Biochemical and Biophysical Research Communications **266,** 135–140 (1999)
Article ID bbrc.1999.1776, available online at http://www.idealibrary.com on IDEAL®

Functional Genomic Analysis Reveals the Utility of the I/LWEQ Module as a Predictor of Protein:Actin Interaction

Richard O. McCann and Susan W. Craig[1]
Department of Biological Chemistry, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205-2185

Received November 3, 1999

The I/LWEQ module is a conserved sequence that we have identified as an actin-binding motif in the metazoan focal adhesion protein talin and the yeast protein Sla2p. Both of these proteins are associated with the actin cytoskeleton in cells. To better establish the value of the I/LWEQ module for prediction of actin-binding function, we have applied a functional genomics approach. Analysis of the 23 available I/LWEQ module sequences supports the division of I/LWEQ protein superfamily into four groups: (1) metazoan talin, (2) *Dictyostelium discoideum* talin homologs TalA/B, (3) metazoan Hip1p, and (4) yeast Sla2p. We show here that I/LWEQ modules from each major group bind to F-actin *in vitro* and that GFP-fusion proteins of the I/LWEQ modules of talin and Sla2p bind to F-actin *in vivo*. Therefore, the presence of an I/LWEQ module is strongly predictive of protein–actin interactions. The structural and functional conservation of the I/LWEQ module across the phylogenetic distance between cellular slime molds and mammals implies that the role of the I/LWEQ module is to connect diverse proteins involved in distinct cellular processes, including cell adhesion, cytoskeletal organization, and cell differentiation, to the actin cytoskeleton.

The focal adhesion protein talin (1–3) has actin-binding regions in its N-terminus, central rod domain, and C-terminus (3). Talin and the yeast cytoskeletal protein Sla2p (4) share a conserved sequence at their C-termini. We showed that this sequence, the I/LWEQ module, specifies F-actin binding for both vertebrate talin and yeast Sla2p (5). The five cDNA sequences then available indicated that I/LWEQ module proteins could be placed into two groups, one including metazoan talin and a talin homolog from *D. discoideum* and the other containing yeast Sla2p and a putative *Caenorhabditis elegans* Sla2p. Subsequent database searches have uncovered 18 additional I/LWEQ module sequences. These include several more talins and an additional *D. discoideum* talin homolog, a putative *D. discoideum* Sla2p, five new fungal Sla2p representatives, and several homologs of metazoan Hip1p, which is similar to Sla2p. Hip1p was identified recently as a binding partner of huntingtin, the protein implicated in the etiology of Huntington disease (6, 7). Based on sequence similarity the 23 I/LWEQ module proteins can now be placed into a superfamily of four distinct groups, which are described along with their inferred phylogenetic distribution in this report. We also show that the I/LWEQ modules of representatives of all four groups bind to F-actin *in vitro* and that the I/LWEQ module also binds to F-actin *in vivo*. Thus, the I/LWEQ module may be thought of as a conserved adaptor module designed to link functionally distinct proteins to the cytoskeleton.

MATERIALS AND METHODS

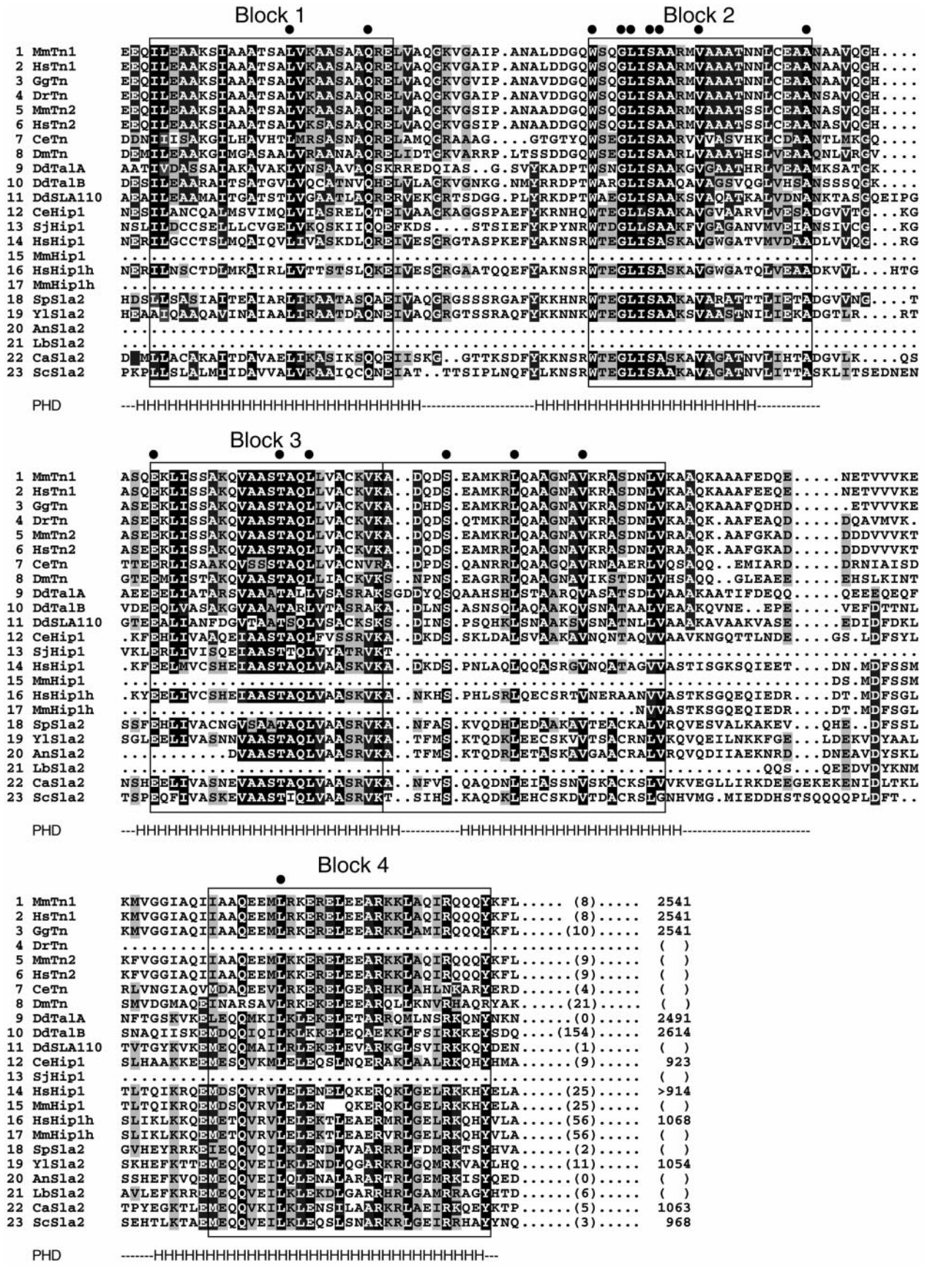
I/LWEQ module His6-fusion proteins from *Mm*Tn1 and *Hs*Tn2 (expression vector: pET-28c, Novagen) or GST-fusion proteins from *Sc*Sla2p and *Dd*TalA (pGEX-2T, Amersham Pharmacia) were prepared and assayed for F-actin binding as in (5). We were unable to prepare Hip1-fusion proteins using these expression systems, so the His6-*Hs*Hip1p and Hip1h constructs (pET-28c) were prepared by *in vitro* transcription/translation (Promega) and assayed similarly. Chinese hamster ovary cells were transiently transfected with either GFP alone, GFP-*Mm*Tn1.2345-2541 or *Sc*Sla2.771-968 (vector: pEGFP-C3, Clontech) and fixed in 4% paraformaldehyde in phosphate buffered saline. F-actin was visualized by staining with rhodamine phalloidin (Molecular Probes), and GFP-fluorescence was observed using epifluorescence.

RESULTS AND DISCUSSION

Novel I/LWEQ modules were identified by searching the nonredundant and expressed sequence tag (EST) databases (GenBank, EMBL, DDJB, PDB; EST; National Center for Biotechnology Information) with BLAST 2.0 (8) using the I/LWEQ modules of *Mm*Tn1 and *Sc*Sla2p as query sequences. Criteria for inclusion

[1] To whom correspondence should be addressed: Fax: 410-955-5759. E-mail: scraig@jhmi.edu.

0006-291X/99 $30.00


Block 1
Block 2
Block 3
Block 4
PHD

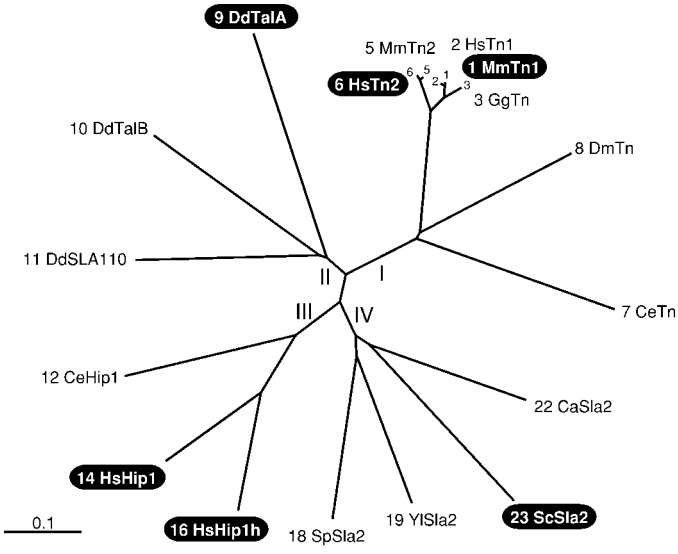


FIG. 2. I/LWEQ module phylogeny. An alignment prepared as in Fig. 1, but lacking the six incomplete sequences, has been displayed using TreeView [(12) http://taxonomy.zoology.gla.ac.uk/rod/rod.html]. Group I: Metazoan talin; Group II: *D. discoideum* TalA, TalB, and SLA110; Group III: Metazoan Hip1p; Group IV: Yeast Sla2p. I/LWEQ modules from each major group (ovals) bind to F-actin *in vitro* as either GST- or histidine-tagged fusion proteins. A congruent phylogeny is obtained when the 11 complete protein sequences are analyzed similarly.

as an I/LWEQ module included overall sequence similarity (BLAST E values < 0.001), maintenance and spacing of Blocks 1–4, and conservation of particular residues (5). The results are shown in Fig. 1, where 23 complete or partial I/LWEQ modules have been aligned using CLUSTAL W (9). The consensus length of the I/LWEQ module is 186 amino acids (*Mm*Tn1.2345–2541) with 16 invariant residues. Blocks 1–4 were originally identified using Block Maker (5, 10) and have remained essentially the same as additional I/LWEQ modules have been identified, except that in the absence of *Dd*TalA, which contains two small insertions, Block 3 extends to the second conserved valine in the region between Blocks 3 and 4 (Fig. 1). Secondary structure predictions using PHD (11) indicate that Blocks 1–4 are coextensive with putative α-helices while the inter-helical regions are loops.

To illustrate the evolutionary relationships of the I/LWEQ module, an alignment prepared as that in Fig. 1, but lacking the six partial I/LWEQ module sequences, was converted into an unrooted tree using TreeView (12). This result (Fig. 2) shows that I/LWEQ module proteins can be placed into four primary groups consisting of (I) metazoan talin; (II) *D. discoideum* TalA, TalB, and SLA110; (III) metazoan Hip1p; and (IV) fungal Sla2p. This phylogeny agrees well with the traditional classification of these taxa. For example, within the metazoan talin branch the vertebrate talins are most closely related to one another than to the other talins. The I/LWEQ modules of the *D. discoideum* talin homologs TalA and TalB are more similar to one another and to SLA110 (which may represent a *D. discoideum* Hip1p or Sla2p) than to the metazoan talins. Similarly the fungal Sla2p sequences are grouped together, as are the metazoan Hip1p sequences.

Sequence conservation among the I/LWEQ modules implies conservation of function. The *Mm*Tn1 and *Sc*Sla2p I/LWEQ modules co-sediment with F-actin *in vitro* (5). Using the same assay we have determined that I/LWEQ modules representing the other major groups of I/LWEQ module proteins, including *Hs*Hip1p, *Hs*Hip1h (Hip1p homolog; KIAA0665, cDNA provided by the Kazusa DNA Research Institute), *Dd*TalA (cDNA provided by G. Gerisch), and *Hs*Tn2 (KIAA0320, Kazusa DNA Research Institute), also bind to F-actin (Fig. 3a). Mutation of the conserved arginine at position 25 in Block 4 of *Sc*Sla2p(R958G) abolishes F-actin binding and alters the electrophoretic mobility of the mutant

FIG. 1. I/LWEQ module sequence alignment. 23 I/LWEQ modules (17 complete, 6 partial, including 3 residues on either side of the I/LWEQ module), which begin with the first residue of Block 1 (5, 10) and end at the last residue of Block 4, were aligned using CLUSTAL W (9) followed by minor manual adjustments. The most conserved residues are in black, conserved similar residues are in gray. Filled circles denote invariant residues. The secondary structure prediction (11) is that of *Mm*Tn1 (similar results were obtained with the other sequences) and shows that Blocks 1–4 are coextensive with 4 putative α-helices (network output for helix greater than 9 out of 10). The extension of Block 3 in the absence of *Dd*TalA is indicated by the dashed line. Sequence identifications: *Mm*Tn1 (accession number: X56123); *Hs*Tn1 (AF007828); *Gg*Tn, (3); *Dr*Tn (AI384570; AI331130); *Mm*Tn2 (AA387755, AA474115, AA218133, AA032400, AA733912, AA039387); *Hs*Tn2 [AB002318; KIAA0320 (24)]; *Ce*Tn (L46861); *Dm*Tn (AI256966); *Dd*TalA (U14576); *Dd*TalB (AB023655); *Dd*SLA110 (C25499); *Ce*Hip1p (M98552); *Sj*Hip1p (AA182338); *Hs*Hip1p (U79734); *Mm*Hip1p (AA139803, AA388714); *Hs*Hip1h [Hip1p homolog; AB014555; KIAA0655 (25)]; MmHip1h (AI426522); *Sp*Sla2p (AA182338); *Yl*Sla2p (U65409); *An*Sla2p (AA786107); *Lb*Sla2p (U93507); *Ca*Sla2p (AJ009556); *Sc*Sla2p (Z22811). Species abbreviations: *Mm, Mus musculus; Hs, Homo sapiens; Dr, Danio rerio; Gg, Gallus gallus; Ce, C. elegans; Sj, Schistosoma japonicum; Dd, D. discoideum; Sp, Schizosaccharomyces pombe; An, Aspergillus* (*Emericella*) *nidulans; Ca, Candida albicans; Yl, Yarrowia lipolytica; Lb, Laccaria bicolor; Sc, Saccharomyces cerevisiae.* All of the fungal species are ascomycetes except for *L. bicolor*, which is a basidiomycete. The number of additional residues at the C-terminus of each protein is indicated, as is the size in amino acids for complete sequences. Blocks were identified using the Blocks server (www.blocks.fhcrc.org). The CLUSTAL W alignment was performed at the European Bioinformatics Institute server (http://www.ebi.ac.uk) and prepared for publication using MacBoxShade (ftp://ulrec3.unil.ch/pub/boxshade/MacBoxshade). The secondary structure prediction was calculated using the PHD server (dodo.cpmc.columbia.edu/predictprotein). The *Mm*Tn2 sequence was constructed using the indicated series of overlapping ESTs. The human and mouse Tn2 (*Hs*Tn2, *Mm*Tn2) orthologs represent novel talin sequences (R. McCann and S. Craig, in preparation).

fusion protein (5). This mutation also has the same effect upon the mouse talin I/LWEQ module (Fig. 3a), which indicates that similar structural determinants in Block 4 mediate actin binding across the I/LWEQ module superfamily.

To determine whether the I/LWEQ module interacts with F-actin *in vivo*, we transfected Chinese Hamster Ovary (CHO) cells with several different GFP-I/LWEQ module constructs. Figure 3b shows that the I/LWEQ module of *Mm*Tn1 interacts with F-actin stress fibers in CHO cells *in vivo*. At the high expression level shown actin stress fibers form large dysregulated bundles and aggregates. Similar results were obtained with the homologous *Sc*Sla2p construct (not shown). Neither the GFP-I/LWEQ construct containing the R958G mutation that abolishes F-actin binding *in vitro* nor the GFP-only control had any effect on actin organization when expressed at similar levels. Thus, based on its phylogenetic distribution and activity, the I/LWEQ module represents an ancient, conserved actin-binding element.

The distribution of I/LWEQ module proteins is congruent with recent views of phylogenetic relationships among animals, fungi, and cellular slime molds (13). So far, I/LWEQ modules have been found only in these taxa. No examples have been identified among the available plant or early diverging eukaryote (protist) sequences. The presence of I/LWEQ module proteins in *D. discoideum* is consistent with the conclusion, based on comparison of protein sequences, that the Mycetezoa (e.g., *D. discoideum, Physarum*) emerged among the multicellular eukaryotes, and are more closely related to animals and fungi than to plants (13–15).

Interestingly, there is no strict correlation between divergence time and sequence similarity among I/LWEQ modules. For example, among fungi the *Sp*Sla2p (class Archiascomycetes) and *Yl*Sla2p (Hemiascomycetes) I/LWEQ modules are 59% identical while sharing a last common ancestor 350 million years ago (mya). The more closely related *Ca*Sla2p and *Sc*Sla2p (both Hemiascomycetes) are even less similar at 53% identity despite sharing a last common ancestor only 150 mya (16). In marked contrast, *Mm*Tn1 (Mammalia) and *Gg*Tn (Aves) are 95% identical even though they shared a last common ancestor 300 mya. Similarly *Mm*Tn1 and *Dr*Tn (Actinopterygii) are 92% identical over the available I/LWEQ module sequence while diverging from a common ancestor 405 mya (17). Thus, the rate of change in the fungal Sla2p sequences is much greater than that of the vertebrate talin sequences, perhaps indicating greater constraints on the evolution of the vertebrate talin I/LWEQ modules than on the fungal Sla2p I/LWEQ modules. This high sequence similarity is confined to vertebrate talin. It is not observed in the other metazoan talins nor in metazoan Hip1p (Fig. 2).

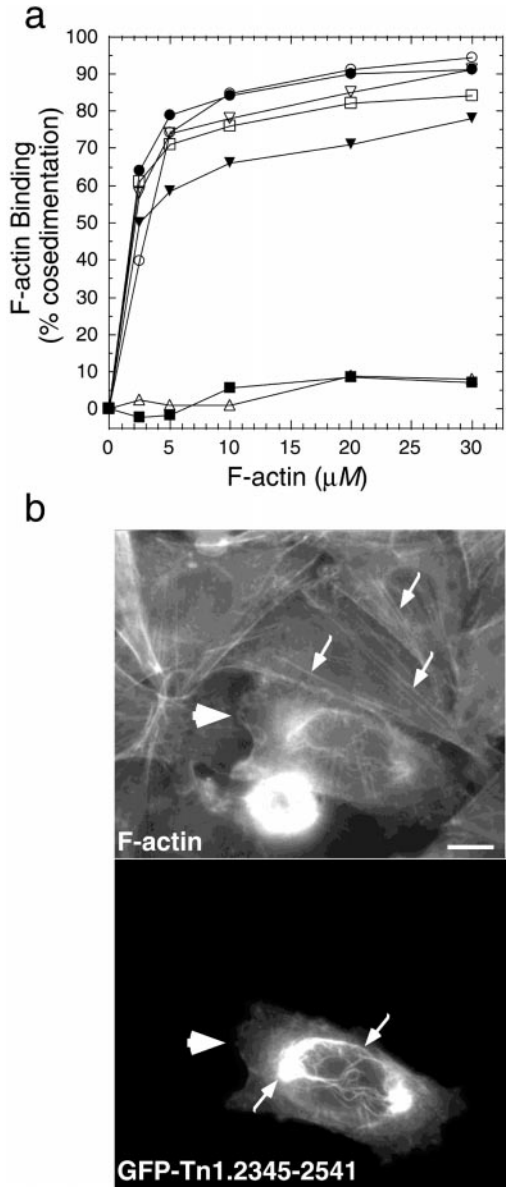


FIG. 3. (a) F-actin binding *in vitro*. I/LWEQ module fusion proteins were assayed for F-actin binding (5). Representatives of each group (Fig. 2) bind to F-actin. Mutation of the Block 4 conserved arginine of talin (R2526) and Sla2p (R958) essentially abolishes F-actin binding. H6-*Mm*Tn1.2345–2541, ○; H6-*Hs*Tn2.2345–2541, ●; GST-*Sc*Sla2.771–968, ▽; GST-*Dd*TalA.2277–2491, □; H6-*Hs*Hip1.693–914, ▼; H6-*Mm*Tn1.2345–2541(R2526G), ■; GST-*Sc*Sla2.771–968(R958G), △. (b) F-actin binding *in vivo*. Top: Transiently transfected CHO cells were fixed and stained with rhodamine phalloidin to visualize F-actin. Bottom: GFP fluorescence due to GFP-*Mm*Tn1.2345–2541. Large arrow: Transfected CHO cell; small arrows: F-actin. Untransfected cells display normal stress fibers (small arrows, top). Expression of GFP-*Mm*Tn1-.2345–2541 dysregulates stress fiber formation and promotes the formation of F-actin aggregates (small arrows, bottom). Neither GFP alone nor the mutant (R958G) GFP-I/LWEQ construct had any effect on F-actin organization when expressed at similar levels (not shown). Scale bar: 10 μm.

The structural relationships among I/LWEQ module proteins are illustrated in Fig. 4, where representatives of the four major branches of the I/LWEQ protein

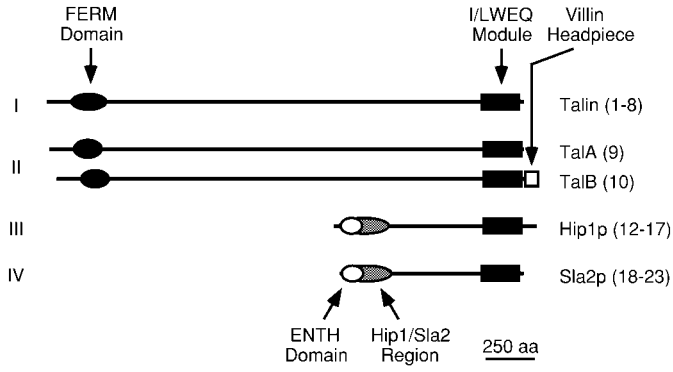


FIG. 4. I/LWEQ module superfamily. Representatives of I/LWEQ module proteins are drawn to relative scale showing their modular structures. The I/LWEQ module is very near the C-terminus except for TalB, which has an additional C-terminal villin headpiece domain. The ENTH domain has been identified in several different proteins which are components of the cytoskeletal and endocytic machinery (26). Hip1p also contains an ENTH domain and is associated with membrane fractions (7). The function of the N-terminal Hip1p/Sla2p region (sequence similarity: 37.1%, including the ENTH domain) is unknown, but the N-terminus of Hip1p is responsible for the interaction of huntingtin and Hip1p (6, 7). Both metazoan talin and *D. discoideum* TalA and TalB contain an N-terminal FERM (band 4.1) domain (27). This domain is found in a number of different proteins where it mediates an interaction with the plasma membrane. FERM domain proteins include the related cytoskeletal proteins ezrin, moesin, and radixin; several protein tyrosine phosphatases; and merlin, the product of the neurofibromatosis-2 gene; and myosin-VIIA, an unconventional myosin expressed in sensory epithelia; focal adhesion kinase (FAK); and several Janus tyrosine kinases (JAKs) (27, 28). Detailed sequence information on the N-terminal FERM and ENTH domains can be found in the Pfam database (http://pfam.wustl.edu/).

family are diagrammed to relative scale. The I/LWEQ module is near the C-terminus except in the case of *Dd*TalB, which has an additional C-terminal villin headpiece domain containing the determinants responsible for a putative interaction with F-actin (18). *Dd*TalA and *Dd*TalB are clearly related to metazoan talin by virtue of their structural similarities and by their presumptive cellular roles (18, 19).

The exact physiological role of the actin-binding activity of the I/LWEQ module has not yet been determined, but the modular organization of the known I/LWEQ module proteins is suggestive (Fig. 4). Huntingtin has been described as a cytoskeleton-associated protein (7), but whether the actin-binding activity of the Hip1p I/LWEQ module connects the huntingtin/Hip1p complex to the actin cytoskeleton *in vivo* is unknown. *Sc*Sla2p is required for organization of the yeast cortical actin cytoskeleton (4, 20), and although Sla2p is also involved in endocytosis and proper expression of the H^+-ATPase Pma1p, the I/LWEQ module is dispensable for these functions (21, 22). Nevertheless, Yang *et al.* (23) have shown that deletion of residues 768–968 of Sla2p, which are coterminous with the I/LWEQ module (Fig. 1), causes subtle defects in cytoskeletal morphology and that overexpression of this region is lethal and results in the formation of thick actin cables similar to those we see when the I/LWEQ module of talin is expressed in CHO cells (Fig. 3).

We proposed that the role of the I/LWEQ module is to serve as an adaptor to the actin cytoskeleton (5). The modular structure of I/LWEQ module proteins along with the conservation of the I/LWEQ module and its distribution among several groups of divergent proteins, each of which has been associated with the cytoskeleton, strengthens that hypothesis. Our results also show that the presence of an I/LWEQ module in novel protein sequences is a robust predictor of actin binding. Further characterization of the I/LWEQ module and I/LWEQ module proteins at the molecular, cellular, and genomic levels will delineate the roles of this module more completely.

ACKNOWLEDGMENTS

We thank G. Gerisch, the Kazusa DNA Research Institute, and the IMAGE consortium (Research Genetics) for cDNAs. The research was supported by a grant from NIH to S.W.C. and a fellowship from the American Heart Association to R.O.M.

REFERENCES

1. Rees, D. J., Ades, S. E., Singer, S. J., and Hynes, R. O. (1990) *Nature* **347,** 685–689.
2. Moulder, G. L., Huang, M. M., Waterston, R. H., and Barstead, R. J. (1996) *Mol. Biol. Cell* **7,** 1181–1193.
3. Hemmings, L., Rees, D. J. G., Ohanian, V., Bolton, S. J., Gilmore, A. P., Patel, B., Priddle, H., Trevithick, J. E., Hynes, R. O., and Critchley, D. R. (1996) *J. Cell Sci.* **109,** 2715–2726.
4. Holtzman, D. A., Yang, S., and Drubin, D. G. (1993) *J. Cell Biol.* **122,** 635–644.
5. McCann, R. O., and Craig, S. W. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 5679–5684.
6. Wanker, E. E., Rovira, C., Scherzinger, E., Hasenbank, R., Walter, S., Tait, D., Colocelli, J., and Lahrach, H. (1997) *Hum. Mol. Genet.* **6,** 487–495.
7. Kalchman, M. A., Koide, H. B., McCutcheon, K., Graham, R. K., Nichol, K., Nishiyama, K., Kazemi-Esfarjani, P., Lynn, F. C., Wellington, C., Metzler, M., Goldberg, Y. P., Kanazawa, I., Gietz, R. D., and Hayden, M. R. (1997) *Nature Genet.* **16,** 44–53.
8. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
9. Thompson, J. D., Higgins, J. D., and Gibson, T. J. (1994) *Nucl. Acids. Res.* **22,** 4673–4680.
10. Henikoff, S., Henikoff, J. G., Alford, W. J., and Pietrokovski, S. (1995) *Gene-COMBIS* **163,** GC17–GC26.
11. Rost, B. (1996) *Methods Enzymol.* **266,** 525–539.
12. Page, R. D. M. (1996) *Comp. Appl. Biosci.* **12,** 357–358.
13. Baldauf, S. L., and Doolittle, W. F. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 12007–12012.
14. Roger, A. J., Smith, M. W., Doolittle, R. F., and Doolittle, W. F. (1996) *J. Eukaryot. Micriobiol.* **43,** 475–485.
15. Kuma, K., Nikoh, N., Iwabe, N., and Miyata, T. (1995) *J. Mol. Evol.* **41,** 238–246.

16. Berbee, M. L., and Taylor, J. W. (1994) *Can. J. Bot.* **71,** 1114–1127.
17. Feng, D.-F., Cho, G., and Doolittle, R. F. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 13028–13033.
18. Tsujioka, M., Machesky, L. M., Cole, S. L., Yahata, K., and Inouye, K. (1999) *Curr. Biol.* **9,** 389–392.
19. Niewohner, J., Weber, I., Maniak, M., Muller-Taubenberger, A., and Gerisch, G. (1997) *J. Cell Biol.* **138,** 349–361.
20. Li, R., Zheng, Y., and Drubin, D. G. (1995) *J. Cell Biol.* **128,** 599–615.
21. Wesp, A., Hicke, L., Palecek, J., Lombardi, R., Aust, T., Munn, A. L., and Riezman, H. (1997) *Mol. Biol. Cell* **8,** 2291–2306.
22. Na, S., Hincapie, M., McCusker, J. H., and Haber, J. E. (1995) *J. Biol. Chem.* **270,** 6815–6823.
23. Yang, S., Cope, M. J., and Drubin, D. G. (1999) *Mol. Biol. Cell* **10,** 2265–2283.
24. Nagase, T., Ishikawa, K., Nakajima, D., Ohira, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., Ohara, O., and Seki, N. (1997) *DNA Res.* **4,** 141–150.
25. Ishikawa, K., Nagase, T., Suyama, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., and Ohara, O. (1998) *DNA Res.* **5,** 169–176.
26. Kay, B. K., Yamabhai, M., Wendland, B., and Emr, S. D. (1999) *Protein Sci.* **8,** 435–438.
27. Chishti, A. H., Kim, A. C., Marfatia, S. M., Lutchman, M., Hanspal, M., Jindal, H., Liu, S. C., Low, P. S., Rouleau, G. A., Mohandas, N., Chasis, J. A., Conboy, J. G., Gascard, P., Takakuwa, Y., Huang, S. C., Benz, E. J., Bretscher, A., Fehon, R. G., Gusella, J. F., Ramesh, V., Solomon, F., Marchesi, V. T., Tsukita, S., Arpin, M., Louvard, D., Tonks, N. K., Anderson, J. M., Fanning, A. S., Bryant, P. J., Woods, D. F., and Hoover, K. B. (1998) *Trends Biochem. Sci.* **23,** 281–282.
28. Girault, J.-A., Labesse, G., Mornon, J.-P., and Callebaut, I. (1999) *Trends Biochem. Sci.* **24,** 47–53.