

X-ray crystallography and NMR studies of domain-swapped canecystatin-1

Napoleão F. Valadares¹, Rodrigo de Oliveira-Silva¹, Italo A. Cavini¹, Ivo de Almeida Marques², **Humberto D'Muniz Pereira¹**, Andrea Soares-Costa³, Flavio Henrique-Silva³, Hans R. Kalbitzer⁴, **Claudia E. Munte¹** and **Richard C. Garratt¹**

¹ Center for Structural Molecular Biotechnology, Department of Physics and Informatics, Physics Institute of São Carlos, University of São Paulo, São Carlos-SP, Brazil

² Physics Institute, Federal University of Goiás, Brazil

³ Laboratory of Molecular Biology, Department of Genetic and Evolution, Federal University of São Carlos, Brazil

⁴ Institute of Biophysics and Physical Biochemistry, University of Regensburg, Germany

Keywords

domain-swapped; NMR; phytocystatin; X-ray crystallography

Correspondence

R. C. Garratt, Center for Structural Molecular Biotechnology, Department of Physics and Informatics, Physics Institute of São Carlos, University of São Paulo, Av. Trabalhador são-carlense 400, São Carlos-SP 13566-570, Brazil
Fax: +55 16 3373 9881
Tel: +55 16 3373-9881
E-mail: richard@ifsc.usp.br

(Received 15 September 2012, revised 29 November 2012, accepted 3 December 2012)

doi:10.1111/febs.12095

The three-dimensional structure of canecystatin-1, a potent inhibitor of cysteine proteases from sugarcane (*Saccharum officinarum*), has been solved in two different crystal forms. In both cases, it is seen to exist as a domain-swapped dimer, the first such observation for a cystatin of plant origin. Size exclusion chromatography and multidimensional NMR spectroscopy show the dimer to be the dominant species in solution, despite the presence of a measurable quantity of monomer undergoing slow exchange. The latter is believed to be the active species, whereas the domain-swapped dimer is presumably inactive, as its first inhibitory loop has been extended to form part of a long β -strand that forms a double-helical coiled coil with its partner from the other monomer. A similar structure is observed in human cystatin C, but the spatial disposition of the two lobes of the dimer is rather different. Dimerization is presumably a mechanism by which canecystatin-1 can be kept inactive within the plant, avoiding the inhibition of endogenous proteases. The structure described here provides a platform for the rational design of specific cysteine protease inhibitors for biotechnological applications.

Database

The coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes [3UL5](#) and [3UL6](#).

Structured digital abstract

- [Canecystatin-1](#) and [Canecystatin-1](#) bind by [molecular sieving](#) (View Interaction: [1](#), [2](#))
- [Canecystatin-1](#) and [Canecystatin-1](#) bind by [nuclear magnetic resonance](#) (View interaction)
- [Canecystatin-1](#) and [Canecystatin-1](#) bind by [dynamic light scattering](#) (View interaction)
- [Canecystatin-1](#) and [Canecystatin-1](#) bind by [x-ray crystallography](#) (View interaction)

Abbreviations

CSI, chemical shift index; DLS, dynamic light scattering; DSS, 4,4-Dimethyl-4-silapentane-1-sulfonic acid; K_{av} , Distribution coefficient = $(V_0 - V_e)/(V_0 - V_c)$, where V_0 is column void volume, V_e is elution volume, and V_c is geometric column volume; PDB, Protein Data Bank; SEC, size exclusion chromatography.

Introduction

The superfamily of cystatins comprises evolutionarily related cysteine protease inhibitors that share a conserved tertiary fold, and are found in vertebrates, insects, and plants [1,2]. They play an important role as regulators of enzymes of the papain subfamily, including the mammalian cathepsins. Many studies have probed their activities [3], and crystallographic and NMR studies have confirmed that their conserved sequences and motifs translate into a conserved fold [4–6]. Crystallographic studies have been of particular importance in defining a common inhibitory mechanism, by showing that mammalian cystatins bind papain-like cysteine proteases in the same fashion as plant cystatins [6–9]. All of these complexes show conserved motifs in two hairpin loops, which, together with the N-terminal region, form the inhibitory site and interact extensively with the protease.

Owing to the importance of cathepsins in human disease, particularly in cancer [10] and neurodegenerative diseases [11], human cystatins and their interaction with cathepsins have been extensively characterized [6–8]. On the other hand, plant cystatins or phytocystatins are inhibitors of cysteine proteases that are putatively involved in plant defense, as endogenous regulators of protein turnover, and by regulating plant protein catabolism and inhibiting extracellular cysteine proteases from pathogens and herbivorous insects [1,12,13]. Owing to their inhibitory activities against both endogenous and exogenous proteases, phytocystatins have multiple applications with great potential economic impact, such as in the regulation of seed germination, modulating fruit ripening, and inhibiting digestive proteases of herbivores [14–16]. Regarding the latter, recent studies have shown promising results regarding the development of transgenic plants that overexpress cystatins in order to generate improved resistance towards insects, nematodes, and phytopathogens [17–21]. Despite these potential applications, there is relatively little structural information concerning plant cystatins available. Oryzacystatin I from rice has been the most thoroughly investigated to date, including the determination of its three-dimensional structure by NMR [4].

Owing to the economic significance of sugarcane (*Saccharum officinarum*) as a renewable source of biofuel, one of the aims of our ongoing research is to structurally characterize sugarcane cystatins. In the present study, we report crystal structures for sugarcane cystatin 1 (canecystatin-1) in different space groups, and reveal the phenomenon of domain-swapping, a previously unreported oligomerization mechanism for plant cystatins. Complementary to this,

high-resolution NMR measurements of the $^{13}\text{C}/^{15}\text{N}$ -labeled canecystatin-1 allowed the complete sequential assignment of the protein [22], and revealed the presence of two conformations undergoing slow exchange in solution. The analysis of the chemical shift index is consistent with the presence of a poorly populated monomeric state, and a highly populated domain-swapped state, in agreement with the observed size exclusion chromatography (SEC) profile. The association of both techniques allowed a comprehensive structural characterization of canecystatin-1 that may well be extrapolated to other phytocystatins.

Results and Discussion

Canecystatin-1 is present as both monomers and domain-swapped dimers in solution

Nickel affinity chromatography and SEC allowed the isolation of pure heterologous canecystatin-1. The SEC chromatograms showed three distinct peaks (Fig. 1), which were analyzed by denaturing PAGE (Fig. S1). Peaks 1, 2 and 3 correspond to canecystatin-1, and migrate as monomers on SDS/PAGE.

The expected molecular mass of His-tag-free canecystatin-1, based on its amino acid sequence, is 11.9 kDa. Therefore, it is likely that the third (poorly resolved) peak observed on gel filtration, with an

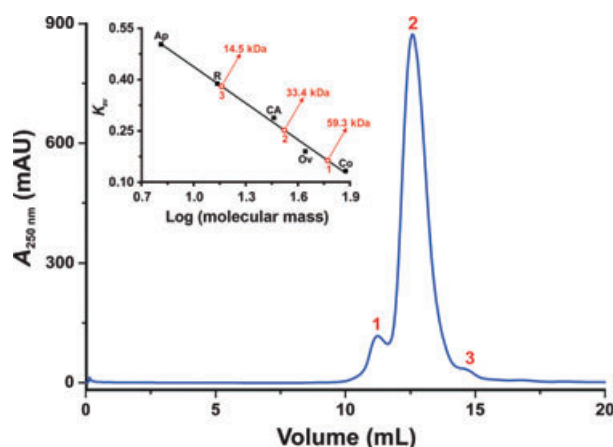


Fig. 1. SEC of canecystatin-1. Three labeled peaks can be observed, corresponding to molecular masses of 59.3, 33.4 and 14.5 kDa. The inset shows the calibration curve, in which the red circles represent the eluted peaks. Aprotinin (Ap, 6.5 kDa), ribonuclease A (R, 13.7 kDa), carbonic anhydrase (CA, 29 kDa), ovalbumin (Ov, 44 kDa) and conalbumin (Co, 75 kDa) were used as standards for column calibration. A solution containing 100 mM NaH_2PO_4 (pH 7.0) and 100 mM NaCl was used as buffer.

estimated molecular mass of 14.5 kDa, corresponds to the monomeric species. On the other hand, the predominant second peak, eluting with an estimated mass of 33.4 kDa, is probably a homodimer. This is consistent with the results of dynamic light scattering (DLS), which yielded a single peak corresponding to a mass of 38.5 ± 11.9 kDa (Fig. S2). The fact that this dimer is expected to be domain-swapped in solution (see below), and therefore elongated in shape, probably explains the overestimation of its molecular mass by both techniques. The first peak (59.3 kDa) most likely represents a small population of tetramers.

Structure description

The four previously resolved phytocystatin structures, from rice [4], potato [23], taro [9], and pineapple [24] [Protein Data Bank (PDB) codes [1EQK](#), [2W9Q](#), [3IMA](#), and [2L4V](#), respectively], are all monomeric, and present a highly conserved fold with a straightforward topology consisting of an α -helix followed by four β -strands that form an antiparallel β -sheet that is curved around the helix (Fig. 2A).

The crystal structures of canecystatin-1 in both orthorhombic and hexagonal space groups were readily solved with the molecular replacement method. After the first refinement cycle, in both structures the electron densities of the first inhibitory hairpin loop (Val60–Gly63, located between β -strands 2 and 3) unambiguously showed that the individual subunits are present in the form of domain-swapped dimers, as can be seen from the omit map shown in Fig. S3. The asymmetric unit of the orthorhombic structure contains two domain-swapped dimers, whereas that of the hexagonal structure contains only one. These dimers present a good overall superposition on one another, with rmsd values varying between 0.8 and 2.1 Å, and their overall fold is similar to that of other cystatins, consisting of a four-stranded antiparallel β -sheet (composed of three strands from one subunit and one from the other) curved around an α -helix (Fig. 2). Larger rmsd values were observed when the two lobular halves of the dimer showed different degrees of closure. In all of the final models (four monomers in the orthorhombic crystal form and two in the hexagonal crystal form) the first 20 residues are disordered. Thereafter, continuous electron density is observed up to within a few residues of the C-terminus. Full data collection and refinement statistics are given in Table 1.

The observed domain swap can be pictured as the pairing of β -strands 3, 4 and 5 from one subunit with the N-terminal region, the α -helix and β -strand 2 from the other, leading to conservation of the phytocysta-

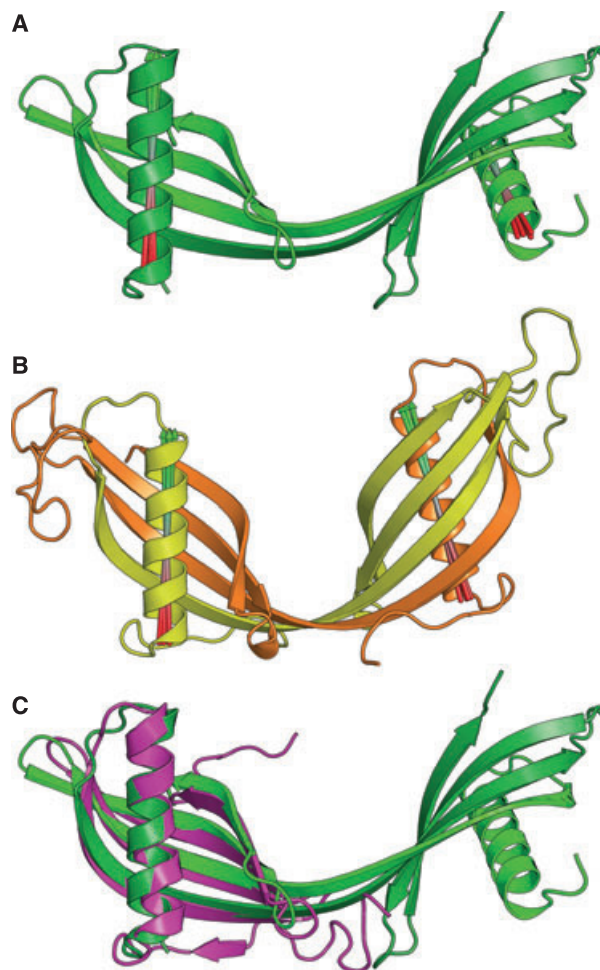


Fig. 2. (A, B) A domain-swapped dimer of canecystatin-1 as observed in the hexagonal crystal form (PDB code [3UL6](#), colored in dark green and green) (A) compared with human cystatin C (PDB code [1G96](#), colored in yellow and orange) (B). Arrows indicating the helical axes are used to define the relative orientation of the two halves of the structure, and their values are listed in Table S1. Four different definitions of the helical axes were used, corresponding to the slightly different orientations of the arrows shown in the figure. (C) Superposition of the oryzacystatin monomer, colored in magenta (PDB code [1EQK](#)), over the domain-swapped dimer of canecystatin-1.

tin fold. It should be pointed out that, here, we use the strand nomenclature originally established for oryzacystatin [4], in which the N-terminal region is considered to form an additional strand (β 1) that is absent from the structures reported here. Although this is the first description of domain-swapping in a phytocystatin dimer, a similar phenomenon has been reported previously in cystatins from other species, including human cystatin C, suggesting that it is not a crystallization artefact (Fig. 2) [25–28]. In the case of canecystatin-1, the ‘open’ interface (defined as being

Table 1. Full data collection and refinement statistics for canecystatin-1 (PDB codes [3UL5](#) and [3UL6](#)). The numerals in parentheses are from the highest-resolution shells.

Data collection		
Resolution (Å)	2.3	2.63
Space group	C222 ₁	P6 ₄ 22
Cell dimensions (Å): <i>a</i> , <i>b</i> , <i>c</i>	99.77, 113.99, 86.78	83.80, 83.80, 142.57
Detector	MarMosaic 225	MarMosaic 225
X-ray source	LNLS-MX2	LNLS-MX2
Wavelength (Å)	1.45	1.45
Resolution range (Å)	56.8–2.30 (2.42–2.30)	50.1–2.63 (2.77–2.63)
Redundancy	3.9 (3.8)	5.4 (5.3)
<i>R</i> _{meas} (%) ^a	13.4 (61.4)	12.1 (73.0)
Completeness (%)	100.0 (100.0)	98.4 (97.3)
Total reflections	87 939 (12 086)	48 078 (6865)
Unique reflections	22 328 (3185)	9139 (1270)
<i>I</i> / <i>σ</i> (<i>I</i>)	12.2 (2.1)	9.0 (2.5)
Refinement parameters		
Reflections used for refinement	21 386	8884
<i>R</i> (%) ^b	0.225	0.246
<i>R</i> _{free} (%) ^b	0.25	0.28
Overall averaged <i>B</i> -factor (Å ²)	34.3	59.4
No. of protein atoms	2611	1219
No. of water molecules	130	21
No. of ligand atoms	14	18
Ramachandran plot		
Most favored region (%)	96.5	95.5
Residues in disallowed regions (%)	0.3	0.6
Rmsd from ideal geometry		
rms bond lengths (Å)	0.003	0.009
rms bond angles (°)	0.677	1.162

^a $R_{\text{meas}} = \frac{\sum_{hkl} n(n-1)/2 \sum_i |I(hkl)_i - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I(hkl)_i}$, where $I(hkl)_i$ is the *i*th measurement of the intensity of reflection *hkl*, $\langle I(hkl) \rangle$ is the mean intensity of reflection *hkl*, and *n* is the number of observations of intensity *I(hkl)*. ^b *R* is the conventional crystallographic *R*-factor, $\sum_i |F_{\text{obs}} - F_{\text{calc}}| / \sum_i F_{\text{obs}}$, where *F*_{obs} and *F*_{calc} are the observed and calculated structure factors, respectively. Five per cent of the reflections that were excluded from the refinement were used in the *R*_{free} calculation.

that which arises on formation of the domain-swapped dimer [29]) is composed of residues Val60–Gly63. These form a new β-strand segment connecting β-strands 2 and 3, effectively transforming them into a single contiguous element of secondary structure. This elongated β-strand pairs with its equivalent from the second monomer to form an antiparallel double-helical coiled coil running from one lobe of the dimer to the other. Intriguingly, in active monomeric cystatins, the residues of the open interface form the first inhibitory hairpin loop, and the domain-swapped dimer is therefore expected to be inactive.

Canecystatin-1 NMR experiments resulted in high-quality data, exemplified by the ¹H/¹⁵N-HSQC spectrum shown in Fig. 3. The consensus chemical shift index (CSI) (www.bionmr.ualberta.ca/bds/software/csi) [30] plotted in Fig. 4A revealed the presence of four structured elements: an α-helix (α₁, residues 27–43) followed by three β-strands (β_{2,3}, residues 48–74; β₄, residues 78–89; and β₅, residues 94–104). This pattern is consistent with the presence of a domain-swapped protein in solution, in agreement with the crystal structures, the SEC profile and the DLS data described above.

However, a careful analysis of the NMR spectra revealed the presence of a significant number of correlated weak peaks in specific regions of the protein, indicating the presence of a second, less abundant conformation in slow exchange with the very abundant domain-swapped protein, depicted in Fig. 3. These additional peaks are present in all three-dimensional spectra, allowing for their sequential assignment. A similar observation has been made for domain-swapped cyanovirin-N, a protein that presents a completely different fold [31]. The CSI analysis of the additional weak peaks showed a break in the first β-strand involving residues 60–63 (Fig. 4B). This pattern of secondary structure is consistent with a monomer, as observed in other phytocystatins [4,9,24], and is probably the active species, in which the first inhibitory loop assumes a conformation compatible with protease binding.

A detailed comparison of the two forms reveals substantial resonance shifts induced by dimerization. To investigate this further, we applied chemical shift mapping, using amino acid-specific calculations of the combined chemical shift perturbation $\Delta\delta_{\text{comb}}$ as implemented in AUREMOL [32]. Backbone atoms (¹HN, ¹⁵NH, ¹³Cα, and ¹³CO) were used for this calculation. As shown in Fig. 5, values above the cut-off of σ_0 are limited to four regions: residues 6–8, together with residues 20 and 21, from the unstructured N-terminal region; residues 61–64 involving the first inhibitory hairpin loop; and residues 88–90 and 94 from the second inhibitory hairpin loop. The shift differences between both states were analyzed and correlated with the conformational changes that are expected between the unknown monomeric structure and that of the crystallographically solved domain-swapped dimer. For these comparisons, the structures of oryzacystatin-1 and pineapple cystatin were used as models for the monomeric state, following the sequence alignment shown in Fig. S4 and based on the similarities found in their consensus CSIs [22,24].

The observed chemical shift differences in the first inhibitory loop are consistent with the structural changes

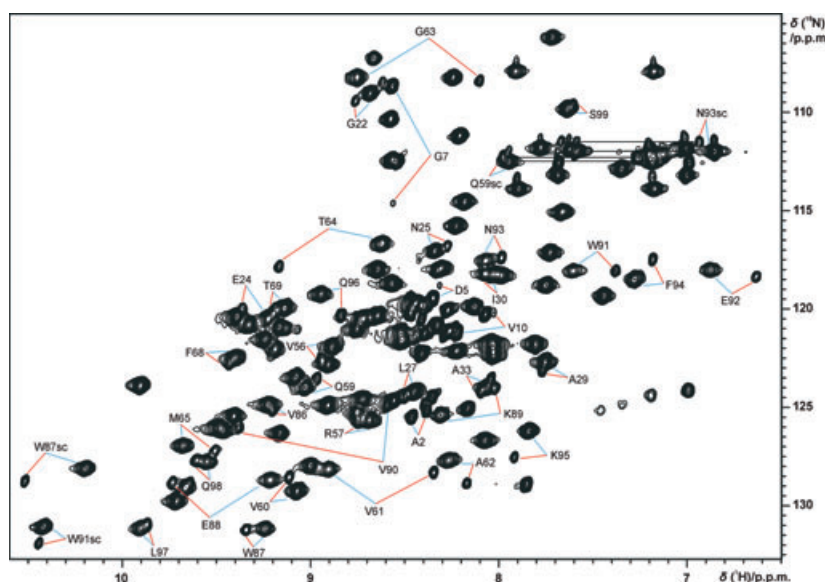


Fig. 3. Assigned high-resolution $^1\text{H}/^{15}\text{N}$ -HSQC spectrum of canecystatin-1 measured at a concentration of 1 mM. Only peaks corresponding to residues that present both conformations (monomer in red and dimer in blue) are indicated. The spectrum was obtained with an 800-MHz spectrometer equipped with a cryoprobe, at 303 K. sc, side chain NH or NH₂ peaks.

that arise from domain-swapping. Residues 61–64 show significant $\text{C}\alpha$ and CO upfield shifts, as well as HN , $\text{H}\alpha$ and $\text{C}\beta$ downfield shifts, indicating a transition from coil to β -sheet on going from the less to the more abundant state. A particularly strong effect could be observed for the amide proton and nitrogen of Thr64, suggesting perturbations arising from reorientation of the neighboring Trp87, owing to ring current effects (see below).

The conformational change to the first inhibitory loop (residues 61–64) seems to promote a knock-on effect on the adjacent second inhibitory loop (residues 88–94). However, in the latter case, the shift perturbations do not indicate a secondary structure change, as no combined upfield/downfield shift is present, as can be seen in Fig. 4. Rather, they suggest a rearrangement of the loop in order to accommodate or stabilize the global structural change to the protein. This appears also to be the case for the N-terminal region, which is unstructured in both crystal forms. Overall, all of the NMR data are consistent with a transition between an active monomeric state, in which the inhibitory hairpin loops are appropriately folded for protease binding, and a domain-swapped dimer, as seen in the crystal structures (Fig. 6). In Fig. 6, the first inhibitory loop (residues 61–64), which is part of the inhibitory triad, is shown in red, and undergoes a significant conformational change on domain-swapped dimerization. In the latter case, this region becomes what is known as the ‘open’ interface (that which is

observed only after domain-swapping). The schematic shown in Fig. 6 includes a partially unfolded monomer (B) in which the ‘closed’ interface between the helix and the β -sheet has been disrupted. It is assumed that this intermediate must exist in order for domain-swapped dimers to form.

Comparisons with other cystatins

Several differences were observed between the crystal structures of canecystatin-1 and those of other cystatins, one of the most apparent being the lack of electron density for the first 20 residues. Some other cystatin structures also lack electron density in this region, but this is usually restricted to the first 10 residues. This suggests that the N-terminal region of canecystatin-1 has increased flexibility, although we cannot rule out the possibility that this apparent discrepancy is the result of different crystal packing. A second interesting feature is the relative orientation of the two α -helices, which can be used as a measure of the torsion between the two lobes of the domain-swapped dimer. The visual observation that the angle between these helices is greater in canecystatin (Fig. 2) was quantified by using four different methods for defining the helical axes and then determining the angle between them. The results are shown in Fig. 2 and Table S1. On average, the interhelical angle in canecystatin is $> 60^\circ$, which is considerably larger than that observed for human cystatin C ($\sim 20^\circ$). This

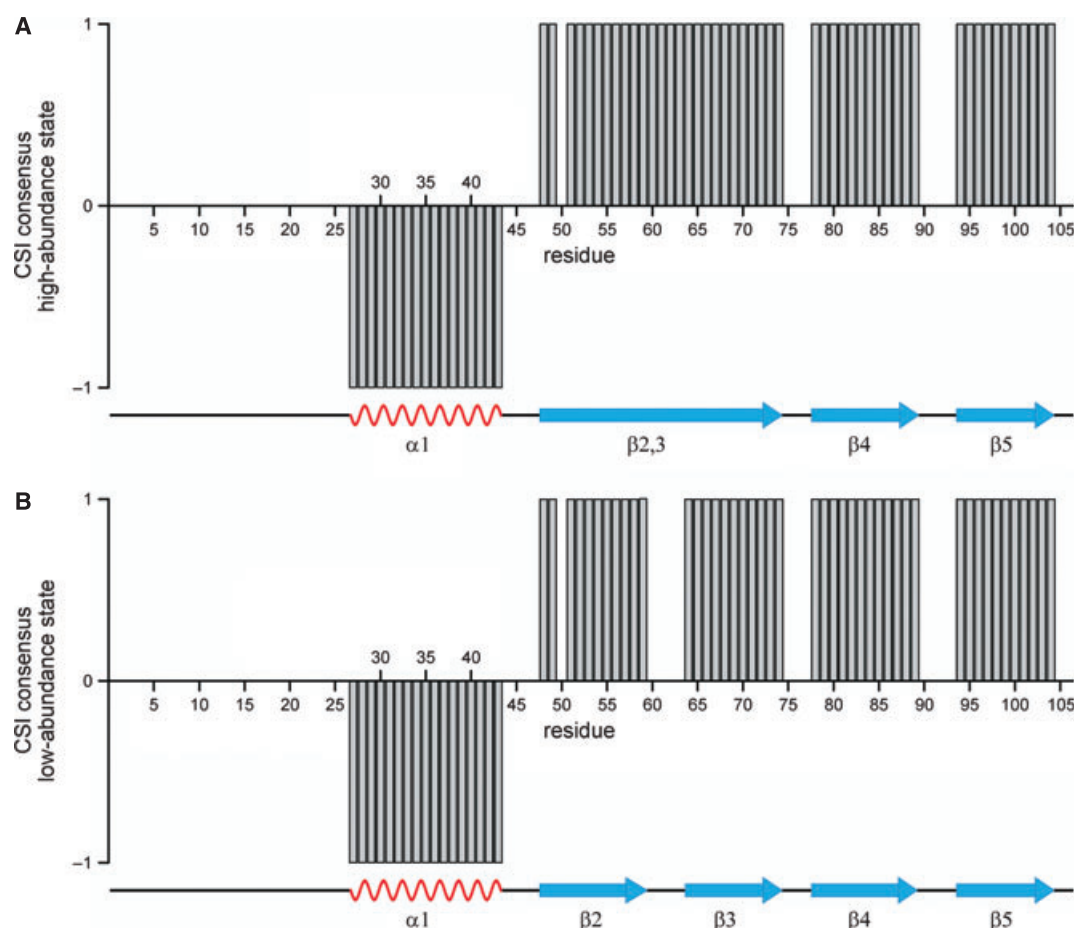


Fig. 4. Consensus CSI [30] of canecystatin-1 derived from the $^1\text{H}_\alpha$, $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ and ^{13}CO chemical shifts. (A) Plot obtained for the highly abundant state. (B) Plot obtained for the less abundant state. Regions identified as having α -helical and β -strand secondary structure are indicated.

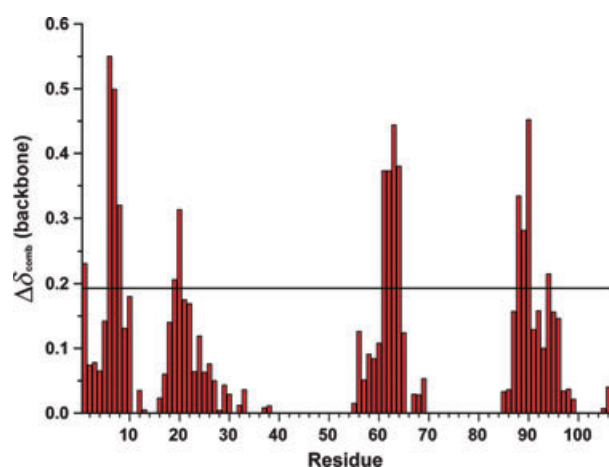


Fig. 5. Combined chemical shift perturbation [32] of canecystatin-1 derived from backbone atoms (^1HN , ^{15}NH , $^{13}\text{C}_\alpha$, and ^{13}CO). Residues with values above the cut-off σ_0 (black line) have significant resonance shifts as a consequence of the dimerization.

corresponds to greater torsion of the central coiled coil of β -strands in the latter than in the former (Fig. 2).

The crystal packing leads to the formation of an interface between two domain-swapped dimers from adjacent asymmetric units, which is present in both crystal forms (Fig. 7A). In this arrangement, Trp91 of each monomer, which projects from the convex face of the extended β -strand, is tightly buried within a pocket formed by Val60, Ala62, Gly63, Thr64 and Val90 of the neighboring domain-swapped dimer (Fig. 7B; see Fig. S4 for a sequence alignment). This arrangement is reminiscent of the previously described ‘handshake’ that occurs in the tetrameric structure of stefin B (PDB code [2OCT](#)) [33]. However, in this case, the tetrameric arrangement is retained in solution, and the interactions that occur in the handshake are believed to contribute to its stability. Furthermore, the stability of the stefin B tetramer in solution is believed to be

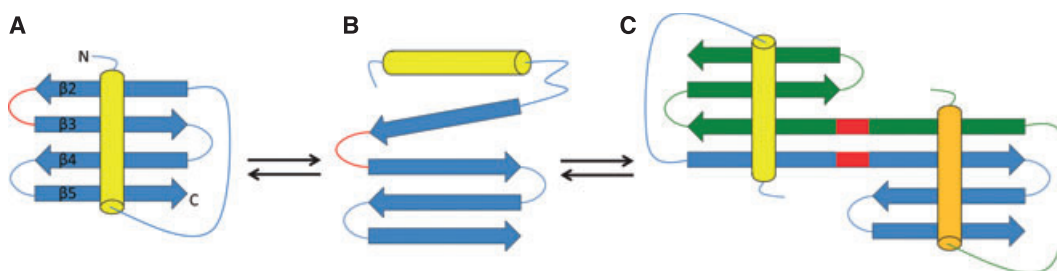


Fig. 6. Simplified scheme of the transition between a closed monomer and a domain-swapped dimer. The yellow and orange cylinders represent α -helices, and the blue and green arrows represent β -strands. (A) Closed monomer. (B) Partially unfolded monomer. (C) Domain-swapped dimer. In both (A) and (C), 'closed interfaces' are formed by the positioning of the α -helices over the β -sheets. The 'open interface', present only in (C), is represented by the two red segments in the long β -strands of the domain-swapped dimer, which correspond to the first inhibitory loop in (A).

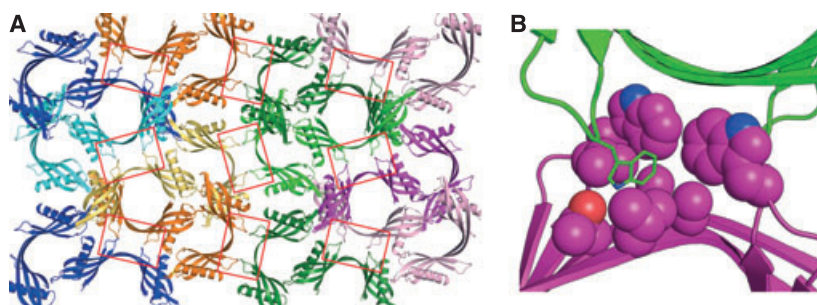


Fig. 7. (A) Crystal packing of canecystatin-1 in space group C222₁ (PDB code [3UL5](#)). The handshake-like contacts are highlighted by red rectangles. Each asymmetric unit contains two domain-swapped dimers, which are colored the same. (B) The chemical environment of Trp91. The interface between two domain-swapped dimers is shown, and each dimer is shown in a different color. The Trp91 in each monomer is < 4 Å away from the side chain of its preceding residue (Val90), as well as from Val60, Val61, Ala62 and Trp91 from one of the monomers from the adjacent dimer, and from Thr64, Val90 and Trp91 from the other.

related to the presence of Pro74, which is absent from canecystatin-1. Moreover, in stefin B, the interface appears to be tighter and the dimers closer together (Fig. S5). Our gel filtration experiments suggest that, in the case of canecystatin-1, a small population of tetramers may coexist with dimers and monomers in solution, but these are clearly less abundant than in the case of stefin B. In the case of human cystatin C, the crystal packing is somewhat different, leading to an interface between domain-swapped dimers in which the tryptophans are exposed at its periphery rather than buried at its center.

A feature that may be related to the observed torsion between the two lobes of the canecystatin-1 domain-swapped dimer is the result of the extended β -strands (the result of the contiguity of β -strands 2 and 3). Together, these two extended strands (one from each monomer) form a 26-residue double-stranded antiparallel coiled coil in all six independent polypeptide chains of the crystal structures of canecystatin-1. This elegant tertiary structural feature is rarely described, and results from the systematic

oscillation of the Φ, Ψ coordinates between two points in Ramachandran space [34]. This can be readily appreciated from Fig. S6, which shows regular alteration of both torsion angles for the greater part of the strand. However, towards the N-terminus, two classic-type β -bulges [35] accentuate the curvature of one strand while maintaining good hydrogen bonding geometry with its partner (Fig. S6B).

Implications for biological activity

Because of the high sequence identity between canecystatin-1 and tarocystatin (which is monomeric), the structures superpose very well (rmsd of 0.30 Å), allowing for the identification of critical residues responsible for conformational differences between these two forms of phytocystatin. The contacts at the so-called 'closed' interface (essentially that between the α -helix and the β -sheet) are preserved as two previously monomeric molecules exchange identical domains to form a domain-swapped dimer. As mentioned above, the small segment corresponding to the first inhibitory

hairpin loop in monomeric cystatins forms the ‘open’ interface of the dimer, with Val61 also contacting Trp87 of the opposite subunit. The conformation of this tryptophan is markedly different in the monomeric and dimeric forms, owing to its participation in this interface, consistent with the dramatic differences in chemical shift observed for this residue in the NMR experiments. All of the residues involved (Val61, Met65, and Trp87) are conserved in tarocystatin, suggesting that it, too, may display domain-swapping under appropriate physicochemical and/or biological conditions (Fig. 8).

Domain-swapping may be a mechanism by which canecystatin-1 activity is regulated. Owing to the absence of the first inhibitory loop (which is essential for the formation of the inhibitory binding site) [25,26], the domain-swapped dimer is expected to have no activity against cysteine proteases, and this may therefore represent a mechanism for the storage of canecystatin-1 in an inactive form, thus avoiding the inappropriate inhibition of endogenous cysteine proteases. During a period of physiological demand or when the plant is attacked by insects or pathogens, temporary fluctuations in the environment, such as pH variations, could favor the formation of the active monomers.

The A10 mutant

In a recent study, we reported a mutant protein, denominated A10, which shows greater inhibitory activity towards cathepsin B than the wild type [36]. It was generated by DNA shuffling, and is similar in sequence to canecystatin-1, except that it possesses the N-terminal region from oryzacystatin-1 (including a seven-residue deletion) and has two point mutations, I30T and L97Q. The latter participates in the closed interface between the α -helix and the β -sheet, and the glutamine substitution would be expected to destabilize

this interface and expose the hydrophobic core to solvent. A10 may therefore resemble the previously described human cystatin C L68Q mutant in some respects, by presenting a reduced energy barrier for reaching the partially unfolded state, as a consequence of the loss of a hydrophobic residue [37]. The reason that most cystatins have low inhibitory activity towards cathepsin B is that there is a large occluding loop in the protease, which leads to steric hindrance with the inhibitor when present as a correctly folded monomer. Perturbation of the closed interface may favor a partially unfolded monomer, in which the three components of the inhibitory site are uncoupled, leading to sufficient structural malleability to reduce steric hindrance and allow access to the protease active site [36]. Moreover, our data show that canecystatin-1 is present in solution predominantly as domain-swapped dimers, which are unable to inhibit cysteine proteases, owing to the absence of the appropriate conformation for inhibitory loop 1. Therefore, activity measurements of wild-type canecystatin-1 may represent the activity of the small population of active monomers, which would produce a lower apparent inhibitory activity. In the case of the A10 mutant, on the other hand, enzyme inhibition would be attributable to a partially unfolded intermediate that retains inhibitory activity.

Cystatin-based inhibitors of cathepsin B are of great biotechnological interest, because of the latter's involvement in tumor development and neurodegenerative diseases [2,10,11]. The information on the three-dimensional structure of canecystatin that we report here allows us to imagine the design of mutations affecting the closed interface, leading to partially destabilized inhibitors. In so doing, it may be possible to generate mutants with more attractive inhibitory properties towards cathepsin B while improving their

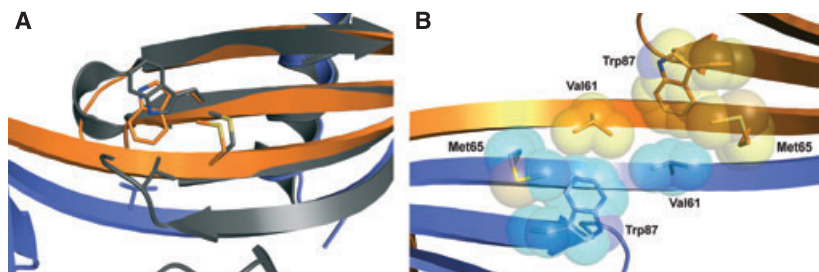


Fig. 8. The tryptophan switch. (A) Superposition of tarocystatin, colored in gray (PDB code [3IMA](#)), and canecystatin-1, colored in blue and orange (PDB code [3UL5](#)). Val61, Met65 and Trp87 are depicted as sticks, and the different conformation of Trp87 is clearly evident. (B) The tryptophan switch in a canecystatin domain-swapped dimer. Trp87 from β -strand 4 participates in the open interface by undergoing novel van der Waals interactions with Val61. Met65 also interacts with both Val61 and Trp87, possibly contributing to the stability of the domain-swapped form. All mentioned residues are labeled, and shown as sticks and transparent spheres.

solubility with respect to A10 itself, which has a tendency to aggregate.

Experimental procedures

Recombinant protein expression and purification

The cDNA coding for canecystatin-1, contained within clone SCCRZ2001G09 (accession no. AY119689), originated from the Sugarcane Genome Project-SUCEST (FAPESP). The coding region for the protein was obtained by amplification by PCR with the following primers: CaneF, 5'-GGG ATG GCC CAT ATG GCC GAG GCA CACA AC GGG-3'; and CaneR, 5'-CC GAA TTC TTA GG C GTC CCC GAC CGG CT-3'. The amplification product was cleaved with *NdeI* and *EcoRI*, and ligated into pET28a cleaved with the same enzymes. The recombinant expression and purification was carried out as previously described [38]. Briefly, an Ni^{2+} -nitrilotriacetic acid superflow column (Qia-gen, Valencia, USA) was used in the first step of purification to capture the His-tagged canecystatin-1, which was eluted in 250 mM imidazole after several wash steps with variable imidazole concentrations. Subsequently, SEC was performed on a Superose 12 10/300 GL column, to obtain high-purity canecystatin-1. For the isotopic $^{13}\text{C}/^{15}\text{N}$ -labeled samples, M9 minimal medium prepared with $^{15}\text{NH}_4\text{Cl}$ and $[^{13}\text{C}]\text{glucose}$ was used from the beginning of cell culture. The His-tag was cleaved overnight with thrombin ($20 \text{ U}\cdot\text{mg}^{-1}$), at 4°C .

Crystallization and X-ray data collection

Canecystatin-1 crystals were grown by use of the sitting drop diffusion vapor method at 18°C , with His-tagged protein. One microliter of a $10 \text{ mg}\cdot\text{mL}^{-1}$ protein sample was mixed with an equal volume of the reservoir solution. Diffraction-quality crystals were obtained in two distinct conditions. The first comprised 100 mM Bis/Tris (pH 7.0) and 3.5 M sodium formate, and the second comprised 100 mM Hepes (pH 7.75), 25% poly(ethylene glycol) 4000, and 200 mM MgCl_2 . A 1 : 1 mixture of the well solution with glycerol was used as cryoprotectant. Two datasets were collected at 100 K with the MX2 beamline of the LNLS synchrotron (Campinas, Brazil).

Data processing, structure determination, and refinement

Diffraction data up to $2.3\text{-}\text{\AA}$ resolution for the face-centered orthorhombic crystal (space group C222_1) and $2.63\text{-}\text{\AA}$ resolution for the primitive hexagonal crystal (space group P6_422) were indexed and integrated with MOSFLM [39], and scaled with SCALA. A solution for the orthorhombic dataset was obtained by molecular replacement with PHASER [40] and the structure of tarocystatin extracted from

its complex with papain (PDB code [3IMA](#)) as the search model. A search for four copies yielded a solution with an log-likelihood gain (LLG) of 929.39, an *R*-factor of 49.32, and a translation function Z-score (TFZ) of 19.5. Refinement was performed with COOT [41] and PHENIX [42]. After the first refinement cycle, it was unambiguous that the four monomers within the asymmetric unit were, in fact, two domain-swapped dimers. The connectivity, chain labels and residue numbers were corrected by manual model rebuilding. In the late refinement stages, solvent molecules were introduced manually with COOT. Model quality was assessed by monitoring R_{work} and R_{free} , as well as the results of MOLPROBITY. A molecular replacement solution for the primitive hexagonal dataset was obtained with PHASER and the domain-swapped dimer formed by chains A and B of a partially refined model of the orthorhombic form.

DLS

DLS measurements were performed for canecystatin-1 at 25°C with a Zetasizer Nano μV (Malvern Instruments, Worcestershire, UK) equipped with a laser ($\lambda = 830 \text{ nm}$) and a detector at 90° . Three measurements were performed for the purpose of averaging on a sample containing 1 mM protein in 50 mM potassium phosphate buffer containing 100 mM NaCl. The results showed a highly monodisperse sample with a polydispersity index of 12.3% and a hydrodynamic radius of $2.8 \pm 0.4 \text{ nm}$, consistent with a predominantly dimeric species.

NMR

For NMR studies, the unlabeled and the $^{13}\text{C}/^{15}\text{N}$ -labeled purified proteins were concentrated up to 1.0 mM by ultrafiltration with an Amicon Ultra 3 kDa MWCO (Millipore, Billerica, MA, USA), and dialyzed against either a 50 mM potassium phosphate buffer in H_2O or a 50 mM potassium phosphate buffer in 99.95% D_2O , both containing 100 mM NaCl and 1 mM NaN_3 . DSS was added to the samples to a final concentration of 0.1 mM, and 5% D_2O was added to the sample in H_2O . NMR data were recorded at 303 K on Bruker DRX-600 and DRX-800 spectrometers, both equipped with a cryoprobe. ^1H , ^{13}C and ^{15}N sequential resonance assignments were obtained with the following standard experiments: HNCA, HNCO, CBCA(CO)NH and ^1H - ^{15}N -HSQC ($^{13}\text{C}/^{15}\text{N}$ -labeled sample in H_2O buffer); HCCH-TOCSY and $^1\text{H}/^{13}\text{C}$ -HSQC ($^{13}\text{C}/^{15}\text{N}$ -labeled sample in D_2O buffer); and $^1\text{H}/^1\text{H}$ -TOCSY and $^1\text{H}/^1\text{H}$ -NOESY (unlabeled sample in H_2O buffer). Proton chemical shifts were referenced to the ^1H resonance frequency of the methyl groups in DSS; ^{13}C and ^{15}N resonances were indirectly calibrated according to IUPAC recommendations [43]. Data were processed in TOPSPIN 3.0 (Bruker BioSpin, Rheinstetten, Germany) and evaluated in AUREMOL (www.

auremol.de). The complete sequential assignments of both the dimeric [22] and monomeric forms have been deposited in the BioMagResBank under accession numbers 18317 and 18839, respectively.

Acknowledgements

This work was supported by the State of São Paulo Research Foundation (FAPESP, grants 08/58316-5 and 10/09100-0) and CNPq. N. F. Valadares received a postdoctoral fellowship from FAPESP.

References

- Benchabane M, Schlüter U, Vorster J, Goulet MC & Michaud D (2010) Plant cystatins. *Biochimie* **92**, 1657–1666.
- Turk V, Stoka V & Turk D (2008) Cystatins: biochemical and structural properties, and medical relevance. *Front Biosci* **13**, 5406–5420.
- Oliveira JP, Magliarelli HF, Pereira FV, Gianotti A, Soares-Costa A, Henrique-Silva F, Wakamatsu A, Soares IC, Nonogaki S, Travassos LR, *et al.* (2011) Sugarcane cystatin CaneCPI-4 inhibits melanoma growth by angiogenesis disruption. *J Cancer Sci Ther* **3**, 161–167.
- 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.
- 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.
- 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.
- Jenko S, Dolenc I, Guncar G, Dobersek A, Podobnik M & Turk D (2003) 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**, 875–885.
- Renko M, Požgan U, Majera D & Turk D (2010) 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**, 4338–4345.
- Chu MH, Liu KL, Wu HY, Yeh KW & Cheng YS (2011) Crystal structure of tarocystatin–papain complex: implications for the inhibition property of group-2 phytocystatins. *Planta* **234**, 243–254.
- Mohamed MM & Sloane BF (2006) Cysteine cathepsins: multifunctional enzymes in cancer. *Nat Rev Cancer* **6**, 764–775.
- Hook VY (2006) Unique neuronal functions of cathepsin L and cathepsin B in secretory vesicles: biosynthesis of peptides in neurotransmission and neurodegenerative disease. *Biol Chem* **387**, 1429–1439.
- Arai S, Matsumoto I, Emori Y & Abe K (2002) Plant seed cystatins and their target enzymes of endogenous and exogenous origin. *J Agric Food Chem* **50**, 6612–6617.
- Margis R, Reis EM & Villeret V (1998) Structural and phylogenetic relationships among plant and animal cystatins. *Arch Biochem Biophys* **359**, 24–30.
- Hwang JE, Hong JK, Je JH, Lee KO, Kim DY, Lee SY & Lim CO (2009) Regulation of seed germination and seedling growth by an *Arabidopsis* phytocystatin isoform, AtCYS6. *Plant Cell Rep* **28**, 1623–1632.
- Neuteboom LW, Matsumoto KO & Christopher DA (2009) An extended AE-rich N-terminal trunk in secreted pineapple cystatin enhances inhibition of fruit bromelain and is posttranslationally removed during ripening. *Plant Physiol* **151**, 515–527.
- Goulet MC, Dallaire C, Vaillancourt LP, Khalf M, Badri AM, Preradov A, Duceppe MO, Goulet C, Cloutier C & Michaud D (2008) Tailoring the specificity of a plant cystatin toward herbivorous insect digestive cysteine proteases by single mutations at positively selected amino acid sites. *Plant Physiol* **146**, 1010–1019.
- Carrillo L, Martinez M, Alvarez-Alfageme F, Castañera P, Smagghe G, Diaz I & Ortego F (2011) A barley cysteine-proteinase inhibitor reduces the performance of two aphid species in artificial diets and transgenic *Arabidopsis* plants. *Transgenic Res* **20**, 305–319.
- Senthilkumar R, Cheng CP & Yeh KW (2010) Genetically pyramiding protease-inhibitor genes for dual broad-spectrum resistance against insect and phytopathogens in transgenic tobacco. *Plant Biotechnol J* **8**, 65–75.
- Kiggundu A, Muchwezi J, Van der Vyver C, Viljoen A, Vorster J, Schlüter U, Kunert K & Michaud D (2010) Deleterious effects of plant cystatins against the banana weevil *Cosmopolites sordidus*. *Arch Insect Biochem Physiol* **73**, 87–105.
- Samac DA & Smigocki AC (2003) Expression of oryzacystatin I and II in alfalfa increases resistance to the root-lesion nematode. *Phytopathology* **93**, 799–804.
- Chan YL, Yang AH, Chen JT, Yeh KW & Chan MT (2010) Heterologous expression of taro cystatin protects transgenic tomato against *Meloidogyne incognita* infection by means of interfering sex determination and suppressing gall formation. *Plant Cell Rep* **29**, 231–238.
- Cavini IA, de Oliveira-Silva R, de Almeida Marques I, Kalbitzer HR & Munte CE (2012) Chemical shift assignments of the canecystatin-1 from *Saccharum*

- officinarum*. *Biomol NMR Assign*. doi:10.1007/s12104-012-9401-2.
- 23 Nissen MS, Kumar GN, Youn B, Knowles DB, Lam KS, Ballinger WJ, Knowles NR & Kang C (2009) Characterization of *Solanum tuberosum* multicystatin and its structural comparison with other cystatins. *Plant Cell* **21**, 861–975.
 - 24 Irene D, Chen BJ, Lo SH, Liu TH, Tzen JT & Chyan CL (2012) Resonance assignments and secondary structure of a phytocystatin from *Ananas comosus*. *Biomol NMR Assign* **6**, 99–101.
 - 25 Ekiel I & Abrahamson M (1996) Folding-related dimerization of human cystatin C. *J Biol Chem* **271**, 1314–1321.
 - 26 Ekiel I, Abrahamson M, Fulton DB, Lindahl P, Storer AC, Levadoux W, Lafrance M, Labelle S, Pomerleau Y, Groleau D, *et al.* (1997) NMR structural studies of human cystatin C dimers and monomers. *J Mol Biol* **271**, 266–277.
 - 27 Janowski R, Kozak M, Abrahamson M, Grubb A & Jaskolski M (2005) 3D domain-swapped human cystatin C with amyloidlike intermolecular beta-sheets. *Proteins* **61**, 570–578.
 - 28 Janowski R, Kozak M, Jankowska E, Grzonka Z, Grubb A, Abrahamson M & Jaskolski M (2001) Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nat Struct Biol* **8**, 316–320.
 - 29 Bennett MJ, Schlunegger MP & Eisenberg D (1995) 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci* **4**, 2455–2468.
 - 30 Wishart DS, Sykes BD & Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647–1651.
 - 31 Barrientos LG, Louis JM, Botos I, Mori T, Han Z, O’Keefe BR, Boyd MR, Wlodawer A & Gronenborn AM (2002) The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure* **10**, 673–686.
 - 32 Schumann FH, Riepl H, Maurer T, Gronwald W, Neidig K-P & Kalbitzer HR (2007) Combined chemical shift changes and amino acid specific chemical shift mapping of protein–protein interactions. *J Biomol NMR* **39**, 275–289.
 - 33 Jenko Kokalj S, Guncar G, Stern I, Morgan G, Rabzelj S, Kenig M, Staniforth RA, Waltho JP, Zerovnik E & Turk D (2007) Essential role of proline isomerization in stefin B tetramer formation. *J Mol Biol* **366**, 1569–1579.
 - 34 Salemme FR (1983) Structural properties of protein beta-sheets. *Prog Biophys Mol Biol* **42**, 95–133.
 - 35 Richardson JS, Getzoff ED & Richardson DC (1978) The beta bulge: a common small unit of nonrepetitive protein structure. *Proc Natl Acad Sci USA* **75**, 2574–2578.
 - 36 Valadares NF, Dellamano M, Soares-Costa A, Henrique-Silva F & Garratt RC (2010) Molecular determinants of improved cathepsin B inhibition by new cystatins obtained by DNA shuffling. *BMC Struct Biol* **10**, 30.
 - 37 Palsdottir A, Abrahamson M, Thorsteinsson L, Arnason A, Olafsson I, Grubb A & Jensson O (1988) Mutation in cystatin C gene causes hereditary brain haemorrhage. *Lancet* **2**, 603–604.
 - 38 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.
 - 39 Leslie AGW & Powell HR (2007) Processing diffraction data with mosflm. In *Evolving Methods for Macromolecular Crystallography* (Read RJ & Sussman JL, eds), pp. 41–51. Springer, Dordrecht, The Netherlands.
 - 40 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.
 - 41 Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **D66**, 486–501.
 - 42 Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **D66**, 213–221.
 - 43 Markley JL, Bax A, Arata Y, Hilbert CW, Kaptein R, Sykes BD, Wright PE & Wüthrich K (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *Pure Appl Chem* **70**, 117–142.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher’s web site:

Fig. S1. SDS/PAGE of the eluted peaks obtained in the SEC experiment shown in Fig. 1.

Fig. S2. DLS data for the His-tag-free protein.

Fig. S3. Omit map of Val60, Val61, Ala62, and Gly63, which form the open interface in domain-swapped dimers and the first inhibitory loop in monomers.

Fig. S4. Sequence alignment of the phytocystatins canecystatin-1, tarocystatin, oryzacystatin-1, and pineapple cystatin, together with human cystatin C.

Fig. S5. Comparison between canecystatin-1, stefin B, and human cystatin C.

Fig. S6. Particularities within the canecystatin-1 β -sheet.

Table S1. Angle between the helices in domain-swapped cystatin dimers.