

# Complete amino acid sequence of the sarcoplasmic calcium-binding protein (SCP-I) from crayfish (*Astacus leptodactylus*)

J. Jauregui-Adell, W. Wnuk\* and J.A. Cox\*

Centre de Recherches de Biochimie Macromoléculaire du Centre National de la Recherche Scientifique, Boîte Postale 5051, F-34033 Montpellier Cedex, France and \*Département de Biochimie, Université de Genève, Quai Ernest-Ansermet 30, CH-1211 Genève 4, Switzerland

Received 24 November 1988

The complete amino acid sequence of the alpha chain of the dimeric sarcoplasmic  $\text{Ca}^{2+}$ -binding protein (SCP-I =  $\alpha_2$ ) from crayfish (*Astacus leptodactylus*) has been determined by partial automatic sequencing of the peptides derived from tryptic digests of the protein after citraconylation or treatment with 1,2-cyclohexanedione. Overlapping peptides were obtained by cleavage with *o*-iodosobenzoic acid, or digestion with *Staphylococcus aureus* protease, thermolysin and pepsin. The acetylated N-terminus was identified by fast atom bombardment mass spectrometry. The monomeric protein contains 192 amino acids and has an  $M_r$  of 21 643. The sequence shows the presence of three calcium-binding sites and perhaps of two others that may be degenerated.

Chain,  $\alpha$ -; Sarcoplasmic protein;  $\text{Ca}^{2+}$ -binding protein; Amino acid sequence; (*Astacus leptodactylus*)

## 1. INTRODUCTION

Sarcoplasmic calcium-binding proteins are soluble proteins present in both vertebrate and invertebrate muscles. Their function is not yet fully understood. Parvalbumins (PV) which are found in vertebrate sarcoplasm, appear to function as a relaxing factor in fast skeletal muscle contraction [1,2]. Interest in PV led to the first amino acid sequence [3] and crystal structure [4] among the calcium-modulated proteins.

In invertebrate muscles, where PV are not found, other acidic calcium-binding proteins have been isolated [5–9]. Proteins from crustaceans exist as dimers and dissociate in the presence of SDS into subunits of 22 kDa. In other invertebrates, only monomeric isotypes are found [10].

Among the sarcoplasmic calcium-binding proteins from invertebrates (SCP), crayfish (*Astacus leptodactylus*) SCPs were the first studied [11]. Their isolation showed a polymorphism with two distinct polypeptide chains: SCP-I ( $\alpha_2$ ), SCP-II ( $\alpha, \beta$ ), SCP-III ( $\beta_2$ ). These isotypes are found in three protein peaks in the proportion 14:1.5:1 according to their absorbance values after elution from DE-52 cellulose column [10]. An analogous polymorphism was observed after the isolation of SCPs from shrimp (*Penaeus* sp.) tail muscle [6].

Of the three isotypes from crayfish, SCP-I is the best characterized. This protein binds 6  $\text{Ca}^{2+}$  per dimeric molecule [12]. The dimer has two  $\text{Ca}^{2+}$ -specific and four  $\text{Ca}^{2+}$ - $\text{Mg}^{2+}$  sites and structural changes induced by the binding of these metals have been reported [13].

A better knowledge of their properties requires determination of the primary and tertiary structures. Whereas work on the three-dimensional structure is in progress (Kretsinger, R.H., personal communication), we describe in this report the complete amino acid sequence of the  $\alpha$ -chain from

Correspondence address: J. Jauregui-Adell, CRBM, CNRS, PO Box 5051, F-34033 Montpellier-Cedex, France

Abbreviations: PV, parvalbumin; SCP, sarcoplasmic calcium-binding protein; CaBP, calcium-binding protein

crayfish SCP-I. A comparison of this sequence with other well-known calcium-modulated proteins has recently been made [14].

## 2. MATERIALS AND METHODS

### 2.1. Purification of protein SCP-I

Protein SCP-I was purified from crayfish muscle as described in [11].

### 2.2. Peptide preparation after enzymatic and chemical cleavages

The protein was carboxymethylated according to [15] and allowed to react with citraconic anhydride [16] or 1,2-cyclohexanedione [17]. After tryptic digestion, peptides were separated by chromatography on two columns (0.9 × 200 cm) of Sephadex G-50 fine, or on one column (0.9 × 140 cm) of Bio Gel P4, both equilibrated with 0.5% ammonium bicarbonate.

In one experiment, the tryptophanyl bonds were cleaved by the *o*-iodosobenzoic acid reagent (Pierce) on the carboxymethylated protein, according to reaction 3 in [18]. Then, insoluble peptides were citraconylated and fractionated as above.

Subdigestion, when necessary, involved *Staphylococcus aureus* V8 protease (Pierce), thermolysin (Serva) or pepsin (Merck), and the subpeptides were fractionated by chromatography on two columns (1.3 × 120 cm) of Bio Gel P10, or a column (0.46 × 25 cm) of  $\mu$ -Bondapak-phenyl. In the latter case, the peptides were eluted at a flow rate of 2 ml/min with a linear gradient from 0 to 80% (0.09% TFA, 90% acetonitrile) for 55 min, and monitored at 220 and 280 nm.

The performic acid oxidized protein or the heat denatured protein as described in [19], were also digested by trypsin and the tryptic peptides were fractionated on a column of Dowex 50 (0.9 × 20 cm) with a pyridine gradient or a  $\mu$ -Bondapak-phenyl reverse phase column.

### 2.3. Sequence determination

Amino acid analysis, automatic Edman degradation and phenylthiohydantoin derivative identification were performed as described in [15].

### 2.4. Molecular mass determination

The molecular mass of some peptides was determined by fast atom bombardment, to determine the global composition in amino acids [20].

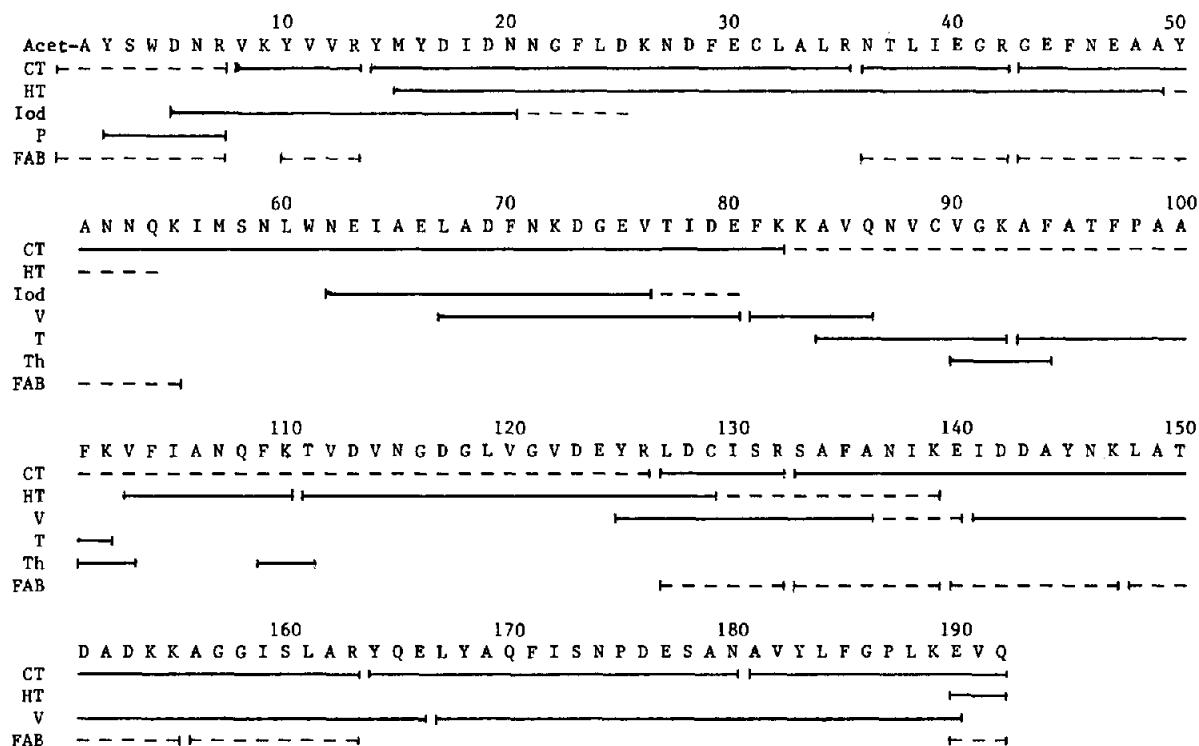


Fig.1. The amino acid sequence of the  $\alpha$ -chain protein. T, CT, HT: tryptic peptides resulting from the denatured, the citraconylated or the cyclohexanedione-treated protein; Iod, V, Th, P: subpeptides derived from cleavage with iodosobenzoic acid, *Staphylococcus aureus* V8 protease, thermolysin and pepsin, respectively; — represents the portion of peptides determined by automated Edman degradation and - - - the portion derived from amino acid composition; FAB: peptides with global composition obtained by fast atom bombardment.

### 3. RESULTS AND DISCUSSION

#### 3.1. Sequence determination

The complete amino acid sequence of protein SCP-I is presented in fig.1. Most of the sequence was easily deduced from automatic Edman degradation of the long peptides obtained after citraconylation or reaction of the protein with 1,2-cyclohexanedione. Overlapping peptides were obtained from cleavage with *o*-iodosobenzoic acid or from *S. aureus* protease digestion. Only the 'core' of the protein (positions 83–102) was difficult to obtain directly, due to the accumulation of hydrophobic residues responsible for peptide loss on the different columns. Therefore, it was necessary to use the tryptic peptides which were overlapped by sequencing subpeptides from thermolysin digestion.

As the N-terminal amino acid was blocked, the sequence was obtained from a peptic subpeptide, the acetyl-alanine terminus from the amino acid composition and the molecular mass determined unambiguously by fast atom bombardment [20].

The C-terminal sequence was derived directly from subpeptides and confirmed by results of carboxypeptidase A digestion.

During this work, a microheterogeneity in positions 78 and 112 with both valine and isoleucine in erratic quantities depending on the sequenced peptide was observed. In an attempt to quantify these amino acids, the tryptic peptides from the heat denatured protein were fractionated by HPLC. A peptide, T9<sub>73–83</sub>, was isolated that gave in position 78: isoleucine 68%, valine 32%; two other peptides, T14<sub>111–126</sub> and T14<sub>111–126</sub>, in the ratio 3:1 showed only valine or isoleucine, respectively, in position 112. We think that the isolated protein (SCP-I) was in fact a mixture of two chains,  $\alpha_a$  and  $\alpha_b$ , the latter different at least in two residues. An analogous microheterogeneity was observed by Takagi and Konishi who characterized two  $\alpha$ -chains from shrimp tail muscle, differing only by four residues [6].

#### 3.2. Characteristics of the sequence

Monomeric SCP-I contains 192 amino acids and its calculated molecular mass is 21643 Da. The value is in agreement with the data obtained previously by gel filtration and equilibrium sedimentation [10,11]. The amino acid composi-

tion as derived from the sequence agrees also with the composition obtained from the amino acid analysis (table 1). Minor discrepancies could be explained by the mixture of  $\alpha_a$  and  $\alpha_b$  chains.

Investigation of calcium-binding sites in the protein SCP-I amino acid sequence shows the presence of three loops (fig.2) with the typical EF hand calcium-binding site [22]. Loops IV and V, if true,

Table 1  
Amino acid composition of the  $\alpha$ -chain from SCP-I

Amino acid	Residues	
	From analysis [10]	From sequence (this work)
Asp	35	17
Asn	—	18
Thr	5	5
Ser	8	7
Glu	21	13
Gln	—	6
Pro	3	3
Gly	12	11
Ala	23	23
Cys	3	3
Val	13	14
Met	2	2
Ile	11	11
Leu	13	13
Tyr	9	10
Phe	13	13
Trp	2	2
Lys	14	14
His	0	0
Arg	7	7

		X	Y	Z		(Y)		X		Z					
		1	2	3	4	5	6	7	8	9	10	11	12		
Calmodulin															
Loop	I	20	D	K	D	G	D	G	T	I	T	T	K E	31	
SCP-I															
Loop	I	17	D	I	D	N	N	G	F	L	D	K	N	D	28
Loop	II	69	D	F	N	K	D	G	E	V	T	I	D	E	80
Loop	III	113	D	V	N	G	D	G	L	V	G	V	D	E	124
Loop	IV	140	E	I	D	D	A	Y	N	K	L	A	T	D	151
Loop	V	174	N	P	D	E	S	A	N	A	V	Y	L	F	185

Fig.2. Calcium-binding sites in  $\alpha$ -chain protein (SCP-I) compared with loop I from calmodulin [21]. Possible calcium coordinating positions are boxed.

cannot bind the metal. They have lost some of the amino acids with oxygen-containing side chains that are essential to bind calcium, and the glycine at position 6 of the loop which is necessary to form a sharp bend. Protein SCP-I is indeed isolated as a dimer which binds 6  $\text{Ca}^{2+}$ .

Finally, a study about the homology of the calcium-modulated proteins has been made recently, where the protein SCP-I is included in a complex subfamily containing also the aequorins and the luciferin-binding protein [14].

More precise comparison of the structural features of sarcoplasmic calcium-binding proteins with other CaBPs will await determination of the crystal structure.

**Acknowledgements:** We thank Dr T.C. Vanaman for participation in the first phase of this work and Dr J.P. Capony for carrying out the amino acid analysis. The work of J.J.A. was supported by the 'Centre National de la Recherche Scientifique' and the 'Institut National de la Santé et de la Recherche Médicale'. W.N. and J.A.C. were supported by the Swiss National Science Foundation, grant no.3.577-0.87, and by the Muscular Dystrophy Association of America.

## REFERENCES

- [1] Pechère, J.F., Derancourt, J. and Haiech, J. (1977) *FEBS Lett.* 75, 111–114.
- [2] Gillis, J.M., Thomason, D.B., Lefèvre, J. and Kretsinger, R.H. (1982) *J. Mus. Res. Cell Motil.* 3, 377–398.
- [3] Pechère, J.F., Capony, J.P., Ryden, L. and Demaille, J. (1971) *Biochem. Biophys. Res. Commun.* 43, 1106–1111.
- [4] Kretsinger, R.H., Nockolds, C.E., Coffee, C.J. and Bradshaw, R.A. (1971) *Cold Spring Harbor Symp. Quant. Biol.* 36, 217–220.
- [5] Wnuk, W., Cox, J.A. and Stein, E.A. (1982) in: *Calcium and Cell Function* (Cheung, W.Y. ed.) vol.2, pp.243–278, Academic Press, New York.
- [6] Takagi, T. and Konishi, Y. (1984) *J. Biochem.* 95, 1603–1615.
- [7] Takagi, T. and Konishi, Y. (1984) *J. Biochem.* 96, 59–67.
- [8] Takagi, T., Kobayashi, T. and Konishi, K. (1984) *Biochim. Biophys. Acta* 787, 252–257.
- [9] Kobayashi, T., Takasaki, Y., Takagi, T. and Konishi, K. (1984) *Eur. J. Biochem.* 144, 401–408.
- [10] Wnuk, W. and Jauregui-Adell, J. (1983) *Eur. J. Biochem.* 131, 177–182.
- [11] Cox, J.A., Wnuk, W. and Stein, E.A. (1976) *Biochemistry* 15, 2613–2618.
- [12] Wnuk, W., Cox, J.A., Kohler, L.G. and Stein, E.A. (1979) *J. Biol. Chem.* 254, 5284–5289.
- [13] Wnuk, W., Cox, J.A. and Stein, E.A. (1981) *J. Biol. Chem.* 256, 11538–11544.
- [14] Kretsinger, R.N., Moncrief, N.D., Goodman, M. and Czelusniak, J. (1988) in: *The Calcium Channel* (Kusche, H. ed.) in press.
- [15] Jauregui-Adell, J., Pechère, J.F., Briand, G., Richet, C. and Demaille, J. (1982) *Eur. J. Biochem.* 123, 337–345.
- [16] Van Eerd, J.P. and Takahashi, K. (1976) *Biochemistry* 15, 1171–1180.
- [17] Dognin, M.J. and Wittmann-Liebold, B. (1980) *Eur. J. Biochem.* 112, 131–151.
- [18] Fontana, A., Dalzoppo, D., Grandi, C. and Zamboni, M. (1981) *Biochemistry* 20, 6997–7004.
- [19] Jauregui-Adell, J. and Pechère, J.F. (1978) *Biochim. Biophys. Acta* 536, 275–282.
- [20] Aubagnac, J.L., Salesse, A. and Jauregui-Adell, J. (1988) *Biomed. Environ. Mass Spectrom.*, in press.
- [21] Watterson, D.M., Sharief, F.S. and Vanaman, T.C. (1980) *J. Biol. Chem.* 255, 962–971.
- [22] Kretsinger, R.H. (1976) *Annu. Rev. Biochem.* 45, 239–266.