Phytochemistry 69 (2008) 2482–2489

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Serine proteinase inhibitors in seeds of *Cycas siamensis* and other gymnosperms

Alexander V. Konarev [a,*], Alison Lovegrove [b], 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

ARTICLE INFO

Article history:
Received 30 April 2008
Received in revised form 3 July 2008
Available online 26 August 2008

Keywords:
Cycas siamensis
Cycadales
Coniferales
Gymnosperms
Kunitz serine proteinase inhibitor
Chymotrypsin
Subtilisin
Trypsin

ABSTRACT

Seeds of 32 species selected from two of the four major groups of gymnosperms, the ancient Cycadales and the economically important Coniferales, were analysed for inhibitors (I) of the serine proteinases trypsin (T), chymotrypsin (C), subtilisin (S) and elastase (E) using isoelectric focusing (IEF) combined with gelatin replicas. Subtilisin inhibitors were detected in 17 species, being particularly active in the Cycadales. Several species of the genera *Cephalotaxus, Pseudotsuga* and *Cycas* contained inhibitors active against elastase while strong CSTIs and CSIs were also present in *Cycas pectinata* and *C. siamensis*. No inhibitors were detected in seeds of *Chamaecyparis, Thuja, Abies, Larix, Picea* and *Pinus* spp. Serine proteinase inhibitors were purified from seeds of *C. siamensis* by affinity chromatography using trypsin and chymotrypsin, IEF and SDS–PAGE. Several CSTI components with M_r ranging from 4000 to 18,000 were partially sequenced using Edman degradation and mass spectrometry. Most of the sequences were similar to a hypothetical protein encoded by an mRNA from sporophylls of *C. rumphii* which in turn was similar to Kunitz-type proteinase inhibitors from flowering plants. Analysis of expressed sequence tag (EST) databases confirmed the presence of mRNAs encoding Kunitz-type inhibitors in the Cycadales and Coniferales and also demonstrated their presence in a third major group of gymnosperms, the Ginkgoales. This is the first report of Kunitz-type serine proteinase inhibitors from plants other than Angiosperms.

1. Introduction

Comparative studies of the seed proteins of gymnosperms and angiosperms are of interest in relation to the evolution of plant proteins and may also be relevant to wider aspects of plant evolution (Shutov et al., 1995). They also have nutritional significance as seeds of some cycads (*Cycas media* R. Br.) and pines (*Pinus sibirica* Du Tour) are consumed by humans in SE Asia and Siberia, respectively, while seeds of the Mediterranean pine *P. pinea* L. are used in confectionery. Proteinase inhibitors are widely distributed in angiosperm seeds and may play multiple roles as storage proteins, in defence against pests and pathogens and in regulation of endogenous enzymes (Shewry and Lucas, 1997; Shewry, 1999, 2003). They have been studied in most detail in the major economic groups of flowering plants (Poaceae, Fabaceae, Solanaceae and Compositae), with inhibitors of serine proteinases such as trypsin, chymotrypsin, elastase and subtilisin being particularly widespread (Shewry, 1999; Konarev et al., 2002, 2004). In some cases they have been demonstrated to inhibit enzymes produced by fungal pathogens that infect the species of origin, for example, inhibitors from sunflower seeds are active against extracellular serine proteinases from the sunflower head pathogen *Sclerotinia sclerotiorum* (Konarev et al., 1999).

Many plant taxa have still not been studied for inhibitors and wider studies could therefore identify inhibitors with novel properties and potential applications in crop protection and biomedicine. For example, a small cyclic inhibitor of serine proteinases in seeds of sunflower (Luckett et al., 1999; Konarev et al., 2000, 2002, 2004) has been used as a scaffold to design novel peptide-based drugs (Korsinczky et al., 2004). Similarly, more recent studies have identified a novel type of trypsin inhibitor in seeds of *Veronica* (Conners et al., 2007) which could also provide a scaffold for rational drug design.

Few studies of proteinase inhibitors in gymnosperms have been reported previously. Salmia and Mikola (1980) reported the presence of inhibitors of endogenous proteinases in seeds of *P. sylvestris* L. while Sawano et al. (2007) isolated an antifungal protein with weak inhibitory activity against pepsin (an aspartyl proteinase) from seeds of *Ginkgo biloba* L. Despite their wide distribution in angiosperms, inhibitors of serine proteinases have not been characterised from gymnosperms, although analysis of cDNAs from white spruce (*Picea glauca* (Moench) Voss) seeds showed the presence of transcripts related to the 2S albumin storage proteins (McInnis et al., 1997) which belong to a wider protein superfamily which includes cereal seed inhibitors of α-amylase and trypsin (Shewry, 2003).

* Corresponding author. Tel.: +7 812 4704384; fax: +7 812 4705110.
E-mail addresses: al_konarev@hotmail.com (Al.V. Konarev), peter.shewry@bbsrc.ac.uk (P.R. Shewry).

0031-9422/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2008.07.001

Similarly, analysis of EST sequences (corresponding to mRNAs) revealed the presence of a putative subtilisin-chymotrypsin (SCI) inhibitor with 49% amino-acid sequence identity with a wheat subtilisin-chymotrypsin inhibitor related to potato inhibitor 1 family in the spores of lycophytes (club mosses), the most primitive of vascular plants (Weng et al., 2005).

We have therefore determined the distribution of inhibitors of trypsin, chymotrypsin, subtilisin and elastase in seeds of 32 species from two major groups of gymnosperms (Chaw et al., 2000; Bowe et al., 2000), the cycads which are the most ancient extant group, and the conifers which include many species with economic importance as sources of wood and chemicals for medicine and industry. We have also isolated and partially characterised a serine proteinase inhibitor from *Cycas siamensis* seeds. *Cycas* species are of particular interest since these plants represent the most primitive living seed plants and are sometimes considered a "missing link" between vascular non-seed plants and the more advanced seed plants (Brenner et al., 2003a).

2. Results

2.1. The distribution of serine proteinase inhibitors in gymnosperms

The distribution of inhibitory activities in gymnosperms seeds was determined by screening aqueous extracts using IEF followed by "gelatin replicas" (see Section 4.2.3) developed with trypsin (Fig. 1A), subtilisin (Fig. 1B), chymotrypsin (Fig. 1C) and elastase (not shown). The results are summarised in Table 1. Clear differences were observed in the presence and activity of inhibitors in the different groups, with higher activities being present in the Cycadales than in the Coniferales. Within the latter only seven out of 21 species contained detectable levels of inhibitors, all of which were subtilisin inhibitors. These subtilisin inhibitors showed only weak activity in four species (*Juniperus* spp., *Taxus cuspidata* Siebold et Zucc. ex Endl.) and medium activity in *Cephalotaxus harringtonia* (Knight ex Forbes) K. Koch (the only member of the Cephalotaxaceae that was analysed) and *Pseudotsuga menziesii* (Mirb.) Franco and *Taxus baccata* L. In *Pseudotsuga* and *Cephalotaxus* the same proteins also inhibited porcine elastase. Inhibitors of trypsin and chymotrypsin were not detected in seeds of any of the Coniferales.

With the exception of *Cycas* species, members of the Cycadales contained strong subtilisin inhibitors but no other inhibitor-types. The four *Cycas* species that were analysed all contained subtilisin inhibitors and in *C. michotiizii* Dyer some of these also appeared to inhibit elastase. In *C. pectinata* Buch.-Ham. and *C. siamensis* Miq. these strong SI were accompanied by TI and CI, with some appearing to inhibit either both trypsin and subtilisin or all three enzymes (trypsin, subtilisin and chymotrypsin).

2.2. Analysis of C. siamensis inhibitors

C. siamensis was selected for more detailed studies because it contained a complex spectrum of inhibitors, including inhibitors of trypsin, chymotrypsin and subtilisin (Fig. 1, tracks 8 and 9) but not elastase (Table 1). Proteins were therefore extracted from milled seeds with water and separated by affinity chromatography with the target enzymes linked to Sepharose 4B.

The fractions eluted from the trypsin affinity column (called Af-T) were then separated by IEF in Immobiline DryStrip gels pH 3–10. One of two gel separations of the Af-T fraction was fixed in 10% TCA revealing the proteins as opaque bands under scattered light with a black background. Three gelatin replicas were also made from a second separation of the same fraction and inhibitors were detected using trypsin (T), chymotrypsin (C) and subtilisin (S)

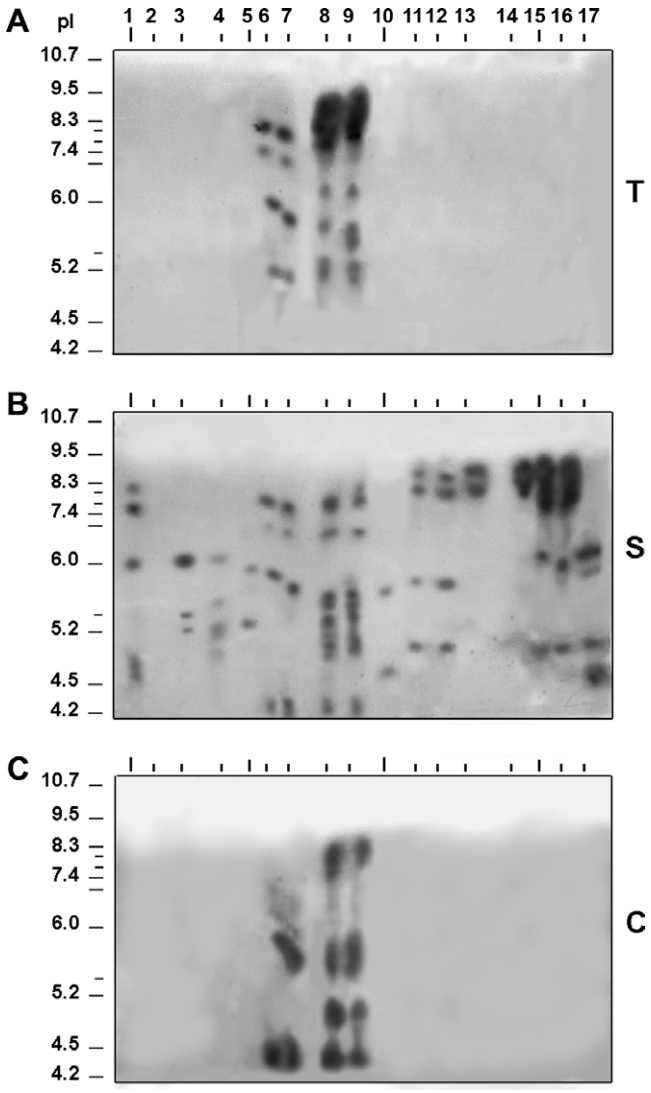


Fig. 1. (A–C) Distribution of serine proteinase inhibitors in seeds of gymnosperms. Water-soluble seed proteins were separated by IEF and proteinase inhibitors were detected using the "gelatin replicas" method. Replicas from the same gel were detected with trypsin (T), subtilisin (S) and chymotrypsin (C). (1) *Cephalotaxus harringtonia*; (2) *Pseudotsuga taxifolia*; (3) *Taxus baccata*; (4) *Cycas micholitzii*; (5) *C. media*; (6 and 7) *C. pectinata*; (8 and 9) *C. siamensis*; (10) *Dioon edule*; (11) *Macrozamia communis*; (12) *M. miquelii*; (13) *M. riedlii*; (14) *Zamia ficheri*; (15 and 16) *Z. furfuracea*; and (17) *Z. pumila*.

(Fig. 2). The inhibitory activity appears as dark bands of undigested gelatin in the replicas labelled T, C and S. The majority of the bands in the Af-T fraction were active against all three proteinases, although they differed in their relative activities.

In order to characterise the individual inhibitors, the proteins labelled as bands a, b and c (the latter comprising two close bands) were extracted from the IEF gel and further separated using SDS–PAGE. They showed similar patterns, with bands varying in mass from about 18,000 to 3500 although the components varied in proportions. Six major bands were sequenced from the SDS–PAGE separations of IEF bands b (Fig. 3, track b) and c (Fig. 3, track c), either by mass spectrometry of peptides obtained from in-gel trypsin digests (bands 1 and 2) or Edman degradation of bands transferred onto PVDF membrane (bands 1, 3–6) (Table 2).

Affinity chromatography was also carried out on columns of chymotrypsin (Af-C) and of immobilised methyl-chymotrypsin (Af-MetC) (ie inactive chymotrypsin). These fractions gave similar patterns to each other and to the Af-T fraction on IEF, showing that

Table 1
Distribution of serine proteinase inhibitors in seeds of gymnosperms

Order, family, species	Presence of inhibitory activity towards proteinases				Inhibitors present
	T	S	C	E	
Coniferales					
Cephalotaxaceae					
Cephalotaxus harringtonia	–	+	–	w	SI, wE/SI
Cupressaceae					
Chamaecyparis pisifera	–	–	–	–	–
Chamaecyparis lawsoniana	–	–	–	–	–
Juniperus communis	–	w	–		wSI
Juniperus phoenicia	–	w	–		wSI
Juniperus virginiana	–	w	–		wSI
Thuja occidentalis	–	–	–	–	–
Pinaceae					
Abies koreana	–	–	–	–	–
Abies concolor	–	–	–	–	–
Abies fraseri	–	–	–	–	–
Abies veitchii	–	–	–	–	–
Larix decidua	–	–	–	–	–
Larix kaempferi	–	–	–	–	–
Larix x lubarskii	–	–	–	–	–
Picea abies	–	–	–	–	–
Pinus sibirica	–	–	–	–	–
Pinus sylvestris	–	–	–	–	–
Pseudotsuga menziesii	–	+	–	w	sSI, E/SI
Taxus canadensis	–	–	–	–	–
Taxus cuspidata	–	w	–	–	wSI
Taxus baccata	–	+	–	–	SI
Cycadales					
Cycadaceae					
Cycas micholitzii	–	w	–	+	wSI, E/SI
Cycas media	–	w	–	–	wSI
Cycas pectinata	+	+	+	–	sTSI, TCI, sSI, CSI
Cycas siamensis	+	+	+	–	sTSCI, CSI,
Zamiaceae					
Dioon edule	–	w	–	–	wSI
Macrozamia communis	–	+	–	–	SI
Macrozamia miquelii	–	+	–	–	SI
Macrozamia riedli	–	+	–	–	sSI
Zamia fisheri	–	+	–	–	sSI
Zamia furfuraceae	–	+	–	–	sSI
Zamia pumila	–	+	–	–	SI

T, trypsin; S, subtilisin; C, chymotrypsin; and E, elastase. "+" or "–": presence or absence of inhibitory activity and "s" or "w": strong or weak activity of inhibitor.

the heterogeneity did not arise from proteolysis during purification (data not shown). The fractions were also separated by SDS–PAGE. A single M_r 22,000 component from Af-C was sequenced. This was not present in Af-T and the sequence showed that it was not a protease inhibitor. A M_r 3500 component from Af-MetC gave the same sequence as component 6 (also M_r 3500) from the Af-T fraction (data not shown).

2.3. Comparison of *C. siamensis* components with sequences in databases

Comparisons with protein sequence databases showed that the sequences of bands 1 and 3 (Fig. 3), from *C. siamensis* were similar to those of a putative Kunitz trypsin inhibitor from *Populus tremula* L. (GenBank accession no. CAI77767; Ingvarsson, 2005) and to other Kunitz-type trypsin inhibitors from angiosperms (Table 3). Furthermore, although no related sequences from gymnosperms are available in protein databases, they are present in expressed sequence tag (EST) databases. Thus, the sequences of all six bands (Fig. 3) are similar to the protein encoded by an EST from *C. rumphii* Miq. sporophylls (DR061860, Brenner et al., 2003b) (Table 3) which itself has high homology with Kunitz-type inhibitors. For example, with TIs from *Prosopis juliflora* (Negreiros et al., 1991), *P. tremula* (Ingvarsson, 2005) and *Theobroma bicolor* (Silva and Figueira, 2001), a cathepsin D/trypsin inhibitor from potato (Ritonja et al., 1990), a bifunctional alpha-amylase/subtilisin inhibitor from *Hordeum vulgare* seeds (called BASI) (Leah and Mundy, 1989) and a tumor-related protein from *Nicotiana tabacum* (Karrer et al., 1998). Comparison of the sequence of the putative Kunitz inhibitor-like protein from *C. rumphii* with those of putative proteins encoded by ESTs from other gymnosperms, including other cycad species gave the following percentage identities and similarities, respectively: *Zamia vazquezii* (42% identical and 59% similar), ginkgo *G. biloba* L. (41% and 57%), Sitka spruce *Picea sitchensis* (Bong.) Carr. (34% and 49%), white spruce *P. glauca* (35% and 50%), and loblolly pine *Pinus taeda* L. (32% and 48%) (Table 3).

The *N*-terminal sequence reported here for the CSTI from *C. siamensis* seeds corresponds to regions close to the *N*-termini of several mature Kunitz-type inhibitors, including BASI (Leah and Mundy, 1989). Similarly, the internal sequences DLNRGLPVRLTPLSG and GLPVLFTPLSGESE (from SDS–PAGE bands 1 and 2) correspond to the reactive site loop regions of many Kunitz-type inhibitors (highlighted in blue, Table 3). For example, the P1 valine specific for subtilisin in BASI (Leah and Mundy, 1989) corresponds to valine 110 in *C. rumphii* protein and to a leucine residue in the sequences of SDS–PAGE bands 1 and 2. The known P1 residues of various Kunitz inhibitors are shown in red in Table 3.

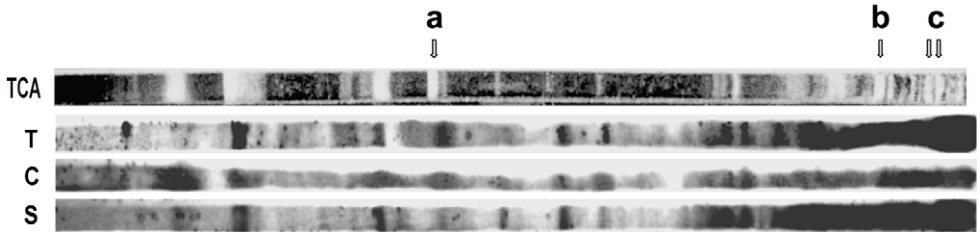


Fig. 2. Heterogeneity of serine proteinase inhibitors from seeds of *Cycas siamensis*. Proteins extracted with water were loaded onto a trypsin affinity column and the 0.1 M HCl eluate separated by IEF. One gel was fixed in TCA, photographed and protein bands a, b and c extracted for subsequent SDS–PAGE and sequencing. Three "gelatin replicas" were obtained from unfixed gels and inhibitors were detected by trypsin (T), chymotrypsin (C) and subtilisin (S).

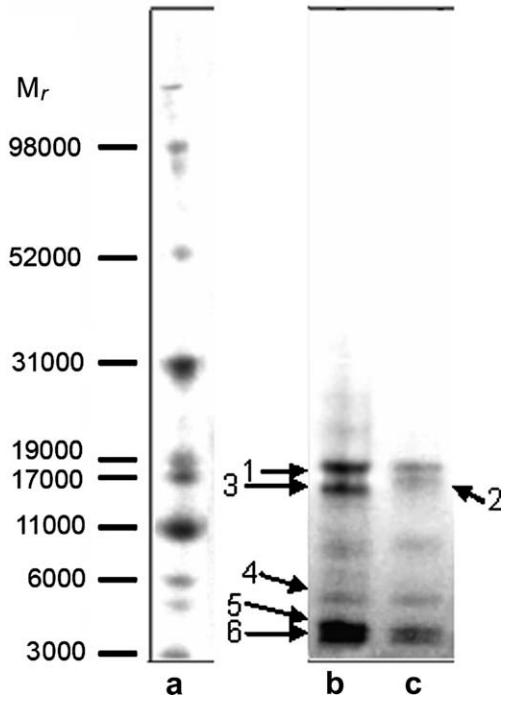


Fig. 3. SDS–PAGE of *C. siamensis* seed protein fractions enriched with serine proteinase inhibitors extracted from IEF gels. (a) M_r marker proteins; (b) IEF band b; (c) IEF band c. Bands 1–6 were sequenced by mass spectrometry or by Edman degradation.

Table 2
Partial amino-acid sequences of serine proteinase inhibitor fractions separated by SDS–PAGE from *Cycas siamensis* seeds

SDS–PAGE band (Fig. 3)	M_r	Partial sequence	Sequencing method
1	18,000	FDENSKVVDLEGXTLKA	E
1	894 peptide[a]	GLPVLFTPLSGESE	MS
2	17000	No sequence	E
2	783 peptide[a]	DLNRGLPVRLTPLSG	MS
3	16,000	FDENSKVVDLEG	E
4	5000	DSEIRVGVDLKVETVALTTXAQ	E
5	4000	PERTFVATGH	E
6	3500	PIXDKLGIYFD	E

M_r, approximate M_r relatively to marker proteins on SDS–PAGE (Fig. 3).
[a] Peptides obtained by trypsin digestion. MS and E, sequenced using mass spectrometry and Edman degradation, respectively.

The *N*-terminal sequence DSEIRVGVDLKVETVAFTTXAQ of SDS–PAGE band 4 overlaps the *C*-terminal end of the reactive loop of Kunitz-type inhibitors while the sequence of SDS–PAGE band 5 (PERTFVATGH) overlaps the reactive site of the cathepsin D inhibitor from *Solanum tuberosum* L., which includes the P1 methionine (Ritonja et al., 1990) specific for chymotrypsin. The sequence of the smallest component (band 6) PIXDKLGIYFD corresponds to the *C*-terminal part of the *C. rumphii* sequence and comparison with other sequences in Table 3 indicates that the unidentified third residue (X) is a cysteine.

These comparisons demonstrate that all of the components sequenced from the Af-T fraction were related in sequence. It is also likely that the smaller components (4–6) were derived from proteolytic processing of the larger M_r 16,000–18,000 components. Indeed, it is possible that all of components were derived from one or more M_r 18,000 precursors. However, the similarity between the patterns of the Af-T, Af-C and Af-MetC fractions indicates that little if any of the proteolysis occurred during affinity purification, despite the use of active enzymes.

3. Discussion

We have isolated a CSTI from *C. siamensis* seeds (called Cs-CSTI) which is highly homologous to a protein encoded by an EST from sporophylls of *C. rumphii* (Brenner et al., 2003b). Both the directly determined sequence of Cs-CSTI and the deduced sequence of the *C. rumphii* protein showed clear homology with typical representatives of the soybean Kunitz inhibitor family (Table 3) and lower but still significant levels of similarity with other proteins known to be related to Kunitz inhibitors, including the taste-modifying protein miraculin (32% identity; Masuda et al., 1995), a miraculin-like protein up-regulated by nematode invasion of potato roots (36% identity; Brenner et al., 1998) and the storage protein sporamin from sweet potato (29% identity; Murakami et al., 1986).

Kunitz-type serine proteinase inhibitors occur widely in various organs of gymnosperms, including seeds (*C. siamensis, Z. vazquezii*), reproductive organs (*C. rumphii* and *G. biloba*) and vegetative parts (*Picea* and *Pinus* species). Also, as in angiosperms, they may also be expressed constitutively in these tissues (*C. siamensis* seeds) or their expression may be induced as a response to infection by pathogens (Pratt et al., 2005) or damage by pests.

The protein characterisation reported here combined with analyses of protein and EST databases demonstrate the presence of Kunitz-type inhibitors in seeds, generative and vegetative organs of representatives of three of the four main classes of gymnosperms: the Cycadopsida, Coniferopsida and Ginkgoopsida. In contrast, EST sequences available in databases for the fourth main class, the Gnetopsida (Ephedrales, Gnetales and Welwitschiales), do not show clear similarity to Kunitz inhibitors.

Analysis of EST databases shows that sequences related to those of the *C. rumphii* Kunitz-type inhibitor and the Cs-CSTI are also present in the prothallium of maidenhair fern (*Adiantum capillus-veneris* L.) (Yamauchi et al., 2005), a non-vascular plant. This suggests that Kunitz inhibitor-type proteins may be widely distributed in more primitive plant taxa as well as in advanced seed plants.

Differences were observed in the range of specificity of the proteinase inhibitors (PIs) in the gymnosperm species studied here. For example, whereas all species in the order Cycadales contained SI, only *C. pectinata* and *C. siamensis* also contained CSTI and CSI. In the Coniferales relatively weak SIs were detected in *Juniperus, Pseudotsuga* and *Taxus* species, some of which also showed weak activity against elastase. No serine PIs were detected in seeds of *Chamaecyparis, Thuja, Abies, Larix, Picea* and *Pinus* species, although analyses of databases indicate that PIs are present in other organs of *Picea* and *Pinus* species. Similar variation in the presence or absence of serine proteinase inhibitors occurs in seeds of angiosperms, within and between species, genera and families (Konarev et al., 2002, 2004).

The Cs-CSTI characterised here represents a single isoform from a complex mixture of forms that differ in their activities against serine proteinases. It is possible that the other isoforms present are also related to Kunitz inhibitors but this remains to be established. Their range of specificity, to subtilisin alone, chymotrypsin and subtilisin, and subtilisin, trypsin and chymotrypsin is also consistent with the specificities of other Kunitz inhibitors (Terada et al., 1994). However, a similar range of specificity has also been reported for another major family of PIs, the potato inhibitor 1

Table 3
Alignment of the amino-acid sequence of the putative Kunitz inhibitor-like protein from sporophylls of *C. rumphii* (encoded by EST DR061860) with those of proteinase inhibitors isolated from *C. siamensis* seeds, known Kunitz-type proteinase inhibitors from various angiosperms and proteins encoded by ESTs of gymnosperms

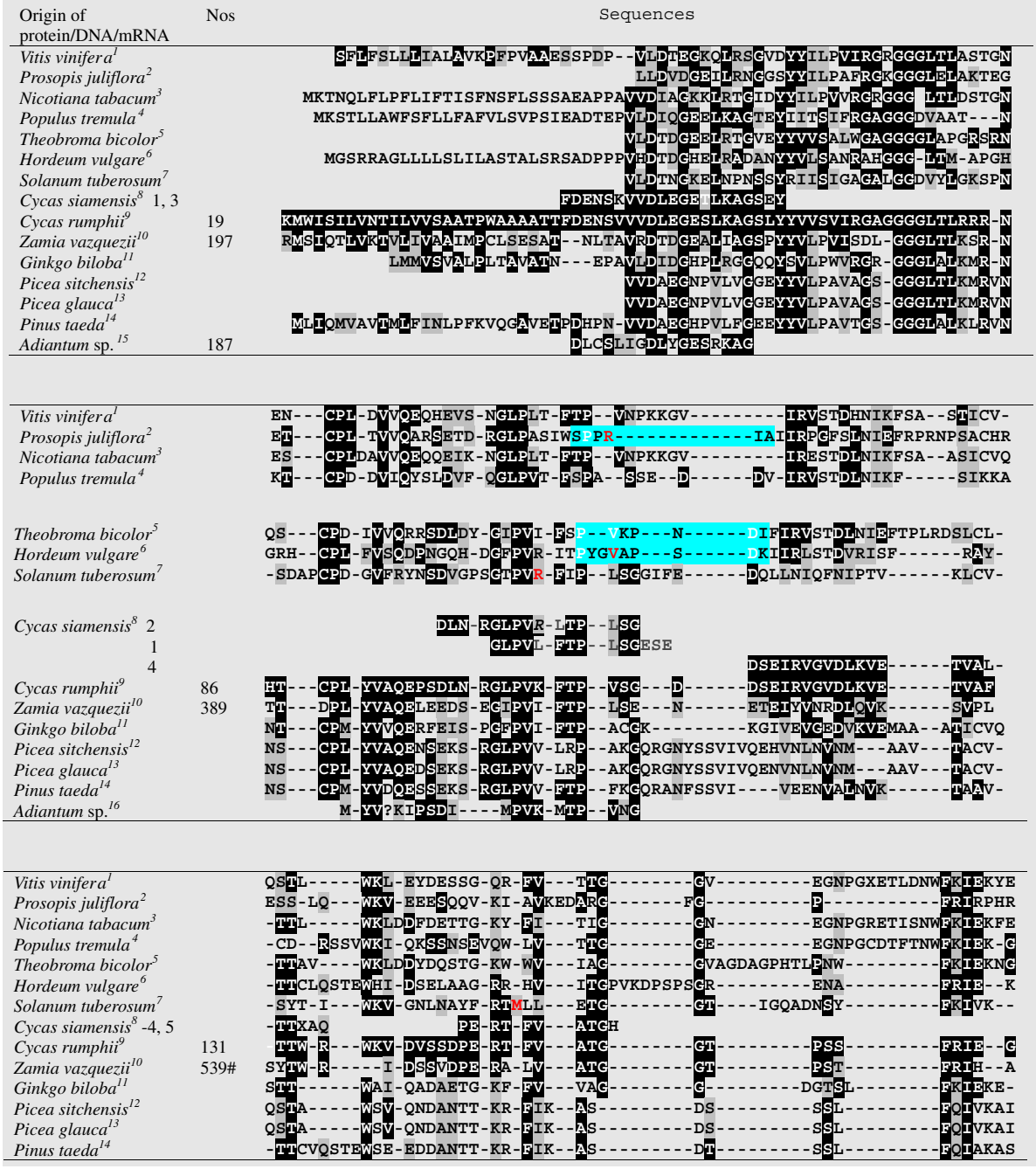
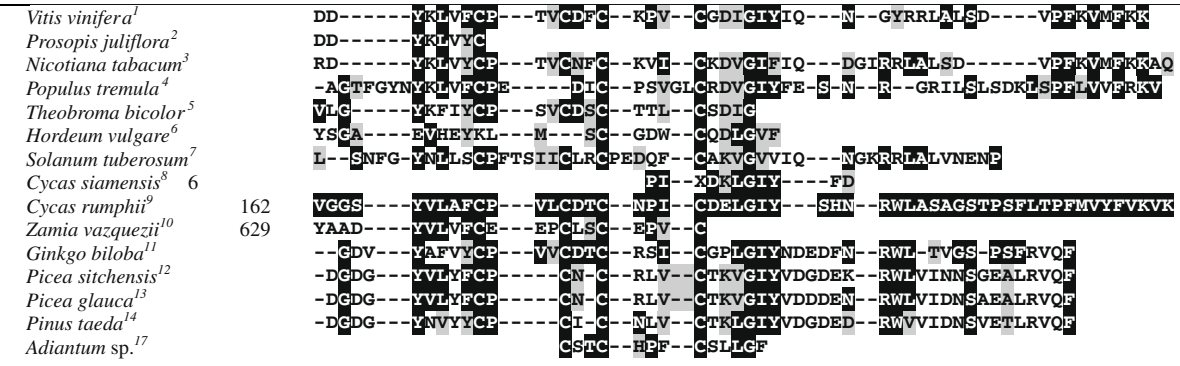


Table 3 (*continued*)



Residues identical to those of the *Cycas rumphii* protein (DR061860) are shown in black boxes; conservative substitutions are black letters in grey boxes; non-related residues are in black letters in white boxes; known P1 residues in the reactive sites of inhibitors are shown in red. Reactive site loops are highlighted in blue. Nos.: residues numbers of the putative *Cycas rumphii* Kunitz inhibitor-like protein and nucleotide numbers for the *Zamia vazquezii* sequence. 1–6 SDS–PAGE bands (Fig. 3 and Table 2).

[1]CAN81015, hypothetical protein of *Vitis vinifera* (Velasco et al., 2006).

[2]P32733 Kunitz-type trypsin inhibitor (alpha chain) from seeds of *Prosopis juliflora* (Negreiros et al., 1991).

[3]AAC49969 tumor-related protein of *Nicotiana tabacum* (Karrer et al., 1998).

[4]CAI77767 Kunitz trypsin inhibitor from *Populus tremula* (Ingvarsson, 2005).

[5]AAL85661 trypsin inhibitor from *Theobroma bicolor* (Silva and Figueira, 2001).

[6]CAA78305 bifunctional alpha-amylase/subtilisin inhibitor from *Hordeum vulgare* seeds (Leah and Mundy, 1989).

[7]P58521 cathepsin D inhibitor from potato (Ritonja et al., 1990).

[8]Trypsin/chymotrypsin/subtilisin inhibitor from *Cycas siamensis* seeds (present study a and b Fig. 3).

[9]DR061860 *Cycas rumphii* sporophyll mRNA/cDNA translated sequence (Brenner et al., 2003b).

[10]FD767966 *Zamia vazquezii* female gametophytes mRNA/cDNA translated sequence (dePamphilis et al., 2008).

[11]EX935030 *Ginkgo biloba* microsporophyll mRNA/cDNA translated sequence (Brenner et al., 2005).

[12]ES856607 *Picea sitchensis* mRNA/cDNA translated sequence (Ralph et al., 2007).

[13]EX381505 *Picea glauca* mRNA/cDNA translated sequence (Boyle et al., 2007).

[14]DR093270 *Pinus taeda* mRNA/cDNA translated sequence (Pratt et al., 2005).

[15–17]BP920714, BP921733 and BP914097 *Adiantum capillus-veneris* prothallium mRNA/cDNA translated sequences (Yamauchi et al., 2005). In cDNA sequence DR061860 reading frame "+3" for nucleotides from 1 to 415 was changed to "+1" for nucleotides from 415 to 670 possibly due to error at sequencing so amino-acid residues 137 and 138 are under question.

[#]Residues NISSQ corresponding to nucleotides 524–538 not included.

family. Our previous studies have shown that this is the most widely distributed inhibitor family in angiosperms (Konarev et al., 2004) and analyses of EST databases indicate that related proteins are indeed present in all four classes of gymnosperms (the Cycadopsida, Coniferopsida, Ginkgoopsida and Gnetopsida) and also in more primitive plants, including the maidenhair fern (*A. capillus-veneris*, Moniliformopses, Filicophyta) and members of the Lycopodiophyta, Bryophyta and Marchantiophyta (alignments not shown). This confirms the wide distribution of this family of inhibitors and demonstrates an ancient origin. However, our studies have failed to show their presence in mature gymnosperm seeds.

Our results therefore demonstrate that seeds of many cycads and conifers contain inhibitors of serine proteinases, but with the exception of *Cycas* spp. these usually inhibit only subtilisin. This specificity may account for their failure to have been reported previously, with most studies focusing on inhibitors of trypsin and chymotrypsin. The dominance of subtilisin inhibitors in seeds of conifers and cycads, and their presence also in seeds of many Asterids could relate to the presence of subtilisin-like enzymes in major pathogens of these species.

4. Conclusions

Representatives of two gymnosperm orders, Coniferales and Cycadales differ in the presence of inhibitors of serine proteinases in seeds. Inhibitory activity towards subtilisin was most widespread with only some *Cycas* species, also showing the presence of inhibitory activity towards trypsin and chymotrypsin. Partial sequencing of an isoform from *C. siamensis* seeds (Cs-CSTI) showed similarity to Kunitz-type inhibitors from angiosperms. Comparison of the partial sequence of Cs-CSTI with those in EST databases revealed the presence of a highly homologous (83% identity) protein expressed in sporophylls of *C. rumphii*, as well as related proteins in various organs of *Z. vazquezii* (52% identity), *G. biloba* (38% identity), *Picea* (32% identity) and *Pinus* species. Thus, Kunitz-type inhibitors or closely related proteins are present in representatives of three of the four main classes of gymnosperms: the Cycadopsida, Coniferopsida and Ginkgoopsida.

5. Experimental

5.1. Plant material

Seeds of various gymnospermae were obtained from the Komarov Botanical Institute (St. Petersburg, Russia) and Chiltern Seeds, Cumbria, UK.

Order Coniferales: Family Cephalotaxaceae: *C. harringtonia* (Knight ex Forbes) K. Koch. Cupressaceae: *Chamaecyparis pisifera* (Siebold et Zicc.) Endl.; *Chamaecyparis lawsoniana* (A. Murr.) Parl.; *Juniperus communis* L.; *J. phoenicea* L.; *J. virginiana* L.; *Thuja occidentalis* L. Pinaceae: *Abies koreana* Wils.; *A. concolor* (Gord.et Glend.) Lindl.ex Hildebr.; *A. fraseri* (Pursh) Poir.; *A. veitchii* Lindl.; *Larix decidua* Mill.; *L. kaempferi* (Lamb.) Carr.; *Larix x lubarskii* Sucacz.; *Picea abies* (L.) Karst.; *Pinus sibirica* Du Tour; *P. sylvestris* L.; *P. menziesii* (Mirb.) Franco var. glauca; *Taxus canadensis* Marsh; *T. cuspidata* Siebold et Zucc. ex Endl.; *T. baccata* L.

Order Cycadales. Cycadaceae: Cycas micholitzii Dyer; *C. media* R. Br.; *C. pectinata* Buch.-Ham.; *C. siamensis* Miq. Zamiaceae: *Dioon edule* Lindl.; *Macrozamia communis* L. A. S. Johnson; *M. miquelii*; *M. riedlei*; *Zamia fisheri* Miq.; *Z. furfuraceae* L.f.; *Z. pumila* L.

5.2. Proteins

5.2.1. Protein extraction

Seeds were dehulled, ground, dried and defatted with ice-cold acetone. Meal was extracted with water (1:4, w/v), followed by centrifugation for 5 min at 12,000*g* and the clear supernatant was used for isoelectric focusing (IEF).

5.2.2. Separation

Analytical IEF was carried out on Servalyt Precotes pH 3–10 gels (Serva) with 4.5 or 10 cm between electrodes (Konarev et al., 2000) and preparative IEF on Immobiline DryStrip gels pH 3–10 NL (Amersham Pharmacia Biotech). Marker proteins with pI ranging from 3.5 to 10.7 (Serva) were used to calculate p*I* values.

5.2.3. Detection of inhibitors

The "gelatin replicas" method was used to detect inhibitors separated by IEF (Konarev et al., 2000, 2004). Briefly, the IEF gel was applied sequentially to layers of gelatin on undeveloped non-transparent photographic film (for 2, 5, 20 and 30 min, respectively). The "gelatin replicas" containing inhibitors absorbed from the IEF gel were then applied to 0.8% (w/v) agarose gels containing 0.1 M Na_2HPO_4 (pH 9) and one of four serine proteinases (Sigma): trypsin (1 μg/ml), chymotrypsin (10 μg/ml), elastase (4 μg/ml) and subtilisin (0.3 μg/ml). Incubation for 30 min at 45 °C allowed the proteases to digest the gelatin on the photographic film with the positions of inhibitors being detected as dark "islands" of undigested gelatin on the photographic layer. The volume of extract required to give clear banding patterns were used to classify the activity as strong (below 0.3 μl), weak (up to 2.5 μl) and low or absent (no bands observed at 5 μl).

5.2.4. Purification of inhibitors

Affinity chromatography on trypsin, chymotrypsin and inactived methyl-chymotrypsin linked to CNBr-activated Sepharose 4B according to the manufacturer's protocol was used to isolate proteinase inhibitors. Inactive methylated chymotrypsin was prepared by treatment of native chymotrypsin with methyl 4-nitrobenzenesulfonate (Ryan and Feeney, 1975). Fifteen grams of *C. siamensis* seed flour was extracted with water (1:4), centrifuged and ammonium acetate added to the extract to a final concentration of 0.2 M. The extract was then applied to 10 ml of affinity matrix and washed with of 0.2 M ammonium acetate, 0.1 M Na_2HCO_3 or 0.1 M Na acetate and water (50 ml of each) and then eluted with 0.015 M HCl. With the methyl-chymotrypsin affinity column, the proteinase inhibitors were eluted with water, without prior washing and freeze-dried.

For micropreparative IEF, fractions were loaded onto two Immobiline DryStrip pH 3–10 NL gels and inhibitor bands located using gelatin replicas of one gel. The second gel was fixed in 10% (v/v) TCA revealing proteins as opaque bands. Zones containing protein were excised and TCA was removed by washing with cold acetone. Proteins were then extracted from the gel with 20% (v/v) ethanol containing 0.1% (v/v) TFA, followed by 30% (v/v) acetonitrile in 0.03 M ammonium bicarbonate and 4% (v/v) formic acid (for 1 h each). Extracts were freeze-dried and separated by SDS–PAGE using NuPage 4–12% gradient gels (Invitrogen). MultiMark colored standards (Invitrogen) were used to estimate the M_r on SDS–PAGE. Gels were stained either with silver (Yan et al., 2000) or Colloidal Coomassie G-250 (Sigma).

5.2.5. Sequence analysis

For sequencing by mass spectrometry proteins were excised from the gel, destained, and in-gel digested with trypsin (Jensen et al., 1999). The digests were concentrated and desalted using Zip-Tips (Millipore) and speed vacuumed to dryness. The peptides were then dissolved in 70% (v/v) methanol containing 1% (v/v) formic acid and sonicated for 3 min. The peptides were loaded into nanoflow tips (Waters) using gel loader tips (Eppendorf). ESI-MS was performed as described previously (Conners et al., 2007).

For *N*-terminal sequencing proteins were transferred to PVDF membrane using a semi-dry method and CAPS buffer. *N*-terminal sequences were determined by automated Edman degradation at the University of Cambridge PNAC facility (Cambridge, 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 (2002) and Fellowship Grants from Rothamsted International (2005) and the BBSRC Underwood fund (2006). The authors are grateful to Dr. G.Yu. Konechnaya, Dr. N.N. Arnautov, Dr. E.M. Arnautova, Dr. I.A. Pautova, Dr. N.G. Kutyavina and other specialists from V.L. Komarov Botanical Institute for providing plant accessions and to Professor Alexei V. Konarev (Vavilov Institute of Plant Industry) for providing access to equipment.

References

Bowe, L.M., Coat, G., de Pamphilis, C.W., 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. Proc. Natl. Acad. Sci. USA 97, 4092–4097.

Boyle, B., Rigault, P., Jacobi, V., Gibbs, C., Villeneuve, A., Dionne, J.P., Cooke, J., Mackay, J., 2007. Arborea Phase 2 – EST sequencing in *Picea glauca* (white spruce), unpublished. EX381505.

Brenner, E.D., Lambert, K.N., Kaloshian, I., Williamson, V.M., 1998. Characterization of LeMir, a root-knot nematode-induced gene in tomato with an encoded product secreted from the root. Plant Physiol. 118, 237–247.

Brenner, E.D., Stevenson, D.W., McCombie, R.W., Katari, M.S., Rudd, S.A., Mayer, K.F., Palenchar, P.M., Runko, S.J., Twigg, R.W., Dai, G., Martienssen, R.A., Benfey, P.N., Coruzzi, G.M., 2003a. Expressed sequence tag analysis in *Cycas*, the most primitive living seed plant. Genome Biol. 4, R78.

Brenner, E.D., Twigg, R.W., Runko, S.J., Katari, M.S., Dedhia, N.N., O'Shaughnessy, A.L., Balija, V., Martienssen, R.A., McCombie, R.W., Benfey, P., Coruzzi, G., Stevenson, D., 2003b. Expressed tag sequences from *Cycas rumphii* sporophyll (w/o ovule), unpublished. DR061860.

Brenner, E.D., Twigg, R.W., Runko, S.J., Katari, M.S., Dedhia, N.N., O'Shaughnessy, A.L., Balija, V., Martienssen, R.A., McCombie, R.W., Benfey, P., Coruzzi, G., Stevenson, D., 2005. Expressed tag sequences from *Ginkgo* microsporophyll, unpublished. EX935030.

Chaw, S.M., Parkinson, C.L., Cheng, Y., Vincent, T.M., Palmer, J.D., 2000. Seed plant phylogeny inferred from all three plant genomes: monophyll of extant gymnosperms and origin of Gnetales from conifers. Proc. Natl. Acad. Sci. USA 97, 4086–4091.

Conners, R., Konarev, A.V., Forsyth, J., Lovegrove, A., Marsh, J., Joseph-Horne, T., Shewry, P.R., Brady, R.L., 2007. An unusual helix-turn-helix protease inhibitory motif in a novel trypsin inhibitor from seeds of *Veronica* (*Veronica hederifolia* L.). J. Biol. Chem. 282, 27760–27768.

de Pamphilis, C.W., Leebens-Mack, J.H., Ma, H., Soltis, D.E., Soltis, P.S., Clifton, S., 2008. Ancestral Angiosperm Genome Project, unpublished. FD767966.

Ingvarsson, P.K., 2005. Molecular population genetics of herbivore-induced protease inhibitor genes in European aspen (*Populus tremula* L., Salicaceae). Mol. Biol. Evol 22, 1802–1812 CAI77767.

Jensen, O.N., Wilm, M., Shevchenko, A., Mann, M., 1999. Sample preparation methods for mass spectrometric peptide mapping directly from 2DE gels. In: Link, A. (Ed.), 2-D Proteome Analysis Protocols Methods in Molecular Biology Series, vol. 112. Humana Press, pp. 513–530.

Karrer, E.E., Beachy, R.N., Holt, C.A., 1998. Cloning of tobacco genes that elicit the hypersensitive response. Plant Mol. Biol. 36, 681–690 AAC49969.

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., 2002. Serine proteinase inhibitors in the Compositae: distribution, polymorphism and properties. Phytochemistry 59, 279–291.

Konarev, Al.V., Griffin, J., Konechnaya, G.Yu., Shewry, P.R., 2004. Serine proteinase inhibitors in seeds of the Asteridae. Phytochemistry 65, 3003–3020.

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

Korsinczky, M.L.J., Schirra, H.J., Craik, D.J., 2004. Sunflower trypsin inhibitor-1. Curr. Prot. Peptide Sci. 5, 351–364.

Leah, R., Mundy, J., 1989. The bifunctional alpha-amylase/subtilisin inhibitor of barley: nucleotide sequence and patterns of seed-specific expression. Plant Mol. Biol. 12, 673–682 CAA78305.

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.

Masuda, Y., Nirasawa, S., Nakaya, K., Kurihara, Y., 1995. Cloning and sequencing of a cDNA encoding a taste-modifying protein, miraculin. Gene 161, 175–177.

McInnis, S.M., Newton, C.H., Sutton, B.S.C., 1997. Molecular characterization of a white spruce (*Picea glauca*) 2S albumin gene and related pseudogene, unpublished. AAD10958.

Murakami, S., Hattori, T., Nakamura, K., 1986. Structural differences in full-length cDNAs for two classes of sporamin, the major soluble protein of sweet potato tuberous roots. Plant Mol. Biol. 7, 343–355.

Negreiros, A.N., Carvalho, M.M., Xavier Filho, J., Blanco-Labra, A., Shewry, P.R., Richardson, M., 1991. The complete amino acid sequence of the major Kunitz trypsin inhibitor from the seeds of *Prosopsis juliflora*. Phytochemistry 30, 2829–2833.

Pratt, L., Cordonnier-Pratt, M.M., Morse, A., Davis, J., Lorenz, W.W., Neale, D., Dean, J.F.D., 2005. An EST database from pitch canker resistant loblolly pine (*Pinus taeda*) shoot tips, unpublished. DR093270.

Ralph, S., Kolosova, N., Oddy, C., Chun, E., Reid, K.E., Liao, N., Jancsik, S., Moore, R., Mayo, M., Wagner, S., King, J., Yanchuk, A., Holt, R., Jones, S., Marra, M., Ritland, C.E., Ritland, K., Bohlmann, J., 2007. Expressed sequence tags from multiple spruce cDNA libraries, unpublished. ES856607.

Ritonja, A., Krizaj, I., Mesko, P., Kopitar, M., Lucovnik, P., Strukelj, B., Pungercar, J., Buttle, D.J., Barrett, A.J., Turk, V., 1990. The amino acid sequence of a novel inhibitor of cathepsin D from potato. FEBS Lett. 267, 13–15.

Ryan, D.S., Feeney, R.E., 1975. The interaction of inhibitors of proteolytic enzymes with 3-methylhistidine-57-chymotrypsin. J. Biol. Chem. 250, 843–847.

Salmia, M.A., Mikola, J.J., 1980. Inhibitors of endogenous proteinases in the seeds of Scots pine, *Pinus sylvestris*. Physiol. Plant 48, 126–130.

Sawano, Y., Miyakawa, T., Yamazaki, H., Tanokura, M., Hatano, K., 2007. Purification, characterization, and molecular gene cloning of an antifungal protein from *Ginkgo biloba* seeds. Biol. Chem. 388, 273–280.

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. In: Callow, J. (Ed.), Advances in Botanical Research, vol. 26. Academic Press, pp. 135–192.

Shutov, A.D., Kakhovskaya, I.A., Braun, H., Baumlein, H., Muntz, K., 1995. Legumin-like and vicilin-like seed storage proteins: evidence for a common single-domain ancestral gene. J. Mol. Evol. 41, 1057–1069.

Silva, C.R.S., Figueira, A., 2001. Phylogeny of genus *Theobroma* and *Herrania* (Sterculiaceae) based on sequences of the nuclear gene trypsin inhibitor, unpublished. AAL85661.

Terada, S., Katayama, H., Noda, K., Fujimura, S., Kimoto, E., 1994. Amino acid sequences of Kunitz family subtilisin inhibitors from seeds of *Canavalia lineata*. J. Biochem. 115, 397–404.

Velasco, R., Zharkikh, A., Troggio, M., Bhatnagar, S., Pindo, M., Cartwright, D., Coppola, G., Eldredge, G., Vezzulli, S., Malacarne, G., Mitchell, J., Stefanini, M., Segala, C., Gutin, N., Grando, M.S., Pruss, D., Dematte, L., Cestaro, A., Toppo, S., Fontana, P., Skolnick, M., Gutin, A., Salamini, F., Viola, R., 2006. The first genome sequence of an elite grapevine cultivar (Pinot noir *Vitis vinifera* L.): coping with a highly heterozygous genome, unpublished. CAN81015.

Weng, J.-Ke., Tanurdzic, M., Chapple, C., 2005. Functional analysis and comparative genomics of expressed sequence tags from the lycophyte *Selaginella moellendorffii*. BMC Genom. 6, 85.

Yamauchi, D., Sutoh, K., Kanegae, H., Horiguchi, T., Matsuoka, K., Fukuda, H., Wada, M., 2005. Analysis of expressed sequence tags in prothallia of *Adiantum capillus-veneris*. J. Plant Res. 118, 223–227.

Yan, J.X., Wait, R., Berkelman, T., Harry, R.A., Westbrook, J.A., Wheeler, C.H., Dunn, M.J., 2000. A modified silver staining protocol for visualization of proteins compatible with matrix-assisted laser desorption/ionization and electrospray ionization-mass spectrometry. Electrophoresis 21, 3666–3672.