Complete Nucleotide Sequence and Deduced Polypeptide Sequence of a Nonmuscle Myosin Heavy Chain Gene from *Acanthamoeba*: Evidence of a Hinge in the Rodlike Tail

John A. Hammer III, Blair Bowers, Bruce M. Paterson, and Edward D. Korn

Laboratory of Cell Biology, National Heart, Lung, and Blood Institute, and Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20893

Abstract. We have completely sequenced a gene encoding the heavy chain of myosin II, a nonmuscle myosin from the soil ameba *Acanthamoeba castellanii*. The gene spans 6 kb, is split by three small introns, and encodes a 1,509-residue heavy chain polypeptide. The positions of the three introns are largely conserved relative to characterized vertebrate and invertebrate muscle myosin genes. The deduced myosin II globular head amino acid sequence shows a high degree of similarity with the globular head sequences of the rat embryonic skeletal muscle and nematode *unc* 54 muscle myosins. By contrast, there is no unique way to align the deduced myosin II rod amino acid sequence with the rod sequences of these muscle myosins. Nevertheless, the periodicities of hydrophobic and charged residues in the myosin II rod sequence, which dictate the coiled-coil structure of the rod and its associations within the myosin filament, are very similar to those of the muscle myosins. We conclude that this ameba nonmuscle myosin shares with the muscle myosins of vertebrates and invertebrates an ancestral heavy chain gene. The low level of direct sequence similarity between the rod sequences of myosin II and muscle myosins probably reflects a general tolerance for residue changes in the rod domain (as long as the periodicities of hydrophobic and charged residues are largely maintained), the relative evolutionary "ages" of these myosins, and specific differences between the filament properties of myosin II and muscle myosins. Finally, sequence analysis and electron microscopy reveal the presence within the myosin II rodlike tail of a well-defined hinge region where sharp bending can occur. We speculate that this hinge may play a key role in mediating the effect of heavy chain phosphorylation on enzymatic activity.

MYOSIN II from the soil ameba *Acanthamoeba castellanii* is composed of a pair of 175-kD heavy chains and two pairs of light chains (17 and 17.5 kD) (26). Electron microscopy reveals that these subunits are arranged into a structure similar to that of other myosins, i.e., a highly asymmetric molecule possessing two globular heads and an extended rodlike tail (25, 34). The tail mediates the self-assembly of myosin II molecules into small bipolar filaments containing 16–40 monomers depending on the polymerization conditions (34). The myosin II actin-activated Mg^{2+}-ATPase activity, which resides in the globular head (1), is regulated by phosphorylation of the heavy chain (9). Unphosphorylated myosin II is activated 40-fold by F-actin while myosin II, which has three phosphates incorporated into each heavy chain by a specific kinase, is not actin activatable. Interestingly, the phosphorylation sites reside within a ~30-residue nonhelical tailpiece at the very carboxyl-terminal end of the rodlike tail (10), nearly 100 nm away from the catalytic site. Recent evidence indicates that heavy chain phosphorylation exerts its effects on ATPase activity by altering the conformation of the small bipolar filaments formed by myosin II (23, 24), although what occurs at the molecular level remains unknown.

In an effort to understand the structural, regulatory, and self-assembly properties of this nonmuscle myosin at a molecular level, we have characterized an ameba gene encoding a myosin II heavy chain. We previously reported the isolation of a recombinant phage clone which contains this gene (18). We now provide the complete nucleotide sequence of the gene and the deduced amino acid sequence of the heavy chain polypeptide.[1] Comparison of our results with those for vertebrate and invertebrate muscle myosins (22, 40) has allowed us to draw several conclusions concerning the possible evolution of this myosin II heavy chain gene and the structural requirements for myosin function. Furthermore, we have identified a hinge or flexible point in the myosin II rodlike tail, first at the sequence level and subsequently by rotary-shadowed electron microscopy of myosin II molecules. This interesting structural feature may play a role in

1. These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00624.

mediating the effect of heavy chain phosphorylation on filament conformation and, consequently, enzymatic activity.

Materials and Methods

General Methods

General procedures, including plasmid subcloning, restriction enzyme digestion, agarose and acrylamide gel electrophoresis, and purification of DNA fragments from gels were performed using standard methods (27). *Acanthamoeba* polyadenylated (poly A^+)[2] RNA was prepared as described previously (18). Restriction and modification enzymes and polymerases were purchased from Bethesda Research Laboratories (Gaithersburg, MD) unless indicated otherwise.

DNA Sequencing

Ameba genomic DNA restriction fragments were subcloned in phage M13 mp18 and/or mp19 (using host *Escherichia coli* strain JM 101) and single-stranded recombinant M13 templates were prepared as described previously (31). The nucleotide sequences were determined by the dideoxy chain termination method of Sanger et al. (36) using Klenow fragment and [^{35}S]dATPαS. Sequencing reactions were resolved on 6% polyacrylamide/0.3% bisacrylamide/8.3 M urea denaturing gels containing a gradient of Tris-borate/EDTA buffer from 0.5 to 2.5× (1× buffer is 0.089 M Tris-borate, 0.1 mM EDTA [pH 8.3]). Either a universal M13 primer (Collaborative Research, Inc., Waltham, MA) or unique deoxyoligonucleotides (17-19 nucleotides long), which were complementary to previously determined ameba DNA insert sequences, were used to prime the sequencing reactions. The later deoxyoligonucleotides were synthesized by the methoxy phosphoramidite method on a DNA synthesizer (Model 380B; Applied Biosystems, Inc., Foster City, CA), and were kindly provided by Dr. M. Brownstein (National Institute of Mental Health, Bethesda, MD). Primer extension products and a 51-bp Xho I fragment from the gene (nucleotides 4,918–4,969; Figs. 1 and 2) were sequenced by the base-specific chemical cleavage reactions according to Maxam and Gilbert (28).

Primer Extension Analysis

To prepare the 62-bp Hha I/Hpa II restriction fragment used as the primer, a pUC 18 plasmid subclone that contained a 340-bp Xho I fragment (nucleotides 541–881; Fig. 1) inserted at the Sal I site was digested with Pst I and Xba I to release the insert. The insert fragment was digested with Hpa II, treated with calf intestinal phosphatase, 5′-end-labeled with [γ^{32}P]dATP and T4 polynucleotide kinase, digested with Hha I, the digest fractionated on a 13% polyacrylamide nondenaturing gel, and the ^{32}P-labeled 62-bp Hha I/Hpa II fragment (nucleotides 700–762; Fig. 1) obtained by electroelution. For each analytical primer extension reaction, 6 μg of ameba poly A^+ RNA and 0.5 × 10^6 dpm of primer were dried, solubilized in 20 μl of hybridization buffer (0.4 M NaCl, 1 mM EDTA [pH 8.0], 40 mM Pipes [pH 6.4], 80% [vol/vol] deionized formamide), brought to 100°C for 2 min, and incubated at 48°C for 16 h. To each sample was added 80 μl of stop solution (0.375 M sodium acetate [pH 5.2], 12.5 mM magnesium acetate, 125 μg/ml tRNA) and 300 μl of ethanol and the DNA precipitated on dry ice/ethanol (2 h). To the dried precipitate was added 20 μl of primer extension buffer (50 mM Tris [pH 7.8], 60 mM KCl, 6 mM $MgCl_2$, 10 mM dithiothreitol, 150 μg/ml bovine serum albumin, and 1.25 mM each of dGTP, dATP, dCTP, and dTTP) and 400 U of Maloney murine leukemia virus reverse transcriptase. After incubation for 60 min at 37°C, the RNA was degraded by addition of 2 μl each of 2 N NaOH and 250 mM EDTA (pH 8.0) and incubation for 30 min at 37°C. To this material was added 176 μl of a solution containing 70 mM Tris (pH 7.5), 0.1% sodium dodecyl sulfate, and 50 μg/ml tRNA. After two phenol extractions and two chloroform extractions, the DNA was collected by ethanol precipitation and the extension products resolved on 7% polyacrylamide/0.35% bisacrylamide/8.3 M urea denaturing gels. For size markers, Hpa II-digested, ^{32}P-end-labeled fragments of pBR 322 were used. To prepare primer extension products for chemical sequencing, individual reactions were scaled up fourfold. The extension products from 10 such reactions were pooled after elution from the gel and sequenced by the method of Maxam and Gilbert (28).

S1-Nuclease Protection Analysis

To prepare the S1 probe, a 718-bp Sal I/Cla I fragment (nucleotides 421–1,139; Fig. 1) was digested with Bst EII, treated with calf intestinal phosphatase, end labeled with [γ^{32}P]dATP and T4 polynucleotide kinase, and the labeled 423-bp Sal I/Bst EII fragment (nucleotides 421–844; Fig. 1) isolated from an acrylamide gel. This fragment was digested with Hinf I and the asymmetrically labeled 388-bp Hinf I/Bst EII fragment (nucleotides 456–844; Fig. 1) was isolated from an acrylamide gel and used as the S1 probe. For individual reactions, 8 μg of ameba poly A^+ RNA and 30,000 dpm of S1 probe in 30 μl of hybridization buffer (same as for primer extension) were brought to 100°C for 3 min and incubated at 49°C for 16 h. To this material was added 420 μl of S1 buffer (0.2 M NaCl, 50 mM sodium acetate [pH 4.5], 1 mM zinc acetate, 10 μg/ml denatured salmon sperm DNA) containing 150 U of *Aspergillus* S1 nuclease (Boehringer Mannheim Diagnostics, Inc., Houston, TX). After incubation for 30 min at 45°C, 10 μl of 0.5 M EDTA (pH 8.0) was added, the sample was extracted with phenol/chloroform, precipitated with ethanol, and the precipitated DNA resolved on a 7% acrylamide/0.3% bisacrylamide/8.3 M urea denaturing gel.

Rotary-shadowed Electron Microscopy

Myosin II "as isolated," i.e., partially phosphorylated (26) (gift of R. Scharff, National Heart, Lung, and Blood Institute, Bethesda, MD), was dialyzed for 20 h against 0.4 M ammonium acetate (pH 7.2). This material was mixed with 2 vol of glycerol and sprayed against freshly cleaved mica (44). The molecules were rotary shadowed at an angle of 6° in a 360 M vacuum evaporator (Balzers, Hudson, NH) fitted with an electron gun for Pt evaporation. Micrographs of the shadowed molecules were taken at primary magnifications of 38,000 or 58,400 in a Philips 410 electron microscope. Length measurements were made to the nearest micrometer from negatives enlarged 10-fold with a comparator (Nikon, Inc., Garden City, NY).

Computer Programs

Nucleotide sequences were analyzed with the aid of the DEC-10 DNA:SEQ and PRT:ALN programs (Division of Computer Research and Technology, National Institutes of Health, Bethesda, MD). Dot matrix analyses were performed on a VAX-11 computer using the DOT MATRIX program of the Protein Identification Resource, National Biomedical Research Foundation, Georgetown University Medical Center (Washington, DC). The specific parameters used for the alignments are presented in the legend to Fig. 5. The secondary structure predictions for the myosin II rod amino acid sequence were performed using the statistical method of Garnier et al. (14), in which each residue is unambiguously assigned one of four conformational states: α-helix, β-sheet, reverse turn, or random coil. The decision constants used were the standard values for "helix rich" (>50%) proteins (14).

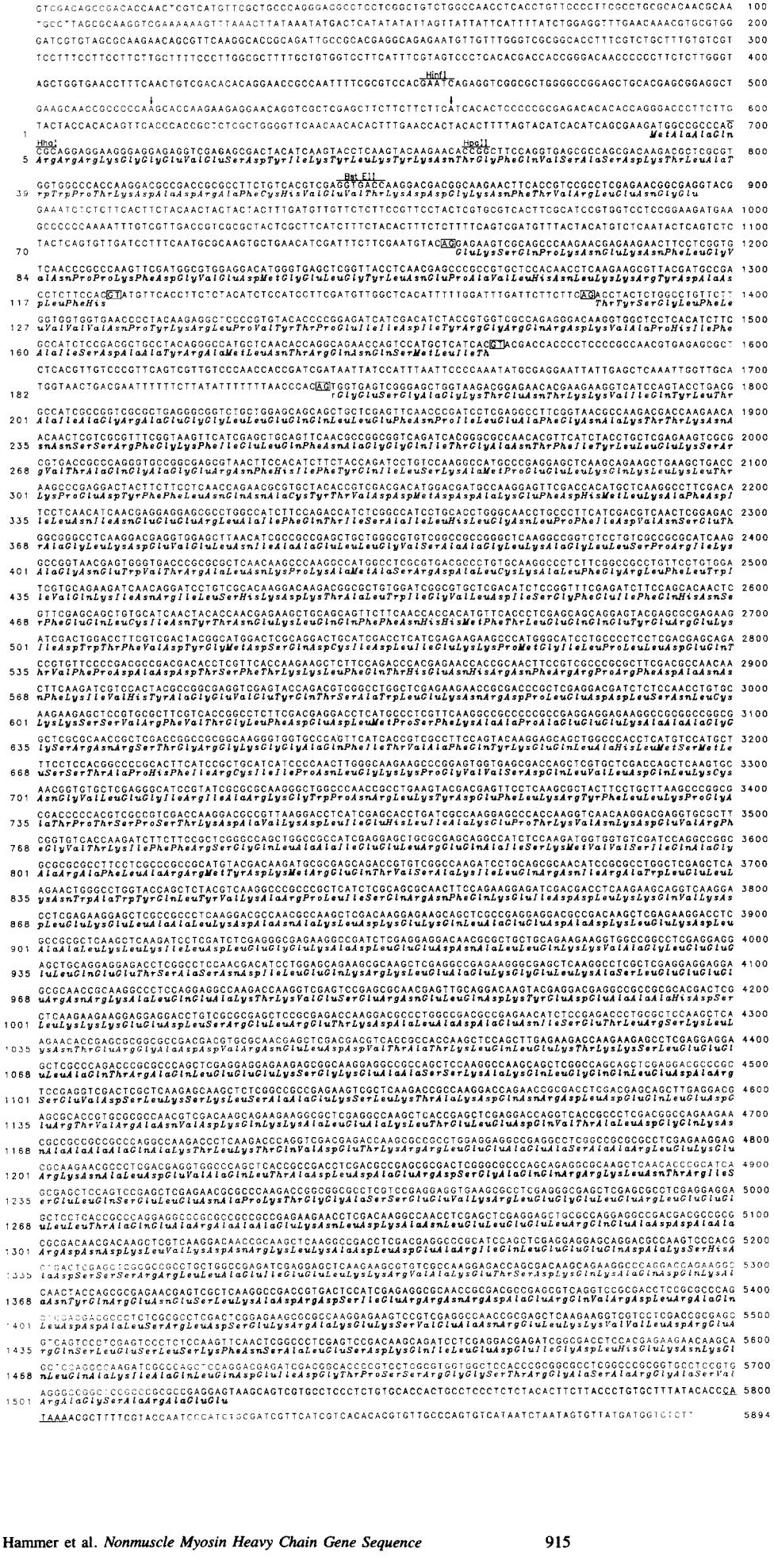
Results

Nucleotide Sequence, Gene Structure, and Deduced Polypeptide Sequence

We previously reported the isolation of recombinant phage clone λ4·13 from a genomic library of *Acanthamoeba* DNA

2. *Abbreviations used in this paper*: IVS, intervening sequence; S1 and S2, subfragment 1 and 2, respectively.

Figure 1. Complete nucleotide sequence of the myosin II heavy chain gene and deduced amino acid sequence of the heavy chain polypeptide. The nucleotide sequence is numbered on the right. The splice donor and acceptor sites are boxed and the putative polyadenylation signal, CATAAA (nucleotides 5,799–5,804), is underlined. The Hinf I and Bst EII sites used to make the probe for S1 nuclease analyses are overlined, as are the Hha I and Hpa II sites used to make the primer for primer extension analysis. The two arrows mark the positions of the major and minor transcription start sites. The amino acids are numbered on the left (each number refers to the amino acid that corresponds to the first complete codon on the line). Nucleotides 1,239–2,074 and 5,470–5,747, along with the deduced amino acid sequence for these regions, were presented previously (18).


HinfI
HhaI
HpaII
BstEII

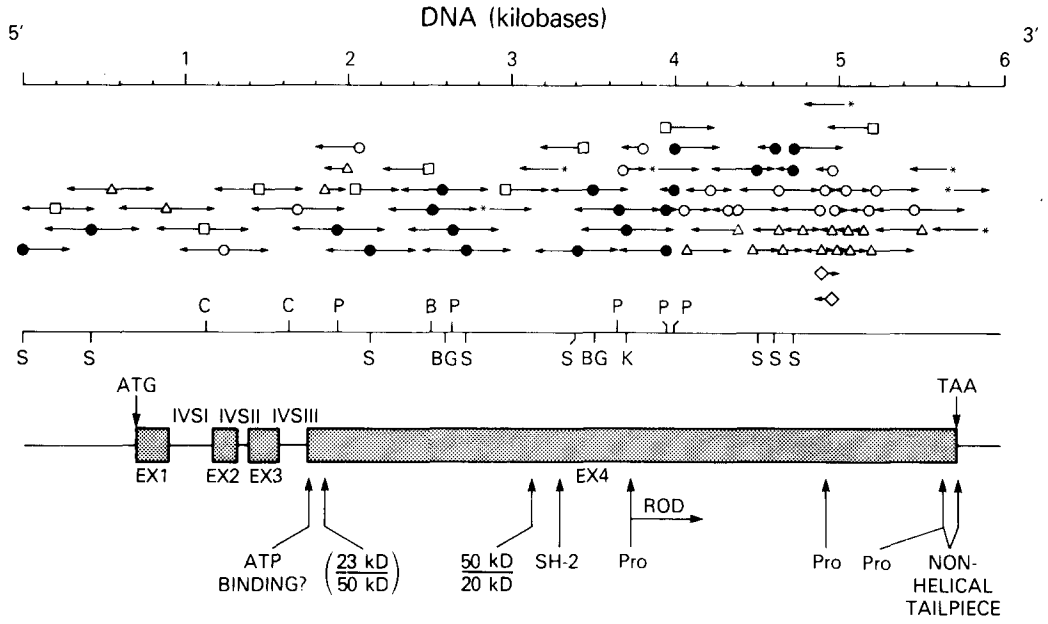


Figure 2. Restriction enzyme map, sequencing strategy, and schematic of the intron/exon structure. The mapped restriction sites are Bam HI (*B*), Bgl II (*BG*), Cla I (*C*), Kpn I (*K*), Pst I (*P*), and Sal I (*S*). DNA sequencing by the dideoxy chain termination method (14) was performed from mapped restriction sites (●), Sau 3a sites (□), Sst I sites (○), and Xho I sites (△), and from partially sequenced templates using synthetic deoxyoligonucleotides to extend the sequences (*). A 51-bp Xho I fragment (nucleotides 4,918–4,969), which contained the hinge region of the rod, was also sequenced (◇) by the chemical method of Maxam and Gilbert (28). (*Arrows*) The direction and extent of each sequencing effort. (*Gray boxes*) The four protein encoding exons (EX1-4). The three introns are labeled intervening sequence I–III (*IVS I–III*). The positions within the gene of various landmarks discussed in the text are shown.

using as a heterologous probe a portion of the nematode *unc* 54 myosin heavy chain gene (18). Phage clone λ4·13 was identified as containing a myosin II heavy chain gene based on hybrid selection analysis, Northern blot analysis, and partial DNA sequence analysis (18). We now present the complete nucleotide sequence of the gene (Fig. 1), which was determined on both strands using the sequencing strategy shown in Fig. 2. In sequencing the portion of the gene encoding the rodlike tail, we took advantage of the plethora of Sst I and Xho I restriction sites. The recognition sequences of these enzymes encode the dipeptides Glu/Leu and Leu/Glu, respectively, which appear repetitively in the amino acid sequence of the tail.

The 5′ end of the gene was identified by mapping the transcription initiation site using a combination of primer extension analysis, S1 nuclease protection analysis, and sequencing of primer extension products (Fig. 3). For primer extension analysis, a 64-bp Hha I/Hpa II restriction fragment was used as the primer. This fragment corresponds to positions +12 to +76, with the A residue of the putative initiator methionine codon ATG as residue +1. Reactions containing ameba poly A^+ RNA yielded a major extension product of 205 bp (corresponding to a possible transcription start site at position −129) and a minor extension product of 248 bp (corresponding to a possible start site at position −172) (Fig. 3 *A*, lane *2*). Control reactions using either rat brain poly A^+ RNA or rat liver tRNA yielded no extension products (Fig. 3 *A*, lanes *3* and *4*, respectively). The existence of these two putative transcription initiation sites was supported by S1 nuclease protection analysis using as a probe a 388-bp Hinf I/Bst EII restriction fragment corresponding to positions −230 to +161. Reactions containing ameba poly A^+ RNA yielded a major protected fragment of 290 bp (corresponding to the transcription start site at −129) and a minor protected fragment of 333 bp (corresponding to the start site at −172) (Fig. 3 *B*, lane *1*). Furthermore, the ratio in the yield of the major and minor protected fragments was very similar to the ratio in the yield of the major and minor extension products (~10:1). To confirm these results, the sequences of the major and minor extension products were determined by the method of Maxam and Gilbert (28) after purification of the fragments from the gel. Both sequences were colinear throughout with the sequence of the genomic DNA from λ4·13 (data not shown). Taken together, these observations indicate that transcripts from this myosin II gene initiate at two sites (−129 and −172), with the majority of the transcripts (~ 90%) initiating at position −129 (see Fig. 3 *C*). The first ATG in the sequence 3′ of both initiation sites marks the start of the protein encoding sequence. The possibility that one of these two transcription start sites represents the transcript of another myosin II gene (with an identical nontranslated RNA leader sequence) cannot be totally ruled out. However, this possibility seems very remote in light of the fact that within multigene families the nontranslated RNA leader sequences of different members are usually quite divergent (6).

The approximate 3′ end of the gene was localized previously by identification of the heavy chain termination codon (10, 18). While we have not mapped the 3′ end of the transcript, we estimate that the 3′ nontranslated sequence and the poly A^+ tail comprise a total of ~250 nucleotides. This estimate is based on a mRNA size of ~4,900 nucleotides (determined using higher resolution RNA blots than used previ-

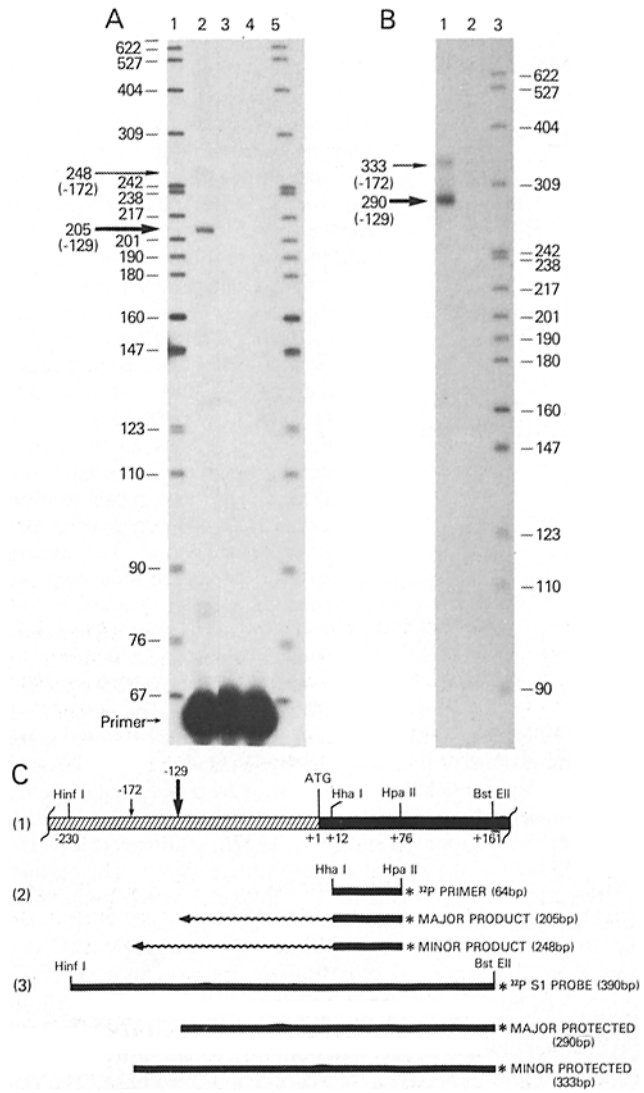


Figure 3. Mapping the transcription initiation sites by primer extension and S1 nuclease protection analyses. (*A*) Primer extension analysis. A 64-bp, ^{32}P-labeled Hha I/Hpa II restriction fragment (corresponding to positions +12 to +76) was used to prime the extension reactions. The extension products were subjected to electrophoresis on a 7% polyacrylamide/8.3 M urea gel. Lane *1*, size markers (Hpa II-digested, ^{32}P-labeled pBR 322); lane *2*, extension reaction using 6 μg of ameba poly A$^+$ RNA; lane *3*, extension reaction using 6 μg of rat brain poly A$^+$ RNA; lane *4*, extension reaction using 6 μg of rat liver tRNA; lane *5*, size markers. The major, 205-bp extension product and the minor, 248-bp extension product are indicated by large and small arrows, respectively (lane *2*). The estimated transcription start sites based on these extension products are in parentheses. The minor extension product appeared on longer exposures to be two or three closely spaced bands. (*B*) S1 nuclease protection analysis. A 390-bp, ^{32}P-labeled Hinf I/Bst EII restriction fragment (corresponding to positions −230 to +161) was used as the probe. The protected fragments were resolved as above. Lane *1*, reactions using 8 μg of ameba poly A$^+$ RNA; lane *2*, reactions using 8 μg of rat liver tRNA; lane *3*, size markers. The major, 290-bp protected fragment and the minor, 333-bp protected fragment (lane *1*) are indicated by large and small arrows, respectively. The estimated transcription start sites based on these protected fragments are in parentheses. (*C*) Schematic of these results. The numbers given for the sizes of the extension products, protected fragments, and transcription start sites are best estimates and could be in error by several bases.

ously [18]; data not shown), a 5′ nontranslated leader of 129 nucleotides, and a protein-encoding sequence of 4,527 nucleotides (see below). Interestingly, the sequence CATAAA, which is very similar to the consensus polyadenylation signal AATAAA (13), is found 68 nucleotides 3′ of the termination codon (Fig. 1).

Between the ATG encoding the initiator methionine and the TAA stop codon are 5,636 nucleotides of genomic DNA sequence (Fig. 1). To identify the protein-encoding sequences, the three reading frames were compared with the published amino acid sequences of the nematode *unc* 54 muscle myosin heavy chain (22), the rat embryonic skeletal muscle myosin heavy chain (40), and fragments of the rabbit skeletal muscle myosin heavy chain (12, 41). For the portion of the gene encoding the rodlike tail, where the level of amino acid sequence similarity with other myosins was low, the identification of the correct reading frame was greatly facilitated by the characteristic pattern of the amino acid sequence in this domain, i.e., the regular repeats of hydrophobic and charged residues (see below). Codon usage, which is strongly biased in this gene, as it is in an *Acanthamoeba* actin gene (33), was also useful for identifying the correct reading frame. This bias is reflected in an extraordinary preference for C and G rather than A and T residues in the third position of the codon (Table I). Using the above guidelines, we found that 4,527 nucleotides of protein-encoding sequence are split by three small introns located near the 5′ end of the gene (Figs. 1 and 2). The identification of these three small introns (intervening sequence [IVS] I, 265 bp; IVS II, 70 bp; IVS III, 177 bp) was supported by the presence of typical 5′ donor (GT) and 3′ acceptor (AG) splice site signals (32) and long

Table I. Codon Usage in the Acanthamoeba Myosin II Heavy Chain Gene

Amino acid	First + second base	Third base			
		G	C	A	T
Ser	TC	42	29	0	1
	AG	–	12	–	0
Leu	CT	44	118	0	5
	TT	3	–	0	–
Arg	CG	0	76	0	12
	AG	10	–	0	–
Phe	TT	–	51	–	0
Tyr	TA	–	28	–	0
Cys	TG	–	7	–	1
Pro	CC	10	22	0	1
His	CA	–	20	–	0
Gln	CA	80	–	0	–
Ile	AT	–	62	0	0
Thr	AC	12	52	0	1
Asn	AA	–	71	–	0
Lys	AA	151	–	0	–
Val	GT	31	28	0	1
Ala	GC	16	135	0	5
Asp	GA	–	102	–	6
Glu	GA	166	–	0	–
Gly	GG	1	44	4	21
Trp	TG	11	–	–	–
Met	AT	17	–	–	–

Values of zero indicate that there are no examples of the particular codon in the myosin II gene. Dashes indicate third base nucleotides that are not allowed for the particular amino acid.

Figure 4. Comparison of the deduced amino acid sequence of the myosin II (*MII*) globular head region with the globular head sequences of the nematode *unc 54* myosin (*NEM*) and the rat embryonic skeletal muscle myosin (*RAT*). The numbering refers to the initiator methionine as residue 1. Bars connect exact residue matches and colons connect conservative residue changes. Spacing adjustments, which were introduced by the alignment program (see Materials and Methods), are indicated by dashes. Almost half of the total spacing adjustments occur at or very near the characteristic sites of tryptic cleavage (see below). These regions are thought to represent surface loops in the native protein (22) and in previous myosin sequence comparisons (22, 40) were found to contain insertions/deletions. The percent similarities between the sequences (see text) equal the number of actual residue matches expressed as a percentage of the total number of possible matches (positions containing dashes were not counted as possible matches; if they are counted, the percent similarities are reduced by ~3%). Slightly different spacing adjustments are possible, especially in the large gaps at the tryptic cleavage sites. Nevertheless, such changes would result in only small changes in the degrees of similarity and would not alter the conclusions. The asterisk above residue 190 in the myosin II sequence marks the glutamate that is cross-linked to a UTP analog by photoaffinity labeling (2). The sequence of a 6-residue myosin II peptide containing this photoaffinity label, which was determined by protein chemical methods (2), is identical to amino acid residues 189–194 in the myosin II sequence above. 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 skeletal muscle myosin are shown. The position of the "50 kDa/20 kDa" site (over Arg 642) is the actual site of tryptic cleavage in the myosin II head (1). The positions where the three introns in the myosin II gene interrupt the protein encoding sequence are indicated (▼).

tracts of pyrimidines just 5′ of the acceptor splice sites (32) (Fig. 1). The 4,527 nucleotides of protein-encoding sequence are therefore spread over four exons and yield a deduced heavy chain polypeptide sequence containing 1,509 amino acid residues with a calculated M_r of 170,966 D, which is in good agreement with the M_r of the myosin II heavy chain as determined by gel electrophoresis (35).

Conservation of Intron Positions

Only two other myosin heavy chain genes have been completely characterized, the rat embryonic skeletal muscle myosin gene (40) and the nematode *unc* 54 muscle myosin gene (22). The structures of both the rat gene (40 introns; 17.5 kb total intron DNA) and the nematode gene (eight introns; 1.4 kb total intron DNA) contrast with the relatively simple structure of the ameba myosin II gene (three introns; 0.5 kb total intron DNA). These differences not withstanding, we carefully compared the intron/exon structure of the myosin II gene with the structures of these two muscle myosin genes, since genes possessing a common evolutionary origin often possess a subset of introns whose positions are identical (15). The positions of all three of the myosin II introns (referred to below as IVS I, II, and III) are conserved relative to the rat embryonic gene. Specifically, the positions of IVS II and IVS III (which both interrupt regions of amino acid sequence that are well conserved between these myosins; see below and Fig. 4) are identical to the positions of introns 4 and 6 in the rat gene, respectively (40). This conservation of position even includes the fact that both IVS III and intron 6 interrupt a Thr codon between the second and third base. The position of IVS I (which falls within a region of amino acid sequence that is poorly conserved between myosins; see below and Fig. 4) is identical to the position of intron 3 in the rat gene (but the rat sequence has an apparent deletion of three amino acid residues relative to the myosin II sequence at the intron position corresponding to IVS I; see Fig. 4). The comparison with the nematode *unc* 54 gene reveals a similar conservation of intron positions, with the positions of IVS I and IVS II being identical to the positions of introns 2 and 3 in the *unc* 54 gene, respectively (40). There is, however, no intron in the *unc* 54 gene corresponding to the position of IVS III. Comparisons can also be made with three other muscle myosin genes, the embryonic and adult chicken fast white skeletal muscle myosin genes (16) and the rat α cardiac muscle myosin gene (39), although these comparisons are limited to the positions of IVS I and IVS II because of the paucity of sequence information available for these genes. All three vertebrate genes contain introns at positions corresponding to IVS I and IVS II (but all three of these myosin sequences have an apparent deletion of three residues relative to the myosin II sequence at the intron position corresponding to IVS I). While the positions of the three introns in the myosin II gene are clearly conserved relative to these various muscle myosin genes, the sequences of the corresponding introns do not show any striking similarities.

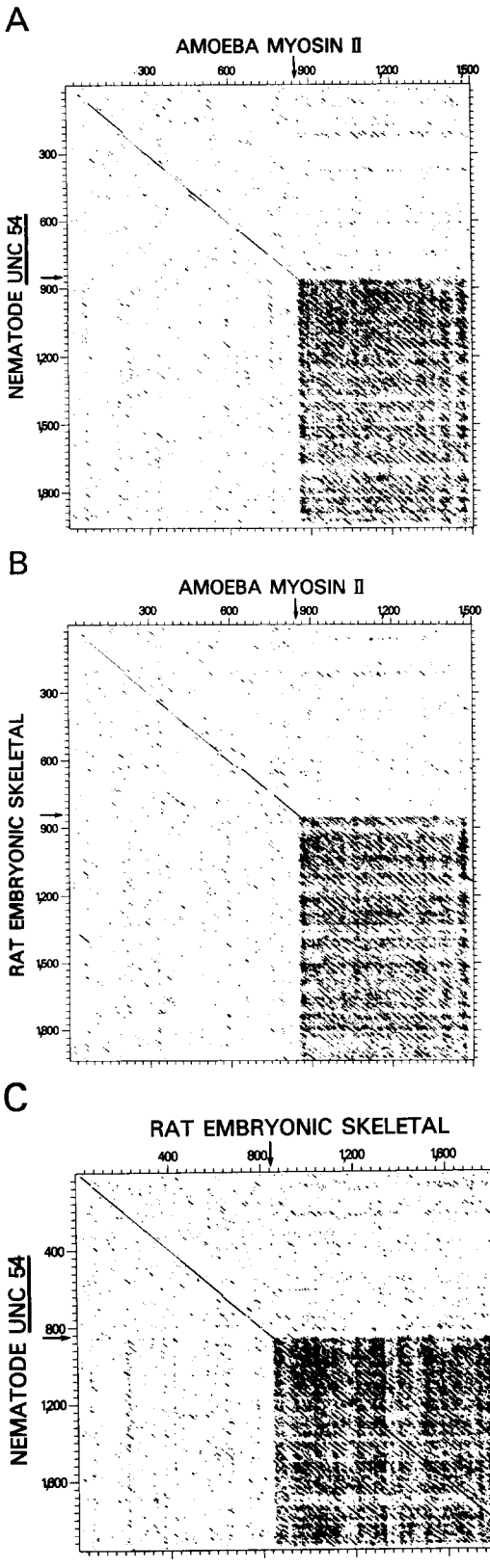


Figure 5. Comparison of myosin heavy chain amino acid sequences by dot matrix analyses. Dot matrix comparisons of the heavy chain amino acid sequences between (*A*) myosin II and the nematode *unc* 54 myosin, (*B*) myosin II and the rat embryonic skeletal muscle myosin, and (*C*) the rat embryonic skeletal muscle myosin and the nematode *unc* 54 myosin. The arrow in each sequence marks the boundary between the globular head and rodlike tail domains. For all three comparisons, the mutational index was used (counts conservative substitutions as well as exact matches), the window size was 25, and the minimum score was 20.

Amino Acid Sequence of the Globular Head Region

The boundary between the two structurally distinct portions of the myosin II molecule, the globular head region and the rodlike tail, is marked in the sequence by the proline residue at position 847. All myosin heavy chains sequenced to date (22, 40) contain a proline at this position, and in every case this proline residue marks the start of the α-helical coiled-coil rod sequence. In Fig. 4, the portion of the deduced myosin II heavy chain amino acid sequence comprising the globular head region (residues 1 through Pro 847) is aligned with the globular head sequences of the nematode *unc* 54 muscle myosin (residues 1 through Pro 850) and the rate embryonic skeletal muscle myosin (residues 1 through Pro 838). The sequences are connected by bars at exact residue matches and by colons at conservative amino acid changes. Very few adjustments in the spacing of the sequences (indicated by *dashes*) are required to produce the alignments. Overall, the myosin II head sequence shows a 43.3% exact match with the nematode myosin sequence and a 45.2% exact match with the rat embryonic myosin sequence. Taking into consideration conservative amino acid changes, the similarity is 55.1 and 56.7% with the nematode and rat sequences, respectively. The alignments with both of the muscle myosin sequences are characterized by stretches of strong sequence similarity interspersed with areas of relatively little similarity. The regions of strong similarity are dispersed across virtually the entire globular head sequence and their positions with respect to the linear sequence of the head are strongly conserved in all three myosins. This conservation of sequence topology is shown graphically by the dot matrix analyses in Fig. 5. Alignment of the myosin II heavy chain amino acid sequence with the nematode muscle myosin sequence (Fig. 5 *A*) and with the rat muscle myosin sequence (Fig. 5 *B*), as well as alignment of the two muscle myosin sequences (Fig. 5 *C*), all produce a single strong diagonal spanning virtually the entire globular head region (residues 1–~850).

The regions of strong similarity between the ameba and nematode myosin head sequences coincide in most cases with the regions of strong similarity between the ameba and rat myosin head sequences (Fig. 4). These highly conserved sequences are probably involved in the invariant functions of the myosin head, i.e., binding of nucleotide, interactions with actin and light chains, catalysis, and conformational changes linked to cross-bridge movement (19). For example, the conserved region between residues 175–200 (Fig. 4) has been implicated in formation of the nucleotide-binding site (2, 22). Similarly, the conserved region containing the reactive cysteines (Fig. 4) is thought to lie close to the catalytic site in the folded protein (42); interestingly, in the myosin II sequence sulfhydryl 1 has been replaced by an alanine residue. Other conserved sequences may be critical in the folding of the polypeptide chain, since it is very likely that the head sequences of all these myosins fold to form similar tertiary structures (3).

Several regions of very low sequence similarity in the head domain are also common amongst these myosins (see also

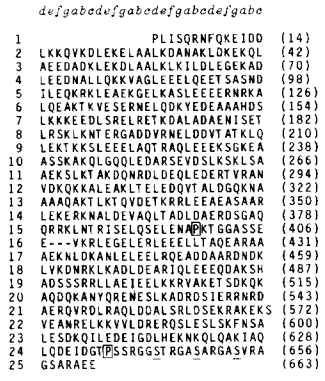


Figure 6. The myosin II rod sequence is built of repeating 28-residue zones. The 28-residue zones were initially identified by analysis of the distribution of charged residues and subsequently confirmed by Fourier transform analysis. The 28 vertical columns of sequence formed by the zones are demarcated by letters across the top in the same fashion used previously for the nematode sequence (30). Tail residues are numbered in parentheses on the right. The 28-residue zones are numbered on the left. The three spaces (*hyphens*) in zone 16 were introduced to maintain the periodicities; the possible significance of this gap will be discussed elsewhere.[3] Proline 398 (in the center of the hinge) and proline 637 (at the start of the nonhelical tailpiece) are boxed. The three phosphorylatable serines within the tailpiece (10) are underlined. The sequence of a carboxy-terminal, chymotryptic peptide of the myosin II heavy chain, determined by protein chemical methods (10), is identical to residues 598–654 of the deduced sequence above.

references 22, 39, 40). These regions include the amino-terminal ~80 residues of sequence and the characteristic sites of tryptic cleavage within the myosin head (Fig. 4; the positions of the "23 kD/50 kD" and "50 kD/20 kD" cleavage sites in the myosin II sequence were located by homology with the known cleavage sites in rabbit skeletal muscle myosin [12, 41]). Interestingly, at the position of the 23 kD/50 kD cleavage site of muscle myosin, the myosin II sequence has deletions of 11 and 9 residues relative to the nematode and rat sequences, respectively, and these deletions remove all of the characterisitic lysine residues at this site (Fig. 4). This observation agrees with the fact that the myosin II protein is not cleaved by trypsin at this position (1). On the other hand, at the position of the 50 kD/20 kD cleavage site, the myosin II sequence has insertions of 7 and 14 residues relative to the nematode and rat sequences, respectively, and is rich in basic residues (Fig. 4). This region has recently been shown to be the site of cleavage by trypsin to generate 73 and 112 kD proteolytic fragments of the myosin II molecule (1).

Amino Acid Sequence of the Rodlike Tail

The remainder of the myosin II heavy chain sequence, from proline 847 to the carboxyl terminus (residue 1509), comprises the 663-residue rodlike tail portion of the molecule (Fig. 6). This rod sequence contains ~40% fewer residues than the rod sequences of the nematode myosin (1,117 residues) and the rat embryonic myosin (1,103 residues). This ~450-residue difference accounts for virtually the entire difference in the heavy chain sizes (myosin II, 1,509 residues; nematode, 1,966 residues; rat, 1,940 residues) and in the lengths of the rodlike tails as measured by electron microscopy (myosin II, ~90 nm [25, 34]; muscle myosin, ~160 nm [29]).

The shorter myosin II rod sequence demonstrates all of the distinctive features of the α-helical coiled-coil rod sequence of muscle myosins, which were revealed previously by a detailed analysis of the nematode *unc* 54 myosin rod sequence (30). Like the nematode sequence, secondary structure calculations predict that the amino acid sequence of the myosin II rod would adopt a virtually uninterrupted α helix (with two important exceptions in the myosin II rod; see below). Like the nematode sequence, Fourier transform analysis of the myosin II rod sequence[3] indicates that it is made up of repeating 28-amino acid–residue zones (Fig. 6). A total of 25 full and partial 28-residue zones comprise the shorter myosin II tail, as compared with the 41 zones in the nematode tail sequence (30). Histograms of the amino acid distributions in the myosin II 28-residue zones (Fig. 7) reveal strong periodicities of hydrophobic and charged residues whose patterns are very similar to those observed for the nematode sequence (30). Hydrophobic residues (Fig. 7 *B*) are concentrated at alternating intervals of three and four residues in positions *a* and *d* of a "heptad" repeat (*a b c d e f g*; four heptad repeats per 28-residue zone). This heptad repeat is common to all sequences that form α-helical coiled-coils (7), where positions *a* and *d* occur at the interface between the two polypeptide chains in the coiled-coil (see Fig. 7 *A*) and hydrophobic residues in these positions serve to stabilize the association of the two chains. In terms of charged residue periodicity (Fig. 7 *C*), a typical myosin II 28-residue zone contains six alternating bands of positive and negative charge (indicated by *arrows*) that are spaced so that the strongest peak of positive charge (at *1b*) is 14 residues (or 1/2 zone) away from the strongest peak of negative charge (at *3b*). These charged residues fall primarily in positions *b*, *c*, and *f* so that the assembled rod (see Fig. 7 *A*) has strong bands of charge of alternating sign displayed across its surface at 14-residue intervals. A very similar charge distribution in the nematode 28-residue zones has been shown to dictate the specific axial stagger and electrostatic attraction of rods within the muscle myosin filament (30). These strong similarities in chemical nature between the myosin II and nematode myosin rod sequences are very apparent when the consensus sequences for the 28-residue zones in both myosins are compared (Fig. 7 *D*). The two consensus sequences (derived by taking the most common residue at each position) show a 68% exact match and an 82% match counting conservative substitutions.

Despite this similarity in chemical nature and in contrast to the strong similarity between the myosin II and nematode myosin globular head sequences, the myosin II and nematode myosin rod amino acid sequences cannot be aligned in any compelling and unique way. This fact is shown best by dot matrix analysis. In Fig. 5 (*A*), where the myosin II and nematode myosin heavy chain amino acid sequences are aligned by the dot matrix analysis, the entire area of the matrix spanning the two rod sequences is covered with short diagonals. These short stretches of contiguous sequence similarity result from the repetitive nature of the rod sequences and the similarity between these repeats in the two sequences. However, there is no long continuous diagonal appearing anywhere in this portion of the dot matrix. Therefore, when both rod sequences are compared in their entirety, no long

3. In preparation is a manuscript (McLachlan, A., and J. A. Hammer III) providing a detailed description of the myosin II tail sequence, including the Fourier transforms used to precisely define the 28-residue zones, determination of the optimal axial staggers for parallel and antiparallel interactions of the rods, and a model describing the assembly of the small myosin II bipolar filaments.

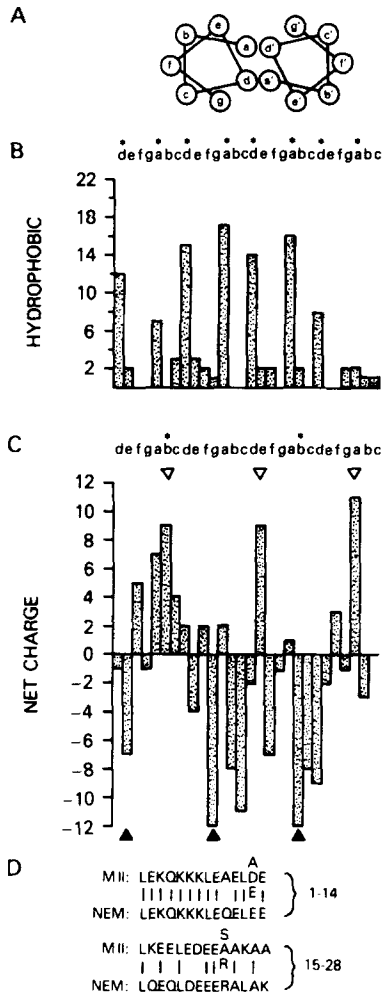


Figure 7. Histograms of the amino acid distributions in the 28-residue zones. The data is compiled from zones 2–23 (see Fig. 6; excludes zone 1, which is partial, and zones 24 and 25, which contain the nonhelical tailpiece). (*A*) A schematic cross section of the coiled-coil showing where positions a–g of the heptad repeat occur in the paired α helices. (*B*) Periodicity of hydrophobic residues. The height of the bars indicates the total number of hydrophobic residues (maximum of 22) in each of the 28 columns of sequence in Fig. 6. The core positions (*a* and *d*), where hydrophobic residues predominate, are indicated with asterisks. (*C*) Periodicity of charged residues. The heights of the bars indicate the net charge (positive minus negative) for each of the 28 columns of sequence in Fig. 6. (*Open triangles*) The major positive peaks and (*solid triangles*) negative charge peaks within the average 28-residue zone of the nematode myosin rod (30). (*D*) Comparison of the consensus sequences (the most common residue at each position) for the 28-residue zones of myosin II (*MII*) and the nematode myosin (*NEM*).

stretches of contiguous sequence similarity are observed between any portions of the two sequences. Similarly, the comparison of the myosin II rod sequence with the rat myosin rod sequence (Fig. 5 *B*) indicates that there is also no compelling and unique way to align these two rod sequences. These results contrast markedly with the control comparison of the nematode and rat muscle myosin heavy chain sequences (Fig. 5 *C*), where the matrix area spanning the two rod sequences contains, in addition to the short diagonals across the entire matrix area, a single strong diagonal spanning the entire rod sequences. This single continuous diagonal, which is contiguous with the diagonal across the globular head sequence, shows graphically the fact that the entire rod sequences of these two muscle myosins can be aligned without introducing a single spacing adjustment to yield a 45% exact match (40).

Evidence of a Hinge in the Rodlike Tail

Secondary structure calculations by the method of Garnier et al. (14) were performed on the myosin II rod amino acid sequence (data not shown). The calculations predict that the α-helix conformation is at least five times more probable than any other structure (β sheet, reverse turn, or random coil) for virtually the entire rod sequence. In only two regions of the rod sequence are conformations other than α helix preferred. One region (tail residues 631–660) occurs at the very carboxyl-terminal end of the rod. This region is characterized by a preference for random coil and reverse turn conformations and the complete absence of the hydrophobic-residue heptad repeat after the proline at tail residue position 636 (see Fig. 6). This region corresponds to the previously identified ~30-residue nonhelical tailpiece where the regulatory phosphorylation sites reside (identification based on an analysis of the amino acid sequence of a ~6-kD carboxyl-terminal chymotryptic peptide of the myosin II heavy chain [10]). The other region occurs within the body of the rod sequence between tail residues 381 and 406 (see Fig. 6). This region also possesses several features that clearly distinguish it from typical rod sequence. First, reverse turn and random coil conformations are more likely than an α-helical conformation in this region. Second, this region contains a proline residue, the classic α-helix breaker, at position 398. All other myosin rod sequences determined to date are devoid of proline residues (with the exception of the nonhelical tailpiece in the nematode *unc* 54 myosin and in myosin II). Proline breaks α helices because the dihedral angle φ is fixed to $-60° \pm 20°$ by the interaction of the ring side chain with the peptide nitrogen; no regular helix in the α-helix region ($\psi = 60°$) is possible because of steric hindrance (37). Third, this region contains a disruption in the hydrophobic heptad repeat. Specifically, the three consecutive core positions immediately following proline 398 contain polar residues (Thr 400, Ser 404, and Glu 407), which will not stabilize the chain/chain interaction like the typical hydrophobic residues (there are no other places in the entire rod sequence, excluding the tailpiece, where more than two consecutive core positions contain polar residues). We predict that in this region of the tail the α-helical coiled-coil structure is significantly weakened or even absent. Therefore, in the protein this region might represent a localized flexible point or hinge where sharp bending can occur in an otherwise uninterrupted rigid-rod structure (which begins at approximately proline 847 and ends at the nonhelical tailpiece).

We examined the shape of the myosin II tail by rotary-shadowed electron microscopy to look for evidence of the flexible point or hinge predicted by the sequence analysis. The myosin II tails were rarely perfectly straight and many were gently curved, suggesting that most parts of the tail are somewhat flexible. However, ~45% of the myosin molecules had tails containing a single sharp bend or kink (Fig. 8 *a*).

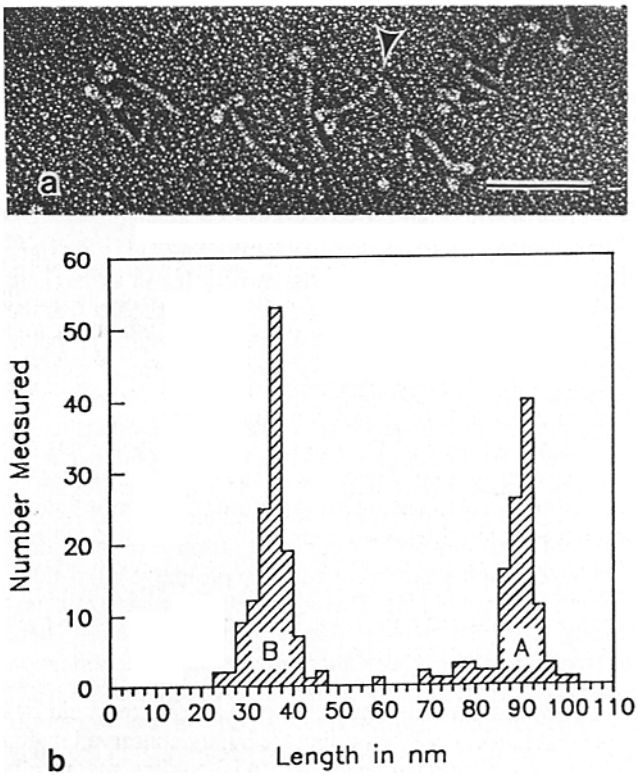


Figure 8. Rotary-shadowed electron microscopy of myosin II molecules supports the existence of a hinge in the rod. (*a*) About 45% of myosin II molecules contain a single sharp bend or kink in the rod. The arrowhead points to one such bend. Bar, 100 nm. (*b*) Histograms of the length distributions from the head/tail junction to the free end of the tail (*A*) and from the position of the bend to the free end of the tail (*B*). The mean and mode values are given in the text.

To define the position(s) of these sharp bends, we measured the total length of the tail and the distance from the free end of the tail to the sharp bend (Fig. 8 *b*). The median value for the total tail length was 88.0 (± 6.0) nm and the mode was 90.1 nm. The measured lengths from the free end of the tail to the sharp bend fell in a narrow distribution with a median value of 35.3 (± 4.0) nm and a mode of 36.4 nm. Therefore, the sharp bends occur at a single, well-defined position in the tail. When expressed as a percentage of the total tail length, this position corresponds to 40% of the way in from the free end of the tail.

How does the position of sharp bend seen in the electron microscope agree with the position of the proline-containing hinge identified in the sequence? To relate the position of the proline-containing hinge within the linear amino acid sequence to the real dimensions of the rod, we assume that the axial translation per residue is essentially constant throughout the entire tail sequence. This assumption seems reasonable since the tail sequence almost certainly forms a virtually uninterrupted coiled-coil structure with a constant helical rise per residue (30). Using proline 398 as the center of the hinge, and expressing residue 398 as a percentage of the total number of residues in the tail (663), the hinge would be predicted to fall ~40% of the way in from the free end of the tail. Because this predicted position is in very good agreement with the actual position of the sharp bend observed in the electron microscope, and because the sequence analysis does not reveal any "weak" regions other than the one around proline 398, we conclude that this latter region of sequence is the actual site of flexibility.

Discussion

The sequence of this myosin II heavy chain gene does not contain any aberrant features characteristic of nonfunctional pseudogenes such as premature stop codons, missing or improper splicing signals, insertions, or deletions (21). Furthermore, the deduced heavy chain amino acid sequence matches exactly the 64 residues of myosin II heavy chain sequence determined previously by protein chemical methods (2, 10). These facts, and especially the fact that the sequence of the genomic DNA 5′ of the ATG is identical to the sequence of the 5′ nontranslated leader of the mRNA (determined by sequencing the primer extension products), strongly support our conclusion that this gene is functional in the ameba. The remote possibility that this gene might not be functional must be considered, however, since it appears that this gene is not the only myosin II heavy chain gene in the ameba; preliminary data indicate that the ameba genome may contain two or perhaps three myosin II heavy chain genes (18).

Evolution of the Myosin II Gene

Strehler et al. (39, 40) previously reported that the positions of five out of the eight introns in the nematode *unc* 54 muscle myosin gene are conserved in the rat embryonic skeletal muscle myosin gene. They speculated that the putative ancestral myosin heavy chain gene from which vertebrate and invertebrate muscle myosin genes evolved must have contained introns in at least these five conserved positions at the time of the split between invertebrates and vertebrates (~5×10^8 yr ago). Furthermore, they hypothesized that this prevertebrate myosin gene was in fact highly split (otherwise, the large number of additional introns in present-day vertebrate myosin genes would have to be the result of later intron insertions events). They suggested that the different lineages arising from this highly split ancestral gene subsequently lost and/or added specific introns (intron loss being prominent in simpler invertebrates like the nematode), which would explain the presence of both unique intron positions and subsets of conserved intron positions in present-day myosin genes. Two lines of evidence lead us to conclude that the ameba nonmuscle myosin described here shares with the muscle myosins of vertebrates and invertebrates an ancestral heavy chain gene. First, the positions of the three introns in the myosin II gene are largely conserved relative to all characterized vertebrate and invertebrate muscle myosin genes. Second, the amino acid sequences that comprise the globular head regions of myosin II and muscle myosins show a high degree of similarity. *Acanthamoeba castellanii* is estimated to have diverged from the main line of eukaryotic descent sometime between the divergence of yeast (~1.2×10^9 yr ago) and the divergence of plants and animals (~1×10^9 yr ago) (15, 38). Therefore, we speculate that as far back as 1×10^9 yr ago the ancestral myosin gene contained introns in at least the three positions that interrupt the present-

day amoeba gene. These observations push back by at least a factor of two the previous estimate of the minimum age of these three intron positions (5×10^8 yr ago; based on the comparison between the nematode and rat genes). If, as Strehler et al. (39, 40) propose, the ancient myosin heavy chain gene was highly split, then during the course of evolution the ameba myosin II gene must have lost all but the three widely conserved introns that split the present-day gene. This would not be too surprising since protozoan genes are usually simpler than their counterparts in higher eukaryotes (15). This feature is thought to reflect the loss of introns resulting from the pressure to minimize the genome size for rapid DNA replication and cell division (the *Acanthamoeba* genome size is $\sim$ 1/100 that of humans [5]).

Structural and Evolutionary Implications of the Myosin II Rod Sequence

Comparisons between different muscle myosin heavy chain amino acid sequences indicate that rod sequences are more divergent than globular head sequences (22, 40). Nevertheless, the rod sequences of muscle myosins from invertebrates to mammals remain very similar; the alignment of the nematode *unc* 54 and rat embryonic skeletal muscle myosin rod sequences does not require a single amino acid insertion or deletion and yields 45% amino acid identity (reference 40 and Fig. 5 *C*). In sharp contrast, the myosin II rod sequence cannot be aligned in any compelling and unique way with either of these muscle myosin rod sequences (see Fig. 5, *A* and *B*). Despite this fact, the rod sequences of myosin II and muscle myosins are comprised of the same chemical motifs (i.e., the heptad repeat of hydrophobic residues and the superimposed 28-residue periodicity of charged residues) and form the same protein structure (the α-helical coiled-coil). These facts suggest that much of the rod sequence divergence between myosin II and muscle myosins, as well as between different muscle myosins, might be due to a general tolerance for amino acid changes so long as the periodicities of hydrophobic and charged residues, which dictate the structure of the rod and its associations in the myosin filament, are largely conserved.

Other factors may be responsible for the rod sequence divergence being greater between myosin II and muscle myosins than between different muscle myosins. First, this ameba myosin gene may be significantly "older" than all muscle myosin genes (see previous section) and consequently its sequence would have diverged over a longer period of time. Second, differences in filament properties between myosin II and muscle myosins probably account for some of the divergence between their rod sequences. The very small bipolar filaments formed by myosin II (and other nonmuscle myosins) contrast sharply with the large thick filaments formed by muscle myosins (19, 34). These differences presumably reflect the different assembly states in vivo of muscle versus nonmuscle myosins, i.e., muscle myosin filaments are essentially fixed in the highly ordered lattice of the sarcomere while nonmuscle myosin filaments are very likely in a dynamic state both temporally and spatially. Further analysis of the myosin II rod sequence[3] and comparison of its features with those of muscle myosins may reveal how the differences in sequence dictate the differences in filament properties.

Despite the divergence in their rod sequences, two related observations indicate that in at least one important aspect of filament structure, the axial stagger between parallel molecules, myosin II and muscle myosins are very similar if not the same. First, Pollard (34) reported that the stagger between the heads of parallel myosins within the small myosin II filament ($\sim$150 Å; measured by electron microscopy) is very similar to the stagger between the heads of parallel myosins in the muscle thick filament (143–146 Å; measured by X-ray diffraction [20]). Second, we find a striking overall similarity between myosin II and muscle myosins in the 28-residue periodicity of charged rod residues, which for muscle myosins has been shown to dictate the 143–146-Å axial stagger (30). Clearly, this repeating 28-residue structural unit is not a feature of just sarcomeric myosins.

The possibility that the myosin II rod sequence has converged with the rod sequences of muscle myosins provides an alternative explanation for the low level of rod amino acid sequence similarity between these myosins. It does seem unlikely, however, that convergent evolution would result in the myosin II rod sequence adopting the same overall pattern of charged residues and the same axial stagger of parallel molecules in the filament as do muscle myosins.

From a practical standpoint, the myosin heavy chain amino acid sequence alignments presented here would indicate that DNA probes encoding the highly conserved regions of the myosin globular head (e.g., ATP binding site, reactive thiol region) are much more suitable as heterologous probes for identifying other myosin genes than are DNA probes encoding the rod sequence. In point of fact, we (17, 18) and others (4) have successfully identified myosin heavy chain genes using heterologous DNA probes that encode portions of the myosin head.

The Hinge and Its Possible Role in Enzymatic Regulation

The myosin II rod sequence is the only completely characterized myosin rod sequence to contain the helix-breaking proline residue (with the exception of the nonhelical tailpiece in the nematode *unc* 54 gene, as well as in myosin II). It will be interesting to see whether vertebrate smooth muscle and nonmuscle myosins, which adopt the 10S monomer conformation requiring sharp bends at two constant positions within the rod (11, 43), have similar proline-containing hinges. For striated muscle myosins, a hinge region has been postulated to exist about one-third of the way down the length of the rod, based on sequence analysis, electron microscopy, protease sensitivity, and biophysical measurements (for review see reference 19). This hinge is thought to divide the muscle myosin rod into two portions, the amino-terminal subfragment 2 (S2) region and the carboxyl-terminal light meromyosin region (27, 32). The S2 region is thought to be only loosely associated with the thick filament backbone, allowing the S2/globular head domain to swing out from the thick filament surface and contact actin filaments (cross-bridges). The light meromyosin region is thought to be responsible for the rod/rod electrostatic associations that anchor the myosin in the filament backbone. An active role in force generation for the muscle myosin hinge (through a helix to coil transition in this region) has been postulated (19).

For myosin II, it has been shown in vitro that phosphoryla-

tion/dephosphorylation of the heavy chain within the nonhelical tailpiece shifts the enzyme between non-actin-activatable (phosphorylated) and actin-activatable (unphosphorylated) forms (9). The observation that phosphorylated molecules inhibit the actin-activated Mg^{2+}-ATPase activity of unphosphorylated molecules when they are present in the same filament (23) has lead to the proposal that changes in the heavy chain phosphorylation state exert their effect on enzymatic activity through intermolecular interactions within the bipolar filament, which in some way alter the conformation of the filament as a whole. We speculate that the hinge we have identified in the myosin II rod might play a key role in these intermolecular interactions. Specifically, we suggest that, much as in muscle myosins, the hinge divides the myosin II rod into S2 and light meromyosin portions. Furthermore, we speculate that the nonhelical tailpiece (whose net charge varies between +2 and −4 depending on the phosphorylation state) acts as an electrostatic catch/release mechanism for the S2/globular head domain of an adjacent molecule in the filament. In this way, active and inactive filament conformations could be generated depending on the heavy chain-phosphorylation state. The enzymatically inactive filament conformation might result from electrostatic association between the phosphorylated, negatively charged tailpiece and a patch of positive charge (of which there are numerous candidates) on the S2 region of an adjacent molecule. This could serve to hold the globular heads of the adjacent molecule down against the filament backbone. The enzymatically active conformation might result when the dephosphorylated, positively charged tailpiece releases the adjacent S2 region, allowing it (along with the attached globular heads) to swing out from the filament surface on the passive proline-containing hinge. Since the interaction of myosin II with F-actin does not appear to be significantly altered by heavy chain phosphorylation (8), the difference between these filament conformations might be in the ability of the myosin to undergo a conformational change resulting in movement of the cross-bridge. With the availability of the cloned myosin II gene, we can now use expression and site-directed mutagenesis to test directly this hypothetical model.

J. Hammer thanks Dr. G. Michaels for assistance in running the dot matrix analyses, Dr. R. Feldman for help in running the secondary structure program and analyzing the results, Mr. M. Shapiro for assistance with the DNA:DRAW program, Dr. E. Eisenberg for discussions concerning myosin structure, and Dr. C. Schmidt for advice on technical matters and many helpful discussions during the course of this work.

Received for publication 8 January 1987, and in revised form 3 April 1987.

Note Added in Proof: Warrick et al. (45) recently published the complete nucleotide sequence and deduced polypeptide sequence of a nonmuscle myosin heavy chain gene from the slime mold *Dictyostelium discoideum*. The gene does not contain introns and encodes a 2,116-residue heavy chain polypeptide (M_r 243 kD). Virtually the entire difference in size between the heavy chain of this myosin and *Acanthamoeba* myosin II is in the sizes of the rod sequences, with the *Dictyostelium* myosin rod sequence containing 1,297 residues as compared with 663 residues in the ameba myosin rod. The *Dictyostelium* myosin rod sequence is devoid of prolines and does not possess an obvious nonhelical tailpiece (45). The *Dictyostelium* myosin globular head amino acid sequence shows a high degree of similarity with the globular head sequences of muscle myosins. The amino acid sequence of the *Dictyostelium* myosin rod is poorly conserved relative to muscle myosins, but demonstrates the typical periodicities of hydrophobic and charged residues (45). Comparison of the *Dictyostelium* myosin heavy chain amino acid sequence with the *Acanthamoeba* myosin II heavy chain sequence by dot matrix analysis (using the same parameters as in Fig. 5; data not shown) reveals that (*a*) their globular head sequences are highly homologous and (*b*) the smaller myosin II rod sequence cannot be aligned in any compelling and unique way with the larger *Dictyostelium* myosin rod sequence.

References

1. Atkinson, M. A. L., and E. D. Korn. 1986. The purification and characterization of a globular subfragment of *Acanthamoeba* myosin II that is fully active when cross-linked to F-actin. *J. Biol. Chem.* 261:3382–3388.
2. Atkinson, M. A. L., E. A. Robinson, E. Appella, and E. D. Korn. 1986. Amino acid sequence of the active site of *Acanthamoeba* myosin II. *J. Biol. Chem.* 261:1844–1848.
3. Bajaj, M., and T. Blundell. 1984. Evolution and the tertiary structure of proteins. *Annu. Rev. Biophys. Bioeng.* 13:453–492.
4. Bernstein, S. I., K. Mogami, J. J. Donady, and C. P. Emerson, Jr. 1983. *Drosophila* muscle myosin heavy chain encoded by a single gene in a cluster of muscle mutations. *Nature (Lond.).* 302:393–397.
5. Bohnert, H. J., and R. J. Herrmann. 1974. The genomic complexity of *Acanthamoeba castellanii* mitochondrial DNA. *Eur. J. Biochem.* 50:83–90.
6. Buckingham, M. E. 1985. Actin and myosin multigene families: their expression during formation of skeletal muscle. *Essays Biochem.* 20:77–109.
7. Cohen, C., and D. A. D. Perry. 1986. α-Helical coiled-coils: a widespread motif in proteins. *Trends Biochem. Sci.* 11:245–248.
8. Collins, J. H., J. Kuznicki, B. Bowers, and E. D. Korn. 1982. Comparison of the actin binding and filament formation properties of phosphorylated and dephosphorylated *Acanthamoeba* myosin II. *Biochemistry.* 21:6910–6915.
9. Cote, G. P., J. H. Collins, and E. D. Korn. 1981. Identification of three phosphorylation sites on each heavy chain of *Acanthamoeba* myosin II. *J. Biol. Chem.* 256:12811–12816.
10. Cote, G. P., E. A. Robinson, E. Appella, and E. D. Korn. 1984. Amino acid sequence of a segment of the *Acanthamoeba* myosin II heavy chain containing all three regulatory phosphorylation sites. *J. Biol. Chem.* 259: 12781–12787.
11. Craig, R., R. C. Smith, and J. Kendrick-Jones. 1983. Light-chain phosphorylation controls the conformation of vertebrate non-muscle and smooth muscle myosin molecules. *Nature (Lond.).* 302:436–439.
12. Elzinga, M., and B. Trus. 1980. Sequence and proposed structure of a 17 kDa fragment of myosin. *In* Methods in Peptide and Protein Sequence Analysis. C. Birr, editor. Elsevier/North Holland, Amsterdam. 213–224.
13. Fitzgerald, M., and T. Shenk. 1981. The sequence AAUAAA forms part of the recognition site for polyadenylation of SV40 late RNAs. *Cell.* 24:251–260.
14. Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120:97–120.
15. Gilbert, W., M. Marchionni, and G. McKnight. 1986. On the antiquity of introns. *Cell.* 46:151–154.
16. Gulick, J., K. Kropp, and J. Robbins. 1985. The structure of two fast-white myosin heavy chain promoters: a comparative study. *J. Biol. Chem.* 260: 14513–14520.
17. Hammer, J. A., III, G. Jung, and E. D. Korn. 1986. Genetic evidence that *Acanthamoeba* myosin I is a true myosin. *Proc. Natl. Acad. Sci. USA.* 83:4655–4659.
18. Hammer, J. A., III, E. D. Korn, and B. M. Paterson. 1986. Isolation of a non-muscle myosin heavy chain gene from *Acanthamoeba. J. Biol. Chem.* 261:1949–1956.
19. Harrington, W. F., and M. E. Rodgers. 1984. Myosin. *Annu. Rev. Biochem.* 53:35–73.
20. Huxley, H. E., and W. J. Brown. 1967. A low-angle X-ray diagram of vertebrate striated muscle and its behavior during contraction and rigor. *J. Mol. Biol.* 30:383–434.
21. Karin, M., and R. I. Richards. 1982. Human metallothionein genes: primary structure of the metallothionein-II gene and a related processed gene. *Nature (Lond.).* 299:797–802.
22. Karn, J., S. Brenner, and L. Barnett. 1983. Protein structural domains in *Caenorhabditis elegans unc*-54 myosin heavy chain gene are not separated by introns. *Proc. Natl. Acad. Sci. USA.* 80:4253–4257.
23. Kuznicki, J., J. P. Albanesi, G. P. Cote, and E. D. Korn. 1983. Supramolecular regulation of the actin-activated ATPase activity of filaments of *Acanthamoeba* myosin II. *J. Biol. Chem.* 258:6011–6014.
24. Kuznicki, J., M. A. L. Atkinson, and E. D. Korn. 1984. Effects of limited tryptic cleavage on the physical and enzymatic properties of myosin II from *Acanthamoeba castellanii. J. Biol. Chem.* 259:308–9313.
25. Kuznicki, J., G. P. Cote, B. Bowers, and E. D. Korn. 1985. Filament formation and actin-activated ATPase activity are abolished by proteolytic removal of a small peptide from the tip of the tail of the heavy chain of *Acanthamoeba* myosin II. *J. Biol. Chem.* 260:1967–1972.
26. Kuznicki, J., and E. D. Korn. 1984. Interdependence of factors affecting

the actin-activated ATPase activity of myosin II from *Acanthamoeba castellanii*. *J. Biol. Chem.* 259:9302–9307.
27. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. 76–85.
28. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific cleavage reactions. *Methods Enzymol.* 65:499–560.
29. McLachlan, A. D. 1984. Structural implications of the myosin amino acid sequence. *Annu. Rev. Biophys. Bioeng.* 13:167–189.
30. McLachlan, A. D., and J. Karn. 1982. Periodic charge distributions in the myosin rod amino acid sequence match cross-bridge spacings in muscle. *Nature (Lond.).* 299:226–231.
31. Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* 101:28–78.
32. Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10:459–472.
33. Nellen, W., and D. Gallwitz. 1982. Actin genes and actin messenger RNA in *Acanthamoeba castellanii*: nucleotide sequence of the split actin gene I. *J. Mol. Biol.* 159:1–18.
34. Pollard, T. D. 1982. Structure and polymerization of *Acanthamoeba* myosin II filaments. *J. Cell Biol.* 95:816–825.
35. Pollard, T. D., W. F. Stafford, III, and M. E. Porter. 1978. Characterization of a second myosin from *Acanthamoeba castellanii*. *J. Biol. Chem.* 253:4798–4808.
36. Sanger, F. S., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.
37. Schulz, G. E., and R. H. Schirmer. 1979. Patterns of folding and association of polypeptide chains. *In* Principles of Protein Structure. C. R. Cantor, editor. Springer-Verlag. New York. 79–83.
38. Sogin, M. L., H. J. Elwood, J. H. Gunderson. 1986. Evolutionary diversity of eukaryotic small subunit rRNA genes. *Proc. Natl. Acad. Sci. USA.* 83:1383–1387.
39. Strehler, E. E., V. Mahdavi, M. Periasamy, and B. Nadal-Ginard. 1985. Intron positions are conserved in the 5′ end region of myosin heavy chain genes. *J. Biol. Chem.* 260:468–471.
40. Strehler, E. E., M. Strehler-Page, J. Perriard, M. Periasamy, and B. Nadal-Ginard. 1986. Complete nucleotide sequence and encoded amino acid sequence of a mammalian myosin heavy chain gene: evidence against intron-dependent evolution of the rod. *J. Mol. Biol.* 190:291–317.
41. Tong, S. W., and M. Elzinga. 1983. The sequence of the NH_2-terminal 204-residue fragment of the heavy chain of rabbit skeletal muscle myosin. *J. Biol. Chem.* 258:13100–13110.
42. Tregear, R. T. 1986. Crossbridges, force and motion. *Nature (Lond.).* 321:563.
43. Trybus, K. M., T. W. Huiatt, and S. Lowey. 1982. A bent monomeric conformation of myosin from smooth muscle. *Proc. Natl. Acad. Sci. USA.* 79:6151–6155.
44. Tyler, J. M., and D. Branton. 1980. Rotary shadowing of extended molecules dried from glycerol. *J. Ultrastruc. Res.* 71:95–102.
45. Warrick, H. M., A. Dehozanne, L. A. Leinwand, and J. A. Spudich. 1986. Conserved protein domains in a myosin heavy chain gene from *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. USA.* 83:9433–9437.