

Amino acid sequence of rabbit fast-twitch skeletal muscle calsequestrin deduced from cDNA and peptide sequencing

(calsequestrin/sarcoplasmic reticulum/cDNA cloning)

LARRY FLIEGEL*, MAMORU OHNISHI†, MICHAEL R. CARPENTER†, VIJAY K. KHANNA*, REINHART A. F. REITHMEIER†, AND DAVID H. MACLENNAN*‡

*Banting and Best Department of Medical Research, University of Toronto, Charles H. Best Institute, 112 College Street, Toronto, ON, Canada M5G 1L6; and †Department of Biochemistry, University of Alberta, Edmonton, AB, Canada T6G 2E1

Communicated by R. H. Wasserman, October 27, 1986

ABSTRACT Partial amino acid sequence analysis of rabbit fast-twitch skeletal muscle calsequestrin permitted the construction of synthetic oligonucleotides that were used as both primers and probes for the synthesis and isolation of cDNAs encoding calsequestrin from neonatal rabbit skeletal muscle libraries. The cDNA sequence encodes a processed protein of 367 residues with a M_r of 42,435 and a 28-residue amino-terminal signal sequence. The deduced amino acid sequence agreed closely with the portions of the mature protein that were sequenced using standard protein sequencing. The neonatal protein, however, contains an acidic carboxyl-terminal extension not present in the adult protein, suggesting that the cDNA sequence may have arisen from an alternatively spliced neonatal transcript. A single transcript of 1.9–2.0 kilobases was seen in neonatal skeletal muscle mRNA. A glycosylation site and two potential phosphorylation sites were detected. Although the protein contains about two acidic residues for each Ca^{2+} bound, there is no repeating distribution of acidic residues and no evidence of EF hand structures. Hydropathy plots show no transmembrane sequences, and structural analyses suggest that less than half of the protein is likely to be highly structured. This sequence defines the characteristics of a class of high-capacity, moderate-affinity, Ca^{2+} binding proteins.

Calsequestrin is a high-capacity, moderate-affinity, Ca^{2+} binding protein (1) localized in luminal spaces of the terminal cisternae of the sarcoplasmic reticulum of muscle cells (1–4). The acidic protein binds about 1 μmol of Ca^{2+} per mg with a K_d of ≈ 1 mM, resulting in dramatic changes in its conformation (1, 5–9). These properties suggest that calsequestrin acts as a Ca^{2+} buffer inside the sarcoplasmic reticulum, lowering free Ca^{2+} concentrations and thereby lowering the gradient against which the Ca^{2+} -ATPase must pump Ca^{2+} (4). Studies of Ca^{2+} concentrations in muscle using x-ray microprobe analysis (10) have demonstrated that Ca^{2+} , presumably as a Ca^{2+} -calsequestrin complex, is localized in the terminal cisternae during relaxation and is lost from that region during tetanus. Thus, a second function of calsequestrin may be to localize Ca^{2+} near the junctional face of the terminal cisternae, the point from which it is released to initiate muscle contraction (4).

We have obtained partial amino acid sequences of the protein (4, 11), and we have used these sequences to construct oligonucleotides to clone cDNA-encoding calsequestrin. In this communication, we report the amino acid sequence of rabbit fast-twitch skeletal muscle calsequestrin deduced from the sequence of complementary DNA. We also demonstrate that this sequence corresponds to our partial amino acid sequences, except for the carboxyl-terminal region.

MATERIALS AND METHODS

Amino Acid Sequence Analysis. Calsequestrin was isolated from adult rabbit skeletal muscle by the method of MacLennan and Wong (1) and cleaved with cyanogen bromide, trypsin, or chymotrypsin using standard techniques (12). Peptides were purified by a combination of gel filtration, ion-exchange, and high-pressure liquid chromatography. Amino acid compositions were determined using a Durrum (Palo Alto, CA) model D-500 amino acid analyzer after hydrolysis of the samples in 6 M HCl at 110°C for 24, 48, and 72 hr. Automated sequence analyses were done on an Applied Biosystems (Foster City, CA) model 470A protein sequencer. Phenylthiohydantoin-amino acid derivatives were analyzed by on-line high-pressure liquid chromatography. Between 50 pmol and 3 nmol of protein or peptide was subjected to sequence analysis.

Construction, Screening, and Sequencing of cDNA Clones. mRNA was isolated from 14-day-old rabbit back and leg muscles as described previously (13). cDNA was synthesized by a primer extension method (14) using a group of synthetic oligonucleotides 5'-TC^AGT^ACGTC^AGT^ACGTC^AGC-3' based on the carboxyl-terminal sequence determined for adult calsequestrin, -Asp-Asp-Asp-Asp-Asp-Glu-COO⁻ (11). Second-strand synthesis, tailing with deoxycytidine, hybridizing to (dG)-tailed pBR322, and transformation of *Escherichia coli* strain RR1 (Bethesda Research Laboratories) were done as described previously (13, 15). The resulting library of 3000 colonies was screened as described earlier (13) with a group of synthetic oligonucleotides of base composition 5'-TC^AGTCCAT^CTCCATCCA-3' based on the upstream sequence -Trp-Met-Glu-Met-Asp-Asp-. This screening yielded 18 calsequestrin clones, the longest of which was 1.1 kilobase (kb); all lacked the 3' untranslated region.

A restriction endonuclease fragment from the 5' end of the longest clone was used to probe the neonatal rabbit skeletal muscle cDNA library described previously (13). Two clones of 1.8 and 1.6 kb were isolated and mapped for restriction endonuclease sites. Fragments were cloned into vectors M13, mp18, or mp19, or into the pTZ vector (Pharmacia) to obtain single-stranded cDNAs oriented in both directions (Fig. 1). Single-stranded cDNA was sequenced by the dideoxy chain-termination method of Sanger *et al.* (16). RNA blot analysis was done as described previously (17).

Structural Predictions. The amino acid sequence was analyzed for hydropathy (18), helical structure, β -sheet structure, and β -turns using computer analyses (19–21). Not all predicted structures derived from these analyses are shown.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡To whom requests for reprints should be addressed.

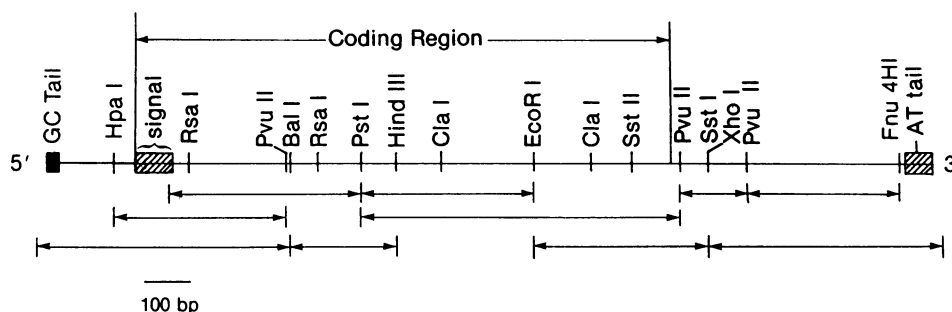


FIG. 1. Sequencing strategy for cloned cDNA encoding calsequestrin of neonatal rabbit skeletal muscle sarcoplasmic reticulum. The restriction map displays five- and six-base restriction endonuclease sites used in sequencing.

RESULTS

The longest cDNA clone encoding calsequestrin contained 1848 base pairs (bp) of which 1185 were in coding sequences, 493 were in the 3' untranslated region, and 170 were in the 5' untranslated region. Calsequestrin is synthesized with a cleavable amino-terminal signal sequence with methionine residues at positions 1 and 7 (22). Thus the location of the initiator methionine was readily identified 27 residues before the glutamic residue at the beginning of the mature sequence. The presence of a glycine residue at the site of cleavage of the signal peptide is consistent with the properties of the signal peptidase that cleaves after small aliphatic residues. The 3' untranslated sequence ended in a poly(dA)⁺ tail 15-bp downstream from the end of the polyadenylation signal A-A-T-A-A-A. The intervening bases contained the sequence T-G-T-G-G-G-G-C that composes part of the polyadenylation signal.

The deduced amino acid sequence of neonatal calsequestrin and the partial sequences obtained by direct amino acid determination of the mature protein are also presented in Fig. 2. The mature protein begins with the sequence ³H₃N-Glu-Glu- and ends with a long acidic sequence. There was excellent agreement between the two sequences with the exception of the carboxyl-terminus. The adult amino acid sequence ends with Glu-(Asp)₅-Glu-COO⁻; the neonatal deduced sequence in the same region is -Glu-(Asp)₄-Glu-(Asp)₈-COO⁻. [The oligonucleotide used as a primer was based on the sequence -(Asp)₅-Glu-, but because only the first two residues of the codon for glutamate were used, the primer would recognize codons for (Asp)₆ as well. In fact, reverse transcription was initiated in the (Asp)₈ region, and the cloned sequence encoded -Glu-(Asp)₄-Glu-(Asp)₈-COO⁻.] The adult protein would be predicted to be comprised of 360 residues with a *M_r* of 41,630 and to contain 44 aspartate and 59 glutamate residues. The neonatal protein, defined by cDNA sequencing, has 367 residues with a *M_r* of 42,435 and contains 51 aspartate and 59 glutamate residues. Both proteins are predicted to contain 24 lysine and 6 arginine residues.

Mature rabbit calsequestrin exists in two forms that can be distinguished by NaDodSO₄/PAGE (23). Most rabbits contain only form 1, whereas some contain forms 1 and 2 in equal concentrations, and a few contain only form 2. All amino acid sequencing was done with form 1 of calsequestrin, but amino acid compositions were obtained for both forms (data not shown). Because form 2 does not contain significantly more aspartate residues than does form 1, the cDNA sequence that we obtained from neonatal rabbit muscle mRNA does not correspond to form 2 of calsequestrin.

In 20 locations two or more acidic residues are juxtaposed. This alignment can be accounted for by random association in the first 325 residues, but 28 of the last 42 residues are acidic, suggesting that they have a functional role. The protein did not contain any sequences that could be con-

strued as forming EF hand types of Ca²⁺ binding sites (24). Ten of the 30 basic residues are paired, and the sequence -Val-Ala-Lys-Lys-Leu-X-Leu- occurs twice. This pairing of basic residues is more than random chance and may also suggest that they have a functional role. Residues 117-128 have a similar sequence to residues 317-328. This is the only other sequence that appears to repeat within the protein.

The five tryptophan residues are located in the carboxyl-terminal third of the molecule. The protein does not contain cysteine residues, consistent with recent amino acid analyses of the protein (5, 25). The protein is relatively rich in proline residues in the carboxyl-terminal half of the protein.

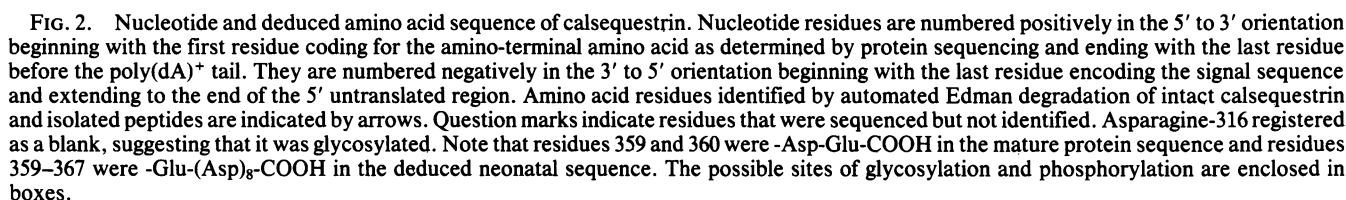
Calsequestrin is a glycoprotein containing an extensively processed carbohydrate component comprised of two moles of *N*-acetylglucosamine and three moles of mannose per mole of protein of mass 40 kDa (26). The only potential glycosylation site, -Asn-Val-Thr-, found at residues 316-318 is almost certainly used, because a blank in the amino acid sequence was registered where a phenylthiohydantoin-asparagine should have been recorded. Whether calsequestrin is phosphorylated *in vivo* is not clear, but it appears that calsequestrin can be phosphorylated *in vitro* (4). Potential phosphorylation sites were noted at residues 189 and 229.

Structural predictions (18-21) for calsequestrin are provided in Fig. 3. The hydropathy plot emphasizes its hydrophilic character and confirms that there are no membrane-spanning regions in this protein. About six regions centering around residues 50-70, 76-88, 124-151, 156-171, 215-230, and 320-350 regions are strongly predicted to be α -helices, whereas another five regions centering around residues 28-36, 41-49, 178-205, 238-258, and 262-273 predict more weakly as α -helices. Three regions centering around residues 13-18, 282-287, and 312-317 predict strongly as β -strands by all predictive methods. The Chou-Fasman (19) predictions (Fig. 3) for β -strands are more extensive, however, than those predicted by the method of Garnier *et al.* (20). While β -turns are predicted throughout, the carboxyl-terminal third of the molecule is exceptionally rich in sequences predicted to form β -turns.

Because cardiac calsequestrin appears to differ from skeletal muscle calsequestrin (27), and because the slow-twitch/cardiac as well as the fast-twitch Ca²⁺-ATPase are expressed in neonatal rabbit skeletal muscle, we looked for cross-hybridizing calsequestrin transcripts in neonatal muscle. An RNA blot, presented in Fig. 4, demonstrates that only one calsequestrin transcript of 1.9-2.0 kb hybridized strongly with a probe from the coding sequence of fast-twitch skeletal muscle calsequestrin. The size of the mRNA suggests that a full-length cDNA clone might be no more than 100-200 bp longer than the calsequestrin cDNA reported here.

DISCUSSION

We have demonstrated that calsequestrin has a *M_r* of $\approx 42,000$ (+ ≈ 1000 due to carbohydrate). This value is consistent with



The amino acid sequence confirms the acidic nature of the protein. Amidated derivatives account for only 15% of the total Asx or Glx residues as predicted earlier (1). The net charge of adult calsequestrin at neutral pH would be -75 and the net charge of the neonatal form would be -82. There is no evidence of a repeating distribution of negative charges such as might be found in an EF hand or other regular structure. Forty-two of the 110 acidic residues in the neonatal form and 42 of the 103 acidic residues in the mature protein are in pairs or triplets, whereas two carboxyl-terminal sequences contain groups of 5 and 14 (7 groups in the adult) acidic residues, respectively. These residues are, no doubt, involved in Ca^{2+} binding, but the exact structure of Ca^{2+}

Calsequestrin contains a single high-affinity binding site for the phenothiazine antipsychotic drug trifluoperazine, and this binding site is contained in a CNBr fragment encompassing residues 203–237 (R.A.F.R., unpublished observations). Trifluoperazine also binds to calmodulin and troponin C (31), and because activation of phosphodiesterase by calmodulin is blocked by phenothiazine antipsychotics, the drug-binding sites may be sites of interaction between

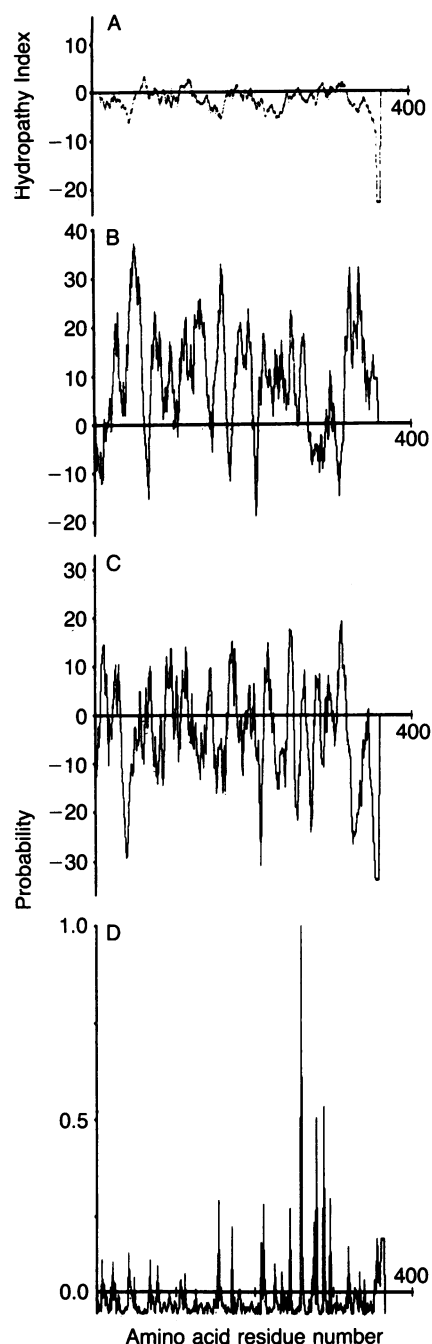


FIG. 3. (A) The hydropathic character of residues 1–367 of the processed form of neonatal calsequestrin was evaluated using the procedure of Kyte and Doolittle (18) with a span of 19 residues. (B) The probability of α -helical structure was determined by the method of Levitt (21) using a span of nine residues. (C) The probability of β -sheet structure was determined by the method of Levitt (21) using a span of nine residues. (D) The probability of β -turns was evaluated using the method of Chou and Fasman (19). The solid line corresponds to a cut-off value of $P = 7.5 \times 10^{-5}$ for analysis of four residues. The highest probability of a β -turn was assigned a value of 1.0 and was equal to 1.20×10^{-3} .

calmodulin and regulated proteins. The sites of drug interaction contain the following sequences:

K D T D S E E E I R E A F R V F D K⁹⁴ (bovine brain calmodulin),
 A K G K S E E E L A E C F R I F D R¹⁰⁴ (rabbit skeletal muscle troponin C),
 D K P N S E E E I V N F V E E H R²²⁶ (rabbit skeletal muscle calsequestrin).

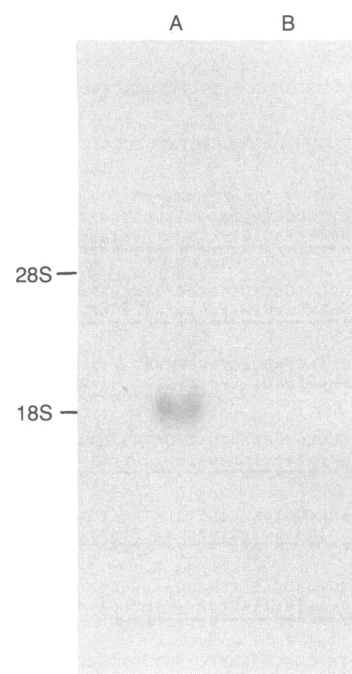


FIG. 4. RNA blot analysis of neonatal rabbit skeletal muscle mRNA. Residues 1–1094 of the coding region of calsequestrin cDNA were used to probe polyadenylated RNA isolated from day-old skeletal muscle. Positions of 28S and 18S ribosomal subunits are indicated. Lane A, 10 μ g of poly(A)⁺ RNA; lane B, 10 μ g of eluate from the oligo(dT)–cellulose column. The size of mRNA coding for calsequestrin was estimated at between 1.9 and 2.0 kb.

These binding sequences all contain an invariant core sequence -Ser-(Glu)₃- followed by a hydrophobic residue (isoleucine or leucine) and are flanked upstream by a basic residue and downstream by a phenylalanine residue and another basic residue. Thus calsequestrin has evolved or retained a binding site for hydrophobic organic compounds shared by ancient high-affinity Ca²⁺ binding proteins; the most likely explanation for this sequence is that it is functional and involved in specific protein–protein interactions.

In this respect, the pairing of basic residues within a hydrophobic setting in the calsequestrin sequence attracts interest. These sequences are similar to sequences containing paired basic residues in calmodulin receptor proteins that have been implicated in the binding sites between calmodulin and these receptors (32). While we do not wish to imply that calsequestrin is an activator molecule or that it is, itself, activated by calmodulin, we point out that these sites of potential interaction between the trifluoperazine binding site and the dibasic sites may be important either in intramolecular bonding to form three-dimensional structures or in intermolecular bonding. Calsequestrin forms aggregates in the presence of excess Ca²⁺ (1, 5), and it forms a matrix within the terminal cisternae (2). These associations might be mediated through Ca²⁺-promoted interactions similar to those occurring between calmodulin and its receptors.

Structural predictions clearly show that calsequestrin does not contain transmembrane sequences. Thus, whereas calsequestrin is labeled by lactoperoxidase in intact vesicles (33), there is no chance that this labeling occurs because the protein contains a transmembrane segment. It might, however, be anchored to the junctional face of the terminal cisternae through an interaction with junction proteins, an interpretation deducible from morphological studies (34).

Although extensive portions of calsequestrin are predicted to be helical in conformation (Fig. 3), circular dichroism studies of calsequestrin have shown that the apoprotein contains only 10–20% α -helix (5, 8, 9, 35). Calsequestrin is a

very acidic protein, with many of the acidic residues clustered. Thus, whereas glutamate is a strong helix former, closely positioned negative charges will repel, creating a random coil conformation. Calsequestrin also shows a dramatic increase in helical content at low pH or when binding Ca^{2+} (5, 8, 9, 35). Circular dichroism studies of calsequestrin (M.O. and R.A.F.R., unpublished data) have indicated that this protein contains substantial β -sheet structure in the presence and absence of Ca^{2+} . β -Sheet structure is predicted to occur in at least three regions of the protein (Fig. 3). A high probability of β -turns in the carboxyl-terminal half of calsequestrin is predicted. This may indicate that this portion of calsequestrin is highly folded.

Tryptophan groups have been used as reporters of extensive structural change in calsequestrin, induced by Ca^{2+} binding and probably leading to helix formation (5, 8, 35). All five tryptophan residues in the molecule are located between residues 242 and 342, a region that has a high potential for β -turns. These residues are likely to be involved in conformational changes as their environment is altered upon Ca^{2+} binding.

In our studies of Ca^{2+} -ATPase expression (13, 17) in neonatal muscle, we found two transcripts, a fast-twitch form and a slow-twitch/cardiac form. We expected to find similar results for calsequestrin, but instead, found only a single strongly hybridizing band (Fig. 4). There are three possible explanations: (i) there is only one calsequestrin transcript in neonatal muscle; (ii) there are two or more transcripts, but they have an identical mobility; (iii) there are two or more transcripts, but they do not cross-hybridize, or they cross-hybridize very weakly. We suspect the latter possibility, because the amino-terminal amino acid sequences of fast-twitch and cardiac calsequestrins are quite different (M.O. and R.A.F.R., unpublished data) and because we have not detected cross-hybridization of our cDNA with any clones in a cardiac cDNA library or in RNA blots of cardiac mRNA.

We predict that similarities will emerge in the molecular genetics of the Ca^{2+} -ATPase and calsequestrin. We are convinced that there will be at least two calsequestrin genes encoding fast-twitch and cardiac (possibly slow-twitch/cardiac) (27) forms of the protein. Moreover, our observations on different, developmentally regulated carboxyl-termini in the fast-twitch Ca^{2+} -ATPase transcripts (30) will almost certainly be repeated in the fast-twitch form of calsequestrin, where we have seen a discrepancy between adult and neonatal carboxyl-terminal sequences.

We are grateful to Dr. C. J. Brandl for advice and assistance in recombinant DNA technology and for providing one of the cDNA libraries used in this study, and to Dr. H. Okayama for his assistance in preparation of this cDNA library. The advice of Dr. E. Holowachuk and the assistance of S. de Leon are also gratefully acknowledged. L.F. is a Postdoctoral Fellow of the Heart and Stroke Foundation of Ontario; M.O. is a Postdoctoral Fellow of the Alberta Heritage Foundation for Medical Research; and R.A.F.R. is a Scholar of the Alberta Heritage Foundation for Medical Research. This work was supported by grants to D.H.M. from the Medical Research Council of Canada, The Muscular Dystrophy Association

of Canada, and the Heart and Stroke Foundation of Ontario, and to R.A.F.R. from the Muscular Dystrophy Association of Canada.

- MacLennan, D. H. & Wong, P. T. S. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 1231-1235.
- Meissner, G. (1975) *Biochim. Biophys. Acta* **389**, 51-68.
- Jorgensen, A. O., Shen, A. C.-Y., Campbell, K. P. & MacLennan, D. H. (1983) *J. Cell Biol.* **97**, 1573-1581.
- MacLennan, D. H., Campbell, K. P. & Reithmeier, R. A. F. (1983) in *Calcium and Cell Function*, ed. Cheung, W. Y. (Academic, New York), Vol. 4, pp. 151-173.
- Ikemoto, N., Nagy, B., Bhatnagar, G. M. & Gergely, J. (1974) *J. Biol. Chem.* **249**, 2357-2365.
- Meissner, G., Conner, G. E. & Fleischer, S. (1973) *Biochim. Biophys. Acta* **298**, 246-269.
- Cozens, B. & Reithmeier, R. A. F. (1984) *J. Biol. Chem.* **259**, 6248-6252.
- Ostwald, T. J., MacLennan, D. H. & Dorrington, K. J. (1974) *J. Biol. Chem.* **249**, 5867-5871.
- Aaron, B.-M. B., Oikawa, K., Reithmeier, R. A. F. & Sykes, B. D. (1984) *J. Biol. Chem.* **259**, 11876-11881.
- Somlyo, A. V., Gonzalez-Serratos, H., Shuman, H., McClellan, G. & Somlyo, A. P. (1981) *J. Cell. Biol.* **90**, 577-594.
- Ohnishi, M. & Reithmeier, R. A. F. (1985) *Biophys. J.* **47**, 284a (abstr.).
- Gross, E. (1967) *Methods Enzymol.* **11**, 238-255.
- MacLennan, D. H., Brandl, C. J., Korczak, B. & Green, N. M. (1985) *Nature (London)* **316**, 696-700.
- Nathans, J. & Hogness, D. S. (1983) *Cell* **34**, 807-814.
- Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263-269.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
- Brandl, C. J., Green, N. M., Korczak, B. & MacLennan, D. H. (1986) *Cell* **44**, 597-607.
- Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105-132.
- Chou, P. Y. & Fasman, G. D. (1979) *Biophys. J.* **26**, 367-384.
- Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97-120.
- Levitt, M. (1978) *Biochemistry* **17**, 4277-4285.
- Reithmeier, R. A. F., de Leon, S. & MacLennan, D. H. (1980) *J. Biol. Chem.* **255**, 11839-11846.
- MacLennan, D. H. (1974) *J. Biol. Chem.* **249**, 980-984.
- Kretsinger, R. H. (1976) *Annu. Rev. Biochem.* **45**, 239-266.
- Caudwell, B., Antoniw, J. F. & Cohen, P. (1978) *Eur. J. Biochem.* **86**, 511-518.
- Jorgensen, A. O., Kalnins, V. I., Zubrzycka, E. & MacLennan, D. H. (1977) *J. Cell Biol.* **74**, 287-298.
- Campbell, K. P., MacLennan, D. H., Jorgensen, A. O. & Mintzer, M. C. (1983) *J. Biol. Chem.* **258**, 1197-1204.
- Weber, K. & Osborn, M. (1969) *J. Biol. Chem.* **244**, 4406-4412.
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680-685.
- Brandl, C. J., de Leon, S., Martin, D. R. & MacLennan, D. H. (1987) *J. Biol. Chem.*, in press.
- Garipey, J. & Hodges, R. S. (1983) *Biochemistry* **22**, 1586-1594.
- Blumenthal, D. K., Takio, K., Edelman, A. M., Charbonneau, H., Titani, K., Walsh, K. A. & Krebs, E. G. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3187-3191.
- MacLennan, D. H., Yip, C. C., Iles, G. H. & Seeman, P. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **37**, 469-477.
- Franzini-Armstrong, C. (1980) *Fed. Proc. Am. Soc. Exp. Biol.* **39**, 2403-2409.
- Ikemoto, N., Bhatnagar, G. M., Nagy, B. & Gergely, J. (1972) *J. Biol. Chem.* **247**, 7835-7837.