International Journal of Biological Macromolecules 102 (2017) 29–41

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

Expression in *Escherichia coli* of cysteine protease inhibitors from cowpea (*Vigna unguiculata*): The crystal structure of a single-domain cystatin gives insights on its thermal and pH stability

José Edvar Monteiro Júnior[a], Napoleão Fonseca Valadares[b], Humberto D'Muniz Pereira[c], Fábio Henrique Dyszy[c], Antônio José da Costa Filho[d], Adriana Ferreira Uchôa[e], Adeliana Silva de Oliveira[e], Cristina Paiva da Silveira Carvalho[f], Thalles Barbosa Grangeiro[a,*]

[a] Departamento de Biologia, Centro de Ciências, Universidade Federal do Ceará (UFC), Fortaleza, CE, Brazil
[b] Laboratório de Biofísica Molecular, Departamento de Biologia Celular, Universidade de Brasília, Brasília, DF, Brazil
[c] Instituto de Física de São Carlos, Universidade de São Paulo, Brazil
[d] Departamento de Física, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Brazil
[e] Departamento de Bioquímica, Centro de Biociências, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil
[f] Departamento de Bioquímica e Biologia Molecular, Centro de Ciências, UFC, Fortaleza, CE, Brazil

ARTICLE INFO

Article history:
Received 12 December 2016
Received in revised form 26 March 2017
Accepted 3 April 2017
Available online 4 April 2017

Keywords:
Phytocystatin
Recombinant
Domain-swapped dimer

ABSTRACT

Two cysteine proteinase inhibitors from cowpea, *Vu*Cys1 and *Vu*Cys2, were produced in *E. coli* ArcticExpress (DE3). The recombinant products strongly inhibited papain and chymopapain as well as the midgut proteases from *Callosobruchus maculatus* larvae, a bruchid that uses cysteine proteases as major digestive enzymes. Heat treatment at 100 °C for up to 60 min or incubation at various pH values caused little reduction in the papain inhibitory activity of both inhibitors. Moreover, minor conformational variations, as probed by circular dichroism spectroscopy, were observed after *Vu*Cys1 and *Vu*Cys2 were subjected to these treatments. The crystal structure of *Vu*Cys1 was determined at a resolution of 1.95 Å, revealing a domain-swapped dimer in the asymmetric unit. However, the two lobes of the domain-swapped dimer are positioned closer to each other in *Vu*Cys1 in comparison to other similar cystatin structures. Moreover, some polar residues from opposite lobes recruit water molecules, forming a hydrogen bond network that mediates contacts between the lobes, thus generating an extended open interface. Due to the closer distance between the lobes, a small hydrophobic core is also formed, further stabilizing the folded domain-swapped dimer. These structural features might account for the extraordinary thermal and pH stability of *Vu*Cys1.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Cystatins comprise a superfamily of evolutionarily related proteins widely distributed in nature, occurring in Archaea, Bacteria and Eukaryota [1]. These proteins are competitive inhibitors, which are able to reversibly inhibit the cysteine proteinases of the C1A papain-like and C13 legumain-like families [2,3]. In humans, for example, papain-like lysosomal cysteine proteases, generally known as cathepsins, are important targets of endogenous cystatins [4], and disturbances of the normal regulation of these proteases may lead to several pathological conditions, such as neurodegenerative disorders, cardiovascular and inflammatory diseases and cancer [5]. Cystatins act like pseudo-substrates, binding tightly to the active site cleft of cysteine proteases and inactivating their target enzymes. Members of this superfamily have the typical cystatin fold, which is characterized by one N-terminal α-helix wrapped by a five-stranded antiparallel β-sheet [3]. Moreover, as observed in the three-dimensional structures of inhibitor-enzyme complexes, the cystatin contact points with the target papain-like protease involve a tripartite wedge-shaped structure, which includes one or two conserved Gly residues, located in the flexible N-terminal region, and two β-hairpin loops, the first loop containing the QxVxG

* Corresponding author at: Departamento de Biologia, Centro de Ciências, Campus do Pici, Universidade Federal do Ceará, Fortaleza, Ceará CEP: 60440-900, Brazil.
E-mail address: tbgrangeiro@gmail.com (T.B. Grangeiro).

http://dx.doi.org/10.1016/j.ijbiomac.2017.04.008
0141-8130/© 2017 Elsevier B.V. All rights reserved.

motif and the second one containing a conserved aromatic (Trp or His) residue [6–8].

The cystatin superfamily was initially divided into three major families of animal cysteine proteinase inhibitors: family I (stefins), family II (cystatins) and family III (kininogens) [9]. Stefins are small proteins (~11 kDa) devoid of disulfide bonds and carbohydrate content, whereas cystatins are low molecular mass proteins (13–14 kDa) with disulfide bonds but lacking any carbohydrate moiety. Kininogens are large glycoproteins (60–120 kDa) containing disulfide linkages and multiple cystatin domains [3]. Although plant cysteine proteinase inhibitors are structurally related to the animal proteins from families I and II of the cystatin superfamily, they are recognized as an independent branch of this superfamily. Therefore, these proteinase inhibitors are usually referred to as phytocystatins [10]. Phytocystatins lack disulfide bridges and besides the three important regions that interact with the target protease, they contain a conserved sequence, [LVI-AGT-RKE-FY-AS-VI]-x-[EDQV-HYFQ] –N, which is not found in animal cystatins [10]. Moreover, phytocystatins are clustered into 3 subfamilies or groups. Group I includes members with only one cystatin domain and molecular masses ranging from 12 to 16 kDa. Group II phytocystatins have a molecular mass of approximately 23 kDa and possess a conserved cystatin domain at their N-terminal region and an extended C-terminal region that can inhibit C13 legumain-like cysteine proteases [11]. Group III is composed by multicystatins, which contain two or more cystatin domains [12].

Phytocystatins are versatile proteins that act as regulators of the activities of both endogenous and exogenous cysteine proteases. They are involved in a variety of physiological processes, including seed development and germination, fruit development and programmed cell death [13]. In addition to endogenous physiological roles, phytocystatins are also involved in defense mechanisms against abiotic and biotic stresses, such as cold, drought, heat, mechanical wounding, nematode, bacterial and fungal parasitism as well as insect predation [14,15]. These findings have raised the possibility to use phytocystatins as molecular tools for pest control and protection against abiotic stresses as well as nutraceuticals and stabilizers of recombinant proteins produced in plant-based platforms [16]. Taking into consideration the potential biotechnological applications that phytocystatins might have, the discovery of molecules with high stability to temperature and pH variations is of great interest.

Cowpea (*Vigna unguiculata*) is an important legume crop that is cultivated in tropical and subtropical regions. cDNA sequences encoding phytocystatins belonging to groups I and III have been obtained from cowpea [17,18]. cDNA clones encoding a group I cowpea cystatin were first identified in developing seeds of a genotype resistant to the cowpea weevil (*Callosobruchus maculatus*) [17], whereas those encoding a group III multicystatin were obtained from leaves of cowpea plants subjected to drought stress [18]. However, the proteins encoded by these genes have not been characterized in detail. In the present work, these two cowpea cystatins, containing one (*Vu*Cys1; group I member) and two (*Vu*Cys2; group III member) cystatin domains, were expressed in *E. coli* and the functionality of the recombinant products was assessed, showing that they have high thermal and pH stability. Furthermore, the crystal structure of *Vu*Cys1 was determined, revealing structural features that might account for its thermal and pH stability.

2. Materials and methods

2.1. Plant material

Seeds of cowpea genotype CE-31 were provided by Dr. F. R. Freire-Filho (Embrapa Meio-Norte, Teresina, Piauí, Brazil). Cowpea plants were grown hydroponically in ¼ strength Hoagland's nutrient solution [19] under greenhouse conditions. Leaves from 12-day-old plants were harvested, frozen in liquid nitrogen and stored at −80 °C until used.

2.2. Plasmids, bacterial strains and reagents

The plasmids pGEM-T Easy and pET302/NT-His were purchased from Promega (Madison, WI, USA) and Invitrogen (Carlsbad, CA, USA), respectively. *Escherichia coli* strains TOP10F′ and ArcticExpress (DE3) were purchased from Invitrogen and Stratagene (La Jolla, CA, USA), respectively. Nα-benzoyl-DL-arginine β-naphthylamide hydrochloride (BANA), 4-(dimethylamino)-cinnamaldehyde (DMACA), Nα-benzoyl-L-arginine 4-nitroanilide hydrochloride (BApNA) and all the proteolytic enzymes used in this work, including the cysteine proteinases papain (EC 3.4.22.2), chymopapain (EC 3.4.22.6), bromelain (EC 3.4.22.32.) and cathepsin B (EC 3.4.22.1) and the serine proteinase trypsin (EC 3.4.21.4), were from Sigma-Aldrich (St. Louis, MO, USA). All other reagents were of an analytical grade.

2.3. Nucleic acids purification

Total RNA was isolated using the Concert Plant RNA Reagent (Invitrogen), whereas genomic DNA was purified using a CTAB (cetyl trimethyl ammonium bromide)-based protocol [20]. Nucleic acids integrity and concentration were evaluated as described by Landim et al. [21].

2.4. cDNA synthesis, PCR amplification and cloning

Treatment of total RNA with DNase I and cDNA synthesis were performed as previously described [22]. The first-strand cDNA products were then amplified by polymerase chain reaction (PCR) using specific oligonucleotide primers, which were designed according to the sequences deposited in the GenBank DNA database (accession numbers Z21954 and AF278573, encoding *Vu*Cys1 and *Vu*Cys2, respectively). The *Vu*Cys1 coding sequence was amplified using the primers 5′-CCGCTCGAGATGGCAGCACTCGGTGGCAAT-3′ (forward) and 5′-CGCGGATCCCTATGCAGGTGCATCTCCAACATG-3′ (reverse). For the amplification of the *Vu*Cys2 coding sequence, the primers were: 5′-CCGCTCGAGATGGCAACAGCAACAGTAACT-3′ (forward) and 5′-CGCGGATCCTCAAACAGCAACATCCACAGG-3′ (reverse). Sites for restriction endonucleases (Fermentas Life Sciences, Ontario, Canada) were incorporated in the forward (*Xho*I) and reverse (*Bam*HI) primers (underlined), respectively, to allow for the further manipulation of the PCR products. Amplifications were performed in a final volume of 10 μL, containing first-strand cDNA (640 ng), 1 × GoTaq reaction buffer (Promega), 1.5 mM $MgCl_2$, 200 μM of each dNTP, 0.5 μM of each primer, and 1.25 U of GoTaq DNA Polymerase (Promega). The reactions were performed in a Mastercycler gradient thermocycler (Eppendorf, Hamburg, Germany) using the following cycling parameters: an initial denaturation step (2 min at 95 °C) followed by 35 cycles of 1 min at 95 °C, 1 min at 65 °C and 1 min at 72 °C. After the last cycle, the reactions were further incubated for 10 min at 72 °C. Amplifications using genomic DNA as template (500 ng per reaction) were also performed under the same conditions as described above. The amplification of DNA bands with the expected sizes was checked by agarose gel electrophoresis [23]. Cloning of PCR products into the pGEM-T Easy vector and selection of *E. coli* TOP10F′ cells transformed with the recombinant plasmids were performed as described by Freire et al. [22].

2.5. DNA sequencing and sequence analysis

PCR products cloned into pGEM-T Easy were sequenced using the Sanger's dideoxy chain termination method. DNA sequencing and sequence assembly were performed as described in detail by Freire et al. [22]. Multiple sequence alignments were generated and manually edited using the programs Clustal W [24] and BioEdit [25], respectively.

2.6. Construction of the expression vector

PCR products cloned into pGEM-T Easy were recovered by digestion with *Xho*I and *Bam*HI, and the released fragments were purified from the agarose gel slices using the GFX purification kit (GE Healthcare, Buckinghamshire, UK). Purified products were ligated into the pET302/NT-His expression vector, which was previously digested with the same restriction enzymes, using T4 DNA Ligase (Fermentas Life Sciences) at 4 °C for 16 h. Recombinant plasmids were first propagated and selected in *E. coli* TOP10F′ and then introduced in *E. coli* ArcticExpress (DE3) cells by electroporation. Transformants were selected on LB agar containing 20 μg mL^{-1} gentamycin and 100 μg mL^{-1} carbenicillin. Clones containing the recombinant plasmids were identified as previously described [22]. Each protein coding sequence was cloned in frame with the N-terminal 6 × His tag, and its expression was under the control of the T7*lac* promoter. The recombinant plasmids were designated pET-*Vu*Cys1 and pET-*Vu*Cys2.

2.7. Protein expression and purification

A single colony of ArcticExpress (DE3) cells harboring a recombinant plasmid was inoculated in 5 mL of LB broth containing 20 μg mL^{-1} gentamycin and 100 μg mL^{-1} carbenicillin, and the culture was incubated with vigorous shaking (180 rpm) at 37 °C for 16 h. An aliquot (2.5 mL) of this culture was then inoculated in 250 mL of LB broth supplemented with the same antibiotics and further incubated at 180 rpm (37 °C) until the OD_{600} reached 0.4–0.6. The cultures were then incubated at 12 °C, and isopropyl-β-D-thiogalactoside (IPTG) was added to a final concentration of 0.5 mM. The cells were incubated again at 12 °C with continuous shaking (180 rpm) for 24 h. Induced cultures were centrifuged (7000 × *g*, 10 min, 4 °C), and the cell pellets were resuspended in lysis buffer [50 mM Tris-HCl (pH 8.0), 0.5 M NaCl, 2 mM EDTA and 1 mM PMSF] containing 100 μg mL^{-1} lysozyme (USB Chemical, Cleveland, OH, USA). After 20 min on ice, the mixture was frozen in liquid nitrogen and thawed twice. The resultant viscous lysate was treated with DNase I for 30 min at room temperature. The crude cell extract was centrifuged (12,000 × *g*, 15 min, 4 °C), and the clear supernatant was loaded onto a Ni Sepharose 6 Fast Flow (GE Healthcare) column, which had been equilibrated with 50 mM Tris-HCl (pH 8.0) containing 0.5 M NaCl and 10 mM imidazole. Unbound proteins were removed by washing the matrix with the equilibration buffer, and the retained proteins were then eluted using the equilibration buffer supplemented with100 mM imidazole. Protein fractions with $A_{280} > 0.050$ were pooled, dialyzed against distilled water, freeze-dried and used in further analyses.

2.8. Soluble protein content

The concentration of soluble protein was determined using the method of Bradford [26] and bovine serum albumin as a standard.

2.9. SDS-Polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis in the presence of SDS (SDS-PAGE) was performed as described by Laemmli [27] using 17.5% slab gels. Samples preparation and staining and destaining of protein bands were performed as described by Lobo et al. [28].

2.10. Proteinase inhibition assay

Inhibitory activity against papain, chymopapain, cathepsin B and bromelain was evaluated as described by Abe et al. [29] using BANA as the substrate and one unit of inhibitory activity was defined as previously described [30]. Inhibitory activity against trypsin was also determined according to the protocol described by Oliveira et al. [31]. The appropriate amounts of cysteine and serine proteinases used in these assays were previously established in standard curves performed in the absence of inhibitors. All reactions were repeated at least three times and graphically expressed as the relative residual proteolytic activities after incubation with *Vu*Cys1 or *Vu*Cys2.

2.11. Thermal and pH stabilities of the recombinant cystatins

The thermal stability was assessed by incubating a solution [1 mg.mL^{-1} in 50 mM sodium phosphate buffer (PB) (pH 6.0)] of each inhibitor at varying temperatures (24, 30, 45, 60, 75, 90 and 100 °C) for 10 min in a water bath. Prolonged incubations (up to 60 min) at a boiling water temperature (100 °C) were also performed. After heating, the samples were cooled at 4 °C for 30 min, and the residual inhibitory activity against papain was determined under standard assay conditions. To evaluate the effect of pH on the stability of the purified recombinant cystatins, aliquots (5 μL) of *Vu*Cys1 and *Vu*Cys2 [1 mg mL^{-1} in sodium PB (pH 6.0)] were diluted 10-fold in different buffer systems, incubated at 24 °C for 30 min and then subjected to a standard papain inhibition assay. The buffer systems used were: 50 mM glycine-HCl (pH 2.2), 50 mM sodium acetate (pH 5.2), 50 mM sodium phosphate (pH 6.0 and 7.3), 50 mM Tris-HCl (pH 8.0) and 50 mM glycine-NaOH (pH 11.0).

2.12. Circular dichroism (CD) measurements

CD spectra were recorded using a JASCO J-815 (Jasco Inc., Maryland, USA) spectropolarimeter calibrated with (+) −10-camphorsulphonic acid. Measurements in the far-UV region (195–250 nm) were made using a 0.1 cm cell under the following conditions: data pitch, 0.2 nm; scan speed, 50 nm min^{-1}; response, 1 s; band width, 1 nm. The protein concentration used was 0.2 mg mL^{-1} [dissolved in 250 mM sodium PB (pH 6.0) with 3 mM EDTA and 1 mM DTT]. The CD values were reported as the mean residue ellipticity, [θ]. The proteins were heated at 10, 20, 30, 40, 50, 60, 70, 80 and 90 °C for 1 min. After each heat treatment, the thermal denaturation of both cystatins was recorded as an average of 5 scans. pH stability was recorded as an average of 10 scans for a protein solution dissolved in 10 mM phosphate-borate-citrate supplemented with 100 mM NaCl at pH values 2.0, 5.3, 7.2, 9.0 and 11.0. The proteins were maintained under each pH condition for 2 min before recording the scans.

2.13. Preparation of insect midgut enzymes and protease inhibition assay

The bruchid beetles *Callosobruchus maculatus* and *Zabrotes subfasciatus* were reared on seeds of a susceptible cowpea genotype at 24 °C, 70% RH and a photoperiod of 12 h (light/dark). Crude midgut extracts from 3rd instar larvae were obtained according to the procedure described by Oliveira et al. [31] and used in the enzymatic assays, according to the protocol of Xavier-Filho et al. [32]. Protease inhibitory activity was determined by the addition of increasing amounts of *Vu*Cys1 or *Vu*Cys2 to the midgut enzymes before adding the azocasein solution. Residual proteolytic activity

was expressed as the percentage of the blank reaction (without the inhibitor). A mixture of gut homogenate enzymes, which provided an absorbance of 0.25-0.3 at 440 nm after the hydrolysis of 1% azocasein in standard proteolytic reaction curves, was used in all assays [33].

2.14. Effects of the recombinant cystatins on the development of C. maculatus

The individual effects of *Vu*Cys1 and *Vu*Cys2 on the development of *C. maculatus* were examined using the artificial seed system described by Leite et al. [34]. All obtained data were subjected to an analysis of variance (ANOVA), and the means were compared using Tukey's test.

2.15. Crystallization, structure resolution and refinement

Purified proteins (20 mg mL^{-1} in distilled water) were subjected to crystallization trials as previously described [21]. X-ray diffraction data were collected at 100 K using copper radiation, MicroMax 007 HF generator (Rigaku Corporation, Tokyo, Japan) and R-AXIS IV++ imaging plate detector (Rigaku). Data were indexed, integrated and scaled using the program XDS [35]. The structure of *Vu*Cys1 was solved by molecular replacement employing program Phaser [36] using the tarocystatin structure (PDB ID: 3IMA) as search model (sequence identity between *Vu*Cys1 and tarocystatin is 43%), which was modified using the CHAINSAW program [37]. The refinement was carried out using the Phenix [38] and Coot [39] programs, using σ_a-weighted 2Fo-Fc and Fo-Fc electron density maps. After the first refinement cycle, as the electron density unambiguously showed that the content of the asymmetric unit was in fact a domain-swapped dimer, the region corresponding to the open interface (residues Val^{50}-Gly^{53}) was deleted and manually rebuilt using Coot. The behavior of R and R_{free} were used as the principal criterion for validating the refinement. The stereochemical quality of the model was evaluated with PROCHECK [40] and MolProbity [41]. Data collection and final refinement statistics are shown in Table 1. The coordinates and structure factors were deposited in the Protein Data Bank (PDB) under the ID 4TX4. Molecular images were prepared using the PyMOL Molecular Graphics System v1.2r1 (DeLano Scientific LLC, CA, USA).

Table 1
Summary of data collection (a) and final refinement statistics (b).

(a) Data collection[a]	Value
Space Group	$P2_12_12_1$
Cell dimensions (Å) *a, b, c*	41.83, 64.64, 87.11
Detector	RIGAKU RAXIS IV++
X-ray source	RIGAKU MICROMAX-007 HF
Wavelength (Å)	1.541
Resolution range (Å)	30.0–1.95 (2.07–1.95)
Redundancy	3.8 (3.4)
R_{meas} (%)[b]	6.4 (40.7)
Completeness (%)	98.5 (96.6)
Total reflections	66981 (9487)
Unique reflections	17574 (2734)
I/σ(I)	16.79 (3.42)
(b)	
Refinement parameters	Value
Reflections used for refinement	17573
R (%)[c]	17.30
R_{free} (%)[c]	19.51
Number of protein atoms	1298
Number of ligand atoms	0
Wilson B-factor (Å^2)	24.29
Coordinate Error (ML based) (Å)	0.19
Phase error (°)	18.96
Ramachandran Plot	
Favored (%)	98.73
Allowed (%)	1.27
Outliers (%)	0.00
All-atom Clash score	1.55
RMSD from ideal geometry	
r.m.s. bond lengths (Å)	0.014
r.m.s. bond angles (°)	1.322

r.m.s.: root mean square.

[a] Values in parentheses are for the outer resolution shell.

[b] $Rmeas = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n}|I_{hkl,j}-I_{hkl}|}{\sum_{hkl}\sum_{j}I_{hkl,j}}$.

[c] $R = Rfree = \frac{\sum_{hkl}|F_{hkl}^{obs}-F_{hkl}^{calc}|}{\sum_{hkl}F_{hkl}^{obs}}$.

3. Results and discussion

3.1. Cloning, sequence analysis and functional features of Vu*Cys1* and Vu*Cys2*

Specific primers were used to amplify the coding sequences of *Vu*Cys1 and *Vu*Cys2 from genotype CE-31. When total cDNA was used as a template, PCR fragments of approximately 310 and 601 bp were produced (Fig. S1, lanes "1" and "3"), in agreement with the sizes of the mRNA coding sequences for *Vu*Cys1 and *Vu*Cys2, respectively, as previously determined [17,18]. The complete sequences of these amplified partial cDNA fragments were 294 (*Vu*Cys1) and 588 nucleotides (nt) long (*Vu*Cys2), encoding proteins of 97 (*Vu*Cys1) and 195 (*Vu*Cys2) amino acid residues. The cDNA sequence encoding *Vu*Cys1 was identical to the sequence in the GenBank database, which was from cowpea genotype IT81D-1045 (accession number: Z21954) [17]. On the other hand, the cDNA sequence coding for *Vu*Cys2 was determined to contain 10 different bases when compared to the cDNA sequence from genotype EPACE-1 (GenBank accession number: AF278573). Six out of these 10 mutations produced the following amino acid differences between the *Vu*Cys2 sequences from genotypes CE-31 (this work) and EPACE-1 [18]: Val8Leu, Ser17Ala, Gly65Glu, Glu102Gly, Ala158Thr and Val172Ala. However, none of these substitutions occurred in the functional motifs of the cystatin domains (Fig. 1A).

In contrast, PCR amplifications using cowpea genomic DNA (genotype CE-31) as a template yielded larger bands of approximately 1317 (*Vu*Cys1) and 743 bp (*Vu*Cys2), suggesting the presence of intervening sequences in the coding regions of both cowpea cystatin genes (Fig. S1, lanes "2" and "4"). The complete sequences of these genomic fragments contained 1289 (*Vu*Cys1) and 679 nt (*Vu*Cys2) (GenBank accession numbers: KY040303 and KY040304). The genomic fragment encoding *Vu*Cys1 (97 residues) exhibited one phase-0 intron of 995 nt, located between the codons for residues Gln^{34} and Asn^{35}. The genomic sequence encoding *Vu*Cys2 (197 residues) contained a small phase-0 intron of 91 nt, inserted between the codons encoding the residues Glu^{133} and Asn^{134}. In both genomic sequences, the dinucleotides GT and AG were found at the start (splice donor) and at the end (splice acceptor) of the intron (Figs. S2 and S3), which conforms with the GT/AG rule [42]. Moreover, the exon sequences of both genomic fragments were identical to the corresponding cDNA sequences. Therefore, the primary structures of *Vu*Cys1 and *Vu*Cys2, deduced from these genomic fragments, were identical to those deduced from the cognate cDNA fragments. The primary structure of *Vu*Cys1 (calculated molecular mass = 10,758.1 Da) is characterized by two Gly residues (^{5}GG^6) in the vicinity of the N-terminus as well as the reactive site ^{49}QVVSG^{53}, which agrees with the consensus motif QxVxG, and the dipeptide ^{79}PW^{80} in the C-terminal region of the polypeptide chain. These three conserved regions are implicated in the interaction between the inhibitor and its target enzyme and are common

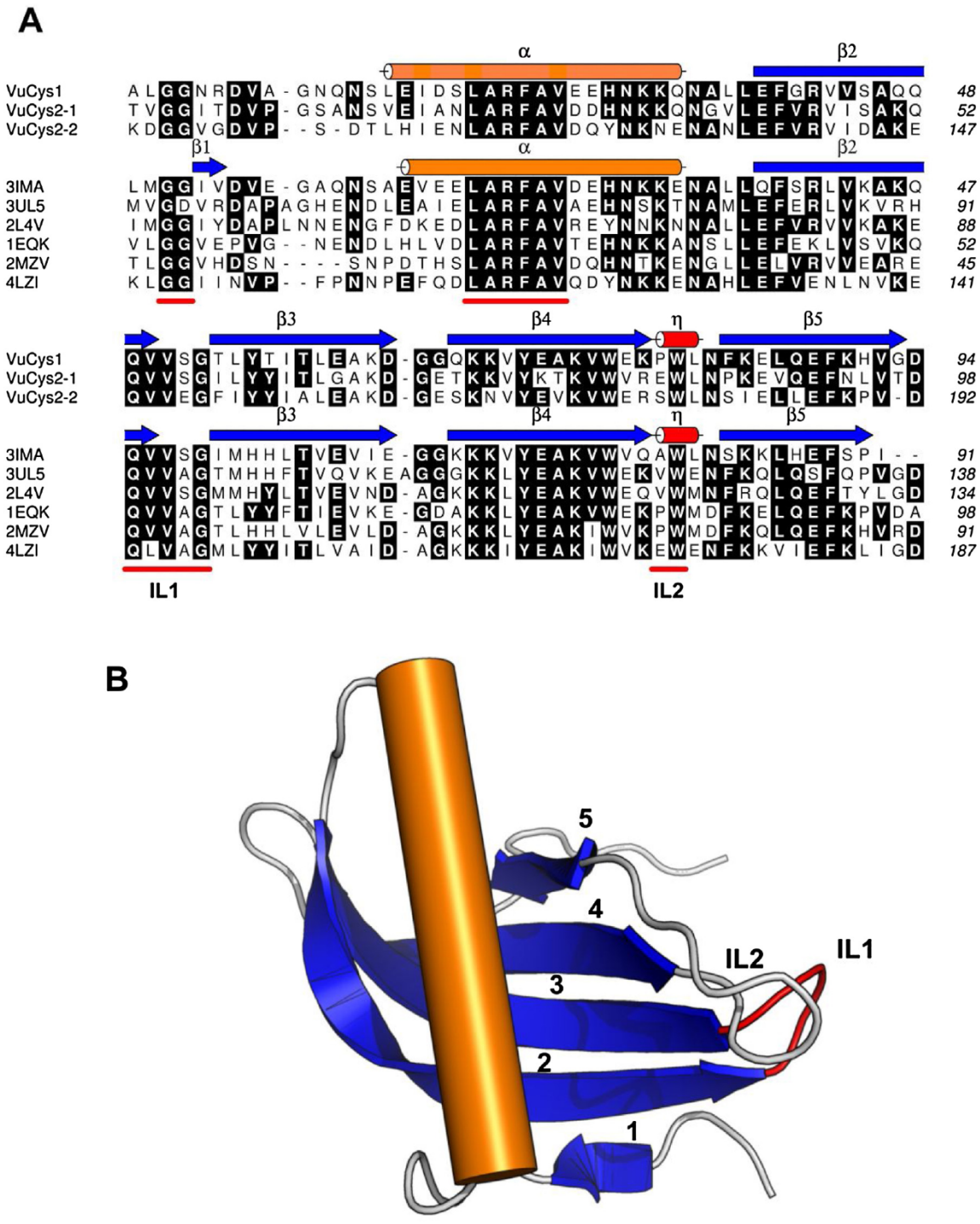

Fig. 1. Multiple alignment of the amino acid sequences of *Vu*Cys1 and *Vu*Cys2 cystatin domains with the primary structures of phytocystatin domains with known three-dimensional structures. (A) The single domain of *Vu*Cys1 and the 2 domains of *Vu*Cys2 (*Vu*Cys2-1 and *Vu*Vys2-2) were aligned with the structures of tarocystatin (PDB ID: 3IMA) [59], canecystatin-1 (PDB ID: 3UL5) [63], pineapple cystain (PDB ID: 2L4 V) [68] and oryzacystatin-I (PDB ID: 1EQK) [60]. The domain 1 of sesame cystatin (PDB ID: MZV) [69] and domain 6 of potato cystatin (PDB ID: 4LZI) [62] were also included. Secondary structure elements (SSEs) of *Vu*Cys1 and tarocystatin, as deduced from their crystal structures (PDB IDs: 4TX4 and IMA, respectively) and represented by blue arrows (β-strands) and orange (α-helices) and red (3_{10}-helices) cylinders, are shown above the corresponding amino acid sequences. The conserved functional motifs that characterize genuine phytocystatins are underlined by red lines, including the inhibitory loops (IL1 and IL2). (B) The cystatin fold, as observed in the three-dimensional structure of oryzacystatin-I (PDB ID: 1EQK), is shown as a ribbon diagram. The β-strands that form the antiparallel β-sheet are labeled (1–5) and the first inhibitory loop (IL1) is colored in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to all cystatins [3]. The consensus sequence LARFAV, which is specific to phytocystatins [10], is also present in the first half of the molecule. With regards to gene structure, the N-terminal half (residues 1–34) of *Vu*Cys1, containing the consensus motif LARFAV, is encoded by the first exon, whereas the C-terminal region (residues 35–97) of the protein, in which the active site motif QxVxG is found, is encoded by the second exon. The position of the intron in the genomic sequence encoding *Vu*Cys1 is similar to

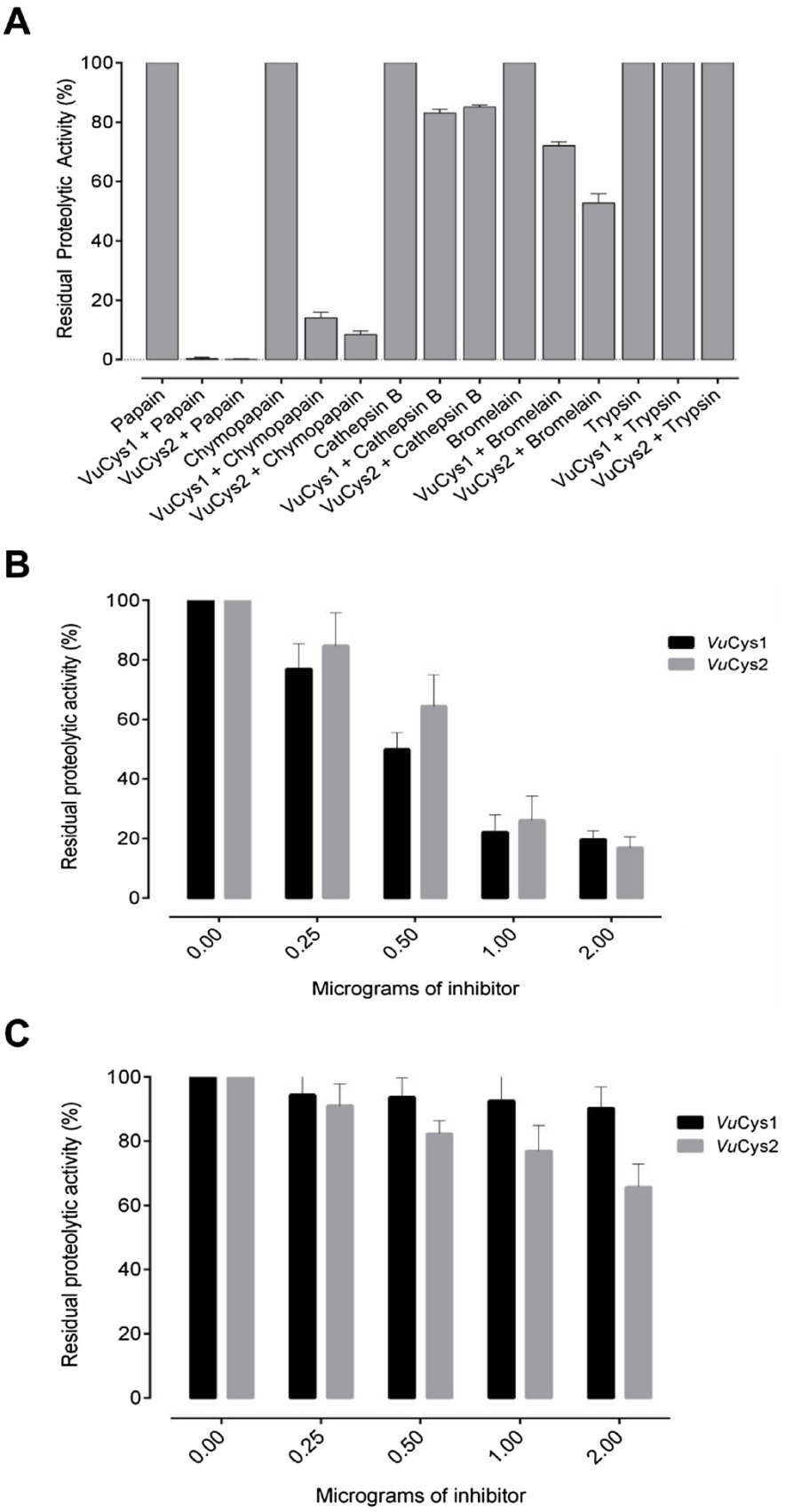

Fig. 2. Inhibitory profiles of *Vu*Cys1 and *Vu*Cys2. The ability of the recombinant proteins to inhibit different cysteine proteinases (A) and the proteolytic enzymes found in larval midgut homogenates of *Callosobruchus maculatus* (B) and *Zabrotes subfasciatus* (C) were evaluated as described in the methods section. Residual proteolytic activity is expressed as the percentage of the control reaction, which contained no inhibitor. The values are an average of three independent assays ± SEM.

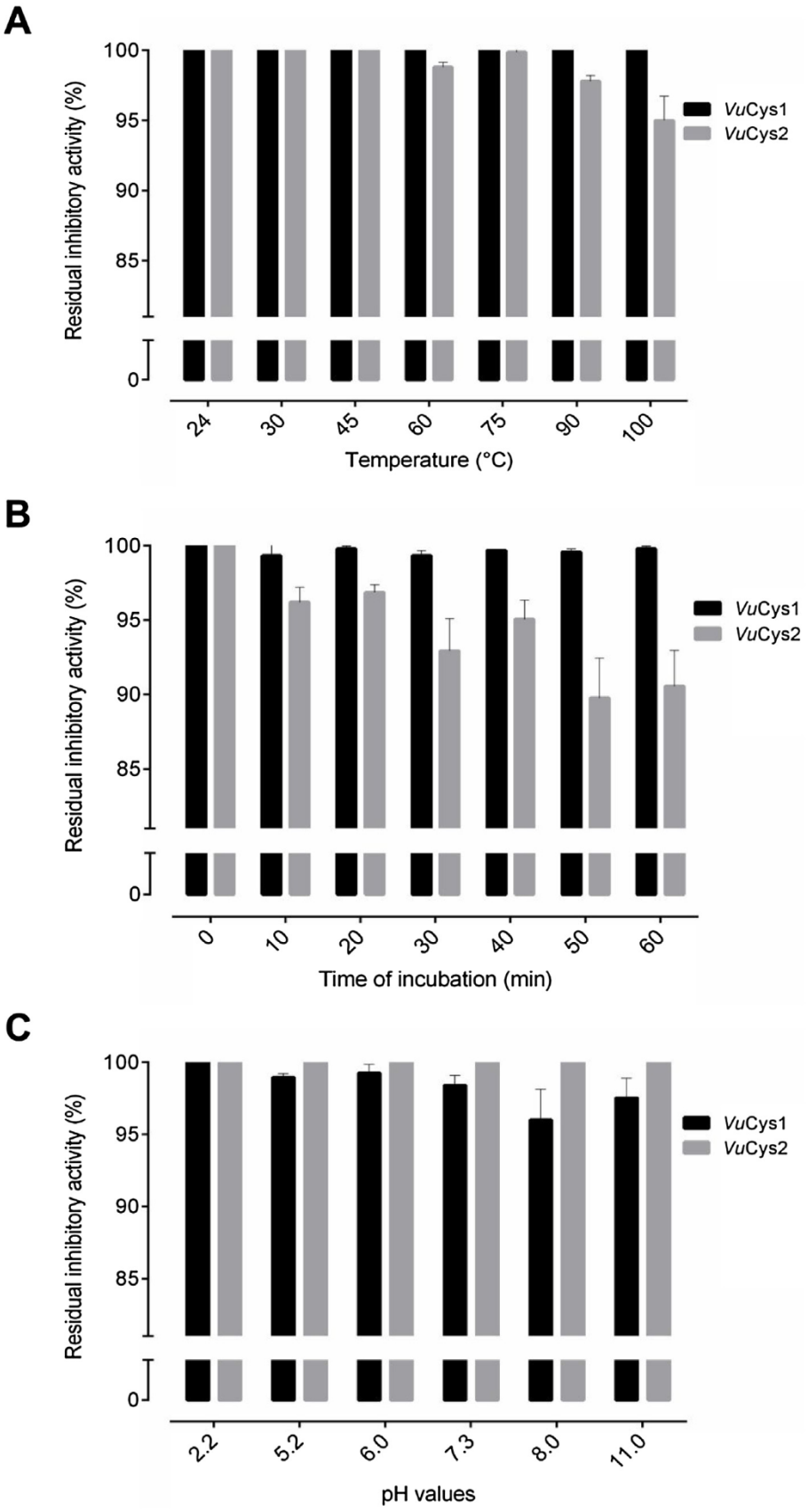

Fig. 3. Thermal and pH stability of *Vu*Cys1 and *Vu*Cys2. Thermal stability was evaluated by incubating the recombinant proteins at increasing temperatures for 10 min et each temperature (A) or incubating them at 100 °C for increasing time periods (B), whereas pH stability was done by incubating each inhibitor at a certain pH for 30 min at room temperature (C). After each treatment, the residual inhibitory activity was determined as described in the methods section and expressed as a percentage. The values are averages of three independent assays ± SEM.

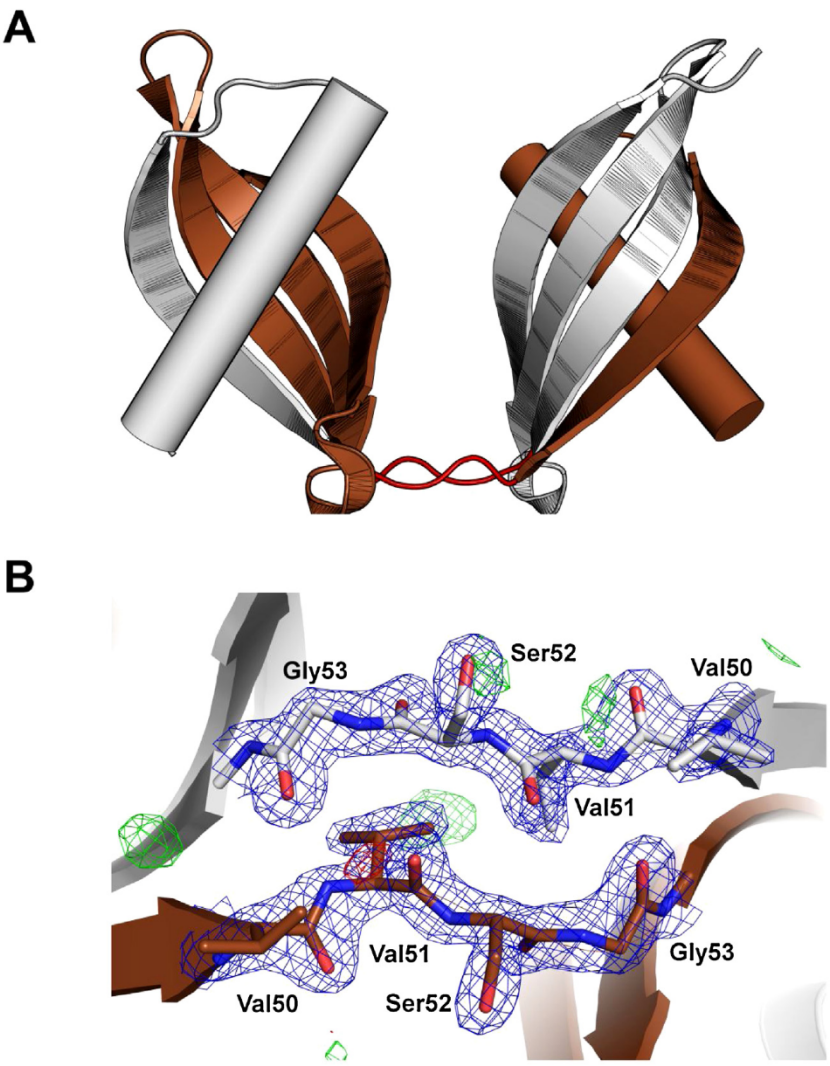

Fig. 4. Three-dimensional structure of the domain-swapped dimer of *Vu*Cys1 and its open interface. (A) Ribbon diagram of the domain-swapped dimer observed in the crystal structure of *Vu*Cys1 (PDB ID: 4TX4). Chain A is colored in light gray, chain B is colored in brown and the residues of the first inhibitory loop, which form the open interface, are colored in red. (B) Electron density of the open interface of the domain-swapped dimer. A $2F_O$-F_C omit map contoured at 2.0 σ is colored in blue, and positive density in the F_O-F_C map contoured at 3.0 σ is colored in green, whereas negative density is colored in red. The chain A is colored in light gray and chain B is colored in brown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that found in other genes encoding single-domain plant cystatins, which also contain only one intervening sequence, such as those characterized in the genomes of barley, rice and *Arabidopsis thaliana* [11]. Furthermore, the position of the single intron in the *Vu*Cys1 genomic sequence corresponds to the insertion site of the first intron when one-domain phytocystatin genes possessing two or three introns are considered [43–46]. The *Vu*Cys2 protein sequence (calculated molecular mass = 21,884.7 Da) comprises two cystatin domains in tandem, corresponding to the segments ^{9}Gly-Ile^{101} (domain 1; 93 residues) and ^{106}Gly-Val^{195} (domain 2; 90 residues) (Fig. S4). The two domains of *Vu*Cys2 share 62.2% sequence identity (excluding positions with gaps in the pairwise alignment) and each one exhibits the conserved motifs that characterize typical phytocystatins: ^{9}GG^{10}, ^{26}LARFAV^{31}, ^{53}QVVSG^{57} and ^{83}EW^{84} in the first domain and ^{106}GG^{107}, ^{121}LARFAV^{126}, ^{148}QVVEG^{152} and ^{178}SW^{179} in the second domain. The first domain and the N-terminal half of the second domain (including the LARFAV sequence) of *Vu*Cys2 are encoded in the first exon, whereas the C-terminal half of the second domain (containing the consensus motif QxVxG and the conserved W) is encoded in the second exon. Moreover, the significant sequence identity between the two cystatin domains of *Vu*Cys2 suggests that its gene likely arose by a duplication event and further divergence. These sequence analyses indicated that the cloned DNA fragments (both genomic and cDNA) encode genuine phytocystatins. To validate this assumption, the proteins were expressed in *E. coli* and the functionality of the recombinant products was evaluated.

3.2. Expression, purification and yield of the recombinant cowpea cystatins

The cowpea cystatins *Vu*Cys1 and *Vu*Cys2 were produced in the *E. coli* strain ArcticExpress (DE3), an expression host designed for increased yields of active, soluble recombinant protein. These cells co-express the cold-adapted chaperonin Cpn10 and co-chaperonin Cpn60 from the psychrophilic bacterium *Oleispira antarctica*, and these chaperonins display high protein refolding activities at temperatures of 4–12 °C [47]. When analyzed by SDS-PAGE, the purified recombinant proteins showed apparent molecular masses of approximately 14 (*Vu*Cys1) and 26 kDa (*Vu*Cys2) (Fig. S5). Taking into account the extra residues (MHHHHHHVNSLE; ~1.5 kDa) at the N-terminal region of each protein (encoded by the expres-

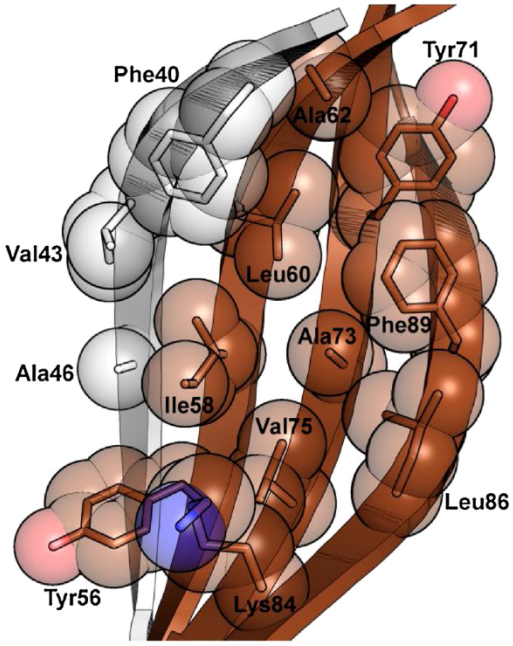

Fig. 5. Amino acid residues of the antiparalle β-sheet of the *Vu*Cys1 structure, which contribute to the formation of the hydrophobic core that characterizes the cystatin fold. The α-helix positioned over these residues is not shown.

sion vector) as well as their calculated molecular masses (~10.8 kDa for *Vu*Cys1 and 21.9 kDa for *Vu*Cys2), obtained from their amino acid sequences (deduced from the corresponding cDNAs), the expected molecular masses for the fusion cystatins would be 12.3 kDa (*Vu*Cys1) and 23.4 kDa (*Vu*Cys2), which are close to the values determined by SDS-PAGE. Using the ArcticExpress (DE3) *E. coli* expression cells, the yields of purified proteins, recovered exclusively from the intracellular soluble fraction, were 10 mg (*Vu*Cys1) and 22 mg (*Vu*Cys2) per 1 L of culture. These expression levels are very similar to those reported for other recombinant phytocystatins, such as the sugarcane cystatin, canecystatin, and its variants caneCPI-2 caneCPI-3, which were expressed at levels of 10, 20 and 22 mg L^{-1} culture, respectively [48,49].

3.3. Functionality and stability of the recombinant cowpea cystatins

Representative cysteine proteases were differently inhibited by *Vu*Cys1 and *Vu*Cys2 (Fig. 2A). Papain was fully inhibited by both recombinant cystatins, whereas chymopapain presented a residual proteolytic activity of only 14.1 and 8.5% when incubated with *Vu*Cys1 and *Vu*Cys2, respectively. Both cystatins also inhibited the proteolytic activity of bromelain, but the inhibitory activities of *Vu*Cys1 and *Vu*Cys2 against this protease (27.7% and 47.3%, respectively) were significantly lower than those observed against papain and chymopapain. Human cathepsin B, on the other hand, was very poorly inhibited by both cystatins. In this case, *Vu*Cys1 and *Vu*Cys2 led to only 9.0 and 12.5% inhibition of the enzymatic activity of human cathepsin B, respectively. Furthermore, the recombinant cystatins did not display any inhibitory activity against trypsin, a serine protease. To further demonstrate the specificity of the recombinant inhibitors, their effects on the total proteinase activity found in midgut homogenates of *C. maculatus* and *Z. subfasciatus* larvae were investigated. As depicted in Fig. 2B and C, both cystatins were able to inhibit the total proteolytic activity detected in the bruchid midguts. However, *Vu*Cys1 and *Vu*Cys2 were more active against the midgut proteinases of *C. maculatus* larvae (Fig. 2B) than those found in *Z. subfasciatus* larvae (Fig. 2C). For example, at a dose of 1 μg, *Vu*Cys1 caused 77.9 and 7.6% inhibition of the proteinase activity observed in midgut crude extracts from *C. maculatus* and *Z. subfasciatus* larvae, respectively. Indeed, although both *C. maculatus* and *Z. subfasciatus* larvae use cysteine and aspartic proteinases as major digestive enzymes [50,51], the relative amounts of cysteine proteinases are higher in *C. maculatus* in comparison to those found in *Z. subfasciatus*, which has higher levels of aspartic proteinases [52,53]. Another reasonable explanation for the different inhibitory profiles of *Vu*Cys1 and *Vu*Cys2 towards the bruchid gut proteases would be that *Z. subfasciatus* larvae express a set of digestive cysteine proteinases that are more resistant to cowpea cystatins than those produced by *C. maculatus* larvae. Because the midgut proteinases from 3rd instar larvae of *C. maculatus* were strongly inhibited by the recombinant cowpea cystatins, the effects of *Vu*Cys1 and *Vu*Cys2 on the development of *C. maculatus* were investigated in a feeding bioassay. Contradictorily, the percentages of larval mortality and adult emergence, average weight of newly emerged adults and mean developmental time of *C. maculatus* reared in the presence of *Vu*Cys1 or *Vu*Cys2 were not significantly ($P > 0.05$) different from those recorded for insects reared on artificial seeds without the recombinant inhibitors (Table S1 and Fig. S6). Indeed, these results agree with previous investigations that have shown that, when exposed to a soybean cysteine protease inhibitor (soyacystatin N, scN), *C. maculatus* rapidly adapts to the challenge diet, thus recovering from the initial growth suppression caused by scN [54]. This adaptation involves, for example, the expression of a *C. maculatus* cathepsin B-like cysteine protease 1 (CmCatB1), whose enzymatic activity is resistant to scN inhibition [55,56], and induction of gut proteases able to degrade the inhibitor [57,58].

The inhibitory activity of each cowpea cystatin against papain was poorly affected when the recombinant products were incubated for 10 min at increasing temperatures, ranging from 24 to 100 °C (Fig. 3A). Even longer incubation (up to 60 min) of each inhibitor at 100 °C was not able to reduce their papain inhibitory activities (Fig. 3B). A similar profile of stability was observed when *Vu*Cys1 and *Vu*Cys2 were pre-incubated at different pH values and then subjected to an *in vitro* inhibition assay for papain activity at pH 6.0 (Fig. 3C). The high thermal and pH stabilities of *Vu*Cys1 and *Vu*Cys2 were further investigated by CD spectroscopy analysis. Visible alterations on the CD spectra of *Vu*Cys1 were only observed when the inhibitor was incubated at 80 or 90 °C (Fig. S7A). Moreover, the *Vu*Cys2 CD spectra varied slightly less than those observed for *Vu*Cys1, even when the inhibitor was incubated at 80 or 90 °C (Fig. S7B). Near identical spectroscopic results were obtained when *Vu*Cys1 and *Vu*Cys2 were incubated in solutions with pH values varying from 2 to 11. The CD spectra of *Vu*Cys1, determined in the pH range 2–11, did not display any apparent alterations (Fig. S7C), and this finding agreed with the high *in vitro* inhibitory activity of *Vu*Cys1 against papain at pH values from 2.2 to 11 (Fig. 3C). The CD spectra of *Vu*Cys2 revealed an alteration in its folding state at pH 2.0. However, at all other pH values, the protein appeared to be folded (Fig. S7D) and, consequently, active, as demonstrated by the *in vitro* inhibition assay with papain (Fig. 3C). To provide insights about the thermal and pH stability of the cowpea cystatins, crystallization trials were performed aiming to solve their three-dimensional structures.

3.4. The crystal structure of *Vu*Cys1

Orthorhombic crystals of *Vu*Cys1 were obtained after two weeks in condition E10 of the PEG II kit, which consisted of 200 mM ammonium sulphate and 20% PEG 4000. Crystals of *Vu*Cys2 were not observed in any of the conditions tested. *Vu*Cys1 crystals diffracted at a resolution of 1.95 Å and a final atomic model of good stereochemistry was obtained (Table 1). The corresponding electron density for the first 14 amino acid residues at the N-terminal region

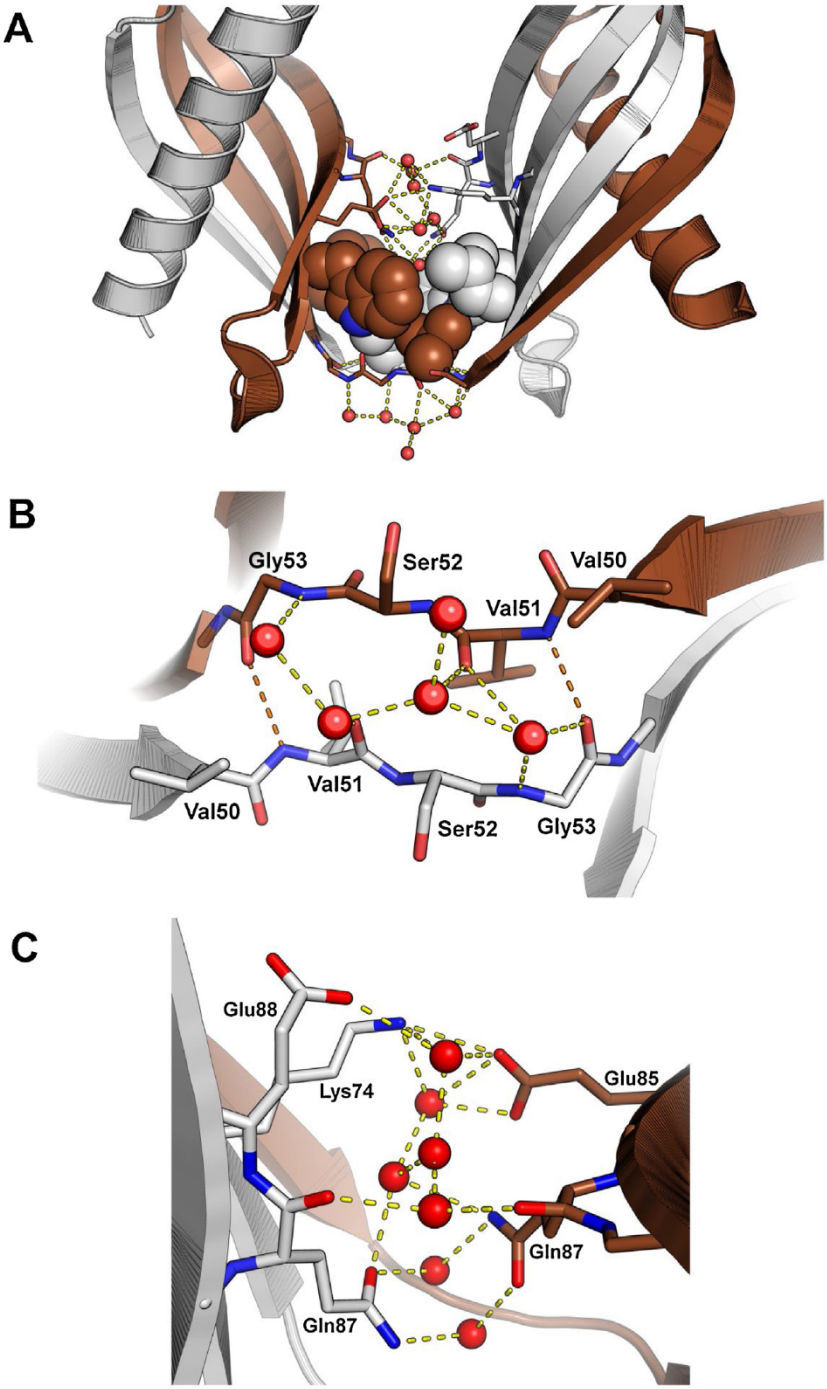

Fig. 6. Overall and detailed views of the extended open interface found in the domain-swapped dimer of *Vu*Cys1. (A) A novel small hydrophobic core is observed between two hydrogen bond networks. The side chains of the residues that constitute the small hydrophobic core are shown as spheres. (B) Hydrogen bond network observed between water molecules and the residues Val^{50}, Val^{51}, Ser^{52} and Gly^{53} in the open interface of the domain-swapped dimer. (C) The novel hydrogen bond network observed in the structure of the domain-swapped dimer of *Vu*Cys1. Several water molecules mediate interactions between the two polypeptide chains, and Lys^{74} from chain A (colored in light gray) and Glu^{85} from chain B (colored in brown) interact, forming a salt bridge. Residues are labeled and water molecules are represented as red spheres. The yellow dashes represent polar contacts within distances of 3.2 Å or less. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was not visible. Thus, the final atomic model of *Vu*Cys1 contained 83 amino acid residues, from Asn^{15} to Ala^{97}. Furthermore, the asymmetric unit contained two *Vu*Cys1 molecules arranged in a domain-swapped dimer. Each monomer in the domain-swapped dimer shows an overall fold similar to that of other cystatins, comprising a central five-turn α-helix wrapped by an antiparallel β-sheet (Fig. 4A). In the classical cystatin fold (Fig. 1B), the antiparallel β-sheet has five β-strands (β1-β5), as observed in taro-cystatin [59] and oryzacystatin-I [60]. But in some structures, the N-terminal region, including the residues of the strand β1, is disordered and not present in the atomic model [61–63]. Therefore, the domain-swapped dimer of *Vu*Cys1 consists of a four-stranded antiparallel β-sheet, including the strands β2-β5, in which three strands come from one subunit and the fourth strand comes from

the other monomer (Fig. 4A). In this structure, the first inhibitory hairpin loop, which connects the strands β2 and β3, is in a stretched conformation (Fig. 4B), forming what is known as the open interface in domain-swapped dimers [64]. The outstretched loop forms a new segment of β-strand in the domain-swapped dimer, connecting the strand β2 from one monomer to the strand β3 from the other monomer and transforming these segments in a contiguous element of secondary structure. These prolonged β-strands pair to each other, forming an antiparallel double-helical coiled coil that goes from one lobe of the dimer to the other, as previously observed in canecystatin-I, for example [63].

The interface formed by the α-helix and the β-sheet of *Vu*Cys1 is maintained by the interaction of the side chains of hydrophobic residues on the same side of the β-sheet (Fig. 5). The strand β2 contributes with Phe^{40}, Val^{43} and Ala^{46}, an arrangement that is made possible due to the presence of two β-bulges, one formed by Gly^{41} and Arg^{42} and the other by Val^{44} and Ser^{45}. These two consecutive β-bulges impose an augmented twisted conformation to the strand β2, allowing this strand to wrap around the α-helix. Tyr^{56}, Ile^{58}, Leu^{60} and Ala^{62} in the strand β3and Tyr^{71}, Ala^{73} and Val^{75} in the strand β4 also contribute to the hydrophobic core of the β-sheet, as does the aliphatic parts of Lys^{84}, Leu^{86} and Phe^{89} from the strand β5. Gln^{87} and Glu^{88} form a β-bulge, allowing Leu^{86} and Phe^{89} to be on the same side of the β-sheet. The α-helix contributes with Ile^{19}, Ala^{23}, Phe^{25}, Ala^{26}, Val^{27}, and the aliphatic part of the side chains of Glu^{29} and Lys^{33}. These residues extensively interact with the mentioned residues from the β-sheet, assembling the foundation of the cystatin fold, a hydrophobic core that is observed both in free monomers and in domain-swapped dimers. Only two direct polar interactions are observed between the α-helix and the β-sheet: Asn^{15} in the beginning of the helix forms a hydrogen bond with Glu^{48} from the strand β2, and Asn^{31} in the C-terminal part of the helix hydrogen bonds both main chain polar atoms of Leu^{38}. All other polar interactions between the α-helix and the curved β-sheet are mediated by water molecules.

Each asymmetric unit contains one domain-swapped dimer, and the use of symmetry operations allowed the analysis of the crystal packing, in which only three regions account for all polar contacts observed between symmetry related molecules. The first region involves α-helices from two domain-swapped dimers, which interact through residues Arg^{24} and Asn^{35} (Fig. S8). The angle between the helices is roughly 80°, considerably larger than those observed in most domain-swapped cystatins, including canecystatin-1 [63], sialostatin L [65], stefin B [66] and human cystatin C [67]. A second region with extensive intermolecular polar contacts exists between two symmetry related chain A molecules (Fig. S9). In this region, Asp^{94} interacts with the nitrogen atom of Leu^{37}, Lys^{69} forms s hydrogen bond with the main chain of Asn^{35}, and Gln^{67} interacts with the side chain of Gln^{34}. An interaction between Glu^{61} and Trp^{80}, both from chain A, is the only other polar contact between symmetry related molecules.

3.5. *The domain-swapped dimer of* Vu*Cys1 recruits water molecules and additional residues to its open interface*

In most domain-swapped cystatins, the open interface is formed by the four residues of the first inhibitory hairpin loop, which adopt a β-strand conformation effectively connecting the strand β2 of one monomer to the strand β3 of the other monomer. In the *Vu*Cys1 structure, these residues correspond to Val^{50}, Val^{51}, Ser^{52} and Gly^{53}. However, instead of presenting the classical hydrogen bond pattern observed in antiparallel β-sheets, some of their main chain polar atoms interact with water molecules, forming a hydrogen bond network that connects the two antiparallel chains (Fig. 6A and B). To the best of our knowledge this feature was not observed in any other domain-swapped cystatin dimer structure.

The two lobes of the domain-swapped dimer are positioned closer to each other in the cowpea cystatin than in other cystatins (Fig. 6A), as indicated by the radius of gyration of these crystal structures (Table S2). This proximity allows polar residues from both lobes to coordinate water molecules, forming a hydrogen bond network that mediates contacts between the lobes. The only direct inter chain contact is a salt bridge formed between Lys^{74} from chain A and Glu^{85} from chain B (Fig. 6C). A stereo view of these interactions is presented in Figs. S10 and S11 . Six water molecules mediate interactions between the two polypeptide chains, and both the side chains and main chain carbonyls of Glu^{87} residues from each chain indirectly interact with each other through hydrogen bonds with water molecules. Additionally, Glu^{88} from chain A hydrogen bonds a water molecule that is bound to Glu^{85} from chain B. Additionally, a small hydrophobic core is formed due the proximity of the two lobes of the domain-swapped dimer, involving residues Val^{51}, Leu^{55} and Trp^{76} (Fig. 6A). This novel molecular feature that employs water molecules and additional amino acid residues to extend the open interface, allowing the formation of a small hydrophobic core, might account for part of the extraordinary thermal and pH stability of *Vu*Cys1.

4. Conclusions

Two cowpea cystatins, containing one (*Vu*Cys1) and two (*Vu*Cys2) inhibitory domains, were produced in *E. coli* cells. The recombinant proteins had high solubility and were able to inhibit papain-like cysteine proteases. Moreover, their inhibitory activities were resistant to high temperatures and pH variations. Finally, the crystal structure of *Vu*Cys1 was determined, revealing structural features that might explain its high thermal and pH stability.

Acknowledgments

This work was supported by research grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP). TBG is a senior researcher of CNPq.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijbiomac.2017.04.008.

References

[1] D. Kordis, V. Turk, Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes, BMC Evol. Biol. 9 (2009) 266, http://dx.doi.org/10.1186/1471-2148-9-266.
[2] A. Barrett, The cystatins—a new class of peptidase inhibitors, Trends Biochem. Sci. 12 (1987) 193–196, http://dx.doi.org/10.1016/0968-0004(87)90092-2.
[3] V. Turk, V. Stoka, D. Turk, Cystatins: biochemical and structural properties and medical relevance, Front. Biosci. 13 (2008) 5406–5420.
[4] V. Turk, B. Turk, D. Turk, Lysosomal cysteine proteases: facts and opportunities, EMBO J. 20 (2001) 4629–4633, http://dx.doi.org/10.1093/emboj/20.17.4629.
[5] V. Stoka, V. Turk, B. Turk, Lysosomal cathepsins and their regulation in aging and neurodegeneration, Ageing Res. Rev. 32 (2016) 22–37, http://dx.doi.org/10.1016/j.arr.2016.04.010.
[6] M.T. Stubbs, B. Laber, W. Bode, R. Huber, R. Jerala, B. Lenarcic, V. Turk, The refined 2.4 A 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 (1990) 1939–1947.
[7] S. Jenko, I. Dolenc, G. Guncar, A. Dobersek, M. Podobnik, D. Turk, Crystal structure of stefin A in complex with cathepsin H: N-terminal residues of inhibitors can adapt to the active sites of endo- and exopeptidases, J. Mol. Biol. 326 (2003) 875–885.

[8] M. Renko, U. Požgan, D. Majera, D. Turk, Stefin A displaces the occluding loop of cathepsin B only by as much as required to bind to the active site cleft, FEBS J. 277 (2010) 4338–4345, http://dx.doi.org/10.1111/j.1742-4658.2010.07824.x.
[9] A.J. Barrett, H. Fritz, A. Grubb, S. Isemura, M. Järvinen, N. Katunuma, W. Machleidt, W. Müller-Esterl, M. Sasaki, V. Turk, Nomenclature and classification of the proteins homologous with the cysteine-proteinase inhibitor chicken cystatin, Biochem. J. 236 (1986) 312.
[10] R. Margis, E.M. Reis, V. Villeret, Structural and phylogenetic relationships among plant and animal cystatins, Arch. Biochem. Biophys. 359 (1998) 24–30, http://dx.doi.org/10.1006/abbi.1998.0875.
[11] M. Martínez, Z. Abraham, P. Carbonero, I. Díaz, Comparative phylogenetic analysis of cystatin gene families from arabidopsis, rice and barley, Mol. Genet. Genomics 273 (2005) 423–432, http://dx.doi.org/10.1007/s00438-005-1147-4.
[12] T.A. Walsh, J.A. Strickland, Proteolysis of the 85-kilodalton crystalline cysteine proteinase inhibitor from potato releases functional cystatin domains, Plant Physiol. 103 (1993) 1227–1234.
[13] J. Szewińska, J. Simińska, W. Bielawski, The roles of cysteine proteases and phytocystatins in development and germination of cereal seeds, J. Plant Physiol. 207 (2016) 10–21, http://dx.doi.org/10.1016/j.jplph.2016.09.008.
[14] K.J. Kunert, S.G. van Wyk, C.A. Cullis, B.J. Vorster, C.H. Foyer, Potential use of phytocystatins in crop improvement, with a particular focus on legumes, J. Exp. Bot. 66 (2015) 3559–3570, http://dx.doi.org/10.1093/jxb/erv211.
[15] M. Martinez, M.E. Santamaria, M. Diaz-Mendoza, A. Arnaiz, L. Carrillo, F. Ortego, I. Diaz, Phytocystatins: defense proteins against phytophagous insects and acari, Int. J. Mol. Sci. 17 (2016) 1747, http://dx.doi.org/10.3390/ijms17101747.
[16] S.G. van Wyk, K.J. Kunert, C.A. Cullis, P. Pillay, M.E. Makgopa, U. Schlüter, B.J. Vorster, Review: the future of cystatin engineering, Plant Sci. 246 (2016) 119–127, http://dx.doi.org/10.1016/j.plantsci.2016.02.016.
[17] K.V. Fernandes, P.A. Sabelli, D.H. Barratt, M. Richardson, J. Xavier-Filho, P.R. Shewry, The resistance of cowpea seeds to bruchid beetles is not related to levels of cysteine proteinase inhibitors, Plant Mol. Biol. 23 (1993) 215–219.
[18] N.N. Diop, M. Kidric, A. Repellin, M. Gareil, A. d'Arcy-Lameta, A.T. Pham Thi, Y. Zuily-Fodil, A multicystatin is induced by drought-stress in cowpea (Vigna unguiculata (L.) Walp.) leaves, FEBS Lett. 577 (2004) 545–550, http://dx.doi.org/10.1016/j.febslet.2004.10.014.
[19] D.R. Hoagland, D.I. Arnon, The water-culture method for growing plants without soil, Circ. Calif. Agric. Exp. Stn. 347 (1950) 1–32.
[20] S.A.J. Warner, Genomic DNA isolation and lambda library construction, in: G.D. Foster, D. Twell (Eds.), Plant Gene Isol. Princ. Pract, John Wiley & Sons, West Sussex, 1996, pp. 51–73.
[21] P.G.C. Landim, T.O. Correia, F.D.A. Silva, D.R. Nepomuceno, H.P.S. Costa, H.M. Pereira, M.D.P. Lobo, Moreno F.B.M.B, J. Brandão-Neto, S.C. Medeiros, I.M. Vasconcelos, J.T.A. Oliveira, B.L. Sousa, I.L. Barroso-Neto, V.N. Freire, C.P.S. Carvalho, A.C.O. Monteiro-Moreira, T.B. Grangeiro, Production in Pichia pastoris, antifungal activity and crystal structure of a class I chitinase from cowpea (Vigna unguiculata): insights into sugar binding mode and hydrolytic action, Biochimie 135 (2017) 89–103, http://dx.doi.org/10.1016/j.biochi.2017.01.014.
[22] J.E.C. Freire, I.M. Vasconcelos, F.B.M.B. Moreno, A.B. Batista, M.D.P. Lobo, M.L. Pereira, J.P.M.S. Lima, R.V.M. Almeida, A.J.S. Sousa, A.C.O. Monteiro-Moreira, J.T.A. Oliveira, T.B. Grangeiro, Mo-CBP3, an antifungal chitin-binding protein from Moringa oleifera seeds, is a member of the 2S albumin family, PLoS One 10 (2015) e0119871, http://dx.doi.org/10.1371/journal.pone.0119871.
[23] J. Sambrook, E. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989.
[24] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[25] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symp. Ser. 41 (1999) 95–98.
[26] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254.
[27] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.
[28] M.D.P. Lobo, F.D.A. Silva, P.G. de C. Landim, P.R. da Cruz, T.L. de Brito, S.C. de Medeiros, J.T.A. Oliveira, I.M. Vasconcelos, H.M. Pereira, T.B. Grangeiro, Expression and efficient secretion of a functional chitinase from Chromobacterium violaceum in Escherichia coli, BMC Biotechnol. 13 (2013) 46, http://dx.doi.org/10.1186/1472-6750-13-46.
[29] M. Abe, K. Abe, M. Kuroda, S. Arai, Corn kernel cysteine proteinase inhibitor as a novel cystatin superfamily member of plant origin. Molecular cloning and expression studies, Eur. J. Biochem. 209 (1992) 933–937.
[30] L.B. da, S. Andrade, A.S. Oliveira, J.K.C. Ribeiro, S. Kiyota, I.M. Vasconcelos, J.T.A. de Oliveira, M.P. de Sales, Effects of a novel pathogenesis-related class 10 (PR-10) protein from Crotalaria pallida Roots with papain inhibitory activity against root-knot nematode Meloidogyne incognita, J. Agric. Food Chem. 58 (2010) 4145–4152, http://dx.doi.org/10.1021/jf9044556.
[31] A.S. Oliveira, L. Migliolo, R.O. Aquino, J.K.C. Ribeiro, L.L.P. Macedo, L.B.S. Andrade, M.P. Bemquerer, E.A. Santos, S. Kiyota, M.P. de Sales, Purification and characterization of a trypsin-papain inhibitor from Pithecelobium dumosum seeds and its in vitro effects towards digestive enzymes from insect pests, Plant Physiol. Biochem. 45 (2007) 858–865, http://dx.doi.org/10.1016/j.plaphy.2007.08.002.
[32] J. Xavier-Filho, F.A.P. Campos, M.B. Ary, C.P. Silva, M.M.M. Carvalho, M.L.R. Macedo, F.J.A. Lemos, G. Grant, Poor correlation between the levels of proteinase inhibitors found in seeds of different cultivars of cowpea (Vigna unguiculata) and the resistance/susceptibility to predation by Callosobruchus maculatus, J. Agric. Food Chem. 37 (1989) 1139–1143, http://dx.doi.org/10.1021/jf00088a071.
[33] C.L. Siqueira-Júnior, K.V.S. Fernandes, O.L.T. Machado, M. da Cunha, V.M. Gomes, D. Moura, T. Jacinto, 87 kDa tomato cystatin exhibits properties of a defense protein and forms protein crystals in prosystemin overexpressing transgenic plants, Plant Physiol. Biochem. 40 (2002) 247–254, http://dx.doi.org/10.1016/S0981-9428(02)01364-5.
[34] Y.F.M.M. Leite, L.M.C.M. Silva, R.C. das, N. Amorim, E.A. Freire, D.M. de Melo Jorge, T.B. Grangeiro, N.M.B. Benevides, Purification of a lectin from the marine red alga Gracilaria ornata and its effect on the development of the cowpea weevil Callosobruchus maculatus (Coleoptera: Bruchidae), Biochim. Biophys. Acta. 1724 (2005) 137–145, http://dx.doi.org/10.1016/j.bbagen.2005.03.017.
[35] W. Kabsch, XDS, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 125–132, http://dx.doi.org/10.1107/S0907444909047337.
[36] A.J. McCoy, Solving structures of protein complexes by molecular replacement with Phaser, Acta Crystallogr. D Biol. Crystallogr. 63 (2007) 32–41, http://dx.doi.org/10.1107/S0907444906045975.
[37] N. Stein, CHAINSAW: A program for mutating pdb files used as templates in molecular replacement, J. Appl. Crystallogr. 41 (2008) 641–643, http://dx.doi.org/10.1107/S0021889808006985.
[38] P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive python-based system for macromolecular structure solution, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 213–221, http://dx.doi.org/10.1107/S0907444909052925.
[39] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 2126–2132, http://dx.doi.org/10.1107/S0907444904019158.
[40] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291, http://dx.doi.org/10.1107/S0021889892009944.
[41] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 12–21, http://dx.doi.org/10.1107/S0907444909042073.
[42] R. Breathnach, P. Chambon, Organization and expression of eucaryotic split genes coding for proteins, Annu. Rev. Biochem. 50 (1981) 349–383, http://dx.doi.org/10.1146/annurev.bi.50.070181.002025.
[43] H. Kondo, K. Abe, Y. Emori, S. Arai, 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 (1991) 87–90.
[44] M. Abe, C. Domoto, H. Watanabe, K. Abe, S. Arai, Structural organization of the gene encoding corn cystatin, Biosci. Biotechnol. Biochem. 60 (1996) 1173–1175, http://dx.doi.org/10.1271/bbb.60.1173.
[45] T. Misaka, M. Kuroda, K. Iwabuchi, K. Abe, S. Arai, 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 (1996) 609–614.
[46] S. Ohtsubo, H. Kobayashi, W. Noro, M. Taniguchi, E. Saitoh, Molecular cloning and characterization of oryzacystatin-III, a novel member of phytocystatin in rice (Oryza sativa L. japonica), J. Agric. Food Chem. 53 (2005) 5218–5224, http://dx.doi.org/10.1021/jf050348j.
[47] M. Ferrer, T.N. Chernikova, M.M. Yakimov, P.N. Golyshin, K.N. Timmis, Chaperonins govern growth of Escherichia coli at low temperatures, Nat. Biotechnol. 21 (2003) 1266–1267, http://dx.doi.org/10.1038/nbt1103-1266.
[48] A. Soares-Costa, L.M. Beltramini, O.H. Thiemann, F. Henrique-Silva, A sugarcane cystatin: recombinant expression, purification, and antifungal activity, Biochem. Biophys. Res. Commun. 296 (2002) 1194–1199.
[49] A. Gianotti, W.M. Rios, A. Soares-Costa, V. Nogaroto, A.K. Carmona, M.L.V. Oliva, S.S. Andrade, F. Henrique-Silva, Recombinant expression, purification, and functional analysis of two novel cystatins from sugarcane (Saccharum officinarum), Protein Expr. Purif. 47 (2006) 483–489, http://dx.doi.org/10.1016/j.pep.2005.10.026.
[50] F.A.P. Campos, J. Xavier-Filho, C.P. Silva, M.B. Ary, Resolution and partial characterization of proteinases and α-amylases from midguts of larvae of the bruchid beetle Callosobruchus maculatus (F.), Comp. Biochem. Physiol. Part B Comp. Biochem. 92 (1989) 51–57, http://dx.doi.org/10.1016/0305-0491(89)90312-X.
[51] F. Lemos, F. Campos, C. Silva, J. Xavier-Filho, Proteinases and amylases of larval midgut of Zabrotes subfasciatus reared on cowpea (Vigna unguiculata) seeds, Entomol. Exp. Appl. 56 (1990) 219–227.
[52] C.P. Silva, J. Xavier-Filho, Comparison between the levels of aspartic and cysteine proteinases of the larval midguts of Callosobruchus maculatus (F.) and Zabrotes subfasciatus (BOH.) (Coleoptera: bruchidae), Comp. Biochem. Physiol. Part B Comp. Biochem. 99 (1991) 529–533, http://dx.doi.org/10.1016/0305-0491(91)90333-9.

[53] C.P. Silva, W.R. Terra, R.M. Lima, Differences in midgut serine proteinases from larvae of the bruchid beetles Callosobruchus maculatus and Zabrotes subfasciatus, Arch. Insect Biochem. Physiol. 47 (2001) 18–28, http://dx.doi.org/10.1002/arch.1031.
[54] K. Zhu-Salzman, R.S. Zeng, Molecular mechanisms of insect adaptation to plant defense: lessons learned from a Bruchid beetle, Insect Sci. 15 (2008) 477–481, http://dx.doi.org/10.1111/j.1744-7917.2008.00236.x.
[55] J. Moon, R.A. Salzman, J.-E. Ahn, H. Koiwa, K. Zhu-Salzman, Transcriptional regulation in cowpea bruchid guts during adaptation to a plant defence protease inhibitor, Insect Mol. Biol. 13 (2004) 283–291, http://dx.doi.org/10.1111/j.0962-1075.2004.00485.x.
[56] Y.D. Koo, J.-E. Ahn, R.A. Salzman, J. Moon, Y.H. Chi, D.-J. Yun, S.Y. Lee, H. Koiwa, K. Zhu-Salzman, Functional expression of an insect cathepsin B-like counter-defence protein, Insect Mol. Biol. 17 (2008) 235–245, http://dx.doi.org/10.1111/j.1365-2583.2008.00799.x.
[57] K. Zhu-Salzman, H. Koiwa, R.A. Salzman, R.E. Shade, J.-E. Ahn, Cowpea bruchid Callosobruchus maculatus uses a three-component strategy to overcome a plant defensive cysteine protease inhibitor, Insect Mol. Biol. 12 (2003) 135–145.
[58] J.-E. Ahn, K. Zhu-Salzman, CmCatD, a cathepsin D-like protease has a potential role in insect defense against a phytocystatin, J.Insect Physiol. 55 (2009) 678–685, http://dx.doi.org/10.1016/j.jinsphys.2009.04.016.
[59] M.-H. Chu, K.-L. Liu, H.-Y. Wu, K.-W. Yeh, Y.-S. Cheng, Crystal structure of tarocystatin-papain complex: implications for the inhibition property of group-2 phytocystatins, Planta 234 (2011) 243–254, http://dx.doi.org/10.1007/s00425-011-1398-8.
[60] K. Nagata, N. Kudo, K. Abe, S. Arai, M. Tanokura, Three-dimensional solution structure of oryzacystatin-I, a cysteine proteinase inhibitor of the rice Oryza sativa L. japonica, Biochemistry (Mosc.) 39 (2000) 14753–14760.
[61] M.S. Nissen, G.N.M. Kumar, B. Youn, D.B. Knowles, K.S. Lam, W.J. Ballinger, N.R. Knowles, C. Kang, Characterization of Solanum tuberosum multicystatin and its structural comparison with other cystatins, Plant Cell 21 (2009) 861–875, http://dx.doi.org/10.1105/tpc.108.064717.
[62] A.R. Green, M.S. Nissen, G.N.M. Kumar, N.R. Knowles, C. Kang, Characterization of Solanum tuberosum multicystatin and the significance of core domains, Plant Cell 25 (2013) 5043–5052, http://dx.doi.org/10.1105/tpc.113.121004.
[63] N.F. Valadares, R. de Oliveira-Silva, I.A. Cavini, I. de, A. Marques, H.D. Pereira, A. Soares-Costa, F. Henrique-Silva, H.R. Kalbitzer, C.E. Munte, R.C. Garratt, X-ray crystallography and NMR studies of domain-swapped canecystatin-1, FEBS J. 280 (2013) 1028–1038, http://dx.doi.org/10.1111/febs.12095.
[64] M.J. Bennett, M.P. Schlunegger, D. Eisenberg, 3D domain swapping: a mechanism for oligomer assembly, Protein Sci. 4 (1995) 2455–2468, http://dx.doi.org/10.1002/pro.5560041202.
[65] M. Kotsyfakis, H. Horka, J. Salat, J.F. Andersen, The crystal structures of two salivary cystatins from the tick Ixodes scapularis and the effect of these inhibitors on the establishment of Borrelia burgdorferi infection in a murine model, Mol. Microbiol. 77 (2010) 456–470, http://dx.doi.org/10.1111/j.1365-2958.2010.07220.x.
[66] S. Jenko Kokalj, G. Guncar, I. Stern, G. Morgan, S. Rabzelj, M. Kenig, R.A. Staniforth, J.P. Waltho, E. Zerovnik, D. Turk, Essential role of proline isomerization in stefin B tetramer formation, J. Mol. Biol. 366 (2007) 1569–1579, http://dx.doi.org/10.1016/j.jmb.2006.12.025.
[67] R. Janowski, M. Kozak, E. Jankowska, Z. Grzonka, A. Grubb, M. Abrahamson, M. Jaskolski, Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping, Nat. Struct. Biol. 8 (2001) 316–320, http://dx.doi.org/10.1038/86188.
[68] D. Irene, T.-Y. Chung, B.-J. Chen, T.-H. Liu, F.-Y. Li, J.T.C. Tzen, C.-I. Wang, C.-L. Chyan, Solution structure of a phytocystatin from Ananas comosus and its molecular interaction with papain, PLoS One 7 (2012) e47865, http://dx.doi.org/10.1371/journal.pone.0047865.
[69] Y.-J. Hu, D. Irene, C.-J. Lo, Y.-L. Cai, T.-C. Tzen, T.-H. Lin, C.-L. Chyan, Resonance assignments and secondary structure of a phytocystatin from Sesamum indicum, Biomol. NMR Assign. 9 (2015) 309–311, http://dx.doi.org/10.1007/s12104-015-9598-y.