

RESEARCH ARTICLE

TcCYPR04, a Cacao Papain-Like Cysteine-Protease Detected in Senescent and Necrotic Tissues Interacts with a Cystatin TcCYS4

Thyago Hermylly Santana Cardoso¹, Ana Camila Oliveira Freitas¹, Bruno Silva Andrade², Aurizangela Oliveira de Sousa¹, André da Silva Santiago¹, Daniela Martins Koop¹, Karina Peres Gramacho³, Fátima Cerqueira Alvim¹, Fabienne Micheli^{1,4}, Carlos Priminho Pirovani^{1*}

1 Centro de Biotecnologia e Genética, Universidade Estadual de Santa Cruz, UESC, Rodovia Ilhéus-Itabuna, Km 16, Ilhéus/BA, Brasil, **2** Universidade Estadual do Sudoeste da Bahia, UESB, Jequié-BA, Brazil, **3** Centro de Pesquisa do Cacau (CEPEC), Comissão Executiva do Plano da Lavoura Cacaueira (CEPLAC), Fitopatologia, Ilhéus, BA, Brazil, **4** CIRAD, UMR AGAP, Montpellier, France

* pirovani@uesc.br



OPEN ACCESS

Citation: Cardoso THS, Freitas ACO, Andrade BS, Sousa AOd, Santiago AdS, Koop DM, et al. (2015) TcCYPR04, a Cacao Papain-Like Cysteine-Protease Detected in Senescent and Necrotic Tissues Interacts with a Cystatin TcCYS4. PLoS ONE 10(12): e0144440. doi:10.1371/journal.pone.0144440

Editor: Eugene A. Permyakov, Russian Academy of Sciences, Institute for Biological Instrumentation, RUSSIAN FEDERATION

Received: July 20, 2015

Accepted: November 18, 2015

Published: December 7, 2015

Copyright: © 2015 Cardoso et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: THSC was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). CPP was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [<http://www.cnpq.br>], process number 305309/2012-9). The authors thank Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) [<http://www.fapesb.ba.gov.br/>] and the CNPq (PRONEM, process number PNE0005/

Abstract

The interaction amongst papain-like cysteine-proteases (PLCP) and their substrates and inhibitors, such as cystatins, can be perceived as part of the molecular battlefield in plant-pathogen interaction. In cacao, four cystatins were identified and characterized by our group. We identified 448 proteases in cacao genome, whereof 134 were cysteine-proteases. We expressed in *Escherichia coli* a PLCP from cacao, named TcCYSPR04. Immunoblottings with anti-TcCYSPR04 exhibited protein increases during leaf development. Additional isoforms of TcCYSPR04 appeared in senescent leaves and cacao tissues infected by *Moniliophthora perniciosa* during the transition from the biotrophic to the saprophytic phase. TcCYSPR04 was induced in the apoplastic fluid of Catongo and TSH1188 cacao genotypes, susceptible and resistant to *M. perniciosa*, respectively, but greater intensity and additional isoforms were observed in TSH1188. The fungal protein MpNEP induced PLCP isoform expression in tobacco leaves, according to the cross reaction with anti-TcCYSPR04. Several protein isoforms were detected at 72 hours after treatment with MpNEP. We captured an active PLCP from cacao tissues, using a recombinant cacao cystatin immobilized in CNBr-Sepharose. Mass spectrometry showed that this protein corresponds to TcCYSPR04. A homology modeling was obtained for both proteins. In order to become active, TcCYSPR04 needs to lose its inhibitory domain. Molecular docking showed the physical-chemical complementarities of the interaction between the cacao enzyme and its inhibitor. We propose that TcCYSPR04 and its interactions with cacao cystatins are involved in the senescence and necrosis events related to witches' broom symptoms. This molecular interaction may be the target for future interventions to control witches' broom disease.

2011) for the financial support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Proteases (also known as peptidases, proteinases or proteolytic enzymes) are mainly divided in two groups: i) exoproteases, which cleave at the ends of the protein and are subdivided into amino- or carboxyproteases; and ii) endoproteases, cleaving on the interior of the protein. The classification of endoproteases depends on the type of residue at the active site and according to MEROPS database (<http://merops.sanger.ac.uk/index.shtml>) it includes: aspartic (Asp residue), cysteine (Cys), glutamic (Glu), metallo (metal ion), asparagine (Asn), serine (Ser), and threonine (Thr) proteases, and mixed (Cys, Ser, Thr) and unknown (unknown catalytic type) classes [1]. Cysteine-proteases consist of a single chain folded to form two domains—an α -helix-rich and a β -barrel-like domain—separated by a cleft containing the active site formed by the Cys and His residues, each one on each domain [2]. For proteolysis, this Cys acts as a nucleophile activated by His in the active site [3, 4]. To date, proteases are represented by 220 families belonging to 13 clans, whereof 72 families and 8 clans correspond to cysteine-proteases [1]. Papain-like cysteine-proteases (PLCPs) are grouped in clan CA, family C1. PLCPs are usually 23–30 kDa in size, and contain an inhibitory pro-region folded back onto the substrate-binding cleft; this inhibitory region needs to be removed for protease activation.

Many plant PLCPs are produced with a signal peptide indicating that they are secreted in the endomembrane system [4]. In plants, PLCPs are involved in pathogen perception, disease resistance, defense against insects and senescence [4]. Proteolysis during plant-pathogen interactions probably promotes the selection of counteracting inhibitors, non-cleavable substrates and other means to evade proteolysis [5, 6]. Therefore, the interaction of proteases with their substrates and inhibitors can be seen as a molecular battlefield [7]. The structure of the complex between exogenous papain and cystatin is well analyzed [8, 9], but little is known about the interaction between PLCP and cystatin in the same organism.

Cystatins are divided into four families: i) family-1 cystatins (stefin family) composed of proteins devoid of sulfur bridges. This group has a molecular mass of about 11 kDa and is generally present in the cytosol; ii) family-2 cystatins (cystatin family), comprising proteins that have sulfide bridges; iii) family-3 cystatins (kininogen family) characterized by high molecular-weight glycoproteins with three repetitions. These proteins contain tandem domains that result from the gene duplication of the family-2 cystatins, and finally; iv) family-4 cystatins (phyto-cystatins), which includes all cystatins described for plants to date [10, 11].

Phytocystatins are specific plant proteins that inhibit PLCPs; they are small, mostly ranging from 12 to 16 kDa. Several roles have been attributed to phytocystatins, since the regulation of various endogenous proteolytic processes to the inhibition of exogenous cysteine-proteases secreted by predatory or pathogenic organisms during herbivory or infection [12, 13]. Previous work performed by our group identified and characterized four phytocystatins [14] from *Theobroma cacao* L.-*M. perniciosa* interaction cDNA libraries [15]. These cystatins, named TcCys1 to TcCys4, were expressed in *E. coli* and the recombinant proteins displayed inhibitory activity *in vitro* against commercial papain, and antifungal activity against *M. perniciosa* [14]. TcCYS4 accumulates when green broom symptoms occur, probably inhibiting the cysteine-proteases involved in programmed cell death in infected tissues [16].

In this work we have produced a polyclonal antibody against a PLCP (TcCYSPR04—E.C. 3.4.22.16) identified in cDNA libraries from the interaction between *T. cacao* and *M. perniciosa*, and analyzed the accumulation of the protein in different cacao tissues by Western blot. The TcCYS04 accumulation was also investigated in tobacco leaves treated with recombinant ethylene-inducing protein from *M. perniciosa* (MpNEP) [17, 18]. Recombinant sepharose-immobilized cacao cystatin captured an active PLCP from cacao tissue, whose was confirmed by mass spectrometry as TcCYSPR04. The structures of TcCYSPR04 and TcCYS4 were

obtained by homology modeling method. The interaction of proteins was predicted by computational docking. Our data indicate that the interaction between protease (TcCYSPR04) and cystatin (TcCYS4) are involved in the senescence and necrosis events that can be triggered in development of witches' broom symptoms; therefore, the balance between these proteins could be exploited for controlling the disease.

Material and Methods

Plant material and Protein extraction

Primordial leaves of Catongo cacao susceptible genotype (seedlings aged 20 days after plantation) were inoculated with a suspension containing 200.000 basidiospores \times mL⁻¹ of *M. perniciosa* [19] or water (control), and kept in a greenhouse until the end of the experiment. Leaves were harvested at three stages after inoculation: E1—leaves up to 10 days; E2—leaves for 15–25 days (without necrosis spots visible); E3—leaves with more than 35 days (necrosis spots visible). Leaves without visible injury (young leaves) and leaves with chlorotic appearance (senescent leaves), as well as mature leaves infected with *M. perniciosa* with absence of necrosis (green brooms) from Catongo variety were collected in an experimental field at UESC—Ilhéus, Bahia, Brazil. All the collected samples were used to total protein extraction [20].

At experimental field, we also collected leaves from healthy and infected Catongo (CAT) and TSH1188 (TSH), a resistant genotype, to perform the apoplastic fluid proteins extraction, according to Pirovani et al [20]. Tobacco (*Nicotiana tabacum*) plants were used to necrosis assay induced by NEP [17] in leaves. For this, tobacco leaves maintained under aseptic conditions were sprayed with 1.4 μ M of recombinant NEP from *M. perniciosa* (MpNEP) [18] and collected at 24, 36, 40, 60 and 72 hours after spraying. As experimental control were used tobacco leaves maintained in the same conditions, but it not treated with MpNEP. Total protein extracts were obtained from these leaves [20].

Quantitative evaluation and classification of proteases in the genome of COCOA

Proteases sequences were retrieved using the database of the genome sequence of a Belizian Criollo genotype (B97-61/B2) provided by the International Cocoa Genome Sequencing Consortium [21]. The identification of proteases was done by searching for “keywords” in the database containing the BLAST analysis results of total genomic sequences of *T. cacao* against *Arabidopsis thaliana*, *Populus trichocarpa*, *Glycine max* and *Vitis vinifera*. Initially, the words “protease”, or “proteinase” or “peptidase” have been used and then a screening was performed by analyzing each result in order to eliminate those corresponding to other proteins, but containing the keywords, as in the case of protease inhibitors and proteases targets.

Sequence analysis

The fragment of cysteine-protease pre-pro-TcCYSPR04 (E.C. 3.4.22.16) and the cystatin TcCYS4 sequences were identified in expressed sequence tags (EST) libraries from the cacao-*M. perniciosa* interaction [15]. The complete sequence of TcCYSPR04 was identified in the EST data bank from Cirad/France (<http://esttik.cirad.fr/>) [21] and confirmed in CocoaGenDB (<http://cocoagendb.cirad.fr/gbrowse/cgi-bin/gbrowse/theobroma/>) [22]. The sequences were compared for putative function determination and annotation to the public sequence database (<http://www.ncbi.nih.gov/BLAST/>) using BLASTX and TBLASTX. Alignments showing similarity to an expected value lower than 1.10^{-4} were considered significant.

Protein translation was carried out using Translate tool (<http://au.expasy.org/tools/dna.html>); the open reading frame (ORF) was defined using the ORF Finder program (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>); peptide signal was predicted using SignalP (<http://www.cbs.dtu.dk/services/SignalP/>); and the functional domains of the proteins were identified using the Pfam software (<http://www.ebi.ac.uk/Tools/InterProScan/>) and the MEROPS database (<http://merops.sanger.ac.uk/index.shtml>) [23]. The ExPASy tools (<http://www.expasy.org>) were used to determine the molecular mass and pI of the proteins (http://www.expasy.org/cgi-bin/pi_tool), the putative phosphorylation sites (NetPhos 2.0, <http://www.cbs.dtu.dk/services/NetPhos/>) and the putative glycosylation sites (NetGlycate 1.0 Server and OGPET v1.0, http://www.geneinfinity.org/sp/sp_proteinptmodifs.html; YinOYang 1.2, <http://www.cbs.dtu.dk/services/YinOYang/>). Prediction of the secondary structures of proteins was obtained using the PSIPRED-GenThreader (<http://bioinf.cs.ucl.ac.uk/index.php?id=780>).

Expression of recombinant cysteine-proteases and antibody production

The ORFs encoding cysteine-protease proteins were obtained by amplification, using the following forward and reverse primers: TcCYSPRot_CatF GTTTCAGAAACATATGTTGG GAGCTGC and TcCYSPRotR AACCTCAACCCTCGAGATGGACCAACTAC, which contained restriction sites for *Xho*I and *Nde*I, respectively, for cloning into pET28a, according to standard techniques for cloning [24]. Transformed cells (*E. coli* Rosetta—DE3) containing the recombinant plasmids were grown at 37°C until reaching OD_{600nm} = 0.7, induced with 0.4 mM IPTG (isopropyl-β-D-thiogalactopyranoside) for 4 h, harvested, and processed. The lysate was centrifuged at 13,000 g, 4°C, for 15 min and soluble and insoluble fractions were obtained. Fusion proteins with a histidine tail were purified using a His-Trap FF Crude column (GE Healthcare), following the manufacturer's instructions. Insoluble recombinant cacao cysteine-proteases were dissolved with buffer solution 6 M urea prior to loading onto the column, and eluted in lyses buffer containing 250 mM imidazole and 6 M urea. Protein concentration was determined by the Bradford method [25].

Antibody production

Anti-TcCYSPR04 polyclonal antibody was obtained by rabbit immunization with the purified recombinant protein (His-tagged TcCYSPR04), according to Pirovani et al [14]. The rabbit was maintained under standard conditions in the animal house at UESC. All immunizations were performed in accordance with protocols approved by the Ethics Commission in Animal Experimentation (CEUA—UESC) [Protocol number 025/09].

The antibody obtained was purified by affinity using the His-tagged TcCYSPR04 antigen immobilized in nitrocellulose membrane [24]. The specificity of the anti-TcCYSPR04 serum was previously evaluated with protein extracts to determine its titer, whereas immunoblotting experiments were conducted at a 1:4000 dilution of this serum.

Immunodetection of cysteine-proteases in cacao and tobacco tissues

Total proteins were extracted under native conditions, as described by Pirovani et al [20]. For Western blot analyses, equal amounts (5 μg) of each protein samples were separated by 15% SDS-PAGE and electroblotted onto Hybond-C Extra membrane (Amersham Biosciences, Buckinghamshire, UK). The protein blot was blocked with 5% casein in TBS-T buffer (20 mM Tris-HCl pH 7.6; 0.8% NaCl; 0.1% Tween 20) and incubated with specific polyclonal anti-TcCYSPR04 antibody, for 1 h, at room temperature. An alkaline phosphatase-labeled anti-rabbit antibody (Life Technologies, Carlsbad, CA, USA) was used as secondary antibody (dilution 1:5,000) [24]. The detection system was NBT/BCIP (Promega, Madison, Wisconsin, USA).

Protease trap

Capture system were used with BSA (control) and recombinant cystatins TcCYS3 and TcCYS4 previously cloned and expressed [14] coupled to CNBr-activated Sepharose TM 4 fast Flow (GE Healthcare), according to the manufacturer instructions. The protease capture was performed using total proteins extracted of the tissues cacao from Catongo variety as described by Pirovani et al [14]. The captured proteases were analyzed regarding protease activity by a qualitative assay. For this, 15% SDS-PAGE and 7% PAGE with ampholytes (IEF—pH 4.0–8.0) were used. After migration, the gels were washed in Triton X-100, at room temperature, for 30 min for SDS removal and protease renaturation. After then, the gels were overlaid with 8% polyacrylamide-0.1% gelatin-gel in the presence of protease activity buffer at room temperature, for 12h, as described above [26, 27]. Activity gels were stained with 0.01% Coomassie blue G-250 [28].

Mass Spectrometry Identification of the captured protease

The captured proteases were mixed with 25 mM of NH_4HCO_3 and 10 mM of dithiothreitol (DTT), and kept at 60°C for 30 min. Then, 50 mM of iodoacetamide were added to the sample, which was kept at room temperature, for 30 min, in the dark, and finally digested with trypsin (25 ng/ μl) at 37°C, for 12h, according to the manufacturer's instructions (Promega). The resulting tryptic digests were vacuum concentrated (Concentrator 5301, Eppendorf), desalted using a pre-Symmetry column (Waters, Mildford, MA, USA) C18 (5 μm , 180 μm in inner diameter x 20 mm long), and then fractionated by C18 reverse phase chromatography column (100 mm x 100 μm , 1.7 μm particles) on the nanoAcquity UPLC (WATERS), for 50 min, under 0.6 $\mu\text{L}\cdot\text{min}^{-1}$ acetonitrile flux. The following gradient was used: 1% for 1 min, 1%-50% in 40 min, 50%-85% in 5 min, 85% for 2 min, 85%-1% in 1 min, 1% for 2 min. Afterwards, the peptides were deionized at 3,000 V and split on positive mode, with a minimum relative intensity of 10 counts on Micromass ESI-Q-TOF (WATERS). Spectra were analyzed using the ProteinLynx Global Server 4.2 (WATERS) and compared with the *T. cacao* CocomGenDB genomic database (<http://cocoagendb.cirad.fr/gbrowse/cgi-bin/gbrowse/theobroma/>).

3D model building

The 3D models of TcCYSPR04 and TcCYS4 were built using a comparative modeling approach. The sequences were subjected to the BLASTP (<http://www.ncbi.nlm.nih.gov/BLAST>) [29] restricted to the Protein Data Bank (PDB, <http://www.pdb.org/>), and used the Blossum62 align matrix [30] for template identification. Alignments showing similarities higher than 44% and 38.42%, respectively, and an expected value lower than 1.10^{-5} were considered significant. The 3D models were constructed using the SwissPdb Viewer v.3.7 [31, 32], following the standard protocol: i) load the template PDB file; ii) align the primary target sequence with template; iii) submit modeling request to the Swiss Model Server [33].

Structure refinement and molecular dynamics simulation

For structure refinement and molecular dynamics (MD) simulation, the LEAP [34] and SANDER [35] utilities of AMBER v.14 have been used [36]. The initial models were neutralized using LEAP, and submitted to SANDER for structure refinement. In addition, the model structures were fully minimized with 1000 steps of steepest descent minimization followed by 1000 more steps of conjugate gradient to an RMSD gradient of 0.01 kcal/2.71Å in vacuum, and then in water, for 1000 steps of steepest descent followed by 1000 more steps of conjugate gradient to an RMSD gradient of 0.01 kcal/2.71Å. After that, MD simulations of the refined structures

were performed in water, using ff12SB force field at 300 K for 5 ns. All MD simulations were carried out without constrain methods. The cutoff value of 10 Å was used for minimization of geometry and MD simulations. All the MD steps were visualized on PyMol 1.7.4.4 and VMD 1.9.2 [37].

Validation of the 3D structure

PROCHECK 3.4 [38] and ANOLEA (Atomic Non-Local Environment Assessment) [39] were used to evaluate both final models by evaluating their Ramachandran plots [40] and the energy plots from each heavy atom in the molecule, respectively [41].

Docking

The docking between 3D models of TcCYSPR04 and TcCYS4 was made using the ClusPro 2.0 web Server (<http://cluspro.bu.edu/>) [42] and the FFT Docking [43] softwares, which are specific for protein-protein docking. Cluspro is the first fully automated web-based program for protein docking and is considered one of the best software by the international scientific community (e.g. CAPRI-Critical Assessment of Predicted Interactions) [44]. After estimating the energy (van der Waals, electrostatic and hydrophobic forces) and the size of the cluster—preferring lower energies and larger cluster sizes—we selected one of the returned models (among 85). The interactions were refined by analyzing the TcCYS4 residues exposition, the orientation of TcCYS4 in relation to TcCYSPR04, and the conformation energy profile. Finally, the best model of the TcCYSPR04-TcCYS4 complex was compared with the PDB file of the resolved structure of cathepsin/stefin complex [45].

Results

Proteases in the genome of cocoa and sequence analysis of TcCYSPR04

In the 28,798 proteins encoded by the cacao genome [22], 448 proteases have been identified. These proteases were distributed in five classes recognized by the KEGG database (Kyoto Encyclopedia of Genes and Genomes) according to the amino acid present in the active site. For class metallo-, aspartic-, serine-, cysteine- and threonine-proteases, 67, 63, 157, 134 and 27 proteins have been identified, respectively.

The TcCYSPR04 fragment was previously identified in a cacao-*M. perniciosus* interaction cDNA library [15]. The complete nucleotide and amino acid sequences identified [22] revealed an ORF of 1068 nucleotides encoding a protein of 355 amino acid residues (Fig 1). A predicted hydrophobic N-terminal region corresponding to the signal peptide in the protein sequence has been detected between the amino acids M1 and A19, and an inhibitory region (pro-region) was found to exist between the amino acids A20 and I137.

The pro-region contained the strongly conserved SNPIR, ERFNIN (EX₃RX₃FX₂NX₃IX₃N) and ANFD (AXNXFXD) motifs. The catalytic region of the protein was comprised between the amino acids L138 and A355. The predicted molecular mass and pI of the different protein forms were: i) 39 kDa and 5.43 for the pre-pro-TcCYSPR04 (355 amino acids); ii) 37 kDa and 5.32 for the pro-TcCYSPR04 (336 amino acids); iii) 23.4 kDa and 4.63 for the TcCYSPR04 (218 amino acids).

Fourteen putative phosphorylation sites were detected: six on serine residues (S55, S69, S184, S238, S271 and S279), five on threonine residues (T21, T90, T108, T289, T296), and three on tyrosine residue (Y67, Y226 and Y333). Seven putative glycosylation sites were detected: five on serine (S41, S55, S271, S290 and S350) and two on threonine (T98 and T278).

```

1  cgggatcctaattattccgaatcaccaaccactactgctttaattctctccacaaaa 60
61  cctgaaacggggagaacagagagagaaccaacaacatagtcATGGCTCGCTTGACGTT 120
      M A R L T L
121  GTTGTCTCGATCGTCTCCTCGTTGCTCCGCCGCGCGCAACAATCTTCGGAGACTC 180
      L S S I V L L V C S A A A A T I F E D S
181  CAACCCATCCGGATGGTCCCGATGACCTCCGTGGCTTGGAGTCTTGTCTCAGTGT 240
      N P I R M V P D D L R G F E S F V L S V
241  CCTGGTCCACAGCGCCACGCTTCTCCTTTGCCCGCTTTGCTTACAGGCATGGGAAGAC 300
      L C H T R H A F S F A R F A Y R H G K T
301  GTACGAGCGGTGGAGGAGATGAAGCTTCGGTTTCAGATTTTAAAGACAATTTGGATTT 360
      Y E S V E E M K L R F Q I F K D N L D L
361  GATCAGATCCACTAACGAAGGGATTGCCTTACACTCTTGCTGTTAATCAGTTTGCCTGA 420
      I R S T N K K G L P Y T L A V N Q F A D
421  TTGGACCTGGGAAGAGTTTCAGAAACACAGTTGGGAGCTGCTCAAACTGCTCTGCTAC 480
      W T W E E F Q K H R L G A A Q N C S A T
481  CGCAAAGGCAATCACAGCTCACTGATGTATCCTTCTGAGTCAAGAGCTGGAGGGA 540
      A K G N H Q L T D V I L P E S K D W R E
541  AGTGGGCATAGTCACTCTGTCAAAAATCAGGGCTCTGTGGATCTTGTGGACATTCAG 600
      V G I V S P V K N Q G S C G S C * W T F S
      -----Peptide1----->
601  TACCACCTGGAGCTCTTGGAGCAGCTTACCATCAGGCATTTGGGAAGGGAATCCCTTTC 660
      T T G A L E A A Y H Q A F G K G I S L S
661  TGAGCAGCAGCTTGTGACTGCGCCGGAGCCTTTAATAACTTTGGATGCAATGGTGGGTT 720
      E Q Q L V D C A G A F N N F G C N G G L
      -----Peptide2----->
721  GCCATCCCAAGCCTTTGAATACATAAATAATGGTGGCCTTGATACAGAGGGGCATA 780
      P S Q A F E Y I K Y N G G L D T E E A Y
781  CCCTTACACAGCAAAAATGGTGAATGCAAAATCTCATCTGAAAATGTTGGGTCCAAGT 840
      P Y T A K N G E C K F S S E N V G V Q V
      -----Peptide3----->
841  CATTGACTCTGTCAACATTACCTTGGGTGCTGAAGATGAATTAAGCATGCAGTTGGCTT 900
      I D S V N I T L G A E D E L K H A V G L
901  GGTTGGCCAGTTAGTGTGATTTGAAGTTGCTACTAGTTTCCGATTTTCAAGAATGG 960
      V R P V S V A F E V V T S F R F Y K N G
      -----Peptide4----->
961  GGTTTACACAGTACGTATGTTGGTAGCACTTCCATGGATGTGAACCACGCCGTTCTTGC 1020
      V Y T S D V C G S T S M D V N H * A V L A
1021  AGTTGGGTATGGAGTTGAAAACGGTGTTCATATTGGCTCATCAAGAACTCATGGGGAGA 1080
      V G Y G V E N G V S Y W L I K N * S W * G E
      -----Peptide5----->
1081  AGATTGGGGGATGATGGCTACTTCAAGATGGAGATGGGGAAAAACATGTGTGGTGTCCG 1140
      D W G D D G Y F K M E M G K N M C G V A
1141  GACATGTGCATCATACCCTATTGTCGCTTGAatctgattgaagaactctatggattaggtc 1200
      T C A S Y P I V A
1201  ggtgctttgtgtagaagaagccgtctagttatggtctcaaacggtgtagttggtccatc 1260
      tcaaaagttgaggttatggcttgatgtacaataaggttttagacgcatactataaaaatg
1321  caaacgggttgtagacacacaaaagcgtatgt.aaagttatctatgaatgatcgtgttg 1380
1381  taatcacttgcaaacattggaatcaattcaatatattaatcttttaaaaaaaaaa 1440
1441  aaaaaaaaaa 1450

```

Fig 1. Nucleotide and amino acid sequences of pre-pro-TcCYS4. Nucleotides of ORF and non-coding regions are shown in uppercase and lowercase letters, respectively. Signal peptide is underlined. White letters on dark gray background indicate the pro-region of the protein. The light gray region represents the catalytic region of the protein. Putative phosphorylated amino acids are indicated in boxes. The double underline shows the putative glycosylated amino acids. ■ indicates the vacuolar targeting motif SNPIR. ▲ and • indicates the ERFNIN and ANFD motifs of the pro-region. Amino acids from the catalytic site are indicated by *. Peptides identified by mass spectrometry after capture by TcCYS4 immobilized on CNBr-sepharose resin are indicated. ♦ indicates the stop codon.

doi:10.1371/journal.pone.0144440.g001

Immunodetection of cystein proteinases in cacao tissues

TcCYS4 was immunodetected in cacao tissues of the susceptible variety (Catongo), control and inoculated with *M. pernicioso* on three stages after inoculation (E1, E2 e E3). A band with an expected size of ~30 kDa corresponding to the mature protein was observed in all stages analyzed for control and inoculated plants (Fig 2a). The intensity of the bands increased from E1 to E3 stages in control plants (healthy). Already for the infected plants, we observed an inverse situation for TcCYS4 accumulation. The band intensity decreased from E1 to E3 stages in infected plants; however, an additional band with ~27 kDa was detected in E3 (Fig 2a, arrow).

Immunodetection performed to protein extracts of plants from experimental field revealed bands in senescent leaves and green broom but not in young leaves, (Fig 2b). Senescent leaves

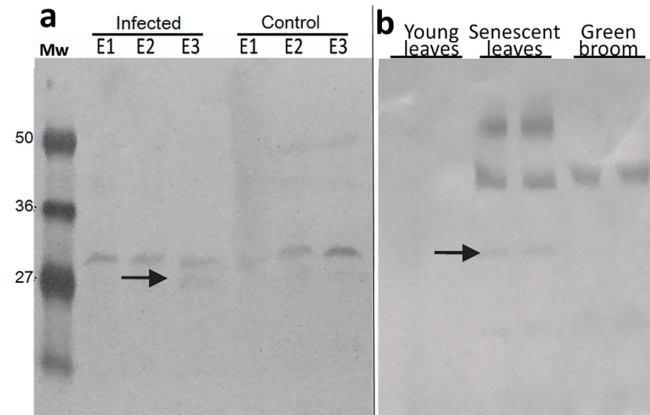


Fig 2. TcCYSPR04 in protein extracts from cacao. **a** Immunodetection of TcCYSPR04 at different plant development stages reported to Catongo cultivated in greenhouse. “Infected” denotes plants whose the leaf primordia were inoculated with basidiospores of the fungus *M. perniciosa*, whereas “control” stands for healthy plants. E1—leaves up to 10 days after inoculation (DAI); E2—leaves 15–25 DAI; E3—leaves with more than 35 DAI (early necrosis). The arrow indicates an additional protein band with molecular mass ~27 kDa, exclusively detected in E3 stage of the infected plants. **b** Immunodetection of TcCYSPR04 in extracts of cacao grown under field conditions. Young leaves—no visible lesions; senescent leaves—leaves with chlorotic appearance; green broom—mature leaves infected with *M. perniciosa* with absence of necrosis, collected in the field. The arrow indicates protein band with molecular mass ~30 kDa detected in senescent leaves from *T. cacao*. Mw—protein molecular weight marker in kDa.

doi:10.1371/journal.pone.0144440.g002

showed the expected band with ~30 kDa, but other bands were observed also in this sample (Fig 2b, arrow). High intensity bands with molecular mass higher than 36 kDa and 50 kDa were detected in senescent leaves, whereas in green broom only the band higher than 36 kDa was observed (Fig 2b).

In the apoplastic fluid, bands of about 50 kDa were detected in both conditions, healthy and infected, analyzed from susceptible (Catongo) and resistant (TSH1188) varieties (Fig 3). The protein bands were more intense in the apoplastic fluid extracted from infected leaves as compared to the healthy leaves. However, TSH1188 apoplastic fluids (infected and healthy) showed bands more intense than Catongo apoplastic fluids. In the infected apoplastic fluid from TSH1188, a band with ~27 kDa was found (Fig 3, arrow).

TcCYSPR04 was immunodetected in tobacco leaves treated with MpNEP. The cross-reaction showed bands with molecular mass between 36 and 50 kDa in all samples analyzed, inclusive in control sample (Fig 4). An exclusive band with the expected size for mature protease between 25–35 kDa was observed in sample extracted at 72 hours after treatment with MpNEP. At 72 hours after treatment with MpNEP, also was detected a band with molecular mass less than 27 kDa (Fig 4).

Protease trap and activity analysis on gel

The capture systems were used to analyze the interaction between proteases from cacao and two recombinants cystatins—TcCYS3 and TcCYS4. Specific proteases in extracts of cacao were captured by cystatins coupled to CNBr-Sepharose. The captured protein showed protease activity (Fig 5). It is interesting to note that the same bands were observed on SDS-PAGE and zymogen gel electrophoresis for the two capture systems (TcCYS3 and TcCYS4), but were not detected bands to BSA-CNBr-Sepharose. Two bands were observed on gelatin/SDS-PAGE, a band less intense with molecular weight ~30 kDa and other more intense with ~45 kDa (Fig 5a). On zymogram gel electrophoresis were observed bands with pH 4.0 to two protease traps (Fig 5b).

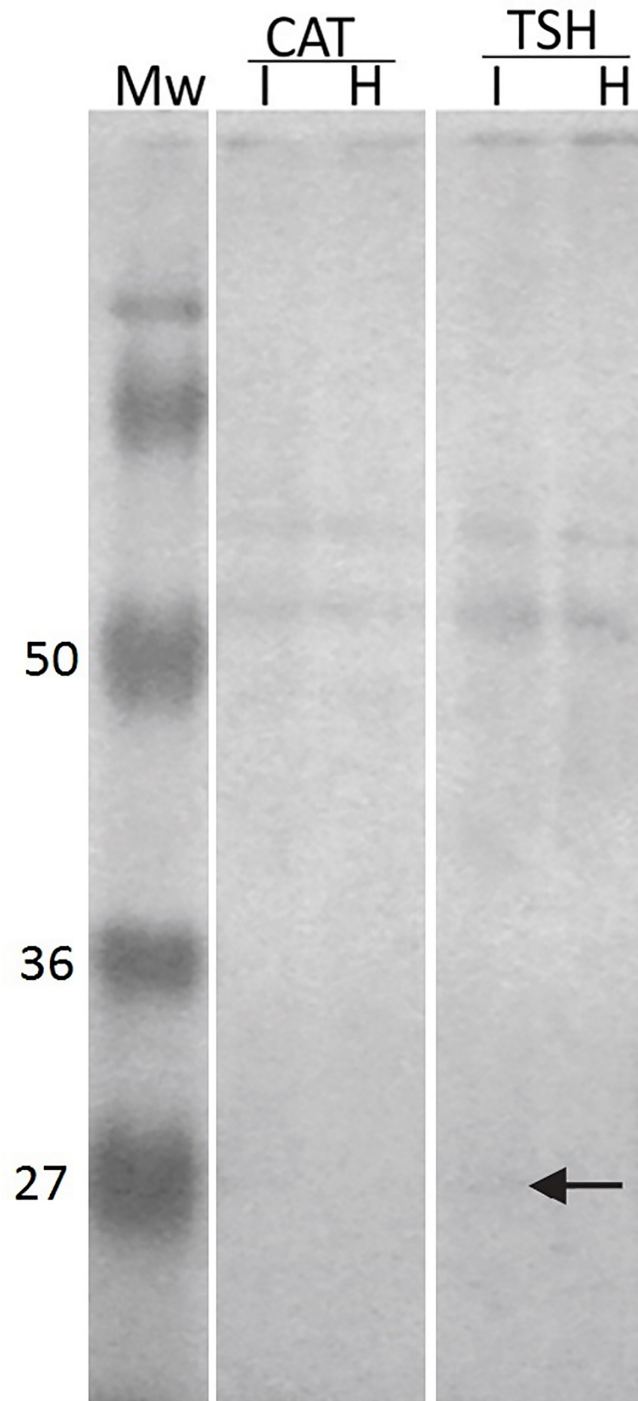


Fig 3. TcCYSPR04 in apoplastic fluid from cacao. Immunodetection of TcCYSPR04 in apoplastic fluid of leaves from susceptible and resistant varieties of cacao. CAT—cacao genotype Catongo susceptible to *M. perniciosa*. TSH—cacao genotype TSH1188 resistant to *M. perniciosa*. I—apoplastic fluid extracted from leaves infected with *M. perniciosa*. H—apoplastic fluid extracted from healthy leaves. Mw—protein molecular weight marker in kDa. The arrow indicates protein band with molecular mass ~27 kDa detected in infected apoplastic fluid from TSH1188.

doi:10.1371/journal.pone.0144440.g003

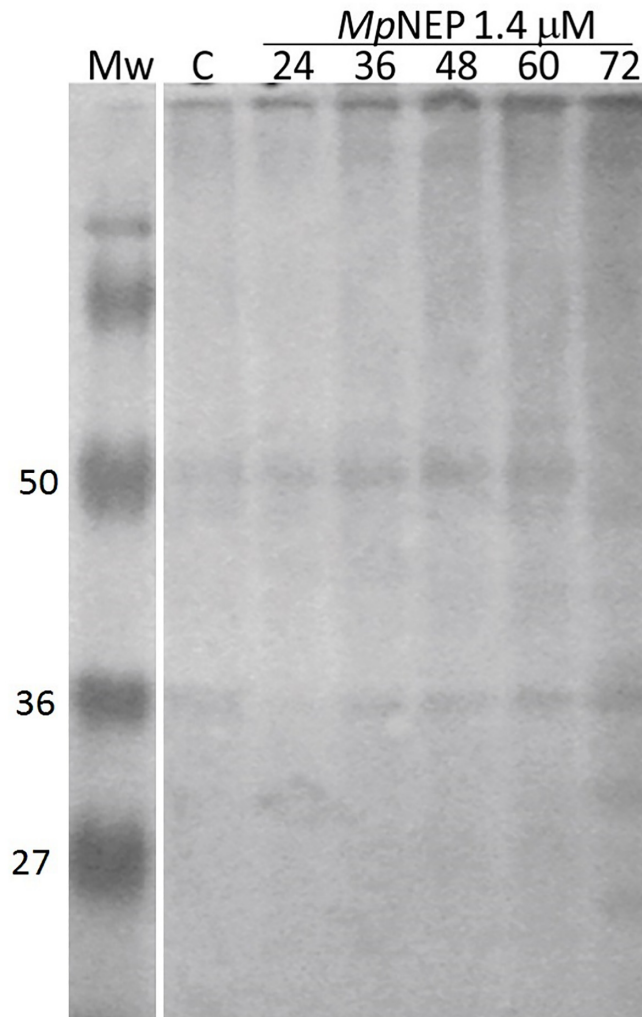


Fig 4. TcCYSPR04 in tobacco leaves treated with MpNEP. Immunodetection of TcCYSPR04 in protein extract from tobacco leaves treated with 1.4 μM of recombinant protein NEP from *M. perniciosa*. C—protein extract from untreated leaves. 24, 36, 40, 60 and 72—hours after spraying the leaves with MpNEP. Mw—protein molecular weight marker in kDa.

doi:10.1371/journal.pone.0144440.g004

Identification of the captured proteases by mass spectrometry

The proteins captured by TcCYS4 were sequenced by mass spectrometry; 5 proteins were encountered, chlorophyll A and B binding protein (3 peptides), glyceraldehyde 3-phosphate dehydrogenase (2 peptides), rubisco (2 peptide), aldolase-like protein (1 peptide), cysteine-protease (5 peptides) (Table 1). After BLAST analysis against public databanks, the last 5 peptides showed 100% identity with TcCYSPR04 (Fig 1), which was previously identified in cacao cDNA libraries [15]. Peptides 1, 2, 3, 4 and 5 had M/Z ratios of 464.2711, 1193.2236, 1332.1821, 724.0779 and 838.3265, respectively (Table 1). The peptides blasted against ESTik [21] and CocoaGenDB [22] identified (100% identity) only one protein encoded by CL83Contig1, which occurs in only one chromosome region (gene: Tc09_g006900 from 3.964.339 to 3.965.747 bp of chromosome 9).

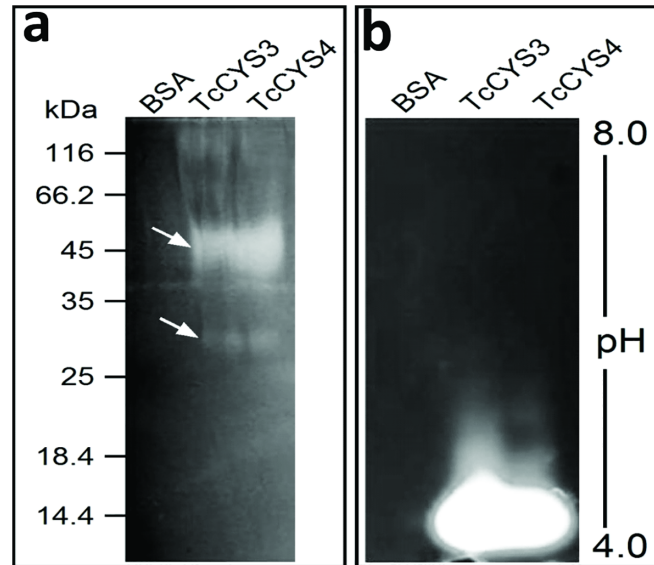


Fig 5. Protease activity of protein captured by TcCYS3 and TcCYS4 coupled to CNBr-Sephacryl. Qualitative assay to test protease activity of samples obtained by capturing system. **a** SDS-PAGE gel showing bands of protease activity of about 30 and 45 kDa (arrows). **b** IEF gel showing clear bands of protease activity at about pH 4.0. BSA—sample captured by BSA-CNBr-Sephacryl (control). TcCYS3—sample captured by recombinant TcCYS3-CNBr-Sephacryl. TcCYS4—sample captured by recombinant TcCYS4-CNBr-Sephacryl.

doi:10.1371/journal.pone.0144440.g005

Building and validation of the 3D model of TcCYSPR04

BLASTP results regarding the cysteine protease of cacao showed 1CS8 (pro-cathepsin L) from *Homo sapiens* as the best template for modeling, with 44% identity. The initial model of

Table 1. Peptides obtained after capture using TcCYS4 and mass spectrometry. MW: molecular weight; M/Z: mass-to-charge ratio.

Protein match	M/Z	Charge	Peak MW (Da)	Peptide MW (Da)	Start*	End*	Sequence
Chlorophyll A and B binding protein	1378.0233	3	4131.0	4130.9	54	92	(R) VLYLGPLSGDPPSYLTGFEFPGDYGWDTAGLSADPETFAR (N)
	528.268	2	1054.5	1054.5	95	102	(R)ELEVIHCR(W)
	492.2465	2	982.4	982.4	124	131	(K)FGEAVWFK(A)
Glyceraldehyde 3-phosphate dehydrogenase	807.4199	2	1612.8	1612.8	84	99	(K)DSPLDVIAINDTGGVK(Q)
	627.8088	2	1253.6	1253.6	312	322	(K)TFAEEVNAAFR(D)
Rubisco	727.8732	2	1453.7	1453.7	58	69	(K)FETLSYLPDLTR(E)
	470.7253	2	939.4	939.4	104	111	(R)IPGYYDGR(C)
	564.7853	2	1127.5	1127.5	170	179	(R)TAAYYQQGAR(F)
Cysteine protease	464.2711	2	926.5	926.5	146	154	(R)EVGIVSPVK(N)
	1193.2236	3	3576.6	3576.6	182	215	(K)GISLSEQQLVDCAGAFNFGCNGGLPSQAFEYIK(Y)
	1332.1821	2	2662.3	2662.3	237	261	(K)FSSENVGVQVIDSVNITLGAEDLK(H)
	724.0779	3	2169.2	2169.2	262	281	(K)HAVGLVRPVSVAFEVVTSEFR(F)
	838.3265	2	1674.6	1674.6	322	335	(K)NSWGEDWGGDGYFK(M)

* In relation to the target protein.

doi:10.1371/journal.pone.0144440.t001

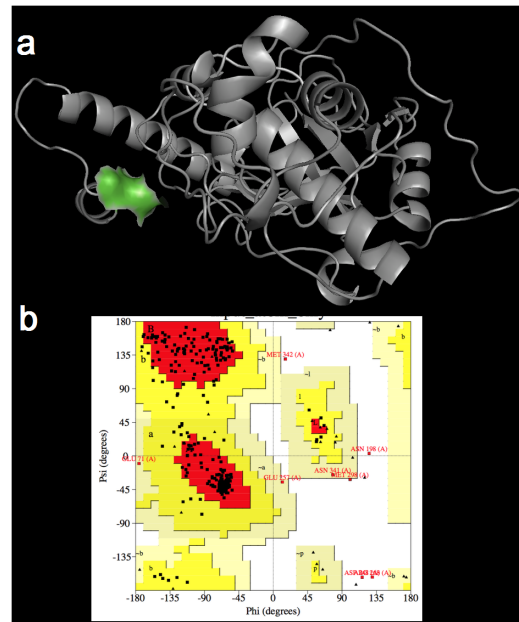


Fig 6. 3D model of the pro-TcCYSPR04. **a** 3D model after MD simulations with detailed active site (in green) containing the C25, H165 and N185 catalytic triad plus the W187. **b** Ramachandran plot of 3D model after MD simulations with all amino acids from active site in allowed positions.

doi:10.1371/journal.pone.0144440.g006

TcCYSPR04 has 354 residues, and presenting all previously described amino acids from catalytic site (Fig 6a). The removal of the inhibitory region has exposed the catalytic cleft of the protein, which contains the catalytic triad formed by the C25, H165 and N185 plus a W187. After refinement and MD, the 3D model of TcCYSPR04 (Fig 6b) showed a Ramachandran plot with 96.9% of residues in allowed regions.

Building and validation of the 3D model of TcCYS4

A clone from cacao-*M. perniciosus* cDNA library has been identified as a putative cystatin named TcCYS4, which was fully described by Pirovani et al [14]. Briefly, the initial model of TcCYS4 showed 196 amino acids with an inhibitory region containing the QVVAG and PWMN motives as well as the conserved phytocystatin (from L23 to K33) and legumain (from S143 to L146) domains (Fig 7a). BLASTP results for the cacao cystatin (TcCYS4) showed one reliable template (4LZI) from *Colocasia esculenta* with 38.42% of identity. After refinement and MD, the 3D model of TcCYS4 showed a Ramachandran plot (Fig 7b) with 95.4% of residues in allowed regions.

TcCYSPR04-TcCYS4 interaction model

The docking between TcCYSPR04 and TcCYS4 has resulted in various possible conformations, whereof only one—the most favorable and probable according to the selection criteria (see [Material and Methods](#))—was chosen. This complex showed energy of -802.7 E/kT. The inhibitory QVVAG domain appears interacting with the catalytic triad of TcCYSPR04 (Fig 8). The interaction was obtained through van der Waals forces and hydrogen bonds.

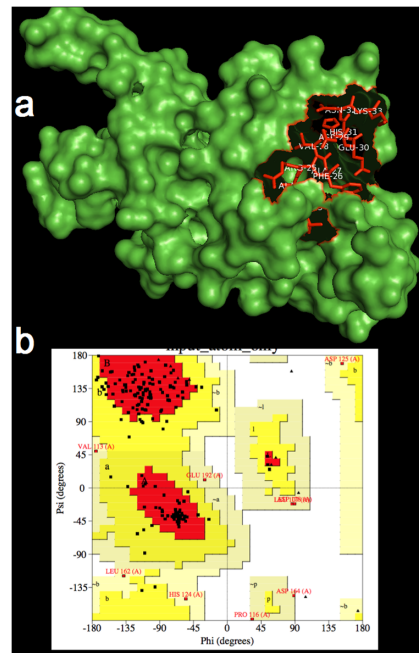


Fig 7. 3D model of TcCYS4. **a** 3D model after MD simulations presenting the inhibitory region (in red). **b** Ramachandran plot of 3D model after MD simulations with amino acids in allowed positions, including inhibitory region.

doi:10.1371/journal.pone.0144440.g007

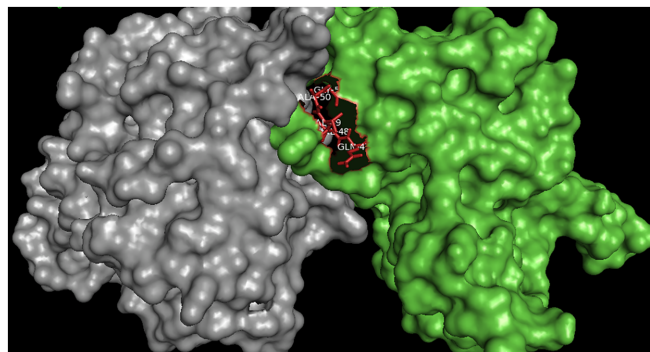


Fig 8. Docking between TcCYS4PR04 and TcCYS4. Complex showing the catalytic and inhibitory (in red) region of TcCYS4PR04 and TcCYS4, respectively.

doi:10.1371/journal.pone.0144440.g008

Discussion

We have identified 448 genes encoding sequences of the proteases of all classes in cacao genome [22]; cysteine-protease corresponds to 30%. The gene TcCysPR04, previously identified in the cDNA library from *T. cacao* and *M. pernicioso* interaction [15], represents one of these sequences. It has an ORF with 1,068 nucleotides in length encoded a polypeptide of 355 amino acids with a theoretical molecular mass of 39 kDa, which is in accordance with the general characteristics of the PLCPs [4].

Sequence analysis has revealed that TcCYS4PR04 is a pre-pro-protein, i.e., containing a peptide signal, an inhibitory pro-region and the mature enzyme (Figs 1 and 6). The presence of peptide signal and the low pI value found (Fig 5) indicates that TcCYS4PR04 is a secreted

protein, as observed for many plant PLCPs [4]. Fourteen and seven putative phosphorylation and glycosylation sites, respectively, were encountered along the pro-TcCYSPR04 (Fig 1). The presence of such post-transcriptional modifications and formation of protein-protein complex may explain the differences observed between theoretical and experimental protein size observed band on SDS-PAGE gel (Figs 2–5).

The 3D model has demonstrated that the putative sites to post-transcriptional modifications were exposed on the protein surface. Post-transcriptional modifications of the pro-protein may also be involved in the pro-region cleavage and mature protein activation, as well as in the interaction of the active TcCYSPR04 with its inhibitor, TcCYS4 [45, 46]. The identification of immunoreactive bands in senescent leaves with expected size for the mature form between 27 and 36 kDa (Fig 2b) highlights the involvement of this protease in the degradation and mobilization of cellular proteins characteristic of this plant development phase [46, 47]. Senescence is controlled by gene regulation and a series of signals that start in the final phase of leaf development [48]. The regulation of senescence is explained by complex external and internal stimuli, for example, signal transduction to suppress genes responsible for photosynthesis, and activation of genes involved in the disassembly of cell structures [49].

The cysteine proteases can be synthesized as monomers and oligomerized to form dimers, trimers and tetramers during senescence, or may be complexed with cystatin [50]. The presence of bands with molecular weight higher than 50 kDa in the leaf extract of *T. cacao* suggests that cysteine-proteases can form protein-protein complex in addition to post-translational modification (PTM), which could increase its molecular weight. Thus, the bands with molecular weight higher than 50 kDa corroborate the predictions of glycosylation shown in Fig 1. A similar result was obtained for the electrophoretic migration of cysteine-proteases in tissues of senescent spinach, where the author suggests that the increased molecular weight of the enzyme is due to PTM [46]. This process tends to lower the protein pI and it is usually involved in activation and deactivation of enzymes. So, this inhibitor should be released in order that the enzymes undergo the PTM and/or acidification and become active by triggering senescence or resistance and program cell death (PCD) responses [14, 27].

Plants of Catongo cacao variety showed a gradual increase of the enzyme (a band of ~ 30 kDa) during leaf ontogeny in control plants. The increase of this specific band was not observed in the infected tissue, but an additional band of ~ 27 kDa appears in stage E3 (mature leaves with more than 35 days after infection) (Fig 2a). Plants may have a localized response to the infection site that prevents either the colonization of the pathogen or even a response to the mismatch, in the case of biotrophic pathogens [51]. In the interaction between *T. cacao* and *M. pernicioso*, there is evidence that the biological process of PCD could be controlled by the pathogen during infection [16]. Thus, the fungus takes advantage of the situation and suppresses the cell death [52]. The appearance of a ~ 27 kDa band in the advanced stage of infection (Fig 2a) may be related to the fungal shift from biotrophic to saprophytic, when occur the plant cell death [53].

Bands with molecular weight higher than 50 kDa were immunodetected in leaf apoplastic fluid of Catongo and TSH1188, but a band with expected molecular weight (~27 kDa) was detected only in infected sample from TSH118 (Fig 3). Such observations indicate that TcCYSPR04 is a vacuolar or apoplastic protein, as was predicted in the TargetP analyses (Fig 1). Besides this, the molecular weight of the protein detected in apoplastic fluid indicates that the protein can pass by PTM after its biosynthesis. However, the exclusive band observed in infected TSH (Fig 3) can be result of the accumulation protein unmodified, a differential response expressed by resistant genotype [19, 20]. The modifications increases the molecular weight and forming mature protease complexes (Fig 5a), as formerly proposed [50]. Furthermore, the captured protease showed activity in pI value near 4 (Fig 5b). The pI value for mature

form of the protein was close to pH of the apoplast and vacuole. The apoplast is an important molecular barrier, cystatin (CC9) from maize when is transcriptionally induced is directed to apoplast where it inhibits apoplastic papain-like cysteine proteases [5].

Cysteine proteases have also responded to the treatment with MpNEP in tobacco leaves (Fig 4). MpNEP is a protein identified in the genome of *M. perniciosa*, in the library of the interaction between *T. cacao* and *M. perniciosa* [15]. By inducing the production of ethylene in the plant, this fungal protein induces the premature senescence process and triggers all cell death mechanisms in the plant [18, 52]. There is a series of bands in the gel immunodetected by anti-serum against TcCYSPR04, suggesting, mainly 72 hours after the interaction MpNEP-plant, the participation of various isoforms of cysteine-proteases in physiological events probable related to necrosis events triggered by MpNEP. Overexpression of TcCYS4 in tobacco affected the response of plants inoculated with MpNEP2 through the action of cystatins, which inhibit cysteine proteases activated during the PCD process triggered by MpNEP2 [52]. Some of these protein-protein complexes may have been identified in our immunoblottings with anti-TcCYSPR04 and this explaining the different molecular weights detected in our samples.

The molecular battle that occurs during the development of witches' broom symptoms begins to be clarified. Our data, together with the results of Pirovani et al [14], which analyzed the accumulation of cystatin in healthy and infected tissues, have indicated that balance between cysteine-protease and cystatin in *T. cacao* might be a determining factor in the development of witches' broom symptoms as suggested by van der Hoorn [3] for different pathosystems. Recently, it was shown that variations in pH and temperature can affect the oligomerization of TcCYS4 modifying the inhibitory activity this cystatin against PLPCs engaged in PCD, and affect symptoms of witches' broom disease of cocoa, caused by the fungus *M. perniciosa* [54]. Cystatin levels also decrease in mature tissues, while the main difference concerns the protein levels between infected and uninfected tissues during the transition from green broom to dry broom, which corresponds to the transition from biotrophic to necrotrophic mycelium of *M. perniciosa* [16].

We have also reported the capture of an active cysteine-protease using two recombinant cystatins (TcCYS3 and TcCYS4), both from cacao (Fig 5). It is noteworthy that this system allowed capture of the active enzyme after their interaction (Fig 5), unlike what was published to another capture system [55]. The captured proteins were sequenced and the peptides obtained showed 100% of identity to TcCYSPR04 with 48% coverage of its sequence (105 out of 218 amino acids; Fig 1). The alignment of the peptides in databanks has allowed the identification of only one cDNA and one region of the genome. Considering 134 cysteine-protease in the cacao genome, the capture of only one protease has indicated that it is probably the most abundant enzyme found in leaves or the one with the highest affinity to TcCYS4 (or both). *TcCYSPR04* was located on the chromosome 9. Interestingly, on this chromosome are located the major QTL (quantitative trait loci) of witches' broom [56, 57] and most of the defense genes [22].

The 3D models of pro-TcCYSPR04 and TcCYSPR04 obtained by comparison between the 1CS8 and 8PCH templates, respectively, revealed that the pro-region covered the catalytic cleft of the catalytic region (Fig 6). The pro-region contained the consensus sequence SNPIR, which is known to be involved in the protein targeting to the vacuole, as observed for aleurain [58]. The pro-region also contained the consensus sequence ERFNIN present in the second α -helix and known to function as an inhibitory region [59]. These observations reinforce the assumption that the pro-region acts as an inhibitor of the catalytic region, avoiding an unduly activation of the mature protein during its transport to the apoplast [60]. Another highly conserved motif, ANFD (AXNFXD), was encountered in one of the loops of the pro-region, between

the second and the third α -helices (Fig 1). In other cysteine-proteases, this motif, which may appear as GNFD, was essential for the correct processing of the protease precursor [61].

The replacement of Asp by Asn, Tyr, Met, Val or Glu resulted in non-functional papain [62]. This indicated the essential role of Asp in protease processing. Likewise, the mutation in the conserved GNFD motif of cathepsin L1 of *Fasciola hepatica* has reduced the folding function of the pro-enzyme [63]. Here, we verified that the pro-region cleavage has influenced the mature region conformation; the amino acids W163, EEA223-225, R281, E312, V315 and V354 were involved in different secondary structures before and after TcCYSPR04 pro-region cleavage (e.g. EEA223-225 was involved in a α -helix in the pro-TcCYSPR04 and in a loop in the TcCYSPR04), suggesting that the pro-region was effectively involved in the mature protein (Fig 6) folding [64–66]. The cleavage site of the pro-region of pro-TcCYSPR04 was predicted by Pfam analysis, by aligning and overlapping the 3D structures obtained from 8PCH and 1CS8 (Fig 6a) and by multiple alignment with other cysteine-proteases from plant, animal and microorganisms. The cleavage site corresponded to the LP138-139 (Figs 1 and 6a). The proteolytic activity of all cysteine-proteases arises from the presence of the catalytic Cys and His residues in the mature enzyme's active center, in TcCYSPR04, C25 and H165 were encountered in the catalytic site (Fig 6a). The catalytic site also contained N185, which, along with C25 and H165 formed the catalytic triad encountered in other cysteine-proteases [3], and W187 (Fig 6a). The latter seemed to participate in the interaction between TcCYSPR04 and TcCYS4 (Fig 8).

The TcCYS4 model obtained by comparison with the 3IMA template showed one α -helix and 4 β -strands forming an antiparallel β -sheet (Fig 6). The conserved QVVAG and PW regions are known to constitute the inhibitory site of the cystatin [14,38]; also, these regions are located on the third and fifth loops of the β -sheet (Fig 6) and were observed to interact with the catalytic site (C, H, N triad plus W) of the cysteine-protease (Fig 7). Likewise, the docking of TcCYSPR04 and TcCYS4 has showed electrostatic complementarities of the surface contact (Fig 8). The same was observed for papain and cystatin S from saliva [9].

Both 3D models of TcCYSPR04 and TcCYS4 were validated using the PROCHECK and ANOLEA programs. The Ramachandran plots obtained for TcCYSPR04 and TcCYS4 before refinement showed that 91.8 and 81.9% of residues in energetically most favorable regions, respectively. After refinement, the percentage of residues in the energetically most favorable regions has decreased (81% and 75% for TcCYSPR04 and TcCYS4, respectively), but still remained high. In compensation, after refinement, the validation using ANOLEA showed a better score, indicating that the 3D models obtained were very similar to the structure of the biological active molecule [67–70]. The accuracy of the results revealed that the models had appropriate stereochemical and thermodynamic values.

Conclusions

Physical interaction between the cysteine-protease (TcCYSPR04) and the cystatin (TcCYS4) from cacao was demonstrated *in vitro* by capture and mass spectrometry (for TcCYSPR04 identification). The balance between cystatin and cysteine protease is part of the cellular responses and both are related to the plant defense response to attack of the *M. pernicioso*. Furthermore, cysteine-protease may be involved in the senescence process and isoforms of TcCYSPR04 immunodetection in tissues during the phase change of the disease suggest that this protein may be associated with the PCD process that occurs in this phase. Our results can corroborate with development of the biotechnological strategy aiming disease control, as well as to improve the understanding mechanisms of the interaction. The homology modeling was obtained for both proteins, where molecular docking showed that the physical-chemical parameters estimated favors the interaction between the cacao enzyme and its inhibitor. The

models developed in this study may be used for improving cacao resistance by de novo design methods.

Acknowledgments

The work of THS Cardoso and DM Koop was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) and the FAPESB, respectively. We are thankful to Xavier Sabau (Cirad, France) for granting TcCYSPR04 cDNA clones, Claudia Fortes Ferreira (Embrapa, Cruz das Almas-BA, Brazil) for the critical reading of the manuscript, and Juliano Oliveira Santana for figure edition.

Author Contributions

Conceived and designed the experiments: THSC ACOF BSA CPP FM. Performed the experiments: THSC ACOF ASS DMK KPG FCA. Analyzed the data: THSC ASS BSA CPP. Contributed reagents/materials/analysis tools: KPG FM CPP. Wrote the paper: THSC FM CPP AOS.

References

1. Rawlings ND. Peptidase inhibitors in the MEROPS database. *Biochimie*. 2010; 92: 1463–1483. doi: [10.1016/j.biochi.2010.04.013](https://doi.org/10.1016/j.biochi.2010.04.013) PMID: [20430064](https://pubmed.ncbi.nlm.nih.gov/20430064/)
2. Salas CE, Gomes MTR, Hernandez M, Lopes MTP. Plant cysteine proteinases: Evaluation of the pharmacological activity. *Phytochemistry*. 2008; 69: 2263–2269. doi: [10.1016/j.phytochem.2008.05.016](https://doi.org/10.1016/j.phytochem.2008.05.016) PMID: [18614189](https://pubmed.ncbi.nlm.nih.gov/18614189/)
3. Van der Hoorn RAL. Plant Proteases: From Phenotypes to Molecular Mechanisms. *Annu Rev Plant Biol*. 2008; 59: 191–223. doi: [10.1146/annurev.arplant.59.032607.092835](https://doi.org/10.1146/annurev.arplant.59.032607.092835) PMID: [18257708](https://pubmed.ncbi.nlm.nih.gov/18257708/)
4. Shindo T, Van Der Hoorn RAL. Papain-like cysteine proteases: key players at molecular battlefields employed by both plants and their invaders. *Mol Plant Pathol*. 2008; 9: 119–125. doi: [10.1111/j.1364-3703.2007.00439.x](https://doi.org/10.1111/j.1364-3703.2007.00439.x) PMID: [18705889](https://pubmed.ncbi.nlm.nih.gov/18705889/)
5. Van der Linde K, Mueller AN, Hemetsberger C, Kashani F, van der Hoorn RAL, Doehlemann G. The maize cystatin CC9 interacts with apoplastic cysteine proteases. *Plant Signaling & Behavior*. 2012; 7(11):1397–1401. doi: [10.4161/psb.21902](https://doi.org/10.4161/psb.21902)
6. Cambra I, Martinez M, Dáder B, González-Melendi P, Gandullo J, Santamaría ME, et al. A cathepsin F-like peptidase involved in barley grain protein mobilization, HvPap-1, is modulated by its own propeptide and by cystatins. *Journal of Experimental Botany*. 2012; 63: 4615–4629. doi: [10.1093/jxb/ers137](https://doi.org/10.1093/jxb/ers137) PMID: [22791822](https://pubmed.ncbi.nlm.nih.gov/22791822/)
7. Van der Hoorn RAL, Jones JDG. The plant proteolytic machinery and its role in defence. *Current Opinion in Plant Biology*. 2004; 7: 400–407. PMID: [15231262](https://pubmed.ncbi.nlm.nih.gov/15231262/)
8. Chu MH, Liu KL, Wu HY, Yeh KW, Cheng YS. Crystal structure of tarocystatin-papain complex: implications for the inhibition property of group-2 phytocystatins. *Planta*. 2011; 234: 243–254. doi: [10.1007/s00425-011-1398-8](https://doi.org/10.1007/s00425-011-1398-8) PMID: [21416241](https://pubmed.ncbi.nlm.nih.gov/21416241/)
9. Nandy SK, Bhuyan R, Seal A. Modelling family 2 cystatins and their interaction with papain. *J Biomol Struct Dyn*. 2013; 31(6): 649–664 doi: [10.1080/07391102.2012.706403](https://doi.org/10.1080/07391102.2012.706403) PMID: [22881286](https://pubmed.ncbi.nlm.nih.gov/22881286/)
10. Kordis D, Turk V. Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes. *BMC Evolutionary Biology*. 2009; 9: 266. doi: [10.1186/1471-2148-9-266](https://doi.org/10.1186/1471-2148-9-266) PMID: [19919722](https://pubmed.ncbi.nlm.nih.gov/19919722/)
11. Grzonka Z, Jankowska E, Kasprzykowski F, Kasprzykowska R, Lankiewicz L, Wiczek W, et al. Structural studies of cysteine proteases and their inhibitors. *Acta Biochim Pol*. 2001; 48: 1–20. PMID: [11440158](https://pubmed.ncbi.nlm.nih.gov/11440158/)
12. Arai S, Matsumoto I, Emori Y, Abe K. Plant seed cystatins and their target enzymes of endogenous and exogenous origin. *Journal of Agricultural and Food Chemistry*. 2002; 50: 6612–6617. PMID: [12381160](https://pubmed.ncbi.nlm.nih.gov/12381160/)
13. Chan YL, Yang AH, Chen JT, Yeh KW, Chan MT. Heterologous expression of taro cystatin protects transgenic tomato against *Meloidogyne incognita*; infection by means of interfering sex determination and suppressing gall formation. *Plant Cell Reports*. 2010; 29: 231–238. doi: [10.1007/s00299-009-0815-y](https://doi.org/10.1007/s00299-009-0815-y) PMID: [20054551](https://pubmed.ncbi.nlm.nih.gov/20054551/)
14. Pirovani CP, Santiago AS, Santos LS, Micheli F, Margis R, Gesteria AS, et al. *Theobroma cacao* cystatins impair *Moniliophthora perniciosa* mycelial growth and are involved in postponing cell death symptoms. *Planta*. 2010; 232: 1485–1497. doi: [10.1007/s00425-010-1272-0](https://doi.org/10.1007/s00425-010-1272-0) PMID: [20859638](https://pubmed.ncbi.nlm.nih.gov/20859638/)

15. Gesteira AS, Micheli F, Carels N, Da Silva AC, Gramacho K, Schuster I, et al. Comparative analysis of expressed genes from cacao meristems infected by *Moniliophthora perniciosa*. *Ann Bot*. 2007; 100: 129–140. PMID: [17557832](#)
16. Ceita GO, Macêdo JNA, Santos TB, Alemanno L, Gesteira AS, Micheli F, et al. Involvement of calcium oxalate degradation during programmed cell death in *Theobroma cacao* tissues triggered by the hemibiotrophic fungus *Moniliophthora perniciosa*. *Plant Science*. 2007; 173:106–117
17. Garcia O, Macedo JAN, Tibúrcio R, Zapparoli G, Rincones J, Bittencourt LM, et al. Characterization of necrosis and ethylene-inducing proteins (NEP) in the basidiomycete *Moniliophthora perniciosa*, the causal agent of witches' broom in *Theobroma cacao*. *Mycological Research*. 2007; 111: 443–455. PMID: [17512713](#)
18. Santana LS, Costa MG, Pirovani NM, Almeida AF, Alvim FC, Pirovani CP. TcCYS4, a cystatin from cocoa, reduces necrosis triggered by MpNEP2 in tobacco plants. *Genetic and Molecular Research*. 2014; 13(3):7636–48.
19. Surujdeo-Maharaj S, Umaharan P, Butler DR, Sreenivasan TN. An optimized screening method for identifying levels of resistance to *Crinipellis perniciosa* in cocoa (*Theobroma cacao*). *Plant Pathology*. 2003; 52: 464–475.
20. Pirovani CP, Carvalho HA, Machado RC, Gomes DS, Alvim FC, Pomella AW, et al. Protein extraction for proteome analysis from cacao leaves and meristems, organs infected by *Moniliophthora perniciosa*, the causal agent of the witches' broom disease. *Electrophoresis*. 2008; 29: 2391–2401. doi: [10.1002/elps.200700743](#) PMID: [18435495](#)
21. Argout X, Fouet O, Wincker P, Gramacho K, Legavre T, Sabau X, et al. Towards the understanding of the cocoa transcriptome: Production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* L. generated from various tissues and under various conditions. *BMC Genomics*. 2008; 9: 512. doi: [10.1186/1471-2164-9-512](#) PMID: [18973681](#)
22. Argout X, Salse J, Aury J-M, Guitinan MJ, Droc G, Gouzy J, et al. The genome of *Theobroma cacao*. *Nat Genet*. 2011; 43: 101–108. doi: [10.1038/ng.736](#) PMID: [21186351](#)
23. Barrett AJ, Rawlings ND, O'Brien EA. The MEROPS database as a protease information system. *J Struct Biol*. 2001; 134: 95–102. PMID: [11551172](#)
24. Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*. 2 st ed. New York: Cold Spring Harbor Laboratory Press; 1989.
25. Bradford MM. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*. 1976; 72: 248–254. PMID: [942051](#)
26. Pirovani CP, Macedo JN, Contim LA, Matrangolo FS, Loureiro ME, Fontes EP. A sucrose-binding protein homologue from soybean exhibits GTP-binding activity that functions independently of sucrose transport activity. *Eur J Biochem*. 2002; 269: 3998–4008. PMID: [12180976](#)
27. Michaud D. Gel electrophoresis of proteolytic enzymes. *Analytica Chimica Acta*. 1998; 372: 173–185.
28. Neuhoff V, Arold N, Taube D, Ehrhardt W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis*. 1988; 9: 255–262. PMID: [2466658](#)
29. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25: 3389–3402. PMID: [9254694](#)
30. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci*. 1992; 89: 10915–10919. PMID: [1438297](#)
31. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997; 18: 2714–2723. PMID: [9504803](#)
32. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*. 2009; 30(1): 162–173.
33. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*. 2003; 31: 3381–3385. PMID: [12824332](#)
34. Schafmeister CEAF, Ross WS, Romanovski V, LeaP. San Francisco: University of California; 1995.
35. Orlikowska M, Jankowska E, Kolodziejczyk R, Jaskólski M, Szymanska A. Hinge-loop mutation can be used to control 3D domain swapping and amyloidogenesis of human cystatin C. *Journal of Structural Biology*. 2011; 173: 406–413. doi: [10.1016/j.jsb.2010.11.009](#) PMID: [21074623](#)
36. Weiner SJ, Kollman PA, Nguyen DT, Case DA. An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry*. 1986; 7: 230–252.
37. Humphrey W, Dalke A, Schulten K. VMD—Visual Molecular Dynamics. *J Molec Graph*. 1996; 33–38.

38. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*. 1993; 26: 283–291.
39. Melo F, Feytmans E. Assessing protein structures with a non-local atomic interaction energy. *Journal of Molecular Biology*. 1998; 277: 1141–1152. PMID: [9571028](#)
40. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol*. 1963; 7: 95–99. PMID: [13990617](#)
41. Melo F. *ANOLEA*: a www server to assess protein structures. *Intell Sys Mol Biol*. 1997; 97: 110–113.
42. Kozakov D, Hall DR, Beglov D, Brenke R, Comeau SR, Shen Y, et al. Achieving reliability and high accuracy in automated protein docking: Cluspro, PIPER, SDU, and stability analysis in CAPRI rounds 13–19. *Proteins: Structure, Function, and Bioinformatics*. 2010; 78: 3124–3130.
43. Ritchie DW, Venkatraman V. Ultra-fast FFT protein docking on graphics processors. *Bioinformatics*. 2010; 26: 2398–2405. doi: [10.1093/bioinformatics/btq444](#) PMID: [20685958](#)
44. Vajda S, Kozakov D. Convergence and combination of methods in protein-protein docking. *Current Opinion in Structural Biology*. 2009; 19: 164–170. doi: [10.1016/j.sbi.2009.02.008](#) PMID: [19327983](#)
45. Jenko S, Dolenc I, Guncar G, Dobersek A, Podobnik M, Turk D. 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. *Journal of Molecular Biology*. 2003; 326: 875–885. PMID: [12581647](#)
46. Van Doorn WG, Balk PA, van Houwelingen AM, Hoeberichts FA, Hall RD, Vorst O, et al. Gene expression during anthesis and senescence in Iris flowers. *Plant Molecular Biology*. 2003; 53: 845–863. PMID: [15082930](#)
47. Tajima T, Yamaguchi A, Matsushima S, Satoh M, Hayasaka S, Yoshimatsu K, et al. Biochemical and molecular characterization of senescence-related cysteine protease–cystatin complex from spinach leaf. *Physiologia Plantarum*. 2011; 141: 97–116. doi: [10.1111/j.1399-3054.2010.01425.x](#) PMID: [21044083](#)
48. Noodén LD, Guiamét JJ, John I. Senescence mechanisms. *Physiologia Plantarum*. 1997; 101: 746–753.
49. Hayden DM, Christopher DA. Characterization of senescence-associated gene expression and senescence-dependent and -independent cysteine proteases differing in microsomal processing in *Anthurium*. *Plant Science*. 2004; 166: 779–790.
50. Kawai-Yamada M. Mammalian Bax-induced plant cell death can be down-regulated by overexpression of Arabidopsis Bax Inhibitor-1 (AtBI-1). *Proceedings of the National Academy of Sciences*. 2001; 98: 12295–12300.
51. Staskawicz B, Ausubel F, Baker B, Ellis J, Jones J. Molecular genetics of plant disease resistance. *Science*. 1995; 268: 661–667. PMID: [7732374](#)
52. Greenberg JT, Yao N. The role and regulation of programmed cell death in plant-pathogen interactions. *Cellular Microbiology*. 2004; 6: 201–211. PMID: [14764104](#)
53. Adebbrhan T, Figueira A, Yamada MM, Cascardo J, Furtek DB. Molecular fingerprinting suggests two primary outbreaks of witches' broom disease (*Crinipellis perniciososa*) of *Theobroma cacao* in Bahia, Brazil. *European Journal of Plant Pathology*. 1999; 105: 167–175.
54. Freitas AC, Souza CF, Monzani PS, Garcia W, de Almeida AA, Costa MG, et al. The activity of TcCYS4 modified by variations in pH and temperature can affect symptoms of witches' broom disease of cocoa, caused by the fungus *Moniliophthora perniciososa*. 2015; 10(4):e0121519. doi: [10.1371/journal.pone.0121519](#)
55. van der Hoorn RA, Leeuwenburgh MA, Bogyo M, Joosten MH, Peck SC. Activity profiling of papain-Like cysteine proteases in plants. *Plant Physiology*. 2004; 135: 1170–1178. PMID: [15266051](#)
56. Faleiro FG, Queiroz VT, Lopes UV, Guimarães CT, Pires JL, Yamada MM, et al. Mapping QTLs for witches' broom (*Crinipellis perniciososa*) resistance in cacao (*Theobroma cacao* L.). *Euphytica*. 2006; 149: 227–235.
57. Brown J, Schnell R, Motamayor J, Lopes U, Kuhn D, Borrone JW. Resistance gene mapping for witches' broom disease in *Theobroma cacao* L. in an F2 population using SSR markers and candidate genes. *J Amer Soc Hort Sci*. 2005; 130: 366–373.
58. Eason JR, Ryan DJ, Watson LM, Hedderley D, Christey MC, Coupe SA. Suppression of the cysteine protease, aleurain, delays floret senescence in *Brassica oleracea*. *Plant Molecular Biology*. 2005; 57: 645–657. PMID: [15988561](#)
59. Grudkowska M, Zagdańska B. Multifunctional role of plant cysteine proteinases. *Acta Biochimica Polonica*. 2004; 51: 609–624. PMID: [15448724](#)
60. Pietra D, Borghini A, Ricci C, Bianucci AM. Enzyme Kinetics Studies on 29-kDa Human Liver Cathepsin L. *Chemical Biology & Drug Design*. 2014; 84(6): 648–658.

61. Wiederanders B. Structure-function relationships in class CA1 cysteine peptidase propeptides. *Acta Biochim Pol.* 2003; 50: 691–713. PMID: [14515150](#)
62. Vernet T, Berti P, de Montigny C, Musil R, Tessier D, Ménard R, et al. Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. *J Biol Chem.* 1995; 270: 10838–10846. PMID: [7738022](#)
63. Cappetta M, Roth I, Díaz A, Tort J, Roche L. Role of the prosegment of *Fasciola hepatica* cathepsin L1 in folding of the catalytic domain. *Biol Chem.* 2002; 383: 1215–1221. PMID: [12437108](#)
64. Mans BJ, Louw AI, Neitz AW. Amino acid sequence and structure modeling of savignin, a thrombin inhibitor from the tick, *Ornithodoros savignyi*. *Insect Biochem Mol Biol.* 2002b; 32: 821–828.
65. Cygler M, Sivaraman J, Grochulski P, Coulombe R, Storer AC, Mort JS. Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. *Structure.* 1996; 4: 405–416. PMID: [8740363](#)
66. McIntyre GF, Erickson AH. The lysosomal proenzyme receptor that binds procathepsin L to microsomal membranes at pH 5 is a 43-kDa integral membrane protein. *Proc Natl Acad Sci.* 1993; 90: 10588–10592. PMID: [8248149](#)
67. Pearlman DA. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp Phys Commun.* 1995; 19: 1–41.
68. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang Ji, Duke RE. AMBER 8. San Francisco: University of California; 2004.
69. Hayik SA, Dunbrack R, Merz KM. A mixed QM/MM scoring function to predict protein-ligand binding affinity. *J Chem Theory Comput.* 2010; 6: 3079–3091. PMID: [21221417](#)
70. Kiss G, Rothlisberger D, Baker D, Houk KN. Evaluation and ranking of enzyme designs. *Protein Sci.* 2010; 19: 1760–1773. doi: [10.1002/pro.462](#) PMID: [20665693](#)