Current Topics in Biochemical Research
Vol. 14, *1*, 2012

Review

The occurrence of proteins with multiple cystatin domains in eukaryotic organisms

Henning Scholze*

University of Osnabrueck, Faculty of Biology, Biochemistry, Barbarastr. 13, 49069 Osnabrueck, Germany

ABSTRACT

Proteases are widespread enzymes that control protein functions inside and outside a cell. Their well directed action is predominantly ensured by the availability of various more or less specific inhibitors, which themselves may be proteins. Cysteine proteases namely the thiol-dependent cathepsins are the most important intracellular proteolytic enzymes. The control of their activities is realized by limited proteolysis of their polypeptide chain, a change of the redox potential of their environment or by the inhibition via small peptide analogues as well as proteinaceous inhibitors. The by far most important protein inhibitors of cysteine proteases widely distributed are those of the chagasin and the cystatin family, respectively, well-known as relatively small (~10-12 kDa) and very robust monomeric proteins. Orthologous sequences are also found as oligo- and multi-domain single chain proteins consisting of up to eight cystatin domains that probably arise as a result of multiple events of gene duplications. Moreover, single and multiple cystatin domains that precede the prosequences of particular cathepsins, known as cathepsin F, can be identified in data bases. In these predicted and partially characterized proteins up to 17 cystatin domains are found, which in most cases are recognizable due to their typically conserved protease binding motifs. Although it may suggest itself that these domains are involved in the control of enzyme activities of endogenous and exogenous proteases for the purpose of cell protection, the significance of multiple cystatin domains in these proteases is still completely unknown.

KEYWORDS: cystatin, cysteine protease inhibitor, multi-domain protein, I29, cathepsin F

INTRODUCTION

Cystatins are well studied cysteine protease inhibitors found in all kingdoms of life, implying eukaryotes and *Bacteria*, but not *Archaea* [1]. This widespread protein family features relatively small, single chain proteins (M_r 11,000 to 18,000) characterized by the presence of a conserved QXVXG motif in their primary structure functioning as the main protease binding region [2]. The functions of members of this protein class are primarily the inhibition of proteolytic activities of cysteine peptidases of the papain-type which represent a major protease class in all kingdoms of organisms, both intra- and extracellularly [3]. In addition to the monomeric cystatins, several proteins containing more than one cystatin domains are found to be involved in a lot of physiological processes, such as blood coagulation, or are components of the kallikrein-kinin-system involved in inflammation or blood pressure regulation [4]. Those important members of the cystatin-superfamily are the kininogens and other serum proteins, namely the fetuins participating in transport processes of several substances in the blood [5, 6]. Although diverse studies concerning the biological roles of these multi-domain proteins have already

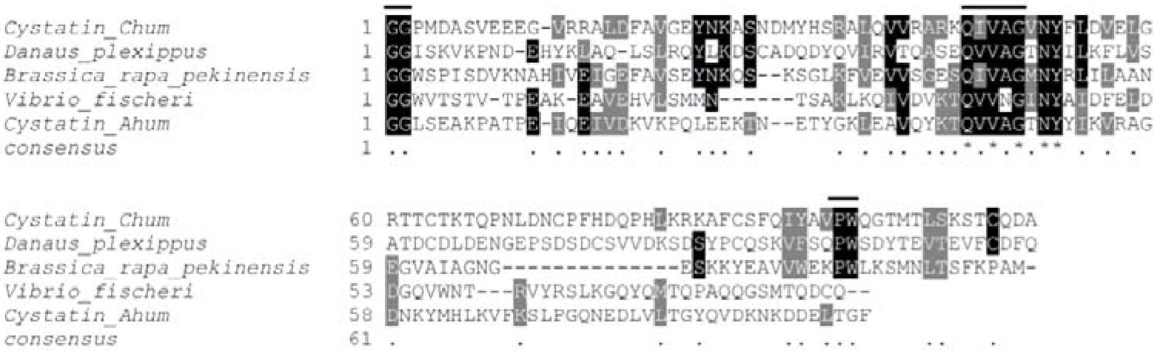
*scholze@biologie.uni-osnabrueck.de

been done, the detailed contributions of the cystatin domains within these processes are still uncertain [3, 7]. The upcoming knowledge of sequence data available from the vast number of genome sequencing projects enables the identification of further putative proteins containing single- and polycystatin sequences. These in turn are found as discrete open reading frames or as constituents of various cysteine peptidase encoding genes whose real functions wait to be discovered. Ideas to this are described in the following.

Monocystatins

Monomeric cystatins are subdivided into type I cystatins (stefin A and B *syn*. cystatin A and B) and type II cystatins (the true cystatins) [2]. Solely true cystatins (for example cystatin C) possess two intramolecular disulphide bonds near their C-termini [8]. Apart from the stefins [9], all monocystatin sequences encoded in animal genes (except protists), and almost all examples of plant and bacterial origins, respectively, are equipped with signal sequences suggesting that these proteins fulfil their main tasks non-cytoplasmatically or extracellularly [2]. The most important protease binding motif of cystatins, QXVXG and relatives, is known to build up the first hairpin loop representing the central part of the wedge-shaped structure of the cystatin molecule, as evaluated by X-ray crystallography [10]. The protease binding region also includes the N-terminal glycine residue and, in case of type II cystatins, the RP-sequence close to the C-terminus designing the third minor binding site [10]. In all examples of monocystatin/ protease interactions studied so far, the molar ratio of protease and bound inhibitor observed is 1:1. The binding of a cystatin molecule to its target protease is very tight, but not covalent, with K_i's in the picomolar range [11]. Forming of the protease/inhibitor complex, in which the active site of the protease is completely masked by the inhibitor molecule thereby blocking the access of a substrate, does not induce large conformational changes within both binding partners [12].

As already mentioned, cystatins are only found in eukaryotes and *Bacteria*; out of which bacterial cystatins never contain more than one cysteine residue in their amino acid sequence, so that they are unable to form intramolecular disulphide bonds (Fig. 1). This fact, together with the comparatively low molecular sizes (less than 90 residues) and the absence of the PW-motif in their primary structures [9], suggests that bacterial cystatins are more related to type I cystatins (stefins) than to type II cystatins. By contrast, phytocystatins (exclusively from angiosperms) are more related to animal cystatin C than to cystatin A and B, respectively, which is reflected in the presence of a sequence triplett near their C-terminus forming the second hairpin loop (Fig. 1) [13]. Furthermore, it is obvious that those residues preceding the PW-motif in the second



```
Cystatin_Chum                 1 GGPMDASVEEEG-VRRALDFAVGEYNKASNDMYHSRALQVVRARKQIVAGVNYFLDVELG
Danaus_plexippus              1 GGISKVKPND-EHYKLAQ-LSLRQYLKDSCADQDYQVIRVTQASEQVVAGTNYILKFLVS
Brassica_rapa_pekinensis      1 GGWSPISDVKNAHIVEIGEFAVSEYNKQS--KSGLKFVEVVSGESQIVAGMNYRLILAAN
Vibrio_fischeri               1 GGWVTSTV-TPEAK-EAVEHVLSMMN------TSAKLKQIVDVKTQVVNGINYAIDFELD
Cystatin_Ahum                 1 GGLSEAKPATPE-IQEIVDKVKPQLEEKTN--ETYGKLEAVQYKTQVVAGTNYYIKVRAG
consensus                     1 ..          . . ....  .  ... .    . . .. ...*.*.*.**  . . .

Cystatin_Chum                60 RTTCTKTQPNLDNCPFHDQPHLKRKAFCSFQIYAVPWQGTMTLSKSTCQDA
Danaus_plexippus             59 ATDCDLDENGEPSDSDCSVVDKSDSYPCQSKVFSQPWSDYTEVTEVFCDFQ
Brassica_rapa_pekinensis     59 EGVAIAGNG--------------ESKKYEAVVWEKPWLKSMNLTSFKPAM-
Vibrio_fischeri              53 DGQVWNT---RVYRSLKGQYQVTQPAQQGSMTQDCQ--
Cystatin_Ahum                58 DNKYMHLKVFKSLPGQNEDLVLTGYQVDKNKDDELTGF
consensus                    61 .         .         .  .      .. ...      ..   .
```

Fig. 1. Alignment of mono-cystatin sequences from plant, insect and bacterium compared to human cystatin A and C. Searches for mono-cystatin sequences was performed by BLAST searches of the NCBI database (http://www.ncbi.nlm.nih.gov/blast). Multiple alignments were performed using the ClustalW algorithm (http://mobyle.pasteur.fr/cgi-bin/portal.py) [29]. Identical residues are shown in black, strongly similar ones in grey; bars above the sequences indicate the three main protease binding motifs.

hairpin loop of plant cystatins in most cases are lysines, occasional arginines, whereas this position in animal cystatins is frequently occupied by glutamine residues. This observation generally emphasizes the importance of a polar residue at this position; in case of plant cystatins, positively charged amino acids are favoured here.

Oligocystatins

Gene sequences deposited in the data base encoding polypeptide chains composed of two or more cystatin domains so far are only found in eukaryotes, but not in prokaryotes. Predicted proteins consisting of solely two cystatin monomers (dicystatins) can be identified in genes of a minor number of insects and molluscs such as the water flea *Daphnia pulex* and *Daphnia magna*, respectively, and the mollusc *Haliotis diversicolor supertexta*. Additionally, numerous different plant species as well as the green alga *Coccomyxa* are equipped with genes encoding dicystatins [14]. Whereas the individual cystatin domains of animal dicystatins may be evolved by simple gene duplications, the situation in the plant counterparts looks quite differently. Comparisons of the individual cystatin domains of dicystatins from different plant sources reveal that either of their two cystatin domains altogether show greater discrepancies among each other than the respective domains from different dicystatins (Fig. 2). This manifests itself in that sequence alignments of the first domains from different plant species alone reveal more than 80% identity, whereas alignments of each of the anterior and posterior domains of a particular dicystatin display sequence identities of only about 20%. As can further be taken from Fig. 2, the glutamine residues occupying the first position of the pentapeptide motif in the first cystatin domain of many plant dicystatins each are frequently replaced by glutamates followed by two aliphatic and two acidic residues in the respective motif of the second domain. The fact that the sequences of the second cystatin domains of most plant dicystatins in each case deviate more considerably from those of the first domains is also expressed in the absence of the turn-forming proline and the following tryptophan residue forming the second hairpin loop. From this, it

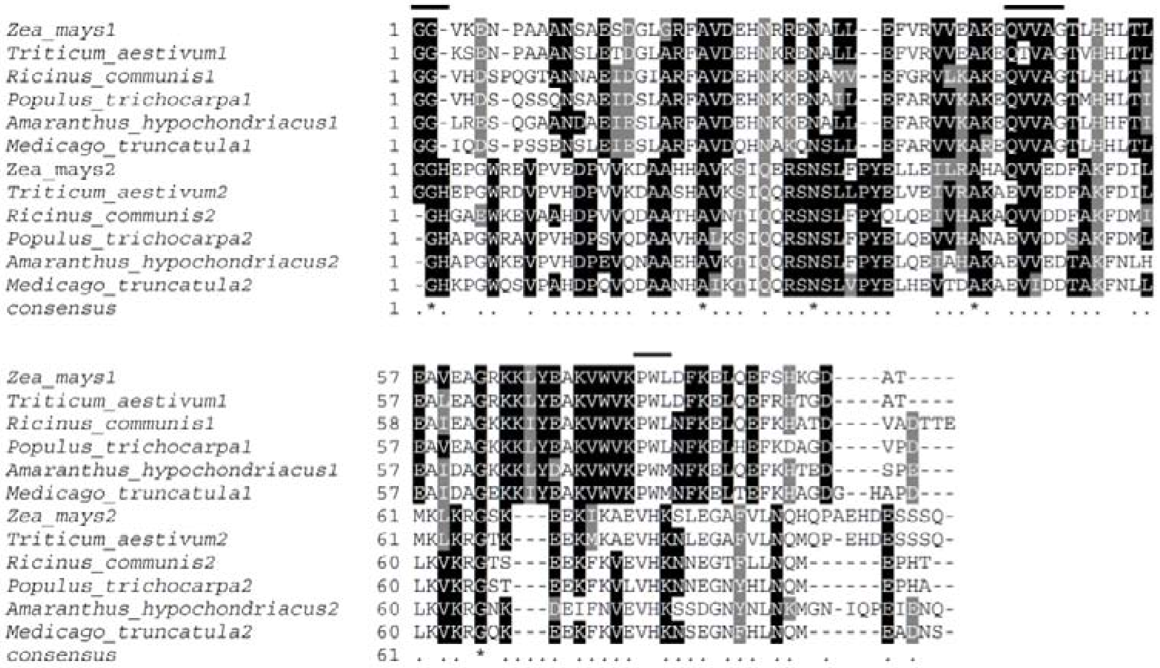

Zea_mays1 1 GG-VKDN-PAAANSAESDGLGRFAVDEHNRRENALL--EFVRVVEAKEQVVAGTLHHLTL
Triticum_aestivum1 1 GG-KSDN-PAAANSLETDGLARFAVDEHNKRENALL--EFVRVVEAKEQIVAGTVHHLTL
Ricinus_communis1 1 GG-VHDSPQGTANNAEIDGIARFAVDEHNKKENAMV--EFGRVLKAKEQVVAGTLHHLTI
Populus_trichocarpa1 1 GG-VHDS-QSSQNSAEIDSIARFAVDEHNKKENAIL--EFARVVKAKEQVVAGTMHHLTI
Amaranthus_hypochondriacus1 1 GG-LRES-QGAANDAEIESLARFAVDEHNKKENALL--EFARVVKAKEQVVAGTLHHFTI
Medicago_truncatula1 1 GG-IQDS-PSSENSLEIESLARFAVDQHNAKQNSLL--EFARVVKAREQVVAGTLHHLTL
Zea_mays2 1 GGHEPGWREVPVEDPVVKDAAHHAVKSIQERSNSLFPYELLEILRAHAQVVEDFAKFDIL
Triticum_aestivum2 1 GGHEPGWRDVPVHDPVVKDAASHAVKSIQQRSNSLLPYELVEIVRAKAEVVEDFAKFDIL
Ricinus_communis2 1 -GHGAEWKEVAAHDPVVQDAATHAVNTIQQRSNSLFPYQLQEIVHAKAQVVDDFAKFDMI
Populus_trichocarpa2 1 -GHAPGWRAVPVHDPSVQDAAVHALKSIQQRSNSLFPYELQEVVHANAEVVDDSAKFDML
Amaranthus_hypochondriacus2 1 -GHAPGWKEVPVHDPEVQNAAEHAVKTIQQRSNSLFPYELQEIAHAKAEVVEDTAKFNLH
Medicago_truncatula2 1 -GHKPGWQSVPAHDPQVQDAANHAIKTIQQRSNSLVPYELHEVTDAKAEVIDDTAKFNLL
consensus 1 .*. .. . . *... . ..*..... ...*. ....
Zea_mays1 57 EAVEAGRKKLYEAKVWVKPWLDFKELQEFSHKGD----AT----
Triticum_aestivum1 57 EALEAGRKKLYEAKVWVKPWLDFKELQEFRHTGD----AT----
Ricinus_communis1 58 EAIEAGKKKLYEAKVWVKPWLNFKELQEFKHATD----VADTTE
Populus_trichocarpa1 57 EAVEAGKKKLYEAKVWVKPWLNFKELHEFKDAGD----VPD---
Amaranthus_hypochondriacus1 57 EAIDAGKKKLYDAKVWVKPWMNFKELQEFKHTED----SPE---
Medicago_truncatula1 57 EAIDAGEKKIYEAKVWVKPWMNFKELTEFKHAGDG--HAPD---
Zea_mays2 61 MKLKRGSK---EEKIKAEVHKSLEGAEVLNQHQPAEHDESSSQ-
Triticum_aestivum2 61 MKLKRGTK---EEKVKAEVHKNLEGAEVLNQMQP-EHDESSSQ-
Ricinus_communis2 60 LKVKRGTS---EEKFKVEVHKNNEGTELLNQM------EPHT--
Populus_trichocarpa2 60 LKVKRGST---EEKFKVLVHKNNEGNYHLNQM------EPHA--
Amaranthus_hypochondriacus2 60 LKVKRGNK---DEIFNVEVHKSSDGNYNLNKMGN-IQPEIENQ-
Medicago_truncatula2 60 LKVKRGQK---EEKFKVEVHKNSEGNEHLNQM------EADNS-
consensus 61 . .. * ....

Fig. 2. Alignment of sequences of single cystatin domains from plant dicystatins. Identical residues are shown in black, strongly similar ones in grey; bars above the sequences indicate the three main protease binding motifs. Sequence searches and alignments were performed as described in the legend of Fig. 1.

can be concluded that the primary structures of phytodicystatins are obviously not the products of single gene duplications, but may rather be derived from fusions of separately developed genes (convergent evolution) after lateral gene transfer (Fig. 3).

Gene sequences deposited in the data base encoding polypeptide chains consisting of more than two cystatin monomers can be found in the genomes of the mentioned avertebrates as well as of angiosperms. These include open reading frames from the water flea *Daphnia pulex* and the abalone *Haliotis discus hannai*, respectively. Tricystatins have been identified in the genomes of the crop plants tomato and bean, and have already been purified from seeds of sun flower [15]. In this regard, it can be seen straight away that the animal examples are distinguished from their plant counterparts in a particular exchange of the glutamine residue by a glutamate at the first position of the pentapeptide motif constituting the first hairpin loop (Fig. 4). It can generally be observed that the particular first positions of the tripeptide motifs in each of the three cystatin domains of animal tricystatins forming the second hairpin loop are occupied by glutamines, whereas those positions in their herbal counterparts are exclusively faced by lysine/arginine residues.

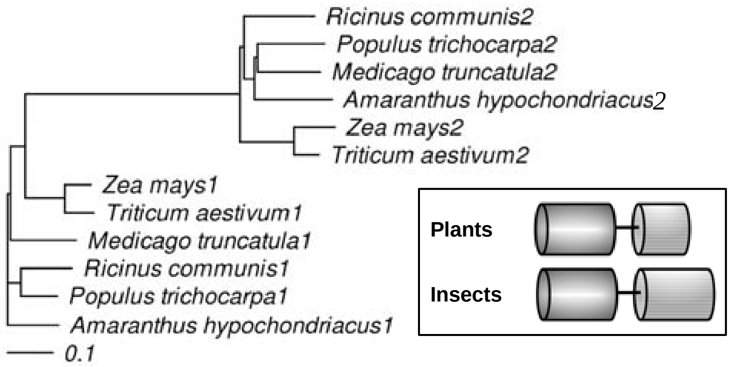


Fig. 3. Horizontal cladogram of sequences of both cystatin domains from plant dicystatins. Each cystatin domain occupy separate and trans-species clades. The dendrogram was created by the Neighbour-Joining method [30]. Insert shows a schematic representation of tricystatin molecules from plants and insects.

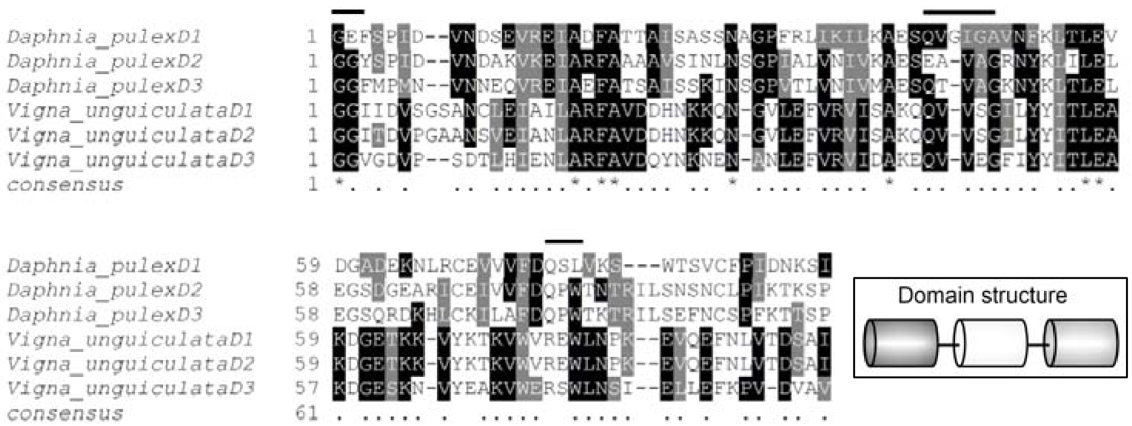

Daphnia_pulexD1 1 GEFSPID--VNDSEVREIADFATTAISASSNAGPFRLIKILKAESQVGIGAVNEKLTLEV
Daphnia_pulexD2 1 GGYSPID--VNDAKVKEIARFAAAAVSINLNSGPIALVNIVKAESEA-VAGRNYKLILEL
Daphnia_pulexD3 1 GGFMPMN--VNNEQVREIAEFATSAISSKINSGPVTLVNIVMAESQT-VAGKNYKLTLEL
Vigna_unguiculataD1 1 GGIIDVSGSANCLEIAILARFAVDDHNKKQN-GVLEFVRVISAKQQV-VSGILYYITLEA
Vigna_unguiculataD2 1 GGITDVPGAANSVEIANLARFAVDDHNKKQN-GVLEFVRVISAKQQV-VSGILYYITLEA
Vigna_unguiculataD3 1 GGVGDVP--SDTLHIENLARFAVDQYNKNEN-ANLEFVRVIDAKEQV-VEGFIYYITLEA
consensus 1 *. . . ..*.**.... .. * .... * ... .**.
Daphnia_pulexD1 59 DGADEKNLRCEVVVFDQSLVKS---WTSVCFPIDNKSI
Daphnia_pulexD2 58 EGSDGEARICEIVVFDQPWTNTRILSNSNCLPIKTKSP
Daphnia_pulexD3 58 EGSQRDKHLCKILAFDQPWTKTRILSEFNCSPFKTTSP
Vigna_unguiculataD1 59 KDGETKK-VYKTKVWVREWLNPK--EVQEFNLVTDSAI
Vigna_unguiculataD2 59 KDGETKK-VYKTKVWVREWLNPK--EVQEFNLVTDSAI
Vigna_unguiculataD3 57 KDGESKN-VYEAKVWERSWLNSI--ELLEFKPV-DVAV
consensus 61 . ...
Domain structure

Fig. 4. Alignment of sequences of single cystatin domains from plant tricystatins. Identical residues are shown in black, strongly similar ones in grey; bars above the sequences indicate the three main protease binding motifs. Sequence searches and alignments were performed as described in the legend of Fig. 1. Insert shows a schematic representation of a tricystatin molecule.

This in turn suggests that neutral or positively charged polar residues are important for protease binding at this site. Note that oligocystatins in most cases are equipped with signal sequences suggesting that their main functions are extracellular. These tasks rather imply the defence of pests than the performance of own housekeeping processes.

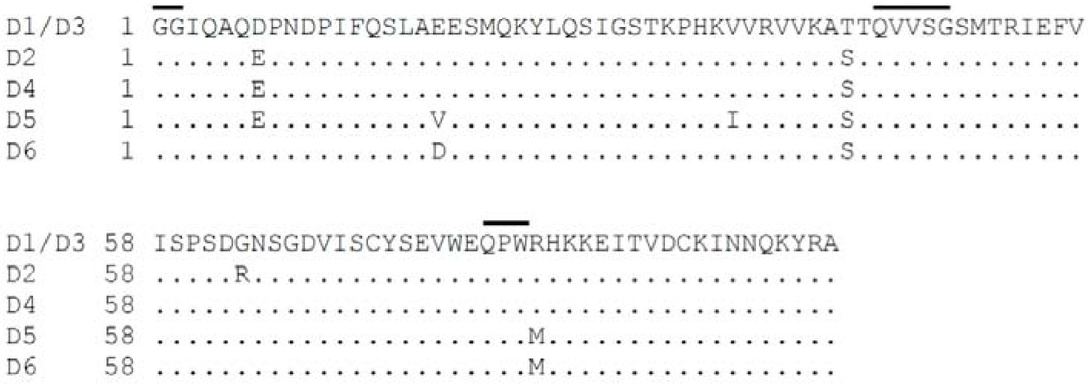
Polycystatins

A gene encoding a single protein solely composed of six cystatin units is found in the genome of the tobacco hornworm *Manduca sexta*. The polypeptide chain of this hexacystatin (M_r 70 kD, pI 8.8) is recognizable by containing not less than six exemplars of the strictly conserved canonical sequence motifs, QVVSG. In this polycystatin, the cystatin domains 1 and 3 are completely identical, the sequences of the other cystatin units differ only marginally from each other. Typical features of the individual cystatin modules of this hexacystatin comprise two successive glycine residues close to the individual N-terminal regions of each cystatin unit (Fig. 5); solely glutamine residues precede the PW motifs as parts of the second hairpin loops near their C-*termini*. The high similarity of the respective cystatin domains in this multi-domain protein is also expressed by the phylogenetic tree, as shown in Fig. 6 (upper scheme), which gives an impression of the branching events during evolution. The canonical and symmetrical appearance of multiple and strongly conserved cystatin domains in this protein suggests its emergence from a single precursor by simple gene duplications at recent times.

A gene encoding a protein composed of actually eight cystatin repeats is found as elicitor-inducible protein in the tomato *Solanum lycopersicum* and in the potato *Solanum tuberosum* [16, 17, 18, 19]. The primary structures of their various central pentapeptide motifs together with the C-terminal canonical tripeptides in each cystatin domain of these octameric cystatins differ more considerably from each other than those of the mentioned hexacystatin of *M. sexta*, although the repetitive occurrences of the typical canonical cystatin motifs (QXVXG{24}R/KEW) in these angiosperms are undoubtedly discernible (Fig. 6, lower scheme) [18, 19]. The major sequence divergence of the individual cystatin units of the potato and tomato octacystatins indicate major distances among incidents of gene duplications and subsequent mutations than in the *M. sexta* protein (Fig. 6). The physiological role of this polycystatin has been supposed to be protease inhibition during tuber growth or a general protective role against proteolytic attacks [17, 20, 21].

Cystatin/cathepsin F-chimaera

Further cystatin sequences are deposited in the sequence data base encoding proteins which represent chimera of cystatins and papain-like cysteine proteases. Cysteine proteases of the papain-type, classified as Clan CA, family C1A,



```
D1/D3  1  GGIQAQDPNDPIFQSLAEESMQKYLQSIGSTKPHKVVRVVKATTQVVSGSMTRIEFV
D2     1  ......E....................................S.............
D4     1  ......E....................................S.............
D5     1  ......E..........V.................I.......S.............
D6     1  .................D.........................S.............

D1/D3 58  ISPSDGNSGDVISCYSEVWEQPWRHKKEITVDCKINNQKYRA
D2    58  .....R....................................
D4    58  ..........................................
D5    58  .......................M..................
D6    58  .......................M..................
```

Fig. 5. Alignment of sequences of cystatin domains from the hexacystatin of *M. sexta*. Different residues are designated, identical residues are denoted by a dot. Sequence alignments of the cystatin domains were performed as described in the legend of Fig. 1. Domains 1 and 3 are identical.

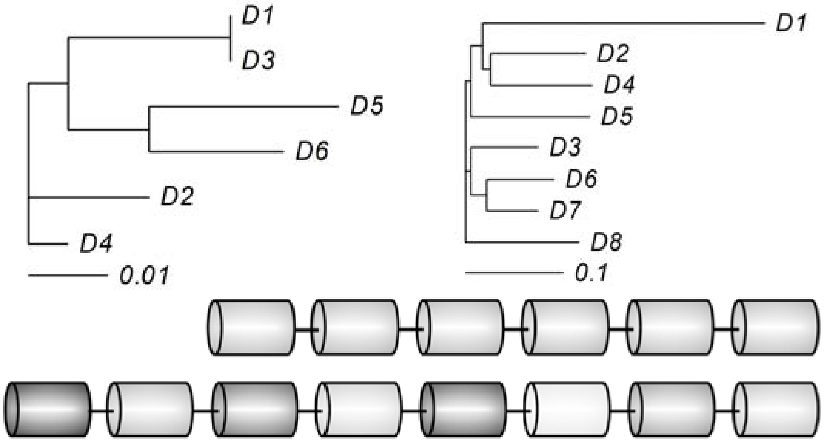


Fig. 6. Phylogenetic relationship of cystatin domains of poly-cystatins. Upper schemes**,** dendograms of cystatin domains from *M. sexta* (left) and *S. tuberosum* (right). The dendrograms were created as described in the legend of Fig. 3. Lower schemes, domain structures of poly-cystatins from *M. sexta* (upper string) and *S. tuberosum* (lower string).

in the MEROPS database, are hydrolytic enzymes assuming a vesicular and extracellular position of action [2, 22]. These enzymes are usually synthesized as zymogens or proenzymes and later on are converted to their active forms by limited proteolyses. The precursors of papain-like proteases usually contain N-terminal extensions of about 60 amino acid residues each forming α-helical domains that cover the substrate-binding site, thereby preventing the approach of a substrate before the enzyme has arrived at its target location [23, 24]. During zymogen activation, this peptide region is removed by proteolytic cleavage either autocatalytically or by the action of another protease activity resulting in a conformational change that facilitates the access of a substrate to the active site [25]. The propeptide sequences in turn in some cases are homologous to another inhibitor type of cysteine peptidases known as I29 that also exists as individual protein [9, 26]. A protease comprising a papain-type protease sequence at its C-terminal section and an I29 domain at its N-*terminus* is classified as cathespin F, an enzyme which plays an important role in the formation of peptides for antigen presentation in mammals [27, 28].

Surprisingly, a lot of genes encoding complete cystatin modules arranged N-terminally to the I29 propeptide region of cathepsins F can be traced in data bases. In these derived proteins, the cystatin domains are parts of the prosequence of the particular cysteine protease and may be enclosed as single or as multiple copies. The calculated molecular masses of the predicted protein products range from 50 to 300 kD, and many of them are additionally equipped with signal sequences. Monocystatin domains residing N-terminally to cathepsin F are found in mammalians, including *Homo sapiens*, and in amphibians, fishes, insects and worms; their domain organization is depicted in Fig. 7 (upper pictogram). Chimeric proteins containing two successive cystatin domains N-terminally connected to their particular cathepsin F are solely found in insects of the order *Hymenoptera*, such as ants, bees and wasps, and in some *Hemiptera* and *Diptera*. The primary structures of the individual cystatin domains of these chimeric proteins often deviate more strongly from each other than those observed in the mentioned polycystatins. That manifests itself in the nature of their respective first cystatin domains, whose sequence tripletts forming the second hairpin loop each are generally more related to plant than to animal cystatins, an observation which suggests the adoption of these cystatin genes by various lateral gene transfers from multiple sources at different times.

Open reading frames encoding chimeric proteins composed of cathepsin F and three or more cystatins are found in insects, such as the jumping

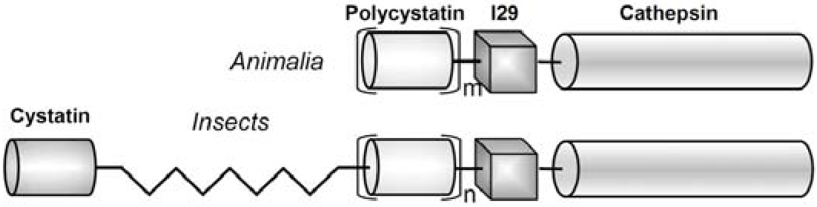


Fig. 7. Schematic representation of domain structures of cystatin/cathepsin F-chimaera. Upper pictogram, Chimeric protein consisting of one or more cystatin domains (short tube), an I29 inhibitor domain (cube) and a cathepsin unit (long tube). Lower pictogram, the same as in the upper one plus an undefined polypeptide region and an N-terminal cystatin unit; m denotes values from 1 to 17, n is up to 5.

ant *Harpegnathos saltator* (3 cystatin domains), the stink bug *Plautia stali* (4), the malaria transmitting mosquito *Anopheles gambiae* (5), the red flour beetle *Tribolium castaneum* and the silkworm *Bombyx mori* (9), respectively. An example of a gene coding for a chimeric protein containing not less than 17 cystatin domains located N-terminally to its cathepsin F sequence is found in the mentioned moth *M. sexta*. In this predicted multi-domain protein, a block of nine canonical cystatin sequences correspond to that found in the hexacystatin protein from the same organism, whereas the residual cystatin domains more strongly diverge from each other (Fig. 7, lower pictogram). In other examples of multi-domain proteins found in the data base containing both cathepsin F and cystatin units, the sections containing the cystatin domains may be interrupted by large segments of sequences of unknown significance. These segments characterized by a high content of diglycine motifs are found in some mosquitoes, such as *Anopheles* and *Culex* species. In these predicted proteins the undefined sequence segments always start after the end of the first cystatin domain and may consist of up to 900 residues; further up to five cystatin domains follow and reside directly in front of the I29 sequence. The significance of all these chimeric proteins is still unknown.

CONCLUSION

The protein group of cystatins comprise an important class of cysteine protease inhibitors and is found in all kingdoms of life except *Archaea*. Sequence comparisons reveal that cystatin monomers of vertebrate and plant origins are more related to cystatin C than to cystatin A or B (stefins), whereas monomeric cystatins from protists as well as from bacteria rather resemble stefins A/B. However, cystatins not only are found as individual and relatively small proteins, but also in terms of large single-chain proteins, in which several cystatin domains are lined up like pearls on a string. The evolutional divergence of the single cystatin domains in these proteins is quite different, ranging from rather identity in the *Manduca* hexacystatin to greater variety in divalent plant cystatins. Multi-domain cystatins may either be composed of solely up to eight individual cystatin units, or may be made up of one or many cystatin units N-terminally attached to the proform of cathepsin F, which itself consists of a particular papain-like cysteine protease and the cysteine protease inhibitor I29. In some of these multi-domain cysteine proteases, the stretch of multiple cystatin sequences is interrupted by large regions of diglycine-rich regions of unknown significance. It is supposed that the cystatin units found in plant and insect proteins are involved in the defense of pathogens.

REFERENCES

1. Kordis, D. and Turk, V. 2009, BMC Evol. Biol., 9, 266.
2. Turk, V., Stoka, V. and Turk, D. 2008, Front. Biosci., 13, 5406.
3. Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B. and Turk, D. 2012, Biochim. Biophys. Acta, 1824, 68.
4. Campbell, D. J. 2003, Int. J. Biochem. Cell. Biol., 35, 784.
5. Rawlings, N. D. and Barrett, A. J. 1990, J. Mol. Evol., 30, 60.

6. Lee, C., Bongcam-Rudloff, E., Sollner, C., Jahnen-Dechent, W. and Claesson-Welsh, L. 2009, Front. Biosci., 14, 2911.
7. Brown, W. M. and Dziegielewska, K. M. 1997, Protein Sci., 6, 5.
8. Turk, V. and Bode, W. 1991, FEBS Lett., 285, 213.
9. Rawlings, N. D., Tolle, D. P. and Barrett, A. J. 2004, Biochem. J., 378, 705.
10. Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. and Turk, V. 1988, EMBO J., 7, 2593.
11. Lindahl, P., Alriksson, E., Jornvall, H. and Björk, I. 1988, Biochemistry, 27, 5074.
12. Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. and Turk, V. 1990, EMBO J., 9, 1939.
13. Martinez, M., Diaz-Mendoza, M., Carrillo, L. and Diaz, I. 2007, FEBS Lett., 581, 2914.
14. Pirovani, C. P., da Silva Santiago, A., dos Santos, L. S., Micheli, F., Margis, R., da Silva Gesteira, A., Alvim, F. C., Pereira, G. A. and de Mattos Cascardo, J. C. 2010, Planta, 232, 1485.
15. Kouzuma, Y., Inanaga, H., Doi-Kawano, K., Yamasaki, N. and Kimura, M. 2000, J. Biochem., 128, 161.
16. Wu, J. and Haard, N. F. 2000, Comp. Biochem. Physiol. C Toxicol. Pharmacol., 127, 209.
17. Girard C., Rivard, D., Kiggundu, A., Kunert, K., Gleddie, S. C., Cloutier, C. and Michaud, D. 2007, New Phytol., 173, 841.
18. Walsh, T. A. and Strickland, J. A. 1993, Plant Physiol., 103, 1227.
19. Annadana, S., Schipper, B., Beekwilder, J., Outchkourov, N., Udayakumar, M. and Jongsma, M. A. 2003, J. Biosci. Bioeng., 95, 118.
20. Nissen, M. S., Kumar, G. N., Youn, B., Knowles, D. B., Lam, K. S., Ballinger, W. J., Knowles, N. R. and Kang, C. 2009, Plant Cell, 21, 861.
21. Weeda, S. M., Mohan Kumar, G. N. and Richard Knowles, N. 2009, Planta., 230, 73.
22. Brömme, D. 2001, Curr. Protoc. Protein Sci., 21, unit 21.2.
23. Vernet, T., Khouri, H. E., Laflamme, P., Tessier, D. C., Musil, R., Gour-Salin, B. J., Storer, A. C. and Thomas, D. Y. 1991, J. Biol. Chem., 266, 21451.
24. Baker, E. and Drenth, J. 1987, Biological Macromolecules and Assemblies, Junak, F. McPherson eds., 3, 313.
25. Brocklehurst, K., Willenbrock, F. and Salih, E. 1987, Hydrolytic Enzymes, Neuberger, A. and Brocklehurst, K. (Eds.), Elsevier, Amsterdam, 39.
26. Kurata, M., Yamamoto,Y., Watabe, S., Makino, Y., Ogawa, K. and Takahashi, S. Y. 2001, J. Biochem., 130, 857.
27. Santamaria, I., Velasco, G., Pendas, A. M., Paz, A. and Lopez-Otin, C. 1999, J. Biol. Chem., 274, 13800.
28. Shi, G. P., Bryant, R. A., Riese, R., Verhelst, S., Driessen, C., Li, Z., Brömme, D., Ploegh, H. and Chapman, H. A. 2000, Exp. Med., 191, 1177.
29. Thompson, J. D., Higgins, D. G. and Gibson, T. J. 1994, Nucleic Acids Res., 22, 4673.
30. Felsenstein, J. 1989, Cladistics, 5, 164.