https://doi.org/10.56093/ijas.v90i9.106624

Indian Journal of Agricultural Sciences **90** (9): 1763–8, September 2020/Article

Identification and *in-silico* characterization of Serpin genes in legumes genomes

PAWAN S MAINKAR[1], MANOJ M L[2], DEEPANSHU JAYASWAL[3], YAMINI AGARWAL[4],
RAKESH K PRAJAPAT[5] and REKHA KANSAL[6]

ICAR-National Institute for Plant Biotechnology, Lal Bahadur Shastri Building, Pusa Campus,
New Delhi 110 012

Received: 19 December 2019; Accepted: 2 January 2020

ABSTRACT

Serine protease inhibitors (serpins) are a unique family of protease inhibitor containing mobile reactive center loop. The availability of genome sequences of pigeonpea, soybean, commanbean, cowpea, mungbean and adzuki bean provided an opportunity to search for the serpin genes. A total of seven serpin genes were identified in the above legume genomes. Phylogenetic analysis of legume serpins, devided them into two major clades (Clade-I:VanSerpin, VraSerpin, VunSerpin, PvuSerpin and clade-II: GmaSerpin1, GmaSerpin2 and CcaSerpin) based on conserved reactive center (P2-P1'), domains and exon-intron boundary in the gene structure. The common reactive center of all the seven serpins indicated their ability to target cognate proteases and play an important role in response to insects, fungal pathogens and also the drought.

Key words: Legumes, Proteases, Reactive center loop, Serine protease inhibitor, Serpins

Serpins are a unique family of serine protease inhibitor and belongs to inhibitor family I4 (Rawling *et al.* 2018). The name was derived from the **ser**ine protease inhibitor as most of them target the chymotrypsin type of serine proteases (Carrell *et al.* 1985, Silverman *et al.* 2001).They are large protein molecules, with size ranging from 330-500 amino acid residues and molecular mass of 40 to 60 kDa (Gettins 2002) and are mostly present in the storage tissue such as seeds.In wheat grain they provide direct defense against insect pests and other organism proteases that attack the endosperm and are present up to 4% of the total protein (Hejgaard *et al.* 2007 and Roberts *et al.* 2008). The native structure of serpins composed of 9α helices, 3β sheets and a mobile reactive center loop (RCL) of 20 amino acid peptide containing reactive site for proteases (Gooptu and Lomas 2009). The mobile RCL exhibits an exposed bait loop sequence on the surface core of serpin proteins that forms a conformational change to trap the target proteases (Silverman *et al.* 2001, Whisstock and Bottomley 2006). Hence the inhibitory serpins act as suicide protease inhibitors.It inserts an extra anti parallel strand into the main β-sheet A of serpins and flings the protease 70 A^0 to the bottom of the serpin and deforms the active site of proteases (Huntington *et al.* 2000, Dementiev *et al.* 2006).The protease inhibitors can interact with proteases in two different ways, the canonical inhibitors like, Bowman Birk, Potato inhibitor I/II and Kunitz, squash and Kazal inhibitor adopt the standard mechanism for their inhibitory action against target proteases (Krowarsch *et al.* 2003). Their protein structure consists of a rigid bait loop and helps to form non-covalent, reversible complex with proteases. Therefore, the inhibitors in the complex can be dissociated to its intact or its modified form. However the serpins uses the nonstandard mechanism for their action against proteases, they form a covalent, irreversible interaction to trap and inhibit the proteases (Farady *et al.* 2010, Whisstock *et al.* 2010, Fluhr *et al.* 2012). This unique mechanism of action enable serpins to form a dead-end complex that induces the cleavage of an internal peptide bond to trigger a conformational change and the inhibitor never recover its initial structure and leads to trap the proteases. Hence the inhibitory serpins act as a suicide protease inhibitors.The most common reactive center found in almost all the plant serpins at P2-P1-P1' is Leu-Arg-Xaa, where Xaa is a (Ala, Cys, Gly, and Ser) and the serpins containing this reactive center are known as the 'LR serpins' (Hejgaard and Roberts 2007, 2008).They are also present in *Arabidopsis* serpin, AtSerpin1 (Lampl *et al.* 2010) and barley serpin, BSZx (Dahl *et al.* 1996). Most of the serpins have role as a serine protease inhibitor, however some of them play role as cross-class inhibitor as they inhibit the cysteine proteases such as caspases and

[1]Pawanmainkar5@gmail.com, [2]manoj7manikandan@gmail.com, [3]jayaswaldeepanshu@gmail.com, [4]agarwalyamini53@gmail.com, [5]rkp123000@gmail.com, [6]rekhakansal@hotmail.com.

cathepsin L, S, K (Komiyama *et al.* 1994, Takeda *et al.* 1995, Schick C *et al.*1998). *In vitro* and *in vivo* studies showed that the cystein proteases are the target for the serpins, as *Arabidopsis* serpin, AtSerpin1 inhibits Metacaspase 9, an endogenous cystein proteinaseand papain-like cysteine proteinase RESPONSIVE TO DESICCATION-21 (RD21) (Vercammen *et al.* 2006, Lampl *et al.* 2010).

The insect orders such as Lepidoptera and Diptera, uses chymotrypsin serine proteases as digestive enzymes (Jongsma and Bolter 1997) and are likely targets for plant serpins. The pumpkin serpin CmPS-1 negatively correlated the survival of *Myzus persicae,* green peach aphid (Yoo *et al.* 2000).The growth of *Spodoptera littoralis* larvae was reduced by 65% and high mortality levels in nymphs of *Acyrthosiphon pisum* were observed when both the insects were exposed to artificial diets containing AtSerpin1 protein at LC50 = 637 ug/ml (Alvarez-Alfageme *et al.* 2011). The tomato serpin inhibited serine proteases secreted by *Fusarium oxysporum* in response to chitinase (Jashni *et al.* 2015). *Medicago truncatula* serpin6 expressed in nodule mediated drought-induced senescence by regulating cysteine protease activity (Ramadoss *et al.* 2018).

Several members of serpins have been identified and characterized in a variety of organisms such as, animals, viruses, fungus, archaea, bacteria and plants (Silverman *et al.* 2001, Gettins 2002, Irving *et al.* 2002, Roberts and Hejgaard 2008). The eight and 14 genes of serpins family protein have been identified and characterized in the genome sequence of *Arabidopsis* and rice, respectively. In the wheat grains six serpins were identified and showed their inhibitory activity towards chymotrypsin and cathepsinproteases (Ostergaard *et al.* 2000).In response to fungal diseases 37 serpin genes were identified in the *Triticum aestivum* (Harriet *et al.* 2019). A single gene was reported in the unicellular green alga, *Chlamydomonas reinhardtii*. Since no such data is available for legumes therefore using computational analysis of the refseq genome of pigeonpea, soybean, mungbean, commanbean, cowpea and adzukibean, 7 serpin genes have been identified and characterized in the reported present study.

MATERIALS AND METHODS

To find all the serpin gene sequences in six legume crops, BlastP (protein blast) search with an evalue cutoff 1e-5 (1×10^{-5}, i.e 0.00001) was performed against the pigeonpea (*Cajanus cajan*), soybean (*Glycine max*), cowpea (*Vigna unguiculata*), mungbean (*Vigna radiata*), commanbean (*Phaseolus vulgaris*) and adzukibean (*Vigna angiculata*) refseq genome database using *Arabidopsis thaliana*, AtSerpin1 (AT1G47710) conserved domain (cd02043) as a query. All the identified sequences were confirmed for the presence of conserved domain (cd02043) using Batch web cd search tool (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi).Consensus motifs sequences were predicted in all the serpins using MEME online tool (http://meme-suite.org/).The target peptide sequence was predicted using Predisi (http://www.predisi.de/) and Subcellular localization was determined using CELLO v.2.5 (http://cello.life.nctu.edu.tw/).

The predicted sequences were aligned with the well characterized Atserpin1 protein sequence using MUSCLE programme (https://www.megasoftware.net/home) implemented in the MEGA 6.0 tool.The sequence aligned file were subjected for the reconstruction of phylogenetic tree using neighbor joining method, with 10000 bootstrap trails, poisons correction, and pairwise deletion. Separately, RCL loop of 7 legume serpins were aligned and used to generate the 'Weblogo' (http://weblogo.berkeley.edu/). The 'Weblogo' represents the height of the stack of amino acids at each site on Y axis and X-axis represents RCL residue numbers P17 to P35.

Presence and absence of introns in the gene structure of serpins were predicted using GSDS, gene structure display server tool (http://gsds.cbi.pku.edu.cn/).

Three dimensional structures of all the putative serpins were predicted using Atserpin1 (3LE2), X-ray crystal structure as template in homology modeling using Swiss model server (https://swissmodel.expasy.org) and models were validated using Ramchandran plot (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php).

RESULTS AND DISCUSSION

Identification of Serpin members

Analysis of available refseq genome data revealed one serpin gene in pigeonpea, commanbean, mungbean, cowpea, adzukibean and two serpin genes in soybean (Table 1). A single serpin gene was also reported in the *Chlamydomonas reinhardtii* (Location: Chlre3/scaffold_28:1327628–1332277). However, 3 serpin genes were identified in chickpea genome database (Sharma *et al.* 2015), whereas in cereals higher numbers of serpins were identified.In rice 14 serpin genes were identified, three of which encoded non inhibitory serpins but in *Arabidopsis* genome 7 serpin genes of inhibitory type were present (Francis *et al.* 2012). A total of 189 serpin genes were reported in wheat genome database (Appels *et al.* 2018, Ramirez Gonzalez *et al.* 2018, Borrill *et al.* 2016). A large number of serpin genes reported in wheat is because of the hexaploid nature (Harriet *et al.* 2019). Afew numbers of serpin genes (1 or 2 serpins) in the studied genome of legumes may be because of their multifunctional and cross class inhibitory role due to the conformational change in the reactive center loop. The nucleotide and protein sequences ranged in size from 1170 to 1278nt and 389 to 425aa in the legumes included in the present study. The isoelectric point of serpins ranged from 5.1 to 8.7 and the instability index was below 40 and ranged from 26.06 (VanSerpin) to 29.69 (GmaSerpin2) predicting them as stable proteins (Table 2).

Domain, Motif, Signal peptide and Subcellular localization

The conserved plant_serpin domain (cd02043) with the size of ~380aa was present in all the seven serpins and its position was between 8-386aa except in mungbean,

Table 2 Primary sequence analysis of serpin genes

Gene name	NCBI Reference Sequence Accession ID	CDS size (bp)	Size (aa)	Mw (kDa)	pI	Instability index
CcaSerpin	NC_033810.1	1170	389	42.8	5.2	28.67
GmaSerpin1	NC_016091.3	1170	389	42.7	6.1	28.20
GmaSerpin2	NC_038242.1	1170	389	42.8	5.16	29.69
VanSerpin	NC_028360.1	1170	389	42.9	5.41	26.06
VraSerpin	NC_023751.1	1278	425	46.7	8.79	27.69
VunSerpin	NC_040287.1	1275	424	46.8	7.11	26.61
PvuSerpin	NC_030645.1	1275	424	46.8	7.76	28.59

Table 3 Location and size of plant_serpin domain (cd02043) present in the 7 serpins of six legumes

Gene name	Domain position	Domain length (aa)
CcaSerpin	8-386	378
GmaSerpin1	8-386	378
GmaSerpin2	8-386	378
VanSerpin	8-386	378
VraSerpin	42-422	380
VunSerpin	43-421	378
PvuSerpin	43-421	378

Table 4 Location and reliability score of legumes Serpins

Gene ID	Most likely location	Reliability score
CcaSerpin	Cytoplasmic	1.39
GmaSerpin1	Extracellular	1.40
GmaSerpin2	Cytoplasmic	1.22
VanSerpin	Cytoplasmic	1.75
VraSerpin	Mitochondrial	1.39
VunSerpin	Mitochondrial	1.30
PvuSerpin	Mitochondrial	1.51

commanbean and cowpea serpin protein which was between 43-421aa (Table 3).

The signature motifs present in the 7 serpin proteins were analysed for 10 conserved motifs and found that four serpins (VanSerpin, GmaSerpin1, GmaSerpin2 and CcaSerpin) contained 9 out of 10 motifs (motif #1- 9) and all are conserved.The motif#10 was present only in 3 serpins (VraSerpin, VunSerpin, and PvuSerpin). Phylogeneticaly, VanSerpin is a clade 1 member devoid of motif#10 (VFKVLKYPKIGILKLASLRASVHRRVEIT) however the other members of clade 1 have an extra motif#10 and is localised upstream outside of plant_serpin domain (Fig 1).

In silico prediction of all the serpins sub-cellular localization based on protein sequence was performed using the program CELLO. The VraSerpin, VunSerpin and PvuSerpin were localized to mitochondria, whereas CcaSerpin, GmaSerpin2 and VanSerpin were localized to cytoplasm and GmaSerpin1 was predicted to be extracellular (Table 4). The plants serpins are distinct from animal serpins, as plant serpins lack N-terminal target peptide sequence but animal's serpins contained signal peptides for their export to cell surface and secretion (Vercammen *et al.* 2006). In *Arabidopsis* GFP tagged serpins, AtSRP2 and AtSRP3 were localized to the nucleus and cytosol respectively (Ahn *et al.* 2009).

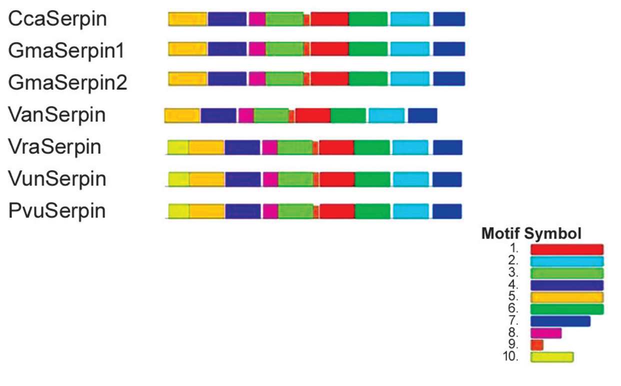


Fig 1 Motif analysis of Serpin genes. The parameters used for motif discovery were minimum width 6, maximum width 50 and maximum number of motifs to find is 10 and 0 order model of sequence.

Sequence alignment and phylogenetic analysis

A phylogenetic analysis and sequence alignment of all the studied legume serpins revealed that the identified legumes serpins are closely related at nucleotide as well as amino acid level.

They contain few critical differencesin reactive site in RCL loop that is likely to be involved in targeting different proteases. The phylogenetic analysis results provided that 7 serpin genes formed two main clades in the neighbor joining tree. CladeI contained members of VanSerpin, VraSerpin, VunSerpin and PvuSerpin and the CladeII consisted of GmaSerpin1, GmaSerpin2 and CcaSerpin (Fig 2).

Alignment and conservation of amino acid residues in the RCL loop

The RCL loop of the 7 legume serpins were aligned and showed that the putative residues P17E (Glu), P14T (Thr) and P8T (Thr), P11A (Ala), P10A (Ala) at the N-terminal of reactive site were conserved in all the seven serpins. Five serpins (CcaSerpin, GmaSerpin1, PvuSerpin,VunSerpin and VanSerpin) were found to be conserved with P2-P1

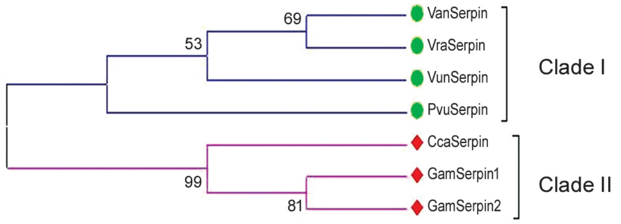
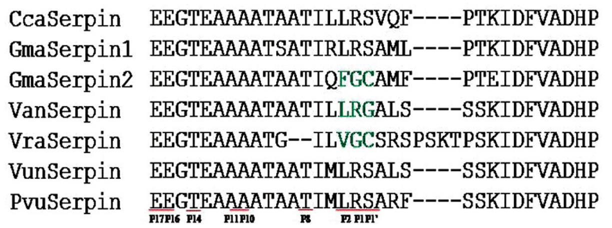


Fig 2 Phylogenetic analysis of serpins genes identified in the genome sequence of pigeonpea, soybean, mungbean, cowpea, commanbean and adzuki bean.

reactive center sequence, 'LR' while two of serpins have the unique reactive center such as 'FG' (GmaSerpin2) and 'VG' (VraSerpin) (Fig 3). At C-terminal of reactive center, P1' Ser (S) was present in 4 serpins (CcaSerpin, GmaSerpin1, PvuSerpin, VunSerpin) while Cys (C) was present in 2 serpins (GmaSerpin2 and VraSerpin) and Gly (G) was present in VanSerpin. The residues (IDFVADHP) are highly conserved among all the serpins at P12'- P19'positions. Based on the conserved residues pattern of all the serpins they were predicted to be inhibitory in functions and diversity at P1' residue indicated its inhibitory specificities and distinct functions.

The frequency of amino acids residues at each position in the mobile reactive center of legume serpins was defined in RCL alignment 'Weblogo' (Fig 4). The plant serpins with conserved reactive center (P2-P1) were called as LR serpins (Hejgaard *et al.* 2007, Roberts *et al.* 2008). These were also found in the model plant *Arabidopsis*, Atserpin1 (Lampl *et al.* 2010) and Barley serpin, BSZx (Dahl *et al.* 1996).



```
CcaSerpin    EEGTEAAAATAATILLRSVQF----PTKIDFVADHP
GmaSerpin1   EEGTEAAAATSATIRLRSAML----PTKIDFVADHP
GmaSerpin2   EEGTEAAAATAATIQFGCAMF----PTEIDFVADHP
VanSerpin    EEGTEAAAATAATILLRGALS----SSKIDFVADHP
VraSerpin    EEGTEAAAATG--ILVGCSRSPSKTPSKIDFVADHP
VunSerpin    EEGTEAAAATAATIMLRSALS----SSKIDFVADHP
PvuSerpin    EEGTEAAAATAATIMLRSARF----SSKIDFVADHP
             P17P16 P14 P11P10  P8  P2 P1P1'
```

Fig 3 Alignment of the amino acid residues in the RCL of pigeonpea with other legume serpins. RCL sequence numbering is that of Schechter *et al.* (1967) like at N-terminal of reactive site, P4……. P1 and at C-terminal of reactive site P1'……P4'.

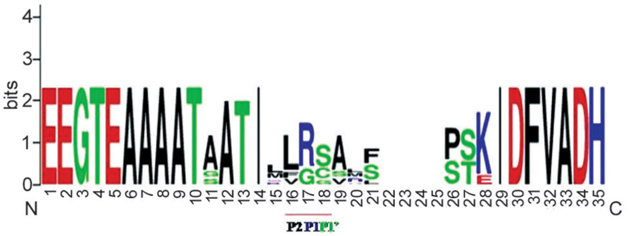


Fig 4 Weblogo represents the conservation of amino acid residues in the reactive center loop of 7 legume serpins.

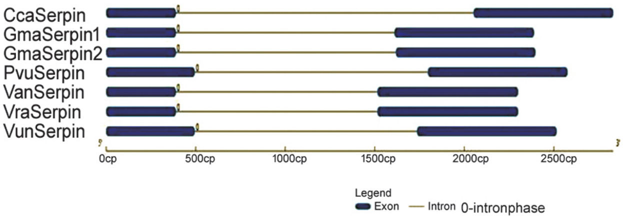


Fig 5 Exon-Intron organization of serpin genes from pigeonpea, soybean, commanbean, adzukibean, mungbean, cowpea.

Exon-intron organization

The detailed analysis of gene structure of legume serpins found that all the serpins contained the single intron. The existing introns of phase 0 type falling between two succeeding codons are present in all the 7 legume serpins (Fig 5). The exon-intron boundaries were conserved in all the serpins present in legumes as well as cereals like rice (Roberts *et al.* 2008) and barley (Brandt *et al.* 1990). The obtained result suggested the standard gene structure for legume serpins.

Molecular modeling of Serpins

To build the three dimensional structure of legume serpins blastP search was performed against the PDB database. The legumes serpins sequences possessed maximum identity with well characterized homologous protein from *Arabidopsis* Atserpin1 (PDB ID 3LE2 and resolution: 2.2Å).It was found that the identity with respect to Atserpin1 was like 61.18%, 60.15%, 59.64%, 60.41%, 58.72%, 61.41, 61.67% for CcaSerpin, GmaSerpin1, GmaSerpin2, VanSerpin, VraSerpin, VunSerpin and PvuSerpin, respectively.The maximum identity(>30%) with template indicated the legume serpins were suitable for homology modeling and all the proteins models were built using Atserpin1 (3LE2) as template.The reactive center loop (RCL) and reactive sites P2-P1 are represented as Leu-Arg in the structure of a, b, d, e, f, Phe-Gly in c and Val-Gly in g (Fig 6a-g).

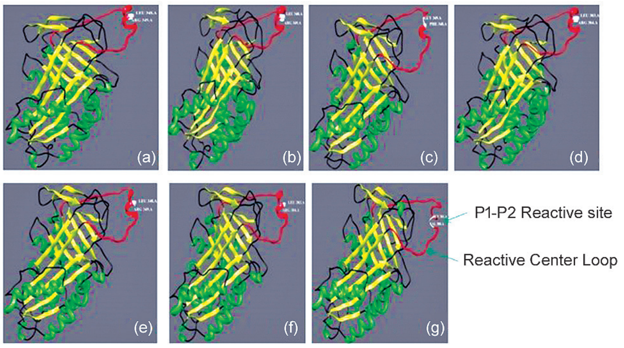


Fig 6 a-g Homology modeling modelled three dimensional structures of legume serpins. (Legend: a-CcaSerpin, b-GmaSerpin1, c-GmaSerpin2, d-PvuSerpin, e-VanSerpin, f-VunSerpin, g-VraSerpin. Red color –Reactive centre loop (RCL) and white color: reactive site (P2-P1, Leu-Arg_a, b,d, e, f, Phe-Gly_c and Val-Gly_g).

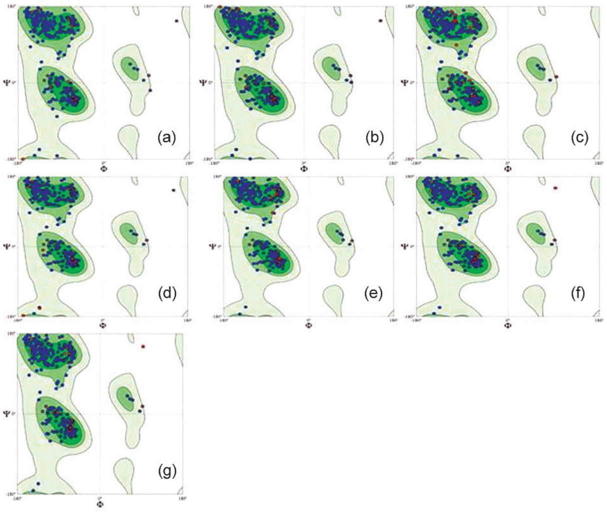


Fig 7 a-g Ramachandran Plot of homology built models of legume serpins. (a-CcaSerpin, b-GmaSerpin1, c-GmaSerpin2, d-PvuSerpin, e-VanSerpin,

The generated 3D models were validated using Ramachandran plot to determine the structure stability based on the position of amino acid in the structure.The built models of legume serpins revealed that more than 97% of the residues were occupied in the allowed region (Fig 7a-g). The PDB database shows that if more than 90% of the residues fall in the allowed region, it indicated the good stereochemical quality of structure and the proteins can be used for the *in-silico* studies.

f-VunSerpin, g-VraSerpin

The above computational studies of the identified seven Serpins have classified them into two major clades with inhibitory activities. Like other plant Serpins, these Serpins also have the conserved reactive Center loop and hence share a common function. Since they are present in the storage tissue, they are also thought to be involved in the control of insect pests and pathogens of the crop plants.

ACKNOWLEDGEMENTS

The work was supported by grants from DBT (authors 1, 2 and 5), CSIR (author 3) and ICAR-NPTC (authors 4 and 6).

REFERENCES

Ahn J W, Atwell B J and Roberts T H.2009. Serpin genes AtSRP2 and AtSRP3 are required for normal growth sensitivity to a DNA alkylating agent in Arabidopsis. *BMC Plant Biology* **9**(1):52.

Alvarez-Alfageme F, Maharramov J, Carrillo L, Vandenabeele S, Vercammen D, Van Breusegem F and Smagghe G.2011. Potential use of a serpin from Arabidopsis for pest control. *PLOS ONE* **6** (5): e20278.

Appels R, Eversole K, Feuillet C, Keller B and Rogers J.2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**: 661.

Borrill P, Ramirez-Gonzalez R and Uauy C.2016.ExpVIP: a customizable RNA-seqdata analysis and visualization platform. *Plant Physiology* **170**: 2172–2186.

Brandt A, Svendsen I B and Hejgaard J.1990. A plant serpin gene: structure, organization and expression of the gene encoding barley protein Z4. *European Journal of Biochemistry* **194** (2):499-505.

Carrell R and Travis J.1985. α1-Antitrypsin and the serpins: Variation and counter variation. *Trends in Biochemical Sciences* **10**:20-24.

Dahl S W, Rasmussen S K, Petersen L C and Hejgaard J.1996. Inhibition of coagulation factors by recombinant barley serpin BSZx. *FEBS Letters* **394**(2):165-168.

Dementiev A, Dobo J and Gettins P G. 2006. Active site distortion is sufficient for proteinase inhibition by serpins structure of the covalent complex of α1-proteinase inhibitor with porcine pancreatic elastase. *Journal of Biological Chemistry* **281** (6):3452-3457.

Francis S E, Ersoy R A, Ahn J W, Atwell B J and Roberts T H.2012. Serpins in rice: protein sequence analysis, phylogeny and gene expression during development. *BMC Genomics* **13** (1): 449.

Gettins P G.2002. Serpin structure, mechanism, and function. *Chemical Reviews* **102** (12): 4751-4804.

Gooptu B and Lomas D A.2009. Conformational pathology of the serpins: themes, variations, and therapeutic strategies. *Annual Review of Biochemistry* 78:147-176.

Harriet R B, Lars S J and Fiona M D.2019.Serpins: genome-wide characterization and expression analysis of the serine protease inhibitor family in *Triticum aestivum*. *G3:Genes|Genomes|Genetics* **9** (8): 2709-2722.

Hejgaard J and Roberts T H.2007.Plant Serpins.*Molecular and Cellular Aspects of the Serpinopathies and Disorders in Serpin Activity*, pp 279-300. Silverman G A and Lomas D A (Eds). New Jersey: World Scientific.

Huntington J A, Read R J and Carrell R W.2000. Structure of a serpin–protease complex shows inhibition by deformation. *Nature* **407**(6806): p 923.

Irving J A, Steenbakkers P J, Lesk A M, Opden Camp H J, Pike R N and Whisstock J C.2002. Serpins in prokaryotes. *Molecular Biology and Evolution* **19**:1881-90.

Jashni M K, Mehrabi R, Collemare J, Mesarich C H and De Wit P J.2015. The battle in the apoplast: further insights into the roles of proteases and their inhibitors in plant–pathogen interactions. *Frontiers in Plant Science* 6: 584.

Jongsma M A and Bolter C.1997. The adaptation of insects to plant protease inhibitors. *Journal of Insect Physiology* **43**(10):885-95.

Komiyama T, Ray C A, Pickup D J, Howard A D, Thornberry N A, Peterson E P and Salvesen G J.1994. Inhibition of interleukin-1 beta converting enzyme by the cowpox virus serpinCrmA.An example of cross-class inhibition. *Biological Chemistry* **269**:19331-7.

Krowarsch D, Cierpicki T, Jelen F and Otlewski J.2003.Canonical protein inhibitors of serine proteases. *Cellular and Molecular Life Sciences* ***60*** (11):2427-2444.

Lampl N, Budai H O, Davydov O, Joss T V, Harrop S J, Curmi P M, Roberts T H and Fluhr R.2010. Arabidopsis AtSerpin1: crystal structure and *in vivo* interaction with its target protease responsive to desiccation-21 (RD21). *Journal of Biological Chemistry* **285**: 13550-13560.

Ostergaard H, Rasmussen S K, Roberts T H and Hejgaard J.2000. Inhibitory serpins from wheat grain with reactive centers resembling glutamine-rich repeats of prolamin storage proteins - Cloning and characterization of five major molecular forms. *Journal of Biological Chemistry* **275**: 33272-33279.

Ramadoss D, Cory M, Michael T M and Paul P D. 2018. Drought-induced senescence of *Medicago truncatula* nodules involves serpin and ferritin to control proteolytic activity and iron levels. *New Phytologist* **200** (1): 196-208.

Ramirez Gonzalez R H, Borrill P, Lang D, Harrington S A, Brinton J *et al.*2018. The transcriptional landscape of polyploid wheat. *Science* **361**: 6403.

Rawlings N D, Barrett A J, Thomas P D, Huang X, Bateman A and Finn R D.2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research***46** (D1): D624-32.

Roberts T H and Hejgaard J.2008.Serpins in plants and green algae. *Functional & Integrative Genomics* 8:1–27.

Roberts T H, Hejgaard J, Saunders N F, Cavicchioli R and Curmi P M.2004.Serpins in unicellular Eukarya, Archaea, and Bacteria: sequence analysis and evolution. *Journal of Molecular Evolution* **59:**437-47.

Schechter I and Berger.1967.On the size of the active site in proteases. I. Papain. *A Biochemical and Biophysical Research Communications* **27**(2):157-62.

Schick C, Pemberton P A, Shi G P, Kamachi Y, Cataltepe S, Bartuski A J, Gornstein E R, Bromme D, Chapman H A and Silverman G A.1998. *Biochemistry* **37**:5258–5266.

Sharma R and Suresh C G.2015. Genome-wide identification and structure-function studies of proteases and protease inhibitors in *Cicer arietinum* (chickpea). *Computers in Biology and Medicine* **56**: 67–81.

Silverman G A, Bird P I, Carrell R W, Church F C, Coughlin P B, Gettins PGW, Irving J A, Lomas D A, Luke C J and Moyer R W.2001. The serpins are an expanding superfamily of structurally similar but functionally diverse proteins - Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *Journal of Biological Chemistry* **276**: 33293-33296.

Takeda A, Yamamoto T, Nakamura Y, Takahashi T and Hibino T.1995. Squamous cell carcinoma antigen is a potent inhibitor of cysteine proteinase cathepsin L. *FEBS Lett* **359:78**-80.

Vercammen D, Belenghi B, van de Cotte B, Beunens T, Gavigan J A, De Rycke R, Brackenier A, Inze D, Harris J L and Van Breusegem F.2006. Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for Metacaspase 9. *Journal of Molecular Biology* **364**: 625-636.

Whisstock J C and Bottomley S P. 2006. Molecular gymnastics: serpin structure, folding and misfolding. *Current Opinion in Structural Biology* **16**(6):761-768.

Whisstock J C, Silverman G A, Bird P I, Bottomley S P, Kaiserman D, Luke C J, Pak S C, Reichhart J M and Huntington J A.2010. Serpins flex their muscle: II. Structural insights into target peptidase recognition, polymerization, and transport functions. *Journal of Biological Chemistry* **285**: 24307-24312.

Yoo B C, Aoki K, Xiang Y, Campbell L R, Hull R J, Xoconostle-Cazares B, Monzer J, Lee J Y, Ullman D E and Lucas W J.2000.Characterization of *Cucurbita maxima* phloem serpin-1 (CmPS-1) – a developmentally regulated elastase inhibitor. *Journal of Biological Chemistry* **275**: 35122–35128.