Original Article

cDNA cloning of myosin heavy chain from white croaker fast skeletal muscle and characterization of its complete primary structure

Sung Ho Yoon,1 Makoto Kakinuma,1,a Yasushi Hirayama,1 Tsuneji Yamamoto2 AND Shugo Watabe1,*

1Laboratory of Aquatic Molecular Biology and Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo, Tokyo 113-8657, and 2The Association for Fish Gel Processing Industry of Japan, Minato, Tokyo 104-1124, Japan

SUMMARY: A cDNA library constructed from the dorsal fast skeletal muscle of white croaker Pennahia argentata was screened for myosin heavy chain using antibody raised against carp fast skeletal myosin. A full-length cDNA was further cloned by reverse transcription–PCR and 5’-RACE using first-strand cDNA as a template together with appropriate sets of primers. The entire cDNA consisted of 5986 nucleotides (nt) with 64 nt 5’-untranslated and 129 nt 3’-untranslated regions. This full-length cDNA had an open reading frame encoding a polypeptide of 1930 amino acid residues. Amino acid alignment with myosins from various vertebrates revealed some striking differences between fish and mammalian sequences which could be due to their position in the vertebrate evolutionary process. Hydrophilicity analysis revealed two different features of the myosin molecule: S1 heavy chain showed a mixed profile of hydrophobicity and hydrophilicity, whereas rod had a profile of only hydrophilicity. Comparison of amino acid sequences in the rod region between white croaker and walleye pollack showed a markedly high identity of 92%. White croaker myosin rod had a characteristic seven-residue (heptad) repeat (a, b, c, d, e, f, g), where positions a and d were normally occupied by hydrophobic residues, and positions b, c and f by oppositely charged residues, which may lead to interhelical electrostatic attractions stabilizing the coiled-coil of α-helices.

KEY WORDS: amino acid sequence, fast skeletal muscle, hydrophilicity, myosin heavy chain, nucleotide sequence, Pennahia argentata, white croaker.

INTRODUCTION

Myosin is one of the principal protein components of numerous contractile systems and comprises almost 50% of the myofibrillar protein in skeletal muscle of vertebrates.1 Its structural and functional properties have been intensively studied due to its importance in the contractile process in living muscle. The myosin molecule consists of two heavy chains, each of approximately 200 kDa, and four light chains of approximately 20 kDa. The N-terminal half of each heavy chain is folded into a globular head or subfragment-1 (S1), which contains actin and ATP binding sites. The remainder participates in a long rod-like structure of α-helical coiled-coil with a light meromyosin (LMM) region, responsible for thick filament formation of muscle, and a subfragment-2 (S2) region between S1 and LMM.2,3 Myosins derived from fish muscle are known to be thermally less stable than their mammalian counterparts and this thermostability may be related to habitat temperatures of fish, an observation that has considerable implications for fish processing and storage.4,5 However, the molecular basis of this variable stability is unknown.

White croaker is one of the most diversified bottom fish and its related species are distributed over regions as diverse as temperate to tropical waters. In addition, its importance for the food industry cannot be underestimated. Food technologists working in surimi production claim that myofibrillar proteins from white croaker are...
highly stable in chilling and frozen storage, making it a good candidate for surimi processing. Although not well demonstrated, white croaker meat has a reputation of being endowed with the best gel-forming ability among all fish species. Thus, white croaker is the primary species used in processing for Japanese high-class surimi, although the meat has only a moderately good flavor. It is well known that myosin, the most abundant muscle protein, is primarily responsible for the gel-forming ability of meat. However, it has remained ambiguous whether or not the structural properties of myosin from white croaker are different from those of fish species currently most utilized for surimi production such as walleye pollack.

As a first step toward resolving such ambiguity, this study was undertaken to clone and sequence a full-length myosin heavy chain cDNA from white croaker fast skeletal muscle.

MATERIALS AND METHODS

Materials

Specimens of white croaker *Pennahia argentata* were caught in Tokyo Bay. Dorsal fast skeletal muscle of the fish was carefully dissected while avoiding contamination with slow muscle, immediately frozen with liquid nitrogen and stored at –80°C until use.

Construction of cDNA library

Total RNA was prepared from 0.5 g of dorsal muscle of white croaker by the acid guanidinium thiocyanate-phenol-chloroform method. Poly(A)+ mRNA was isolated from total RNA using an oligo d(T) cellulose column (Amersham Pharmacia Biotech, Buckinghamshire, UK). Five micrograms of mRNA was used for double-stranded cDNA synthesis, which was carried out using Pharmacia cDNA synthesis kits with an oligo dT primer. The cDNA library was constructed using synthesized cDNA in a phage vector λZAP II according to the supplier’s instructions (Stratagene, La Jolla, USA), after ligating cDNA to a NotI/EcoRI linker (Amersham Pharmacia Biotech).

**cDNA cloning**

The cDNA library was screened using anti-carp myosin polyclonal antibody9 by the method of Sambrook et al.10 Vector pBluescript SK – and *Escherichia coli* strain XL1-Blue were purchased from Stratagene. *Escherichia coli* infected with the λZAP II cDNA library was cultured on an agar plate at 42°C for 4 h, and a nitrocellulose filter containing 10 mM isopropyl-β-d-thiogalactopyranoside was overlaid to the plate and incubated at 37°C for 4 h. The nitrocellulose filter replica was screened with anti-carp myosin polyclonal antibody. In vivo excision was carried out according to the manufacturer’s instructions (Stratagene). The DNA fragments were subcloned into a plasmid vector, pBluescript II SK+, using *E. coli* strain XL1-Blue as a host bacterium.

**Reverse transcription–PCR**

Reverse transcription–polymerase chain reaction (PCR) was performed to isolate cDNA clones encoding N-terminal portions of S1 heavy chain and C-terminal portions of S2. The PCR amplifications were performed using two sets of primers containing *Bam*HI recognition sequences (Table 1). The conditions for PCR were

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverse transcription–PCR</td>
<td></td>
</tr>
<tr>
<td>5MHC-RT1</td>
<td>5′-dTATGGTGTGCTGGAGGGTA-3′</td>
</tr>
<tr>
<td>3MHC-RT1</td>
<td>5′-dCTCCTGAAACAGTGACACCA-3′</td>
</tr>
<tr>
<td>5MHC-R-T2</td>
<td>5′-dCCAGAGTCAGAGACCTCAGGTCAACAA-3′</td>
</tr>
<tr>
<td>3MHC-RT2</td>
<td>5′-dAGCGGATCCAGAGCTTGTGTGGTCTAG-3′</td>
</tr>
<tr>
<td>5MHC-RT3</td>
<td>5′-dCCGGAGATCCGAAAGACTGTGAACACCAAGC-3′</td>
</tr>
<tr>
<td>3MHC-RT3</td>
<td>5′-dGATTGGATCCCAACATGTCAGGAGACCCGATG-3′</td>
</tr>
<tr>
<td>5MHC-SPACER</td>
<td>5′-dAGCTGGGATCCAGATATGAAACTCCAGCAG-3′</td>
</tr>
<tr>
<td>3MHC-SPACER</td>
<td>5′-dCTTGGGATCTCAGGACTCAGTACGGTG-3′</td>
</tr>
<tr>
<td>5′-End amplification</td>
<td></td>
</tr>
<tr>
<td>5RPN</td>
<td>5′-dGTCTCAATATCAGCAGTGCC-3′ (phosphorylation primer)</td>
</tr>
<tr>
<td>5RS1</td>
<td>5′-dGGGATCCAACTCTGAGCAGGAGCC-3′</td>
</tr>
<tr>
<td>5RS2</td>
<td>5′-dGGGATCCCTTTCCTTCTCCACCACACAC-3′</td>
</tr>
<tr>
<td>5RA1</td>
<td>5′-dGGGATCCAGATCGCTTCTTGCTTCTTGCTTCACCAC-3′</td>
</tr>
<tr>
<td>5RA2</td>
<td>5′-dGGGATCCGTAACGCGGAAGACTTGAGG-3′</td>
</tr>
</tbody>
</table>

---

**Table 1**  Nucleotide sequences of primers used for PCR amplifications
as follows: each of the forward and reverse primers (20 pmol) and cDNA library template were combined with 4 μL 25 mM dNTP mixture and 5 μL 10× PCR buffer containing 100 mM Tris-HCl (pH 8.3), 500 mM KCl, and 15 mM MgCl₂. The final volume was adjusted to 49 μL with sterilized water, and 1 μL containing 1 U Pfu DNA polymerase was then added. The PCR was carried out with denaturation at 94°C for 1 min, annealing at 58°C for 1.5 min, and polymerization at 72°C for 2 min. This cycle was repeated 30 times with a final extension step at 72°C for 10 min. The DNA fragments successfully amplified by PCR were subcloned into a plasmid vector, pBluescript II SK–. The 5’ and 3’ primers for PCR were synthesized referring to DNA nucleotide sequences reported for myosins of various animals. The 5’ and 3’ primers used are listed in Table 1.

**5’-End amplification of cDNA**

A 5’-Full RACE Core set (TaKaRa, Otsu, Japan) was used for determination of the 5’ end of myosin heavy chain. We devised a simple method for identification of the 5’ end of mRNA in which the first-strand cDNA was circularized and/or joined into a concatenate form by T4 RNA ligase. The resulting single-stranded DNA was subsequently used as a template for amplification of the 5’ end by PCR with gene-specific primers (Fig. 1; Table 1).

**Plasmid extraction and DNA sequencing**

Plasmid extraction was performed as described by Sambrook et al. Purified plasmid DNA was sequenced by Dye Deoxy™ (Applied Biosystems, Foster City, USA) termination cycle sequence kits with an Applied Biosystems DNA sequencer model 373S.

**Sequence analysis**

Sequence entry and analysis were performed using a Perkin Elmer Applied Biosystems model SeqEd version 1.0.3 sequence analysis system with suites of computer
Hydropilicity analysis was based on the method of Hopp and Woods.\textsuperscript{11}

RESULTS AND DISCUSSION

cDNA cloning of white croaker myosin heavy chain

At first, a cDNA library was constructed from poly(A)$^+$ mRNA of the dorsal muscle of white croaker in $\lambda$ZAP II vector, whereas cDNA clones, pMHC-53 and pMHC-8, were isolated by probing the cDNA library with the anticarp myosin polyclonal antibody. Next, five cDNA fragments were obtained by means of reverse transcription–PCR using first-strand cDNA as a template and appropriate sets of primers. These fragments were subcloned into a pBluescript vector and the clones thus obtained were called pMHC-SPACER, pMHC-RT1, pMHC-RT2, pMHC-RT3, and pMHC-RACE (Fig. 1). The DNA fragments obtained by restriction endonuclease digestion of pMHC-8 with EcoRI, HindIII, and PstI were subcloned into pBluescript vectors and nucleotide sequences were determined. The pMHC-8 contained an approximately 2 kb insert and encoded a part of the S2 region and the entire LMM fragment. A DNA fragment isolated from clone pMHC-53 using HincII digestion was subcloned into a pBluescript vector and nucleotide sequence was analyzed. pMHC-53 had a 1.4 kb insert and contained an almost entire S2 region and a C-terminal part of S1 heavy chain. The pMHC-SPACER, pMHC-RT1, pMHC-RT2, pMHC-RT3, and pMHC-RACE had inserts of 505, 525, 927, 790, and 701 bp, respectively. The following seven overlapping clones, pMHC-RT1, pMHC-RT2, pMHC-RT3, pMHC-SPACER, pMHC-RACE, pMHC-53, and pMHC-8, contained the entire coding sequence for myosin heavy chain. Through sequencing of the clones, their nucleotide and deduced amino acid sequences were successfully determined. The entire cDNA consisted of 5986 nucleotides (nt) with 64 nt 5'-untranslated and 129 nt 3'-untranslated regions (Fig. 2). The stop codon, TAA, was found at the 5791st nt from the 5' end, whereas at the 3' side the non-coding region and polyadenylation signal was included. The overall GC content was 49.5%.

Primary structure of white croaker myosin heavy chain

The full-length cDNA of 5986 nt encoding white croaker myosin heavy chain had an open reading frame encoding a polypeptide of 1930 amino acid residues. The S1 heavy chain, S2, and LMM were assigned from 1 to 2499 nt, from 2500 to 3846 nt, and from 3847 to 5790 nt, respectively (Fig. 2). The molecular weight of this translated peptide was 221 198, and isoelectric point 5.38. The ratios of acidic, basic, and hydrophilic amino acid residues compared to the total were 18.7, 17.8, and 36.2%, respectively. Hydrophilicity analysis revealed two different features of the myosin molecule (Fig. 3): S1 heavy chain had a mixed profile of hydrophobicity and hydrophilicity, whereas rod showed a profile of only hydrophilicity.

The homology of the deduced amino acid sequences between white croaker and other species is shown in Table 2. The overall sequence identity of white croaker myosin heavy chain to walleye pollack was 89.6%, which was the highest among those with vertebrates compared in this study. The homology between white croaker and other vertebrates was lowest with S1 heavy chain among the three myosin heavy chain fragments including S2 and LMM. It is well known that S1 heavy chain contains functionally important domains for ATP splitting and actin binding. It seems that S1 heavy chain must have changed its structure according to species-specific requirements to facilitate the above functions under various circumstances. In contrast, S2 sequence was highly homologous among vertebrate species compared in the present study.

When compared in respective myosin heavy chain fragments, white croaker S1 heavy chain showed 79–88% sequence identity to those of various vertebrates including walleye pollack,\textsuperscript{12} carp acclimated to 10°C and 30°C,\textsuperscript{13} chicken,\textsuperscript{14} and rabbit (GenBank/EMBL/DDBJ accession number U32574), which showed walleye pollack had the highest homology. Interestingly, white croaker S1 heavy chain contained some amino acid replacements at the conserved regions of other species (data not shown), which might have significant effects on protein structure or biochemical properties. While white croaker S2 exhibited 85–94% amino acid sequence identity with those of the above vertebrates including walleye pollock,\textsuperscript{15} carp acclimated to 10°C and 30°C,\textsuperscript{16} chicken,\textsuperscript{14} and rabbit (GenBank/EMBL/DDBJ accession number U32574), it showed higher homology with those from fish than homiotherms. Finally, the amino acid sequence identity of white croaker LMM with those from walleye pollack,\textsuperscript{15} carp acclimated to 10°C and 30°C,\textsuperscript{16} chicken,\textsuperscript{14} and rabbit\textsuperscript{18} were 79–92%. The identity was around 90% with those from other fish and higher than those from homiotherms of 80%. As was the case for S1 heavy chain and S2, these differences in the primary structure of LMM were supposed to be implicated in differences in its thermal stability and function. However, it remains unclear at the present stage without further comparison of higher structures to elucidate such implications.

Besides S1 heavy chain, prevalence of low homology between poikilotherms and homiotherms for both S2 and LMM suggests that the amino acid sequence is altered according to the body temperature of the animal. The S2 as well as S1 is suggested to contribute to force...
Fig. 2  Nucleotide and deduced amino acid sequences of a full-length cDNA clone encoding white croaker myosin heavy chain. A total number of 5986 nucleotides (nt) were sequenced. The 5‘-untranslated region of 64 nt and 3‘-untranslated region of 129 nt are underlined. The start and stop codons are denoted with an asterisk. The nucleotide sequence will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession number AB039672.
production in muscle contraction, whereas myosin rod consisting of S2 and LMM is of fundamental importance to assemble into a functional form, the thick filament.\textsuperscript{19,20} It seems that LMM requires more amino acid substitutions to adapt its assembly characteristics to different environmental temperatures than S2.\textsuperscript{28}

28-residue repeats and predicted coiled-coil of $\alpha$-helical structures of white croaker myosin rod

Figure 4a shows the amino acid sequence of white croaker S2 arranged in 28-residue repeat zones. The sequence was divided into 17 zones. The letters, d, e, f, g, a, b, and c, indicate positions in the seven-residue repeat of coiled-coil, where positions a and d are marked by asterisks. Hydrophobic amino acids are shown in bold letters. A skip residue was located at the 343rd amino acid from the N-terminus.

Figure 4b shows the histograms of amino acid distributions in 28-residue repeat zones of white croaker S2. Two skip residues were shown at the 211th and 436th positions. Figure 5a shows the amino acid sequence of white croaker LMM arranged in 28-residue repeat zones. Two skip residues were shown at the 211th and 436th positions. Figure 5b shows the histograms of amino acid distributions in 28-residue repeat zones of white croaker LMM. The distribution patterns of hydrophobic and charged amino acid residues were very similar to those of S2, which is in favor of the coiled-coil structure. Hydrophobic residues of white croaker myosin rod were much more concentrated as described above at positions a and d than b, c, e, f, and g. The frequency with which hydrophobic residues appeared at core positions a and d in white croaker rod were calculated.\textsuperscript{25,26} From the total hydrophobic residues in S2, 34.7% were at position a and 40.6% at positions d, while for LMM, 33.6% were at both positions a and d, indicating that S2 may form a more stable $\alpha$-helix.

There were considerable charged residues at core position a, about 18.8% in white croaker S2 and 9.9% in the LMM counterpart. According to Lu and Wong,\textsuperscript{19} these

| Table 2 | Amino acid sequence identity of three fragments of white croaker myosin heavy chain with those of other fish species and higher vertebrates |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | Walleye pollack | 10-Carp         | 30-Carp         | Chicken         | Rabbit          |
| Myosin heavy chain | 89.6           | 87.1            | 88.0            | 80.2            | 80.8            |
| Subfragment-1 heavy chain | 88.0           | 81.7            | 83.2            | 79.5            | 78.8            |
| Subfragment-2 | 91.6           | 92.8            | 93.9            | 84.8            | 85.3            |
| Light meromyosin | 91.0           | 88.8            | 91.5            | 79.1            | 80.1            |

Data for comparison were cited: walleye pollack from Ojima et al.\textsuperscript{12} for subfragment-1 (S1) and from Togashi et al.\textsuperscript{13} for subfragment-2 (S2) and light meromyosin (LMM); carp acclimated to 10°C (10-carp) and 30°C (30-carp) from Hirayama and Watabe\textsuperscript{14} for S1, Watabe et al.\textsuperscript{16} for S2, and from Imai et al.\textsuperscript{17} for LMM; chicken from Maita et al.\textsuperscript{14} for S1, S2, and LMM; rabbit from GenBank/EMBL/DDJB accession number U32574 for S1 and S2, and from Maeda et al.\textsuperscript{18} for LMM.
these thiols are distributed in different regions of rod, Pliszka and Lu\textsuperscript{28} examined the capacity of rod to form filament at various conditions and claimed that thiols may play a role in filament formation and movement of cross-bridges. However, in white croaker LMM, only two out of three cysteine residues were conserved. This replacement might have significant effect on protein structure and biochemical properties of white croaker LMM.

Considered together, it was found in the present study that white croaker myosin heavy chain showed frequent amino acid replacements in the primary structure when compared with those not only from homoiotherms, but also from other fish species including walleye pollack. As charged residues at position $a$ may lead to destabilization of the coiled-coil interaction. More charged residues at positions $a$ in S2 would loosen the binding between the two S2 chains more easily compared to those in LMM or tropomyosin.\textsuperscript{19,26}

There are six cysteine residues in myosin rod, three highly conserved in myosin S2 and three in LMM.\textsuperscript{27} It is known that there are three highly conserved cysteine residues at position $a$ in myosin rod. Considering that these thiols are distributed in different regions of rod, Pliszka and Lu\textsuperscript{28} examined the capacity of rod to form filament at various conditions and claimed that thiols may play a role in filament formation and movement of cross-bridges. However, in white croaker LMM, only two out of three cysteine residues were conserved. This replacement might have significant effect on protein structure and biochemical properties of white croaker LMM.

Considered together, it was found in the present study that white croaker myosin heavy chain showed frequent amino acid replacements in the primary structure when compared with those not only from homoiotherms, but also from other fish species including walleye pollack. As...
noted before, white croaker and walleye pollack muscles are both very important materials for surimi-based products. However, both fish are apparently different from each other in gel-forming ability. We still need further investigation to elucidate the meaning of amino acid replacements in terms of functional properties which may be related to different thermal stabilities and gel-forming abilities between white croaker myosin and other fish species or vertebrates.

ACKNOWLEDGMENT

This study was funded in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan.

REFERENCES

3. Lowey S, Slayter HS, Weeds AG, Baker H. Substructure of the myosin molecule. I. Subfragments of myosin by enzymic degrada-
5. Johnston IA, Frearson N, Goldspink G. The effects of environmental temperature on the properties of myofibrillar adenosine
15. Ojima T, Kawashima N, Inoue A, Amauchi A, Tagoshi M, Watabe S, Nishita K. cDNA cloning of myosin rod and the com-
19. Lu RC, Wong A. The amino acid sequence and stability predic-
21. McLachlan AD, Stewart M. The 14-fold periodicity in alpha-
24. Moitra J, Sulak L, Krylov D, Vinson C. Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine
27. Pliiska B, Lu RC. Reactivities of thiols in myosin rod: effect of magne