Proc. Nat. Acad. Sci. USA
Vol. 69, No. 3, pp. 581–584, March 1972

Structure of a Calcium-Binding Carp Myogen

(muscle/x-ray analysis/amino-acid sequence/troponin)

CLIVE E. NOCKOLDS†*, ROBERT H. KRETSINGER†, CAROLE J. COFFEE‡, AND RALPH A. BRADSHAW‡

† Department of Biology, University of Virginia, Charlottesville, Va. 22903, and ‡ Department of Biological Chemistry, Washington University School of Medicine, St. Louis, Missouri 63110

Communicated by F. A. Cotton, December 15, 1971

ABSTRACT **The amino-acid sequence and three-dimensional structure of a calcium-binding protein prepared from carp muscle has been determined. This protein, designated carp-muscle calcium-binding protein B, is one of three closely related parvalbumins found in this tissue. The electron density map, calculated by heavy-atom substitution crystallographic methods to 2.0-Å resolution, reveals the orientation of most of the amino-acid side chains. The calcium coordination site consists of one glutamic- and three aspartic-acid carboxyl groups in a tetrahedral arrangement. The core of this spherical molecule is remarkably hydrophobic, with 8 of its 10 phenylalanine side chains packed in an approximate herringbone pattern. 52 of the 108 residues are in six α-helixes; there is no β-pleated sheet. The acetylated amino-terminal alanine appears not to be accessible to solvent. All of the heavy-atom derivatives are bound at the sole cysteine. The properties of this protein suggest a relationship to troponin A of mammalian tissue.**

Fish and amphibian white muscle contain several closely related proteins called parvalbumins, which account for 20–30% of the myogen fraction. They are defined by the following characteristics: (*a*) high solubility in water, (*b*) precipitation between 70- and 95%-saturated ammonium sulfate, (*c*) unusual amino-acid compositions, with 10% phenylalanine, 20% alanine, and only one or no tryptophan, tyrosine, arginine, histidine, cysteine, proline, or methionine residues, (*d*) isoelectric points between pH 4.0 and 4.5, and (*e*) molecular weights about 12,000 (1–3). Pechère (4) has shown that one of the hake parvalbumins binds calcium, as does the carp protein described here. In view of this property, this protein has been designated carp-muscle calcium-binding protein. These proteins apparently are not involved in glycolysis or osmotic regulation, nor do they serve any known nutritional role (2, 3). However, several features suggest that they are related to mammalian troponin A (4). This relation will be described below.

Preparation. The myogens were extracted from ground carp muscle with water and dialyzed against water, which precipitated all proteins except the albumins. Most of these albumins are precipitated when ammonium sulfate is added to 70% saturation, and the remaining proteins in solution were easily divided into a high- and a low-molecular-weight fraction by chromatography on Sephadex G-75. The low molecular weight myogens or parvalbumins were separated by DEAE-cellulose ion-exchange chromatography (5) into three fractions—A, B, and C—corresponding to components 5, 3, and 2 obtained by Konosu *et al.* (2) by moving boundary electrophoresis. The B component was used in the structural studies because of its preferable crystallographic properties, although preliminary sequence analysis of the A and C fractions indicates they are very similar to the B proteins (unpublished results). All preparative work was done in a cold room at 4°C.

Amino-Acid Sequence Determination. The general approach was to purify the 15 tryptic peptides by ion-exchange chromatography. The sequence of each peptide was determined by a combination of Edman degradation, aminopeptidase M digestion, and carboxypeptidase A and/or B digestion. The amino-terminal peptide, amino acids 1–19, was hydrolyzed into four peptides by thermolysin digestion, and subsequently sequenced as were the tryptic peptides. The amino-terminal *N*-acetylalanine was identified with a mass spectrometer. Similar results for hake parvalbumin have been obtained by nuclear magnetic resonance (NMR) analyses (Dr. J. F. Pechère, personal communication).

Ordering of Peptides. These peptides, without any overlap data, were sequenced at the same time as the first 2.0-Å resolution electron density map was calculated. With just these results, six of the sequenced peptides, containing 53 residues, could be placed with certainty. Five additional peptides, with 32 residues, were positioned with less confidence. Even though the remaining four peptides, containing 23 amino acids, could not be assigned unambiguously, it is clear that close coordination of crystallographic and sequence studies certainly facilitates the determination of protein structure.

After these analyses, Dr. Pechère kindly sent us the sequence for the homologous calcium-binding protein from hake (4). With this information, the unassigned sequences could be placed and the initial ordering of peptides was thus confirmed. Both proteins begin with *N*-acetylalanine, and are 108 residues long. As can be seen in Table 1, there are 29 differences between the hake and carp proteins, all of which are easily accommodated in the carp structure. There are two differences in amide assignments—carp: Asp-61, Gln-68; hake: Asn-71, Glu-68. Both positions are at the surface, and either the acid or amide groups could fit into the tertiary structure. We find five (17%) of the replacements occur at internal sites: Val-33 → Ile, Val-43 → Ile , Ala-46 → Val, Ile-58 → Val, and Leu-105 → Met.

Crystal Structure Determination. Our studies were done on crystals in 4.0 M phosphate (pH 7) space group C2 (a = 28.2 Å, b = 61.0 Å, c = 54.3 Å, β = 94°37′) (5). All three

* Present address: Electron Microscopy Unit, University of Sydney, Sydney, Australia.

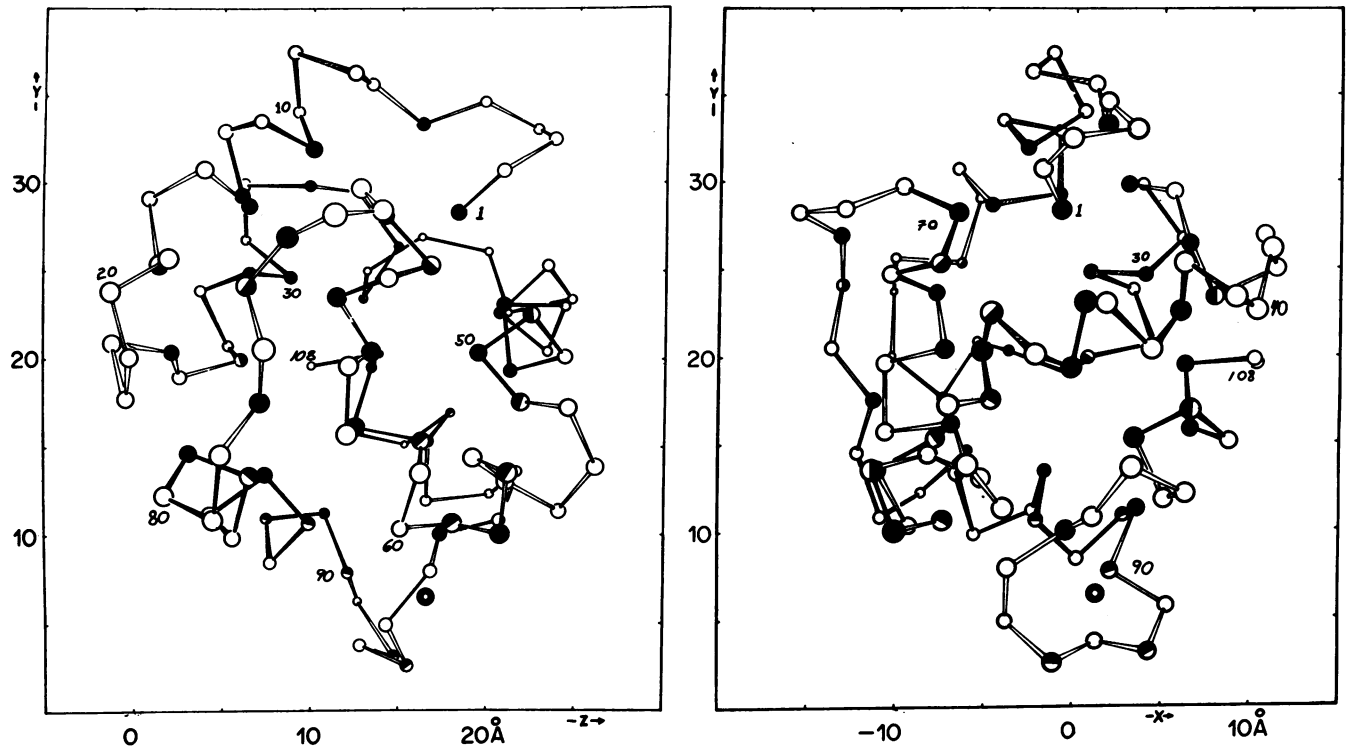

FIG. 1

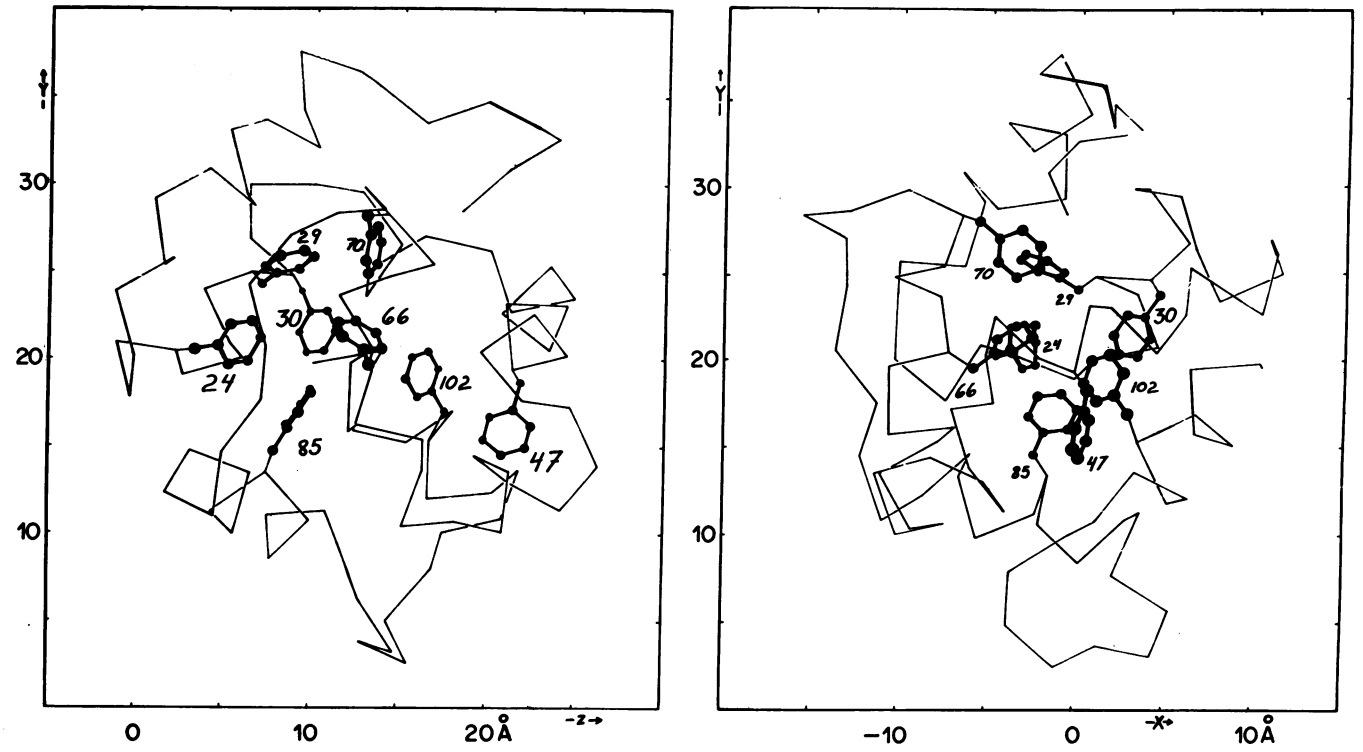

FIG. 2

heavy-atom derivatives (chloromercurymethoxypropyl urea, $HgBr_2$, and C_2H_2HgCl) are bound to the same major site near Cys-18. The phase ambiquity resulting from having only a single site was resolved by the use of anomalous dispersion data, and by finding a minor binding site for $HgBr_2$ and C_2H_2-HgCl on the other side of the cysteine from the major site. The quality of the electron density map is quite good, since about 90 of the carbonyl oxygen atoms can be identified as distinct knobs of density, and several of the phenylalanine rings show a slight dimple at their center.

The structure

The most striking feature of this protein is the extent to which it conforms to the generalization that hydrophobic groups are internal and hydrophilic groups are exposed to solvent. Of the 28 side chains judged to be internal, 26 residues—8 phenyl-

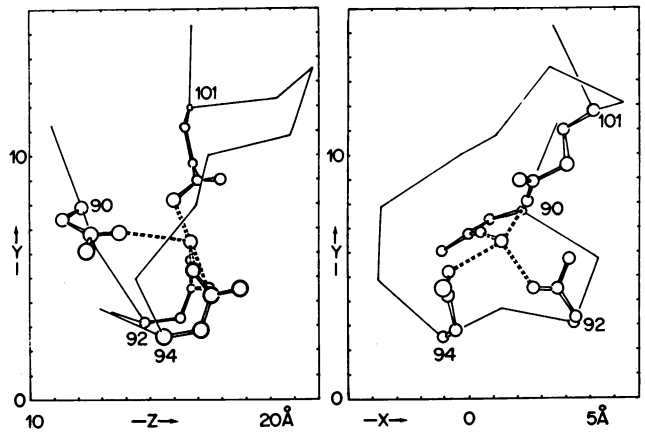

FIG. 3

FIGS. 1 and 2 (*facing page*) show the same two mutually perpendicular views of the α-carbon skeleton of carp calcium-binding protein, fraction B. The x axis corresponds to the crystallographic a axis; z axis to $c \cdot \sin \beta$. Their common origin is the crystallographic 2-fold axis, y, in drawing. In Fig. 1, *open circles* represent α-carbons whose side chains are exposed to solvent; *solid circles*, buried side chains; *half circles*, partially exposed. *Hatched bonds* represent α-helixes. In Fig. 2, the α-carbons are lightly traced to emphasize the core of eight phenylalanine side groups. Fig. 3 shows the coordination of the calcium ion by Asp-90, Asp-92, Asp-94, and Glu-101.

TABLE 1. *The amino-acid sequence of carp myogen*

Residue no.		Carp	Hake	Residue no.		Carp	Hake	Residue no.		Carp	Hake
1	I	Ala		37	S	Ser	Gly	73	S	Asp	Gly
2	S	Phe		38	S	Lys		74	I	Ala	
3	S	Ala		39	S	Ser		75		Arg	
4	S	Gly		40 H	S	Ala		76	S	Ala	
5	S	Val	Ile	41 H	S	Asp	Ala	77	I	Leu	
6	I	Leu		42 H	S	Asp		78 H	S	Thr	
7 H	S	Asn	Ala	43 H	I	Val	Ile	79 H	S	Asp	
8 H	S	Asp		44 H	S	Lys		80 H	S	Gly	Ala
9 H	S	Ala		45 H	S	Lys		81 H	I	Glu	
10 H	S	Asp		46 H	I	Ala	Val	82 H		Thr	
11 H	I	Ile		47 H	I	Phe		83 H	S	Lys	Ala
12 H	S	Ala	Thr	48 H	S	Ala	Gly	84 H	S	Thr	
13 H	S	Ala		49 H		Ile		85 H	I	Phe	
14 H	I	Ala		50 H	I	Ile		86 H		Leu	
15 H	I	Leu		51 H		Asp		87 H	S	Lys	
16	S	Glu	Ala	52	S	Gln		88 H		Ala	
17	S	Ala		53	S	Asp		89 H	I	Gly	
18		Cys		54	S	Lys		90		Asp	
19	S	Lys		55	S	Ser		91	S	Ser	
20	S	Ala		56	S	Gly	Asp	92		Asp	
21	S	Ala	Glu	57		Phe		93	S	Gly	
22	S	Asp	Gly	58	I	Ile	Val	94		Asp	
23	S	Ser		59	S	Glu		95	S	Gly	
24	I	Phe		60	S	Glu		96	S	Lys	
25	S	Asn	Lys	61	S	Asp	Asn	97	I	Ile	
26 H		His		62		Glu		98	S	Gly	
27 H	S	Lys	Gly	63	I	Leu		99	S	Val	
28 H	S	Ala	Glu	64	S	Lys		100	S	Asp	Glu
29 H	I	Phe		65	S	Leu		101	S	Glu	
30 H	I	Phe		66	I	Phe		102 H	I	Phe	
31 H	S	Ala	Thr	67 H	I	Leu		103 H		Thr	Ala
32 H	S	Lys		68 H	S	Gln	Glu	104 H	S	Ala	
33 H	I	Val		69 H		Asn		105 H	I	Leu	Met
34	S	Gly		70 H	I	Phe		106 H	I	Val	
35	I	Leu		71 H	S	Lys	Ser	107 H	S	Lys	
36		Thr	Lys	72	S	Ala		108	S	Ala	Gly

The sequence of hake parvalbumin (4) is listed at the nonidentical positions. Both proteins contain 108 residues and an N-terminal acetyl group. In the second column, *H* indicates that the residue is in an α-helix. In the third column, *I* means the side chain is internal, inaccessible to solvent; *S* means the side chain is at the surface; a blank means partially buried. The designations in columns 2 and 3 refer only to the carp protein.

alanine, 7 leucine, 4 isoleucine, 3 valine, and 4 alanine are very nonpolar. There is one internal glycine. One glutamic acid, position 81, is internal and forms a hydrogen bond to Arg-75, thus compensating the two charges. The cluster of eight phenyl rings forming the core of the molecule (Fig. 2) is in a general herringbone pattern characteristic of crystals of many benzene derivatives (6). There is no π–π stacking. Even among those residues that have been tentatively designated as partially exposed to solvent, the majority that have a significant dipole moment place this dipole toward the surface. The two exceptions are Thr-82, where no potential hydrogen bond is seen, and Arg-75, which forms an internal hydrogen bond to Glu-81. The carboxyl group of Glu-62 is partially buried, but this is one of the three short regions where the interpretation is still uncertain. Even the main chain (see Fig. 1) appears to keep its carbonyl and amide groups near the surface, except for the α-helical regions, where of course half of the chain is obliged to be on the inside.

The calcium ion is coordinated by the four carboxyl groups of aspartate residues 90, 92, and 94 and glutamic acid 101 (Fig. 3), at an average calcium–oxygen distance of 2.5 A. In a calcium–glycylglycylglycine complex (7), the calcium is liganded by seven oxygen atoms with distances ranging from 2.30 to 2.50 Å. In *Staphylococcal* nuclease, calcium is coordinated by two aspartic- and one glutamic-acid residues (8). Even with an estimated error of 0.4 Å in the oxygen–calcium distance, Asp-90 is further away from the calcium than are the other three. The tetrahedral coordination is distorted in such a way that with only 45° rotation of the α-β bonds of aspartic acids 92 and 94, the calcium could easily dissociate in the direction distal to Asp-90. With the calcium not present, the carboxyl groups would all be at the surface of the protein. The function of the protein must ultimately be explained in terms of the structures with and without calcium. Determination of the calcium-free structure is in progress.

In contrast to the situation found in other proteins of known structure, the binding site of the carp myogen consists entirely of residues 90, 92, 94, and 101, which are grouped together in the sequence. The sequence in this region, beginning with Gly-89, is Gly-Asp-Ser-Asp-Gly-Asp-Gly. This segment contains only those three residues that show the highest probability, all over 50% of total occurrence, to exist in β-bends (9). However, the only configuration resembling a β-bend involves the α-carbon of Asp-90, being about 4.5 Å from the α-carbon of Gly-93. Yet, a hydrogen bond, characteristic of the β-bend, could not be formed without rotation of the carbonyl group of

Asp-90 by 90°. It is certainly possible that the removal of calcium causes a localized structural change in which one or two hydrogen-bonded β-bends are formed.

52 of the 108 residues are in six α-helixes (Fig. 1 and Table 1). Each of these shows some distortion from the canonical α-helix, in that all six have one or several nonlinear hydrogen bonds. Helices A, C, E, and F would be considered to be one residue longer if either a carbonyl or amide group were rotated some 45° to bring them into correct hydrogen-bonding position.

In only three regions are there significant uncertainties in the interpretation of the electron density map. The amino-terminal *N*-acetylalanine is tucked back into the molecule, while the Phe-2 side chain is at the surface. Their positions might be reversed, thus adding another benzene ring to the well-defined group of eight. The loop from Ser-55 to Glu-62 may be interpreted in several ways.† The present model is certainly not correct in this region in detail, although the alternative interpretations do not significantly effect the present description. Finally, the region from Gly-34 to Lys-38 contains a large peak of electron density that is still unexplained.

In examination of this structure, no grooves or pits of the type characterizing the active sites of the various enzymes whose structures have been solved were found. This lack of a definable catalytic site supports the idea that this protein plays some role in regulating muscle contraction, and is perhaps analogous to troponin A or calcium-sensitizing factor of mammalian muscle. Troponin A avidly binds calcium (10–13), which presumably is the basis of its mechanism of action in releasing the inhibition of the myosin ATPase by the tropomyosin–troponin B complex (10). Troponin A, like calcium-binding protein, is acidic and has a high phenylalanine content, with no tryptophan, as judged by spectral and amino-acid analyses (11–13). It has a molecular weight between 17,000 (13) and 22,000 (11), about twice that of the carp myogen, which has a molecular weight, from sequence data, of 11,489. It is clear that considerable further studies will be necessary to establish whether calcium-binding protein from fish and mammalian troponin A share any common functional and/or structural features. If such should prove to be the case, the knowledge of the complete structure of carp-muscle calcium-binding protein should provide an excellent basis for clarifying the nature of the regulatory processes of muscle contraction.

† A preliminary examination of the electron density map, calculated from tangent formula refinement phases, has resolved the ambiguity in the interpretation of the region about Glu-62. Carp myogen contains a second cation coordinated by the four carboxyl groups of asparate residues 51 and 53 and of glutamate residues 59 and 62.

A preliminary description of this structure was presented at the 1971 Cold Spring Harbor Symposium, and will be included in Volume XXXVI. At that time, the *N*-acetylalanine had not been identified. Hence, the numbering of residues in the sequence has been increased by one. A detailed description of the structure and experimental techniques will be published when the electron density map, calculated from tangent formula refinement phases, is fully interpreted. Accurate α-carbon coordinates and approximate complete coordinates are available from R.H.K.

We thank Dr. J. F. Pechère for sending us the sequence data for hake parvalbumin and for his determination of the *N*-acetylated alanine. The research has been supported by NIH grants GM-15786, AM-13362, and NSF-GB29306. R.A.B. is a Research Career Development Awardee of the NIH, AM-23968. C.J.C. is a U.S. Public Health Service postdoctoral fellow, AM-37344.

1. Hamoir, G. & Konosu, S. (1965) "Carp myogens of white and red muscles: General composition and isolation of low-molecular-weight components of abnormal amino acid composition", *Biochem. J.* **96**, 85–97.
2. Konosu, S., Hamoir, G. & Pechère, J. -F. (1965) "Properties and amino acid composition of the main-low-molecular-weight components of white muscle", *Biochem. J.* **96**, 98–112.
3. Pechère, J. -F. & Focant, B. (1965) "Gross isolation on sephadex columns of the low-molecular-weight components and examination of their participation in anaerobic glycogenolysis", *Biochem. J.* **96**, 113–118.
4. Pechère, J. -F., Capony, J. -P. & Ryden, L., "The primary structure of the major parvalbumin from hake muscle: Isolation and general properties of the protein", *Eur. J. Biochem.*, in press.
5. Kretsinger, R. H., Dangelat, D. & Bryan, R. F. (1971) "Crystal data for low molecular weight albumins of carp", *J. Mol. Biol.* **59**, 213–214.
6. Wyckoff, R. W. G. (1969) in "Crystal Structures", *The Structure of Benzene Derivatives* (Interscience Publishers, New York), Vol. 6.
7. Van Der Helm, D. & Willoughby, T. V. (1969) "The crystal structure of $CaCl_2 \cdot$ glycylglycylglycine $\cdot 3H_2O$", *Acta. Cryst.* **B25**, 2317–2326.
8. Arnone, A., Bier, C. J., Cotton, F. A., Day, V. W., Hazen, E. E., Jr., Richardson, D. C. & Yonath, A. (1971) "A high resolution structure of an inhibitor complex of the extracellular nuclease of *Staphylococcus aureas*", *J. Biol. Chem.* **246**, 2302–2316.
9. Lewis, P. N., Momany, F. A. & Scheraga, H. A. (1971) "Folding of polypeptide chains in proteins: A proposed mechanism for folding", *Proc. Nat. Acad. Sci. USA* **68**, 2293–2297.
10. Ebashi, S., Endo, M. & Ohtsuki, I. (1969) "Control of muscular contraction", *Quant. Rev. Biophys.* **2**, 351–384.
11. Greaser, M. L. & Gergely, J. (1971) "Reconstitution of troponin activity from three protein components", *J. Biol. Chem.* **246**, 4226–4233.
12. Hartshorne, D. J., Theiner, M. & Mueller, H. (1969) "Studies on troponin", *Biochim. Biophys. Acta* **175**, 320–330.
13. Ebashi, S., Wakabayashi, T. & Ebashi, F. (1971) "Troponin and its components", *J. Biochem.* **69**, 441–445.