Space-filling representation of myosin subfragment-1, a molecular motor. The colors define the global organization of the polypeptide chains that constitute this molecule. This structure provides the basis for hypotheses for the molecular mechanism of muscle contraction. See pages 50 and 58 and the Perspective on page 35. [Graphics: Adam Steinberg and Ivan Rayment]

Three-Dimensional Structure of Myosin Subfragment-1: A Molecular Motor

Ivan Rayment,* Wojciech R. Rypniewski,† Karen Schmidt-Bäse,‡ Robert Smith, Diana R. Tomchick,§ Matthew M. Benning, Donald A. Winkelmann, Gary Wesenberg, Hazel M. Holden

Directed movement is a characteristic of many living organisms and occurs as a result of the transformation of chemical energy into mechanical energy. Myosin is one of three families of molecular motors that are responsible for cellular motility. The three-dimensional structure of the head portion of myosin, or subfragment-1, which contains both the actin and nucleotide binding sites, is described. This structure of a molecular motor was determined by single crystal x-ray diffraction. The data provide a structural framework for understanding the molecular basis of motility.

Motility is one of the characteristic features of many living organisms and involves the transduction of chemical into mechanical energy. Only a limited number of strategies have evolved to accomplish this task. At present, three major classes of molecular motors have been identified, myosin, dynein, and kinesin, and all are important in cellular movement (*1*). Of these three proteins, the most abundant is myosin, which plays both a structural and an enzymatic role in both muscle contraction and intracellular motility.

I. Rayment, W. R. Rypniewski, K. Schmidt-Bäse, R. Smith, D. R. Tomchick, M. M. Benning, G. Wesenberg, and H. M. Holden are in the Department of Biochemistry and Institute for Enzyme Research, University of Wisconsin, 1710 University Avenue, Madison, WI 53705. D. A. Winkelmann is in the Department of Pathology, Robert Wood Johnson Medical School, Piscataway, NJ 08854.

*To whom all correspondence should be addressed.
†Present address: European Molecular Biology Laboratory, Notkestrasse 85, 2000 Hamburg 52, Germany.
‡Present address: Max Planck Institute for Biochemistry, 8033 Martinsried, Germany.
§Present address: Department of Biological Sciences, Purdue University, West Lafayette, IN 47907.

The role of myosin in movement has been most clearly defined from the study of cross-striated skeletal muscle, which shows a high degree of structural organization. In striated muscle the basic contractile unit is the sarcomere, which consists of overlapping arrays of thick and thin filaments. During contraction, these filaments, which are composed primarily of myosin and actin, respectively, slide past one another, thereby shortening the length of the sarcomere (*2*). Electron micrographs of muscle in rigor have revealed connections between the filaments in the overlap region, the so-called crossbridges. These crossbridges are formed by the globular regions of the myosin molecule and are responsible for force generation in the contractile process through the hydrolysis of adenosine triphosphate (ATP).

Myosin, which has a molecular size of about 520 kilodaltons, consists of two 220-kD heavy chains and two pairs of light chains that vary in molecular size depending on the source but are usually between 15 and 22 kD (*3*, *4*). The molecule is highly

asymmetric, consisting of two globular heads attached to a long tail. Each heavy chain forms the bulk of one head and intertwines with its neighbor to form the tail. Limited proteolytic digestion has shown that the myosin head, or subfragment-1 (S1), contains an ATP, actin, and two light chain binding sites and that the myosin rod, which is formed by a coiled coil of two α helices, accounts for the self-association of myosin at low ionic strength and the formation of the thick filament backbone (*3*). Spudich and co-workers have demonstrated that the globular head portions of myosin are sufficient to generate movement of actin in an in vitro motility assay (*5*).

Each globular head, derived from limited proteolysis, consists of a heavy chain fragment having a molecular size of 95 kD and two light chains yielding a combined molecular size of ~130 kD (*6*). The two light chains differ in their structure and properties and are known by a variety of names. In this article they are referred to as the regulatory and essential light chains. Neither type is required for the adenosine triphosphatase (ATPase) activity of the head (*7*). In some species, however, these chains regulate or modulate the ATPase activity of myosin in the presence of actin (*8, 9*). Amino acid sequence analyses reveal that both light chains share considerable sequence similarity with calmodulin and troponin C although most of the divalent cation binding sites have been lost during evolution (*10*).

During the last 40 years, enormous effort has been expended toward understanding the structure and function of the myosin head (*11*). Measurements from electron micrographs have suggested that the myosin head is pear-shaped, about 190 Å long and 50 Å wide at its thickest point (*12*). Molecular dimensions subsequently derived from studies of fixed thin sections cut from crystals of myosin S1 were consistent with these observations (*13*).

Although much biochemical and physical information has accumulated for myosin since 1950, structural knowledge of this protein or any other molecular motor has been lacking. We now describe the tertiary structure of the myosin head and suggest how this protein may serve to transduce energy from the hydrolysis of ATP into directed movement. We present the three-dimensional structure of myosin S1 at a nominal resolution of 2.8 Å and refinement *R* factor of 22.3 percent for all x-ray data recorded in that range.

Crystallization of myosin subfragment-1. Myosin is an abundant protein that can be easily prepared in gram quantities. Likewise, the myosin head, which is readily cleaved from the rest of the molecule by mild proteolysis, can be prepared in large quantities. This soluble subfragment has been known for approximately 30 years and has resisted crystallization despite numerous attempts. In view of its central importance for understanding the molecular basis of muscle contraction, we undertook an alternative approach to the usual ways of obtaining x-ray quality crystals. The protein was first subjected to mild chemical modification of the lysine residues by reductive methylation. This chemical modification has long been used as a gentle way to introduce a radioactive label into a protein (*14*).

Considerable effort was expended to determine the optimal procedure for modifying the protein since it was recognized that complete, homogeneous modification of the molecule was essential for obtaining high-quality crystals (Table 1). Many of the experiments necessary to derive the optimal protocol for methylation were performed in a parallel study on hen egg white lysozyme (*15*). In that study the three-dimensional structure of the modified protein was determined and refined to 1.8 Å resolution and shown to be essentially identical to that of the native protein except for the modified lysine residues. Modification of the lysine residues in lysozyme dramatically changed its crystallization properties. The kinetic and structural effects of this treatment on myosin S1 are discussed below.

Table 1. Amino acid analysis of modified and native myosin S1 (*60*). Prior to modification, the protein, at 5 mg/ml, was dialyzed against 200 mM potassium phosphate, pH 7.5, 1 mM $MgCl_2$. The protein was reductively methylated at 4°C by the sequential addition of 1 M dimethylamine borane complex dissolved in water (20 μl per milliliter of protein) and 1 M formaldehyde (40 μl per milliliter of protein) with rapid stirring. This process was repeated after 2 hours; a further portion (10 μl/ml) of dimethylamine borane complex was added after 2 hours and the reaction mixture was kept overnight at 4°C in the dark. The reaction was quenched by the addition of 3.8 M ammonium sulfate to a final concentration of 1 M and then dialyzed for 48 hours against 2.5 M ammonium sulfate, 50 mM potassium phosphate at pH 6.7 to precipitate the protein (*15, 49*). All except three to four of the lysine residues were modified. Discrepancy between the total number of lysine residues in the native and modified protein may have arisen from a calibration error in the dimethyllysine standard. The analyses for histidine, methionine, and arginine are shown as controls.

Amino acid	Residues (no.)		
	Theoretical	Native	Modified
Lysine	103	96.2	4.2
Me_1-Lys	0	0	0
Me_2-Lys	0	0.6	96.7
Me_3-Lys	3	3.6	3.4
Total lysine	106	100.4	104.3
Histidine	24	23.7	23.2
Methionine	39	39.4	38.9
Arginine	46	47.5	47.2

Myosin isolated from chicken pectoralis muscle consists of a mixed population of two isozymes caused by the existence of two species of the essential light chain (*16*). These light chains are referred to as A1 (21 kD) and A2 (16 kD). Amino acid sequence studies of the light chains have demonstrated that A1 and A2 are identical over their 142 residues at the COOH-terminus. The size difference is caused by an additional 41 amino acids present at the NH_2-terminus of A1. These isozymes arise by alternative transcription and two modes of splicing from a single gene (*17*).

The crystals used in our study contained both isoforms of the essential light chain. Myosin S1 was prepared by digestion with papain in the presence of $MgCl_2$ because the fragment produced under these conditions contained both the regulatory and essential light chains. The major drawback of papain as a proteolytic enzyme, however, was its lack of specificity. Apart from cleaving the heavy chain at the head-rod junction, additional proteolytic breaks were introduced into both the regulatory and A1 essential light chains. Also, there was partial phosphorylation of the regulatory light chain by endogenous myosin light chain kinase. The myosin S1 was prepared by an improved purification protocol that removed the heterogeneity arising from both proteolysis of the light chains and phosphorylation of the regulatory light chain (*18*).

Crystals were grown by batch methods from 1.35 M ammonium sulfate, 500 mM potassium chloride, and 50 mM potassium phosphate (pH 6.7) in the presence of 5 mM dithiothreitol and 0.5 mM sodium azide at a final protein concentration of 8 to 12 mg/ml. Crystallization was initiated by microseeding, and the crystals grew as thick rods to a length of 1 to 2 mm and a width and thickness of 0.4 and 0.3 mm, respectively, over a period of 2 to 3 months at 4°C. They belonged to the space group $C222_1$ with unit cell dimensions of $a = 98.4$, $b = 124.2$, $c = 274.9$ Å, and one molecule in the asymmetric unit. These crystals were different from those originally reported (*19*) and arose from improvements in both the chemical modification procedure and the protein homogeneity.

Structure determination. The x-ray data were collected in two stages (*20*). First, x-ray data sets to 4.5 Å resolution for the native and heavy atom–containing crystals were recorded by an area detector with the goal of determining the positions of the metal binding sites. These data were then extended to 2.8 Å resolution with

synchrotron radiation at Stanford University (SSRL) and Cornell University (CHESS). We recognized early that x-ray data collection and determination of the protein phases by multiple isomorphous replacement would be difficult unless care was taken to minimize the systematic errors introduced by differences between the successive protein preparations. Consequently, for each stage in the heavy atom derivative data collection, a corresponding native data set was recorded from the same protein preparation. For each purification trial, approximately 700 mg of myosin S1 was prepared and set up for crystallization. Many attempts were made before a single preparation yielded sufficient crystals for x-ray data collection.

The structure was determined by a combination of multiple isomorphous replacement and solvent flattening. The first derivative solved was obtained from crystals soaked in trimethyllead acetate and proved to be highly isomorphous with only four binding sites (*21*). It was used to determine the positions of the other heavy atom binding sites by difference Fourier techniques (Table 2). The positions and occupancies of the heavy atom sites were refined according to the origin-removed Patterson-function correlation method by the program HEAVY (*22*). The overall figures of merit for the area detector, CHESS, and SSRL synchrotron data were 0.47, 0.58, and 0.42, respectively.

The higher resolution x-ray data collected at SSRL were placed on the same scale as the area detector data and included as a block from 4.5 to 2.8 Å. Efforts to merge the overlapping data between the area detector and synchrotron data were unsatisfactory. However, the phase information from all three sources was combined throughout the entire resolution range via the phase probability coefficients (*23*).

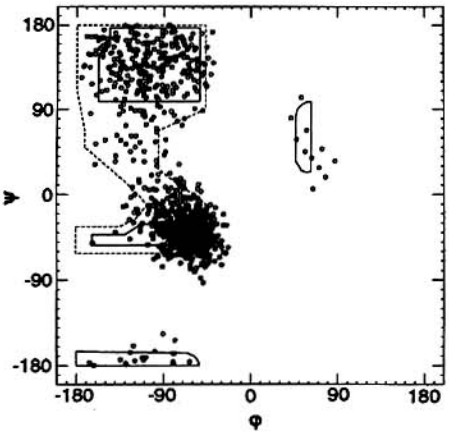

Fig. 1. Ramachandran plot of the main chain dihedral angles of all non-glycinyl residues in the model presented.

These protein phases were improved by solvent flattening (*24*). The positions and occupancies of heavy atom binding sites were further refined against these modified phases (*25*). This gave an improved electron density map into which approximately 550 alanine residues were built with the program FRODO (*26*). The map showed good connectivity and many well-defined side chains.

Once several long segments were connected, the positions of these alanine residues were matched to the known amino acid sequence (*27, 28*). At this stage phase information from the partial model was combined with the heavy atom derivative phases by the program SIGMAA (*29*). The structure was refined concurrently with the model building process by the program package TNT (*30*). Once the model building was near completion, a cycle of refinement with X-PLOR (*31*) was performed to improve the conformations of the side chains. The strategy of alternate model building and refining proved successful and constantly improved the estimation of the protein phases. Toward the end of the analysis there were clear segments of electron density corresponding to portions of the light chains that were completely missing in the original maps phased with heavy atom derivatives alone.

At present, 1072 residues (of a total of 1157) have been built into the electron density map. The model was refined to an *R* factor of 22.3 percent for all measured x-ray data between 30 to 2.8 Å with root-mean-square deviations from ideal geometry of 0.018 Å for bond lengths, 2.5° for bond angles, and 0.013 Å for groups of atoms expected to be coplanar. No solvent molecules have yet been built into the electron density (Figs. 1 and 2).

Structure description. In a space-filling representation of all atoms in the myosin S1 model (Fig. 3), the green, red, and blue segments represent parts of the heavy chain and the yellow and magenta stretches cor-

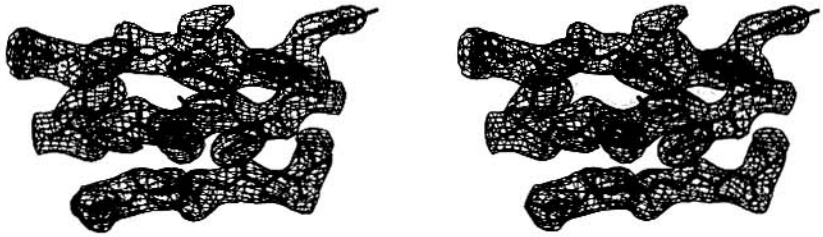

Fig. 2. A stereo view of a representative section of electron density located in the seven-stranded β sheet motif of the heavy chain calculated with SIGMAA coefficients (*29*). The phases and weights used to calculate the electron density were obtained by combining the information from the heavy atom phases and those derived from the atomic model.

Table 2. Heavy atom derivatives used in the structure determination and their data collection statistics.

Derivative	Conditions*		Method	R_{sym}†	Reflections (no.)	Resolution (Å)	R_{scale}‡	Sites (no.)	Phasing power§
	Concentration (mM)	Time (days)							
Trimethyllead acetate	20	21	Area detector	6.7	13,394	3.5	24.5	4	1.01
$KAu(CN)_2$	1	5	Area detector	5.8	18,082	3.5	18.2	6	1.14
K_2OsO_4/pyridine	2–20	2	Area detector	6.7	11,804	4.0	24.5	4	1.02
$K_3UO_2F_5$	3	4	Area detector	6.7	11,688	4.0	19.1	7	1.06
Trimethyllead acetate	20	21	CHESS	11.5	34,498	2.8	28.2	4	1.19
$KAu(CN)_2$	1	5	CHESS	9.7	36,339	2.8	17.0	8	1.23
K_2OsO_4/pyridine	2–20	2	CHESS	10.0	32,554	2.8	32.2	8	0.99
Cis-$Pt(NH_3)_2Cl_2$	2	3	CHESS	10.3	36,108	2.8	21.0	12	1.12
$K_3UO_2F_5$	3	4	CHESS	10.9	36,419	2.8	22.7	9	0.98
Trimethyllead acetate	15	21	SSRL	13.0	31,667	2.8	27.1	4	1.11
$K_3UO_2F_5$	2	3	SSRL	11.8	33,043	2.8	22.2	6	0.88

*The heavy atom derivatives were prepared at 4°C by first slowly transferring the crystals to a synthetic mother liquor composed of 1.5 M ammonium sulfate, 500 mM KCl buffered with 20 mM Pipes at pH 6.7. †$R_{sym} = \Sigma\Sigma[|I_{hi}| - |I_h|]/\Sigma_{ni}I_{hi} \times 100$, where I_{hi} and I_h are the intensities of the individual and mean structure factors. ‡$R_{scale} = \Sigma[|F_h| - |F_n|]/\Sigma_n F_n \times 100$, where F_h and F_n are the heavy atom and native structure factors. §The phasing power is defined as the mean value of the heavy atom structure factor divided by the residual lack-of-closure error.

respond to the essential and regulatory light chains, respectively. As can be seen, the myosin head is highly asymmetric with a length of 165 Å, a width of 65 Å, and a thickness of approximately 40 Å.

Previous knowledge of the organization of the heavy chain in the myosin head was derived from proteolytic studies. Limited tryptic digestion of vertebrate skeletal S1 indicated that the head contained three major regions: a 25-kD NH_2-terminal nucleotide binding region (*32*), a central 50-kD segment, and a 20-kD COOH-terminal segment; the last two were shown to bind to actin (*33, 34*). These proteolytic segments are displayed in green, red, and blue, respectively (Fig. 3); the light chains abut one another and are wrapped around a single α helix of the heavy chain but do not overlap to any significant extent.

The secondary structure of the myosin head is dominated by α helices with approximately 48 percent of the amino acid residues in this conformation (Figs. 4 and 5). One key structural feature is the long (approximately 85 Å) α helix which extends from the thick part of the head down to the COOH-terminus of the heavy chain. This α helix constitutes the light chain binding region of the heavy chain. There is a bend, delineated by amino acid residues Trp^{829}, Pro^{830}, Trp^{831}, and Met^{832}, which connects this long α helix to a short COOH-terminal α helix of the 95-kD heavy chain fragment. A brief description of the three polypeptide chains constituting the myosin head is given below.

The regulatory light chain is located at the end of the molecule distal from the nucleotide binding site (Figs. 4 and 5). It consists of two domains and shares considerable structural homology with calmodulin and troponin C except that the long connecting helix observed in calmodulin and troponin C is distorted (*35, 36*). A comparison of the regulatory light chain with calmodulin is shown in Fig. 6A where the eight helices that comprise the two domains have been labeled A through H. The regulatory light chain is arranged such that its NH_2-terminal domain wraps around the COOH-terminus of the heavy chain between amino acid residues Asn^{825} and Leu^{842} whereas its COOH-terminal domain interacts with the heavy chain in the region defined by amino acid residues Glu^{808} to Val^{826}. The interaction of the NH_2-terminal domain with the heavy chain is stabilized by a cluster of hydrophobic residues including nine phenylalanines, two tryp-

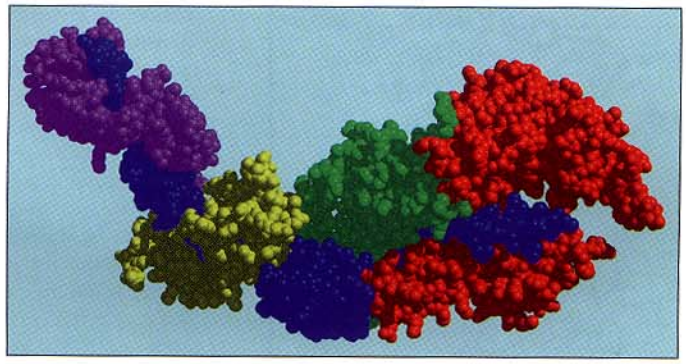

Fig. 3. A space-filling representation of all of the atoms in the current model of myosin S1. The model is oriented such that the actin binding surface is located at the lower right-hand corner. The 25-, 50-, and 20-kD segments of the heavy chain are colored in green, red, and blue, respectively, whereas the essential and regulatory light chains are shown in yellow and magenta, respectively. In this orientation the prominent horizontal cleft that divides the central 50-kD segment of the heavy chain into two domains (upper and lower defined by this orientation) is clearly visible. This figure was prepared with the molecular graphics program MIDAS (*61*).

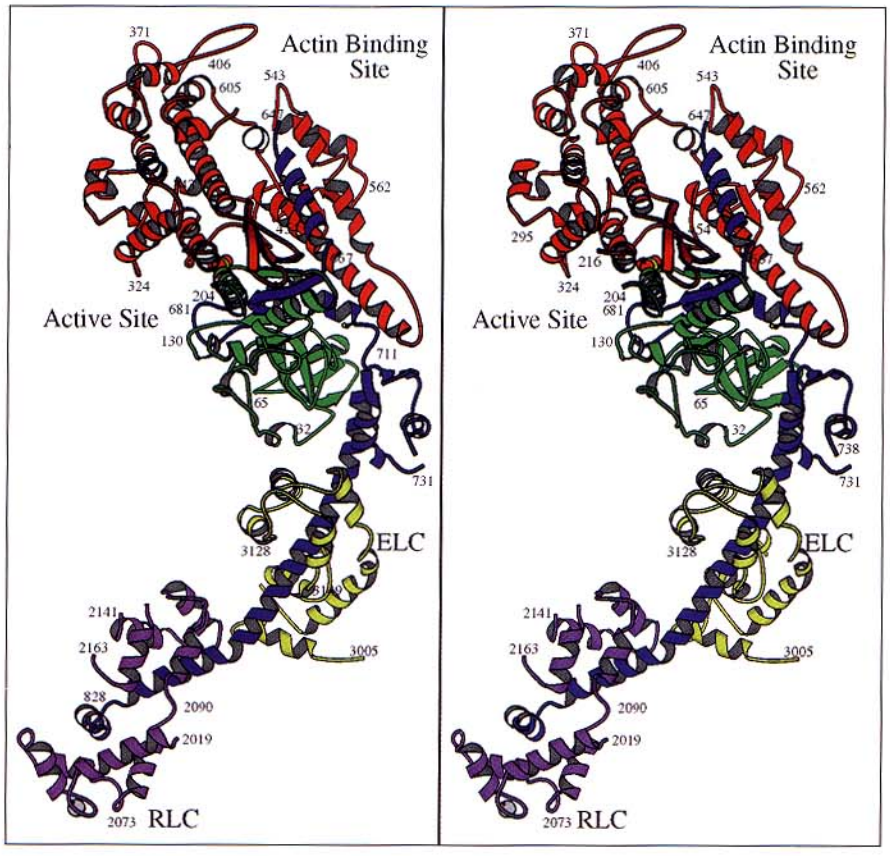

Fig. 4. A ribbon representation of the entire model for myosin S1. In this and all successive figures, 2000 and 3000 have been added to the residue numbers of the regulatory and essential light chains, respectively, to distinguish these from the heavy chain. Heavy chain residues Asp^4 to Glu^{204}, Gly^{216} to Tyr^{626}, and Gln^{647} to Lys^{843} are colored in green, red, and blue, respectively. These segments are separated by disordered loops for which no density is evident in the current map. There are two additional segments in the heavy chain for which the density is weak or disordered. These include residues Lys^{572} to Lys^{574} and Ile^{732} to Phe^{737}. The A2 isozyme of the essential light chain, shown in yellow, theoretically contains 149 amino acid residues. In the model it extends from residue Asp^5 to Val^{149} and contains one ill-defined region that includes residues Leu^{50} to Ala^{60}. The regulatory light chain, which is colored in magenta, theoretically consists of 166 amino acid residues. In the current model it extends from residue Phe^{19} to Lys^{163} but is disordered between residues Pro^{142} and Asn^{147}. In this figure the molecule is oriented perpendicular to its long axis and rotated to view along the active site pocket. A sulfate ion, shown here in a space-filling representation, is located at the base of the pocket. The actin binding surface has been defined as indicated on the figure by the location of the 50- to 20-kD junction (residues Tyr^{626} and Gln^{647}) and by its interaction with actin (*46*). Figures 4 to 7 were prepared with the molecular graphics program MOLSCRIPT (*62*).

tophans, and four methionines. Five of these residues are contributed by the heavy chain. Superposition of the NH_2-terminal domains of the regulatory light chain and calmodulin reveals an rms difference in the positions of 59 equivalent residues of only 1.3 Å. By contrast the COOH-terminal domain is less similar to the structure observed in calmodulin and is due to a difference in the positions of the F and G helices in the regulatory light chain that have moved to accommodate the heavy chain. In addition, the COOH-terminal domain as a whole has rotated, relative to calmodulin, about the midpoint between the two domains in order to form a tight complex with the heavy chain.

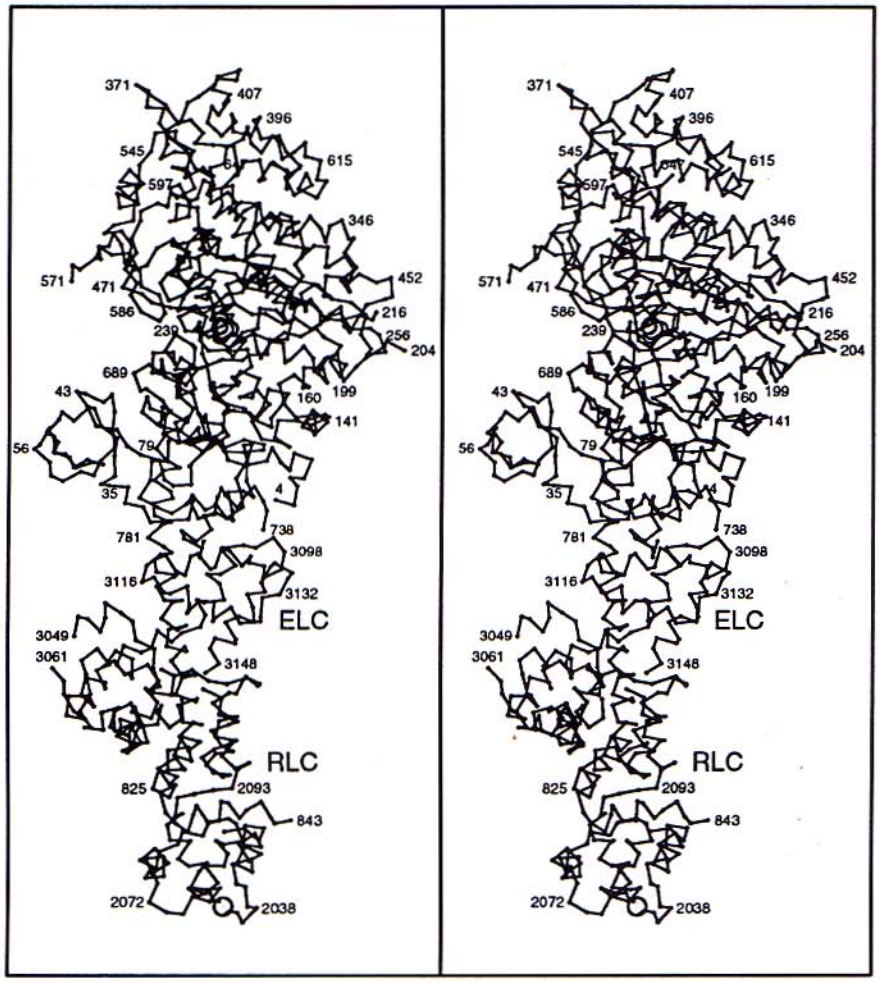

Fig. 5. A stereo α carbon plot of the entire myosin head in which the view has been rotated 90° with respect to Fig. 4. In this view, the active site pocket is seen as a wide depression. Selected residues have been labeled to allow the path of the chain to be followed and to identify the start and end of the secondary structural elements.

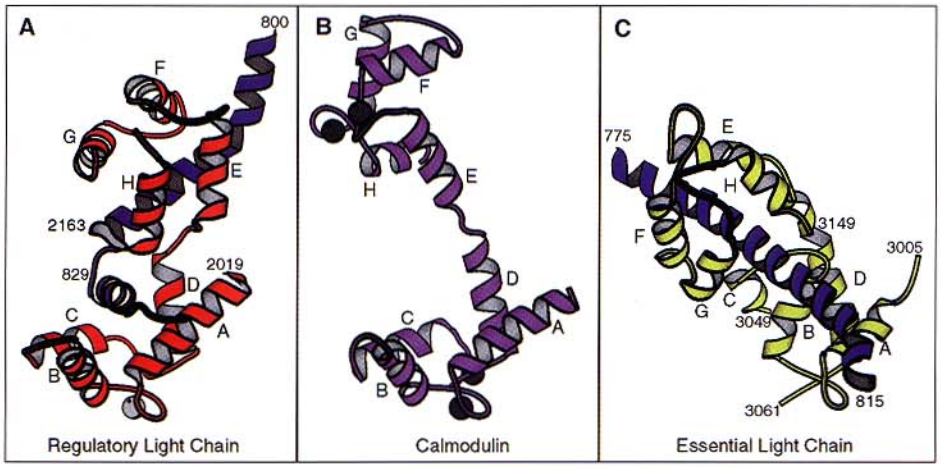

Fig. 6. (**A**) and (**C**) show ribbon representations of the regulatory and essential light chains together with the segment of the heavy chain with which they interact. The light chains are oriented such that the NH_2-terminal domains have the same orientation as calmodulin shown in (**B**). The coordinates for calmodulin were taken from the Brookhaven Protein Data Bank (file 3CLN) from the structure determined by Cook and co-workers (*63*).

The divalent cation binding site is located in the first helix-loop-helix motif observed in the amino acid sequence and, as indicated above, has a conformation similar to that observed in calmodulin. A divalent cation is clearly evident in the electron density and is most likely Mg^{2+} in that this was a minor constituent of the crystallization buffer. In our model, no electron density was observed for the first 18 amino acid residues in the regulatory light chain. This includes Ser^{13} and is, by sequence homology to rabbit myosin, the site of phosphorylation by myosin light chain kinase (*37*). Presumably this portion of the polypeptide chain is flexible in S1 and perhaps only plays a functional role when the head is attached to the remainder of the molecule. The observed NH_2- and COOH-terminal residues of the regulatory light chain lie close to the interface between the two domains.

The essential light chain interacts with the long α helix of the heavy chain through amino acid residues Leu^{783} to Met^{806} (Fig. 6C). Likewise, it wraps around the heavy chain α helix but in a manner different from that observed for the regulatory light chain. Its arrangement resembles that for the interaction of calmodulin with a target peptide from myosin light chain kinase (*38*). It differs in that the second and third helices in the NH_2-terminal domain abut the heavy chain with their external surfaces, whereas the corresponding secondary structural elements in calmodulin enclose the respective target peptide. The electron density for this part of the molecule is the least well ordered of the entire map. Indeed, very little of the essential light chain was visible in the original electron density map and only appeared after the phase information from the rest of the molecule was included. This could be due to either lack of isomorphism in the heavy atom derivative phases or conformational flexibility of the polypeptide chain. It is difficult to distinguish between these two possibilities because the crystals contain both classes of essential light chain isoforms. As with the regulatory light chain, the NH_2- and COOH-terminal residues lie close to the interface between the two domains.

The heavy chain constitutes the entire thick portion of the myosin head and con-

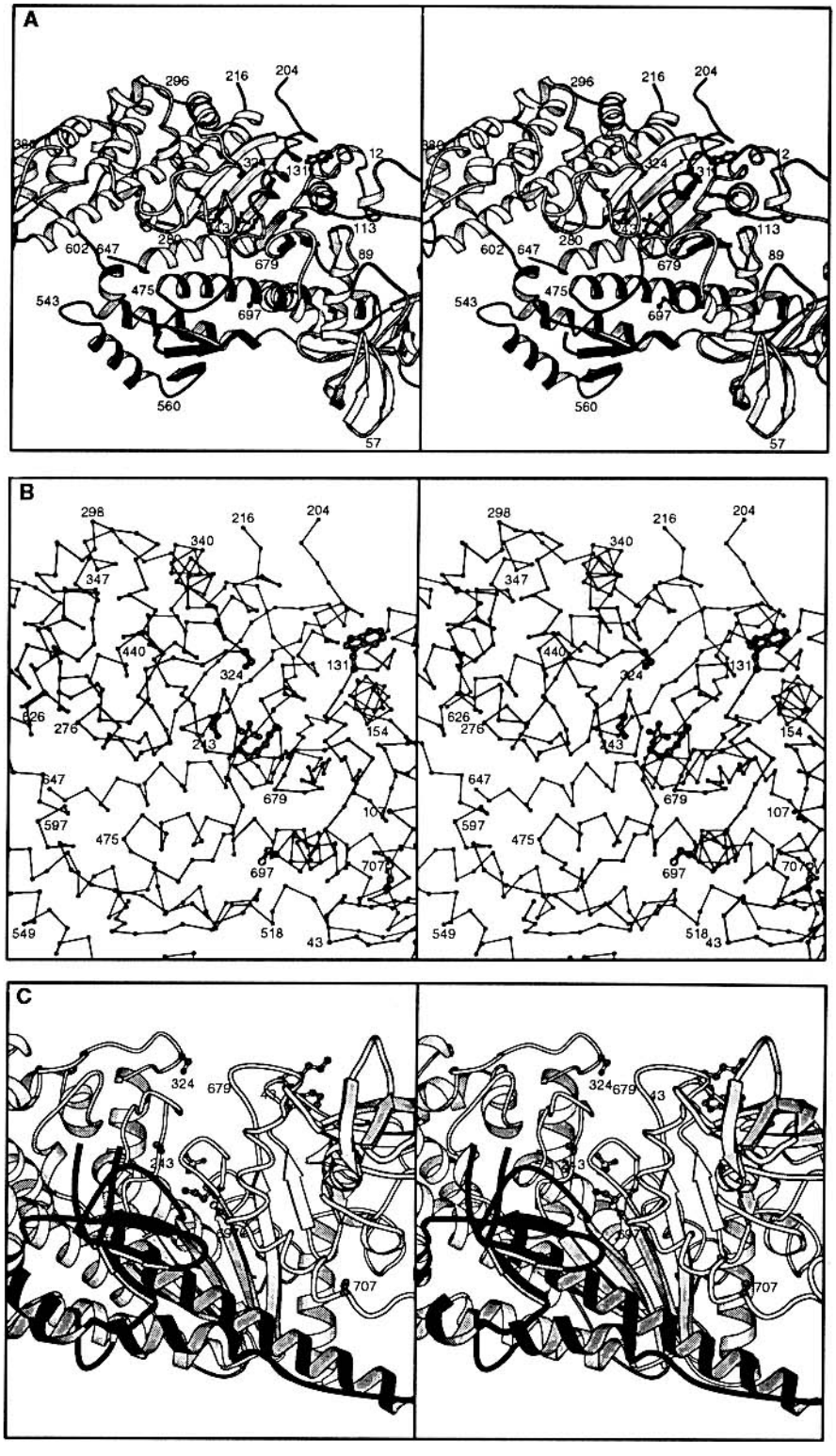

Fig. 7. A stereo ribbon and α carbon plots of the catalytic portion of the myosin head centered on the active site. In (A) and (B) the actin binding face, as defined by the position of the 50- to 20-kD junction, is located on the far side of the molecule. In (C) the molecule has been rotated 90° about the horizontal axis to reveal more clearly the relation between the active site pocket and the reactive cysteine residues. (**A**) A larger segment of the myosin head that reveals the overall disposition of the secondary structural elements around the nucleotide binding site. The upper domain of the 50-kD segment is shaded in gray whereas the lower domain is shaded in black to emphasize the narrow cleft that divides them. (**B**) A more detailed view of the residues that form the interface between the upper and lower domains of the 50-kD segment. Marker residues are identified that allow all other residues to be located. In addition, a few of the side chains for the residues that have been implicated to be important in the catalytic mechanism, from amino acid sequence analyses and from chemical studies, have been included. Residues Trp^{131} and Ser^{324} that have been identified from photolabeling studies lie on opposite sides of the nucleotide binding pocket. (**C**) The helix connecting the reactive cysteines, Cys^{707} and Cys^{697}, lies at the base of a cleft at the junction between the lower domain of the 50-kD segment and the NH_2-terminal 25-kD segment.

tains both the nucleotide binding site and actin binding region. These are located on opposite sides of the protein. This part of the molecule contains a complex arrangement of secondary structural elements centered mainly around a large, mostly parallel, seven-stranded β sheet motif. The topology of this β sheet is such that strands one and six run in the opposite direction to the other five strands. The central strand corresponds to the strand-loop-helix binding motif, which has the sequence GESGAGKT (*39*), observed both in adenylate kinase and the Ras protein (*40*). The topology and organization of the heavy chain are described below in terms of the three major tryptic fragments. However, these fragments arise from proteolytic cleavage at flexible loops and do not represent discrete structural domains.

The first observed residue at the NH_2-terminus of the heavy chain is Asp^4 and is located close to the essential light chain at the approximate center of the entire myosin molecule (Figs. 4 and 5). From here the heavy chain crosses the width of the molecule and forms a small six-stranded antiparallel β sheet motif (Lys^{35} to Met^{80}), which is fairly independent of the rest of the head and protrudes from the molecule as a whole. The function of this domain is unknown although it does not appear essential for motility in that it is missing in several single-headed myosin I–type molecules (*41*). The topology of this sheet is similar to that of the Src–homology 3 domain observed in spectrin (*42*). After this motif, the heavy chain forms three strands of the large β sheet motif that are connected by a series of α helices. The first two strands extend from Tyr^{116} to Tyr^{118} and from Cys^{123} to Val^{126} and are connected by a β turn. Thereafter there are three short helices prior to the fourth β strand in the sheet that extends from Gln^{173} to Gly^{179}. The third strand belongs to the COOH-terminal 20-kD fragment of the heavy chain fragment. The fourth or central strand precedes the phosphate binding loop and is followed by a helix, Lys^{185} to Ile^{199}, which forms the base of the nucleotide binding pocket. The topology of this loop is essentially identical to that in the Ras protein and adenylate kinase (*40*). A sulfate ion is embedded in the phosphate binding loop and is located close to the position of the β phosphate observed in the complex between Ap_5A [P^1,P^5,bis-(adenosine-5′-) pentaphosphate] and adenyl-

ate kinase (Figs. 4 and 5). It is perhaps not surprising to find a sulfate ion in the nucleotide binding site because ammonium sulfate is a competitive inhibitor of the ATPase activity (*43*). A break in the electron density is observed between Glu^{204} and Gly^{216} at the far end of the active site pocket. The missing segment, which contains six charged residues, occurs at the 25- to 50-kD junction and is most likely a constitutively flexible loop.

The 50-kD fragment has a complex topology that can be described as two major domains separated by a long narrow cleft as is evident in the space-filling drawing (Fig. 3). This cleft divides the distal one-third of the myosin head into two regions, which are referred to as the upper and lower domains of the 50-kD segment (Fig. 3).

Electron density for the polypeptide chain resumes at Gly^{216} as the start of an α helix (Leu^{218} to Gly^{233}). This helix forms part of the nucleotide binding pocket. Thereafter, the chain loops around close to the phosphate binding site and connects up to β strands six and seven of the large β sheet motif that extend from Gly^{247} to His^{254} and Leu^{260} to Tyr^{268}, respectively. Strand seven terminates in a domain composed of random coil and several short helices and extending from Glu^{271} to Asp^{327}. This region is located close to the nucleotide binding site and contains Ser^{324} which had been previously identified by photolabeling to be an active site residue (*44*) (Fig. 7). An α helix extending from Asp^{327} to Ile^{340} forms the top of the nucleotide binding pocket. After this domain, the polypeptide chain forms the end of the myosin head through a series of long α helices. The longest of these is 45 Å in length and extends from Val^{419} to Leu^{449}. Strand five of the large mixed β sheet follows this helix and extends from Tyr^{457} to Ala^{465}. This strand terminates in a random coil that drops from the "upper" to "lower" domains of the 50-kD fragment. The midpoint between the upper and lower domains is located close to Gly^{466} and occurs in a region of the sequence (Tyr^{457} to Gly^{516}) that is highly conserved in all myosins (*45*). Furthermore, the cleft itself contains many individual highly conserved residues that extend into the space between the two domains.

The lower domain is built from several long α helices (Phe^{475} to Lys 505 and Met^{517} to Glu^{539}), the last of which contains a hydrophobic bulge at Pro^{529}. After another helix (Asp^{547} to His^{558}) there is a three-stranded antiparallel β sheet, which includes residues Asn^{564} to Lys^{567}, Phe^{579} to Val^{582}, and Thr^{587} to Tyr^{590}. The electron density for Lys^{572}, Gly^{573}, and Lys^{574} is very weak, and therefore these residues have been excluded from the model. The segment between Pro^{529} and Lys^{553} is one component of the actin binding surface as defined by Rayment *et al.* (*46*). A single segment of random coil (Lys^{600} to Leu^{603}) passes from the lower domain and across the cleft to form a helix-loop-helix motif on the outer face of the upper 50-kD domain and terminating at Tyr^{626}. There is no electron density corresponding to amino acid residues Gly^{627} to Phe^{646}. This particular stretch contains the second major site of trypsin proteolysis and is the junction between the 50- and 20-kD fragments. The primary sequence in this disordered region contains nine glycine and five lysine residues, suggesting that it may be a flexible region in the molecule. This site is resistant to proteolysis in the actomyosin complex and as such may contribute to the actin binding interface of myosin (*33*). In addition, this region has also been implicated in actin binding from crosslinking and kinetic studies of proteolytically cleaved protein (*34, 47*).

Electron density for the polypeptide chain resumes at Gln^{647} and proceeds as a long α helix (Ser^{650} to Arg^{665}) across the flat face of the molecule toward the light chain binding region and lies between the upper and lower domains of the 50-kD fragment. This helix is part of a highly conserved segment that runs from Leu^{658} to Asn^{678}. At the end of the helix, the polypeptide chain turns into the center of the molecule and forms the third strand of the mixed β sheet (His^{668} to Ile^{675}). Thus the major tertiary motif of the head contains contributions from all three of the tryptic fragments. After leaving the β sheet, the polypeptide chain proceeds through the large surface loop defined by Thr^{667} to Glu^{687}, which caps one end of the nucleotide binding site pocket. Subsequently, the polypeptide chain forms two α helices lying under the nucleotide binding site and delineated by His^{688} to Asn^{698} and Val^{700} to Arg^{708}. This highly conserved segment in the sequence contains the two sulfhydryl groups, Cys^{707} and Cys^{697}, which are more reactive than the other 11 in the molecule and have been given the names SH1 and SH2, respectively, in the order of their chemical reactivity. These two thiols can be crosslinked by oxidation and a wide variety of bifunctional chemical reagents differing in length from 14 to 3 Å but only in the presence of nucleotide (*48*). Indeed, formation of a covalent link between these two groups serves to trap Mg^{2+}-ADP (adenosine diphosphate) in the active site. Although these reactive sulfhydryls have been thought to reside in a flexible loop, the discovery that these two residues are separated by an α helix was surprising (Fig. 7C). This is a well-defined region of the electron density map. The fact that the α carbons of Cys^{697} and Cys^{707} are approximately 18 Å apart suggests that a rearrangement or conformational change in this area must occur upon nucleotide binding. This point is further emphasized by the observation that SH1 and SH2 both lie in small clefts that face out toward the solvent on opposite sides of the molecule. The functional significance of this region is also indicated by the very high degree of amino acid sequence conservation in this area of the molecule.

The segment that follows the reactive sulfhydryl groups consists of a small three-stranded antiparallel β sheet that includes residues Arg^{714} to Tyr^{717}, Tyr^{758} to Gly^{761}, and Lys^{764} to Phe^{767}, and is associated with two short helices. This domain is separated from the adjacent NH_2-terminal domain of the 25-kD fragment of the heavy chain by a distinct cleft and shows a greater association with the COOH-terminal domain of the essential light chain. Thereafter the heavy chain continues as a long α helix that shows distinct curvature beginning at Leu^{771} and ending at Val^{826}. There is a decided bend in the course of the polypeptide chain resulting from the Trp^{829}, Pro^{830}, Trp^{831} sequence (Figs. 4 and 5). The heavy chain terminates at residue Lys^{843} after a small α helix that lies nearly at right angles to the preceding long helix.

Effect of the reductive methylation on the protein structure and function. One question that must be addressed is the effect, if any, of reductive methylation on the conformation of the protein. An examination of the kinetic properties of modified myosin S1 reveals that the protein is enzymatically active (*49*). There are changes in the kinetic parameters that are similar to those observed when only the reactive sulfhydryl groups are alkylated (*50*). The results do not suggest any major changes in the overall conformation of the molecule since these would be expected to abolish its enzymatic activity. Myosin from most sources already contains several post-translationally modified amino acid residues. For example, in chicken skeletal myosin S1, Lys^{35} is monomethylated, Lys^{130} and Lys^{551} are trimethylated, and His^{757} contains a 3-*N*-methylated side chain (*27*). Although the role of these modified residues is unknown, it has been suggested that methylation of Lys^{130} provides a permanent positive charge that may become buried when nucleotide is bound (*51*). In our structure, Lys^{130} is exposed to the solvent at the edge of the nucleotide binding pocket. However, this region of the protein probably rearranges when nucleotide binds because the adjacent Trp^{131} is photolabeled by two purine ATP analogues (*52*).

The structure of methylated lysozyme is essentially identical to that of the native protein (*15*). From this it is not expected that the folding motifs in myosin S1 will be significantly affected by this treatment. There is, however, the possibility that the relation between the various domains could be altered. Our data reveal that almost all of

the lysine residues are located at the surface and hence would not be expected to influence the structure in any major way. The A2 isozyme of chicken myosin S1 contains 102 lysine residues (*27*), of which 85 have been built into the model. Of those, 67 are located at the surface of the protein and only 18 participate in salt bridges and can be considered buried. Five of these lysines participate in crystalline contacts. The remaining 17 lysine residues in the A2 isozyme are located in disordered loops; in our structure all except 4 lysine residues are reproducibly modified under conditions where 100 percent dimethylation is expected (Table 1) (*15*). Thus, the unmodified lysine residues are most likely located in salt bridges where they would be expected to have a higher pK_a. Mass for the additional methyl groups on the lysine residues is evident in the electron density map for most of the well-ordered side chains. However at this resolution it difficult to categorically decide if a residue has been modified based on the density alone. Even so, it appears that Lys^{185}, which resides in the phosphate binding loop, is not modified.

Active site and possible mechanism for muscle contraction. The catalytic site of the myosin head was identified by analogy to the phosphate binding loop in both the Ras protein and adenylate kinase and by the position of the amino acid residues previously identified by chemical studies with ATP analogues (*52*). The nucleotide binding pocket is located on the opposite side of the head from the proposed actin binding site and is in an open conformation (Figs. 4, 5, and 7). The view in Fig. 7 shows the position of the sulfate ion in the phosphate binding loop and a few of the amino acid residues that have been chemically labeled, including Trp^{131}, Ser^{181}, Ser^{243}, and Ser^{324} (*44*, *52*, *53*). The width of the nucleotide binding pocket at its surface is approximately 15 Å as measured between α carbons. Since the binding constant of myosin for Mg^{2+}-ATP is about 3×10^{11} (*54*) and residues on both sides of the cleft have been photochemically labeled, it is likely that the pocket closes when nucleotides bind in the active site. The pocket is approximately 13 Å wide and 13 Å deep with an angle between the faces of the pocket of ~40°. The base of the cleft is located 90 Å from the COOH-terminus of the myosin head. If the binding face to actin remains essentially stationary, closure of the nucleotide binding cleft could produce a movement at the COOH-terminus of the myosin head of approximately 60 Å. How this rearrangement is actually accomplished cannot be easily predicted from our structure.

The orientation of the molecule in Fig. 5 is rotated such that the actin binding surface is approximately perpendicular to the page (*46*). Closure of the nucleotide binding pocket would rotate the COOH-terminal end of the heavy chain that carries the light chains toward the viewer, which is consistent with that expected for the start of the power stroke. From this perspective it appears that a major function of the light chains is to create a longer molecule and hence amplify the conformational changes associated with the active site.

Muscle contraction consists of the cyclic attachment and detachment of the myosin head to the actin filament with the concomitant hydrolysis of ATP. From the extensive kinetic studies on the interaction of myosin with actin (*55*), a general picture of the sequence of kinetic events occurring during muscle contraction has emerged.

Transient kinetic measurements originally demonstrated that transduction of the chemical energy released by the hydrolysis of ATP into directed mechanical force occurred during product release rather than during the hydrolysis step itself (*56*). The cycle of events was summarized as follows: Mg^{2+}-ATP rapidly dissociates the actomyosin complex by binding to the ATPase sites of myosin; free myosin then hydrolyzes ATP and forms a relatively stable myosin-products complex; actin recombines with this complex and dissociates the products, thereby forming the original actin-myosin complex. Presumably, force is generated during the last step. Although this model provided an important conceptual framework for studies of the contractile cycle, it soon became clear that the interactions between myosin, actin, and the substrate and products were more complex (*55*).

Structural information on the conformational changes that occur during the actomyosin interactions is limited. Addition of ATP causes no significant change in the amount of secondary structure as assessed by circular dichroism (*57*). The changes observed in tryptophan fluorescence are typical of most enzymes whose active sites are induced to fit around their substrates. However, significant movement within the myosin head must occur during the ATPase activity because of the large change in distance between the two reactive cysteine residues (Cys^{707} and Cys^{697}) that is induced when nucleotide binds (*48*, *58*). Recent low-angle x-ray scattering studies also suggest a large-scale movement during ATP hydrolysis (*59*).

In formulating a model for muscle contraction from the structure of myosin S1 presented here, it must be understood that it neither contains nucleotide nor is bound to actin. Most likely the crystal structure is an intermediate between these two extremes, although probably closer to the actin bound state. Preliminary attempts to dock myosin (*46*) to actin suggest that a better fit to the image reconstructions of S1-decorated actin would be obtained if the long narrow cleft between the upper and lower 50-kD domains were to close, thus implying that this is an important structural feature of the molecule. In addition, the preliminary fit implies that the actin binding site contains components from both the upper and lower 50-kD domains and the first α helix from the 20-kD region. From the location of residues Tyr^{626} and Gln^{647}, the positively charged disordered segment at the 50- to 20-kD junction could readily interact with the negatively charged amino acids at the NH_2-terminus of actin.

All the current kinetic models for the mechanism of muscle contraction require a change in the binding affinity of myosin for actin when ATP binds to the active site. Although it is difficult to predict how this effect can be communicated to the actin binding site, the structure suggests that this might be generated by changes in the relation between the upper and lower domains of the 50-kD segment prompted by binding of the γ phosphate. Examination of Fig. 7 reveals that the potential binding site for the γ phosphate would be located close to the confluence of the upper and lower domains of the 50-kD region below the current location of the sulfate ion. These observations together with the information from docking myosin onto actin provide the information necessary to formulate a basic structural model for muscle contraction (*46*).

The three-dimensional model of the myosin S1 presented in this article provides a molecular framework that can be used to address the issues of conformational changes during the contractile cycle and suggests how this molecule functions as a molecular motor. By a combination of molecular biology, in vitro motility assays, and chemical and kinetic studies, it should be possible to test these hypotheses concerning the molecular basis of motility.

Finally, it is appropriate to consider why reductive methylation allows this molecule to crystallize. Examination of the structure reveals that it contains elements of flexibility that might lead to multiple conformations in solution, which in turn might prevent the formation of a crystalline lattice. It is conceivable that reductive methylation serves to stabilize one of these conformations in solution. Alternatively, reductive methylation may serve only to reduce the solubility of the protein.

REFERENCES AND NOTES

1. R. D. Vale and L. S. B. Goldstein, *Cell* **60**, 883 (1990).
2. A. F. Huxley and R. Niedergerke, *Nature* **173**, 971 (1954); H. Huxley and J. Hanson, *ibid.* p. 973.
3. S. Lowey, H. S. Slayter, A. G. Weeds, H. Baker, *J. Mol. Biol.* **42**, 1 (1969).
4. A. G. Weeds and S. Lowey, *ibid.* **61**, 701 (1971).
5. Y. Y. Toyoshima *et al.*, *Nature* **328**, 536 (1987).
6. S. S. Margossian, W. F. Stafford III, S. Lowey, *Biochemistry* **20**, 2151 (1981).
7. P. D. Wagner and E. Giniger, *Nature* **292**, 560 (1981).

8. S. Citi and J. Kendrick-Jones, *BioEssays* **7**, 155 (1987).
9. J. S. Sellers, *Curr. Opin. Cell Biol.* 1, 98 (1991).
10. J. H. Collins, *J. Muscle Res. Cell Motil.* **12**, 3 (1991).
11. S. Lowey, in *Myology*, A. G. Engel and B. Q. Banker, Eds. (McGraw-Hill, New York, 1986), vol. 19, pp. 563–586; P. Vibert and C. Cohen, *J. Muscle Res. Cell Motil.* **9**, 296 (1988).
12. A. Elliott and G. Offer, *J. Mol. Biol.* **123**, 505 (1978).
13. D. A. Winkelmann, H. Mekeel, I. Rayment, *ibid.* **181**, 487 (1985); D. A. Winkelmann, T. S. Baker, I. Rayment, *J. Cell Biol.* **114**, 701 (1991).
14. R. H. Rice and G. E. Means, *J. Biol. Chem.* **246**, 831 (1971).
15. W. R. Rypniewski, H. M. Holden, I. Rayment, *Biochemistry*, in press.
16. L. Silberstein and S. Lowey, *J. Mol. Biol.* **148**, 153 (1981).
17. Y. Nabeshima, Y. Fujii-Kuriyama, M. Muramatsu, K. Ogata, *Nature* **308**, 333 (1984).
18. R. Smith, W. R. Rypniewski, I. Rayment, in preparation.
19. I. Rayment and D. A. Winkelmann, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4378 (1984).
20. The data to 3.5 Å resolution were recorded at 4°C on a Siemens X1000D area detector; 13 crystals were used to collect the native data. Most of these crystals were translated to expose a new region after 10 to 11 hours in the x-ray beam such that the data were collected from 35 segments. A total of 94,265 reflections were measured, which reduced to 21,370 unique reflections (Theoretical 24,556) with an R_{merge} of 5.3 where

$$R_{merge} = \Sigma\Sigma \{|I_{hi}| - |I_h|\}/\Sigma_m I_{hi} \times 100$$

I_{hi} and I_h are the intensities of the individual and mean structure factors, respectively. These data were recorded with the goal of obtaining a complete, accurate low-resolution x-ray data set that could be used to determine the positions of the heavy atoms in the derivatives. The frame data were processed by the program XDS (*64*) and scaled with the Fox and Holmes algorithm as implemented by P. Evans in the programs Rotavata and Agrovata (*65*). The x-ray data between 3.5 and 2.8 Å were recorded on film at the synchrotron sources located at Cornell (CHESS) and Stanford (SSRL). Data were processed with the software developed by M. Rossmann, modified to operate on a VAX (*66*). The data were merged and scaled with the same software used for the area detector data, except for the inclusion of post-refinement to utilize the partial data (*66*). A total of 178,986 measurements were recorded on 97 films to yield 36,781 independent reflections in the SSRL native data set with an R_{merge} of 9.2. The final native data set consisted of structure factors from 100 to 4.5 Å recorded on the area detector and data from 4.5 to 2.8 Å recorded at SSRL.
21. H. M. Holden and I. Rayment, *Arch. Biochem. Biophys.* **291**, 187 (1991).
22. M. G. Rossmann, *Acta Crystallogr.* **13**, 221 (1960); T. C. Terwilliger and D. Eisenberg, *Acta Crystallogr. Sect. A* **39**, 813 (1983).
23. W. A. Hendrickson and E. E. Lattman, *ibid. Acta Crystallogr. Sect. B* **26**, 136 (1970).
24. B. C. Wang, *Methods Enzymol.* **115**, 90 (1985). The algorithm was written by W. Kabsch (Heidelberg, Germany).
25. M. A. Rould, J. J. Perona, D. Söll, T. A. Steitz, *Science* **246**, 1135 (1989).
26. T. A. Jones, *Methods Enzymol.* **115**, 157 (1985).
27. T. Maita *et al.*, *J. Biochem.* **110**, 75 (1991); G. Matsuda, *Adv. Biophys.* **16**, 185 (1983).
28. The size of the side chains observed in the electron density map was matched to the sequence with the program FITSEQ, available from I. Rayment on request.
29. R. J. Read, *Acta Crystallogr. Sect. A* **42**, 140 (1986).
30. D. E. Tronrud, L. F. Ten Eyck, B. W. Matthew, *ibid.* **43**, 489 (1987).
31. A. T. Brunger, *X-PLOR Manual Version 2.1* (Yale University, New Haven, CT, 1990).
32. L. Szilagyi, M. Balint, F. A. Sreter, J. Gergley, *J. Biochem. Biophys. Res. Commun.* **87**, 936 (1979).
33. D. Mornet, R. Bertrand, P. Pantel, E. Audemard, R. Kassab, *Biochemistry* **20**, 2110 (1981).
34. K. Sutoh, *ibid.* **21**, 4800 (1982).
35. Y. S. Babu *et al.*, *Nature* **315**, 37 (1985).
36. O. Herzberg and M. N. G. James, *ibid.* **313**, 653 (1985).
37. G. Matsuda, Y. Suzuyama, T. Maita, T. Umegane, *FEBS Lett.* **84**, 53 (1977).
38. M. Ikura *et al.*, *Science* **256**, 632 (1992); W. E. Meador, A. R. Means, F. A. Quiocho, *ibid.* **257**, 1251 (1992).
39. Abbreviations for the amino acids residues are A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
40. C. W. Muller and G. E. Schulz, *J. Mol. Biol.* **224**, 159 (1992); E. F. Pai *et al.*, *EMBO J.* **9**, 2351 (1990).
41. T. D. Pollard, S. K. Doberstein, H. G. Zot, *Annu. Rev. Physiol.* **53**, 653 (1991).
42. A. Musacchio *et al.*, *Nature* **359**, 851 (1992).
43. C. Tesi, T. Barman, F. Travers, *FEBS Lett.* **236**, 256 (1988).
44. R. Mahmood, M. Elzinga, R. G. Yount, *Biochemistry* **28**, 3989 (1989).
45. H. M. Warrick and J. A. Spudich, *Annu. Rev. Cell Biol.* **3**, 379 (1987).
46. I. Rayment *et al.*, *Science* **261**, 58 (1993).
47. K. Yamamoto, *J. Mol. Biol.* **217**, 229 (1991).
48. M. Burke and E. Reisler, *Biochemistry* **16**, 5559 (1977); J. A. Wells and R. G. Yount, *Methods Enzymol.* **85**, 93 (1982).
49. H. White and I. Rayment, *Biochemistry*, in press.
50. J. A. Sleep, K. M. Trybus, K. A. Johnson, E. W. Taylor, *J. Muscle Res. Cell Motil.* **2**, 373 (1981).
51. Y. Okamoto and R. G. Yount, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1575 (1985).
52. R. G. Yount, C. R. Cremo, J. C. Grammer, B. A. Kerwin, *Philos. Trans. R. Soc. London Ser. B* **336**, 55 (1992).
53. C. R. Cremo, J. C. Grammer, R. G. Yount, *J. Biol. Chem.* **264**, 6608 (1989); J. C. Grammer and R. G. Yount, *Biophys. J.* **59**, 226a (1991).
54. R. S. Goody, W. Hofmann, H. G. Mannherz, *Eur. J. Biochem.* **78**, 317 (1977).
55. E. W. Taylor, *CRC Crit. Rev. Biochem.* **6**, 103 (1979); R. S. Adelstein and E. Eisenberg, *Annu. Rev. Biochem.* **49**, 921 (1980); M. G. Hibberd and D. R. Trentham, *Annu. Rev. Biophys. Biochem.* **15**, 119 (1986); M. A. Geeves, *Biochem. J.* **274**, 1 (1991).
56. R. W. Lymn and E. W. Taylor, *Biochemistry* **10**, 4617 (1971).
57. W. B. Gratzer and S. Lowey, *J. Biol. Chem.* **244**, 22 (1969).
58. E. E. Huston, J. C. Grammer, R. G. Yount, *Biochemistry* **27**, 8945 (1988).
59. K. Wakabayashi *et al.*, *Science* **258**, 443 (1992).
60. Amino acid analyses were performed by L. Mende-Mueller at the Protein and Nucleic Acid Facility, Medical College of Wisconsin, Milwaukee, WI 53226.
61. T. E. Ferrin *et al.*, *J. Mol. Graphics* **6**, 13 (1988).
62. P. J. Kraulis, *J. Appl. Cryst.* **24**, 946 (1991).
63. Y. S. Babu, C. E. Bugg, W. J. Cook, *J. Mol. Biol.* **204**, 191 (1988).
64. W. Kabsch, *J. Appl. Cryst.* **21**, 67 (1988); *ibid.*, p. 916.
65. G. C. Fox and K. C. Holmes, *Acta Crystallogr.* **20**, 886 (1966).
66. M. G. Rossmann, *Methods Enzymol.* **114**, 237 (1985).
67. We thank the co-directors (P. A. Frey, W. W. Cleland, H. Lardy, and G. H. Reed) at the Institute for Enzyme Research for support and discussion, K. Johnson and J. Dewane for technical assistance, and J. Sakon and J. Wedekind for help in the synchrotron data collection and processing. This project was initiated at Brandeis University and continued at the University of Arizona; we thank D. L. D. Caspar, J. H. Law, and M. A. Wells at those institutions for their support and encouragement. Supported by NIH grants (I.R., H.M.H., and D.A.W.). I.R. thanks R. Yount for support and encouragement during the long process of solving this structure.

6 April 1993; accepted 4 June 1993