THE AMINO ACID SEQUENCE OF THE MAJOR PARVALBUMIN FROM HAKE MUSCLE *

J.-F. PECHERE, J.-P. CAPONY, L. RYDEN ** and J. DEMAILLE***

Département de Biochimie Macromoléculaire du CNRS, B. P. 1018, 34-Montpellier, France

Received April 20, 1971

SUMMARY - A sequence of 107 amino acids has been determined in the single chain of the major parvalbumin (pI 4.36) from hake (Merluccius merluccius) muscle as a result of studies on peptides obtained after tryptic, chymotryptic, peptic and cyanogen bromide cleavages of the protein.

Muscular parvalbumins are a family of homologous proteins (1-3) which are found abundantly in the muscles of lower vertebrates (4). Their biological function remains uncertain. Recent observations in this laboratory, however, suggest that they might be evolutionary precursors of the calcium sensitizing factor (troponin A) which is present in the muscles of higher vertebrates (5-9). The small size and the great variability of muscular parvalbumins - in number as well as in proportion and in composition- suggest anyway that they could be very suitable for the study of protein evolution. Work in this direction was thus started by determining the complete

---

* Contribution N°55 from the Department. This investigation was supported in part by contract N° 68 01207 with the Délégation Générale à la Recherche Scientifique et Technique. A preliminary report of some of the results has been made at the Meeting of the Société de Chimie biologique, Montpellier, May 12th 1970.

** Permanent address : Biokemiska Institutet, University of Uppsala, Sweden.

*** Permanent address : Laboratoire de Biochimie B, UER de Médecine, Montpellier, France.

amino acid sequence of the major component from the muscles of a widely available species, the hake, the three parvalbumins of which can be isolated in a simple manner.

The major component was obtained by a slight modification of the procedure described previously (10). It was found to be homogenous by disc electrophoresis in 12 % polyacrylamide gel at pH 9.4, by cellulose acetate electrophoresis at pH 5.5, by sedimentation and diffusion. The amino acid composition was found to be : Lys_{12}, His, Arg, Asx_{12-13}, Thr_5, Ser_5, Glx_{10}, Gly_{12}, Ala_{19}, 1/2 Cys, Val_4, Met, Ile_7, Leu_8, Phe_{10}, with 3 amide groups and 1 free -SH. No free N-terminus was available, which is also the case with several other parvalbumins, but hydrazinolysis (11) indicated Gly to be C-terminal.

About 4.5 μmoles of the S-sulfoderivative, prepared in the presence of 6M guanidine hydrochloride by the method of Chan (12), were digested for 5 hrs at 38° by 2.5 mg TPCK-trypsin in the presence of 2 M urea and 0.01 M Ca^{++}. The resulting peptides were isolated on a column of Sephadex G-10 in 0.1 M ammonium acetate pH 9.25, and subsequently separated by chromatography on Dowex 1-X2 in a linear acetic acid gradient and on Dowex 50-X4 in a linear pyridine gradient (13,14). Further purification of some of the peptides was achieved by paper electrophoresis at pH 6.4 or 3.6 (15) and paper chromatography (16). Tryptic peptides were also obtained from 4.9 μmoles performic acid oxidized (17) parvalbumin by digestion for 8 hrs, with 2.3 mg TPCK-trypsin, of the oxidized protein after its denaturation in 6M guanidine hydrochloride, and separation of the peptides on Dowex 50-X4 and paper as above. Chymotryptic peptides were obtained in the same way from 4.0 μmoles protein and from 2.2 μmoles performic acid oxidized protein, after their denaturation in 6M guanidine hydrochloride, by digestion for 8 hrs, with 2.2 mg and 1.2 mg chymotrypsin, respectively.

The diagonal method for cysteine peptides was performed as described by Brown and Hartley (18) on 0.5 umole parvalbumin digested with 0.25 mg pepsin. Cleavage at the single arginyl residue was effected with 0.4 mg TPCK-trypsin in presence of 0.01M Ca^{++} after maleylation (19) of 1.3 μmoles protein in presence of 6M guanidine hydrochloride, and the resulting fragments were separated at 3° on DEAE-cellulose in a chloride gradient at pH 7.5 (20). Cyanogen bromide cleavage (21) was effected in 75 % formic

acid on 0.6 μmoles parvalbumin and the peptides generated were fractionated on Sephadex G-25 in 0.2M acetic acid.

Sequential degradations were conducted either by the method of Blombäck et al. (22), or by the procedure of Gray (23). The identification of the PTH-amino acids was made by thin-layer chromatography (24-26); that of the DNS-amino acids by thin-layer chromatography (27,28) or paper electrophoresis (29). Carboxypeptidase digestions were performed according to standard procedures (30,26). The location of amide groups resulted from the Edman degradations and the electrophoretic mobility of the peptides (31), or from their digestion with aminopeptidase M (32).

The results of the sequence studies are summarized in Fig. 1.

They are in agreement with the amino acid composition found for the whole protein, except for the number of Ala residues, of which only 18, instead of 19, were found in the sequence. Also, of the two possible values (12-13) for the number of Asx residues, the sequence studies seem to indicate that the lower value is the correct one.

The nature of the N-terminal blocking group cannot be ascertained at this moment. The presence of at least one covalently bound acetyl group in the major parvalbumin from hake is apparent from its NMR spectrum (33). However, although phenylalanine is the only amino acid liberated from peptide C1b after acid hydrolysis, the mobility of the intact peptide at pH 6.4 is distinctly different from that of acetylphenylalanine and, also, from that of formyl-phenylalanine, a fact which argues against the presence of either of these two substituents alone at the N-terminus of the molecule. Further work is thus needed to elucidate this point and confirm the suggestion of Gerday and Rao (3) that parvalbumins have acetylated N-terminal amino acids.

Basic amino acids appear close to each other in positions 35-37 and 43-44 and clusters of acidic amino acids are found in positions 50-52, 58-61, 78-80 and 89-93. The abundance of hydrophobic residues is particularly noteworthy between the N-terminus of the chain and the single cysteine residue. This might indicate that, in the native protein, this portion of the molecule is buried - a situation which could explain why the S-sulfonation of the cysteine is quantitative only in the presence of guanidine hydrochloride, and why the introduction of the bulky substituent renders the protein much more susceptible to trypsin digestion.

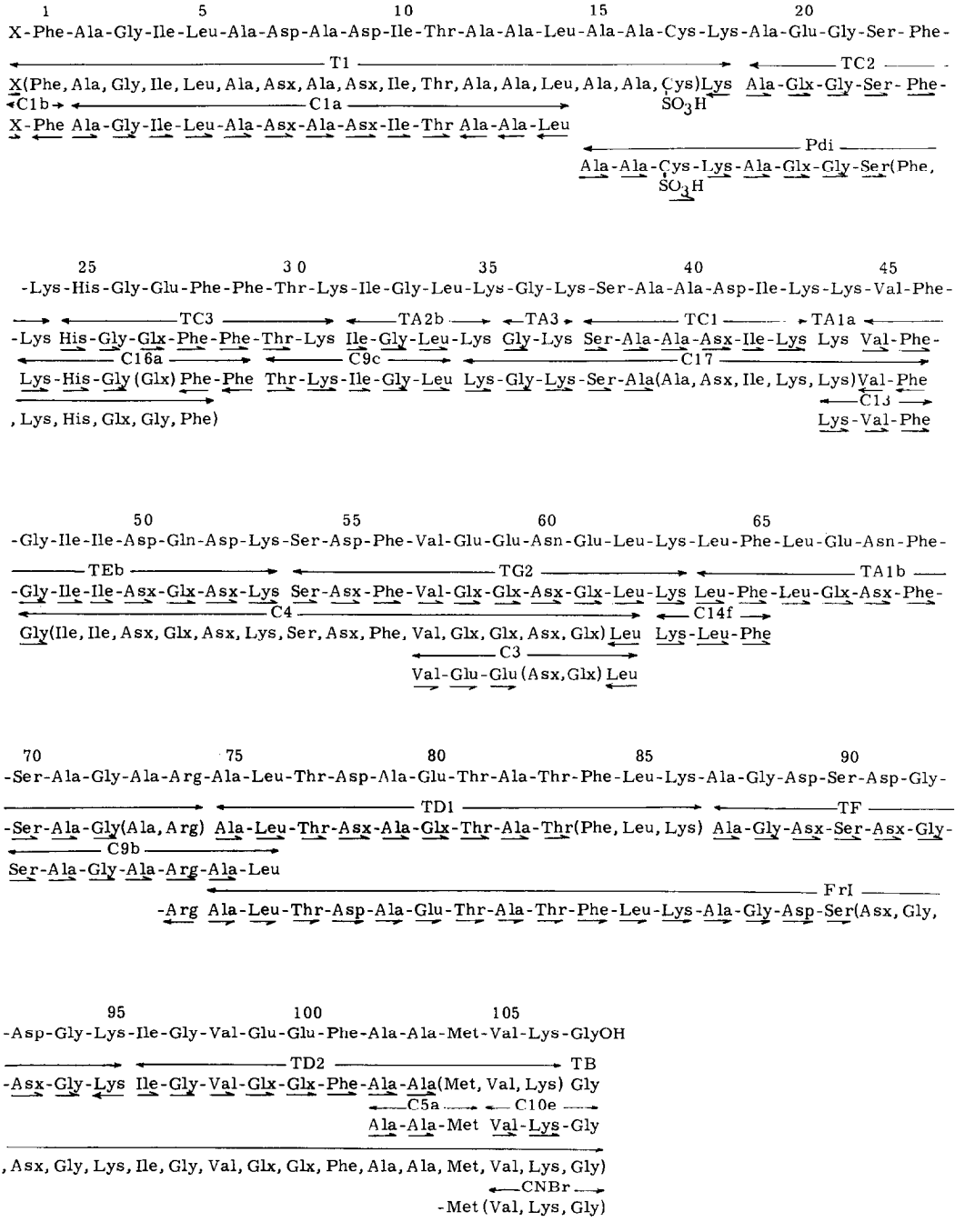

Fig. 1. Results of amino acid sequence studies on the major parvalbumin from hake muscle. The top line represents the proposed complete sequence; underneath are the peptides which provide evidence for it. Peptides whose designation starts with a T, a C and a P represent tryptic, chymo tryptic and peptic peptides, respectively. The occasional second capital letter indicates a fraction from a Dowex-1 column, while figures and small letters indicate a fraction from a Dowex-50 column and a further fractionation by paper methods, respectively. The diagonal peptide is labeled di, the fragment obtained after the Arg-split is designated FrI, and the cyanogen bromide peptide CNBr.

—→ Edman degradation; ⟶ Edman-dansyl degradation; ←— carboxypeptidase digestion. X represents the N-terminal blocking group.

A certain number of residues occur in doublets : Phe-Phe (28-29), Lys-Lys (43-44), Ile-Ile (48-49), Glu-Glu (58-59 and 99-100) and chiefly Ala-Ala, which appears four times (12-13, 15-16, 39-40, 102-103). Their frequency corresponds to what is expected statistically from the amino acid composition of the protein. Such is the case, also, for the twenty-three other heterologous dipeptide sequences which are found at least twice in the molecule. Two tripeptide sequences, Lys-Ile-Gly (31-33 and 95-97) and Val-Glu-Glu (57-59 and 98-100), appear twice, but with different relative spacings, so that an internal duplication in the molecule cannot be established on this basis. Application of the diagram test (34) at first sight yields a similar, negative result. However, the striking pattern of distribution along the chain of the Ala residues - which are present exclusively in two well defined regions of approximately equal length (2-40 and 71-103) each ending with a doublet - perhaps indicates that the possibility of such a duplication is not to be rejected completely.

Acknowledgements

The help of Mr. R. Bertrand in peptide column chromatography and of Mr. E. Audemard with amino acid analyses is gratefully acknowledged.

References

(1) J. F. Pechère, Comp. Biochem. Physiol., 24, 289 (1968)
(2) J. F. Pechère & J. P. Capony, Comp. Biochem. Physiol., 28, 1089 (1969)
(3) C. Gerday & K. S. P. B. Rao, Comp. Biochem. Physiol., 36, 229 (1970)
(4) B. Focant & J. F. Pechère, Arch. Intern. Physiol. Biochem., 73, 334 (1965)
(5) D. J. Hartshorne, M. Theiner & H. Mueller, Biochim. Biophys. Acta, 175, 320 (1969)
(6) M. C. Schaub & S. V. Perry, Biochem. J., 115, 993 (1969)
(7) M. L. Graeser & J. Gergely, Fed. Proc., 29, 463 (1970)
(8) S. Graffin & S. Watanabe, Fed. Proc., 29, 714 (1970)
(9) W. Drabikowski, R. Drabikowska & R. Barylko, FEBS-Lett., 12, 148 (1971).
(10) J. F. Pechère, J. Demaille & J. P. Capony, Biochim. Biophys. Acta, in press.
(11) B. C. Carlton & C. Yanowsky, J. Biol. Chem., 238, 636 (1963)
(12) W. W. C. Chan, Biochem., 7, 4247 (1968)
(13) W. A. Schroeder, R. T. Jones, J. Cormick & K. McCalla, Anal. Chem., 34, 1570 (1962)
(14) R. A. Bradshaw, D. R. Rabin, M. Nomoto, N. G. Shrinivasin, L. H. Ericsson, K. A. Walsh & H. Neurath, Biochem., 8, 3859 (1969)
(15) R. P. Ambler, Biochem. J., 89, 349 (1963)
(16) S. G. Waley & J. Watson, Biochem. J., 55, 238 (1953)
(17) C. H. W. Hirs, in Methods in Enzymology, 11, C. H. W. Hirs edit., Academic Press, New York & London (1967) p. 197

(18) J. R. Brown & B. S. Hartley, Biochem. J., 101, 214 (1966)
(19) P. J. G. Butler, J. I. Harris, B. S. Hartley & R. Lebermann, Biochem. J., 112, 679 (1969)
(20) B. E. Davidson, M. Sajgo, H. F. Noller & J. I. Harris, Nature 216, 1181 (1967)
(21) E. Gross, in Methods in Enzymology, 11, C. H. W. Hirs edit., Academic Press, New York & London (1967) p. 238
(22) B. Blombäck, M. Blombäck, P. Edman & B. Hessel, Biochim. Biophys. Acta, 115, 371 (1966)
(23) W. R. Gray, in Methods in Enzymology, 11, C. H. W. Hirs ed., Academic Press, New York & London (1967) p. 469
(24) M. Brenner, A. Niederwieser & G. Pataki, Experientia 17, 145 (1961)
(25) J. O. Jeppson & J. Sjöquist, Anal. Biochem., 18, 264 (1967)
(26) F. J. Morgan & A. Henschen, Biochim. Biophys. Acta, 181, 93 (1969)
(27) D. Morse & B. L. Horecker, Anal. Biochem., 14, 429 (1966)
(28) K. Crowshaw, S. J. Jessup & P. W. Ramswell, Biochem. J., 103, 79 (1967)
(29) W. R. Gray, in Methods in Enzymology, 11, C. H. W. Hirs edit., Academic Press, New York & London (1967) p. 139
(30) R. P. Ambler, in Methods in Enzymology, 11, C. H. W. Hirs edit., Academic Press, New York & London (1967) p. 155
(31) R. E. Offord, Nature 211, 591 (1966)
(32) C. Bennett, W. H. Konigsberg & G. M. Edelman, Biochem., 9, 3181 (1970)
(33) J. Parello & J. F. Pechère, in preparation
(34) A. J. Gibbs & G. A. McIntyre, Eur. J. Biochem., 16, 1 (1970)