Volume 295, number 1,2,3, 93–96 FEBS 10517 December 1991

Molecular cloning and expression of the cDNA coding for a new member of the S100 protein family from porcine cardiac muscle

Hisataka Ohta[1], Toshiya Sasaki[1], Michiko Naka[1], Osamu Hiraoka[2], Chikara Miyamoto[2], Yasuhiro Furuichi[2] and Toshio Tanaka[1]

[1]*Department of Molecular and Cellular Pharmacology, Mie University School of Medicine, Edobashi, Tsu, Mie 514, Japan* and [2]*Department of Molecular Genetics, Nippon Roche Research Center, Kamakura, Kanagawa 247, Japan*

Received 13 September 1991

We isolated a new calcium-binding protein from porcine cardiac muscle by calcium-dependent hydrophobic and dye-affinity chromatography. It showed an apparent molecular weight of 11 000 on SDS-PAGE. Amino acid sequence determination revealed that the protein contained two calcium-binding domains of the EF-hand motif. The cDNA gene coding for this protein was cloned from the porcine lung cDNA library. Sequence analysis of the cloned cDNA showed that the protein was composed of 99 amino acid residues and its molecular weight was estimated to be 11 179. Immunological and functional characterization showed that the recombinant S100C protein expressed in *Escherichia coli* was identical to the natural protein. Homologies to calpactin light chain, S100α and β protein were 41.1%, 40.9% and 37.5%, respectively. The protein was expressed at high levels in lung and kidney, and low levels in liver and brain. The tissue distribution was apparently different from those of the other S100 protein family. These results indicate that this protein represents a new member of the S100 protein family, and thus we refer to it as S100C protein.

S100C; cDNA; Expression; S100 protein; Calcium-binding protein; Porcine cardiac muscle

1. INTRODUCTION

An increase in intracellular calcium triggers a variety of cellular events, and the calcium ion is mediated by calcium-binding proteins. Most well known calcium-binding proteins are groups which have the 'EF-hand' motif in their structure such as calmodulin, troponin C and S100 proteins. The S100 protein was first identified as a low molecular mass, acidic calcium-binding protein present in brain tissue [1]. The S100 proteins have two EF-hand structures within the molecule and most of these proteins exist as homo- or heterodimers [2]. Recently, new members of S100 related proteins were found in various tissues and cell lines [3]. Some of these members have been implicated to be involved in cell differentiation and cell cycle regulation [3], but the exact function of most S100 proteins is still uncertain. During our search for new calcium-binding proteins, we found in porcine cardiac muscle a protein which shared some of the structural features with the known S100 proteins. This novel S100-related calcium binding protein had a molecular weight of 11 000, as measured by SDS-PAGE, and thus, was termed S100C protein.

In this paper, we describe the cloning of cDNA coding for the S100C protein, the determination of the complete nucleotide sequence, the characterization of the recombinant S100C protein expressed in *Escherichia coli* and the tissue distribution by Northern and Western blotting analyses.

Correspondence address: T. Tanaka, Dept. Molecular and Cellular Pharmacology, Mie University School of Medicine, Edobashi, Tsu, Mie 514, Japan.

2. MATERIALS AND METHODS

2.1. *Purification of calcium-binding protein, S100C protein*

The calcium-binding protein S100C was purified from porcine heart by the procedure of Naka et al. (to be published elsewhere).

2.2. *Amino acids sequence analysis*

The purified S100C protein was digested by either trypsin or *Staphylococcus aureus* V8 protease digestion at an enzyme to substrate ratio 1:100 (w/w) for 3 h. The resulting peptides were separated by reverse-phase HPLC using Waters C_R columns with a linear gradient of 0–80% (v/v) acetonitrile in 0.05% trifluoroacetic acid. Peptides selected at random were sequenced using an Applied Biosystems model 473A amino acid micro-sequencer.

2.3. *Construction of porcine heart cDNA libraries*

Total RNA was isolated by guanidinium thiocyanate methods described by Okayama et al. [4]. The poly(A) containing RNA was purified by the batch method that used Oligotex(dT)^{30} (JSR/Nippon Roche, Tokyo Japan) [5]. Double-stranded cDNA was prepared from the poly(A) RNA using reverse transcriptase with oligo(dT) 25-mer as a primer, and the library was constructed in λgt 10 [6]. The commercially available λgt 10 library of porcine lung (Clontech) was also used for the cloning.

2.4. *Expression of S100C protein*

S100C protein was produced in *E. coli* as a hybrid protein which contains the extra polypeptide (Met-Arg-Gly-Ser-(His)$_6$-Gly-Ser) at the N-terminus to make the purification of the recombinant S100C easy and prompt [7].

3. RESULTS AND DISCUSSION

3.1. *Amino acid and cDNA sequencing*

In an effort to determine the complete amino acid sequence of S100C, we first carried out sequence analysis of peptides obtained by digestion with cyanogen bromide, trypsin and *Staphylococcus aureus* V8 protease. Our initial work suggested that the S100C protein contained a blocked N-terminus. The peptides produced by limited digestion were purified by reverse-phase HPLC column chromatography and were individually sequenced. A total of 82 residues of the amino acid sequences were thus determined. The sequences (more than ten) were aligned with the other members of the S100 protein family and comparative studies were carried out (Fig. 1). Several oligonucleotides were prepared on the basis of the amino acid sequence, and used as primers for amplification of the cDNA fragment that encoded the protein. A fragment of about 80 bp was obtained by the polymerase chain reaction (PCR) from a porcine cDNA library [8]. The nucleotide sequence of this 80-bp fragment was determined, which matched well with the amino acid sequence obtained from peptides (Fig. 1 boxed). A λgt 10 library prepared from poly(A)-containing RNA of porcine lung, a rich source of this protein (Fig. 2), was screened for the full-size cDNA clone using the 80-bp probe DNA which was labeled at the 5′-terminus by [α-^{32}P]ATP and polynucleotide kinase. The nucleotide sequence was determined by the dideoxy method of Sanger [9]. We obtained a clone which contained the full-length open reading frame. From the sequence data (Fig. 1) obtained with the cloned cDNA, the following points were clarified: (a) the cDNA contains 508 nucleotides (including the poly(A) tail and the poly(A) addition signal [10]); (b) there is one open reading frame that encodes 99 amino acids and a strong initiation codon for protein synthesis which completely matches the consensus sequence, GACATGG [11]; (c) the molecular mass of the predicted polypeptide is calculated to be 11 179 Da which fits well with our previous estimation by SDS-PAGE; (d) as anticipated, the protein contains two calcium-binding sites of the EF hand motif — the normal EF hand with 28 amino acids is located at the

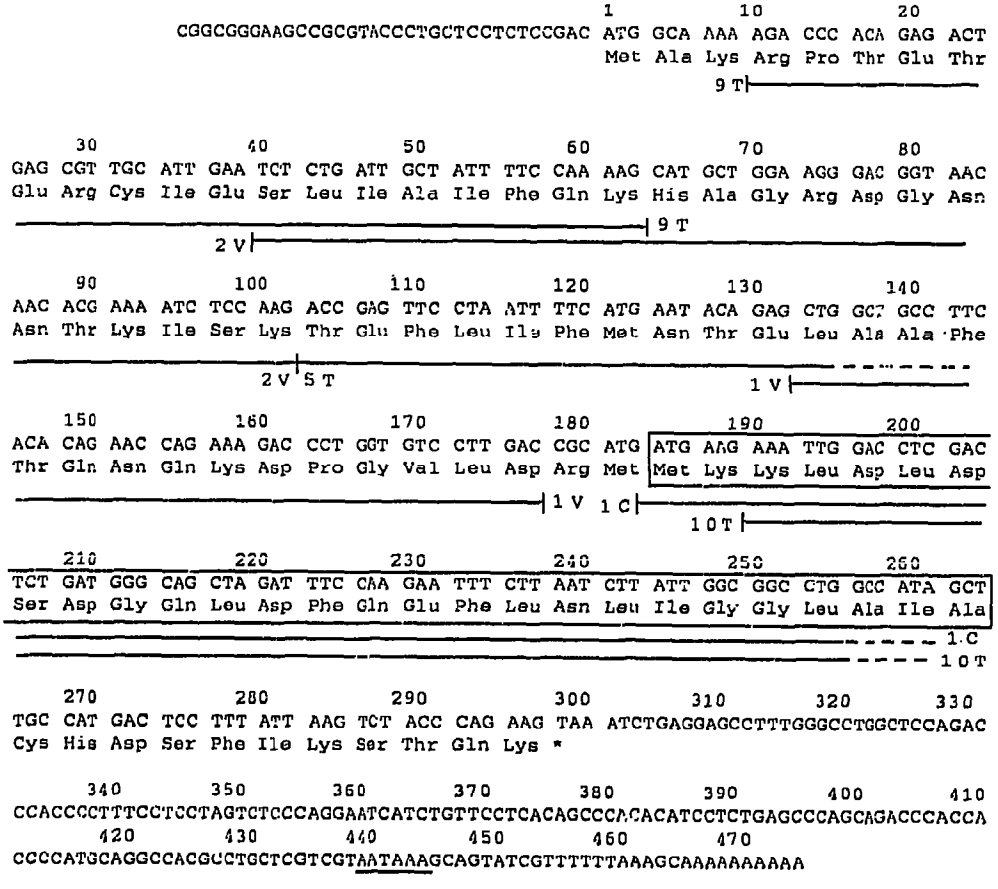

Fig. 1. Nucleotide and deduced amino acid sequences of the porcine S100C protein. The sequences obtained by micro-sequencing of the peptides of the S100C protein are shown (underlined). Also underlined is the polyadenylation signal [10]. The 80-bp DNA fragment used as a probe is indicated by the open box. T, trypsin; V, *Staphylococcus aureus* V8 protease; C, cyanogen bromide.

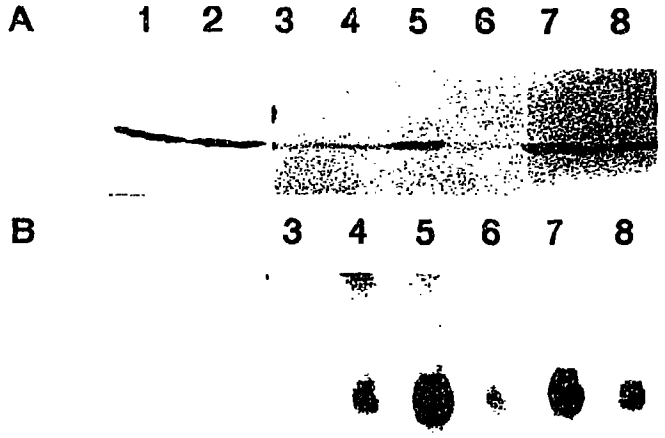

Fig. 2. A. Immunoblot analysis of the various tissue extracts with polyclonal antibody raised against the purified S100C protein from porcine cardiac muscle. Tissue extracts (150 μg each) were separated by SDS-PAGE (15%) and electrophoretically transferred to PVDF membranes: native protein from porcine cardiac muscle (lane 1); the expressed protein in *E. coli* (lane 2); liver (lane 3); heart (lane 4); kidney (lane 5); brain (lane 6); lung (lane 7); adrenal gland (lane 8). B. Distribution of S100C protein mRNA. RNA was isolated from the indicated mouse tissues and 20 μg of RNA per lane was hybridized, under stringent conditions [6], to the S100C specific DNA probe: liver (lane 3); heart (lane 4); kidney (lane 5); brain (lane 6); lung (lane 7); adrenal gland (lane 8).

C-terminal part of the protein and the variant EF hand with 30 amino acids is found in the N-terminal part of the protein [12,13].

3.2. *Expression of the S100C protein and its calcium binding*

An expression vector coding for the S100C protein was constructed in pDS56/RBS II, 6HIS plasmid and the fusion S100C protein containing 6 histidine residues at the N-terminus was produced in *Escherichia coli*. Polyclonal antibodies made against the purified porcine heart S100C protein was immunoreactive to this fusion protein on Western blot experiments (Fig. 2). The ability of the recombinant fusion protein (rS100C) to bind Ca^{2+} was examined directly by the dot-blot assay. The membrane which retained the rS100C or native S100C proteins were incubated in a solution containing $^{45}Ca^{2+}$. We could readily detect the Ca^{2+} binding to the expressed rS100C or to the purified native S100C proteins.

A computer-assisted search of protein sequence libraries showed that the predicted amino acid sequence of the S100C protein shared a significant homology with several other Ca^{2+} binding proteins. As shown in Fig. 3, the S100C protein was highly homologous to the calpactin light chain, bovine S100α and S100β subunit with homologies of 41.1%, 40.9% and 37.5%, respectively. The homology strongly suggested that the S100C protein is structurally related to the S100 protein family. This conclusion was further supported by three additional observations. First, the best homology matching of the amino acid sequence was obtained without the introduction of gaps for any sequence compared. Second, many of the amino acid differences were conservative substitutions. Third, the homology between the S100C protein and the bovine S100α subunit was particularly striking (100%) in the regions where the highest conservation between the α and β subunits of the S100 protein occurred. These homology analyses indicated that the S100C protein is located in the earlier branch of S100 subfamily lineage, between the S100α subunit and the calpactin light chain [13].

3.4. *Tissue distribution*

The presence of S100C was examined in several tissues by Northern and Western blotting analyses. As shown in Fig. 2, S100C specific mRNAs were most abundant in lung and kidney. Intermediate levels were found in heart and adrenal gland. In brain and liver, the levels were very low, since Northern blot analysis using 20 μg of total RNA from the tissue did not give rise to any detectable signals. These data were also confirmed by Western blotting analysis (Fig. 2). It should be noted that the tissue distribution of S100C protein was different from those of other S100 protein family members. Therefore, it is probable that the S100C protein may play a role in cell regulatory function distinct from other S100 proteins.

At the moment, the precise role of the S100 proteins is still unclear, although circumstantial evidence implies that the S100C protein may act as a vehicle for Ca^{2+} ions, both intracellularly and extracellularly. A member

```
                    L··LL··L+··OGO·+·LC··OL··LL··O          L··LL··LO·O·GG·LO··OL··LL··L
S100C:P     MAKRPTETERCIESLIAIFQKHAGRDGHNTKISKTEFLIFMNTELAAFTQNQKDPGVLDRMMKKLDLDSDGQLDFQEFLNLIGGLAIACHDSFIKSTQK
p11:P         PSQMEHAMETMMFTFHKFA   GDKGYLTKEDLRVLMEKEFPGFLENQKDPLAVDKIMKDLDQCRDGKVGFQSFFSLIAGLTIACNDYFVVHMKQKGK      41.1 %
S100a:B       GSELETAMETLINVFHAHSGKEGDKYKLSKKELKELLQTELSGFLDAQKDADAVDKVMKELDENGDGEVDFQEYVVLVAALTVACNNFFWENS          40.9 %
S100β:B        SELEKAVVALIDVFHQYSGREGDKHKLKKSELKELINNELSHFLEEIKEQEVVDKVMETLDSDGDGECDFQEFMAFVAMITTACHEFFEHE           37.5 %
S100L:B     MSSPLEQALAVMVATFHKYSGQEGDKFKLSKGEMKELLHKELPSFVGEKVDEEGLKKLMGDLDENSDQQVDFQEYAVFLALITIMCNDFFQGSPARS      33.0 %
2A9:H       MACPLDQAIGLLVAIFHKYSGREGDKHTLSKKELKELIQKEL TIGSKLQDAEI ARLMEDLDRNKDQEVNFQEYVTFLGALALIYNEALKG          30.6 %
ICaBP:P       SAQKSPAELKSIFEKYAAKEGDPNQLSKEELKQLIQAEFPSL   LKGPRTLDDLFQELDKNGNGEVSFEEFQVLVKKISQ                   28.0 %
pEL98:M     MARPLEEALDVIVSTFHKYSGKEGDKFKLNKTELKELLTRELPSFLGKRTDEAAFQKVMSNLDSNRDNEVDFQEYCVFLSCIAMMCNEFFEGCPDKEPRKK   27.5 %
```

Fig. 3. Comparison of amino acid sequences of different members of the S100 protein family. Shaded areas indicate positions of which at least four sequences share the same amino acid residue. O stands for residues EDQNST (oxygen-containing residues), L stands for residues LVIFM (hydrophobic amino acids), and + indicates inserted amino acids. p11, calpactin light chain; ICaBP, intestinal calcium binding protein; P, porcine; B, bovine; H, human; M, mouse. Homology % to S100C protein are shown on the right side.

of the S100 proteins, calpactin light chain, interacts in a 1:1 molar ratio with p36, a substrate of tyrosine protein kinase and induces in p36 a calcium sensitivity [14]. It is therefore possible that S100C modulates the activity of other intracellular proteins involved in cellular signal transduction. The identification of a target protein of the S100C protein will be important in the understanding of the biological function of S100C protein.

REFERENCES

[1] Moore, B.W. (1965) Biochem. Biophys. Res. Comm. 19, 739–744.
[2] Isobe, T. and Okuyama, T. (1978) Eur. J. Biochem. 89, 379–388.
[3] Kligman, D. and Hilt, D.C. (1988) Trends Biochem. Sci. 13, 437–443.
[4] Okayama, H., Kawaichi, M., Brownstein, M., Lee, F., Yokota, T. and Arai, K. (1987) Methods Enzymol. 154, 3–28.
[5] Kuribayashi, K., Hikata, M., Hiraoka, O., Miyamoto, C. and Furuichi, Y. (1988) Nucleic Acids Res. Symposium Series 19, 61–64.
[6] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning, A laboratory manual (2nd edn.) Cold Spring Harbor Laboratory.
[7] Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R. and Stüber, D. (1988) Bio/Technology 6, 1321–1325.
[8] Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) Science 230, 1350–1354.
[9] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[10] Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41, 349–359.
[11] Kozak, M. (1986) Cell 44, 283–292.
[12] Szebenyi, D.M.E., Obendorf, S.K. and Moffat, K. (1981) Nature 294, 327–332.
[13] Kretsinger, R.H., Moncrief, N.D., Goodman, M. and Czelusniak, J. (1988) in: Homology of calcium-modulated proteins: their evolutionary and functional relationships (M. Morad, W. Nayler, S. Kazda and M. Schramm, eds.) The calcium channel: structure, function and implications, pp. 16–35, Springer, Berlin.
[14] Gerke, V. and Weber, K. (1985) J. Biol. Chem. 260, 1688–1695.