CrossMark

ORIGINAL RESEARCH ARTICLE

In Silico Analysis for Five Major Cereal Crops Phytocystatins

Rupal Chauhan[1] · Yogesh Jasrai[1] · Himanshu Pandya[1]

Abstract Five major cereal crops such as rice, wheat, maize, barley and sorghum are continuously threatened by a multitude of pathogens and other disorders. Cystatins offers a pivotal role in deciding the promising plant response. The use of bioinformatics tools for phylogenetic relationships of five major cereal crop (rice, wheat, maize, barley and sorghum) phytocystatins based on amino acid sequence information was elucidated, and their secondary and tertiary structures were investigated for structural comparisons. Twenty-eight distinct phytocystatins from 28 plant species were investigated. Phytocystatins could be divided into five distinct phylogenetic groups. Five major cereal crops their structural features were highly conserved, and their amino acid sequence similarities ranged from 48 to 86 %. A new highly conserved amino acid sequence motif, YEAKxWxKxF, in the C-terminal end being unique to phytocystatins was identified. The predicted 3D homology models showed a high conservation of the general central structure of the phytocystatins, i.e., the 4–5 anti-parallel β-sheets, wrapping halfway round a single central α-helix and particularly the three active site regions, the N-terminal, the first and second hairpin loops. Any structural differences seem to be mainly in the length of the N- and C-terminal, the length of the second hairpin loop and the fifth β-sheet. Via docking experiments, small heterogeneities were observed in the vicinity of the OC-I active sites that seemed to be influential in the binding process and stability of the resultant inhibitor–protease complex.

Keywords Five major cereal crop phytocystatin · Phylogenetic relationship · Structural comparisons

1 Introduction

Phytocystatins are proteinacious inhibitors of plant origin that inhibit specifically cysteine proteases by forming tight reversible bonds thus preventing the hydrolysis of proteins by protease [1]. The cystatin super family is subdivided into three families based mainly on the three criteria sequence homology, presence of disulfide bonds and the molecular mass of the protein. These families are the stefins, cystatins and kininogens [2]. Many different phytocystatins have been isolated from different plants and their gene sequences available on public databases.

Phylogenetic analysis provides an insight into the molecular evolution of proteins. Numerous bioinformatics and computational tools provide automated analysis of relationships of proteins at molecular and structural level. Public sequence databases have also provided a very useful and wide range of resources to perform such analyses. One of the key ideas in genomic bioinformatics is the concept of homology. This is used to predict the function of genes and proteins. This is followed by a next level where not only protein function can be predicted but also the primary, secondary and tertiary structures of a protein also can be predicted. This is achieved through powerful computation methods referred to as *in silico* analysis. Such analysis provides a better understanding of the microstructures on the protein surface that contribute or may even hinder its proper function.

✉ Rupal Chauhan
chauhanrupal757@gmail.com

[1] Applied Botany Center, Department of Botany, University School of Sciences, Gujarat University, Ahmadabad, Gujarat 380 009, India

Phylogenetic relationship of phytocystatins based on available amino acid sequence information was one of the primary objectives. This was carried out by a comparative study on the primary, predicted 2D and 3D structures of known phytocystatins. In particular the 3D positions of the amino acid involved in binding, structures of active sites and local structural variation among members of the proposed phytocystatin family were studied.

2 Materials and Methods

2.1 Sequence Analysis

Amino acid sequences of phytocystatins were derived from at least 28 different plant species from NCBI databases (http://www.ncbi.nlm.nih.gov/). GeneBank accession number with the references of all the phytocystatins are represented in Table 1.

Multiple alignments were performed using the program CLUSTALX [26] with default setting, and the alignments were edited manually. Long sequences were truncated both at the N- and the C-terminal to include only the domain region, and the alignment was repeated.

Phylogenetic inference was performed using the PHYLIP version 3.5.7 suit [27]. First a distance matrix was generated using the PROTDIST program followed by the neighbor-joining method using NEIGHBOR programmer. A consensus tree was derived after 1000 bootstraps through the program BOOTST and CONSES. An un-rooted phylogenetic tree was contracted using TREEVIEW program [28] and MEGA5 [29].

Analysis of conserved motif was performed by MEME (Multiple Em for Motif Elicitation) software version 3.5.4

Table 1 Phytocystatins obtained from sequence databases

Code	Common name	Species name	Database (Acc No.)	References[a]
Apple	Apple	*Malus domestica*	EMBL: AY173139	Ryan et al. [3]
Ara	Arabidopsis	*Arabidopsis thaliana*	EMBL: AAM64985	Haas et al. [4]
Bar	Barley	*Hordeum vulgare*	EMBL: Y12068	Gaddour et al. [5]
Bit	Bitter Dock	*Rumex obtusifolius*	EMBL: AJ428415	Tinney et al. (unpub.)
Broc	Broccoli	*Brassica oleracea*	EMBL: AY065838	Watson and Coupe 2001 (unpub.)
Cab	Chinese Cabbage	*Brassica rapa*	EMBL: L41355	Lim et al. [6]
Car	Carnation (Clove pink)	*Dianthus caryophyllus*	EMBL: AY028994	Sugawara et al. [7]
Cast	Castor	*Ricinus communis*	EMBL: Z49697	Szederkenyi et al. [8]
Cau	Cauliflower	*Brassica oleracea*	TrEMBL: Q8VYX5	Watson and Coupe 2001 (unpub.)
ChesI	European Chestnut	*Castanea sativa*	EMBL: AJ224331	Pernas et al. [9]
ChesII	American Chestnut	*Castanea dentate*	EMBL: AF480168	Connors et al. (unpub.)
ChrB	Christmas Bells	*Sandersonia aurantiaca*	EMBL: AF469485	Eason 2002 (unpub.)
Cock	Cockscomb	*Celosia cristata*	EMBL: AJ535712	Gholizadeh et al. [10]
Cor	Corn	*Zea mays*	EMBL: D10622	Abe et al. [11]
Cow	Cowpea	*Vigna unguiculata*	EMBL: Z21954	Fernandes et al. [12]
Faba	Faba bean	*Vicia faba*	EMBL: AY237958	Berdnikov et al. (unpub.)
Kiwi	Kiwi	*Actinidia deliciosa*	GB: AY390353	Rassam and Laing [13]
Mugw	Mugwort	*Artemisia vulgaris*	EMBL: AF143677	Hubinger et al. [14]
Rice	Rice	*Oryza sativa*	EMBL: J03469	Abe et al. [15]
Pap	Papaya	*Carica papaya*	EMBL: X71124	Song et al. [16]
Pot	Potato	*Solanum tuberosum*	GB: AAB29661	Walsh et al. [17]
Sorg	Sorghum	*Sorghum bicolor*	EMBL: X87168	Li et al. [18]
Soy	Soyabean	*Glycine max*	EMBL: U51855	Zhao et al. [19], Botella et al. [20]
Sug	Sugarcane	*Saccharum officinarum*	NCBI: AAM78598	Soares-Costa et al. (unpub.)
Swe	Sweet potato (Batate)	*Ipomoea batatas*	EMBL: AF117334	To et al. [21], Huang et al. [22]
Taro	Taro	*Colocasia esculenta*	EMBL: AF525880	Yang et al. [23]
Tom	Tomato	*Solanum lycopersicum*	EMBL: AF198388	Girarad et al. [24]
Whe	Wheat	*Triticum aestivum*	EMBL: AB038393	Kuroda et al. [25]

[a] Years on unpublished references indicate date sequences were submitted in the database

(http://meme.sdsc.edu) using minimum and maximum motif width of 8 and 15, respectively, and a maximum number of 15 motifs, keeping the rest of the parameters at default.

2.2 Protein Structure Modeling and Analysis

The coordinate file (pdb) for OC-I [30] and papain [31] were obtained from the protein data bank (PDB) database (www.rcsb.org/pdb/) [32]. The OC-I pdb file was used to predict the 3D structures of selected representative from each of the phylogenetic groups. Structure modeling to predict the unknown structures was done using the I-TASSER [33] that determines structure using the satisfaction of spatial constraints. The input files consisted of the pdb file of OC-I and the amino acid sequence alignment between OC-I with the unknown sequence at greater that 30 % sequence similarity with OC-I. Predicted model was evaluated for energy distribution. Stereochemical quality of the predicted structures was tested using the energy minimization online tool on Chiron (http://troll.med.unc.edu/) and PROCHECK programs (http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) [34], respectively. Structures were visualized using both SWISS-PDB Viewer [35] and PyMOL (www.pymol.org).

3 Results and Discussion

3.1 Sequence Analysis

The multiple sequence analysis of the phytocystatins using ClustalX to shows high levels of sequence homology and conservation especially around the regions involved in function and important structural features (Fig. 1). The conserved glycine (G/GG) residue in the N-terminal region is known to be characteristic to this group of proteins and involved in N-terminal binding. The QVVxG motif characteristic of all members of the cystatin super family and responsible for the second binding site (located in the 2nd hairpin loop) was clearly identified in the multiple alignments (Fig. 1).

The LARFAV motif was also found in the N-terminal corresponds to the alpha-helix structure and is characteristic to phytocystatins [36]. A new YEAKxKxWxKxF was identified in the C-terminal of the phytocystatins. This
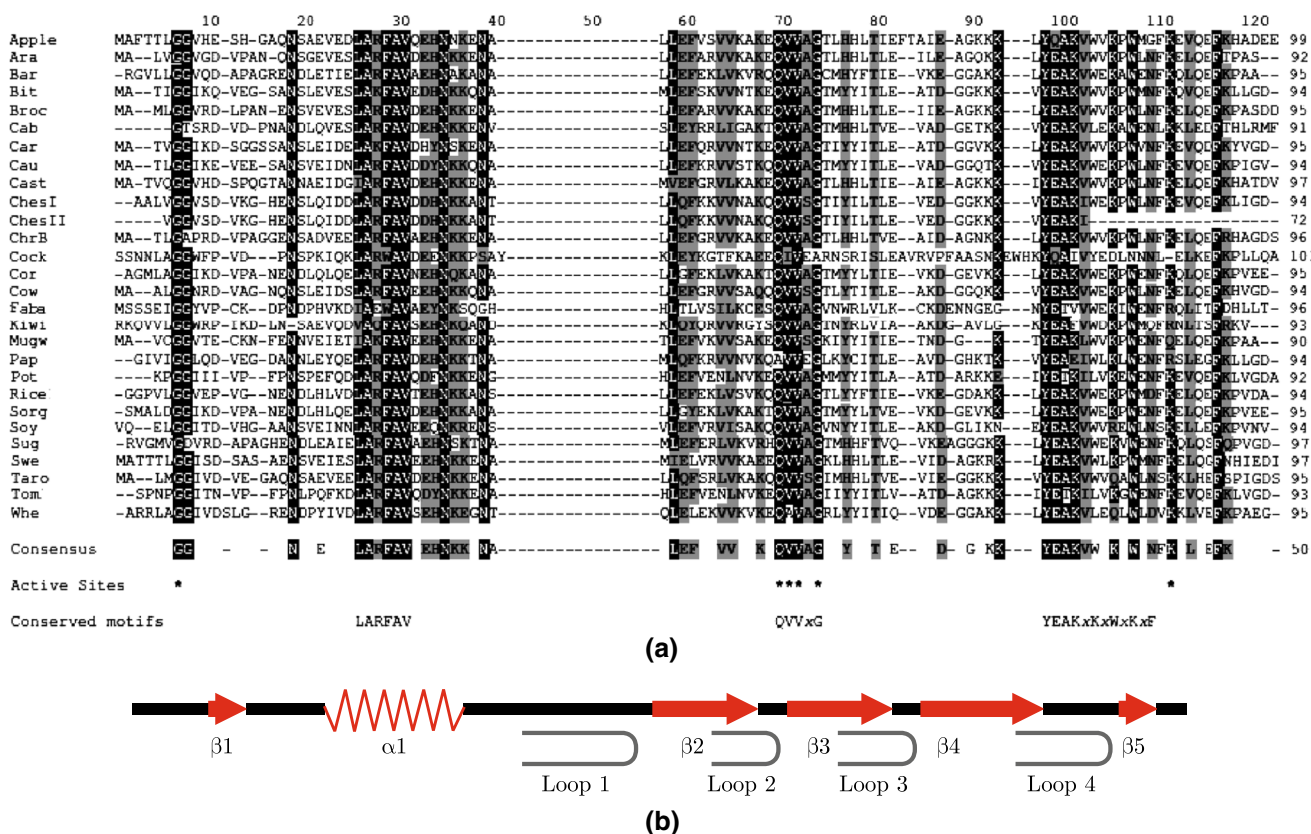
Fig. 1 a Amino acid sequence alignment of known phytocystatins showing residue conservation across the different cystatins studies. A consensus sequence was also generated. Identical amino acids are highlighted in *black*, while similar ones are in *gray*. **b** Cartoon of the generalized secondary structural elements of phytocystatins. The *orange arrows* are the β-sheet (numbered from 1 to 5) and red spiral representing the single α-helix. The positions were the loops occur are indicated with a *gray paper clip mark* and labeled 1–4

Fig. 2 Phylogenetic tree for known phytocystatins based on the neighbor-joining method using PROTDIST and NEIGHBOR program available in the PHYLIP (Phylogeny Inference Package) Version 3.5.7 suit. *Circled numbers* indicate the five clades that were obtained
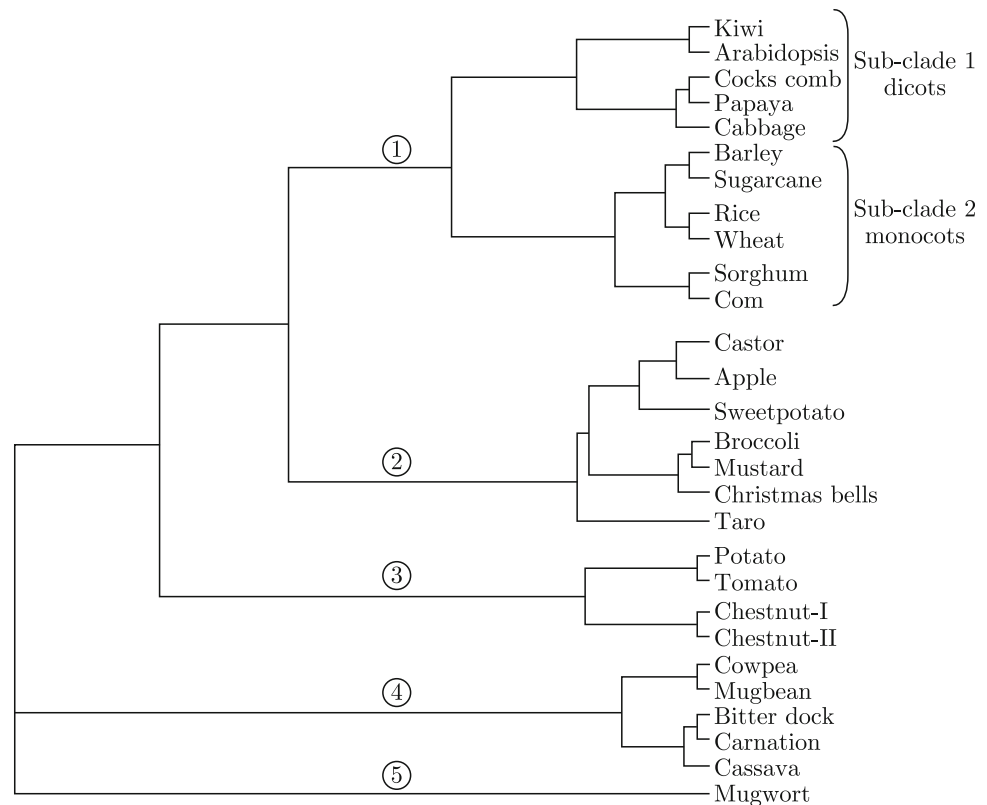
motif being unique to phytocystatins further adds to their qualification for a separate subfamily. This region is not as highly conserved as in animal cystatins. It has been respected to constitute the third binding region but with less binding capacity and probably more important in stabilizing the complex with proteases. Since this region is characteristic only to phytocystatins, several workers have proposed that this group of proteins may constitute a separate subfamily within the cystatin family. From the multiple alignments, it is also clear that there is high correlation of conserved region to important structural features used for either binding or structural conformity of the protein (Fig. 1).

To determine the evolutionary the phylogenetic relationships of five major cereal crop (rice, wheat, maize, barley and sorghum) cystatins with other cystatins for the based on the neighbor-joining phylogenetic tree that was generated, phytocystatins could be separated into five distinctive clades (Fig. 2). Clades 4 and 5 seemed to be more primitive and may be progenitor of all the other group of phytocystatins.

The biggest clade, clade1 could further be divided into two subclade, subclade 1 and 2 with the entire monocot cystatins grouping together in subclades 2. Subclade 1 included a rather more diverse group of phytocystatins. Other clades also showed high plant taxa relationships, for example in clade 3 are members of the Tomato and Potato

multi-cystatin, both plants belong to *Solanaceae* family and both phytocystatins are characterized by multiple domains. Clade 4 includes mostly members of the Fabaceae (Leguminosae) family for example the Mugbean and Cowpea cystatin. This clade seems to be evolutionary primitive. However, one of the domains clustered in clade 3, shows significant difference from its other domain cousins. The highest similarity percentage was found between rice, wheat, maize, barley and sorghum cystatins, and this suggests that these are orthologs, genes that have maintained sequence and functional similarity ranged from 48 to 86 % (Table 2).

ScanProsite tool for EXPASY was used to scan the cereal crop cystatins sequence against PROSITE patterns and profiles. PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them. It also used to identify the presence of Cysteine Proteases Inhibitor signature pattern (PS00287–[GSTEQKRV]–Q-[LIVT]–[VAF]–[SAGQ]–G-{DG}–[LIVMNK]–{TK}–X-[LIVMFY]–{S}–[LIVMFYA]–[DENQKRHSIV]). On examination we found that though this pattern was present in rice, wheat, maize, barley and sorghum phytocystatins protein sequence (Table 3). For further analysis of motifs in rice, wheat, maize, barley and sorghum, the MEME program was used. A motif is a sequence pattern that occurs repeatedly in a group of related protein or DNA sequences. MEME analysis

Table 2 Similarity between five major cereals crops (rice, wheat, maize, barely and sorghum) in percentages

Similarity between cereal crops		Similarity in percentages (%)
Rice	Wheat	65
Rice	Maize	73
Rice	Barley	57
Rice	Sorghum	72
Wheat	Maize	62
Wheat	Barley	48
Wheat	Sorghum	61
Barley	Maize	54
Barley	Sorghum	52
Maize	Sorghum	86

revealed the presence of highly conserved five motifs common in rice, wheat, maize, barley and sorghum. In barley motif 5 is absent. The twelve-residue-long YEAKxKxWx KxF motif that was highly conserved in the C-terminal end being unique to phytocystatins was identified in motif 1 and 3. The three-residue-long LARFAV motif in N-terminal corresponds to the alpha-helix structure was present in motif 2, and the five-residue-long QVVxG motif in second binding site which was located in the second hairpin loop was present in motif 4 (Fig. 3).

3.2 Structure Analysis

NMR structure of oryzacystatins (OC-I, PDB accession No. 1eqk) which is the only plant phytocystatin whose crystal structure has been determined so far, was the only template structure used for the comparative modeling to determine the 3D models of five major cereal crops (rice, wheat, maize, barley and sorghum) cystatins using I-TASSER. For each molecule, five structures were generated in the database, out of which the minimized average model with maximum score was selected. The energies of the designed structures were minimized using the energy minimization online tool of Chiron. The minimized structures were finally saved as .pdb files, which were validated online by PROCHECK software (data not shown).

The core structure of the OC-I involving an α-helix and four antiparallel β-strands is well preserved in all the major cereal crops cystatin with slight differences that do not bring any major changes in the 3D structures (Fig. 4). The hairpin loops L1 and L2, which are known to play an important role in interaction with the cysteine proteases and contain the highly conserved residues of QVVXG and P/AW, possess the same number of residues in five major cereal crops as in OC-I. Fig. 5 shows the similarity and differences among the five major cereal crops (rice, wheat, maize, barley and sorghum) cystatins, with respect to the secondary structure of OC-I. It also show the diagrammatic representation of the secondary structures of five major cereal crops and OC-I.

In silico docking experiments involving OC-I and papain revealed that OC-I attaches onto the active side cleft of papain and possibly in the same way with five major cereal crops (rice, wheat, maize, barley and sorghum) cysteine proteases (Fig. 6). Due to electrostatic forces along these two molecules, an average distance of 1.8 Å separates them from each other. During the docking process, one residue in the N-terminal of OC-I, aspartic acid (Asp4), prevented the inhibitor from docking closer into the papain active site (Fig. 6).

4 Discussion

Phytocystatins are inhibitors of cysteine proteinases that are used as potent weapons by phytopathogens and pests for invasion and colonization. In such a system of interacting proteins, each partner coevolves in response to the changes occurring in the cognate molecule [37]. Cystatins in each species have conserved residues particularly the ones involved in maintaining structure and related biochemical function. Despite the conservation of the sequences, sufficient variations are also present, which do not bring gross alterations in structure and function but do introduce changes in specificity and inhibitory activity. This is desirable for improving the repertoire of cysteine proteinase inhibitors in plants against offenders.

Table 3 Cysteine proteases inhibitor sequence motif in five major cereal crops (rice, wheat, maize, barley and sorghum)

Cereal crops	Cysteine proteases inhibitor	Cysteine proteases inhibitor no. of amino acid	Total no. of amino acid in sequences
Rice	QQVVAGTLYyFTIE	90–103	140
Wheat	EQAVAGRLYyITIQ	91–104	142
Maize	TQVVAGTMYyLTIE	84–97	134
Barley	QQVVAGCMHyFTIE	62–75	107
Sorghum	TQVVAGTMYyLTVE	80–93	130

Fig. 3 *Block diagram* represent and distribution of different motifs identified by MEME software in cystatins of rice, wheat, maize, barley and sorghum

Sequence	E-value
Maize	8.8e-104
Sorghum	5.2e-101
Rice	7.2e-97
Barley	4.3-80
Wheat	1.5e-73
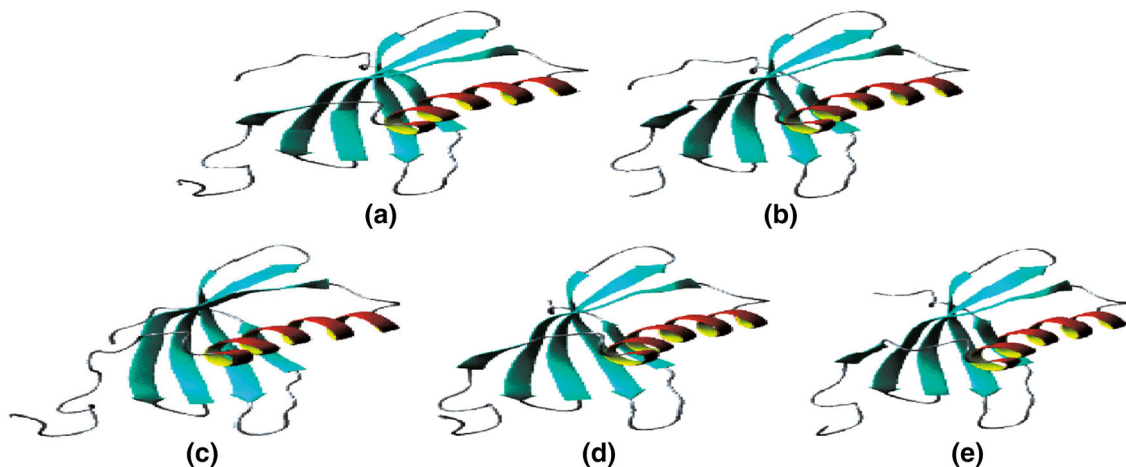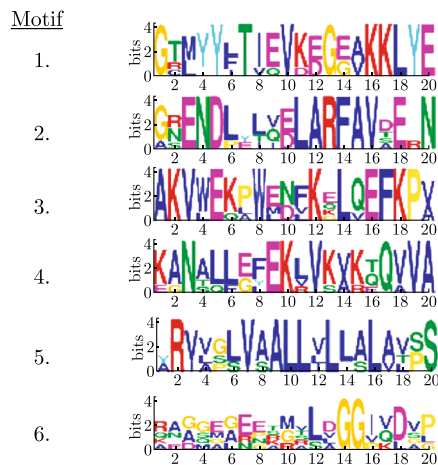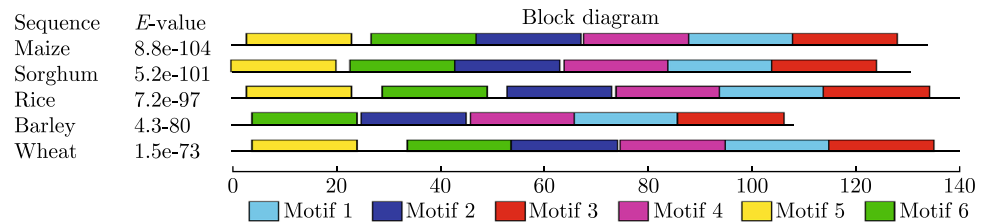
Fig. 4 a 3D structure of oryzacystatin. Predicted three-dimensional structures of five cereal crops **b** Wheat, **c** Maize, **d** Barley, **e** Sorghum and showing the secondary structure elements; five anti-parallel *β*-strands (*blue*), one α-helix (*red*), three hairpin loops, a long N-terminal trunk and short C-terminal. 3D structures were constructed by I-TASSER software using oryzacystatin as the template

In five major cereal crops, cystatins have been experimentally identified and characterized. Each of them display characteristic features of cystatins but with variations depending on spatial, temporal and conditional stimuli regulating their expressions. This study has provided new information about the structure of phytocystatins. Firstly, despite high structural similarity of phytocystatins, there is very wide variation in inhibition potential meaning that biochemical screening could yield selection of cystatins targeting a wide range of pests as well as other uses [1]. It has been shown in this study that the experimentally determined structure of OC-I can effectively and successfully be used to predict structural conformations of unknown cystatins. Docking and inhibitor–protease complex prediction was possible for the first time using phytocystatins from the analyses of binding candidate residues to engineer for improved binding were inferred.

Evolutionary relationships among phytocystatins were inferred using an unrooted phylogenetic tree. As expected, most phytocystatins grouped together to reflect the plant taxonomic groups. However, some members of the multidomain cystatins tended to occur in distinctly different
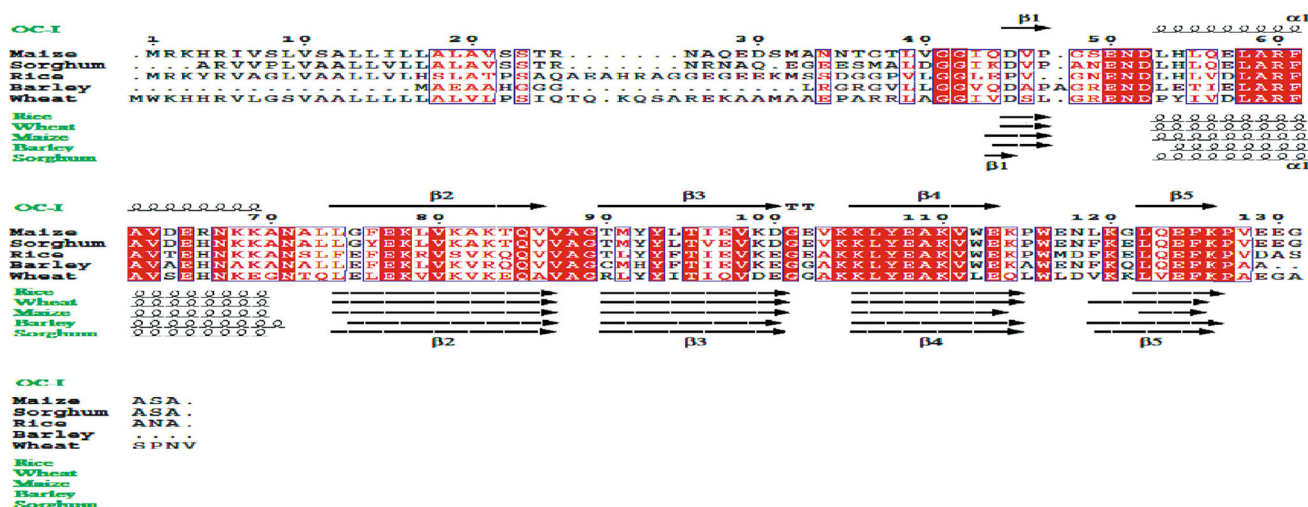
Fig. 5 Amino acid sequence alignment of oryzacystatin (OC-I) and five major cereal crops (rice, wheat, maize, barley and sorghum) cystatins. *Red box sequence* show highly similarity. The diagrammatic representation of secondary structures of five major cereal crops (rice, wheat, maize, barley and sorghum) cystatins
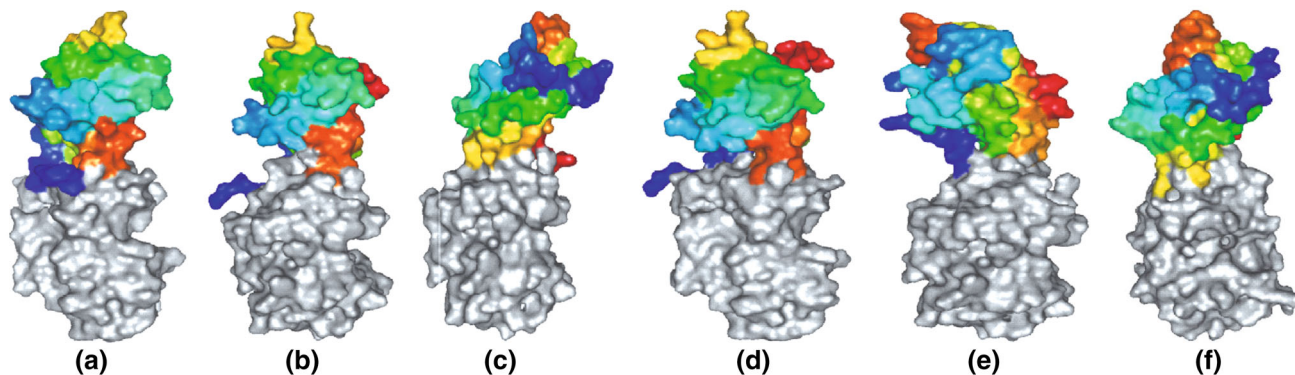
Fig. 6 **a** Modeled complexes between OC-I (*top*) and papain (*bottom*) in front and side view. OC-I is shown in spheres and its structure colored by rainbow. The N-terminal is blue toward the C-terminal *red*. The surface of papain is colored *light gray*. The active site of papain appears as a trench extending from the front to the back of the molecule. **b–f** The modeled complexes between OC-I (*top*) and **b** rice, **c** wheat, **d** maize, **e** barley and **f** sorghum (*bottom*). The complex was initially manually and the model refined using MULTIDOCK program based on minimization algorithms. Visualization and rendering graphics were done using PYMOL program

grouping. This suggests that plants, such as tomato, soybean and sugarcane, contain complete cystatin-coding genes that may have distinctly different evolutionary origin. The twelve residues long YEAKxKxWxKxF motif was highly conserved in the C-terminal end being unique to phytocystatins was identified in motif 1 and 3.

In silico observations in this study have shown the possible formation of a bond between residues GLY6 and VAL8 forming a loop structure in long enough N-terminals. This probably stabilizes the trunk allowing a more precise binding more and rendering the complex more stable. Cystatins bind to proteases with a 1:1 stoichiometry and with varying affinities. The whole phytocystatin molecule is wedge shaped with the N-terminal, and the two hairpin loops forming the sharp edge in some case the

N-terminal protrudes out into a long arm extending outwards from the rest of the structure forming what has been referred to as a trunk (Fig. 3). This sharp edge is highly hydrophilic and complimentary to the active cleft of cysteine proteases forming the active site region. The active site itself is composed of a glycine residue N-terminal, and this appears to be the most important binding site in many cystatins although its removal or absence does not seem to affect binding by other types of phytocystatins.

In this study, bioinformatics tools have been successfully used to predict the inhibitor–protease (OC-I and papain) complex. In general, the binding in the predicted complex was in agreement with that of the experimental complex structures previously reported between stefin-B and papain and stefin-A with cathepsin-H. In docking OC-I

and papain, it was difficult to dock two residues of aspartic acid ASP4 in the N-terminal trunk and ASP86 in the 2nd binding hairpin loop of the C-terminal region. These two residues are close to the active site and seemed to prevent closer binding of the active sited to the target papain. Therefore, these sites appear to be potential targets for site-directed at improving binding and therefore potency of OC-I to papain and probably to other cysteine protease. The wide variation in affinities found for phytocystatins in this study, and indeed also in the other animal cystatins, is not explainable by a simple structural difference. For example, an inhibitor with highly similar structural features has a great difference in affinity. 1st binding hairpin loop with the QVVAG highly conserved motif was not essential for cysteine protease inhibitor activity in cystatin-A. Therefore, it is possible that such functional differences may be explainable by small structural features at residue level that result in great differences in affinity of the inhibitor.

References

1. Shriparna D, Singh VK, Soma SM, Anil K (2010) In silico analysis of sequential, structural and functional diversity of wheat cystatins and its implication in plant defense. Genomic Proteomics Bioinform 8(1):42–56

2. Turk V, Bode W (1991) The cystatins: protein inhibitors of cysteine proteinases. FEBS Lett 285(2):213–219

3. Ryan SN, Liang WA, McManus MT (1989) A cysteine proteinase inhibitor purified from apple fruit. Phytochemistry 49:957–963

4. Haas BJ, Volfovsky N, Town CD, Troukhan M, Alexandrov N, Feldmann KA, Flavell RB, White O, Salzberg SL (2002) Full-length messenger RNA sequences greatly improve genome annotation. Genome Biol 3(6):1465–6906

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

6. Lim CO, Lee SI, Chung WS, Park SH, Hwang I, Cho MJ (1996) Characterization of a cDNA encoding cysteine proteinase inhibitor from Chinese cabbage (Brassica campestris L. ssp. pekinensis) flower buds. Plant Mol Biol 30(2):373–379

7. Sugawara H, Shibuya K, Yoshioka T, Hashiba T, Satoh S (2002) Is a cysteine proteinase inhibitor involved in the regulation of petal wilting in senescing carnation (Dianthus caryophyllus L.) flowers? J Exp Bot 53(368):407–413

8. Szederkenyi J, Komor E, Schobert C (1995) cDNA expressed in Ricinus cotyledons. Plant Physiol 109:721–722

9. Pernas M, Sanchez-Monge R, Gomez L, Salcedo G (1998) A chestnut seed cystatin differentially effective against cysteine proteinases from closely related pests. Plant Mol. Biol. 38(6):1235–1242

10. Gholizadeh A, Santha IM, Kohnehrouz BB, Lodha ML, Kapoor HC (2005) Cystatins may confer viral resistance in plants by inhibition of a virus-induced cell death phenomenon in which cysteine proteinases are active: cloning and molecular characterization of a cDNA encoding cysteine-proteinase inhibitor

11. (celostatin) from Celosia cristata (crested cock's comb). Appl Biochem 42(3):197–204

11. Abe M, Abe K, Kuroda M, Arai S (1992) Corn kernel cysteine proteinase inhibitor as a novel cystatin superfamily member of plant origin. Molecular cloning and expression studies. Eur J Biochem 209(3):933–937

12. Fernandes KV, Sabelli PA, Barratt DH, Richardson M, Xavier-Filho J, Shewry PR (1993) The resistance of cowpea seeds to bruchid beetles is not related to levels of cysteine proteinase inhibitors. Plant Mol Biol 23(1):215–219

13. Rassam M, Laing WA (2004) Purification and characterization of phytocystatins from kiwifruit cortex and seeds. Phytochemistry 65(1):19–30

14. Hubinger G, Gruber P, Vollmann U, Kraft D, Ferreira F, Himly M (1999) Molecular cloning and nucleotide sequence analysis of a cDNA (Accession No. AF143677) encoding the cystatin homolog from Mugwort Artemisia vulgaris) Pollen. Plant Physiol 120(2):634

15. Abe K, Emori Y, Kondo H, Suzuki K, Arai S (1987) Molecular cloning of a cysteine proteinase inhibitor of rice (oryzacystatin). Homology with animal cystatins and transient expression in the ripening process of rice seeds. J Biol Chem 262(35): 16793–16797

16. Song I, Taylor M, Baker K, Bateman RC Jr (1995) Inhibition of cysteine proteases by Carica papaya cystatin produced in Escherichia coli. Gene 162(2):221–224

17. Walsh TA, Strickland JA (1993) Proteolysis of the 85-kilodalton crystalline cysteine proteinase inhibitor from potato releases functional cystatin domains. Plant Physiol 103(4):1227–1234

18. Li Z, Sommer A, Dingermann T, Noe CR (1996) Molecular cloning and sequence analysis of a cDNA encoding a cysteine proteinase inhibitor from Sorghum bicolor seedlings. Mol Gen Genet 251(4):499–502

19. Zhao Y, Botella MA, Subramanian L, Niu X, Nielsen SS, Bressan RA, Hasegawa PM (1996) Two wound-inducible soybean cysteine proteinase inhibitors have greater insect digestive proteinase inhibitory activities than a constitutive homolog. Plant Physiol 111(4):1299–1306

20. Botella MA, Xu Y, Prabha TN, Zhao Y, Narasimhan ML, Wilson KA, Nielsen SS, Bressan RA, Hasegawa PM (1996) Differential expression of soybean cysteine proteinase inhibitor genes during development and in response to wounding and methyl assonate. Plant Physiol 112(3):1201–1210

21. To KY, Suen DF, Chen SCG (1999) A sweet potato leaf cDNA (Accession No. AF117334) encoding cysteine proteinase inhibitor. Plant Physiol 119(4):1568

22. Huang YJ, To KY, Yap MN, Chiang WJ, Suen DF, Chen SC (2001) Cloning and characterization of leaf senescence up-regulated genes in sweet potato. Physiol Plant. 113(3):384–391

23. Yang AH, Yeh KW (2005) Molecular cloning, recombinant gene expression, and antifungal activity of cystatin from taro (Colocasia esculenta cv. Kaosiung no.1). Planta 221:493–501

24. Girarad C, Rivard D, Kiggundu A, Kunert K, Gleddie SC, Cloutier C, Michaud D (2007) A multicomponent, elicitor-inducible cystatin complex in tomato, Solanum lycopersicum. New Phytol 173(4):841–851

25. Kuroda M, Kiyosaki T, Matsumoto I, Misaka T, Arai S, Abe K (2001) Molecular cloning, characterization, and expression of wheat cystatins. Biosci Biotechnol Biochem 65(1):22–28

26. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

27. Felsenstein J (1989) PHYLIP-phylogeny inference package (version 3.2). Cladistics 5:164–166

28. Page RDM (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12(4):357–358

29. Koichiro T, Daniel P, Nicholas P, Glen S, Masatoshi N, Sudhir K (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739

30. Nagata K, Kudo N, Abe K, Arai S, Tanokura M (2000) Three-dimensional solution of oryzacystatin-I, a cysteine proteinase inhibitor of the rice, *Oryza sativa* L. japonica. Biochemistry 39:14753–14760

31. 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 interactionD. EMBO J 9:1939–1947

32. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissing H, Shindyalov IN, Bourne PE (2000) The protein data bank. nucleic Acids Res 28:235–242

33. Ambrish R, Alper K, Yang Z (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Nat Protoc 5:725–738

34. Laskowski RA, MacArthur MW, Moss DS, Thomton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291

35. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdfViewer. An environment for comparative protein modeling. Electrophoresis 18:2714–2723

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

37. Martinez M, Diaz Z (2008) The origin and evolution of plant cystatins and their target cysteine proteinases indicate a complex functional relationship. BMC Evol Biol 8:198