CYANOGEN BROMIDE FRAGMENTS OF THE CARDIAC I LIGHT CHAIN FROM BOVINE MYOSIN: EVIDENCE FOR SEQUENCE HOMOLOGY WITH RABBIT SKELETAL MYOSIN ALKALI LIGHT CHAINS

Alan G. WEEDS
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

Received 21 August 1975

1. Introduction

It has recently been shown that there is extensive homology of amino acid sequence between rabbit skeletal muscle myosin alkali light chains, rabbit troponin C and a calcium binding parvalbumin from carp [1–3], which indicates that these proteins may have evolved from a common precursor. Internal repeats within all three sequences suggest that gene duplication has occurred during their evolution. The structure of carp parvalbumin, which contains two calcium binding sites, has been determined by X-ray analysis [4]. Each site is located in a corner between two helical regions, giving a basic structural unit of the type: Helix A–Site S–Helix B (A–S–B). The calcium ions are ligated to a number of acidic residues within each site [4]. Duplication of this unit produces the two-site structure A1–S1–B1–L–A2–S2–B2, where L is a link region. The parvalbumin contains two additional helices at its N-terminal end, and the six helices pack together to form a hydrophobic core which stabilises the three-dimensional structure. The existence of this core depends on certain apolar residues located at specific positions within each helix forming van der Waals' interactions between the different helices. Similar groupings have been observed in the apolar residues of both troponin C and the alkali light chains, which suggest that these proteins may contain similar structural elements. Troponin C binds four calcium ions [5], and if the sequence is drawn out to correspond to the structural features of the parvalbumin, four units of the A–S–B type are produced [2]. Furthermore, within each site so generated acidic residues are found corresponding to the calcium ligating positions of the parvalbumin. The sequence of the alkali light chains has also been drawn out to correspond to the same structural features [2], but in this case the 'site' regions do not contain the requisite numbers of acidic residues found in the other two proteins, and there is no evidence that isolated alkali light chains bind calcium. Nevertheless the preservation of the pattern of apolar residues together with the extensive sequence homology, suggest strong structural similarities between these proteins. Table 1 shows this sequence drawn out to display these structural features.

One of the light chains from beef cardiac myosin contains an identical thiol sequence to that of the alkali light chains [6] but it has two other thiol groups which are absent from the rabbit protein. Cyanogen bromide fragments have been isolated from the cardiac light chain to investigate the location of these thiol groups within the sequence, and tryptic peptides from these fragments analysed. Although a complete sequence has not been produced, these preliminary studies provide evidence for extensive sequence homology between the cardiac and skeletal light chains. Furthermore, it is possible to order many of the tryptic peptides on the basis of these homologies, and the same specific patterns of apolar residues are observed within the sequence so generated.

Table I

Amino acid sequence of rabbit myosin alkali light chain and partial sequences from bovine cardiac light chain

	Pro *Pro*	*Lys* *Ala*	*Lys* *Pro*	*Asn* *Ala*	*Val* *Pro*	*Lys* *Lys*	*Lys* *Glu*	*Pro* *Glu*	*Ala* *Lys*	*Ala* *Ile*	*Ala* *Asp*	*Ala* *Leu*	*Ala* *Ser*	*Pro* *Ala*	*Ala* *Ile*	*Pro* *Lys*	*Lys* *Ile*	*Ala* *Glu*	*Pro* *Phe*	*Ala* *Ser* ASP	*Pro* *Lys* ARG	*Ala* *Glu* THR	*Pro* *Gln* PRO	*Ala*	
HELIX																									SITE
A1	*Gln* LYS	*Asp* CYS	*Glu* Glu	*Phe* MET/	*Lys* Lys	*Glu*	*Ala*	*Phe*	*Leu*	*Leu*	*Tyr*	———	LYS ►	*Asp*• (GLX	*Arg* GLX	*Thr*• GLX	*Gly* LEU	*Asp*• Asp	*Ser* Ser	*Lys*• Lys)	*Ile* (Ile	*Thr*• Thr			S1
B1	*Leu* TYR	*Ser* GLY	*Gln*• Gln	*Val* CYS	*Gly* Gly	*Asp* Asp	*Val* Val	*Leu* Leu	*Arg* Arg)	*Ala* (Ala	*Leu* Leu	———		*Gly* Gly	*Thr* GLN	*Asn* Asn	*Pro* Pro	*Thr* Thr							
A2	*Asn* GLN	*Ala* Ala	*Glu* Glu	*Val* Val	*Lys* LEU	*Lys* ARG)	*Val* (Val	*Leu* Leu	*Gly* Gly	*Asn* LYS	*Pro* Pro	*Ser* LYS)	— MET/	*Asp*•	*Glu*	*Gln*•	*Met*	*Asn*•	*Ala*	*Lys*•	*Lys*	*Ile*	*Glu*• / ASP ►		S2
B2	*Phe* Phe	*Glu* ASP	*Gln*• THR	*Phe* Phe	*Leu* Leu	*Pro* Pro	*Met* Met/	*Leu* (Leu ►	*Gln* Gln	*Ala* HIS	*Ile* Ile	———	———	*Ser* Ser	*Asn* LYS)	*Asn* (Asn	*Lys* Lys)	*Asp* (Asp	*Gln* THR	*Gly* Gly	*Thr* Thr					
A3	*Tyr* Tyr	*Glu* Glu	*Asp* Asp	*Phe* Phe	*Val* Val	*Glu* Glu	*Gly* Gly	*Leu* Leu	*Arg* Arg)	*Val* (Val	*Phe* Phe	———	———	*Asp*• Asp	*Lys* Lys)	*Glu*• (Glx	*Gly* Gly	*Asp*• Asx,	*Thr* Thr,	*Val*• Val,	*Gly* GLX,	*Met*• LYS)	Met/		S3
B3	*Gly* Gly ►	*Ala* Ala	*Glu*• Glu	*Leu* Leu	*Arg* Arg)	*His* (His	*Val* Val	*Leu* Leu	*Ala* Ala	*Thr* Thr	*Leu* Leu	———	———	*Gly* Gly	*Glu* Glu	*Lys* Lys)	*Met* LEU	*Lys* MET/							
A4	*Glu*	*Glu*	*Glu*	*Val*	*Glu*	*Ala* LYS	*Leu* Leu	*Met* Met/	*Ala* Ala ►	*Gly* Gly	———	———	———	*Gln*• Gln	*Glu* Glu	*Asp*• Asp	*Ser* Ser	*Asn*• Asn	*Gly* Gly	*Cys*• Cys	*Ile* Ile	*Asn*• Asn			S4
B4	*Tyr* Tyr	*Glu* Glu	*Ala*• Ala	*Phe* Phe	*Val* Val	*Lys* Lys	*His* His	*Ile* Ile	*Met* Met/	*Ser* ALA ►	*Ile* GLY														

Lengend for table 1

The sequence of the Alkali 1 light chain [8] is given in italics. With the exception of the N-terminal proline-rich region, the sequence has been drawn out to corresponds to the structural features of carp parvalbumin [2], with four structural units of the type A–S–B. The eight helix regions are indicated and the four sequences corresponding to the calcium sites shown. Residues underlined correspond to the positions of the buried apolar residues of carp parvalbumin, and the pattern of these hydrophobic residues has been emphasised previously [2]. Residues marked • correspond to the positions of the calcium ligating residues of the parvalbumin, and in the case of both troponin C and parvalbumin there are four acidic groups among the six marked residues in each site. Horizontal lines indicate that the sequence is continuous between the helices and sites. The sequences given below the italics are those of the tryptic peptides from the cardiac CNBr fragments. Capital letters mark amino acid replacements. The beginning of each CNBr fragment is indicated ► and brackets are used to define each tryptic peptide, the order based on the sequence homology. Within each bracket the sequences are complete except where commas have been used. The end of each CNBr fragment is indicated/. (N.B.: there is one change in the alkali light chain sequence as shown here from that previously published [8]. In site S3 residues 4 and 5 were originally reported to be Asp–Gly, though the cardiac T5 peptide is Gly–Asp. Further analysis of the chymotryptic peptide from the alkali light chain showed that residues S3(4) gave dansyl glycine much stronger than aspartic acid, and residue S3(5) gave dansyl aspartic acid stronger than glycine. The author favours this sequence as being correct.)

2. Experimental and results

Light chains were dissociated from beef cardiac myosin in 4M urea as described previously and the two components separated on DEAE-cellulose [7]. The larger cardiac 1 light chain was reacted with [^{14}C]-iodoacetic acid and the incorporation of radioactivity corresponded to 2.83 mol of carboxylmethyl cysteine/21 000, showing that all three thiol groups had reacted. Cyanogen bromide fragmentation was carried out as previously [8], the fragments being separated on Sephadex G-50 in 1 M sodium propionate. Fig.1 shows the elution profile. The three radioactive fragments are well separated, and five fractions were taken as indicated. After freeze drying, the fractions were desalted on a 23 × 2.8 cm column of Sephadex G-15. Elution was carried out with 5% pyridine for the radioactive fragments, where detection was by scintillation counting, and with 0.1 M ammonia for fractions C and E where detection was by absorbance.

Fractions B and C were further purified on cellulose phosphate under conditions described previously [8], though in each case a single peak was obtained. The smaller fragments in fractions D and E were purified by paper electrophoresis. Fraction D produced two radioactive bands at pH 6.5, but further electrophoresis at pH 2 showed that these peptides had identical mobilities, suggesting homoserine and homoserine lactone forms of the same sequence, a conclusion confirmed by amino acid analysis (table 2). Staining the papers with ninhydrin showed no other peptides present in this fraction. Fraction E yielded four ninhydrin staining spots, two of which were strong. E1, with a mobility of 0. 58 with respect to aspartic acid [9], was eluted directly, while E2, the neutral band, was submitted to further electrophoresis at pH 2, giving two new bands

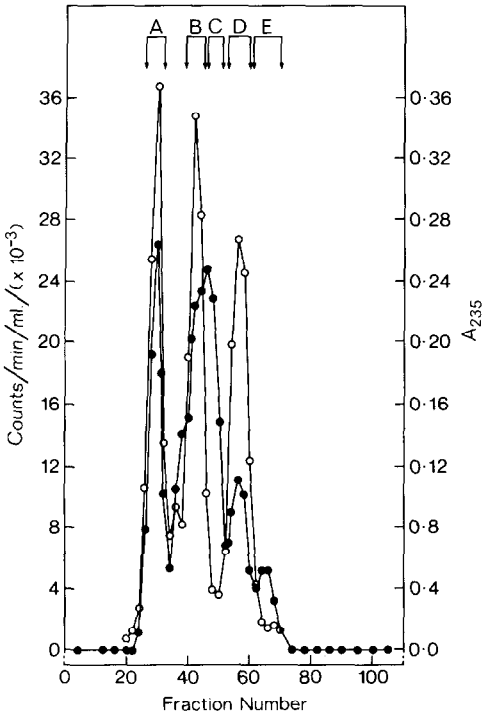

Fig.1. Fractionation of CNBr Fragments on Sephadex G-50. The CNBr fragments were applied to a column of 90 × 2.5 cm and eluted with 1.0 M sodium propionate, 50 mM Tris pH 9.5. The fraction size was 6 ml. Absorbance at 235 nm (•–•–•) and radioactivity (○–○–○). Five fractions were taken as indicated.

Table 2
Amino acid composition of cyanogen bromide fractions

Amino Acid	A	B	C	D	E 1	E2b	Sum	Cardiac Light Chain
Cys	0.9	1.0 (1)		0.8 (1)			3	3.0
Asp	3.3	3.2 (3)	5.9 (5)	3.2 (3)	2.1 (2)		16	17.7
Thr	2.3	2.1 (2)	4.4 (4)		1.0 (1)		9	10.8
Ser	1.7	0.9 (1)	1.0 (1)	1.0 (1)			5	5.1
Glu	8.5	6.7 (7)	9.3 (7)	3.3 (3)	0.4 (–)		26	27.6
Pro	10.0	1.7 (2)			0.9 (1)		13	17.3
Gly	1.4	4.0 (4)	5.5 (5)	2.1 (2)		1.1	13	13.7
Ala	8.2	2.3 (2)	2.0 (2)	2.0 (2)		1.0	15	18.0
Val	0.7	2.7 (3)	4.1 (4)	1.0 (1)			9	8.8
Met	(1)	(1)	(2)	(1)	(1)		6	6.8
Ile	2.0	0.9 (1)	1.0 (1)	1.9 (2)			6	7.0
Leu	1.3	5.0 (5)	6.5 (6)		1.0 (1)		13	13.7
Tyr		1.0 (1)	1.0 (1)	1.0 (1)			3	3.0
Phe	3.3	0.1	1.9 (2)	1.0 (1)	2.0 (2)		8	9.7
Lys	7.8	4.0 (4)	5.4 (5)	1.0 (1)	0.4 (–)		18	19.9
His		0.2	1.9 (2)	1.0 (1)			3	3.0
Arg	1.0	2.0 (2)	2.0 (2)				5	5.0
Total	53	39	49	20	8	2	171	190

The figures in parentheses are the compositions calculated from the tryptic peptides. The composition of the light chain was published previously [11].

Table 3
Amino acid sequences of tryptic peptides from fraction B

Peptide	Mobility	Sequence
T1	0.35	Glx–Glx–Glx–Leu–Asp–Ser–Lys
T2	0.23	Ile–Thr–Tyr–Gly–Gln–Cys–Gly–Asp–Val–Leu–Arg
T3	0.0	Ala–Leu–Gly–Gln–Asn–Pro–Thr–Gln–Ala–Glu–Val–Leu–Arg
T4	–0.68	Val–Leu–Gly–Lys–Pro–Lys
T5	0.0	Homoserine

The sequence of peptide T2 had been determined previously [11]. Peptide T3 was further degraded with chymotrypsin, which cleaved after both the leucine and theonine residues.

Table 4
Amino acid sequences of tryptic peptides from fraction C

Peptide	Mobility	Sequence
T1	–0.52	Leu–Gln–His–Ile–Ser–Lys
T2	–0.62	Asn–Lys
T3	0.55	Asp–Thr–Gly–Thr–Tyr–Glu–Asp–Phe–Val–Glu–Gly–Leu–Arg
T4	0.0	Val–Phe–Asp–Lys
T5	0.68	Glx–Gly–Asx–(Thr,Val,Glu)Lys
T6	0.0	Gly–Ala–Glu–Leu–Arg
T7	0.0	His–Val–Leu–Ala–Thr–Leu–Gly–Glu–Lys
T8	0.0	Leu–homoserine and free homoserine

E2a and E2b. E2a had the same mobility as serine at pH 2, and amino acid analysis indicated that this was free homoserine, a result confirmed by dansylation. Although this result suggested the presence of a Met–Met sequence in the protein, the recovery of this component was very much lower than E2b. The compositions of all the fractions are shown in table 2.

Digestions with trypsin and chymotrypsin were carried out as reported earlier [8], and sequences determined using the Edman procedure as described by Hartley [10], the N-terminal residues being identified by dansylation. Amide assignments were based on peptide mobilities at pH 6.5 [9]. The order of the peptides within the CNBr fragments is based on sequence homology with the alkali light chain, and the peptides have been numbered in tables 3 and 4 according to this order. Similarly the order of the CNBr fragments themselves is based on the extensive homology with the alkali light chain sequence.

Fragment A is rich in proline, alanine and lysine residues (table 2) indicating that it probably corresponds to the proline-rich sequence at the N-terminus of alkali 1, which is absent from alkali 2 [8]. The composition differs from that calculated from the alkali 1 sequence, showing that there must be sequence variation. In view of the considerable difficulties encountered in the determination of the proline-rich sequence [8], no attempt has been made to date to make detailed comparisons within this part of the molecule, which does not correspond to the structural elements discussed above (see table 1). However, the labelled thiol peptide produced by chymotryptic digestion of this fragment was isolated using gel filtration and paper electrophoresis. The composition of the purified peptide corresponded to that of Peptide C published previously [6,11], except for the presence of homoserine and its lactone in place of the methionine residue. Thus this peptide occurs at the C-terminal end of the proline-rich fragment, and its position in the sequence may be assigned approximately on the basis of the size of the fragment. It appears to coincide with the difference region in the sequence of alkali 1 and alkali 2 [8] which occurs at the beginning of helix A1 (table 1). Matching of the sequence of this peptide produces only very limited evidence of homology with the alkali light chain, but with the exception of the cysteine residue, the other substitutions occur at an average or above average frequency in homologous proteins [12]. The lysine residue at the C-terminus of this peptide (table 1) may be placed on the basis of the tryptic sequence published previously [11].

Fragment B gave integral values for many of the amino acids shown in its composition (table 2), evidence that it was reasonably homogeneous. Dansylation showed N-terminal lysine as the main component, but there was also a weak dansyl-glycine spot. Tryptic digestion produced four major peptides, but also free lysine and homoserine. Sequences were determined as shown in table 3. The sum of the residues in these four peptides, together with the free lysine and homoserine, equals the composition of the fragment as shown in table 2. Thus the free lysine may be located at the N-terminus of the fragment as anticipated from the dansylation, while the C-terminus is Lys–Met. This fragment spans the region from site S1 through helix B1 to the end of helix A2 (table 1).

Fraction C, which appeared to be larger than Fragment B, had the composition shown in table 2. Dansylation showed both glycine and leucine as N-terminal residues, suggesting that two fragments were present, although they had not separated on the cellulose phosphate column. No further fractionation was attempted, but tryptic peptides were isolated by paper electrophoresis and most of the sequences established (table 4). Peptide T5 was not completed, and its sequence is tentative, because the Edman degradation failed after the third residue. Furthermore, the composition did not give integral values in all cases (Lys(0.63), Asp(1.0), Thr(0.9), Glu(2.3), Gly(1.1), Val(1.0), Leu(0.6)). Lysine and homoserine lactone elute in the same position on the Locarte analyser, so it was possible that the lysine assignment could be wrong. However, the dansyl chromatograms showed strong ϵ-DNS lysine indicating that at least some lysine was present. Both leucine and glutamic acid were identified in the N-terminal position of T5, showing that it was impure. The peptide has been placed in site S3 (table 1), but this should be regarded as provisional.

The neutral peptides shown in table 4 were purified by further electrophoresis at pH 2 and pH 3.5. T8 was eluted after pH 2 electrophoresis, while T4, T6 and T7 required a further electrophoresis step. The composition of T8 indicated the presence of homoserine, leucine and lysine but not in integral yields. Dansylation showed only leucine and homoserine lactone, with no trace of ϵ-DNS lysine. Thus T8 appears to be a

mixture of two components. One is homoserine, confirmed by analysis of T8 without acid hydrolysis, but the other component did not coincide with free leucine on this analysis, and is probably a dipeptide Leu–homoserine, though this has not been confirmed.

With the exception of these uncertainties, the sequences of the remaining six tryptic peptides show a close resemblance to the alkali light chain sequence from helix B2 through to the end of helix B3 (table 1). As already stated, fraction C appears to contain two fragments, both because of its elution position on the Sephadex column and from the N-terminal results. The dansyl–leucine may be accounted for in peptide T1, while the dansyl–glycine arises from peptide T6. Ordering the peptides according to the sequence homologies places the incomplete T5 in site S3 where a methionine residue occurs in the alkali light chain. The presence of homoserine in T8 is consistent with a Lys–Met sequence here, but the evidence is tenuous. Suggestive evidence for leucine–homoserine also in T8 supports the existence of a second CNBr fragment, but further experiments will be needed to confirm this.

Fragment D contains the remaining thiol group whose sequence has been published previously [6,11]. Chymotryptic digestion produced the expected radioactive peptide and two small peptides, Glu–Ala–Phe, and Val–Lys–His–Ile–homoserine. Thus this fragment exactly matches site S4 and helix B4 (table 1).

Fragment E1 gave the sequence shown in table 1: Asp–Phe–Asp–Thr–Phe–Leu–Pro–homoserine, corresponding to the beginning of helix B2 (table 1). E2b was Ala–Gly, and the absence of homoserine in this peptide indicated that it might be the C-terminus of the protein. This was confirmed by the isolation of a histidine containing tryptic peptide whose sequence was His–Ile–Met–Ala–Gly.

When all the peptides in these fractions are ordered as shown in table 1, it is evident that there is extensive sequence homology between the two proteins. Furthermore the pattern of apolar residues within the helix regions is completely preserved in all the sequences shown, which suggests that the structural similarities hitherto noted [2] may be preserved in the cardiac light chain also. The sum of the compositions of the fractions falls somewhat short of the composition of the protein shown in table 2, which indicates that at least one small fragment is missing. This is also indicated by the presence of gaps in the sequence alignment in table 1. Although it is possible to speculate about possible overlaps between these CNBr fragments on the basis of existing evidence, further experiments are needed to establish the presence of missing sequences and determine the methionine overlaps.

References

[1] Collins, J. H. (1974) Biochem. Biophys. Res. Comm. 58, 301–308.
[2] Weeds, A. G. and McLachlan, A. D. (1974) Nature, 252, 646–649.
[3] Tufty, R. M. and Kretsinger, R. H. (1975) Science, 187, 167–169.
[4] Kretsinger, R. H. and Nockolds, C. E. (1973) J. Biol. Chem. 248, 3313–3326.
[5] Potter, J. D. and Gergely, J. (1974) Fed. Proc. 33, 1465.
[6] Weeds, A. G. and Pope, B. (1971) Nature, 234, 85–88.
[7] Weeds, A. G. and Lowey, S. (1971) J. Mol. Biol. 61, 701–725.
[8] Frank, G. and Weeds, A. G. (1974) Eur. J. Biochem. 44, 317–334.
[9] Offord, R. E. (1966) Nature, 211, 591–593.
[10] Hartley, B. S. (1970) Biochem. J. 119, 805–822.
[11] Weeds, A. G. and Frank, G. (1972) Cold Spring Harb. Symp. Quant. Biol. 37, 9–14.
[12] McLachlan, A. D. (1972) J. Mol. Biol. 64, 417–437.