Predicting allosteric switches in myosins

KENT KIRSHENBAUM,¹ MALIN YOUNG,¹ AND STEFAN HIGHSMITH²

¹Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

²Department of Biochemistry, University of the Pacific School of Dentistry, San Francisco, California 94115

(RECEIVED January 8, 1999; ACCEPTED May 21, 1999)

Abstract

The sequences of several members of the myosin family of molecular motors are evaluated using ASP (Ambivalent Structure Predictor), a new computational method. ASP predicts structurally ambivalent sequence elements by analyzing the output from a secondary structure prediction algorithm. These ambivalent sequence elements form secondary structures that are hypothesized to function as switches by undergoing conformational rearrangement. For chicken skeletal muscle myosin, 13 discrete structurally ambivalent sequence elements are identified. All 13 are located in the heavy chain motor domain. When these sequence elements are mapped into the myosin tertiary structure, they form two compact regions that connect the actin binding site to the adenosine 5'-triphosphate (ATP) site, and the ATP site to the fulcrum site for the force-producing bending of the motor domain. These regions, predicted by the new algorithm to undergo conformational rearrangements, include the published known and putative switches of the myosin motor domain, and they form plausible allosteric connections between the three main functional sites of myosin. The sequences of several other members of the myosin I and II families are also analyzed.

Keywords: ASP; allosteric switch; conformational rearrangement; PHD; sequence; structure prediction

Myosin is an actin-based motor protein. The intramolecular conformational changes that occur during the force-generating cycle are not well characterized. Low- and high-resolution structural studies suggest that large-scale bending of the 135 kDa heterotrimeric myosin motor domain is part of the mechanism. But there are only hints as to what structural changes might coordinate the bending. Identifying the internal structural changes, which are required for allosteric communication between the ATP and actin binding sites and which control the bending of the motor domain, is an important goal.

Results obtained using several independent techniques indicate that bending of the motor domain is part of the mechanism. A comparison of the crystal structures of isolated skeletal and smooth muscle myosin motor domains, with different nucleotides bound, indicate that the regulatory subdomain rotates by 70° relative to the catalytic subdomain (Rayment et al., 1993b; Dominguez et al., 1998). This comparison of high-resolution structures of skeletal and smooth muscle myosins is consistent with results from lower resolution image reconstruction of electron micrographs of smooth

muscle motor domain bound to actin, which indicate its regulatory domain rotates when ADP is removed (Whittaker et al., 1995). In solution, transient electric birefringence measurements of the hydrodynamic size of isolated skeletal muscle myosin motor domain indicate it bends near the center of its long axis, when phosphate is removed (Highsmith & Eden, 1990).

The atomic structure of chicken smooth muscle myosin motor domain shows that bending involves rotations about residue 699 and residue 710 (Dominguez et al., 1998), using skeletal muscle sequence numbers. This fulcrum site is located about 4 and 8 nm, respectively, away from the ATP and actin binding sites, which are about 6 nm apart from each other (Fig. 1A). It is the internal structural changes involved in allosteric communication between these three functional sites that we seek to identify. Comparison of high-resolution crystal structures of *Dictyostelium* myosin II-truncated motor domain with several different nucleotide analogs bound (Fisher et al., 1995) indicate small ligand-dependent shifts in subdomain positions that are consistent with movement near the fulcrum. But regions that might provide internal communication between the actin and ATP binding sites, or between either of those and the fulcrum site, have almost superimposable secondary structures for all the nucleotide-myosin motor domain complexes investigated (Rayment et al., 1993b; Fisher et al., 1995; Smith & Rayment, 1995, 1996; Bauer et al., 1997; Gulick et al., 1997). The smooth muscle myosin structure, which has the regulatory domain rotated, also has very few differences in secondary structure, compared to the skeletal muscle myosin structure. Despite the similarity of these crystal structures, it is likely that allosteric communication between the actin, ATP, and fulcrum sites, as well

Reprint requests to: Stefan Highsmith, Department of Biochemistry, University of the Pacific, 2155 Webster St., San Francisco, California 94115-2399; e-mail: shighsmith@uop.edu.

Abbreviations: AR, ambivalent region; ASP, the computer algorithm Ambivalent Structure Predictor; ATP, adenosine 5'-triphosphate; $\Delta Pr(k)$, residue ambivalence score for residue k ; $\mu(k)$ [or $\mu(k)$ in some instances], mean ambivalence score for residue k as part of a windowed subsequence; PHD, the computer algorithm Profile network prediction HeiDelberg; S1, the myosin motor domain called myosin subfragment 1; SASE, structurally ambivalent sequence element; z , z -score cutoff.

as the bending transition itself, involve structural rearrangements. These structural changes may be transient and produce intermediates that are not sufficiently stable to crystallize, but they are crucial to the mechanism of force generation.

Here we evaluate myosin sequences using a novel computational method called ASP (Ambivalent Structure Predictor), which predicts regions of protein structure more likely to undergo conformational rearrangement (Young et al., 1999). First, an ambivalence score $\Delta Pr(k)$ is determined for each residue k (Equation 1 in Methods). The lower the value of $\Delta Pr(k)$, the more ambivalent a residue is about what secondary structure it will assume. The $\Delta Pr(k)$ values are used to evaluate the residues as members of windowed subsets to obtain a mean ambivalence score $\mu(k)$ (Equation 2 in Methods). Finally, using a z -score cutoff of -1.75 standard deviations compared to the mean ambivalence score for the entire sequence (Equation 3 in Methods), we define structurally ambivalent sequence elements (SASEs). These subsequences are predicted to function as switches or allosteric connectors by undergoing conformational rearrangement in solution, although they may maintain stable secondary structures in crystalline environments.

The ASP algorithm has been tested by analyzing sequences of proteins known to undergo backbone rearrangements as part of their biological activity. It reliably identified known and putative sites of conformational switches and related conformational rearrangements from the primary sequences of four members of the serpin family, transducin- α , adenylyl cyclase activator α subunit, elongation factor Tu, NF- κ B p50, $\gamma\delta$ resolvase, TATA-box binding protein, DNA topoisomerase I, prion protein, aldose reductase, hemagglutinin, and chicken skeletal muscle myosin (Young et al., 1999).

Here we use ASP to analyze myosin sequences in greater detail to identify structurally ambivalent sequence elements that may function in the tertiary structure as switches or as the “moving parts” of the allosteric connections between the ATP, actin, and fulcrum sites (Kirshenbaum et al., 1998). Putative switch and allosteric regions in myosin have been implicated by other methods, and those results are used below to evaluate the predictions made using this new algorithm.

Results

Skeletal muscle myosin

The profile of $\Delta Pr(k)$ values for chicken skeletal muscle myosin heavy chain indicates that the range of scores is broad with many low values, indicating secondary structure assignments that are highly ambivalent (Fig. 2A). The $\mu(k)$ profile obtained from $\Delta Pr(k)$ using a window size equal to 21 has 13 clusters of amino acids scoring below the -1.75 standard deviation cutoff (Fig. 2B; Table 1). The average size is 13 ± 5 residues. All 13 of these SASEs are in the motor domain of the heavy chain (amino acids 1–843 out of 1938 total), which is usually called subfragment 1 (S1).

Although the SASEs are dispersed in the primary sequence, they exist as a nearly contiguous structure comprising two regions (ambivalent region I, ARI, and ambivalent region II, ARII) at the tertiary level. ARI includes the 67 amino acids in subsequences 235–247, 279–296, 310–316, 357–372, and 404–416. This region begins close to the ATP site [which itself does not include low $\mu(k)$ residues] and extends as a curved column of tertiary structure to the actin binding site (Fig. 1B). Starting at the ATP site, sub-

Table 1. Skeletal muscle heavy chain structurally ambivalent sequence elements^a

SASE	Residues/PHD assignment	Size	A.K.A.
38–5	VFVVPKESFVDGT EEE-----HHHHEE	14	
235–247	AKTVRNDNSSRFG -----	13	Switch I
251–256	RIHFGA EEEE--	6	
279–296	LPAERSYHIFYQIMSNNK HHH----EEHHHHHHHH--	18	
310–316	YDYHYVS --EEEE-	7	
357–372	AVMHYGNLKFQKQRE HHH-E-E-----	16	
404–416	PRVKVGNFVTKG --EE--HHHHH--	13	
459–481	IGVLDIAGFEIFDFNSFEQLCIN EEEE-----E-----HHHHE-	23	Switch II rigid relay loop
484–489	NEKLQQ -HHHHH	6	Switch II rigid relay loop
501–508	EEYKKEGI HHHHH---	8	Switch II rigid relay loop
580–587	SLVHYAGT EEEE--E	8	
691–708	VLHQLRCNGVLEGIRICR HHHHH-HHHHHHHHHHHH	18	SH1-SH2 fulcrum, converter
754–769	DHTQYRFGHTKVFFKA -----EEHHHH	16	Converter

^aThe computational method is described in the text. The $\Delta Pr(k)$ and Z window size (w) is 21 for the above analysis. The secondary structure assignments, H for helix and E for extended, are by PHD (Rost & Sander, 1993). The column headed A.K.A. indicates functional assignments for a SASE made by others (see text).

sequences 235–247, 279–296, 310–316, and 357–372 about one another in that order. The region from 357–372 is separate from but very close to subsequence 404–416, which is part of the actin binding site (Rayment et al., 1993a).

ARII has a core of 61 amino acids in sub-sequences 251–256, 459–481, 484–489, 580–587, and 691–708. These contiguous tertiary regions form a globular substructure with an extension ending with residue 256 (Fig. 1C). The 459–481 subsequence is located near the ATP site. The remainder of ARII consists of the 14 amino acids in sequence 38–51, which is separate from but near residue 691 of the core, and the 24 amino acids in the combined sequences 501–508 and 754–769, which are separate from but near residue 708 of the core (Fig. 1C).

When chicken skeletal muscle myosin light chain sequences are analyzed using a window size equal to 21, none of the residues have $\mu(k)$ scores less than 9 (Fig. 3). This result suggests that the skeletal muscle light chains do not contain conformational switches (Young et al., 1999).

Fig. 1. See caption on facing page.

Fig. 2. **A:** Residue ambivalence scores $\Delta Pr(k)$ obtained using Equation 1, for chicken skeletal muscle myosin heavy chain (1,938 residues). For these unwindowed raw values, the scale is from 0 for the least reliable secondary structure assignment (the most ambivalent) to 18 for the most reliable assignment (the least ambivalent). **B:** Mean ambivalence scores $\mu(k)$ obtained using Equation 2, with a window size equal to 21. The subsequences with $\mu(k)$ greater than 1.75 standard deviations from the average value for the entire sequence (e.g., those below the dashed line) are identified as structurally ambivalent and predicted to form secondary structures that function as switches (see Equation 3).

The myosin II family

Members of the myosin II family, which includes chicken skeletal muscle myosin, have two motor domains joined by a long rod of the α -helical coiled coil that aggregates them into bipolar arrays in vivo. These myosins have related functions, to move, or to move on, actin filaments, although there are differences in their supra-molecular structures. If the ambivalent regions identified above for chicken skeletal muscle myosin have a functional role, similar regions of conformationally plastic secondary structure should exist in other myosin IIs. A selection of nonskeletal muscle myosin II sequences was analyzed for comparison to the chicken case, and their $\mu(k)$ profiles are shown in Figure 4. The profiles for chicken skeletal, chicken smooth, human cardiac, scallop adductor, and

insect flight muscle myosins are strikingly similar. These myosins have a high percent sequence identity, which undoubtedly contributes to similarities in the PHD output and, therefore, in the profiles. Nonetheless, the similar locations of ambivalent subsequences and the likely similar tertiary structures of the various muscle myosins suggests that each of them may have two regions of conformationally labile secondary structure neighboring the ATP site, as described above for chicken skeletal muscle myosin (Fig. 1).

There are crystal structures of chicken skeletal muscle myosin with the full regulatory domain (Rayment et al., 1993b) and of chicken smooth muscle myosin with a truncated regulatory domain (Dominguez et al., 1998). In addition, the crystal structure of the truncated heavy chain catalytic domain of *Dictyostelium discoideum* myosin II, which has no regulatory domain, is solved (Fisher

Fig. 1. (facing page) The structurally ambivalent sequence elements identified by ASP form ambivalent regions (ARs) when they are mapped into the tertiary structure of the motor domain of chicken skeletal muscle myosin heavy chain (Fig. 2B; Table 1). **A:** The catalytic subdomain of the motor domain of the myosin heavy chain (residues 1–843) is rendered blue. ATP, the actin binding site, and the fulcrum sites are rendered as solid surfaces and rendered yellow, magenta, and orange, respectively. The catalytic subdomain rotates by 70° about the fulcrum site (see text). The rod portion of the heavy chain and the light chains, which make up the regulatory subdomain, are not shown. The structure is from Rayment et al. (1993b). **B:** The structurally ambivalent sequence elements (Table 1) that make up ARI, which connects the ATP site to the actin binding site, are rendered in red. The magenta actin binding site in 1A is part of ARI and is red. ATP is yellow. **C:** The structurally ambivalent sequence elements (Table 1) that make up ARII, which connects the ATP site to the fulcrum site, are rendered in red. The orange residue 699 in A is part of ARII and is red. ATP is yellow.

Fig. 3. Profiles of $\mu(k)$ (window size = 21) for chicken skeletal muscle light-chain sequences. (A) Light chain 1 (also called essential, or alkali, light chain one); (B) light chain 2 (also called regulatory, or DTNB, light chain); (C) light chain 3 (also called essential, or alkali, light chain two). The dotted line is the $z = -1.75$ cutoff, but the important feature is that none of the sequences have $\mu(k)$ values below 9. These proteins have no SASEs.

et al., 1995). The Dictyostelium myosin II heavy chain and the chicken skeletal muscle myosin have 36% sequence identity (Warwick et al., 1986). Analysis of the Dictyostelium sequence, using a window size of 21 and a -1.75 cutoff, identifies subsequences that map into the crystal structure at many of the same locations as those identified for chicken skeletal muscle myosin. If the cutoff is reduced to -1.70 to make the SASEs larger by including slightly less ambivalent residues, the conformationally labile secondary structures in the Dictyostelium structure become very similar with those identified in the chicken structure (Fig. 5). Nine additional residues are identified as ambivalent, when z is increased from -1.75 to -1.70 . As is typical when the z -cutoff is changed in an ASP analysis, the length of some or all of the existing SASEs change. For the Dictyostelium myosin II heavy chain, when $z = 1.75$ the SASEs are 232–234, 269–291, 304–306, 358–364, 396–411, 457–469, 492–505, 582–585, 672–691, and 736–747, and when $z = -1.70$ the SASEs are 232–234, 268–295, 303–307, 358–364, 396–411, 457–469, 492–505, 581–586, 672–691, and 736–747.

The myosin I family

Members of the myosin I family have smaller motor domains and little or no α -helical rod portion. They function, unaggregated, as monomeric motors. They bind and move, or move on, actin, but have a much wider range of functions than members of the myosin II family (Pollard et al., 1991). Mean residue ambivalence profiles

for several myosin I heavy chain sequences, which were selected to represent higher and lower organisms (Coluccio, 1997), are shown in Figure 6. The percent sequence identity of these myosin I family members is below 50. Each protein has several SASEs, and the profiles have many similarities. In all cases, the negative peaks in the $\mu(k)$ profiles are in almost identical locations in the 1–620 region, taking into account small shifts of the starting positions. This similarity suggests that the positions of the conformationally labile switch elements of the myosin I motor domains are conserved. Beyond residue 620, which is likely to be beyond the ATP and actin binding site-containing motor domain, the similarities decrease. The function of the regions of the myosin I sequences beyond the motor domain is to bind to actin in an ATP-independent fashion or to bind to membrane. What is bound by the different myosin I species differs, and there is no reason to expect SASEs to be in the same location.

Discussion

The method

The computational method used to analyze the proteins in this study is based on several assumptions. One assumption is that there are conformational switches in proteins, which depend on secondary structural lability. A low preference for a particular secondary structure is presumed to increase the probability of transi-

Fig. 4. Profiles of $\mu(k)$ (window size = 21) are shown for muscle myosin II heavy chains. (A) Chicken skeletal; (B) chicken smooth; (C) scallop adductor; (D) human cardiac; (E) insect flight muscle. The dotted line is at $z = 1.75$.

tions to other conformations, which provides the switching property. A second assumption is that labile secondary structures can be predicted from primary sequences, an idea first hypothesized by Chou and Fasman (1974, 1975). The accuracy of secondary structure prediction has improved since then. Using the algorithm PHD, the accuracy is greater than 70% (Rost & Sander, 1993). This level of reliability strengthens the idea that the PHD analysis of the predictions, which is the input for ASP, is also reliable. Finally, it is assumed that regardless of which secondary structures are most stable in the various protein crystals in the currently available databases, the primary sequences will be assigned a secondary structure type with an associated low reliability by prediction algorithms.

Skeletal muscle myosin

The results obtained for skeletal muscle myosin are remarkable in several ways, and they suggest that the approach developed here may be valuable for predicting switch elements of functional significance (Young et al., 1999). All the identified SASEs are in the catalytic portion of the motor domain, which is the functional enzymic subdomain of myosin. The average size of a sub-sequence predicted to be a labile secondary structural unit is 13 residues (Table 1). Although the SASEs exist as 13 discrete segments in the sequence, they form two compact and coherent tertiary regions (ARI and ARII) within the motor domain (Fig. 1). The location of these regions is nearly identical in chicken skeletal and Dictyostelium myosins (Fig. 5). It is also striking that the location of the labile substructure within the myosin motor domain makes it plau-

sible that by concerted backbone displacements it could serve as the structural basis for allosteric communication between the ATP site, the actin binding site, and the fulcrum site for lever arm motion. Further experimental evidence will be needed to test this hypothesis rigorously. But at least a beginning can be made here by determining whether the several switch-like sequences in myosin, which have been reported from experimental studies, are among those identified here as ARI and ARII.

Two sites, called switch I and switch II, which are close to the ATP site in myosin, were tentatively identified by comparison of the topological similarities between the active site regions of myosin and the G-protein family (Smith & Rayment, 1996). By analogy to the G-proteins, switch I in myosin is subsequence 235–245. This subsequence is part of ARI, which connects the ATP site to the actin binding site (Table 1; Fig. 1B). The other G-protein-derived ATP-sensitive site, switch II, is subsequence 466–506 in myosin (Kull et al., 1996; Sablin et al., 1996; Smith & Rayment, 1996). Almost all of this switch element is part of ARII, which connects the ATP site to the part of myosin catalytic domain that contains the fulcrum site and abuts the regulatory domain lever arm (Table 1; Fig. 1C).

The crystal structure of chicken smooth muscle myosin motor domain with MgADP·AlF₄ bound has the regulatory domain rotated compared to the skeletal muscle myosin motor domain with no nucleotide bound. The interpretation is that the MgADP·AlF₄ complex approximates the structure before force is produced, and the nucleotide-free motor domain, solved by Rayment et al. (1993b), approximates the structure after force is produced (Dominguez et al., 1998). The rotation is 10° about residue 699, which is iden-

Fig. 5. The structurally ambivalent sequence elements are mapped into the tertiary structures of the truncated motor domain of *D. discoideum* ($z \leq -1.70$), above in green, and chicken skeletal muscle myosin motor domain heavy chain ($z \leq -1.75$), below in blue, and rendered as red. Tertiary structures are from the Brookhaven Protein Data Bank (Rayment et al., 1993b; Fisher et al., 1995).

tified by ASP (Table 1), and 60° about residue 710, which is not identified by ASP, but is located between the two extensions of SASEs that are part of ARII (Fig. 1). In the interpretation of the smooth muscle myosin structure, two functional elements are invoked, a rigid relay loop consisting of residues 465–512, and a converter domain consisting of residues 699–767 (Dominguez et al., 1998). The rigid relay loop is proposed to connect the ATP site to the fulcrum site. ASP identified most of the rigid relay loop as SASEs (459–581, 484–489, and 501–508 in Table 1). This consistency supports the validity of the ASP prediction, although we suggest that the rigid relay loop may not be rigid as it appears to provide an allosteric connector between the ATP site and the fulcrum site (Fig. 1C). ASP also identifies portions of the converter domain, specifically subsequences 691–708 and 754–769. These SASEs correspond to the N- and C-termini of the converter, and we suggest that they function as switches during the rotation of the regulatory domain.

Another approach to identifying internal elements of communication is mutational analysis. Null phenotypes are produced when multiple mutations are introduced at locations critical for allosteric communication (Ruppel & Spudich, 1996b). The proteins from the null phenotypes are isolated, purified, and analyzed. Based on mutational analysis of *D. discoideum*, mutations at positions 470,

475, and 487 (using chicken skeletal muscle myosin sequence) produce a null phenotype, apparently by interfering with critical structural changes on the lower side of the cleft of the 50 kDa subdomain (Ruppel & Spudich, 1996a). These residues are in subsequences identified here to be conformationally flexible (Table 1) and are part of ARII that connects the ATP site to the fulcrum region (Fig. 1C). Single site-specific mutations in smooth muscle myosin at locations that correspond to residues 468 or 470 in skeletal muscle myosin (in ARII, Table 1) have been shown to inhibit ATP hydrolysis and actin activation of ATPase activity, without inhibiting ATP-induced actin dissociation (Onishi et al., 1997). This result is consistent with the prediction that ARII connects the ATP site to the fulcrum region (Fig. 1C), because lever arm rotation about the fulcrum appears to be coupled to ATP hydrolysis (Highsmith & Eden, 1990; Rayment et al., 1993b; Highsmith et al., 1998).

The region of myosin containing the chemically reactive cysteines at locations 697 and 707 residues (Reisler et al., 1974; Wells & Yount, 1979) has been suggested to undergo nucleotide-dependent switch-like behavior to be cross-linked (Polosukhina & Highsmith, 1997). These residues are part of ARII. The Dictyostelium equivalent helix-loop-helix structure, which includes residues 697 and 707 in skeletal myosin, changes position as part of a sub-

Fig. 6. Profiles of $\mu(k)$ (window size = 21) for members of the myosin I heavy chains: (A) *Drosophila melanogaster* 1A; (B) *D. melanogaster* 1B; (C) *Acanthamoeba castellanii* 1B; (D) *D. discoideum* 1A; (E) *D. discoideum* 1B; (F) *D. discoideum* 1C; (G) rat brush-border myosin 1. The dotted line in at $z = 1.75$.

domain motion, but not secondary structure, when $\text{Mg} \cdot \text{ADP} \cdot \text{BeF}_x$ is substituted for $\text{Mg} \cdot \text{ADP} \cdot \text{AlF}_4$ in truncated catalytic domain crystals (Fisher et al., 1995). This subdomain movement is thought to be coupled to the conversion of ATP to its hydrolysis transition state in the active site and to be part of the allosteric communication in myosin (Fisher et al., 1995). Conformational switching may involve transient intermediate structures that are not stable enough to crystallize. But if the subdomain movement that is observed in crystals were to occur in solution, it would increase solvent accessibility and would lower the energy of activation for a transient conformational change. The observed subdomain movement (Fisher et al., 1995) and the secondary structure lability predicted here by ASP are consistent with the chemical cross-linking data (Reisler et al., 1974; Wells & Yount, 1979). They are also consistent with a helix to flexible loop mechanism hypothesized to store energy transiently from ATP binding to myosin (Polosukhina & Highsmith, 1997).

There are also functional sites on myosin, which have been predicted by others not to have switch properties. The structural changes in the ATP site itself are very limited (Rayment et al., 1996). The sequences that make up the ATP site are not predicted to be structurally ambivalent. The segments 204–216 and 627–646 are thought to be flexible loops on the surface of myosin, because they do not appear in the crystal structure (Rayment et al., 1993b). They are not predicted to be ambivalent secondary structures, which is reasonable because, although they assume many conformations, all of them are in the loop (L) category of the PHD prediction. None of the five heavy-chain tryptophans is among the predicted switch structures. There is an increase in intrinsic fluorescence

intensity when ATP is hydrolyzed (Werber et al., 1972), but to the best of our knowledge, no tryptophan has been implicated as being part of structure that has a switch function.

There are no ambivalent subsequences identified in the 800–1938 region (Fig. 2), which is an α -helical structure, most of which is a coiled coil (rod). Coiled coils are usually highly stable α -helical structures, and the high $\mu(k)$ scores observed for this region (Fig. 2B) may be serving as an internal control of the ASP computational method (Young et al., 1999). However, segmental flexibility has been identified at two locations in the myosin rod in earlier experiments (Mendelson et al., 1973; Highsmith et al., 1977), centered around residues 840 and 1290 in chicken skeletal muscle myosin sequence (Maita et al., 1991). There are two possible reasons that ASP does not identify these bending sites as conformationally labile. One is that changes in secondary structure may not be involved. Bending can be localized to one or two amino acids, which would not be predicted by ASP. The second possible reason is that the bending could be highly delocalized and not involve any changes in secondary structure, as has been suggested for the bending in the 1,290 region (Hvidt et al., 1984). In this case as well, ASP would not predict ambivalence (Young et al., 1999).

Myosin I and II family results

The four myosin II $\mu(k)$ profiles are very similar (Fig. 4), suggesting similar allosteric switch pathways for these myosins. The locations of the conformational switch sequences of the myosin I family members analyzed here are different from those in the myosin II family, but are in register with one another (Fig. 6). The

myosin I structures were chosen as examples to include unconventional myosins from the lower organisms and from the three phylogenetically derived subcategories for the vertebrates (Coluccio, 1997). The similarity of the profiles is consistent with members of the myosin I family having conserved allosteric connections between the interacting ATP, actin, and fulcrum sites. The dissimilarities of the myosin I and myosin II profiles suggest that the allosteric connections may be different in the two families.

It was noted above that several of the myosin I profiles indicate SASEs in regions beyond the motor domain. These regions are not involved in force generation. They are responsible for anchoring or for binding cargo. The existence of SASEs in these regions suggests conformational switch function may be involved in binding, perhaps with some regulatory function.

Conclusions

The results obtained from the analyses of sequences of several members of the myosin family, a group of large enzymes with complex mechanochemical function, are consistent with the algorithm successfully identifying secondary structures that undergo conformational rearrangement. The myosin results appear to parallel those obtained for several other proteins (Young et al., 1999), in which functional switches have been characterized by X-ray crystallography and by NMR. Taken together, the data suggest that ASP may be able to identify functional switch regions from sequence for a wide variety of proteins.

The ambivalent subsequences, when mapped into the tertiary structures of chicken skeletal muscle and Dictyostelium myosins, include the published known and putative switch structures in their motor domains. More interestingly, the identified subsequences provide plausible allosteric connections between the ATP site, the actin binding site, and the fulcrum site. This predicted allosteric connection may be useful for the design of experiments that will unambiguously determine the nature of the connection for members of the myosin family.

Methods

ASP identifies those sequences of amino acids lacking a strong intrinsic preference for a single type of secondary structure (Young et al., 1999). The first step is to use the algorithm PHD (Rost & Sander, 1993) to predict the secondary structural preference for each amino acid in the sequence. PHD uses profile-based neural network methodology (Rost, 1996) to make three possible secondary structure element assignments: helical (*H*), extended (*E*), or loop (*L*), where *H* is α -helical, *E* is β -sheet, and *L* is defined as secondary structure that is neither *H* nor *E*. For each amino acid, PHD predicts not only its secondary structural element, but also estimates the reliability of the prediction, by computing the relative probabilities (*prH*, *prE*, and *prL*) that it will assume a particular secondary structure. The three relative probability values are scaled such that their sum is 9 (Rost, 1996).

The next step is to use the PHD output to compute a residue ambivalence score, $\Delta Pr(k)$, for each amino acid in the sequence, using Equation 1.

$$\Delta Pr(k) = |prH - prE| + |prH - prL| + |prE - prL|. \quad (1)$$

The $\Delta Pr(k)$ value rates the ambivalence of the PHD secondary structural element prediction for each residue *k*. The minimum

possible value of $\Delta Pr(k)$ is 0. This value is obtained for an amino acid that was assigned equal values of *prH* = *prE* = *prL* = 3 by the PHD analysis, and it indicates the lowest reliability (highest ambivalence) for a secondary structural element prediction. The maximum possible value of $\Delta Pr(k)$ is 18 for the highest reliability (lowest ambivalence) for a prediction. Our hypothesis is that high ambivalence in the prediction of secondary structure correlates with structures that are likely to exist as more than one conformation.

We next evaluate the residues as members of a subset defined by a particular window size. The window is centered about each amino acid in the sequence, and the mean ambivalence score $\mu(k)$ is calculated using Equation 2:

$$\mu(k) = \Sigma \Delta Pr(i) / S \quad (2)$$

where *S* is the window size, and where $k - [(S - 1)/2] \leq i \leq k + [(S - 1)/2]$.

Finally, we use a two-step process to identify structurally ambivalent sequence elements (SASEs). First, a coarse analysis is used to discriminate which proteins are expected to display backbone conformational rearrangements. For this analysis, we arbitrarily define subsequences with $\mu(k)$ scores of 9 or less to be potentially conformationally switching regions, whereas subsequences with globally high $\mu(k)$ scores are defined to not have this property. Only proteins that contain subsequences with mean residue scores of 9 or less are analyzed further.

For those proteins, a statistical criterion is applied to identify consistently the residues of a particular protein that are most likely to undergo changes in secondary structure. A simple *z*-score cutoff for each windowed score $\mu(k)$, relative to the mean for the entire sequence $\mu(seq)$, is used to identify the most ambivalent residues. The *z*-score is calculated according to Equation 3, so that the maximal values corresponded to regions of minimum ambivalence.

$$z = [\mu(k) - \mu(seq)] / \sigma(seq) \quad (3)$$

where $\sigma(seq)$ is the standard deviation of $\mu(seq)$. Thus, we can calculate the *z*-score for each residue and identify those regions that are significantly ambivalent. A cutoff of -1.75 is used to identify a SASE. The derivation of ASP is discussed in greater detail in the accompanying paper (Young et al., 1999).

Acknowledgments

We thank Sandy Bernstein and Michael Reedy for giving us a correct sequence of the insect flight muscle myosin heavy chain, to replace the confusing GenBank Protein Sequence version that has multiple inserts in the motor domain region. We also thank the Computer Graphics Laboratory at UCSF. K.K. thanks Chiron Corp. and Ronald Zuckermann for their generosity. M.Y. is supported by NIH Biotechnology Training Grant GM08388. K.K. is supported by NIH Training Grant GM08388. S.H. is supported by NIH Grant AR42895. This manuscript is in partial requirement for the PhD theses of K.K. and M.Y. The first two authors contributed equally to this paper.

References

- Bauer CB, Kuhlman PA, Bagshaw CR, Rayment I. 1997. X-ray crystal structure and solution fluorescence characterization of Mg 2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the *Dictyostelium discoideum* myosin motor domain. *J Mol Biol* 274:394–407.
- Chou PY, Fasman GD. 1974. Prediction of protein conformation. *Biochemistry* 13:222–245.
- Chou PY, Fasman GD. 1975. The conformation of glucagon: Predictions and consequences. *Biochemistry* 14:2536–2541.

- Coluccio LM. 1997. Myosin I. *Am J Physiol [Cell Physiol]* 42:C347–C359.
- Dominguez R, Freyzon Y, Trybus KM, Cohen C. 1998. Crystal structure of a vertebrate smooth muscle myosin motor domain and its complex with the essential light chain: Visualization of the pre-power stroke state. *Cell* 94:559–571.
- Fisher A, Smith CA, Thoden JB, Smith R, Sutoh K, Holden HM, Rayment I. 1995. X-ray structures of the myosin motor domain of *Dictyostelium discoideum* complexed with MgADP BeFx and MgADP ALF4-. *Biochemistry* 34:8960–8972.
- Gulick AM, Bauer CB, Thoden JB, Rayment I. 1997. X-ray structures of the MgADP, MgATP γ S, and MgAMPPNP complexes of the *Dictyostelium discoideum* myosin motor domain. *Biochemistry* 36:11619–11628.
- Highsmith S, Duignan K, Franks-Skiba K, Polosukhina K, Cooke R. 1998. Reversible inactivation of myosin subfragment 1 activity by mechanical immobilization. *Biophys J* 74:1465–1472.
- Highsmith S, Eden D. 1990. Ligand-induced myosin subfragment 1 global conformational change. *Biochemistry* 29:4087–4093.
- Highsmith S, Kretschmar KM, O'Konski CT, Morales MF. 1977. The flexibility of myosin rod, light meromyosin and subfragment-2 in solution. *Proc Natl Acad Sci USA* 74:4986–4990.
- Hvidt S, Chang T, Yu H. 1984. Rigidity of myosin and myosin rod by electric birefringence. *Biopolymers* 23:1283–1294.
- Kirshenbaum K, Young M, Highsmith S. 1998. Identifying ambimorphic structures in the myosin sequence. *Biophys J* 74:A282.
- Kull FJ, Sablin EP, Lau R, Fletterick RJ, Vale RD. 1996. Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature* 380:550–555.
- Maita T, Yajima E, Nagata S, Miyanishi T, Nakayama S, Matsuda G. 1991. The primary structure of skeletal muscle myosin heavy chain. 4. Sequence of the rod, and the complete 1,938-residue sequence of the heavy chain. *J Biochem (Tokyo)* 110:75–87.
- Mendelson RA, Morales MF, Botts J. 1973. Segmental flexibility of the S-1 moiety of myosin. *Biochemistry* 12:2250–2255.
- Onishi H, Morales MF, Kojima S, Katoh K, Fujiwara K. 1997. Functional transitions in myosin: Role of highly conserved Gly and Glu residues in the active site. *Biochemistry* 36:3767–3772.
- Pollard TD, Doberstein SK, Zot HG. 1991. Myosin-I. *Annu Rev Physiol* 53:653–681.
- Polosukhina K, Highsmith S. 1997. Kinetic investigation of the ligand dependence of rabbit skeletal muscle myosin subfragment 1 Cys-697 and Cys-707 reactivities. *Biochemistry* 36:11952–11958.
- Rayment I, Holden HM, Whittaker M, Yohn CB, Lorenz M, Holmes KC, Milligan RA. 1993a. Structure of the actin–myosin complex and its implications for muscle contraction. *Science* 261:58–65.
- Rayment I, Rypniewski WR, Schmidt-Base K, Smith R, Tomchick DR, Benning MM, Winkelmann DA, Wesenberg G, Holden HM. 1993b. 3-Dimensional structure of myosin subfragment-1—A molecular motor. *Science* 261:50–58.
- Rayment I, Smith C, Yount RG. 1996. The active site of myosin. *Annu Rev Physiol* 58:671–702.
- Reisler E, Burke M, Himmelfarb S, Harrington WF. 1974. Spatial proximity of the two essential sulfhydryl groups of myosin. *Biochemistry* 13:3837–3840.
- Rost B. 1996. PHD: Predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol* 266:525–539.
- Rost B, Sander C. 1993. Prediction of protein structure at better than 70% accuracy. *J Mol Biol* 232:584–599.
- Ruppel KM, Spudich JA. 1996a. Structure–function analysis of the motor domain of myosin. *Annu Rev Cell Dev Biol* 12:543–573.
- Ruppel KM, Spudich JA. 1996b. Structure–function studies of the myosin motor domain: Importance of the 50-kDa cleft. *Mol Biol Cell* 7:1123–1136.
- Sablin EP, Kull FJ, Cooke R, Vale RD, Fletterick RJ. 1996. Crystal structure of the motor domain of the kinesin-related motor ncd. *Nature* 380:555–559.
- Smith CA, Rayment I. 1995. X-ray structure of the magnesium(II)–Pyrophosphate complex of the truncated head of *Dictyostelium discoideum* myosin to 2.7 angstrom resolution. *Biochemistry* 34:8973–8981.
- Smith CA, Rayment I. 1996. X-ray structure of the Mg(II)–ADP–vanadate complex of the *Dictyostelium discoideum* myosin motor domain to 1.9 angstrom resolution. *Biochemistry* 35:5404–5417.
- Warrick HM, DeLozanne A, Leinwand LA, Spudich JA. 1986. Conserved protein domains in a myosin heavy chain gene from *Dictyostelium discoideum*. *Proc Natl Acad Sci USA* 83:9433–9437.
- Wells JA, Yount RG. 1979. Active site trapping of nucleotides by crosslinking two sulfhydryls in myosin subfragment 1. *Proc Natl Acad Sci USA* 76:4966–4970.
- Werber MM, Szent-Gyorgyi AG, Fasman GD. 1972. Fluorescence studies on heavy meromyosin–substrate interaction. *Biochemistry* 11:2872–2883.
- Whittaker M, Wilson-Kubalek EM, Smith JE, Faust L, Milligan RA, Sweeney HL. 1995. A 35-angstrom movement of smooth muscle myosin on ADP release. *Nature* 378:748–751.
- Young M, Kirshenbaum K, Dill KA, Highsmith S. 1999. Predicting conformational switches in proteins. *Protein Sci* 8:1752–1764.