Molecular evolution of the actin family

Holly V. Goodson* and William F. Hawse

Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA

*Author for correspondence (e-mail: hgoodson@nd.edu)

Summary

Members of the actin family have well-characterized cytoskeletal functions, but actin and actin-related proteins (ARPs) have also been implicated in nuclear activities. Previous analyses of the actin family have identified four conserved subfamilies, but many actin-related proteins (ARPs) do not fall into these groups. A new systematic phylogenetic analysis reveals that at least eight ARP subfamilies are conserved from humans to yeast, indicating that these ARPs are part of the core set of eukaryotic proteins. Members of at least three subfamilies appear to be involved in chromatin remodeling, suggesting that ARPs play ancient, fundamental roles in this nuclear process.

The actin family

The actin family is a diverse and evolutionarily ancient group of proteins. Conventional actin appears to be ubiquitous in eukaryotes and, recently, cell-division proteins bearing a striking structural similarity to actin have been identified in eubacteria (van den Ent et al., 2001). Conventional actin is one of the principal components of the eukaryotic cytoskeleton, and it has a central role in cellular processes ranging from cell motility to intracellular transport and cell organization. In the early 1990s, researchers realized that most, if not all, eukaryotic cells also contain actin-related proteins, or 'ARPs'. Some of these ARPs have well-characterized roles in cytoskeletal functions, including actin polymerization (ARP2/3) and dynein motor activity (ARP1) (reviewed by Machesky and May, 2001; Schafer and Schroer, 1999). In addition, both conventional actin and specific ARPs have been strongly implicated in the initially surprising functions of chromatin remodeling and/or transcription regulation (Machesky and May, 2001; Schafer and Schroer, 1999; Sheterline et al., 1998).

The ARP proteins have been named on the basis of their similarity to conventional actin (Schroer et al., 1994). Members of the ARP1, ARP2, ARP3,
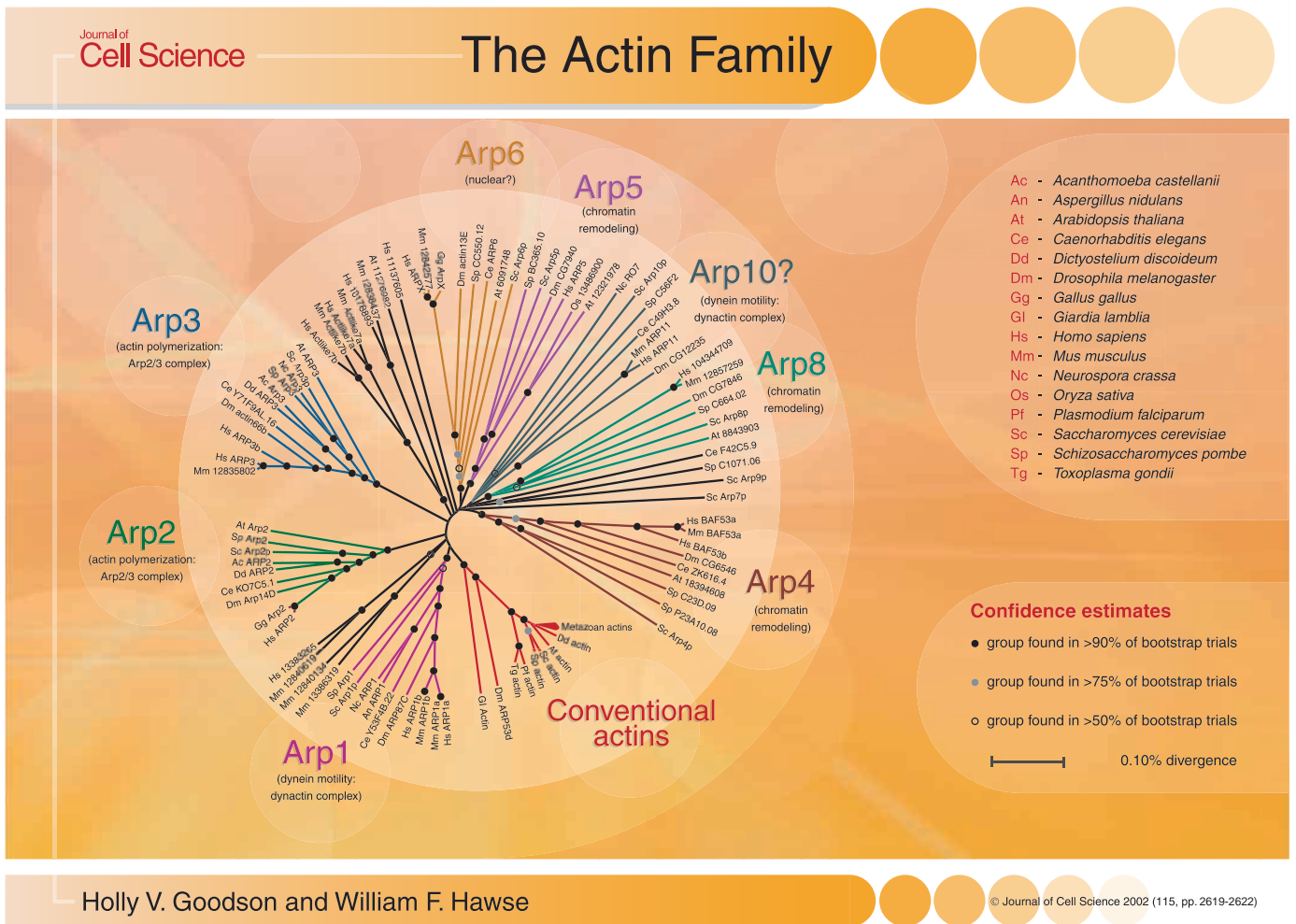
(See poster insert)

and conventional actin subfamilies have been found in organisms ranging from humans to fungi. Additional ARPs that do not fall into these subfamilies have been identified in a number of organisms, but the relationship between these proteins has thus far been unclear (Machesky and May, 2001; Schafer and Schroer, 1999). Moreover, genome-sequencing projects have produced a large number of new ARP sequences. Some are closely related to characterized proteins, but many are not. What are the functions of these novel ARPs? One of the best ways to provide a necessary set of initial hypotheses is phylogenetic analysis.

Phylogenetic analysis

Phylogenetic analysis attempts to reconstruct evolutionary relationships between proteins by studying sequence relationships. These evolutionary relationships contain a wealth of functional information. For example, protein subfamilies that predate the divergence of animals and fungi probably act in fundamental cellular processes, given that they apparently existed in the common ancestor and were important enough to be retained for the billion or so years separating these organisms. By contrast, proteins specific to particular phyla, orders or species appeared relatively recently and are likely to have more specialized functions. Defining evolutionary relationships also allows a researcher to gauge the likelihood that an uncharacterized protein has functions similar to those of a particular characterized protein. For example, proteins that are 'orthologs' (proteins related by species divergence) are likely to have similar functions, while 'paralogs' (proteins related by gene duplication) are more likely to have functions that have themselves diverged.

To address these questions as they relate to the actin superfamily, we have performed a systematic phylogenetic analysis of actin-related proteins in all fully sequenced organisms, specifically *Homo sapiens, Drosophila melanogaster, Caenorhabditis elegans, Saccharomyces cerevisiae* and *Arabidopsis thaliana.* Sequences from mouse, *Schizosaccharomyces pombe*

and selected additional organisms were included to help define branch points. The poster shows the resulting unrooted phylogenetic tree (neighbor-joining method as implemented by ClustalX, bootstrapped 1000-times; to avoid crowding, some closely related ARPs were not included).

Examination of the topology of this tree and the confidence estimates provided by the bootstrapping analysis[†] reveals that the actin superfamily contains at least eight subfamilies that have been conserved from humans to yeast (see poster and Table 1; we define groups as 'subfamilies' if they contain sequences from divergent organisms and are found in >90% of bootstrap trials). Consistent with previous studies (Poch and Winsor, 1997; Schroer et al., 1994), we propose that these subfamilies be named by the yeast ARP contained in them[‡]. By this convention, the conserved actin subfamilies are: conventional actin, Arp1, Arp2, Arp3, Arp4, Arp5, Arp6 and Arp8. Experimental evidence supports the existence of an additional subfamily conserved from humans to fungi (Arp10), although bootstrapping support for this group is weak (<50%). Five of the subfamilies (Arp4, Arp5, Arp6, Arp8, Arp10) have not been rigorously defined before, although homologies between some *Drosophila*, mammalian and fungal proteins have previously been

recognized (Eckley et al., 1999; Kato et al., 2001; Lee et al., 2001). Most subfamilies contain at least one protein that has been at least partially characterized in mammalian cells, but no members of the Arp5 or Arp8 subfamilies have been identified outside of yeast, except as 'hypothetical proteins'. All subfamilies except Arp1 and Arp10 have recognizable members in the *Arabidopsis* genome. A number of organisms, particularly mammals, possess additional 'orphan' ARPs that did not group into any of these subfamilies.

Functional predictions

The existence of members of ARP subfamilies in widely divergent organisms indicates that these proteins are ancient and suggests that they have conserved roles in fundamental aspects of cell biology. It is interesting and perhaps surprising to note that the existing experimental evidence implicates the characterized members of most of the novel ARP subfamilies in nuclear functions (Table 1). The sum of this data suggests that two of these conserved ARP subfamilies have a role in actin polymerization (Arp2, Arp3), two have a role in dynein motor function (Arp1 and the tentative subfamily Arp10), and four appear to act in the nucleus, particularly in chromatin

[†]The branches of the tree in the poster represent individual sequences, and branch lengths are proportional to distance between sequences (sequence nonidentity). The tree is 'unrooted', meaning that there is no assumption as to which sequence is closest to the common ancestor. 'Nodes' (connection points between branches) indicate sequence divergence. This divergence can occur either by gene duplication or species divergence, and the only way to resolve this ambiguity is to compare the observed protein divergence pattern with the expected organism divergence pattern (the existence of multiple proteins in the same organism is a clear indication of gene duplication). There is a positive relationship between time and branch length (ancient times are towards the center of the tree) but, this relationship is not defined and probably varies over different parts of the tree. Like any type of scientific analysis, phylogenetic trees require controls before they can be interpreted. The data rarely support different parts of a tree equally well. Which parts of the tree can be trusted? One of the best ways to address this problem is 'bootstrap analysis', a statistical tool that works by taking a sample of the aligned sequence data, building a tree, and then repeating this process many (100-1000) times (Felsenstein, 1985). Generally speaking, groups found in >90% of bootstrap trials are regarded as being 'strongly supported', those found in >75% of trials 'moderately supported', and those in >50% of trials 'suggestive'. Groupings found in less than 50% of trials are generally regarded to be uninterpretable. We have used this standard for our analysis here, defining subgroups by the deepest strongly supported node.

[‡]Nomenclature of actin-related proteins (ARPs) was originally defined by the degree of relatedness to actin itself, with increasing numbers referring to increasingly divergent actins (Schroer et al., 1994). However, this approach leads to ambiguities since many ARPs have similar levels of divergence. We propose that ARP subfamilies be named by the *S. cerevisiae* genes included in them, and that otherwise uncharacterized ARPs be given names based on the subfamily to which they belong. This approach is consistent with most of the established nomenclature and allows unambiguous naming of most uncharacterized sequences. To avoid future ambiguity, we suggest that ARPs that do not yet group into one of the defined subfamilies be given alternative ARP names (for example, based on functional characteristics or chromosomal loci) until further analyses clarify the evolutionary relationships. The poster uses previously established gene names where they exist, and either chromosomal loci/gi numbers where they do not.

Table 1. Summary of functional information for actin subfamilies

Actin subfamily	Organisms?	Localization	Essential In yeast?	Function
Actin	Human→Giardia (Ubiquitous?)	Cytoplasm[MFY], nucleus[M]	Y	Cell motility/transport[MFY], polarity[MFY], chromatin remodeling[MY]
Arp1	Human→yeast	Cytoplasm[MFY]	N (spindle alignment)	Dynein motor function (dynactin complex[MFY])
Arp2	Human→yeast, *Arabidopsis*	Cytoplasm[MFY]	Y	Actin polymerization (Arp2/3 complex[MFY])
Arp3	Human→yeast, *Arabidopsis*	Cytoplasm[MFY]	Y	Actin polymerization (Arp2/3 complex[MFY])
Arp4	Human→yeast, *Arabidopsis*	Nucleus[Y]	Y	Chromatin remodeling (SWI/SNF[M], INO80[Y], NuA4[Y], histone acetyltransferase[YM])
Arp5	Human→yeast, *Arabidopsis*	Nucleus[Y]	Y	Chromatin remodeling (INO80 complex[Y])
Arp6	Human→yeast, *Arabidopsis*	Nucleus[FY]	N	Localized to heterochromatin[F]
Arp8	Human→yeast, *Arabidopsis*	Nucleus[Y]	N	Chromatin remodeling (INO80 complex[Y])
Arp10	Likely human→yeast	Cytoplasm[MY]	?	Dynein motor function (dynactin complex[M, *N. crassa*])
Orphans				
Arps 7,9	*S. cerevisiae*, *S. pombe* (Arp 9)	Nucleus[Y]	Y/Y	Chromatin remodeling (SWI/SNF[Y], RSC complex[Y])

Superscripts ([M,F,Y]) refer to the type of organism (mammal, fly, yeast) in which the observation was made. The ARP10 subfamily is italicized to indicate that bootstrap confidence in the validity of the observed grouping is low. However, the observations that mammalian Arp11 is part of dynactin and that *Neurospora* RO7 protein is required for proper localization and function of dynein (Lee et al., 2001) support the existence of a conserved subfamily.

Yeast localization and deletion phenotype data are summarized by Harata et al., 2000 and the Saccharomyces Genome Database (http://genome-www.stanford.edu/Saccharomyces). No Arp10 deletions have been reported in the literature. Characterization of biochemical associations of conventional actin, ScARP4, ScARP5, ScARP5, ScARP7/9 and HsBAF53 with transcription machinery can be found in Cairns et al., 1998; Galarneau et al., 2000; Harata et al., 1999; Park et al., 2002; Peterson et al., 1998; Zhao et al., 1998; Shen et al., 2000. Other primary references for characterization of actin, Arp1, Arp2, Arp3 and individual Arp subfamily members can be found referenced in Machesky and May, 2001; Schafer and Schroer, 1999. Our apologies to those authors we were not able to cite due to space restrictions.

remodeling (Arp4, Arp5, Arp6, Arp8). Actin itself has been found to be a stable stoichiometric component of several chromatin remodeling complexes (see references in Table 1), which suggests that actin's participation in chromatin remodeling predates the gene duplications leading to the chromatin remodeling ARP subfamilies. While most of the 'orphan' ARPs have unknown functions, the *S. cerevisiae* proteins ARP7 and ARP9 are well-characterized members of chromatin remodeling complexes (Cairns et al., 1998; Peterson et al., 1998). It is possible that some of the orphan proteins are pseudogenes, but all of the human orphans have at least five 'hits' in the human expressed sequence tag (EST) database, and some have more than 100 hits (H.V.G. and W.F.H., unpublished). As is true for actin, it is possible that some ARPs have multiple functions. For example, recent proteomic analysis of the nucleolus has suggested that Arp2, Arp3 and actin itself may be part of this structure (Andersen et al., 2002).

It is interesting to note that Arp1 and Arp10 are not obviously present in the *Arabidopsis* genome. Are members of these subfamilies missing, unsequenced, or just unrecognizable in flowering plants? Given that *Arabidopsis* also appears to lack cytoplasmic dynein (Lawrence et al., 2001), and that

dynactin is an activator of cytoplasmic dynein, it is tempting to speculate that plants lack dynactin (Lawrence et al., 2001) and therefore lack the dynactin-associated ARPs (Arp1, Arp10).

Both actin and ARPs have previously been implicated in nuclear activities (reviewed by Machesky and May, 2001; Schafer and Schroer, 1999; Sheterline et al., 1998), although most attention by the cell biology community has focused on cytoskeletal functions of actin and ARPs. Our phylogenetic analysis shows that representatives of four apparently nuclear ARP subfamilies exist in organisms as divergent as humans, yeast and plants, and suggests that these ARPs and actin itself play ancient, fundamental and under-appreciated roles in the nucleus.

Methods

The public protein and nucleic acid databases (November 2001) were scanned for actin-related proteins using either PSI-BLAST (protein databases) or tBLASTn (nucleotide databases) (Altschul et al., 1990; Altschul et al., 1997). After initial sequence collection, databases were probed with individual yeast ARP sequences to enhance the chances of finding sequences related to these divergent proteins; all 'orphan' ARPs were also used to individually probe the databases. A final set of sequences was obtained by choosing only those sequences <95% identical and including only one sequence from organisms with multiple conventional actins. Sequences were aligned using ClustalX (Thompson et al., 1997) with default

alignment parameters. Adjustments were made in the resulting initial alignment by asking ClustalX to realign specified sequences (across the entire length) or regions (all sequences were realigned in the specified region), resulting in an otherwise good alignment that contained unnecessary gaps. The final alignment was obtained by realigning the adjusted alignment after resetting the gaps (in this procedure the guide tree is calculated before the gaps are removed).

Phylogenetic analysis was performed on the conserved core of this alignment (corresponding to residues of human β-actin) by the neighbor-joining algorithm of ClustalX (Thompson et al., 1997) using default parameters (gapped regions were included). Bootstrap analysis (1000 trials) provided a measure of confidence for the detected relationships as described above. The resulting tree was graphed by the program 'Unrooted' provided with ClustalX, and was prepared for presentation by Adobe Illustrator 7.0 (it should be noted that Illustrator 7.0 handles the pict file output from the tree graphing program much better than does Illustrator 9.0). Phylogenetic analysis was also performed using the neighbor-joining algorithm as implemented by the PHYLIP package [J. Felsenstein, PHYLIP (Phylogeny Inference Package) version 3.5c, Department of Genetics, University of Washington, Seattle, 1993]. Distance measurements were based on the PAM250 matrix instead of an identity matrix, and sequence addition was randomized to control for additional order bias. No significant changes in the topology of the tree or bootstrap values were observed.

Sequence references: GenBank gi numbers for the protein sequences used are as follows and are listed by subfamily. It should be noted that unannotated ARP sequences are designated by their chromosomal locus and/or gi number both in this list and on the tree. **Conventional actin:** Gl Actin, gi1703155; DmArp53d, gi7302881; Tg Actin, gi1703160; Pf Actin, gi5911379; Eh Actin,

gi113294; Sp Actin, gi113303; Sc Actin, gi170986; AtActin, gi6598382 (one of several At actin genes); Dd actin, gi4093161 (one of several Dd actin genes). **Arp1:** SpArp1, gi7490069; ScArp1, gi6321921; NcArp1, gi728797; AnArp1, gi4731565; CeY53F4B.22, gi17537473; DmArp87C, gi1168334; MmArp1b, gi18606465; HsArp1b, gi11342680; MmArp1a, gi8392847; HsArp1a, gi625520. **Arp2:** At Arp2, gi3818624; Sp Arp2, gi6650375; Sc Arp2, gi6320175; Ac Arp2, gi1703144; Dd Arp2, gi4093161; Ce K07C5.1, gi7505422; DmArp14D, gi1168330; GgArp2, gi806554; HsArp2, gi5031571. **Arp3:** AtArp3, gi4850401; ScArp3, gi6322525; NcArp3, gi11276973; SpArp3, gi416581; AcArp3, gi703143; DdArp3, gi1168328; CeY71F9AL.16, gi7105615; Dmactin66b, gi168329; HsArp3b, gi9966913; Mm 12835802, gi12835802; Hs Arp3a, gi5031573; **Arp4:** Sc Arp4, gi6322380; Sp P23A10.08, gi11276974; Sp C23D3.09, gi1351610; At 18394608, gi18394608; Ce ZK616.4, gi7332261; Dm CG6546, gi7302793; Hs BAF53b (also called 'Arp6'), gi 7705294; Mm BAF53a, gi4001805; Hs BAF53a, gi4757718. **Arp5:** At 12321978, gi12321978; Os 13486900, gi13486900; HsArp5, gi13396318; Dm CG7940, gi7300345; ScArp5, gi6324269; SpBC365.10, gi7490072. **Arp6:** Scarp6, gi6323114; At 6091748, gi6091748; CeARP6, gi14916971; SpCC550.12, gi7490073; Dmactin13E, gi1168327; GgArpX, gi12082091; Mm 12842577, gi12842577; Hs ArpX, gi12082089. **Arp8:** At 8843903, gi8843903; Sc Arp8, gi6324715; Sp C664.02, gi692009; Dm CG7846, gi7293397*;* Mm 12857259, gi12857259; Hs 10434709, gi10434709. **Arp10:** Dm CG12235, gi7293622; Hs Arp11, gi8923712; Mm Arp11, gi6176554; Ce C49H3.8, gi7497696; Sp C56F2, gi3116133; Sc Arp10, gi6320311; NcRo7, gi8347739. **Orphans** (listed by group): Hs 13383265, gi13383265; Mm 12840619, gi12840619; Mm 12840134, gi12840134; Mm 13386316, gi13386316; /Hs 11137605, gi11137605; /At 11276982, gi11276982; /Mm 12838437, gi12838437; Hs 10178893, gi10178893; Mm Actlike7a, gi6752956; Hs Actlike7a, gi5729720; Mm Actlike7b, gi6580806; Hs Actlike7b, gi5729722; /ScArp7, gi6325291; /ScArp9, gi6323676; Sp C1071.06, 7490070; /CeF42C5.9, 17540400.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol*. **215**, 403-410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.

Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K., Lam, Y. W., Steen, H., Mann, M. and Lamond, A. I. (2002). Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12**, 1-11.

Cairns, B. R., Erdjument-Bromage, H., Tempst, P., Winston, F. and Kornberg, R. D. (1998). Two actin-related proteins are shared functional components of the chromatin-remodeling complexes RSC and SWI/SNF. *Mol. Cell* **2**, 639-651.

Eckley, D. M., Gill, S. R., Melkonian, K. A., Bingham, J. B., Goodson, H. V., Heuser, J. E. and Schroer, T. A. (1999). Analysis of dynactin subcomplexes reveals a novel actin-related protein associated with the arp1 minifilament pointed end. *J. Cell Biol.* **147**, 307-320.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-791.

Galarneau, L., Nourani, A., Boudreault, A. A., Zhang, Y., Heliot, L., Allard, S., Savard, J., Lane, W. S., Stillman, D. J. and Cote, J. (2000). Multiple links between the NuA4 histone acetyltransferase complex and epigenetic control of transcription. *Mol. Cell* **5**, 927-937.

Harata, M., Oma, Y., Mizuno, S., Jiang, Y. W., Stillman, D. J. and Wintersberger, U. (1999). The nuclear actin-related protein of Saccharomyces cerevisiae, Act3p/Arp4, interacts with core histones. *Mol. Biol. Cell* **10**, 2595-2605.

Harata, M., Oma, Y., Tabuchi, T., Zhang, Y., Stillman, D. J. and Mizuno, S. (2000). Multiple actin-related proteins of Saccharomyces cerevisiae are present in the nucleus. *J. Biochem. (Tokyo)* **128**, 665-671.

Kato, M., Sasaki, M., Mizuno, S. and Harata, M. (2001). Novel actin-related proteins in vertebrates: similarities of structure and expression pattern to Arp6 localized on Drosophila heterochromatin. *Gene* **268**, 133-140.

Lawrence, C. J., Morris, N. R., Meagher, R. B. and Dawe, R. K. (2001). Dyneins have run their course in plant lineage. *Traffic* **2**, 362-363.

Lee, I. H., Kumar, S. and Plamann, M. (2001). Null mutants of the neurospora actin-related protein 1 pointed-end complex show distinct phenotypes. *Mol. Biol. Cell* **12**, 2195-2206.

Machesky, L. M. and May, R. C. (2001). Arps: actin-related proteins. *Results Probl. Cell Differ.* **32**, 213-229.

Park, J., Wood, M. A. and Cole, M. D. (2002). BAF53 forms distinct nuclear complexes and functions as a critical c-Myc-interacting nuclear cofactor for oncogenic transformation. *Mol. Cell Biol.* **22**, 1307-1316.

Peterson, C. L., Zhao, Y. and Chait, B. T. (1998). Subunits of the yeast SWI/SNF complex are members of the actin-related protein (ARP) family. *J. Biol. Chem.* **273**, 23641-23644.

Poch, O. and Winsor, B. (1997) Who's who among the Saccharomyces cerevisiae actin-related proteins? A classification and nomenclature proposal for a large family. *Yeast* **13**, 1053-1058.

Schafer, D. A. and Schroer, T. A. (1999). Actin-related proteins. *Annu. Rev. Cell Dev. Biol.* **15**, 341-363.

Schroer, T. A., Fyrberg, E., Cooper, J. A., Waterston, R. H., Helfman, D., Pollard, T. D. and Meyer, D. I. (1994). Actin-related protein nomenclature and classification. *J. Cell Biol.* **127**, 1777-1778.

Shen, X., Mizuguchi, G., Hamiche, A. and Wu, C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. *Nature* **406**, 541-544.

Sheterline, P., Clayton, J. and Sparrow, J. (1998). *Actin*. Oxford: Oxford University Press.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-4882.

van den Ent, F., Amos, L. A. and Lowe, J. (2001). Prokaryotic origin of the actin cytoskeleton. *Nature* **413**, 39-44.

Zhao, K., Wang, W., Rando, O. J., Xue, Y., Swiderek, K., Kuo, A. and Crabtree, G. R. (1998). Rapid and phosphoinositol-dependent binding of the SWI/SNF-like BAF complex to chromatin after T lymphocyte receptor signaling. *Cell* **95**, 625-636.

Cell Science at a Glance on the Web
Electronic copies of the poster insert are available in the online version of this article at jcs.biologists.org. The JPEG images can be downloaded for printing or used as slides.