HOMOLOGY OF MYOSIN LIGHT CHAINS, TROPONIN-C AND PARVALBUMINS DEDUCED FROM COMPARISON OF THEIR AMINO ACID SEQUENCES

John H. Collins

Department of Muscle Research, Boston Biomedical Research Institute, Boston, Mass. 02114. U.S.A.

Received March 26,1974

The "alkali" light chains of rabbit skeletal muscle myosin have been compared with troponin-C from the same source, and both of these are compared with parvalbumins from pike, hake and carp. The similarities in amino acid sequence indicate that these proteins all evolved from a common ancestor.

Parvalbumins are small (about 108 residues) calcium-binding proteins found in the skeletal muscle of aquatic lower vertebrates, but not in mammals, reptiles or birds (1,2). Although their physiological function is still a subject of speculation, their structure has been thoroughly characterized. Complete amino acid sequences of parvalbumins from pike (3), hake (4) and carp (5), and partial sequences of several others (2) have been reported. The three-dimensional structure of a carp parvalbumin has been determined at 1.85-A resolution by X-ray diffraction (6). There are two calcium-binding sites, and each is confined to an 11-residue segment along the polypeptide chain. Kretsinger (6,7) found that parvalbumins can be divided into three segments, each 33-37 residues long. Homology between two of the segments (each of which contains a calcium-binding site) was inferred principally from similarities in their three-dimensional structures.

Troponin is a complex of three proteins (8) found in the thin filaments of muscle of a wide variety of species (9), which, in combination with tropomyosin, regulates muscle contraction by conferring a calcium sensitivity on the interaction of actin and myosin (for review, see ref. 10). The

calcium-binding component of the troponin complex, TN-C*, is a polypeptide chain of molecular weight about 18,000. The amino acid sequence of rabbit skeletal muscle TN-C shown in Fig. 1 differs from that previously reported (11,12) in two respects: an additional residue of methionine occurs at position 79, and residue 141 is asparagine rather than aspartic acid. Sequence assignments of residues no. 1-4 and 63-69 are still somewhat tentative. A detailed description of the sequence determination of TN-C will be published when the order of these residues has been definitively established.

Our previous results (11,12) have borne out predictions by others (7,13) that the amino acid sequence of TN-C is similar to those of parvalbumins. We divided the sequence of TN-C into four segments (36-38 residues long) similar in sequence to each other and to the two calcium-binding regions of parvalbumins. From comparison of the sequences and knowledge of the three-dimensional structure of parvalbumins (6), we located four apparent calcium-binding sites in TN-C. The prediction of four calcium-binding sites in TN-C agreed with some (14,15), but not all (8,16,17) experimental binding studies. We further proposed that the calcium-binding regions of TN-C had three-dimensional structures similar to those of parvalbumins, and that the two proteins were probably homologous, each having evolved by replication of a small (about 35 residues) precursor with a single calcium-binding site.

Myosins from a variety of species and muscle types (9,18,19) are hetero-multimers consisting of two heavy chains (molecular weight about 200,000) and three or four light chains (molecular weight about 20,000). In some primitive species, the troponin-mediated calcium sensitivity of muscle contraction is replaced by a regulatory system associated with the light chains of myosin (18,19). Other species appear to have both myosin-linked and troponin-linked regulatory systems (9). It is thus of great interest to study the evolutionary relationship of troponin and myosin light chains. The best

*Non-standard abbreviations used in this paper are: TN-C, troponin-C; LC-1 and LC-3, the two "alkali" light chains of rabbit skeletal muscle myosin.

way to do this is by comparing amino acid sequences. The only myosin light chains for which significant amounts of sequence have been reported are the so-called "alkali" light chains of rabbit skeletal muscle myosin (20), LC-1 and LC-3 (21). These light chains, either purified or in whole myosin, do not bind calcium at physiological concentrations of magnesium (14), but have been implicated in the magnesium-ATP-binding site of myosin (22). Weeds and Frank (23) have shown that LC-3 contains 152 amino acid residues and LC-1 contains 193 residues. The sequence of LC-1 is very similar (but not identical) to that of LC-3, except for an additional 41-residue segment at the amino-terminus of LC-1 which is unusually rich in alanine and proline residues. Sequences of the amino-terminal cyanogen bromide peptides of LC-1 (residues 1-99) and LC-3 (residues 1-58), and of a 27-residue thiol peptide common to both LC-1 and LC-3 have been reported (23,24).

Fig. 1 shows the alignment of the sequences of TN-C, LC-3 (and LC-1) and parvalbumins. The "diagram" procedure (25) was used for individual comparisons to produce in each case the maximum number of identical residues while introducing as few deletions as possible. Similarity between a pair of sequences was measured simply by calculating the number of identical residues as a percentage of total residues (including deletions, if any) compared. A more sophisticated analysis which also takes into account the degree of similarity between nonidentical residues (26) must await the definitive completion of sequence studies on TN-C, LC-1 and LC-3. For the present study, the following sequences were compared: the two halves of TN-C with each other; each half of TN-C with the published partial sequence of LC-3; each half of TN-C with pike parvalbumin; the partial sequence of LC-3 with pike parvalbumin. The unusual amino-terminal segment of LC-1 (23) was not similar to any other sequence mentioned in this report. Pike parvalbumin (3), not previously available for comparison with TN-C (11,12), proved to be more similar to both TN-C and LC-3 than are parvalbumins from carp and hake. The sequences of the hake (4) and carp (5) parvalbumins were aligned with the pike

Table 1

TN-C (1-83)	TN-C (84-159)	Light Chains LC-3	Light Chains LC-1	Light Chains (SH)	Parvalbumins Pike	Parvalbumins Hake	Parvalbumins Carp
asp			LYS		25 asn	LYS	asp
thr			ile		his		
gln			asp		lys	gly	
gln			leu		ala	glu	
5 ala			ser		phe		
glu			ala		30 phe		
ala			ile		ala	thr	
arg	LYS		LYS		LYS		
ser	85 glu		ILE		val	ILE	
10 tyr	asp	ser	glu		gly		
leu	ala	phe			35 leu		
SER	lys	SER			lys		thr
glu	gly	ALA	lys		ALA	gly	ser
GLU	90 lys	5 asx	GLU		met	lys	lys
15 met	ser	glx	gln		ser		
ILE	glu	ILE	gln		40 ala		
ALA	glu	ALA	ASP		asn	ALA	ASP
GLU	GLU	GLX	GLU		asp		
PHE	95 leu	10 PHE			val	ile	
20 LYS	ala	LYS			LYS		
ala	GLU	GLU			45 lys		
ALA	cys	ALA			val		ALA
PHE	PHE	PHE			PHE		
asp	100 arg	15 leu			lys	gly	ala
25 met	ile	leu			ala	ile	ile
phe	phe	tyr			50 ile		
ASP*	ASP*	ASP*			ASP*		
ala	ARG	ARG			ala	gln	gln
asp*	105 asn	20 thr			asp*		
30 GLY	ala	GLY			ala	lys	lys
gly	ASP	ASP			55 ser*		
gly	gly	ser			gly	asp	
asp	tyr	lys			phe		
ILE	110 ILE	25 ILE			ILE	val	
35 ser	asp	thr			glu*		
val	ala	leu			60 glu		
lys	glu	ser			glu	asp	asp
glu*	glu*	gln			glu*		
leu	115 leu	30 val			leu		
40 GLY	ala	GLY			lys		
thr	glu	asp			65 phe	leu	leu
VAL	ile	VAL			VAL	phe	phe
met	-	LEU			LEU		
-	-	-			lys	gln	gln
-	-	-			ser	asn	asn
-	phe	-			70 phe		
ARG	120 ARG	35 ARG			ala	ser	lys
45 met	ALA	ALA			ALA		
LEU	ser	LEU			asp	gly	
GLY	GLY	GLY			GLY	ala	ala
gln	glu	thr			75 arg		

Figure 1 (contd)

TN-C (1-83)	TN-C (84-159)	Light Chains LC-3	Light Chains LC-1	Light Chains (SH)	Parvalbumins Pike	Parvalbumins Hake	Parvalbumins Carp
thr	125 his	40 asn			asp	ala	ala
50 PRO	val	PRO		met	leu		
THR	THR	THR		lys	THR		
lys	ASP	ASX		glu	ASP		
GLU	GLU	ALA		GLU	80 ALA		gly
GLU	130 GLU	45 GLX		GLU	GLU		
55 leu	ile	val		val	thr		
asp	GLU	LYS		GLU	LYS	ala	
ALA	ser	lys		ALA	ALA	thr	thr
ile	LEU	val		LEU	85 phe		
ile	135 MET	50 LEU		MET	LEU		
60 glu	lys	gly		-	lys		
glu	asp	asn		ALA	ALA		
val	GLY	pro		GLY	ala	GLY	GLY
asp*	asp*	ser		gln	90 asp*		
GLU	140 lys	55 asp		GLU	lys	ser	ser
65 ASP*	asn	glu		ASP*	ASP*		
gly	asn	gln		ser	gly		
ser	asp*	met		asn	asp*		
GLY	GLY			GLY	95 GLY		
thr	145 arg			cys	lys		
70 ILE	ILE			ILE	ILE		
asp	asp			asn	gly		
phe	phe			tyr	ile	val	val
GLU	asp			GLU	100 asp	GLU	
glu*	150 glu*			ala	glu*		
75 PHE	PHE			PHE	PHE		
leu	leu			val	glu	ala	thr
val	LYS			LYS	thr	ala	ala
met	met				105 leu	met	
met	155 met				val		
80 val	glu				his	lys	lys
arg	gly				glu	gly	ala
gln	val				ala	-	-
met	gln						

Figure 1. Comparison of the sequences of TN-C, LC-3, LC-1 and parvalbumins from three species. Identities which involve LC-3 or LC-1 are capitalized. Asterisks (*) are used to indicate parvalbumin residues whose side chains are involved in calcium binding (7), and corresponding identical residues in TN-C, LC-3 and LC-1. Residues in LC-1 are given only where they differ from the corresponding residues of LC-3 (23). The sequence of the thiol (SH) peptide is common to both LC-1 and LC-3 (24). Residues of hake (4) and carp (5) parvalbumin are given only where they differ from the corresponding residues of pike (3) parvalbumin. For more details, see text.

sequence as shown by Frankenne, et al. (3). The amino-terminal part of parvalbumin sequences, which does not include a calcium-binding site (6), is not similar to any part of TN-C or LC-3, and is not discussed here. The

most favorable alignments deduced from individual comparisons were all mutually consistent.

As shown in Fig. 1, most of the TN-C sequence can be aligned to produce two halves with 30% identical residues, suggesting some symmetry in the three-dimensional structure of TN-C. In the three parvalbumin sequences considered here, an average of 32% of the residues are identical with those in the carboxyl-terminal half of TN-C, and 26% with those in the amino-terminal half. Thus, TN-C and parvalbumins are most similar at their carboxyl termini. Residues 1-58 of LC-3 (or residues 42-99 of LC-1) are most similar to a region near the amino terminus of TN-C, while the thiol sequence most closely resembles a segment near the carboxyl terminus of TN-C. There are 33% identical residues between TN-C and LC-3. It would not be surprising to find that the 67 additional residues (23) of LC-3 are similar to the remaining 64 residues of TN-C. Comparison of the sequences of LC-3 and parvalbumins gives about 5% fewer identical residues than does comparison of the corresponding segments of TN-C with parvalbumins. LC-3 does not contain specific, high-affinity calcium-binding sites of the type found in TN-C and parvalbumins. One would predict, then, that LC-3 is less similar than is TN-C to the parvalbumin segments containing the calcium-binding sites (see fig. 1).

The extent of similarity between TN-C, LC-3 and parvalbumins is too great to have occurred merely by chance (26). This implies that they are homologous, i.e., all the proteins have evolved from a common ancestor. However, the possibility of convergent evolution (the independent evolution of similar structures to perform similar functions) must be considered. In arguing for homology between the two calcium-binding regions of parvalbumins, Kretsinger (6,7) pointed out that calcium-binding sites in other proteins are quite different in structure. Furthermore, since all the residues involved in each parvalbumin calcium-binding site are close together in the sequence, gene replication of a smaller precursor (containing about 35 residues and

one calcium-binding site) would be a very favorable way of producing a protein capable of binding more than one calcium ion. These arguments can be applied even more forcefully in the case of TN-C, where there are four apparent calcium-binding regions and much stronger internal sequence repeats (11,12). TN-C is obviously a product of multiple gene duplication. Convergent evolution, by definition, is unlikely if two proteins are dissimilar. The physiological functions of TN-C and parvalbumins are clearly very different (2,5). For example, parvalbumins do not bind to myofibrils and are unable to replace TN-C in the regulation of muscle contraction. The properties of the "alkali" light chains of myosin and TN-C are so different that similarities in their sequences can not be due to convergent evolution.

Internal sequence repeats are much stronger in TN-C than in LC-3, i.e. TN-C is more closely related to the hypothetical precursor mentioned above. This may mean that LC-3 evolved from TN-C (or a precursor very similar to TN-C). The findings of Lehman, et al. (9) are not necessarily inconsistent with this proposal. It is possible that parvalbumins may also have evolved from a TN-C-like precursor. TN-C seems to be much more widely distributed among different species and muscle types than are parvalbumins (2,9), suggesting that TN-C is a more primitive type of protein.

Additional comparative sequence studies are required to clarify the evolutionary relationships discussed here. Such studies should also be helpful in understanding the properties of calcium-binding proteins and the mechanism of calcium-mediated regulation of muscle contraction.

Acknowledgments. This study was supported by a grant from the National Institutes of Health (no. AM-14728). Dr. M. L. Greaser and others (see refs. 11 and 12) have participated in the determination of the sequence of TN-C. I thank Drs. M. Elzinga and J. Gergely for their interest and helpful criticism.

References

1. Hamoir, G. (1957) Biochem. J. 66, 29P.
2. Demaille, J., Dutruge, E., Capony, J.-P. and Pechere, J.-F. (1973) Symposium on Calcium Binding Proteins (Drabikowski, W. and Carafoli, E., eds.), in press, Elsevier, Amsterdam.
3. Frankenne, F., Joassin, L. and Gerday, Ch. (1973) FEBS Letters 35, 145-147.
4. Capony, J.-P., Ryden, L., Demaille, J. and Pechere, J.-F. (1973) Eur. J. Biochem. 32, 97-108.
5. Coffee, C.J. and Bradshaw, R.A. (1973) J.Biol. Chem. 248, 3305-3312.
6. Kretsinger, R.H. and Nockolds, C.E. (1973) J. Biol. Chem. 248, 3313-3326.
7. Kretsinger, R.H. (1972) Nature 240, 85-87.
8. Greaser, M.L. and Gergely, J. (1973) J. Biol. Chem. 248, 2125-2133.
9. Lehman, W., Kendrick-Jones, J. and Szent-Gyorgyi, A.G. (1972) Cold Spring Harbor Symp. Quant. Biol. 37, 319-330.
10. Weber, A. and Murray, J.M. (1973) Physiol. Rev. 53, 612-673.
11. Collins, J.H., Potter, J.D., Horn, M.J., Wilshire, G. and Jackman, N. (1973) FEBS Letters 36, 268-272.
12. Collins, J.H., Potter, J.D., Horn, M.J., Wilshire, G. and Jackman, N. (1973) Symposium on Calcium Binding Proteins (Drabikowski, W. and Carafoli, E., eds.), in press, Elsevier, Amsterdam.
13. Benzonana, G., Capony, J.-P. and Pechere, J.-F. (1972) Biochim. Biophys. Acta 278, 110-116.
14. Bremel, R.D. and Weber, A. (1972) Nature 238, 97-101.
15. Potter, J.D., Seidel, J.C., Leavis, P., Lehrer, S.S. and Gergely, J. (1973) Symposium on Calcium Binding Proteins (Drabikowski, W. and Carafoli, E., eds.), in press, Elsevier, Amsterdam.
16. Drabikowski, W., Barylko, B., Dabrowska, R., Nowak, E. and Szpacenko, A. (1973) Symposium on Calcium Binding Proteins (Drabikowski, W. and Carafoli, E., eds.), in press, Elsevier, Amsterdam.
17. Hartshorne, D.J. and Pyun, H.Y. (1971) Biochim. Biophys. Acta 229, 698-711.
18. Kendrick-Jones, J., Szentkiralyi, E.M. and Szent-Gyorgyi, A.G. (1972) Cold Spring Harbor Symp. Quant. Biol. 37, 47-53.
19. Szent-Gyorgyi, A.G., Szentkiralyi, E.M. and Kendrick-Jones, J. (1973) J. Mol. Biol. 74, 174-203.
20. Weeds, A.G. and Lowey, S. (1971) J. Mol. Biol. 61, 701-725.
21. Sarkar, S. (1972) Cold Spring Harbor Symp. Quant. Biol. 37, 14-17.
22. Dreizen, P. and Richards, D.H. (1972) Cold Spring Harbor Symp. Quant. Biol. 37, 29-45.
23. Weeds, A.G. and Frank, G. (1972) Cold Spring Harbor Symp. Quant. Biol. 37, 9-18.
24. Weeds, A.G. (1967) Biochem. J. 105, 25C-27C.
25. Gibbs, A.J. and McIntyre, G.A. (1970) Eur. J. Biochem. 16, 1-11.
26. Barker, W.C. and Dayhoff, M.O. (1972) Atlas of Protein Sequence and Structure (Dayhoff, M.O., ed.), pp. 101-110, National Biomedical Research Foundation, Washington, D.C.