

Flying Through the *Drosophila* Cytoskeletal Genome[Ⓞ]

Lawrence S.B. Goldstein and Shermali Gunawardena

Howard Hughes Medical Institute, Department of Cellular and Molecular Medicine and Department of Pharmacology, School of Medicine, University of California San Diego, La Jolla, CA 92093-0683

The well-developed genetics of *Drosophila* make it an excellent system to understand mechanisms of cytoskeletal function and organization and how the cytoskeleton is coupled to development and differentiation. Many years of work have revealed that *Drosophila* has a set of cytoskeletal structural proteins and protein motors qualitatively similar to most other eukaryotes. The full toolbox of cytoskeletal elements, and the extent to which they resemble cytoskeletal proteins in other organisms has, however, remained unknown owing to the lack of a complete genome sequence. With the recent arrival of the complete sequence of the euchromatic genome of *Drosophila* (Adams et al., 2000; Rubin et al., 2000), we can begin to compile a complete list of cytoskeletal components. We can also answer once and for all the question of what cytoskeletal proteins discovered in mammals and other eukaryotes are present in the fly, and what variations on these known proteins exist. This article provides a brief review of the cytoskeletal genes of *Drosophila*, and highlights several interesting features of their nature and organization not previously known.

As described (Adams et al., 2000) several different methods were used to predict the set of proteins encoded in the *Drosophila* genome. These methods relied on a combination of coupling the sequence to known genetic loci, EST sequences, and gene prediction programs. The predicted proteins were searched using greater than 1,000 cytoskeletal query sequences (see Table SI at <http://www.jcb.org/cgi/content/full/150/2/F63/DC1>) from vertebrates, *C. elegans*, and *S. cerevisiae*. Queries were chosen by selecting representative sequences from each class of known cytoskeletal proteins described in Kreis and Vale (1999), as well as selections of proteins classified as cytoskeletal on the *C. elegans* and *S. cerevisiae* proteome Web sites. In addition, consensus sequence elements from the BLOCKS database were also used (Henikoff et al., 1999a,b). Searches were done primarily using BLASTP

against the set of predicted proteins. In a few cases, TBLASTN was used to search the entire nucleotide sequence when an exhaustive classification of protein families was needed, or where a rigorous case for absence of a particular protein, or protein class was being established. G proteins and kinases were not analyzed since other workers were analyzing these proteins separately. Therefore, attention was effectively restricted to cytoskeletal structural proteins, non-enzymatic regulatory proteins, or proteins that moved, bound, cross-linked, or severed cytoskeletal polymers. Search results were inspected and evaluated manually; many false positive hits were found and eliminated from the classification as cytoskeletal.

Overall, 262 genes (see Table SII at <http://www.jcb.org/cgi/content/full/150/2/F63/DC1>) were found to give moderate to completely convincing homology with some member of the query set. This result corresponds to ~2% of the estimated 13,000 proteins encoded in the *Drosophila* genome. This 2% estimate is probably an underestimate since there are no doubt many proteins not yet recognized to have cytoskeletal functions, and it is also possible that some weakly homologous proteins were incorrectly eliminated from consideration. This 2% estimate is about half that reported for *C. elegans* (Chervitz et al., 1998) although the precise source of the estimate in this case is unclear. A few systematic sources of potential error were also identified in these analyses. In particular, proteins having coiled-coil, WD-40, or ankyrin repeat regions often gave significant BLAST scores with a variety of cytoskeletal queries, but examination of the actual sequence alignments revealed that they were not true homologues or relatives. For example, searches with the myosin II tail or intermediate filament proteins (see below) yield apparently significant hits with clearly unrelated α -helical coiled domains of myosins, kinesins, lamins, restin, and other unrelated α -helical coiled-coil containing proteins.

Among the ~261 genes encoding cytoskeletal structural or motor proteins 95 genes encode polypeptides belonging to the kinesin, dynein, or myosin motor superfamilies, or accessory/regulatory polypeptides known to interact with the motor polypeptide subunits. Approximately 80 genes encode actin-binding proteins of various types, including proteins belonging to the spectrin/ α -actinin/dystrophin

[Ⓞ]The online version of this article contains supplemental material.

Address correspondence to Dr. Lawrence S.B. Goldstein, HHMI/CMM Room 334, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0683. Tel.: (858)-534-9702. Fax: (858)-534-9701. E-mail: lgoldstein@ucsd.edu

superfamily of membrane cytoskeletal and actin-cross-linking proteins. 26 genes encode proteins likely to bind microtubules including relatives of microtubule binding proteins found in other organisms. 14 genes encode members of the actin superfamily, 12 genes encode members of the tubulin superfamily, and 5 genes encode septins.

Dystrophin, Spectrin, and Related Cross-linking Proteins: Links to Human Disease

Most multicellular eukaryotes encode multiple proteins related to spectrin and other proteins of the membrane cytoskeleton. In humans, not only are these proteins important for cell shape, but muscular dystrophy and various hemolytic anemias in humans appear to be caused by cellular fragility owing to a weakened or disrupted membrane cytoskeleton in specific cell types. For example, hemolytic anemias are often caused by mutations in genes encoding spectrins, ankyrins, or interacting proteins while Duchenne muscular dystrophy is caused by mutations in a gene encoding a protein related to spectrin called dystrophin.

The possible existence of dystrophin and its associated proteins in *Drosophila* has been discussed in the literature for many years, and other than one recent report (Roberts and Bobrow, 1998), have been thought not to exist. The sequence of *Drosophila* has revealed the clear existence of a dystrophin homologue (Fig. 1 A; unfortunately spread over adjacent gene pieces in the database of predicted genes). In addition, a clear homologue of dystrobrevin (Fig. 1 B), and reasonable matches to the dystrophin binding proteins syntrophins were also found (see Table SII at <http://www.jcb.org/cgi/content/full/150/2/F63/DC1>). Thus, *Drosophila* may prove a good model for understanding the normal functions of dystrophin, the proteins it interacts with, and the role of this system in human disease.

A class of proteins that has spectrin repeats, actin-binding motifs, and that appear to cross-link various filamentous classes has recently been found. In mammals, this class includes BPAG (Yang et al., 1999) and MACF (Leung et al., 1999). In *Drosophila*, other than the spectrins

themselves, kakapo (CG18076) and MSP-300 (CG18252) are the clearest members of this group. Strikingly, even after analysis of the complete genomic sequence, there is still only a single α -actinin gene. This result is puzzling since mutants lacking α -actinin function have a relatively muscle-specific phenotype (Fyrberg et al., 1998), yet α -actinin is thought to be an important component of adhesion plaques (Hynes and Zhao, 2000). Clearly, this conclusion is either incorrect, or other proteins can substitute for this function of α -actinin. Finally, several band 4.1 relatives were found as were two clear homologues of ankyrin.

Tubulin and Actin Polymers and Their Binding Proteins

Work in *Drosophila* has played a major role in our understanding of the actin and tubulin multi-gene families, the regulation of microtubule or microfilament polymerization, and the interactions of microtubules and microfilaments with other cellular components. Many genes have been identified by mutation in *Drosophila*, and found to encode important components of these systems previously discovered by biochemical or other approaches in various organisms.

Before the completion of the genome sequence, *Drosophila* was known to have four conventional α -, four conventional β -, and two γ -tubulin genes. The complete sequence revealed the presence of two additional, well-conserved conventional tubulin genes. One encodes a protein most like α -tubulin, the other encodes a protein most like β -tubulin (Fig. 2). Surprisingly, no proteins similar to the recently discovered δ - or ϵ -tubulins (Dutcher and Trabuco, 1998; Chang and Stearns, 2000) were found (both BLASTP and TBLASTN were used to ensure that the genes are truly missing, although it is formally possible that one or both could still be lurking in the few small sequence gaps, or heterochromatic DNA). The lack of δ is particularly surprising given that δ -tubulin has been found both in humans and algae. In view of the proposal that δ -, and perhaps ϵ -tubulins, are important components of centrioles and centrosomes respectively, these data suggest

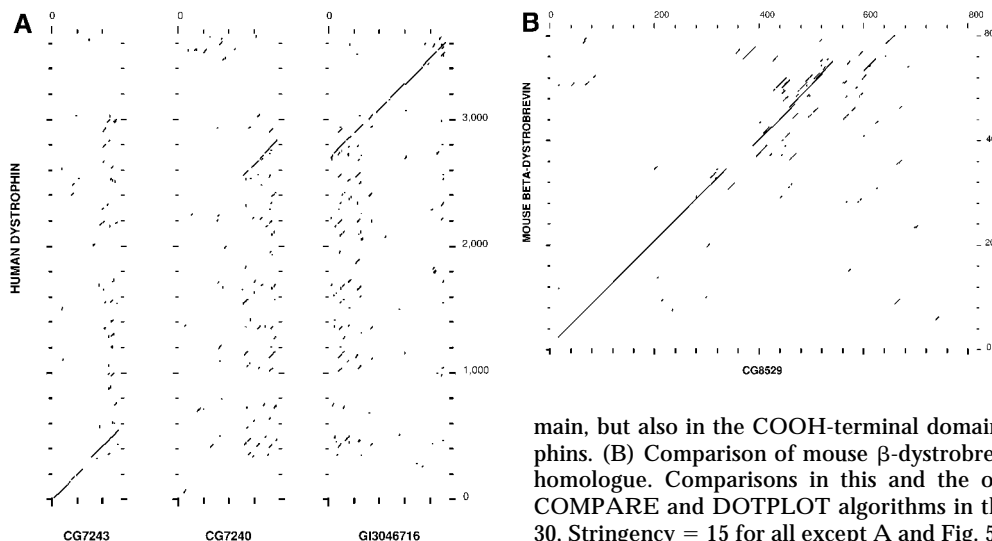


Figure 1. Dot matrix analysis of *Drosophila* homologues of dystrophin and dystrobrevin. (A) Comparison of human dystrophin (gi30846) with three sequences from *Drosophila* that make up portions of a likely dystrophin homologue. These three sequences are currently not properly assembled in the genome database. Note the significant homology both in the NH₂-terminal actin-binding domain, but also in the COOH-terminal domain unique to dystrophins and utrophins. (B) Comparison of mouse β -dystrobrevin (gi3702159) and a *Drosophila* homologue. Comparisons in this and the other figures was done using the COMPARE and DOTPLOT algorithms in the UWGCG package; Window = 30, Stringency = 15 for all except A and Fig. 5 where Stringency = 25.

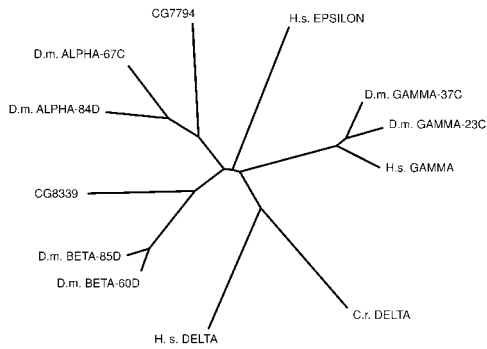


Figure 2. Unrooted tree showing relationships of new *Drosophila* tubulin genes. The tree was generated using DrawTree with a ClustalW alignment of representative tubulin family members.

that other tubulins, or other non-tubulin proteins can substitute for these functions.

A long-standing question about microtubule-binding proteins in the fly has concerned whether homologues of the repeat-containing MAPs that are typical of many mammalian MAPs were present in the fly (Cambiazo et al., 1995). Previously, a 205K MAP lacking such a motif was found

(Irminger-Finger et al., 1990). The sequence now reveals a protein that is a bit larger than mammalian tau protein, but has excellent sequence similarity to the repeat domains (Fig. 3). Whether it is a true homologue of tau, or whether it is a smaller version of MAP2 remains to be determined. It is intriguing however, that the fly may have a τ -like, but not a MAP2-like protein since tau is known to be enriched in mammalian axons while MAP2 is enriched in mammalian dendrites. This feature could be correlated with the general rarity of typical dendrites in fly neurons (Kandel et al., 1991). *C. elegans* also has a protein like tau and CG5606, but no MAP2-like molecule has been reported in that organism either.

Drosophila also has a variety of other well-known microtubule or microfilament binding proteins that were previously known or can now be inferred based on sequence homology (see Table SII at <http://www.jcb.org/cgi/content/full/150/2/F63/DC1>). The most striking new additions are homologues of the microtubule-severing protein called katanin, which is composed of two subunits, p60 and p80 (Fig. 4). *Drosophila* has genes encoding both subunits but has the added wrinkle of two genes encoding p60. Whether this allows novel regulatory properties awaits functional studies. Finally, *Drosophila* has well-conserved homologues of all of the recently identified components of the ARP2/3 complex

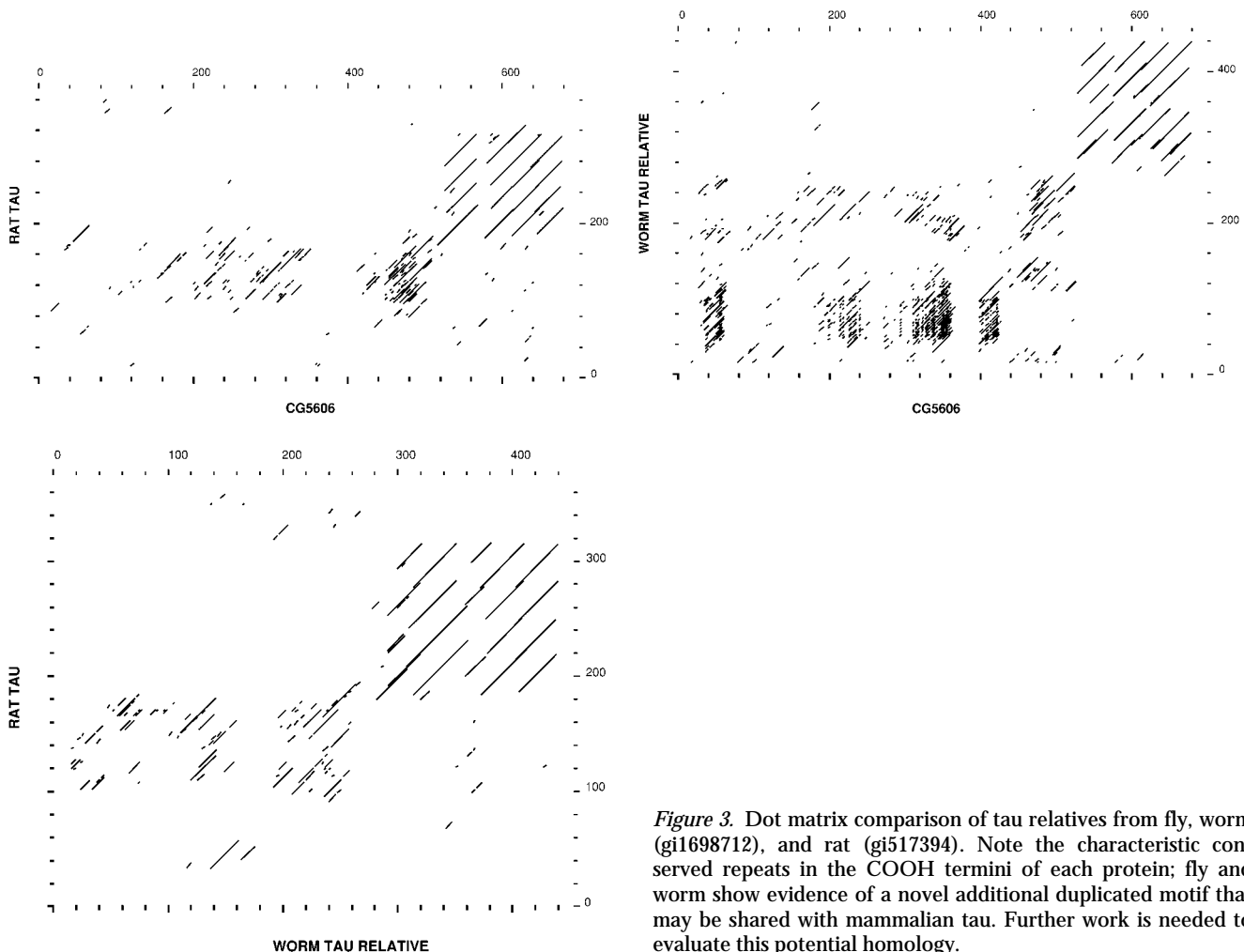


Figure 3. Dot matrix comparison of tau relatives from fly, worm (gi1698712), and rat (gi517394). Note the characteristic conserved repeats in the COOH termini of each protein; fly and worm show evidence of a novel additional duplicated motif that may be shared with mammalian tau. Further work is needed to evaluate this potential homology.

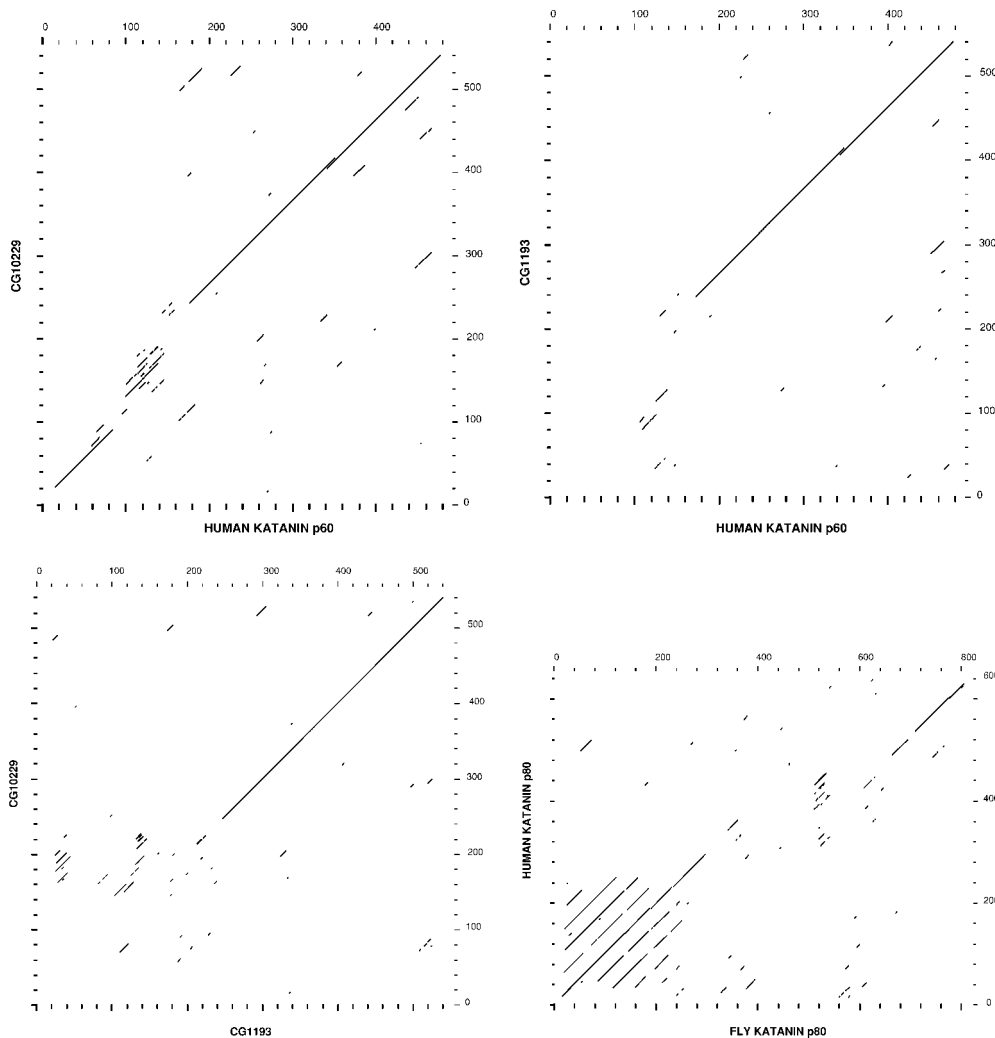


Figure 4. Dot matrix comparison of fly and human katanin subunits. Human katanin p60 (gi3283072) has two clear homologues in fly. There appears to be some variability in the NH₂-terminal domains (or CG1193 may have an assembly error). Human katanin p80 (gi3005599) also has a clear homologue in the fly, CG9910, that is currently truncated. A complete sequence in the GenBank database (gi6979998) was used instead.

as well as proteins homologous to WASP and Scar and most other known actin binding or bundling proteins.

Motors

Three major superfamilies of cytoskeletal motor proteins and associated proteins have been identified in all organisms: kinesins, dyneins, and myosins. Among the 95 genes encoding components of these systems are 24 genes with sequence similarity to the motor domain of kinesins. Many of these were previously described, although eight are new. Among these eight are two with unusual sequence features. One has a domain in its tail region with sequence similarity to G_γ, while another has similarity to RAS-GAP domains. Two had close relationships to so-called chromokinesins. Intriguingly, there are three members of the central motor domain class of microtubule-catastrophe kinesins exemplified by XKCM1 and KIF2 (Desai et al., 1999). Two of these in the fly appear to be present as a local duplication. Finally, no new COOH-terminal kinesins other than NCD were found, although there are a few predicted kinesins whose sequence is still partial. Several homologues of kinesin-II raft components were also found

as was the kinesin-II KAP subunit. Finally, a clear homologue of the mammalian protein PAT1 (Zheng et al., 1998), which has sequence similarity to kinesin light chain and binds to the amyloid precursor protein and microtubules, was found. Strikingly absent was any gene with significant sequence similarity to kinectin (Kumar et al., 1998), the proposed kinesin receptor for ER or ER-derived membranes. The kinectin sequence was searched exhaustively using both BLASTP and TBLASTN and nothing other than unrelated α-helical coiled-coil proteins were found among the top hits.

Of the 12 dynein heavy chains, 3 appeared to be new. No new cytoplasmic heavy chains were found; most appear to be axonemal although further work will be needed to categorize them rigorously. In addition to the gene already known to encode the p150 glued component of dynein, a new relative was also found (Fig. 5). Finally, all previously identified dynein subunits were also found.

Among the 13 myosin genes, two genes were found with significant sequence similarity to myosin VIIA, the myosin defective in the human hereditary deafness and blindness disease, Usher's type IB. One gene was previously known; the other is new (Fig. 6). There are also several genes with

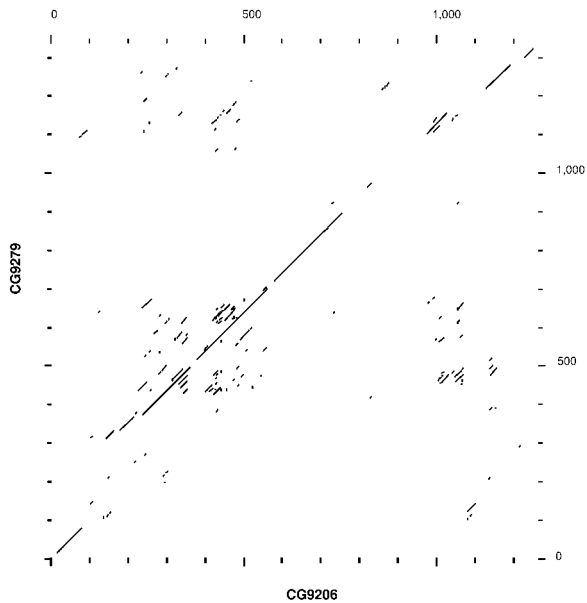


Figure 5. Dot matrix comparison of the known p150 Glued gene with a new potential homologue.

homology to myosin light chains including several not previously known.

Spindle Poles

A great deal is now known about the composition of centrosomes or spindle poles. Work from *Drosophila* has revealed a group of proteins called GRIPs that interact with γ -tubulin to form a nucleation complex (Oegema et al., 1999). Other proteins called CP60/CP190 were also previously found. Work in vertebrates has identified two large α -helical coiled-coil proteins called pericentrin and NUMA that are important structural components of centrosomes. Extensive searching revealed no true homologue of these two proteins recognizable by simple sequence similarity. However, a protein called centrosomin previously identified by mutation has a similar predicted structure to pericentrin and NUMA, and the phenotype of *cnn* mutants (Megraw et al., 1999) is consistent with the notion that *cnn* encodes the functional homologue of NUMA or pericentrin. Further work is needed to test this hypothesis.

Intermediate Filaments

Intermediate filaments exist in virtually every well-studied eukaryotic organism. *Drosophila*, however, has long been argued to lack such proteins based on morphological grounds (reviewed in Fyrberg and Goldstein, 1990). In addition, the lack of any genes encoding homologous proteins has bedeviled and perplexed many of us. Searches of the complete genomic sequence with many different intermediate filament probes, both full-length, and non-helical fragments using both BLASTP and TBLASTN, failed to yield any convincing matches other than nuclear lamins. Comparable searches of the *C. elegans* genomic sequence with the same query set easily identified *C. elegans* intermediate filament proteins with scores better than those

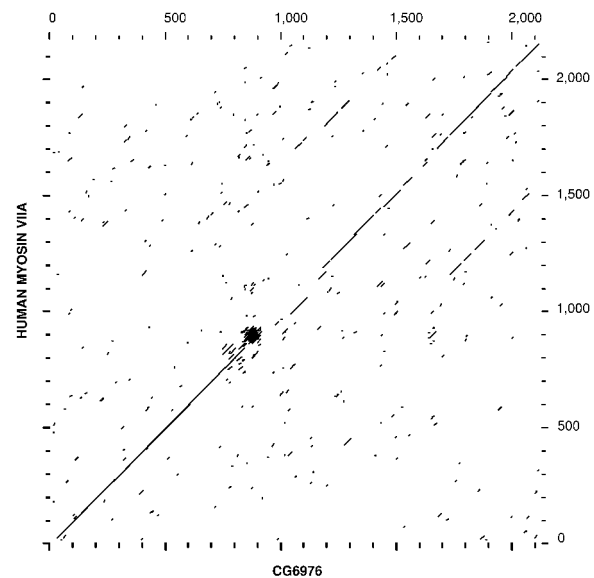
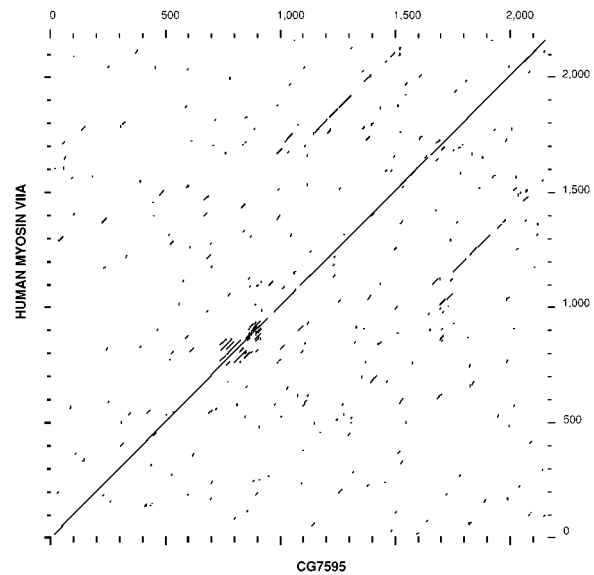


Figure 6. Dot matrix comparison of myosin VIIA homologues. Two fly homologues were compared with human myosin VIIA (gi1613788). CG7595 corresponds to the previously recognized *ck* gene.

garnered with nuclear lamins. In addition, *Drosophila* appears to lack an intermediate filament binding domain in the kakapo protein, which is a relative of mammalian proteins that can link actin filaments and intermediate filaments (Gregory and Brown, 1998). How *Drosophila* lives without intermediate filaments remains a puzzle since the organism clearly has epithelia and other tissues under mechanical stress. While part of the answer may lie in the hard chitinous exoskeleton of the fly, other cytoskeletal systems may have acquired the mechanical functions played by cytoplasmic intermediate filaments in other organisms (discussed in Gregory and Brown, 1998). Further work is needed to understand this feature of *Drosophila* cell biology.

In conclusion, for studies of cytoskeletal function, and the role of the cytoskeleton in development and differentiation, the sequence of the major components provides an invaluable starting point. These data, coupled to the ever expanding array of P-element mutations and other genetic manipulations will enable a complete elucidation of cytoskeletal biology and the beginnings of cytoskeletal genomics. Given the inherent redundancy of some cytoskeletal subsystems, the ability to manipulate genetically the dosage or content of several cytoskeletal genes simultaneously will be informative. Similarly, new genetic screens with enhancement or suppression of cytoskeletal mutant phenotypes, or overexpression screens using EP collections and cytoskeletal phenotypes will catapult us to a new era of understanding of cytoskeletal biology. Finally, cytoskeletal chips that allow monitoring of cytoskeletal and other gene function will be an invaluable tool for relating the overt phenotype of cytoskeletal mutants to cytoskeletal expression patterns in various developmental mutants. As the ultimate phenotypic readout of determinative decisions, an understanding of global patterns of cytoskeletal gene expression will be essential to elucidating the program and principles of eukaryotic development.

The authors would like to thank Celera Genomics for their kind assistance in the analysis of the data discussed in this paper. In particular, we are grateful to Wenyan Zhong for her excellent and tireless help in the early phases of this work. Lawrence S.B. Goldstein is an Investigator of the Howard Hughes Medical Institute.

Submitted: 2 June 2000

Revised: 23 June 2000

Accepted: 23 June 2000

References

- Adams, M.D., S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatide, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science*. 287:2185–2195.
- Cambiazio, V., M. Gonzales, and R.B. Maccioni. 1995. DMAP-85: a tau-like protein from *Drosophila melanogaster* larvae. *J. Neurochem.* 64:1288–1297.
- Chang, P., and T. Stearns. 2000. Delta-Tubulin and epsilon-tubulin: two new human centrosomal tubulins reveal new aspects of centrosome structure and function. *Nat. Cell Biol.* 2:30–35.

- Chervitz, S.A., L. Aravind, G. Sherlock, C.A. Ball, E.V. Koonin, S.S. Dwight, M.A. Harris, K. Dolinski, S. Mohr, T. Smith, et al. 1998. Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science*. 282:2022–2028.
- Desai, A., S. Verma, T.J. Mitchison, and C.E. Walczak. 1999. Kin I kinesins are microtubule-destabilizing enzymes. *Cell*. 96:69–78.
- Dutcher, S.K., and E.C. Trabuco. 1998. The UNI3 gene is required for assembly of basal bodies of *Chlamydomonas* and encodes delta-tubulin, a new member of the tubulin superfamily. *Mol. Biol. Cell*. 9:1293–1308.
- Fyrberg, C., A. Ketchum, E. Ball, and E. Fyrberg. 1998. Characterization of lethal *Drosophila melanogaster* alpha-actinin mutants. *Biochem. Genet.* 36: 299–310.
- Fyrberg, E.A., and L.S.B. Goldstein. 1990. The *Drosophila* cytoskeleton. *Annu. Rev. Cell Biol.* 6:559–596.
- Gregory, S.L., and N.H. Brown. 1998. *kakapo*, a gene required for adhesion between and within cell layers in *Drosophila*, encodes a large cytoskeletal linker protein related to plectin and dystrophin. *J. Cell Biol.* 143:1271–1282.
- Henikoff, J.G., S. Henikoff, and S. Pietrokovski. 1999a. New features of the Blocks Database servers. *Nucleic Acids Res.* 27:226–228.
- Henikoff, S., J.G. Henikoff, and S. Pietrokovski. 1999b. Blocks+: a non-redundant database of protein alignment blocks derived from multiple compilations. *Bioinformatics.* 15:471–479.
- Hynes, R.O., and Q. Zhao. 2000. The evolution of cell adhesion. *J. Cell Biol.* 150:F89–F95.
- Irminger-Finger, I., R.A. Laymon, and L.S. Goldstein. 1990. Analysis of the primary sequence and microtubule-binding region of the *Drosophila* 205K MAP. *J. Cell Biol.* 111:2563–2572.
- Kandel, E.R., J.H. Schwartz, and T.M. Jessell, editors. 1991. Principles of Neural Science. 3rd edition. Appleton & Lange, East Norwalk, CT.
- Kreis, T., and R. Vale. 1999. Guidebook to the Cytoskeletal and Motor Proteins. Oxford University Press, Oxford, UK.
- Kumar, J., H.P. Erickson, and M.P. Sheetz. 1998. Ultrastructural and biochemical properties of the 120-kDa form of chick kinectin. *J. Biol. Chem.* 273: 31738–31743.
- Leung, C.L., D. Sun, M. Zheng, D.R. Knowles, and R.K.H. Liem. 1999. Microtubule actin cross-linking factor (MACF): a hybrid of dystonin and dystrophin that can interact with the actin and microtubule cytoskeletons. *J. Cell Biol.* 147:1275–1285.
- Megraw, T.L., K. Li, L.-R. Kao, and T.C. Kaufman. 1999. The centrosomin protein is required for centrosome assembly and function during cleavage in *Drosophila*. *Development.* 126:2829–2839.
- Oegema, K., C. Wiese, O.C. Martin, R.A. Milligan, A. Iwamatsu, T.J. Mitchison, and Y. Zheng. 1999. Characterization of two related *Drosophila* gamma-tubulin complexes that differ in their ability to nucleate microtubules. *J. Cell Biol.* 144:721–733.
- Roberts, R.G., and M. Bobrow. 1998. Dystrophins in vertebrates and invertebrates. *Hum. Mol. Genet.* 7:589–595.
- Rubin, G.M., M.D. Yandell, J.R. Wortman, G.L. Gabor Miklos, C.R. Nelson, I.K. Hariharan, M.E. Fortini, P.W. Li, R. Apweiler, et al. 2000. Comparative genomics of the eukaryotes. *Science*. 287:2204–2215.
- Yang, Y., C. Bauer, G. Strasser, R. Wollman, J.P. Julien, and E. Fuchs. 1999. Integrators of the cytoskeleton that stabilize microtubules. *Cell*. 98:229–238.
- Zheng, P., J. Eastman, S. Vande Pol, and S.W. Pimplikar. 1998. PAT1, a microtubule-interacting protein, recognizes the basolateral sorting signal of amyloid precursor protein. *Proc. Natl. Acad. Sci. USA.* 95:14745–14750.