S2).

Each of the ciliated protists listed above use clustered cilia (polykinetids) for food collection and/or locomotion. Given the significantly low E values of their matches, it seems reasonable to predict that the sequences displayed represent proteins that, like cagein, serve roles in linking basal bodies into the orderly patterns characteristic of polykinetids. In this context, it is worth noting that immunoblots utilizing affinity-purified antibodies to *E. aediculatus* cagein protein showed significant cross reactivity with whole-cell and cortical protein preparations from both *Paramecium* and *Tetrahymena* (Kloetzel and Brann 2012). And an oxytrichid ciliate isolated from local waters near Baltimore showed reactions with the same purified anti-cagein antibody as well, both by immunofluorescence (Kloetzel and Brann 2012) and in immunoblots. Two bands were seen in reactions against whole-cell protein extracts (approximately 40 and 50kDa; Kloetzel, pers. observ.). Cortical preparations were not possible to assess, since THS extraction dissolves the entire cell in oxytrichids.

How cagein fits into the larger picture of basal-body connectors and cortical anchoring is an open question. The *Euplotes* cage framework is structurally more elaborate than those found in many other ciliates, with two distinct cross-linked layers (Kloetzel and Brann 2012). Euplotids lack, in the mature structures, prominent striated kinetodesmal fibers (rootlets) and related appendages found in most ciliates (see Lynn and Corliss 1991); their cage structures thus may provide additional support and anchoring against the mechanical stresses of ciliary activity. Since basal body cages can be isolated as intact “organelles”, free of basal bodies (Kloetzel and Brann 2012), additional proteins may be presumed to play roles as the “glue” anchoring the many individual kinetosomes into the cage frameworks. Such proteins may be related to those already identified as kinetosome-associated in earlier investigations of the ciliate cortex (Heydeck et al. 2016; Kilburn et al. 2007) and other systems

(Jana et al. 2018). The most plausible relationships to cagein assemblies would be to those ‘fibrous’ structures seen microscopically linking kinetosomes in a variety of ciliates (de Puytorac et al. 1976; Grim 1972; Grimes 1972; Jerka-Dziadosz 1980; Matsusaka et al. 1984).

Protist cytoskeletons and cellular morphogenesis

The assembly of structural elements in higher cells remains a topic of continued interest. Studies of cellular morphogenesis have often taken advantage of particular features found in model protistan organisms. Protists are typically high on the scale of eukaryotic cell size, and have evolved an impressive array of strategies to reinforce their large surface areas -- to accommodate their equally impressive variety of lifestyles. These strategies typically invoke arrangements of multiple cytoskeletal proteins (and sometimes membranes) to allow flexibility and resilience, while protecting against varied environmental insults and challenges. Goodenough et al. (2018) and Preisner et al. (2018) provide extensive catalogs and discussion of these manifold cytoskeletal proteins in recent reviews.

Among protist models, ciliates have been especially well studied. These large cells typically feature highly organized arrays of cytoskeletal elements localized at the cell periphery, in a cortical complex referred to as the pellicle (Adoutte and Fleury 1996; Lynn and Corliss 1991). One singular advantage of using ciliates to study cellular morphogenesis is the frequency with which they remodel and duplicate their cortical cytoskeletons (more than once daily, for many), in a predictable, stereotypical series of steps. Additionally, because elements of the pellicle can be readily stained and visualized by a variety of increasingly sensitive techniques, ciliates continue to be used as favored models. These studies have gained more significance as the wider taxonomic relatedness of many of the proteins involved has become better appreciated (Bengueddach et al. 2017; Lutz et al. 2001; Nabi et al. 2019; see Preisner et al. 2018, Soares et al. 2020). The identification of cagein adds a significant new item to the “parts list” of proteins used in the construction of at least some ciliate cortical cytoskeletons.

For certain cytoskeletal proteins, detailed knowledge has been achieved in moving from individual molecules to supramolecular/subcellular structures. Keratin provides one good example (Bragulla and Homberger, 2009; Snider et al. 2013). In particular model ciliates, notably

Paramecium, significant progress has been made in relating individual cytoskeletal protein subunits to the assembly of higher-order cortical structures (Aubusson-Fleury et al. 2013; Bengueddach et al. 2017; Nabi et al. 2019).

Cytoskeletal elements need to be dynamic to serve various cellular functions and shape changes; thus controls over these elements' assembly and disassembly states necessarily have evolved. Cycles of phosphorylation and dephosphorylation have been shown to regulate the state of assembly of cytoskeletal elements of many kinds from their component protein subunits (Bragulla and Homberger 2009; Omary et al. 2006; Sihag et al. 2018; Snider and Omary 2014). In the ciliate *Paramecium*, Sperling et al. (1991) were the first to demonstrate that hyperphosphorylation of several ciliary rootlet proteins was tightly correlated with the disassembly of the rootlet fibers, major elements of the pellicular cytoskeleton. Such modulation of cortical structure accompanied the duplication of the entire cell surface, both in time and space.

In *Euplotes*, we have discussed (Kloetzel et al. 2003b) evidence suggesting that cycles of phosphorylation/dephosphorylation of the platein proteins could control the disassembly/reassembly cycles of the alveolar plates during duplication of the cell cortex accompanying vegetative growth. Based on the presence of a significant number of high-likelihood phosphoacceptor STY residues in the cagein protein, as well as in plateins, it seems reasonable to suggest that the assembly dynamics of basal body cages could be similarly regulated. If phosphorylation were involved in loosening the coiled-coil backbone of native cagein, the released hydrophilic cagein subunits would be readily soluble in the cytoplasmic matrix. Such a pool of subunits could be subsequently reused in new cage assembly. And, as discussed in Preisner et al. (2018; see also Pinto et al. 2014), the propensity of IF-type subunit proteins to self-assemble into subcellular structural elements has been amply demonstrated.

ACKNOWLEDGMENTS

We wish to thank Anne Baroin-Tourancheau (Laboratoire de Biologie Cellulaire 4, Université Paris-Sud, 91405 Orsay, France) for her indispensable help with expression library screening. The assistance of Marilis Rodriguez and Andrea Molinaro, New York Blood Center, as well as Nick Ambulos, University of Maryland, Baltimore, for DNA sequencing, is gratefully acknowledged.

Cristina Miceli (Università di Camerino) aided considerably in sequence analysis. The MARC Program (grant T34 GM136497 from NIGM) provided support for Maurice Butler during his undergraduate research at UMBC. Other support from UMBC included a Designated Research Initiative Fund award to JAK. The residence of JAK in Orsay was supported by a grant from the CNRS, and in Camerino through funds provided by the University of Camerino.

LITERATURE CITED

- Adoutte, A. & Fleury, A. 1996. Cytoskeleton of ciliates. *In*: Hausmann, K. & Bradbury, P. (ed), Ciliates: Cells as Organisms. Gustav Fischer Verlag, New York. pp 41-49.
- Aubusson-Fleury, A., Bricheux, G., Damaj, R., Lemullois, M., Coffe, G., Donnadieu, F., Koll, F., Viguès, B. & Bouchard, P. 2013. Epiplasmins and epiplasm in *Paramecium*: The building of a submembraneous cytoskeleton. *Protist*, 164:451-469.
- Bengueddach, H., Lemullois, M., Aubusson-Fleury, A. & Koll, F. 2017. Basal body positioning and anchoring in the multiciliated cell *Paramecium tetraurelia*: roles of OFD1 and VFL3. *Cilia*, 6:6 DOI 10.1186/s13630-017-0050-z
- Bondarenko, V. S. & Gelfand, M. S. 2016. Evolution of the exon-intron structure in ciliate genomes. *PLoS ONE*, 11(9):e0161467. DOI:10.1371/journal.pone.0161476
- Bragulla, H. H. & Homberger, D. G. 2009. Structure of keratins and keratin filaments in simple, stratified, keratinized and cornified epithelia. *J. Anat.*, 214:516-559.
- de Puytorac, P., Grain, J. & Rodrigues de Santa Rosa, M. 1976. A propos de l'ultrastructure corticale du cilie hypotriche *Stylonychia mytilus* Ehrbg., 1838: les caracteristiques du cortex buccal adoral et paroral des polyhymenophora Jankowski, 1967. *Trans. Amer. Micros. Soc.*, 95:327-345.
- Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. 2015. JPred4: a protein secondary structure prediction server. *Nucl. Acids Res.*, 43, W1:W389-W394
- Gao, F., Warren, A., Zhang, Q., Gong, J., Miao, M., Sun, P., Xu, D., Huang, J., Yi, Z. & Song, W. 2016. The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the Phylum Ciliophora (Eukaryota, Alveolata). *Sci. Rep.*, 6:24874. DOI:10.1038/srep24874

- Ghosh, S., Jaraczewski, J. W., Klobutcher, L. A. & Jahn, C. L. 1994. Characterization of transcription initiation, translation initiation, and poly(A) addition sites in the gene-sized macronuclear DNA molecules of *Euplotes*. *Nucl. Acids Res.*, 22:214-221
- Goodenough, U., Roth, R., Kariyawasam, T., He, A. & Lee, J.-H. 2018. Epiplasts: membrane skeletons and epiplastin proteins in Euglenids, Glaucophytes, Cryptophytes, Ciliates, Dinoflagellates, and Apicomplexans. *MBio* 9:e02020-18
- Grim, J. N. 1972. Fine structure of the surface and infraciliature of *Gastrostyla steinii*. *J. Protozool.*, 19:113-126.
- Grimes, G. 1972. Cortical structure in non-dividing and cortical morphogenesis in dividing *Oxytricha fallax*. *J. Protozool.*, 19:428-445.
- Heydeck, W., Stemm-Wolf, A. J., Knop, J., Poh, C. C. & Winey, M. (2016) Sfr1, a *Tetrahymena thermophila* Sfr1 repeat protein, modulates the production of cortical row basal bodies. *mSphere*, 1(6):e00257-16
- Hoffman, D. C., Anderson, R. C., DuBois, M. L. & Prescott, D. M. 1995. Macronuclear gene-sized molecules of hypotrichs. *Nucl. Acids Res.*, 23:1279-1283.
- Jana, S. C., Mendonça, S., Machado, P., Werner, S., Rocha, J., Pereira, A., Maiato, H. & Bettencourt-Dias, M. 2018. Differential regulation of transition zone and centriole proteins contributes to ciliary base diversity. *Nature Cell. Biol.*, 20:928-941.
- Jerka-Dziadosz, M. 1980. Ultrastructural study on development of the hypotrich ciliate *Paraurostyla weissei*. I. Formation and morphogenetic movements of ventral ciliary primordia. *Protistologica*, 16:571-589.
- Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A. et al. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol*, 12:e1001889.
- Kilburn, C. L., Pearson, C. G., Romijn, E. P., Meehl, J. B., Giddings, Jr., T. H., Culver, B. P., Yates, III, J. R. & Winey, M. 2007. New *Tetrahymena* basal body protein components identify basal body domain structure. *J. Cell Biol.*, 178:905-912.

- Klobutcher, L. A., Gyga, S. E., Podoloff, J. D., Vermeesch, J. R., Price, C. M., Tebeau, C. M. & Jahn, C. L. 1998. Conserved DNA sequences adjacent to chromosome fragmentation and telomere addition sites in *Euplotes crassus*. *Nucl. Acids Res.*, 26:4230-4240.
- Kloetzel, J. A. 1991. Identification and properties of plateins, major proteins in the cortical alveolar plates of *Euplotes*. *J. Protozool.*, 38:392-401.
- Kloetzel, J. A., Butler, M. D., Baroin-Tourancheau, A., Fleury-Aubusson, A. & Banerjee, D. 2002. Sequencing and analysis of the gene encoding cagein, a novel cytoskeletal protein in the ciliate *Euplotes*. Abstract 13. *J. Euk. Microbiol.*, 49:3A.
- Kloetzel, J. A., & Butler, M. D. 2003. Cagein, a novel cytoskeletal protein in *Euplotes*: sequence analysis suggests an assembly-disassembly mechanism. Abstract 21. *J. Euk. Microbiol.*, 50:7A.
- Kloetzel, J. A., Baroin-Tourancheau, A., Miceli, C., Barchetta, A., Farmar, J., Banerjee, D. & Fleury-Aubusson, A. 2003a. Cytoskeletal proteins with N-terminal signal peptides: plateins in the ciliate *Euplotes* define a new family of articulins. *J. Cell. Sci.*, 116:1291-1303.
- Kloetzel, J. A., Baroin-Tourancheau, A., Miceli, C., Barchetta, A., Farmar, J., Banerjee, D. & Fleury-Aubusson, A. 2003b. Plateins: a novel family of signal peptide-containing articulins in euplotid ciliates. *J. Eukaryot. Microbiol.*, 50:19-33.
- Kloetzel, J. A. & Brann, T. W. (2012) Structure and protein composition of a basal-body scaffold ('cage') in the ciliated protozoan *Euplotes*. *J. Euk. Microbiol.*, 59:587-600.
- Lobanov, A. V., Heaphy, S. M., Turanov, A. A., Gerashchenko, M. V., Pucciarelli, S., Devaraj, R., Xie, F., Petyuk, V. A., Smith, R. D., Klobutcher, L. A. et al. 2017. Position-dependent termination and widespread obligatory frameshifting in *Euplotes* translation. *Nat. Struct. Mol. Biol.*, 24:61-68.
- Lutz, W., Lingle, W. L., McCormick, D., Greenwood, T. M. & Salisbury, J. L. 2001. Phosphorylation of centrin during the cell cycle and its role in centriole separation preceding centrosome duplication. *J. Biol. Chem.*, 276:20774-20780.
- Lynn, D. H. 2008. The Ciliated Protozoa: Characterization, Classification, and Guide to the Literature. 3rd ed. Springer-Verlag, New York.
- Lynn, D. H., Corliss, J. O. 1990. Ciliophora. In: Harrison, F. W. (ed), Microscopic Anatomy of Invertebrates. Wiley-Liss, New York. 1:333-467.

- Mason, J. M. & Arndt, K. M. 2004. Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem*, 5:170-176.
- Matsusaka, T., Nakamura, T. & Nagata, K. 1984. Ultrastructure, disintegration and formation of a cirrus in the vegetative, encysting and excysting ciliate *Histriculus muscorum*. *J. Electron Microsc. Japan*, 33:217-229.
- Miceli, C., Ballarini, P., Di Giuseppe, G., Valbonesi, A. & Luporini, P. 1994. Identification of the tubulin gene family and sequence determination of one beta-tubulin gene in a cold- poikilotherm protozoan, the Antarctic ciliate *Euplotes focardii*. *J. Euk. Microbiol.*, 41:420– 427.
- Nabi, A., Yano, J., Valentine, M. S., Picariello, T. & Van Houten, J. L. 2019. SF-assemblin genes in *Paramecium*: phylogeny and phenotypes of RNAi silencing on the ciliary striated rootlets and surface organization. *Cilia.*, doi.org/10.1186/s13630-019-0062-y
- Omary, M. B., Ku, N. O., Tao, G. Z., Toivola, D. M. & Liao, J. 2006. 'Heads and tails' of intermediate filament phosphorylation: multiple sites and functional insights. *Trends Biochem. Sci.*, 31: 383-394.
- Pinto, N., Yang, F.-C., Negishi, A., Rheinstädter, M. C., Gillis, T. E. & Fudge, D. S. 2014. Self-assembly enhances the strength of fibers made from vimentin intermediate filament proteins. *Biomacromolecules*, 15:547-581.
- Preisner, H., Habicht, J., Garg, S. G. & Gould, S. B. 2018. Intermediate filament protein evolution and protists. *Cytoskeleton*, 75:231-243.
- Prescott, D. M. 1994. The DNA of ciliated protozoa. *Microbiol. Rev.*, 58:233-267.
- Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. & Deber, C. M. 2009. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc. Nat. Acad. Sci.*, 106:1760-1765.
- Ricci, F., Candelori, A., Brandi, A., Alimenti, C., Luporini, P. & Vallesi, A. 2018. The sub-chromosomal macronuclear pheromone gene of the ciliate *Euplotes raikovi*: comparative structural analysis and insights into the mechanism of expression. *J. Euk. Microbiol.*, 66:376-384.
- Sambrook, J. R., Fritsch, E. F. & Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

- Sihag, R. K., Inagaki, M., Yamaguchi, T., Shea T. B. & Pant, H. C. 2007. Role of phosphorylation on the structural dynamics and function of types III and IV intermediate filaments. *Exp. Cell Res.*, 313:2098-2109.
- Snider, N. T. & Omary, M. B. 2014. Post-translational modifications of intermediate filament proteins: mechanisms and functions. *Nat. Rev. Mol. Cell Biol.*, 15:163-177.
- Snider, N. T., Park, H. & Omary, M. B. 2013. A conserved rod domain phosphotyrosine that is targeted by the phosphatase PTP1B promotes keratin 8 protein insolubility and filament organization. *J. Biol. Chem.*, 288:31329-31337.
- Soares, H., Carmona, B., Nolascom, S. & Viseu Melo, L. 2019. Polarity in ciliate models: from cilia to cell architecture. *Front. Cell Dev. Biol.*, 7:240. doi: 10.3389/fcell.2019.00240.
- Soares, H., Sunter, J. D., Wloga, D., Joachimiak, E. & Miceli, C. 2020. *Trypanosoma*, *Paramecium* and *Tetrahymena*: From genomics to flagellar structures and cytoskeleton dynamics. *Eur. J. Protistol.*, 76.doi.org/10.1016/j.ejop.2020.125722
- Sperling, L., Keryer, G., Ruiz, F. & Beisson, J. 1991. Cortical morphogenesis in *Paramecium*: a transcellular wave of protein phosphorylation involved in ciliary rootlet disassembly. *Dev. Biol.*, 148:205-218.
- Wang, R., Liu, J., Di Giuseppe, G. & Liang, A. 2020. UAA and UAG may encode amino acid in Cathepsin B gene of *Euplotes octocarinatus*. *J. Eukaryot. Microbiol.*, 67:144-149.
- Wang, R., Miao, W., Wang, W., Xiong, J. & Liang, A. 2018. EOGD: the *Euplotes octocarinatus* genome database. *BMC Genomics*, 19:63.
- Wang, R., Xiong, J., Wang, W., Miao, W. & Liang, A. 2016. High frequency of +1 programmed ribosomal frameshifting in *Euplotes octocarinatus*. *Sci. Rep.*, 6: 21139.
- Williams, N. E., Honts, J. E., Lu, Q., Olson, C. L. & Moore, K. C. 1989. Identification and localization of major cortical proteins in the ciliated protozoan, *Euplotes eurystomus*. *J. Cell Sci.*, 92:433-439.
- Youens-Clark, K., Bomhoff, M., Ponsero, A.J., Wood-Charlson, E.M., Lynch, J., Choi, I., Hartman, J.H. & Hurwitz, B.L. 2019. iMicrobe: Tools and data-driven discovery platform for the microbiome sciences. *Giga Science*, 8: <https://doi.org/10.1093/gigascience/giz083>

FIGURE LEGENDS

Figure 1. The cagein gene. **A.** Complete nucleotide sequence (1538 nt), including telomeres (lowercase blue; a presumptive 3'-telomeric overhang is not shown). The open reading frame is 1417 nt, running from position 63 through 1479 (start and termination codons indicated in underlined red italics). Nucleotides 243-266 (highlighted in yellow) represent an intron that we presume is spliced out of the pre-mRNA (see text). Shown in blue italics, wavy underline, is the hexanucleotide thought to signal a +1 coding frameshift following the TAA at nucleotides 1125-1127. Excision signals 17 nt internal to the telomeres are italicized and underlined. Presumptive poly(A) addition signal and poly(A) addition site are highlighted in gray. **B.** Location of the presumptive +1 frameshift signaled by the hexanucleotide (red) at nt 1122-1127. The nucleotide sequence of this gene segment is in lowercase blue; the three possible coding frames are shown above. The terminus of the cagein polypeptide sequence encoded by the upstream ORF is highlighted in yellow; the sequence initiating the downstream ORF following frameshift is highlighted in gray.

Figure 2. Southern hybridization using a cagein-specific gene construct to probe electrophoretically separated whole-cell DNA isolated from *Euplotes* (E) or *Tetrahymena* (T). λ E/H spots are size markers (kb) derived from a restriction-enzyme digest of Lambda phage DNA. Note the strong hybridization band at approximately 1.58 kb. The probe is nonreactive with the *Tetrahymena* DNA.

Figure 3. Derived amino acid sequence of the cagein protein, representing the ORF in Fig. 1A minus the presumptive intron. The sequences deduced using our PCR approach are highlighted in yellow. The red residues at positions 346-7 indicate the junctions of the ORF +1 frameshift. Overlain on this sequence are the results of analyzing the protein for predicted coiled-coil sequences (Coils 2.2, ExPASy; 0.5 threshold) and predicted phosphorylation sites (NetPhos3.1a; probabilities of >0.85). Polypeptide segments predicted to adopt coiled-coil structures are wavy underlined (see Fig. S1). High likelihood phosphoacceptor residues (see Fig. S2) are highlighted in color (serines in green, threonine in gray, tyrosines in pink).

Figure 4. Suggested model for cagein tertiary structure, obtained by fold-recognition based on multiple-threading alignments (I-TASSER server). This best fit is that of a helix-loop-helix hairpin.

Figure 5. Multiple sequence alignment (Clustal Omega 1.2.4) of *E. aediculatus* cagein protein with candidate cagein sequences from three other *Euplotes* species, *E. octocarinatus*, *E. crassus* and *E. focardii*. Extensive regions of identity are found at both N- and C-terminal ends of the respective proteins. *E. aediculatus*, *E. octocarinatus* and *E. focardii* have predicted frameshifts during translation. The respective upstream ORFs are highlighted in yellow; downstream ORFs are highlighted in gray. The XX residues in red indicate the region of sequence uncertainty in the *E. octocarinatus* protein.

Figure 6. Multiple sequence alignment of N-terminal ends of *E. aediculatus* cagein protein with candidate cagein homologues derived from two distantly related ciliates, *Paramecium tetraurelia* (XP_001446876.1) and *Tetrahymena pyriformis* (XP_001020877.2). The region of maximum conservation is underlined and highlighted in yellow.

Figure 7. Multiple sequence alignment of the short cagein region demonstrating a high level of sequence homology in 5 diverse ciliate species: *Euplotes aediculatus*, *Stylonychia lemnae*, *Oxytricha trifallax*, *Tetrahymena thermophila* and *Ichthyophthirius multifiliis*. Red indicates identical residues, blue indicates similar residues for all 5 species. Note that the pairs of oxytrichids (rows 2 and 3) and oligohymenophoreans (rows 4 and 5) each have identical sequences through this region.

SUPPORTING INFORMATION

Methods S1. PCR strategy for completing the cagein gene sequence.

Methods S2. Three-dimension modeling of cagein, using the I-TASSER server.

Figure S1. Plot of the predicted coiled-coil domains along the length of the cagein polypeptide.

Figure S2. Predicted phosphorylation sites within the cagein protein.

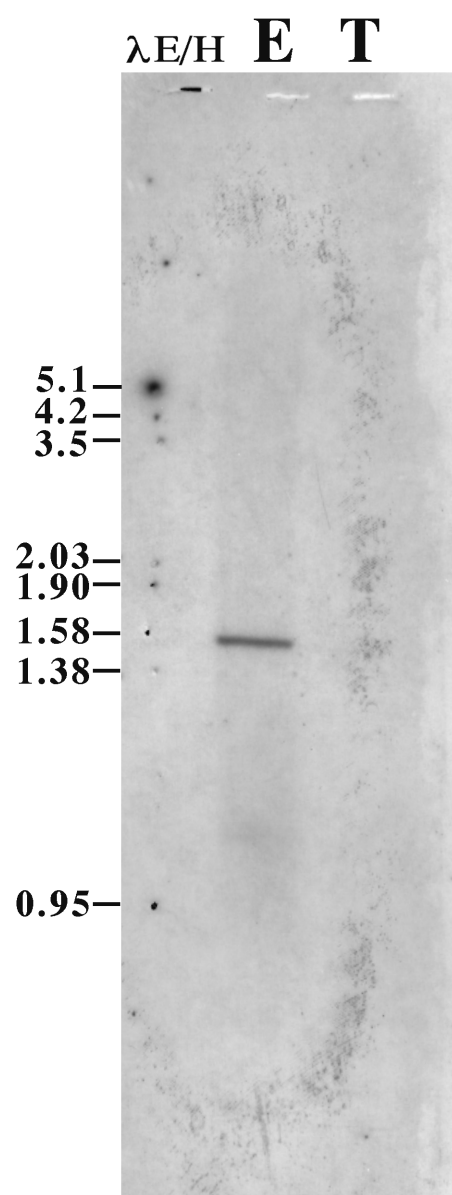
Figure S3. Pairwise alignment of the *E. aediculatus* cagein protein with the homologous protein of *E. octocarinatus*.

Figure S4. Pairwise alignment of the *E. aediculatus* cagein protein with the homologous protein of *E. crassus*.

Figure S5. Pairwise alignment of the *E. aediculatus* cagein protein with the homologous protein of *E. focardii*.

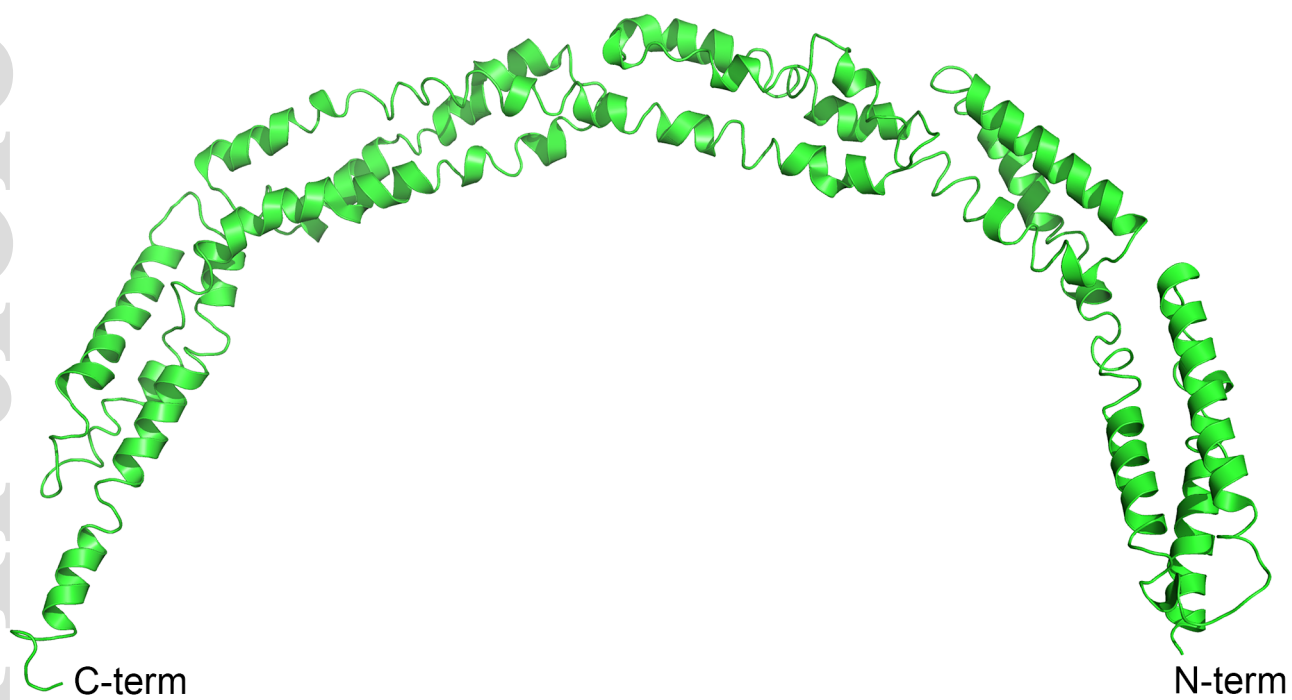
1 ccccaaaaacc ccaaaaacccc aaaaccccTA TTTAAAGATT AAAATTTGAA AGAAATTTGT 60
61 ATATGCTCTGA GCAAGTACTG AAGCACCCGA AAGTGCAGGA AGAGATAAAG AAGATCAAAG 120
121 AGGAGAACTC TGCAAAGACC AAGCAGCTGA TGGCCTCGAT CAACAAACTC AAAAGACGACC 180
181 TCGAGAAGGA GAAGTACCAA AACCAAGACA ACGTGCGGGC CAAAGTCATC GAGCGCATGA 240
241 AGGTACGCTC TCCCCTAATC CTTTAGAAAAG ATATTGCCGA CAATGAAACT GTCATTGAAA 300
301 TGATGCGCAC GATGATCAAC GATGATGATG CCATCAATAA GGAAATCGTC AAAGTTCTTT 360
361 CGAAAGGACC TGAAAGGGTC AGAGTCCTTT CCCGGAAGA ACTCAGAATG GAGATTAAAC 420
421 GCCTTACTGG CGAGGTCAAC TCGTTAAAGC TTAATGGCGC GACGAAGAAG AAAAAGTCGA 480
481 AGAAGGGATC TGAGAAGGGA ACTGAAGAAG ATGTCGCTTT GGAAGAGAGC TTTGATAATA 540
541 TGACTTTTCGA GAAGTTTAAA GAACTTAAAG TTGCTTATGA TCTCTTACAA AGCCAATTAG 600
601 ACGCAAAAAG TAAACAGTTG GTTGATGCTC AGATTACCAT AAGCAATATG AGGGATGATA 660
661 TGAAGGGATT CCAGATGAGC CAACAGACAC TTATGGACAA AGTTAAGGAA ATGGGCGAGG 720
721 ATAAGAAGAT CTTTGAACAG GTTAAGGCAA AACTTAGAGA ATTCAAAGCA AGTGGCCTGT 780
781 TCAAAGGCAT TGAATTAGAT GTTTCAGAAG AGCATGATAA TGATGCAATT AATCTTCTCA 840
841 ATAAGTTCAA CAAACTAATT GACAAAATAA AGGACGAGAT TGATCAAAGA GATAAAGAGA 900
901 ACTATGCCTT CAAAAATAAG CTAGAACAGT CAGAAAAAGA TAAGCTTAAA GAGGTAGATG 960
961 AACTCATTTGA CGAATACAAT CAAAAAGACG AAGCCTTTCA ATCGCTTAAT GATAAGTACA 1020
1021 TCGCTCTTAA GAGGGAAATG AAGATGAAAA TGCAGGAGTT AGCAAAATGC AGAGAAGAAC 1080
1081 TACCAAGAAG AAGAAAAATT AGTCCTTGAT TAAAAAGATG AATATAAGG AAAAGCTGAGC 1140
1141 AGTAAAAAGG AAGAGATGTC TAAAAATGAT AGCATAGTTG CTGAGTTAAA AGAAGAAATT 1200
1201 GCAAGCAGAG ATAAAAAGAT TGAAGAGCTA AAGATTGAAC TAGAAACTCA ACTTCAAAGG 1260
1261 CCAGGAGGAT ATAGTGAAGA TAGCAGTATG GAGTTTGACC CTCTTGCTCT TGCTAATAAT 1320
1321 CTTTGTAAAG AAAAAAGAAA AGTTTCTAAA TTGACTGAAG AAATCAAGAA ATTATACGAC 1380
1381 GATATTGACA AGTTGAAGCT CCAAGTGAAT GTTGAGCATA AACAGGTTGA AAATATGACA 1440
1441 CTTGAAAAGG TTCGAAGTGC TTCAAAGAA AGATTAATAA ATAGTAAATT CAATTGATTG 1500
1501 TTTAAAAAAT ggggttttgg gggttttgggg ttttgggg 1538

L K D E Y K G K L S S K K E E M S K N D S I V
F K R C I * R K A E Q * K G R D V * K C * H S
I * K M N I K E S C A V K R K R C L K M I A *
1109 attttaaagatgaatataaaggaaagctgagcagtaaaaaaggaagagatgtctaaaaatgatagcatag



jeu_12850_f2.tif

1 MSEQVLKHPK VQEEIKKIKE ENSAKTKQLM ASINKLKDDL EKEYONQDN VRAKVIERMK
61 KDIADNETVI EMMRTMINDD DAINKEIVKV LSKGPERVRV LSREELRMEI KRLTGEVNSL
121 KLNGATKKKK SKKGSEKGTE EDVALEESFD NMTFEKFKEL KVAYDLLO SQ LDAKDKQLVD
181 AQITISNMRD DMKGFQMSQQ TLMDKVKEMG EDKKIFEQVK AKLREFKASG LFKGIELDV S
241 EEHDNDAINL LNKFNKLIDK IKDEIDORDK ENYAFKNKLE QSEKDKLKEV DELIDEYNQK
301 DEAFOSLNDK YIALKREMKM KMOELAKCRE ELPTRTKI SP CFKRCIKGKL SSKKEEM SKN
361 DSIVAELKEE IASRDKKIEE LKIELETQLQ RPGGYSED SS MEFDPLALAN NLCKEKRKVS
421 KLTEEIKKLY DDIDKLKLOV NVEHKOVENM TLEKVRSA SK ERL*



jeu_12850_f4.tif

<i>E. aediculatus</i>	MSEQVLKHPKVQEEIKKIKEENSAKTKQLMASINKLKDDLEKEKYQNQDNVRAKVIERMK	60
<i>E. octocarinatus</i>	MSDPVLKHKKVQEEIKRIKDEHSAKTKQLMASINKLKEDIEKEKYQNQDNVRAKIIERMK	60
<i>E. crassus</i>	MADP-MKSKRVIAEIKRIKDEHSLKTKSLMKSINKLKDELEKEKYQKQDNVRAKIIERMK	59
<i>E. focardii</i>	MADP-MKSKRVKEELKHLKDEHSSCKSLMKSINKLKDDLEKEKYQNQDNVRAKIIERMK	59
	:: : :* *:*:*:*: * *.** *****:::*****:*****:*****	
<i>E. aediculatus</i>	KDIADNETVIEMMRTMINDDDAINKEIVKVLSSKGPFRVRLSREELRMEIKRLTGEVNSL	120
<i>E. octocarinatus</i>	KDIADNETVIEMMRGMINDDDAINKEIVKVLSSKGPFRVRLSREELRMEIKRLTAEVNAL	120
<i>E. crassus</i>	KDIGDNETVIELMREIINDDDQIDKAIVKVLKGGIDRARVLSREELRMEIKRLNGELQKL	119
<i>E. focardii</i>	KDIQDNETIIEMMRGMVNDDDAINEGIIISILKGVNKAARVLSREELRMEIKKLINHINSI	119
	*** *****:*.** ::***** *: :*. * . :: *****:*.** ::* *	
<i>E. aediculatus</i>	KLNGATKKKKSKKGSEKGTEDVALEESFDNMTFEKFELKVAYDLLQSQDLAKDKQLVD	180
<i>E. octocarinatus</i>	KQGGATKKKXXKKGTEKETEDVALENSFDNMTFDKFKELKADYDLLKGQDLGKQKELED	180
<i>E. crassus</i>	KTQGVVRRKKKKVDNEEKDEDEQKLENSFEGLTFEHFKELKEQIDRLQAELESKQDELDK	179
<i>E. focardii</i>	KAGGATKRKKKSLKSE-ISEQDKALENSFDHLTFDHFKDLKIEQDKLKEELDDCKKELDK	178
	* *...:* . . * *: : **::*: :*:*:*: * *: :*: * .:*	
<i>E. aediculatus</i>	AQITISNMRDDMKGFQMSQQTLMDKVKEMGEDKKIFEQVKAKLREFKASGLFKGIELDVS	240
<i>E. octocarinatus</i>	AYQTIANMRDDMKGLQMSQQTLMDKVRDMGEDKKIFEQVKGLKEFKAKGLFKGIELDIS	240
<i>E. crassus</i>	ANDQIEELKEDKKVTGISLHSLMDRVKNMDSDKKAFEKVKIKLKEFKNQGIFKGIDLVD	239
<i>E. focardii</i>	AIAQIDEMRDDKKVTGISLHSLMDKVKNMDSDKQAFGRVKVKNLNEFKKEGIFKGIELDSD	238
	* * :*:*: * *: :*:*:*:*:*.** * :** **.*** .*:*****.* *	
<i>E. aediculatus</i>	EEHDNDAINLLNKFNLIDKIKDEIDQRDKENYAFKNKLEQSE----KDKLKEVDELIDE	296
<i>E. octocarinatus</i>	EEHDNETINSNLNKLNLIDKIKEEMDQRDKHQHFKNKLEQNE----KDQAKEVDELIDE	296
<i>E. crassus</i>	EEDEETVNLNRLVNLIDKTKADIENQDKYKEFKDSHTRKEQEREKREKDEIDGLVDE	299
<i>E. focardii</i>	EEEDDETVNLNKLNLVRLIDRTKKEIEKRDKETKEMRETHEKTEQNRKVNDAEIDDLVDE	298
	.*:*:*: **: :*:*: * :*:*: :*: .:* . *:* :***	
<i>E. aediculatus</i>	YNQKDEAFQSLNDKYIALKREMKMKMQELAKCREELPRTRKISPCFKRCIKGKLSSKKEE	356
<i>E. octocarinatus</i>	YNQKDEAFQSLNDKYIALKREMKMKMLELAKCREEMEEAQNLANELKDEYKGKLDKKDE	356
<i>E. crassus</i>	YNQRDDEFHEIEDKYIALKKDMKKKMYEVAKYKEELEELQEVADRLKEENKVIKKDKQ-	358
<i>E. focardii</i>	YNQRDDEFHDINDKYVQLKKDMKKKMYLVAKYKEEMEDAQALADGLKEEYTKLKDRA-	357
	***:*: *:*:*:*: *:*:** ** :** **: : :*: * .. *	
<i>E. aediculatus</i>	MSKNDSIVAEELKEEIASRDKKIEELKIELETQLQRPGGYSEDSSMEFDPLALANNLCKEK	416
<i>E. octocarinatus</i>	FSKFDSQIADLKQEIITNRDKKIEELKIELEQQLRPGGQSEDSSMEIDSYAL----TKER	412
<i>E. crassus</i>	-----RIIDLEADVKNKNKILELGVIEELLQRTPTKS-DTSTEFDSFAL----VKER	407
<i>E. focardii</i>	-----KIIIELEDQLKSKDIRIINLQNDMDEYLQRSPSKS-DTSTEMDSYAL----TNER	406
	: :*: :* :*: :* :*: ** * *: * *: * ** :*	
<i>E. aediculatus</i>	RKVSKLTTEEIKKLYDDIDKLKLQVNVEHKQVENMTLEKVRSAKERL*-	463
<i>E. octocarinatus</i>	KKVARLTTEEIKKLYDIDKLNLQVDVEHKQVEHMTLERARSASRER---	458
<i>E. crassus</i>	KKVSKLTDEIKKLYDEIDKLSLKNEELVQKNRESAFGKIRSTSRERMK*	455
<i>E. focardii</i>	RKVAKLTTEEIKKLYDQIDKLNLRLKGENIKNRDNTLERLRNISKERQN*	454
	:*:*:*:***** *****.*: . : .. : : * .*:**	

E. aediculatus -----MSEQV-----LKHPKVQEEIKKIKEENSAKTKQLMA 31
P. tetraurelia --MQDQOKIVILANQLEELRKLNKDQKELIFQYEEQNGAAEKKINALNQIHQDKVRNLMN 58
T. pyriformis MSAFDREKFLALKQRNKDLELKLESIEEVITQYKANESAADQKIKALNSIHQDKIRSLMN 60
 ::: :. :*: :. :. * :. **

E. aediculatus SINKLKDDLEKEYQNQDNVRAKVIERMKKDIADNETVIEMMRTMINDD----DAINKEI 87
P. tetraurelia SITLLKKENAQLKNMNKEHKRSQLEQLNQEIADQDLVIQTLRSIVNND----QRCDNQI 114
T. pyriformis SIQLLKKENAQLEKLTKEHKRSELITQLQKDVSDQDVIIETIREVCKNRNISDEVLDNAI 120
 ** **.: : : :*: *: *: : : : : *

E. aediculatus VKVLSKGPFRVRVLSREELRMEIKRLTGEVNSLKL- 122
P. tetraurelia IEALNKGPPKIKALTREELKIENKRLDNQLRALKEQ 150
T. pyriformis VDALNKGPPRIRAPTREELKMEIKKHKATI----- 150
 :..*.*** :. :. :****:* *: :

<i>E.aedic.</i>	91	LSKGPERVRVLSREELRMEIK
<i>S.lemn.</i>	121	LDKGPKRIRVASREELKMDI-
<i>O.trif.</i>	121	LDKGPKRIRVASREELKMDI-
<i>T.therm.</i>	124	LNKGPPRIRAPTREELKMEIK
<i>I.multi.</i>	104	LNKGPPRIRAPTREELKMEIK