ORGANIZATION AND EXPRESSION OF CALMODULIN GENES IN PLANTS

Sang Hyoung Lee, Jong Cheol Kim, Mal Soon Lee, Young Ju Choi, Jong Chan Hong, Jeong Dong Bahk, and Moo Je Cho

Plant Molecular Biology and Biotechnology Research Center, Gyeongsang National University, Chinju 660-701, Korea.

I. Introduction

Calmodulin (CaM) is a ubiquitous, highly conserved, calcium binding protein which interacts with a number of enzymes and mediates their activities in eukaryotic cell (VanEldik et al., 1982 ; Poovaiah and Reddy, 1987). Since the discovery of CaM in plants (Anderson et al., 1978 ; Muto and Miyachi, 1977) the role of CaM in plant metabolism, organization and expression of the genes has been under active investigation.

Plant responses to various signals such as light, stress, phytohormones, and gravity control, and the processes involved in the transduction of these signals have been a great interests during the past decade (Hepler et al., 1985). Accumulating evidence strongly suggest the involvement of calcium ion as a main secondary signal tranducer in plants cells as well as in animals, and many of these calcium- involved signal transductions are shown to be mediated by calcium binding proteins such as CaM (Leonard and Hepler, 1990).

CaM is a small, acidic and heat-stable protein with four calcium binding domain. The structure of calcium binding domain has the unique helix-loop-helix

structure called EF-hand (Strynadka and James, 1989). The Ca2+- binding domain I, II and III, IV are linked by the flexible central a - helix, and constitute the dumbell shape of CaM (Babu et al., 1988).

One of the remarkable characteristics of CaM is its highly conserved primary structure. Amino acid sequences of CaM are nearly identical among mammals and avian species, and higher plant CaM shares more than 95% amino acid sequence homology with vertebrate CaM.

In human and rats, CaM is encoded by a family of at least three genes (Fisher et al., 1988; Nojima et al., 1987). The human CaM proteins encoded by each gene family member, have an identical amino acid sequence, although their respective nucleotide sequences are diverged by approximately 20%. In plants, CaM is encoded by a single gene in barley (Ling et al., 1989) and potato (Jena et al., 1989), and by six genes in Arabidopsis (Perera and Zielinski, 1992).

cDNAs encoding plant CaM were isolated from various sources including potato (Jena et al., 1989), barley (Ling et al., 1989), Arabidopsis (Ling et al., 1991; Perera and Zielinski, 1992) and Alfalfa (Barnette and Long, 1990). Their deduced amino acid sequences have very high sequence identity (more than 90% except for potato CaM). One noticeble characteristics of reported plant CaM is the absence of tyrosine residue in the 3rd Ca2+- binding domain. Although no information is available up to now about phosphorylation of plant CaM, several reports are present in the case of animal CaM. Interestingly, alteration of CaM activity due to the phosphorylation of tyrosine residue by src kinase is reported in RSV-transformed chicken cell system (Fukami et al., 1986).

The expression of CaM or CaM-related genes is induced by environmental stimuli, including touch, rain, wind, wounding and darkness (Janet and Davis, 1977). These results suggest that CaM promoters have multiple regulatory element (Zambryski et al., 1989).

Our laboratory is concerned with several questions regarding with CaM :

1) What is the extent of amino acid sequence divergence for CaM among higher plant.
2) What is the physiological role of the individual CaM encoded by multigene family.
3) Has the individual CaM a specific binding protein ?
4) What is the genetic regulatory mechanism for CaM gene expression and how these controls interface with the mechanisms governing plant growth and development.

In this report, we will present some preliminary information about the questions. We have found that CaM is expressed in at least five isomeric forms in soybean. The five genes could be classified into two subfamiles SCaM-I and SCaM-II which encodes structurally distint CaM protein. The members of SCaM-II have novel distinct protein coding sequences and highly diverged from other plant and animal CaM sequences. The expression patterns of these two subfamilies are observed to be different and indicates differential regulation mechanism of these genes. We also show characteristics of the CaM protein produced by E.coli.

II. Nucleotide sequence of soybean CaM cDNA clones and deduced amino acid sequence of the CaM isoforms

The cDNAs encoding soybean CaM was isolated from the cDNA library constructed from the half-apical and half-elongating hypocotyl regions of the 4 day-old etiolated seedlings by using rice genomic CaM clone, cam-2, as a probe DNA (Choi, Y.J., unpublished data). Total 62 positive clones were identified from 3.5 ¥ 104 recombinant phage. Positively hybridizing plaques were characterized

by restriction enzyme mapping and partial nucleotide sequencing of each 5' and 3' terminal. From strongly hybridizing plaques, three different soybean CaM cDNA clones, designated as SCaM-1, -2 and -3, were isolated. Two additional CaM cDNA clones, designated as SCaM-4 and -5, were isolated from very weekly hybridizing plaques. Nucleotide sequencing was performed by both orientation.

All of the isolated clones are full-length clones. SCaM-1, -2 and -3 have nucleotide length around 850 bp. SCaM-4 has an abnormally long 5' untranslated leader sequence, which spans 656 bp long. Nucleotide sequence identity within protein coding sequence among SCaM clones is higher than 80 % except for SCaM-4 and -5 which has nucleotide sequence identity around 72 % with SCaM-1,-2, -3. However, most of the divergence among SCaM-1, -2 and -3 is found in the third positions of codons. The 5' leader and 3' untranslated sequences of each SCaM clone have no significant homology, indicating that these cDNA clones are derived from different gene transcripts. Putative polyadenylation signals, i.e., AATAAA or TATAAA or ATTAAA, are found in SCaM-2, SCaM-3 and SCaM-4. The structure of each SCaM member is summarized in the Table. 1.

Table 1. The structure of each SCaM cDNA.

	SCaM-1	SCaM-2	SCaM-3	SCaM-4	SCaM-5
5' leader	1-78	1-58	1-40	1-56	1-68
CDS	79-528	50-508	41-490	57-1109	1-207,69-518
3' UTR	529-795	509-877	491-875	1110-1421	208-586,529-916
Poly-adenylation signal	ND	TATAAA	AATAAA	ATTAAA	ND
% GC	48.4 %	51.8 %	48.9 %	42.2 %	40.2 %
Total length	795	877	875	1421	916

All sequences are numbered by base pairs and %GC is calculated from protein coding sequences.Abbreviations are: CDS; protein coding sequence, UTR; untranslated region, ND; not determined.

Comparison of the deduced amino acid sequences among SCaM clones clearly indicates that SCaM-4 and -5 are distinct not only from other SCaMs but also from other plant and animal CaM (Fig. 1A). SCaM-1 and -3 have an identical amino sequence which is the same as that of alfalfa CaM (Barnette and Long, 1990). SCaM-2 is differed by two amino acid substitutions to SCaM-1 and -3. However, the amino acid exchanges are E8D (Glu to Asp exchange at 8th residue) and S11A, and seems to have no significant effect on the structure or function of CaM. Furthermore SCaM-2 has an identical amino acid sequences to the barley CaM (Ling et al., 1989). SCaM-1, -2 and -3 show more than 95 % amino acid sequence homology with other plant CaMs. The SCaM-4 and -5 are highly diverged from other SCaMs and plant CaMs. Total 32 amino acid substitutions to SCaM-1 are found in SCaM-4. At least 17 of them are observed to be SCaM-4 specific, which means these substitutions have not been found in all reported plant and animal CaM sequences. SCaM-5 has eighteen amino acid exchanges to SCaM-4. These exchanges include two Asp to Glu exchanges as found in the case of SCaM-1 and -2. One of the noticeable exchanges of amino acid residues found in SCaM-4 and -5 is the Tyr resiudue located in the 3rd Ca2+-binding domain. The Tyr residue at the 3rd Ca2+-binding domain is also found in animal CaM (Roberts et al., 1986). The Tyr residue has been thought to be a candidate for phosphorylation target by src kinase in animal system (Fukami et al, 1986). As clearly indicated from the sequence comparison, the SCaM-4 and -5 are new CaM sub-family. To further investigate the relative relatedness of each SCaM member with other CaM, a pairwise cluster alignment of the deduced amino acid sequences of various CaM was performed based on the strategy UPGMA (Sneath

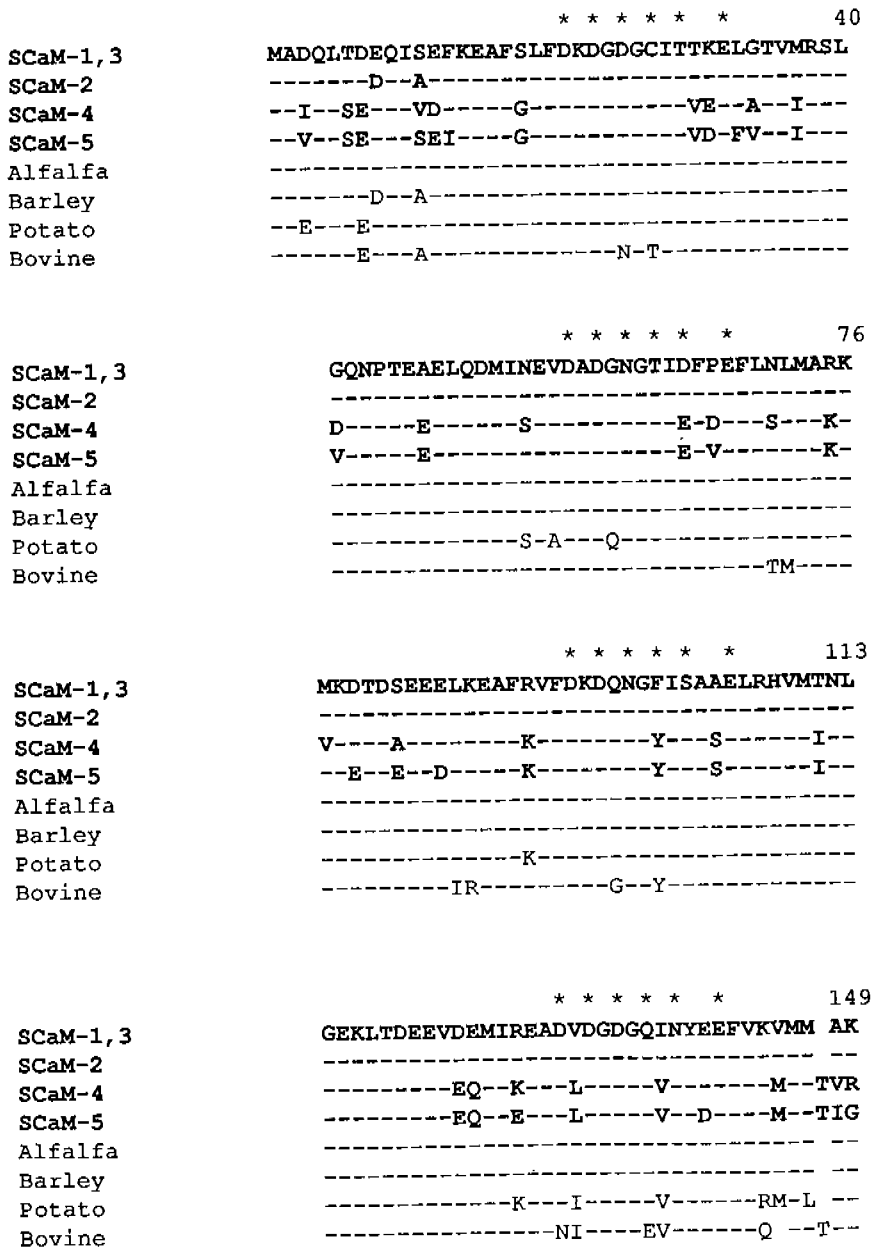

Fig. 1. Comparison of SCaM members with other CaMs.

(A) Comparison of the deduced amino acid sequences of higher plants and bovine CaM polypeptides. Residues marked with asterisks are those which act as Ca2+-binding ligands.

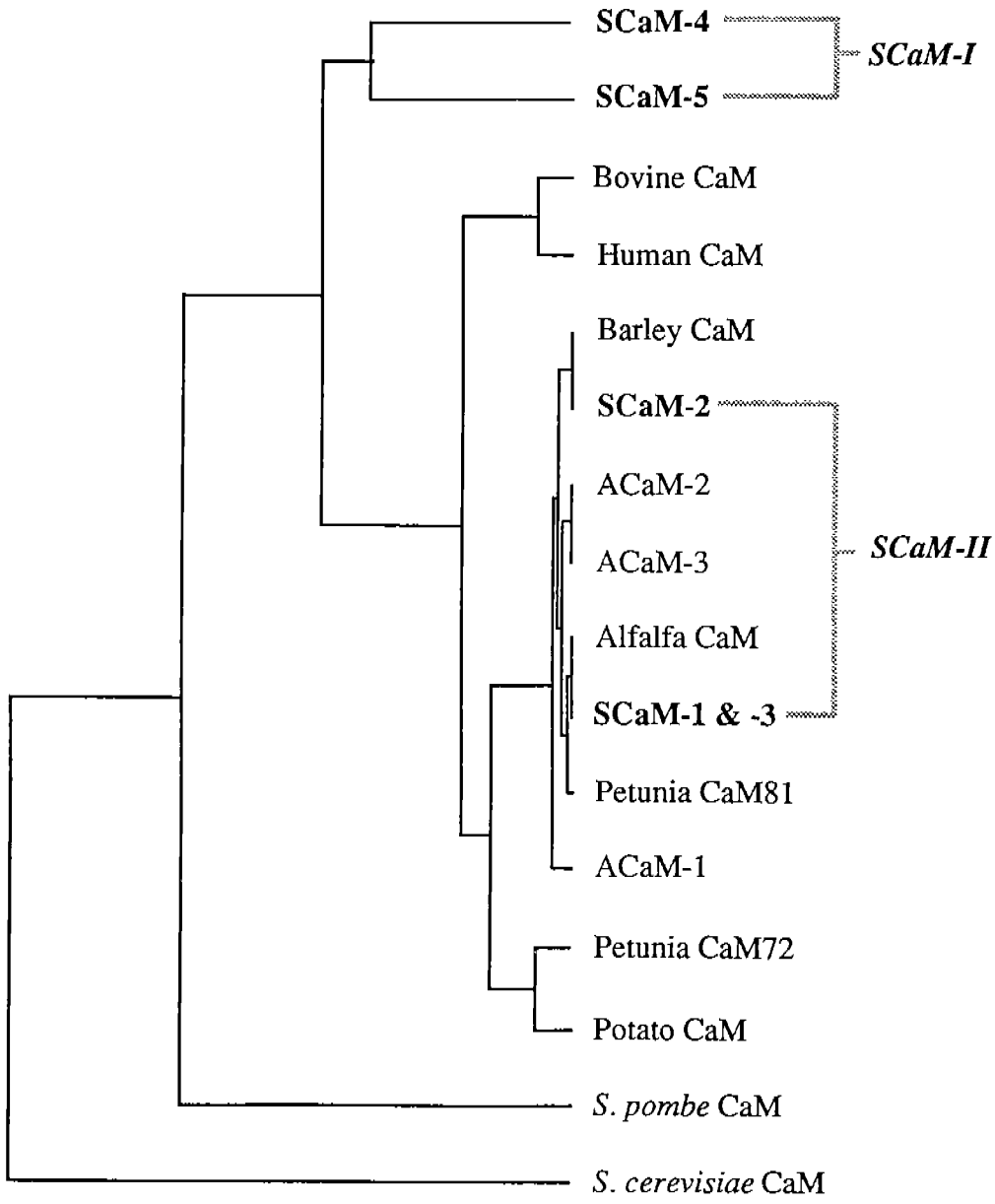

Fig. 1. Comparison of SCaM members with other CaMs.

(B) Dendrogram which shows a relative relatedness of SCaM members to other CaM proteins. A pairwise cluster analysis using each deduced amino acid sequence of CaM was performed with PILEUP software in GCG sequence analysis package.

and Sokal, 1973). The final alignment was illustrated by a dendrogram to clearly indicate the relationship among different CaMs. As shown in Fig. 1(B), SCaM-1, -2 and -3 belong to the CaM group which represents majority of plant CaMs. Bovine and human CaM constitute a different CaM group to plant CaM and the two groups are branched from a common prototype CaM. However, SCaM-4 and -5 do not belong to the two CaM groups and represent a new CaM sub-family. Furthermore SCaM-4 and -5 are branched from the prototype CaM from which animal and plant CaMs evolved. So the SCaM members are divided into two sub-family, designated as SCaM-I and SCaM-II. SCaM-1, -2 and -3 consitute the SCaM-I and the other two are members of the SCaM-II.

III. Analysis of Genomic DNA

Genomic Southern blot analysis was performed to examine the copy number of genes encoding soybean CaM. Soybean genomic DNA digested with EcoRI, BamHI, or HindIII was fractionated on a 0.8 % agarose gel and the Southern blot was probed with 32P-labelled 678 bp EcoRI fragment of SCaM-1 which contained the entire CaM coding sequence. As shown in Fig. 2, more than five strongly hybridizing bands were observed in each digestion with several additional weekly hybridizing bands. These results indicate that soybean CaM is encoded by a small multigene family consist of at least five members. However, it is not clear whether weekly hybridizing bands represent additional CaM genes or other CaM-related EF-hand homologs (Strynadka and James, 1989).

IV. Tissue/Organ-specific expression of SCaM

To determine whether the SCaM genes are expressed in a developmental stage-, tissue-, or organ-specific manner, total RNA or poly (A)+ RNA was

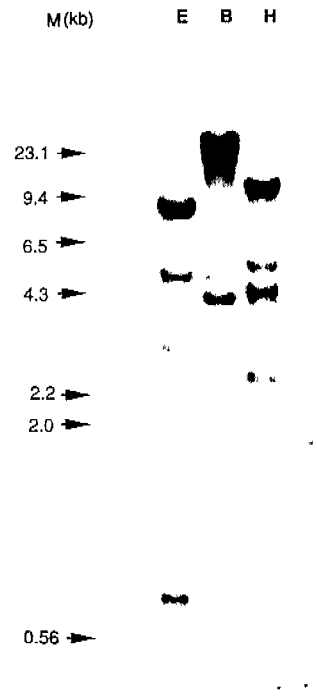

Fig. 2. Genomic Southern blot analysis of SCaM gene in soybean.

Ten micrograms of soybean genomic DNA was digested with EcoRI (E), BamHI (B), or HindIII (H), fractionated on 0.8 % agarose gel and transfered to nylon membrane. A 32P-labelled 678 bp EcoRI fragment of SCaM-1 was used as a probe. M indicates molecular size marked in kb.

isolated from various parts of four day-old etiolated seedlings or from several tissues of three week old mature plant. Northern blot analysis was performed either with coding sequence probe (678 bp EcoRI fragment of SCaM-1) or with each gene-specific probe made from 3' UTR of each SCaM. As shown in Fig. 3 (A), the members of SCaM-I are highly expressed in apical, elongating hypocotyl and plumule tissue sections, and are nearly undetectable in mature hypocotyl and cotyledon tissue sections. These results may indicate the possible involvement of CaM in cell division and elongation and show developmentally regulated expression of SCaM-I members. In mature plant, expression of SCaM-I is abundant in stem and root rather than in leaf. The transcript sizes of SCaM-I members are estimated to be around 0.85 kb and nearly identical to the size of cDNA clone, which indicate these cDNAs are full-length clones. The SCaM-4 was very poorly expressed. The size of major transcript was observed to be around 0.85 kb. As shown in Fig. 3 (B), a faint band of approximately 1.6 kb was also found when 8 μg of poly (A)+ RNA from half-apical and half-elongating region was used. The presence of multiple transcript size may due to the presence of multiple transcription initiation site in a single gene or as a result of duplicate gene transcript. To examine these possiblity, genomic Southern blot using SCaM-4 gene-specific probe was performed and revealed that SCaM-4 has a duplicate gene (data not shown). SCaM-5 is poorly expressed in all seedling tissues examined except for root and nodule of mature plant (data not shown). The expression of SCaM-II is somewhat different to SCaM-I in their transcription level and do not show developmental stage-specific expression pattern.

V. Production of SCaM isoforms in E.coli

The presence of two distint SCaM sub-families in the soybean genome was clearly demonstrated through the analysis of each SCaM members by comparison

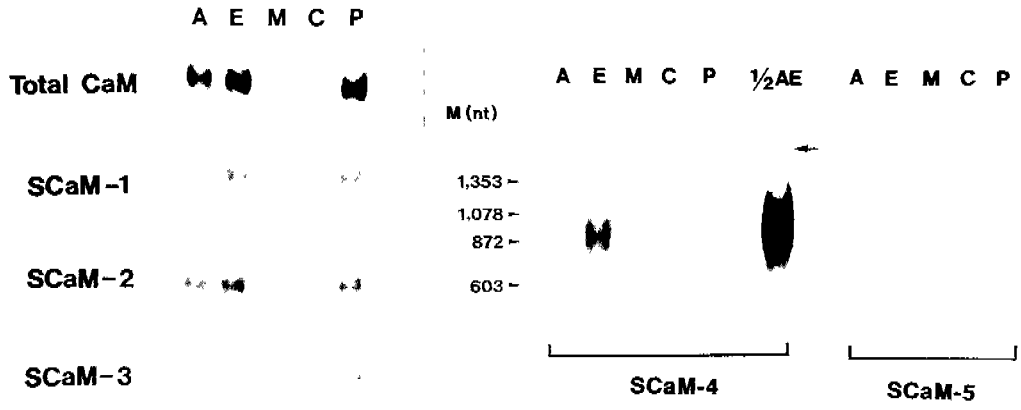

Fig. 3. Tissue-specific expression of SCaMs.

Four day-old etiolated seedlings were divided into apical hypocotyl (A), elongating hypocotyl (E), mature hypocotyl (M), cotyledon (C), and plumule (P) sections and total or poly (A)+ RNA was prepared from each sections.

(A) Expression of SCaM-I members.Twelve micrograms of total RNA were fractionated, transfered onto nylon membranes, and hybridized with each gene-specific probes. For detection of total CaM expression, 678 bp EcoRI fragment of SCaM-1 was used as a probe.

(B) Expression of SCaM-II members. Three micrograms of poly (A)+ RNA was used for analysis except for half-apical and half-elongating hypocotyls (1/2AE) in which 8 μg of poly (A)+ RNA was used. M indicates relative size of transcript determined by the comparison with radiolabelled FX174 DNA digested with HaeIII.

of nucleotide and respective deduced amino acid sequences and Northern blootting. However, the functional differences between the SCaM isoforms encoded by each sub-family remain to be solved. To further investigate the functional differences at the protein level, we construct SCaM expression vectors designed to over-produce SCaM isoforms in E.coli. Two SCaM cDNAs, SCaM-1 and SCaM-4, which are representatives of two SCaM sub-families were modified by subcloning and PCR with a mutagenic oligo primer to produce clonable ends to pET-3d T7 expression vector. The cloning strategy of the SCaM-1 and SCaM-4 in pET-3d T7 expression vector are shown in Fig 4. The SCaM isoforms produced by these expression vectors in E.coli were intact CaM proteins rather than a fusion molecules. The two SCaM isoforms were successfully over-produced in E.coli. SCaM isoforms accumulates up to 10 to 13 % of total soluble protein in E.coli. SCaM-4 isoform produced in E.coli was purified to homogeniety through the standard CaM purification procedure. The purification procedure employes two critical step , i.e, heating at 90 ℃, and phenyl-sepharose hydrophobic interaction column chromatography. With this procedure SCaM -4 was succesfully purified from culture extract of E.coli. SCaM-4 isoform seems to have heat stability and calcium-dependent conformational changes despite of its deduced amino acid sequence divergency to standard bovine CaM. SDS-PAGE analysis of purified SCaM-4 , shows a Ca 2+ - dependent electrophoretic mobility shifting which is one of the diagnostic characteristic of CaM. Purification will provide a good evidence to understand the functional difference of two SCaM sub-families.

VI. Summary

We have cloned five different CaM cDNAs from soybean, and characterize them by nucleotide sequencing and northern blot analysis. Analysis of nucleotide

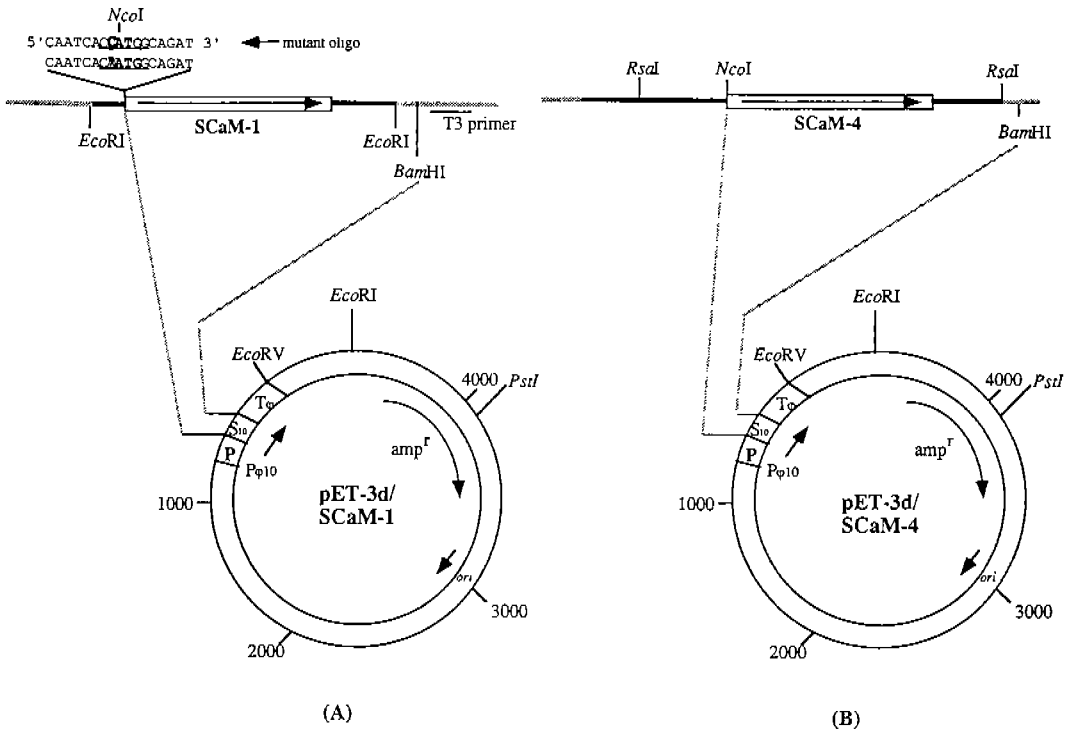

Fig. 4. Construction strategy for the expression of SCaM-1 (A) andSCaM-4 (B) in E.coli by using pET-3d T7 expression vector.

and the deduced amino acid sequences of each SCaMs revealed two structurally different SCaM sub-family, SCaM-I and SCaM-II. Comparison of nucleotide sequences among the members of SCaM-I revealed higher than 83 % identity within protein coding regions. Furthermore, deduced amino acid sequences of SCaM-1 and -3 are identical to that of previously reported alfalfa CaM and SCaM-2 is identical to that of barley CaM.

SCaM-4 and -5 are different from SCaM-I members by more than 30 % nucleotide sequence diversity and 32 amino acid exchanges. SCaM-4 and -5 have a very distinct amino acid sequence even from animal and yeast CaM. Interestingly, SCaM-4 and -5 have a tyrosine residue in the 3rd Ca2+-binding domain which has not been found in plant CaM yet. The Tyr-100 is phosphorylation target by the insulin receptor kinase, which results in the modification of the biological activity of CaM (Laurino et al., 1988). One additional repoprt is phosphorylation of tyrosine residue of CaM by src kinase (Fukami et al., 1986). The finding of this Tyr residue in SCaM-4 and -5 suggest that similar post-translational regulation mechanism of CaM by phosphorylation in plant system may occur. Multiple amino acid sequence analysis of several CaM using UPGMA produces a dendrogram which clearly indicates the SCaM-II constitues another CaM sub-family. Furthermore SCaM-II is seems to be derived from other prototype CaM which constitute most of higher eukaryotic CaM. The members of SCaM-I belongs to the same group which all plant CaM sequences belong to.

In summary, soybean CaM is encoded by a multigene family and we have cloned five members of it. Three of them have similar sequences and properties to other plant CaM. The other two have novel distinct sequences which has not been found in plant and animal CaM. Considering the various role of CaM in calcium-regulated signal transduction pathway, these results may provide a new insight to understand the function and regulation mechanism of CaM. Overproduction of

the SCaM-1 and SCaM-4 in E.coli may facilitate understanding of biochemical and biophysical information of the two distinct SCaM Subfamilies. CaM-binding proteins will be isolated and characterized.

References

Babu, Y.S., Bugg, C.E., Cook, W.J. (1988) J. Mol. Biol. 204: 191-204.

Babu, Y,S., Sack, J.S., Greenhough, T.J., Bugg, C.E., Means, A.R., Cook, W.J. (1985) Nature 315: 37-40.

Barnette, M.J., Long, S.R. (1990) Nucleic Acid Res. 18, 3395.

Fisher, R., Koller, M., Flura, M., Mathews, S., Strehler-Page, M., Krebs, J., Penniston, J.T., Carafoli, E., Strehler, E. (1988) J Biol. Chem. 263: 17055-17062.

Fukami, Y., Nakamura, T., Nakayama, A., Kanehisa, T. (1986) Proc. Natl. Acad. Sci. USA 83: 4190-4193.

Hepler, P.K., Wayne, R.O. (1985) Annu. Rev. Plant. Physiol. 36: 397-439.

Jena, P.K., Reddy, A.S.N., Poovaiah, B.W. (1989) Proc. Natl. Acad. Sci. USA 86: 3644-3648.

Laurino, J.P., Colca, J.R., Pearson, J.D., DeWard, D.B., McDonard, J.M.(1988) Arch. Biochem. Biophys. 265: 8-21.

Leonard, R.T., Helpler, P.K. (1990) Calcium in Plant Growth and Development, American society of Plant Physiologists, Maryland.

Ling, V., Zielinski, R.E. (1989) Plant Physiol. 90: 714-719.

Ling, V., Perera, I., Zielinski, R.E. (1991) Plant Physiol. 96, 1196-1202.

Nojima, H., Kishi, K., Sokabe, H. (1987) Mol. Cell Biol. 7, 1873-1880.

Perera, I.Y., Zielinski, R.E. (1992) Plant Mol. Biol. 19: 649-664.

Poovaiah, B.W., Reddy, A.S.N. (1987) CRC Crit. Rev. Plant Sci. 6: 47-103.

Roberts, D.M., Lukas, T.J., Watterson, D.M. (1986) CRC Crit. Rev. Plant Sci. 4: 311-339.

Sneath, P.H.A., Sokal, R.R. (1973) Numerical taxanomy, pp.230-234. W.H. Freeman, San Franscisco, California, USA

Strynadka, N.C.J., James, N.G.(1989) Annu. Rev. Biochem. 58: 951-998 .