Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 11386–11390, December 1993
Genetics

The *Drosophila melanogaster* flightless-I gene involved in gastrulation and muscle degeneration encodes gelsolin-like and leucine-rich repeat domains and is conserved in *Caenorhabditis elegans* and humans

(behavioral mutant/transgenic organism/cDNA cloning/nucleotide sequence/protein sequence)

H. D. CAMPBELL*, T. SCHIMANSKY†, C. CLAUDIANOS*, N. OZSARAC‡, A. B. KASPRZAK*, J. N. COTSELL*, I. G. YOUNG§, H. G. DE COUET¶, AND G. L. GABOR MIKLOS*||

*Research School of Biological Sciences and §John Curtin School of Medical Research, The Australian National University, Canberra, ACT 2601, Australia; †Biologie III, University of Freiburg, Freiburg, D-7800 Germany; ‡Biological Sciences, University of New South Wales, Sydney, NSW 2033, Australia; and ¶Zoology Department, University of Hawaii, Honolulu, HI 96822

Communicated by M. M. Green, September 7, 1993

ABSTRACT **Mutations at the flightless-I locus (*fliI*) of *Drosophila melanogaster* cause flightlessness or, when severe, incomplete cellularization during early embryogenesis, with subsequent abnormalities in mesoderm invagination and in gastrulation. After chromosome walking, deficiency mapping, and transgenic analysis, we have isolated and characterized flightless-I cDNAs, enabling prediction of the complete amino acid sequence of the 1256-residue protein. Data base searches revealed a homologous gene in *Caenorhabditis elegans*, and we have isolated and characterized corresponding cDNAs. By using the polymerase chain reaction with nested sets of degenerate oligonucleotide primers based on conserved regions of the *C. elegans* and *D. melanogaster* proteins, we have cloned a homologous human cDNA. The predicted *C. elegans* and human proteins are, respectively, 49% and 58% identical to the *D. melanogaster* protein. The predicted proteins have significant sequence similarity to the actin-binding protein gelsolin and related proteins and, in addition, have an N-terminal domain consisting of a repetitive amphipathic leucine-rich motif. This repeat is found in *D. melanogaster*, *Saccharomyces cerevisiae*, and mammalian proteins known to be involved in cell adhesion and in binding to other proteins. The structure of the maternally expressed flightless-I protein suggests that it may play a key role in embryonic cellularization by interacting with both the cytoskeleton and other cellular components. The presence of a highly conserved homologue in nematodes, flies, and humans is indicative of a fundamental role for this protein in many metazoans.**

A number of mutations causing disturbances of behavior have been mapped to a genetically well-characterized region of the *Drosophila melanogaster* X chromosome (1–3). To further define loci of interest, a chromosomal segment was microdissected and cloned (4), chromosomal walks were performed, breakpoint rearrangements and mutational lesions were mapped, and transgenic organisms were constructed (5, 6). One of these loci is flightless-I (*fliI*), some mutations of which lead to loss of flight ability due to myofibrillar abnormalities in the indirect flight muscles; more severe alleles cause lethality in early embryogenesis (2, 3). Importantly, only some parts of these embryos are cellularized (3) and ventral furrow formation and gastrulation are abnormal. Cellularization during early embryogenesis is critically dependent on the actin-based cytoskeleton. The formation of membrane furrows, which come down between nuclei, is brought about by a hexagonal array of actin and myosin filaments that form around the nuclei in the cortex of the syncytial blastoderm (7–10).

Here we report the molecular cloning and characterization of *D. melanogaster fli* cDNAs and of homologous cDNAs from *Caenorhabditis elegans* and from humans.** The predicted proteins contain a gelsolin-related area likely to be involved in actin binding, as well as an amphipathic leucine-rich repeat (LRR) region likely to be involved in protein–protein interactions. These findings have significant implications for our understanding of early embryogenesis and muscle development.

MATERIALS AND METHODS

Germ-Line Transformation and Genetic Crossing Programs. Chromosome walking following chromosome microdissection and cloning was as described (5, 6). A 10.2-kb genomic *Xho* I fragment was ligated into the *Xho* I site of pW8 (11), a P-element vector that carries a white$^+$ (w^+) marker, to yield the construct pFLI (see Fig. 1). This was injected into *w* embryos as described (5). Standard genetic crosses introduced the transgene into different genomes harboring chromosomal rearrangements or alleles of the *fli* locus. Two of these rearrangements, *Df 17-257* and *Df 16-129*, are deficiencies that remove the *fli* locus. Crosses of the type w^1 *Df 17-257*/y^+ *Ymal*$^+$; pFLI/+ male by w^1 *Df16-129/FM7* female yielded the appropriate experimental and control genotypes. Tests of viability and flight capacity were conducted (2). The *FM7* and y^+ *Ymal*$^+$ chromosomes are as described (12).

Molecular Cloning of *D. melanogaster* cDNAs. *D. melanogaster* cDNAs were isolated from λgt10 libraries (13) using the two *Eco*RI fragments that detected the 5.1-kb *fli* transcript (see Fig. 1) on Northern blots (data not shown). cDNA fragments were subcloned in pGEM1. Two *Eco*RI fragments of 2.3 and 1.5 kb from the phage with the largest insert, together with a third, overlapping 1.5-kb fragment from another phage, were sequenced on both strands (6) using Sequenase (United States Biochemical). Compressions were resolved with dITP. A 420-bp *Bam*HI/*Sal* I fragment spanning the internal *Eco*RI site was subcloned from both cDNA and genomic clones and sequenced to confirm the *Eco*RI junction.

Molecular Cloning of *C. elegans* cDNAs. *C. elegans* cDNAs were isolated from a mixed stage library in λZAPII (Strata-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: LRR, leucine-rich repeat.

||To whom reprint requests should be addressed.

**The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U01182–U01184).

gene, no. 937006) using synthetic oligonucleotides corresponding to GenBank entry CELB0303 nucleotides 1478–1557 and to nucleotides 2231–2175 (complement) joined to 1565–1543 (complement). The two 80-mers were annealed and extended with deoxynucleoside triphosphates and DNA polymerase I (Klenow fragment). The resulting 145-bp duplex was gel-purified, labeled by random priming, and used to screen the library. Inserts from two positive clones were subcloned in pBluescript SK− (Stratagene) using ExAssist helper phage. The 4.6-kb insert from the larger clone was isolated and the coding region was sequenced completely on both strands.

Molecular Cloning of *Homo sapiens* cDNAs. Degenerate oligonucleotide primers including the following were synthesized (14): outer primers, 5′-TGYGTIGTITAYTTYTG-GCA-3′ and 5′-TTRTCIARDATCATDATRTC-3′; inner primers, 5′-WSICAYTTYAARMGIAARTT-3′ and 5′-ARRTCRTCYTGRCARAARTC-3′ (where Y is T or C, R is G or A, D is G or A or T, W is A or T, S is G or C, and M is A or C). The PCR was carried out on a GeneAmp PCR system 9600 (Perkin–Elmer) using GeneAmp PCR reagents. Each oligonucleotide was present at a final concentration of 2 μM and 0.5 μg of genomic DNA was added per reaction mixture (100 μl). After 1.5 min at 95°C, PCR was carried out for 40 cycles (95°C, 50°C, and 72°C, each for 30 sec) followed by 4.5 min at 72°C. An aliquot (1 μl) of the first PCR was used as template in the second reaction. Aliquots of the second reaction mixture were end-repaired with DNA polymerase I (Klenow fragment) and phosphorylated with T4 polynucleotide kinase. Fragments were isolated by electrophoresis and cloned into M13mp10 for sequencing. PCR fragments were used to screen human hippocampal and fetal brain cDNA libraries (Stratagene, nos. 936205 and 936206). Inserts were subcloned in pBluescript SK− and the longest one of 4.1 kb was sequenced on both strands.

Computer Methods. Sequences were assembled and analyzed with Staden (15) and Genetics Computer Group (GCG) (16) software. Nucleotide sequence data bases were searched by using TFASTA (17) or TBLASTN at the National Center for Biotechnology Information (Bethesda, MD). Two-way and multiple amino acid sequence alignments were obtained using the GCG programs GAP and PILEUP, with gap weight 3.0 and gap length weight 0.1. End gaps were weighted like other gaps.

RESULTS AND DISCUSSION

Genetic Localization of the *fli* Locus and Transgenic Rescue. Extensive molecular genetic analysis of a 250-kb chromosomal walk narrowed the *fli* locus to a 16-kb region. A restriction map of the region containing the *fli* gene, together with the positions of the other transcription units in this area, is shown in Fig. 1. Germ-line transformation (11) using the 10.2-kb *Xho* I fragment (denoted pFLI) yielded a single transformant line, in which the transgene mapped to chromosome 4 (data not shown). This transforming fragment contains only one complete transcription unit (Fig. 1), and it completely restores flight ability to flightless mutants—e.g., *fli*^{3} (alleles described in refs. 2 and 3). Two lethal alleles at this locus, *D44* and *2/19B*, correspond, respectively, to a small deficiency and an insertion in one of the two genomic *Eco*RI fragments (Fig. 1) that detect this transcription unit (unpublished results). This transcription unit thus corresponds to the *fli* gene.

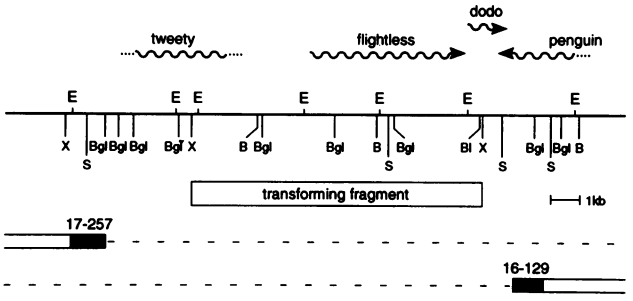

FIG. 1. Restriction map of the *fli* region. Restriction sites are shown for *Eco*RI (E), *Bam*HI (B), *Bgl* II (Bgl), *Sal* I (S), and *Xho* I (X). The breakpoints of chromosomal deficiencies are shown. Four transcription units (tweety, flightless, dodo, and penguin) and their direction of transcription (where known) are indicated. The 10.2-kb *Xho* I fragment used for germ-line transformation is indicated.

Overlapping chromosomal deficiencies (Fig. 1) have allowed us to delete the *fli* landscape and simultaneously introduce the transgene so that the appropriate genotypes for testing are all generated from the same crosses (e.g., genotypes *A*, *B*, and *C*; Table 1). Flies that are homozygous deficient for the landscape are white-eyed (genotype *A*; Table 1) and all die by the midpupal stage. Homozygous deficient individuals carrying a wild-type copy of pFLI are viable, with orange colored eyes (genotype *B*; Table 1). Their viability is about two-thirds that of their full sibs (genotype *C*; Table 1). All deficiencies in this region are lethal in the homozygous condition (e.g., genotype *D*).

When flies of the various genotypes were tested, one copy of the *fli*$^+$ locus was sufficient to give normal flight ability (genotype *E*; Table 1). On the other hand, a single copy of the transgene is insufficient to restore flight ability to homozygous deficient flies (genotype *B*; Table 1; 6% of these individuals exhibit some flight ability). By contrast, 99% of their full sibs (genotype *C*) fly normally. Part of the regulatory region of the *fli* gene may be missing from pFLI, or its activity may be reduced because this insertion is on the β-heterochromatin-like fourth chromosome (18). Alternatively, the lack of full restoration of flight ability by the transgene could be due to the deletion of part or all of three neighboring transcription units (Fig. 1).

Molecular Cloning of *fli* cDNAs. To characterize the *fli* transcription unit, 13 cDNAs were isolated, and the longest

Table 1. Survival to adulthood and flight ability of flies of various genotypes bearing deficiencies (*Df*) for the *fli* locus in the presence or absence of a single wild-type transgenic copy of the locus

	Genotype				
	A	*B*	*C*	*D*	*E*
	w^1 *Df 16–129* / w^1 *Df 17–257*	w^1 *Df 16–129* / w^1 *Df 17–257* plus transgene	w^1 *Df 17–257* / *w* Balancer plus transgene	w^1 *Df 16–129* / w^1 *Df 16–129*	w^1 *Df 16–129* / *w* Balancer
No. of surviving adult females	0	249	378	0	393
Flight ability	—	6%*	99%†	—	95%‡

*One hundred and six tested.
†Two hundred and thirteen tested.
‡One hundred and nineteen tested.

clone and another that overlapped its 3′ end were sequenced. The composite sequence consisted of 4672 bases plus a poly(A) tail and contained a single large open reading frame. The predicted protein (M_r 143,683) (Fig. 2) contains neither a signal sequence nor any potential transmembrane hydrophobic regions nor a phosphatidylinositol anchor site.

Molecular Cloning of a *fli* Homologue from *C. elegans*. Data base searches revealed significant sequence similarity between the *D. melanogaster fli*-encoded protein and predicted protein sequences from the *C. elegans* cosmid B0303 (19). B0303 appeared to encode amino acids 1–442 (Fig. 2) of a putative fli protein homologue, allowing for introns. Two *C. elegans* cDNAs were isolated by using a probe based on cosmid B0303 (19). The larger was sequenced (4529 bp), enabling prediction of the complete amino acid sequence of the protein (M_r 144,756). Alignment of the *C. elegans* protein (1257 amino acids) with the fli protein (1256 amino acids) revealed 49% identity (69% similarity when conservative substitutions are considered).

Comparison of the cDNA sequence with the complete genomic sequence (GenBank entries CELB0303 and CELB0523) shows that the *C. elegans* gene consists of 14 exons and 13 introns covering 8647 bp from the 5′ end of the cDNA to the poly(A) attachment site. For comparison, the *D. melanogaster fli* gene consists of four exons and three introns, covering 5195 bp (unpublished results). Only the first intron in the *D. melanogaster* gene corresponds exactly in position to one of the *C. elegans* introns (intron 2). A gene encoding $tRNA^{Trp}$ (*sup-5*, present in CELB0523) lies in intron 9 of the *C. elegans fli* gene homologue.

Molecular Cloning of a Human *FLI* Homologue. Alignment of the *D. melanogaster* and *C. elegans* protein sequences revealed stretches of completely conserved sequence (Fig. 2). Degenerate oligonucleotide primers were designed by reverse translation of these sequences, and PCR was used to detect *fli* homologues in DNA from various organisms (14). One set of four primers gave a discrete band of 850 bp with human genomic DNA. These primers correspond to residues 963–969, 1013–1019, 1178–1172, and 1188–1182 of the fly protein (Fig. 2). The sequence of this fragment suggested that it encoded a human homologue of the fli protein. This fragment was used to screen human brain cDNA libraries; six positive clones were isolated, and the longest of these (4.1 kb) was sequenced. The encoded protein had 58% identity to the *D. melanogaster* protein (74% similarity when conservative substitutions are considered). Southern blotting experiments using the cDNAs as probes indicate that a single copy of the gene is present in both the haploid human and fly genomes (unpublished results).

An alignment of the three proteins is shown in Fig. 2 and suggests that the human coding region is almost complete. Data base searches using the three sequences revealed sig-

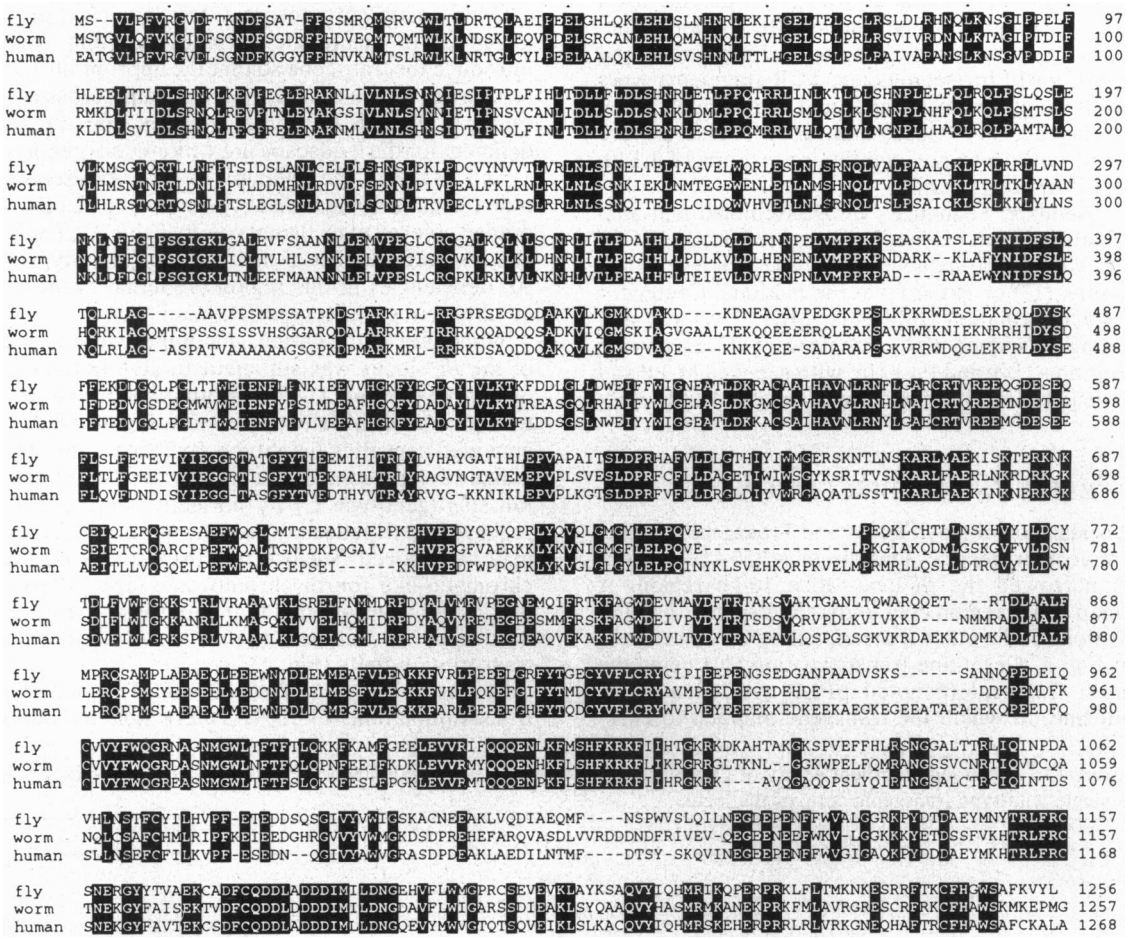

FIG. 2. Alignment of the amino acid sequences of the predicted *D. melanogaster* fli protein and its *C. elegans* and human homologues. Residues identical in all three are highlighted. Amino acid 154 is a phenylalanine residue encoded by TTC in both the *C. elegans* cDNAs, whereas in the genomic sequence (CELB0303) the serine codon TCC is present. We have isolated additional human cDNAs but have not yet obtained a clone encoding the initiation codon. The *D. melanogaster* LRRs are at amino acids 1–21, 22–44, 45–67, 68–92, 93–115, 116–139, 140–162, 163–186, 187–211, 212–234, 235–257, 258–280, 281–305, 306–328, 329–351, and 352–374 and the LRRs of the *C. elegans* and human proteins are at 1–24, 25–47, 48–70, 71–95, 96–118, 119–142, 143–165, 166–189, 190–214, 215–237, 238–260, 261–283, 284–308, 309–331, 332–354, and 355–377. The consensus sequence for the LRRs of the predicted proteins is P--(L/I)--L--L--L(D/N)L(S/N)-N-L--(L/V).

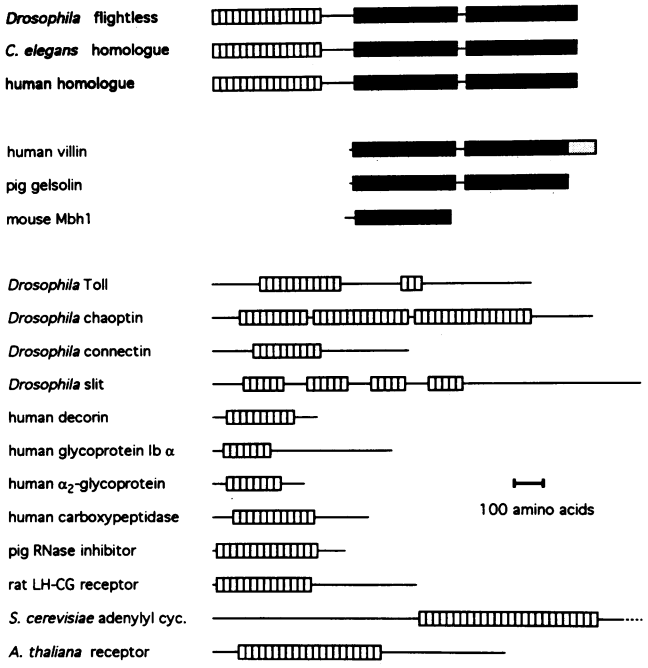

FIG. 3. Schematic view of the predicted *D. melanogaster*, *C. elegans*, and human proteins together with proteins containing related domains (refs. 20–37; GenBank entry ATRLPKC). Hatched box, LRRs; solid box, domain present in Mbh1, gCap39, fragmin, and severin, and as two copies in gelsolins, villins, and the fli protein homologues; shaded box, villin headpiece.

nificant similarity to two known protein families. The first contains actin-binding proteins such as gelsolins and villins, mouse Mbh1, and the slime mold proteins severin and fragmin, all of which cap or sever actin filaments (20–24). The second family consists of proteins containing various numbers of a LRR (25–37). These relationships are illustrated schematically in Fig. 3.

Alignments with the Gelsolin Family. Alignment of the human FLI protein with human gelsolin reveals 31% identity (52% similarity). Alignment of the *D. melanogaster* fli protein and *D. melanogaster* gelsolin (24) reveals 27% identity (48% similarity). The region of similarity between the fli proteins and members of the gelsolin family extends from near residue 500 to the C termini of the fli proteins (Fig. 3).

Alignments with the LRR Family. All three proteins possess 16 LRRs at their N termini (Fig. 3). In general, LRR domains are thought to have a role in specific protein–protein interactions (25, 26). For example, the LRR-containing human glycoprotein 1b α and β heterodimer is involved in the adhesion of human platelets to vascular subendothelium mediated by the von Willebrand factor (27), while the LRR of the rat lutropin–choriogonadotropin receptor putatively forms the extracellular hormone-binding domain (28). LRRs also appear to be involved in the interaction of yeast adenylyl cyclase (29) with the ras protein (30). Other protein–protein interactions involving LRRs are exhibited by the pig RNase inhibitor, which forms a 1:1 complex with RNase (31), and the human carboxypeptidase subunit which forms a tetrameric complex with the catalytic subunit (32). Four LRR family members from *D. melanogaster* are chaoptin [a photoreceptor cell-specific glycoprotein known to be involved in cell adhesion (33)], connectin [a cell adhesion molecule expressed on particular muscles and their innervating motoneurons during embryogenesis (25)], the Toll transmembrane protein [involved in the specification of dorsal–ventral polarity (34)], and the slit protein [involved in the development of glial cells and commissural axons (26, 35)]. These and other proteins (36, 37) containing LRRs are shown schematically in Fig. 3. Since repeats of this type are present in a variety of proteins involved in protein–protein interactions, it is likely that the LRRs of the three fli protein homologues are also involved in such interactions.

Actin-Binding Aspects of the Gelsolin and fli Families. The functional implications of the structure of the *Drosophila* fli protein and its *C. elegans* and human homologues are intriguing. All three exhibit extensive similarity to the gelsolin-like family over the six segments of gelsolin known to be involved in actin binding. In human gelsolin, actin monomer binding sites occur in segments 1 and 4–6, while F-actin binding sites occur in segments 2 and 3 (Fig. 4). Segment 1 of human gelsolin, which contains a high-affinity monomeric actin-binding site, has been examined by site-directed mutagenesis (39). To examine functional aspects of the fli proteins in the context of these mutagenesis results, we performed multiple alignments of gelsolins, including *D. melanogaster* gelsolin (24), and gelsolin-like proteins, from various organisms. Segment 1 of such an alignment is shown in Fig. 4,

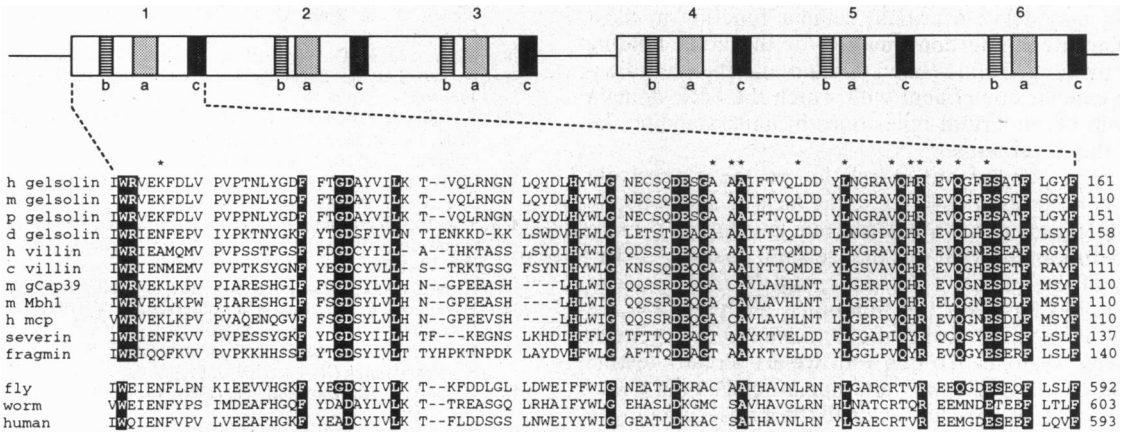

FIG. 4. Alignment of segment 1 of human gelsolin with members of the gelsolin family of actin-binding proteins and comparison with the corresponding regions of the three fli protein homologues. Family members include villin, gCap39, Mbh1, macrophage capping protein (mcp), fragmin, and severin (refs. 20–24 and 38). h, Human; m, mouse; p, pig; d, *D. melanogaster*; c, chicken. Residues identical in all the gelsolin family are highlighted, as are residues in the three fli proteins that match them. Residues where at least one substitution tested by *in vitro* mutagenesis alters actin binding are marked with an asterisk (39). Two large boxes indicate the two domains present in gelsolin and villin and the three shaded areas in each of the six segments (numbered 1–6) represent repetitive, conserved motifs. Hatched bars, motif b; shaded bars, motif a; solid bars, motif c.

together with the sites where mutagenesis alters actin binding. In members of the gelsolin family not including the fli protein homologues, 20 residues are absolutely conserved in segment 1, and many others are highly conserved over this wide phylogenetic spectrum. Of the invariant residues, 15 are conserved in the *D. melanogaster* protein, 12 are conserved in *C. elegans*, and 13 are conserved in humans. Of the 12 residues where some mutagenesis-induced changes have been found to alter actin binding (Fig. 4), 8 are conserved in *D. melanogaster*, and 6 are conserved in both *C. elegans* and humans. The differences that do occur naturally in these residues have yet to be tested by *in vitro* mutagenesis (39). It seems likely from the sequence similarities and the *in vitro* mutagenesis results that the actin-binding properties of the three fli proteins are similar to those of gelsolin.

Cellularization Processes in Early *Drosophila* Embryogenesis. How do these different properties of the actin-binding and LRR domains relate to the biological context in which the fli protein functions? A number of alleles of varying severity (*W-2, EN3, EF498, HC183*, and *DA534*) have been analyzed by germ-line clone analysis (3) where various amounts of wild-type fli product are presumably available for the developing egg. In the most extreme case (*W-2*), only some parts of the embryo cellularize correctly (figure 7 in ref. 3), gastrulation is abnormal, and the mesoderm does not invaginate.

The formation of membrane furrows which come down between nuclei to complete cellularization at stage 14 is controlled by the organization and dynamics of the actin-based cytoskeleton. This process is predominantly under the influence of maternal products and it appears that the actin–myosin network draws the membranes around the nuclei (7–10). Since the fli protein is clearly involved in cellularization, has potential actin-binding and LRR protein-binding domains, and is a maternal product, it could play a critical role in the cytoskeletal processes of early embryogenesis by acting as a bifunctional linking protein that tethers actin to other cellular components, possibly including membrane proteins. It is also possible that an effector molecule could interact with the LRR domain of the fli protein, modulating its actin-binding activity and thereby regulating the behavior of the cytoskeleton. In later development of the indirect flight muscles, the fli protein may again be needed for similar events of muscle differentiation. Alternatively, it may play a structural role in indirect flight muscle. The conservation of the fli protein in distantly separated organisms indicates that this protein may serve a critical cellular function in most higher eukaryotes. The confirmation of the actin binding properties of the gelsolin-related domain and the identification of the cellular component with which the LRR domain interacts will be important milestones in understanding the biology of the fli proteins.

As has been pointed out in relation to the extensively studied gelsolin protein (24), ". . . the lack of a genetically manipulable organism has hindered our developing a more complete definition of what it is that gelsolin does. . . ." The availability of the human, *C. elegans*, and *D. melanogaster fli* genes should facilitate the functional analysis of these gelsolin family members using the *D. melanogaster* genetic and transgenic systems. To this end, we are already testing the function of the *C. elegans* and human cDNAs in transgenic flies. In addition, we have cloned the mouse homologue of *fli* (unpublished results) and are using gene targeting to examine the role of this gene in mammalian systems. Finally, chromosome mapping of the human *FLI* gene homologue will allow investigation of its possible role in inherited disorders involving developmental abnormalities or muscle degeneration.

We thank L. Hillier and A. Coulson for sequence data and helpful discussions; P. Milburn for synthesis of oligonucleotides; and J. Whitehead, J. Wilson, and M. Whittaker for illustrations and photography. H.G.d.C. was supported by National Science Foundation Grant DCB-9106129.

1. Miklos, G. L. G., Kelly, L. E., Coombe, P. E., Leeds, C. & Lefevre, G. (1987) *J. Neurogenet.* **4**, 1–19.
2. Miklos, G. L. G. & de Couet, H. G. (1990) *J. Neurogenet.* **6**, 133–151.
3. Perrimon, N., Smouse, D. & Miklos, G. L. G. (1989) *Genetics* **121**, 313–331.
4. Miklos, G. L. G., Yamamoto, M., Davies, J. & Pirrotta, V. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2051–2055.
5. Delaney, S. D., Hayward, D. C., Barleben, F., Fischbach, K.-F. & Miklos, G. L. G. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7214–7218.
6. Hayward, D. C., Delaney, S. J., Campbell, H. D., Ghysen, A., Benzer, S., Kasprzak, A. B., Cotsell, J. N., Young, I. G. & Miklos, G. L. G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2979–2983.
7. Foe, V. E. & Alberts, B. M. (1983) *J. Cell Sci.* **61**, 31–70.
8. Karr, T. M. & Alberts, B. M. (1986) *J. Cell Biol.* **102**, 1494–1509.
9. Miller, K. G., Field, C. M. & Alberts, B. M. (1989) *J. Cell Biol.* **109**, 2963–2975.
10. Simpson, L. & Wieschaus, E. (1990) *Development* **110**, 851–863.
11. Klemenz, R., Weber, U. & Gehring, W. J. (1987) *Nucleic Acids Res.* **15**, 3947–3959.
12. Lindsley, D. L. & Zimm, G. G. (1992) *The Genome of Drosophila melanogaster* (Academic, San Diego).
13. Poole, S. J., Kauvar, L. M., Drees, B. & Kornberg, T. (1985) *Cell* **40**, 37–43.
14. McPherson, M. J., Jones, K. M. & Gurr, S. J. (1991) in *PCR: A Practical Approach*, eds. McPherson, M. J., Quirke, P. & Taylor, G. R. (IRL, Oxford), pp. 171–186.
15. Staden, R. (1987) in *Nucleic Acid and Protein Sequence Analysis: A Practical Approach*, eds. Bishop, M. J. & Rawlings, C. J. (IRL, Oxford), pp. 173–217.
16. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
17. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
18. Miklos, G. L. G. & Cotsell, J. N. (1990) *BioEssays* **12**, 1–6.
19. Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R. & Waterston, R. (1992) *Nature (London)* **356**, 37–41.
20. Hartwig, J. H. & Kwiatkowski, D. J. (1991) *Curr. Opin. Cell Biol.* **3**, 87–97.
21. Friederich, E., Pringault, E., Arpin, M. & Louvard, D. (1990) *BioEssays* **12**, 403–408.
22. Vandekerckhove, J. (1990) *Curr. Opin. Cell Biol.* **2**, 41–50.
23. Prendergast, G. C. & Ziff, E. B. (1992) *EMBO J.* **10**, 757–766.
24. Heintzelman, M. B., Frankel, S. A., Artavanis-Tsakonas, S. & Mooseker, M. S. (1993) *J. Mol. Biol.* **230**, 709–716.
25. Nose, A., Mahajan, V. B. & Goodman, C. S. (1992) *Cell* **70**, 553–567.
26. Rothberg, J. M. & Artavanis-Tsakonas, S. (1992) *J. Mol. Biol.* **227**, 367–370.
27. Lopez, J. A., Chung, D. W., Fujikawa, K., Hagen, F. S., Papayannopoulou, T. & Roth, G. J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5615–5619.
28. McFarland, K. C., Sprengel, R., Phillips, H. S., Köhler, M., Rosemblit, N., Nikolics, K., Segaloff, D. L. & Seeburg, P. H. (1989) *Science* **245**, 494–499.
29. Kataoka, T., Broek, D. & Wigler, M. (1985) *Cell* **43**, 493–505.
30. Field, J., Xu, H.-P., Michaeli, T., Ballester, R., Sass, P., Wigler, M. & Colicelli, J. (1990) *Science* **247**, 464–467.
31. Hofsteenge, J., Kieffer, B., Matthies, R., Hemmings, B. A. & Stone, S. R. (1988) *Biochemistry* **27**, 8537–8544.
32. Tan, F., Weerasinghe, D. K., Skidgel, R. A., Tamei, H., Kaul, R. K., Roninson, I. B., Schilling, J. W. & Erdös, E. G. (1990) *J. Biol. Chem.* **265**, 13–19.
33. Krantz, D. E. & Zipursky, S. L. (1990) *EMBO J.* **9**, 1969–1977.
34. Hashimoto, C., Gerttula, S. & Anderson, K. V. (1991) *Development* **111**, 1021–1028.
35. Rothberg, J. M., Jacobs, J. R., Goodman, C. S. & Artavanis-Tsakonas, S. (1990) *Genes Dev.* **4**, 2169–2187.
36. Krusius, T. & Ruoslahti, E. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7683–7687.
37. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1906–1910.
38. Dabiri, G. A., Young, C. L., Rosenbloom, J. & Southwick, F. S. (1992) *J. Biol. Chem.* **267**, 16545–16552.
39. Way, M., Pope, B. & Weeds, A. G. (1992) *J. Cell Biol.* **116**, 1135–1143.