Amino Acid Sequence of Myosin Essential Light Chain from the Scallop *Aquipecten irradians*†

John H. Collins,*,‡ Ross Jakes,§ John Kendrick-Jones,§ John Leszyk,‡ Winifred Barouch,‡ Janet L. Theibert,‡ Judith Spiegel,∥ and Andrew G. Szent-Györgyi∥

Departments of Biology and Chemistry, Clarkson University, Potsdam, New York 13676, Laboratory of Molecular Biology, Medical Research Council Centre, University Medical School, Cambridge CB2 2QH, England, and Department of Biology, Brandeis University, Waltham, Massachusetts 02254

Received July 11, 1986; Revised Manuscript Received August 22, 1986

ABSTRACT: The amino acid sequence of the scallop myosin essential light chain (SELC) was determined from analysis of the intact, S-carboxymethylated protein and peptides produced by cleavage at its four methionine residues by cyanogen bromide digestion and at its six arginine residues by citraconylation and tryptic digestion. SELC contains 156 amino acid residues, including three cysteines, four tyrosines, one tryptophan, two histidines, and an unblocked amino-terminal proline. The protein has a calculated M_r of 17 616. SELC is an acidic protein, with a net charge of 18– at physiological pH. Comparative analysis reveals four homologous domains (I–IV), which arose by reduplication of a gene for a small, ancestral calcium binding protein. Each domain has a helix–loop–helix structure, with all the ligands for calcium binding located within a 12-residue segment that spans the loop and the first turn of the following helix. Potential calcium binding sequences were found in the ancestral sites III (residues 94–105) and IV (residues 132–143). Mutations in critical positions in domains I and II seem to preclude the possibility of calcium binding in the amino-terminal half of SELC. An unexpected third potential calcium binding segment (at residues 119–130, predicted to be in helical conformation) was found in domain IV. A reactive thiol group (Cys-78) that is involved in binding of regulatory light chains was tentatively located in an extended "linker region", which connects the two halves of the molecule.

Contraction of molluscan muscles is triggered by the direct binding of calcium to myosin. Molluscan myosins contain specific, high-affinity calcium binding sites that must be saturated for the actin-activated Mg-ATPase to function (Kendrick-Jones et al., 1970). The myosin subunits responsible for regulation are the regulatory light chains (RLCs)[1] and the essential light chains (ELCs). The RLCs of scallop myosin (SRLC) can be removed directly without loss of contractile activity, resulting in a loss of specific calcium binding and abolishing the calcium sensitivity of the actin–myosin interaction. Readdition of SRLC restores all the calcium-dependent functions (Szent-Györgyi et al., 1973; Chantler & Szent-Györgyi, 1980). Similarly, calcium dependence of tension generation in skinned scallop fibers requires the presence of SRLC (Simmons & Szent-Györgyi, 1978, 1985).

The suggestion that scallop ELCs (SELC) are also regulatory subunits is based on findings that antibodies specific to SELC desensitize scallop myosin (Walliman & Szent-Györgyi, 1981). Further support for this hypothesis comes from several lines of evidence indicating that SELC and SRLC are close together in the neck region of the myosin molecule (Bagshaw, 1977; Walliman et al., 1982; Walliman & Szent-Györgyi, 1981; Flicker et al., 1983; Szentkiralyi, 1984; Winkelman et al., 1984). Cross-linking studies show that the amino-terminal portion of SRLC is separated from SELC by a distance of more than 10 Å at rest, but the separation decreases during contraction (Hardwicke et al., 1983). The change is apparently due to the movement of SELC (Hardwicke & Szent-Györgyi, 1985) and takes place on myosin in the absence of actin (Ashiba et al., 1980; Wells & Bagshaw, 1985). SRLC protects SELC from digestion by papain (Stafford et al., 1979) and abolishes the reactivity of the thiol groups of SELC (Hardwicke et al., 1982). The recent finding (Ashiba & Szent-Györgyi, 1985) that SRLC blocks the exchange of SELC in intact myosin preparations suggests that parts of the SELC may lie between SRLC and the heavy chain of myosin.

The calcium ions that trigger the myosin ATPase are bound at specific sites on the molecule that are unique to calcium-regulated myosins. These sites, which have not yet been localized, are not the same as the divalent cation binding sites on SRLC (Bagshaw & Kendrick-Jones, 1979). Although the removal of SRLC lowers their affinity and selectivity, the specific sites are retained on the desensitized SELC–heavy chain complex (Chantler & Szent-Györgyi, 1980). Therefore, it is possible that SELC are responsible for calcium binding by desensitized myosin (Kendrick-Jones & Scholey, 1981).

In order to more precisely define the myosin subunit interactions that are responsible for calcium regulation in scallop muscle, it is necessary to elucidate the structures of the subunits. The amino acid sequences of SRLC (Kendrick-Jones & Jakes, 1976) and several closely related RLCs (Maita et

† This work was supported by grants from the NIH (AM35120 and AM15963), the NSF, and the Muscular Dystrophy Association.

* Address correspondence to this author at the Department of Biology, Clarkson University.
‡ Clarkson University.
§ University Medical School.
∥ Brandeis University.

[1] Abbreviations: SELC, scallop essential light chain; ELC, myosin essential light chain; SRLC, scallop regulatory light chain; RLC, regulatory light chain; CNBr, cyanogen bromide; TFA, trifluoroacetic acid; PTH, phenylthiohydantoin; PTC, phenylthiocarbamyl; HPLC, high-performance liquid chromatography; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid.

0006-2960/86/0425-7651$01.50/0

al., 1984; Miyanishi et al., 1985) are known. In this paper we report the amino acid sequence of SELC. By comparative analysis with troponin C and other related calcium-binding proteins, we have formulated several testable hypotheses regarding the three-dimensional structure and the functional properties of SELC.

Experimental Procedures

Materials. All reagents were of the highest grade commercially available. Sequencer reagents were obtained from Applied Biosystems. PTH-amino acid standards were obtained from either Pierce or Applied Biosystems. Amino acid standards, phenyl isothiocyanate (for nonsequencer use), constant-boiling HCl, urea, and citraconic anhydride were purchased from Pierce. TPCK-treated trypsin was obtained from Worthington. HPLC-grade water, TFA, and methanol were obtained from Fisher. HPLC-grade acetonitrile was obtained from Burdick & Jackson. Triethylamine was purchased from Aldrich, and iodoacetic acid was purchased from Sigma. All other chemicals were obtained from Fisher. Sephadex was obtained from Pharmacia.

Preparation of Protein. SELC was prepared by the method of Kendrick-Jones et al. (1976). For alkylation, a single preparation of 125 mg of lyophilized protein was dissolved to a volume of 42 mL in 5 mM Tris buffer (pH 8.0) containing 8 M urea, 1 mM EDTA, and 0.01% sodium azide. The solution was cloudy but immediately became clear when 3.16 mL of 2-mercaptoethanol was added. After 30 min, 3.23 g of iodoacetic acid (freshly dissolved in a minimum volume of alkylation buffer) was added over a period of 15 min. After the alkylation reaction was completed (as judged by a constant pH), another 3.16 mL of 2-mercaptoethanol was added, and the sample (final volume 60 mL) was immediately applied to a desalting column. Throughout the entire procedure, the protein solution was stirred, saturated with nitrogen, and kept away from sunlight, and the pH was maintained between 7.5 and 8.5. For desalting, a 5 cm × 55 cm column of Sephadex G-25 (fine), equilibrated with 25% acetic acid, was used. The recovered *S*-(carboxymethyl)-SELC was concentrated by rotary evaporation, diluted with water, and lyophilized.

Preparation of CNBr Peptides. For cleavage at methionine residues, 60 mg of S-carboxymethylated SELC was dissolved to a clear solution of ~2 mL with 70% formic acid. Then, 600 mg of CNBr (freshly dissolved in 0.60 mL of 70% formic acid) was added, and digestion was allowed to take place at room temperature (in a closed 13 mm × 125 mm screw-cap tube) for 17 h with gentle stirring. After digestion, the resulting clear and colorless solution was diluted with 100 mL of water and then concentrated by rotary evaporation and transferred (with several washings) to a 10-mL graduated cylinder. The final volume of the sample was 4.2 mL. After aliquots were taken for analysis, partial separation of the CNBr peptides was achieved by Sephadex G-50 chromatography as shown in Figure 5 (see paragraph at end of paper regarding supplementary material). Final purification of the CNBr peptides was accomplished by HPLC, as shown in Figures 6 and 7 (supplementary material).

Preparation of Tryptic Peptides. For cleavage at arginine residues, 240 nmol of S-carboxymethylated SELC was dissolved to a clear solution with 3 mL of 20 mM sodium phosphate buffer (pH 8.5) containing 4 M urea. Over a period of 30 min, a total of 6.6 μL of citraconic anhydride was added, with stirring, while maintaining the pH between 7.0 and 8.0. The sample (final volume 3.7 mL) was then desalted on a 1 cm × 47 cm column of Sephadex G-25 (coarse) equilibrated in 50 mM sodium bicarbonate (pH 8.0) containing 0.01% sodium azide. Then, 50 μg of trypsin (freshly dissolved in 50 μL of water) was added to the desalted, citraconylated protein (volume 7.5 mL), and digestion was allowed to take place for 4 h at room temperature. Then, 2.5 mL of glacial acetic acid was added to stop the digestion and to remove the citraconyl groups, and the digest was gently stirred overnight at room temperature. The tryptic peptides were separated by HPLC, as shown in Figure 8 (supplementary material).

Preparative HPLC. Preparative HPLC was performed on one of two systems. The first system contained a Waters U6K injector, two Waters M510 pumps, a Waters M680 controller, a Waters M480 variable-wavelength absorbance detector, and a Linear dual-channel recorder. The second system contained a Glenco SV-3 injector, a Waters M6000A pump, a Waters M45 pump, a Waters M660 gradient maker, a Gilson Holochrome variable-wavelength absorbance monitor equipped with an HPLC flow cell, and a Linear dual-channel recorder. Equivalent results were obtained with both systems. Peptide mixtures were dissolved in a minimum volume of 70% formic acid and then diluted with 25% acetic acid to achieve the desired volume prior to injection. All separations were carried out by gradient elution on a 4.6 mm × 25 cm Vydac 218TP54 reverse-phase column. Solvent A was 0.1% TFA in acetonitrile–water (5:95 v/v), and solvent B was 0.1% TFA in acetonitrile–water (95:5 v/v). The flow rate was 1.0 mL/min, and the absorbance at 220 nm was monitored. Peptide peaks were collected manually and used directly for amino acid analysis and sequencing.

Amino Acid Compositions and Sequence Analysis. Amino acid compositions were determined by analysis of their PTC derivatives (Heinrikson & Meredith, 1984; Bidlingmeyer et al., 1984; Cohen et al., 1986). Phenylthiocarbamylated acid hydrolysates of samples were prepared with the Waters PICO-Tag work station. PTC-amino acids were analyzed by reverse-phase HPLC on a Waters PICO-Tag column, using the gradient elution system recommended by the manufacturer. Amino acid sequences were determined on an Applied Biosystems Model 470A gas-phase protein sequencer, essentially as described by Hewick et al. (1981). PTH-amino acids obtained from the sequencer were also analyzed by reverse-phase HPLC, using a Waters Nova-Pak column and the gradient elution system described in Waters Associates Applications Brief M3500. The recovery of PTH-amino acids at each cycle was measured quantitatively. PTH-Ser and PTH-Thr were usually obtained in low yields and sometimes were not detectable at all; they could always be identified, however, from the appearance of breakdown peaks that absorbed at 313 nm. PTH-Arg and PTH-His were also often recovered in low yields. A Waters HPLC system including two M510 pumps, a M721 system controller, a WISP 710B autoinjector, a temperature control module, a M440 dual-channel absorbance detector, and a M730 integrative recorder was used for both PTH (sequence) and PTC (composition) amino acid analyses. The detector was set to measure the sum of the absorbances at 254 nm (for quantitative measurement of PTH- and PTC-amino acids) and 313 nm (for qualitative detection of breakdown products of PTH-Ser and PTH-Thr).

Results and Discussion

Thanks mainly to sequencer runs that permitted the unambiguous identification of up to 58 amino acid residues, the sequence determination of SELC was quite straightforward (see Figure 1). Nearly all of the sequence was obtained from the five CNBr peptides. The alignments of the CNBr peptides and the remainder of the sequence were determined from the overlapping tryptic peptides R4 and R7. Unlike all other

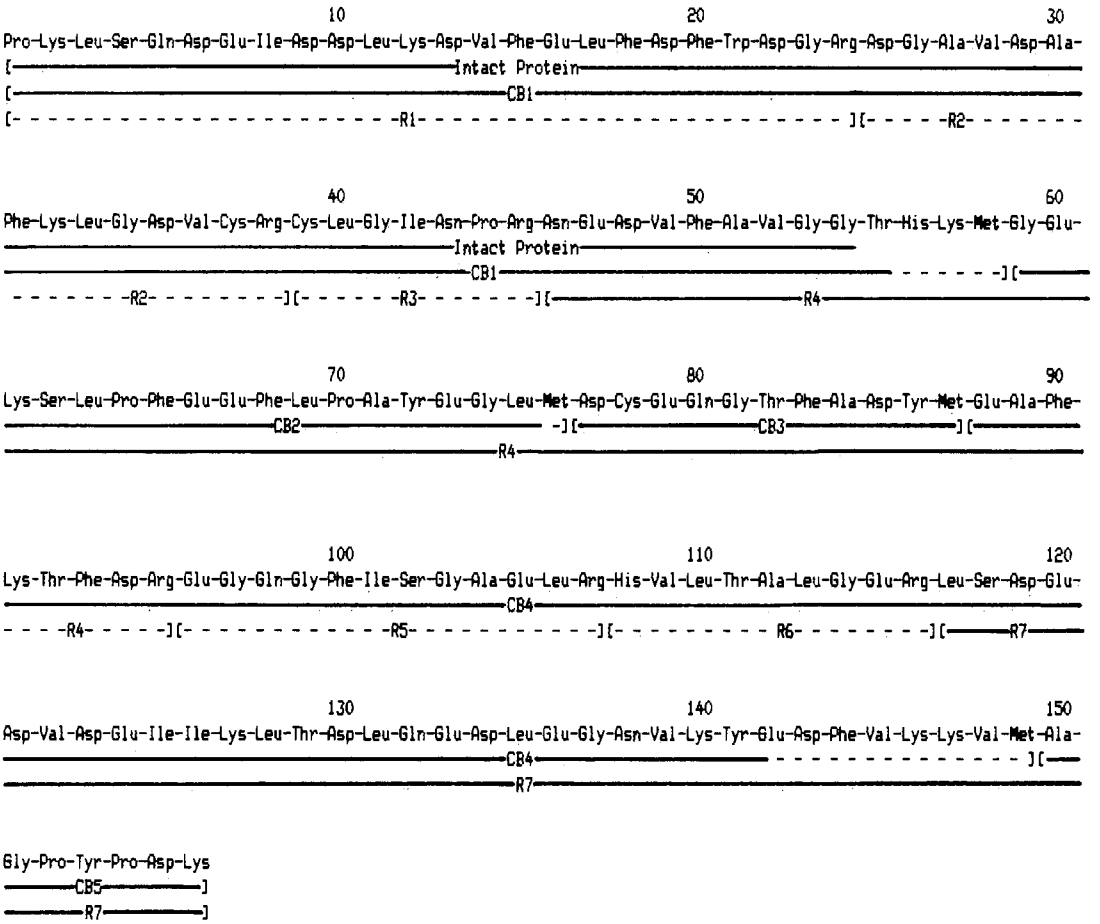

FIGURE 1: Amino acid sequence of scallop (*Aquipecten irradians*) myosin essential light chain. Peptides obtained from CNBr cleavage (CB1 to CB5) and tryptic cleavage of the citraconylated protein (R1 to R7) are shown. Solid lines indicated residues identified by sequence analysis.

myosin light chains that have been studied to date, SELC has an unblocked amino terminus, and we were able to confirm residues 1–54 by sequence analysis of the intact protein. The sequence X-Pro-Lys, where X is α-*N*,*N*,*N*-trimethylalanine, is present at the amino termini of many vertebrate myosin light chains but appears to be absent in invertebrate muscles (Henry et al., 1985). A blocked amino acid or peptide may have been removed from SELC by proteolysis during purification. Another possibility is that of mild acid cleavage of an Asp–Pro bond (Landon, 1977) near the amino terminus. Such a cleavage could have occurred during alkylation, since 25% acetic acid was used as the solvent for desalting. Neither possibility seems likely, since we have used the same or very similar procedures to prepare several other muscle proteins for sequencing and have found that amino-terminal blocking groups and Asp–Pro bonds remain intact under these conditions.

The original comparative analyses (Collins et al., 1973; Collins, 1974, 1976a,b) of the sequences of troponin C and myosin light chains with the sequence and crystal structure of a carp parvalbumin (Kretsinger & Nockolds, 1973) established that all these proteins had evolved from a common ancestor. It was suggested that this ancestor would contain two identical halves (I–II and III–IV), which arose by gene duplication. As a result of an even earlier gene duplication, each half would contain two very similar Ca binding domains. Each domain would consist of a pair of helices and a central, 12-residue calcium binding site. Evolutionary changes in the amino acid sequences of the ancestral sites of many present-day proteins would result in a loss of calcium binding ability, without markedly changing the overall structure of the domain. As more protein sequences have become available, these proposals have been repeatedly confirmed and extended to include calmodulin and other calcium binding proteins (Baba et al., 1984). The early structural predictions have been further confirmed by the recently established crystal structures of troponin C (Herzberg & James, 1985a; Sundralingam et al., 1985a) and calmodulin (Babu et al., 1985). An unexpected structural feature of these two proteins, however, was a long, exposed, central helix that connected the two halves of each molecule.

A comparison of the sequence of SELC and other ELCs with a reconstructed ancestral sequence (see Figure 2) provides some insight concerning the structure and function of SELC. The two invertebrate ELCs (scallop and drosophila) have common features that distinguish them from the vertebrate ELCs: a three-residue deletion following helix C, a three-residue insertion within helix G, and a carboxyl terminus that is elongated by five amino acid residues. These sequence similarities may be a reflection of common features in the three-dimensional structures of invertebrate ELCs. A long helix in SELC (residues 65–93, connecting the D and E helices) would be shortened (or at least distorted) by the invariant Pro-70 in the middle of helix D. Sundralingam et al. (1985b) have postulated that several intrahelical salt bridges between acidic and basic residues are needed to stabilize the long helix in troponin C and calmodulin. SELC contains only a single basic residue (Lys-91) in this segment, and so only one such salt bridge (with Glu-88) could be formed. Thus, it would seem that a long helix is not likely to be present in isolated SELC, although such a structure could be stabilized by interaction with other subunits in the intact myosin mol-

```
Domain I:                    ├─────A─────┤          ├─────B─────┤
scallop (1-44)       pklsqdeiddlkdvfe--lfdfwdgrdgavdafkl-GDvcRcLgiNP
drosophila           madvpkrevenvefvf--evmgspge--gidavdl-GDalRaLnlNP
chicken skeletal      sfspdeindfkeafl--lfdrtgda--kitlsqv-GDivRaLgqNP
chicken cardiac      veftpdqieefkeafs--lfdrtpksemkityaqc-GDvlRaLgqNP
chicken gizzard      cdfseeqtaefkeafq--lfdrtgdg--kilysqc-GDvmRaLgqNP
ancestral protein        ekseeelkevfk--vfdedgdg--kinyeel-kkimktleadek
                                      X Y Z  -Y-X -Z

Domain II:                   ├─────C─────┤          ├─────D─────┤
scallop (45-80)       rnedvfavgGthk---mgeK--slpfeeFLPayeglm-dceq
drosophila            tlalieklgGtkk---rneK--kikldeFLPiysqvkkekeq
chicken skeletal      tnaeinkilGnpskeemnaK--kitfeeFLPmlqaaannkdq
chicken cardiac:      tqaevmkvlGrpkqeemnsK--midfetFLPmlqhisktkdt
chicken gizzard       tnaevmkvlGnpksdemnlK--tlnfeqFLPmmqtiaknkdq
ancestral protein    ekteeelkevfk--afdedgdg--kinyeeflkmimktleddek
                                      X Y Z  -Y-X -Z

Domain III:                  ├─────E─────┤          ├─────F─────┤
scallop (81-116)      GtfaDymEafk--tfDrEgqg--fisgAEl-rHvLtaL--GEr
drosophila            GcyeDfiEclk--lyDkEeng--tmllAEl-qHaLlaL--GEs
chicken skeletal      GtfeDfvEglr--VFDkEgng--tvmgAEl-rHvLatL--GEk
chicken cardiac       GtyeDfvEglr--vfDkEgdt--tvmgAEl-rHvLatL--GEr
chicken gizzard       GcfeDyvEglr--vfDkEgng--tvmgAEi-rHvLvtL--GEk
ancestral protein    ekseeelkevfk--vfdedgdg--kinyeel-kkimktleadek
                                      X Y Z  -Y-X -Z

Domain IV:                   ├─────G─────┤          ├─────H─────┤
scallop (117-156)     lsdEdVdeiikltdlqEDleG--nvkYedfVkkvmagpypdk
drosophila            lddEqVetlfadcmdpEDdeG--fipYsqfVqrlmsdpvvfd
chicken skeletal      mteEeVeelmk---gqEDsnG--cinYeafVkhimsv
chicken cardiac       lteEeVdklma---gqEDanG--cinYeafVkhiman
chicken gizzard       mteEeVeqlva---ghEDsnG--cinYeelVrmvlsg
ancestral protein    ekteeelkevfk--afdedgdg--kinyeeflkmimktleddek
                                      X Y Z  -Y-X -Z
```

FIGURE 2: Comparison of the sequences of SELC, a representative group of other essential light chains (ELCs), and a reconstructed "ancestral protein A" (Baba et al., 1984), thought to be very similar to the common ancestor of myosin light chains, troponin C, calmodulin, and other calcium binding proteins. The only other published invertebrate ELC sequence is that of *Drosophila* (Falkenthal et al., 1984). Chicken heart (Maita et al., 1980) and chicken gizzard (Matsuda et al., 1981a; Grand & Perry, 1983) are the only sources of complete ELC sequences from cardiac and smooth muscles, respectively. Skeletal muscle ELC sequences from chicken (Maita et al., 1981; Matsuda et al., 1981b; Nabeshima et al., 1984), rat (Periasamy et al., 1984), mouse (Benoit et al., 1984), and rabbit (Frank & Weeds, 1974) are known, but since they are all very similar, only the smaller of two variants of the chicken ELC is shown. All of the sequences shown in Figure 2 are complete, with the exception that an additional 41 residues at the amino terminus of the chicken cardiac ELC have been omitted. The four domains of the ancestral protein are aligned with each other as shown by Baba et al. (1984). Each ELC sequence was compared with the ancestral sequence by the Protein Identification Resource ALIGN program (Dayhoff et al., 1983). Within each domain, X, Y, Z, –Y, –X, and –Z represent six Ca-coordinating oxygen ligands, arranged octahedrally in a 12-residue binding site. The predicted helices are designated A–H (Collins, 1976a,b; Herzberg & James, 1985a). Residues that are identical in all five ELC sequences are capitalized.

ecule. This may be a special feature of molluscan ELCs, since other ELCs (Figure 2) have additional lysine residues in the "linker region" between helices D and E. On the other hand, hydrodynamic and other physical studies have shown that SELC and other myosin light chains are very asymmetric molecules, suggesting that they are at least partially elongated (Stafford & Szent-Györgyi, 1978; Alexis & Gratzer, 1978).

A hydropathy plot of SELC (see Figure 3) yields results that are in good general agreement with the structure predicted on the basis of sequence homology. Each of the predicted helices corresponds to a hydrophobic segment (expected to be buried in the interior of the molecule), while the ancestral calcium binding sites are in hydrophilic segments (expected to be at or near the surface). This plot is also helpful in predicting the behavior of certain reactive amino acid side chains in spectroscopic and chemical modification studies. SELC is particularly rich in this regard, containing three cysteines, two histidines, four tyrosines, and one tryptophan.

It would be expected that the most reactive thiol group of SELC would be at Cys-78, which may occupy a key position in the linker region between the two halves of the molecule.

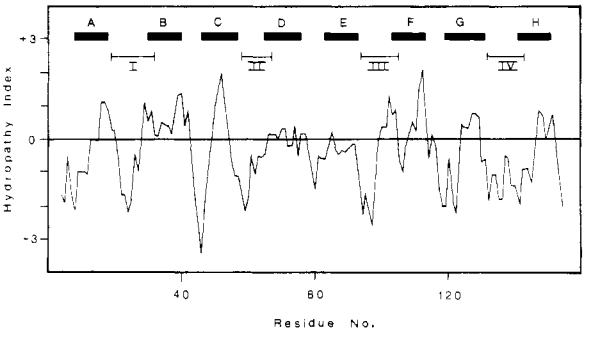

FIGURE 3: Hydropathy plot of SELC, showing the locations of predicted helices (heavy bars) and ancestral Ca binding sites (bracketed lines). Consecutive hydropathy averages are plotted for a seven-residue window advancing from the amino to the carboxyl terminus. The method of Kyte and Doolittle (1982) was used. Positive values indicate hydrophobic regions, while negative values indicate hydrophilic regions.

Cys-37 and Cys-39 are predicted to be buried in helix B. This would explain why, in the absence of denaturants, isolated SELC appears to have a single thiol group that is reactive to

```
site III (94-105): Asp-Arg-Glu-Gly-Gln-Gly-Phe-Ile-Ser-Gly-Ala-Glu

helix G (119-130): Asp-Glu-Asp-Val-Asp-Glu-Ile-Ile-Lys-Leu-Thr-Asp

site IV (132-143): Gln-Glu-Asp-Leu-Glu-Gly-Asn-Val-Lys-Tyr-Glu-Asp

test sequence:     Asp- x -Asp-Gly-Asx-Gly- x -Ile-Asx- x - x -Glu

coordinates:        X       Y       Z       -Y      -X          -Z
```

FIGURE 4: Potential calcium binding sites of SELC. The degree of similarity between the test sequence shown and each 12-residue segment within the sequence of SELC was scored with the Protein Identification Resource mutation data matrix (Dayhoff et al., 1983). Numbers in parentheses are the residue numbers within SELC.

iodoacetic acid (Ashiba & Szent-Györgyi, 1985). This thiol group is protected from alkylation when SELC is combined with regulatory light chains (Hardwicke et al., 1982). It thus appears that Cys-78 is located in a region that is exposed on the surface of the molecule and is involved in the binding of regulatory light chains. Experiments to identify the reactive thiol(s) of SELC are now in progress.

The side chain of Trp-21 should be near the surface of the molecule as part of an ancestral site that has lost its ability to bind calcium. His-56 appears to be in a hydrophilic region near the end of helix C, while His-108 in the center of helix F is invariant in the ELCs and so may serve an essential function. Tyr-72 and Tyr-86 (both near to the potentially reactive Cys-78) occupy central positions in helices D and E, respectively, and they may be useful markers for investigating the structure of the presumed extended segment that connects the two halves of SELC. For example, it is well established that regulatory light chains protect SELC from chymotryptic digestion (Stafford et al., 1979; Konno & Watanabe, 1985). Tyr-72 and Tyr-86 would be susceptible to chymotryptic digestion if they were exposed in the linker segment of SELC but could be protected if regulatory light chains bind to this region. Experiments to test this hypothesis are now in progress. Tyr-141 is invariant in all known ELC sequences, and so may play a crucial role in the function of these proteins. Burke and Wang (1982) found that in rabbit skeletal muscle this Tyr is unreactive to tetranitromethane. Therefore, it may reasonably be expected that the side chain of Tyr-141 in SELC would be buried in the interior of the molecule. On the other hand, Tyr-154 should be quite reactive, since it appears to be in a hydrophilic region following helix H. The reactivities of Tyr-141 and Tyr-154 may be altered when SELC is bound to myosin heavy chains. In rabbit skeletal muscle myosin, it has been shown that a 14-residue COOH-terminal segment (homologous to residues 138–151 of SELC) of the ELC is essential for binding to the heavy chain (Ueno et al., 1985; Morita et al., 1985).

SELC plays a special role in the myosin-linked calcium regulation of scallop muscle contraction (Ashiba & Szent-György, 1985). This behavior apparently is not shared by the vertebrate muscle ELCs, perhaps because of differences in the calcium binding properties of the ELCs. To test this hypothesis, a detailed comparative analysis of the ELC sequences shown in Figure 2 was carried out. By use of the Protein Identification Resource RELATE program (Dayhoff et al, 1983), each sequence was searched for 12-residue segments similar to a hypothetical, ancestral Ca binding sequence (see Figure 4). It should be noted that this type of analysis by itself cannot reliably estimate the binding affinity or specificity of a particular site for calcium, magnesium, or other divalent cations (Herzberg & James, 1985b). Furthermore, analysis of the crystal structure of intestinal calcium binding protein (Szebenyi et al., 1981) has shown that calcium binding can be maintained even when an ancestral site has been significantly altered by mutations. Experimental binding studies on ELCs are in progress, but the accuracy of the predictions may be fully tested only when the crystal structures become available.

With the procedure described above, three potential calcium binding sites were identified in SELC (see Figure 4). The most likely candidate for binding is the ancestral site III. This is also the case in the other ELCs, as has been pointed out previously (Matsuda, 1983; Matsuda et al., 1981b). However, the SELC site may have an increased affinity for calcium because it is the only known ELC that contains an amino acid residue (serine-102) capable of contributing a side-chain oxygen at the –X axis of the ancestral site in domain III. This may be a unique and functionally important feature of molluscan ELCs: our preliminary sequence results (unpublished) on the ELC of the clam *Mercenaria mercenaria* show that Ser-102 is retained. The SELC sequence also shows some potential for calcium binding at the ancestral site IV. The other ELCs would seem to be less likely to bind calcium at site IV, since they have fewer charged and oxygen-containing side chains in this segment. Ancestral sites I and II of the ELCs appear to have lost the ability to bind calcium, with the possible exceptions of site I of the vertebrate skeletal and gizzard (but not cardiac) ELCs. The third potential calcium binding site of SELC is in an unexpected location: the G helix, immediately preceding ancestral site IV. This would seem to be an example of convergent evolution, since the ancestral protein and most of its present-day descendencts do not have a potential binding sequence in the G helix. It is tempting to speculate that during the course of evolution invertebrate ELCs developed an elongated G helix (see Figure 2) and that further mutations in the SELC sequence resulted in the formation of a new calcium binding site. It may be possible that two weak and/or incomplete binding sites in domain IV somehow cooperate (and perhaps influence the site in domain III) to produce a specific, high-affinity calcium binding site that is responsible for myosin-linked regulation. Comparative sequence studies on ELCs from other calcium-sensitive myosins will reveal how highly conserved (and therefore how functionally important) this region is. The G helix of the *Drosophila* ELC probably does not bind calcium, since it has two fewer negative charges than that in SELC and also lacks the critical oxygen-containing side chain at the –Z position (Collins, 1976b).

Supplementary Material Available

Tables I and II containing amino acid composition and amino acid sequence data and Figures 5–8 showing chromatograms of the separation of the cyanogen bromide and tryptic

peptides (9 pages). Ordering information is given on any current masthead page.

Registry No. Myosin (*Aquipecten irradians* essential light chain), 104848-54-6; Ca, 7440-70-2.

REFERENCES

Alexis, M. N., & Gratzer, W. B. (1978) *Biochemistry 17*, 2319–2325.

Ashiba, G., & Szent-Györgyi, A. G. (1985) *Biochemistry 24*, 6618–6623.

Ashiba, G., Asada, T., & Watanabe, S. (1980) *J. Biochem. (Tokyo) 88*, 837–846.

Baba, M. L., Goodman, M., Berger-Cohn, J., Demaille, J. G., & Matsuda, G. (1984) *Mol. Biol. Evol. 1*, 442–455.

Babu, Y. S., Sack, J. S., Greenhough, T. G., Bugg, C. E., Means, A. R., & Cook, W. J. (1985) *Nature (London) 315*, 37–40.

Bagshaw, C. R. (1977) *Biochemistry 16*, 59–67.

Bagshaw, C. R., & Kendrick-Jones, J. (1979) *J. Mol. Biol. 130*, 317–330.

Benoit, R., Daubas, P., Akimenko, M.-A., Cohen, A., Garner, I., Guenet, J.-L., & Buckingham, M. (1984) *Cell (Cambridge, Mass.) 39*, 129–140.

Bidlingmeyer, B . A., Cohen, S. A., & Tarvin, T. L. (1984) *J. Chromatogr. 336*, 93–104.

Burke, M., & Wang, H. L. (1982) *Eur. J. Biochem. 124*, 177–182.

Chantler, P. D., & Szent-Györgyi, A. G. (1980) *J. Mol. Biol. 138*, 473–492.

Cohen, S. A., Bidlingmeyer, B. A., & Tarvin, T. L. (1986) *Nature (London) 320*, 769–770.

Collins, J. H. (1974) *Biochem. Biophys. Res. Commun. 58*, 301–308.

Collins, J. H. (1976a) *Nature (London) 259*, 699–700.

Collins, J. H. (1976b) *Symp. Soc. Expt. Biol. 30*, 303–334.

Collins, J. H. Potter, J. D., Horn, M. J., Wilshire, G., & Jackman, N. (1973) *FEBS Lett. 36*, 268–272.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524–545.

Falkenthal, S., Parker, V. P., Mattox, W. W., & Davidson, N. (1984) *Mol. Cell. Biol. 4*, 956–965.

Flicker, P. F., Walliman, T., & Vibert, P. (1983) *J. Mol. Biol. 169*, 723–741.

Frank, G., & Weeds, A. G. (1974) *Eur. J. Biochem. 44*, 317–334.

Grand, R. J. A., & Perry, S. V. (1983) *Biochem. J. 211*, 267–272.

Hardwicke, P. M. D., & Szent-Györgyi, A. G. (1985) *J. Mol. Biol. 183*, 203–211.

Hardwicke, P. M. D., Walliman, T., & Szent-Györgyi, A. G. (1982) *J. Mol. Biol. 156*, 141–152.

Hardwicke, P. M. D., Walliman, T., & Szent-Györgyi, A. G. (1983) *Nature (London) 249*, 478–482.

Heinrikson, R. L., & Meredith, S. C. (1984) *Anal. Biochem. 136*, 65–74.

Henry, G. D., Trayer, I. P., Brewer, S., & Levine, B. A. (1985) *Eur. J. Biochem. 148*, 75–82.

Herzberg, O., & James, M. N. G. (1985a) *Nature (London) 313*, 653–659.

Herzberg, O., & James, M. N. G. (1985b) *Biochemistry 24*, 5298–5302.

Hewick, R. M., Hunkapillar, M. W., Hood, L. E., & Dreyer, W. J. (1981) *J. Biol. Chem. 256*, 7990–7997.

Kendrick-Jones, J., & Jakes, R. (1976) in *Myocardial Failure* (Ricker, G., Weber, A., & Goodwin, J., Eds.) International Boehringer Symposium, Munich, pp 28–40, Springer-Verlag, Berlin.

Kendrick-Jones, J., & Scholey, J. M. (1981) *J. Muscle Res. Cell Motil. 2*, 347–372.

Kendrick-Jones, J., Lehman, W., & Szent-Györgyi, A. G. (1970) *J. Mol. Biol. 54*, 313–326.

Kendrick-Jones, J., Szentkiralyi, E. M., & Szent-Györgyi, A. G. (1976) *J. Mol. Biol. 104*, 747–775.

Konno, K., & Watanabe, S. (1985) *J. Biochem. (Tokyo) 97*, 1645–1651.

Kretsinger, R. H., & Nockolds, C. E. (1973) *J. Biol. Chem. 248*, 3313–3326.

Landon, M. (1977) *Methods Enzymol. 48*, 145–149.

Maita, T., Umegane, T., Kato, Y., & Matsuda, G. (1980) *Eur. J. Biochem. 107*, 565–575.

Maita, T., Umegane, T., & Matsuda, G. (1981) *Eur. J. Biochem. 114*, 45–49.

Maita, T., Konno, K., Ojima, T., & Matsuda, G. (1984) *J. Biochem. (Tokyo) 95*, 167–177.

Matsuda, G. (1983) *Adv. Biophys. 16*, 185–218.

Matsuda, G., Maita, T., & Umegane, T. (1981a) *FEBS Lett. 126*, 111–113.

Matsuda, G., Maita, T., Kato, Y., Chen, J.-I., & Umegane, T. (1981b) *FEBS Lett. 135*, 232–236.

Miyanishi, T., Maita, T., Morita, F., Kondo, S., & Matsuda, G. (1985) *J. Biochem. (Tokyo) 97*, 541–551.

Morita, F., Kondo, S., Tomari, K., Minowa, O., Ikura, M., & Hikichi, K. (1985) *J. Biochem. (Tokyo) 97*, 553–561.

Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M., & Ogata, K. (1984) *Nature (London) 308*, 333–337.

Periasamy, M., Strehler, E. E., Garfinkel, L. I., Gubits, R. M., Ruiz-Opazo, N., & Nadal-Ginard, B. (1984) *J. Biol. Chem. 259*, 13595–13604.

Simmons, R. M., & Szent-Györgyi, A. G. (1978) *Nature (London) 273*, 62–64.

Simmons, R. M., & Szent-Györgyi, A. G. (1985) *J. Physiol. (London) 358*, 47–64.

Stafford, W. F., & Szent-Györgyi, A. G. (1978) *Biochemistry 17*, 607–614.

Stafford, W. F., Szentkiralyi, E. M., & Szent-Györgyi, A. G. (1979) *Biochemistry 18*, 5273–5280.

Sundralingam, M., Bergstrom, R., Strasburg, G., Rao, S. T., Roychowdhury, P., Greaser, M., & Wang, B. C. (1985a) *Science (Washington, D.C.) 227*, 945–948.

Sundralingam, M., Drendel, W., & Greaser, M. (1985b) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7944–7947.

Szebenyi, D. M. E., Obendorf, S. K., & Moffat, K. (1981) *Nature (London) 294*, 327–332.

Szent-Györgyi, A. G., Szentkiralyi, E. M., & Kendrick-Jones, J. (1973) *J. Mol. Biol. 74*, 179–203.

Szentkiralyi, E. M. (1984) *J. Muscle Res. Cell Motil. 5*, 147–164.

Ueno, H., Katoh, T., & Morita, F. (1985) *J. Biochem. (Tokyo) 97*, 1785–1793.

Walliman, T., & Szent-Györgyi, A. G. (1981) *Biochemistry 20*, 1188–1197.

Walliman, T., Hardwicke, P. M. D., & Szent-Györgyi, A. G. (1982) *J. Mol. Biol. 156*, 153–173.

Wells, C., & Bagshaw, C. R. (1985) *Nature (London) 313*, 696–697.

Winkelman, D. A., Almeda, S., Vibert, P., & Cohen, C. (1984) *Nature (London) 307*, 758–760.