A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Molecular cloning and characterization of novel cysteine protease inhibitors from *Calotropis procera* R. Br.

Calotropis procera R. Br. 유래 신규 cysteine protease 저해제의 유전자 클로닝 및 특성 규명

February, 2017

Department of Agricultural Biotechnology

Seoul National University

Yeo, Subin

석사학위논문

Molecular cloning and characterization of novel cysteine protease inhibitors from *Calotropis procera* R. Br.

지도교수 장 판 식

이 논문을 석사학위 논문으로 제출함

2017년 2월

서울대학교 대학원

농 생 명 공 학 부

여 수 빈

여수빈의 석사 학위논문을 인준함

2017년 2월

위원장 문 태 화 (인)

부위원장 장 판 식

(인) 위원 서 진 호 (인)

Abstract

The cysteine protease is an important enzyme in industrial, physiological, and therapeutic uses. This enzyme was found out to be a significant role in the autolysis in foods and the proliferation of tumor cells. Therefore, it needs to be inhibited during food processing and can be the target enzyme for drugs. For these reasons, effective cysteine protease inhibitors from natural sources have been studied for decades.

Calotropis procera R. Br. was selected as an inhibitor source. It is a medicinal and edible tropical plant. In the preliminary study, its cDNA sequence and putative cysteine proteases were revealed by RNA sequencing. Propeptides of cysteine proteases from *Calotropis procera* R. Br. were considered to be efficient inhibitors against cathepsin L.

In this study, the molecular cloning, expression, and characterization of several candidate cysteine protease inhibitors were described, and the cause that occur the inhibitory activity difference was presumed by bioinformatic analysis.

Firstly, eight kinds of cysteine protease propeptides were selected as candidate inhibitors by sequence analysis with cathepsin L. They were cloned

and expressed in *Escherichia coli* BL21(DE3). Three inhibitors (SnuCalCpI02, SnuCalCpI03, and SnuCalCpI15) were overexpressed in soluble form, and two inhibitors (SnuCalCpI12 and SnuCalCpI16) were overexpressed in mostly insoluble form. The others (SnuCalCpI08, SnuCalCpI14, and SnuCalCpI17) were not expressed.

Then, the characterization of five expressed inhibitors was performed. Only two (SnuCalCpI03 and SnuCalCpI15) exhibited inhibitory activity against papain, and their inhibitory activity against human cathepsin L was investigated. The half maximal inhibitory concentrations (IC_{50}) of SnuCalCpI03 and SnuCalCpI15 were 18.58 nM and 17.50 nM, respectively. They acted as the competitive inhibitors. The inhibitor proteins were stable at a high temperature and a wide range of pH except pI point.

Lastly, to analyze the cause of the different inhibitory activity of inhibitors, bioinformatic analysis was carried out. Consequently, it is presumed that the inhibitory activity can be different by the tendency that inhibitors interact with the enzyme.

Novel cysteine protease inhibitors from *Calotropis procera* R. Br. are expected to be used as efficient anti-tumor drugs or food stabilizers.

Keywords: *Calotropis procera* R. Br., cysteine protease inhibitor, molecular

cloning, inhibitory activity, inhibition mechanism, stability, bioinformatic analysis

Student number: 2015-21781

Contents

List of tables

List of figures

Fig. 8. Inhibitory activity against human cathepsin L of SnuCalCpI03 and SnuCalCpI15.

Fig. 9. The Lineweaver-Burk plot for the determination of inhibition type (a) SnuCalCpI03 and (b) SnuCalCpI15.

Fig. 10. Thermal stability of SnuCalCpI03 and SnuCalCpI15.

Fig. 11. pH stability of SnuCalCpI03 and SnuCalCpI15.

Fig. 12. Protein-protein docking simulation of SnuCalCpIs and papain (a) Papain zymogen, (b) Papain-SnuCalCpI02 complex, (c) Papain-SnuCalCpI03 complex, (d) Papain-SnuCalCpI12 complex, (e) Papain-SnuCalCpI15 complex, and (f) Papain-SnuCalCpI16 complex.

1. Introduction

The cysteine protease is a class of protease enzymes that are significant to industrial, physiological and therapeutic applications. They are good targets of therapeutic drugs or food industrial agents. For example, many cathepsins, like cathepsin L, B or S, are expressed at high level in tumor cells and they are involved with the progression, metastasis and angiogenesis of tumor cells (Olson, et al., 2015; Sudhan, et al., 2015). Therefore, cathepsins would be target enzymes of anti-tumor drugs (Sudhan, et al., 2015). In addition, cysteine proteases excessively degrade seafood and meat during meat or surimi processing, and have a negative effect on food quality (An, et al., 1996; Hopkins, et al., 2002; Jiang, 2000; Shyu, et al., 2004). In this case, cysteine proteases need to be regulated. Thus, the effective cysteine protease inhibitors have been researched for decades.

Cysteine protease inhibitors are inhibitory substances that regulate the activity of cysteine proteases. There are many types of inhibitors including chemical small molecules or proteinous inhibitors. Among them, natural propeptides of cysteine proteases can be used for inhibitors toward cysteine proteases. Proteases are usually synthesized as precursor proteins called

zymogens. Prodomains of cysteine proteases are regulatory propeptides that make proteases inactive. They are usually block a protease active site to prevent substrate access. By removing propeptides, cysteine proteases change into active enzymes (Verma, et al., 2016). These propeptides can function as inhibitors for other homologous endogenous and exogenous cysteine proteases. It is reported that cysteine protease propeptides exist in various plants and animals.

According to the previous studies, cysteine protease inhibitors have various physiological functions. They have anti-viral (Aoki, et al., 1995), anti-fungal (Martinez, et al., 2005; Popovic, et al., 2012), and anti-parasitic activity by inhibiting microbial cysteine proteases. And they are involved in the inhibition of tumor cells, anti-inflammation, and immunomodulatory responses (Vray, et al., 2002). Moreover, cysteine protease inhibitors from plants have been supplemented as stabilizers in food processing (Shyu, et al., 2004).

Calotropis procera R. Br. (Asclepiadaceae family) is a medicinal and edible tropical plant. In the preliminary studies, the latex from this plant has various therapeutic functions such as hepatoprotective, anti-arthritic, anti-inflammatory, antipyretic, and anticancer treatments (Choedon, et al., 2006; Dewan, et al., 2000; V. Kumar, et al., 1994; V. L. Kumar, et al., 2007; Shivkar, et al., 2003). And the full cDNA sequence of *Calotropis procera* R. Br. was

revealed, and twenty cysteine proteases from *Calotropis procera* R. Br. were identified by RNA sequencing (Kwon, et al., 2015). It was presumed that its propeptides can inhibit papain-like cysteine proteases and be applied as potent functional substances. I tried to find efficient novel cysteine protease inhibitors from *Calotropis procera* R. Br. to apply food and drugs.

In this study, I described the molecular cloning, expression and characterization of several novel cysteine protease inhibitors obtained from *Calotropis procera* R. Br. (SnuCalCpI). And the cause that make a difference in inhibitory activity of inhibitors was inferred by the bioinformatics tools.

2. Materials and Methods

2.1. Strains, plasmids, and media

Escherichia coli DH5α obtained from Invitrogen (California, USA) was grown in LB medium at 37°C and used as a strain for plasmids with genes encoding recombinant proteins. pET29b(+) was used for recombinant protein expression in *E. coli* BL21(DE3). LB medium was used for growing *E. coli* strains at 37°C. LB agar medium containing 50 μg/mL kanamycin was used to screen the recombinant clones.

2.2. Selection of candidate cysteine protease inhibitors from *Calotropis procera* R. Br.

There are twenty kinds of cysteine proteases in *Calotropis procera* R. Br. (Table 1). Homology modeling of twenty cysteine proteases from *Calotropis procera* R. Br. was performed by SWISS-MODEL tool (http://swissmodel.expasy.org/), and sequences and structures of target proteins were compared to template proteins. Then, eight propeptides that have the highest identity (>40%) with cathepsin L propeptide were selected as candidate inhibitors (Table 2).

Table 1. List of cysteine proteases encoding transcripts from *Calotropis procera* R. Br.

Unigene ID	Protein length (amino acids)	BLASTP best match (Species, Gene bank accession ID)	Identity	E-value	Putative domains contained
SnuCalCp01	354	Procerain B, partial (*Calotropis procera*, AGI59309.1)	208/212 (98%)	1e-151	Signal sequence, I29, peptidase C1A
SnuCalCp02	461	Cysteine protease Cp4 (*Actinidia deliciosa*, ABQ10202.1)	333/460 (72%)	0.0	Signal sequence, I29, peptidase C1A
SnuCalCp03	344	Cysteine protease CP15 (*Nicotiana tabacum*, AGV15823.1)	189/354 (53%)	7e-117	Signal sequence, I29, peptidase C1A
SnuCalCp04	373	Cysteine proteinase 15A-like	281/344	0.0	Signal sequence, I29,

		(*Citrus sinensis*, XP_006473584.1)	(82%)		peptidase C1A
SnuCalCp05	458	Cysteine protease CP6 (*Nicotiana tabacum*, AGV15820.1)	337/443 (76%)	0.0	Signal sequence, I29, peptidase C1A
SnuCalCp06	297	Cathepsin L-like proteinase (*Medicago truncatula*, XP_003626102.1)	224/326 (69%)	7e-175	Peptidase C1A
SnuCalCp07	388	Papain-like cysteine proteinase (*Ipomoea batatas*, AAF61440.1)	295/372 (79%)	0.0	Signal sequence, I29, peptidase C1A
SnuCalCp08	363	Cysteine protease (*Nicotiana tabacum*, BAA96501.1)	286/363 (79%)	0.0	Signal sequence, I29, peptidase C1A
SnuCalCp09	362	Xylem cysteine proteinase 1-like (*Solanum tuberosum*,	278/362 (77%)	0.0	Signal sequence, I29, peptidase C1A

		XP_006342169.1)			
SnuCalCp10	317	Pro-asclepain f (*Gomphocarpus fruticosus*, CAR31335.1)	253/316 (80%)	5e-176	I29, peptidase C1A
SnuCalCp11	367	Cysteine proteinase-like (*Vitis vinifera*, XP_002279940.1)	245/329 (74%)	0.0	Signal sequence, I29, peptidase C1A
SnuCalCp12	349	Cysteine protease CP15 (*Nicotiana tabacum*, AGV15823.1)	190/360 (53%)	5e-119	Signal sequence, I29, peptidase C1A
SnuCalCp13	382	Papain family cysteine protease (*Arabidopsis thaliana*, NP_567010.5)	246/351 (70%)	4e-180	Signal sequence, I29, peptidase C1A
SnuCalCp14	455	Cysteine protease family protein (*Populus trichocarpa*,	280/417 (67%)	0.0	Signal sequence, I29, peptidase C1A

		XP_002307688.2)			
SnuCalCp15	366	Cysteine protease CP15 (*Nicotiana tabacum*, AGV15823.1)	186/359 (52%)	7e-118	Signal sequence, I29, peptidase C1A
SnuCalCp16	361	Cysteine protease (*Nicotiana tabacum*, BAA96501.1)	241/360 (67%)	0.0	Signal sequence, I29, peptidase C1A
SnuCalCp17	469	Cysteine protease Cp6 (*Actinidia deliciosa*, ABQ10204.1)	221/321 (69%)	3e-153	Signal sequence, I29, peptidase C1A
SnuCalCp18	356	Vignain-like (*Solanum lycopersicum*, XP_004239155.1)	280/353 (79%)	0.0	Signal sequence, I29, peptidase C1A
SnuCalCp19	363	Cys endopeptidase family protein (*Populus trichocarpa*,	273/360 (76%)	0.0	Signal sequence, I29, peptidase C1A

		XP_002321654.1)			
SnuCalCp20	390	KDEL-tailed cysteine endopeptidase CEP1-like (*Solanum lycopersicum*, XP_004252607.1)	237/320 (74%)	6e-167	I29, peptidase C1A

Table 2. List of cysteine protease propeptides of over 40% identity with cathepsin L propeptide

Unigene ID	Template protein	Identity
SnuCalCp02	*Toxoplasma gondii* cathepsin L	42.11%
SnuCalCp03	*Toxoplasma gondii* cathepsin L	42.11%
SnuCalCp08	*Toxoplasma gondii* cathepsin L	45.61%
SnuCalCp12	*Toxoplasma gondii* cathepsin L	42.11%
SnuCalCp14	*Toxoplasma gondii* cathepsin L	43.86%
SnuCalCp15	*Toxoplasma gondii* cathepsin L	40.35%
SnuCalCp16	*Toxoplasma gondii* cathepsin L	43.86%
SnuCalCp17	*Toxoplasma gondii* cathepsin L	40.35%

2.3. Total RNA extraction and cDNA synthesis

Total RNA was extracted from *Calotropis procera* R. Br. leaf by RNeasy® plant mini kit (Qiagen, Valencia, CA, USA). Two hundred mg of *Calotropis procera* R. Br. leaf was prepared by grinding in liquid nitrogen. All RNA extraction procedures followed the manufacturer's protocol. RNA was eluted with 30 µL of RNase-free water. cDNA was synthesized with 5 µg of RNA using PrimeScript™ 1st strand cDNA synthesis kit (Takara Korea Biomedical Inc., Seoul, Korea). Synthesis of cDNA was carried out with total RNA, oligo dT primer, RNase inhibitor, 5X PrimeScript buffer, dNTP mixture, and PrimeScript RTase.

2.4. cDNA cloning and sequencing of SnuCalCpI genes

To amplify the cysteine protease inhibitor genes, specific primers were designed (Table 3). cDNA cloning was carried out using a synthetic single-stranded cDNA as a template with 2X TOPsimple™ PreMIX-Forte (Enzynomics, Daejeon, Korea). Thermocycler (Bio-Rad (California, USA), T100™) was used for the PCR. PCR cycle was performed as follows: initial denaturation at 95°C for 2 min, 35 cycles of denaturation at 95°C for 30 sec, annealing at 53°C for 30 sec, elongation at 72°C for 1 min, and final elongation at 72°C for 5 min. PCR products were purified from 1% agarose gel using

MEGAquick-spin^{TM} Total Fragment DNA Purification Kit (iNtRON Biotechnology, Seongnam, Korea). DNA sequencing was performed by ABI 3730xl DNA analyzer.

Table 3. Primers used for cloning of recombinant cysteine protease inhibitors

Primer name	Sequence (5’ to 3’)
SnuCalCpI02_F	AAGGAGATATACATATGTTAGACATGTCCATTATCAGT
SnuCalCpI02_R	GGTGGTGGTGCTCGAGATCATCACCAGCATTAAAAGA
SnuCalCpI03_F	AAGGAGATATACATATGAAAATCATATCCATTGCCGAT
SnuCalCpI03_R	GGTGGTGGTGCTCGAGGTTAAGGTCAACTTCAGAAAA
SnuCalCpI08_F	AAGGAGATATACATATGGTTGACGACGGATCATCAG
SnuCalCpI08_R	GGTGGTGGTGCTCGAGGACAACGTTGGTTAGCTTG
SnuCalCpI12_F	AAGGAGATATACATATGATTGCCGATGAATTAGTCCG
SnuCalCpI12_R	GGTGGTGGTGCTCGAGAGAAAGGTTATAGTTAACTTGG
SnuCalCpI14_F	AAGGAGATATACATATGTCATTTTCATCTTCTTCTTCTT
SnuCalCpI14_R	GGTGGTGGTGCTCGAGATCAAAATCATCAAAGACATCT

SnuCalCpI15_F	AAGGAGATATACATATGATCATCACTACTAGCCTCC
SnuCalCpI15_R	GGTGGTGGTGCTCGAGGCTTTCAGATCCAACTTTGT
SnuCalCpI16_F	AAGGAGATATACATATGGACCGTTCATCATTCTCCG
SnuCalCpI16_R	GGTGGTGGTGCTCGAGAGCAACGTTGTTGAGCTTTA
SnuCalCpI17_F	AAGGAGATATACATATGTCTGAGATCACGTCGGTTA
SnuCalCpI17_R	GGTGGTGGTGCTCGAGATTATCTTCGGCTTTAGGAAG

2.5. Construction of expression plasmids in *E. coli* system

To construct the expression plasmids pET29b(+)-SnuCalCpI-C6His (vector-gene-tag), eight kinds of SnuCalCpI genes (320~380 bp) were amplified by PCR and extracted from agarose gel to obtain a correct size of fragments. pET29b(+) were digested by *Nde*I and *Xho*I, and all SnuCalCpI DNA fragments from PCR products were inserted into the *Nde*I and *Xho*I site of pET29b(+) expression vector by In-Fusion® HD Cloning Kit (Takara). Recombinant plasmids were transformed into *E. coli* DH5α. Recombinant strains were screened by kanamycin LB agar medium and verified with DNA sequencing. Verified recombinant plasmids were transformed into *E. coli* BL21(DE3) for cysteine protease inhibitor expression.

2.6. Expression of cysteine protease inhibitors in *E. coli* system

Each recombinant strains of *E. coli* cells were grown in 5 mL of LB medium containing 50 μg/mL kanamycin at 37°C with shaking overnight. Culture medium were inoculated to 200 mL of LB medium containing 50 μg/mL kanamycin at 37°C with shaking until the optical density at 600 nm (OD_{600}) was approximately 0.6. Subsequently, expression of the SnuCalCpI genes was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM and incubating at 20°C overnight. After

centrifugation at 10,000xg for 10 min at 4°C, cell pellet was suspended in 50 mM Tris-HCl buffer (pH 8) with 300 mM NaCl and 10 mM imidazole and then disrupted by sonication. Sonication was conducted with 10 cycles of 10 sec pulse, 10 sec pause, and 35% power. The cell lysates were separated into supernatant and cell pellet by centrifugation at 12,000xg for 20 min. The supernatant and cell pellet were used for SDS-PAGE analysis.

2.7. SDS-PAGE

SDS-PAGE was carried out using 12% polyacrylamide gels followed by the method of Laemmli (Laemmli, 1970). Protein concentration was determined according to Bradford method (Kruger, 1994) using BSA as a standard. 20 µL of expressed soluble fractions were mixed with the same volume of the Laemmli sample buffer and boiled for 3 min before loading. The protein size marker (ELPIS Biotech, Daejeon, Korea) was used as a mid-range protein standard to estimate the molecular weight of proteins. Protein samples were isolated on Hoefer SE 250 mini-gel (GE Healthcare, Seoul, Korea) at room temperature at 20 mA current. Proteins were stained with Coomassie Brilliant Blue G-250 (Bio-Rad).

2.8. His-tagged protein purification by Ni-NTA column

Expressed soluble recombinant proteins were purified by Ni-NTA column. Ni-NTA slurry was packed in econo-columns and the resin was washed with the lysis buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, and 10 mM imidazole). Supernatant fraction of expressed proteins in the lysis buffer were loaded on the resin after shaking slowly for 30 min. Unbound proteins were washed with the washing buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, and 20 mM imidazole). Target proteins were eluted with the elution buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, and 250 mM imidazole) and the eluent was collected.

2.9. Inhibition assay

2.9.1. Inhibition assay against papain

Papain activity was measured using 1 mM N-α-Benzoyl-DL-arginine-β-naphthylamide (BANA) as a substrate. Various concentrations (0, 20, 40, 60, 80, and 100 μg/mL) of recombinant SnuCalCpI samples in 0.2 mL were mixed with 0.1 mL sodium phosphate buffer (0.5 M sodium phosphate with 10 mM EDTA, pH 6.0), 0.1 mL of 50 mM 2-mercaptoethanol, and 0.1 mL of 25 μg/mL papain (Sigma, Missouri, USA) solution. The mixture was incubated at 37°C for 10 min. After pre-incubation, 0.2 mL of 1 mM BANA was added as a substrate and the mixture was incubated for 20 min at 37°C. The enzyme reaction was terminated by adding 1 mL of 2% (v/v) HCl/ethanol and 1 mL of

0.06% (w/v) *p*-dimethylaminocinnamaldehyde/ethanol, and the mixture was put at room temperature for 30 min for color development. After that, the absorbance of products was measured at 540 nm with a spectrophotometer. The relative residual activity of papain was calculated using the following,

Equation: Relative residual activity (%) $= \frac{T*}{T} X\ 100$

where T is the absorbance at 540 nm in the absence of inhibitors, and T* is the one in the presence of inhibitors.

2.9.2. Inhibition assay against cathepsin L

Cathepsin L (BioVision Inc., California, USA) inhibition assay was conducted following description. Various concentrations of SnuCalCpI03 and SnuCalCpI15 samples in 20 µL were mixed with 40 µL of 0.5 M sodium phosphate buffer (pH 6.0) with 10 mM EDTA, 20 µL of 100 mM DTT (dithiothreitol), and 20 µL of 2 nM cathepsin L. The mixture was incubated at 30^{o}C for 10 min. After that, 0.1 mL of 1.2 µM Z-Phe-Arg-7-amino-4-methylcoumarin hydrochloride (from a 1 mM stock solution in DMSO). Cathepsin L activity was detected by the fluorescence of 7-amino-4-methylcoumarin (AMC) with excitation at 355 nm and emission at 460 nm in 96-well microplate. The relative residual activity of cathepsin L was calculated using the following,

Equation: Relative residual activity (%) $= \frac{T^*}{T} X\ 100$

where T is the initial velocity in the absence of inhibitors and T* is the initial velocity in the presence of inhibitors. The initial velocity was the slope of linear curve for initial 2 min of the reaction. The half maximal inhibitory concentration, IC_{50}, was determined from the fit curve equation using SigmaPlot.

2.10. Inhibition type analysis

The inhibition type of inhibitors was analyzed using cathepsin L inhibition assay. Inhibitor samples were added 0, 20, and 40 nM respectively and concentrations of the substrate were 0.5, 0.8, 1 and 2 μM. Each cathepsin L activity was measured following above. The concentrations of substrate and the initial velocity were converted into Lineweaver-Burk equation, and Lineweaver-Burk plots were drawn to determine the inhibition mechanism.

2.11. Stability test

Stability test was evaluated with cathepsin L. Experiment processes were the same with the procedure described in ‘2.9.2. Inhibition assay against cathepsin L’.

2.11.1. Thermal stability

To examine the thermal stability, inhibitor samples were incubated in various temperatures (30, 40, 50, 60, 70, 80, 90, and 100^{o}C) for 30 min in 50 mM Tris-HCl buffer (pH 8). After the samples were cooled at 4°C, cathepsin L inhibitory activity was evaluated as described above.

2.11.2. pH stability

Inhibitor samples were incubated in 100 mM glycine-HCl buffer at pH 2.0~3.0, 100 mM sodium acetate buffer at pH 4.0~5.0, 100 mM sodium phosphate buffer at pH 6.0, 100 mM Tris-HCl buffer at pH 7.0~8.0, and 100 mM glycine-NaOH buffer at pH 9.0~10.0 at 4^{o}C for 2 hr. The inhibitory assay toward cathepsin L was performed.

2.12. Bioinformatic analysis

To analyze the reason why the difference of inhibitors occurred, firstly, homology modeling of papain and SnuCalCpIs was performed by SWISS-MODEL. Then protein-protein interaction residues of each proteins were predicted by meta-PPISP (http://pipe.scs.fsu.deu/meta-ppisp.html/) which was the program for predicting protein-protein interaction sites. Lastly, the protein-protein docking simulation of papain-inhibitor complexes was conducted by

ZDOCK (http://zdock.umassmed.edu/).

3. Results and Discussion

3.1. Molecular cloning and sequencing of SnuCalCpI genes

After cDNA synthesis, nucleotide sequences of eight SnuCalCpI genes including SnuCalCpI02, SnuCalCpI03, SnuCalCpI08, SnuCalCpI12, SnuCalCpI14, SnuCalCpI15, SnuCalCpI16, and SnuCalCpI17 were amplified by PCR. Amplification of SnuCalCpI genes using specific-designed primers yielded fragments with sizes of 320~380 bp approximately (Fig. 1). Gel extraction and DNA clean-up process were carried out, and DNA fragments were cloned into pET29b(+) vector. Recombinant plasmids were transformed into *E. coli* DH5α. The presence of SnuCalCpI genes in recombinant plasmids was confirmed by colony PCR and sequencing for comparing with the verified cDNA sequences of *Calotropis procera* R. Br. (Fig. 2). The DNA sequences of SnuCalCpIs were determined by ABI 3730xl DNA analyzer. pET29b confirm primers were used for initial sequencing. Expected protein molecular weights are 13~16 kDa.

Then, SnuCalCpI genes were multiple aligned with papain propeptide gene and human cathepsin L propeptide gene (Fig. 3). SnuCalCpIs showed the high homology with papain propeptide and human cathepsin L propeptide. SnuCalCpI genes contain ‘ERFNIN (EXXXRXXXFXXNXXXIXXXN)’

motif and ‘GNFD (GXNXFXD)’ motif which are highly conserved regions (Wiederanders, et al., 2003). These motifs are important regulatory elements of papain-like cysteine protease prodomains. ERFNIN motif is present in the long α-helix which contributes a major part of the propeptide. GNFD motif is located at the kink of the β-sheet ahead of the short α-helix capping the opening of the interdomain cleft. The presence of ERFNIN-GNFD motif suggests that SnuCalCpI is related to cathepsin L group and can have high possibility to inhibit human cathepsin L (Kwon, et al., 2015; Martinez, et al., 2008; Roy, et al., 2012).

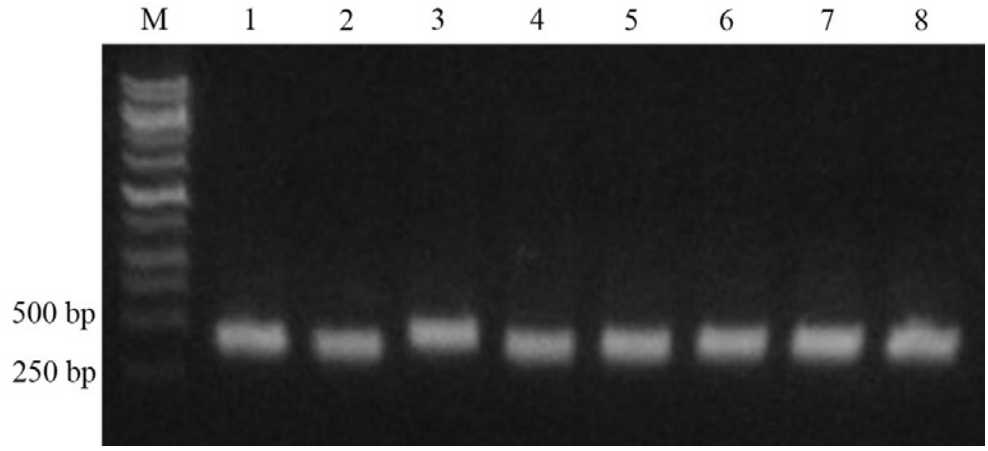

Fig. 1. Gel electrophoresis of amplified SnuCalCpI genes by RT-PCR; lane M, DNA size marker; lane 1, SnuCalCpI02; lane 2, SnuCalCpI03; lane 3, SnuCalCpI08; lane 4, SnuCalCpI12; lane 5, SnuCalCpI14; lane 6, SnuCalCpI15; lane 7, SnuCalCpI16; lane 8, SnuCalCpI17 (95°C, 30 sec denaturation; 53°C, 30 sec annealing; 72°C, 1 min elongation; 35 cycles, 2X TOPsimpleTM PreMIX-Forte).

(a)

```
ttagacatgtccattatcagttatgataatgaccatggtcagatggttaggtctgatgat
 L  D  M  S  I  I  S  Y  D  N  D  H  G  Q  M  V  R  S  D  D
gaggttaggtctttgtatgaatcttggcttgttaagcatgggaaagcttacaatgcttta
 E  V  R  S  L  Y  E  S  W  L  V  K  H  G  K  A  Y  N  A  L
ggggagaaagagaaaaggtttgaaattttcaaagataatcttcagttcattgacgaacat
 G  E  K  E  K  R  F  E  I  F  K  D  N  L  Q  F  I  D  E  H
aactctaagaacctttcttacaaacttggccttaatcgtttctcggatctgagtcacgag
 N  S  K  N  L  S  Y  K  L  G  L  N  R  F  S  D  L  S  H  E
gagtttcggtccatttttgtgagtggtcgaatggatcggaaggctaggttgatgaagggt
 E  F  R  S  I  F  V  S  G  R  M  D  R  K  A  R  L  M  K  G
aaggttggggatcgttattcttttaatgctggtgatgat
 K  V  G  D  R  Y  S  F  N  A  G  D  D
```

(b)

```
atatccattgccgatgaatccgtgtggcgaaccgaagaagaagtcatggcaatatacgag
 I  S  I  A  D  E  S  V  W  R  T  E  E  E  V  M  A  I  Y  E
gaatggatagtgaaacacgggaaatcatacaacgcattaggagaggagaaatttaaaaga
 E  W  I  V  K  H  G  K  S  Y  N  A  L  G  E  E  K  F  K  R
ttcgaaatatttaaggataatcttaaatatatcgagaaacacaacagtcttcccaatcaa
 F  E  I  F  K  D  N  L  K  Y  I  E  K  H  N  S  L  P  N  Q
atctacaaactcggtttgaatcaattttcggatcttacctttgacgagtttaaatccatt
 I  Y  K  L  G  L  N  Q  F  S  D  L  T  F  D  E  F  K  S  I
tatttaagtagcatccctatggatacttcgttaagtgagtccaaaattgacttttctgaa
 Y  L  S  S  I  P  M  D  T  S  L  S  E  S  K  I  D  F  S  E
gttgaccttaactttccc
 V  D  L  N  F  P
```

(c)

gttgacgacggatcatcagcctattttgcccaggtgaatccgatcaggcaagtcgtgtcc
V D D G S S A Y F A Q V N P I R Q V V S
gacggtctgcgtgaattagagaattcttttgttcaggttattggaaatactcgccatgtg
D G L R E L E N S F V Q V I G N T R H V
ctctcctttgctcgctttgctcataggtatggaaagaggtacgagactgctgaggagata
L S F A R F A H R Y G K R Y E T A E E I
aaagtgaggttcgacatattcagggacaatctgcggatgattaaatcgcataacaagaag
K V R F D I F R D N L R M I K S H N K K
ggactgtcattcagtcttggtgttaatgcattttctgatttgacatgggaggaattccgt
G L S F S L G V N A F S D L T W E E F R
aagcataggttgggagctgcccagaactgttcagctaccacaaagggaaatctcaagcta
K H R L G A A Q N C S A T T K G N L K L
accaacgttgtc
T N V V

(d)

attgccgatgaattagtccggcgaactgacgaagaagtcatgtcaatatacgaggaatgg
I A D E L V R R T D E E V M S I Y E E W
Atggtggaatacaggaaatcctacgacgcattaggagtggagaaattaaagagattcgaa
M V E Y R K S Y D A L G V E K L K R F E
atatttaaggataatcttaagtatatggaagagcacaacagtcttcccaatcaaacttac
I F K D N L K Y M E E H N S L P N Q T Y
aagctcggtttgaaccaattttccgatcttactcttcgcgagtttaaatccatctattta
K L G L N Q F S D L T L R E F K S I Y L
agcagcagccctattgatactttgttagatgagtccgaaattgacttttcctattttccc
S S S P I D T L L D E S E I D F S Y F P
caagttaactataacctttct
Q V N Y N L S

(e)

```
tcattttcatcttcttcttcttcttcttcgccgagtactgctgatttattcgagaat
 S  F  S  S  S  S  S  S  S  S  S  P  S  T  A  D  L  F  E  N
tggtgtcgagaatatggaaaaacatactcttctgaacaagaaaaacagtacagacatgga
 W  C  R  E  Y  G  K  T  Y  S  S  E  Q  E  K  Q  Y  R  H  G
gtatttaaagataactatgattacattacccagcataacagtaagggtaattcgacctgt
 V  F  K  D  N  Y  D  Y  I  T  Q  H  N  S  K  G  N  S  T  C
actctttcccttaacgcctttgctgatctcacccaccatgagtttaaatcccagttcttg
 T  L  S  L  N  A  F  A  D  L  T  H  H  E  F  K  S  Q  F  L
ggtctctccgcttctctcaatagtcccattcgattgaatcgaggttcctcctctgctatt
 G  L  S  A  S  L  N  S  P  I  R  L  N  R  G  S  S  S  A  I
gggacattagatgtctttgatgattttgat
 G  T  L  D  V  F  D  D  F  D
```

(f)

```
atcatcactactagcctccataattctcagaataaacttgtttggcgtactaatgatgaa
 I  I  T  T  S  L  H  N  S  Q  N  K  L  V  W  R  T  N  D  E
gtcatttcattatttgaggaatggttagttaaacataggaaggtatataatgctatagga
 V  I  S  L  F  E  E  W  L  V  K  H  R  K  V  Y  N  A  I  G
gaaaaagaaaagagattcgagatctttaagaataatcttaaatttattgatgagcacaat
 E  K  E  K  R  F  E  I  F  K  N  N  L  K  F  I  D  E  H  N
attagatatccaaacaagacttacacacttggcctaaatgtgtttgctgatcttactgat
 I  R  Y  P  N  K  T  Y  T  L  G  L  N  V  F  A  D  L  T  D
gatgagtaccaatccaagtatttaggtacccgtattcatccaaagagaaagtattttgca
 D  E  Y  Q  S  K  Y  L  G  T  R  I  H  P  K  R  K  Y  F  A
tctcacagtagtgatgatgatgagtatcttcacaaagttggatctgaaagc
 S  H  S  S  D  D  D  E  Y  L  H  K  V  G  S  E  S
```

(g)

```
gaccgttcatcattctccgaggagaatccgatcagacaagtcgtagcggaccgcctgcgt
 D  R  S  S  F  S  E  E  N  P  I  R  Q  V  V  A  D  R  L  R
gaactagaagcttctttccttgaagtcgtcggagacaaccgccacgtcctctcatttgct
 E  L  E  A  S  F  L  E  V  V  G  D  N  R  H  V  L  S  F  A
agctttgctcttaagtacggaaagaagtacgaaactgatggggagattaagaaaaggttc
 S  F  A  L  K  Y  G  K  K  Y  E  T  D  G  E  I  K  K  R  F
gaaatattcagagagaacttatggaagattgaatcccataacaagaagggactctcatat
 E  I  F  R  E  N  L  W  K  I  E  S  H  N  K  K  G  L  S  Y
actctcggcattaatgaattttctgatatgacatgggaggaattcagcaagcgtaggttg
 T  L  G  I  N  E  F  S  D  M  T  W  E  E  F  S  K  R  R  L
ggagcaccgcagcactgttcagcaaccaaatattcgactctaaagctcaacaacgttgct
 G  A  P  Q  H  C  S  A  T  K  Y  S  T  L  K  L  N  N  V  A
```

(h)

```
tctgagatcacgtcggttagctccgctggaaacaagaaaactaggccgccgtcgcagcga
 S  E  I  T  S  V  S  S  A  G  N  K  K  T  R  P  P  S  Q  R
accgatgatgaagttgctgcattatacgaggaatggcttgtcaaacacggcaaggcgtat
 T  D  D  E  V  A  A  L  Y  E  E  W  L  V  K  H  G  K  A  Y
aatgccttgggagagaaatccaatagattcgcaatatttaaggataatcttaggtatata
 N  A  L  G  E  K  S  N  R  F  A  I  F  K  D  N  L  R  Y  I
gatgatcacaacagcaatgctaaccggacttacaaacttggtctgaaccaatttgccgat
 D  D  H  N  S  N  A  N  R  T  Y  K  L  G  L  N  Q  F  A  D
cttaccaatagtgagttccggtcaatgtatttagacacacgaatcgatgctaaaattagg
 L  T  N  S  E  F  R  S  M  Y  L  D  T  R  I  D  A  K  I  R
aggagatcatcaaagttcgtcgtcagcgatcgatatcttcctaaagccgaagataat
 R  R  S  S  K  F  V  V  S  D  R  Y  L  P  K  A  E  D  N
```

Fig. 2. The nucleotide sequences and deduced amino acid sequences of SnuCalCpI genes (a) SnuCalCpI02, (b) SnuCalCpI03, (c) SnuCalCpI08, (d)

SnuCalCpI12, (e) SnuCalCpI14, (f) SnuCalCpI15, (g) SnuCalCpI16, and (h) SnuCalCpI17.

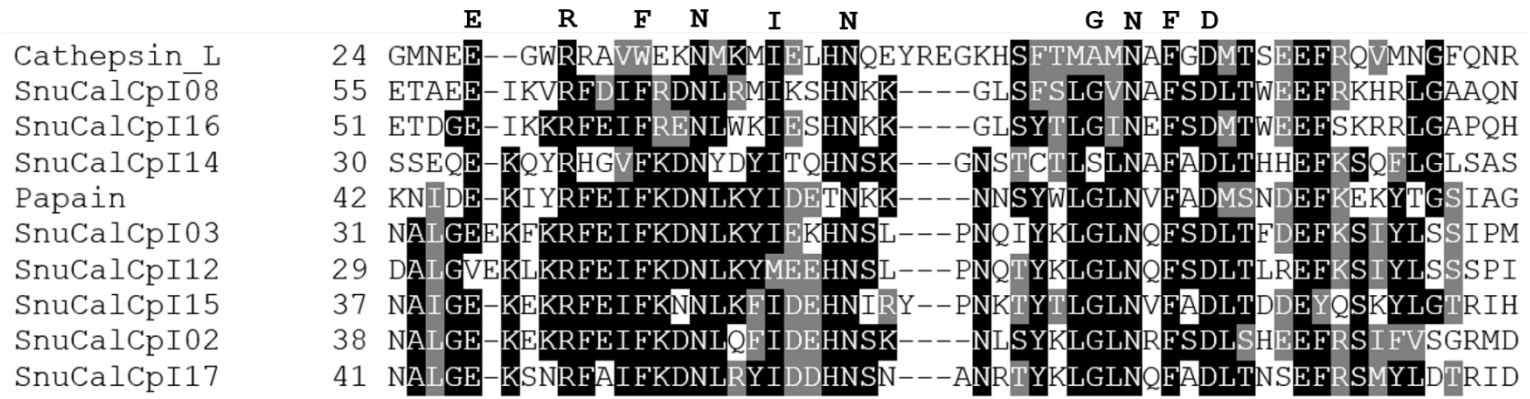

Fig. 3. Multiple protein sequence alignment of SnuCalCpIs with papain propeptide and human cathepsin L propeptide by Clustal Omega.

3.2. Expression of SnuCalCpIs in *E. coli* system and purification

pET29b(+)-SnuCalCpI-C6His (Fig. 4) were transformed into *E. coli* BL21(DE3) for protein expression. *E. coli* is the most popular protein production system because its cell factory is well-established and economical. Moreover, there are many molecular tools and protocols for the high-level production of heterologous proteins (Rosano, et al., 2014). Proteins were expressed by T7 promoter, and 0.5 mM IPTG was used as an inducer. 6X histidine was tagged at C-terminal of SnuCalCpIs for the purification. SDS-PAGE analysis (Fig. 5) showed that soluble SnuCalCpI02, SnuCalCpI03, and SnuCalCpI15 were expressed at high level and SnuCalCpI12 and SnuCalCpI16 were expressed but most of them were insoluble. And SnuCalCpI08, SnuCalCpI14, and SnuCalCpI17 were not expressed in *E. coli* system. The reason why SnuCalCpI08, SnuCalCpI14, and SnuCalCpI17 were not overexpressed in *E. coli* system is not clear. SnuCalCpI12 and SnuCalCpI16 were obtained through the scale-up culture. Unexpressed proteins, SnuCalCpI08, SnuCalCpI14, and SnuCalCpI17, were excluded in the next experiments.

Expressed proteins were purified by Ni-NTA column. SDS-PAGE analysis showed that target proteins were purified by 6X His-tag (Fig. 6).

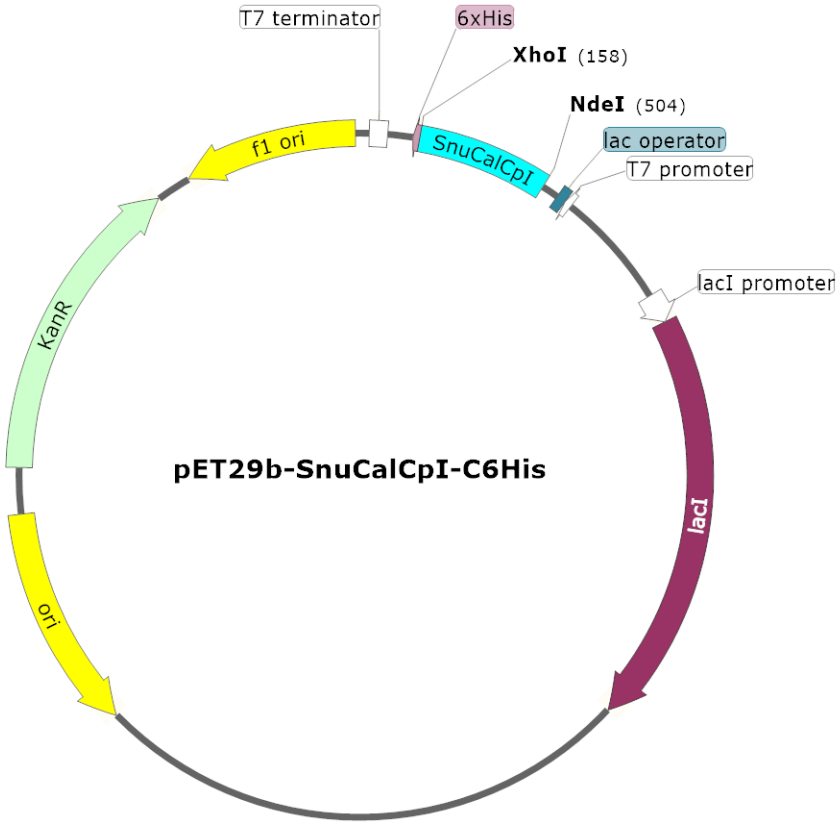

Fig. 4. Construction map of expression vector, pET29b(+)-SnuCalCpI-C6His.

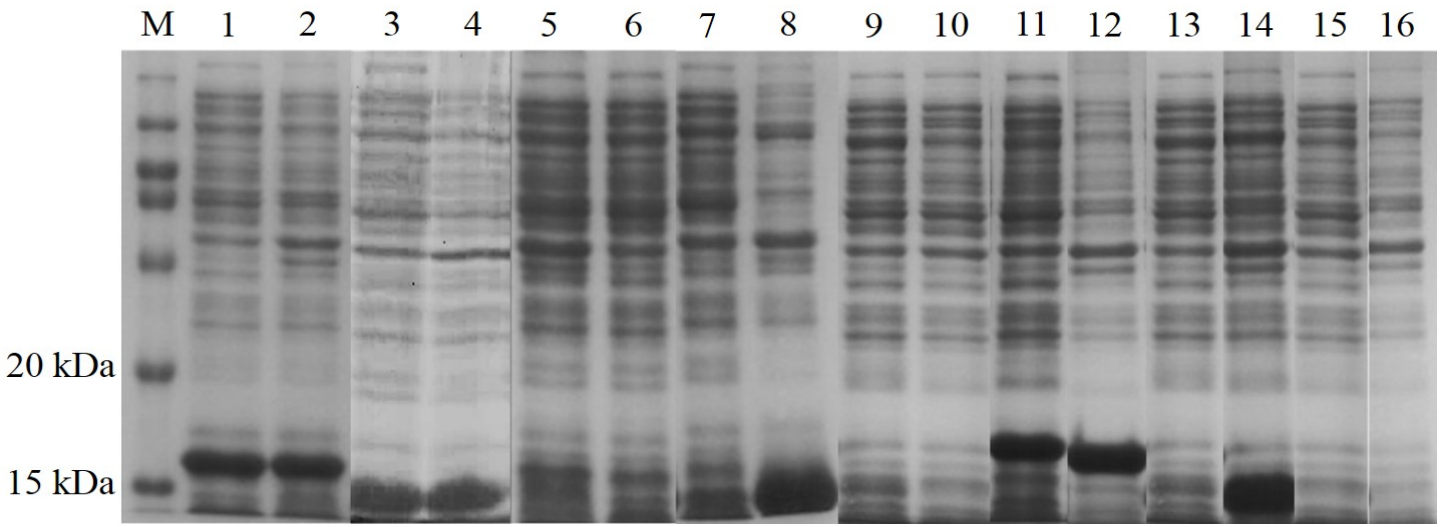

Fig. 5. SDS-PAGE analysis of recombinant SnuCalCpIs expressed in *E. coli* system; lane M, protein size marker; lane 1-2, supernatant and pellet of SnuCalCpI02; lane 3-4, supernatant and pellet of SnuCalCpI03; lane 5-6, supernatant and pellet of SnuCalCpI08; lane 7-8, supernatant and pellet of SnuCalCpI012; lane 9-10, supernatant and pellet of SnuCalCpI14; lane 11-12, supernatant and pellet of SnuCalCpI15; lane 13-14, supernatant and pellet of SnuCalCpI16; lane 15-16, supernatant and pellet of SnuCalCpI17.

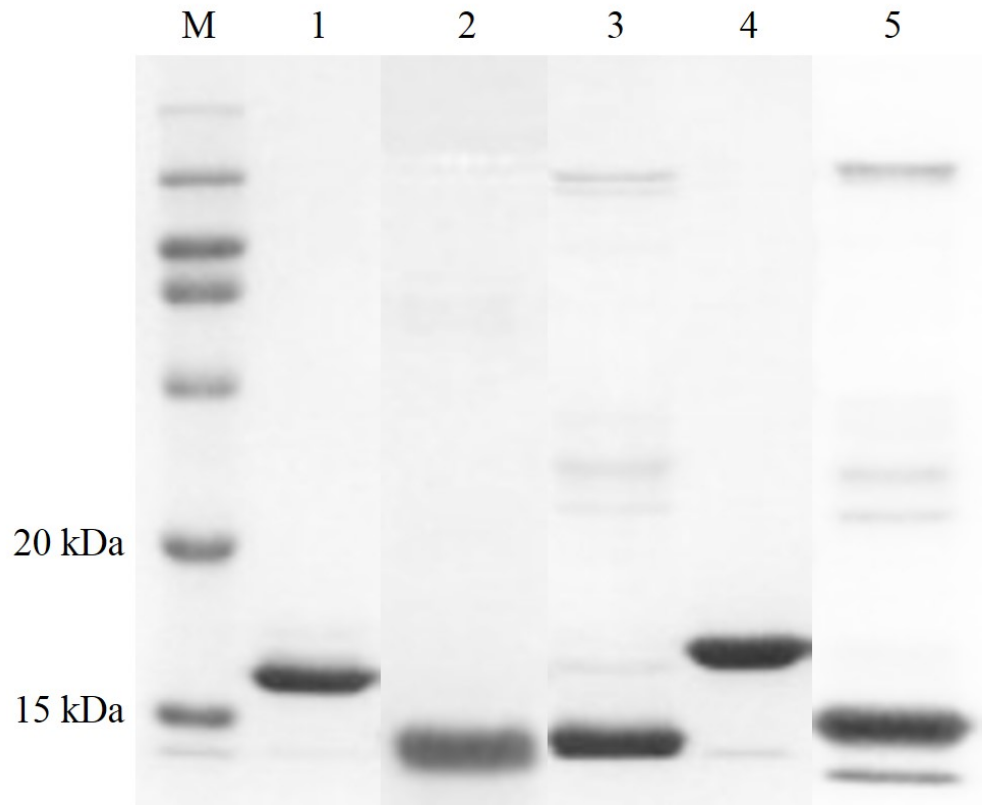

Fig. 6. SDS-PAGE analysis of affinity-purified recombinant SnuCalCpIs; lane M, protein size marker; lane 1, SnuCalCpI02; lane 2, SnuCalCpI03; lane 3, SnuCalCpI12; lane 4, SnuCalCpI15; lane 5, SnuCalCpI16.

3.3. Characterization

3.3.1. Inhibitory activity against papain

The characterization was carried out using the affinity-purified recombinant inhibitor samples. As a pre-screening, the inhibitory activity of five kinds of expressed SnuCalCpIs (SnuCalCpI02, SnuCalCpI03, SnuCalCpI12, SnuCalCpI15, and SnuCalCpI16) were determined against papain from papaya latex. The papain inhibition assay was followed previous method (Abe, et al., 1994). EDTA and 2-mercaptoethanol were used as the activator and the reducing agent for enzyme reaction. BANA, the substrate, was degraded into β-naphthylamine by papain. β-naphthylamine absorbs the visible light of 540 nm wavelength. The lower the absorbance of products was, the more inhibition occurred. As a result, SnuCalCpI03 and SnuCalCpI15 showed inhibitory activity against papain, while SnuCalCpI02, SnuCalCpI12, and SnuCalCpI16 could not inhibit papain (Fig. 7). It could be concluded that two inhibitors with inhibitory activity against papain were expected to have inhibitory activity against cathepsin L. Three inhibitors without the inhibitory activity against papain won’t be able to inhibit cathepsin L or have very low inhibitory activity.

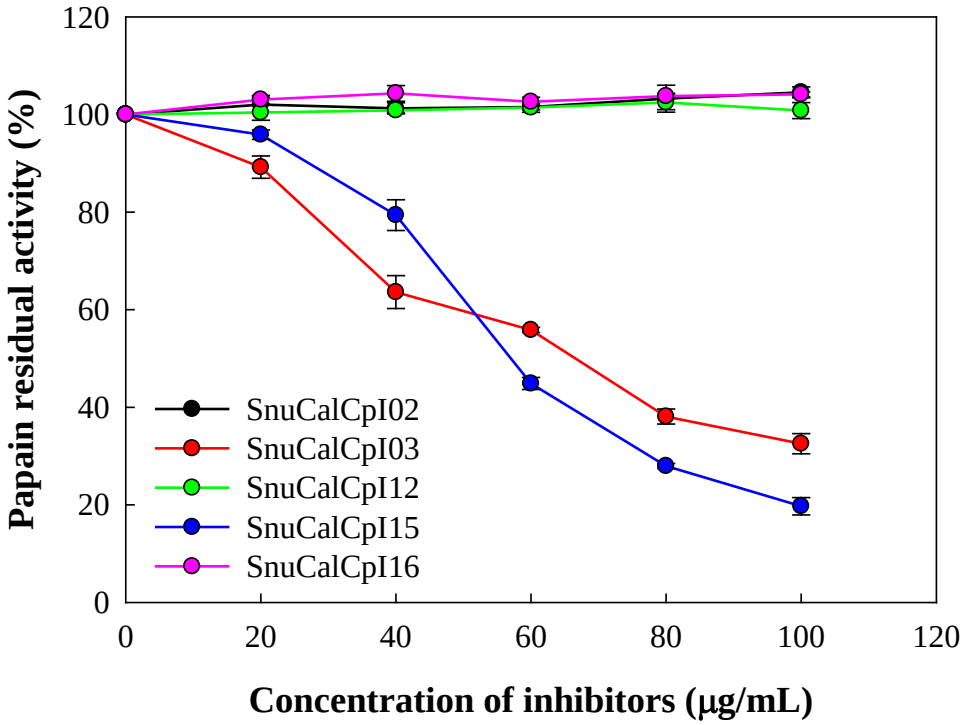

Fig. 7. Inhibitory activity of SnuCalCpIs against papain; ●: SnuCalCpI02, ●:SnuCalCpI03, ●: SnuCalCpI12, ●: SnuCalCpI15, ●: SnuCalCpI16.

3.3.2. Inhibitory activity against human cathepsin L

Then, the inhibition assay against human cathepsin L was conducted with SnuCalCpI03 and SnuCalCpI15. Recombinant human cathepsin L was used. Cathepsin L recognize a cleavage site that are the next of Phe and Arg of Z-Phe-Arg-AMC, and the product, 7-methylcoumarin was fluorescent. It was detected by the fluorophotometer for 10 min. The inhibitory activity was evaluated with the initial velocity.

Both of SnuCalCpI03 and SnuCalCpI15 exhibited inhibitory activity against recombinant human cathepsin L (Fig. 8). IC_{50} values were determined by drawing fit curves. IC_{50} means the half maximal inhibitory concentration. IC_{50} values of SnuCalCpI03 and SnuCalCpI15 were 18.58 nM and 17.50 nM respectively. Their inhibitory activity toward human cathepsin L was similar. It was more sensitive than against papain. The low nanomolar IC_{50} values of SnuCalCpI03 and SnuCalCpI15 means that they are more efficient inhibitors compared to other inhibitors of previous studies. For example, IC_{50} value of thiocarbazate (PubChem SID 26681509) against cathepsin L was 155 nM (Shah, et al., 2008), and FgStefin-1 have IC_{50} value of 213.3 nM against cathepsin L (Tarasuk, et al., 2009).

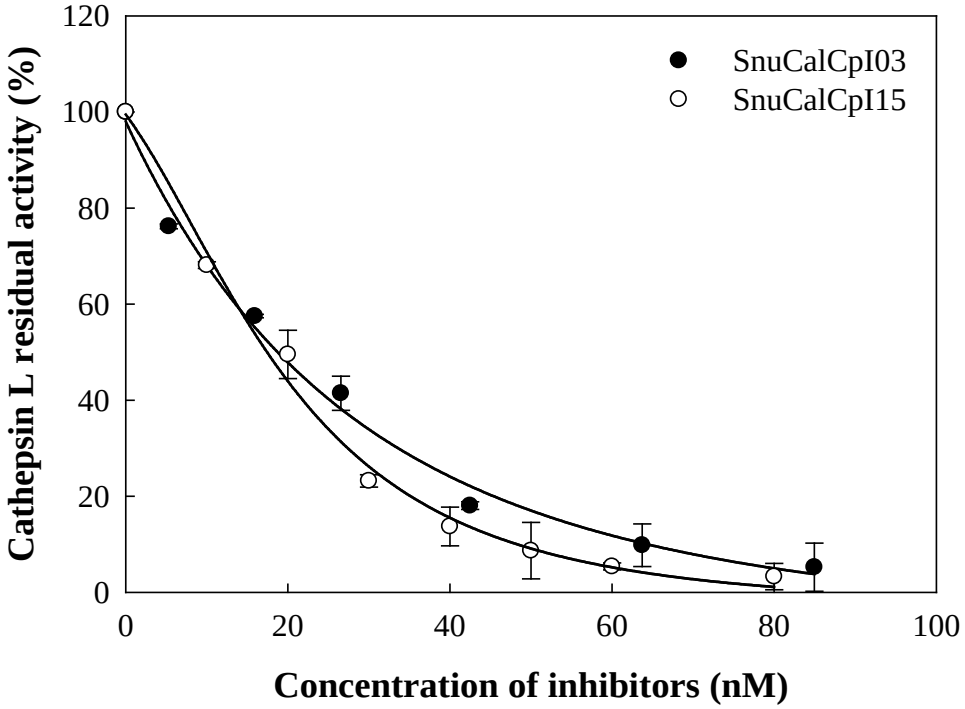

Fig. 8. Inhibitory activity against human cathepsin L of SnuCalCpI03 and SnuCalCpI15; ●: SnuCalCpI03, ○: SnuCalCpI15.

3.3.3. Inhibition type analysis

Inhibition type of SnuCalCpI03 and SnuCalCpI15 was analyzed by the Lineweaver-Burk plot. The Lineweaver-Burk plot is widely used to determine important terms in enzyme kinetics (Lineweaver, et al., 1934). The y-intercept is equivalent to the inverse of V_{max} and the x-intercept represents $-1/K_m$. It also gives a quick, visual impression of the different forms of enzyme inhibition. In the case of competitive inhibition, K_m increases and V_{max} is fixed. Non-competitive inhibitors have the same K_m but V_{max} reduces. And uncompetitive inhibition causes both of K_m and V_{max} to reduce. According to the Lineweaver-Burk plot (Fig. 9), it could be concluded that SnuCalCpI03 and SnuCalCpI15 are competitive inhibitors. Their K_m values increased along with the increase of inhibitor concentrations, and V_{max} values were constant.

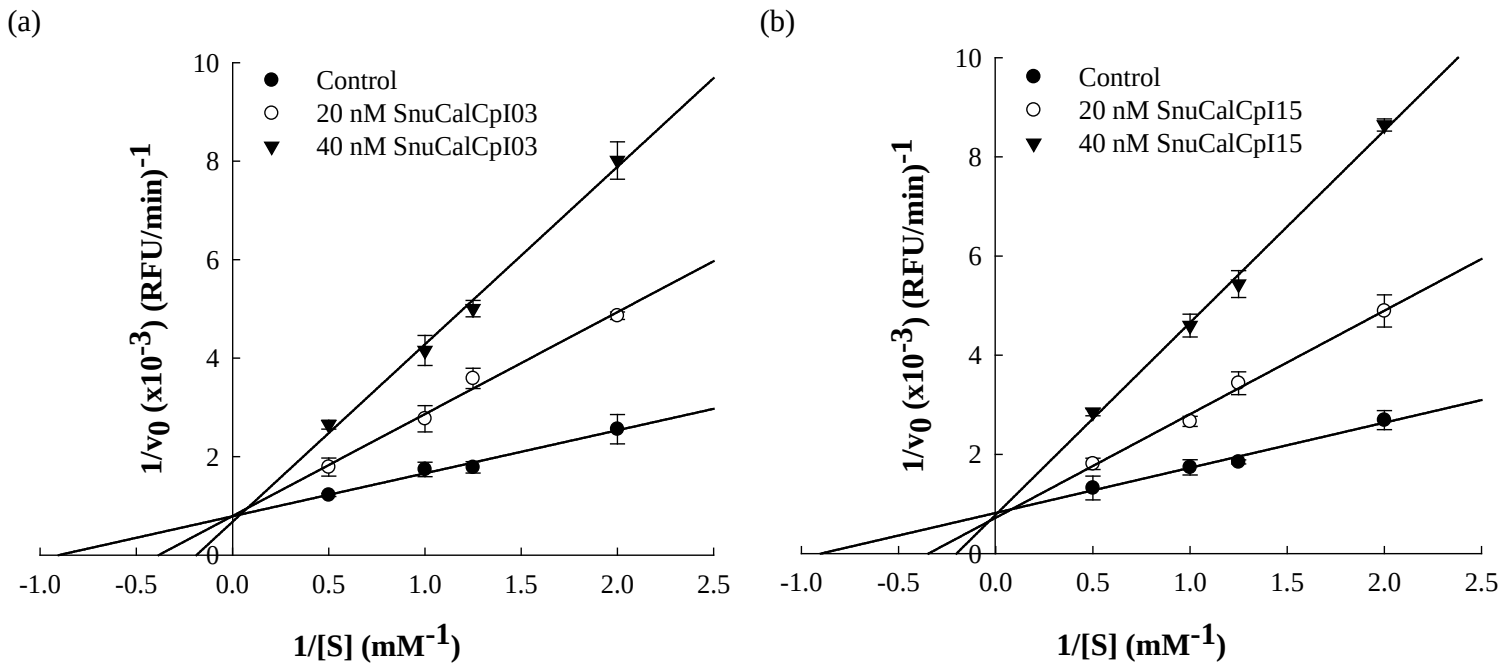

Fig. 9. The Lineweaver-Burk plot for the determination of inhibition type (a) SnuCalCpI03 and (b) SnuCalCpI15.

3.3.4. Thermal stability

To assess of the thermal stability of SnuCalCpI03 and SnuCalCpI15, the recombinant proteins were treated at 30~100°C for 30 min. Both inhibitors were thermally stable at the temperature below 70°C maintaining more than 90% inhibitory activity (Fig. 10). They slightly lost their inhibitory activity to 70~80% at the higher temperature. SnuCalCpI03 and SnuCalCpI15 are more thermostable than other cysteine protease inhibitors. It could be due to their simple secondary structure, the compactness of protein structure (Thompson, et al., 1999), or intra-helical salt bridges (Das, et al., 2000).

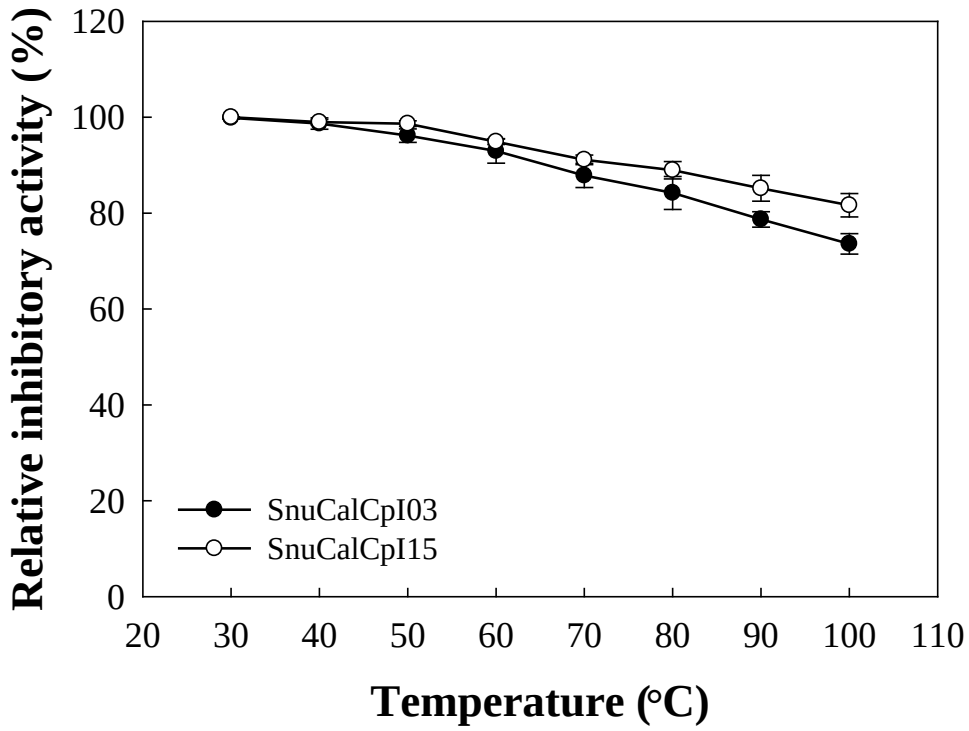

Fig. 10. Thermal stability of SnuCalCpI03 and SnuCalCpI15; ●: SnuCalCpI03, ○: SnuCalCpI15.

3.3.5. pH stability

To examine pH effect on SnuCalCpI03 and SnuCalCpI15, recombinant proteins were incubated at different pH values (pH 2~10) for 2 hr. Both inhibitors maintained their inhibitory activity against cathepsin L even at strongly acidic or basic environment comparing to other cysteine protease inhibitors (Fig. 11). However, SnuCalCpI03 lost its inhibitory activity around pH 5 and SnuCalCpI15 lost its inhibitory activity around pH 7. It is assumed that isoelectric point (pI) of SnuCalCpIs affected the result. The theoretical pI point of SnuCalCpI03 is 5.2 and the one of SnuCalCpI15 is 6.9. At the pI point, the charge of inhibitor proteins changed and solubility was lower to affect their inhibitory activity. According to the result of stability test, SnuCalCpI03 and SnuCalCpI15 are expected to be applied or stored at diverse environments, because they are stable at a high temperature and a wide range of pH except pI point.

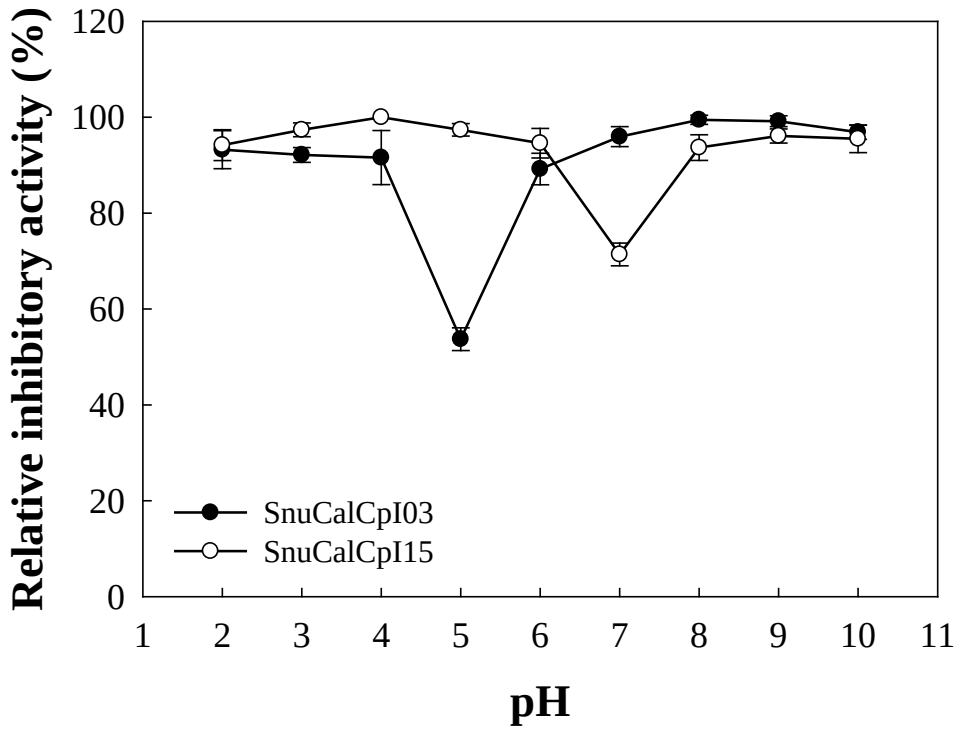

Fig. 11. pH stability of SnuCalCpI03 and SnuCalCpI15; ●: SnuCalCpI03, ○: SnuCalCpI15.

3.4. Bioinformatic analysis

To analyze of the reason why some inhibitors have the inhibitory activity and some have not, comparative bioinformatics analysis was conducted. According to previous study, ERFNIN and GNFD motifs are important elements to inhibit papain-like cysteine proteases (Pandey, et al., 2009). Some residues of these motifs interact with residues around a catalytic site of cysteine protease mature domain and block a catalytic site. It was expected that the structural or charge differences affected this blocking form.

Using meta-PPISP, residues that can interact with other proteins were analyzed. meta-PPISP predicts the protein-protein interaction sites through the combination of three prediction programs including cons-PPISP, PINUP, and Promate (Qin, et al., 2007). cons-PPISP is a neural network predictor that uses sequence profiles and solvent accessibilities of neighboring residues as input (Chen, et al., 2005). PINUP is based on an empirical energy function consisting of a side-chain energy term, a term proportional to solvent accessible area, and a term accounting for sequence conservation (Liang, et al., 2006). Promate uses a composite probability calculated from properties such as secondary structure, atom distribution, amino acid pairing, and sequence conservation (Neuvirth, et al., 2004). meta-PPISP combines three methods in a linear regression analysis with the raw scores as input, and it is found to

consistently outperform the three individual methods. If meta-PPISP prediction score of a residue was over 0.34, that residue is in the interface.

Then, ZDOCK (Pierce, et al., 2014), protein-protein docking simulation, performed 3D-modeling of enzyme-inhibitor complexes (Fig. 12). ZDOCK are Fast Fourier Transform based protein docking programs. ZDOCK searches all possible binding modes in the translational and rotational space between the two proteins and evaluates each pose using an energy-based scoring function. It provides prediction models in order of the suitability. The condition to select blocking residues and binding site residues was set up according to the result of meta-PPISP. The top prediction models were used to comparison. The protein-protein docking simulation result showed that SnuCalCpI02, SnuCalCpI12, and SnuCalCpI16 cannot block the active site (green-circled) of the papain. There is high probability that inhibitors interact with other sites of papain. On the other hand, SnuCalCpI03 and SnuCalCpI15 tend to interact with residues around the active site and block the active site more. This is similar with papain zymogen structure. That is, the inhibitory activity can be different by the tendency that inhibitors interact with the enzyme, although the sequences of inhibitors are similar and have same regulatory regions.

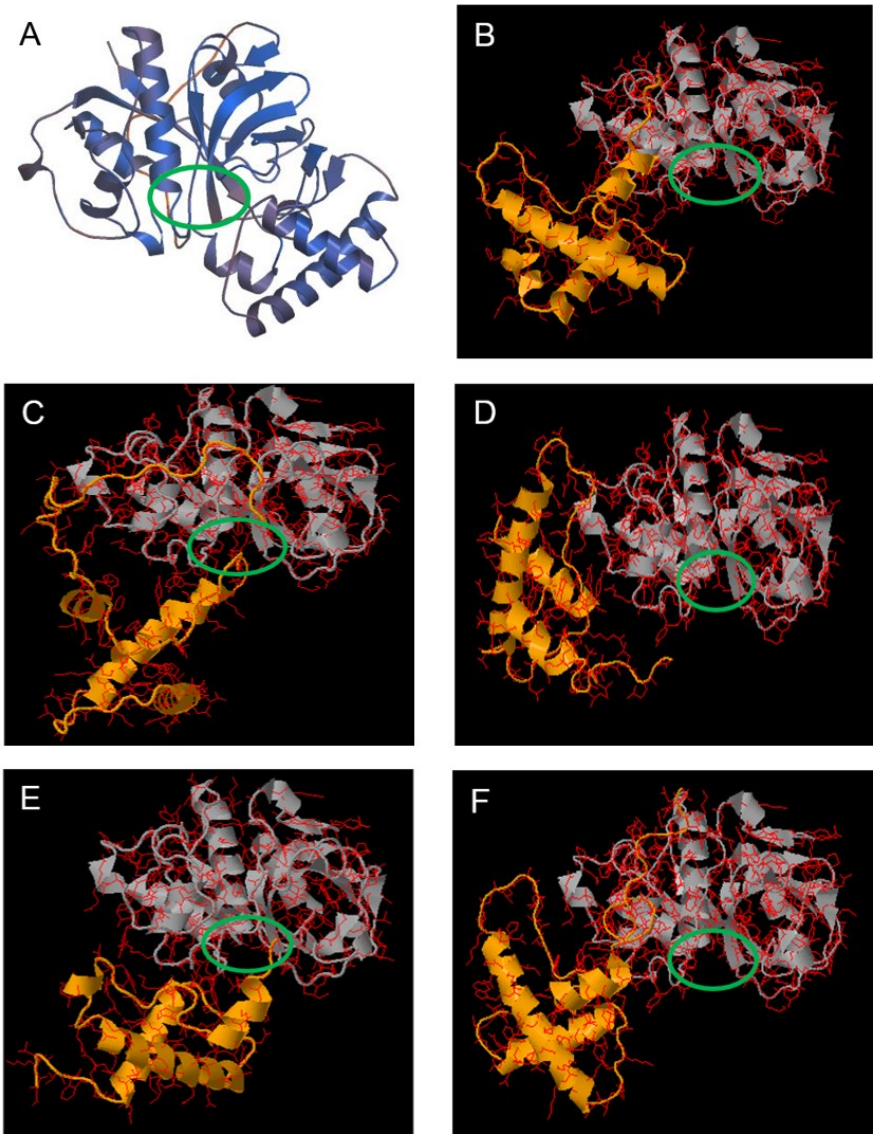

Fig. 12. Protein-protein docking simulation of SnuCalCpIs and papain; (a) Papain zymogen, (b) Papain-SnuCalCpI02 complex, (c) Papain-SnuCalCpI03 complex, (d) Papain-SnuCalCpI12 complex, (e) Papain-SnuCalCpI15 complex, and (f) Papain-SnuCalCpI16 complex.

4. Conclusion

In this study, novel eight cysteine protease propeptides from *Calotropis procera* R. Br. were cloned and five of them were expressed in *E. coli* system. SnuCalCpI03 and SnuCalCpI15 have the effective inhibitory activity toward papain and cathepsin L. Their inhibition type was competitive inhibition. They are stable at a high temperature and a wide range of pH except pI point. And it is presumed that the inhibitory activity can be different by the tendency that inhibitors interact with the enzyme. These novel cathepsin L inhibitors would be anti-tumor drugs, microbial agents or food stabilizers.

5. References

Abe, M., Abe, K., Iwabuchi, K., Domoto, C., & Arai, S. (1994). Corn cystatin I expressed in Escherichia coli: investigation of its inhibitory profile and occurrence in corn kernels. *Journal of Biochemistry, 116*(3), 488-492.

An, H., Peters, M. Y., & Seymour, T. A. (1996). Roles of endogenous enzymes in surimi gelation. *Trends in Food Science & Technology, 7*(10), 321-327.

Aoki, H., Akaike, T., Abe, K., Kuroda, M., Arai, S., Okamura, R.-i., Negi, A., & Maeda, H. (1995). Antiviral effect of oryzacystatin, a proteinase inhibitor in rice, against herpes simplex virus type 1 in vitro and in vivo. *Antimicrobial Agents and Chemotherapy, 39*(4), 846-849.

Chen, H., & Zhou, H. X. (2005). Prediction of interface residues in protein–protein complexes by a consensus neural network method: test against NMR data. *Proteins: Structure, Function, and Bioinformatics, 61*(1), 21-35.

Choedon, T., Mathan, G., Arya, S., Kumar, V. L., & Kumar, V. (2006). Anticancer and cytotoxic properties of the latex of *Calotropis procera* in a transgenic mouse model of hepatocellular carcinoma. *World*

Journal of Gastroenterology, 12(16), 2517.

Das, R., & Gerstein, M. (2000). The stability of thermophilic proteins: a study based on comprehensive genome comparison. *Functional & Integrative Genomics, 1*(1), 76-88.

Dewan, S., Kumar, S., & Kumar, V. L. (2000). Antipyretic effect of latex of *Calotropis procera*. *Indian Journal of Pharmacology, 32*(3), 252-252.

Hopkins, D., & Thompson, J. (2002). Factors contributing to proteolysis and disruption of myofibrillar proteins and the impact on tenderisation in beef and sheep meat. *Crop and Pasture Science, 53*(2), 149-166.

Jiang, S.-T. (2000). Effect of proteinases on the meat texture and seafood quality. *Food Science and Agricultural Chemistry, 2*(2), 55-74.

Kruger, N. J. (1994). The Bradford method for protein quantitation. In *Basic Protein and Peptide Protocols*, (pp. 9-15): Springer.

Kumar, V., & Basu, N. (1994). Anti-inflammatory activity of the latex of *Calotropis procera*. *Journal of Ethnopharmacology, 44*(2), 123-125.

Kumar, V. L., & Roy, S. (2007). *Calotropis procera* latex extract affords protection against inflammation and oxidative stress in Freund's complete adjuvant-induced monoarthritis in rats. *Mediators of Inflammation, 2007*.

Kwon, C. W., Park, K.-M., Kang, B.-C., Kweon, D.-H., Kim, M.-D., Shin, S.

W., Je, Y. H., & Chang, P.-S. (2015). Cysteine Protease Profiles of the Medicinal Plant Calotropis procera R. Br. Revealed by De Novo Transcriptome Analysis. *PloS ONE*, *10*(3), e0119328.

Laemmli, U. (1970). Denaturing (SDS) discontinuous gel electrophoresis. *Nature*, *277*, 680-685.

Liang, S., Zhang, C., Liu, S., & Zhou, Y. (2006). Protein binding site prediction using an empirical scoring function. *Nucleic Acids Research, 34*(13), 3698-3707.

Lineweaver, H., & Burk, D. (1934). The determination of enzyme dissociation constants. *Journal of the American Chemical Society, 56*(3), 658-666.

Martinez, M., Abraham, Z., Gambardella, M., Echaide, M., Carbonero, P., & Diaz, I. (2005). The strawberry gene Cyf1 encodes a phytocystatin with antifungal properties. *Journal of Experimental Botany, 56*(417), 1821-1829.

Martinez, M., & Diaz, I. (2008). The origin and evolution of plant cystatins and their target cysteine proteinases indicate a complex functional relationship. *BMC Evolutionary Biology, 8*(1), 1.

Neuvirth, H., Raz, R., & Schreiber, G. (2004). ProMate: a structure based prediction program to identify the location of protein–protein binding sites. *Journal of Molecular Biology, 338*(1), 181-199.

Olson, O. C., & Joyce, J. A. (2015). Cysteine cathepsin proteases: regulators of cancer progression and therapeutic response. *Nature Reviews Cancer, 15*(12), 712-729.

Pandey, K. C., Barkan, D. T., Sali, A., & Rosenthal, P. J. (2009). Regulatory elements within the prodomain of Falcipain-2, a cysteine protease of the malaria parasite Plasmodium falciparum. *PloS ONE, 4*(5), e5694.

Pierce, B. G., Wiehe, K., Hwang, H., Kim, B.-H., Vreven, T., & Weng, Z. (2014). ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics, 30*(12), 1771-1773.

Popovic, M. M., Bulajic, A., Ristic, D., Krstic, B., Jankov, R. M., & Gavrovic-Jankulovic, M. (2012). In vitro and in vivo antifungal properties of cysteine proteinase inhibitor from green kiwifruit. *Journal of the Science of Food and Agriculture, 92*(15), 3072-3078.

Qin, S., & Zhou, H.-X. (2007). meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics, 23*(24), 3386-3387.

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology, 5*, 172.

Roy, S., Choudhury, D., Aich, P., Dattagupta, J. K., & Biswas, S. (2012). The

structure of a thermostable mutant of pro-papain reveals its activation mechanism. *Acta Crystallographica Section D: Biological Crystallography, 68*(12), 1591-1603.

Shah, P. P., Myers, M. C., Beavers, M. P., Purvis, J. E., Jing, H., Grieser, H. J., Sharlow, E. R., Napper, A. D., Huryn, D. M., & Cooperman, B. S. (2008). Kinetic characterization and molecular docking of a novel, potent, and selective slow-binding inhibitor of human cathepsin L. *Molecular Pharmacology, 74*(1), 34-41.

Shivkar, Y., & Kumar, V. (2003). Anthelmintic activity of latex of *Calotropis procera*. *Pharmaceutical Biology, 41*(4), 263-265.

Shyu, D. J., Chyan, C.-L., Tzen, J. T., & Chou, W.-M. (2004). Molecular cloning, expression, and functional characterization of a cystatin from pineapple stem. *Bioscience, Biotechnology, and Biochemistry, 68*(8), 1681-1689.

Sudhan, D. R., & Siemann, D. W. (2015). Cathepsin L targeting in cancer treatment. *Pharmacology & Therapeutics, 155*, 105-116.

Tarasuk, M., Grams, S. V., Viyanant, V., & Grams, R. (2009). Type I cystatin (stefin) is a major component of Fasciola gigantica excretion/secretion product. *Molecular and Biochemical Parasitology, 167*(1), 60-71.

Thompson, M. J., & Eisenberg, D. (1999). Transproteomic evidence of a loop-

deletion mechanism for enhancing protein thermostability1. *Journal of Molecular Biology, 290*(2), 595-604.

Verma, S., Dixit, R., & Pandey, K. C. (2016). Cysteine Proteases: Modes of Activation and Future Prospects as Pharmacological Targets. *Frontiers in Pharmacology, 7*(107).

Vray, B., Hartmann, S., & Hoebeke, J. (2002). Immunomodulatory properties of cystatins. *Cellular and Molecular Life Sciences CMLS, 59*(9), 1503-1512.

Wiederanders, B., Kaulmann, G., & Schilling, K. (2003). Functions of propeptide parts in cysteine proteases. *Current Protein and Peptide Science, 4*(5), 309-326.

국문초록

Cysteine protease는 단백질 가수분해 효소의 한 종류로 의약품이나 식품 등의 산업에서 중요한 역할을 한다. 이 효소는 약물의 목적 효소가 되기도 하며, 식품의 품질에 부정적인 영향을 끼칠 경우 저해될 필요가 있다.

따라서 효과적인 cysteine protease 저해제를 찾기 위한 연구가 많이 진행되어 왔다. 선행연구에서 *Calotropis procera* R. Br.이라는 식물의 cDNA 서열이 밝혀졌으며 20가지 papain family의 cysteine protease가 존재하는 것을 밝혔다. 본 연구에서는 *Calotropis procera* R. Br.로부터 cysteine protease를 효과적으로 저해할 수 있는 신규 저해제를 찾기 위한 연구를 진행하였다.

우선 *Calotorpis procera* R. Br. 잎으로부터 RNA를 추출하였고, 이로부터 cDNA를 합성하였다. 서열 분석을 통해 후보로 선정한 propeptide 8가지에 대한 각각의 primer들을 디자인하였고 이를 사용해 cDNA를 증폭시켜 cDNA library를 구성하였다. 8가지 propeptide 유전자들을 각각 pET29b(+) 벡터에 삽입하여 재조합

플라스미드를 구성하였고, *E. coli* BL21(DE3)에 transformation하였다. 발현시킨 결과, 3가지는 soluble하게 과발현이 되었고, 2가지는 과발현 되었으나 대부분 insoluble하게 존재하였으며, 나머지 3가지는 발현이 되지 않았다.

이 중 발현이 되는 5가지를 가지고 특성 분석을 진행하였다. 5가지 중 2가지가 papain에 대한 저해효과를 가졌다. 이 2가지의 cathepsin L에 대한 저해활성을 측정한 결과, IC_{50} 값은 각각 18.58 nM, 17.50 nM이었다. 2가지 저해제는 경쟁적 저해제로 작용하는 것으로 나타났다. 또한 70°C 이하의 온도에서 안정하였으며, pI를 제외한 pH 2~10 범위에서 안정했다.

마지막으로 서열이 유사함에도 불구하고 5가지 저해제들의 저해효과의 차이가 나타나는 이유를 밝히기 위해 생물정보분석을 진행하였다. Protein-protein interaction site prediction을 통해 inhibitor와 papain의 residue들이 다른 protein과 상호작용을 할 수 있는지를 분석하였고, 이를 토대로 protein-protein docking simulation을 진행하였다. 그 결과, 저해제가 효소와 상호작용할 때 복합체를 이루는 구조적 경향성의 차이 때문에 저해능력의 차이가 나타나는 것으로 추론할 수 있었다.

본 연구는 cathepsin L에 대해 높은 저해활성을 갖는 새로운 소재의 저해제를 찾아냈다는 것에 의의가 있으며, 본 연구의 저해제들은 자가소화가 일어나기 쉬운 식품의 안정제나 항암 소재의 개발 등에 적용될 수 있을 것으로 예상된다.

주요어: *Calotropis procera* R. Br., cysteine protease 저해제, 유전자 클로닝, 저해활성, 저해메커니즘, 안정성, 생물정보분석

학번: 2015-21781