International Journal of Plant & Soil Science

34(22): 1110-1117, 2022; Article no.IJPSS.91244
ISSN: 2320-7035

In-silico Analysis and Functional Characterization of *Rhizoctonia solani* Effector Proteins

K. Prudhvi Raj Kumar [a], P. Vignesh [b], A. Shanthinie [a], D. Sudhakar [a], E. Kokiladevi [a], R. Raghu [a], M. Jayakanthan [c] and S. Varanavasiappan [a*]

[a] Department of Plant Biotechnology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India.
[b] Department of Plant Pathology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India.
[c] Department of Plant Molecular Biology and Bioinformatics, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/IJPSS/2022/v34i2231474

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/91244

Original Research Article

Received 15 June 2022
Accepted 26 August 2022
Published 29 August 2022

ABSTRACT

Sheath blight is considered the second most prevalent disease in rice with no proper resistance genes identified for its resistance. 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, function 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, criteria like *In-silico* analysis and validation of the proteins are to be analyzed. The reactivity and stability of these proteins were evaluated physicochemically, domain prediction, and secondary structure prediction using bioinformatics methods such as Protparam, Pfam, and SOPMA, respectively. These bioinformatics techniques were found to be remarkably suitable for characterizing the protein's function. 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.

**Corresponding author: E-mail: shanvaran@gmail.com;*

Keywords: Effector binding element; In-silico analysis; PROCHECK; R. solani; Robetta.

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 [1]. 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 [2]. According to reports, sheath blight disease causes yield losses ranging from 8 % to 50 % [3]. Although some rice genotypes confer partial genetic resistance to sheath blight through genes/QTL with a small effect, no major gene governing resistance has been identified so far [2]. The conventional breeding method for developing Sheath blight resistance in rice varieties is complicated as the trait shows polygenic inheritance.

R. solani infects about one hundred species of crops, pastures, and horticultural plants. Strains of *R. solani* are capable of infecting cereals, potatoes, beans, cotton, sugar beets, lettuce, cantaloupe, forest trees, and ornamental plants [4,5]. As sclerotia or mycelia, this soil-borne disease attacks rice plants during irrigation and flooding. Because they are inefficient and their excessive use is detrimental to humans, cattle, and the environment, chemical control options for sheath blight are neither sustainable nor practicable. To prevent diseases caused by fungi, genetic engineering can be used to modify plants with genes for fungal resistance.

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 [6,7].

Plant immunity enhances resistance to sheath blight [8,9,10] *OsWRKY4, OsWRKY13, OsWRKY30, and OsWRKY80* modify resistance to sheath blight. SWEET11 (sugar transporter 11) [11] negatively affects rice's susceptibility to sheath blight. Recent research indicates that SWEETs, the targets of pathogen effector proteins during host-microbe interactions, make many plant species susceptible to disease. *Xanthomonas oryzae pv. oryzae* (Xoo) strain PXO99A produces PthXo1, which binds directly to the *OsSWEET11* promoter [12,13]. Delivering TAL effectors to the nucleus activates SWEET genes, ensuring that colonized cell apoplasts get sucrose [12].

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 [14].

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. MATERIALS 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 [15].

2.2 Physio-chemical Characterization

The ProtParam server from Expasy [16] 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 [17]. 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 [18].

2.4 Secondary Structure Prediction

Secondary structural characteristics of effector proteins was calculated using SOPMA (Self Optimized Prediction Method with Alignment) [19] 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.

2.5 Model Building, Evaluation

These protein three-dimensional structures were not available in PDB. Therefore, a web server Robetta [20] was used to model the proteins three-dimensional. The SAVES v6.0 (PROCHECK) [21] 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). There is a range of 350–688 amino acid residues with differing molecular weights for the three investigated effector proteins. The isoelectric point is important for the estimation of protein solubility, electrophoresis, and electrophoretic separation [22]. 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 [23]. 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 [24]. 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 [25].

3.2 Functional Characterization

Using pattern, profile, and Pfam analysis, the functional characterization of effector proteins was carried out (Table 3). Expasy's Prosite was utilised to assess the pattern and profile of the protein. Annotations for motif descriptors are stored in the PROSITE database, which is used to identify protein families and domains [26]. 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. Pfam analyses proteins to increase genome annotation efficiency by semi-automatically curating information on known protein families [27]. Therefore, if the protein families and domains are known, it is possible to correlate their characteristics by comparing them to those of other proteins with similar domains and families.

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

**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*

3.3 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 [28].

3.4 Model Building and Validation

The three-dimensional structure of the proteins was modelled by Robetta, an *abinitio* approach [29] (Fig. 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 (Fig. 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 [30] (Table 5).

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

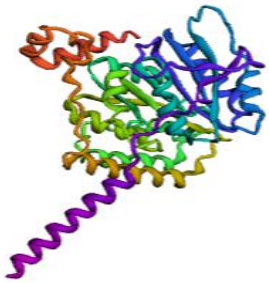
Protein ID	Name of protein (as available on NCBI)	Pfam Analysis	Patterns by Prosite	Profile by Prosite
ELU36809.1	glycosyltransferase family 2 protein	Glycosyl transferase family 2(Family)	Casein kinase II phosphorylation site N-myristoylation site N-glycosylation site Protein kinase C phosphorylation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

Table 4. Predicted secondary structures present in proteins

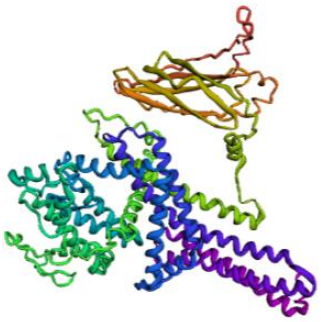
Proteins	Amino acids	Alpha helix	Extended sheets	Beta turns	Random coils
ELU36809.1	350	46.00%	15.71%	5.14%	33.14%
ELU40661.1	688	25.87%	23.69%	6.54%	43.9%
ELU38182.1	405	30.37%	23.95%	7.16%	38.52%

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%



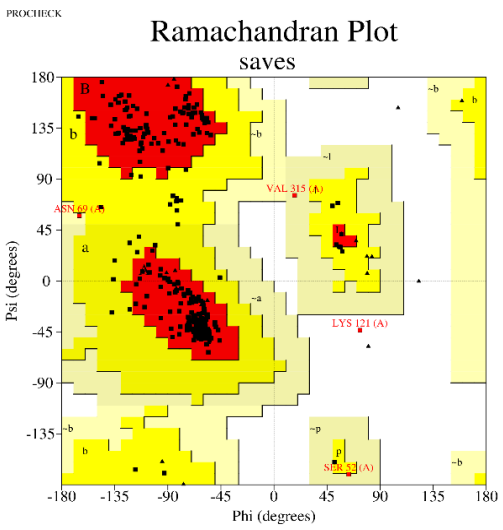
(a) ELU360809.1



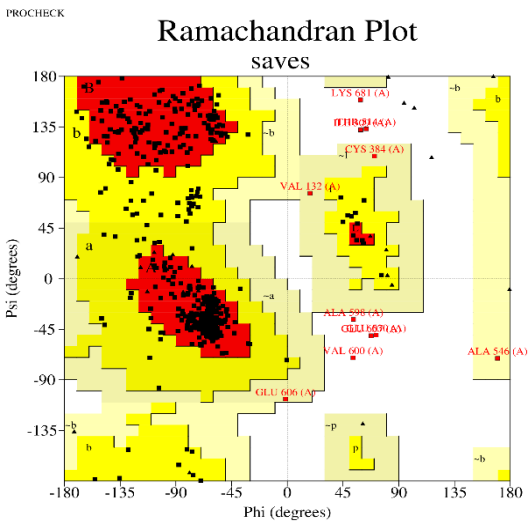
(b) ELU40661.1

(c) ELU38182.1

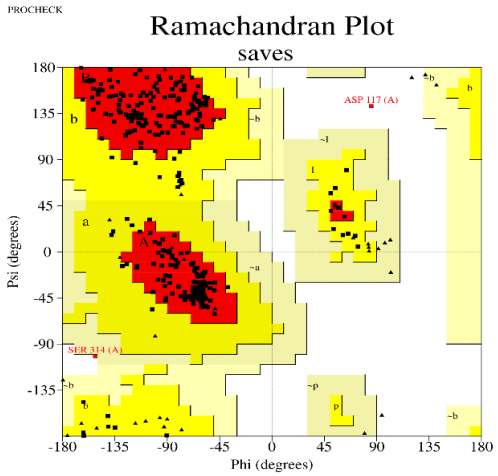
Fig. 1. Models of effector proteins predicted by the Robetta server



(a) ELU36809.1



(b) ELU40661.1



(c) ELU38182.1

Fig. 2. The Ramachandran plot generated by the PROCHECK server using the modelled effector proteins

4. CONCLUSION

The purpose of this article is to inform readers about the effector proteins of *R. solani*, which pose a severe threat to a variety of crops, particularly rice, by generating sheath blight, one of the most devastating diseases. The AG1 group of effector proteins of sheath blight is thought to be the primary cause of the infection. A few details regarding effector proteins are provided to aid in understanding the characteristics of certain AG1 effector proteins. Recent examination of physicochemical properties, the AG1 group effector proteins of *R. solani* are clearly stable 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 validated protein information is required for interaction studies between the protein and the promoter region in order to discover the effector binding element in the host genes. These effector binding elements should be known to knock off, perhaps resulting in sheath blight resistance.

ACKNOWLEDGEMENTS

The authors are grateful to the Department of Plant Biotechnology at Tamil Nadu Agricultural University in Coimbatore for providing the research facilities. The authors express their heartfelt appreciation to the Department of bioinformatics for giving enormous knowledge and support for this research.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Khush GS. Origin, dispersal, cultivation and variation of rice. Plant Molecular Biology. 1997;35(1):25-34.
2. Dasgupta MK. Rice sheath blight: The challenge continues. Plant Diseases of International Importance. 1992;1:130-157.
3. Savary S, Willocquet L, Elazegui FA, Castilla NP, Teng PS. Rice pest constraints in tropical Asia: Quantification of yield losses due to rice pests in a range of production situations. Plant Disease. 2000;84(3):357-369.
4. Anderson NA. The genetics and pathology of *Rhizoctonia solani*. Annual Review of Phytopathology. 1982;20(1):329-347.
5. Sneh B, Burpee L, Ogoshi A. Identification of *Rhizoctonia* species. 1991: APS press.
6. Taheri P, Gnanamanickam S, Höfte M. Characterization, genetic structure, and pathogenicity of *Rhizoctonia* spp. associated with rice sheath diseases in India. Phytopathology. 2007;97(3):373-383.
7. Das S, Shah FA, Butler RC, Falloon RE, Stewart A, Raikar S, et al. Genetic variability and pathogenicity of *Rhizoctonia solani* associated with black

scurf of potato in New Zealand. Plant Pathology. 2014;63(3):651-666.
8. Peng X, Hu Y, Tang X, Zhou P, Deng X, Wang H, et al. Constitutive expression of rice WRKY30 gene increases the endogenous jasmonic acid accumulation, PR gene expression and resistance to fungal pathogens in rice. Planta. 2012; 236(5):1485-1498.
9. Wang H, Meng J, Peng X, Tang X, Zhou P, Xiang J, et al. Rice WRKY4 acts as a transcriptional activator mediating defense responses toward *Rhizoctonia solani*, the causing agent of rice sheath blight. Plant Molecular Biology. 2015;89(1):157-171.
10. Lilly JJ, Subramanian B. Gene network mediated by WRKY13 to regulate resistance against sheath infecting fungi in rice (*Oryza sativa* L.). Plant Science. 2019; 280:269-282.
11. Gao Y, Zhang C, Han X, Wang ZY, Ma L, Yuan DP, et al. Inhibition of OsSWEET11 function in mesophyll cells improves the resistance of rice to sheath blight disease. Molecular Plant Pathology. 2018;19(9): 2149-2161.
12. Chen LQ, Hou BH, Lalonde S, Takanaga H, Hartung ML, Qu XQ, et al. Sugar transporters for intercellular exchange and nutrition of pathogens. Nature. 2010; 468(7323):527-532.
13. Yang B, Sugio A, White FF. Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. Proceedings of the National Academy of Sciences. 2006; 103(27):10503-10508.
14. Arulganesh T, Kumam Y, Kumar KK, Arul L, Kokiladevi E, Nakeeran S, Varanavasiappan S, Sudhakar D. Genome editing of elite rice cultivar CO51 for bacterial leaf blight resistance. Electronic Journal of Plant Breeding. 2021; 12(4):1060-1068.
15. Available:http://www.ncbi.nlm.nih.gov/
16. Available:https://web.expasy.org/protparam
17. Available:https://prosite.expasy.org
18. Available:http://pfam.xfam.org/search/sequence
19. Available:https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa%20_sopma.html
20. Available:https://robetta.bakerlab.org/
21. Available:https://saves.mbi.ucla.edu/
22. Audain E, Ramos Y, Hermjakob H, Flower DR, Perez-Riverol Y. Accurate estimation of isoelectric point of protein and peptide based on amino acid sequences. Bioinformatics. 2016;32(6): 821-827.
23. Sivakumar K, Balaji S. In silico characterization of antifreeze proteins using computational tools and servers. Journal of Chemical Sciences. 2007; 119(5):571-579.
24. Gamage DG, Gunaratne A, Periyannan GR, Russell TG. Applicability of instability index for *in vitro* protein stability prediction. Protein and Peptide Letters. 2019;26(5): 339-347.
25. Magdeldin S, Yoshida Y, Li H, Maeda Y, Yokoyama M, Enany S, et al. Murine colon proteome and characterization of the protein pathways. BioData Mining. 2012; 5(1):1-14.
26. Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, et al. PROSITE: A documented database using patterns and profiles as motif descriptors. Briefings in Bioinformatics. 2002;3(3):265-274.
27. Liu Z, Ma H, Goryanin I. A semi-automated genome annotation comparison and integration scheme. BMC Bioinformatics. 2013;14(1):1-12.
28. Errington N, Iqbalsyah T, Doig AJ. Structure and Stability of the α-Helix. In Protein Design. 2006;3-26. Humana Press.
29. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Research. 2004;32(Suppl_2):W526-W531.
30. Sateesh P, Rao A, Sangeeta SK, Babu MN, Grandhi RS. Homology modeling and sequence analysis of anxC3. 1. International Journal of Engineering Science and Technology. 2010;2(5):1125-30.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/91244