International Journal of Genetic Engineering and Biotechnology.
ISSN 0974 3073 Volume 5, Number 1 (2014), pp. 63-70

http://www.irphouse.com

In Silico Analysis of Cysteine Protease Sequences Imparting Senescence

Anukriti Verma[1] and Smriti Gaur*[1]

[1]Department of Biotechnology,
Jaypee Institute of Information Technology,
(Deemed University), A-10, Sector-62, Noida, U.P., India.

Abstract

Senescence is the prepared depreciation that can cause death in plants. Senescence occurrence depends upon certain hormones and genes including cysteine protease genes that increase the levels of senescence and certain proteins that are involved in macromolecular degradation, detoxification of oxidative metabolites, induction of defence mechanisms and signalling and regulatory events. Leaf senescence can hinder the crop yields and biomass production. Senescence associated proteases are involved in nutrient recycling and nitrogen remobilization.

In this study, the comparative characterization using various bio-computational tools to show the physiochemical and structural properties has been performed involving the species containing cysteine protease and showing senescence in their leaves. Result showed that all proteases are acidic except Q40143. Q8S929, Q9M1Y0 and Q9ZSC2 are found to be unstable whereas all other sequences are stable. All proteases are thermostable. Maximum number of disulfide bonds and motifs that provide stability and possess large amount of functional information are present only in B9A123. So B9A123 of *Spinacia oleracea* can be used as a suitable candidate for further studies related to senescence. These findings will provide more insight to the function of genes associated with senescence as well as will be helpful to further studies on senescence at molecular or structural levels.

Keywords- senescence; cysteine protease; in silico analysis.

1. Introduction

Senescence is planned deterioration within the plant species leading to death (Nooden et al, 1997). It occurs in both annual as well as perennial plants. Arrival of senescence is time and again triggered by hormones including ethylene, cytokinins, abscisic acid, jasmonic acid, salicylic acid and auxins. Cytokinins inhibit leaf senescence and cytokinin level drops with the advancement of leaf senescence. Few genes in arabidopsis that are shown to be affected by cytokinins are the SAG12, AHK3 and IPT (Zwack and Rashotte, 2013). It has also been reported that amount of ethylene increase during senescence. One of the genes that control ethylene induced senescence in arabidopsis leaves is EDR1 (Tang et al, 2005). Abscisic acid genes are upregulated during occurrence of senescence that include NECD and aldehyde oxidase genes. Other genes concerned in the regulation of leaf senescence by auxin in arabidopsis include ARF2. Genes such as SAG12 and PR1a are found when salicyclic acid is present in leaves.

High carbon to nitrogen ratio is observed during senescence period (Parrott et al, 2010). Nutrient relocation from leaves to reproducing seeds is also achieved during senescence. Leaf senescence involves a synchronized and regulated action at the cellular, tissue, organ, and organism levels. Proteins, antioxidants and other nutritional compounds are degraded during senescence. Leaf senescence is disadvantageous to crop yield and biomass production and postharvest loss in vegetables. Senescence is a fatal stage of plant development that involves processes leading to degradation of macromolecules and mobilization of components such as nitrogen, carbon and minerals from mature leaf to other parts of plant. During senescence, intense changes occur in leaf cells in the metabolism and the degeneration of cellular structures. The most major change in cell structure is the breakdown of the chloroplast that lowers the levels of proteins such as Rubisco and chlorophyll a/b binding (CAB). Lowering of polysomes and ribosomes is also observed. Biochemical changes help in the removal of nutrients from the cell. A fast decline in photosynthesis is also observed (Watling et al, 2000).

Some of the environmental factors are also involved including abiotic factors such as drought, temperature, light, ozone and availability of nutrients. Biotic factors that lead to senescence include shade and exposure of a pathogen. Leaf senescence is escorted by changes in gene expression. Around 100 genes whose activity is increased during leaf senescence have been isolated from a variety of plant species such as arabidopsis, rapeseed, tomato, maize, rice, tobacco, potato, and bean. There are 6 classes of genes involved including housekeeping genes, genes expressed in green leaves, regulatory genes and genes involved in mobilization of compounds. Senescence enhanced genes including protease and nuclease play a key role in senescence (Wollaston, 1997). Several genes show increase in level of transcription like cysteine protease genes (SAG2,SAG12,See1,See2,LSC7 and LSC790) or glutamine synthetase (Atgsr2) present in plant species like radish, maize etc. (Wollaston, 1997). Senescence associated genes contain proteins that participate in macromolecular degradation, detoxification of oxidative metabolites, induction of defence mechanisms and signalling and regulatory events (Gepstein et al, 2003). Senescence associated proteases are involved in nutrient recycling and nitrogen

remobilization (Roberts et al, 2012).

In this paper the comparative characterization using various bio-computational tools has been performed of species containing cysteine protease and showing senescence in their leaves so that most appropriate plant species in terms of stability or species that possess most biological information for further studies related to senescence can be obtained. These findings will provide more insight to the function of genes associated with senescence as well as will be helpful to further studies on senescence at molecular or structural levels.

2. Methodology

2.1 Sequence Retrieval

Expasy (uniprot KB) that provides protein sequences and annotation data (Jain et al, 2009) was used to retrieve the sequences of species showing senescence in their leaves and containing cysteine protease that is involved in the process of senescence. Out of these sequences well annotated and curated sequences were obtained by removing the partial, precursor, putative and fragment sequences. These were downloaded in FASTA format to be used for further analysis (http://www.uniprot.org).

2.2 Physiochemical Characterization

Physiochemical parameters including Theoretical isoelectric point (pI), Instability index and Aliphatic index were computed with the help of Expasy Protparam bioinformatic tool (Gasteiger et al, 2005) (http://www.expasy.org/tools/protparam .html) that are deduced from protein sequences.

2.3 Secondary Structure Prediction

SOPMA tool (Self-Optimized Prediction Method with Alignment) (Geourjon and Deleage, 1995) was applied to extract the information regarding the secondary structures that consist of Alpha helix, 310 helix, Pi helix, Beta bridge, Extended strand, Beta turn, Bend region, Random coil, Ambiguous states and Other states.

2.4 Functional Analysis

CYS_REC tool was able to predict number of cysteines, disulfide bonds at a particular cysteine position and the probable pattern (Sivakumar and Balaji, 2007). Motif search (www.genome.jp/tools/motif) was used to find the number of motifs, motif ID, description and position of the motif found. The results were obtained in the PROSITE pattern which is a manually derived motif library (Ogiwara et al,1996). SOSUI sever discriminates membrane proteins from amino acid sequences (Shigeki, 2002) that was applied to find transmembrane region, type and length of transmembrane protein.

3. Result and Discussion

3.1 Sequence Retrieval

It is seen that the stimulation of cysteine proteases is an excellent marker for senescence (Kardailsky and Brewin, 1996) so 19 species containing cysteine protease involved in imparting senescence in their leaves were retrieved and 13 annotated

protein sequences were obtained that contained 4 sequences of *arabidopsis thaliana* (Mouse-ear cress), 2 sequences each of *ipomoea batatas* (Sweet potato), *spinacia oleracea* (Spinach) and *nicotiana tabacum* (Common tobacco) and 1 sequence each of *solanum lycopersicum* (Tomato), *zea mays* (Maize) and *pisum sativum* (Garden pea). Their further physiochemical characterization and structure prediction was done to see the effect of various parameters on senescence (table 1).

Table 1. Sequences of cysteine proteases of various plant species showing senescence

	Gene names	**Organism**
Q8H166	ALEU AALP SAG2 At5g60360 MUF9.1	*Arabidopsis thaliana* (Mouse-ear cress)
Q8S929	ATG4A APG4A At2g44140 F6E13.27	*Arabidopsis thaliana* (Mouse-ear cress)
Q9M1Y0	ATG4B APG4B At3g59950 F24G16.220	*Arabidopsis thaliana* (Mouse-ear cress)
Q40143	CYP-3	*Solanum lycopersicum* (Tomato) (Lycopersicon esculentum)
Q9AUC7	SPG31	*Ipomoea batatas* (Sweet potato) (Convolvulus batatas)
Q38886	SAG12	*Arabidopsis thaliana* (Mouse-ear cress)
Q949A2	Elsa	*Pisum sativum* (Garden pea)
Q9ZSC2		*Ipomoea batatas* (Sweet potato) (Convolvulus batatas)
Q43705	see1	*Zea mays* (Maize)
B9A123	SoCP3-41	*Spinacia oleracea* (Spinach)
C0STW6	CPI	*Spinacia oleracea* (Spinach)
E7D4U5	SAG12	*Nicotiana tabacum* (Common tobacco)
Q9LRI2	NTCP-23	*Nicotiana tabacum* (Common tobacco)

3.2 Physiochemical Characterization

Table 2 shows various physiochemical parameters. It is seen that photosynthetic activity decreases during leaf senescence (Arora et al, 2009). Rubisco level increases and senescence decreases in acidic pH. All proteases are acidic except Q40143. If instability index is above 40 it indicates the instability. It is seen that Q8S929, Q9M1Y0 and Q9ZSC2 are unstable having values between 45-52 whereas all other sequences are stable having instability index below 40 in the range of 16-36. Aliphatic index shows relative volume occupied by aliphatic side chains.The aliphatic index is of high range indicating that proteases are thermostable. These properties could be useful to understand the mechanism related to senescence.

Table 2. Physiochemical Properties

	Theoretical isoelectric point (pI)	**Instability index**	**Aliphatic index**
Q8H166	6.26	26.62	78.44
Q8S929	5.13	46.46	74.11
Q9M1Y0	4.94	52.55	78.09
Q40143	8.59	28.64	81.38
Q9AUC7	5.12	27.03	68.39
Q38886	7.02	31.66	68.5
Q949A2	6.21	34.9	74.14
Q9ZSC2	5.76	46.8	84.45
Q43705	7.55	23.57	77.56
B9A123	5.13	36.06	65.79
C0STW6	5.03	16.27	82.6
E7D4U5	7.97	27.46	69.08
Q9LRI2	6.94	25.3	79.64

3.3 Secondary Structure Prediction

Alpha helix, beta strands or random coils (table 3) are predicted using the secondary structure prediction approach. Random coil dominates alpha helix followed by beta turn in all the sequences except Q9ZSC2 where alpha helix value of 44.49% dominates random coil value of 37% followed by beta turn value of 4.33%.

Table 3. Secondary Structures

	Alpha helix	Extended strand	Beta turn	Random coil
Q8H166	33%	20.11%	5%	42%
Q8S929	27%	16%	3%	53%
Q9M1Y0	26%	18%	3%	53%
Q40143	34%	17%	6%	42%
Q9AUC7	35.19%	16.72%	6.45%	42%
Q38886	35.26%	16.18%	5.78%	43%
Q949A2	38%	17%	5%	40%
Q9ZSC2	44.49%	14.17%	4.33%	37%
Q43705	39%	16%	5%	40%
B9A123	24%	18%	5%	53%
C0STW6	38%	17%	6%	38%
E7D4U5	36%	16%	6%	42%
Q9LRI2	37%	16%	5%	42%

3.4 Functional Analysis

CYS_REC tool predicted Q8H166, Q40143, Q9AUC7, E7D4U5 and Q9LRI2 to contain 9 cysteine residues with 5, 4, 8, 5 and 5 cysteine residues to have disulfide bond respectively. Q8S929 contain 12 cysteine residues with 2 cysteine residues to have disulfide bond. Q9M1Y0 contain 14 cysteine residues with 2 cysteine residues to have disulfide bond. Q38886 and Q949A2 contain 10 cysteine residues with 5 cysteine residues to have disulfide bond each. Q43705 contain 8 cysteine residues with 3 cysteine residues to have disulfide bond. C0STW6 does not contain any cysteine residues nor any disulfide bond. B9A123 contain 21 cysteine residues with 18 cysteine residues to have disulfide bond. Maximum disulphide bonds were present in B9A123 that provide stability to protein (Trivedi et al, 2009). No motifs were found in Q8S929 and Q9M1Y0. Q8H166, Q40143, Q9AUC7, Q38886, Q949A2, Q43705, E7D4U5 and Q9LRI2 have 3 motifs belonging to Eukaryotic thiol (cysteine) proteases cysteine active site, Eukaryotic thiol (cysteine) proteases asparagine active site and Eukaryotic thiol (cysteine) proteases histidine active site families. Q9ZSC2 and C0STW6 contain 1 motif of Cysteine proteases inhibitors signature family. B9A123 include 4 motifs where 3 belong to Eukaryotic thiol (cysteine) proteases family and 1 to Phospholipase A2 histidine active site. B9A123 was the selected protease containing highest number of motifs that are portions of protein that fold independently and contain the functional information. Q8S929, Q9M1Y0, Q43705, C0STW6 and Q9LRI2 were found to be soluble in nature and rest as membrane proteins. B9A123 is also a primary membrane protein having 1 transmembrane helix of length 23 having average hydrophobicity -0.428753 (table 4). This sequence contains gene SoCP3-41 (Spinacia oleracea cysteine protease) that is involved in senescence of spinach leaves (Tajima et al, 2011). The maximum amount of cysteine residues, the stability of these residues, the presence of highest amount of motifs and the membrane protein property of B9A123 makes it a suitable candidate for further studies with respect to senescence.

Table 4. *Cysteine-Cysteine Binding, Motif Prediction and Transmembrane Range Prediction*

	Cys residues	**Status**	**Average hydrophobicity**	**Type of protein**	**Transmembrane region**	**Length**	**Type**
B9A123	21	cys168 is s-s bound	-0.428753	MEMBRANE	ASLSLLLLFSLLAVSSAVDMSII	23	PRIMARY
		cys171 is not s-s bound					
		cys202 is s-s bound					
		cys210 is s-s bound					
		cys243 is probably not s-s bound					
		cys301 is s-s bound					
		cys353 is probably not s-s bound					

		cys389 is probably s-s bound					
		cys395 is s-s bound					
		cys401 is s-s bound					
		cys402 is s-s bound					
		cys403 is s-s bound					
		cys411 is s-s bound					
		cys416 is probably s-s bound					
		cys417 is s-s bound					
		cys424 is s-s bound					
		cys425 is s-s bound					
		cys431 is s-s bound					
		cys432 is s-s bound					
		cys439 is s-s bound					
		cys446 is s-s bound					

4. Conclusion

Senescence can lead to death in plants. So studies to control senescence are required. The *in silico* approach has been used for comparing 13 species showing senescence in their leaves and containing cysteine protease that is a suitable marker for senescence based on physiochemical parameters and structure prediction tools to study the effect of these parameters on senescence. By studying various physiochemical parameters, it is seen that all proteases are acidic except Q40143. Q8S929, Q9M1Y0 and Q9ZSC2 are found to be unstable whereas all other sequences are stable. All proteases are thermostable. Random coil dominates other structures in all sequences excluding Q9ZSC2 where alpha helix value is more. B9A123 contains maximum number of disulfide bonds. B9A123 contains highest number of motifs and is found to be primary transmembrane protein involving SoCP3-41 gene that imparts senescence in spinach leaves and can be used as an appropriate plant model thus B9A123 of *Spinacia oleracea* can be used as a suitable candidate for further studies related to senescence.

These results will provide more insight to the function of genes associated with senescence. These findings will be helpful to further studies on senescence at molecular or structural levels. Further studies may help to understand the mechanism of protein degradation in chloroplast, macromolecules degradation, and the functional implications to identify the mechanism of induction and progress of senescence etc.

5. References

[1] A Ogiwara, I Uchiyama, T Takagi, M Kanehisa (1996). Construction and analysis of a profile library characterizing groups of structurally known proteins. *Protein Sci.,* **5**, pp. 1991-1999.

[2] C Geourjon, G Deleage (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*, **11**, pp. 681-684.

[3] D L Parrott, J M Martin, A M Fischer (2010). Analysis of barley (Hordeum vulgare) leaf senescence and protease gene expression: a family C1A cysteine protease is specifically induced under conditions characterized by high carbohydrate, but low to moderate nitrogen level. *New Phytol*, **187**, pp. 313–331.

[4] D Tang, K M Christiansen, R W Innes (2005). Regulation of Plant Disease Resistance, Stress Responses,Cell Death, and Ethylene Signaling in Arabidopsis by the EDR1 Protein Kinase. *Plant Physiol.*, **138,** pp. 1018–1026.

[5] E Gasteiger, C Hoogland, A Gattiker, S Duvaud, M R Wilkins, R D Appel, A Bairoch (2005). Protein Identification and Analysis Tools on the ExPASy Server, In *The Proteomics Protocols Handbook* (John M. Walker, Eds), Humana Press, pp. 571-607.

[6] E Jain, A Bairoch, S Duvaud, I Phan, N Redaschi, B E Suzek, M J Martin , P McGarvey, E Gasteiger (2009). Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics*, **10**.

[7] I N Roberts, C Caputo, M V Criado, C Funk (2012). Senescence-associated proteases in plants. *Physiol Plantarum*, **145**, pp.130–139.

[8] I V Kardailsky, N J Brewin (1996). Expression of cysteine protease genes in pea nodule development and senescence. *Mol. Plant Microbe Inter.*, **9**, pp. 689–695.

[9] *J R Watling, M C Press, W P Quick (2000). Elevated CO_2 Induces Biochemical and Ultrastructural Changes in Leaves of the C_4 Cereal Sorghum.* Plant Physiol*, **123,** pp. 1143-1152.*

[10] K Sivakumar, S Balaji (2007). Biocomputational analysis and characterization Of some antifreeze proteins. *J Theor Comput Chem*, **6**, pp. 127–140.

[11] L D Nooden, J J Guiaet, I John (1997). Senescence mechanisms. *Physiol Plantarum*, **101**, pp.746–753.

[12] M Shigeki (2002). Computational Analysis of Genome Information Using Membrane Prediction System SOSUI. **42**, pp. 104-109.

[13] M V Trivedi, J S Laurence, T J Siahaan (2009).The Role of Thiols and Disulfides on Protein Stability. *Curr.Protein Pept. Sci.*, **10**, pp. 614-625.

[14] N Arora, A K Banerjee, S Mutyala, U S Murty (2009). Comparative characterization of commercially Important xylanase enzymes. *Bioinformation*, **3**, pp. 446-453.

[15] P J Zwack and A M Rashotte (2013). Cytokinin inhibition of leaf senescence. *Plant Signal Behav.* **8.**

[16] S Gepstein, G Sabehi, M J Carp, T Hajouj, M F Nesher, I Yariv, C Dor, M Bassani (2003). Large-scale identification of leaf senescence- associated genes. *Plant J.*, **36**, pp. 629-642.

[17] T Tajima, AYamaguchi, S Matsushima, M Satoh, S Hayasaka, K Yoshimatsu, Y Shioi (2011). Biochemical and molecular characterization of senescence-related cysteine protease-cystatin complex from spinach leaf. *Physiol Plantarum*, **141,** pp. 97-116.

[18] V B Wollaston (1997). The molecular biology of leaf senescence. *J Exp Bot*, **48**, pp. 181-199.