Nucleic Acids Research, 2002, Vol. 30, No. 1 347–348

PLANT-PIs: a database for plant protease inhibitors and their genes

F. De Leo, M. Volpicella, F. Licciulli[1], S. Liuni[2], R. Gallerani and L. R. Ceci[2,*]

Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, Via Amendola 165/A, Bari, Italy, [1]Area di Ricerca, CNR, Via Amendola, Bari, Italy and [2]Centro di Studio sui Mitocondri e Metabolismo Energetico, CNR, Via Amendola 165/A, 70126 Bari, Italy

Received August 21, 2001; Revised and Accepted September 20, 2001

ABSTRACT

PLANT-PIs is a database developed to facilitate retrieval of information on plant protease inhibitors (PIs) and related genes. For each PI, links to sequence databases are reported together with a summary of the functional properties of the molecule (and its mutants) as deduced from literature. PLANT-PIs contains information for 351 plant PIs, plus several isoinhibitors. The database is accessible at http://bighost.area.ba.cnr.it/PLANT-PIs.

INTRODUCTION

Plant protease inhibitors (PIs) are small proteins, generally present at high concentration in storage tissues (up to 10% of protein content), but also detectable in leaves in response to the attack of insects and pathogenic microorganisms (1). PIs' contribution to plant defense mechanisms relies on inhibition of proteases present in insects' guts or produced by micro-organisms, causing a reduction in the availability of amino acids necessary for their growth and development. As the role of inhibitors is simply achieved by the activation of single genes, several transgenic plants expressing PIs have been produced in the last 15 years and tested for enhanced defensive capacities, with particular efforts against pest insects.

Among plant PIs, inhibitors active toward the four mechanistic classes of proteases have been described. PIs active against serine, cysteine and metallocarboxy-proteases are ubiquitous, while inhibitors active towards aspartic proteases have not been detected in seeds (2). The activity of PIs is due to their capacity to form stable complexes with target proteases, blocking, altering or preventing access to the enzyme active site.

The presence of multigene families has been reported for several plant PIs (1).

A considerable number of reviews have been written on PIs and plant PIs, covering specific areas of studies on these molecules. The most recent articles on PIs structure and their interactions with proteases are by Laskowski and Qasim (3), dealing specifically with serine PIs, and by Bode and Huber (4), also exemplifying possible mechanisms of inhibition for cysteine and metallo-protease inhibitors. For plant PIs, in addition to the specific reviews by Valueva and Mosolov (2,5) on seed PIs, other articles with large sections describing PIs characteristics, functions and utilization are those by Shewry and Lucas (6), Jouanin *et al.* (7), Schuler *et al.* (8), Hilder and Boulter (9) and Shewry (10). In these reports, other references for previous reviews can be found.

Table 1. Families of plant protease inhibitors reported in PLANT-PIs and their links to InterPro

Plant protease inhibitor family	PLANT-PIs code	InterPro accession no.
Bowman–Birk serine proteinase inhibitors	BBI	IPR000877
Cereal trypsin/α-amylase inhibitors	BRI	IPR001768
Cysteine proteinase inhibitors	CYS	IPR000010
Metallocarboxypeptidase inhibitors	MCI	Not available
Mustard trypsin inhibitors	MSI	Not available
Potato type I inhibitors	PI1	IPR000864
Potato type II proteinase inhibitors	PI2	IPR003465
Serpin	SPI	IPR000215
Soybean trypsin inhibitors (Kunitz)	KNI	IPR002160
Squash inhibitors	SQI	IPR000737

The capacity of some insects to overcome the anti-nutritional activity of PIs by over-expression of insensitive proteases was first demonstrated 6 years ago (11,12), and has been continuously investigated ever since (13 and references therein). Studies on PI–protease interaction and their evolution constitute another attractive field of investigations for scientists interested in research on plant PIs. The possibility of designing new PIs with higher or different activities is now being exploited (14–16).

The large amount of sequences for plants' PIs (either as polypeptides or nucleic acids) and the ever-more increasing amount of data concerning activity, structure, mutational analysis and expression in transgenic plants and heterologous systems, prompted us to develop a database correlating information contained in primary sequence databases (EMBL and SWISS-PROT) to functional analysis of the proteins reported in literature.

DESCRIPTION OF THE DATABASE

Annotation of PIs in the database is organized according to the well-known classification of PI families (Table 1) as also reported in InterPro (InterPro, Integrated Resource of Protein Domains and Functional Sites, is available at http://www.ebi.ac.uk/interpro/). This classification is based on the

*To whom correspondence should be addressed. Tel: +39 080 5443311; Fax: +39 080 5443317; Email: luigi.ceci@area.ba.cnr.it

protein primary structure. With the exception of the Cysteine protease inhibitor and Metallocarboxypeptidase inhibitor families, all the reported families of PIs contain inhibitors of serine proteases, the most diffused and studied PIs. It must be noted however that these families can also contain PIs that are not active toward serine proteases. This is the case of inhibitors of aspartic proteases, which are present in the Kunitz and in the Cysteine protease inhibitors families and some potato cysteine protease inhibitors that belong to the Kunitz family. Also proteins with different function from inhibition of proteases have been described, but they have not been reported in this compilation.

Each entry in the database contains links to available sequence informative databases: EMBL (http://www.ebi.ac.uk/), SWISS-PROT (http://www.ebi.ac.uk/swissprot) and the structural database PDB (Protein Data Bank at http://pdb-browsers.ebi.ac.uk/).

Additional information in each entry

Inhibited protease. Literature data about inhibitory activity of PIs against specific proteases are reported. Activity is usually described as the equilibrium dissociation constant (K_i) of the complex between the PI and the inhibited protease. Other activity values eventually used are also reported.

Reactive site. PIs reacting with proteases in a substrate-like mechanism possess the so-called reactive site: a scissile peptide bond indicated as P_1-P_1' (17). Identified or putative reactive site(s) are reported in the database.

Canonic serine PIs exhibit an external loop acting as the primary binding segment with the protease and containing the reactive site. Identity of P_1 residues determines the specificity of inhibited serine protease: P1 = Arg or Lys is specific of inhibitors of trypsin-like enzymes; P1 = Trp, Phe, Tyr, Leu, Met can be found in inhibitors of chymotrypsin-like enzymes; P1 = Ala specifies for inhibitors of elastase-like enzymes; etc. Note that some PIs may exhibit several reactive sites able to interact with additional copies of the same protease or with different proteases. 'Multi-headed' PIs arise either from multimeric association of single-chain inhibitors (as in the Potato I family) or from the presence of different reactive sites on a single-polypeptide derived structure. This is the case of the Bowman–Birk double-headed inhibitors, derived from tandem homology regions on a single polypeptide, and of the Potato II double-headed inhibitors in which a second reactive domain is obtained by proper folding of the two polypeptide termini (18).

TransPlant expression. Expression of PI genes in transgenic plants is reported.

Heterologous expression. In vitro expression of plant PIs has often been performed to study their activities toward specific insects. Available data on *in vitro* expression of plant PIs are reported.

Mutational analysis. Mutational analysis of PIs is largely used to study alterations in the specificity toward target enzymes and/or variations in inhibitor activity. Mutated residues and eventually new activity values are reported.

The PLANT-PIs database has been developed by analysis of the literature and of sequences deposited at the EMBL and SWISS-PROT databases. Sequences retrieved by means of the sequence retrieval system (SRS) service at the European Bioinformatics Institute (EBI, http://srs.ebi.ac.uk), have been catalogued in the database according to their family classification, together with data from literature. To complete the analysis of sequence databases, a sequence from each family has also been used for FastA analysis of the whole EMBL database at the address http://www2.ebi.ac.uk/fasta3/. Updating of the database was in July 2001.

DATABASE AVAILABILITY AND CITATION

Retrieval of information for specific entries from the database is possible by the SRS facility at http://bighost.area.ba.cnr.it/srs/ (SeqRelated field).

Users of the database should cite the present publication as reference. Comments, corrections and new entries are welcome.

REFERENCES

1. Ryan,C.A. (1990) Protease inhibitors in plants: genes for improving defenses against insects and pathogens. *Annu. Rev. Phytopath.*, **28**, 425–449.
2. Valueva,T.A. and Mosolov,V.V. (1999) Protein inhibitors of proteinases in seeds: 1. Classification, distribution, structure, and properties. *Russian J. Plant Physiol.*, **46**, 362–378.
3. Laskowski,M.,Jr and Qasim,M.A. (2000) What can the structures of enzyme–inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim. Biophys. Acta*, **1477**, 324–337.
4. Bode,W. and Huber,R. (2000) Structural basis of the endoproteinase–protein inhibitor interaction. *Biochim. Biophys. Acta*, **1477**, 241–252.
5. Valueva,T.A. and Mosolov,V.V. (1999) Protein inhibitors of proteinases in seeds: 2. Physiological functions. *Russian J. Plant Physiol.*, **46**, 379–387.
6. Shewry,P.R. and Lucas,J.A. (1997) Plant proteins that confer resistance to pests and pathogens. *Adv. Botan. Res.*, **26**, 135–192.
7. Jouanin,L., Bonade-Bottino,M., Girard,C., Morrot,G. and Giband,M. (1998) Transgenic plants for insect resistance. *Plant Sci.*, **131**, 1–11.
8. Schuler,T.H., Poppy,G.M., Kerry,B.R. and Denholm,I. (1998) Insect-resistant transgenic plants. *Trends Biotechnol.*, **16**, 168–175.
9. Hilder,V.A. and Boulter,D. (1999) Genetic engineering of crop plants for insect resistance – a critical review. *Crop Protection*, **18**, 177–191.
10. Shewry,P.R. (1999) Enzyme inhibitors of seeds: types and properties. In Shewry,P.R. and Casey,R. (eds), *Seed Protein.* Kluwer Academy Publishers, Amsterdam, The Netherlands, pp. 587–615.
11. Broadway,R.M. (1995) Are insects resistant to plant proteinase inhibitors? *J. Insect Physiol.*, **41**, 107–116.
12. Jongsma,M.A., Bakker,P.L., Peters,J., Bosch,D. and Stiekema,W.J. (1995) Adaptation of *Spodoptera exigua* larvae to plant proteinase inhibitors by induction of gut proteinase activity insensitive to inhibition. *Proc. Natl Acad. Sci. USA*, **92**, 8041–8045.
13. Mazumdar-Leighton,S. and Broadway,R.M. (2001) Identification of six chymotrypsin cDNAs from larval midgets of *Helicoverpa zea* and *Agrotis ipsilon* feeding on the soybean (Kunitz) trypsin inhibitor. *Insect Biochem. Mol. Biol.*, **31**, 633–644.
14. Koiwa,H., Shade,R.E., Zhu-Salzman,K., Subramanian,L., Murdock,L.L., Nielsen,S.S., Bressan,R.A. and Hasegawa,P.M. (1998) Phage display selection can differentiate insecticidal activity of soybean cystatins. *Plant J.*, **14**, 371–379.
15. Beekwilder,J., Shipper,B., Bosh,D. and Jongsma,M.A. (2000) Characterization of potato proteinase inhibitor II reactive site mutants. *Eur. J. Biochem.*, **267**, 1975–1984.
16. Volpicella,M., Ceci,L.R., Gallerani,R., Jongsma,M.A. and Beekwilder,J. (2001) Functional expression on bacteriophage of the mustard trypsin inhibitor MTI-2. *Biochem. Biophys. Res. Commun.*, **280**, 813–817.
17. Schechter,I. and Burger,A. (1967) On the size of the active site in proteases. I. Papain, *Biochem. Biophys. Res. Commun.*, **27**, 157–162.
18. Lee,M.C.S., Scanlon,M.J., Craik,D.J. and Anderson,M.A. (1999) A novel two-chain protease inhibitor generated by circularisation of a multidomain precursor protein. *Nature Struct. Biol.*, **6**, 526–530.