Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278–6007
Vol 5, Issue3, 2014, pp150-162
http://www.bipublication.com

In-Silico Analysis of Structural Properties of Pathogen-Related Protein (PR1) in Potato Somatic Hybrid

Ritu Singh[1], Shashi Rawat[1]*, Jagesh K. Tiwari[1], Vinay Sharma[2] and B. P Singh[1]

[1]Central Potato Research Institute, Shimla-171 001, Himachal Pradesh, India
[2]Department of Bioscience and Biotechnology, Banasthali University, Rajasthan 304022, India
Corresponding author: Email: shashi29@yahoo.com

[Received-30/04/2014, Accepted-01/08/2014]

ABSTRACT

Pathogen-related protein (PR) are induced in the host plants as a defense response during infection with oomycetes, fungi, bacteria, viruses, insect attack, wounding, salinity, exposure to harsh chemicals and UV light. PR proteins are used as marker gene for systemic acquired resistance. The three-dimensional structure of pathogen related protein from somatic hybrid of *Solanum tuberosum* (+) *S. pinnatisectum* was constructed from using the crystal structure of P14A from *Solanum lycopersicum* (PDB ID: 1CFE) as template using Modeller9.11. The best model was further assessed by PROCHECK, ProSA, and ERRAT plot in order to analyze the quality and reliability of generated model. The overall quality of computed model from Ramachandran plot revealed that 99.1% amino acid sequence in favored and allowed regions. The StPR1 forms α-β-α sandwich fold with four α helices (α1-α4) and four stranded β sheets (β1-β4) which forms a tight topology stabilized by hydrogen bonding. The three-dimensional structure of PR from the somatic hybrid has a conserved cleft with two histidine and two glutamate residues which forms a putative active site in the model protein. This model might be helpful in developing new strategies to develop transgenic plants with increased fungal resistance in the field in response to disease and stress.

Keywords: Homology modeling, pathogen-related protein, protein structure, potato somatic hybrid, *Solanum tuberosum*, ProCheck validation.

Abbreviations

PR	Pathogen-related protein
SAR	systemic acquired resistance
PGSC	Potato Genome Sequencing Consortium
CRISP	cysteine-rich secretory protein
PDB	Protein Data Bank
EP	electrostatic potential

INTRODUCTION

Plant diseases caused by wide range of pathogens and during stress are responsible for causing huge economic loss in crop productivity. Therefore, plant defenses responses against pathogen infection are crucial for the plant survival. Plants in response to pathogen infection possess both basal and inducible mechanisms which provides defense against variety of pathogens. However the activation of defense-related genes encoding pathogenesis-related (PR) proteins is one of the most important and effective defense mechanism. The PR proteins belong to "stress-inducible" or "defense-related" proteins family which decreases plant susceptibility by forming a protective barrier against pathogens at infection sites. The PR proteins are widely distributed in the plant, animal and fungal kingdoms proteins. In the plant kingdom, PR proteins are often used as molecular marker for systemic acquired resistance (SAR) in plants [1]. PR proteins were first discovered in *Nicotiana tabacum* showing hypersensitivity to tobacco mosaic virus infection [2]. PR proteins have anti-fungal or anti-bacterial activity against a wide range of pathogens.

In plants, PR proteins are synthesized by plants during pathogen infection (virus, viroids, bacteria, fungi, nematodes, insects and herbivores), wounding, osmotic stress, salinity, exposure to harsh chemicals and other stress-related responses [3, 4]. PR proteins employ various function in cellular defense and plant development such as leaf senescence, pollen maturation, as well as environmental factors such as cold, salinity, osmotic stress and UV-light [5]. The PR proteins have low molecular weight, highly resistant to proteases, possess low pH and are localized predominantly in the intercellular spaces of leaves. Presently PR proteins are classified into 17 families (PR-1 to PR-17) based on their function, biochemical properties and immunological relationship [6]. PR-1 family is the first discovered PRs having molecular mass 15-17 kDa whose biological activity is still unknown. The PR-1 proteins are the most abundantly accumulated in response to pathogen infection and stress and also play a vital role in plant development. Alexender et al. (1990) [7] demonstrated that PR-1a is a defense protein by expressing PR-1a in transgenic tobacco confers resistance to fungal pathogens. Anti-fungal activity of PR-1 proteins have been detected in tobacco and tomato which inhibit the zoospore germination and mycelial growth of *P. infestans* [8]. PR-1 proteins have been identified in Arabidopsis, wheat, rice, tobacco, pepper barley, maize, tomato, potato, broad bean, *Gomghrena globosa*, Chenopodium and *Camellia sinensis*.

Potato (*Solanum tuberosum* L.) is the third most important non-cereal food crop after rice and wheat (http://faostat.fao.org) which is grown throughout the globe. The completion of the potato genome (Potato Genome Sequencing Consortium, PGSC) helps to identify all the PR-1 protein at genomic level. In this paper, we elucidate the three dimensional structure along with physicochemical properties of PR1 protein from the late blight resistant potato somatic hybrid of *Solanum tuberosum* (+) *S. pinnatisectum* [9] which would provide insights into the functional mechanism. However, 3D crystal structure of PR1 protein from potato (*S. tuberosum*) remains unknown. In the present study, we generate the 3D structure of the PR1 protein M1A2A4 gene id PGSC0003DMG400005111 from interspecific potato somatic hybrid of *Solanum tuberosum* (+) *S. pinnatisectum*. The model structure was further validated using various standard parameters like PROCHECK, ProSA, Verify3D. This study is useful in functional

characterization of PR-1 gene from *Solanum tuberosum* (StPR1) in response to defense.

MATERIAL AND METHODS

Template identification and alignment

Based on the previous study of global gene expression profile of highly up-regulated gene PR1 (gene id PGSC0003DMG400005111) from interspecific potato somatic hybrid of *Solanum tuberosum* (+) *S. pinnatisectum* (unpublished data) was retrieved from Uniprot containing 179 amino acid residues. A sequence similarity NCBI-Blast (http://www.ncbi.nlm.nih.gov/Blast) [10] was used to retrieve the corresponding template structure from the Protein Data Bank (PDB) [11] using Blossum 62 substitution matrix with e-value cut-off of 10. The solved structure with high percentage identity and resolution were selected as template. A further step consisted of performing a multiple sequence alignment of the selected templates and target sequence using the program ClustalW with the Blossum 62 substitution matrix with gap penalties of gap start 1 and gap extension 0.1 was used for alignment [12].

Model building and energy refinement

Molecular modelling and computations in this study were carried out on Z800 workstation on windows 7 platform by using the comparative protein modelling program Modeller9.11 [13]. The program uses an automated approach for comparative protein structure modelling. In brief, the modelling procedure begins with an alignment of the sequence to be modelled (target) with related known 3D structures (templates). The output is a 3D model for the target sequence containing all main chain and side chain non-hydrogen atoms.

Model quality and evaluation

The stereochemistry of the models was further gauged by Ramachandran plot using PROCHECK program accessible at SAVES Server (http://nihserver.mbi.ucla. edu/SAVES). The PROCHECK program provides the information about the stereo chemical quality of a given protein structure and was used to generate Ramachandran plot. All final models were inspected for the DOPE (discrete optimized potential energy) score of modeller output per residues of the model. DOPE score calculated by Modeller program is the distance dependent statistical potential based on probabilistic theory. ProSA-Web-server (https://prosa.services.came.sbg.ac.at/prosa.php) was used to test the local and overall quality of the developed models from Modeller. The quality of the fold was inspected with PROSA [14] which allows all residues in negative energy regions very similar to the template protein, indicating the possible correctness of the modelled structures.

Electrostatic potential

To understand more characteristic features of StPR1, MEP analysis was performed. The electrostatic interaction is a crucial part of the non-covalent interaction energy between molecules. The electrostatic potential (EP) on the surface is generally colored according to the sign and magnitude of the potential. The color ramp for EP ranges from red (most positive) to blue (most negative).

Physiochemical parameter analysis

The physico-chemical characterization of StFT protein were computed using the Expasy's ProtParam server (http://web.expasy.org/protparam/) [15] which computes theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average hydropathy (GRAVY).

Studying solvent accessibility of the protein model

Further the solvent accessibility of the amino acid residues in the modelled protein was

determined by using ASA-view (http://gibk26.bse.kyutech.ac.jp/~shandar/netasa/ asaview/) [16].

Domain prediction

Domain prediction was done by InterProScan (V4.8; http://www.ebi.ac.uk/Tools/pfa/iprscan/). InterProScan is a sequence analysis application which provides functional analysis of proteins by classifying them into families and predicting domains and functional sites.

Transmembrane domain prediction

Transmembrane domain prediction was done using SOSUI server (http://bp.nuap.nagoya-u.ac.jp/sosui/sosuiG/sosuigsubmit.html) [17]. SOSUI distinguishes between membrane and soluble proteins from amino acid sequences and predicts the transmembrane helices for the membrane proteins.

RESULTS AND DISSCUSSION

The first step toward constructing the homology model of StPR1 (M1A2A4_SOLTU; gene id PGSC0003DMG400005111) was a template search, carried out in NCBI-Blast using Blossum 62 substitution matrix with e-value cut-off of 10. The program provided the most promising template structures for homology modelling. The NMR solution structure of pathogenesis-related protein P14A from *Solanum lycopersicum* (PDB ID: 1CFE) was selected as template to build the model with high percentage identity was selected as template for alignment. The template protein is 135 amino-acids in length which shares 90% identity with StPR1 protein. The selected templates (1CFE) were aligned pairwise (globally) to the query sequence StPR1 using needle program in EMBOSS-GUI (http://bips.ustrasbg.fr/EMBOSS) revealed that the identity and similarity between the target and template protein was 50.8% and 55.9%, respectively. Further the selected template (1CFE) were aligned pairwise (locally) to the query sequence StPR1 using water program in EMBOSS-GUI (http://bips.ustrasbg.fr/EMBOSS) revealed the identity and similarity between the target and template protein was 66.4% and 73.0% respectively. A multiple sequence alignment was carried out with StPR1 and the selected template by using multiple sequence algorithm ClustalW (Fig. 1). The structurally conserved regions (SCR's) and structurally variable regions (SVR's) were defined and model was built using Modeller9.11. The program deduces the distance and angle constraints from a template structure and combines them with energy terms for adequate stereochemistry in an objective function that is later optimised in Cartesian space with conjugate gradients and molecular dynamic methods. Ten sets of models were generated. The 8th model has lowest DOPE score so this model was selected for further study.

The quality of the model was further assessed using Ramachandran plot in PROCHECK validation package. In this model, 89.5% of the amino acid residues were in the favourable regions of the plot for the protein, whereas 9.6% of the residues were in the additionally allowed region and no residue in generously allowed region. There is only one residue (Asn 27) in disallowed region contributing to 0.9%. The phi and psi distributions of the Ramachandran plot of non-Glycine, non-Proline residues are summarized in (Fig. 2). Altogether 99.1% of the residues was in favoured and allowed regions. The Verify3D results show that 99.28% of the residues have an averaged 3D-1D score > 0.2. The overall quality factor from Errat is 65.116.

Further refinement of the final StPR1 model was done by energy minimization of the selected outlier residues using protein report tool. The final and refined three dimensional

structure of the StPR1 superimposes well on the crystal structure of the templates taken for model building. The alpha-trace of StPR1 generated was found to be similar to resolved crystal structure of the 1CFE in the same orientation (Fig. 3). This indicates that the model was sufficiently accurate for the structure based analysis. ProSA-Web-server analysis revealed that the modelled structure occupies a region of NMR predicted native protein structures of same size with Z score of -5.39 (Fig. 4). This supports that the StPR1 model is sufficiently accurate. The solvent accessibility of the StPR1 protein model is shown in Fig. 5. The positive charged residues (R, K, H) are shown in blue, negative charged residues (D, E) are in red, polar uncharged residues (G, N, Y, Q, S, T, W) are shown in green whereas grey color determines the hydrophobic residues.

At sequence level comparison of StPR1 along with the template reveals a putative N-terminal (1-23aa length) signal peptide for secretion and C-terminal elongated polypeptide 19 aa in length. Further analysis from the PROSITE database revealed that the StPR1 protein have conserved two cysteine-rich secretory protein (CRISP) family namely CRISP1 from 117-127 namely GHYTQVVWRNS and CRISP2 FITCNYDPPGNW at the in C-terminal of the protien. This shows that there is one amino acid difference in CRISP1 and two amino acid differences in CRISP2 between the template and StPR1 protein, respectively (Table 1). CRISP play various functions in plants in relation to innate host defence and the blockage of ion channels. However, in plants, PR proteins with a CRISP domain are synthesized under abiotic stress playing defensive role against pathogens [18]. These proteins are highly expressed and conserved in animals, plants, fungi, humans, insect and reptile venoms. The CRISPs domain of PR1 proteins exhibit diverse biological functions and high amino acid sequence similarity to protein DE in rats [19], acidic epididymal glycoprotein (AEG) in mice [20] and glioma pathogenesis-related protein (GLIPR) in humans [21] indicating an important functional role of this domain between the human immune system and a plant defense system [22].

The topology of StPR1 protein model consists of a CAP protein domain (S21-Y160 amino acids) comprising of cysteine-rich secretory proteins (CRIPS), allergen V5 and PR1 protein (Fig. 6). The StPR1 protein consists of four antiparallel β strands and four alpha helices α strands which form the conserved three stacked layers of alpha-beta-alpha sandwich core architecture stabilized by a buried hydrogen bonding network. This α-β-α sandwich fold in modelled protein StPR1 has a conserved structure characterized in tomato P14a, vespid Ves v 5, golgi associated pathogenesis related-1 (GAPR1) and snake venom stecrisp. The four α helices consists of residues α1 4-18, α2 28-43, α3 64-73, α4 94-98 which is a single turn of 3_{10}-helix. The four antiparallel β sheet β1 24-25, β2 53-59, β3 106-112, β4 115-121 (Fig. 7).

The short helix α4 is completely buried between the β sheet (β1-β4) and the two α1 and α3. The tertiary fold of StPR1 is stablized by six cysteines residues which are highly conservation throughout plant families. All cysteines residues in StPR1 protein are involved in disulfide bonds except for Cys11 which is not included in the structure. The StPR1 protein form three conserved disulfide bonds namely Cys87-Cys93, Cys109-123, Cys44-Cys114. The Cys87-Cys93 is located in the loop connecting the α3 and α4 where as Cys109-Cys123 directly connects the antiparallel β3 and β4. The Cys44-Cys114 connects the loop connecting the α2 and the β2 with the loop between the β3 and β4.

The central part of the StPR protein is formed two anti-parallel beta sheets β3 and β4 connect by a disulfide bond Cys109-123, the β2 strand is anti-parallel to β4 and the β1 is parallel to β3. The β-sheets in the modelled protein show connectivity of +3x, −2 x, +1 forming a topology in Richardson notation [23]. The StPR1 forms a complex α+β topology where the central beta strands are arranged between alpha helices (α1, α3 and α4) on one side and α2 on the other side respectively forming a tight topology which is stabilized by hydrophobic interactions and hydrogen bonding (Fig. 8). The 3D model of StPR1 has a conserved cleft with putative active site center consisting of two histidine and two glutamate residues namely His (H48), Glu (E55), Glu (E76) and His (H95) which are conserved in the appropriate positions in the model protein structure revealing that PR1 proteins are strongly conserved throughout the evolutionary process. However the exact roles of these residues are still unknown. The StPR1 homology model shows many structural differences with the solvent models of P14a. The alpha helix 3 and beta strand 2 are smaller in length in StPR1 as compared to the crystal structure P14a. Superposition of the active sites residues reveals that the active site residues represent similar but the orientations of the active sites residues with solvent models of P14a. The two histidines are oriented to the center of the cleft directly forms hydrogen bonds with the acidic amino acid glutamic acid (Fig. 9). The Cuc m 3 is a plant allergen from muskmelon revealing its involvement in food allergy. The Cuc m 3 shows 47.7% of sequence identity with PR-1 members from *S. tuberosum*. Moreover StPR1 model has a Ves allergen (Ves v 5) domain comprising of 115-134 aa which is one of the major allergen of vespid wasp venom shows high amino acid sequence identity to human, mouse, rat and PR from *N. tabacum* and *S. lycopersicum* leaves. It has been observed that plant PR-1 proteins is involved in systemic acquired resistance which are highly up-regulated proteins after pathogen attack and suppression of infection is mainly induced targeting their cell wall. The growing fungal hyphae tip a high Ca2+ gradient and low Ca2+ concentrations result in inhibiting hyphal growth and development. The StPR1 shows similarity to Mexican banded lizard or snake venoms helothermine protein which interacts with membrane-channel proteins of target cell and and function as Ca2+ ion channel blocker [24]. However determining the electronegative character of StPR1 regions may be crucial in conferring antifungal activity. The transgenic tobacco plants transformed with pepper PR-1 protein (CABPR-1) confers enhanced resistance to the oomycete pathogen *Phytophthora nicotianae* and the bacterial pathogens *Ralstonia solanacearum* and *Pseudomonas syringae* pv. tabaci along with partial tolerance to heavy metal stress [25]. Based on the structure of pathogenesis-related protein P14A from *S. lycopersicum* (1CFE) it was suggested that all pathogen related proteins have an electrostatically polarized surface that may be critical for antifungal activity of this group of proteins. Here the electrostatic surface potential of StPR1 varies from -58.277 to -58.277 (Fig. 10). However in broad bean and rice, the induction of PR-1 protein are often used as markers for acquired resistance with potential applications in plant defense responses exhibiting strong inhibiting activity against the rust fungus *Uromyces fabae* in broad bean [26, 27, 28].

The physicochemical properties deduced ProtParam shows that the StPR1 protein sequence has high percentage of Ala (A) and Asn (R) 10.6%, 10.1% and least for Glu (E), Lys (K), Met (M), His (H) (2.2 %) respectively. The total number of positively charged residues (+R) (Arg+Lys) is 15 and

negatively charged residues (-R) (Asp+Glu) is 12. The protein has molecular weight 20223.6 with theoretical pI 8.48 showing its basic nature. The aliphatic index (AI) is defined as the relative volume of a protein occupied by aliphatic side chains (A, V, I and L) which is regarded as a positive factor for the increase of thermal stability of globular proteins. The thermo stability of the protein is assessed by the aliphatic index. The aliphatic index of StPR1 protein sequence is 60.61 showing that the StPR1 protein is stable at variable temperature. The stability of a protein is calculated via the instability index which indicates the extent of stability of the proteins. However it is reported that proteins having instability index of less than 40 is considered stable while more than 40 is unstable. The instability index values of StPR1 protein sequence were found to be 32.36 revealing that this protein is stable at various temperatures. The Grand Average hydropathy (GRAVY) value for a protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. The Grand Average hydropathy index (GRAVY) indicates the solubility of proteins. A GRAVY index of StFT is -0.525. A negative GRAVY value for StFT describes its hydrophilic nature (Table 2).

Functional analysis of StPR1 proteins includes prediction of one transmembrane helix region. The transmembrane domain requires at least 18 residues to span the membrane. The region of the transmembrane domain in StPR1 has length approximately 20 residues from AAMGYSNIALIICFLIFAIFHSS (2-23) at N terminal region. The average of hydrophobicity calculated by SOSUI is -0.524581. The transmembrane regions and their length predicted in StPR1 is described in Table 3 and hydropathy profile is shown in Fig 11.

CONCLUSION

In this study, we have developed a homology model of StPR1 protein based on known crystal structure of PR protein P14A from *Solanum lycopersicum* based on sequence similarity. Structural and functional features of StPR1 protein have been well-characterized by three dimensional modelling. The StPR1 show homologies and structural motifs in common with proteins from fungi, animals and humans. It also provides a platform for a systematic mutagenesis strategy. The antifungal property of PR proteins isolated could be used in agribusiness to create genetically modified potato plants with increased fungal resistance in the field. The StPR1 homology model provides the insight into the role played by PR proteins in response to disease and stress.

ACKNOWLEDGEMENT

The authors are grateful to the Director, Central Potato Research Institute (ICAR), Shimla, for providing the necessary facilities during the work.

REFERENCES

1. Klessig, D.F., Durner, J., Noad, R., Navarre, D.A., Wendehenne, D., Kumar, D., Zhou, J.M., Shah, J., Zhang, S., Kachroo, P., Trifa, Y., Pontier, D., Lam, E., Silva, H. (2000) Nitric oxide and salicylic acid signaling in plant defense. Proc Natl Acad Sci USA 97: 8849-55.
2. van Loon, L.C., van Kammen, A. (1970) Polyacrylamide disc electrophoresis of the soluble leaf proteins from *Nicotiana tabacum* var. 'Samsun' and 'Samsun NN': II. Changes in protein constitution after infection with tobacco mosaic virus. Virology 40: 199-01.
3. Warner, S.A.J., Scott, R., Draper, J. (1992) Characterization of a wound-induced transcript from the monocot asparagus that shares similarity with a class of intracelluar pathogenesis-related (PR) proteins. Plant Mol Biol 19: 555-61.
4. Iturriaga, E.A., Leech, M.J., Barratt, D.H.P., Wang, T.L. (1994) Two ABA-responsive proteins from pea (*Pisum sativum* L) are

closely related to intracellular pathogenesis-related protein. Plant Mol Biol 24: 235-40.

5. Kitajima, S., Sato, F. (1999) Plant pathogenesis-related proteins: molecular mechanisms of gene expression and protein function. J Biochem (Tokyo) 125: 1-8.
6. van Loon, L.C., Rep, M., Pieterse, C.M. (2006) Significance of inducible defense-related proteins in infected plants. Annu Rev Phytopathol 44: 135-62.
7. Alexander, D., Goodman, R.M., Gut-Rella, M., Glascock, C., Weymann, K., Maddox, F.L.D., Ahl-Goy, P., Luntz, T., Ward, E., Ryals, J. (1990) Increased tolerance to two oomycete pathogens in transgenic tobacco expressing pathogenesis-related protein 1a. Proc Natl Acad Sci USA 90: 7327-31.
8. Niderman, T., Genetet, I., Bruyère, T., Gees, R., Stintzi, A., Legrand, M., Fritig, B., Mösinger, E. (1995) Pathogenesis-related PR-1 proteins are antifungal. Isolation and characterization of three 14-kilodalton proteins of tomato and of a basic PR-1 of tobacco with inhibitory activity against *Phytophthora infestans*. Plant Physiol 108: 17-27.
9. Sarkar, D., Tiwari, J.K., Sharma, S., Poonam, Sharma, S., Gopal, J., Singh, B.P., Luthra, S.K., Pandey, S.K., Pattanayak, D. (2011) Production and characterization of somatic hybrids between *Solanum tuberosum* L. and *S. pinnatisectum* Dun. Plant Cell Tiss Organ Cult 107: 427-40.
10. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. J Mol Biol 215: 403-10.
11. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000) The Protein Data Bank. Nucleic Acids Research 28: 235-42.
12. Thompson, J.D., Higgins, D.G., Gibson, T. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673-4680.
13. Sali, A., Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234: 283-291.
14. Wiederstein, M., Sippl, M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research, 35: W407-W410.
15. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker (eds) The Proteomics Protocols Handbook. Humana Press, pp 571-607.
16. Ahmad, S., Gromiha, M., Fawareh, H., Sarai, A. (2004) ASAView: Database and tool for solvent accessibility representation in proteins. BMC Bioinformatics 5: 51.
17. Hirokawa, T., Boon-Chieng, S., Mitaku, S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics 14: 378-9.
18. Fernández, C., Szyperski, T., Bruyère, T., Ramage, P., Mösinger, E., Wüthrich, K. (1997) NMR solution structure of the pathogenesis-related protein P14a. J Mol Biol 266: 576-93.
19. Brooks, D.E. (1982) Purification of rat epididymal proteins 'D' and 'E,' demonstration of shared immunological determinants, and identification of regional synthesis and secretion. Int J Androl 5: 513-24.
20. Kasahara, M., Hayashi, M., Yoshida, M.C., Nadeau, J.H., Fujimoto, S., Ishibashi, T. (1995) Mapping of acidic epididymal glycoprotein (Aeg) genes to mouse chromosome 17. Mamm Genome 6: 52-54.
21. Hayashi, M., Fujimoto, S., Takano, H., Ushiki, T., Abe, K., Ishikura, H., Yoshida, M.C., Kirchhoff, C., Ishibashi, T., Kasahara, M. (1996) Characterization of a human glycoprotein with a potential role in sperm-egg fusion: cDNA cloning, immunohistochemical localization, and chromosomal assignment of the gene (AEGL1). Genomics 32: 367-74.
22. Szyperski, T., Fernandez, C., Mumenthaler, C., Wuthrich, K. (1998) Structure comparison of human glioma pathogenesis-related protein GliPR and the plant pathogenesis-related

protein P14a indicates a functional link between the human immune system and a plant defense system. Proc Natl Acad Sci USA 95: 2262-66.

23. Richardson, J.S. (1981) The anatomy and taxonomy of protein structure. Advan Protein Chem 34: 167-39.
24. Morrissette, J., Kratzschmar, J., Haendler, B., el-Hayek, R., Mochca-Morales, J., Martin, B.M., Patel, J.R., Moss, R.L., Schleuning, W.D., Coronado, R., Possani, L.D. (1995) Primary structure and properties of helothermine, a peptide toxin that blocks ryanodine receptors. Biophys J 68: 2280-88.
25. Sarowar, S., Kim, Y.J., Kim, E.N., Kim, K.D., Hwang, B.K., Islam, R., Shin, J.S. (2005) Overexpression of a pepper basic pathogenesis-related protein 1 gene in tobacco plants enhances resistance to heavy metal and pathogen stresses. Plant Cell Rep 24: 216-24.
26. Rauscher, M., Adam, A.L., Sabine, W., Richard, G., Kurt, D., Holger, B.D. (1999) PR-1 protein inhibits the differentiation of rust infection hyphae in leaves of acquired resistant broad bean. Plant J 19: 625-33.
27. Agrawal, G.K., Jwa, N.S., Rakwal, R. (2000a) A novel rice (*Oryza sativa* L.) acidic PR1 gene highly responsive to cut, phytohormones, and protein phosphatase inhibitors. Biochem Biophys Res Commun 274: 157-65.
28. Agrawal, G.K., Rakwal, R., Jwa, N.S. (2000b) Rice (*Oryza sativa* L.) OsPR1b gene is phytohormonally regulated in close interaction with light signals, Biochem Biophys Res Commun 278: 290-98

Figures

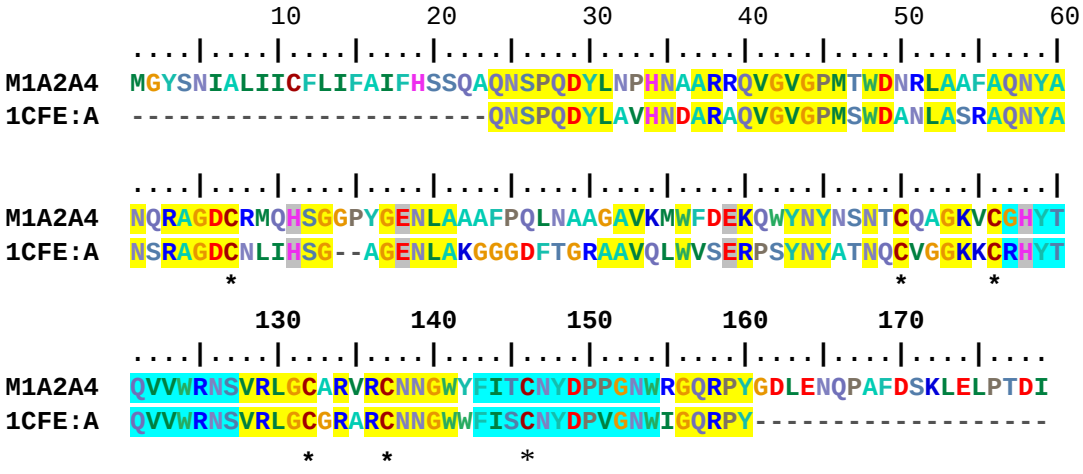
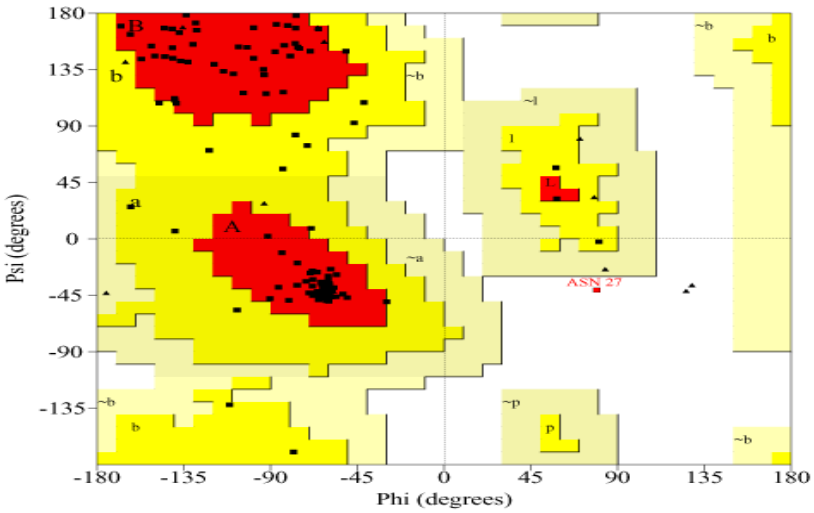


Fig. 1 Multiple-sequence alignment of StPR1 with the crystal structure of pathogenesis-related protein P14A from *Solanum lycopersicum* (1CFE) as template. The one-letter amino acid code is used for the alignment. Yellow color indicates the highly conserved regions, the CRIPS motifs are shown with cyan color. The dash represent gaps inserted to maximize the extent of homology among sequences. The histidine and glutamate amino acids involved the putative active site are indicated in gray color. The positions of the conserved cysteine residues forming disulphide bonds in StPR1 are marked by *.The alignment was produced with BioEdit.



Ramachandran plot summary

Residues in most favoured regions	102	89.5%
Residues in additional allowed regions	11	9.6%
Residues in generously allowed regions	0	0.0%
Residues in disallowed regions	1	0.9%
Number of non-glycine and nonproline residues	114	100%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	13	
Number of proline residues	8	
Total number of residues	**137**	

Fig. 2 PROCHECK-Ramchandran plot and summary for the homology model of StFT.

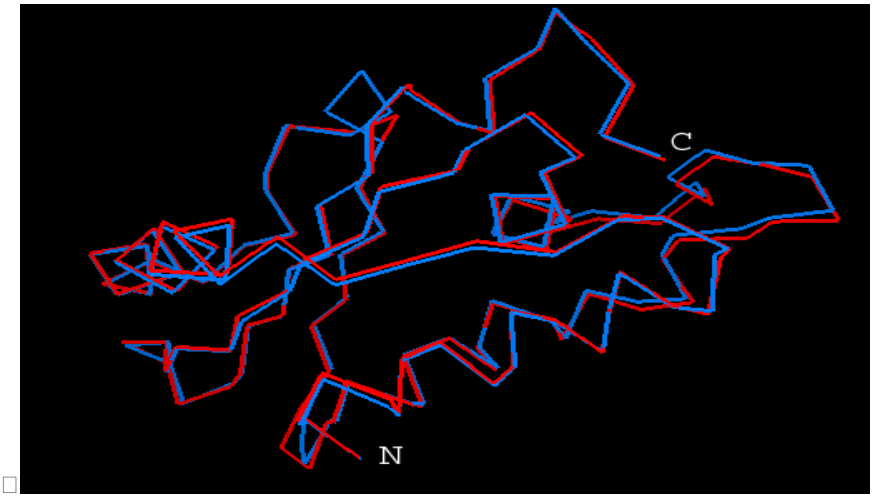


Fig. 3 Superposition of alpha-trace of StPR1 (red) and template 1CFE (blue). Figure is prepared with PyMol.

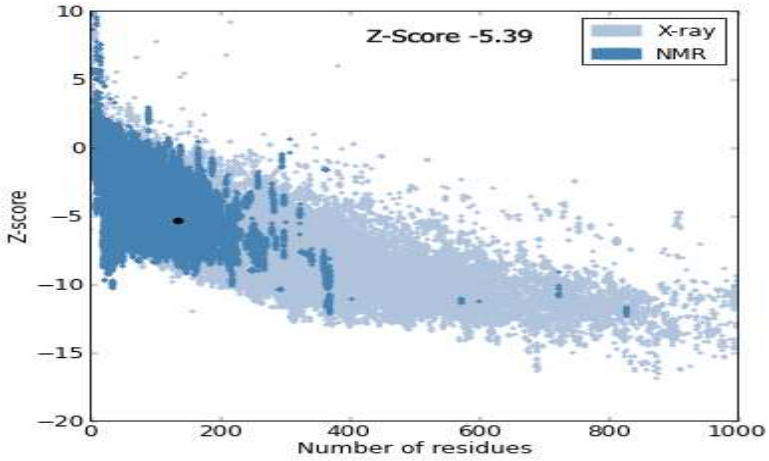


Fig. 4 Prosa-Plot for the homology model of StPR1 displaying Z-score value. Evaluation of the StPR1 using ProSa web server (https://prosa.services.came.sbg.ac.at/prosa.php) displaying Z-Score value indicating nearness of the model structure with native structure.

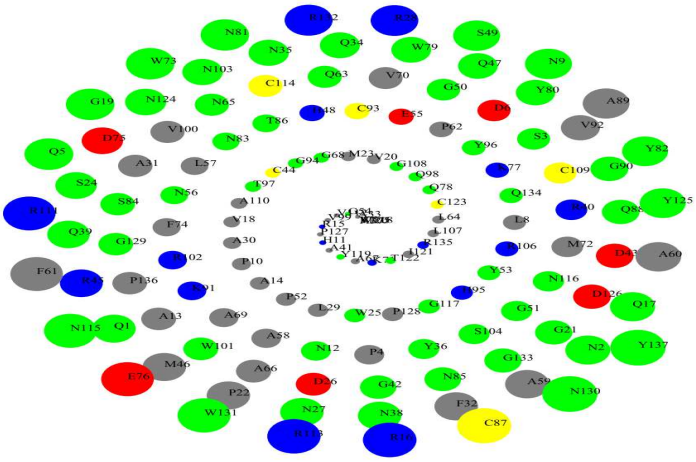


Fig. 5 Solvent accessibility plot of the StPR1 protein.

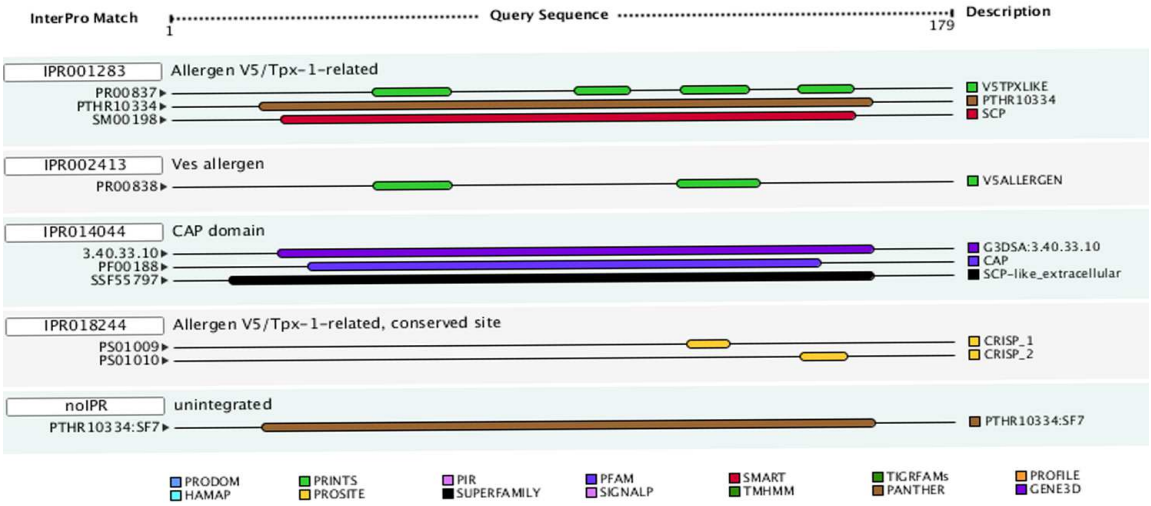


Fig. 6 InterProScan output for *Solanum tuberosum* (Accession number: M1A2A4).

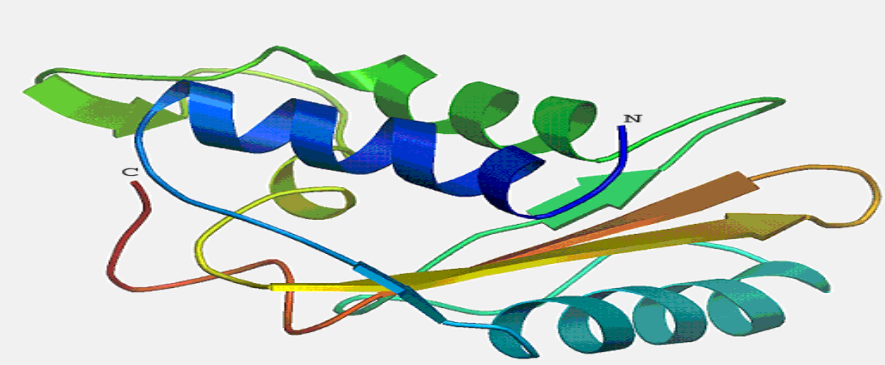


Fig. 7 Overall topology of the theoretical structure of StPR1 showing four α-helices and β-strands along with the N-terminal and C-terminal of the protein. The figure is prepared with pymol.

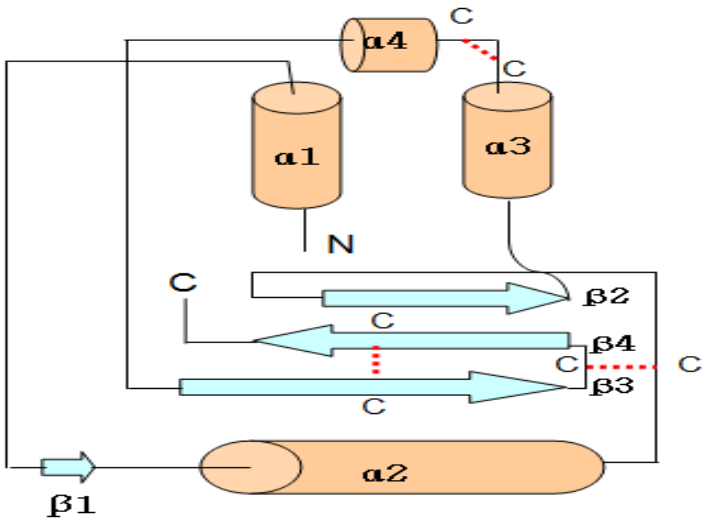


Fig. 8 Systematic representation of secondary structure of StFT showing central four antiparallel β sheets (β1, β2, β3, β4) shown in cyan and four helices (α1, α2, α3, α4) along with the N-terminal and C-terminal of the protein. The red dotted lines shows the three conserved disulfide bonds.

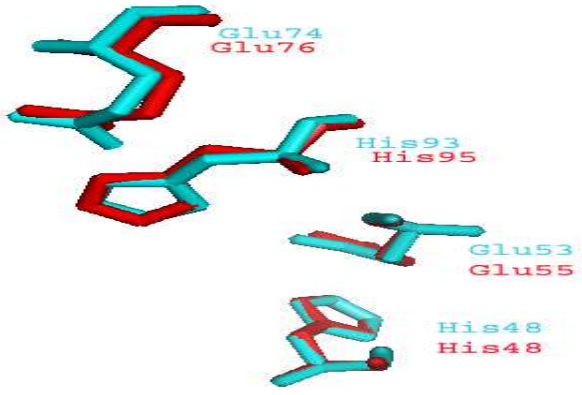


Fig. 9 A superposition of the four conserved residues of StPR1 (red) and template P14A (cyan). Three letter amino acid abbreviations are used with position numbers.

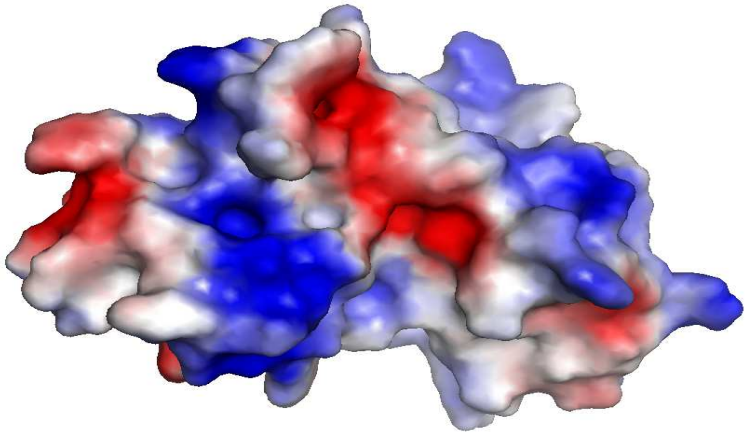


Fig. 10 Accessible surface map representation of electrostatic potential distribution of StPR1. Positive and negative electrostatic potentials are colored blue and red respectively.

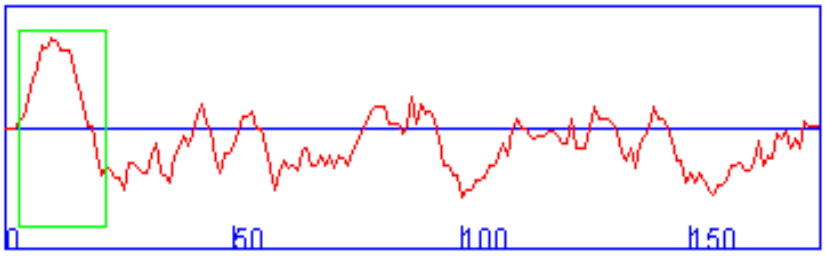


Fig. 11 Hydropathy profile of StPR1 along the length of the sequence. The green box shows the transmembrane region.

Table 1: Analysis of the StPR1 and template 1CEF of CRISP proteins in correlation with the amino acid sequence of their signature regions. The amino acid differences are underlined

	StPR1	Template (1CEF)	No. of different amino acid compared
CRISP1	GHYTQVVWRNS	RHYTQVVWRNS	1/11
CRISP2	FITCNYDPPGNW	FISCNYDPVGNW.	2/12

Table 2: Various physicochemical properties of StFT protein sequence computed by ProtParam.

Name	Source	Mol.wt	Theoretical pI	-R	+R	Instability index	Aliphatic index	GRAVY
M1A2A4	Potato	20223.6	8.48	12	15	32.36	60.61	-0.525

Table 3: Transmembrane region of StFT protein sequence computed by SOSUI.

Uniprot Id	N-terminal position	Transmembrane region	C terminal position	Type of helix	Length
StPR1	3	YSNIALIICFLIFAIFHSSQ	22	Primary	20