Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 6720–6724, October 1987
Biochemistry

The heavy chain of *Acanthamoeba* myosin IB is a fusion of myosin-like and non-myosin-like sequences

(low molecular weight non-muscle myosin/gene sequence/myosin structure/intron position)

GOEH JUNG, EDWARD D. KORN, AND JOHN A. HAMMER III

Laboratory of Cell Biology, National Heart, Lung, and Blood Institute, National Institutes of Health, Building 3, Room B1-22, Bethesda, MD 20892

Communicated by William F. Harrington, June 19, 1987 (received for review April 16, 1987)

ABSTRACT ***Acanthamoeba castellanii*** **myosins IA and IB demonstrate the catalytic properties of a myosin and can support analogues of contractile and motile activity *in vitro*, but their single, low molecular weight heavy chains, roughly globular shapes, and inabilities to self-assemble into filaments make them structurally atypical myosins. We now present the complete amino acid sequence of the 128-kDa myosin IB heavy chain, which we deduced from the nucleotide sequence of the gene and which reveals that the polypeptide is a fusion of myosin-like and non-myosin-like sequences. Specifically, the amino-terminal ≈76 kDa of amino acid sequence is highly similar to the globular head sequences of conventional myosins. By contrast, the remaining ≈51 kDa of sequence shows no similarity to any portion of conventional myosin sequences, contains regions that are rich in glycine, proline, and alanine residues, and lacks the distinctive sequence characteristics of an α-helical, coiled-coil structure. We conclude, therefore, that the protein is composed of a myosin globular head fused not to the typical coiled-coil rod-like myosin tail structure but rather to an unusual carboxyl-terminal domain. These results support the conclusion that filamentous myosin is not required for force generation and provide a further perspective on the structural requirements for myosin function. Finally, we find a striking conservation of intron/exon structure between this gene and a vertebrate muscle myosin gene. We discuss this observation in relation to the evolutionary origin of the myosin IB gene and the antiquity of myosin gene intron/exon structure.**

Acanthamoeba castellanii myosins IA and IB are incapable of self-assembly into bipolar filaments but, like conventional myosins, they can support analogues of contractile and motile activity *in vitro* (for review, see ref. 1). The only well-developed model of actomyosin-dependent contractile and motile activities is the sliding filament model, which depends specifically on the ability of myosins to self-assemble into bipolar filaments (2). Clearly, the properties of myosins IA and IB are at odds with this model. It would seem, therefore, that a full understanding of the structure and function of these nonfilamentous, actin-activated ATPases might provide important insights into the general mechanism of actomyosin-dependent movement. Toward this end, we present the complete amino acid sequence of the myosin IB heavy chain, which was deduced from the nucleotide sequence of the gene.* The results reveal that the myosin IB heavy chain polypeptide is a fusion of myosin-like and non-myosin-like sequences. The structure of the protein suggested by this sequence determination—i.e., a myosin globular head fused to an unusual carboxyl-terminal domain—provides a rationale for understanding the paradoxical properties of myosin I and presents a further perspective on the structural requirements for myosin function.

MATERIALS AND METHODS

DNA sequencing was performed by the dideoxy chain-termination method (3) and by the chemical cleavage method (4). S1 nuclease protection analysis, primer extension analysis, and computer-based sequence analyses were performed as described elsewhere (5).

RESULTS AND DISCUSSION

Nucleotide Sequence, Gene Structure, and Deduced Polypeptide Sequence. We previously reported the isolation of phage clone λ3.9 from a genomic library of *Acanthamoeba* DNA (6). The clone was shown to contain a myosin IB heavy chain gene by hybrid selection analysis, RNA transfer blot analysis, and partial DNA sequencing (6). We now present the complete nucleotide sequence of the gene (Fig. 1), which necessitated the isolation of an overlapping phage clone, λ3.10 (Fig. 2*A*). The gene was sequenced on both strands by the sequencing strategy shown in Fig. 2*B*, and the approximate 5′ and 3′ ends of the mRNA transcript were determined by primer extension and S1 nuclease protection analyses (data not shown; see Fig. 2*C* for summary). For the portion of the gene that encodes the myosin-like amino acid sequence, the intron/exon structure was determined by comparison of the three translation frames with known myosin heavy chain amino acid sequences. For the portion of the gene that encodes the unique, carboxyl-terminal domain (see below), identification of the protein-encoding sequences required the mapping of all intron splice sites by S1 nuclease analysis and careful sequencing of the (G+C)-rich regions (see Fig. 2*B*). The complete structure of the gene is presented in schematic form in Fig. 2*C*. The gene spans ≈6 kb and contains 3504 nucleotides of protein-encoding sequence spread over 24 exons. The splice signals in all 23 introns conform to the GT . . . AG rule (8). The deduced polypeptide sequence contains 1168 residues with a calculated mass of 127.8 kDa, which is in good agreement with the estimated mass of the purified protein (125 kDa, ref. 1). The approximate transcript size of ≈3900 nucleotides, determined by RNA transfer blot analysis (data not shown), is close to the size calculated from the data in Fig. 2.

Amino Acid Sequence Comparisons. The myosin IB heavy chain amino acid sequence was compared, using a dot matrix program, with the heavy chain sequences of the nematode *unc*54 muscle myosin (9) (Fig. 3*A*) and the rat embryonic skeletal muscle myosin (7) (Fig. 3*B*). In both matrices, there is a long, virtually uninterrupted diagonal spanning the first ≈670 residues of the myosin IB sequence (and residues ≈100 through ≈770 of the conventional myosin sequences); this is followed by the complete absence of a long diagonal in the portion of the matrix spanning the remaining ≈500 residues

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.

*This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J02974).

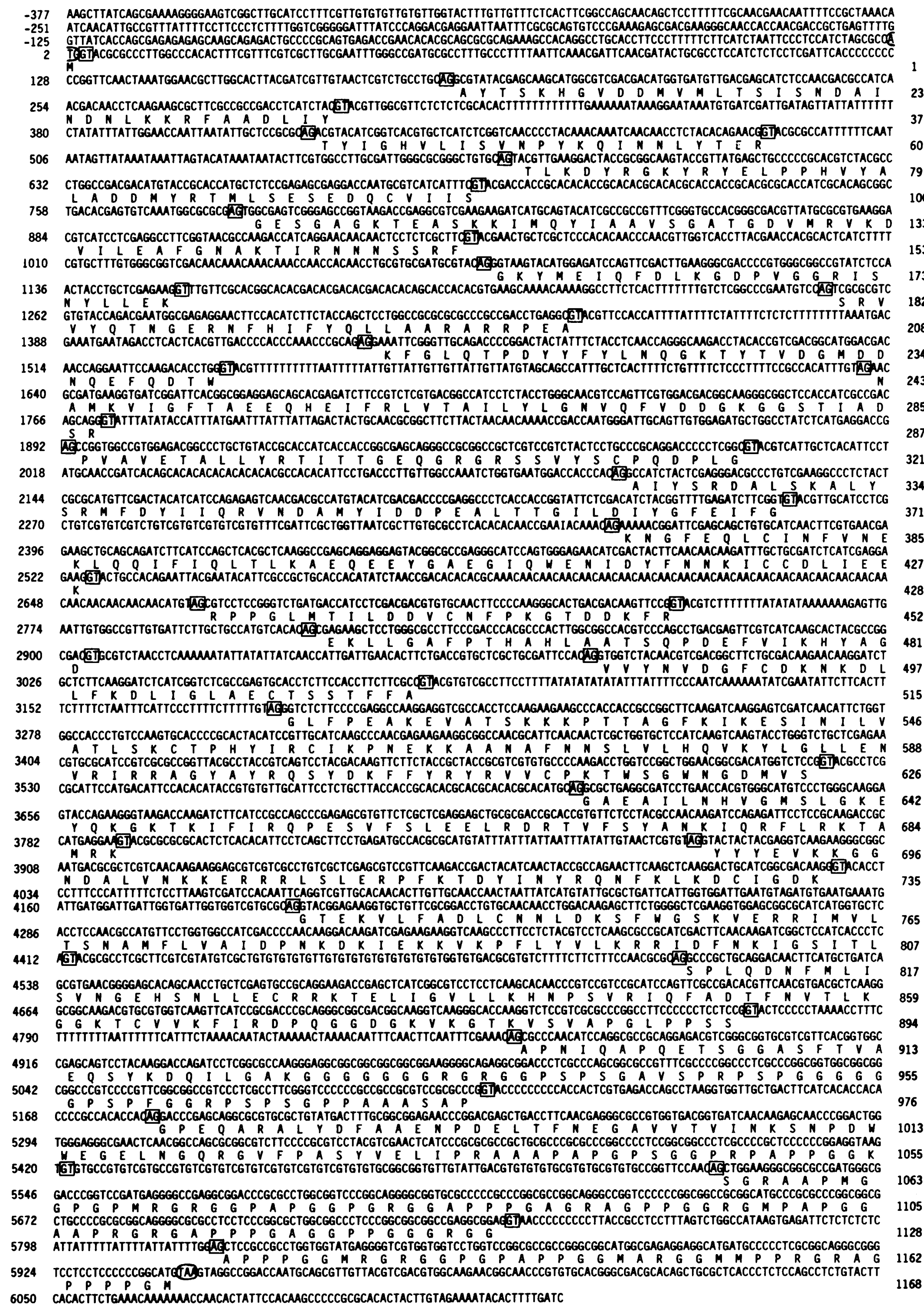

FIG. 1. Complete nucleotide sequence (numbered on left) and deduced amino acid sequence (numbered on right) of the myosin IB heavy chain gene and polypeptide. The initiation (ATG) and termination (TAA) codons are circled; splice donor (GT) and acceptor (AG) sites are boxed. Nucleotides 208–1149 and 3195–3494 were presented previously (6).

of the myosin IB sequence. Therefore, the amino-terminal ≈76 kDa of the myosin IB heavy chain shows strong conservation of sequence and sequence topology relative to most of that portion of conventional myosin sequences that forms the globular head domain (residues 1 through ≈850). Conversely, the carboxyl-terminal ≈51 kDa of the myosin IB heavy chain shows no significant similarity to any portion of conventional myosin sequences. These comparisons indicate that the myosin IB heavy chain is a fusion of conventional myosin head-like sequences and non-myosin sequences (Fig. 2*D*).

Globular Head Domain. Fig. 4 shows a direct comparison of the myosin-like portion of the myosin IB heavy chain amino acid sequence (residues 1–670) with the globular head sequences of the rat and nematode muscle myosins. In total, this portion of the myosin IB sequence shows a 36% exact match (55% similarity, counting conservative substitutions) with both muscle myosin sequences. The alignments require

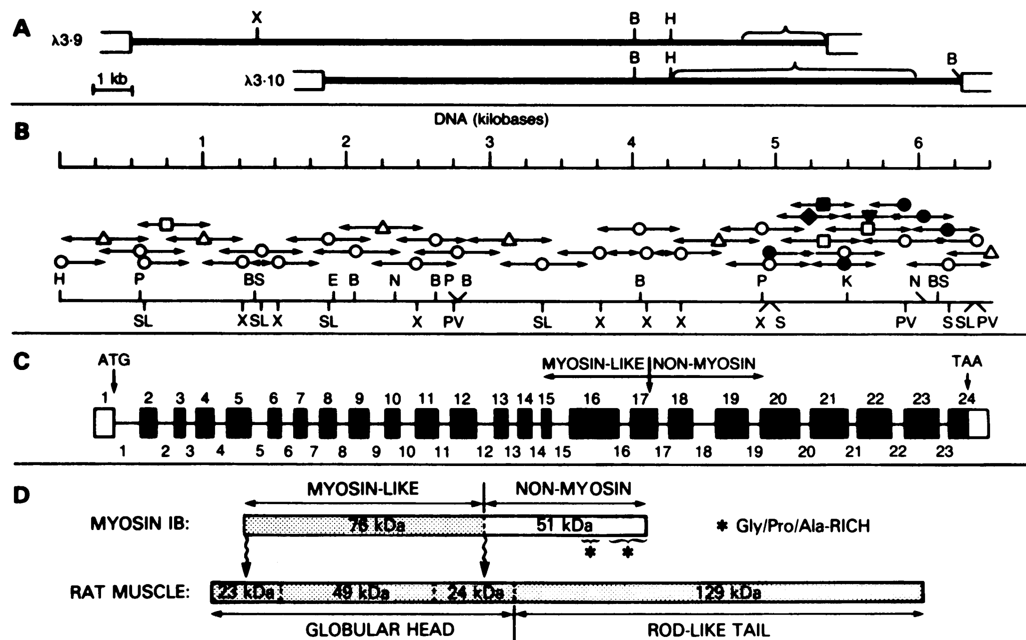

FIG. 2. (A) Restriction enzyme maps (B, *Bam*HI; H, *Hin*dIII; X, *Xba* I) of phage clones λ3.9 (6) and λ3.10. Phage clone λ3.10 was isolated from an *Acanthamoeba* genomic DNA library using as a probe a 2.2-kilobase (kb) *Eco*RI/*Xba* I fragment from λ3.9 (bracketed). The bracketed area of the λ3.10 insert is expanded in *B*. (*B*) Mapped restriction enzymes are *Bgl* II (B), *Bst*EII (BS), *Eco*RI (E), *Hin*dIII (H), *Kpn* I (K), *Not* I (N), *Pst* I (P), *Pvu* II (PV), *Sal* I (SL), *Sst* I (S), and *Xho* I (X). DNA sequencing by the dideoxy method (open symbols) was performed from mapped restriction sites (○) and *Sau*3A sites (□) and from partially sequenced templates using synthetic deoxyoligonucleotides as primers (△). The (G+C)-rich regions of the gene were also sequenced by the chemical cleavage method (closed symbols) from mapped sites (●), *Nar* I sites (◆), *Sau*3A sites (■), and *Taq* I sites (▼). (*C*) Schematic of the gene structure. Exons, filled boxes, numbered above; introns, solid lines, numbered beneath; 5′ and 3′ nontranslated sequences, open boxes. The approximate position of the 5′ end of the transcript was obtained by S1 nuclease protection analysis [the probe, 286-base-pair (bp) *Alu* I/*Nco* I restriction fragment corresponding to nucleotides −288 to −2, yielded a single major protected fragment of ≈128 bp, corresponding to a start site at ≈−130] and primer extension analysis (the primer, a 111-bp *Aha* II/*Sau*3A fragment corresponding to nucleotides 244–355, yielded a major extension product of ≈310 bp, which also corresponds to a start site at ≈−130) (data not shown). The approximate position of the 3′ end of the transcript (≈150 nucleotides 3′ of the TAA) was obtained by S1 nuclease analysis (the probe, a 420-bp *Sal* I/*Eco*RI fragment corresponding to nucleotide 5982 to a 3′ *Eco*RI site, yielded a single major protected fragment of ≈112 bp) (data not shown). (*D*) Schematic of the myosin IB heavy chain polypeptide structure versus the rat embryonic skeletal muscle myosin heavy chain polypeptide structure (7).

few spacing adjustments and, in most cases, the regions of strong similarity between the myosin IB and nematode myosin sequences coincide with the regions of strong similarity between the myosin IB and rat myosin sequences. Some of these regions have been linked to the invariant functions of the myosin head—e.g., binding of nucleotide, interaction with actin, and catalysis (10). For example, the region of conventional myosin sequences homologous to myosin IB residues ≈75–100 (Fig. 4) has been implicated in the formation of the nucleotide binding site (9, 10). Similarly, the conserved region containing the reactive cysteines in muscle myosins (Fig. 4) is thought to lie close to the catalytic site in the folded protein (11); interestingly, in myosin IB, sulfhydryls 1 and 2 have been replaced by arginine and tyrosine residues, respectively. Overall, these results suggest that this portion of the myosin IB heavy chain sequence folds into a three-dimensional structure that is very similar to the globular head domain of conventional myosins (12). This conclusion is consistent with the structural and enzymatic properties of myosin IB.

The regions of sequence within this myosin IB globular head domain that show little similarity to the sequences of conventional myosins include the characteristic sites (for muscle myosins) of tryptic cleavage within the myosin head [Fig. 4; the "23 kDa/50 kDa" and "50 kDa/20 kDa" cleavage sites are those of muscle myosin (13)]. The myosin IB sequence also has small deletions at both of these sites relative to conventional myosins (Fig. 4). Similar observations have been made in previous myosin sequence comparisons (5, 7, 9). The myosin IB sequence also shows no significant similarity to the carboxyl-terminal ≈5% of conventional myosin globular head sequences (residues ≈800–850). This later region is thought to contain the flexible "swivel" that attaches the head to the rod-like tail (10). Finally, the myosin IB globular head sequence is missing ≈80 amino-terminal residues relative to conventional myosins. These ≈80 residues are highly divergent between conventional myosins (5, 7, 9).

Carboxyl-Terminal Domain. In conventional myosins, the globular heads are joined to the α-helical, coiled-coil, rod-like tail, which mediates the self-assembly of myosin molecules into bipolar filaments (10). This myosin rod sequence contains throughout a 7-residue repeat of hydrophobic residues. The carboxyl-terminal ≈500 residues of the myosin IB heavy chain, in addition to showing no similarity to any portion of conventional myosin sequences (Fig. 3), also shows no evidence of this hydrophobic residue repeat. Furthermore, much of this portion of the myosin IB sequence is rich in proline residues, which are incompatible with the α-helical conformation. These results indicate conclusively that myosin IB has no coiled-coil rod-like tail, which is in agreement with its monomeric nature and inability to self-assemble into filaments. These results support the conclusion, drawn from previous study of the protein (1), that filamentous myosin is not required for force generation. This conclusion is consistent with a number of recent findings (reviewed in ref. 11) suggesting that the conformational change(s) in the myosin molecule that results in crossbridge movement occurs within the myosin head.

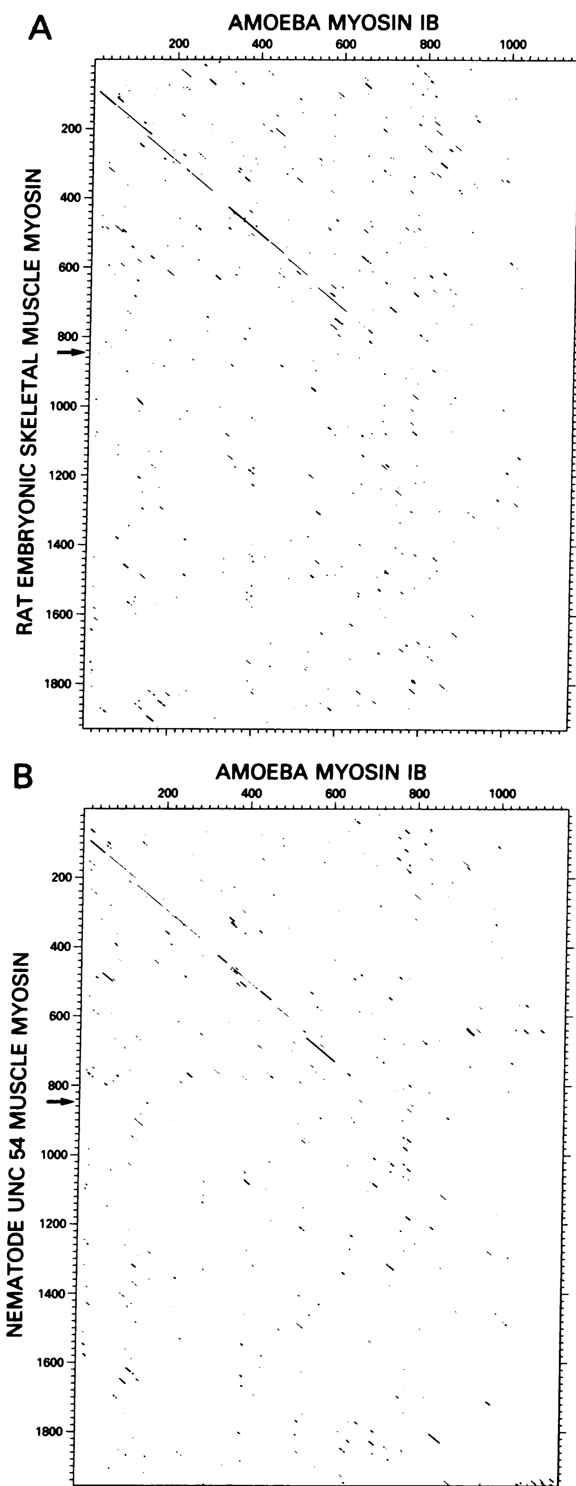

FIG. 3. Dot matrix comparisons between the myosin IB heavy chain amino acid sequence and the heavy chain sequences of the nematode *unc*54 myosin (*A*) and the rat embryonic skeletal muscle myosin (*B*). The arrows in the conventional myosin sequences mark the boundary between the globular head region and rod-like tail. The mutational index was used (counts conservative substitutions as well as exact matches), the window size was 25 residues, and the minimum score was 20. A dot appears in the matrix when 20 or more residues out of 25 match.

The unusual, carboxyl-terminal sequence of myosin IB contains two regions that are rich in glycine, proline, and alanine residues (Fig. 5). The first region (region A) spans 56 residues (residues 923–978) and is 39% glycine, 23% proline, and 11% alanine. The second region (region B) spans the carboxyl-terminal 135 residues of the heavy chain (residues 1034–1168) and is 36% glycine, 30% proline, and 15% alanine. These unusual amino acid compositions are very similar to the amino acid composition of a tryptic peptide containing the carboxyl-terminal 26 kDa of the myosin IA heavy chain (34% glycine, 21% proline, and 12% alanine; ref. 14). Despite their glycine- and proline-rich nature, neither region A nor B contains the repeating 3-residue motif of the collagen helix (Gly-Xaa-Yaa$_n$; Xaa is usually proline and Yaa is usually hydroxyproline). However, regions A and B contain numerous, short, imperfect repeats. For example, region B contains 11 copies of a 7-residue repeat whose consensus sequence is Gly-Gly-Pro-Pro-Pro-Gly-Gly (Fig. 5). We also note that region B contains 17 basic residues and no acidic residues.

The myosin I molecule has been shown to contain two actin binding sites (1). One site is located within the globular head domain, whereas the second site is located within the unusual carboxyl-terminal domain. Based on these observations, a model of how this nonfilamentous myosin might support contractile activity has been proposed (1). Interestingly, we did not find any significant homology between the two distinct portions of the myosin IB heavy chain. Therefore, these two actin binding sites are apparently quite distinct at the sequence level. Finally, we also have not found any significant similarity between the sequence of the unusual carboxyl-terminal domain and any sequence in current protein and DNA databanks.

Conservation of Intron/Exon Structure. We observed a striking conservation of intron positions between the portion of the myosin IB gene that encodes the globular head domain (interrupted by introns 1–16) and the analogous regions of the rat embryonic skeletal muscle myosin gene (interrupted by introns 3–21) and the nematode *unc*54 muscle myosin gene (interrupted by introns 1–4). Specifically, the positions where myosin IB introns 2, 4, 5, 8, 10, 11, and 12 interrupt the protein-encoding sequence are exactly conserved relative to introns 4, 6, 8, 11, 13, 14, and 15, respectively, in the rat gene. In addition, the positions of myosin IB introns 7 and 9 are only very slightly shifted relative to introns 10 and 12 in the rat gene [the shifts are 1 and 4 bases, respectively, and may have arisen by junctional sliding (15)]. The positions of myosin IB introns 2 and 12 are also identical to the positions of introns 2 and 4 in the nematode gene. This remarkable conservation of intron/exon structure, together with the observed conservation of protein sequence (Fig. 4), leads us to speculate that the myosin-like portion of the myosin IB molecule shares an ancestral gene with vertebrate and invertebrate muscle myosins.

A. castellanii is estimated to have diverged from the main line of eukaryotic descent sometime between the divergence of yeast and the divergence of plants and animals (i.e., $\approx$1–1.2 $\times$ 10^9 years ago) (15, 16). Our results suggest, therefore, that as far back as 1 $\times$ 10^9 years ago the putative ancestral myosin gene contained introns in at least the nine positions that are conserved between the present-day amoeba myosin IB and rat myosin genes. These observations provide additional support for the hypothesis of Strehler *et al.* (7) that this putative ancestral myosin gene was highly split. Our observations also concur with recent comparisons of gene structures (15) that favor the idea that the first genes were split rather than the alternative view that they were continuous, with introns inserted at random later. The very narrow size distribution for exons in the myosin IB gene [158 $\pm$ 15 bp (mean $\pm$ SEM)] is, furthermore, inconsistent with random insertion of introns into a preexisting, continuous structure (15).

In summary, we conclude that myosin IB is composed of a myosin globular head fused not to the typical rod-like myosin tail structure but rather to an unusual, glycine-,

FIG. 4. Comparison of the myosin IB globular head amino acid sequence (MIB) with the globular head sequences of the rat embryonic skeletal muscle myosin (RAT) and the nematode *unc*54 myosin (NEM). The numbering refers to the initiator methionine as residue 1. The similarity with the rat and nematode sequences begins at residue 80 and residue 78, respectively. Bars, exact matches; colons, conservative amino acid changes; dashes, spacing adjustments. The positions of sulfhydryl 1 (SH1), sulfhydryl 2 (SH2), and the characteristic tryptic cleavage sites ("23 kDa/50 kDa" and "50 kDa/20 kDa") of muscle myosin are shown. The positions where introns interrupt this portion of the gene are indicated (▼).

proline-, and alanine-rich carboxyl-terminal domain (Fig. 2*D*). This polypeptide structure and the intron/exon structure of the gene together suggest that the gene might have been generated by a recombination event in which exons encoding the myosin globular head domain were fused to exons encoding the unusual carboxyl-terminal domain. Though this seems plausible, the origin of the exons encoding the carboxyl-terminal domain remains unknown. Nevertheless, a protein with very similar properties to *Acanthamoeba* myosin I has been purified from *Dictyostelium discoideum* (17). Furthermore, the 110-kDa protein from avian intestinal brush border microvilli has been shown to have myosin-like properties (18). These facts, together with the recent report (19) that *Acanthamoeba* myosin I might function as a motor for moving intracellular vesicles on actin filaments, indicate that a careful search for myosin I-like molecules in other organisms is warranted.

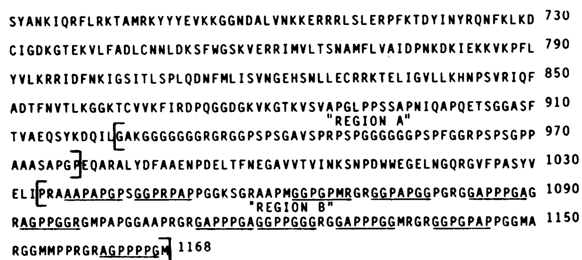

FIG. 5. Deduced amino acid sequence of the unusual carboxyl-terminal domain. Regions A and B are bracketed and 11 copies of a repeating sequence are underlined (see text).

1. Korn, E. D., Atkinson, M. A. L., Brzeska, H., Hammer, J. A., III & Jung, G. (1987) *J. Cell. Biochem.*, in press.
2. Huxley, H. E. (1974) *J. Physiol. (London)* **243**, 1–43.
3. Sanger, F. S., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **83**, 5463–5467.
4. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
5. Hammer, J. A., III, Bowers, B., Paterson, B. M. & Korn, E. D. (1986) *J. Cell Biol.*, in press.
6. Hammer, J. A., III, Jung, G. & Korn, E. D. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4655–4659.
7. Strehler, E. E., Strehler-Page, M., Perriard, J., Periasamy, M. & Nadal-Ginard, B. (1986) *J. Mol. Biol.* **190**, 291–317.
8. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
9. Karn, J., Brenner, S. & Barnett, L. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4253–4257.
10. McLachlan, A. D. (1984) *Annu. Rev. Biophys. Bioeng.* **13**, 167–189.
11. Tregear, R. T. (1986) *Nature (London)* **321**, 563.
12. Bajaj, M. & Blundell, T. (1984) *Annu. Rev. Biophys. Bioeng.* **13**, 453–492.
13. Tong, S. W. & Elzinga, M. (1983) *J. Biol. Chem.* **258**, 13100–13110.
14. Lynch, T. J., Albanesi, J. P., Korn, E. D., Robinson, E. A., Bowers, B. & Fujisaki, H. (1986) *J. Biol. Chem.* **261**, 17150–17162.
15. Gilbert, W., Marchionni, M. & McKnight, G. (1986) *Cell* **46**, 151–154.
16. Sogin, M. L., Elwood, H. J. & Gunderson, J. H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1383–1387.
17. Cote, G. P., Albanesi, J. P., Ueno, T., Hammer, J. A., III & Korn, E. D. (1985) *J. Biol. Chem.* **260**, 4543–4546.
18. Collins, J. A. & Borysenko, C. W. (1984) *J. Biol. Chem.* **259**, 14128–14135.
19. Adams, R. J. & Pollard, T. D. (1985) *Nature (London)* **322**, 754–756.