J. Biochem. 128, 161–166 (2000)

# Molecular Cloning and Functional Expression of cDNA Encoding the Cysteine Proteinase Inhibitor with Three Cystatin Domains from Sunflower Seeds[1]

**Yoshiaki Kouzuma,[2] Hideko Inanaga, Keiko Doi-Kawano, Nobuyuki Yamasaki, and Makoto Kimura**

*Laboratory of Biochemistry, Department of Bioscience and Biotechnology, Faculty of Agriculture, Graduate School, Kyushu University, Hakozaki 6-10-1 Higashi-ku, Fukuoka 812-8581*

Received March 8, 2000; accepted May 2, 2000

**Two cysteine proteinase inhibitors, cystatins Sca and Scb, were previously isolated from sunflower seeds [Kouzuma *et al. J. Biochem.* 119 (1996) 1106–1113]. A cDNA clone encoding a novel phytocystatin with three repetitive cystatin domains was isolated from a cDNA library of sunflower seeds using the Sca cDNA fragment as a hybridization probe. The cDNA insert comprises 1,093 bp and encodes 282 amino acid residues. The deduced amino acid sequences of the domains are highly similar to each other (66–81%), sharing 65–90% identical residues with Sca. The cDNA was expressed in *Escherichia coli* cells, and then the recombinant sunflower multicystatin (SMC) was purified and its inhibitory activity toward papain was examined. SMC exhibited strong inhibitory activity toward papain, with a stoichiometry of 1:3, indicating that each cystatin domain independently functions as a potent cysteine proteinase inhibitor. Proteolysis of SMC with Asn-specific proteinase suggested that post-translational processing by an Asn-specific proteinase may give rise to mature Sca-like phytocystatins.**

**Key words: cDNA cloning, cysteine proteinase, overexpression, phytocystatin, sunflower.**

A large number of cysteine proteinase inhibitors have been found in various plant organs, and information on their amino acid sequences has recently been obtained through either protein or cDNA sequencing. These cysteine proteinase inhibitors of plant origin have been grouped into a fourth cystatin family, the "phytocystatins," based on the sequence similarity and absence of disulfide bonds. Phytocystatins have common structural properties as follows: (i) they are single polypeptide chains with molecular masses ranging from 10 to 12 kDa, and (ii) they share three conserved sequence motifs: Gly in the vicinity of the N-terminal region, Gln-Xaa-Val-Xaa-Gly in the first hairpin loop, and Pro-Trp in the second hairpin loop (*1–9*). Although their physiological functions remain obscure, the gene expression pattern, as well as the properties of the proteins suggest that they may play a role in the plant's defense system.

We earlier isolated two distinct phytocystatins, Sca and Scb, from sunflower seeds and characterized their inhibitory specificities as well as their primary structures (*10*). Sca and Scb are single polypeptide chains comprising 83 and 101 amino acid residues, respectively, and share 43% identical residues. They exhibit inhibitory activity toward papain with $K_i$ values of $5.6 \times 10^{-9}$ and $1.7 \times 10^{-10}$ M, respectively, and, in addition, Scb exhibits inhibitory activity toward cathepsins B, L, and H. Although both phytocystatins in general have structural features in common with other phytocystatins, they have characteristic features as compared with other phytocystatins. First, Sca lacks about 10–25 amino acid residues at the N-terminus as compared with other phytocystatins, resulting in the absence of Gly in the N-terminal region. Furthermore, Scb is a highly basic protein with a pI value of 9.6, and has an additional amino acid sequence in the middle of the molecule as compared with those of other phytocystatins.

For a better understanding of the structure-function relationship of the two phytocystatins from sunflower seeds, we cloned a cDNA encoding Scb and functionally expressed it in *Escherichia coli* cells (*11*). Furthermore, a mutagenesis study on Scb demonstrated the importance of the N-terminal Gly residue for the inhibitory activity toward cysteine proteinases (*11*). In the present study, we amplified cDNA fragments encoding Sca and used them as probes for screening the full length of its cDNA clone. Unexpectedly, a cDNA clone encoding a novel phytocystatin consisting of

[1] This study was partly supported by a Grant-in-Aid for the Encouragement of Young Scientist from the Japan Society for the Promotion of Science. The DNA sequence in this paper has been submitted to the DDBJ (DNA Data Bank of Japan) database under accession number AB039673.

[2] To whom correspondence should be addressed. Tel: +81-92-642-4215, Fax: +81-92-642-2864, E-mail: kouzuma@agr.kyushu-u.ac.jp

Abbreviations: C-, carboxy-; IPTG, isopropyl-β-D-thiogalactopyranoside; $K_i$, inhibition constant; N-, amino-; PCR, polymerase chain reaction; PMC, potato multicystatin; Pyr-, L-pyroglutamyl; RP-HPLC, reverse-phase high-performance liquid chromatography; rSMC, recombinant sunflower multicystatin; Sca, sunflower cystatin a; Scb, sunflower cystatin b; SDS-PAGE, sodium dodecylsulfate polyacrylamide gel electrophoresis; SMC, sunflower multicystatin; TFA, trifluoroacetic acid; Xaa, not conserved amino acid residue.

three repetitive cystatin domains was discovered. We report here the molecular cloning and functional expression of the cDNA encoding the novel sunflower multicystatin (SMC) in *E. coli* cells. We also discuss the physiological relationship between SMC and Sca.

## MATERIALS AND METHODS

*Materials*—A sunflower maturing seed cDNA library was constructed previously (*11*). The oligonucleotides used in this study were purchased from Amersham-Pharmacia Biotech. *Taq* DNA polymerase, a random primer DNA labeling kit, a DNA ligation kit, and asparaginylendopeptidase were obtained from Takara Shuzo, and used as recommended by the supplier. Restriction endonucleases and DNA modifying enzymes were purchased from either Toyobo or MBI Fermentas. The plasmid vectors used in this work were as follows: pUC18 from MBI Fermentas, pGEM™-T EASY vector from Promega, and pET-22b expression vector from Novagen. Papain and pronase were obtained from Sigma Chemicals. All other chemicals were of analytical grade for biochemical use.

*Preparation of the Sca Gene Fragment by PCR*—Four oligonucleotides were synthesized for use as primers corresponding to the amino acid sequence of the Sca protein: forward primer 1, 5′-TTCGCCGTCGACGARCAYAAYAARAARC-3′ (FAVDEHNKK at positions 9–17); back primer 1, 5′-TCYTTRAAGCTTYTCCCANGGYTTNACCC-3′ (WVKPWENFKE at positions 60–69); forward primer 2, 5′-TGGCNGAATTCGCNGTNGAYGARCA-3′ (LARFAVDEH at positions 7–15); and back primer 2, 5′-GCNGCRTCNACNGGYTTGAATTCYTG (QEFKPVDAA at positions 72–80), where R represents A and G; Y, C, and T; N, A, C, G, and T. PCR was performed with sunflower developing seed cDNAs, as a template, and an automated thermal circular (ASTEC, Program Temp Control System PC-700). The PCR products were purified on a gel and then subcloned into the pGEM™-T EASY vector. After confirmation of its sequence, the resulting fragment was used as a hybridization probe to screen the cDNA library.

*Screening and Nucleotide Sequencing*—The sunflower developing seed cDNA library was screened by plaque hybridization utilizing the PCR product labeled by the random priming method, as described in Ref. *12*. Through the second and third plaque hybridizations, a positive clone was isolated and phage DNA was prepared from it. The cDNA insert was excised with *Eco*RI and then subcloned into plasmid vector pUC18, and then sequencing was performed with a DNA Sequencer DSQ-1000 (Shimadzu) using a Thermo Sequenase Cycle Sequencing kit (Amersham-Pharmacia Biotech).

*Expression of SMC in E. coli Cells*—To obtain the SMC cDNA fragment with restriction sites *Nde*I and *Bam*HI at the 5′- and 3′- ends, respectively, PCR was performed with the SMC cDNA as a template with forward primer 5′-CTAACATATGTCACTTGTTGG-3′ and back primer 5′-GGATTGGATCCTCAATAAAGGTGCTTGAA-3′. The PCR products were ligated into the pGEM™-T EASY vector. After confirmation of the DNA sequence, the DNA fragment encoding SMC was excised by digestion with *Nde*I and *Bam*HI, and then ligated into expression vector pET-22b, previously digested with the same enzymes. The resulting plasmid, pET-SMC, was introduced into *E. coli* BL21(DE3), and then the recombinant protein was induced with 1 mM IPTG according to the supplier's instructions.

*SDS–Polyacrylamide Gel Electrophoresis*—Overproduction of SMC in *E. coli* cells was analyzed by SDS-PAGE on 15% acrylamide gels (*13*).

*Purification of the Recombinant SMC*—Purification of the recombinant protein was performed as previously described (*11*). The insoluble inhibitor proteins (inclusion bodies) in *E. coli* were dissolved in 30 mM Tris-HCl buffer (pH 7.5) containing 30 mM NaCl and 5 M guanidine hydrochloride (Gdn-HCl). The Gdn-HCl concentration in the protein solution was decreased by dialysis and the solution was finally dialyzed against 30 mM Tris-HCl buffer (pH 7.5) containing 30 mM NaCl. To avoid precipitation of the protein, an appropriate concentration (0.1 to 1 M) of arginine hydrochloride was added to the dialysis buffer. Proteins were purified by gel filtration on a Sephadex G-75 column, followed by RP-HPLC on a YMC-GEL C4 (300 Å) column (4.6 × 250 mm). The N-terminal amino acid sequence of the protein was determined with a gas phase sequencer PSQ-1 (Shimadzu).

*Protease Inhibition Assay*—The papain inhibition assay was performed as described in Ref. *10* using Pyr-Phe-Leu-*p*-nitroanilide. The concentrations of the recombinant proteins were determined by amino acid analysis. Active site titration of papain was performed using E-64 (*14*).

*Proteolysis of SMC*—Digestion of the recombinant SMC protein was performed using asparaginylendopeptidase at an enzyme–substrate ratio of 1:400 to 1,000 at 25°C for 24 h in 50 mM acetate buffer, pH 5.0, containing 10 mM DTT and 1 mM EDTA.

## RESULTS

*Cloning and Nucleotide Sequencing*—To obtain a hybridization probe for the screening of a cDNA clone encoding Sca, we amplified cDNA fragments with two sets of degenerated primers based on the amino acid sequences of Sca, as described under "MATERIALS AND METHODS." Amplification with 1 and 2 sets of primers produced 180 and 200 bp cDNA fragments, respectively. These were the expected sizes based on the positions of primers in the amino acid sequence in Sca. Nucleotide sequence analysis of the PCR products revealed that the deduced amino acid sequence of the 180 bp fragment was in perfect agreement with the protein sequence of Sca, while that of the 200 bp fragment exhibited high, but not perfect, sequence identity with that of Sca, suggesting the presence of isophytocystatins of Sca. Each fragment was used to screen Sca cDNA clones. Screening of approximately $2.4 \times 10^5$ independent recombinant clones from a λgt11 cDNA library using the 180 or 200 bp fragment gave an identical positive clone. The insert of the positive clones was subcloned into the pUC 18 vector and sequenced using a DNA sequencer DSQ-1000.

The nucleotide sequence of the cDNA and its deduced amino acid sequence are shown in Fig. 1. The cDNA clone comprises 1,093 bp with an open reading frame of 849 nucleotides, this being extremely longer than that predicted from the Sca protein sequence. Comparison of the amino acid sequence deduced from the cDNA sequence with that of Sca revealed that this cDNA encodes a novel phytocystatin with three repetitive cystatin domains per molecule. The deduced amino acid sequences of the three domains as

```
ATT CGC GGC CGC TAA AAG ATG TCA CTT GTT GGA GGA TTT ACA GAA GTA AAG GAC TTT GCA     60
                        M   S   L   V   G   G   F   T   E   V   K   D   F   A      14

AAC AGC ATT GTG ATC GAC GAT CTC GCT CGA TTC GCC GTC GAT GAA TAC AAC AAG AAG CAG    120
N   S   I   V   I   D   D   L   A   R   F   A   V   D   E   Y   N   K   K   Q      34

AAT ACC CTG CTG GAA TTT AGG AAG GTG CTG AAT GCA AAG GAG CAG ATA GTT TCT GGT ACA    180
N   T   L   L   E   F   R   K   V   L   N   A   K   E   Q   I   V   S   G   T      54

TTG TAT TAT ATC ACA CTT GAT GCA GCT AAT GGT GGT ATC ATA AAG ACT TAT GAA GCC AAG    240
L   Y   Y   I   T   L   D   A   A   N   G   G   I   I   K   T   Y   E   A   K      74

GTT TGG GTT AAG AAA TGG GAA AAC TTA AAA GAA TTG CAA GAG TTC AAG CCT GTT GAT GCT    300
V   W   V   K   K   W   E   N   L   K   E   L   Q   E   F   K   P   V   D   A      94

GCC ACC TCA GTT ATT GGA GGA ATT ACA GAA GTA AAG GAC TTT GCA AAC AGC CTT GAG ATC    360
A   T   S   V   I   G   G   I   T   E   V   K   D   F   A   N   S   L   E   I     114

GAA GAT CTC GCT CGA TTC GCC GTA GAT GAA CAC AAC AAG AAG CAG AAT ACC CTG CTG GAA    420
E   D   L   A   R   F   A   V   D   E   H   N   K   K   Q   N   T   L   L   E     134

TTT GGG AAA GTA CTG AAT GCA AAG GAG CAG ATA GTT GCT GGT AAA TTG TGT TAT ATC ACA    480
F   G   K   V   L   N   A   K   E   Q   I   V   A   G   K   L   C   Y   I   T     154

CTT GAA GCA ACT GAT GGT GGT GTC AAA AAG ACT TAT GAA GCC AAG GTT TGG GTT AAG CCA    540
L   E   A   T   D   G   G   V   K   K   T   Y   E   A   K   V   W   V   K   P     174

TGG GAA AAC TTC AAA GAA TTG CAG GAG TTC AAG CCT GTT GAT GCT GCC ACC CCA GTC ATT    600
W   E   N   F   K   E   L   Q   E   F   K   P   V   D   A   A   T   P   V   I     194

GGA GGA ATT ACA GAA GTA AAG GAC TTT GCA AAC AGC CTT GTG ATC GAC GAT CTC GCT CGA    660
G   G   I   T   E   V   K   D   F   A   N   S   L   V   I   D   D   L   A   R     214

TTC GCC GTC GAT GAA TAC AGC AAG AAG CAG AAT ACC CTG CTG GAA TTT GAG AGG GTA CTG    720
F   A   V   D   E   Y   S   K   K   Q   N   T   L   L   E   F   E   R   V   L     234

GAT GCA AAG CAG CAG ATA GTT GCT GGT ACA ATG TAT TAT TTC ATA CTT GAA GCA ACT GTT    780
D   A   K   Q   Q   I   V   A   G   T   M   Y   Y   F   I   L   E   A   T   V     254

GGT GGT GTC AAA AAC ACT TAC GTA GCC AAG GTT TTG CTT AAG CCG GAC AAC TCC AAA GAA    840
G   G   V   K   N   T   Y   V   A   K   V   L   L   K   P   D   N   S   K   E     274

TTG AAG GAA TTC AAG CAC CTT TAT TGA AGC AGT AAT GGT GTA TGT TCT ATG ATG CAG GCG    900
L   K   E   F   K   H   L   Y   *                                                 282

AGA GTG CCA GGA GAC CGC TAC CAG GGT GTG AAC CTT AAT CTT ATA AAT AAT ATC GTT TTG    960

GTT AAT AAT AAA CGG TTC TTG CAT TAC ATT GCA AGT TTG TGC CGC TGC TTA TGT TTG CTT   1020

AAA TAA TTT CGT TTG GTT AGT AAT AAA CGG TTC TTG CAT TAC ATT CAA GTT AAA AAA AAA   1080

AAA AAA AAA AAA A                                                               1093
```

Fig. 1. **Nucleotide and deduced amino acid sequences of the cDNA clone.** The predicted amino acid sequence is shown below the nucleotide sequence. The termination codon is indicated by an asterisk. The conservative Gln-Xaa-Val-Xaa-Gly sequences are underlined. The putative poly (A) signals are enclosed in boxes.

well as that of Sca are aligned so as to give maximum homology in Fig. 2. The three repetitive domains are homologous to each other, having 66–81% identical residues; domains 1 and 2 exhibit the highest homology (81%), while domains 1 and 3 exhibit the lowest homology (66%), and also share 65–90% identical residues with Sca. The second domain is the most homologous to Sca: there are identical 72 residues in the two sequences, that is, 90% of the residues compared. In contrast, the third domain is the most divergent from Sca (65%). Essential structural motifs in cystatin families: Gly in the N-terminal region, Gln-Xaa-Val-Xaa-Gly in the middle region, and the Trp residue in the C-terminal region, are conserved in all three domains except for the third one, the conserved Trp being absent in the third domain. It should also be noted that a single Cys (Cys-151) is present in the second cystatin domain. We refer to this novel cystatin as sunflower multicystatin, SMC.

*Expression and Inhibitory Activity of the Recombinant Protein*—To characterize the properties of SMC, we attempted the expression of SMC in *E. coli*. Expression plasmid pET-SMC was introduced into *E. coli* strain BL21-(DE3) and then expression of the protein was induced by adding IPTG to the culture medium. Subsequently, an extract of *E. coli* expressing SMC was prepared by sonication, and the soluble and insoluble fractions were analyzed by SDS-PAGE. This analysis showed that the recombinant SMC (rSMC) predominantly occurred in the form of inclusion bodies (Fig. 3).

The rSMC was solubilized, refolded as described under

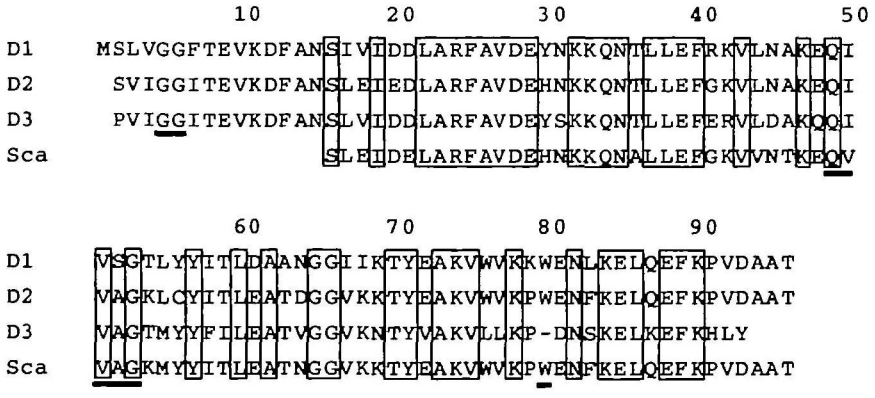


Fig. 2. **Comparison between the deduced amino acid sequences of the three phytocystatin-domains of SMC and sunflower cystatin Sca.** The three cystatin domains (D1–D3) of SMC and Sca are aligned to maximize similarities. The identical residues are enclosed in boxes. The three regions involved in papain-binding are underlined.

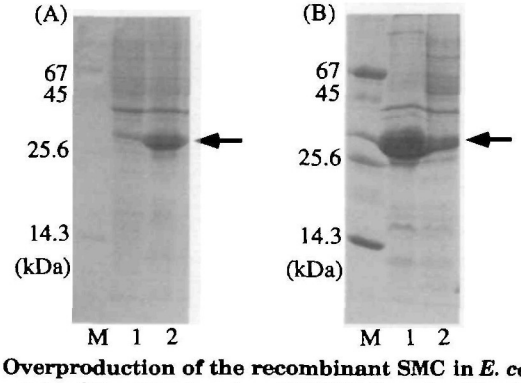


Fig. 3. **Overproduction of the recombinant SMC in *E. coli*.** (A) Total protein of *E. coli* harbouring pET-SMC. Lane 1, without induction; lane 2, induction by IPTG. (B) Soluble or insoluble fraction of *E. coli* after sonication. Lane 1, insoluble fraction; lane 2, soluble fraction. M, molecular weight markers: Bovine serum albumin (67 kDa), ovalbumin (45 kDa), α-chymotrypsinogen (25 kDa), and lysozyme (14 kDa). Arrows indicate the position of the expressed SMC protein.

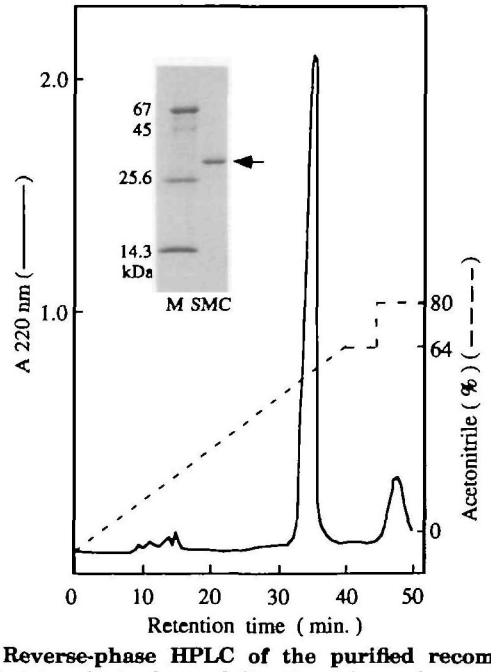


Fig. 4. **Reverse-phase HPLC of the purified recombinant SMC.** Proteins obtained on gel filtration column chromatography were subjected to RP-HPLC on a Wakosil $5C_{18}$ column (4 × 250 mm) with 0.1% TFA at the flow rate of 1 ml/min. The protein was eluted with a linear gradient of 0 to 64% acetonitrile in 0.1% TFA. The eluted protein fraction (black bar) was recovered and subjected to SDS-PAGE. The arrow indicates the position of the recombinant SMC protein.

"MATERIALS AND METHODS," and then purified by RP-HPLC on a YMC-GEL C4 column (Fig. 4). The major peak material exhibited a strong papain-inhibitory activity and the molecular mass was estimated to be 32 kDa on SDS-PAGE. This value was in agreement with the calculated molecular mass of SMC (31,831 Da). The yields were about 20 mg/liter of induced culture. Sequence analysis of the rSMC gave the N-terminal sequence: Phe-Thr-Glu-Val-Lys-Asp, indicating that rSMC was processed at the C-terminal side of a conserved glycine (Gly-6) by endogenous proteinases in *E. coli*.

The inhibitory activity of rSMC toward papain is shown in Fig. 5. The SMC could inhibit papain with a stoichiometry of 1:3, suggesting that each of the three phytocystatin domains in the SMC molecule independently functions as a cysteine proteinase inhibitor. The inhibition constant, $K_i$, of SMC for papain was calculated to be $4.0 \times 10^{-8}$ M, assuming that SMC has three equivalent reactive sites for papain. This $K_i$ value was almost the same as that of Sca ($5.6 \times 10^{-9}$ M). The reduction of the affinity may be caused by the steric hindrance between SMC and the papain molecule.

*Characterization of the Recombinant SMC*—Comparison of the deduced amino acid sequence of SMC with that of Sca suggested that mature Sca-like phytocystatins could be produced through cleavage at the conserved Asn-Ser sites in the N-terminal regions of each of the three cystatin domains in SMC. Since the proteinases responsible for cleavage at an Asn-Xaa linkage in plants are well known (*15*), this finding suggested that post-translational processing of a large protein like SMC by an Asn-specific proteinase may give rise to Sca-like phytocystatin molecules. This assumption was addressed to see whether or not treatment of the recombinant SMC with a proteinase would result in fragmentation of the SMC into about 10 kDa polypeptides similar to the Sca protein. When SMC was digested with an Asn-specific proteinase with various SMC/enzyme ratios at 37°C, increases of polypeptides estimated on a gel to be 20 and 10 kDa in size were observed (Fig. 6). Thus, the recombinant SMC was digested and the resulting peptides, which were separated by SDS-PAGE and blotted onto PVDF membranes, were sequenced using a gas-phase sequencer.

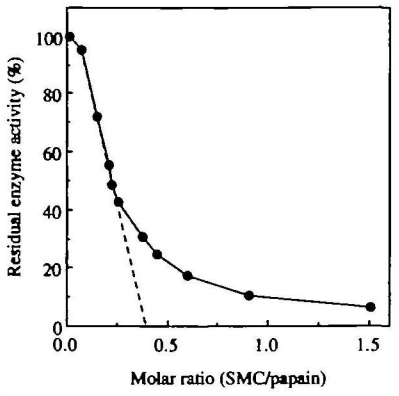


Fig. 5. **Inhibition of papain by increasing amounts of the recombinant SMC.** A fixed amount of papain was mixed with increasing amounts of the inhibitors, and then residual enzyme activities were assayed.

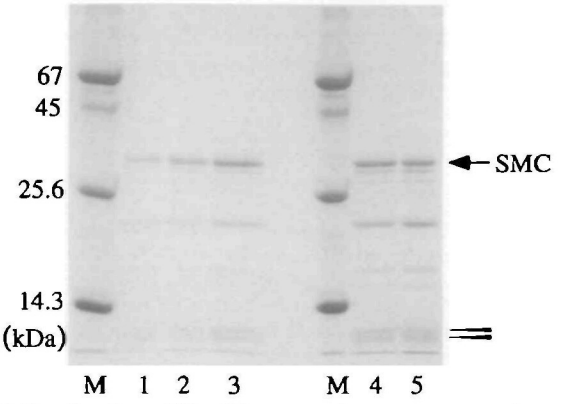


Fig. 6. **Production of Sca-like cystatins on proteolytic digestion of SMC.** The recombinant SMC protein (lane 1, 3.4 μg; lane 2, 8.5 μg; lanes 3–5; 17 μg) was digested with asparaginylendopeptidase (lanes 1–3, 0.5 μU; lane 4, 1 μU; lane 5, 2 μU) at 25°C for 24 h. The arrows indicate Sca-like cystatins.

The N-terminal sequences of polypeptides (~10 kDa) were found to be Ser-Ile-Val-Ile and Ser-Leu-Val-Ile, which were similar to that of Sca. A similar result was obtained on pronase digestion of the SMC molecule (data not shown). These results suggested that linker regions between the individual Sca-like domains within the SMC molecule may form surface-exposed loops, and that Sca-like proteins may be produced through post-translational processing of a large precursor, the SMC molecule.

Next, we examined whether or not the SMC molecule is present in mature seeds by immunoblot analysis with polyclonal anti-SMC antibodies. The results showed that there were no bands corresponding to approximately 32 kDa were recognized on immunoblots of the seed extracts (data not shown).

## DISCUSSION

Cystatins inactivate their target proteinases by trapping them in a reversible, tight equimolar complex. The crystal structures of chicken cystatin, and of a complex between cystatin B and papain have provided considerable information on the nature of these interactions. A crystallographic study of chicken cystatin suggested that the proteinase binding site comprises three regions, the N-terminal region around Gly-9, the Gln-Leu-Val-Ser-Gly sequence at residues 53–57, and the region around Trp-104 (*16*). This proposal was confirmed by a crystallographic study on a complex composed of cystatin B and papain, in which cystatin B interacts with papain to form a wedge composed of three motifs corresponding to those found in the structure of chicken cystatin (*17*). Site-directed mutagenesis of cystatin C further showed that Trp-106 is involved in the binding to proteinases and its substitution by Gly reduced the affinity for papain by 900-fold (*18*). In the present study, we isolated a cDNA clone encoding the novel phytocystatin SMC from developing sunflower seeds. SMC possesses three repetitive cystatin domains per molecule. Comparison of the deduced amino acid sequences of the three repetitive cystatin domains in SMC revealed the absence of the conserved Trp residue in the third domain. The lack of this Trp residue has little effect on its papain-inhibitory activity, because the three cystatin domains in the recombinant SMC protein can each independently function as a potent cysteine proteinase inhibitor (Fig. 5), and the $K_i$ value ($4.0 \times 10^{-8}$ M) of SMC toward papain is relatively the same as that ($5.6 \times 10^{-9}$ M) of Sca. This finding indicated that the Trp plays no essential role in the inhibitory activity of the Sca-like phytocystatin toward papain. This result is consistent with the result we obtained on recent mutagenesis of Scb: the replacement of Trp-85 in Scb had a little influence on the inhibitory activity toward papain (Doi-Kawano *et al.*, unpublished results).

Cysteine proteinase inhibitors with tandem repeats of the cystatin domains in their molecules are known as kininogens, members of the animal cystatin family (*19*). Kininogens, which are glycoproteins with molecular masses of 50–114 kDa, comprise three cystatin domains in the N-terminal region and an additional extension in the C-terminal region, from which a peptide kinin is derived. In contrast, SMC found in sunflower seeds has no additional sequence in the molecule, indicating that it plays no role aside from its function as a cysteine proteinase inhibitor. In the phytocystatin family, a potato multicystatin (PMC) was reported (*20–22*). This PMC is a phytocystatin found in potato tubers in the form of proteinaceous crystals. It consists of a polypeptide of about 85 kDa and inhibits eight papain molecules simultaneously. Genomic DNA sequence analysis indicated that PMC has eight repeats of cystatin domains and no extra sequences in the molecule. Immuno blot analysis of potato tubers revealed a group of protein bands corresponding to approximately 85 kDa, but no evidence for major accumulation of smaller $M_r$ fragments was obtained. These findings suggested that PMS is dominantly present in potato tubers. A similar observation was reported for the serine proteinase inhibitors (PIs) in *Nicotiana alata* stigmas (*23*). It was described that a long precursor (42 kDa) containing five tandem repeats of the PI unit is translated and, unlike PMC, processed in the tissue, resulting in five homologous PIs with a $M_r$ of 6 kDa. Since the intact SMC or smaller SMC fragments were not detected for the mature sunflower seeds, it could not be concluded that SMC was a long precursor for Sca-like phytocystatins. It was, however, found that SMC was readily processed into phytocystatin molecules, similar to Sca, through proteoly-

sis. The present study thus suggested that post-translational processing may be a potential mechanism for the production of proteinase inhibitors with diverse proteinase specificities in the plant kingdom.

## REFERENCES


1. Abe, K., Emori, Y., Kondo, H., Suzuki, K., and Arai, S. (1987) Molecular cloning of a cysteine proteinase inhibitor of rice (Oryzacystatin). *J. Biol. Chem.* **262**, 16793–16797
2. Kondo, H., Abe, K., Nishimura, I., Watanabe, H., Emori, Y., and Arai, S. (1990) Two distinct cystatin species in rice seeds with different specificities against cysteine proteinases. *J. Biol. Chem.* **265**, 15832–15837
3. Abe, M., Abe, K., Kuroda, M., and Arai, S. (1992) Corn kernel cysteine proteinase inhibitor as a novel cystatin superfamily member of plant origin. *Eur. J. Biochem.* **209**, 933–937
4. Fernandes, K.V.S., Sabelli, P.A., Barratt, D.H.P., Richardson, M., Xavier-Filho, J., and Shewry, P.R. (1993) The resistance of cowpea seeds to bruchid beetles is not related to levels of cysteine proteinase inhibitors. *Plant Mol. Biol.* **23**, 215–219
5. Song, I., Taylor, M., Baker, K., and Bateman Jr, R.C. (1995) Inhibition of cysteine proteinases by *Carica papaya* cystatin produced in *Escherichia coli*. *Gene* **162**, 221–224
6. Kimura, M., Ikeda, T., Fukumoto, D., Yamasaki, N., and Yonekura, M. (1995) Primary structure of a cysteine proteinase inhibitor from the fruit of avocado (*Persea americana* Mill). *Biosci. Biotechnol. Biochem.* **59**, 2328–2329
7. Li, Z., Sommer, A., Dingermann, T., and Noe, C.R. (1996) Molecular cloning and sequence analysis of a cDNA encoding a cysteine proteinase inhibitor from *Sorghum bicolor* seedlings. *Mol. Gen. Genet.* **251**, 499–502
8. Lim, C.O., Lee, S.I., Chung, W.S., Park, S.H., Hwang, I., and Cho, M.J. (1996) Characterization of a cDNA encoding cysteine proteinase inhibitor from Chinese cabbage (*Brassica campestris* L. ssp. *pekinensis*) flower buds. *Plant Mol. Biol.* **30**, 373–379
9. Rogelj, B., Popovic, T., Ritonja, A., Strukelj, B., and Brzin, J. (1998) Chelidocystatin, a novel phytocystatin from *Chelidonium majus*. *Phytochemistry* **49**, 1645–1649
10. Kouzuma, Y., Kawano, K., Kimura, M., Yamasaki, N., Kadowaki, T., and Yamamoto, K. (1996) Purification, characterization, and sequencing of two cysteine proteinase inhibitors, Sca and Scb, from sunflower (*Helianthus annuus*) seeds. *J. Biochem.* **119**, 1106–1113
11. Doi-Kawano, K., Kouzuma, Y., Yamasaki, N., and Kimura, M. (1998) Molecular cloning, functional expression, and mutagenesis of cDNA encoding a cysteine proteinase inhibitor from sunflower seeds. *J. Biochem.* **124**, 911–916
12. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
13. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685
14. Zucker, S., Buttle, D.J., Nicklin, M.J.H., and Barrett, A.J. (1985) The proteolytic activities of chymopapain, papain, and papaya proteinase III. *Biochim. Biophys. Acta* **828**, 196–204
15. Abe, Y., Shirane, K., Yokosawa, H., Matsushita, H., Mitta, M., Kato, I., and Ishii, S. (1993) Asparaginyl endopeptidase of jack bean seeds. Purification, characterization, and high utility in protein sequence analysis. *J. Biol. Chem.* **268**, 3525–3529
16. Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., and Turk, V. (1988) The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *EMBO J.* **7**, 2593–2599
17. Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., and Turk, V. (1990) The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *EMBO J.* **9**, 1939–1947
18. Björk, I., Brieditis, I., Raub-Segall, E., Pol, E., Håkansson, K., and Abrahamson, M. (1996) The importance of the second hairpin loop of cystatin C for proteinase binding. Characterization of the interaction of Trp-106 variants of the inhibitor with cysteine proteinases. *Biochemistry* **35**, 10720–10726
19. Barret, A.J., Rawlings, N.D., Davis, M.E., Machleidt, W., Salvesen, G., and Turk, V. (1986) *Proteinase Inhibitors*, pp. 515–569, Elsevier Science Publishers BV, Amsterdam, New York
20. Rodis, P. and Hoff, J.E. (1984) Naturally occurring protein crystals in the potato. *Plant Physiol.* **74**, 907–911
21. Waldron, C., Wegrich, L.M., Merlo, P.A.O., and Walsh, T.A. (1993) Characterization of a genomic sequence coding for potato multicystatin, an eight-domain cysteine proteinase inhibitor. *Plant Mol. Biol.* **23**, 801–812
22. Walsh, T.A. and Strickland, J.A. (1993) Proteolysis of the 85-kilodalton crystalline cysteine proteinase inhibitor from potato releases functional cystatin domains. *Plant Physiol.* **103**, 1227–1234
23. Heath, R.L., Barton, P.A., Simpson, R.J., Reid, G.E., Lim, G., and Anderson, M.A. (1995) Characterization of the protease processing sites in a multidomain proteinase inhibitor precursor from *Nicotiana alata*. *Eur. J. Biochem.* **230**, 250–257