The Sequence of an Embryonic Myosin Heavy Chain Gene and Isolation of Its Corresponding cDNA*

(Received for publication, December 12, 1986)

Marcia I. Molina, Keith E. Kropp‡, James Gulick, and Jeffrey Robbins§

From the Department of Pharmacology and Cell Biophysics, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267-0575

The complete sequence of an embryonic chicken myosin heavy chain has been determined. Introns and exons were identified by comparison with the corresponding cDNA. The cDNA contains 5,962 bases, of which 85 bases constitute the poly(A) tail. The cDNA represents the entire mRNA transcript, except for 90 bases at the 5'-coding terminus and 101 bases of the 5'-untranslated region. The gene's coding region is split by 37 introns; two additional introns split the 101 base pairs which make up the 5'-untranslated region. The complete gene is approximately 23,000 base pairs and encodes a protein whose molecular weight is 222,559 and consists of 1,940 amino acids. Analysis of the protein and comparison with other myosin sequences reveal that certain regions have been conserved; those amino acids which have been postulated to participate in the ATPase and actin-binding activities of the molecule are highly homologous. These comparisons have allowed the identification of isolated regions within the myosin heavy chain that appear to be essential for the molecule's function.

By mass, the myosin heavy chain is the major protein found in the contractile apparatus of the cell and together with associated proteins forms the characteristic thick filament (1-3). In the sarcomere, myosin exists as a complex of proteins, consisting of two heavy chains (Mr, ~220,000) and two pairs of light chains of two types (Mr, ~17,000-20,000) (4). It is within the heavy chain that the actin-activated ATPase resides, and it is this enzymatic activity which underlies the development of mechanical force in the muscle fiber.

The myosin heavy chain must interact with a number of other proteins found in the contractile apparatus. The characterization of the different sequences which participate directly in these interactions have been subject to a number of investigations, the ultimate aim being to elucidate the structure-function relationship of the protein. These studies have shown that the myosin heavy chain is a rather unusual, highly asymmetric molecule, consisting of a globular "head" region at the NH₂ terminus and a fibrous, α-helical domain which makes up the "rod" at the COOH terminus. The ATPase is localized within the head region, as are the sites with which the light chains interact (5, 6). On the basis of digestion with a variety of enzymes, the myosin molecule has been subdivided further into a number of regions which are thought to correspond to structural domains. Treatment with chymotrypsin results in cleavage of the protein between the globular and fibrous domains (7). The resulting globular fragment, termed S-1, contains some 800 amino acids and shows both ATPase activity and the ability to bind actin (8). Limited digestion of this fragment with trypsin results in the production of three polypeptides whose molecular weights, in order from the amino terminus, have been estimated to be 27,000, 50,000, and 20,000 (9, 10); each fragment has been assigned various roles in the molecule's activities. The hypothesis that these three fragments are representative of certain domains has been advanced since cleavage of S-1 with a number of other enzymes also results in the production of these or closely related fragments (11, 12).

Other portions of the myosin molecule have been defined in a similar manner; by controlled proteolysis, the connecting segment between the globular S-1 region and the helical rod portion has been isolated and termed S-2, or "hinge region" (13-15). This fragment, which is highly susceptible to proteolysis (16), has been postulated to be the functional link between the coiled-coil rod structure and the globular, ATPase-containing head region. It has been postulated that the S-2 hinge permits the movement of the head along the actin, relative to the thick filament (17, 18).

The rod region, in contrast, is thought to play a relatively static, structural role in the thick filament. This portion of the molecule, which also can be isolated by controlled proteolysis, is often termed light meromyosin (LMM) and is insoluble at physiological ionic strengths (19, 20). The unique properties of LMM in terms of the charge periodicity needed to maintain its conformation and its interactions with the other myosin rods packed into the thick filament have been long appreciated (21, 22). The existence in this region of a 28-residue repeat has been confirmed by extensive sequencing (23-25).

Although enzymatic dissections have been invaluable in localizing some of the functional properties of the molecule, the resolution of these methodologies is limited. Indeed, a prerequisite for exact assignment of structure to function is the determination of the protein's sequence. Previously, we have cloned and characterized members of the chicken myosin heavy chain gene family (26, 27) and have isolated recombinant bacteriophage which, taken together, encompass a com-

* This work was supported by National Institutes of Health Grant HD 21893, United States Department of Agriculture Grant CRCR-1-729, and a grant from the Missouri Affiliate of the American Heart Association. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Supported by National Institutes of Health Training Grant HL07382.
§ Established Investigator of the American Heart Association. To whom correspondence should be addressed.

1 The abbreviations used are: S-1, subfragment 1; S-2, subfragment 2; LMM, light meromyosin.
plete myosin heavy chain gene and its promoter elements (28). In addition to the structural considerations at the protein level outlined above, we are also interested in defining the gene's structural basis for its regulation; and we have noted that some of these elements may be localized within the interior of the gene (29, 30). Therefore, we have sequenced the entire embryonic myosin heavy chain gene. In addition, we have prepared an essentially full-length cDNA which lacks only 90 base pairs of the coding region and have used this construct to confirm directly the intron-exon junctions of the gene and the protein's sequence.

**EXPERIMENTAL PROCEDURES**

**Sequencing Strategy**—The bacteriophage containing the fragments of the embryonic myosin gene have been described previously (28). Suitable restriction fragments of between 300 and 3500 base pairs were generated and isolated using low-melt agarose. These fragments were cloned into M13mp18 or M13mp19 (31, 32), and both strands of the entire fragment were sequenced directly by generating a series of nested deletions as described by Dale et al. (33). All other sequencing conditions have been described (28).

**Synthesis of Complementary DNA**—Messenger RNA was isolated from 14-day embryonic chick breast muscle as described previously (28). First-strand synthesis was carried out using 5 μg of the RNA in a volume of 50 μl, essentially as described by Okayama and Berg (34). Reverse transcriptase XL (Life Science Inc., St. Petersburg, FL) was found to give superior results in 30-min incubations at 42°C. A parallel reaction, containing 7 μg of RNA but using Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories), was carried out at 37°C for 1 h in 50 μl containing 60 mm KCl, 50 mm Tris-HCl (pH 7.5), 3 mm MgCl₂, 0.5 mm dNTPs, 5 mm dithiothreitol, and 1 μg of oligo(dT)₁₄₋₁₈. The reactions were mixed, quenched by the addition of EDTA to 20 mm, extracted with phenol/chloroform, and precipitated by the addition of an equal volume of 4 M ammonium acetate and 2 volumes of ethanol. The tube was placed at -70°C for 10 min, and the cDNA was collected by centrifugation. Second-strand synthesis was carried out in 250 μl essentially as described by Gubler and Hoffman (35).

**RESULTS AND DISCUSSION**

**Gene's Organization**

As shown in Fig. 1, the gene encompasses approximately 23,000 bases and consists of 40 exons, varying in size from 24 to 399 nucleotides. In contrast to the rather tight constraint on exon sizes, the introns vary widely, the smallest being 74 nucleotides which are translated into 1,940 amino acids to produce a protein (neglecting glycosylation) of M₀ = 222,559. The distal half of the last exon contains the nucleotides which encode the 3'-untranslated region of 146 bases; this exon also contains the polyadenylation signal. The ratio of intron nucleotides to exon nucleotides is approximately 3:1 which, when compared to other eucaryotic genes, is not unusual. In addition, when compared with the other vertebrate myosin genes (25, 37, 50), the organization of the gene is highly conserved. This conservation is maintained across both species (25) and isoforms (60) and indicates that a primordial myosin gene gave rise by a series of duplications to a number of different gene isoforms.

**Determination of Intron-Exon Junctions**

Protein sequence information for the vertebrate myosins, which could be used to confirm intron-exon junctions, is limited. Although canonical splice junctions and the probable introns vary widely, the smallest being 74 nucleotides which are translated into 1,940 amino acids to produce a protein (neglecting glycosylation) of M₀ = 222,559. The distal half of the last exon contains the nucleotides which encode the 3'-untranslated region of 146 bases; this exon also contains the polyadenylation signal. The ratio of intron nucleotides to exon nucleotides is approximately 3:1 which, when compared to other eucaryotic genes, is not unusual. In addition, when compared with the other vertebrate myosin genes (25, 37, 50), the organization of the gene is highly conserved. This conservation is maintained across both species (25) and isoforms (60) and indicates that a primordial myosin gene gave rise by a series of duplications to a number of different gene isoforms.

**TABLE I**

| Exon | Intron |
|------|--------|
| Residues | Size | Residues | Size |
| 1 | 1-24 | 24 | 25-263 | 239 |
| 2 | 264-306 | 42 | 307-2,273 | 1,967 |
| 3 | 2,274-2,611 | 238 | 2,512-2,585 | 74 |
| 4 | 2,596-2,739 | 144 | 2,740-2,836 | 97 |
| 5 | 2,837-2,993 | 157 | 2,994-3,566 | 573 |
| 6 | 3,567-3,594 | 28 | 3,595-3,691 | 97 |
| 7 | 3,692-3,806 | 115 | 3,807-3,880 | 74 |
| 8 | 3,881-3,897 | 95 | 3,897-4,092 | 119 |
| 9 | 4,093-4,156 | 64 | 4,157-4,242 | 86 |
| 10 | 4,243-4,341 | 99 | 4,342-4,441 | 100 |
| 11 | 4,442-4,484 | 44 | 4,546-4,636 | 91 |
| 12 | 4,637-4,775 | 139 | 4,776-4,787 | 100 |
| 13 | 4,792-4,957 | 119 | 4,898-4,928 | 100 |
| 14 | 6,988-7,647 | 658 | 7,648-7,794 | 947 |
| 15 | 7,755-7,965 | 201 | 7,967-8,164 | 682 |
| 16 | 8,308-8,357 | 59 | 8,358-8,377 | 59 |
| 17 | 9,973-9,992 | 74 | 9,993-10,093 | 281 |
| 18 | 10,094-10,181 | 88 | 10,182-10,265 | 84 |
| 19 | 10,266-10,383 | 118 | 10,384-10,688 | 315 |
| 20 | 10,699-10,822 | 124 | 10,823-10,908 | 86 |
| 21 | 10,909-11,045 | 137 | 11,046-11,802 | 757 |
| 22 | 11,030-12,058 | 256 | 12,047-12,932 | 874 |
| 23 | 12,933-13,175 | 243 | 13,176-13,289 | 106 |
| 24 | 13,281-13,457 | 177 | 13,458-13,641 | 84 |
| 25 | 13,542-13,687 | 146 | 13,668-13,773 | 86 |
| 26 | 13,774-13,864 | 91 | 13,865-13,949 | 85 |
| 27 | 13,850-14,339 | 390 | 14,340-14,969 | 650 |
| 28 | 14,970-15,086 | 127 | 15,097-15,816 | 720 |
| 29 | 15,817-15,935 | 119 | 15,936-16,255 | 320 |
| 30 | 16,256-16,452 | 197 | 16,453-16,783 | 421 |
| 31 | 16,874-17,057 | 184 | 17,058-17,135 | 78 |
| 32 | 17,136-17,301 | 166 | 17,302-17,458 | 157 |
| 33 | 17,459-17,583 | 125 | 17,584-17,786 | 203 |
| 34 | 17,727-18,085 | 399 | 18,096-18,192 | 97 |
| 35 | 18,193-18,386 | 206 | 18,387-19,198 | 802 |
| 36 | 19,199-19,324 | 126 | 19,325-19,411 | 87 |
| 37 | 19,412-19,582 | 171 | 19,583-19,668 | 86 |
| 38 | 19,669-19,773 | 105 | 19,774-20,025 | 252 |
| 39 | 20,028-20,121 | 93 | 20,122-22,381 | 2,270 |
| 40 | 22,392-22,687 | 225 |

**FIG. 2. Myosin cDNA.** The 38 exons included in the transcript are shown. The 3' untranslated region is indicated.■.

**FIG. 1. Organization of myosin heavy chain gene.** The locations of the exons (■) are shown. kb, kilobase.
intron-exon organization can be determined by computer analysis, unambiguous assignation of the boundaries requires comparison of the gene sequence to the corresponding mRNA (or cDNA). Alternatively, primer extension analysis of each putative junction can be done. A comparison with the contiguous nucleotides is, however, the most direct approach. To this end, an essentially full-length cDNA was constructed, sequenced, and used to unambiguously confirm the intron-exon organization. The organization of this cDNA is shown in Fig. 2, and the exons it contains are indicated. All exons that are translated are contained within the transcript, the cDNA terminating only 90 bases away from the translational initiation site. The sizes of the exons and introns and their locations are shown in Table I. The globular S-1 region, which makes up approximately 40% of the protein, is encoded by approximately 19 exons; the S-2 region is encoded by approximately seven; and the rod, which contains 33% of the protein's amino acids, is encoded by approximately 11. (None of the domains' boundaries correspond exactly to exon junctions, hence the approximations.)

FIG. 3. Nucleotide sequence
Fig. 3—continued
The entire sequence of the gene is shown in Fig. 3; the nucleotides corresponding to the exons are capitalized. The entire gene, from the transcriptional start site to the site of polyadenylation, consists of 22,684 nucleotides. As previously noted (28), the basal structural elements of a eucaryotic promoter, the CAAT and TATA boxes, are present in their usual positions upstream from the base (+1) at which transcription is initiated.

### Nucleotide Sequence

The sequence of the protein was determined directly from the gene and cDNA sequences. We noted that only three nucleotide changes had occurred between the gene and cDNA. One of these (amino acid 379) shows a GGT → GAT change which results in a glycine → aspartate change. At amino acid residues 1212 (ATT → ATC, isoleucine) and 1824 (GTG → GTA, valine), silent site substitutions have taken place, and the amino acids have been conserved. These changes merely
reflect the extensive polymorphisms present in the myosins which we have previously observed (61) and show the polymorphic variations in different Leghorn flocks. The amino acid sequence deduced from the gene's nucleotide sequence is shown in Fig. 4. The number and percentage in the protein of each amino acid are shown in Table I.

Subfragment 1—As noted above, controlled proteolysis of the intact myosin produces well-defined fragments, and one of these, S-1, corresponds to the globular head region. On the basis of available information in the literature (9, 10, 24, 25) to generation of this fragment at lysine 846, although other investigators (24, 25) have indicated that cleavage might occur at lysine 836 or 842. The resulting fragment has a molecular weight of 96,240 and can be subdivided into three discrete regions whose conformations display a high probability of containing a turn or having a relatively unordered structure (41). Reaction of one of the lysines with 2,4,6-trinitrobenzenesulfate (42, 43) causes a dramatic decrease in the ATPase activity. Previous sequence information (44) placed this residue in the peptide fragment aspartate-proline-proline-lysine, which corresponds exactly to residues 81–84 in the embryonic sequence.

The 60-kDa peptide which encompasses amino acids 214–640 (M, = 47,962) is adjacent to the 23-kDa amino-terminal fragment and is also able to interact with ATP analogues. Korner et al. (46) showed that modification of a carboxyl group in this fragment decreased the ATPase activity. More recently, Mahmood and Yount (47) demonstrated that a trap-group in this fragment decreased the ATPase activity. The positive charge can ionically attract the polyphosphate moiety of the ATP (45). This NH2-terminal fragment is known to contain many other reactive lysine residues, and it is noteworthy that lysines 55, 84, 130, 146, 147, and 185 all occur in or around regions whose conformations display a high probability of containing a turn or having a relatively unordered structure (41). Reaction of one of the lysines with 2,4,6-trinitrobenzenesulfate (42, 43) causes a dramatic decrease in the ATPase activity. Previous sequence information (44) placed this residue in the peptide fragment aspartate-proline-proline-lysine, which corresponds exactly to residues 81–84 in the embryonic sequence.

The 60-kDa peptide which encompasses amino acids 214–640 (M, = 47,962) is adjacent to the 23-kDa amino-terminal fragment and is also able to interact with ATP analogues. Korner et al. (46) showed that modification of a carboxyl group in this fragment decreased the ATPase activity. More recently, Mahmood and Yount (47) demonstrated that a trap-group in this fragment decreased the ATPase activity. The positive charge can ionically attract the polyphosphate moiety of the ATP (45). This NH2-terminal fragment is known to contain many other reactive lysine residues, and it is noteworthy that lysines 55, 84, 130, 146, 147, and 185 all occur in or around regions whose conformations display a high probability of containing a turn or having a relatively unordered structure (41). Reaction of one of the lysines with 2,4,6-trinitrobenzenesulfate (42, 43) causes a dramatic decrease in the ATPase activity. Previous sequence information (44) placed this residue in the peptide fragment aspartate-proline-proline-lysine, which corresponds exactly to residues 81–84 in the embryonic sequence.

The 60-kDa peptide which encompasses amino acids 214–640 (M, = 47,962) is adjacent to the 23-kDa amino-terminal fragment and is also able to interact with ATP analogues. Korner et al. (46) showed that modification of a carboxyl group in this fragment decreased the ATPase activity. More recently, Mahmood and Yount (47) demonstrated that a trap-group in this fragment decreased the ATPase activity. The positive charge can ionically attract the polyphosphate moiety of the ATP (45). This NH2-terminal fragment is known to contain many other reactive lysine residues, and it is noteworthy that lysines 55, 84, 130, 146, 147, and 185 all occur in or around regions whose conformations display a high probability of containing a turn or having a relatively unordered structure (41). Reaction of one of the lysines with 2,4,6-trinitrobenzenesulfate (42, 43) causes a dramatic decrease in the ATPase activity. Previous sequence information (44) placed this residue in the peptide fragment aspartate-proline-proline-lysine, which corresponds exactly to residues 81–84 in the embryonic sequence.

The 60-kDa peptide which encompasses amino acids 214–640 (M, = 47,962) is adjacent to the 23-kDa amino-terminal fragment and is also able to interact with ATP analogues. Korner et al. (46) showed that modification of a carboxyl group in this fragment decreased the ATPase activity. More recently, Mahmood and Yount (47) demonstrated that a trap-group in this fragment decreased the ATPase activity. The positive charge can ionically attract the polyphosphate moiety of the ATP (45). This NH2-terminal fragment is known to contain many other reactive lysine residues, and it is noteworthy that lysines 55, 84, 130, 146, 147, and 185 all occur in or around regions whose conformations display a high probability of containing a turn or having a relatively unordered structure (41). Reaction of one of the lysines with 2,4,6-trinitrobenzenesulfate (42, 43) causes a dramatic decrease in the ATPase activity. Previous sequence information (44) placed this residue in the peptide fragment aspartate-proline-proline-lysine, which corresponds exactly to residues 81–84 in the embryonic sequence.

The 60-kDa peptide which encompasses amino acids 214–640 (M, = 47,962) is adjacent to the 23-kDa amino-terminal fragment and is also able to interact with ATP analogues. Korner et al. (46) showed that modification of a carboxyl group in this fragment decreased the ATPase activity. More recently, Mahmood and Yount (47) demonstrated that a trap-group in this fragment decreased the ATPase activity. The positive charge can ionically attract the polyphosphate moiety of the ATP (45). This NH2-terminal fragment is known to contain many other reactive lysine residues, and it is noteworthy that lysines 55, 84, 130, 146, 147, and 185 all occur in or around regions whose conformations display a high probability of containing a turn or having a relatively unordered structure (41). Reaction of one of the lysines with 2,4,6-trinitrobenzenesulfate (42, 43) causes a dramatic decrease in the ATPase activity. Previous sequence information (44) placed this residue in the peptide fragment aspartate-proline-proline-lysine, which corresponds exactly to residues 81–84 in the embryonic sequence.

The 60-kDa peptide which encompasses amino acids 214–640 (M, = 47,962) is adjacent to the 23-kDa amino-terminal fragment and is also able to interact with ATP analogues. Korner et al. (46) showed that modification of a carboxyl group in this fragment decreased the ATPase activity. More recently, Mahmood and Yount (47) demonstrated that a trap-group in this fragment decreased the ATPase activity. The positive charge can ionically attract the polyphosphate moiety of the ATP (45). This NH2-terminal fragment is known to contain many other reactive lysine residues, and it is noteworthy that lysines 55, 84, 130, 146, 147, and 185 all occur in or around regions whose conformations display a high probability of containing a turn or having a relatively unordered structure (41). Reaction of one of the lysines with 2,4,6-trinitrobenzenesulfate (42, 43) causes a dramatic decrease in the ATPase activity. Previous sequence information (44) placed this residue in the peptide fragment aspartate-proline-proline-lysine, which corresponds exactly to residues 81–84 in the embryonic sequence.

The 60-kDa peptide which encompasses amino acids 214–640 (M, = 47,962) is adjacent to the 23-kDa amino-terminal fragment and is also able to interact with ATP analogues. Korner et al. (46) showed that modification of a carboxyl group in this fragment decreased the ATPase activity. More recently, Mahmood and Yount (47) demonstrated that a trap-group in this fragment decreased the ATPase activity. The positive charge can ionically attract the polyphosphate moiety of the ATP (45). This NH2-terminal fragment is known to contain many other reactive lysine residues, and it is noteworthy that lysines 55, 84, 130, 146, 147, and 185 all occur in or around regions whose conformations display a high probability of containing a turn or having a relatively unordered structure (41). Reaction of one of the lysines with 2,4,6-trinitrobenzenesulfate (42, 43) causes a dramatic decrease in the ATPase activity. Previous sequence information (44) placed this residue in the peptide fragment aspartate-proline-proline-lysine, which corresponds exactly to residues 81–84 in the embryonic sequence.
S-1, encompasses amino acids 641–846 and has an \( M_r = 23,967 \) (Fig. 4). It can interact with the other domains found in the head region and has also been implicated in actin binding (48–50). In addition, this peptide fragment contains two reactive thiol groups, termed SH1 and SH2 (9), corresponding to cysteines 700 and 710 in Fig. 4. Their modification significantly affects ATPase activity. SH2, which is located in a flexible part of the polypeptide, not only can interact with the SH1 site but can be cross-linked to a reactive thiol in the 50-kDa fragment (residue 403, 480, 523, or 541). This cross-linking induces the stable trapping of Mg\(^{2+}\)-ADP (52). These apparent interactions between amino acids which are far apart in the primary sequence serve to emphasize the folding which places these domains in proximity to one another and the globular nature of the entire region in the native protein.

**Subfragment 2**—The residues located between lysine 846 and arginine 1284 make up the central part of the myosin molecule which is designated as subfragment 2. S-2 is the structural head link between the head region and the rigid LMM rod. S-2 can be further divided, on the basis of trypsin digestion, into two fragments: the “short” S-2, located at the NH\(_2\) terminus, and the more flexible hinge comprising the COOH terminus. The sequence of this portion of the molecule predicts the high degree of \( \alpha \)-helix which has been noted previously in the myosins (25, 51). The 7- and 28-residue repeats are maintained as are the positions of the “skip residues,” which have the effect of widening the helix and altering the degree of twist (57). A hydrophilicity profile of the molecule shown in Fig. 6 emphasizes the dichotomy of the two ends of the protein. The globular characteristics of S-1 are graphically reflected in the mixed hydrophobic and hydrophilic natures of the region. A deep hydrophobic pocket is present proximal to photoreactive tryptophan 131, and the COOH terminus of the 20-kDa fragment ends in the “loop,” the most hydrophilic feature in the entire profile. This loop has been observed in previously sequenced myosins (51). In contrast to S-1, the other regions of the protein, S-2 and LMM, show minimal hydrophobicity, this being a reflection of their essentially rod-like natures.

**Myosin Heavy Chain Gene**

![Hydrophobic and basic amino acid distributions in three regions of myosin heavy chain](image)

**TABLE II**

| Amino acid | % | No. |
|-----------|---|-----|
| Asp       | 5 | 101 |
| Glu       | 13| 259 |
| Arg       | 5 | 98  |
| His       | 2 | 38  |
| Lys       | 11| 217 |
| Asn       | 4 | 74  |
| Gln       | 7 | 130 |
| Cys       | 1 | 17  |
| Ser       | 5 | 106 |
| Thr       | 5 | 90  |
| Tyr       | 2 | 42  |
| Gly       | 4 | 74  |
| Ala       | 9 | 170 |
| Leu       | 10| 196 |
| Ile       | 5 | 96  |
| Pro       | 2 | 32  |
| Val       | 4 | 83  |
| Phe       | 3 | 56  |
| Trp       | 0 | 9   |
| Met       | 3 | 52  |

The exact nature of the hinge is unknown. It has been postulated that the flex inherent in the \( \alpha \)-helical coil is sufficient to allow for the movement of the head relative to the rod (51). However, other investigators (54) have noted that the region may undergo melting and have hypothesized that this results in a localized conformational change.

A major contribution to the overall stability of the region is made by the interactions between the two \( \alpha \)-helices. The coiled-coil was first considered in detail by Crick (55) who postulated that the amino acids would have a 7-residue repeat (a-g). Apolar residues in positions a and d would form a “stripe” inclined around the axis of the helix stabilizing the packing. Charged groups would be located preferentially on the surface. The model has been shown to be correct for a large number of proteins (56), and consideration of the implications for myosin has been detailed (25, 51, 57). Fig. 5 shows a residue analysis for the short S-2 and the flexible hinge region, as well as for the highly ordered “rigid” LMM. The data show that the hydrophobic, stabilizing interactions at position a are significantly reduced in the hinge section; this may lead to an overall decrease in the thermal stability of this portion of the molecule. Based on a cautious application of the available algorithms for predicting conformation (43, 53), we think a plausible structural interpretation of the hinge might be the formation of a highly localized region of disorder within residues 1285–1315, which are unable to form any sort of an ordered structure. The resulting conformational changes could then be propagated over short distances, the process being facilitated by the lower degree of stability of this region relative to the short S-2 and LMM.

**Light Meromyosin**—We identify residue 1284 (arginine) as the probable trypsin site which defines LMM. The sequence of this portion of the molecule predicts the high degree of \( \alpha \)-helix which has been noted previously in the myosins (25, 51). The 7- and 28-residue repeats are maintained as are the positions of the “skip residues,” which have the effect of widening the helix and altering the degree of twist (57). A hydrophilicity profile of the molecule shown in Fig. 6 emphasizes the dichotomy of the two ends of the protein. The globular characteristics of S-1 are graphically reflected in the mixed hydrophobic and hydrophilic natures of the region. A deep hydrophobic pocket is present proximal to photoreactive tryptophan 131, and the COOH terminus of the 20-kDa fragment ends in the “loop,” the most hydrophilic feature in the entire profile. This loop has been observed in previously sequenced myosins (51). In contrast to S-1, the other regions of the protein, S-2 and LMM, show minimal hydrophobicity, this being a reflection of their essentially rod-like natures.
Comparison with Other Myosins: Conservation of Basic Gene Structure

Both the nematode and rat myosin sequences have been determined (24, 25). The nematode gene, unc54, encodes a body-wall myosin which contains 1966 amino acids, whereas the rat gene encodes a myosin which is expressed predominantly in embryonic fast-white muscles. The embryonic chicken myosin, as might be expected, is closely related to both of these sequences, but is more homologous to the vertebrate sequence and contains exactly the same number (1940) of amino acids. We compared the amino acids from both the nematode and the rat to the chicken sequence; the overall homology was 48.8 and 83.5%, respectively. Although the exons in the chicken do not appear to define any obvious domains (e.g., no exon junctions are located where enzymatic domains are defined), we arbitrarily chose them as a unit of comparison, and an exon-by-exon analysis was made. Fig. 7 illustrates the results in a histogram showing the percent homology of the amino acids in each exon (or, in the case of the nematode, the amino acids in the analogous regions which correspond to the chicken exons) in relation to the overall homology. The results are striking and demonstrate that certain regions are preferentially conserved, whereas others are hypervariable.

The third exon, which contains the translational initiation codon, significantly deviates from the average homology, being hypervariable both in the nematode and rat. We have noted this previously in a comparison between the chicken, rabbit, and rat myosin genes (28) and have hypothesized that this exon, which encodes the amino terminus of the protein, may partly account for the differing ATPase activities of the different isoforms.

The fifth exon contains a portion of the ATPase site (40) encoding ε-trimethyllysine 130 and tryptophan 131. In both the rat and nematode, this exon (or corresponding region in the nematode) is preferentially conserved, although the homology is much more pronounced between the rat and chicken. The adjoining exons show an interesting pattern in both comparisons: exons 6 and 8 are highly conserved, whereas exon 7 is not. Exons 9 and 10 are also highly conserved. Previous biochemical analyses have indicated that residues 117–195 (contained within exons 5–7) are important in myosin's interaction with ATP; this appears to be reflected by the striking sequence conservation apparent in exons 6 and 8. Exon 7, by comparison, does not appear to be as highly conserved. However, upon closer examination, we observed that this exon is highly conserved at its 5' end and hypervariable in its 3' region. In fact, it is within this exon (7) that the short consensus sequence (62), which is found in a wide...
variety of procaryotic and eucaryotic nucleotide-binding proteins (63), is located. Fig. 8 shows the extent of the homologies in and around this region. Between residues 143 and 200, there is 100% homology between the chicken and rat myosins. The available rabbit sequence (45) also shows 100% homology in this fragment. Whereas there are numerous differences between the chicken and nematode sequences, all but three of the substitutions are conservative and would have little effect, according to the available algorithms (41, 53), on the conformation of this region. It is noteworthy that the homologies extend across and terminate before exon borders. This emphasizes the apparent lack of correspondence between probable protein domains and individual exons.

The conservation observed between amino acids 473 and 529, encoded by exon 13 (see Fig. 7), is quite striking; it is highly conserved between the nematode and chicken, and the chicken and rat sequences are also homologous. Interestingly, both actin-binding and ATP interactions have been localized to these residues (47–50); and it is not surprising, considering the importance of these functions to the protein’s activities, that these residues also have been preferentially conserved.

All of these regions are localized in S-1; the pattern of sequence conservation in both comparisons is less apparent in the remainder of the molecule. However, the analysis does reveal some interesting aspects of LMM. The sequences which encode the rod diverge extensively between the different myosins, and this is reflected in the histogram (Fig. 7); the homologies in exons 27–40 are mostly below the average. However, the amino acids encoded by (chicken) exons 37 and 39 are preferentially conserved between the three myosins. These peptides are located at the end of the rod, and it has been hypothesized that this region may play an important role in the assembly of the thick filament (58). It seems probable that, if this is the case, the peptides important for this function are encoded by these sequences.

It should be emphasized that the unit of comparison, the chicken exon, is an arbitrary one; and we do not mean to imply that the exons define any functional or structural domain. Indeed, in the nematode gene, the coding sequence is interrupted by only eight introns (although five of these are located in the positions occupied by analogous introns in the vertebrate genes); and the histogram was constructed using the regions in the nematode which correspond to exon sequences in the vertebrates. It should also be noted that the sizes of the different exons do bias the analysis to some extent since exons encoding a small number of amino acids have a higher probability of deviating from the average degree of homology. Despite the above caveats, the analysis serves to illustrate which general regions are conserved and which diverge extensively as compared to the overall homologies between the genes.

The conservation of gene structure between the rat (25) and chicken genes is high. Both genes contain two exons which encode the 5′-untranslated end; the methionine initiating codons are both located in the third exons. Both genes encode 1940 amino acids; and, within the coding regions, all intron positions are exactly conserved except for the last. The terminal exon in the chicken is split in two in the rat, the extra mammalian intron occurring 24 bases upstream from the stop codon.

The development of a cDNA which encodes almost the entire polypeptide should be useful in initiating structure–function studies and increasing the biochemical resolution of the system. For example, it should now be possible to perform site-specific mutagenesis of the nucleotides encoding particular amino acids, such as tryptophan 131, or one of the reactive thiols. If these constructs can be expressed in a suitable host and the protein used to reconstitute the myofibril, it will become possible to determine the exact primary structure which underlies the protein’s functions.

**REFERENCES**

1. Lowey, S., and Holt, S. (1982) Cold Spring Harbor Symp. Quant. Biol. 37, 19–28
2. Silberstein, L., and Lowey, S. (1981) J. Mol. Biol. 148, 153–189
3. Oreizeen, P., and Richards, D. H. (1982) Cold Spring Harbor Symp. Quant. Biol. 37, 25–37
4. Katayama, E., and Wakabayashi, T. (1981) J. Biochem. (Tokyo) 90, 703–714
5. Burke, M., Sivaramakrishnan, M., and Kamalakannan, V. (1983) Biochemistry 22, 3046–3053
6. Szalay, L., Balint, M., Szilagyi, L., and Balint, M. (1972) Biochem. Biophys. Res. Commun. 77, 935–945
7. Mueller, H., and Perry, S. V. (1962) Biochem. J. 85, 431–439
8. Balint, M., Wolf, L., Tarscafalvi, A., Gergely, J., and Sretter, A. (1978) Arch. Biochem. Biophys. 190, 793–799
9. Mornet, D., Bertrand, R., Pantele, P., Audemard, E., and Kassab, R. (1981) Nature 292, 301–305
10. Mornet, D., Pantele, P., Bertrand, R., Audemard, E., and Kassab, R. (1981) FEBS Lett. 123, 54–58
11. Mornet, D., Us, K., and Morales, M. F. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 736–739
12. Muhlrad, A., and Morales, M. F. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 1003–1007
13. Huxley, H. E. (1969) Science 164, 1536–1538
14. Thomas, D. D., Seidel, J. C., Hyde, J. S., and Gergely, J. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 1729–1733
15. Pepe, F. A. (1967) J. Mol. Biol. 27, 205–225
16. Harrington, W. F. (1971) Proc. Natl. Acad. Sci. U. S. A. 68, 685–689
17. Tsong, T. Y., Himmelfarb, S., and Harrington, W. F. (1983) J. Mol. Biol. 164, 431–450
18. Lu, R. C., and Wong, A. (1985) J. Biol. Chem. 260, 3456–3461
19. Biro, N. A., Szilagyi, L., and Balint, M. (1972) Cold Spring Harbor Symp. Quant. Biol. 37, 53–65
20. Harrington, W. F., Burke, M., and Barton, J. S. (1972) Cold Spring Harbor Symp. Quant. Biol. 37, 64–85
21. Parry, D. A. D. (1981) J. Mol. Biol. 153, 459–464
22. Trus, B. I., and Elutings, M. (1981) in Structural Aspects of Recognition and Assembly in Biological Macromolecules (Balan, M., Sussman, J. L., Trub, W., and Yonath, A., eds) Vol. 1, p. 361, International Science Services, Boston
23. McLachlan, A. D., and Stewart, M. (1975) J. Mol. Biol. 98, 293–304
24. Karn, J., Brenner, S., and Barnett, L. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 4253–4257
25. Strehler, E. E., Strehler-Page, M. A., Perriard, J. C., Periasamy, M., and Nadel-Ginard, B. (1986) J. Mol. Biol. 190, 291–317
26. Robbins, J., Freyer, G. A., Chisolm, D., and Gilliam, T. C. (1982) J. Biol. Chem. 257, 549–556
27. Robbins, J., Horan, T., Gulick, J., and Kropf, K. (1986) J. Biol. Chem. 261, 6606–6612
28. Gulick, J., Kropf, K., and Robbins, J. (1985) J. Biol. Chem. 260,
Myosin Heavy Chain Gene

29. Wright, S., Rosenthal, A., Flavell, R., and Grosfeld, F. (1984) J. Biol. Chem. 259, 2979-2982
30. Wright, S., Rosenthal, A., Flavell, R., and Grosfeld, F. (1984) J. Biol. Chem. 259, 2983-2987
31. Messing, J. (1983) Methods Enzymol. 101, 20-78
32. Messing, J. (1983) Methods Enzymol. 101, 20-78
33. Dale, R. M., Koolhein, T., and Houchins, J. P. (1985) Plasmid 13, 31-40
34. Okayama, H., and Berg, P. (1982) Mol. Cell. Biol. 2, 161-170
35. Gohier, U., and Hoffman, B. J. (1983) Gene (Amst.) 18, 49-53
36. Freyer, G. A., and Robbins, J. (1983) J. Biol. Chem. 258, 7149-7154
37. Robbins, J., Kropp, K., Glick, J., and Valente, N. (1985) UCLA Symp. Mol. Cell. Biol. 29, 373-385
38. Lu, R., Sosinski, J., Balint, M., and Sreter, F. A. (1978) Fed. Proc. 37, 1695
39. Okamoto, Y., and Yount, R. G. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 1575-1579
40. Wells, J. A., and Yount, R. G. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 4903-4907
41. Finer-Moore, J., and Stroud, R. M. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 1575-1579
42. Hozumi, T., and Muhlrad, A. (1981) Biochemistry 20, 2945-2950
43. Mornet, D., Pantel, R., Bertrand, R., Audemard, E., and Kassab, R. (1980) FEBS Lett. 117, 183-188
44. Kubo, S., Tokuyama, H., and Tomomura, Y. (1965) Biochim. Biophys. Acta 100, 459-470
45. Tong, S. W., and Elzinga, M. (1983) J. Biol. Chem. 258, 13100-13110
46. Kornner, M., Thiem, N. V., Cardinaud, R., and Lacombe, G. (1983) Biochemistry 22, 5843-5847
47. Mahmood, R., and Yount, R. G. (1984) J. Biol. Chem. 259, 12956-12961
48. Sutoh, K. (1983) Biochemistry 22, 1579-1585
49. Mornet, D., Pantel, P., Audemard, E., and Kassab, R. (1979) Biochem. Biophys. Res. Commun. 89, 925-932
50. Yamamoto, K., and Kekine, T. (1979) J. Biochem. (Tokyo) 86, 1863-1868
51. McMahan, W. D. (1984) Annu. Rev. Biochem. Biophys. 13, 167-189
52. Chaussepied, P., Mornet, D., and Kassab, R. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 2037-2041
53. Chou, P. Y., and Fasman, G. D. (1974) Biochemistry 13, 222-245
54. Robbins, J., Kropp, K., Gulick, J., and Valente, N. (1985) UCLA Symp. Mol. Cell. Biol. 29, 373-385
55. Crick, F. H. C. (1953) Acta Crystallogr. 6, 869-979
56. Cohen, C., and Parry, D. A. D. (1986) Trends Biochem.Sci. 2, 145-148
57. McLachlan, A. D., and Karn, J. (1983) J. Mol. Biol. 164, 65-66
58. Rozek, C., and Davidson, N. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 2128-2132
59. Deleted in proof
60. Umeda, P. K., Levin, J. E., Sinha, A. M., Cribbs, L. L., Darling, D. S., Ende, D. J., Hsu, H.-J., Dixon, E., and Jakowicz, S. (1985) UCLA Symp. Mol. Cell. Biol. 29, 809-823
61. Gilliam, T. C. (1982) Ph.D. thesis, University of Missouri-Columbia
62. Walker, J. E., Sarasate, M., Runswick, M. J., and Gay, N. J. (1982) EMBO J. 1, 945-953
63. Higgens, C. F., Hiles, I. D., Salmond, G. P. C., Gill, D. R., Downie, J. A., Evans, I. J., Holland, I. B., Gray, L., Buckel, S. D., Bell, A. W., and Hermodson, M. A. (1986) Nature 323, 445-450
64. Hopp, T. P., and Woods, K. E. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 3624-3628