Original Research Article

In-silico analysis and functional characterization of *Rhizoctonia solani* effector proteins

ABSTRACT

Sheath blight is considered as the second most prevalent disease in rice with no proper resistance genes identified for its defence. Out of the 14 anastomosis groups in *R. solani* (AG1 to AG13 and AGBI), the AG1 group is mostly responsible for the infection in the rice. There are many effector proteins under this group which act as transcription factors. Many bioinformatics tools are available to determine the interaction between the promoter region and effector proteins. But in an attempt to use these effector proteins for molecular docking and simulation studies certain criteria like *In-silico* analysis and validation of the proteins are to be analyzed. The reactivity and stability characteristics of these proteins were assessed physicochemically. In order to characterize the protein's function, bioinformatics tools were utilized to predict secondary structures. Robetta an abinitio approach was used to predict the 3D structure, and the models were validated using the SAVESv6.0 (Procheck) server. The effector protein functional analysis and 3D structure predictions made using empirical data can shed light on the interaction studies for locating effector binding elements in the promoter regions of host genes.

Keywords: *Effector binding element, R. solani, In-silico analysis, Robetta, Procheck.*

1. INTRODUCTION

Rice (*Oryza sativa*) is the most widely grown crop and an important staple food in India, which is grown in large areas in the Southern states of India. In India, the largest area has been occupied by rice which accounts for about 21% of the total cropped area (Khush, 1997) [15]. Rice production is affected by various diseases caused due to fungal, bacterial, and viral pathogens. Rice sheath blight disease is caused by the fungal pathogen, *Rhizoctonia solani Kuhn* (*R. solani*), [Teleomorph stage, *Thanatephorus cucumeris* (Frank) Donk] AG1-IA. It is regarded as the second most important disease after rice blast (Dasgupta, 1992) [6]. Sheath

blight disease causes yield losses ranging from 8% to 50%, according to reports (Savary *et al*., 2000) [20]. Although some rice genotypes confer partial genetic resistance to sheath blight through genes/QTL with a small effect, still no major gene governing resistance identified so far (Dasgupta, 1992) [6]. The conventional breeding method for developing Sheath blight resistance in rice varieties is complicated as the trait shows polygenic inheritance.

More than 100 species of soilborne Basidiomycete fungus *R. solani* are known to infect all varieties of crops, pastures, and horticultural plants. Numerous plant species, including cereals, potatoes, beans, cotton, sugar beet, lettuce, melon, forest trees, and ornamental plants, are susceptible to infection by *R. solani* strains, which often have a broad host range (Anderson 1982; Sneh *et al*., 1991) [1,23]. It is a soil-borne pathogen that survives as sclerotia or mycelia which infects rice plants during irrigation and flooding. The chemical control measures for sheath blight are neither sustainable nor practical as they are not much effective and even the excessive use of these chemical fungicides proves to be deleterious to humans, livestock, and the ecosystem. Through the application of genetic engineering, it is now possible to control fungi caused diseases by modifying crops with fungal resistance genes.

In *R. solani*, there are 14 anastomosis groups (AG1 to AG13 and AGB I). *R. solani* subgroup AG1 IA is mostly responsible for infection in various crop species, including more than 27 families of monocots mainly targeting rice and dicots, and is responsible for diseases like sheath blight, banded leaf, aerial blight, and brown patch1-3. Also, different AG1-IA strains exhibit very diverse clinical characteristics on a given host, including the number of disease lesions, their size, RVSC (relative vertical sheath colonization), disease score, relative lesion length, etc (Taheri *et al*., 2007; Das *et al*., 2013) [24,5].

Generally, sheath blight resistance is majorly assisted by the immune responses in the plant. *OsWRKY4, OsWRKY13, OsWRKY30* and *OsWRKY80* have been reported to positively regulate resistance to sheath blight (Peng *et al*.,2012, 2016; Wang *et al*., 2015; John Lilly and Subramanian 2019) [18,25,16]. On the other extreme, genes such as SWEET11(sugar will eventually be exported by sugar transporter 11) (Gao *et al*., 2018) [9], negatively control the sheath blight resistance in rice. Recent research has demonstrated that SWEETs, which serve as the targets of effector proteins released by pathogens during host-microbe interactions, are responsible for disease susceptibility in many plant species. For instance, the transcription activator-like (TAL) effector, PthXo1 is produced by the *Xanthomonas oryzae pv. oryzae* (Xoo) strain PXO99A and directly binds to the *OsSWEET11* promoter (Chen *et al*., 2010; Yang *et al*., 2006) [4,26]. Delivering TAL effectors to the nucleus causes the expression of certain SWEET genes, ensuring that sucrose is supplied to the colonised cells' apoplasts (Chen *et al.,* 2010) [4].

Similar to this, various *R. solani* effector proteins belonging to the AG1-IA family (AG1IA 09161, AG1IA 05310, and AG1IA 07795) bind to the SWEET gene's promoter region and act as transcription factors, resulting in an excessive amount of sugar being generated. The *R. solani* consumes these generated sugars so order to grow. The term "effector binding element" refers to the area on the promoter where effector proteins bind. To prevent effector proteins from binding to the promoter region, the effector binding element

should be known to knock out. By using genome editing, the effector binding element in the SWEET14 promoter region was altered to confer resistance to bacterial leaf blight (ArulGanesh *et al*.,2021) [2].

The AG1 group effector proteins from *R. solani* are therefore the subject of the current analysis since they confer a broad host range. The majority of protein properties can be explained by a strong structure-function link. Researchers can predict the structure and characteristics of proteins that can be employed for interaction studies using computational approaches. So, the *in silico* identification, and characterization of different effector proteins of *R. solani* have been done in the current study.

2. MATERIAL AND METHODS

2.1 Retrieval of protein sequences

Sequence *R. solani* effector proteins AG1IA-09161-ELU36809.1 (glycosy transferase family 2 protein), AG1IA-05310**-**ELU40661.1(cytochrome oxidase assembly factor), AG1IA-07795**-** ELU38182.1(serine protease) were retrieved in FASTA format from NCBI's protein database (http://www.ncbi.nlm.nih.gov/) [11].

2.2 Physio-chemical characterization

The ProtParam server from Expasy (https://web.expasy.org/protparam) [14] was used to characterize the physio-chemical properties of protein sequences and determine the amino acid composition of the effector proteins. The isoelectric point (pI), the total number of negative (-R) and positive (+R) residues, the extinction coefficients (EC), the instability index (II), the aliphatic index (AI), and the grand average hydropathy were calculated (GRAVY).

2.3 Functional characterization

Functional characterization of effector proteins was done using Expasy's Prosite server (https://prosite.expasy.org) [13]. Protein family, domain, and functional site data were computed using the Prosite server. Pfam analysis was carried out to characterize the chosen effector proteins in relation to the particular protein family (http://pfam.xfam.org/search/sequence) [10].

2.4 Secondary structure prediction

Secondary structural characteristics of effector proteins was calculated using SOPMA (Self Optimized Prediction Method with Alignment) method with their default parameters like similarity threshold 8 and Window width 17. It uses the amino acid sequence to determine the secondary structures such as the beta-strands, alpha-helix, Beta turn, and random coils.
(https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa%20_sopma.html) [12].

2.5 Model building, evaluation

These protein three-dimensional structures were not available in PDB. Therefore, a web server Robetta was used to model the protein's three-dimensional. The SAVES v6.0 (PROCHECK) server was used to validate the model.

3. RESULTS AND DISCUSSION

3.1 Physio-chemical characterization:

The amino acid composition of all effector protein sequences was computed using Expasy's Protparam tool and tabulated in (Table 1). The physio-chemical characterization was calculated for the proteins (Table 2). For the three examined effector proteins, there are an overall range of 350–688 amino acid residues with varying molecular weights. The isoelectric points is important for the estimation of protein solubility, electrophoresis, and electrophoretic separation (Audain *et al*., 2016) [3]. These proteins calculated isoelectric points were greater than 7 which indicates that it has higher protein solubility and electrophoretic separation. The proportion volume filled by aliphatic side chains (alanine, valine, isoleucine, and leucine) in a protein is known as the aliphatic index, and it is a measure of how stable a protein is across a wide temperature range (Sivakumar and Balaji 2007) [22]. ELU40661.1 and ELU38182.1 have lower aliphatic indices. Aliphatic indices of ELU36809.1 is 104.40. The stability of a protein is determined by the Instability Index value, a number below 40 denotes a stable protein (Gamage *et al*., 2019) [8]. Two of the proteins have Instability Index values that fall between 19.34 and 35.52, indicating that they are both stable proteins. With a 45.92 Instability Index, ELU40661.1 is unstable. The effector proteins under study may interact with water molecules more effectively if they have low GRAVY indices, according to this theory**.** Proteins with a GRAVY score above 0 are more likely to be hydrophobic proteins (Magdeldin *et al*., 2012) [17].

Table 1. Amino acid composition of proteins

Amino Acids	ELU36809.1 (glycosyltransferase)	ELU40661.1(cytochrome oxidase assembly factor)	ELU38182.1(serine protease)
Ala (A)	8.3%	8.0%	11.9%
Arg (R)	5.7%	8.1%	6.7%
Asn (N)	4.6%	4.1%	4.2%
Asp (D)	6.0%	3.9%	6.4%
Cys (C)	1.4%	2.8%	0.2%
Gln (Q)	1.4%	3.6%	2.5%
Glu (E)	4.3%	3.6%	2.2%
Gly (G)	5.7%	5.2%	10.9%
His (H)	3.4%	2.3%	2.0%
Ile (I)	4.0%	5.7%	4.9%
Leu (L)	12.6%	7.8%	6.4%
Lys (K)	4.9%	3.5%	4.0%
Met (M)	3.1%	2.2%	1.7%
Phe (F)	3.1%	5.1%	1.7%

Pro (P)	4.9%	7.6%	4.2%
Ser (S)	6.3%	6.5%	7.7%
Thr (T)	6.0%	5.8%	7.7%
Trp (W)	1.7%	2.2%	2.0%
Tyr (Y)	1.7%	3.5%	2.7%
Val(V)	10.9%	8.4%	10.1%
Pyl (O)	0.0%	0.0%	0.0%
Sec (U)	0.0%	0.0%	0.0%

Table 2. Parameters computed using Expasy's ProtParam tool

Protein Name	No. of amino acids	MW	pI	-R	+R	EC	II	AI	GRAVY
ELU36809.1	350	38682.93	7.73	36	37	42190	35.52	104.40	0.089
ELU40661.1	688	77806.04	9.53	52	80	119385	45.92	85.16	-0.107
ELU38182.1	405	42771.27	9.47	35	43	60390	19.34	85.51	-0.140

Where, MW: Molecular weight; pI: Isoelectric Point; -R: Number of negative residues; +R: number of positive residues; EC: Extinction Coefficient at 280 nm; II: Instability Index; AI: Aliphatic Index; GRAVY: Grand Average Hydropathicity

3.2 Functional characterization

Using prediction of pattern, profile, and Pfam analysis, the functional characterization of effector proteins was performed (Table 3). Expasy's Prosite was used to analyse the proteins pattern and profile. Annotations for motif descriptors are stored in the PROSITE database, which is used to identify protein families and domains (Sigrist *et al*., 2002) [21]. Casein kinase II phosphorylation sites, N-myristoylation sites, N-glycosylation sites, and Protein Kinase C phosphorylation sites are among the patterns that were discovered in all three proteins. One protein out of three had a profile. ELU38182.1 contains the serine proteases and subtilase domain profile (serine protease). Three effector proteins were the subject of a Pfam study.

Table 3. Prediction of patterns and profile by using PROSITE and Pfam analysis

Protein ID	Name of protein (as available on NCBI)	Pfam Analysis	Patterns by Prosite	Profile by Prosite
ELU36809.1	glycosyltransferase family 2 protein	Glycosyl transferase family 2(Family)	Casein kinase II phosphorylation site N-myristoylation site N-glycosylation site Protein kinase C phosphorylation site	-

ELU40661.1	cytochrome oxidase assembly factor	Cytochrome c oxidase assembly protein Ct(Family)	N-myristoylation site N-glycosylation site Casein kinase II phosphorylation site Protein kinase C phosphorylation site Cell attachment sequence cAMP- and cGMP-dependent protein kinase phosphorylation site	-
ELU38182.1	serine protease	Subtilase family(Domain), Peptidase inhibitor I9(Domain)	Serine proteases, subtilase family, aspartic acid Serine proteases, subtilase family, histidine active site Serine proteases, subtilase family, serine active site Tyrosine kinase phosphorylation site 2 Protein kinase C phosphorylation site N-myristoylation site cAMP- and cGMP-dependent protein kinase phosphorylation site Casein kinase II phosphorylation site N-glycosylation site Amidation site	Serine proteases, subtilase domain profile

3.3 Secondary structure prediction

Each effector protein's secondary structure was predicted using SOPMA (Table 4). Extended strands and beta turns, followed by random coils and alpha helix, were found to be the most often occurring secondary structural elements, according to SOPMA. The results were shown in a table. When compared to other proteins, ELU36809.1 has a higher percentage of alpha helices, demonstrating the robust nature of proteins (Errington *et al*., 2006) [7].

Table 4. Predicted secondary structures present in proteins

Protein	Amino acids	Alpha helix	Extended sheets	Beta turn	Random coils
ELU36809.1	350	46.00%	15.71%	5.14%	33.14%
ELU40661.1	688	25.87%	23.69%	6.54%	43.9%
ELU38182.1	405	30.37%	23.95%	7.16%	38.52%

3.4 Model building and validation

The three-dimensional structure of the proteins was modelled by Robetta, an abinitio approach (Figure 1). Similarly, the models have been predicted for all the effector proteins of *R. solani*. The residues were categorized in the Ramachandran plot analysis based on their quadrangle regions. The graph's red sections

show the most permitted areas, while the yellow areas show permitted areas. Ramachandran plot generated by PROCHECK for models developed using the Robetta was represented in (Figure 2). Using Ramachandran Map calculations conducted with the aid of the PROCHECK tool, the stereochemical quality of the predicted models and the quality of the protein models were assessed following the refinement process. For all three proteins, the total number of residues scattered in the most distributed area is greater than 85%, demonstrating the accuracy and high quality of the modelled structure (Sateesh *et al*., 2010) [19] (Table. 5).

Figure 1. Predicted models of effector proteins by Robetta

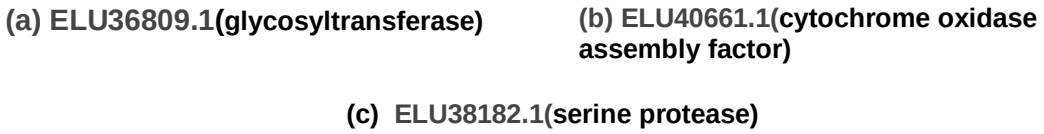

Figure 2: **Ramachandran plot of modelled effector proteins by PROCHECK**

Table 5: Calculation of the Ramachandran plot using the Procheck tool.

Ramachandran plot statistics	Distribution		
	ELU36809.1	**ELU40661.1**	**ELU38182.1**
Residues in most favoured region	89.7%	87.3%	88.3%
Residues in additionally allowed regions	9.0%	10.9%	11.1%
Residues in generously allowed regions	1.0%	0.7%	0.0%
Residues in disallowed regions	0.3%	1.2%	0.6%

4. CONCLUSION

One of the most devastating diseases to damage rice is sheath blight. The AG1 group of effector proteins of sheath blight is thought to be the primary cause of the infection. Here, a few details concerning effector proteins are provided to understand their characteristics. According to current study of physicochemical properties, the AG1 group effector proteins of *R*. *solani* are clearly stable in nature yet frequently reactive and harmful. Robetta is used to secure the PDB format because the proteins do not yet have a PDB format. 3D structure prediction is then employed, and its validation is examined via Ramachandran plot analysis. This information is necessary to perform interaction studies between the protein and promoter region to identify the effector binding element in the host genes.

REFERENCES

1. Anderson, Neil A. The genetics and pathology of Rhizoctonia solani. *Annual review of phytopathology* 20, 1982 no. 1: 329-347.
2. Arulganesh, T., Yaiphabi Kumam, K. K. Kumar, L. Arul, E. Kokiladevi, S. Nakeeran, S. Varanavasiappan, S. Manonmani, and D. Sudhakar. Genome editing of elite rice cultivar CO51 for bacterial leaf blight resistance. *Electronic Journal of Plant Breeding* 12, 2021 no. 4: 1060-1068.
3. Audain, Enrique, Yassel Ramos, Henning Hermjakob, Darren R. Flower, and Yasset Perez-Riverol. Accurate estimation of isoelectric point of protein and peptide based on amino acid sequences. *Bioinformatics* 32, 2016, no. 6: 821-827.
4. Chen, L.Q., Hou, B.H., Lalonde, S., Takanaga, H., Hartung, M.L., Qu, X.Q., Guo, W.J., Kim, J.G., Underwood, W., Chaudhuri, B. and Chermak, D. Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature*, 468, 2010 (7323), pp.527-532.
5. Das, S., F. A. Shah, R. C. Butler, R. E. Falloon, A. Stewart, S. Raikar, and A. R. Pitman. Genetic variability and pathogenicity of *Rhizoctonia solani* associated with black scurf of potato in New Zealand. *Plant Pathology* 63, no. 3 2014: 651-666.
6. Dasgupta, M. K. Rice sheath blight: The challenge continues. Singh, US *et al* (ed.). Plant Diseases of International Importance. Volume I. Diseases of Cereals and Pulses. 1992: 130-157.
7. Errington, Neil, Teuku Iqbalsyah, and Andrew J. Doig. Structure and Stability of the α-Helix. In *Protein Design*, 2006, pp. 3-26. Humana Press.
8. Gamage, Dilani G., Ajith Gunaratne, Gopal R. Periyannan, and Timothy G. Russell. "Applicability of instability index for in vitro protein stability prediction." *Protein and Peptide Letters* 26, 2019, no. 5 : 339-347.
9. Gao, Yue, Chong Zhang, Xiao Han, Zi Yuan Wang, Lai Ma, De Peng Yuan, Jing Ni Wu *et al*. Inhibition of *OsSWEET11* function in mesophyll cells improves resistance of rice to sheath blight disease. *Molecular plant pathology* 19, 2018 no. 9: 2149-2161.
10. http://pfam.xfam.org/search/sequence
11. http://www.ncbi.nlm.nih.gov/
12. https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa%20_sopma.html
13. https://prosite.expasy.org
14. https://web.expasy.org/protparam
15. Khush, Gurdev S. Origin, dispersal, cultivation and variation of rice. *Plant molecular biology* 35, 1997, no. 1: 25-34.
16. Lilly, Jimmy John, and Babu Subramanian. Gene network mediated by *WRKY13* to regulate resistance against sheath infecting fungi in rice (*Oryza sativa L*.). *Plant Science* 280. 2019: 269-282.

17. Magdeldin, Sameh, Yutaka Yoshida, Huiping Li, Yoshitaka Maeda, Munesuke Yokoyama, Shymaa Enany, Ying Zhang *et al*. Murine colon proteome and characterization of the protein pathways. *BioData mining* 5, 2012, no. 1 : 1-14.
18. Peng, Xixu, Yaojun Hu, Xinke Tang, Pinglan Zhou, Xiaobo Deng, Haihua Wang, and Zejian Guo. Constitutive expression of rice WRKY30 gene increases the endogenous jasmonic acid accumulation, PR gene expression and resistance to fungal pathogens in rice. *Planta* 236, 2012, no. 5: 1485-1498.
19. Sateesh, Patchikolla, A. Rao, Suresh Kumar Sangeeta, M. Naresh Babu, and R. S. Grandhi. "Homology modeling and sequence analysis of anxC3. 1." *International Journal of Engineering Science and Technology* 2, 2010, no. 5 : 1125-30.
20. Savary, Serge, Laetitia Willocquet, Francisco A. Elazegui, Nancy P. Castilla, and Paul S. Teng. Rice pest constraints in tropical Asia: quantification of yield losses due to rice pests in a range of production situations. *Plant disease* 84, 2000 no. 3: 357-369.
21. Sigrist, Christian JA, Lorenzo Cerutti, Nicolas Hulo, Alexandre Gattiker, Laurent Falquet, Marco Pagni, Amos Bairoch, and Philipp Bucher. PROSITE: a documented database using patterns and profiles as motif descriptors. *Briefings in bioinformatics* 3, 2002, no. 3 : 265-274.
22. Sivakumar, K., and S. Balaji. "In silico characterization of antifreeze proteins using computational tools and servers." *Journal of Chemical Sciences* 119, 2007, no. 5 : 571-579.
23. Sneh, Baruch, Lee Burpee, and Akira Ogoshi. *Identification of* Rhizoctonia *species*. APS press, 1991.
24. Taheri, Parissa, Sam Gnanamanickam, and Monica Hofte. Characterization, genetic structure, and pathogenicity of Rhizoctonia spp. associated with rice sheath diseases in India. *Phytopathology* 97, 2007 no. 3: 373-383.
25. Wang, Haihua, Jiao Meng, Xixu Peng, Xinke Tang, Pinglan Zhou, Jianhua Xiang, and Xiaobo Deng. Rice WRKY4 acts as a transcriptional activator mediating defense responses toward Rhizoctonia solani, the causing agent of rice sheath blight. *Plant molecular biology* 89, 2015 no. 1: 157-171.
26. Yang, Bing, Akiko Sugio, and Frank F. White. *Os8N3* is a host disease-susceptibility gene for bacterial blight of rice. *Proceedings of the National Academy of Sciences* 103, 2006, no. 27: 10503-10508.