THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 266, No. 28, Issue of October 5, pp. 18469-18476, 1991
Printed in U.S.A.

Complete Primary Structure of a Scallop Striated Muscle Myosin Heavy Chain

SEQUENCE COMPARISON WITH OTHER HEAVY CHAINS REVEALS REGIONS THAT MIGHT BE CRITICAL FOR REGULATION*

(Received for publication, May 13, 1991)

László Nyitray, Elizabeth B. Goodwin‡, and Andrew G. Szent-Györgyi§

From the Brandeis University, Department of Biology, Waltham, Massachusetts 02254

We have determined the primary structure of the myosin heavy chain (MHC) of the striated adductor muscle of the scallop *Aequipecten irradians* by cloning and sequencing its cDNA. It is the first heavy chain sequence obtained in a directly Ca^{2+}-regulated myosin. The 1938-amino acid sequence has an overall structure similar to other MHCs. The subfragment-1 region of the scallop MHC has a 59–62% sequence identity with sarcomeric and a 52–53% identity with nonsarcomeric (smooth and metazoan nonmuscle) MHCs. The heavy chain component of the regulatory domain (Kwon, H., Goodwin, E. B., Nyitray, L., Berliner, E., O'Neall-Hennessey, E., Melandri, F. D., and Szent-Györgyi, A. G. (1990) *Proc. Natl. Acad. Sci. U. S. A.* 87, 4771–4775) starts at either Leu-755 or Val-760. Ca^{2+}-sensitive Trp residues (Wells, C., Warriner, K. E., and Bagshaw, C. R. (1985) *Biochem. J.* 231, 31–38) are located near the C-terminal end of this segment (residues 818–827). More detailed sequence comparison with other MHCs reveals that the 50-kDa domain and the N-terminal two-thirds of the 20-kDa domain differ substantially between sarcomeric and nonsarcomeric myosins. In contrast, in the light chain binding region of the regulatory domain (residues 784–844) the scallop sequence shows greater homology with regulated myosins (smooth muscle, nonmuscle, and invertebrate striated muscles) than with unregulated ones (vertebrate skeletal and heart muscles). The N-terminal 25-kDa domain also contains several residues which are preserved only in regulated myosins. These results indicate that certain heavy chain sites might be critical for regulation. The rod has features typical of sarcomeric myosins. It is 52–60% and 30–33% homologous with sarcomeric and nonsarcomeric MHCs, respectively. A Ser-rich tailpiece (residues 1918–1938) is apparently nonhelical.

Myosin is the best-characterized molecular motor. It generates force during muscle contraction and has a fundamental role in a variety of cellular motilities (for review, see Ref. 1). The conventional myosin molecule is a hexameric protein composed of a pair of heavy chains (≈200 kDa) and two pairs of light chains (≈20 kDa). The N-terminal ≈800 residues of the heavy chains fold into two globular head domains, while the C-terminal halves of the heavy chains dimerize to form an α-helical coiled-coil rod domain. The head contains the active site for myosin ATPase and the actin-binding site. The rod is responsible for the assembly of myosin molecules into filaments.

Contraction is triggered by binding of Ca^{2+} ions to regulatory components. There are two basic types of regulatory systems: actin- and myosin-based. In vertebrate skeletal and cardiac muscles the control is achieved by the troponin-tropomyosin complex associated with actin filaments. In vertebrate smooth muscles, nonmuscle cells and most invertebrate muscles, the myosin molecule itself is regulated; the regulatory components are the light chains. In nonmuscle systems and perhaps in molluscan smooth catch muscles phosphorylation of the myosin rod may also have some regulatory function (1, 2). Some invertebrate muscles are doubly regulated (3), and smooth muscles may also have additional actin-linked regulatory components, *i.e.* the proteins caldesmon and calponin (see Ref. 4). In regulated myosins it is either the Ca^{2+}-dependent phosphorylation of the regulatory light chain (RLC)[1] (smooth muscles, nonmuscle systems) or the direct binding of Ca^{2+} to the myosin molecule (molluscan muscles) which initiates contraction. Despite the difference in triggering events, there is evidence indicating structural similarities in regulated myosins. For instance, smooth muscle, nonmuscle, and molluscan muscle myosins can adopt a folded (10 S) conformation (5–7), in contrast to the nonregulated myosins; phosphorylation of smooth muscle RLC can substitute for direct Ca^{2+}-binding in the "switching on" of hybrid scallop myosin (8).

Both LCs are involved in the regulation of molluscan myosin. The RLC suppresses activity in the absence of calcium, and its presence is also required for Ca^{2+} binding (9, 10). However, the specific Ca^{2+} binding site is most likely located in the ELC. Sequence comparison of different ELCs indicated that only the molluscan ELCs contain a potential Ca^{2+} binding site (11). Although all RLCs contain a nonspecific divalent cation binding site, it has been shown by site-directed muta-

* This work was supported by National Institutes of Health (NIH) Grant AR15963, the Muscular Dystrophy Association, and NIH Shared Instrumentation Grant RR04671. 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.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) X55714.

‡ Present address: Laboratory of Molecular Biology, Univ. of Wisconsin, Madison, WI 53706.

§ To whom correspondence should be addressed. Tel.: 617-736-3133; Fax: 617-736-3107.

[1] The abbreviations used are: RLC, regulatory light chain; ELC, essential light chain; LC, myosin light chain; MHC, myosin heavy chain; RD-HC,heavy chain component of regulatory domain; S1, subfragment-1; 25K, 50K, 20/25K, the three main proteolytic "domains" of S1 with molecular masses of 25, 50, and either 20 or 25 kDa; kb, kilobase(s); SH, reactive thiol.

genesis that it is not the specific triggering site (12).

Recently we have isolated a regulatory complex from scallop myosin, consisting of the two LCs and a portion of the heavy chain. Studies of hybrid complexes with foreign ELCs have directly demonstrated the role of the ELC in the specific Ca^{2+}-binding (13). This Ca^{2+} binding site is unusual,since it requires the interaction of ELC with RLC and the heavy chain to form the specific, high affinity site. The role of the RLC is very specific, since RLCs from nonregulated myosins (which can be inhibitory in scallop myosin hybrids) cannot restore Ca^{2+} binding (14). The specific contribution of the heavy chain in Ca^{2+} binding and regulation is unknown.

Protein sequencing and cloning of cDNA or genomic DNA have elucidated the primary structure of several MHCs and LCs, including LCs but no heavy chain from molluscan muscles (1, 16). It is advantageous to study the regulation of contraction at the molecular level in the molluscan system, since this myosin itself is a Ca^{2+}-regulated molecule. Therefore, we initiated sequence studies of the scallop striated adductor muscle MHC through cDNA cloning. Here we present its complete deduced primary structure. Analysis of the sequence revealed that the overall structure of the scallop MHC is similar to other conventional MHCs. Nevertheless, more detailed comparison with other MHCs delineates heavy chain sites which could be essential in the regulatory function of regulated myosins.

MATERIALS AND METHODS

cDNA Cloning—Bay scallops (*Aequipecten irradians*) were generously supplied by the Nantucket Marine Department. Total RNA was isolated from striated adductor muscle by the single-step extraction method of Chomczynski and Sacchi (17). Messenger RNA was isolated by oligo(dT)-cellulose chromatography (18) or directly from lysed cell with the Fasttrack kit (Invitrogen, San Diego, CA). Complementary DNA was synthesized as described by Gubler and Hoffman (19). Oligo-dT or an internally specific oligonucleotide, LL-60 (5′ACGCTCCTCGAGTTCCTT3′) was used as a primer in the reverse transcriptase reaction. The blunt-ended cDNA was tailed with the nonpalindromic *Bst*XI linker, size-fractionated on agarose gel, and ligated into *Bst*XI-cut pcDNAII phagemid expression vector (Invitrogen). LL-60 (nucleotides 2893–2910 in the full-length cDNA sequence (23)) was designed based on the sequence of a cDNA-fragment isolated from a λgt11 scallop catch muscle library (kind gift of Dr. L. Leinwand, A. Einstein College of Medicine, Bronx, NY). We have completely sequenced this cDNA fragment (nucleotides 1123–4885); it turned out to be identical with the striated MHC sequence. The libraries were screened with an affinity-purified polyclonal rabbit anti-scallop myosin antibody (20). Part of the internally primed library was also screened by colony hybridization. The DNA probe was labeled by random primed incorporation of digoxigenin-labeled dUTP using the Genius Kit (Boehringer-Mannheim). RNA Northern blot analysis was carried out according to Sambrook *et al.* (18).

DNA Sequencing—Double-stranded cDNA was sequenced directly in pcDNAII, using vector- and MHC-specific primers, or after subcloning overlapping restriction fragments into pTZ19R vector (Pharmacia LKB Biotechnology Inc.) or generating nested deletions by exonuclease III digestion (Promega Biotec.). Each base was determined on both strands a minimum of two times. Sequencing by the dideoxy chain-termination method (21) was performed with T7 DNA polymerase (Pharmacia).

DNA and Protein Sequence Analysis—Sequences of DNA fragments were assembled and translated by the Pustell Sequence Analysis Program (IBI). Protein sequences were analyzed by the University of Wisconsin Genetics Computer Groups (GCG) program, Version 6.2 (22). Sequence alignments and comparisons were performed by the GAP, LINEUP and DISTANCES programs. PEPTIDESTRUCTURE and PEPPLOT were used to predict secondary structure. Genebank and EMBL databases were the source of other MHC sequences.

RESULTS AND DISCUSSION

Cloning of Scallop MHC cDNA

We screened 2×10^4 colonies from a striated adductor muscle expression cDNA library with a polyclonal adductor muscle MHC antiserum (20) and isolated 15 positive clones. All of the cDNA fragments had a poly(A) tail and the four longest (3.5 kb) had exactly the same 5′ end, indicating that premature termination occurred during construction of the library. In order to get longer clones we made a second plasmid library in the same vector, primed by an internally specific oligonucleotide, LL-60 (see "Materials and Methods"). 23 additional positives were isolated from the second library by antibody screening and DNA probe hybridization with a 0.45-kb internal *Bam*HI fragment (nucleotides 1942–2388). The longest cDNA clones covered the 5′ end of the coding region and 5′-untranslated region sequences.

Fig. 1 shows a restriction map of the cDNA sequence and the three longest overlapping fragments (AHC-2, EHC-10, R-17) that we used to determine the full-length cDNA sequence. We have also mapped and partially sequenced several shorter cDNA clones and found no sequence difference in the protein coding region. Northern blot analysis, probed with the 3′-end *Eco*RI fragment revealed that a 6.8-kb message is expressed both in striated and catch adductor muscle of the scallop (data not shown).

The total length of the sequenced cDNA is 6751 nucleotides. A 5814-nucleotide open reading frame begins at position 145. Evidence indicates that translation is initiated at this site: it is the first ATG in the determined sequence; there is an in-frame stop codon 12 nucleotides upstream; the encoded N-terminal sequence is similar to other conventional myosins.

The full-length cDNA sequence has been reported (23). In this paper we provide the derived amino acid sequence and compare it with other published MHCs. Although the two light chains are known to be the primary regulatory components, nevertheless analysis of the heavy chain sequence may establish which site(s), if any, contribute to the regulatory function.

Features of the Deduced Heavy Chain Sequence

The calculated molecular weight of the scallop MHC is 222,553. The sequence is similar overall to other MHCs. Simple comparison with fully sequenced heavy chains shows 58 ± 2, 40 ± 1, and 34% identity with sarcomeric (vertebrate striated, cardiac, and invertebrate muscles), nonsarcomeric (smooth muscle and nonmuscle cells), and unicellular (*Dictyostelium*) MHCs, respectively. The globular head region is

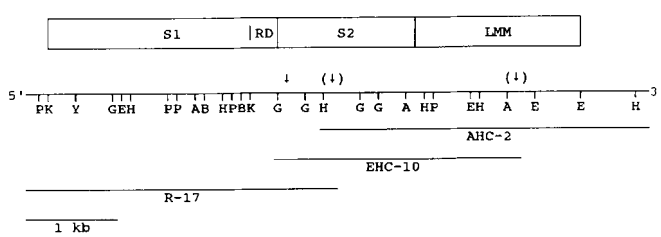

FIG. 1. **Restriction map of scallop striated muscle MHC cDNA.** The three cDNA fragments, shown under the restriction map, were isolated from an oligo-dT (AHC-2) or an internally primed (EHC-10, R-17) pcDNAII plasmid expression library and used for sequencing the full-length transcript. The *down arrows* show the positions of LL-60 18-mer used to initiate the second library. In parentheses are sites with sufficient homologies for LL-60 priming. The *block* at the top illustrates the protein coding region. Restriction sites: *A*, *Apa*I; *B*, *Bam*HI; *E*, *Eco*RI; *G*, *Bgl*II; *H*, *Hin*dIII; *K*, *Kpn*I; *P*, *Pst*I; *Y*, *Aha*II.

apparently more conserved (59–62% and 52–53% identity with sarcomeric and nonsarcomeric MHCs, respectively) than the coiled-coil rod (52–60% and 30–33%), a feature common with other MHCs (1).

Conserved and Variable Regions in the Globular Head—We aligned the scallop MHC sequence with 10 other MHCs by the GCG GAP and LINEUP programs (22). Fig. 2, *rows a-c* show part of the alignment, the scallop sequence (*row a*) in the head region (residues 1–844) with a vertebrate skeletal MHC (chicken embryonic, *row b* (24)) and a vertebrate smooth muscle MHC (chicken gizzard, *row c* (25)). A consensus sequence is shown in *row d*, based on the three sequences above, plus nine other MHCs: rat embryonic skeletal (26), human embryonic and perinatal skeletal (27, 28), rat α- and β-cardiac (29), *Drosophila* sarcomeric *cD301* (30), nematode sarcomeric *unc-54* (31), chicken intestinal epithelial cells (32), and *Drosophila* nonmuscle cells (33) MHCs.[2] Landmark residues and segments, as well as two short sequences previously determined by protein microsequencing (13),[3] are also illustrated in Fig. 2. The least conserved regions, with considerable length and sequence heterogeneity, correlate with the protease-sensitive regions of S1. These regions probably form flexible, surface-loop structures. Limited proteolysis nicks the myosin heads at two "junctions," resulting in three fragments, from the N terminus 25, 50, and 20/25 kDa (the C-terminal fragment is either 20 or 25 kDa depending on the condition of the digestion; see Ref. 34). These fragments form the putative domains of the head. Szentkiralyi (34) has shown that in addition to the 25–50K junction (residues 200–213; all the positions referred to hereafter are in the scallop sequence of Fig. 2) and the 50–20/25K junction (residues 624–640), scallop S1 has a third tryptic site located 65 kDa from the N terminus. The sequence from residue 557 to 575 contains several Pro, Gly, and Lys; this is indicative of a surface-loop structure, and hence presumably corresponds to the protease-sensitive site, the cleavage of which abolishes ATPase activity (35). In rabbit muscle myosin thrombin cleaves after an adjacent lysyl residue (Lys-558), provided thiol groups are modified in the 50K domain, resulting also in the loss of ATPase activity (36).

Highly conserved regions include parts of the nucleotide-binding pocket in the 25K domain (residues 112–132, 159–192) and 50K domain (residues 225–251) (37–39), and the reactive SH region in the 20/25K domain (residues 693–703). Cys-703 (SH-1) was found to be the most reactive cysteine in scallop as well (40). In addition there are short conserved segments scattered in the primary sequence and a long one in the 50K domain (residues 456–500) which could have important yet unidentified structural or functional roles.

At position 128 the rabbit skeletal and gizzard MHCs contain a trimethylated Lys which was thought to participate in the ATPase reaction (1, 25) while residue 129, a Trp in vertebrates, has been cross-linked to a photoaffinity ATP adenine analogue (37). In the scallop these 2 residues are substituted with an Arg-Arg dipeptide, making it unlikely that the modified lysyl side chain plays a critical role in the active center. The intrinsic fluorescence of Trp residue(s) is enhanced upon ATP binding both in rabbit skeletal and scallop myosins (41, 42). It has been assumed that Trp-129 in the rabbit sequence is an ATP-sensitive residue (37). Based on our sequence information, the ATP-sensitive Trp is most presumably located in the 50K fragment; there is no conserved Trp in the 25K, whereas there are three in the 50K domain (Trp-437, -507, and -594).

The precise location of the actin binding site(s) on the myosin molecule is uncertain. Sites in the 50K and 20/25K domains, at the 50–20/25K junction and around the reactive SH groups were shown to interact with actin (see Ref. 15). The 30-kDa C-terminal peptide of scallop S1 has been cross-linked to actin (43).

Heavy Chain Component of the Regulatory Domain—We have recently isolated by proteolytic digestion a regulatory complex (referred to as RD) from scallop myosin, comprised of the two light chains and an approximately 8–10-kDa heavy chain fragment. The RD occupies the neck region of myosin. The N terminus of its heavy chain component (RD-HC) is at either Leu-755 or Val-760 (13). The C-terminal end of the RD-HC peptide has yet to be established but most likely is around Pro-835. This fully conserved residue was believed to be at or near the physical boundary of the head and the rod (1). An alternative view places the head-rod junction near residue 885 (44, 45). In this case the proteolytic separation of S1 from the rod would occur within the head. This possibility together with the fact that scallop S1, unlike myosin or heavy meromyosin, lacks Ca^{2+} sensitivity (although the specific Ca^{2+} binding site is preserved (46)) suggests that residues beyond Pro-835, as part of the neck region, may be essential for the regulatory function (*cf.* Refs. 13 with 15). There is no apparent EF-hand motif in the regulatory domain region which could provide a Ca^{2+} binding site, supporting the view that the triggering Ca^{2+} site is in the ELC (13).

The structure of the RD-HC is presumably quite extended: morphological evidence indicates an extended neck region, attached to a globular distal region (25 + 50K domains) of the head; the two light chains lie close to each other for much of their length; the RD-HC and the LCs mutually protect each other from proteolytic digestion (13, 15). Secondary structure predictions, using the Garnier *et al.* (47) and Chou and Fasman (48) algorithms, showed a consensus α-helix only for residues 770–780, also suggesting a "loose" extended structure for RD-HC, which may be stabilized only in interactions with the light chains.

Electron microscopy studies located the light chains in the neck region of myosin myosin (see Ref. 15). The RD-HC binds both light chains stoichiometrically, demonstrating that the binding sites are within this peptide (13). The most important binding determinants for the RLC presumably reside in the C-terminal 20 or so residues. It is known that in contrast to the longer Mg-S1^{+LC} preparation, the ~5-kDa shorter EDTA-S1^{-LC} does not contain RLC (34, 46). Moreover, the longer scallop S1 contains conformationally sensitive Trp residues, corresponding to a Trp-rich, highly hydrophobic region located immediately before the last Pro of the head (residues 818–834), which responds to Ca^{2+} binding to the regulatory domain (42). Although this aromatic-rich segment is required for regulation, it is not proven that it has a specific role.

The location of the ELC (called alkali LC in vertebrate myosins) binding site has been shown by recombinant DNA techniques to be within the C-terminal 42 residues of the longer S1 (49, 50). The fact that the C terminus of the 20-kDa fragment in trypsin-cut rabbit or chicken skeletal S1 is at residue 805 (1) further narrows the main binding region to within residues 784–805. The finding that after removal of the essential light chain, trypsin cleaves off a 2-kDa peptide from the C-terminal end of the 20-kDa fragment (51) supports this localization. There is relatively low sequence conservation in the light chain binding region, despite the fact that all the

[2] The complete alignment is available (from A. G. S.-G.) upon request.

[3] M. Titus and E. O'Neall-Hennessey, unpublished results.

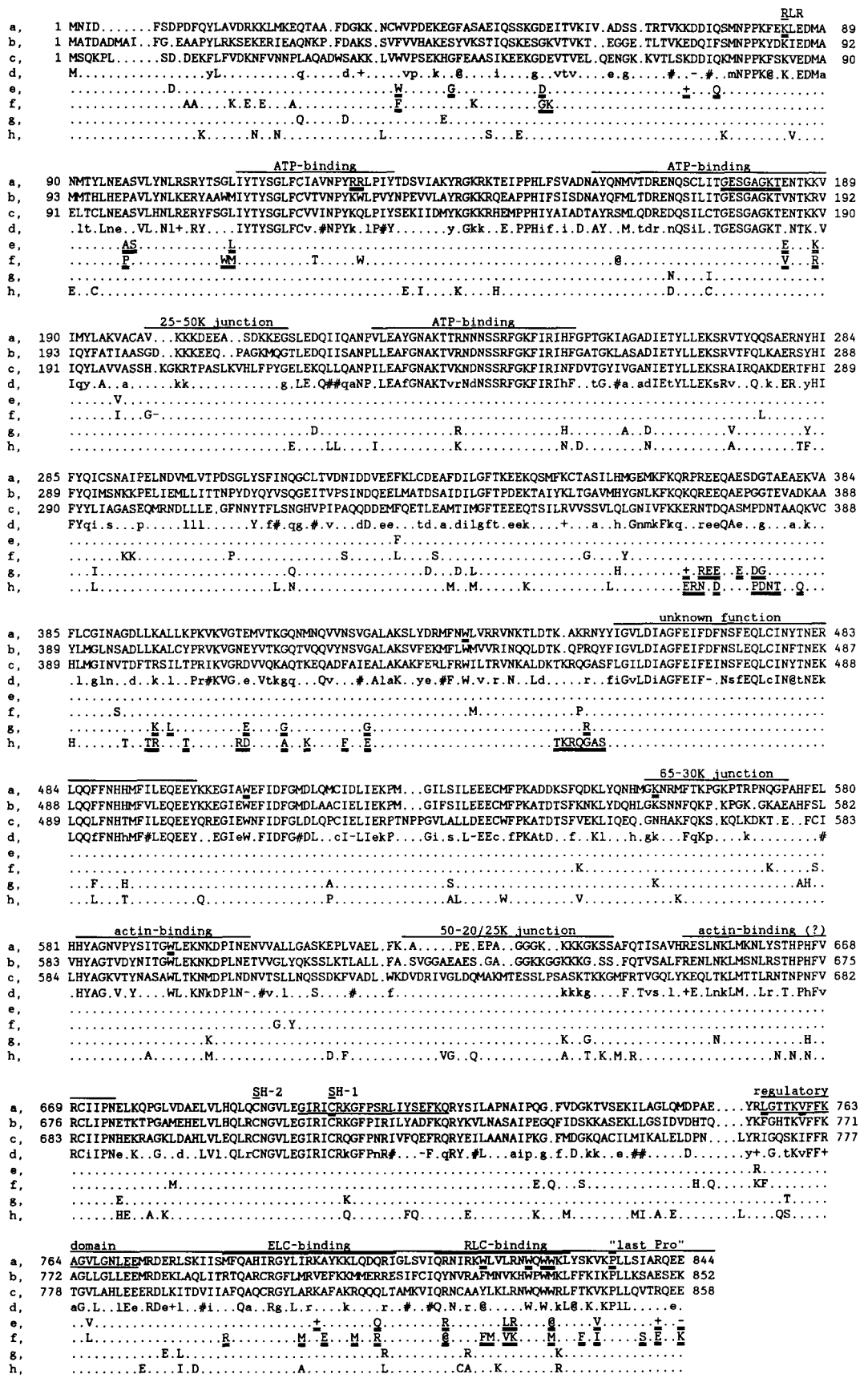

FIG. 2. **Sequence alignment of myosin heads.** The scallop MHC head region was aligned with 11 other MHCs using the GCG GAP and LINEUP programs. *Rows a-c* show part of the alignment: *a*, scallop MHC; *b*, chicken embryonic skeletal muscle MHC (26); *c*, chicken smooth muscle MHC (27). *Periods* are padding characters introduced to optimize the alignment. Major sequence features, such as the main proteolytic junctions, the nucleotide-, actin-, and light chains binding sites, reactive SH-1 and SH-2, a reactive lysyl (RLR), are also indicated

FIG. 3. **Sequence of the scallop MHC rod.** Positions of the 28-residue repeats, the skip residues, the N-terminal end of LMM are indicated. The tailpiece of the *Mytilus* anterior bysuss retractor muscle MHC is shown under the scallop sequence. *Double underlining* refers to a phosphorylated Ser (54).

RLCs tested have a common attachment site on the scallop heavy chain and the same is likely true for the ELCs (13, 14). The high positive charge of this stretch of the sequence in all myosins (residues 784–835, 13:1 positive:negative in scallop) could contribute to the high affinity for the light chains having net negative charge.

It is worthwhile to note that a complex analogous to the RD can be isolated from rabbit skeletal myosin,[4] indicating that the RD is a common structural entity in all myosins. This complex may have lost its regulatory function in vertebrate sarcomeric myosins by key mutations in the heavy chain and/or in the light chains.

Sequence Characteristics of the Rod—The scallop rod sequence has features typical of other sarcomeric MHC rods (Fig. 3). It shows a strong heptad repeat (with hydrophobic residues at positions 1 and 4) which is necessary to form the hydrophobic seam between the two α-helical chain of the coiled-coil, and a superimposed 28-residue repeat, the charge distribution of which is responsible for the electrostatic interactions between myosin molecules in the filamentous state (52). The sequence contains 4 skip residues at strictly conserved positions (Gln-1185, Glu-1382, Glu-1579, and Gly-1804). These residues interrupt the regular heptad repeat, probably modifying the pitch of the coiled-coil (52), and by causing local instability could be responsible for bending of the myosin rod at well-defined positions (45). Comparison with other MHC rods reveals additional conserved regions (not shown). The functions of these regions remain to be established, with the exception of the one which appears to be involved in the assembly of vertebrate sarcomeric myosins into thick filaments (residues around 1749–1770).[5]

After 38 and one-half 28-residue repeats, the rod ends in a 21-residue-long Ser-rich tailpiece. Secondary structure prediction indicates that this stretch of peptide is apparently nonhelical. In contrast to other myosin tailpieces the scallop sequence does not contain proline. The tailpieces have considerable length and sequence diversity, but their functional significance has not been demonstrated. In the last 30 residues of the coiled-coil preceding the nonhelical end the sequence conservation is relatively high, indicating some structural/functional constraint in this region. Tailpieces of several nonsarcomeric MHCs have been shown to be phosphorylatable, including molluscan catch muscle, vertebrate smooth muscle, vertebrate nonmuscle MHC, and protozoan MHCs (see Ref. 2). *Acanthamoeba* and *Dictyostelium* myosin II tailpiece phosphorylation regulates filament assembly and ATPase activity (see Ref. 1), while phosphorylation of catch MHC is thought to be involved in the control of the catch contraction (2). Castellani *et al.* (54) have isolated and sequenced two phosphorylated peptides from a *Mytilus* anterior byssus retractor muscle myosin preparation. Peptide II represents the C-terminal end of the anterior bysuss retractor muscle MHC,[6] has the same length as the scallop adductor myosin tailpiece, and has 57% homology with it. There is a 6-residue stretch around the phosphorylated Ser (see Fig. 3) which is identical in the two sequences. In spite of this similarity the scallop striated muscle MHC is not phosphorylated by endogenous kinase (2).

[4] H. Kwon, unpublished results.

[5] R. L. Sohn, A. G. Szent-Györgyi, and L. A. Leinwand, submitted for publication.

[6] L. Nyitray and W. Yang, unpublished results.

above the sequences (see text). The two short scallop sequences determined by protein sequencing are *underlined*. *Row d* is a consensus sequence based on the three sequences above, plus nine others: rat embryonic skeletal (26), human embryonic and perinatal skeletal (27, 28), rat α- and β-cardiac (29), *Drosophila* sarcomeric *cD301* (30), nematode sarcomeric *unc-54* (31), chicken intestinal epithelial cells (32), and *Drosophila* nonmuscle cells (33). *Upper-case letters*, invariant residues; *lower-case letters*, identical in at least 10 sequences; +, K or R; −, E or D; #, L,I,V,M; @, W,F,Y. *Rows e-f* depict extra conserved residues (in addition to the invariant residues in *row d*) in regulated and nonregulated MHCs, respectively, provided the conserved residue is not found in the opposite group. *Rows g-h* show the extra conserved residues in sarcomeric and nonsarcomeric MHCs, respectively. The regulation-sensitive and sarcomeric/nonsarcomeric-specific residues mentioned in the text are *double underlined*.

TABLE I

Segmental comparison of the scallop MHC with other MHCs

A, the percentage sequence identity was calculated by the GAP program. S1, the head region from residues 5 to 844 (of the scallop sequence); 25K, 25K domain (residues 5–200); 50K, 50K domain (residues 214–623); 20K/N, the N-terminal two-thirds of the 20/25K domain (residues 639–783); RD/LC, light chain binding region of the RD-HC (residues 784–844). The single lines and the double lines divide the columns into two groups: sarcomeric/nonsarcomeric MHCs in S1, 50K, 20K/N, and regulated/nonregulated sequences in 25K and RD/LC, respectively. *B*, the sequences were aligned by the GAP and LINEUP programs, and the similarity matrix calculated by the DISTANCES program.

A. Sequence identity of the scallop segments with other MHCs

	S1	25K	50K	20K/N	RD/LC
RTCA[a]	61.9	53.9	68.8	68.1	47.5
CHSK	60.3	53.6	66.7	66.0	39.3
DRSA	61.0	54.7	64.3	62.9	59.0
CHSM	52.6	57.2	50.5	53.8	60.7
CHNM	52.9	56.0	52.0	53.8	59.0

B. Similarity matrix in the RD/LC (upper half) and 50K (lower half) segments

	RTCA	RTSK	CHSK	SCSA	DRSA	NESA	CHSM	CHNM	DRNM
RTCA	—	85.9	85.9	65.6	65.6	60.6	59.4	57.8	59.4
RTSK	90.2	—	95.3	60.9	60.9	53.1	53.1	53.1	54.7
CHSK	91.9	94.1	—	62.5	64.1	54.7	53.1	53.1	53.1
SCSA	79.7	78.4	79.7	—	78.1	68.8	75.0	73.4	68.8
DRSA	79.7	78.2	79.4	78.0	—	64.1	68.8	70.3	65.6
NESA	77.2	75.3	76.0	73.7	74.6	—	64.1	62.5	60.9
CHSM	68.9	68.9	68.4	66.9	65.7	67.4	—	95.3	87.5
CHNM	67.7	68.1	67.9	65.7	65.0	67.7	92.2	—	87.5
DRNM	70.7	70.2	70.0	67.5	67.0	68.2	82.6	82.9	—

[a] The abbreviations used are: RTCA, rat α-cardiac; RTSK, rat embryonic skeletal; CHSK, chicken embryonic skeletal; SCSA, scallop sarcomeric; DRSA, *Drosophila* sarcomeric; NESA, nematode sarcomeric; CHSM, chicken gizzard; CHNM, chicken intestinal epithelial; DRNM, *Drosophila* nonmuscle (for references see legend to Fig. 2). The similarity groups are boxed: in the 50K domain the sarcomeric and nonsarcomeric sequence pairs, in the RD/LC segment the regulated and nonregulated myosin sequence pairs (double-boxed).

Comparison of Regulated and Nonregulated Myosin Heavy Chains—There is a considerable sequence divergence between sarcomeric and nonsarcomeric MHCs, probably contributing to the differences in the contractile behavior of these myosins (32, 33). The role of the heavy chain in the regulatory function of regulated myosins, whether in the sarcomeric class (*e.g.* scallop muscle) or in the nonsarcomeric class (*e.g.* smooth muscle), is unknown. Sequence comparison of vertebrate smooth muscle MHC with other MHCs suggested that certain heavy chain sites in the C-terminal sequences of the 20/25K domain may be critical for regulation (15). In order to find out whether regulated and nonregulated myosins differ systematically, we have conducted a more detailed comparison of the head domains.

Table IA shows a simple segmental comparison (matching amino acids) of the scallop sequence with three regulated MHCs (chicken smooth muscle and nonmuscle, *Drosophila* sarcomeric) and two nonregulated ones (chicken skeletal and rat α-cardiac) in four regions of the head: 25K (residues 5–200), 50K (214–623) domains, N-terminal two-thirds of 20/25K domain (639–783), and the light chain binding region of the RD-HC (784–844). The junctions between the proteolytic domains and the extreme N-terminal sequence were omitted from the comparisons because of the high degree of length divergence. In the 50K domain and the N-terminal 145 residues of the 20/25K domain, the sequences (as is the case of the whole head region) fall clearly into two groups: sarcomeric and nonsarcomeric myosins. In contrast, the light chain binding region of the RD-HC of the scallop sequence (including the first few residues of the rod) resembles more the corresponding regions of other regulated myosin sequences, with a lesser homology to the nonregulated vertebrate sarcomeric myosins. The 25K N-terminal domain of the scallop myosin is also more similar to those of the regulated MHCs (Table IA). The same conclusion can be drawn when the aligned 50K and RD-HC light chain binding regions are compared pairwise by taking into account conservative amino acid replacements (Table IB).

Fig. 2, *rows e* and *f*, show conserved residues found in the sequence of the regulated and nonregulated classes, respectively, in addition to the conserved positions present in all of the aligned MHCs (*row d*). Likewise, *rows g* and *h* depict the additional conserved residues in sarcomeric or nonsarcomeric sequences, respectively. These additional conserved residues characteristic to a particular class of MHC represent positions where heavy chain classes differ from each other. The residues specific to sarcomeric/nonsarcomeric MHCs are most abundant in the 50K and the N-terminal two-thirds of the 20/25K domains (Fig. 2, *rows g-h*), in agreement with the segmental homologies. There are three clusters with especially high degrees of conservation within these two classes: at residues 366–381 (scallop sequence numbering), 395–424 and 449–455. The last stretch of sequence is flanking a highly conserved region in all MHCs (residues 456–517) with an unknown function. These sequence differences may contribute to the structural and functional differences between sarcomeric and nonsarcomeric myosins.

When the regulated MHCs are matched against nonregulated ones the additional conserved residues appear to be considerably more clustered in the LC binding region of the RD-HC and in the 25K domain, as expected from segmental comparisons. We tentatively identify these residues as being "regulation-sensitive." Based on their number and position we assume that it is unlikely that the members of the two groups are related by chance alone; instead, the conservation within a group probably reflects an important functional

constraint. The conserved residues in regulated myosins may be the result of evolutionary pressure to preserve conformation of critical regions for regulatory function. The different conserved residues in these regions of vertebrate sarcomeric myosins may be partly responsible for the lack of "on" and "off" regulation by light chains, or they may contribute to other functions characteristic to these myosins.

The regulated myosins form, at first sight, a heterogeneous group. The triggering signal is either the phosphorylation of the RLC (smooth muscle myosin) or a direct Ca^{2+} binding to myosin, most likely to the ELC (molluscan muscles). Nevertheless, the basic regulatory mechanism appears to be similar. The RLC is thought to suppress myosin activity by inhibiting a step in the kinetic cycle (55, 56), and this inhibition is reversed by phosphorylation or by direct Ca^{2+} binding. Several lines of evidence indicate structural similarities in molluscan and smooth muscle myosins. In both systems S1, unlike heavy meromyosin, is unregulated (9, 46, 57); removal of the RLC abolishes regulation (10, 58); regulated myosins, unlike nonregulated ones, can adopt a 10 S-folded configuration involving the neck region (5–7); structural changes have been observed by electron microscopy on Ca^{2+} activation or on phosphorylation at the head-rod junction of muscle filaments regulated by myosin (59, 60). All of this evidence implies that the light chains and the head-rod junction of the heavy chain play a crucial role in regulation; hence, it is not surprising that the light chain binding region of the RD-HC contains the largest number of regulation-sensitive residues. It has been suggested that specific interactions between the RLC and the N-terminal region of the rod could restrict the flexibility of the hinge, causing the inhibited state.[7] The two regulation-sensitive residues found beyond the "last Pro" of the head (positions 841 and 844) could be part of the interactive site. Interestingly, regulation-sensitive residues can be found near the ATP binding site of the 25K domain. Conserved residues of regulated MHCs are present: at positions 184 and 188, adjacent to the consensus Gly-rich loop "GESGAGKT" (residues 176–183) which is part of the nucleotide-binding pocket; at positions 98, 99, and 111 flanking part of the adenine binding region (residues 112–132); at several positions in the N-terminal 90 residues, containing nucleotide-modulated proteolytic cleavage sites (25, 53) and a highly reactive lysyl residue (Lys-81). Communication from the regulatory site to the catalytic site should occur over a large distance, involving domain-domain interactions (*cf.* Refs. 15 with 16); some of the regulation-sensitive residues may reside along this communication pathway.

Acknowledgments—We thank Dr. Leslie Leinwand for her continuous help and encouragement during this work and Elizabeth O'Neall-Hennessey, Francesco Melandri, and Hyockman Kwon for reading the manuscript.

REFERENCES

1. Warrick, H. M., and Spudich, J. A. (1987) *Annu. Rev. Cell Biol.* **3,** 379–421
2. Castellani, L., and Cohen, C. (1987) *Science* **235,** 334–337
3. Lehman, W., and Szent-Györgyi, A. G. (1975) *J. Gen. Physiol.* **66,** 1–30
4. Trybus, K. (1991) *Cell Motil. Cytoskeleton* **18,** 81–85
5. Craig, R., Smith, R., and Kendrick-Jones, J. (1983) *Nature* **302,** 436–439
6. Castellani, L., and Cohen, C. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 4058–4062
7. Ankrett, R. J., Rowe, A. J., Cross, R. A., Kendrick-Jones J., and Bagshaw, C. R. (1991) *J. Mol. Biol.* **217,** 323–335
8. Scholey, J. M., Taylor, K. A., and Kendrick-Jones, J. (1981) *Biochimie* **63,** 255–271
9. Szent-Györgyi, A. G., Szentkiralyi, E. M., and Kendrick-Jones, J. (1973) *J. Mol. Biol.* **74,** 179–203
10. Chantler, P. D., and Szent-Györgyi, A. G. (1980) *J. Mol. Biol.* **138,** 473–492
11. Collins, J. H., Jakes, R., Kendrick-Jones, J., Leszyk, J., Barouch, W., Thiebert, J., Spiegel, J., and Szent-Györgyi, A. G. (1986) *Biochemistry* **25,** 7651–7656
12. Goodwin, E. B., Leinwand, L. A., and Szent-Györgyi (1990) *J. Mol. Biol.* **216,** 85–93
13. Kwon, H., Goodwin, E. B., Nyitray, L., Berliner, E., O'Neall-Hennessey, E., Melandri, F. D., and Szent-Györgyi, A. G. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 4771–4775
14. Sellers, J. R., Chantler, P. D., and Szent-Györgyi, A. G. (1980) *J. Mol. Biol.* **144,** 223–245
15. Vibert, P., and Cohen, C. (1988) *J. Muscle Res. Cell Motil.* **9,** 296–305
16. Goodwin, E. B., Szent-Györgyi, A. G., and Leinwand, L. A. (1987) *J. Biol. Chem.* **262,** 11051–11056
17. Chomczynski, P., and Sacchi, N. (1986) *Anal. Biochem.* **162,** 156–159
18. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
19. Gubler, V., and Hoffman, B. J. (1983) *Gene (Amst.)* **25,** 263–269
20. Walliman, T., and Szent-Györgyi, A. G. (1981) *Biochemistry* **20,** 1176–1187
21. Hattari, M., and Sakaki, Y. (1986) *Anal. Biochem.* **152,** 232–238
22. Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395
23. Nyitray, L., Goodwin, E. B., and Szent-Györgyi, A. G. (1990) *Nucleic Acids Res.* **18,** 7158
24. Molina, M. I., Kropp, K. E., Gulick, J., and Robbins, J. (1987) *J. Biol. Chem.* **262,** 6478–6488
25. Yanagishawa, M., Hamada, Y., Katsuragawa, Y., Imamura, M., Mikawa, T., and Masaki, T. (1987) *J. Mol. Biol.* **198,** 143–157
26. Strehler, E. E., Strehler-Page, M. A., Perriard, J. C., Periasamy, M., and Nadal-Ginard, B. (1986) *J. Mol. Biol.* **190,** 291–317
27. Stedman, H. H., Eller, M., Jullian, E. H., Fertels, S. H., Sarkar, S., Sylvester, J. E., Kelly, A. M., and Rubinstein, N. A. (1990) *J. Biol. Chem.* **265,** 3568–3576
28. Kersch-Mizrachi, I., Feghali, R., Shows, T. B., and Leinwand, L. A. (1990) *Gene (Amst.)* **89,** 289–294
29. McNally, E. M., Kraft, R., Bravo-Zehnder, M., Taylor, D. A., and Leinwand, L. A. (1989) *J. Mol. Biol.* **210,** 665–671
30. George, E. L., Ober, M. B., and Emerson, C. P., Jr. (1989) *Mol. Cell. Biol.* **9,** 2957–2974
31. Karn, J., Brenner, S., and Barnett, L. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 4253–4257
32. Shohet, R. P., Conti, M. A., Kawamoto, S., Preston, Y. A., Brill, D. A., and Adelstein, R. S. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 7726–7730
33. Ketchum, A. S., Stewart, C. D., Stewart, M., and Kiehart, D. P. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 6316–6320
34. Szentkiralyi, E. M. (1984) *J. Muscle Res. Cell Motil.* **5,** 147–164
35. Szentkiralyi, E. M. (1987) *J. Muscle Res. Cell Motil.* **8,** 349–357
36. Chaussepied, P., Mornet, D., Audemard, E., Derancourt, J., and Kassab, R. (1986) *Biochemistry* **25,** 1134–1140
37. Okamoto, Y., and Yount, R. G. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 1575–1579
38. Cremo, C. R., Grammer, J. C., and Yount, R. G. (1989) *J. Biol. Chem.* **264,** 6608–6611
39. Grammer, J. C., and Yount, R. G. (1991) *Biophys. J.* **59,** (suppl.) 226
40. Titus, M. A.,and Szent-Györgyi, A. G. (1986) *Biophys. J.* **49,** 187
41. Werber, M. M., Szent-Györgyi, A. G., and Fasman, G. D. (1972) *Biochemistry* **11,** 2872–2883
42. Wells, C., Warriner, K. E., and Bagshaw, C. R. (1985) *Biochem. J.* **231,** 31–38
43. Labbe, J. P., Chaussepied, P., Derancourt, J., and Kassab, R. (1988) *Biochemistry* **27,** 5914–5922
44. Rimm, D. L., Sinard, J. H., and Pollard, T. D. (1989) *J. Cell Biol.* **108,** 1783–1789
45. Offer, G. (1990) *J. Mol. Biol.* **216,** 213–218
46. Stafford, W. F., Szentkiralyi, E. M., and Szent-Györgyi, A. G. (1979) *Biochemistry* **18,** 5273–5280

[7] H. Kwon, F. D. Melandri, and A. G. Szent-Györgyi, submitted for publication.

47. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120
48. Chou, P. Y., and Fasman, G. D. (1978) *Adv. Enzymol.* **47,** 45–148
49. Mitchell, E. J., Karn, J., Brown, D. M., Newman, A., Jakes, R., and Kendrick-Jones, J. (1989) *J. Mol. Biol.* **208,** 199–205
50. McNally, E. M, Bravo-Zehnder, M. M., and Leinwand, L. A. (1991) *J. Cell Biol.* **113,** 585–590
51. Burke, M., and Kamalakannan, V. (1985) *Biochemistry* **24,** 846–852
52. McLachlan, A. D., and Karn, J. (1982) *Nature* **299,** 226–231
53. Mornet, D., Pantel, P., Audemard, E., Derancourt, J., and Kassab, R. (1985) *J. Mol. Biol.* **183,** 479–489
54. Castellani, L., Elliott, B. W., and Cohen, C. (1988) *J. Muscle Res. Cell Motil.* **9,** 533–540
55. Chalovich, J. M., Chantler, P. D., Szent-Györgyi, A. G., and Eisenberg, E. (1984) *J. Biol. Chem.* **259,** 2617–2621
56. Sellers, J. R., Eisenberg, E., and Adelstein, R. S. (1982) *J. Biol. Chem.* **257,** 13880–13883
57. Seidel, J. C. (1980) *J. Biol. Chem.* **255,** 4355–4361
58. Trybus, K. M., and Lowey, S. (1988) *J. Biol. Chem.* **263,** 16485–16492
59. Ikebe, M., and Ogihara, S. (1982) *J. Biochem.* **92,** 1973–1977
60. Vibert, P., and Craig, R. (1985) *J. Cell Biol.* **101,** 830–837