

Novel Ubiquitin Fusion Proteins: Ribosomal Protein P1 and Actin

John M. Archibald^{1*}, Evelyn M. Teh² and Patrick J. Keeling¹

¹Canadian Institute for Advanced Research, Program in Evolutionary Biology
Department of Botany
University of British Columbia
3529-6270 University Boulevard, Vancouver BC V6T 1Z4, Canada

²Department of Biochemistry and Molecular Biology
University of British Columbia
2146 Health Sciences Mall
Vancouver, BC V6T 1Z3
Canada

Ubiquitin is a small, highly conserved protein found in all eukaryotic cells. Through its covalent attachment to other proteins, ubiquitin regulates numerous important cellular processes including apoptosis, transcription, and the progression of the cell cycle. Ubiquitin expression is unusual: it is encoded and expressed as multimeric head-to-tail repeats (polyubiquitins) that are post-translationally cleaved into monomers, or fused with ribosomal proteins L40 and S27a. The ubiquitin moiety is removed from these fusion proteins, but is thought to act as a chaperone in ribosome biogenesis prior to cleavage. Here we show that the chlorarachniophyte algae express several novel ubiquitin fusion proteins. An expressed sequence tag (EST) survey revealed ubiquitin fusions with an unidentified open reading frame (ORF), ribosomal protein P1 and, most interestingly, actin. Actin is an essential component of the eukaryotic cytoskeleton and is involved in a variety of cellular processes. In other eukaryotes, actin genes only exist as stand-alone ORFs, but in all chlorarachniophytes examined, actin is always encoded as a ubiquitin fusion protein. The variety of ubiquitin fusion proteins in these organisms raises interesting questions about the evolutionary origins of ubiquitin fusions, as well as their possible biochemical functions in other processes, such as cytoskeletal regulation.

© 2003 Elsevier Science Ltd. All rights reserved

Keywords: ubiquitin; actin; ribosomal proteins; gene fusions; chlorarachniophyte algae

*Corresponding author

Introduction

The selective degradation of short-lived proteins in eukaryotic cells is mediated by ubiquitin, a 76 amino acid residue globular protein that is one of the most highly conserved eukaryotic proteins known.¹ Ubiquitin exerts its effect through covalent attachment to substrate proteins. Polyubiquitinated substrates are recognized by a multisubunit protease, the proteasome,² resulting in ATP-dependent proteolysis of the protein and release of free ubiquitin monomers. Ubiquitin-dependent protein degradation plays a central role in a large number of fundamental biochemical processes, including regulation of the cell cycle, apoptosis, and signal transduction.³

While the ability of ubiquitin to act as a signal for proteasomal degradation is well established, it

is becoming increasingly clear that ubiquitination also plays a more general role in intracellular signaling. Ubiquitin is conjugated to substrates *via* an isopeptide bond between its C-terminal glycine (G76) residue and the ϵ -amino group of a target lysine, a process carried out by the action of a specific set of enzymes.⁴ Additional ubiquitin moieties are then added to form branched ubiquitin-ubiquitin conjugates, and isopeptide linkages between G76 and lysine 48 (K48) or K29 of ubiquitin target the protein to the proteasome-dependent degradation pathway.⁵ However, alternate linkages between ubiquitin monomers (e.g. between G76 and K63) appear to act as proteolysis-independent signals in the regulation of a variety of different processes, including ribosome function,⁶ transcription,⁷ mitochondrial DNA inheritance,⁸ and DNA repair.⁹ The addition of one or a few ubiquitins to certain proteins is known to function as a signal for endocytosis^{10,11} as well as play a role in chromatin remodeling and transcriptional regulation.¹² A great deal of specificity thus appears to reside in both the length and nature of

Abbreviations used: EST, expressed sequence tag; ORF, open reading frame.

E-mail address of the corresponding author: jarch@interchange.ubc.ca

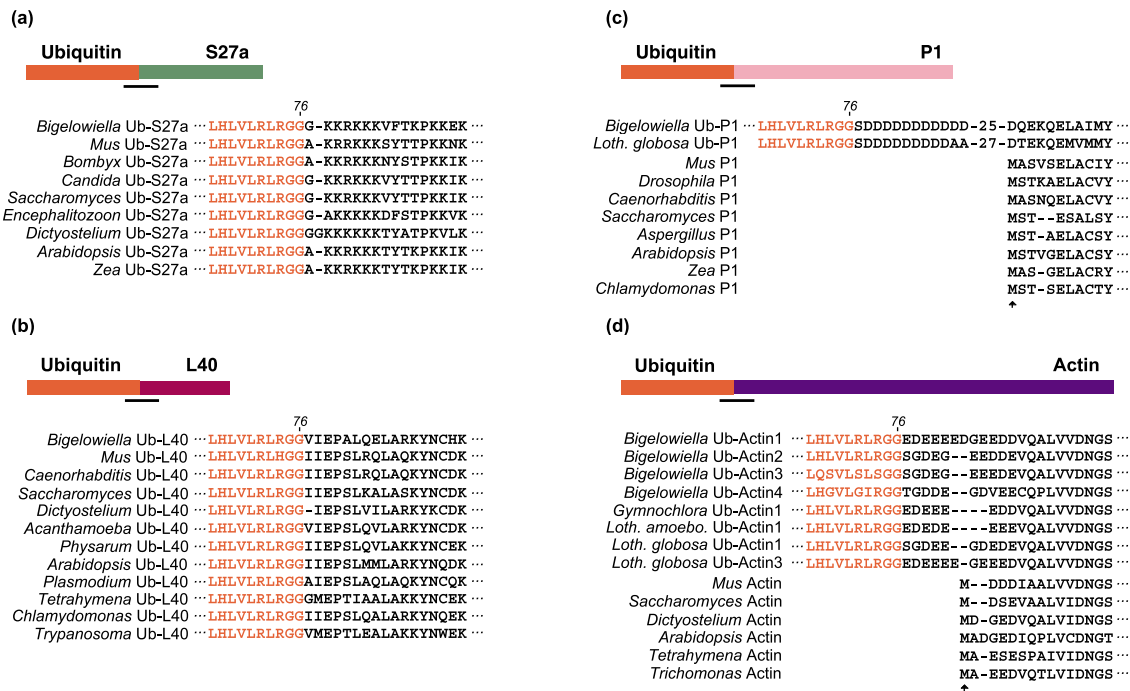


Figure 1. Ubiquitin fusions identified in *B. natans* and other chlorarachniophyte algae. (a) Schematic of the *B. natans* ubiquitin-ribosomal protein S27a fusion and the amino acid sequence of the region flanking the ubiquitin-S27a boundary (roughly corresponding to the filled bar in the schematic). For comparison, the sequence is aligned with homologs from animals, fungi, and plants. (b) Ubiquitin-ribosomal protein L40 fusion. Animal, fungal, plant and protist fusions are shown. (c) Novel ubiquitin-ribosomal protein P1 fusions identified in *B. natans* and *L. globosa*. Both contain a highly acidic region linking ubiquitin to P1. Sequences are shown aligned with the N termini of canonical P1 proteins from other eukaryotes. (d) Novel ubiquitin-actin fusions identified in *B. natans*, *L. amoebiformis*, *L. globosa* and *G. stellata*. Numbers correspond to distinct ubiquitin-actin fusions identified from the same organism. Ubiquitin sequences are colored orange, and glycine 76, an essential residue for ubiquitin conjugation (see the text) is indicated. The positions of the N-terminal methionine residues on the canonical P1 (c) and actin (d) proteins are highlighted by arrows. In the schematics, the lengths of the fusion proteins are shown roughly to scale. New sequences are available at <http://www.botany.ubc.ca/keeling/ubac/suppl.html>. Abbreviations: *Loth*; *Lotharella*; *amoeba*, *amoebiformis*.

the ubiquitin conjugates, targeting substrate proteins for a large number of distinct intracellular pathways.

Ubiquitin genes typically exist as part of a large multigene family and are found in three general forms. In one instance, the ubiquitin monomer is encoded by a stand-alone open reading frame (ORF).¹³ More often, however, they are found as polymers of head-to-tail ubiquitin coding regions (polyubiquitins),¹⁴ or as N-terminal fusions with two different ribosomal proteins, L40 and S27a (S37 in yeast).^{15,16} Free ubiquitin is generated by the post-translational cleavage of ubiquitin polyproteins or ubiquitin extension proteins by specific proteases.^{17,18} In the case of ubiquitin-ribosomal protein fusions, the ubiquitin moieties have been suggested to play a role in ribosome biogenesis,¹⁹ although in yeast, L40 and S37 can function without N-terminal ubiquitins if they are overexpressed.^{19,20}

To date, only S27a and L40 are known to be expressed as ubiquitin fusion proteins in eukaryotic cells, but here we describe three novel ubiquitin fusions in the chlorarachniophyte algae. Chlorarachniophytes are a small group of photosynthetic amoeboid flagellates, characterized by their

long pseudopodia and unusual evolutionary history.^{21,22} Their plastids (chloroplasts) were acquired through a process called secondary endosymbiosis, in which a non-photosynthetic heterotrophic eukaryote engulfed a green alga and retained its photosynthetic apparatus. The nucleus and nuclear genome of the algal endosymbiont also persist, albeit in a severely reduced form.²³ In addition to the well-characterized fusions involving ribosomal proteins S27a and L40, we have found that chlorarachniophytes contain ubiquitin fusions involving an unidentified protein, the acidic ribosomal protein P1 and, most surprisingly, the cytoskeletal protein actin. Actin is a critical component of the eukaryotic cytoskeleton that is involved in diverse processes such as intracellular trafficking, determination of cell shape, cell cycle, and motility. Actin functions by polymerizing and depolymerizing polar microfilaments, one of the central scaffolds of all eukaryotic cells. This process is extremely dynamic and is tightly regulated by the availability of free actin monomers in the cytosol, which are recycled from depolymerizing microfilaments. In chlorarachniophytes, every actin gene appears to be expressed as a ubiquitin fusion protein, raising a number of questions

<i>Bigelowiella</i> polyUb1	[MS]MQIFVKTLTGKTI
<i>Arabidopsis</i> polyUb	MQIFVKTLTGKTI ...
<i>Bigelowiella</i> Ub-S27a	[MS]MQIFVKTLTGKTI ...
<i>Arabidopsis</i> Ub-S27a	MQIFVKTLTGKTI ...
<i>Bigelowiella</i> Ub-L40	[MS]MQIFVKTLTGKTI ...
<i>Arabidopsis</i> Ub-L40	MQIFVKTLTGKTI ...
<i>Bigelowiella</i> Ub-P1	[MS]MQIFVKTLTGKTI ...
<i>Bigelowiella</i> Ub-Actin2	[MS]MQIFVKTLTGKTI ...

Figure 2. Amino acid sequence alignment of the inferred N-terminal regions of the *B. natans* polyubiquitin 1 and ubiquitin fusion proteins. The MS di-peptide extension unique to *B. natans* is highlighted. For reference, the *Arabidopsis thaliana* polyubiquitin, ubiquitin-S27a and ubiquitin-L40 N termini are also shown. The canonical initiation methionine is indicated by an arrow.

regarding potential roles for ubiquitin in cytoskeleton dynamics, as well as how ubiquitin fusion proteins originate and evolve.

Results

Ubiquitin genes in *Bigelowiella natans*

Using an expressed sequence tag (EST) sequencing approach, we have identified ubiquitin and ubiquitin-like genes in a variety of contexts in the chlorarachniophyte *Bigelowiella natans* (also known as *Chlorarachnion* sp. CCMP 621). *B. natans* has previously been shown to possess three polyubiquitin genes containing three or four ubiquitin repeats.²⁴ These genes are similar to other known polyubiquitins, except that they contain a two amino acid residue N-terminal extension and a single amino acid insertion at each of the monomer-monomer junction points. As in other eukaryotes, *B. natans* also expresses ubiquitin fusion proteins involving ribosomal proteins S27a and L40 (Figure 1(a) and (b)), which are not unusual. Both the ubiquitin moiety and ribosomal proteins themselves are highly similar to those found in other organisms (Figure 1(a) and (b)). The *B. natans* S27a begins with a lysine and arginine-rich region, as in animal, fungal and plant homologs (Figure 1(a)). In contrast to other eukaryotes, *B. natans* also possesses an additional ribosomal protein fusion involving the acidic protein P1 (Figure 1(c)). The ubiquitin and P1 moieties are extremely similar to ubiquitin and P1 proteins in other eukaryotes, but the fusion is unique in structure in that it contains a long, highly acidic linker region, which begins with a serine and 11 consecutive aspartic acid residues.

Most interestingly, however, all actin genes expressed by *B. natans* were found to exist as ubiquitin-actin fusions (Figure 1(d)). The *B. natans* host nuclear genome has been shown to encode at least four distinct actin genes,²⁵ but these were partial sequences obtained by PCR, and therefore did not include the N-terminal coding region. Transcripts from two of these genes were identified from ESTs, and both encode N-terminal ubiquitin extensions. The actin-2 gene was by far the most highly expressed, with 23 independent clones identified, while only a single transcript from

actin-3 was sequenced. The actin-2 ubiquitin extension is highly similar to the ubiquitin monomers present in *B. natans* polyubiquitin genes and to the ubiquitin moieties of the S27a and L40 fusion proteins. In contrast, the ubiquitin extension encoded by actin-3 was more divergent at the protein level than is typical of ubiquitin (the cDNA encoding this gene was slightly truncated at its 5' end). To determine whether the other known *B. natans* actins exist as stand-alone genes, ubiquitin-actin fusion genes were amplified from genomic DNA. Using degenerate primers that flank the fusion region, genes encoding additional *B. natans* ubiquitin-actin fusions were amplified and sequenced, and found to correspond to the previously characterized, partial actin-1 and actin-4 genes. Therefore, all known *B. natans* actins contain N-terminal ubiquitin extensions (Figure 1(d)). Even if the *B. natans* genome encodes a stand-alone actin, which appears unlikely, it is not expressed at a significant level, since only fusion proteins were identified from the EST survey.

Interestingly, the MS di-peptide extension found on the first repeat of the *B. natans* polyubiquitin genes²⁴ is also present at the N termini of all ubiquitin moieties in *B. natans*, regardless of their context (Figure 2). This characteristic is unique to chlorarachniophytes, and suggests that the otherwise universally conserved N terminus can tolerate this extension, or that chlorarachniophyte cells utilize a unique processing step in ubiquitin maturation. Apart from this, the ubiquitin moieties on actin and other fusion proteins are predicted to function normally, since they are generally highly conserved (with the exception of the actin-3 ubiquitin extension), and possess the functionally critical C-terminal glycine and internal lysine residues necessary for conjugation.⁴

In addition to the novel ubiquitin fusion proteins, an unusual fusion gene involving a 5'-truncated ubiquitin-like ORF and a 120 amino acid residue ORF sharing no significant similarity with known proteins was also identified (not shown). This ubiquitin-like protein, although clearly related to ubiquitin, is not particularly similar to any of the other well-characterized classes of ubiquitin-like proteins.

Ubiquitin fusions in other chlorarachniophytes

The presence of ubiquitin-actin and ubiquitin-P1 fusion proteins in *B. natans* raises the obvious question of whether they are unique to this organism or not. To address this issue, fusion genes were sought from diverse chlorarachniophyte genera. A ubiquitin-P1 fusion gene was amplified from *Lotharella globosa* and, as is the case in *B. natans*, the *L. globosa* ubiquitin-P1 is characterized by a long, acidic linker region (Figure 1(c)). ubiquitin-actin fusions were also isolated from *L. globosa*, as well as from two other chlorarachniophytes, *L. amoebiformis* and *Gymnochlorella stellata* (Figure 1(d)). Multiple ubiquitin-actin fusion genes were

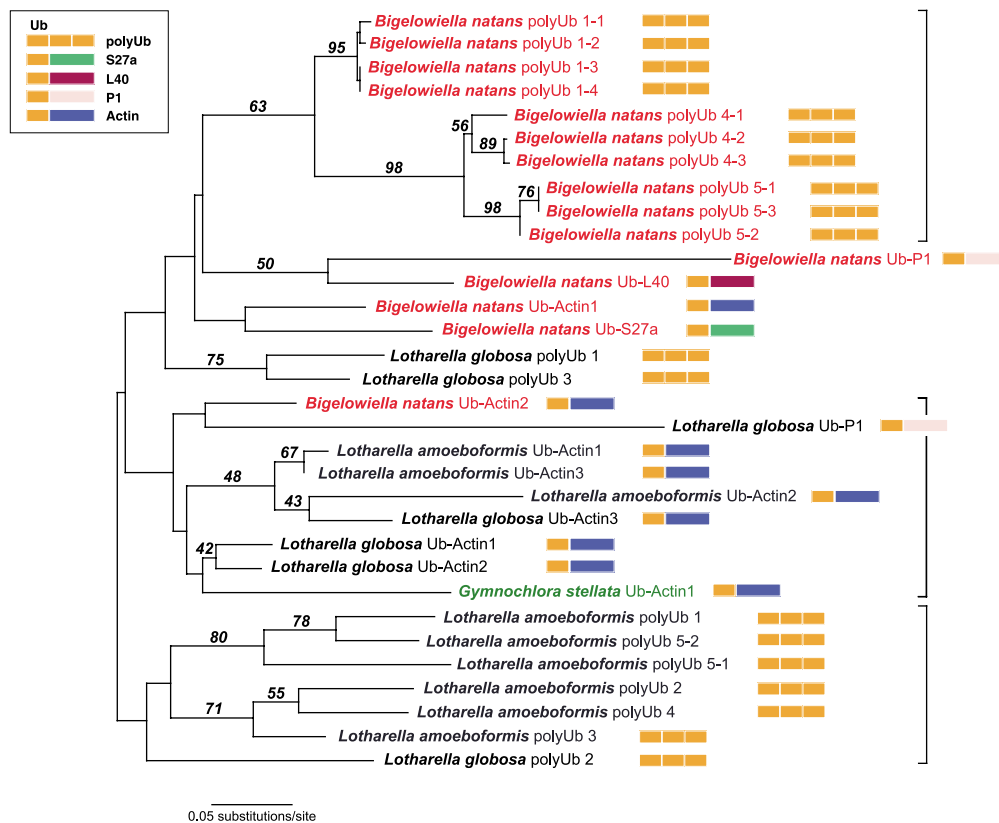


Figure 3. Phylogenetic analysis of chlorarachniophyte ubiquitin monomer DNA sequences derived from polyubiquitin genes and ubiquitin fusions. The tree shown is a maximum likelihood-distance tree inferred from an alignment of 32 ubiquitin monomers and 208 nucleotide positions (see the text). Sequences are colored according to the species from which they were obtained, and colored boxes indicate that the ubiquitin monomer sequence is part of a polyubiquitin gene or an S27a, L40, P1 or actin fusion. Numbers after sequence names correspond to the gene number and, if part of a polyubiquitin gene, the ubiquitin repeat number. Statistical support for individual nodes on the tree (obtained from an analysis of 1000 bootstrap replicates in PAUP⁴¹) are shown where greater than 40%.

present in the two *Lotharella* species: three distinct coding regions were obtained from *L. amoebiformis* and *L. globosa*, presumably corresponding to duplicate genes. Overall, the copies are very similar to one another, for the most part differing only at synonymous sites (data not shown). However, some variability in the length of the region linking the two ORFs was observed among the ubiquitin-actin genes in *L. globosa* (Figure 1(d)). A single ubiquitin-actin fusion gene was also isolated from *G. stellata*.

Using the same approach, we attempted to isolate the ubiquitin-actin fusion from organisms that are known to be related to the chlorarachniophytes, including the cercomonad flagellates, plasmodiophorid plant pathogens, euglyphid amoebae, as well as the foraminiferans, which have recently been shown to be distantly related to these lineages.^{24,25} Despite numerous attempts, no actin fusion proteins were characterized. While this does not prove that the ubiquitin-actin fusion gene is not present in these organisms, these results are consistent with the idea that the ubiquitin-actin gene is specific to chlorarachniophyte algae, and that the fusion proteins were assembled after the origin of chlorarachniophytes.

The post-translational processing of the ubiquitin moieties from ribosomal proteins S27a and L40 has been demonstrated in several different systems (for example, see Finley *et al.*¹⁹), raising the possibility that the novel fusion proteins are also processed. We attempted to determine whether the ubiquitin moiety is cleaved from the actin fusion protein by Western blot analysis using actin antibodies. Whole protein extracts from *B. natans* and *L. amoebiformis* were probed with one polyclonal and two different monoclonal actin antibodies that recognize epitopes at the N or C terminus of actin. Surprisingly, none of the antibodies recognized the chlorarachniophyte actin despite the fact that all of the antibodies did react with a purified actin positive control, and the target epitopes were present in the inferred actin amino acid sequences (not shown). Two other proteins (clathrin and calmodulin) were successfully recognized by monoclonal antibodies against *B. natans* protein extracts (not shown), indicating that the negative results were not strictly the result of the protein isolation. At present, therefore, it is unclear whether the novel ubiquitin moieties on the chlorarachniophyte P1 and actin proteins are removed or exist as integral components of the mature proteins in these

cells, but it seems most likely that they are removed.

Origin and evolution of the novel ubiquitin fusions

The evolution of novel ubiquitin gene fusions is interesting in that it not only involves the duplication of pre-existing ubiquitin genes, but the in-frame addition of the ubiquitin coding region onto the 5'-end of multiple fusion partners. Ubiquitin-L40 and -S27a fusions appear to be universally distributed amongst eukaryotes, indicating that they formed very early in eukaryotic evolution. This is clearly not the case with the chlorarachniophyte P1 and actin fusion proteins. Therefore, we performed phylogenetic analyses of ubiquitin monomers from chlorarachniophyte polyubiquitin and ubiquitin fusion proteins in an attempt to determine where the ubiquitin moieties of the more recent ubiquitin fusion proteins originated and how they evolve within the ubiquitin multigene family.

Figure 3 shows a maximum likelihood-distance tree constructed from an alignment of 32 ubiquitin monomer DNA sequences including polyubiquitins and all four fusion proteins. The ubiquitin monomers making up the polyubiquitin genes of a given organism cluster together to the exclusion of other sequences. This pattern is indicative of concerted evolution, whereby duplicate genes within a genome are repeatedly homogenized by recombination, and thus remain highly similar to one another. The concerted evolution of polyubiquitin genes has been demonstrated in a variety of eukaryotic genomes (for example, see Refs. 26–29). However, the phylogeny also indicates that, in *B. natans*, polyubiquitin monomers are more similar to their intragenic neighbors than to intergenic copies of the repeat. Repeats one to four of the polyubiquitin-1 gene form a highly supported group to the exclusion of all the other ubiquitins in *B. natans* and the other organisms, and the same is true for polyubiquitins-2 and -3. This suggests that recombination is occurring more frequently within a gene than between different genes.

In contrast to the pattern observed for the polyubiquitin genes, the ubiquitin moieties from the actin fusion proteins cluster with one another rather than with ubiquitins from different contexts in the same genome. The single exception is the *B. natans* ubiquitin-actin-1 monomer, which branches with ubiquitin-S27a, perhaps suggesting a recent cross-context recombination event. This “actin” cluster also contained the ubiquitin-P1 from *L. globosa*. Although the grouping of the ubiquitin-actin genes was not well supported, this analysis suggests that, for the most part, the ubiquitin-actins have not recombined with other ubiquitins since the origin of the ubiquitin-actin fusion in the chlorarachniophytes, but it does not suggest an obvious source for the ubiquitin moiety of the newly formed fusion protein.

Regarding the ubiquitin-P1 and -L40 fusions, the *B. natans* ubiquitin-P1 monomer does not branch with the ubiquitin-P1 from *L. globosa*, but instead clusters with the *B. natans* ubiquitin-L40. This surprising relationship is also supported by the presence of a shared intron: one of the four ubiquitin-L40 cDNAs sequenced from *B. natans* contained a 244-nucleotide intron in the ubiquitin coding region, and the genomic copy of the *B. natans* ubiquitin-P1 revealed the presence of a 62-nucleotide intron in the same position and phase. This indicates a cross-context recombination between the ubiquitin moieties of *B. natans* P1 and L40. Curiously, the ubiquitin-P1 genes from *B. natans* and *L. globosa*, which do not branch together in the phylogeny, also share an intron in a different region of the ubiquitin. If gene conversion and/or unequal crossing over were actively homogenizing all ubiquitin repeats in a genome, one would expect all (or none) of the ubiquitin monomers to have introns at the same positions, which is clearly not the case. Moreover, the large size difference and lack of detectable sequence similarity between the putatively homologous *B. natans* ubiquitin-P1 and -L40 introns suggests that, while the ubiquitins appear more similar to each other than to any of the other ubiquitins, they have not recombined recently. Overall, the ubiquitin coding regions in these genomes are all recombining to some extent, but at very different rates. The ubiquitin extensions of fusion proteins are apparently recombining the least, likely due to their restrictive context. The presence of introns within the ubiquitin coding regions likely further reduces their capacity to recombine with ubiquitins in other contexts. However, despite all these potential barriers to recombination the clear relationship between the ubiquitin moieties of the *B. natans* P1 and L40 fusions indicates that cross-context recombination can still occur.

It is also significant that all ubiquitin coding regions in the *B. natans* genome were found to possess the unique MS di-peptide N-terminal extension (Figure 2). Since the ubiquitin-S27a and -L40 fusions clearly originated long before the unique MS extension in *B. natans*, there has almost certainly been at least some recombination between all of the genes, regardless of their context, such that the extension spread to all genes.

Discussion

We have described several novel ubiquitin fusion genes in the chlorarachniophyte alga *Bigeloniella natans*. Unlike the ubiquitin-L40 and -S27a fusions, which appear to be a universal feature of eukaryotic cells, the ubiquitin-actin and ubiquitin-P1 fusions have arisen relatively recently, within the evolution of the chlorarachniophyte algae. The presence of novel ubiquitin fusion proteins in chlorarachniophytes raises intriguing questions about the origin and evolution of ubiquitin fusion

proteins, their processing, and their possible biochemical functions.

In the case of S27a and L40 fusion proteins of other eukaryotes, the ubiquitin moieties are post-translationally cleaved by members of a specific family of proteins: the ubiquitin-specific proteases (ubps).^{17,18} Ubps likely function in chlorarachniophyte cells as well, although no ubp genes were identified in our EST survey. These enzymes appear to be relatively insensitive to the amino acid sequence downstream of the C-terminal glycine residue of ubiquitin.³⁰ Prior to their processing, the S27a and L40 ubiquitins are thought to facilitate the assembly of these proteins into nascent ribosomes.^{19,20} While it has been shown in yeast that both S27a and L40 can function when not translated as a fusion, the stand-alone ribosomal proteins must be over-expressed significantly in order to make up for the loss of the ubiquitin moiety.^{19,20} Furthermore, the maintenance of these proteins as ubiquitin-fusions across the entire diversity of eukaryotes suggests an important role for ubiquitin in S27a and L40 maturation. As P1 is also a ribosomal protein, essentially the same reasoning can be applied to it as to S27a and L40.

Actin, however, is very different. Unlike the ribosomal proteins, actin assembles and disassembles in an extremely dynamic and highly controlled process that is at the foundation of its function in the cell. It is likely that the ubiquitin-actin fusion proteins are processed because ubps cleave ubiquitins regardless of the downstream amino acid sequence,³⁰ so ubiquitin units should be removed from any fusion protein. In addition, the structural and functional constraints on actin monomers are severe, as reflected in the very high degree of conservation of actin sequences across the whole diversity of eukaryotes, including the chlorarachniophytes.²⁵ The folding of actin monomers has been studied in considerable detail, and is known to be dependent on the cytosolic chaperonin complex CCT.^{31–35} The CCT complex is composed of eight distinct subunits,^{33,36} and in the course of our EST survey, we identified *B. natans* cDNAs encoding five of these (alpha, gamma, delta, epsilon, and theta). The inferred protein sequences of these five subunits are highly conserved (not shown), suggesting that in chlorarachniophytes, actin–CCT interactions are similar to those in other eukaryotic cells. It thus seems unlikely that chlorarachniophyte actin monomers could fold and polymerize as ubiquitin fusion proteins, or that ubiquitin-tagged microfilaments could function properly. Therefore, if the actin fusion is processed, what is the role of ubiquitin? Actin monomers are continuously recycled in a well-characterized ATP/ADP nucleotide exchange-dependent process.^{37,38} If, prior to processing, the ubiquitin extension is acting as a chaperone in actin polymerization, actin monomers could not be recycled, and microfilament dynamics would be significantly altered. A more interesting possibility is that the cleavage of ubiquitin exten-

sions from nascent actin proteins plays a more indirect role in the cytoskeleton dynamics of chlorarachniophytes by “activating” monomers and regulating the amount of free actin available for microfilament assembly. Interestingly, a link between ubiquitination and cytoskeleton regulation has recently been suggested in yeast.³⁹

With respect to the origin of ubiquitin fusions in eukaryotic cells, the presence of these diverse fusion proteins and the nature of the ubiquitin processing system raise an interesting possibility. The pre-existence of a system for the post-translational removal of ubiquitin monomers from the amino terminus of other proteins (including itself in the case of polyubiquitins) would allow new fusion proteins with no initial function to arise. Ubiquitin coding regions could fuse with any protein-coding gene as long as processing of their protein product took place before any deleterious effect was felt. Initially, therefore, ubiquitin fusions may simply be accidents that happen to satisfy the expression requirements of both genes. Once established, such “neutral” fusions could later acquire functions that render them indispensable, as has been suggested in other biochemical processes.⁴⁰ At the extreme, it is possible that the fusion proteins never have any function, and that ubiquitin is simply “piggy-backing” on the expression of other genes. The fact that all known fusion proteins encode only one ubiquitin monomer argues against this extreme possibility: this common and significant feature suggests that ubiquitin plays some role in the maturation or activity of its fusion partners.

Materials and Methods

Strains, culture conditions and DNA extractions

A culture of *B. natans* (*Chlorarachnion* sp. CCMP 621) was obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP). Cultures of *L. amoebiformis* (strain CCMP 2058), *L. globosa* (strain CCMP 1729) and *G. stellata* were provided by K. Ishida (Kanazawa University, Japan). All cultures were grown in f/2-Si medium at 20°C with a 16-hour light and eight-hour dark cycle. DNA extractions were performed using the DNeasy Plant Mini Kit (Qiagen) on cells harvested from 75 ml cultures. A *B. natans* lambda Zap II cDNA library was kindly provided by G. I. McFadden and P. Gilson.

Amplification, cloning and sequencing of ubiquitin fusion genes

Ubiquitin fusion genes from *B. natans* were sequenced in the course of an ongoing cDNA sequencing project. Ubiquitin-actin fusion genes were amplified from *L. amoebiformis*, *L. globosa* and *G. stellata* genomic DNAs using a ubiquitin forward primer and two different actin reverse primers (UBIQ1: 5'-GGCCATGCARATHTTYGTNAARAC-3'; ACTR1: 5'-GGCCTGGAARCAYTTNCGRTGNAC-3'; ACTR2: 5'-AGCGCGTANCCYTRTADATNGGNAC-3'). The ubiquitin-P1 fusion gene was

amplified from *G. stellata* and *B. natans* genomic DNAs using UBIQ1 (above) and a reverse primer (P1R1: 5'-CCGAACATCTTAGGCCNRWANGG-3'). Amplifications were carried out as follows: after an initial three minutes denaturation at 94°C, 40 or 45 cycles of 45 seconds at 92°C, one minute at 50°C, and one or 1.5 minute at 72°C were performed. All reactions were finished with a final five minutes extension at 72°C. Fusion products of the expected size were isolated with the Ultra-Clean™15 DNA purification kit (MO BIO Laboratories) and cloned into pCR2.1 using the TOPO TA cloning kit (Invitrogen). Multiple independent clones were sequenced for each gene.

Phylogenetic analyses

The extreme sequence and absolute length conservation of ubiquitin allowed a DNA sequence alignment to be constructed manually. The alignment contained 32 chlorarachniophyte ubiquitin monomer sequences and 208 nucleotide positions, after the removal of regions corresponding to amplification primer sites (*B. natans* ubiquitin-actin genes-3 and -4 were excluded from the analysis, as the former was a 5'-truncated cDNA and the latter was too divergent to be analyzed reliably). Maximum likelihood (ML) and ML-distance (minimum evolution) trees were inferred from all three codon positions of the alignment. Trees were constructed using PAUP* version 4.0b10⁴¹ with a general time reversible plus Γ plus invariable sites (GTR + Γ + P_{INV}) model using the heuristic search option. A Γ distribution was approximated by four rate categories and the Γ shape parameter α , proportion of invariable sites parameter (P_{INV}) and base frequencies were estimated from the data in PAUP* ($\alpha = 0.84$, $P_{INV} = 0.51$). Starting trees for TBR branch swapping were obtained with the neighbor-joining method or by ten random sequence addition replicates. Support for ML and ML-distance trees was determined by bootstrapping with 100 or 1000 resampling replicates, respectively, using PAUP*.

GenBank accession numbers

New sequences were deposited in GenBank under the following accession numbers: AY251792-AY251815.

Acknowledgements

We thank K. Ishida for chlorarachniophyte cultures and G. I. McFadden and P. Gilson for a *B. natans* cDNA library. We acknowledge Dr Les Burtnik (Department of Chemistry, University of British Columbia) for providing an actin sample. N. Fast is also thanked for helpful comments on the manuscript. This work was supported by a grant (227301-00) from the Natural Science and Engineering Research Council of Canada (NSERC) to P.J.K. J.M.A. is supported by postdoctoral fellowships from the Canadian Institutes of Health Research (CIHR) and the Killam Foundation (University of British Columbia). E.M.T. acknowledges postdoctoral support from CIHR and the Canadian Blood Services (CBS). P.J.K. is a scholar of the Canadian Institute for Advanced Research, CIHR,

and the Michael Smith Foundation for Health Research.

References

- Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**, 405–439.
- Baumeister, W., Walz, J., Zuhl, F. & Seemuller, E. (1998). The proteasome: paradigm of a self-compartmentalizing protease. *Cell*, **92**, 367–380.
- Hershko, A. & Ciechanover, A. (1998). The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479.
- Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533.
- Pickart, C. M. (2000). Ubiquitin in chains. *Trends Biochem. Sci.* **25**, 544–548.
- Spence, J., Gali, R. R., Dittmar, G., Sherman, F., Karin, M. & Finley, D. (2000). Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. *Cell*, **102**, 67–76.
- Kaiser, P., Flick, K., Wittenberg, C. & Reed, S. I. (2000). Regulation of transcription by ubiquitination without proteolysis: Cdc34/SCF(Met30)-mediated inactivation of the transcription factor Met4. *Cell*, **102**, 303–314.
- Fisk, H. A. & Yaffe, M. P. (1999). A role for ubiquitination in mitochondrial inheritance in *Saccharomyces cerevisiae*. *J. Cell Biol.* **145**, 1199–1208.
- Spence, J., Sadis, S., Haas, A. L. & Finley, D. (1995). A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol. Cell. Biol.* **15**, 1265–1273.
- Shih, S. C., Sloper-Mould, K. E. & Hicke, L. (2000). Monoubiquitin carries a novel internalization signal that is appended to activated receptors. *EMBO J.* **19**, 187–198.
- Terrell, J., Shih, S., Dunn, R. & Hicke, L. (1998). A function for monoubiquitination in the internalization of a G protein-coupled receptor. *Mol. Cell*, **1**, 193–202.
- Robzyk, K., Recht, J. & Osley, M. A. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. *Science*, **287**, 501–504.
- Krebber, H., Wostmann, C. & Bakker-Grunwald, T. (1994). Evidence for the existence of a single ubiquitin gene in *Giardia lamblia*. *FEBS Letters*, **343**, 234–236.
- Ozkaynak, E., Finley, D. & Varshavsky, A. (1984). The yeast ubiquitin gene: head-to-tail repeats encoding a polyubiquitin precursor protein. *Nature*, **312**, 663–666.
- Ozkaynak, E., Finley, D., Solomon, M. J. & Varshavsky, A. (1987). The yeast ubiquitin genes: a family of natural gene fusions. *EMBO J.* **6**, 1429–1439.
- Redman, K. L. & Rechsteiner, M. (1989). Identification of the long ubiquitin extension as ribosomal protein S27a. *Nature*, **338**, 438–440.
- Baker, R. T., Tobias, J. W. & Varshavsky, A. (1992). Ubiquitin-specific proteases of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **267**, 23364–23375.
- Jonnalagadda, S., Butt, T. R., Monia, B. P., Mirabelli, C. K., Gotlib, L., Ecker, D. J. & Crooke, S. T. (1989). Multiple (alpha-NH-ubiquitin) protein endoproteases in cells. *J. Biol. Chem.* **264**, 10637–10642.
- Finley, D., Bartel, B. & Varshavsky, A. (1989). The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. *Nature*, **338**, 394–401.

20. Warner, J. R. (1989). Ubiquitin. A marriage of convenience or necessity? *Nature*, **338**, 379.
21. Hibberd, D. J. & Norris, R. E. (1984). Cytology and ultrastructure of *Chlorarachnion reptans* (Chlorarachniophyta divisio nova, Chlorarachniophyceae classis nova). *J. Phycol.* **20**, 310–330.
22. McFadden, G. I., Gilson, P. R., Hofmann, C. J. B., Adcock, G. J. & Maier, U. G. (1994). Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. *Proc. Natl Acad. Sci. USA*, **91**, 3690–3694.
23. Gilson, P. R. & McFadden, G. I. (2002). Jam packed genomes—a preliminary, comparative analysis of nucleomorphs. *Genetica*, **115**, 13–28.
24. Archibald, J. M., Longet, D., Pawlowski, J. & Keeling, P. J. (2003). A novel polyubiquitin structure in Cercozoa and Foraminifera: evidence for a new eukaryotic supergroup. *Mol. Biol. Evol.* **20**, 62–66.
25. Keeling, P. J. (2001). Foraminifera and Cercozoa are related in actin phylogeny: two orphans find a home? *Mol. Biol. Evol.* **18**, 1551–1557.
26. Keeling, P. J. & Doolittle, W. F. (1995). Concerted evolution in protists: recent homogenization of a polyubiquitin gene in *Trichomonas vaginalis*. *J. Mol. Evol.* **41**, 556–562.
27. Perelygin, A. A., Kondrashov, F. A., Rogozin, I. B. & Brinton, M. A. (2002). Evolution of the mouse polyubiquitin-C gene. *J. Mol. Evol.* **55**, 202–210.
28. Sharp, P. M. & Li, W.-H. (1987). Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. *J. Mol. Evol.* **25**, 58–64.
29. Sun, C. W., Griffen, S. & Callis, J. (1997). A model for the evolution of polyubiquitin genes from the study of *Arabidopsis thaliana* ecotypes. *Plant Mol. Biol.* **34**, 745–758.
30. Gilchrist, C. A., Gray, D. A. & Baker, R. T. (1997). A ubiquitin-specific protease that efficiently cleaves the ubiquitin-proline bond. *J. Biol. Chem.* **272**, 32280–32285.
31. Frydman, J., Nimmesgern, E., Erdjument-Bromage, H., Wall, J. S., Tempst, P. & Hartl, F. U. (1992). Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits. *EMBO J.* **11**, 4767–4778.
32. Gao, Y., Thomas, J. O., Chow, R. L., Lee, G. H. & Cowan, N. J. (1992). A cytoplasmic chaperonin that catalyzes beta-actin folding. *Cell*, **69**, 1043–1050.
33. Leroux, M. R. & Hartl, F. U. (2000). Protein folding: versatility of the cytosolic chaperonin TriC/CCT. *Curr. Biol.* **10**, R260–R264.
34. Llorca, O., McCormack, E. A., Hynes, G., Grantham, J., Cordell, J., Carrascosa, J. *et al.* (1999). Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. *Nature*, **402**, 693–696.
35. Llorca, O., Martin-Benito, J., Ritco-Vonsovici, M., Grantham, J., Hynes, G. M., Willison, K. R. *et al.* (2000). Eukaryotic chaperonin CCT stabilizes actin and tubulin folding intermediates in open quasi-native conformations. *EMBO J.* **19**, 5971–5979.
36. Kubota, H., Hynes, G. & Willison, K. (1995). The chaperonin containing t-complex polypeptide 1 (TCP-1). Multisubunit machinery assisting in protein folding and assembly in the eukaryotic cytosol. *Eur. J. Biochem.* **230**, 3–16.
37. Bamburg, J. R., McGough, A. & Ono, S. (1999). Putting a new twist on actin: ADF/cofilins modulate actin dynamics. *Trends Cell Biol.* **9**, 364–370.
38. Pollard, T. D., Blanchoin, L. & Mullins, R. D. (2000). Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 545–576.
39. Kaminska, J., Gajewska, B., Hopper, A. K. & Zoladek, T. (2002). Rsp5p, a new link between the actin cytoskeleton and endocytosis in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **22**, 6946–6948.
40. Stoltzfus, A. (1999). On the possibility of constructive neutral evolution. *J. Mol. Evol.* **49**, 169–181.
41. Swofford, D. L. (1998). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Sinauer Associates, Sunderland, Massachusetts.

Edited by J. Karn

(Received 6 January 2003; received in revised form 13 March 2003; accepted 14 March 2003)