BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **223,** 160–164 (1996)
ARTICLE NO. 0862

A 11.5-kb 5′-Terminal cDNA Sequence of Chicken Breast Muscle Connectin/Titin Reveals Its Z Line Binding Region

Hirohiko Yajima,* Hiroshi Ohtsuka,* Yuuki Kawamura,† Hideaki Kume,† Takashi Murayama,‡ Hiroshi Abe,* Sumiko Kimura,* and Koscak Maruyama*,[1]

**Department of Biology, Faculty of Science, Chiba University, Chiba 263, Japan; †National Institute for Physiological Sciences, Okazaki 444, Japan; and ‡Department of Pharmacology, Faculty of Medicine, Juntendo University, Hongo, Tokyo 113, Japan*

Received April 21, 1996

A partial 5′-region cDNA (11.5 kb) of chicken breast muscle connectin/titin encoding 3752 amino acids was sequenced. The predicted amino acid sequence contains 31 immunoglobulin C2 motifs and 10 interdomains. The sequence suggests a skeletal muscle type of connectin isoforms. Immunoelectron microscopic studies using antisera raised against several products of the cDNA fragments expressed in *E. coli* revealed that a region of some 800 amino acids from the N terminus of connectin is involved in its binding to the Z line in a sarcomere.

Connectin, also called titin, is the largest polypeptide known to date: molecular mass, ~3 MDa (1,2). Connectin links the myosin filament to the Z line in a sarcomere of vertebrate striated muscle (3) and it positions the myosin filament at the center of a sarcomere acting as "a safety guard spring" (4). The connectin molecule is as long as 1 μm (5,6) and its length is variable depending on the external force (7). This molecule is thus an elastic filament that is responsible for passive tension generation of striated muscle.

This paper describes the sequence (3752 amino acids) of chicken breast muscle connectin from the N_2 line region in the I band to the N terminal end located at the Z line *in situ.* It is also shown that some 800 amino acids of the N terminal region are involved in the binding to the Z line. While this work was in progress, Labeit and Kolmerer (8) completed the entire sequence determination of human cardiac connectin cDNA. The present results are generally in good agreement with their work and show that this sequence is a skeletal muscle type of connectin.

MATERIALS AND METHODS

Isolation and sequencing of cDNA clones. Poly(A)$^+$RNA was prepared from 7-week old chicken breast muscle. cDNA synthesis was primed with random hexamers. The double stranded cDNA fragments were ligated to λZAP II expression vector arms (Stratagene). For immunoscreening, polyclonal antibody Pc1200 (9) and monoclonal antibody T12 (3) (BioMaker) were used. Both strands of obtained clones, CnA1, CnT1, CnD3, CnD9 and CnH14 were sequenced by the dideoxynucleotide chain-termination method using BcaBEST Sequencing Kit (TAKARA). Deletion mutants for sequencing were generated by the exonuclease III/mung bean nuclease deletion technique.

Fusion proteins and preparation of the antibodies. Each cDNA fragment was ligated inframe to pGEX-2T (Pharmacia) or pRSET (Invitrogen) and the constructs were transformed to the *Escherichia coli* strain DH1 or BL21, respectively. The fusion proteins with glutathione S-transferase or 6×His tag were induced by the addition of isopropyl β-D-thiogalactopyranoside and purified through glutathione Sepharose 4B (Pharmacia) or Ni-NTA-agarose (QIAGEN) columns, respectively. The protein band was cut out from the gels electrophoresed in the presence of SDS (10). The polyclonal antibodies were raised in rabbits by injecting the isolated protein several times.

Immunoblots and immunoelectron microscopy. A total SDS extract of chicken breast muscle was electrophoresed using 2.3–4 % polyacrylamide gels (10). Immunoblot detection was carried out as described (9). Immunoelectron microscopy was performed using JEM 100S electron microscope as previously reported (7).

[1] To whom correspondence should be addressed. FAX: 81-43-290-2812.

0006-291X/96 $18.00

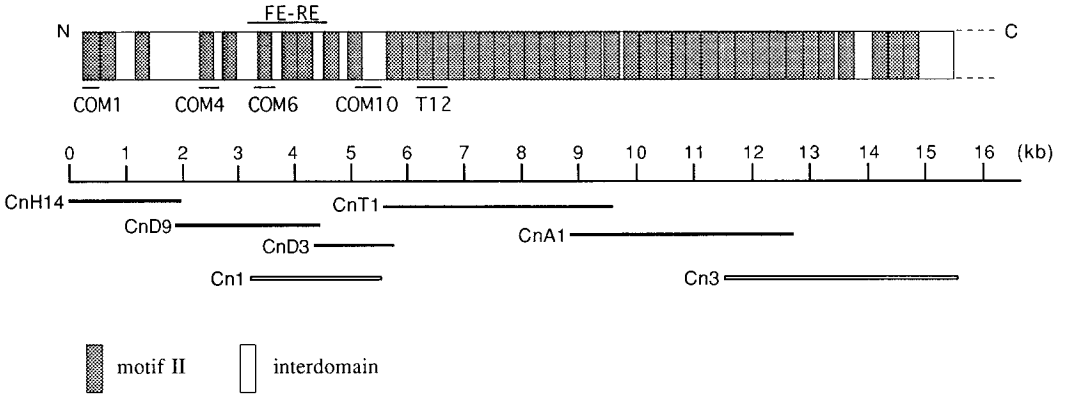

FIG. 1. Cloning, sequencing, and expression of the cDNA coding the N terminal region of connectin. Five cDNA clones, CnH14, CnD9, CnD3, CnT1 and CnA1, are sequenced. Cn1 and Cn3 were sequenced earlier (13, 12). Four connectin fragments, COM1, COM4, COM6, and COM10, are expressed in *E. coli.* For FE-RE and T12, see ref. (15).

RESULTS AND DISCUSSION

Sequencing of the N Terminal Region of Connectin

We screened a cDNA library prepared from poly(A)$^+$ RNA of adult chicken breast muscle using polyclonal antibodies to the 1200 kDa fragment of rabbit skeletal muscle connectin (Pc1200) and monoclonal antibody T12. Pc1200 binds to mainly the Z line (9) and T12 binds to an I band region approximately 0.1 μm apart from the Z line (3). Two clones, CnA1 (Pc 1200) and CnT1 (T12), were obtained. CnA1 (3.8 kb) partially overlapped with CnT1 (4.2 kb) and Cn3 (4 kb), as shown in Figure 1. Sequencing revealed that both CnA1 and CnT1 consisted almost entirely of motif IIs, immunoglobulin C2 domain (cf. 11).

"Walking" toward the N terminus of connectin was then attempted starting from the 5′ terminus of CnT1 using whole CnT1 as probe. Twenty one clones thus obtained were sequenced. A 15.5 kb cDNA of overlapped clones with one open reading frame including Cn3 (12) was obtained (EMBL accession number D83390). Sequencing of CnH14 revealed the presence of stop codon TGA in the

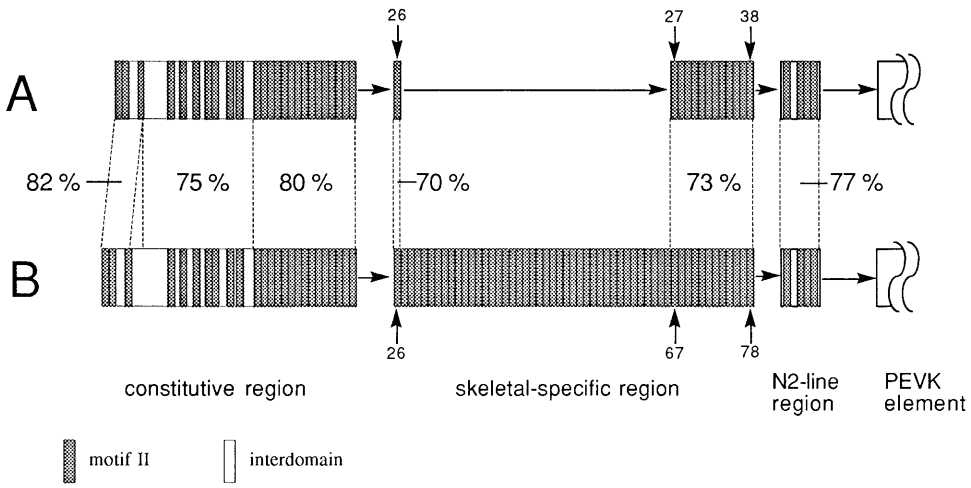

FIG. 2. Comparison of the structure of connectin isoforms of chicken and human skeletal muscles. A, chicken breast muscle; B, human skeletal muscle (8). Percentage indicates homology of amino acid sequence. See Labeit and Kolmerer (8). The 26th, 27th, 38th, 67th, and 78th motif IIs are indicated.

upstream of initiation codon ATG separated by 21 nucleotides. Hence the 5′ terminus of connectin cDNA was determined.

In the present study, 11.5 kb cDNA of connectin corresponding to 3752 amino acids was newly sequenced from its 5′ terminus. There are 31 motif IIs in this region (Figure 1). Two of them are present at the N terminal region followed by a short interdomain and the 3rd motif II. After a long interdomain 7 motif IIs follow with short interdomains; then 21 motif IIs exist in tandem. These domain repeats are very similar to those of human connectin (8).

During the present work, it became apparent that Cn1 previously assigned to the C terminal region of connectin (13) is a part of the N terminal region (14) (Figure 1). Sebestyén *et al.* (15) pointed out that there were three deletions of the triplet code in the region from the 4th interdomain to the 8th motif II (Cn1) of chicken skeletal connectin (13). The present work reexamined the sequence of the region (CnD9 and CnD3) and found that two triplets were actually missing and the other was due to sequencing error. In addition, it was pointed out that an 18 amino acid sequence in the 6th motif II lacked consensus motif II sequence (15). Reexamination showed that there were 3 additional. nucleotides in 60 bp in the 6th motif resulting in homologous sequences with rabbit cardiac connectin cDNA.

Skeletal and Cardiac Isoforms of Connectin

Connectin is an extraordinarily long, filamentous protein (∼3000 kDa) extending from the Z line to the M line in striated muscle sarcomeres (1,2). Within the region of the connectin molecule that is localized in the I band *in situ,* skeletal and cardiac muscle-specific isoforms are expressed by differential splicing (8). Comparison of the sequences of human cardiac and skeletal (8) and chicken skeletal (this work) muscle connectins shows an interesting feature of connectin isoforms (Figure 2). The 3752 amino acid sequence from the N terminus (present work) and the successive 1323 amino acid sequence (12) cover the elastic portion of connectin ranging from the Z line to the N_2 line in the I band of a sarcomere in chicken breast muscle.

As shown in Figure 2, there is high homology (75–82%) between human and chicken connectins between the N terminus and the 25th motif II, except that there are additional 206 amino acids in the second interdomain of human connectin. This follows a tissue-specific splicing domain: there

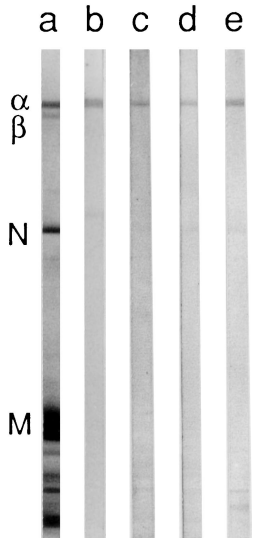

FIG. 3. Specificity of PcCOMs. A total SDS extract of chicken breast muscle was electrophoresed on 2.3–4% polyacrylamide gels (10). a, Amido Black stain; b, reacted with PcCOM1; c, reacted with PcCOM4; d, reacted with PcCOM6; e, reacted with PcCOM10. α, α-connectin; β, β-connectin; N, nebulin; M, myosin heavy chain.

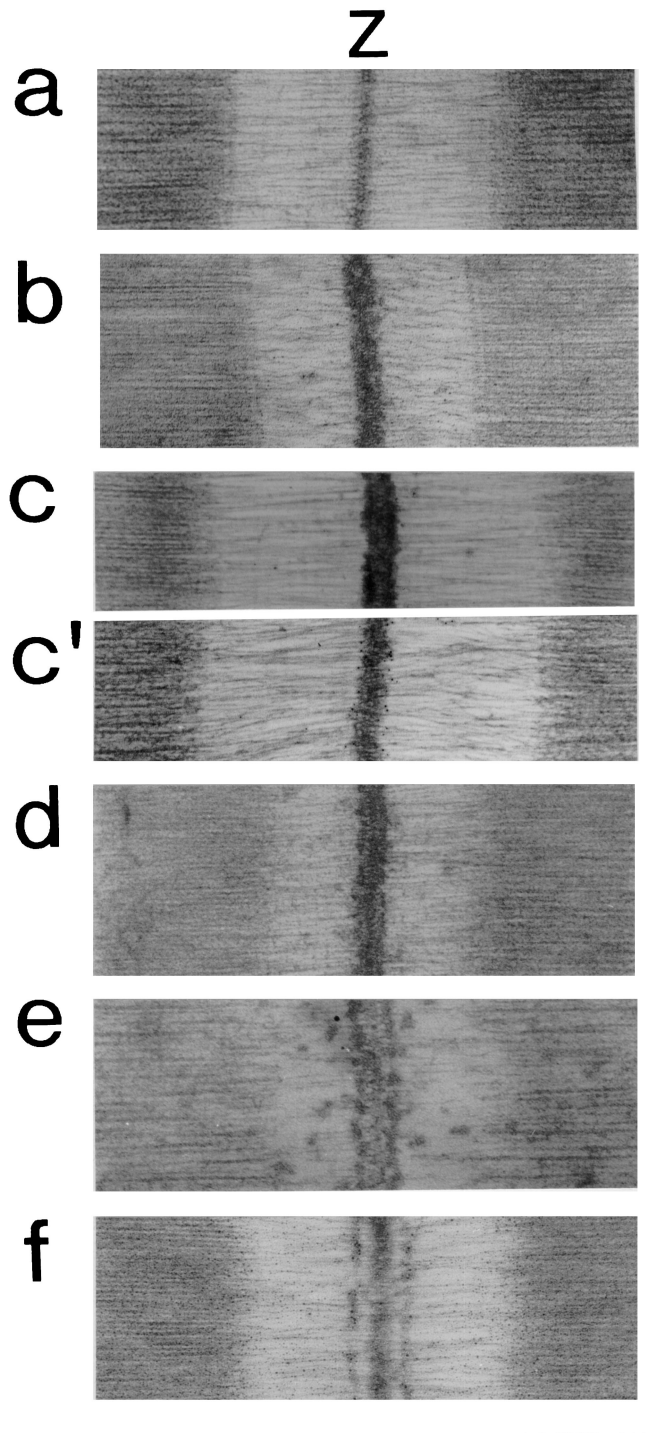

FIG. 4. Immunoelectron micrographs of chicken breast muscle sarcomeres treated with PcCOMs. a, control; b, treated with PcCOM1; c and c′, treated with PcCOM4; d, treated with PcCOM6; e, treated with PcCOM10; f, treated with T12. a–f, FITC-labelled second antibodies were used; c′, Gold-labelled second antibody was used. Z, Z line; bar, 0.5 μm.

are 53 motif IIs in tandem in human skeletal muscle connectin whereas motif IIs from the 27th to 66th are missing in chicken connectin. The homology between corresponding 26th motif IIs was 70%, whereas that between the 26th motif II (chicken) and the 27-66th motif IIs (human) was less than 40%. The homology of 12 motif IIs, 27–38th (chicken) and 67–78th (human) was 73%.

Immunoelectron Microscopic Location of Several Fusion Proteins at the Z Line

Fusion proteins corresponding to the N terminal motif II (COM1, 10 kDa, from amino acid 3 to 100), 4th motif II (COM4, 15 kDa from 684 to 816), 6th motif II (COM6, 15 kDa from 1016 to 1147) and 10th motif II to 8th interdomain (COM10, 16 kDa, from 1622 to 1760) were expressed in *E. coli* (see Figure 1). Polyclonal antibodies to each polypeptide raised in a rabbit reacted with α-connectin (mother molecule) but not with β-connectin (N-terminal region free fragment; cf.9) (Figure 3).

Immunoelectron microscopy using the above antisera as probe clearly indicated that COM1 and COM4 were localized at the Z line of chicken breast muscle sarcomeres (Figure 4b, c). It appears that COM4 was localized at both edges of a Z line (Figure 4c, c′), while COM6 and COM10 were localized very near the Z line in the I band (Figure 4d, e). These observations and the result with T12 (0.1 μm away from the Z line (3); Figure 4f) strongly suggest that some 800 amino acid peptides of the N terminal region of connectin are involved in the binding to the Z line.

Greaser and associates prepared polyclonal antibodies to the 56 kDa peptide corresponding to from the 6th to 9th motif IIs in rabbit cardiac connectin (FE-RE in Figure 1) and showed that under immunofluorescence microscope the portion was localized at the Z line (15). However, in the present immunofluorescence observations, COM6 (6th motif II) appeared to be localized at the Z line (data not shown), although immunoelectron microscopic studies revealed that it was present just outside of the Z line (Figure 4d). Monoclonal antibody T12 to the 13th and 14th motifs (15) binds to the region 0.1 μm apart from the Z line (3). Therefore, it is likely that the portion from the N terminus up to the 5th motif II is involved in the binding of connectin to the Z line (see Figure 1).

REFERENCES

1. Maruyama, K. (1994) *Biophys. Chem.* **50,** 73–85.
2. Trinick, J. (1994) *Trends Biochem. Sci.* **19,** 405–409.
3. Fürst, D. O., Osborn, M., Nave, R., and Weber, K. (1988) *J. Cell Biol.* **106,** 1563–1572.
4. Horowits, R., Maruyama, K., and Podolsky, R. J. (1989) *J. Cell. Biol.* **109,** 2169–2176.
5. Nave, R., Fürst, D. O., and Weber, K. (1989) *J. Cell Biol.* **109,** 2177–2187.
6. Suzuki, J., Kimura, S., and Maruyama, K. (1994) *J. Biochem.* **116,** 406–410.
7. Itoh, Y., Suzuki, T., Kimura, S., Ohashi, K., Higuchi, H., Sawada, H., Shimizu, T., Shibata, M., and Maruyama, K. (1988) *J. Biochem.* **104,** 504–508.
8. Labeit, S., and Kolmerer, B. (1995) *Science* **270,** 293–296.
9. Kimura, S., Matsuura, T., Ohtsuka, S., Nakauchi, Y., Matsuno, A., and Maruyama, K. (1992) *J. Muscle Res. Cell Motil.* **13,** 39–47.
10. Laemmli, U. K. (1970) *Nature* **227,** 680–685.
11. Politou, A. S., Gautel, M., Improta, S., Vangelista, L., and Pastore, A. (1996) *J. Mol. Biol.* **255,** 604–616.
12. Maruyama, K., Endo, T., Kume, H., Kawamura, Y., Kanzawa, N., Nakauchi, Y., Kimura, S., Kawashima, S., and Maruyama, K. (1993) *Biochem. Biophys. Res. Comm.* **194,** 1288–1291.
13. Maruyama, K., Endo, T., Kume, H., Kawamura, Y., Kanzawa, N., Kimura, S., Kawashima, S., and Maruyama, K. (1994) *J. Biochem.* **115,** 147–149.
14. Yajima, H., Ohtsuka, H., Kume, H., Endo, T., Maruyama, K., Kimura, S., and Maruyama, K. (1996) *Zool. Sci.* **13,** 119–123.
15. Sebestyén, M. G., Wolff, J. A., and Greaser, M. L. (1995) *J. Cell Sci.* **108,** 3029–3037.