J. Biochem. 121, 671-676 (1997)

Structure of the Ascidian, *Halocynthia roretzi*, Troponin C Gene[1]

Hajime Julie Yuasa, Shigeru Sato, Hiroaki Yamamoto, and Takashi Takagi[2]

Biological Institute, Graduate School of Science, Tohoku University, Aoba-ku, Sendai, Miyagi 980-77

Received for publication, November 11, 1996

Two distinct cDNAs encoding troponin C (TnC) isoforms were isolated from the ascidian, *Halocynthia roretzi*. One is expressed in adult body wall smooth muscle and heart muscle, and the other in larval striated muscle. The *H. rorezti* gene is composed of 7 exons separated by 6 introns, and Southern blot analysis showed that TnC is a single copy gene product. The two isoforms of TnC were derived through the alternative splicing of the third exon.

Key words: alternative splicing, gene structure, *Halocynthia roretzi*, troponin C.

Troponin C (TnC) is one of the three components of troponin, a main regulatory protein of striated muscle contraction, and belongs to the EF-hand calcium-binding protein family. In mammalian and avian muscles, two isoforms of TnC have been identified, one is fast skeletal TnC (sTnC), which is expressed in fast skeletal muscle, and the other is slow/cardiac TnC (cTnC), which is present in heart muscle and slow skeletal muscle. The gene structures of these human and mouse TnC isoforms have already been determined (*1-4*). Both isoforms possess four EF-hand type calcium binding sites. All EF-hand calcium binding sites of sTnC are functional, *i.e.*, sTnC can bind four Ca^{2+} per molecule, however, in cTnC, the first EF-hand site (site I) is nonfunctional due to amino acid insertion and substitutions, so cTnC binds three Ca^{2+} per molecule.

We have isolated TnCs from two protochordate species, *Halocynthia roretzi* (ascidian), a urochordate, and *Branchiostoma lanceolatum* (amphioxus), a cephalochordate, and determined their amino acid sequences (*5, 6*). Ascidian TnC was obtained from body wall muscle, being a rare instance of a smooth muscle TnC. The only other case is the smooth muscle TnC obtained from scallop adductor muscle (*7*). Site I of ascidian TnC seemed to lack Ca^{2+} binding ability due to a lethal amino acid insertion as in the case of vertebrate cTnCs. The nonfunctional EF-hand sites of TnC were well observed in invertebrate TnCs (*8-12*), but all resulted from unsuitable amino acid substitutions, and not insertion. Amphioxus TnC was obtained from striated muscle and can bind three Ca^{2+} per molecule. Site IV is nonfunctional due to unsuitable amino acid substitutions, but site I is functional as in vertebrate sTnCs.

H. roretzi adult body wall muscle, adult heart muscle, and larval tail muscle consist of multinucleate smooth muscle cells (*13, 14*), unicellular striated muscle cells (*15*), and monocellular striated muscle cells (*16*), respectively. In this study, to elucidate the relationship of adult smooth and cardiac muscle, and larval striated muscle TnCs, we determined cDNAs of these tissues. We identified different cDNAs in adult and larval tissues. On genomic structural analysis, ascidian TnC was revealed to be a single copy gene product, and two isoforms were expressed from the same gene through alternative splicing.

MATERIALS AND METHODS

The solitary ascidian, *Halocynthia roretzi*, was obtained from the Marine Biological Station of Asamushi, Tohoku University, and at a seafood market in Sendai.

Total RNA of body wall muscle and heart muscle was prepared according to the method of Chomczynski and Sacchi (*17*), and mRNA was purified with an Oligotex dT-30 Super (Japan Roche). The single-stranded cDNA was synthesized with moloney murine leukemia virus reverse transcriptase (New England Biolabs) using an oligo-dT adaptor, 5′-GGGATCCGAATTCT$_{17}$-3′, as a primer. The cDNA of body wall muscle TnC was amplified by polymerase chain reaction (PCR) (*18*) using Ex *Taq* DNA polymerase (Takara). The redundant oligomer used for PCR was 5′-CARGARATGATHGARGARGTNGA-3′, where R represents A and G; H, A, C, and T; and N, A, C, G, and T. This was designed based on the amino acid sequence, QEMIEEVD (residues 51-58), of body wall muscle TnC (*5*). The oligo-dT adaptor was used as another primer.

The 5′ end of cDNA was determined as follows. The *Eco*RI-ended double-stranded cDNA was synthesized from mRNA using a TimeSaver cDNA Synthesis Kit (Pharmacia). The *Eco*RI Cassette (Takara) was ligated to each end of the cDNA. The 5′ upstream region was amplified by PCR using cassette-specific primer C1, 5′-GTACATATTGTCGTTAGAACGCG-3′ and R1 (Table I). The heart muscle TnC cDNA was amplified by PCR with primers F4 and R3 (Table I).

The cDNA library of larvae was constructed in λgt10 using mRNA prepared from mid-tailbud stage embryos. The larval TnC cDNA was also amplified by PCR using a cDNA library as a template. The primers used were the λgt10-specific reverse primer and F4 (Table I).

The genomic DNA was prepared from a single specimen of *H. roretzi* by the conventional phenol-chloroform method. Several sets of primers were used to amplify the genomic DNA fragment by PCR. The primers were designed so as to correspond to the cDNA sequence of ascidian

[1] The nucleotide sequences have been submitted to the DDBJ under the accession numbers, D88494 (*H. roretzi* body wall muscle TnC cDNA), D88495 (*H. roretzi* larval TnC cDNA), and D88496 (*H. roretzi* TnC genome).

[2] To whom correspondence should be addressed. Tel: +81-22-217-6677, Fax: +81-22-263-9206, E-mail: a21692@cctu.cc.tohoku.ac.jp

body wall muscle TnC, and the primers used are listed in Table I. The strategy for amplification of the ascidian TnC gene is shown in Fig. 4. The 5′- and 3′-flanking regions were amplified as follows. The *Eco*RI cassette was ligated to *Eco*RI-digested genomic DNA. Then the 5′-upstream and 3′-downstream regions were amplified with cassette-specific primer C1 and R5, and C1 and F3, respectively. Some of them were reamplified by using inner primers. All the amplified products were subcloned into the pCR II plasmid vector (TA-cloning kit; Invitrogen) or pUC18 for sequencing. The sequences of the products were determined by the dideoxy chain termination method with a Dye Primer Cycle Sequencing Kit (Applied Biosystems) and an automated DNA sequencer (Applied Biosystems 373A).

To determine the specificity of expression of TnCs in adults and larvae, RT-PCR was performed using their specific primers. The adult and larval TnC specific forward primers were F2 (Table I) and 5′-AGGACACCGAGGATG-

TABLE I. **Primers used for amplification of the *H. roretzi* TnC genomic DNA fragments.**

Primers (positions of cDNA sequence)	Sequence
F4 (−149 to −130)	TCAGTAACATTGGTAGAAGG
F2 (85 to 104)	ATCAGTTCCAAAGAGTTGGG
F3 (453 to 472)	GACAATGATGAAGTTTGTGC
R5 (7 to 26)	TGATCTTCTGTCAAATGTTC
R4 (123 to 142)	TCTCAGTTGGATTTTGTCCC
R3 (480 to 499)	GAAACACTGTTGCGGGTTTG
R1 (725 to 742)	TCTTAGAATTGACATCTG

```
adult:                -156   GTGATCGTCAGCAACATTGGTAGAAGGTCAGTTAAAGAAAGAAGGTTGATCAGCTCATTGCAGAAT  -91
larva:                -149   ..........................................................  -91

adult:TGAATAAACGACCCTGAAGTATATATTGACGGGAGACTTTCGGATTTACTATTCTTCAATTTCCTAGTTATTTCCAACAAAACAAGTAAA   -1
larva:.....................................A.....................................  -1

      M V E H L T E D Q K S E F R 1 F D I F V A D A K D G T I S S   30
adult:ATGGTGGAACATTTGACAGAAGATCAAAAATCTGAATTTAGAGCTGCGTTTGATATATTCGTTGCTGATGCAAAAGACGGTACCATCAGT   90
larva:.....................................C...A.ATGC..C..C.........AG..CA.CG.G..T..A..G....CG   90
      . . . . . . . . . . . . . . . . T C . . . . . E . T E . . . . T   30

      S K E L G K V M K M L G Q N P T E K D L Q E M I E E V D I D   60
adult:TCCAAAGAGTTGGGTAAAGTAATGAAAATGTTGGGACAAAATCCAACTGAGAAGGATTTACAGGAAATGATCGAAGAAGTCGACATAGAC  180
larva:G.G.....AC.T..A...T.G.............G.........T.C...C....GC.TA.A......G.............T.G..T  180
      A . . . . . L . . . . . . . . . S . Q E . K . . V . . . . L .   60

      G S G T I D F E E F C L M M Y R Q M Q A E E E A K I P E R E   90
adult:GGAAGTGGAACGATCGATTTCGAAGAATTTTGTCTTATGATGTATCGTCAAATGCAAGCGGAAGAAGAAGCGAAAATCCCAGAGAGGGAA  270
larva:.....................................................................................  270
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .   90

      E K E L S E A F R L F D L D G N G L I G W D E L K A A L D G   120
adult:GAAAAAGAACTTTCAGAAGCGTTCCGATTATTCGATTTGGACGGTAACGGTTTAATCGGATGGGATGAATTGAAGGCTGCTCTAGACGGA  360
larva:.....................................................................................  360
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  120

      T G E N V E T W E V D E M M A D G D K N H D S Q I D Y E E W  150
adult:ACGGGTGAGAACGTCGAAACTTGGGAAGTTGACGAAATGATGGCCGATGGGGACAAAAATCACGATTCGCAAATTGATTACGAGGAATGG  450
larva:.....................................................................A.....  450
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  150

      V T M M K F V Q *                                              158
adult:GTGACAATGATGAAGTTTGTGCAGTAAAACAAACCCGCAACAGTGTTTCAATTGAGGATATCTTACACACAAGTTAAAGTATAGGATATA  540
larva:.....................................................................................  540
      . . . . . . . . *                                              158

adult:AACCAAACTTGAAACGAATATTTATCTACAATTACATATTATCGTTATATTTACGAACTGGGTTCAAATATTTTCAACAATGTTTATGAT  630
larva:.....................................................................................  630

adult:CTATGGATGTCCGAAAATATCACATTAAAAACTGTATCATGACCACATGCCTGTTGCATATAATTCTGACTACAAATATTTTTTCTTTCA  720
larva:.....................................................................................  720

adult:AATACAGATGTCAATTCTAAGATAn                                      744
larva:..............                                                 734
```

Fig. 1. **Comparison of the cDNA and derived amino acid sequences of *H. roretzi* TnC isoforms.** Upper lanes: adult body wall muscle TnC cDNA and the deduced amino acid sequence; lower lanes: larval TnC cDNA and the deduced amino acid sequence. Identical nucleotides and amino acids to those in adult muscle are indicated by dots (.). Stop codon is indicated by an asterisk (*). No typical polyadenylation signal was observed. The N-terminal methionine (underlined) must have been removed after translation.

GAACG-3′, based on the sequence of 65 to 84nt of larval TnC cDNA, respectively. R3 (Table I) was used as a common reverse primer.

For Southern blot analysis, the 648 bp body wall muscle TnC cDNA was labeled with the DIG-DNA Labeling Mixture (Boehringer Mannheim) by PCR using primers F4 and R3 (Table I). Ten micrograms of genomic DNA digested with restriction enzymes was separated on a 0.8% agarose gel, and then transferred to a nylon membrane. Hybridization and washing were carried out according to the manufacturer's instructions (Boehringer Mannheim), and the TnC gene was detected with a DIG Luminescent Detection Kit (Boehringer Mannheim).

RESULTS AND DISCUSSION

The cDNA of the smooth muscle TnC was amplified by the PCR method and the complete cDNA sequence of 900 nucleotides was constructed from two overlapping fragments (Fig. 1). The open reading frame was composed of 477 nucleotides and encoded a protein of 158 amino acid residues. Several amino acid substitutions and insertions were observed, on comparison with previously reported *H. roretzi* body wall muscle TnC (5); Gln was Glu in place at residue 8; Glu in place of Gln at residues 11 and 80; Asn in place of Asp at residue 105; and Leu-107 and Trp-110 were inserted. The cDNA sequence was also confirmed by the genomic sequence (see below), thus the sequence difference was presumably due to peptide sequencing errors. According to the cDNA-derived amino acid sequence, the first EF-hand site is unlikely to bind Ca^{2+} due to amino acid substitutions and insertion, as in the case of vertebrate cTnCs. The other three EF-hand sites might have the ability to bind Ca^{2+}, consistent with the results of the Ca^{2+} binding experiment (2.5 mol of Ca^{2+}/molecule) (*19*).

We amplified the cDNA of heart muscle TnC by the PCR method and confirmed the nucleotide sequence of the open reading frame was exactly identical with that of body wall muscle TnC (data not shown).

The cDNA of the larval TnC was composed of 883 nucleotides and the open reading frame comprised 477 nucleotides, encoding a protein of 158 amino acid residues,

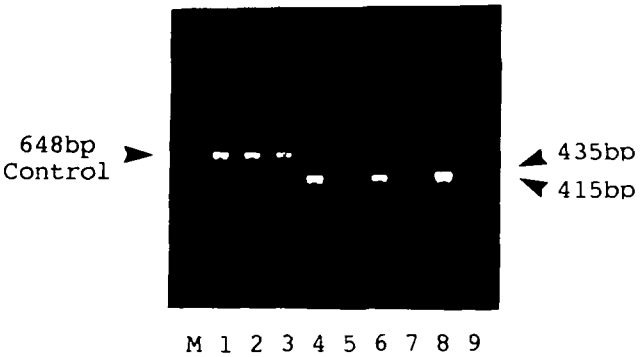

Fig. 2. **Tissue-specific expression of the two TnC isoforms.** Lane M, molecular markers. Lanes 1-3, primers F4 and R3 (larval and adult TnC common primers) were used for PCR. The amplification of a 648 bp-product was expected (control). Lanes 4-6, primers F2 (adult TnC-specific primer) and R3 were used, a 415 bp-product being expected. Lanes 7-9, a larval TnC-specific primer and R3 were used, a 435 bp-product being expected. Lanes 1, 4, and 7, cDNA prepared from adult body wall muscle was used as a template. Lanes 2, 5, and 8, a tailbud embryo cDNA library in λgt10 was used. Lanes 3, 6, and 9, cDNA prepared from adult heart muscle was used.

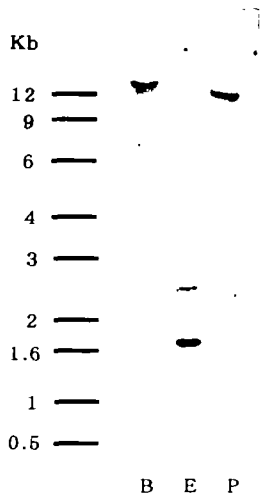

Fig. 3. **Southern blot analysis of the *H. roretzi* TnC gene.** Genomic DNA prepared from body wall muscle of a single specimen of *H. roretzi* was digested with restriction enzyme *Bam*HI (lane B), *Eco*RI (lane E), or *Pst*I (lane P), and then hybridized to the DIG-labeled body wall muscle TnC cDNA. Left, size markers in kb.

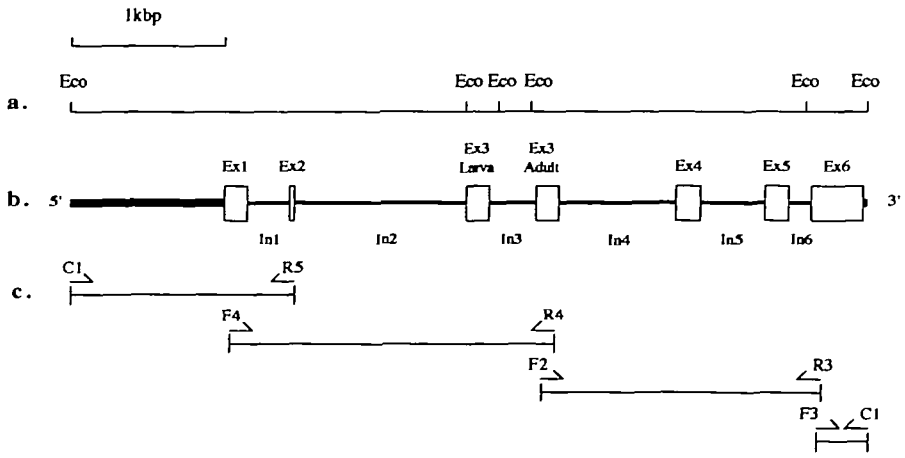

Fig. 4. **Structure of the *H. roretzi* TnC gene and strategies used for PCR amplification of DNA fragments.** a: *Eco*RI cleavage sites. b: Exon/intron map of the *H. roretzi* TnC gene. The 5′- and 3′-flanking regions are indicated by thick bars. Exons (Ex1-Ex6) and introns (In1-In6) are shown as boxes and thin bars, respectively. c: Strategies used for PCR amplification. The primers used for PCR are shown by arrows and their sequences are given in Table I.

the same as the adult TnC, as shown in Fig. 1. There were 43 nucleotide substitutions reflecting 14 amino acid differences between the adult and larval TnCs. All substitutions except for those of two nucleotides were restricted to the segment comprising nucleotides 38 to 180 which was in exon 3 (Figs. 4 and 5). Two nucleotide substitutions, A in

```
gaattcaatcgtgttctgatgtttgatagagaacgccagcgaacaatcaaccgatgccattaaggtgtactggcctgaaaagaaccccagccgctgagctg  100
ccgccctcatacggatcggttatgagttcatccagtcaatcaacggcatcatcagttccagatggcaactctttctcaaaataagtttgaagtatgggat  200
ttatgagttgcattggattattagacgttgtgacctatcgaactgacaatcacagatggtgtattgccggtactttgagccgaacatacataagacatat  300
ccaacaacggcggaaccaacctgtggttaatgaagttttcattgttgttcaacattatgtaaactaataaaaggaataagagtgtaatcgaagaaattca  400
ttoggactttaagttagggttatgatgcctaaaatatcttttatttatttttaatttttcaattccaaccccgaaccccccaagaatgaggtagactgcgcgct  500
aaaatagcgctacatatttgactacatcagtagtataggttatatataacttgctaatatttccatcatcatgggatgtaactgtcacaacttgttttaa  600
aacgtaaatatctgaaactcaataaattctatactacagcaccaaagtatcacgttacctaaaagacatagaacggacttaaactagtgagagaaaatct  700
agatatcgtctcgacaagtgctataataatggtggcaagtttagtgccttcttttgataagagttgtttttttggagatcgtgtgagggtattatcgcccg  800
tattgtgcaaatattccatgtatcaaatcgttaggagaagatgaacgttctaagcgagaacgcgaataagacgtctgcctgcttgctcgccttgctcaag  900
tttgtcagctgctcaacgcgtcaactaattttgattaaattgtccacaaaacaagctactccaactcatactacatt<u>cattcct</u>gagttcagcagctctg 1000
GTGATCGTCAGCAACATTGGTAGAAGGTCAGTTAAAGAAAGAAGGTTGATCAGCTCATTGCAGAATTGAATAAACGACCCTGAAGTATATATTGACGGGA 1100
GACTTTCGGATTTACTATTCTTCAATTTCCTAGTTATTTCCAACAAAACAAGTAAAATGgtaagtacgttttatccttatacacttaggaaagacactaa 1200
agtactttaataaatgtgccgtatttaagtaagttgcatatctaaatttacatagaccgcttgttctttgaagtagatatgaatacattgcatccgcagc 1300
aataagtttaatgatacaacgttattagccagaaatataaaatagtattcacttgtgtgtgattgaaatttaaaacgagattgggaaagtttaggaaatt 1400
atgtcgcatacaatcgatcacatttatttattctttactagGTGGAACATTTGACAGAAGATCAAAAATCTGgtaagtgatttctaatgttttccaaat 1500
ttagataaaatatttgatatttaagttattattttaagaaactcctattctaagtaatagagtattggatgacgcctacattccgaggttagtttgccta 1600
attactgtggcatttaacggactaacgcgaacgaaacattttaaccaatatttttcaacaacttttgacttgtaaggggacttaggtaagattttcgtt 1700
catactcaacagggttatagtgggaaaactgtctagataaggaaccactttctggagttcctttccagcgaccttgtggttctactcttatccatgtat 1800
agaagtacacagctgaatatcctacttgcgctataataagaagtgacgattacatatcacattccctcaggagcgtcgccggctctgttacacagacggt 1900
acgatagcgtattatgggatcagccacaccagttacgcaggaatacttatcttcatcagtacattaaccagttgcctaaatacaacttcagtaaatgtca 2000
ttactcatcactgacaaccctgatgctctttcaacacaagaccaaggagtcgaatcaaattgcgatgcaataacactatacctttcaaatagggactgtag 2100
atcagtacaaacatatttcacgcactagcgacgatattaaaaccagatgtagaccagattaaaactttacacagtaatttaaattatttttattcaccac 2200
atatttgttggtgaacatatctcatttcctttcactatgttatggtgacaattcacatttaacagattggatttttaccatttgcttatatccgtttctg 2300
aacacctcacttccaaatgcttgaagattatttcgtttgggttctgtcatttgtacggatgacggttttattgcgctcaatagctaagtaaagtcagttt 2400
tacgaatggaaatacagtttattcatactgctgtgtgaattatttgaatttttctactcgtaggagccgtgtgtttatgtacagaaacatagaaacaac 2500
acaattaaaaaattaagagttagcagttttttcactacctatcttgctcgcgtattgtatccactgtatgcgcatacattaaaacccgactctttaaaca 2600
atcttctgcctcgaaaatcaaacaccagAATTCAGAACATGCTTCGACATATTCGTTGAGGACACCGAGGATGGAACGATCACGGCGAAAGAACTTGGAA 2700
AATTGATGAAGATGTTGGGGCAAAATCCATCCGAGCAGGAGCTTAAAGAAATGGTCGAAGAAGTCGACTTGGATGgttcgtcgaagaaaattctactctg 2800
tatttaggatgttgttttctgctgaagaatgcgattgctgtggaattctgtagattatttatcccgtgctgtgtggttggcgatttataatgcgaaacgg 2900
aattattgcgttccttttcctaaaacttcttcaacattctatttagctgctgacacaaaatgcgctcgaagaaaacaattcgtccggaatacatacaaca 3000
catataaagcatatctctagggatttttttaaatataatttaacttattcttttgagaattcggccatttattttcatgcagAATTTAGAGCTGCGTTTG 3100
ATATATTCGTTGCTGATGCAAAAGACGGTACCATCAGTTCCAAAGAGTTGGGTAAAGTAATGAAAATGTTGGGACAAAATCCAACTGAGAAGGATTTACA 3200
GGAAATGATCGAAGAAGTCGACATAGACGgtttgtgcaaatcacgcaaatcataacacatcacattatctcatatcacaacacaacatcaatagtcattc 3300
tcactttagaagttgaggtcatttcatcaacctcattacaatggaactatgttggttcactattgagccgaatctctccctttttcaatcaaatccgaaat 3400
tatatacgataatctataactcaatgaaattaaaacgaatttgcgatagaaagcaagttgtaaagaaagagtttacaaaggaaaactaattatcgcggaa 3500
gatattccattaattccagtgttgcacgtttactagttttctgtattgcatgactacttgtatgggtgtccattgagggttacaaactgaggagttgctc 3600
cactggatcataatccatagatatacttcgttatactctgctgatagtccattcatactcggtttcattactgctaagttataatttaatgttttgcacc 3700
tccaaaaagcaatttaaaatatatgataacgaaagtaatgatatgcttgatatctcatgggaaaattgtagcggactctcgtgcctacctatgcatgctt 3800
gagattgtttcactgtgaataatttgtaatgcatggtctatataacatcttgttttagaaattaacatttttttaacccatttcacttagcgcatgacaa 3900
tcaatataagccaatgtcatatcaagttacccatataggtctaagtattggagtctaattaacccatacc<u>ttgaacag</u>GAAGTGGAACGATCGATTTCGA 4000
AGAATTTTGTCTTATGATGTATCGTCAAATGCAAGCGGAAGAAGAAGCGAAAATCCCAGAGAGGGAAGAAAAAGAACTTTCAGAAGCGTTCCGATTATTC 4100
GATTTGGACGGTAACGGTTTAATCGGATGGGATGAATTGAAGgtaataaaaagacgctcaaatgttgcaaatattttatctaacgtacagtatcgagatg 4200
cattaaggatctatatgtatatcgcgtcataagtacttaacaatttggaaacgctccaaaaggttaattatatagcataaagttcaccaataacatcttc 4300
tacagagcaattacgcttaattgctaagactatgttggcgtatgcattagcagttttgtgttgtttctaggcatttcaaataccagcgaaacctcgctaa 4400
atgatataaatatcgcgagaaaactaaatctgttaatgttaagttggaaagaaattggtacatattatggttatcaataatagaatacaatttattcaac 4500
ttactattcacaacaacgttacgacacgaggacttacaaatactattcctaaatacagGCTGCTCTAGACGGAACGGGTGAGAACGTCGAAACTTGGGAA 4600
GTTGACGAAATGATGGCCGATGGGGACAAAAATCACGATTCGCAAATTGATTACGAGGgtaattacttacttcgtcacctgtttttctcagtttcactga 4700
tataaaaaactaaaacattaatctgaatagctaaagctaatacattaatctgaatagctaaagctaatacattaatctgaatagctaaagcgaattctaa 4800
taataaatcttggttcctttacagAATGGGTGACAATGATGAAGTTTGTGCAGTAAAACAAACCCGCAACAGTGTTTCAATTGAGGATATCTTACACACA 4900
AGTTAAAGTATAGGATATAAACCAAACTTGAAACGAATATTTATCTACAATTACATATTATCGTTATATTTACGAACTGGGTTCAAATATTTTCAACAAT 5000
GTTTATGATCTATGGATGTCCGAAAATATCACATTAAAAACTGTATCATGACCACATGTCTGTTGCATATAATTCTGACTACAAATATTTTTTCTTTCAA 5100
ATACAGATGTCAATTCTAAGATatttcggtctgcactgaattc                                                          5143
```

Fig. 5. **Nucleotide sequence of the *H. roretzi* TnC gene.** The exons are indicated by capital letters. The sequences of the 5′- and 3′-flanking regions and introns are shown by small letters. Four E-boxes (CANNTG) are boxed and the M-CAT consensus sequence (CATTCCT) is underlined.

place of C at 50 and A in place of G at 441, had no influence on the amino acid sequence, probably due to individual differences.

As shown in Fig. 2, when the larval TnC specific primer was used for PCR amplification, only larval TnC was amplified, *i.e.* adult TnC was not amplified. In contrast, only adult TnC was amplified when the adult TnC-specific primer was used. These results suggested that two TnC isoforms were expressed specifically in larval and adult muscles, respectively.

Restriction enzyme-digested genomic DNAs were hybridized to the 613-bp body wall muscle TnC cDNA probe (Fig. 3). The ascidian TnC gene has no *Bam*HI and *Pst*I cleavage sites, but has 6 *Eco*RI sites, as shown in Fig. 4. Although only two large fragments (2.6 and 1.7 kbp) were hybridized on *Eco*RI digestion, other fragments were in the intron and outside of the probe (see Figs. 4 and 5). These results were confirmed using five different individuals (data not shown), suggesting that *H. roretzi* TnCs are encoded by a single gene, and two isoforms are expressed from the same gene.

The complete nucleotide sequence of the ascidian TnC gene is shown in Fig. 5, which was constructed from 4 overlapping fragments separately amplified by the PCR method, as shown in Fig. 4. The genomic structure showed that it was composed of 5,143 bp and divided into 7 exons by 6 introns. The third and fourth exons were adult and larva specific ones, respectively. The other exons were common to both TnCs. The nucleotide sequences of the exons were exactly identical with the adult and larval TnC cDNAs except for one nucleotide, *i.e.* the codon encoding larval TnC Lys-39, AAA changing to AAG in the genome. This transition was presumably caused by an individual difference. These results clearly indicate that the *H. roretzi* TnC gene is alternatively selected in adult or larval muscle. Thus, based on the amino acid sequence, both adult and larval TnCs are separated into 6 exons by 5 introns the same as in vertebrate s and cTnCs. All introns start with gt and end with ag. The positions of introns 1, 2, 3, and 5 are identical to those in vertebrate cTnC. However, intron 4 was placed at 3.24/0, which means site III, the 24th amino acid and before the triplet codon, according to the nomenclature of Kretsinger and Nakayama (*20*), the corresponding vertebrate intron being 3.11/2.

As judged on comparing the vertebrate s and cTnC genes, the position of the first intron is at −10/0 (just after the initiator ATG) in sTnC and at −17/0 (21 bp downstream of the initiation codon) in cTnC, and the other 4 introns are at identical positions (*1-4*). The first intron of the ascidian TnC gene is located just after the initiator ATG, like that of vertebrate sTnC. However, the ascidian TnC lacks 7 amino acid residues at the N-terminus compared to vertebrate cTnC, thus the position of the first intron could also be considered to be 21 bp downstream from the initiator ATG, like in the vertebrate cTnC gene. The ascidian TnC is a particular case which is expressed in smooth, cardiac, and striated muscles.

For myofibrillar proteins, including troponin T (*21-24*), myosin heavy chain (*25*), myosin light chain (*26-29*), and tropomyosin (*30*), alternatively spliced isoforms have been observed, however, as for troponin C, *H. roretzi* is the first case of alternative splicing. The spliced region encoded by exon 3 contains the entire site I and a part of site II. Site I of both isoforms is unlikely to bind Ca^{2+} due to amino acid insertions and substitutions, whereas the amino acid substitutions observed in site II have no effect on Ca^{2+} binding. Thus, there seems to be no difference in Ca^{2+} binding function between the two isoforms. Another function of TnC is interaction with troponin I. One of the binding sites for troponin I is the N-terminal helix region of site II (*31*). In this region, two charge difference substitutions were observed at residues 48 (K to Q) and 51 (Q to K), however, the total charge is not different. Thus a functional difference between the two isoforms was not suggested.

In the 5′ upstream region, a typical TATA-box or CCAAT-box sequence was not found, however, muscle specific regulatory elements were present, *i.e.* four E-box sequences (CANNTG) (*32*) and an M-CAT sequence (CATTCCT) (*33*) at −179, as shown in Fig. 5. Other regulatory elements, such as a CArG-box (*34*) and MEF-2 (*35*), were not observed. The absence of these sequences is commonly observed in vertebrate TnC genes.

In this study, we revealed that *H. roretzi* TnC is a single copy gene product, and two isoforms (adult and larval types) are produced through alternative splicing. However, there are little data on the genomic structure of TnCs, particularly nonvertebrate ones, to determine the evolutional relationship.

We wish to thank Dr. T. Numakunai of the Marine Biological Station of Asamushi, Tohoku University, for collecting the *H. roretzi*.

REFERENCES

1. Parmacek, M.S. and Leiden, J.M. (1989) Structure and expression of the murine slow/cardiac troponin C gene. *J. Biol. Chem.* **264**, 13217-13225
2. Gahlmann, R. and Kedes, L. (1990) Cloning, structural analysis, and expression of the human fast twitch skeletal muscle troponin C gene. *J. Biol. Chem.* **265**, 12520-12528
3. Schreier, T., Kedes, L., and Gahlmann, R. (1990) Cloning, structural analysis, and expression of the human slow twitch skeletal muscle/cardiac troponin C gene. *J. Biol. Chem.* **265**, 21247-21253
4. Parmacek, M.S., Bengur, A.R., Vora, A.J., and Leiden, J.M. (1990) The structure and regulation of expression of the murine fast skeletal troponin C gene. *J. Biol. Chem.* **265**, 15970-15976
5. Takagi, T. and Konishi, K. (1983) Amino acid sequence of troponin C obtained from ascidian (*Halocynthia roretzi*) body wall muscle. *J. Biochem.* **94**, 1753-1760
6. Takagi, T., Petrova, T., Comte, M., Kuster, T., Heizmann, C.W., and Cox, J.A. (1994) Characterization and primary structure of Amphioxus troponin C. *Eur. J. Biochem.* **221**, 537-546
7. Ojima, T. and Nishita, K. (1986) Isolation of troponins from striated and smooth adductor muscle of *Akazara scallop*. *J. Biochem.* **100**, 821-824
8. Nishita, K., Tanaka, H., and Ojima, T. (1994) Amino acid sequence of troponin C from scallop striated adductor muscle. *J. Biol. Chem.* **269**, 3464-3468
9. Kobayashi, T., Takagi, T., Konishi, K., and Wnuk, W. (1989) Amino acid sequences of the two major isoforms of troponin C from crayfish. *J. Biol. Chem.* **264**, 18247-18259
10. Collins, J.H., Theibert, J.L., Francois, J.-M., Ashley, C.C., and Potter, J.D. (1991) Amino acid sequences and Ca^{2+}-binding properties of two isoforms of barnacle troponin C. *Biochemistry* **30**, 702-707
11. Garone, L., Theibert, J.L., Miegel, A., Maeda, Y., Murphy, C., and Collins, J.H. (1991) Lobster troponin C: amino acid sequences of three isoforms. *Arch. Biochem. Biophys.* **291**, 89-91
12. Kobayashi, T., Kagami, O., Takagi, T., and Konishi, K. (1989) Amino acid sequence of horseshoe crab, *Tachypleus tridentatus*,

striated muscle troponin C. *J. Biochem.* **105**, 823-828
13. Shinohara, Y. and Konishi, K. (1982) Ultrastructure of the body-wall muscle of the ascidian *Halocynthia roretzi*: smooth muscle cell with multiple nuclei. *J. Exp. Zool.* **221**, 137-142
14. Terakado, K. and Obinata, T. (1987) Structure of multinucleated smooth muscle cells of the ascidian *Halocynthia rorezti*. *Cell Tissue Res.* **247**, 85-94
15. Kalk, M. (1970) The organization of a tunicate heart. *Tissue Cell* **2**, 99-118
16. Satoh, N. (1994) *Developmental Biology of Ascidians*, Cambridge Univ. Press, New York
17. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159
18. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491
19. Endo, T. and Obinata, T. (1981) Troponin and its components from ascidian smooth muscle. *J. Biochem.* **89**, 1599-1608
20. Kretsinger, R.H. and Nakayama, S. (1993) Evolution of EF-hand calcium-modulated proteins. IV. Exon shuffling did not determine the domain compositions of EF-hand proteins. *J. Mol. Evol.* **36**, 477-488
21. Breitbart, R.E. and Nadal-Ginard, B. (1986) Complete nucleotide sequence of the fast skeletal troponin T gene. *J. Mol. Biol.* **188**, 313-324
22. Gahlmann, R., Troutt, A.B., Wada, R.P., Gunning, P., and Kedes, L. (1987) Alternative splicing generates variants in important functional domains of human slow skeletal troponin T. *J. Biol. Chem.* **262**, 16122-16126
23. Cooper, T.A. and Ordahl, C.P. (1989) Nucleotide substitutions with the cardiac troponin T alternative exon disrupt pre-mRNA alternative splicing. *Nucleic Acids Res.* **17**, 7905-7921
24. Jin, J.-P., Huang, Q.-Q., Yeh, H.-I., and Lin, J.J.-C. (1992) Complete nucleotide sequence and structural organization of rat cardiac troponin T gene. *J. Mol. Biol.* **227**, 1269-1276
25. Babij, P. and Periasamy, M. (1989) Myosin heavy chain isoform diversity in smooth muscle is produced by differential RNA processing. *J. Mol. Biol.* **210**, 673-679
26. Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M., and Ogata, K. (1984) Alternative transcription and two modes of splicing result in two myosin light chains from one gene. *Nature* **308**, 333-338
27. Periasamy, M., Strehler, E.E., Garfinkel, L.I., Gubits, R.M., Ruiz-Opazo, N., and Nadal-Inard, B. (1984) Fast skeletal muscle myosin light chains 1 and 3 are produced from a single gene by a combined process of differential RNA transcription and splicing. *J. Biol. Chem.* **259**, 13595-13604
28. Robert, B., Daubas, P., Akimenko, M.-A., Cohen, A., Garner, I., Guenet, J.-L., and Buckingham, M. (1984) A single locus in the mouse encodes both myosin light chains 1 and 3, a second locus corresponds to a related pseudogene. *Cell* **39**, 129-140
29. Lenz, S., Lohse, P., Seidel, U., and Arnold, H.-H. (1989) The alkali light chains of human smooth and nonmuscle myosins are encoded by a single gene. *J. Biol. Chem.* **264**, 9009-9015
30. Tsukahara, T., Casciato, C., and Helfman, D.M. (1994) Alternative splicing of β-tropomyosin pre-mRNA: multiple *cis*-elements can contribute to the use of the 5′- and 3′-splice sites of the nonmuscle/smooth muscle exon 6. *Nucleic Acids Res.* **22**, 2318-2325
31. Ohtsuki, I., Maruyama, K., and Ebashi, S. (1986) Regulatory and cytoskeletal proteins of vertebrate skeletal muscle. *Adv. Prot. Chem.* **38**, 1-67
32. Lassar, A.B., Buskin, J.N., Lockshon, D., Davis, R.L., Apone, S., Hauschka, S.D., and Weintraub, H. (1989) MyoD is a sequence-specific DNA binding protein requiring a region of *myc* homology to bind to the muscle creatine kinase enhancer. *Cell* **58**, 823-831
33. Mar, J.H. and Ordahl, C.P. (1988) A conserved CATTCCT motif is required for skeletal muscle-specific activity of the cardiac troponin T gene promoter. *Proc. Natl. Acad. Sci. USA* **85**, 6404-6408
34. Minty, A. and Kedes, L. (1986) Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. *Mol. Cell Biol.* **6**, 2125-2136
35. Gossett, L.A., Kelvin, D.J., Sternberg, E.A., and Olsen, E.N. (1989) A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. *Mol. Cell Biol.* **9**, 5022-5033