Characterization of myosin heavy chains from the mantle muscles of cephalopods.

Guo-Feng Wang[1], Shugo Watabe[2], Yoshihiro Ochiai[3*]

[1]Graduate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo, Tokyo 113-8657, Japan

[2]School of Marine Bioscience, Kitasato University, Minami-ku, Sagamihara, Kanagawa 252-0373, Japan

[3]Graduate School of Agricultural Science, Tohoku University, Aramaki, Aoba, Sendai, Miyagi 980-0845, Japan

Abstract

Myosin heavy chain (MyHC) genes from the mantle muscles of three cephalopod species, namely, golden cuttlefish *Sepia esculenta*, Japanese common squid *Todarodes pacificus*, and spear squid *Loligo bleekeri* were subjected to synthesis of first strand cDNA by reverse transcription-PCR. As a result of sequencing of the obtained cDNA fragments, the MyHCs from *S. esculenta* and *L. bleekeri* were found to consist of 1936 amino acids, while *T. pacificus* MyHC consisted of 1939 residues. The sequences were found to be 73%, 55%, and 41% identical to those of the Ca^{2+}-regulated scallop MyHCs, human skeletal MyHC, and chicken gizzard (smooth muscle) MyHC, respectively. The cephalopod MyHCs conserved ATP- and actin-binding sites as well as light chain binding sites in the head (S1) region with the clear differences in sequence homology from those of higher organisms, and typical α-helical heptad repeats involved in the formation of coiled-coil structure in the rod region. The sequences of the two surface loops at the 25/50 kDa and 50/20 kDa junctions of the motor domain were found to be conserved among cephalopods compared to those from the other sources. In the phylogenetic tree based on the amino acid sequences, the cephalopod MyHCs formed a cluster, clearly different from the other invertebrate counterparts, in accordance with the results of tertiary structure modeling of the head regions.

Keywords: Myosin, Muscle, Cephalopod, Structure, Phylogeny.

Accepted on 15 November, 2017

Introduction

Among the proteins in muscle, myosin is one of the principal components and occupies the major part of the contraction apparatus, myofibril [1]. Its structural and functional properties have been intensively studied due to its importance in the contractile process in muscle. Muscle myosins and those found in other tissues resemble each other both structurally and functionally, belonging to class II, thus are termed "conventional myosins" [2]. Only these isoforms are able to associate to form thick filaments *via* their long coiled-coil tails. The molecule of conventional myosin is a hexameric protein composed of a pair of heavy chain subunits (approximately 200 kDa) and two pairs of light chain subunits (approximately 20 kDa). The N-terminal halves of the heavy chains fold into two globular head domains, while the C-terminal halves of the heavy chains dimerize to form an α-helical coiled-coil rod domain. The head domains contain the active site for ATPase and the actin-binding site. The rod is responsible for the assembly of myosin molecules into thick filaments under physiological condition. The myosin heavy chain (MyHC) can be fragmented into light meromyosin (LMM), which contains only tail (rod) sequences and heavy meromyosin (HMM), which contains a part of the tail and the two head regions by tryptic digestion. Further digestion cleaves HMM into the subfragment-1 (S1) and subfragment-2 (S2). HMM consists of the head and S2 of the HMM tail portion. The sequence of head region is known to be highly conserved. The S1 region consists of a core motor (catalytic) domain with a relatively open ATP binding pocket and an actin binding interface followed by a 'converter' region which links the core motor domain to the 'lever arm' [3-5]. The rod region is responsible for bending of the molecule and the assembly into filaments, and the sequence is less conserved than the head region.

Muscle contraction is tightly coupled to concentrations of intracellular Ca^{2+}. However, Ca^{2+} regulation is mediated by a number of different means. In vertebrate smooth muscle/non-muscle myosin IIs, phosphorylation of the regulatory light chain (RLC) by a calcium/calmodulin dependent light chain kinase switches on myosin-actin interaction [6,7]. RLC phosphorylation also regulates the ability of these myosins to assemble into filaments, although RLCs of non-regulatory vertebrate striated/cardiac myosins are phosphorylated by a similar kinase. Regulation in these muscles is controlled by the troponin/tropomyosin complex on the actin filaments [8].

MyHCs are worth intensive investigation, because tremendous amounts of data have been accumulated for this protein, making it easy to compare the variation in amino acid sequence related to their functional diversity and molecular evolution. Although the structural and functional properties of vertebrate MyHCs have been studied intensively, the numbers of reports regarding invertebrate counterparts are much less.

Cephalopods such as squids and cuttlefishes are able to cruise at high speed, and show various activities unlike the other classes of the phylum *Mollusca*. They also give rise to unique musculature, namely, obliquely striated muscle, the formation mechanism of which remains unexplained clearly. They are considered to have succeeded in specializing locomotive devices. They also show peculiar regulation system of muscle contraction, i.e., myosin-linked regulation where Ca^{2+} directly binds to RLC, thereby, initiating the interaction of myosin with actin (Ca^{2+} regulated myosin). The information on cephalopod myosins, however, has been fragmentary [9-13]. To investigate the structure of cephalopod muscle MyHCs, the essential study is to determine the cDNA sequences of cephalopod muscle MyHCs. However, it requires a tremendous effort, because MyHCs consist of about 2,000 amino acids. To date, for the molluscs, MyHC cDNAs or genes have been sequenced in scallops (genera *Argopecten*, *Patinopecten*, *Pecten*, and *Placopecten*) and one cephalopoda genus *Loligo*. In the present study, the cDNA nucleotide sequences of MyHCs from three representative cephalopod species have been determined in order to explore new insights into comparative structural features of cephalopod muscle. Estimation of the tertiary structure was performed by homology modelling. It was demonstrated that cephalopod MyHCs generally share the common structure all along the entire molecule together with the dynamic substitutions in the specific regions.

Materials and Methods

Live specimens of golden cuttlefish *Sepia esculenta* (body weight: 267 g, full length 21 cm), Japanese common squid *Todarodes pacificus* (body weight: 303 g, full length 38 cm), and spear squid *Loligo bleekeri* (body weight: 317 g, full length 46 cm) were slaughtered at the Kensaki fishing port, Kanagawa Prefecture, Japan. The mantles were dissected, and the skins were carefully removed. The sampled pieces of mantle were cut into small pieces with a pair of scissors, immediately kept in RNA later (QIAGEN, Hilden, Germany), and transported to the laboratory in ice. All the chemicals used in this study were of reagent grade purchased from Wako Chemicals (Osaka, Japan).

Total RNA isolation

Total RNA extraction was performed by using Isogen (Nippon Gene, Tokyo, Japan) according to the manufacturer's protocols with a slight modification as described as follows. The sample tissues were cut into smaller pieces with dissecting scissors and homogenized at high speed with 1 ml of Isogen in 1.5 ml plastic tubes. The tubes were then swirled using a vortex and stored at room temperature for 5 min. An aliquot of 0.2 ml chloroform (99.5%) was poured into the tubes, which were subsequently agitated vigorously for 15 s and stored at room temperature for 2 to 3 min. The tubes were then centrifuged at 16,000 × g for 15 min at 4℃. The uppermost aqueous layers containing the extracted RNA, were transferred into new tubes, added 0.5 ml of isopropanol (99.5%), and stored at -20℃ for overnight to precipitate RNA, which was subsequently collected by centrifugation at 16,000 × g for 30 min at 4℃. All aqueous phase was discarded and the precipitated RNA was washed with 1 ml of 70% ethanol. The tubes were again centrifuged at 4,600 × g for 10 min at 4℃.

Alcohol was discarded and the tubes were dried briefly till the alcohol residues were completely evaporated. Finally, the RNA pellets were resuspended in 25 µl of sterile distilled water. The concentration of each sample was determined by a spectrophotometer.

First strand cDNA synthesis

The first strand cDNA synthesis was performed by using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific, Waltham, MA, USA). The solution containing 5 µg of RNA was diluted to 7 µl with distilled water, and mixed with 5 µl of 10 µM AP primer and 1 µl of 10 µM dNTP. The mixture was incubated at 65℃ for 5 min and transferred into 0℃ for 2 min. For cDNA synthesis, the mixture was subsequently mixed with 4 µl of 5 x First Strand Buffer, 1 µl of 0.1 mM dithiothreitol (DTT) and 1 µl of Super Script III Reverse Transcriptase, kept at 48℃ for 1.5 h, and then incubated at 70℃ for 15 min. Finally, the solution was digested by 1 µl of RNase H (Thermo Fisher Scientific) at 37℃ for 30-60 min.

cDNA cloning

A PCR-based cloning strategy was used. Initially, several pairs of degenerate primers were designed by referring to the conserved regions of the available MyHC cDNA sequences from GenBank (Table 1). Other degenerate primers and specific primers were designed according to the subsequent sequencing results. For the 3' end, a universal AUAP primer was adopted as the reverse primer. These primers were synthesized by Eurofins (Val Fleuri, Luxemburg).

By using these primers, PCR was performed by Ex taq DNA polymerase (TaKaRa Bio Inc., Kusatsu, Shiga, Japan). The reaction mixture (the total volume of 20 µl) contained 1 µl of first strand cDNA as a template, 2 µl of 10 x Ex Taq buffer, 1.6 µl of 2.5 mM dNTP, 1 µl of 10 µM forward primer, 1 µl of 10 µM reverse primer, 0.2 µl Ex taq and 13.2 µl distilled water. PCR consisted of 30 cycles of denaturation at 94℃ for 5 s, annealing at 55-62℃ according to the melting temperature (Tm) of each pair of primers for 30 s, and extension at 72℃ for 1 min. The last cycle employed an extension of 5 min. For purification, the amplified PCR products were precipitated with 50 µl solution containing 2.5 volume of 95% ethanol and 0.1 volume of 3 M sodium acetate (pH 4.8) by centrifugation at 15,000 rpm for 30 min. The supernatant was removed, and the pellets were washed with 50 µl of 70% ethanol and then centrifuged at 15,000 rpm for 10 min. Ethanol was discarded and the tubes were dried briefly by leaving upside down. Finally, the pellets were resuspended into 10 µl of distilled water. Three microliters of the purified PCR product suspension were subjected to 2% agarose gel electrophoresis, and the bands were stained with SYBR Green (Thermo Fisher Scientific).

Sequencing

The PCR products containing the target bands were subcloned into pGEM-T vectors using pGEM-T Vector Systems (Promega, Fitchberg, WI, USA) according to the manufacturer's instructions. The reaction tubes were incubated overnight at 4℃ to have the maximum number of transformants. The resultant ligation products were then transformed into *Escherichia coli* JM109 competent cells. The cells were subsequently plated onto agar plates and incubated at 37℃ overnight. The individual recombinant white clones were picked up and subjected to insert-check PCR amplification.

The PCR products were run on 2% agarose gel. The positive inserted clones were grown overnight at 37℃ in Luria-Bertani (LB) liquid medium containing ampicillin with vigorous shaking. Isolation and purification of inserted plasmid DNA from the cells were performed using GenElute Plasmid Miniprep Kit (Sigma-Aldrich, St. Louis, MI, USA) according to the manufacturer's protocols.

Prior to sequencing, the purified plasmid DNA were subjected to labeling PCR using BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific), and sequencing was conducted using ABI Prism 3100 Genetic Analyzer. The first strand cDNAs were subsequently synthesized by RT-PCR. Using the first strand cDNAs as the templates, five cDNA fragments, namely F1, F2, F3, F4 and F5 (Table 1), were amplified by PCR using degenerated primers designed according to the conserved regions of the molluscan MyHC cDNA sequences obtained from Gen Bank database.

The sequences of these fragments were determined. Then, to amplify the gap regions between these 5 fragments (gap2~gap5), degenerated primers or specific primers which match the end regions of these fragments were designed. Because the lengths of gap4 and gap5 were too long to be sequenced in a single sequence read, another two degenerated primers were designed according to the conserved regions at the middle of these two gaps. By using these primers, the cDNA fragments of the gap regions were amplified by PCR.

The sequences were determined in succession. For the 3' end, PCR was performed by using the primer 6 gf designed according to the 3' end sequences of the F5 fragments and the AUAP primer. Another forward primer 3 utr was also used for primer walking. For the 5' end, several degenerated primers were designed from the start codon (ATG), according to the conserved regions of the available molluscan MyHC cDNA sequences. Finally, the primer start was shown to be able to amplify the 5' end cDNA fragments succeeded in *L. bleekeri* and *S. esculenta* with the reverse primer (1 gr). Therefore, the sequences of the 3' and 5' regions were also determined. By editing these overlapping sequences, the full length MyHC cDNA sequences could be obtained.

Bioinformatical analyses

BLAST program was used to align the amino acid sequences [14] based on the one of *S. esculenta*. A phylogenetic tree was constructed by a neighbor-joining method based on the amino acid sequences using multiple sequence alignment CLUSTAL W and MEGA 4 programs [15,16].

Table 1. *Locations and nucleotide sequences of the obtained fragments and primers.*

Fragment	Location* (bp)	Length	Primer ID	Nucleotide sequences (5'-3') **
Gap1	1-387	387	Start	ATGAMYATSGAYTWYARTGA
			1 gr	AAGACGTCTGTAGGGGTTGA
F1	368-757	390	1 f	TCAACCCCTACAGACGTCT
			1 r	GGATGAATTTACCGAATC
Gap2	740-1319	580	2 gf	GATTCGGTAAATTCATCC
			2 gr	CAGTTGAACATACGATCGTA
F2	1300-1631	332	2 f	TACGATCGTATGTTCAACTGG
			2 r	TTGGGGAACATGCACTCYTC
Gap3	1591-2167	577	3 gf	GGTATCTTGTCCATCCTTGA
			3 gr	AGTATCTCTGCTTGAACTCA
F3	2138-2454	317	3 f	TSATCTACTCTGAGTTCAARCA
			3 r	CCTGAYGTTACGCTGGAT
Gap4a	2437-3084	648	4 gaf	ATCCAGCGTAACRTCAGG
			4 gar	CTGTTCAAGCTTGGYCTTCA
Gap4b	3065-3644	580	4 gbf	TGAAGRCCAAGCTTGAACAG
			4 gbr	CTGCTCTTMACCTTCTGGA
F4	3574-4136	563	4 f	CGCA AGAAGMACCAGGAWGCC
			4 r	GCACCCTCGCTCTCRAA
Gap5a	4099-4509	411	5 gaf	GTACAACAATGGAGRTCC
			5 gar	GTTCTTGTTCTCCCTCCT
Gap5b	4492-4748	257	5gbf	AGGAGGGAGAACAAGAAC
			5 gbr	TCGAATTCCTCTTCCTTCTC
F5	4729-5189	461	5 f	GAGAAGGAAGAGGAATTCG
			5 r	TCRAGTTTCCTCTTCTGGC
Gap6	5170-end	638	6 gf	AGCCAGAAGAGGAAACT
			3 utr	ARYCGYAGCTCYGTHTCTG
			AUAP	GGCCACGCGTCGACTAGTAC

(*The positions are referred to the cDNA sequences of Loligo pealei myosin heavy chain isoform A. **K=G+T, M=A+C, R=A+G, S=C+G).

The sequence of rabbit uterus and chicken gizzard MyHC was taken as the out group to root the tree. Accession numbers of amino acid sequences of MyHCs are as follows: AF197336 for longfin inshore squid *Loligo pealei*, HG421291 for octopus *Octopus bimaculoides*, U59294 and U59295 for sea scallop *Placopecten magellanicus* striated and catch muscles, respectively, AF134172 for great scallop *Pecten maximus*, X55714 and U09782 for bay scallop *Argopecten irradians* striated and catch muscles, respectively, HE601631 for blood

fluke *Schistosoma mansoni*, M61229 for fruit fly *Drosophila melanogaster*, M74066 for nematode *Onchocerca volvulus* body wall, J01050 for nematode *Caenorhabditis elegans*, AY165122 and AK000947 for human *Homo sapiens* skeletal (adult) and embryo muscles, U32574 for rabbit *Oryctolagus cuniculus* skeletal muscle, D89991 for carp *Cyprinus carpio* skeletal muscle, AB039672 for white *croaker Pennahia argentata*, skeletal muscle, B076182 for chum salmon *Oncorhynchus keta* skeletal muscle, M77812 for rabbit *O. cuniculus* uterus, and X06546 for chicken *Gallus gallus* gizzard.

Modelling of the tertiary structure of S1 domains was carried out with MolFeat program (ver.3.0, FiatLux, Tokyo, Japan), based on the atomic structure of scallop S1 (PDB 1L2O).

Results

The full length cDNA sequences of the three cephalopod species (*S. esculenta, T. pacificus* and *L. bleekeri*) were determined for the first time. The sequences have been registered to DDBJ databank with the accession numbers, EU687253, GU338005 and EU687252, respectively. The MyHC cDNAs from *L. bleekeri* and *S. esculenta* consisted of 6236 and 6258 bp, respectively (data not shown). Both cDNAs contained an open reading frame of 5808 nucleotides, encoding 1936 amino acids. The major part of *T. pacificus* MyHC cDNA contained an additional undetermined stretch of about 90 bp at the 5' end, and the open reading frame consisted of 5817 bp encoding 1939 residues, three residues longer compared to those from the other two species (Figure 1). The sequences contained four skip residues at strictly conserved positions of the coiled-coil region like the reported sequence of *L. pealei* [10]. Moreover, one isoform was found to be present for *T. pacificus* MyHC. The partial cDNA sequences of the F4 and gap6 fragments indicated that this isoform possessed lower occurrence rate and lower homology compared to the other obtained sequences. It was found for the first time that the isoform has an unprecedented prolonged tail at the C terminus (Figure 1). The amino acid sequences of the decapod MyHCs are highly conserved including the reported sequence of *L. pealei* MyHC, giving a cluster in the phylogenetic tree drawn based on the deduced amino acid sequences (Figure 2).

The full sequence of *S. esculenta* MyHC was shown to be around 93-96% identical to those the other cephalopod muscle MyHCs. The full sequences of *L. bleekeri*, and *S. esculenta* MyHCs were shown to be 73% identical to those of the Ca^{2+}-regulated scallop (*Pecten maximus* and *A. irradians*) MyHCs, and 55% identical to that of human skeletal MyHC of none Ca^{2+} regulation, and 41% identical to smooth muscle MyHCs of rabbit uterus and chicken gizzard (Table 2).

The sequence identities were dependent on the respective regions of MyHC (S1, S2 and LMM) (Table 2). The sequence identities of the five ATP-binding sites of MyHCs from various vertebrate and invertebrate species (human, carp, fruit fly, scallop, octopus, nematode together with the three cephalopod species) were 96.4, 76.8, 93.9, 49.4 and 45.6%, respectively, from the N terminus, indicating that the homologies of the sites close to the C terminus are relatively low.

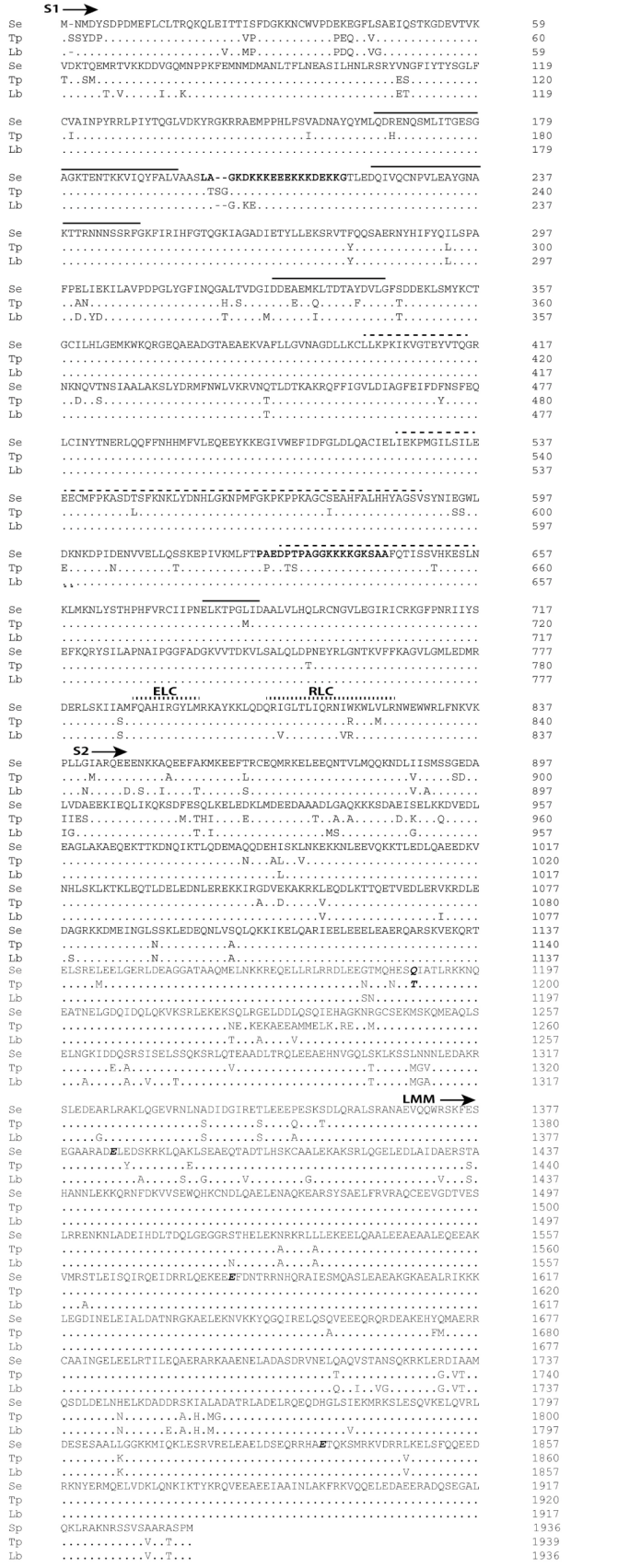

Figure 1. *Alignment of the amino acid sequences of myosin heavy chains from the cephalopods S. esculenta (Se), T. pacificus (Tp) and L. bleekeri (Lb). The alignment was performed based on the sequence of S. esculenta myosin heavy chain, and the same residues were replaced by dots. The residue numbers from the N termini are shown on the right. The junctions regions of 25 K/50 K and 50 K/20 K domains (Loop 1 and Loop 2, respectively) are boldfaced. Actin-binding sites and ATP-binding sites are indicated with solid lines and broken lines just above the sequences, respectively. The binding sites of essential light chain (ELC) and regulatory light chain (RLC) are shown by dotted lines. The arrows indicate the start residues of S1, S2 and LMM regions. The four skip residues are indicated by boldfaced italic letters.*

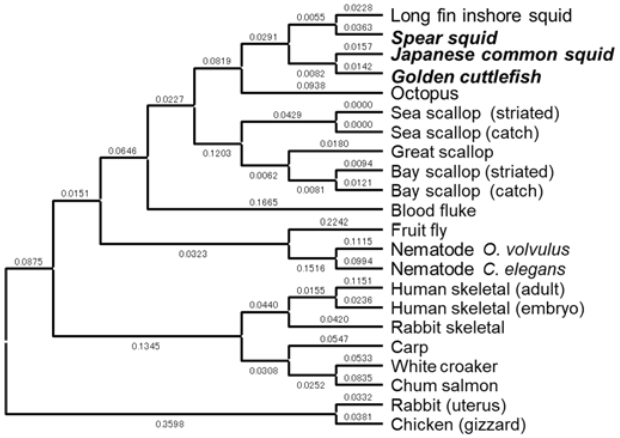

Figure 2. *Phylogeny of myosin heavy chains drawn based on the amino acid sequences of the whole molecule. The bootstrap values from 1000 replicate analyses are indicated at each node of the tree.*

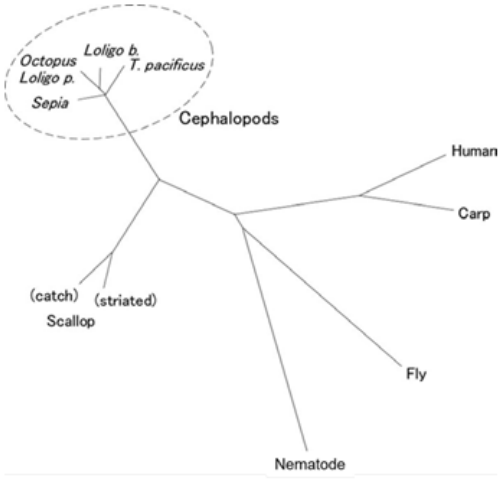

Figure 3. *Unrooted phylogenetic tree drawn based on the amino acid sequences of S1 region of myosin heavy chain.*

On the other hand, the sequence identities of the three actin-binding sites from the N terminus, are 93.9, 57.8 and 35.8%, respectively, indicating the value is quite low for that nearest to the C terminus.

When the phylogeny of S1 regions was examined, the cephalopod were found to form a cluster (Figure 3), suggesting that their muscle myosins have evolved in such a way to adapt to their excellent locomotory activities. The cephalopod S1s were found to be clearly distant from bivalves (scallops) with Ca^{2+}- regulated myosin. The modelled tertiary structure of S1 regions of MyHC based on the scallop structure (PBD 1L2O) is shown in Figure 4. In the figure, the structures of scallop and squid S1s were mostly overlapped, but the local structures were clearly different from each other.

Especially, the extent of bending of the lever arms (corresponding to the light chain binding sites) was clearly different between the two structures. This fact suggests the difference in the contractile mechanism(s) of muscle between the mollusks.

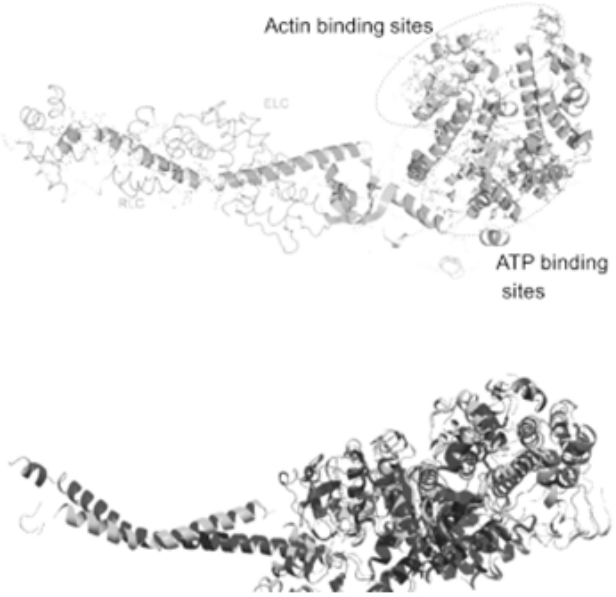

Figure 4. *Tertiary structure of myosin heavy chain S1 region (PDB ID 1L2O). (top) a ribbon model of S1 region with light chain subunits (RLC, ELC) around the lever arm. Actin and ATP binding sites are encircled. (bottom) superimposed images of the S1 regions of scallop (light gray) and spear squid (deep gray).*

Table 2. *Identities of amino acid sequences in the respective regions of myosin heavy chains from various vertebrate and invertebrate species.*

Portion of myosin		Cuttlefish	Inshore squid	Bay scallop	Sea scallop	Croaker	Human	Chicken smooth
Full	Spear squid	94	96	73	73	54	55	41
	Cuttlefish		93	73	73	55	55	40
S1	Spear squid	94	95	76	73	58	56	50
	Cuttlefish		93	76	73	58	57	50
S2	Spear squid	94	95	67	66	54	53	35
	Cuttlefish		92	68	67	55	55	35
LMM	Spear squid	94	95	67	66	54	53	35
	Cuttlefish		92	68	67	55	55	35

Discussion

As shown in Table 2, the squid MyHCs shows 73, 54, and 55% homology with those of scallop, fish (white croaker) and human, respectively. The rod region seems to be slightly less conservative than the head region. The three cephalopod muscle MyHCs of *S. esculenta*, *T. pacificus*, and *L. bleekeri* were 73-76% identical to the scallop counterparts of the S1, 66-67% identical of S2, and 66-67% identical of LMM (Figure 2). Therefore, compared with the scallop MyHCs, the S1 region of cephalopod is the most conserved, and the S2 and LMM are less conserved than the S1 region. This is quite in good accordance with the differences in the tertiary structures as shown in Figure 4. When compared with the vertebrate skeletal MyHCs or smooth muscle MyHCs, the identities reveal that, although the S1 region is still the most conserved, the LMM is less conserved than the S2 region. This result indicates that, MyHCs of different muscle types are more divergent in the rod region, and each portion of the rod possessed different rate of amino acid substitution during molecular evolution and adaptation to the respective habitat and locomotory activities, although the motor domain and regulatory domain are located in S1 [17]. Then, the structures of these cephalopod muscle MyHCs are discussed below for each portion.

S1 region

The motor domain of S1 contains both ATP- and actin-binding sites and acts as the molecular motor that generates the contractile force. The motor domain contains the residues conserved in both the conventional and unconventional myosins [18], including the highly conserved ATP binding pocket, and the upper and lower faces of the 50 kDa cleft domain [19]. The ATP-and actin-binding sites as well as RLC binding site are highly conserved among various myosins. However, the two loops, Loops 1 and 2, known as the 25/50 kDa and 50/20 kDa domain junctions, respectively, and also as flexible and protease-sensitive sites [20] are less conserved. The conformations of these loops are considered to be related with the contractile activity and ATP turnover rate of myosin [21,22].

The N terminal side of Loop 1 shows the greatest divergence in the sequence and the length (Figure 1). Loop 2 also shows considerable variations, namely, 12 amino acids in the stretch of 33 residues (623^{rd}~655^{th} residues in the squid sequence) were replaced. It has been pointed out that these loops of the myosin motor domain are the least conserved regions and may account for the diversity of ATPase activity and motor activity of myosin [23]. Loops 1 and 2 are known to be rich in Gly and Lys residues. There are actually 2-3 Gly in the cephalopod counterparts, while 2-6 Gly residues in Loop 1 of fish (walleye pollack, white croaker and carp) and rabbit myosins. This may indicate that the flexibility of this loop is comparable among invertebrate and vertebrate animals. Ser residue, which is also found in the squid MyHC around the 180^{th} residue, is considered to be one of ATP binding residue [24], while Pro, a helix breaker, found in scallop Loop 1 is absent in the counterparts of squid myosin. Though the physiological role of Pro in Loop 1 has not been established so far, it may be related to the efficiency of ATP hydrolysis and behavior of myosin during muscle contraction [25]. On the other hand, the length of Loop 2 is slightly shorter in cephalopod MyHCs (21 residues) than the counterparts of higher vertebrates (28-31 residues). Therefore, it is likely the diversity in Loop 2 is correlated to actin-activated ATPase activity and muscle contraction speed.

ELC binding site shows relatively lower sequence homology (ca. 25%) compared to RLC binding site (ca. 50%). The light chain binding motif, i.e., IQ-motif (IQ***RG***R**Y) which binds to the EF hands of light chains [26], is partially conserved in the binding sites for RLC. Interestingly, Lys819 is conserved in squid MyHC. The side chain of this Lys (K817 in the scallop MyHC) interacts strongly with the RLC interlobe linker, connecting D and E helices of RLC in the scallop structure.

Although the light chain binding sites are well conserved between the squid and the scallop MyHCs, the only difference is in the ELC binding site, where Ile796 in scallop myosin is replaced by Met in the squid counterpart. The crystal structure of the scallop regulatory domain shows that this residue is present at the interface with ELC [26]. However, these amino acids share very similar profiles, and thus it is likely such a substitution hardly affects the structure and binding ability to ELC. All the variations in the light chain binding sites are considered to be conservative replacements of solvent exposed residues, except Asp833, which is Ser831 in scallop *Argopecten* myosin and Ala831 in that of *Placopecten*.

Rod region

The rod portion of MyHC consists of S2 and LMM. The sequence of squid S2 shows 58% and 56% homology to those of fish and human S2s, respectively (Table 2), indicating that the S2 is not the conservative region in MyHCs [27]. Squid S2 also shows typical features for α-helical coiled-coil motif, i.e. a seven residue repeat and a 28 residue repeating pattern [28]. The LMM portion of the squid MyHC is 66-67% identical to scallops, 53-54% identical to non-regulated vertebrate myosins, and 35% identical to smooth muscle myosins. The rod portion including LMM contains the typical heptad repeats and the skip residues, forming the coiled-coil structure [29,30].

The sequence required for thick filament formation consists of 29 residues located close to the C terminus, called the 'assembly competent domain (ACD)', which is also present in squid MyHC (1870^{th}-1898^{th} residues).

On the other hand, one of the skip residue (Thr) is present at around 1190th residue in the squid sequence. Although the substitutions by Gly are evenly distributed over the rod region, the substitutions seem to be more frequent in the coiled-coil portion. The physiological significance of such a substitution to Gly remains unclear, but it is suggested the substitutions by Gly would result in destabilization of coiled-coil structure. Cephalopod MyHCs seem to have evolved in such a way to adapt to their sophisticated locomotory activity.

Presence of isoforms

By sequencing multiple clones spanning the 25/50 kDa junction region of the *T. pacificus* MyHC, at least two isoforms of the squid MyHC were found (data not shown). One isoform with unprecedented prolonged tail at the C terminus was found to be present solely in *T. pacificus* MyHC. Such a long tail may be involved in the assembly of MyHC into the thick filament [31], and may affect the musculature and behavior of muscle. However, no isoform was detected for MyHCs from the other species. Since the nucleotide sequences of the two MyHCs were identical except the surface loops, it is likely that alternative RNA splicing of a single gene is responsible for both of these isoforms [32]. The overlap of the sequences derived from different clones would strengthen the argument that the spliced variant is considered not to be PCR artifacts. Assuming a similar exon structure in squid MyHC genes to those of the scallop MyHCs, exon 6 could be alternatively spliced between these isoforms. Distinct MyHC isoforms exist in different muscle and non-muscle cells to perform numerous physiological functions. Actually, fast muscle type fiber has been found in squid *L. pealei* [33]. Divergent sequences in specific functional domains of MyHC isoforms could partly determine the differences in the contractile properties of different muscle types.

Acknowledgment

This study was funded by a Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science (KAKENHI # 19380119) to Y.O. The authors are grateful to Dr. Seiji Akimoto, Kanagawa Prefectural Fisheries Research Institute for facilitating the sampling of the specimens used.

References

1. Geeves MA, Holmes KC. Structural mechanism of muscle contraction. Ann Rev Biochem. 1999:68:687-728.
2. Sellers JR. Myosins: a diverse superfamily. Biochim Biophys Acta Mol Cell Res. 2000;1496:3-22.
3. Rayment I, Rypniewski WR, Schmidt-Bäse K, R, et al. Three-dimensional structure of myosin subfragment-1: A molecular motor. Science. 1993;261:50-8.
4. Dominguez R, Freyzon Y, Trybus KM, et al. 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. 1998;94:559-71.
5. Houdusse A, Kalabokis VN, Himmel D, et al. Atomic structure of scallop myosin subfragment S1 complexed with MgADP: a novel conformation of the myosin head. Cell. 1999;97:459-70.
6. Tan JL, Ravid S, Spudich JA. Control of nonmuscle myosin by phosphorylation. Annu Rev Biochem. 1992;61:721-59.
7. Bresnick AR. Molecular mechanisms of nonmuscle myosin-II regulation. Current Opinion Cell Biol. 1999;11:26-33.
8. Farah CS, Reinach FC. The troponin complex and regulation of muscle contraction. FASEB J. 1995;9:755-67.
9. Anderson EJ, DeMont ME. The mechanics of locomotion in the squid Loligo pealei: locomotory function and unsteady hydrodynamics of the jet and intramantle pressure. J Exp Biol. 2000;203:2851-63.
10. Sirokman K, Szent-Györgyi AG, Matulef K, et al. Amino-acid sequence of squid myosin heavy chain. J Muscle Res Cell Motil. 1998;19:705-12.
11. Cortés-Ruiz JA, Pacheco-Aguilar R, Ramírez-Suárez JC, et al. Conformational changes in proteins recovered from jumbo squid (Dosidicus gigas) muscle through pH shift washing treatments. Food Chem. 2016;196:769-75.
12. O'Neall-Hennessey E, Reshetnikova L, Senthil Kumar VS, et al. Purification, crystallization and preliminary X-ray crystallographic analysis of squid heavy meromyosin. Acta Cryst. 2013;F69:248-52.
13. Shaffer JF, Kier WM. Tuning of shortening speed in coleoid cephalopod muscle: no evidence for tissue-specific muscle myosin heavy chain isoforms. Invertebrate Biol. 2016;135:3-12.
14. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nuc Acids Res. 1997;25:3389-402.
15. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nuc Acids Res. 1994;22:4673-80.
16. Tamura K, Dudley J, Nei M, et al. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol. 2007;24:1596-9.
17. Houdusse A, Cohen C. Structure of the regulatory domain of scallop myosin at 2A resolution: Implications for regulation. Structure. 1996;4:21-32.
18. Jamie M, Cope1 TV, Whisstock J, et al. Conservation within the myosin motor domain: implications for structure and function. Structure. 1996;4:969-87.
19. Fisher AJ, Smith CA, Thoden J, et al. Structural studies of myosin:nucleotide complexes: A revised model for the molecular basis of muscle contraction. Biophys J. 1995;68:19-28.
20. Yamamoto K, Sekine T. Interaction of myosin subfragment-1 with actin: I. Effect of actin binding on the susceptibility of subfragment-1 to trypsin. J Biochem. 1979;86:1855-62.
21. Murphy CT, Spudich JA. Dictyosterium myosin 25-50K loop substitutions specifically affect ADP release rates. Biochemistry. 1998;37:6738-44.
22. Perreault-Micale CL, Kalabokis VN, Nyitray L, et al. Sequence variations in the surface loop near the nucleotide binding site modulate the ATP turnover rates of molluscan myosins. J Muscle Res Cell Motil. 1996;17:543-53.
23. Spudich JA. How molecular motors work. Nature. 1994;372:515-8.

24. Hiratsuka T, Katoh T. Chemical identification of Serine 181 at the ATP-binding site of myosin as a residue esterified selectively by the fluorescent reagent 9-anthroylnitrile. J Biol Chem. 2003;278:31891-4.
25. Gauvry L, Ennion S, Ettelaie C, et al. Characterization of red and white muscle myosin heavy chain gene coding sequences from antarctic and tropical fish. Comp Biochem Physiol. 2000;127:575-88.
26. Xie X, Harrison DH, Schlichting I, et al. Structure of the regulatory domain of scallop myosin at 2.8 a resolution. Nature. 1994;368:306-12.
27. Togashi M, Kakinuma M, Hirayama Y, et al. cDNA cloning of myosin rod and the complete primary structure of myosin heavy chain of walleye pollack fast skeletal muscle. Fish Sci. 2000;66:349-57.
28. Offer ADG. Skip residues correlate with bends in the myosin tail. J Mol Biol. 1990;216:213-8.
29. Lachlan Mc, Karn L. Periodic charge distributions in the myosin rod amino acid sequence match cross-bridge spacings in muscle. Nature. 1982;299:226-31.
30. Straussman R, Squire JM, Ben-Ya'acov A, et al. Skip residues and charge interactions in myosin II coiled-coils: Implications for molecular packing. J Mol Biol. 2005;353:613-28.
31. Hodge TE, Cross R, Kendrick-Jones J. Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebrate nonmuscle myosin rod. J Cell Biol. 1992;118:1085-95.
32. Nyitray L, Jancso A, Ochiai Y, et al. Scallop striated and smooth muscle myosin heavy chain isoforms are produced by alternative RNA splicing from a single gene. Proc Natl Acad Sci. 1994;91:12686-90,
33. Kier WM, Curtin NA. Fast muscle in squid (Loligo pealei): contractile properties of a specialized muscle fibre type. J Exp Biol. 2002;205:1907 -16.

*Corresponding author

Yoshihiro Ochiai

Graduate School of Agricultural Science, Tohoku University,

Aramaki, Aoba, Sendai, Miyagi 980-0845,

Japan

E-Mail: yochiai@tohoku.ac.jp