THE AMINO ACID SEQUENCE OF RABBIT SKELETAL MUSCLE TROPONIN C: GENE REPLICATION AND HOMOLOGY WITH CALCIUM -BINDING PROTEINS FROM CARP AND HAKE MUSCLE

John H. COLLINS, James D. POTTER, Marcus J. HORN, Gedelita WILSHIRE and Nora JACKMAN

Boston Biomedical Research Institute, Department of Muscle Research
20 Staniford St., Boston, Mass, 02114 USA

Received 25 June 1973
Revised version received 7 September 1973

1. Introduction

Troponin, a protein found in the thin filaments of muscle of a wide variety of species [1], in combination with tropomyosin regulates muscle contraction by conferring a calcium sensitivity on the interaction of actin and myosin, as originally shown by Ebashi [2]. Troponin is a complex of three proteins [3, 4]: a calcium binding protein (TN-C); an inhibitory protein (TN-I) and a protein (TN-T) which binds to tropomyosin. TN-C is a single polypeptide chain of molecular weight about 18 000 [5].

In lower vertebrates (but not in mammals) a family of low mol. wt (about 12 000), soluble muscle calcium-binding proteins (MCBP), or parvalbumins, are found in abundance [6–11]. Although the biological role of the MCBPs is unknown, it has been speculated that they may be homologous with TN-C [6].

We report here the amino acid sequence of TN-C. The possibility of gene replication in the evolution of TN-C and the relationship of its sequence with that of MCBP are discussed.

2. Results

The amino acid sequence of TN-C from rabbit skeletal muscle, shown in fig. 1, was determined as follows. The polypeptide chain was cleaved at its 9 methionine residues with cyanogen bromide (CB), the resulting peptides were isolated, and their amino acid sequences determined [12]. The larger CB-peptides were further cleaved with proteolytic enzymes. Sequences were determined by conventional procedures of subtractive Edman degradation and digestion with aminopeptidase M or carboxypeptidase A [13], or by means of an automatic, solid-phase peptide sequencer [14, 15]. There remain 10 residues (no. 1–4 and 64–69) whose sequence assignment is tentative. Full documentation of the sequence determination will be published when the order of these residues has been definitively established. The order of the CB peptides was determined by studies on the methionine-containing tryptic peptides of TN-C. Although these studies are still in progress, we have been able to unequivocally identify all of the necessary overlaps.

The calculated mol. wt of the polypeptide chain of TN-C is 17 846, and it has a net negative charge of 30 (assuming complete ionization of all carboxyl, amino, guanidino and imidazole groups). Included in the 158 amino acid residues are one residue each of cysteine, histidine and proline, and two residues each of tyrosine and asparagine. Tryptophan is absent from TN-C.

We have used the 'diagram' method [16] for comparing the sequences of TN-C and MCBP, and also for seeking internal repeats in the TN-C sequence. Similarities between sequences were evaluated both by counting the number of identical residues and by using the '0 to 9' scale proposed by McLachlan [17] for comparing distantly related sequences. TN-C was com-

North-Holland Publishing Company – Amsterdam

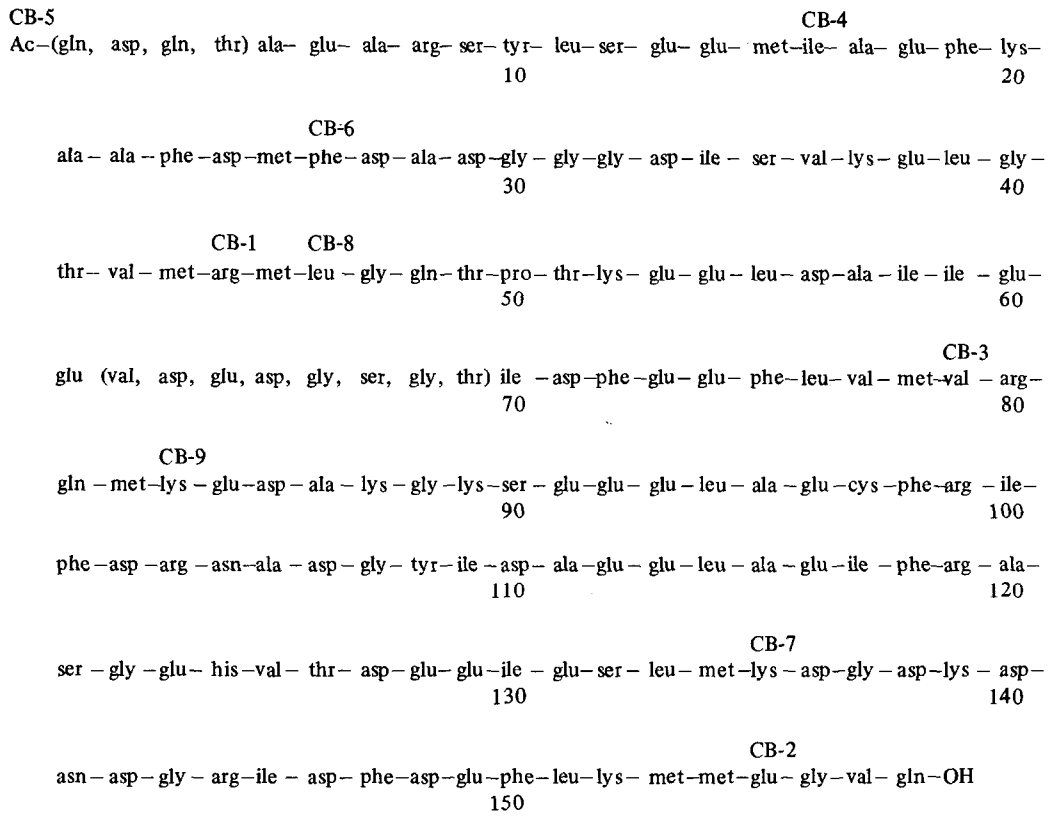

Fig. 1. The amino acid sequence of rabbit skeletal muscle TN-C. The order of the nine CB peptides is shown. The order of residues in parentheses is tentative.

pared with four homologous MCBPs of known sequence: carp components 2, 3, and 5 and the major component of hake [9, 11, 18]. Each of these MCBPs has 108 residues, with 70 positions invariant among them. The 3-dimensional structure of carp component 3 has been determined [19] and the other three MCBPs are probably very similar in structure [18]. The best alignment of the TN-C and MCBP sequences, which gives 35 identities, is shown in fig. 2. TN-C has hydrophobic residues at 17 of the 22 positions where hydrophobic core residues occur in MCBP. The greatest sequence similarity is between the C-terminal regions of TN-C (residues 116–156) and MCBP (residues 68–108), where 17 of 41 residues are identical and all 8 of the hydrophobic core residues of MCBP have corresponding hydrophobic residues in TN-C. It is necessary to place a 3 residue gap between residues 67 and 68 of TN-C to continue the similarity into the middle third of the MCBP sequence. The N-terminal third of MCBP and residues 45–84 of TN-C show very little (if any) similarity.

Kretsinger [18, 19] proposed that gene triplication has occurred to produce three homologous regions (called AB, CD and EF) in MCBP. The sequence similarities are barely detectable and the internal homologies were inferred from similarities in the 3 dimensional structures of these regions. McLachlan [20] also examined the sequence and structure of MCBP and concluded that the CD and EF regions (both of which include one calcium-binding site) are probably homologous, but that there is no evidence for homology of either CD or EF with the AB region (which lacks a calcium-binding site). TN-C can be divided into four regions (I, II, III and IV) which are similar in sequence to each other and to the CD and EF regions of MCBP. These are shown in fig. 3. The AB region is slightly

TN-C met–leu–gly–gln–thr–pro–thr–lys–glu–glu–leu–asp–ALA–ile–ile–GLU–glu–val–
50 60

MCBP Ac–ala–phe–ala–gly–val–leu–asn–asp–ala–asp–ile–thr–ALA–ala–leu–GLU–ala–cys–
10

TN-C asp–glu–asp–GLY–SER–gly–thr–ile–asp–phe–glu–GLU–PHE–leu–val–met–VAL–arg–gln–
70 80

MCBP (glu)ala(glu–GLY)SER–phe–asp– his–lys(GLU)PHE–phe–ala–lys–VAL–
20 31

TN-C met–lys–glu–asp–ala–LYS–GLY–LYS–SER–glu–glu–glu–leu–ala–glu–cys–PHE–arg–ILE–
90 100

MCBP gly–leu(LYS–GLY)LYS–SER–ala–asp–asp–val–lys–lys–ala–PHE–ala–ILE
40

TN-C phe–ASP–arg–asn–ala–asp–GLY–tyr–ILE–asp–ala–glu–GLU–LEU–ala–glu–ile–phe–arg–
111

MCBP ile–ASP–gln–asp–lys–ser–GLY–phe–ILE–glu–glu–asp–GLU–LEU–lys–leu–phe–leu–gln
51 60

TN-C ala–ser– gly–glu–his–val–THR–ASP–glu–GLU–ile–glu–ser–leu–met–LYS–
120 130

MCBP asn–phe–lys–ala–asp–ala–arg–ala–leu–THR–ASP(ala)GLU–thr–lys–thr–phe–leu–LYS–
71 80

TN-C asp–GLY–ASP–lys–ASP–asn–ASP–GLY–arg–ILE–asp–phe–ASP–GLU–PHE–leu–lys–MET–met
141 151

MCBP ala–GLY–ASP–ser–ASP–gly–ASP–GLY–lys–ILE–gly–val–ASP–GLU–PHE–thr–ala(MET)val–
90 100

TN-C glu–GLY–val–gln–OH
MCBP lys(GLY)–OH

Fig. 2. Comparison of the sequences of MCBP and rabbit TN-C. The MCBP sequence is that of carp component 3 except where corresponding residues (given in parentheses) in components 2 or 5 or in the hake protein give identity or closer similarity [11] with TN-C. Identical residues are capitalized. Hydrophobic core residues in MCBP and corresponding hydrophobic residues in TN-C are underlined. Asterisks (*) are used to indicate residues in MCBP that are involved in calcium-binding. Gaps in the MCBP sequence are those introduced by Kretsinger and Nockolds in their comparison of three segments of the MCBP with each other [19].

similar to I, but not at all to the other regions of TN-C. Within TN-C, the greatest similarity is between regions II and IV, and both of these are more similar to EF than to CD. The greatest overall similarity among all regions (except AB) is in the 12 residue segments corresponding to the two calcium-binding sites of MCBP. The distribution of hydrophobic residues among the different regions is also very similar. In the 5 positions where hydrophobic core residues occur in both CD and EF, hydrophobic residues occur in all four regions of TN-C. There are 8 segments in TN-C where helices can be predicted from the sequence by the 'helical wheel' method [21]. These correspond very well with the known α-helices in MCBP (see fig. 3).

3. Conclusions

The sequence similarities shown in fig. 3 lead us to conclude that TN-C has evolved by gene replication, and that it is homologous with MCBP. However, the possibility of convergent evolution between the two cannot be completely excluded. The great similarity in the sequence of the calcium-binding sites of MCBP

MCBP				TN-C			
AB	CD	EF		I	II	III	IV
1 a		72 a		9 s		84 e	
2 f,y	34 g	73 d,g		10 y		85 d	
3 a,s	35 L	74 a		11 L	47 g	86 a	122 g
4 g	36 t,a,k	75 r		12 s	48 q	87 k	123 e
5 v,i	37 s,g	76 a		13 e	49 t	88 g	124 h
6 L	38 k	77 L		14 e	50 p	89 k	125 V
7 n,a	39 s,t	78 t		15 m	51 t	90 s	126 t
8 d	40 a,p	79 d		16 i	52 k	91 e	127 d
9 a	41 d,a	80 g,a		17 a	53 e	92 e	128 e
10 d	42 d	81 e		18 e	54 e	93 e	129 e
11 I	43 V,I	82 t		19 F	55 L	94 L	130 I
12 a,t	44 k	83 k,a		20 k	56 d	95 a	131 g
13 a	45 k	84 t,a		21 a	57 a	96 e	132 s
14 a	46 a	85 F		22 a	58 I	97 c	133 L
15 l	47 F	86 L		23 F	59 I	98 F	134 M
16 e,a	48 a	87 k		24 d	60 e	99 r	135 k
17 a	49 i	88 a		25 m	61 e	100 i	136 d
18 c	50 I	89 g		26 F	62 V	101 F	137 g
19 k,e	51 d*	90 d*		27 d	63 d	102 d	138 d
20 a	52 q	91 s		28 a	64 e	103 r	139 k
21 a,e	53 d*	92 d*		29 d	65 d	104 n	140 d
22 d,g	54 k	93 g		30 g	66 g	105 a	141 n
23 s	55 s*	94 d*		31 g	67 s	106 d	142 d
24 F	56 g,d	95 g		32 g	68 g	107 g	143 g
25 d,n,k	57 f*	96 k*		33 d	69 t	108 y	144 r
	58 I,V	97 I		34 I	70 I	109 I	145 I
	59 e*	98 g		35 s	71 d	110 d	146 d
26 h,e	60 e	99 v		36 v	72 f	111 a	147 f
27 k	61 d	100 d,e		37 k	73 e	112 e	148 d
28 a,e	62 e*	101 e*		38 e	74 e	113 e	149 e
29 F	63 L	102 F		39 L	75 F	114 L	150 F
30 F	64 k	103 t,a	1 q	40 g	76 L	115 a	151 L
31 a,s,t	65 l	104 a	2 d	41 t	77 v	116 e	152 k
32 k	66 F	105 L,M	3 q	42 V	78 M	117 I	153 M
33 V	67 L	106 V	4 t	43 M	79 V	118 F	154 M
	68 q	107 k	5 a	44 r	80 r	119 r	155 e
	69 n	108 a,g	6 e	45 m	81 q	120 a	156 g
	70 F		7 a	46 L	82 M	121 s	157 V
	71 k,s		8 r		83 k		158 q

Fig. 3. Proposed homologous regions in TN-C and MCBP. The alignment of the MCBP regions is according to Kretsinger and Nockolds [19]. The regions of TN-C are aligned on the basis of maximum overall sequence similarity [9, 11]. Residue numbers and single letter abbreviations [23] for the amino acids are given. Residues from carp component 3 in MCBP are listed first, followed by substitutions (if any) occurring in components 2 or 5 or in the hake protein. Hydrophobic core residues in MCBP and corresponding hydrophobic residues in TN-C are capitalized. Residues in MCBP involved in calcium-binding are indicated by asterisks (*). Vertical bars indicate helical regions in MCBP and predicted α-helices [21] in TN-C.

and the four corresponding segments of TN-C suggest that TN-C has four calcium binding sites, located at residues 27–38, 63–74, 102–113 and 138–149. This correlates well with recent Ca^{2+} binding studies on TN-C [22] in which the binding of 4 Ca^{2+} ions per mole is demonstrated. The correspondence of α-helices and hydrophobic residues suggests that each of the four regions of TN-C has a three-dimensional structure very similar to the CD and EF regions of MCBP. The validity of these conclusions may be tested by determining the three-dimensional structure of TN-C.

Acknowledgements

This work was supported by grants from the National Institutes of Health (AM-14728 and HL-05949), the National Science Foundation, the American Heart Association, and the Central Chapter of the Massachusetts Heart Association. J.D.P. was a Postdoctoral Trainee of the National Institutes of Health (HL-05811) during the initial phases of this work and is currently a fellow of the Muscular Dystrophy Associations of America, Inc. The authors wish to thank Dr. Marshall Elzinga for advice and encouragement, and Dr. Marion Greaser for determining the N-terminal amino acetyl group of TN-C.

References

[1] Lehman, W., Kendrick-Jones, J. and Szent-Györgyi,A.G. (1972) Cold Spring Harbor Symp. Quant. Biol. 37, 319.
[2] Ebashi, S. and Endo, M. (1968) Prog. Biophys. Mol. Biol. 18, 123.
[3] Greaser, M.L. and Gergely, J. (1971) J. Biol. Chem. 246, 4226.
[4] Greaser, M.L., Yamaguchi, M., Brekke, C., Potter, J. and Gergely, J. (1972) Cold Spring Harbor Symp. Quant. Biol. 37, 235.
[5] Hartshorne, D.J. and Dreizen, P. (1972) Cold Spring Harbor Symp. Quant. Biol. 37, 225.
[6] Pechere, J.-F., Capony, J.-P., Ryden, L. and Demaille, J. (1971) Biochem. Biophys. Res. Commun. 43, 1106.
[7] Benzonana, G., Capony, J.-P. and Pechere, J.-F. (1972) Biochim. Biophys. Acta 278, 110.
[8] Pechere, J.-F., Capony, J.-P. and Ryden, L. (1971) Eur. J. Biochem. 23, 421.
[9] Capony, J.-P., Ryden, L., Demaille, J. and Pechere, J.-F. (1973) Eur. J. Biochem. 32, 97.
[10] Nockolds, C.E., Kretsinger, R.H., Coffee, C.J. and Bradshaw, R.A. (1972) Proc. Natl. Acad. Sci. U.S., 69, 581.
[11] Coffee, C.J. and Bradshaw, R.A. (1973) J. Biol. Chem. 248, 3305.
[12] Collins, J.H. and Potter, J.D. (1973) Federation Proc. 32, 569.
[13] Elzinga, M. (1970) Biochemistry 9, 1365.
[14] Laursen, R.A. (1971) Eur. J. Biochem. 20, 89.
[15] Laursen, R.A., Horn, M.J. and Bonner, A.G. (1972) FEBS Letters 21, 67.
[16] Gibbs, A.J. and McIntyre, G.A. (1970) Eur. J. Biochem. 16, 1.
[17] McLachlan, A.D. (1971) J. Mol. Biol. 61, 409.
[18] Kretsinger, R.H. (1972) Nature New Biol. 240, 85.
[19] Kretsinger, R.H. and Nockolds, C.E. (1973) J. Biol. Chem. 248, 3313.
[20] McLachlan, A.D. (1972) Nature New Biol. 240, 83.
[21] Schiffer, M. and Edmundson, A.D. (1970) Biophys. J. 10, 293.
[22] Potter, J.D., Seidel, J.C., Leavis, P., Lehrer, S.S. and Gergely, J., Symposium on Calcium Binding Proteins (Drabikowski, W. and Carafoli, E., eds), Elsevier, in press.
[23] Dayhoff, M.O. (1972) Atlas of Protein Sequence and Structure, vol. 5, p. D-2, National Biochemical Research Foundation, Washington, D.C.