식물생명공학회지 제35권 제4호
J Plant Biotechnol
Vol. 35, No. 4, 317-327 (2008)

Identification and characterization of the phytocystatin family from *Brassica rapa*

Joon Ki Hong[1], Jung Eun Hwang[2], Tae-Ho Park[1], Yun-Xiang Zang[1], Sang Choon Lee[1], Soo-Jin Kwon[1], Jeong-Hwan Mun[1], Huyn Uk Kim[1], Jin A Kim[1], Mina Jin[1], Jung Sun Kim[1], Soo In Lee[1], Myung-Ho Lim[1], Yoonkang Hur[3], Chae Oh Lim[2], and Beom-Seok Park[1*]

[1]*Brassica Genomics Team, National Institute of Agricultural Biotechnology, Rural Development Administration, Suwon 447-707, Korea*

[2]*Environmental Biotechnology National Core Research Centre and PMBBRC, Division of Applied Life Science (BK21 Program), Graduate School of Gyeongsang National University, Jinju 660-701, Korea*

[3]*Plant Genome Research Institute, Chungnam National University, Daejeon 305-764, Korea*

ABSTRACT Phytocystatins, which are inhibitors of plant cysteine peptidases, are involved in the regulation of protein turnover and in the defense against insect pests and pathogens. Extensive searches in the *Brassica rapa* genome allowed the prediction of at least eight different phytocystatin genes on seven chromosomes in the *B. rapa* genome. Structure comparisons based on alignments of the all *BrCYS* (*B. rapa* phytocystatin) proteins using the CLUSTALW program revealed conservation of the three consensus motifs known to interact with the active site of cysteine peptidases. According to the phylogenetic analysis based on the deduced amino acid sequences, the eight BrCYS proteins were divided into several clusters related to the orthologous phytocystatin. The predicted three-dimensional structure models of the eight BrCYS proteins demonstrate that all of these proteins are similar to the reported crystal structure of oryzacystatin-I (OC-I). Digital northern and RT-PCR analyses indicated that the eight *BrCYS* genes exhibit different expression patterns in *B. rapa* tissues and respond differently to abiotic stimuli. The differences in gene structure and expression between the eight *BrCYS* genes suggest that these proteins may play diverse physiological roles in *B. rapa* and may interact with cysteine peptidases through different mechanisms.

Introduction

Peptidase inhibitors, which are classified based on the class of peptidase they inhibit, are involved in the control of protein breakdown and are thought to help protect cells from inappropriate proteolysis of endogenous or external peptidases (Turk and Bode 1991). The plant cystatins (Phytocystatins; subsequently abbreviated "PhyCys") are proteinaceous inhibitors of cysteine peptidases (CPs) of the papain family (MEROPS peptidase database, http://merops.sanger.ac.uk). These inhibitors are widely distributed and have been isolated from a number of sources, including rice (Abe et al. 1987; Kondo et al. 1990), maize (Abe et al. 1992), cowpea (Flores et al. 2001), soybean (Misaka et al. 1996), Chinese cabbage (Lim et al. 1996a), barley (Gaddour et al. 2001), and wheat (Kuroda et al. 2001). PhyCys proteins possess a conserved [LVI]-[AGT]-[RKE]-[FY]-[AS]-[VI]-x-[EDQV]-[HYFQ]-N sequence (α1-helix) in their N-terminal regions, and this sequence is absent in animal cystatins (Margis et al. 1998). This unique sequence can be used to identify PhyCys proteins from protein data banks and to distinguish them from other members of the superfamily. In

*Corresponding author Tel 031-299-1631 Fax 031-299-1672
E-mail: pbeom@rda.go.kr

addition to this consensus sequence, PhyCys contain three conserved regions: a QVVAGTMYYIT motif (Loop 1) in the first hairpin loop, an EAKVWVKPW motif (Loop 2) in the second hairpin loop, and a glycine residue in the amino (N)-terminal region, common to almost members of the cystatin superfamily. These conserved sequences are involved in the formation of the tripartite structure of cystatin, which interacts with the active sites of CPs (Machleidt et al. 1989; Margis et al. 1998).

The suggested role of PhyCys is to regulate endogenous and heterologous CPs in a variety of physiological processes (Arai et al. 2002), including defense mechanisms via inhibition of digestive enzymes in the guts of various insect pests and nematodes (Kuroda et al. 1996; Zhao et al.1996; Pernas et al. 1998; Vain et al. 1998; Lecardonnel et al. 1999; Agrawal et al. 2002;), protection of cytosolic metabolism from intracellular peptidases released by incidental rupturing of protein bodies (Kondo et al. 1990; Turk and Bode 1991), regulation of peptidase activity during programmed cell death (Soares-Costa et al. 2002), and tolerance to abiotic stresses such as cold or drought (Gaddour et al. 2001). PhyCys have also been implicated in regulation of endogenous protein turnover by inhibition of endogenous proteolytic activity during seed development and germination (Abe et al. 1987; Arai et al. 1991; Corre-Menguy et al. 2002; Hong et al. 2007).

Recently, the genome sequences of *Arabidopsis thaliana* and *Oryza sativa* were completed, and expressed sequence tag (EST) collections from other plants have become available. Using bioinformatic resources, several investigators annotated 7 different *PhyCys* genes in *Arabidopsis*, 7 in barley, and 12 in rice and have described the phylogenetic relationships between these PhyCys (Martínez et al. 2005; Abraham et al. 2006). This comparative phylogenetic analysis revealed three groups of *PhyCys* genes. The first and second groups were small proteins with molecular masses from 11 to 16 kDa transcribed from genes without introns or genes with a single intron located between sequences encoding the LARFAV of the first hairpin loop and the reactive site QxVxG of the second hairpin loop. The third group included genes with three introns, C-terminal extensions, and molecular masses of 23.5 kDa. In *B. rapa*, three cDNA clones encoding the PhyCys have already been identified and characterized (Lim et al. 1996a, 1996b; Hong et al. 2007, 2008); however, the systematic analysis of the *PhyCys* gene family in the *B. rapa* genome has not been fully investigated.

In the current study, we first annotated the PhyCys family in *B. rapa* and predicted the three-dimensional structure of the deduced proteins via bioinformatics resources. In addition, we provide evidence from digital northern and RT-PCR analyses that these genes have different expression patterns in various tissues and in response to abiotic stresses.

Materials and methods

Sequence data, alignment, and phylogenetic analysis

Searches for *B. rapa PhyCys* genes were performed in publicly available genome databases. To assemble the *PhyCys* genes in *B. rapa*, we adopted an *in silico* approach following Martínez et al. (2005) with minor modifications. Exon-intron splice sites were analyzed and confirmed by comparing the *B. rapa* bacterial artificial chromosomes (BAC) clones with EST and cDNA sequences with GENSCAN (http://genes.mit.edu/GENSCAN.html; Burge and Karlin 1998). Sequence-based genetic mapping was conducted to localize seven BAC clones containing the *BrCYS* genes on the referenced map on the *B. rapa* genome using simple sequence repeat (SSR), intron based-polymorphism (IBP), and EST markers (Kim et al. 2006). BrCYS proteins and the other PhyCys deduced amino acid sequences derived from cDNA or genomic DNA were obtained from the GenBank database (http://www.ncbi.nlm.nih.gov) and the *Brassica rapa* Genome Project (BrGP) database (http://www.brassica-rapa.org/BGP/NC_brgp.jsp). Analysis of DNA and comparisons of deduced amino acid sequences were carried out with the current bioinformatics tools. Sequence alignment was performed by using the GeneDoc program (http://www.pcs.edu/biomed/genedoc). Signal peptide analysis was performed with the SignalP Version 3.0 (http://www.cds.dtu.dk/services/SignalP) program. The phylogenetic tree was created with Molecular Evolutionary Genetic Analysis (MEGA)

software, version 4.0 (http://www.megasoftware.net) using the neighbor-joining method. The complete deletion and Poisson correction were used for the treatment of amino acid gaps and substitution models on the PhyCys multiple alignments (Kumar et al. 2004; Abraham et al. 2006). The sampling variance of the distance values was estimated from a 1,000 bootstrap resampling of the alignment columns. The molecular masses of the eight BrCYS proteins were calculated using the BioEdit sequence Alignment Editor Program (Isis Pharmaceuticals, USA).

Homology model building of BrCYS proteins

The three-dimensional (3D) structures of the BrCYS proteins were generated by the automated SWISS-MODEL program (http://swissmodel.expasy.org) (Peitsch 1996; Abraham et al. 2006). The reported crystal structure of rice oryzacystatin I (OC-I, PDB identifier 1EQK) was used to construct the homology-based models (Nagata et al. 2000). The cartoon ribbon diagrams of the three-dimensional structure were generated using the user-friendly molecular visualization RasMol 2.7 program (http://Bernstein-plus-sons.com/software/ rasmol/) (Sayle and Milner-White 1995; Abraham et al. 2006).

Digital northern analysis

Searches for EST sequences corresponding to the *BrCYS* genes were made in the GenBank EST database (dbEST, http://www.ncbi.nlm.nih.gov/dbEST/index.html) and the *B. rapa* EST and Microarray Database (BrEMD, http://www.brassica-rapa.org/BrEMD/microarray_overview.jsp). The databases were searched with the amino acid sequence of each PhyCys using the tblastn program, and the resulting accessions were classified by the source tissue from which the cDNA libraries were derived (Martínez et al. 2005). In this analysis, the GenBank database and BrEMD contained 182,784 and 127,143 ESTs, respectively.

Plant materials, stress treatments, and reverse transcription (RT)-PCR analysis

B. rapa L. ssp. *perkinensis* cv. Chiifu seeds were grown on MS medium (Murashige and Skoog 1961) containing 3% sucrose and 0.25% phyta-gel (pH 5.8) and imbibed at 4℃ for 5 days in the dark to induce synchronous germination. The plants were then transferred to a growth chamber and grown at 22℃ for 5 days under long-day conditions (16 h light and 8 h dark) as described by Yang et al. (2006). The 5 day-old seedlings that were subjected to treatment with abscisic acid (ABA) or gibberellic acid (GA) were grown hydroponically in solutions containing 100 μM ABA (mixed isomers; Sigma, St. Louis, MO) or 10 μM GA_{4+7} (Duchefa, Haarlem, The Netherlands) for 6 h (Yamaguchi-Shinozaki and Shinozaki 1994), respectively. For heat treatment, 5 day-old seedlings were incubated in a 37℃ chamber (Eyela, Hidekata Chino, Japan) for 90 min (Yang et al. 2006).

Total RNA was isolated from 5-day-old seedlings exposed to abiotic stress, as previously described (Chirgwin et al. 1979). For RT-PCR analysis, cDNA was synthesized from 2 μg of total RNA. Each cDNA sample was diluted 1:10, and 1 μl of the diluted cDNA was used for PCR amplification (94℃ for 1 min, 57℃ for 1 min, 72℃ for 1 min) in a volume of 100 μl with gene-specific primer sets (Table 1). The resulting PCR products (5 μl each) were analyzed by electrophoresis and ethidium bromide staining.

Results and discussion

Identification and genomic organization of *PhyCys* genes from *B. rapa*

In order to assemble a complete and non-redundant set of the *PhyCys* gene sequences from *B. rapa*, we used an *in silico* approach by first screening the genome sequences of *B. rapa*. Full-length sequences of different PhyCys proteins containing the conserved consensus sequences and motifs (the LARFAV-consensus sequence, the QxVxG motif, and the PW motif) were used to perform Blast searches in the NCBI and *B. rapa* database BrGP. Blast searches were repeated with each new PhyCys sequence found in the *B. rapa* databases to complete the PhyCys collections. Eight non-redundant PhyCys sequences were found in seven BAC clones (Figure 1).

The structures of the eight *BrCYS* genes and their

Table 1 Oligonucleotide primers used in RT-PCR

Target genes	Primer sequences	Expected size (bp)
BrCYS-a	5'-ATGAATAACAAAACAACCTTCATTG-3' 5'-AAACTCAAGAGGAAACAACAATTAC-3'	383
BrCYS-b	5'-ATTCTTTTCCTCGCTCTCGTCCTC-3' 5'-AAACTCAAGAGGAAACAAGAGTTAC-3'	329
BrCYS-c	5'-AATGAATAACAAGGCAACATCCATTC-3' 5'-AGACGAAACAGCCATTAATTAAATGG-3'	382
BrCYS-d	5'-AGGTAGAAGCAATATGTCACAAGTC-3' 5'-ACCAGCCACTAACCCACTGTTAATAG-3'	504
BrCYS-e	5'-AGTAAGAAGCAATGTCAAAGGTTTC-3' 5'-AACCCATTGTTAATAGCGATCATATC-3'	530
BrCYS-f	5'-ATAATGGCGGATCAACAAGGAG-3' 5'-ACGTAGAATGTGATGTGCTTATCAG-3'	341
BrCYS-g	5'-ATAATGGCGGATCAACAAGGAG-3' 5'-ATGTTGATGTGTTTTACCAGCATGG-3'	348
BrCYS-h	5'-AGTGAAGATGATGCAAAGCCGTTTC-3' 5'-AGAGTTTATAGTCTGCACCAGAAGC -3'	690
*Bactin**	5'-TGGCATCACACTTTTCTACAA-3' 5'-CAACGGAATCTCTCAGCTCC-3'	515
*BrAct1**	5'-CCCTAAGGCTAACAGGGAGAA-3' 5'-AGCTCCGATGGTGATGACTT-3'	424

* *Bactin* and *BrAct1* were used as internal controls for RNA quantity (Yae et al. 2005 and Lee et al. 2008).

BAC clone	NCBI acc. no.	Gene name	Predicted Structure	Amino acid	kDa	signal peptide	C-term extension
KBrH009C03	AC232398	*BrCYS-a*		120	13.0	+	-
KBrH009C03	AC232398	*BrCYS-b*		120	13.0	+	-
KBrH004M24	AC232396	*BrCYS-c*		120	13.0	+	-
KBrH003D18	AC189536	*BrCYS-d*	77	141	15.5	+	-
KBrB046B21	AC189363	*BrCYS-e*	81	142	15.5	+	-
KBrB080E24	AC189473	*BrCYS-f*	102	101	11.3	-	-
KBrB041J18	AC189340	*BrCYS-g*	74	101	11.3	-	-
KBrB063G23	AC189419	*BrCYS-h*	143 85 108	223	25.3	+	+

Figure 1. Genomic organization of *B. rapa PhyCys* genes. Exons are indicated by boxes and introns by lines. The LARFAVDEHN sequence and the active-site QxVxG motif are represented as cross-hatched and black boxes, respectively. The number of base pairs corresponding to each intron is indicated in the predicted structure. The number of amino acids encoded by the ORFs, the molecular mass (kDa), and the presence of a predicted signal peptide and C-terminal extension are also indicated.

chromosome locations were determined from the genomic sequences of *B. rapa* (Figures 1 and 2). Structural analysis of the genes based on the conservative position of each exon and intron supported the division of the *BrCYS* genes into three distinct groups. Genes encoding BrCYS-a, BrCYS-b, and BrYCS-c possessed no introns in their ORFs. Four PhyCys sequences (BrCYS-d to BrCYS-g) each contained one intron within their ORFs. Although these intron lengths were variable, the introns were present between the sequences encoding the conserved LARFAV consensus sequences and the reactive site QxVxG motifs. The longest encoding sequence, BrCYS-h, contained three introns of 143, 85, and 108 nt. The first intron was located in a similar position to the introns found in the single-intron *PhyCys* genes, and introns 2 and 3 were positioned after the conserved PW motifs located downstream of the reactive site QxVxG motifs. These results revealed that the genomic organization of the *BrCYS* genes was similar in exon-intron structure to the reported *PhyCys* genes from *Arabidopsis*, rice, and barley (Martínez et al. 2005; Abraham et al. 2006).

The molecular mass calculated for most of the BrCYS proteins was in the 11.3 to 15.5 kDa range, while the BrCYS-h protein, which has an extended C-terminal region, has a molecular mass of 25.3 kDa. The predicted sequences of these proteins contained 101 to 223 amino acids, and all PhyCys proteins, with the exception of BrCYS-f and BrCYS-g,

contained a predicted signal peptide of 21 to 27 amino acids (Figure 1). Since the BAC clones containing the *PhyCys* genes are linked to genetic markers, the relative position of these sequences in the genetic map could be estimated. The eight *PhyCys* genes are distributed on seven of the ten *B. rapa* chromosomes (Figure 2). One of the eight annotated BrCYS proteins, BrCYS-f, has been previously cloned and structurally characterized (Lim et al. 1996a; Hong et al. 2008). BrCYS-f occurred only as a 16 kDa monomer and was highly stable over a wide range of pH and temperature values. Results of the genome structural analysis revealed that the *B. rapa* genome allowed prediction of at least eight *PhyCys* genes. Although only eight *PhyCys* genes have been annotated in *B. rapa*, the actual number of expected genes in this member of the Brassicaceae family is probably higher than the number found in rice, barley, or *Arabidopsis* (Martínez et al. 2005; Abraham et al. 2006).

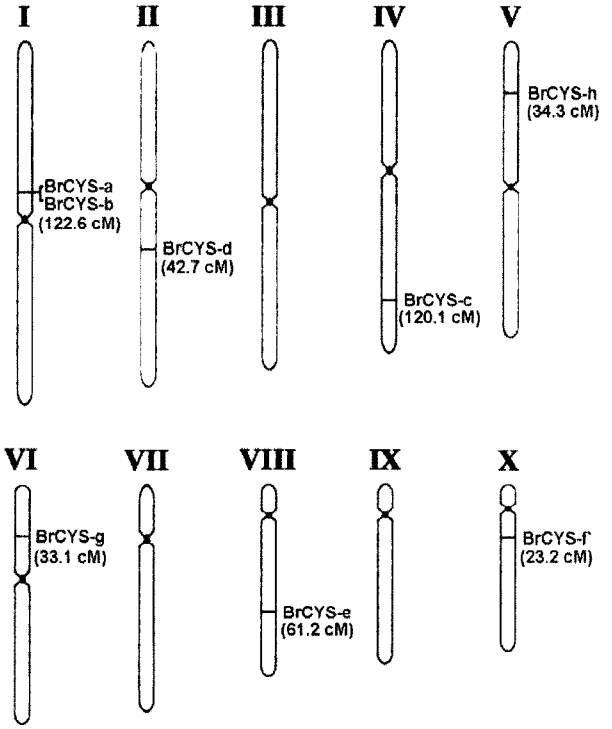

Figure 2. Chromosomal distribution of *BrCYS* genes in the *B. rapa* genome. All positions in centMorgans (cM) were estimated in accordance to genetic markers assigned to the BAC clones. The eight *BrCYS* genes are located throughout the *B. rapa* genome as single genes or in cluster, which is specified with a bracket. The chromosome numbers are indicated by Roman numerals. The centromeres are represented by block boxes.

Comparison and phylogenetic analysis of the BrCYS proteins

To determine which amino acid variants affected the inhibitory capability of the BrCYS proteins, we performed an alignment of the deduced amino acid sequences of the BrCYS proteins with those of PhyCys from other species (Figure 3). The pairwise identity among the deduced amino acid sequences of the eight BrCYS proteins varied from 25.0 to 95.8%, indicating that the *BrCYS* gene family is quite divergent (Table 2). In the conserved consensus sequences, the eight BrCYS proteins showed significant similarity to each other as well as to other members of the PhyCys family. All proteins contain the LARFAVDEHN sequence within their α1-helical regions, which are typical features of members of the PhyCys family (Margis et al. 1998; Kouzuma et al. 2000). The BrCYS proteins also possess the general QVVAGTMYYIT motif (Loop 1) in the first binding site, the EAKVWVKPW motif (Loop 2) in the second binding loops, and two glycine residues at the N-terminal region. All of these sequence motifs are commonly found in members of the cystatin superfamily. These loops form a wedge that fits into the peptidase active site and are indispensable for the inhibitory activity of cystatins (Bode and Huber 1988; Stubbs et al. 1990). In particular, the BrCYS-h protein contains a long C-terminal region that includes a cysteine residue, which may be involved in the formation of intermolecular disulfide bonds or regulation of the control of CP inhibition (Turk et al. 1993; Maris et al. 1998; Hong et al. 2008).

To determine the phylogenetic relationships between the BrCYS proteins and other PhyCys, the deduced amino acid sequences, including the signal peptide, of the eight BrCYS proteins were compared with the amino acid sequences of PhyCys proteins from *Arabidopsis*, rice, and barley. A phylogenetic tree was generated with the MEGA analysis platform using the neighbor-joining method (Figure 4). Bootstrap values greater than 50% allowed the tree to be organized into three major clusters: A, B, and C. These groups correlated with the major clusters obtained previously in phylogenetic analysis of the homologous *PhysCys* genes from

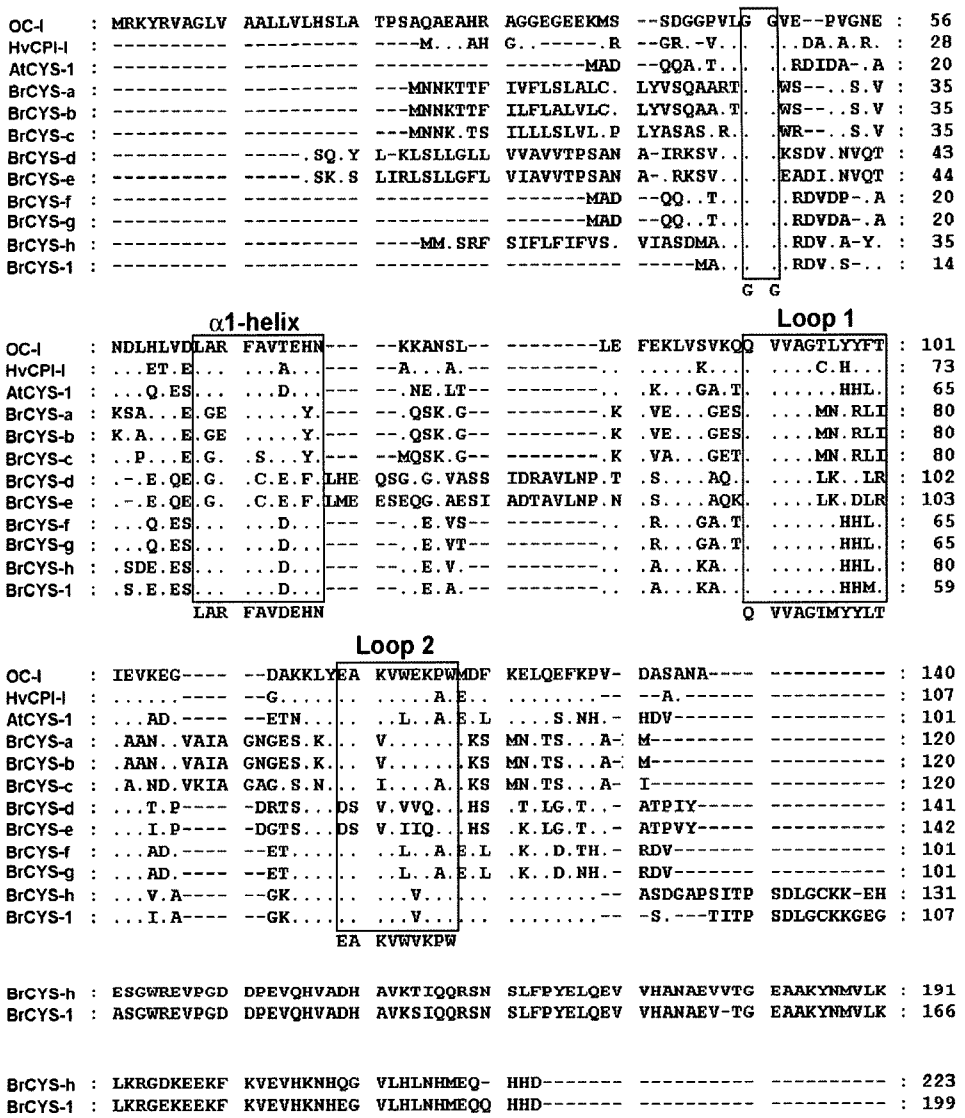

Figure 3. Comparison of the amino acid sequences of the BrCYS proteins. Alignment of the deduced amino acid sequences of eight BrCYSs and the PhyCys from other plant species. The GenBank accession numbers for these sequences are as follows: *Arabidopsis thaliana*, AtCYS-1 (At5g12140); *Brassica rapa*, BrCYS-1 (L41355), BrCYS-a (EE521363), BrCYS-b (ES933634), BrCYS-c (EX125500), BrCYS-d (EX115104), BrCYS-e (EX115104), BrCYS-f (AF408862), BrCYS-g (CV137144), BrCYS-h (EX137144); *Hordeum vulgare*, HvCPI-1 (Y12068); and *Oryza sativa*, OC-I (Os01g58890). The residues that have been identified as critical for PhyCys activity are boxed. Dots indicate identical residues. Dashes represent gaps introduced to maximize the alignment.

Table 2 Pairwise sequence identity values for eight *B. rapa* PhyCys proteins based on their deduced amino acid sequences

	BrCYS-a	BrCYS-b	BrCYS-c	BrCYS-d	BrCYS-e	BrCYS-f	BrCYS-g	BrCYS-h
BrCYS-a	100	95.8	70.8	25.2	25.2	25.0	25.0	33.2
BrCYS-b		100	70.8	26.1	27.0	26.0	25.0	34.2
BrCYS-c			100	26.1	25.2	32.3	32.3	33.3
BrCYS-d				100	78.7	31.2	28.1	31.4
BrCYS-e					100	30.2	27.1	30.5
BrCYS-f						100	93.1	58.4
BrCYS-g							100	59.4
BrCYS-h								100

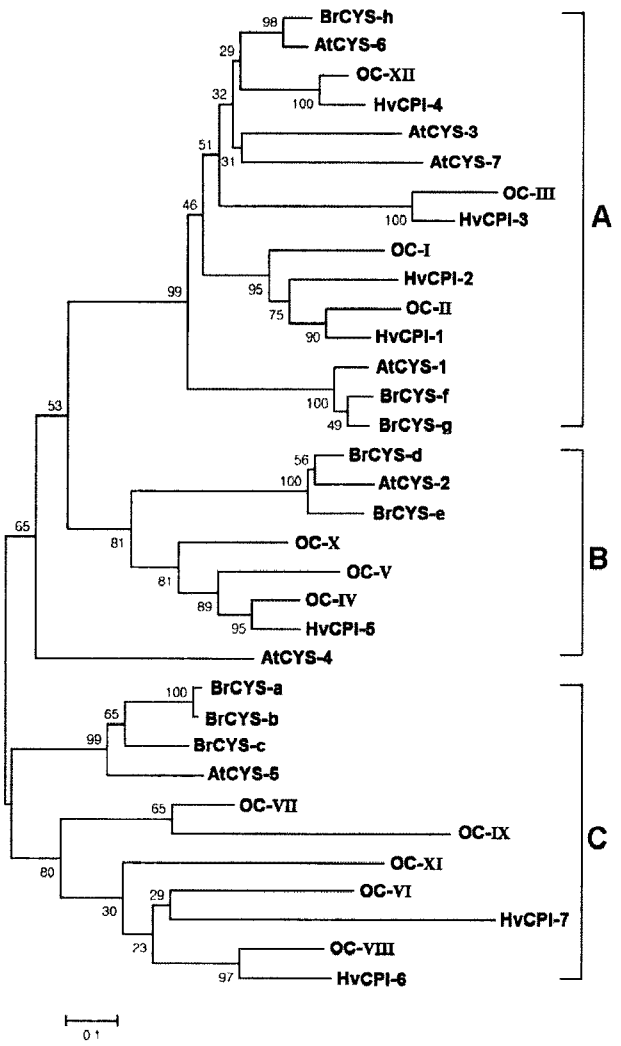

Figure 4. Phylogenetic relationships between BrCYS and other PhyCys. A phylogenetic tree was generated based on the neighbor-joining analysis of the predicted amino acid sequences of 8 BrCYS proteins and another 26 PhyCys (7 from *Arabidopsis*, 7 from barley, and 12 from rice). Bootstrap values are indicated as percentages. The length of each pair of branches represents the phylogenetic distance between sequence pairs. The major clusters of the orthologous genes identified are indicated on the right. Scale bar corresponds to 0.1 amino acid substitution per residue. GenBank accession numbers are as follows: *Arabidopsis thaliana*, AtCYS-1 (At5g12140), AtCYS-2 (At2g31980), AtCYS-3 (At2g40880), AtCYS-4 (At4g16500), AtCYS-5 (At5g47550), AtCYS-6 (At3g12490), and AtCYS-7 (At5g05110); *Brassica rapa*, BrCYS-a (EE521363), BrCYS-b (ES933634), BrCYS-c (EX125500), BrCYS-d (EX115104), BrCYS-e (EX115104), BrCYS-f (AF408862), BrCYS-g (CV137144), BrCYS-h (EX137144); *Hordeum vulgare*, HvCPI-1 (Y12068), HvCPI-2 (AJ748337), HvCPI-3 (AJ748338), HvCPI-4 (AJ748344), HvCPI-5 (AJ748340), HvCPI-6 (AJ748341), and HvCPI-7 (AJ748345); and *Oryza sativa*, OC-I (Os01g58890), OC-II (Os05g41460), OC-III (Os05g3880), OC-IV (Os01g68660), OC-V (Os01g68670), OC-VI (Os03g11180), OC-VII (Os03g11170), OC-VIII (Os03g31510), OC-IX (Os01g11160), OC-X (Os04g2250), OC-XI (Os09g08110), and OC-XII (Os01g16430).

rice, barley, and *Arabidopsis* (Martínez et al. 2005; Abraham et al. 2006). The phylogenetic analysis indicated that the eight BrCYS proteins studied, which likely do not represent the total number of PhyCys of the *B. rapa* genome, are representative of the three different clusters of homologous *PhyCys* genes. The results indicated that BrCYS proteins are most closely related to *Arabidopsis* PhyCys proteins; however, BrCYS and AtCYS proteins were dispersed in the phylogenetic tree, which made it difficult to determine the orthologous partners in the monocot (Martínez et al. 2005). These observations suggest that divergent PhyCys in *B. rapa* may differ in their distinct cellular function. In support of this notion, different animal cystatin proteins have unique properties and are not functionally equivalent (Martínez et al. 2005).

The three-dimensional structures of the BrCYS proteins were predicted using the reported crystal structure of rice oryzacystatin-I (OC-I) as a molecular model (Figure 5). Although the helix-sheet architecture of the BrCYS proteins was predicted with variable accuracy, these proteins contain many residues that are identical to those in OC-I, and thus, the protein structures are likely very similar with a conserved α -helix spanning the LARFAVDEHN sequence and the five-stranded antiparallel β-sheet (β1, β2, β3, β4, and β5) that wraps around a central α-helix (Nagata et al. 2000; Abraham et al. 2006). The wedge-shaped edges of these proteins are formed by the two tight β-hairpin loops (QVVAGTMYYIT and EAKVWVKPW motifs) and the N-terminal region (conserved GG sequence). These loops form a wedge that fits into the peptidase active site and are indispensable for the inhibitory activity of cystatins (Stubbs et al. 1990; Abraham et al. 2006). This report describes the first predicted models of BrCYS proteins, and these models will provide valuable structural information for deciphering the variable mechanisms of inhibitory activity.

Expression analysis of *BrCYS* genes in *B. rapa* tissues

To analyze the expression of *BrCYS* genes in various tissues of *B. rapa*, we searched the EST database in GenBank and

BreMD for each *BrCYS* gene and then conducted a digital northern analysis on expression of these genes. The digital northern analysis is based on the assumption that, for the majority of genes, mRNA expression is reflected by the presence of ESTs, and these sequences provide indirect information on cellular function via spatial and temporal expression profiles of their encoding genes (Audic et al. 1997). The results were summarized based on presence and/or absence of ESTs in seven tissues (flower, leaf, root, seed, silique, zygotic embryogenesis, and primary inflorescence) of *B. rapa*. As shown in Table 3, the eight *BrCYS* genes exhibit different expression patterns, but none of these genes were detected in

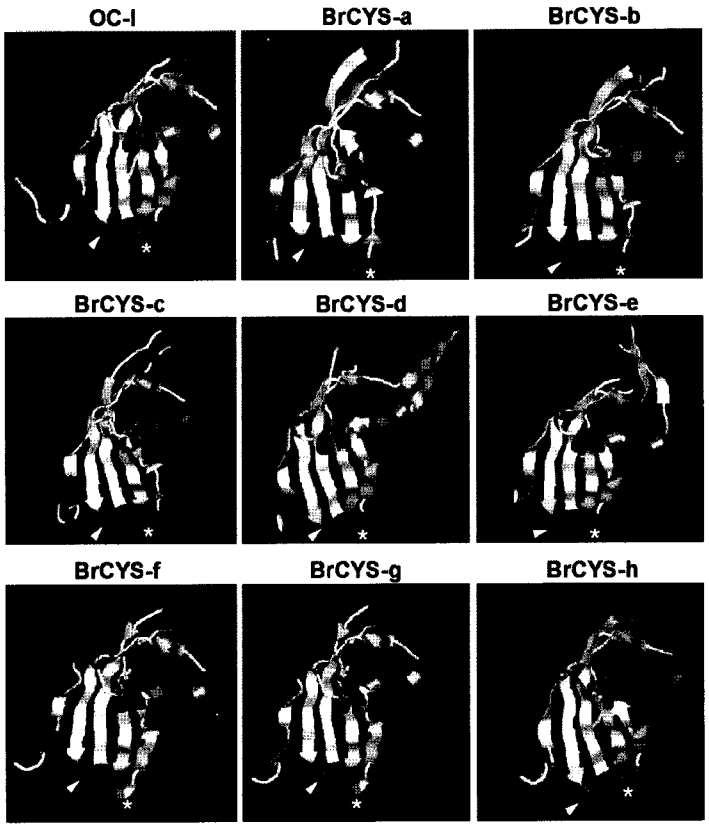

Figure 5. Structural comparison between the BrCYS proteins and oryzacystatin-I (OC-I). The three-dimensional structures of BrCYS proteins were predicted using the automated SWISS-MODEL program with OC-I as a template. The ribbon diagrams of the 3D structures were prepared with RasMol 2.7. The α-helices, β-strand, and loops are colored in red, yellow, and blue, respectively. Two important loops involved in the inhibitor-peptidase interaction site are indicated: the reactive site (arrow heads) and W residue (asterisks).

Table 3 Expression profiles of the eight *BrCYS* genes as determined by EST data

Gene name	Tissue[1]						
	Flower	Leaf	Root	Seed	Silique	Zygotic embryogenesis	Primary inflorescence
BrCYS-a	-	-	-	-	+	+	-
BrCYS-b	-	-	-	-	+	+	-
BrCYS-c	-	+	+	-	-	+	-
BrCYS-d	+	-	-	-	-	+	-
BrCYS-e	+	-	-	-	-	+	-
BrCYS-f	+	-	-	-	+	-	-
BrCYS-g	-	-	-	-	-	-	-
BrCYS-h	+	-	+	-	-	-	+

[1]Presence (+) or absence (-) of *BrCYS* gene sequences in EST collections derived from the indicated tissues (Martínez et al. 2005).

the EST collection from seed. *BrCYS-a* and *-b* were detected in silique and zygotic embryogenesis, whereas *BrCYS-c* was found in leaf, root, and zygotic embryogenesis. *BrCYS-d* and *-e* were expressed in flower and zygotic embryogenesis, and *BrCYS-f* was detected in flower and silique. *BrCYS-g* was not represented in these tissues. *BrCYS-h* was expressed in flower, root, and primary inflorescence. Overall, these studies indicated that expression of the eight *BrCYS* genes is tissue-specific and developmentally regulated.

Expression analysis of *BrCYS* genes in response to abiotic stresses

To determine the responsiveness of *BrCYS* genes to abiotic stresses, we analyzed gene expression using RT-PCR under various stress conditions including 100 μM ABA, 10 μM GA_{4+7}, and 37℃ heat stress. *BrCYS-g* expression was increased in response to ABA, while expression of *BrCYS-h* was increased in response to ABA and GA_{4+7} (Figure 6). In contrast, expression of *BrCYS-d* was significantly decreased by ABA, GA_{4+7}, and 37℃ heat treatments. In addition, *BrCYS-c* and *-f* were expressed at nearly consistent levels following all of the stress treatments examined. Steady-state levels of *BrCYS-a,-b*, and *-e* in these same samples were not detectable by RT-PCR analysis. These data suggest that the different expression patterns of *BrCYS* genes may reflect their roles in response to abiotic stresses.

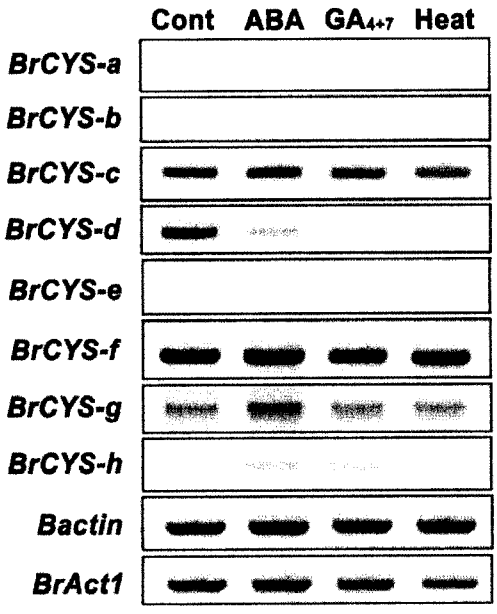

Figure 6. Analysis of the expression patterns of *BrCYS* genes in response to abiotic stress. Total RNA was isolated from 5-day-old *B. rapa* seedlings exposed to 100 μM ABA, 10 μM GA_{4+7}, or heat stress (37℃) as indicated (see Materials and Methods). A 2-μg aliquot of total RNA was reverse-transcribed into first-strand cDNA for RT-PCR for gene expression. Actin (*Bactin* and *BrAct1* genes) were amplified as controls.

In this study, although the precise physiological functions of the eight identified *BrCYS* genes cannot be inferred from these data, the presence of different *PhyCys* genes suggests that a wide range of interactions between CPs and BrCYSs likely exists in *B. rapa*. These interactions may be involved in such physiological functions as control of the degradation of storage proteins, regulation of development, programmed cell death, and/or inhibition of digestive peptidases introduced into plants by insects or pests (Kondo et al., 1991; Zhao et al., 1996). Further studies of these *BrCYS* genes will help determine the biological function of PhyCys in plant development, growth, and response to various stress conditions.

Acknowledgments

This work was supported by grants from the National Institute of Agricultural Biotechnology (Code # 04-1-12-2-3) and the BioGreen21 Program (Code # 20050301034438, Code # 20070301034037), the Rural Development Administration, Republic of Korea.

References

Abe M, Abe K, Kuroda M, Arai S (1992) Corn kernel cysteine proteinases inhibitor as a novel cystatin superfamily member of plant origin. Eur J Biochem 209: 933-937

Abe K, Emori Y, Kondo H, Suzuki K, Arai S (1987) Molecular cloning of a cysteine proteinase inhibitor of rice (oryzacystatin). Homology with animal cystatins and transient expression in the ripening process of rice seeds. J Biol Chem 262: 16793-16797

Abraham Z, Martinez M, Carbonero P, Diaz I (2006) Structural and functional diversity within the cystatin gene family of *Hordeum vulgare*. J Exp Bot 57: 4245-4255

Arai S, Matsumoto I, Kondo H, Emori Y, Abe K (2002) Plant seed cystatins and their target enzymes of endogenous and exogenous origin. J Arg Food Chem 50: 6612-6617

Arai S, Watanabe H, Kondo H, Emori Y, and Abe K (1991) Papain inhibitory activity of oryzacystatin, a rice seed cysteine proteinases inhibitor, depends on the central Gln-Val-Val-Ala-Gly region conserved among cystatin superfamily members. J Biochem 109: 294-298

Audic S, Claverie JM (1997) The significant digital gene expression profiles. Genome Res 7: 986-995

Bode W, Engh R, Musil D, Thiele U, Huber R, Karshikov A, Brzin J, Kos J, 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

Burge CB, Karlin (1998) Finding the genes in genomic DNA. Curr Opin Struc Biol 238: 346-354

Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochem 18: 5294-5299

Flores VMG, Louro RP, Xavier-Filho J, Barratt DHP, Shewry PR. Fernandes KVS (2001) Temporal and tissue localization of a cowpea (*Vigna unguiculata*) cystatin. Physiol Plant 112: 195-199

Gaddour K, Vicente-Carbajosa J, Lara P, Isabel-Lamoneda I, Diaz I. Carbonero P (2001) A constitutive cystatin-encoding gene from barley (*Icy*) responds differentially to abiotic stimuli. Plant Mol Biol 45: 599-608

Hong JK, Hwang JE, Lim CJ, Yang KA, Jin JL, Kim CY, Koo JC, Chung WS, Lee KO, Lee SY, Cho MJ, Lim CO (2007) Over-expression of Chinese cabbage phytocystatin 1 retards seed germination in *Arabidopsis*. Plant Sci 172: 556-563

Hong JK, Hwang JE, Lim CJ, Lee KO, Chung WS, Park B-S, Lim CO (2008) Molecular characterization of phytocystatin isolated from Chinese cabbage flower buds. Gene Genomics 30: 235-243

Kim JS, Chun TY, J.king G, Jin M, Yang TJ, Jin Y-M, Kim H-I, Park B-S (2006) A sequence-tagged linkage map of Brassica rapa. Genetics 174: 29-39

Kondo H, Abe K, Emori Y, Arai S (1991) Gene organization of oryzacystatin-II, a new cystatin superfamily member of plant origin, is closely related to that of oryzacystatin-I but different from those of animal cystatins. FEBS Lett 278: 87-90

Kondo H, Abe K, Nishimure I, Watanabe H, Emori Y, Arai S (1990) Two distinct cystatin species in rice seeds with different specificities against cysteine proteinases: molecular cloning, expression, and biochemical studies on oryzacystatin-II. J Biol Chem 256: 15832-15837

Kouzuma Y, Inanaga H, Doi-Kawano K, Yamasaki N, Kimura M (2000) Molecular cloning and functional expression of cDNA encoding the cysteine proteinase inhibitor with three cystatin domains from sunflower seeds. J Biochem 128: 161-166

Kumar S, Tamura K, Nei M (2004) MEGA3, integrated software for molecular evolutionary genetic analysis and sequence alignment. Brief Bioinform 5: 150-163

Kuroda M, Ishimoto M, Suzuki K, Kondo H, Abe K, Kitamura K, Arai S (1996) Oryzacystatins exhibit growth inhibitory and lethal effects on different species of bean insect pests. *Callosobruchus chinensis* (Coleoptera) and *Riptortus calavatus* (Hemiptera), Biosci. Biotechnol. Biochem. 60: 209-212

Kuroda M, Kiyosaki T, Matsumoto I, Misaka T, Arai S, Abe K (2001) Molecular cloning, characterization, and expression of wheat cystatins. Biosci Biotechnol Biochem 65: 22-28

Lecardonnel A, Chauvin L, Jouanin L, Beaujean A, Prevost G, Sangwan-Norreel B (1999) Effects of rice cystatin I expression in transgenic potato on Colorado potato beetle larvae. Plant Sci 140: 71-79

Lee SC, Lim MH, Kim JA, Lee SI, Kim JS, Jin M, Kwon SJ, Mun JH, Kim YK, Kim HU, Hur Y, Park BS (2008) Transcriptome analysis in *Brassica rapa* under the abiotic stresses using *Brassica* 24K oligo microarray. Mol Cells (in press)

Lim CO, Kim HY, Kim MG, Lee SI, Chung WS, Park SH, Hwang I, Cho MJ (1996a) Expressed sequence tags of Chinese cabbage flower bud cDNA. Plant Physiol 111: 577-588

Lim CO, Lee SI, Chung WS, Park SH, Hwang I. Cho MJ (1996b) Characterization of a cDNA encoding cysteine proteinase inhibitor from Chinese cabbage (*Brassica campestris* L. ssp. *pekinensis*) flower buds Plant Mol. Biol. 30: 373-379

Machleidt W, Thiele U, Laber B, Assfalg-Machleidt I, Esterl A, Wiegand G, Kos J, Turk ,V Bode W (1989) Mechanism of inhibition of papain by chicken egg white cystatin. Inhibition constants of N-terminally truncated forms and cyanogen bromide fragments of the inhibitor. FEBS Lett 243: 234-238

Margis R, Reis EM, Villeret V (1998) Structural and phylogenetic relationships among plant and animal cystatins. Arch Biochem Biophys 359: 24-30

Martínez M, Abraham Z, Carbonero P, Díaz I (2005) Comparative phylogenetic analysis of cystatin gene families from Arabidopsis, rice and barley. Mol Gen Genomics 273: 423-432

Misaka T, Kuroda M, Iwabuchi K, Abe K. Arai S (1996) Soyacystatin, a novel cysteine proteinase inhibitor in soybean, is distinct in protein structure and gene organization from other cystatins of animal and plant origin. Eur J. Biochem 240: 609-614

Nagata K, Kudo N, Abe K, Arai S, Tanokura M (2000)

Three-dimensional solution structure of oryzacystatin-I, a cysteine proteinase inhibitor of the rice, *Oryza sativa* L. japonica. Biochemistry 39: 14753-14760

Peitsch M (1996) ProMod and Swiss-Model: internet-based tools for automated comparative protein modeling. Biochem Soc T 224: 274-279

Pernas M, Sánchez-Monge R, Gómez L, Salcedo G (1998) A chestnut seed cystatin differentially effective against cysteine proteinases from closely related pests. Plant Mol Biol 38: 1235-1242

Sayle R, Milner-White EJ (1995) RasMol: biomolecular graphics for all. Trends Biochem Sci 20: 374

Soares-Costa A, Beltramini LM, Thiemann OH, Henrique-Silva F (2002) A sugarcane cystatin: recombinant expression, purification, and antifungal activity. Biochem Biophys Res Commun 296: 1194-1199

Stubbs MT, Laber B, Bode W, Huber R, Jerala R, Lenarcic B, 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

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. Physiol Plant 15: 473-497

Turk B, Krizaj I, Kralj B, Dolenc I, Popovic T, Bieth JG, Turk V (1993) Bovine stefin C, a new member of the stefin family. J Biol Chem 268: 7323-7329

Turk V, Bode W (1991) The cystatins: Protein inhibitors of cysteine proteinases. FEBS Lett 285: 0231-219

Vain P, Worland B, Clarke MC, Richard G, Beavis M, Liu H, Kohli A, Leech M, Snape J, Christou P, Atkinson H (1998) Expression of an engineered cysteine proteinase inhibitor (Oryzacystatin-IDELTAD86) for nematode resistance in transgenic rice plants. Theor Appl Genet 96: 266-271

Yamaguchi-Shinozaki K, Shinozaki K (1994) A novel *cis*-acting element in Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. Plant cell 6: 251-264

Yang KA, Lim CJ, Hong JK, Park CY, Cheong YH, Chung WS, Lee KO, Lee SY, Cho MJ, Lim CO (2006) Identification of cell wall genes modified by a permissive high temperature in Chinese cabbage. Plant Sci 171: 175-182

Yao K, Lockhart KM, Kalanack JJ (2005) Cloning of dehydrin coding sequences from *Brassica juncea* and *Brassica napus* and their low temperature-inducible expression in germinating seeds. Plant Physiol Biochem 43: 83-89

Zhao Y, Botella MA, Subramanian L, Niu X, Nielsen SS, Bressan RA, Hasegawa PM (1996) Two wound-inducible soybean cysteine proteinase inhibitors have greater insect digestive proteinase inhibitory activities than a constitutive homolog. Plant Physiol 111: 1299-1306

(Received November 21, 2008; Accepted December 4, 2008)