Available online at www.sciencedirect.com
SCIENCE DIRECT®
Phytochemistry 65 (2004) 3003–3020
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

The distribution of serine proteinase inhibitors in seeds of the Asteridae

Alexander V. Konarev [a,*], Jonathan Griffin [b], Galina Yu. Konechnaya [c], Peter R. Shewry [b]

[a] *All-Russian Institute of Plant Protection (VIZR), Podbelsky 3, St. Petersburg 196608, Russia*
[b] *Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*
[c] *V.L. Komarov Botanical Institute, Prof. Popova 2, St. Petersburg 197376, Russia*

Received 14 March 2004; received in revised form 21 July 2004
Available online 8 October 2004

Abstract

The Asteridae is one of the most successful clades of flowering plants comprising some 80,000 species. Despite this diversity, analysis of seeds from 398 species (representing 8 orders, 32 families and 181 genera) showed just two major types of serine proteinase inhibitors (PI). PIs of the potato inhibitor I family were widely distributed. These had M_r of 7000–7500 and were inhibitory to subtilisin and one or more other proteinases (but only rarely elastase). The second major group was TI related to the well-characterised Bowman–Birk inhibitors of legume seeds but these varied widely in their sequences and structure. In addition to these two groups of inhibitors, seeds of the Solanaceae also often contained PI of the potato inhibitor II family while some other asterids contained inhibitors whose relationships were not established.

Keywords: *Anthriscus sylvestris*; Apiaceae; *Veronica hederifolia*; Scrophulariaceae; Asteridae; Trypsin inhibitor; Subtilisin inhibitor; Chymotrypsin inhibitor; Seeds

1. Introduction

Protein and peptide inhibitors of various exogenous (from invertebrates, viruses, fungi, mammals) and endogenous proteinases are widespread in seeds and may prove to be universal. Their main functions are thought to be in plant defence and the regulation of endogenous proteinases (Shewry and Lucas, 1997; Mosolov et al., 2001; Birk, 2003) but they may also function as storage proteins (Shibata et al., 1988; Shewry, 2003). They are of interest as potential sources of resistance against pests and pathogens in transgenic plants (Ryan, 1990; Gutierrez-Campos et al., 1999; Urwin et al., 1998; Lawrence and Koundal, 2002) and as drugs with antiviral and other properties (Korsinczky et al., 2004) as well as providing markers for studies of plant diversity and evolution (Konarev, 1988, 1996; Konarev et al., 2002b; Kollipara and Hymowitz, 1992).

Some 12 families of inhibitors can be recognised based on their amino acid sequences and target proteinases (Shewry, 1999). However, most studies have been carried out on crop plants (cereals, legumes and solanaceous species) with economically unimportant species being comparatively neglected. We have recently carried out an extensive survey of 65 genera of the Compositae,

Abbreviation: CI, chymotrypsin inhibitor; EI, elastase inhibitor; TI, trypsin inhibitor; SI, subtilisin inhibitor; T/C/E/SI, trypsin/chymotrypsin/elastase/subtilisin inhibitor; TLGF, thin layer gel-filtration; RP-H-PLC, reversed-phase HPLC.

* Corresponding author. Tel.: +7 812 465 6617/466 4704; fax: +7 812 465 6617.
E-mail addresses: al_konarev@hotmail.com (Al.V. Konarev), peter.shewry@bbsrc.ac.uk (P.R. Shewry).

0031-9422/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2004.08.022

analysing seed and vegetative tissues for inhibitors (I) of trypsin (T), chymotrypsin (C) and subtilisin (S) (Konarev et al., 1999a, 2000, 2002a). The results included the identification in sunflower seeds of a novel cyclic 14 amino peptide which is the smallest and most potent trypsin inhibitor from a plant source (Konarev et al., 2002a; Luckett et al., 1999). This inhibitor has since been used to design synthetic analogues in order to modify the activity and specificity for biomedical applications (Korsinczky et al., 2004). It is probable that SFTI-1 was responsible for the reported inhibition of *Helicoverpa amigera* (Bhat et al., 1996). However, SFTI-1 has only been detected in species of *Helianthus* (including sunflower and Jerusalem artichoke) and in the related species, *Tithonia diversifolia* (Mexican sunflower), with the most widely distributed proteinase inhibitors in the Compositae being related to the potato inhibitor I family with activity against both bacterial subtilisin and subtilisin-like proteinases from the plant pathogenic fungi including *Sclerotinia sclerotiorum* (Konarev et al., 1999b, 2002a).

It is clear that plant seeds, and particularly those of non-crop species, remain important sources of novel proteinase inhibitors. We have therefore extended our published studies of the Compositae to a wider group of related species, the Asteridae. The Asteridae is one of the largest and most successful clades of flowering plants, comprising more than 100 families and about 80,000 species (about 25–30% of all flowering plants), but only two of these families have been previously analysed for proteinase inhibitors (the Solanaceae and Compositae). The Asteridae also contains four of the 10 largest families of flowering plants: the Compositae (Asteraceae), Rubiaceae, Lamiaceae and Apocynaceae, and has been recognised as a natural grouping since the 18th century with recent confirmation coming from molecular analyses (Bremer et al., 1998, 2002). We have therefore analysed almost 400 species selected from eight of the 10 known orders of asterids, including economically important food, feed, oil, medicinal and decorative species such as campanula, honeysuckle, teasel, valerian, dill, olive, sesame, plantain, veronica, madder, coffee, sweet pepper and rhododendron.

2. Results

2.1. Distribution of serine proteinase inhibitors

The distribution and specificity of proteinase inhibitors in seeds of the asterids were determined using the "gelatin replicas" method. In this method the inhibitors are separated by isoelectric focusing and then transferred to a layer of gelatin and incubated with the target proteinase (trypsin, chymotrypsin, subtilisin, elastase). The presence of an inhibitor prevents digestion of the gelatin resulting in an undigested gelatin "island". The molecular masses of the inhibitors were estimated by comparison of their relative mobility (Rm) on thin layer gel filtration (TLGF) with the mobilities of standard proteins. The results are summarised in Figs. 1 and 2 and Table 1.

2.1.1. Order Asterales

2.1.1.1. Family Compositae. Numerous members of the sub-family Carduoideae were analysed in this and previous studies (Konarev et al., 2002a). The major inhibitors were T/(C)/SI with relative mobilities corresponding to masses of ≈7500 (see Table 1). SI were rarely present (Fig. 1 and Table 1) while representatives of several genera contained no detectable inhibitors. Species of the Cichorioideae all had T/SI ($M_r \approx 7500$) or TI. Members of the Asteroideae (e.g. *Aster*, *Zinnia*, *Kalimeris*, *Gaillardia*, *Silphium*) tended to have high levels of TI of $M_r \approx 10{,}000$–$14{,}000$ with many also containing T/SI and C/SI of $M_r \approx 7500$ (see Table 1 and Konarev et al., 2002a).

However, although a total of 80 Compositae genera have been studied, low M_r TI related to SFTI-1 were only detected in species of *Helianthus* and *Tithonia*.

2.1.1.2. Families Campanulaceae and Lobeliaceae. Most species of these two families (including *Campanula* spp., *Jasione* spp., and most *Lobelia* spp.) contained only weak T/SIs varying in pI from ≈7.8 to 9.8 (Fig. 1, tracks 12–17), with *Lobelia sessilifolia* exhibiting a weak SI band (Fig. 1, track 18).

2.1.2. Order Dipsacales

2.1.2.1. Family Caprifoliaceae. Seven species of *Lonicera* (honeysuckle) had similar T/weakC/SI bands of pI ≈ 7.9 and 9.4 while *L. edulis* had two SI bands (Fig. 1, tracks 20 and 21, respectively). Six other species of *Lonicera* contained no detectable inhibitors (Fig. 1, track 19). Species of *Sambucus* (elder) varied in the presence and composition of T/C/SI with the major bands in *S. coreana* having similar pI to those in *Lonicera* (tracks 22–24). *Symphoricarpos occidentalis* (snowberry) contained T/SI of pI ≈ 4.5 (Fig. 1, track 26) while no inhibitors were detected in species of *Viburnum*, *Triosteum* and *Weigela.*

2.1.2.2. Family Dipsacaceae. Most species contained T/(C)/SI but these varied between genera and, in *Dipsacus* (teasel) and *Scabiosa* (scabious), between species (Fig. 1, tracks 27–35). Some *Cephalaria* and *Dipsacus* species contained high pI (≈10) TSI (Fig. 1, tracks 27–32) and some *Scabiosa* spp. also contained SI. Molecular masses

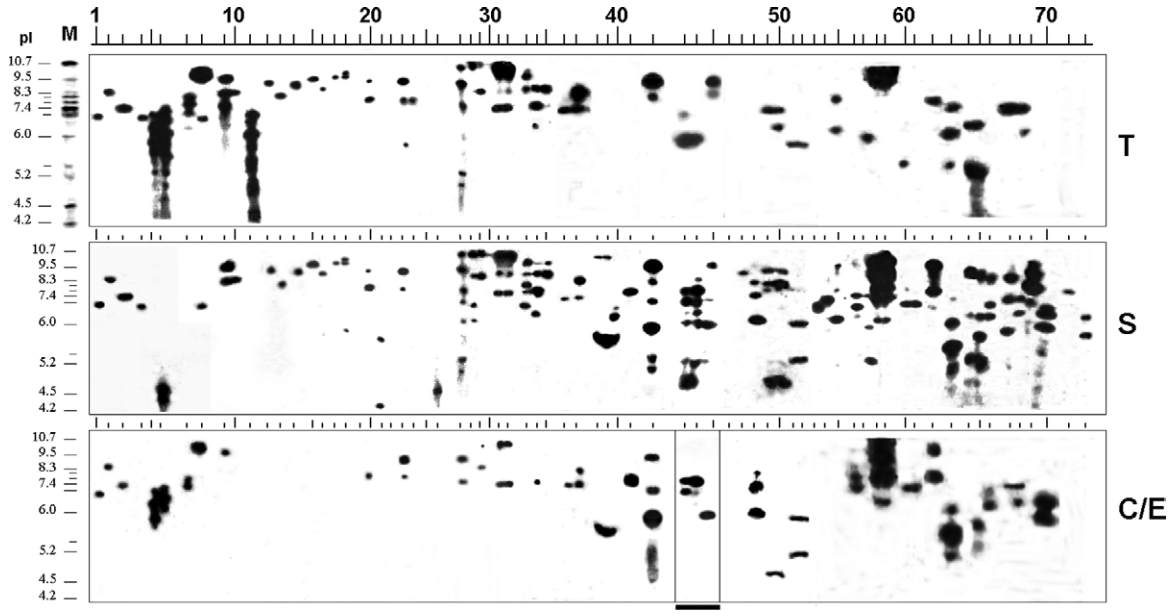

Fig. 1. Polymorphism of proteinase inhibitors in seeds of some representatives of orders **Asterales** (**1**–**18**), **Dipsacales** (**19**–**37**) and **Apiales** (**38**–**72**). Proteins extracted with water were loaded on gels in volumes 0.3–4 μl (according to activity of inhibitors) and separated by IEF in the pH range 3–10. Inhibitors were detected in four gelatin replicas made from each gel and developed with trypsin (T), subtilisin (S), chymotrypsin (C) and elastase (underlined tracks in lower C/E row). **Compositae:** Subfamily Carduoideae: **1**, *Carthamus tinctorius*. **2**, *Centaurea pseudophrygia*. **3**, *Jurinea arachnoidea*. Subfamily Cichorioideae: **4**, *Scorzonera tau-saghys*. **10**, *Andryala agardhii*. Subfamily Asteroideae: **5**, *Kalimeris incisa*. **6**, *Aster alpinus*. **11**, *Crinitaria linosyrus*. **7**, *Madia sativa*. **8**, *Helianthus annuus*. **9**, *Adenostyles glabra*. **Campanulaceae: 12**, *Campanula barbata*. **13**, *Campanula latifolia*. **14**, *Edraianthus graminifolius*. **15**, *Jasione montana*. **16**, *Phyteuma nigrum*. **Lobeliaceae: 17**, *Lobelia inflata*. **18**, *Lobelia sessilifolia*. **Caprifoliaceae: 19**, *Lonicera caucasica* **20**, *Lonicera kamtschatica*. **21**, *Lonicera edulis*. **22**, *Sambucus canadensis*. **23**, *S. coreana*. **24**, *S. racemosa*. **25**, *Symphoricarpos albus*. **26**, *S. occidentalis*. **Dipsacaceae: 27**, *Cephalaria gigantea*. **28** and **29**, *Dipsacus sp*. **30**, *D. fullonum*. **31**, *D. laciniatus*. **32**, *D. sativus*. **33**, *Scabiosa atropurpurea*. **34**, *S. bipinnata*. **35**, *S. graminifolia*. **Valerianaceae: 36**, *Valeriana phu*. **37**, *V. officinalis*. **Apiaceae: 38**, *Ammi majus*. **39**, *A. visnaga* vr4. **40**, *A. visnaga* vr6. **41**, *Anethum graveolens*. **42** and **45**, *Anthriscus cerefolium k-9* and *k-6*. **43**, *Heracleum mantegazzianum*. **44**, *H. pubescens*. **46**, *Astrantia biebersteinii*. **47**, *Athamanta matthioli*. **48**, *Bupleurum aureum*. **49** and **50** *Carum ajovan k-2* and *k-1*. **51** and **52**, *Carum carvi BI and k-45*. **53**, *Chaerophyllum bulbosum*. **54**, *Coriandrum sativum* k-54. **55**, *Cryptotaenia canadensis*. **56** and **57**, *Cuminum cyminum* vr68 and vr86. **58** and **59**, *Daucus sativus*, cv. Shantene 2461 and cv. Nantskaya 4. **60** and **61**, *Foeniculum vulgare* k-76 and k-38. **62**, *Laser trilobium*. **63**, *Pastinaca sativa*. **64**, *Pimpinella anisum*. k-16. **65**, *P. major*. **66**, *Prangos ferulacea*. **67**, *Seseli libanotis*. **68**, *S. peucedanoides*. **69**, *Sium sisarum*. **70**, *Visnaga daucoides*. **Araliaceae: 71**, *Eleutherococcus sessiliflorus*. **Pittosporaceae: 72**, *Pittosporum crassifolium*. M, marker proteins with pI in range 3.5–10.7 stained with Coomassie Blue.

were estimated as ≈7500 for T/SI from *Scabiosa lucida* but ≈4500 for T/C/SI from *D. sativus*.

2.1.2.3. Family Valerianaceae. Three species of *Valeriana* contained T/C/SI of pI ≈ 7.4 and $M_r \approx 7500$ while *V. officinalis* had an additional strong band of pI ≈ 8.3 (Fig. 1, track 37). No inhibitors were detected in *Patrinia gibbosa.*

2.1.3. Order Apiales

2.1.3.1. Family Apiaceae. The distribution of inhibitors in the Apiaceae was consistent with the division of this family into clades based on molecular data (Downie et al., 2000). The "Apium" clade (*Ammi*, *Visnaga*, *Anethum*, *Foeniculum*, and *Prangos* spp.) contained mainly CSI ($M_r \approx 6000$) and SI with *Pimpinella* also having TSI. CSI ($M_r \approx 7500$), TSI ($M_r \approx 7500$) and SI also occurred in members of the "Angelica" clade including coriander, parsnip and other spices, forage and vegetable crops (Table 1 and Fig. 1, tracks 54, 63). Elastase inhibitors were also detected in one accession of *Heracleum mantegazzianum* (T/C/E/SI) and in *H. pubescens* (C/S/SI) (Fig. 1, tracks 43 and 44). *Heracleum* spp. also contained strong SI with pI ≈ 4.7. Carrot and other members of the "Daucus" clade contained highly active T/C/SI and C/SI of $M_r \approx 7500$ (Fig. 1, track 45). Members of other clades ("Aegepodium": *Carum* spp.; "Oenanthe": spp.; "Bupleurum": *Bupleurum* spp.) also contained TSI and/or CSI with *Sium sisarum* (clade "Oenanthe") also containing SI of $M_r \approx 7500$ (Fig. 1, tracks 48–52). The five species studied from the subfamily Saniculoideae (*Astrantia*, Fig. 1, track 46, and *Eryngium*) contained no detectable inhibitors.

2.1.3.2. Families Araliaceae and Pittosporaceae. Weak SIs were detected in some *Eleutherococcus* and *Aralia* spp. with two SI bands in *Pittosporum crassifolium* (Fig. 1, tracks 71 and 72).

2.1.4. Order Lamiales

2.1.4.1. Family Acanthaceae. Crabbea reticulata contained strong TI and heterogeneous T/SI of $M_r \approx 4000$ (Fig. 2, track 1) while *Dicliptera resupinata* contained T/SI together with a T/C/E/SI of pI ≈ 7.8 (Fig. 2, track 2).

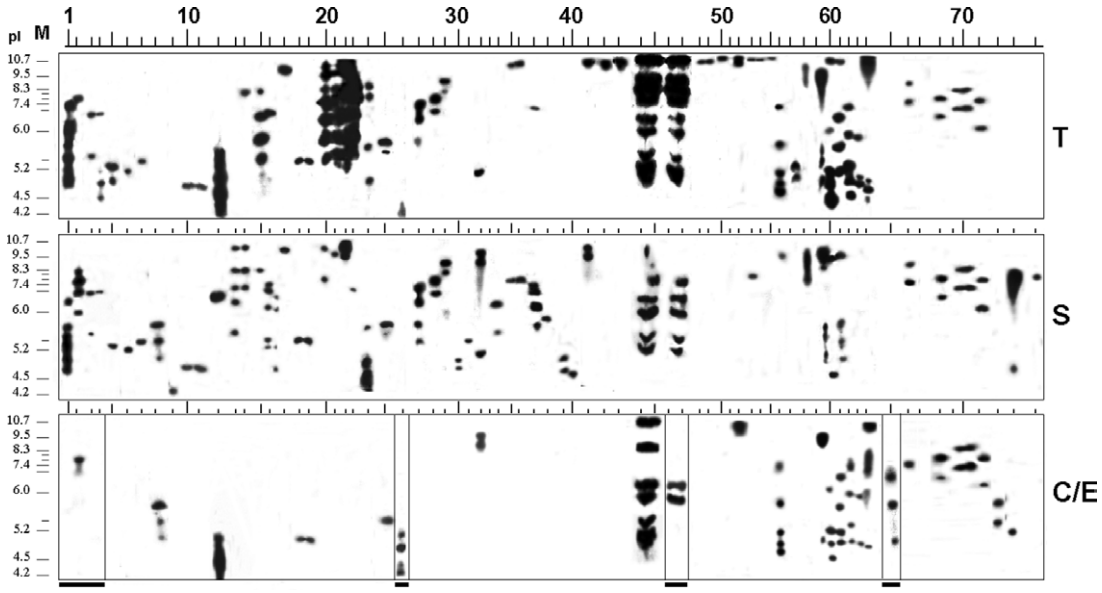

Fig. 2. Polymorphism of proteinase inhibitors in seeds of some representatives of orders **Lamiales (1–26), Gentianales (27–38), Solanales (39–65)** and **Ericales (66–74)** (conditions as in Fig. 1). **Acanthaceae: 1**, *Crabbea reticulata*. **2**, *Dicliptera resupinata*. **Bignoniaceae: 3**, *Catalpa bignonioides*. **4**, *C. ovata*. **5**, *Incarvillea delavayi*. **Lamiaceae: 6**, *Agastache rugosa*. **7**, *Betonica officinalis*. **8**, *Leonurus quinquelobatus*. **Oleaceae: 9**, *Syringa vulgaris*. **Pedaliaceae: 10**, *Sesamum alatum* k-1355. **11**, *S. indicum* k-1212. **Globulariaceae: 12**, *Globularia punctata*. **Scrophulariaceae: 13**, *Digitalis grandiflora*. **14**, *D. lutea*. **15**, *Paederota bonarota*. **16**, *Penstemon congestus*. **17**, *Tetranema mexicana*. **18**, *Verbascum bactriamum*. **19**, *V. thapsus*. **20** and **21**, *Veronica gentianoides*. **22**, *V. prostrata*. **23**, *V. virginica*. **Verbenaceae: 24**, *Lantana kisi*; **25**, *L. camara*; **26**, *Callicarpa americana*. **Apocynaceae: 27**, *Acokanthera oppositifolia*. **28**, *Carissa bispinosa*. **29**, *Rauwolfia tetraphylla*. **Gentianaceae:30**, *Gentiana purpurea*. **31**, *Gentiana septemfida*. **Rubiaceae: 32**, *Borreria verticillata*. **33**, *Coffea canephora*. **34**, *Ophiorrhiza mungos*. **35**, *Psychotria bacteriophila*. **36**, *Psychotria maingayi*. **37**, *Rubia cordifolia*. **38**, *Tarenna asiatica*. **Boraginaceae:39**, *Cordia myxa*. **40**, *Onosma stellulatum*. **Solanaceae: 41**, *Solanum dulcamara*; **42**, *S. integrifolium* k-915; **43**, *S. melongena* k-747. **44**, **45** and **47**, *Capsicum annuum* BI, k-1522 and k-758; **46**, *C. pubescens* k-6690 (tracks 46 and 47 in lower row were developed by elastase) **48**, *Lycopersicon esculentum* var. *elongatum*; **49**, var. *acemigerum*; **50**, *Physalis pubescens* k-1; **51** and **52**, *P. alkekengi* (wild) k-139 and *k-135*. **53** and **54**, *Petunia* x *atkinsiana* cv.Ratsher and cv. Royal Pers. **55**, *Nicotiana rustica*. **56**, *Anisodus stramonifolius*. **57**, *Atropa belladonna*; **58**, *A. belladonna* var. *lutea*. **59**, *Datura stramonium*. **60**, *Hyoscyamus niger*. **61** and **65**, *Physochlaina orientalis*. **62**, *Scopolia caucasica*; **63**, *S. tangutica*. **64**, *Withania frutescens*. **Ericaceae:66**, *Erica tetralix*. **67**, *Rhododendron canadense*; **68**, *R. ferrugineum*. **69**, *Vaccinium myrtillus*; **70**, *V. uliginosum*; **71**, *V. vitis-idaea*. **Polemoniaceae: 72**, *Polemonium caeruleum*; **73**, *Phlox drummondii*. **Primulaceae: 74**, *Primula japonica*.

2.1.4.2. Families Bignoniaceae and Gesneriaceae. These families contained only T/SI (Fig. 2, tracks 3–5).

2.1.4.3. Family Lamiaceae. Many species (including *Perilla*, *Lavandula*, *Hyssopus* and *Ocimum*) contained no detectable inhibitors while others showed only low activity (Fig. 2 and Table 1). The latter included T/SI in two out of nine accessions of *Lallemantia iberica*, traces of TI or SI in sage (*Salvia*) and other members of the subfamily Nepetoideae (*Scutellaria*, *Thymus*, *Zizi-phora*) and C/SI and SI in *Leonurus quinquelobatus* and *Sideritis taurica* (subfamily Lamioideae).

2.1.4.4. Families Oleaceae, Orobanchaceae and Phrymaceae. Two species of *Forsythia* differed in the presence of weak T/SI of pI ≈ 6.3 while *Ligustrum vulgare* (privet) and *Olea* (olive) spp. had no detectable inhibitors. Two out of three species of *Syringa* (lilac) had weak SI with pI ≈ 4.3 and 4.4 (Fig. 2, tracks 9 for *S. vulgaris*). Single members of the Orobanchaceae, *(Aeginetia indica)*, and Phrymaceae (*Phryma leptostachya*) contained no detectable inhibitors.

2.1.4.5. Family Pedaliaceae. Eight accessions of two species of sesame (*Sesamum alatum* and *S. indicum*) were studied, originating from various countries in Africa and Asia. All had T/SI with pI ≈ 4.7 (Fig. 2, tracks 10 and 11).

2.1.4.6. Families Plantaginaceae and Globulariaceae. Plantago (plaintain) species all possessed heterogeneous active TI with pI ranging from ≈6.0–8.0. The major components in *P. psyllum* had M_r of ≈4500 while *Globularia punctata* had heterogeneous T/CI with pI of 4.6–5.6 and a single SI with pI ≈ 6.8 and M_r ≈ 7500 (Fig. 2, track 12).

2.1.4.7. Family Scrophulariaceae. Heterogenous mixtures of SI were present in species of *Digitalis* (foxglove), where they varied in pI from 5.4 to 10 (Fig. 2, tracks 13–14), and in *Penstemon confestus* (track 16) and *Scrophularia* (figwort). *D. lutea* and *P. confestus* also contained single T/SI bands but *D. thaspi* and *Lathraea squamaria* did not contain any clearly resolved inhibitors. Seeds of *Ervinus alpinus* contained three T/SI with pI ranging from 8.3 to 10 while *Tetranema mexicana* possessed a

Table 1
Distribution of serine proteinase inhibitors in seeds of the Asteridae

Order, Family, Subfamily, *Genus*, *Species* (examples)	Number of		Proteinase inhibitors				
	Species/acc	Genera		S/			
			T/(C)	T/(C)	C	E/(C)	–
Asterales[1&2]	166	88	+/–	+/–	+/–	–	+/–
Compositae[1&2]	153	81	+/–	+/–	+/–	–	+/–
Carduoideae[1&2]	32	12	–	+/–	–	–	+/–
Carthamus tinctorius[1]			–	5*	–	–	–
Cichorioideae[1&2]	21	12	+/–	+/–	+/–	–	–
Taraxacum officinale[1]			–5*	5*	–	–	
Lactuca sativa[1]			+	+	–	–	–
Asteroideae[1&2]	101	59	+/–	+/–	+/–	–	–
Aster alpinus			7*!	5*	–	–	–
Helianthus annuus[1&2]		0*!	5*	–	–	–	
Cosmos bipinnatus[1&2]		–	5*	5*	–	–	
Zinnia elegans[1&2]			7*!	+	–	–	–
Campanulaceae + Lobeliaceae	13	7	–	w	–	–	–
Campanula barbata L.			–	w	–	–	–
Dipsacales	53	13	w/–	+/–	–	–	–
Caprifoliaceae	34	7	w/–	+/–	–	–	–
Lonicera kamtschatica			–	!	–	–	–
Sambucus coreana			–	!	–	–	–
Viburnum sp.	3		–	–	–	–	–
Dipsacus sativus			–	3*!	–	–	–
Scabiosa lucida			–	5*!	–	–	–
Valerianaceae	4	2	–	+	–	–	–
Valeriana officinalis			–	5*	–	–	–
Apiales	47	26	+/–	+/–	+/–	+/–	+/–
Apiaceae	38	22	+/–	+/–	+/–	+/–	+/–
Ammi majus			–	–	4*!	–	–
Foeniculum vulgare			–	–	4*	–	–
Pimpinella anisum			–	5*!	–	–	+
Visnago daucoides			–	–	5*!	–	–
Heracleum mantegazzianum	2		–	5*/–	5*/–	+/–	!/–
Pastinaca sativa			–	5*!	5*!	–	–
Seseli libanotis	2		–	+/–	5*/–	–	+/–
Carum ajovan			–	!	!	–	–
Daucus sativus			–	5*	5*!	–	–
Laser trilobium			–	5*	5*	–	–
Anthriscus cerefolium	2		–	!	!	5*!/–	–
Anthriscus sylvestris	2		–	!/–	–	!	–
Sium sisarum.			–	w	–	–	5*/–
Bupleurum aureum			–	–	!	–	–
Subfam. Saniculoideae: *Astrantia* sp.	5		–	–	–	–	–
Araliaceae	8	3	–	–	–	–	w/–
Acanthopanax sessiliflorus			–	–	–	–	w
Aralia sp.	5		–	–	–	–	w/–
Pittosporaceae, *Pittosporum* sp.			–	–	–	–	+
Lamiales	97	52	+/–	+/–	+/–	+/–	+/–
Acanthaceae	3	3	+/–	+	–	+/–	–
Crabbea reticulate			!	3*!	–	–	–
Dicliptera resupinata			–	6*	–	6*	–
Bignoniaceae	2	3	–	+	–	–	–
Gesneriaceae	5	4	–	+/–	–	–	–
Lamiaceae	41	22	w/–	+/–	+/–	–	+/–
Betonica sp.	3		–	+/–	–	–	+/–
Calamintha nepeta			–	w	–	–	–
Coleus blumei			–	–	–	–	–
Hyssopus officinalis			–	–	–	–	–
Lallemantia iberica, acc.	9		–	–	–	–	–
Lallemantia iberica, acc.	2		–	w	–	–	–
Lavandula angustifolia			–	–	–	–	–
Phlomis tuberose			–	+	–	–	–

(*continued on next page*)

Table 1 (*continued*)

Order, Family, Subfamily, *Genus*, *Species* (examples)	Number of		Proteinase inhibitors				
	Species/acc	Genera		S/			
			T/(C)	T/(C)	C	E/(C)	–
Salvia sp.	4		–	+/–	–	–	+/–
Scutellaria woronowii			–	+	–	–	–
Sideritis taurica			–	–	–	–	!
Thymus serpyllum			w	–	–	–	–
Ziziphora bvericalyx			w	–	–	–	–
Oleaceae	10	5	–	+/–	–	–	+/–
Forsythia sp.	2		–	+/–	–	–	+/–
Olea sp.	2		–	–	–	–	–
Syringa sp.	3		–	–	–	–	w/–
Orobanchaceae, *Aeginetia indica*			–	–	–	–	–
Pedaliaceae, *Sesamum* sp.	2	1	–	+	–	–	–
Phrymaceae, *Phryma leptostachya*			–	–	–	–	–
Plantaginaceae	3	1	+	–	–	–	–
Plantago sp.	3		+	–	–	–	–
Plantago psyllium			3*	–	–	–	–
Globulariaceae, *Globularia punctata*			!	–	–	–	5*
Scrophulariaceae	21	8	+/–	+/–	–	–	w/–
Digitalis sp.	3		–	w/–	–	–	w/–
Paederota bonarota			2*!	w	–	–	–
Scrophularia nodosa			–	–	–	–	w
Tetranema sp.	2		–	4*	–	–	–
Verbascum sp.	4		–	4*	–	–	–
Veronica sp.	7		!/–	+	–	–	–
Veronica hederifolia			3*!	+	–	–	–
Veronica persica			–	!	–	–	–
Verbenaceae	6	3	w/–	+/–	+/–	+/–	–
Lantana sp.	4		–	5*/–	+/–	+/–	–
Callicarpa Americana			w	–	–	EI	–
Gentianales	40	17	+/–	+/–	+/–	–	+/–
Apocynaceae	7	5	–	+	–	–	+
Acokanthera oppositifolia			–	5*	–	–	+
Rauwolfia sp.	2		–	+	–	–	–
Gentianaceae	20	2	–	–	–	–	+/–
Gentiana sp.	13		–	–	–	–	+/–
Swertia iberica			–	–	–	–	+
Rubiaceae	13	10	+/–	+/–	+/–	–	+/–
Borreria verticillata			–	+	+	–	–
Coffea sp.	2		–	–	–	–	–
Gardenia thunbergia			–	–	–	–	+
Ophiorrhiza mungos			–	–	–	–	5*!
Psychotria sp.	2		3–4*	–	–	–	5*
Rubia cordifolia			–	+	–	–	5*!
Solanales	46	25	+/–	+/–	–	+/–	w/–
Boraginaceae	9	9	–	–	–	–	w/–
Myosotis sylvatica			–	–	–	–	–
Cordia myxa			–	–	–	–	w
Hydrophyllaceae, *Phacelia tanacetifolia*			–	!	–	–	–
Solanaceae	47	16	+/–	+/–	–	+/–	–
Capsicum sp.	5		^!	+/–	–	TCESI!	–
Lycopersicon sp.	5		^w	–	–	–	–
Schizanthus sp.			–	–	–	–	–
Solanum sp.	16		^/–	–	–	–	–
Solanum auriculatum			^	–	–	–	–
Physalis sp.	3		^/–	–	–	–	–
Petunia sp.	2		^	–	–	–	–
Nicotiana rustica			–	–	–	–	–
Anisodus stramonifolius			^	–	–	–	–
Atropa belladonna	2		w/–	+	–	–	–
Datura stramonium			–	!	–	–	–
Hyoscyamus niger			^	+	–	–	–
Physochlaina orientalis			–	–	–	4*	–

Table 1 (*continued*)

Order, Family, Subfamily, *Genus*, *Species* (examples)	Number of		Proteinase inhibitors				
	Species/acc	Genera		S/			
			T/(C)	T/(C)	C	E/(C)	–
Scopolia sp.	2		+	–	–	–	–
Withania frutescens			^	–	–	–	–
Cornales, Cornaceae, *Swida* sp.	2	1	–	–	–	–	–
Ericales	62	19	–	+/–	–	–	w/–;
Actinidiaceae	2	1	–	–	–	–	–
Ebenaceae, Diospyros ebenum	1	1	–	–	–	–	–
Ericaceae	59	17	–	+/–	–	–	w/–
Erica tetralix			–	w	–	–	–
Rhododendron sp.	13		–	+/–	–	–	w/–
R. x *intermedium*			–	5*	–	–	–
Vaccinium sp.	3		–	+	–	–	–
Myrsinaceae	5	2	–	–	–	–	–
Polemoniaceae	6	2	–	CI/–	–	–	!/–
Polemonium caucasicum			–	CI	–	–	–
Phlox drummondii			–	–	–	–	7*!
Primulaceae	19	5	–	–	–	–	+/–
Primula sp.	12		–	–	–	–	+/–

[1] Konarev et al. (2002a).

[2] Present work; "+" or "–", presence or absence of detectable inhibitor bands in IEF spectrum of seed proteins; "+/–", variability within taxa in the presence or absence of inhibitor; "^", inhibitor with pI above 9.5; "!" and "w", indicate highly active or weak inhibitor bands respectively (see Section 5). EI or CI, inhibitor of elastase or chymotrypsin only. 0–9* are relative mobilities (Rm) compared to those of marker proteins of known molecular mass (see Section 5). They corresponded approximately to the following masses (Rm for marker proteins (see Section 5.2.2) are shown in bold): **0**, $M_r \approx$ **1500 1**, $M_r \approx$ 2300 **2**, $M_r \approx$ **3000 3**, $M_r \approx$ 4500 **4**, $M_r \approx$ 6000 **5**, $M_r \approx$ **7500 6**, $M_r \approx$ 9000 **7**, $M_r \approx$ 11,000 **8**, $M_r \approx$ **12,500 9**, $M_r \approx$ 14,000.

single TS/I of pI ≈ 10 (Fig. 2, track 17) and $M_r \approx 6000$. *Paederota bonarota* contained highly active TI with pI ≈ 5.3, 5.8 and 6.8 (Fig. 2, track 15) and low M_r (≈3000–4000), and a single T/SI with pI of ≈8.0. *Verbascum* species had T/C/SI with pI ≈ 5.3 (Fig. 2, tracks 18 and 19) and $M_r \approx 6000$ while six out of seven species of *Veronica* (speedwell) had active heterogenous TI with pI ranging from ≈5.3–10.5 (Fig. 2, tracks 20–23) and $M_r \approx 4500$. All *Veronica* species also had T/SI with pI ranging from ≈7 to 10.

2.1.4.8. Family Verbenaceae. Species of *Lantana* varied in the presence or absence of weak inhibitors of T, C and S. Seeds of *L. kisii* contained TI of pI ≈ 7.5 and 8.3, SI with pI below 5.0 and T/SI of $M_r \approx 7500$ while *L. camara* contained active TSI with pI ≈ 5.8 (Fig. 2, tracks 24 and 25). *Callicarpa americana* (Fig. 2, track 26) was unusual in that it contained an elastase inhibitor which did inhibit other serine proteinases. Although *Callicarpa* has recently been placed in the Lamiaceae based on molecular markers (NCBI, 2003), the other species analysed from this family also lacked inhibitors of elastase (see above).

2.1.5. Order Gentianales

2.1.5.1. Family Apocynaceae. TSI with pI ≈ 5.5–7.8 and $M_r \approx 7500$ were present in *Acokanthera oppositifolia* and *Carissa bispinosa* (Fig. 2, tracks 27 and 28) while two *Rauwolfia* spp. had TSI of pI ≈ 9.4. *R. tetraphylla* also contained a TSI of pI ≈ 8.3 (Fig. 3, track 29).

2.1.5.2. Family Gentianaceae. Eight species of *Gentiana* (gentian) had no detectable inhibitors while five had weak SI with pI ≈ 4.8–5.3 (Fig. 2, tracks 30–31). *Swertia iberica* contained SI of pI ≈ 4.8.

2.1.5.3. Family Rubiaceae. Members of this family varied in the presence and composition of inhibitors. *Borreria verticillata* had one TI and two CSI, with pI of ≈5.2 and ≈9.0, respectively (Fig. 2, track 32). *Ophiorrhiza*

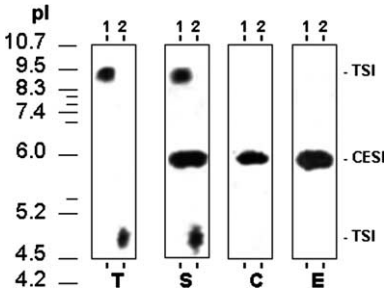

Fig. 3. Heterogeneity of proteinase inhibitors from seeds of *Anthriscus cerefolium* (1) and *A. sylvestris* (2). Water-soluble seed proteins were separated by IEF in Servalyt Precotes pH 3–10 gels. Four gelatin replicas were obtained from the same gel were developed with trypsin (T), subtilisin (S), chymotrypsin (C) and elastase (E). T/SI and C/E/SI, inhibitor types.

mungos contained an SI of pI ≈ 6.5 while *Psychotria* spp. contained TI of pI 10.7 and $M_r \approx 5000$ and SI of pI ≈ 8 (Fig. 3, tracks 34–36) and $M_r \approx 7500$. *Rubia cordifolia* (Indian madder) had weak TSI but strong SI of pH 5.5–7.3 and $M_r \approx 7500$ (Fig. 3, track 37) while *Tarenna asiatica* (Fig. 3, track 38) and *Gardenia thunbergia* contained low pI SI. Other species of the Rubiaceae, including *Asperula scutellaris* and *Coffea* spp., contained no detectable inhibitors.

2.1.6. Order Solanales

2.1.6.1. Families Boraginaceae and Hydrophyllaceae. Although 8 species of the Boraginaceae were analysed, inhibitors (weak SI of pI 4.6–5.1) were only detected in *Cordia myxa* and *Onosma stellulatum* (Fig. 2, tracks 39 and 40). The single studied species of the Hydrophyllaceae, *Phacelia tanacetifolia*, had active TSI of pI ≈ 7.2.

2.1.6.2. Family Solanaceae. Most *Solanum* spp. (including nightshade and aubergine), *Brunfelsia*, *Capsicum* (sweet pepper), *Hyoscyamus* (henbane), *Lycopersicon* (tomato), *Nicotiana* (tobacco), *Petunia*, *Physalis* and *Withania* contained highly alkaline TI, which also inhibited C in some species such as *Physalis* (Fig. 2, track 52). Some nightshade spp. (e.g. *S. dulcamara*, Fig. 2, track 41) also contain alkaline SI (pI > 9.5). However, neutral or weakly acidic T/(C)I were present in some species (*Anisodus stramonifolius*, one *Atropa belladonna* acc, *Scopolia*, some *Solanum* spp. (Fig. 2, tracks 56, 57, 62 and 63)) and TCSI in others (*Datura stramonium*, *Hyoscyamus niger*; Fig. 2, tracks 59 and 60). The widest variation in inhibitor content was in *Capsicum* spp. with TI, T/CI, T/SI and T/C/E/SI (Fig. 2, tracks 44–47). T/C/E/SIs were also present in *Physochlaina orientalis* (Fig. 2, tracks 41 and 45).

2.1.7. Order Cornales

No inhibitors were detected in *Swida alba* and *S. australis* (Cornaceae).

2.1.8. Order Ericales

2.1.8.1. Families Actinidiaceae, Ebenaceae and Myrsinaceae. Weak SI were present in one accession of *Maesa lanceolata* (Myrsinaceae) but no inhibitors were detected in any other accessions of the three families (*Actinidia*, *Diospyros*, *Ardisia*, *Maesa lanceolata*, *M. argentea* acc).

2.1.8.2. Family Ericaceae. Some species and accessions of *Rhododendron* contained T/C/SI of pI ≈ 6.8 and 8.0 (Fig. 2, track 68) while others and a number of other species (*Arctostaphylos uva-ursi*, *Bruckenthalia spiculifolia*, *Chamaedaphne calyculata*, *Gaultheria shallon*, *Ledum groenlandicum*, *Menziesia ferruginea*) contained no detectable inhibitors. *Erica tetralix* (heath), *Vaccinium myrtillus* (bilberry), *M. ulginosum* (blueberry) and *V. vitis idaea* (red bilberry) all contained two T/C/SI of pI of 6.0–9.0 (Fig. 2, tracks 66, 69–71).

2.1.8.3. Families Polemoniaceae and Primulaceae. Polemonium caeruleum and *P. caucasicum* contained weak CI of pI 4.5–5.8 which were unusual in that they did not inhibit other proteinases (Fig. 2, track 72), while *Phlox drummondii* contained highly active SI of pI ≈ 8 and $M_r \approx 9$–10,000 (Fig. 2, track 73). Of the Primulaceae, only one out of 12 *Primula* spp. (*P. japonica*; Fig. 2, track 74) and *Cortusa turkestanica* contained weak SI.

2.2. Purification and characterisation of inhibitors

Several inhibitors were selected for purification based on their distribution, abundance, specificity and molecular mass.

2.2.1. Anthriscus sylvestris (Apiaceae)

A. sylvestris and related species contain a similar C/E/SI band and T/SI bands which vary in their pI (Fig. 3). This species was selected because elastase inhibitors occur only rarely in seeds while T/SI have not been previously characterised from the Apiaceae. An aqueous extract was therefore shaken sequentially with chymotrypsin-Sepharose gel and trypsin-Sepharose gel and absorbed inhibitors eluted as described in 5.2.4. The fractions were then separated by RP-HPLC (Fig. 4) and the eluent monitored for inhibitory activity using IEF and gelatine replicas. The fraction absorbed to chymotrypsin-Sepharose was separated into two inhibitors with pI of ≈6.0 and 7.2 which corresponded to the major protein components revealed by TCA precipitation in the gel. The latter were therefore extracted and analysed by MALDI-TOF MS which gave masses of 7079 (pI 6.0) and 7098 (pI 7.2). The major (pI 6.0) component was also re-separated by HPLC after elution from the IEF gel. Reduction with DTT gave peptides of mass 4517 and 2579. This indicates that the M_r 7098 component was cleaved at the reactive (inhibitory) site during affinity chromatography. This is supported by Edman sequencing which showed homology with sequences adjacent to the PI residues of members of the potato inhibitor I family (Table 2).

RP-HPLC of the fraction eluted from the trypsin-Sepharose column gave a major TSI of M_r 7236 (Fig. 5). Automated Edman degradation failed to give an *N*-terminal sequence so a fraction cleaved with NaI and CNBr (Smith, 1994) was subjected to Q-TOF MS analysis. Peptides of M_r 1452 and 1769 gave overlapping sequences which differed in the absence of three residues from the former and were similar to sequences in potato I inhibitors (Table 2). The presence of tryptophan adjacent to the related sequences in some inhibitors (e.g. *Li-*

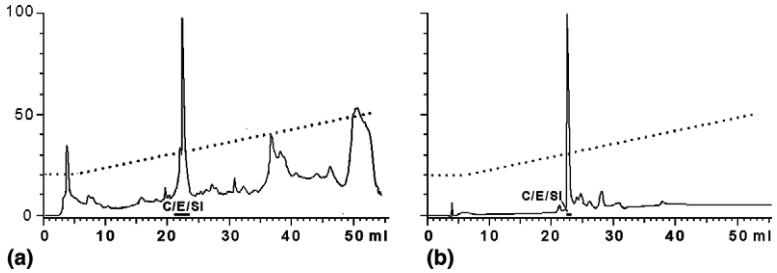

Fig. 4. Purification of elastase (C/E/SI) inhibitor from *Anthriscus sylvestris* seeds. Proteins isolated by affinity chromatography on chymotrypsin-Sepharose were fractionated by RP-HPLC on a Vidac C18 column with a 20–50% acetonitrile gradient (a). Fractions containing C/E/SI were separated by IEF in DryStrips with an immobilised pH 3–10 gradient. Bands with inhibitory activity were excised and subjected to a second RP-HPLC separation (b).

Table 2
Alignment of the amino acid sequences of subtilisin inhibitors isolated from seeds of various members of the Asteridae with those of some members of the potato inhibitor I family

Origin	Inhibitor	Res. Nos.	P		P′																								
			2	*1*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*					*15*					*20*					
Fagopyrum esculentum	PI[1]	44	P	R	D	L	R	C	D	R	V	W	V	F	V	D	E	R	G	V	V	V	D	T	P	V	V	M	
Cosmos bipinnatus	TSI(e)[2]	?	?	?	D	L	R	C	D	R	V	W	V	F	?	?													
Amaranthus hypochondriacus	TI[3]	44	T	K	D	F	R	C	D	R	V	W	V	V	V	D	S	T	G	V	V	V	R	T	P	R	V	T	
Carthamus tinctorius	pept E[2]	1	?	?	D	F	R	C	D	R	V	W	V	W	?	?													
Carthamus tinctorius	SI[2]	?	?	?	D	F	R	C	D	R	V	W	V	W	?	?													
Arabidopsis thaliana	put PI[4]		T	A	D	F	R	C	D	R	V	R	V	F	V	D	G	N	R	I	V	V	K	T	P	K	S	G	
Anthriscus sylvestris	CESI[5]	?	?	?	D	Y	R	(C)	D	R	V	R	V	V	V	N	N	D	G	F	V	V	Q	A	?	?			
Anthriscus sylvestris	#TSI[5]										?	?	V	M	S	V	N	E	A	R	V	V	R	A	P	K	L	T	
Cosmos bipinnatus	CSI(g)[2]	?	?	?	D	Y	R	C	D	R	V	R	V	F	?	?													
Linum usitatissimum	TI[6]		T	K	D	F	R	C	D	R	V	W	V	I	V	N	D	H	G	V	V	T	S	V	P	H	I	T	
Zinnia elegans	PutPI[7]		T	A	D	F	R	C	D	R	V	W	V	W	V	N	S	N	G	V	V	L	R	T	P	S	I	G	
Lycopersicon peruvianum	PI[8]	86	T	K	D	F	R	C	D	R	V	R	L	F	V	N	I	L	G	D	V	V	Q	I	P	R	V	T	
Hordeum vulgare	CI2(CSI)[9]		T	M	E	Y	R	I	D	R	V	R	L	F	V	D	K	L	D	N	I	A	Q	V	P	R	V	G	
Solanum tuberosum	PI[10]	82	T	M	D	Y	R	C	N	R	V	R	L	F	D	N	I	L	G	D	V	V	Q	I	P	R	V	A	
Hirudo medicinalis	E/CtGI[11]	43	T	L	D	L	R	Y	N	R	V	R	V	F	Y	N	P	G	T	N	V	V	N	H	V	P	H	V	
Veronica hederifolia	TSI[5]	?	?	?	D	F	R	P	N	R	V	W	V	W	G	N	D	R	S	V	V	N	A	?	?				
Cucurbita maxima	TI[12]	41	T	K	D	Y	R	P	N	R	V	R	V	F	N	D	D	S	G	K	V	N	S	I	P	R	I	G	
Hordeum vulgare	CI1(CSI)[13]	52	P	L	N	F	N	P	N	R	V	F	V	L	V	H	K	A	T	T	V	A	Z	V	S	R	V	G	
Sambucus nigra	PR-6[14]	54	T	T	D	E	R	C	D	R	V	R	V	W	V	D	E	N	G	I	V	T	R	V	P	V	I	G	

Residues identical to those of the *Anthriscus sylvestris* CESI are shown in black boxes; conserved substitutions are in grey boxes: nonrelated residues are in black letters in white boxes. Res. nos., numbers of first residues of sequences shown in table. P and P′, *1*, *2*,... positions in reactive centre of potato I inhibitor family members. Species names shown in bold are studied in previous ([2]) and present papers. PI, proteinase inhibitor; pept, peptide; put, putative; # , peptide with M_r 1769 obtained after chemical cleavage of TSI with NaI/CNBr; PR-6, pathogenesis-related protein PR-6 type. [1], Belozersky et al. (1995); [2], Konarev et al. (2002a,b); [3], Hejgaard et al. (1994); [4], Lin et al. (1999); [5], present work; [6], Cierpicki and Otlewski (2000); [7], Demura et al. (2002); [8], Wingate and Ryan (1991); [9], Svendsen et al. (1980); [10], Cleveland et al. (1987); [11], Seemuller et al. (1980); [12], Krishnamoorthi et al. (1990); [13], Jonassen and Svendsen (1982); [14], Coupe et al. (1997).

num usitatissimum TI) is consistent with the specificity of the cleavage.

It can be concluded therefore that the C/E/SI and TSI of *A. sylvestris* are related proteins of the potato inhibitor I family.

2.2.2. Solanum auriculatum (Solanaceae)

Seeds of solanaceous species often contain TI with pI of ≈10.7. *S. auriculatum* was selected because of its simple spectrum of inhibitors and availability of seed. Affinity chromatography on trypsin-Sepharose followed by RP-HPLC gave a major peak of M_r 5919. Edman degradation showed a relationship to the potato inhibitor II family, with the sequence determined spanning the inhibitory site (Table 3). This sequence presumably resulted from proteolysis of the inhibitor, but not at the inhibitory site (i.e. not during affinity purification).

2.2.3. Veronica hederifolia (Scrophulariaceae)

V. hederifolia was selected as a member of the Scrophulariaceae which contained strong TI and TSI. Affinity chromatography on trypsin-Sepharose followed

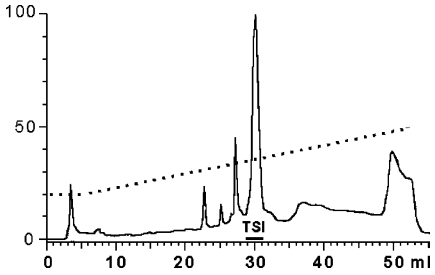

Fig. 5. Purification of T/SI from seeds of *Anthriscus sylvestris*. Proteins were isolated by affinity chromatography on a trypsin-Sepharose column from the extract previously passed through a chymotrypsin-Sepharose gel (see legend to Fig. 4) and were subjected to RP-HPLC on a Vidac C18 column with a 20–50% acetonitrile gradient.

by RP-HPLC (Fig. 6(a)) showed the presence of a mixture of TI and two TSI components (Fig. 6(b)).

MALDI-TOF MS of TI peaks A-E (Fig. 6(b)) showed masses ranging from 4064 to 3842. They were also unusual in being revealed on the basis of activity but not precipitation (with TCA or ammonium sulphate) or staining with Coomassie BBR250. Further analysis of peak D by micropreparative IEF in Dry-Strips gave a series of bands with TI activity and M_r ranging from 3679 to 4048.

Treatment of peak D with DTT gave peptides of M_r 1455, 1671, 1785 and 2280 while Q-TOF analysis of the first three gave overlapping sequences which showed weak similarity with the sequences of Bowman–Birk inhibitors (BBI), including a *C*-terminal arginine residue corresponding to the reactive site (Table 4).

MALDI-MS of fraction I (Fig. 6) containing the major T/SI pI 8 component gave peptides of mass 7322 (major), 4600 and 2800, with the proportion of the former decreasing on incubation with DTT. Edman degradation gave a sequence corresponding to the reactive site of potato inhibitor I. These results are consistent with analyses of other inhibitors of this type (Table 2) (Konarev et al., 2002a) and may result from partial cleavage of an *N*-terminally blocked inhibitor. Similar analysis of fraction G, containing the second (pI 7.5) TSI, showed peptides of similar masses (7361, 4600 and 2816) with the M_r 7361 component being lost on incubation with DTT. It is concluded that both TSI belong to the potato inhibitor I family, with partial proteolysis occurring during preparation.

Table 3
Alignment of the amino acid sequences of the *Solanum auriculatum* TI with those of some members of the potato inhibitor II family

Origin	Inhibitor	Localisation	Meth.	Res. Nos.	P									P′										
					0	9	8	7	6	5	4	3	2	1	1	2	3	4	5	6	7	8	9	0	1	2	3
Capsicum annuum[1]	PI	Seed	Ed	29	D	P	N	N	P	K	A	C	P	R	Y	C	D	T	R	I	A	Y	S	K	C	P	R
Solanum auriculatum[2]	TI	Seed	Ed		?	?	?	?	?	K	A	C	P	R	N	C	D	P	D	V	A	Y	M	V	C	P	?
Nicotiana tabacum[3]	PI	Leaf	Ed	30	D	P	R	N	P	K	A	C	P	R	N	C	D	P	R	I	A	Y	G	I	C	P	L
Capsicum annuum[4]	Put PI	Leaf	Gen	42	H	P	N	N	P	K	A	C	P	R	N	C	D	T	R	I	A	Y	S	K	M	S	T

Residues identical to those of the *Solanum auriculatum* TI are shown in black boxes; conserved substitutions are in grey boxes: nonrelated residues are in black letters in white boxes. Res. nos., numbers of first residues of sequences shown in table. P & P′, *1*, *2*,... positions in reactive site of members of the potato II inhibitor family; put, putative inhibitor of potato inhibitor II family; Ed, sequencing by Edman method; Gen, conceptual translation. [1] Antcheva et al. (1996); [2] present work; [3] Pearce et al. (1993); [4] Shin et al. (2001).

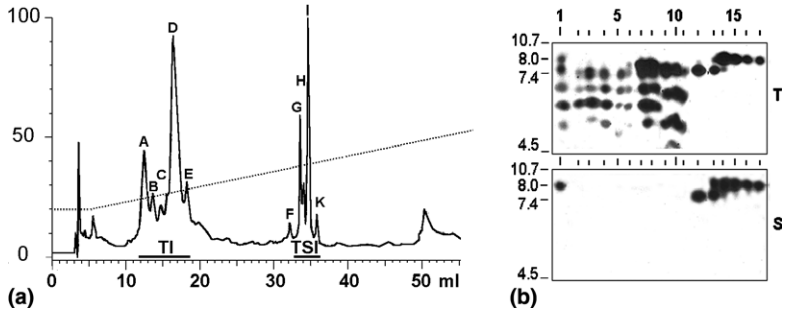

Fig. 6. Purification of serine proteinase inhibitors from *Veronica hederifolia* seeds. Proteins eluted from trypsin-Sepharose with 0.015 M HCl were freeze dried and separated by RP-HPLC with a 20–50% acetonitrile gradient (a). Fractions were freeze dried and analysed by IEF (b). Inhibitors of trypsin (T) and subtilisin (S) were detected using two gelatin replicas obtained from the same gel. 1, water-soluble seed proteins; 2–17, fractions collected from RP-HPLC: 2–4, fractions of peak A; 5 and 6, peak B; 7 and 8, peak D; 9–11, peak E; 12, peak G; 13, peak H; 14–15, peak I; 16–17, peak K.

Table 4
Alignment of the *N*-terminal sequence of *Veronica hederifolia* TI with those of *Zinnia elegans* TI and some members of BBI family

Inhibitor	P	P′
	...98765--432112345678 90..	
SFTI-1[1]	<GRCTKSIPPICFPD>	
IBB1-SOYBN[2&4]	----SKPCCDQC--ACTKSNPPQCRCSD..	
IBBR-ORYSA(I)[2&4]	MEKPW-KCCDNIERLPTKTNPPQWRCND..	
IBBR-ORYSA(II)[2&4]	----WGDCCDKA--FCNKMNPPTCRCMD..	
VhTI	NTDP-EQCKVMC--YAQR-DHFQ-ECND..	
ZTI[3]	PWEQCRSQI--AIEKLNEKQ--MH...	
SACOF-AYO 93808[2]	--SSW-PCCDNC-GVCNKKFPPDCQCND..	

SFTI-1, cyclic sunflower TI; IBB1-SOYBN, soybean BBI (P01055); IBBR-ORYSA, rice BBI (P07084); I and II, the first and the second copies of the tandem repeats; VhTI; *Veronica* TI combining full sequence of peptide of M_r 1785 and partial N-terminal sequence of peptide 2280 beginning from putative P_1' residue; ZTI, *Zinnia elegans* TI; SACOF, sugarcane BBI. 3,2,1 −1,2,3.., P and P′ positions of soybean BBI residues. [1] Luckett et al. (1999); [2] Mello et al. (2003); [3] Konarev et al. (2002a); [4] Internet source http://www.ncbi.nlm.nih.gov/BLAST.

3. Discussion

The asterids comprise about 25–30% of all flowering plants so it is not surprising that there was variation in the spectra of proteinase inhibitors present in the seeds. However, many contained proteins of $M_r \approx 7000$–7500 which inhibited subtilisin-like proteinases of bacteria and fungi. These proteins were often also active against one or more other serine proteinase (trypsin, chymotrypsin, elastase) and in many taxa they were the only detectable inhibitors of serine proteinases. Some other species and taxa appeared to lack serine proteinase inhibitors in their seeds, for example, some species of the Compositae subfamily Carduoideae. Inhibitors of elastase occurred the least frequently among the proteinases tested, being present only in a few members of the families Apiaceae, Acanthaceae, Verbenaceae and Solanaceae.

Trypsin inhibitors which did not also inhibit subtilisin occurred less widely than subtilisin inhibitors, and varied widely in their masses (1500–14,000) and other properties. Types of serine proteinase inhibitors represented in various Asteridae are summarised in Table 5.

It is important to note that only single samples of most species were analysed and these were not grown under a range of conditions. Hence, effects of environmental factors cannot be ruled out completely in relation to the intensities of bands. However, our previous studies included analyses of lines of sunflower and safflower which had been grown under different conditions (Konarev et al., 2000, 2002) while the present study included accessions of the some species (notably *Veronica*) which were obtained from different collections. Since these accessions had similar spectra of serine proteinase inhibitors we conclude that environmental factors had little or no impact on the expression of these proteins. Most seed inhibitors are expressed constitutively rather than being induced and their spectra are often species specific.

Although serine proteinase inhibitors may be present in other plant tissues there is limited information on the Asteridae species studied here. However, analyses of leaves and fruits (Coupe et al., 1997; Pearce et al., 1993; Wingate and Ryan, 1991) showed the presence of the two major inhibitor families (potato I and potato II) which were present in the seeds. In contrast to seeds, these are usually induced by pathogens or other stresses and are absent from intact organs.

3.1. Subtilisin inhibitors

Partial amino acid sequences were obtained from species of several families, including the Compositae (*Carthamus tinctorius* TSI, *Cosmos bipinnatus* TSI and TCSI), Apiaceae (*Anthriscus sylvestris* CESI and TSI) and Scrophulariaceae (*Veronica hederifolia* TSI). All showed relationships with the potato inhibitor I/leech eglin family of inhibitors, with most of the sequences being adjacent to the P_1 residues of the active sites: this presumably resulted from partial cleavage during affinity purification on trypsin-Sepharose or chymotrypsin-Sepharose. With only one exception, the TCSI from *Cosmos bipinnatus*, all of the inhibitors which were studied appeared to have blocked *N*-termini. Although the vast majority of the SI clearly belong to the potato inhibitor I family, exceptions may occur. For example, the SI in seeds of *Phlox drummondii* (Polemoniaceae) have higher M_r (≈9000–10,000) and may belong to a separate inhibitor group.

Although few of the species studied here have been studied at the RNA and genomic levels our data are consistent with the prediction of a putative proteinase inhibitor in *Zinnia elegans* (Demura et al., 2002) and the reported expression of a PR6 (potato inhibitor I) type PR protein in leaves of *Sambucus nigra* (Coupe et al., 1997).

The wide distribution of SI of the potato inhibitor I family may indicate a fundamental biological role, although their absence from some species demonstrates that they are certainly not essential. It is probable that their role relates to interactions with other species, providing defense against pathogenic fungi and regulating interactions with symbionts possessing subtilisin-like enzymes (Reddy et al., 1996). Although endogenous subtilisin-like proteinases are present in some plants (including asterids such as sunflower and plantain) (Bogacheva et al., 2001), their sensitivity to the inhibitors described here has not been established.

3.2. Trypsin inhibitors

BBI are probably the most widely studied group of inhibitors but until recently had only been characterised

Table 5
Summary of the types of serine proteinase inhibitors represented in seeds of various species of the Asteridae

Order, Family, *Species* (examples)	Potato inhibitor I family		Bowman–Birk inhibitor family		Potato inhibitor II family		Non-classified inhibitors	
	Inhibitor	M_r	Inhibitor	M_r	Inhibitor	M_r	Inhibitor	M_r
Asterales[1&2], **Compositae**[1&2]	+		+		–			
Carthamus tinctorius[1&2]	TSI	7555	–		–			
Helianthus annuus[1&2]	TSI[#]	7596	T(C)I	1514	–			
Cosmos bipinnatus[1&2]	TSI CSI	7680 7775	–		–			
Zinnia elegans	+[1,2,3]	≈7500	TI[1##]	11350	–			
Aster, *Kalimeris*, *Gaillardia*, *Silphium*[1&2]	TSI[#]	≈7500	?		–		TI	10,000–14,000
Campanulaceae and **Lobeliaceae**[2]	TSI[#]	n/a	–		–			
Dipsacales[2,4]	+	≈7500	–		–			
Dipsacus sativus[2]	?		–		–		TCSI	≈4500
Apiales, **Apiaceae**[2]	+		–		–			
Anthriscus sylvestris[2]	TSI CESI	7236 7079	–		–			
Sium sisarum[2]	TSI[#] SI[#]	n/a ≈7500	–		–			
Lamiales[2]	+		+		–			
Acanthaceae, *Crabbea reticulata*[2]	?		?		–		TSI	≈4000
Plantaginaceae, *Plantago psyllum*[2]	–		?				TI	≈4000
Pedaliaceae, *Sesamum* sp.[2]	TSI[#]	n/a	–		–			
Scrophulariaceae	+		+		–			
Paederota bonarota[2]	TSI[#]	n/a	?		–		TI	≈3000
Veronica hederifolia[2]	TSI	7322	TI[##]	3900	–			
Gentianales	+[#]		?					
Apocynaceae, *Acokantheras*p.[2]	TSI[#]	≈7500	–		–			
Rubiaceae, *Psychotria* sp.[2]	SI[#]	≈7500	?		–		TI	≈5000
Solanales, Solanaceae	+[2,5]		–		TI[2,6]	6000		
Ericales, Ericaceae	TSI[#]	≈7500	–		–			
Polemoniaceae,*Polemonium* sp.[2]	–		–		–		CI	n/a
Phlox drummondii	–		–		–		SI	≈11,000

"+" or "–", presence or absence of detectable inhibitor; [#], classification based on homology with other inhibitors; [##], inhibitors, which may be distantly related to BBI, based on sequence data. "?", relationship not confirmed by determination of M_r or amino acid sequence. n/a, not analyzed. "≈", approximate M_r according to TLGF; Based on data reported by [1], Konarev et al. (2002a); [2], present work; [3], Demura et al. (2002); [4], Coupe et al. (1997); [5], Cleveland et al. (1987); [6] Antcheva et al. (1996).

from legumes and cereals. Our results demonstrate that proteins related to BBI are present in many asterids but that their degree of relatedness varies. They include the small cyclic SFTI-1 from *Helianthus* (Luckett et al., 1999) which corresponds to part of the inhibitory loop of a BBI and is presumably derived from post-translational processing of a larger BBI-type protein. The ZTI from *Zinnia elegans* (Compositae) and VhTI from *Veronica hederifolia* (Scrophulariaceae) may also be related to the typical BBI (e.g. in the presence of the PWE-C motif and this spacing between C and K residues in the putative reactive site) with VhTI being less distant than ZTI. VhTI also has a similar molecular mass to the thionins, which also exhibit activity as TI (Melo et al., 2002) and have been proposed to be related to BBI.

Our data on the distribution of M_r 3000–5000 TI in species of the Plantaginaceae, Globulariaceae and Scrophulariaceae is also consistent with revised taxonomy which transfers several genera (including *Plantago*, *Globularia*, *Paederota*, *Veronica*) into the Antirrhinaceae (Olmstead et al., 2001; Bremer et al., 2002). Our TLGF studies also demonstrated the presence of M_r 4000–5000 TI in *Psychotria* (Rubiaceae). These may reward further study as *Psychotria* also contains cyclic peptides (cyclotides) of $M_r \approx 3000$ (Witherup et al., 1994) which are related to the cyclotides in *Viola* spp. (Craik et al., 1999) and circular trypsin inhibitors in members of the Cucurbitaceae (Felizmenio-Quimio et al., 2001).

Most species of the Solanaceae also contained alkaline inhibitors of $M_r \approx 6000$ (e.g. SaTI in *S. auriculatum*) which belong to the potato inhibitor II family. Related inhibitors are present in seeds of paprika (*Capsicum annuum*) (Antcheva et al., 1996) and are also induced in leaves by wounding (Moura and Ryan, 2001). These paprika inhibitors are particularly interesting as they are derived by post-translational processing of larger precursor proteins at sites between the inhibitor domains.

4. Conclusions

Although the asterids include a diverse range of species their seeds are characterised by the presence of only two major groups of serine proteinase inhibitors, subtilisin inhibitors of the potato inhibitor I family and trypsin inhibitors of the Bowman–Birk type (Table 5), although the latter may be highly divergent in the se-

quence and structure. In contrast, proteinase inhibitors of the potato inhibitor II family are highly restricted in distribution, to members of the Solanaceae. It is, of course, possible that some of the inhibitors identified in this study may prove to belong to other families and that novel inhibitor types will be identified. However, we consider it unlikely that these broad conclusions will be modified greatly.

5. Experimental

5.1. Seeds

Seeds of various species of the Asteridae were obtained from the world collection of the Vavilov Institute of Plant Industry (VIR, St. Petersburg, Russia), the herbarium and seed collection of the Komarov Botanical Institute (St. Petersburg), Long Ashton Research Station (Bristol, UK) and were purchased from "Herbiseed" (UK) and "Chiltern Seeds" (UK), or were collected by the authors. "k-1, k-2,..." and "vr1, vr2,..." are the numbers of seed accessions in the main and temporal catalogs of VIR; "BI" or default indicates that the acc. was obtained from the Komarov Botanical Institute. Many species were represented by several accessions originating from different collections.

5.1.1. Order Asterales

5.1.1.1. Family Compositae. Tribe Mutisieae Cass.: *Leibnitzia anandria* (L.) Turcz. Subfamily Carduoideae: *Alfredia cernua* (L.) Cass., *Carthamus tinctorius* L. (k-4), *Centaurea alpina* L., *C. atropurpurea* Waldst. et Kit., *C. dealbata* Willd., *C. pseudophrygia* C. A. Mey., *Jurinea arachnoidea* Bunge. Subfamily Cichorioideae: *Scorzonera tau-saghys* Lipsch. et Bosse (k-3), *Achyrophorus maculatus* (L.) Scop., *Achyrophorus uniflorus* (Vill.) Bluff and Fingerh., *Andryala agardhii* Haensel ex DC. Subfamily Asteroideae. Tribe Astereae Cass.: *Aster alpinus* L., *Kalimeris incisa* (Fisch.) DC., *Crinitaria linosyris* (L.) Less., *Kemulariella caucasica* (Willd.) Tamamsch. Tribe Anthemideae Cass.: *Chamomilla recutita* (L.) Rauschert. Tribe Gnaphalieae Benth.: *Antennaria neglecta* Greene, *A. plantaginifolia* Hook., *A. rosea* Greene. Tribe Senecioneae Cass.: *Arnica chamissonis* Less., *A. sachalinensis* (Regel) A.Gray, *Tussilago farfara* L., *Adenostyles glabra* (Mill.) DC. Tribe Eupatorieae Cass.: *Eupatorium micranthum* Less. Tribe Heliantheae Cass., *Guizotia abyssinica* (L. f.) Cass. (k-6), *Madia sativa* Molina, *Helianthus annuus* L., line VIR-130.

5.1.1.2. Family Campanulaceae. Campanula barbata L., *C. latifolia* L., *C. persicifolia* L., *Codonopsis clematidea* (Schrenk) Clarke, *Edraianthus graminifolius* (L.) A.DC., *Edraianthus pumilio* (Port. ex Boem.) A. DC., *E. tenuifolius* (Waldst. et Kit.) A. DC., *Jasione crispa* (Pourr.) G. Sampaio., *J. humilis* Loisel., *J. laevis* Lam., *J. montana* L., *Phyteuma nigrum* F. W. Schmidt.

5.1.1.3. Family Lobeliaceae. Lobelia inflata L., *L. sessilifolia* Lamb., *Pratia begonifolia* Lindl.

5.1.2. Order Apiales

5.1.2.1. Family Apiaceae. Ammi majus L. (vr1 and vr9), *A. visnaga* (L.) Lam. (vr4 and vr6), *Anethum graveolens* L., *Anthriscus* sp. (vr2 and vr3), *Anthriscus cerefolium* (L.) Hoffm. (k-6 and k-9), *A. sylvestris* (L.) Hoffm. (ANRSY), *Astrantia biebersteinii* Trautv., *A. major* L., *A. maxima* Pall., *A. ranunculifolia* Reichenb. f., *A. trifida* Hoffm., *Athamanta matthioli* Wulf., *Bupleurum aureum* Fisch. ex Hoffm., *Carum ajovan* Benth. et Hook. f. (k-1 and 2), *C. carvi* L.(k-45, vr135 and BI), *Chaerophyllum bulbosum* L., *Coriandrum sativum* L. (k-289, k-54 and BI), *Cryptotaenia canadensis* (L.).DC. (vr1), *Cuminum cyminum* L. (vr68 and vr86), *Daucus sativus* (Hoffm.) Roehl.: cv. Shantene 2461 (k-1285), Nantskaya 4 (k-1296) and Moskovskaya zimnyaa (k-1467), *Eryngium agavifolium* Griseb., *E. alpinum* L., *Foeniculum vulgare* Mill. (k-38, k-76), *Heracleum mantegazzianum* Somm. et Levier, *H. pubescens* (Hoffm.) Bieb., *H. stevenii* Manden., *Laser trilobium* (L.) Borkh., *Pastinaca sativa* L., *Peucedanum ostruthium* (L.) Koch, *Pimpinella anisum* L. (k-16 & vr36), *P. major* (L.) Huds., *Prangos ferulacea* (L.) Lindl., *P. pabularia* Lindl., *Seseli libanotis* (L.) Koch, *S. peucedanoides* (Bieb.) K.-Pol., *Sium latifolium* L. (vr4), *S. sisarum* L. (vr2).

5.1.2.2. Family Araliaceae. Acanthopanax sessiliflorus (Rupr.et Maxim.) Seem., *Aralia continentalis* Kitag., *A. cordata* Thunb., *A. elata* (Mig.) Seem., *A. nudicaulis* L., *A. racemosa* L., *Eleutherococcus senticosus* (Rupr.et Maxim.) Maxim., *Pseudopanax crassifolium* C. Koch.

5.1.2.3. Family Pittosporaceae. Pittosporum crassifolium Soland. ex Patterl.

5.1.3. Order Cornales

Family Cornaceae: *Swida alba* (L.) Opiz var. *argenteo-marginata* (Rehd.) H.N. Moldenke, *S. australis* (C. A. Mey.) Pojark. ex Grossh.

5.1.4. Order Dipsacales

5.1.4.1. Family Caprifoliaceae. Diervilla sessilifolia Buckl., D. Lonicera Mill., *D. rivularis* Gatt., *Lonicera alpigena* L., *L. caucasica* Pall., *L. demissa* Rehd., *L. edulis* Turcz. ex Freyn, *L. involucrata* (Richards.) Banks ex Spreng., *L. kamtschatica* (Sevast.) Pojark., *L. ledebourii* Esch., *L. maacki* (Rupr.) Herd., *L. maximowiczii* (Rupr.) Regel, *L. praeflorens* Batal., *L. ruprechtiana* Regel., *L. tatarica* L. v. *rosea* Regel.

L. xylosteum L. Regel, *Sambucs racemosa* L., *S. canadensis* L., *S. coreana* (Nakai) Kom. et Aliss., *Symphoricarpos albus* C. Koch, *S. occidentalis* Hook., *Triosteum hirsutum* Roxb., *Viburnum lantana* L., *V. opulus* L., *V. tinus* L., *Weigela praecox* (Lemoine) Bailey, *W. subsessilis* (Nakai) Bailey.

5.1.4.2. Family Dipsacaceae. Cephalaria gigantea (Ledeb.) Bobr., *Dipsacus* sp. (k-1), *Dipsacus fullonum* L., *D. laciniatus* L., *D. pilosus* L., *D. sativus* (L.) Honck., *Scabiosa atropurpurea* L., *S. bipinnata* C. Koch, *S. graminifolia* L., *S. lucida* Vill., *Succisa pratensis* Moench.

5.1.4.3. Family Valerianaceae. Patrinia gibbosa Maxim., *Valeriana officinalis* L., *V. phu* L., *V. sambucifolia* Mikan f.

5.1.5. Order Ericales

5.1.5.1. Family Actinidiaceae. Actinidia kolomikta (Maxim.) Maxim, *A. polygama* (Siebold et Zucc.) Mig.

5.1.5.2. Family Ericaceae. Arctostaphylos uva-ursi (L.) Spreng., *Bruckenthalia spiculifolia* (Salisb.) Reichenb., *Chamaedaphne calyculata* (L.) Moench, *Erica tetralix* L. *Gaultheria shallon* Pursh, *Ledum groenlandicum* Oeder, *Menziesia ferruginea* Sm., *Rhododendron camtschaticum* Pall., *R. canadense* (L.) Britton Stern et Pogg., *R. catawbiense* Michaux, *R. catawbiense* Michaux (cv.grandiflorum), *R. fauriei* Franch., *R. ferrugineum* L., *R. hirsutum* L., *R. x hybridum* Ker-Gavl, *R. luteum* Sweet, *R. schlippenbachii* Maxim., *R. sichotense* Pojark., *R. smirnowii* Trautv., *R. vaseyi* A. Gray, *R. x intermedium* Tausch., *Vaccinium myrtillus* L., *V. vitis-idaea* L., *V. uliginosum* L.

5.1.5.3. Family Myrsinaceae. Ardisia crenata Sims., *A. japonica* (Hornst.) Blume, *A. wallichii* A. DC., *Maesa argentea* Wall., *M. lanceolata* Voigt.

5.1.5.4. Family Polemoniaceae. Phlox drummondii Hook., *Polemonium caeruleum* L., *P. caucasicum* N. Busch., *P. molle* Greene, *P. racemosum* (Regel) Kitam.

5.1.5.5. Family Primulaceae. Anagallis arvensis L., *Androsace carnea* L. subsp. *brigantica* (Jord.et Fourr.) L.K. Ferguson, *A. septentrionalis* L., *Cortusa turkestanica* Losinsk., *Lysimachia verticillaris* Spreng., *L. vulgaris* L., *Primula auricula* L., *P. chungensis* Balf. et Ward, *P. conspersa* Hort. ex Gard., *P. cordifolia* Rupr., *P. denticulata* Smith, *P. elator* (L.) Hill, *P. florindae* Ward, *P. japonica* A. Gray, *P. macrocalyx* Bunge, *P. paxiana* Gilg, *P. veris* L.

5.1.6. Order Gentianales

5.1.6.1. Family Apocynaceae. Acokanthera oppositifolia (Lam.) Codd., *Alyxia daphnoides* A. Cunn., *Apocynum* sp., *A. androsaemifolium* L., *Carissa bispinosa* (L.) Desf.ex. Brenan, *Rauwolfia tetraphylla* L., *R. verticillata* (Lour.) Baill.

5.1.6.2. Family Gentianaceae. Gentiana acaulis L., *G. asclepiadea* L., *G. burseri* Lapeyr., *G. cruciata* L., *G. decumbens* L.f, *G. frigida* Haenke, *G. kolakovskyi* Doluch., *G. lagodechiana* (Kusn.) Grossh., *G. lutea* L., *G. purpurea* L., *G. septemfida* Pall., *G. tianschanica* Rupr., *G. wutaiensis* Marquand, *Gentiana* sp., *G. frigida* Haenke, *Swertia iberica* Fish. et C.A. Mey.

5.1.6.3. Family Rubiaceae. Asperula scutellaris Vis., *Borreria verticillata* (L.) G.F.W. Mey., *Coccocypselum guianense* K. Schum., *Coffea canephora* Pierre, *C. stenophylla* G. Don, *Gardenia thunbergia* L. f., *Hamelia patens* Jacq., *Ophiorrhiza mungos* L., *Psychotria bacteriophila* Valet, *P. maingayi* Hook. f., *Rubia cordifolia* L., *Tarenna asiatica* Kuntze.

5.1.7. Order Lamiales

5.1.7.1. Family Acanthaceae. Crabbea reticulata Clarke, *Dicliptera resupinata* Juss., *Hemigraphis primulaefolia* Villar.

5.1.7.2. Family Bignoniaceae. Catalpa bignonioides Walt., *C. ovata* C. Don f., *Incarvillea delavayi* Bureau et Franch.

5.1.7.3. Family Gesneriaceae. Alloplectus capitatus Hook., *A. vittatus* Linden et Andre, *Columnea sanguinea* (Pers.) Hanst., *Diastema quinquevulnerum* Planch. et Linden, *Ramondia mycani* (L.) Reichenb.

5.1.7.4. Family Lamiaceae. Agastache rugosa (Fisch. et C.A.Mey.) O. Kuntze, *Betonica alopecuros* L., *B. macrantha* C. Koch, *B. officinalis* L., *Calamintha nepeta* (L.) Savi, *Coleus blumei* Benth., *Dracocephalum integrifolium* Bunge, *D. moldavica* (L.) cv. Album, *D. ruyschiana* L., *Sideritis montana* L., *S. taurica* Steph., *Horminum pyrenaicum* L., *Hyssopus officinalis* L., *Lallemantia iberica* (Bieb.) Fish. et C.A. Mey. (k-13, k-14, k-30, k-44, k-48, k-50, k-56, k-63, k-66, k-75, k-76), *Lavandula angustifolia* Mill., *Leonurus quinquelobatus* Gilib., *Lycopus exaltatus* L. f., *Nepeta cataria* L., *Nepeta grandiflora* Bieb., *Ocimum gratissimum* L., *Perilla ocymoides* L. (k-130), *Phlomis pratensis* Kar. et Kir., *Phlomis tuberosa* L., *Prunella laciniata* (L.) L., *Prunella vulgaris* L., *Salvia glutinosa* L., *S. pratensis* L., *S. verticillata* L., *S. virgata* Jacq., *Scutellaria woronowii* Juz., *Stachys byzanthina* C. Koch, *S. germanica* L., *Thymus serpyllum* L., *Ziziphora bvericalyx* Juz.

5.1.7.5. Family Oleaceae. Forsythia ovata Nakai, *F. x intermedia* Zab., *Ligustrum japonicum* Thunb, *L. vulgare* L., *Olea europaea* L., *O. africana* Mill., *Phillyrea latifolia*

L. v. media (L.) Schneid., *Syringa amurensis* Rupr., *S. vulgaris* L., *S. x prestoniae* McKelvey.

5.1.7.6. Family Orobanchaceae Vent. Aeginetia indica L.

5.1.7.7. Family Pedaliaceae. Sesamum alatum Thorn. (k-1355, k-1111), *S. indicum* L. (k-1060, k-1109, k-1129, k-1212, k-1222, k-1298).

5.1.7.8. Family Phrymaceae. Schauer: Phryma leptostachya L.

5.1.7.9. Family Plantaginaceae Juss. Plantago lanceolata L., *P. major* L. cv. Atropurpurea, *P. psyllium* L.

5.1.7.10. Family Globulariaceae DC. Globularia punctata Lapeyr.

5.1.7.11. Family Scrophulariaceae. Digitalis grandiflora Mill., *D. lutea* L., *D. thapsi* L., *Erinus alpinus* L., *Lathraea squamaria* L., *Paederota bonarota* L., *Penstemon congestus* (M.E. Jones) Peunell, *Scrophularia czerniakowskiana* B. Fedtsch., *S. nodosa* L., *Tetranema mexicana* Benth., *T. roseum* (M. Mart. et Gal.) Standl.et Steyerm. *Verbascum bactriamum* Bunge, *V. gossypinum* Bieb., *V. nigrum* L., *V. thapsus* L., *Veronica anagalloides* Guss., *V. arvensis* L., *V. hederifolia* L., *V. persica* Poir., *V. gentianoides* Vahl, *V. prostrata* L., *V. virginica* L. (last three species both from "Herbiseed" and BI).

5.1.7.12. Family Verbenaceae. Callicarpa americana L., *Clerodendron thomsonae* Balf., *Lantana camara* L., *L. kisi* A.Rich., *L. lilacina* Desf., *L. viburnoides* Vahl.

5.1.8. Order Solanales

5.1.8.1. Family Boraginaceae. Anchusa italica Retz, *Cordia myxa* L., *Cynoglossum officinale* L., *Lithospermum officinale* L., *Myosotis sylvatica* Ehrh. ex Hoffm., *Onosma stellulatum* Waldst.et Kit., *Symphytum caucasicum* Bieb, *Trachystemon orientale* (L.) G.Don fil.

5.1.8.2. Family Hydrophyllaceae. Phacelia tanacetifolia Benth.

5.1.8.3. Family Solanaceae. Atropa belladonna L. var. *lutea* Boel, *Anisodus stramonifolius* G. Don, *Brunfelsia uniflora* (Pohl) D. Don, *Capsicum annuum* L., *C. conicum* G. F. W. Mey., *C. chinense* Jacq., *C. frutescens* L., *C. pubescens* Ruiz et Pav. *Datura stramonium* L., *Hyoscyamus niger* L., *L. cerasiforme* Dun., *Lycopersicon esculentum* Mill., *L. esc*. var. *succentriatum*, *L. esc*. var. *pruniforme*, *L. esc*. var. *elongatum*, *L. racemigerum* Lange, *L. pimpinellifolium* (L.) Mill., *L. pyriforme* Dun., *Nicandra. physaloides (L.) Gaertn.*, *Nicotiana rustica* L., *Petunia x atkinsiana* D. Don (var. *milliflora* and *grandiflora*), *Physalis ixocarpa* Brot. ex Hornem., *P. pubescens* L., *P. alkekengi* L. (k-139, k-135 and k-151), *P. franchetii* Mast., *P. peruviana* L., *Physochlaina orientalis* (Bieb.) G. Don f., *Schizanthus* sp., *Scopolia caucasica* Kolesn. ex Kreyer, *S. tangutica* Maxim., *Scopolia carniolica* Jacq.; *Solanum auriculatum* Ait., *S. capsicastrum* Link, *S. dulcamara* L., *S. guineense* Lam., *S. incanum* L., *S. integrifolium* Poir., *S. kitagawae* Schinbeck-Temesy, *S. mammosum* L., S. melongena L. (k-87, k-114, k-115, k-117, k-171 k-262, k-357, k-509, k-558, k-599, k-747, k-857 and k-930), *S. nigrum* L., *S. persicum* Willd. ex Roem.et Schult., *S. pseudocapsicum* L., *S. pseudopersicum* Pojark., *S. quitoense* Lam., *S. sisymbriifolium* Lam., *S. sodomeum* L., *S. topiro* Humb. et Bonpl. ex Dun., *S. trifolium* Dun., *S. villosum* Mill., *Withania frutescens* (L.) Pauquy, *W. somnifera* (L.) Dun.

5.2. Proteins

5.2.1. Extraction

Proteins were extracted from milled seeds with water (1:4 w/v) for 1 h at 20 °C, based on previously reported studies (Konarev et al., 2000). For preparative separation, the meal was defatted with hexane and then extracted with water (1:10 w/v). The mixture was centrifuged at 20,000 *g* for 30 min at 20 °C.

5.2.2. Analytical separation

Fractions were separated by isoelectric focusing (IEF) in Servalyt Precotes pH 3–10 (Serva) gels with 4.5 cm between electrodes (Konarev et al., 2000) or in 24 cm long DryStrips NL (Pharmacia-Amersham) with an immobilised gradient pH 3–10 and by thin layer gel-filtration (TLGF) in Sephadex G-50 (Konarev et al., 2002a). IEF protein markers with pI values 3.5, 4.2, 4.5, 5.2, 5.3, 6.0, 6.9, 7.4, 7.8, 8.0, 8.3, 9.5 and 10.7 (SERVA) were stained with Coomassie R-250 after IEF and used to estimate the pI values of the inhibitor bands.

TLGF was used to estimate the molecular masses of proteinase inhibitors present in crude extracts, based on their mobility in relation to characterised inhibitors and other proteins with known molecular masses. The relative mobility (Rm, in units 0–9) was calculated based on *H. annuus* TI (SFTI-1, M_r 1514); *Cosmos bipinnatus* TSI (M_r 7500) (Konarev et al., 2002a); *Ecballium elaterium* (L.) A. Rich. TIs ($M_r \approx 3000$), and cytochrome *c* (M_r 12,500) (see footnote to Table 1).

5.2.3. Detection of inhibitors

Proteinase inhibitors were detected after analytical separation of native proteins by the gelatin replicas method using a gelatin layer on photographic film (Konarev, 1986; Konarev et al., 2000). One to four replicas were sequentially placed in contact with the same IEF gel for 2, 5, 20 and 30 min, respectively. The replicas were then applied to 0.8% (w/v) agarose gels containing

0.1 M Na_2HPO_4 (pH 9) and one of the following proteinases (Sigma); trypsin (1 μg/ml), chymotrypsin (10 μg/ml), porcine pancreatic elastase (4 μg/ml) and subtilisin (0.3 μg/ml), and incubated at 45 °C for 30 min.

The activities of inhibitors in extracts were estimated based on the volume of extract required to obtain clear inhibitor banding patterns. Volumes below 0.3 μl, therefore, were defined as "high activity"; volumes up to 2.5 μl as "weak" and the failure to detect digestion with 5 μl extract was defined as "low or absent"

5.2.4. Purification of inhibitors

Single inhibitor components were purified by combining affinity chromatography, reversed-phase (RP) HPLC (Konarev et al., 2000, 2002a) and micropreparative IEF in Servalyt precotes gels or in DryStrips. Ammonium acetate (AmAc) was added to the crude protein extract to 0.2 M and 1 l of the extract was shaken for 30 min with 15 ml of trypsin- or chymotrypsin-Sepharose gel. The gel was then washed with 0.4 l of 0.2 M ammonium acetate, 0.3 l of 0.1 M Na_2HCO_3 pH 8.0, 0.3 l of 0.1 M Na acetate (NaAc) buffer pH 4.2 (in case of high affinity of the purified inhibitor for the ligand) and 0.3 l of water and packed into a column. Inhibitors were eluted with 0.015 M HCl, freeze-dried and subjected to HPLC on a C_{18} RP Vydac column with an acetonitrile (ACN) gradient (various from 5% to 50% (v/v)). Fractions were freeze-dried and inhibitors detected after analytical IEF. HPLC-fractions containing proteinase inhibitors were separated on Servalyt precotes (pH 3–10) or DryStrips gels. Protein bands were visualised by washing in 10% TCA, excised, washed with cold (0 °C) acetone, eluted with 100–200 μl of water and freeze-dried. Bands were also visualised with saturated ammonium sulphate (AmS). The eluate was then subjected to a second separation by RP-HPLC or to mass spectrometry analysis.

5.2.5. Cleavage of inhibitors

Inhibitors were often partially cleaved at their reactive or other sites during purification, particularly by affinity chromatography on trypsin or chymotrypsin-Sepharose. Peptides were also cleaved by modified chemical methods based on CNBr (Smith, 1994), iodosobenzoic acid (IBA) (Fontana et al., 1983), a combination of NaI with CNBr (Huang and Huang, 1994) or CNBr followed by IBA cleavage.

5.2.5.1. Cleavage at methionine residues with CNBr. One to four micrograms of protein were dissolved in 4 μl of 0.4 M ammonium bicarbonate with 5% (v/v) 2-mercaptoethanol and incubated for 18 h under nitrogen. After drying under vacuum 5 μl of CNBr solution (20 mg/ml) in 70% (v/v) formic acid were added to sample. The mixture was incubated in the dark for 24 h and dried under a flow of nitrogen.

5.2.5.2. Cleavage at methionine and tryptophan residues with NaI/CNBr. Two microlitres of NaI (2 mg/ml) and 10 μl of CNBr (20 mg/ml) in 80% (v/v) formic acid were added to 1–4 μg of protein. After overnight incubation the sample was dried under a flow of air.

5.2.5.3. Cleavage at tryptophan and tyrosine residues with iodosobenzoic acid. One to five micrograms of protein were dissolved in 5 μl of IBA (15 mg/ml) in 4 M guanidinium chloride in 80% (v/v) acetic acid and incubated for 24 h in the dark.

5.2.6. Mass spectrometry

Peptides were dissolved in 10 μl of 1% (v/v) formic acid. Samples were prepared using the dried droplet method of crystallization, where 0.75 μl of peptide was mixed with 0.75 μl of matrix on the stainless steel sample probe, then allowed to air dry.

MALDI-TOF MS experiments used a M@LDI L/R (Micromass, Manchester). Data acquisition and processing were performed via the Mass Lynx 3.5 data system.

The machine was calibrated in linear mode across the mass range 2000–16,000 using 10 pmol/μl of adrenocorticotropic hormone (ACTH M_r 2465), aprotinin (M_r 6,514), cytochrome c (M_r 11,694) and myglobin (M_r 16,940). For linear mode 1 mg/ml of 3,5-dimetoxy-4-hydroxysinnamic acid was dissolved in 49.5% (v/v) ethanol, 49.5% (v/v) acetonitrile and 1% (v/v) of a 1% (v/v) TFA solution.

The machine was calibrated in reflectron mode across the mass range 1000–3000 using 0.75 pmol of an alcohol dehydrogenase tryptic digest and tuned to a resolution of greater than 10,000. For reflectron mode 1 mg/ml of α-cyano-4-hydroxycinnamic acid (HCCA) was dissolved in 49.5% (v/v) ethanol, 49.5% (v/v) acetonitrile and 1% (v/v) of a 1% (v/v) TFA solution.

5.2.7. Amino acid sequencing

Automated Edman degradation was carried out on a pulsed-liquid amino acid sequencer (model 477A, Applied Biosystems) equipped with an online phenylthiohydantoin amino acid analyzer (model 120A), by Dr. Mike Naldrett at the JIC, Norwich, UK.

Acknowledgements

Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom. This work was carried out in the Program of Fundamental Investigations on Plant Protection of the Russian Academy of Agricultural Sciences and was supported under an "Ex-Agreement Visit" grant from The Royal Society of London. The authors are grateful to Dr. N.N.

Arnautov, Dr. I.A. Pautova, Dr. N.G. Kutyavina, Dr. O.A. Zvereva, T.E. Vachrusheva and other specialists from V.L. Komarov Botanical Institute and N.I. Vavilov Institute of Plant Industry (VIR) for providing plant accessions. Many thanks to Prof. Alexey V. Konarev (VIR) for providing equipment for research and to Dr. R. Fido for his valuable advice and assistance.

References

Antcheva, N., Patthy, A., Athanasiadis, A., Tchorbanov, B., Zakhariev, S., Pongor, S., 1996. Primary structure and specificity of a serine proteinase inhibitor from paprika (*Capsicum annuum*) seeds. Biochim. Biophys. Acta 14 (1298), 95–101.

Belozersky, M.A., Dunaevsky, Y.E., Musolyamov, A.X., Egorov, T.A., 1995. Complete amino acid sequence of the protease inhibitor from buckwheat seeds. FEBS Lett. 371, 264–266.

Bhat, N.S., Puttarangappa, K., Virupakshappa, Prasad, D.T., 1996. Proteinase inhibitor in sunflower seed and its influence on growth and development of *Capitulum borer*, *Helicoverpa armigera* (Hubner). In: Proceedings of 14th International Sunflower Conference, International Sunflower Association, Beijing-Shenyang, PR China, 12–20 June 1996, pp. 523–532.

Birk, Y., 2003. Plant Protease Inhibitors: Significance in Nutrition, Plant Protection, Cancer Prevention and Genetic Engineering. Springer, Berlin, p. 170.

Bogacheva, A.M., Rudenskaya, G.N., Dunaevsky, Y.E., Chestuhina, G.G., Golovkin, B.N., 2001. New subtilisin-like collagenase from leaves of common plantain. Biochimie 83, 481–486.

Bremer, K., Chase, M.W., Stevens, P.F., 1998. An ordinal classification for the families of flowering plants. Ann. Missouri Bot. Garden 85, 531–553.

Bremer, B., Bremer, K., Heidari, N., Erixon, P., Olmstead, R.G., Anderberg, A.A., Kallersjo, M., Barkhordarian, E., 2002. Phylogenetics of Asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. Mol. Phylogenet. Evol. 24, 274–301.

Cierpicki, T., Otlewski, J., 2000. Determination of a high precision structure of a novel protein, *Linum usitatissimum* trypsin inhibitor LUTI), using computer-aided assignment of NOESY cross-peaks. J. Mol. Biol. 302, 1179–1192.

Cleveland, T.E., Thornburg, R.W., Ryan, C.A., 1987. Molecular characterisation of a wound-inducible inhibitor I gene from potato and the processing of its mRNA and protein. Plant Mol. Biol. 8, 199–207.

Coupe, S.A., Taylor, J.E., Roberts, J.A., 1997. Temporal and spatial expression of mRNAs encoding pathogenesis-related proteins during ethylene-promoted leaflet abscission in *Sambucus nigra*. Plant Cell Environ. 20, 1517–1524.

Craik, D.J., Daly, N.L., Bond, T., Waine, C., 1999. Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. J. Mol. Biol. 294, 1327–1336.

Demura, T., Tashiro, G., Horiguchi, G., Kishimoto, N., Kubo, M., Matsuoka, N., Minami, A., Nagata-Hiwatashi, M., Nakamura, K., Okamura, Y., Sassa, N., Suzuki, S., Yazaki, J., Kikuchi, S., Fukuda, H., 2002. Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells. Proc. Natl. Acad. Sci. USA 99 (24), 15794–15799.

Downie, S.R., Katz-Downie, D.S., Watson, M.F., 2000. A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA rpl16 and rpoC1 intron sequences: towards a suprageneric classification of subfamily Apioideae. Am. J. Bot. 87, 273–292.

Felizmenio-Quimio, M.E., Daly, N.L., Craik, D.J., 2001. Circular proteins in plants: solution structure of a novel macrocyclic trypsin inhibitor from *Momordica cochinchinensis*. J. Biol. Chem. 276, 22875–22882.

Fontana, A., Dalzoppo, D., Grandi, C., Zambonin, M., 1983. Cleavage at tryptophan with *o*-iodosobenzoic acid. Meth. Enzymol. 91, 311–319.

Gutierrez-Campos, R., Torres-Acosta, J.A., Saucedo-Arias, L.J., Gomez-Lim, M.A., 1999. The use of cysteine proteinase inhibitors to engineer resistance against potyviruses in transgenic tobacco plants. Nat. Biotechnol. 17, 1223–1226.

Hejgaard, J., Dam, J., Petersen, L.C., Bjorn, S.E., 1994. Primary structure and specificity of the major serine proteinase inhibitor of amaranth (*Amaranthus caudatus* L.) seeds. Biochim. Biophys. Acta 1204, 68–74.

Huang, S.S., Huang, J.S., 1994. Cleavage of both tryptophanyl and methionyl peptide bonds in proteins. J. Protein Chem. 13, 450–451.

Jonassen, I., Svendsen, I., 1982. Identification of the reactive sites in two homologous serine proteinase inhibitors isolated from barley. Carlsberg Res. Commun. 47, 199–203.

Kollipara, K.P., Hymowitz, T., 1992. Characterisation of trypsin and chymotrypsin inhibitors in the wild perennial *Glycine* species. J. Agri. Food Chem. 40, 2356–2363.

Konarev, Al.V., 1986. Analysis of protease inhibitors from wheat grain by gelatine replicas method. Biochemistry (Moscow) 51, 195–201.

Konarev, Al.V., 1988. Polymorphism of chymotrypsin-subtilisin inhibitors and possibilities of its use in wheat variety identification. In: Biochemical Identification of Varieties. Materials III International Symposium of ISTA, VIR, Leningrad, USSR, pp. 176–181.

Konarev, Al.V., 1996. Interaction of insect digestive enzymes with plant protein inhibitors and host parasite co-evolution. Euphytica 92, 89–94.

Konarev, Al.V., Anisimova, I.N., Gavrilova, V.A., Shewry, P.R., 1999a. Polymorphism of inhibitors of hydrolytic enzymes present in cereal and sunflower seeds. In: Mugnozza, G.T.S., Porceddu, E., Pagnotta, M.A. (Eds.), Genetics and Breeding for Crop Quality and Resistance. Kluwer Academic Publishers, The Netherlands, pp. 135–144.

Konarev, Al.V., Kochetkov, V.V., Bailey, J.A., Shewry, P.R., 1999b. The detection of inhibitors of the *Sclerotinia sclerotiorum* (Lib) de Bary extracellular proteinases in sunflower. J. Phytopathol. 147, 105–108.

Konarev, Al.V., Anisimova, I.N., Gavrilova, V.A., Rozhkova, V.T., Fido, R., Tatham, A.S., Shewry, P.R., 2000. Novel proteinase inhibitors in seeds of sunflower (*Helianthus annuus* L.): polymorphism, inheritance and properties. Theor. Appl. Genet. 100, 82–88.

Konarev, Al.V., Anisimova, I.N., Gavrilova, V.A., Vachrusheva, T.E., Konechnaya, G.Yu., Lewis, M., Shewry, P.R., 2002a. Serine proteinase inhibitors in the Compositae: distribution, polymorphism and properties. Phytochemistry 59, 279–291.

Konarev, Al.V., Tomooka, N., Vaughan, D., 2002b. Proteinase inhibitor polymorphism in genus *Vigna* subgenus *Ceratotropis* and its biosystematic implications. Euphytica 123, 165–177.

Korsinczky, M.L.J., Schirra, H.J., Craik, D.J., 2004. Sunflower trypsin inhibitor—1. Curr. Peptide Prot. Sci., in press.

Krishnamoorthi, R., Gong, Y., Richardson, M., 1990. A new protein inhibitor of trypsin and activated Hageman factor from pumpkin (*Cucurbita maxima*) seeds. FEBS Lett. 273, 163–167.

Lawrence, P.K., Koundal, K.R., 2002. Plant protease inhibitors in control of phytophagous insects. Electron. J. Biotechnol. 5, 93–109 (http://www.ejbiotechnology.info/content/vol5/issue1/index.html).

Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E., Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M., Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., VanAken,

S.E., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M., Venter, J.C., 1994. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 402, 761–768.

Luckett, S., Garcia, R.S., Barker, J.J., Konarev, Al.V., Shewry, P.R., Clarke, A.R., Brady, R.L., 1999. High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. J. Mol. Biol. 290, 525–533.

Melo, F.R., Rigden, D.J., Franco, O.L., Mello, L.V., Ary, M.B., Grossi de Sa, M.F., Bloch Jr., C., 2002. Inhibition of trypsin by cowpea thionin: characterisation, molecular modeling, and docking. Proteins 48, 311–319.

Mello, M.O., Tanaka, A.S., Silva-Filho, M.C., 2003. Molecular evolution of Bowman–Birk type proteinase inhibitors in flowering plants. Mol. Phylogenet. Evol. 27, 103–112.

Mosolov, V.V., Grigor'eva, L.I., Valueva, T.A., 2001. Plant proteinase inhibitors as multifunctional proteins. Appl. Biochem. Microbiol. (Moscow) 37, 643–650.

Moura, D.S., Ryan, C.A., 2001. Wound-inducible proteinase inhibitors in pepper. Differential regulation upon wounding, systemin, and methyl jasmonate. Plant Physiol. 126, 289–298.

NCBI Entrez Taxonomy Homepage (http://www.ncbi.nlm.nih.gov/entrez/).

Olmstead, R.G., DePamphilis, C.W., Wolfe, A.D., Young, N.D., Elisons, W.J., Reeves, P.A., 2001. Disintegration of the Scrophulariaceae. Am. J. Bot. 88, 348–361.

Pearce, G., Johnson, S., Ryan, C.A., 1993. Purification and characterization from tobacco (*Nicotiana tabacum*) leaves of six small, wound-inducible, proteinase isoinhibitors of the potato inhibitor II family. Plant Physiol. 102, 639–644.

Reddy, P.V., Lam, C.K., Belanger, F.C., 1996. Mutualistic fungal endophytes express a proteinase that is homologous to proteases suspected to be important in fungal pathogenicity. Plant Physiol. 111, 1209–1218.

Ryan, C.A., 1990. Protease inhibitors in plants: genes for improving defences against insects and pathogens. Ann. Rev. Phytopathol. 28, 425–449.

Seemuller, U., Eulitz, M., Fritz, H., Strobl, A., 1980. Structure of the elastase-cathepsin G inhibitor of the leech *Hirudo medicinalis*. Hoppe-Seyler's Z. Physiol. Chem. 361, 1841–1846.

Shewry, P.R., 1999. Enzyme inhibitors of seeds: types and properties. In: Shewry, P.R, Casey, R. (Eds.), Seed Proteins. Kluwer Academic Publishers, The Netherlands, pp. 587–615.

Shewry, P.R., 2003. Tuber storage proteins. Ann. Bot. 91, 755–769.

Shewry, P.R., Lucas, J.A., 1997. Plant proteins that confer resistance to pests and pathogens. Adv. Bot. Res. 26, 135–192.

Shibata, H., Hara, S., Ikenaka, T., 1988. Amino acid sequence of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) chymotrypsin inhibitor WCI-3. J. Biochem. Tokyo 104, 537–543.

Shin, R., Lee, G.J., Park, C.J., Kim, T.Y., You, J.S., Nam, Y.W., Paek, K.H., 2001. Isolation of pepper mRNAs differentially expressed during the hypersensitive response to tobacco mosaic virus and characterization of a proteinase inhibitor gene. Plant Sci. 161, 727–737.

Smith, B.J., 1994. Chemical cleavage of proteins. In: Walker, J.M. (Ed.), Basic Protein and Peptide Protocols, Methods in Molecular Biology, 32. Humana Press Inc., Totowa, NJ, pp. 297–309.

Svendsen, I., Martin, B., Jonassen, I., 1980. Characteristics of Hiproly barley III. Amino acid sequences of two lysine-rich proteins. Carlsberg Res. Commun. 45, 79–85.

Urwin, P.E., McPherson, M.J., Atkinson, H., 1998. Enhanced transgenic plant resistance to nematodes by dual proteinase inhibitor constructs. Planta 204, 472–479.

Wingate, V.P.M., Ryan, C.A., 1991. A novel fruit-expressed trypsin inhibitor I gene from a wild species of tomato. J. Biol. Chem. 266, 5814–5817.

Witherup, K.M., Bogusky, M.L., Anderson, P.S., Ramjit, H., Ransom, R.W., Wood, T., Sardana, M., 1994. Cyclopsychotride A, a biologically active 31-residue cyclic peptide isolated from *Psychotria longipes*. J. Nat. Prod. 57, 1619–1625.