

This is the author's final, peer-reviewed manuscript as accepted for publication. The publisher-formatted version may be available through the publisher's web site or your institution's library.

## **Serine protease identification (in vitro) and molecular structure predictions (in silico) from a phytopathogenic fungus, *Alternaria solani***

Murugesan Chandrasekaran, Raman Chandrasekar, Tong-Min Sa, and Muthukrishnan Sathiyabama

### **How to cite this manuscript**

If you make reference to this version of the manuscript, use the following information:

Chandrasekaran, M., Chandrasekar, R., Sa, T., & Sathiyabama, M. (2014). Serine protease identification (in vitro) and molecular structure predictions (in silico) from a phytopathogenic fungus, *Alternaria solani*. Retrieved from <http://krex.ksu.edu>

### **Published Version Information**

**Citation:** Chandrasekaran, M., Chandrasekar, R., Sa, T., & Sathiyabama, M. (2014). Serine protease identification (in vitro) and molecular structure predictions (in silico) from a phytopathogenic fungus, *Alternaria solani*. *Journal of Basic Microbiology*, 54(S1), S210-S218.

**Copyright:** © 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

**Digital Object Identifier (DOI):** doi:10.1002/jobm.201300433

**Publisher's Link:** <http://onlinelibrary.wiley.com/doi/10.1002/jobm.201300433/full>

This item was retrieved from the K-State Research Exchange (K-REx), the institutional repository of Kansas State University. K-REx is available at <http://krex.ksu.edu>

**Title:** Serine protease identification (*in vitro*) and molecular structure predictions (*in silico*) from a phytopathogenic fungus, *Alternaria solani*

**Running head:** An *in vitro* and an *in silico* study of *Alternaria solani* Protease

**Murugesan Chandrasekaran<sup>1, a</sup>, Raman Chandrasekar<sup>2</sup>, Tong-Min Sa<sup>a</sup>, and Muthukrishnan Sathiyabama<sup>1, \*</sup>**

<sup>1</sup>Department of Plant Science, Bharathidasan University, Tiruchirappalli-620 024, Tamil Nadu, India.

<sup>a</sup>Department of Environmental and Biological Chemistry, Chungbuk National University, Cheongju, Chungbuk-361 763, Republic of Korea.

<sup>2</sup>Department of Biochemistry and Molecular Biophysics, 238 Burt Hall, Kansas State University,, Manhattan 66506, KS, USA.

Full name and full address of the \*Corresponding author:

Muthukrishnan Sathiyabama,

Assistant Professor,

Department of Plant Science,

Bharathidasan University,

Tiruchirappalli-620 024, Tamilnadu, India.

e-mail: sathiyabamam@yahoo.com

Tel: +91 431 2407061

Fax: +91 431 2407045

## Abstract

Serine proteases generally share a relatively high degree of sequence identity and play a major role in the diversity of biological processes. Here we focus on three-dimensional molecular architecture of serine proteases from *Alernaria solani*. The difference in flexibility of active binding pockets and electrostatic surface potential distribution of serine proteases in comparison with other fungal species is reported in this study. In this study we have purified a serine protease from the early blight pathogen, *Alernaria solani*. MALDI-TOF-MS/MS analysis revealed that protease produced by *A. solani* belongs to alkaline serine proteases. AsP is made up of 403 amino acid residues with molecular weight of 42.1kDa (Isoelectric point (pI)-6.51) and molecular formula  $C_{1859}H_{2930}N_{516}O_{595}S_4$ . The follow-up research on the molecular structure prediction is used for assessing the quality of *A. solani* Protease (AsP). The AsP protein structure model was built based on its comparative homology with serine protease using the program, MODELER. AsP had 16  $\beta$ -sheets and 10  $\alpha$ -helices, with Ser<sup>350</sup> (G347–G357), Asp<sup>158</sup> (D158–H169) and His<sup>193</sup> (H193–G203) in separate turn/coil structures. Biological metal binding region situated near the 6<sup>th</sup>-helix and His<sup>193</sup> residue is responsible for metal binding site. In addition, the calcium ion is coordinated by the carboxyl groups of Lys<sup>84</sup>, Ile<sup>85</sup>, Lys<sup>86</sup>, Asp<sup>87</sup>, Phe<sup>88</sup>, Ala<sup>89</sup>, Ala<sup>90</sup> (K84-A90) for first calcium (Ca<sup>2+</sup>) binding site and carbonyl oxygen atom of Lys<sup>244</sup>, Gly<sup>245</sup>, Arg<sup>246</sup>, Thr<sup>247</sup>, Lys<sup>248</sup>, Lys<sup>249</sup>, and Ala<sup>250</sup> (K244–A250), for second Ca<sup>2+</sup> binding site. Moreover, Ramachandran plot analysis of protein residues falling into most favored secondary structures were determined (83.3%). The predicted molecular 3D structural model was further verified using PROCHECK, ERRAT and VADAR servers to confirm the geometry and stereo-chemical parameters of the molecular structural design. The functional analysis of AsP 3D

molecular structure predictions familiar in the current study may provide a new perspective in the understanding and identification of antifungal protease inhibitor designing.

**Key words:** *Alternaria solani*, Serine Protease, MODELER, PROCHECK, ERRAT, MALDI-TOF-MS/MS, Ramachandran Plot, 3D Molecular Structural design.

## 1. Introduction

*Alternaria solani* (Ellis and Martin) Sorauer, is an important plant pathogenic fungi causing early blight disease in tomato (*Lycopersicon esculentum* Mill.) [1, 2]. A number of fungal mechanisms and molecules have been shown to contribute to fungal pathogenicity and/or virulence, including cell wall degrading proteins [3], inhibitory proteins [4], and enzymes involved in the toxin synthesis [5]. Many phytopathogenic fungi release an array of cell wall-degrading, hydrolytic enzymes to fragment the plant cell wall polymers, including proteases and glycanases (e.g. endo-polygalacturonases, cellulases, pectin lyase, xylanase,  $\beta$ -galactosidase, galacturonases, xylanases, and glucanases), thus facilitating the colonization of the host cells [6, 7]. *A. solani* uses an infection strategy based on the secretion of numerous cell-wall degrading enzymes [8]. Remarkable progress has been made over the past decade on the contribution of proteolytic enzymes in elucidating plant-pathogen interaction mechanism [9, 10, 11, 12]. Proteases of various phytopathogenic fungi have been detected in infected host plant tissues, but their precise function has not been elucidated yet.

Extracellular proteolytic enzymes of the phytopathogenic fungi are represented to a large degree by serine peptidases. Many distinct families of serine proteases exist; they have been

grouped into six clans, of which the two largest are the chymotrypsin-like and subtilisin-like clans [13]. Subtilisin-like serine proteases play an important role in the pathogenicity of pathogenic fungi. By using subtilisin-like serine proteases, pathogenic fungi disrupt the physiological integrity of the hosts during penetration and colonization [14, 15]. Previous studies have suggested that pathogenic fungi with different life styles utilize subtilisin-like serine proteases as their virulence factor [15, 16]. Serine proteases of subtilisin (S8) family produced by oil seed rape pathogen *Pyrenopeziza brassicae* have been proposed to play a role in host-pathogen interactions [17]. The strong induction of different serine protease inhibitors in a blast-resistant rice mutant during blast infection suggests that these proteins play an important role in blast resistance [18]. Collectively, these studies suggest the existence of a correlation between subtilisin-like protease and pathogenicity of the fungus.

In this study, we examined protease of phytopathogenic fungi *A. solani*, it was identified as subtilisin-like protease through MALDI-TOF-MS/MS fragmentation, designated as *A. solani* Protease (AsP). In addition, we undertook an analysis of AsP sequences with an available serine protease sequence to predict 3D molecular structural design of AsP by homology modeling. The functional analysis of AsP 3D molecular structural predictions may provide a new perspective in the understanding and identification of antifungal protease inhibitor designing.

## **2. Materials and methods**

### **2.1. Identification of serine protease**

An alkaline extracellular serine protease of *Alternaria solani* isolate 4632 ITCC was purified to homogeneity and subjected to MALDI-TOF-MS/MS analysis. Mono-isotopic peptide masses were assigned and used in the database search. The protein identification was

accomplished utilizing the MASCOT database search engine (Matrix Science, London, UK, <http://www.matrixscience.com>) scores >63 were considered to be significant ( $P < 0.05$ ) in the MASCOT search. Hence, the protein identified scores less than the significance level were repeated as unidentified.

## **2.2. Target protein sequence and template selection**

The template protein sequence of AsP from *A. solani* was obtained from M. Chandrasekaran and M. Sathiyabama [11]. It was ascertained that the three-dimensional (3D) structure of the protein was not available in Protein Data Bank (PDB) (<http://www.rcsb.org/pdb>). The NCBI-BLAST was used to identify the template for modeling the three dimensional structure of *Thermus aquaticus* (4DZT). The result of NCBI-BLAST against the PDB database was used for selection of a suitable template for 3D modeling of the target protein.

## **2.3. Sequence alignment and Phylogenetic tree analysis**

AsP amino acid sequences were used for alignment with template protein using PSI-BLAST (<http://blast.ncbi.nlm.nih.gov/Blast>). An *in silico* study, mainly comparative homology modeling, of the target sequence AsP can be helpful to investigate sequential-structural-functional relationship. The 3D structure of AsP was predicted based on available homologous template structure in Protein structure Data Bank (PDB) resources. Template selection was performed using PDB advanced BLAST (<http://www.rcsb.org>). Retrieved template structure was used for comparative homology modeling of AsP. Also, polypeptide hydropathy/amphipathicity was evaluated using the Kyte–Doolittle algorithm as implemented in the GCG computer program (Genetics Computer Group, Wisconsin Package version 8.1, Wisconsin). The GCG program was run on a Macintosh computer with eXodus5.2 software (White Pine Software, Inc.). The

hydropathy/amphipathicity plot obtained from the GCG program was saved as text form and redrawn using Kaleida-Graph software.

Several programs exist for making phylogenetic trees that display the relationship between sequences to calculate all possible tree topologies to find the one that fits best with the sequence data. We used Clustal X to generate alignments of the sequences. Nucleotide/protein sequences were subsequently aligned manually in order to increase alignment (Table 1). A bootstrapped, unrooted Neighbor-Joining tree was generated (pairwise) using the same program. After getting the root distance format using the Clustal W, the code was submitted into the phylo-draw software version 8.2 in NJ plot.

#### **2.4. Homology modeling and structure refinement**

The 3D structure of AsP has been predicted using DS-MODELLER (<http://salilab.org/modeller>) and PHYRE2 (<http://www.sbg.bio.ic.ac.uk/>). From the homology modeling searching, two templates were selected: high-resolution X-ray crystallography structure of the *T. aquaticus* 4DZT [19, 20] and 3F70 [21]. Out of the two models, 4DZT was found to be the best model according to the scoring of PROCHECK, totally non-local energy of the protein ( $E/kT$  units) and overall model quality Z-score. Loop refinement and structural simulation were done using LOOPER and CHARM force field, respectively. Finally, predicated 3D model was subjected to a series of tests for testing its internal consistency and reliability. The quality of the model was checked; VERIFIED 3D [22], PROFILE 3D [23], and ERRAT [24] and the stereo-chemical properties based on backbone conformation were evaluated by inspection of Psi/Phi/Chi/Omega angle using a Ramachandran plot of mol-Probity (<http://molprobity.biochem.duke.edu/>). Further, quantitative analysis was done using accessible surface area prediction using Volume Area Dihedral Angle Reporter (Suppl. Table S2, S3, S4)

by using VADAR (<http://vadar.wishartlab.com/>) online server [25]. Successfully modeled, verified and the most reliable structure of AsP was deposited in PMDB Acc. No. PM0078786 (<http://mi.caspur.it/PMDB/>).

## **2.5. Active binding site and metal binding site detection**

After complete modeling, simulation and refinement of the structure of AsP prediction of the possible metal binding site ~~was~~ were performed using the Q-site Finder (<http://bmbpcu36.leeds.ac.uk/qsitefinder/>). LIGPLOT and FunFOLD is a program (<http://www.reading.ac.uk/bioinf/IntFOLD/>) used to plot schematic diagram of protein-ligand interactions for a given in a PDB file. In this study, this software was used to create a 3D diagram of calcium binding site and the coordinate residues. PyMOL software used to illustrate the binding site of the target protein. These binding sites were further compared to the active sites of the template.

## **3. Results and Discussion**

### **3.1. Identification of serine protease**

Subtilisin-like serine proteases play an important role in the pathogenicity of pathogenic fungi [17, 18]. With the help of these enzymes, pathogenic fungi disrupt the physiological integrity of the hosts during penetration and colonization [14, 15, 16]. The 42kDa protein was excised and in-gel tryptic digestion was carried out and gathered peptide mass finger-printing analysis (PMF). Pooled mono-isotopic masses of 42kDa analyzed using MASCOT software showed that the protein is of alkaline serine protease having the first score of 191. MALDI/TOF/MS data also confirmed AsP is homologous to the alkaline protease of *A. fumigatus* [26, 27]. The PMF analysis of alkaline protease showed fragmentation of nearly 22

peptides [11] of them only one peptide 1485 m/z yielded good fragmentation (Suppl. Table S1). The precursor ion at 1485 m/z was selected by the first TOF and allowed to react with the floating cell with inert gas; the MS/MS reported spectrum was obtained after ionization by second TOF operated in the reflectron mode. Sequence coverage was 55% and identity was 100% (Suppl. Fig. S1). Protein motif identification showed three hits in extracellular alkaline proteases. If a protein includes at least two of the three active site signatures, the probability of it being a serine protease from the subtilase family is 100%. Collectively, the data generated indicate that the purified 42 kDa protein of *A. solani* is a serine protease belonging to the subtilisin family.

### **3.2. Homology modeling**

The success and applicability of homology modeling are steadily increasing due to growing number and availability of experimentally determined protein structures [28]. The subtilisin family (S8 and S53) appears to play many roles in fungal biology. We isolated and identified serine protease from a phytopathogenic fungus, *A. solani* (AsP) and subjected to molecular structural analysis [11]. The AsP enzyme has 403 amino acids in length with a molecular weight of  $\approx$  42.1 kDa. It is >45 % of Ala, Gly, Ser, rich amino acid with theoretical pI 6.51. Out of 403 amino acid residues, serine was found to have the highest number of 39 residues (9.7%), followed by Glycine, and Threonine residues (35, 8.4%) and the amino acids such as Val, Ile, Leu, Asp, Lys with >25 residues (7.5%). The molecular formula of AsP was found to be  $C_{1859}H_{2930}N_{516}O_{595}S_4$  with a total number of 5904 atoms. The instability index (28.37) was computed which classified the protein as stable [29].

The X-ray crystal structure of *T. aquaticus* (4DZT) were specifically selected on the basis of NCBI-BLAST results and was utilized as a template for structure modeling of AsP. The sequence of the target and the template are brought into an optimal alignment. Then, the structure of the target protein is constructed by exploiting the information from the template structure. The modeling steps are: backbone superposition of the atom, loop modeling and orientation of the side chains. The target backbone from N-terminus to C-terminus is built by averaging the backbone atom position of the template structures. The matched structures were superposed with respect to a selected set of AsP enzyme C- $\alpha$  atoms (54% superposition), with the structure the structure (4DZT) template having the best score of 3.5 (RMSD value were between 0.331 and 0.524 Å) which helped to identify common segments corresponding to structurally conserved regions; Fig. 1A). The regions surrounding the putative catalytic residues or homology boxes (boxes 1-3 in Fig. 1C), along with some of the  $\beta$ -strands, particularly (box 2, 3), which is involved in substrate interactions, show a high level of identity. The corresponding sequence of AsP modeled in this paper and secondary structure elements ( $\alpha$ -helices and  $\beta$ -strands) are represented schematically and labeled accordingly to predicated model 3D molecular structure (Fig. 1B).

The 3D predicated model was analyzed using energy minimization, refinement and simulation program of PROCHECK. The PDB Sum server was employed for evaluation by comparing the geometry and stereo chemical parameters ( $r_0$ ,  $\theta$  and  $\Phi$ ) quality of predicted models. A large number of literature related homology modeling were also found to use the PROCHECK for screening the best model [30, 31]. Further Ramachandran plot used to evaluate the values of the dihedral angels agree with the values of allowed conformation for protein backbones (**Fig. 2A**). Ramachandran plot analysis showed 83.3% of amino acid residues within the most favored

and 15% residues in additional and generously allowed regions, whereas 4 residues were found in disallowed regions (Fig. 2A). The comparable Ramachandran plot characteristic and G-factor score confirmed the good quality of the present predicated model. Hydropathy plot is a quantitative analysis of the degree of hydrophobicity or hydrophilicity of amino acids in a protein (Fig. 2B). Those inter  $\alpha$ -helical contacts are mainly from hydrophobic residues such as Val, Ile, Leu, Asp, Lys, Phe, and Ala. These hydrophobic interactions are the major force contributing to stabilization of the helix bundle structure. Hence, it is envisaged that a common mechanism may underlie in the folding and function of AsP and prefer hydrophobic amino acids but exhibits broad substrate specificity [29]. Furthermore, Ji *et al.* [32] state that the hydropathic character of the sequence residue has a larger effect on the sequence's choice for  $\alpha$ -helix or  $\beta$ -sheet, as compared to the intrinsic propensities of the amino acids for a particular secondary structure.

However, further qualitative and quantitative analysis of the predicted AsP model by using VERIFY 3D details lie between 0.01-0.82 representing the best verified and reliable model of AsP. The Overall quality factor was calculated using ERRAT server (Suppl. Table S2, S3, S4) and the model structure were found to have a 95% quality factor. VADAR analysis revealed accessible surface area, excluded volume, backbone and side chain dihedral angles, secondary structure, hydrogen bonding partners, hydrogen bond energies, steric quality, solvation free energy as well as local and overall fold quality yielded good results. Using atomic radii from Sharke method, we observed 23% residues were involved in the formation of  $\alpha$ -helices, 30% in  $\beta$ -sheets, 45% in coils and 20% residues formed turns. The observed mean hydrogen bond (H-bond) distance and energy value were closely similar with expected value in H-bond statistics.

The obtained expected residues with H-bond were 75% and we observed 77% for the predicted model. Dihedral angle statistics also represented approximately similar score with that of the expected values (Suppl. Table S2, S3, and S4). It was found that the overall quality and quantity on the basis of secondary elements of the predicted AsP model was good and reliable. The generated AsP model was successfully deposited in PMDB (<http://www.caspur.it/PMDB>) bearing Model ID: **PM0078786**. The resulting structure of AsP shows a homo-dimer with eight stranded beta sheets surrounded by seven  $\alpha$ -helices and eight-stranded beta-sheets surrounded by three alpha-helices (Fig. 1B).

Phylogenetic analysis of 25 serine protease using MEGA software with NJ method, as described in Materials and Methods. The results reveals that, ascomycete fungi species are phylogenetically (Fig. 3) more similar proteases also have a more similar distribution of major secondary structure elements were well matched but the some turns and loops, especially in human and plant located on the protein surface, showed large structural deviations [33, 34, 35]. This is not surprising, because the secondary structure elements are better conserved and have fewer amino acid insertions and deletions than the loop/turn regions. Comparison of the entire deduced peptide sequence with other serine protease revealed an identity of 80%, 70%, 67%, 59%, 56%, 48% respectively to that of *Aspergillus niger*, *Podospora anserina*, *Veriticillium albo-atrum*, *Trichoderma hamatum*, *Magnaporthe grisea*, *Neurospora crassa* and only 57% and 15% of that of *Bacillus cereus* and *Saccharomyces cerevisiae* (Table 1). Notably, *Arabidopsis* protease shares a relatively high degree of sequence identity (67%), whereas human low percentage of identity (29%). By incorporating an evolutionarily diverse range of serine proteases, our analysis indicates that although the core structures deviated considerably during evolution, the relative positions of the catalytic triad C $\alpha$  atoms maintained very close relative

distances and were potent stabilized by other highly conserved residues (Fig. 3). The phylogenetic tree presented here agrees with a previous attempt to segregate proteases into functional groups based on phylogenetic analysis [34, 36]. However, the subtle differences in amino acid components of the substrate-binding regions and other regions, such as loops and turns, can result in differences in the architecture of the substrate-binding pockets and other structural regions, leading to changes in their physicochemical properties, such as optimal reaction temperature, thermal stability, and electrostatic surface potential of these proteases.

### **3.3. Active site for metal binding region**

Active binding site identification of 3D predicated model AsP with PMDB ID: PM0078786 was done using Q-site finder. In AsP, the metal binding site is situated near the top of a  $(\beta\alpha)_6$  barrel and is assembled by Histidine (His<sup>193</sup>) residues perched at the C-terminal of inner barrel strands (Fig. 4A). Based on predicated binding site and the work conducted by others, we suggested that AsP may have a metal cofactor for binding and/or activating the AsP enzyme. Protein alignments of AsP and 4DZT confirmed the catalytic site of His<sup>193</sup> residues. These were found to be prominent active binding sites for metal and protein-cofactors interactions (Fig. 4B). S1-4 substrate-binding pockets are large and have been considered as the primary determinants for substrate specificity and affinity [35, 37]. This leads to the speculation that alkaline proteases have higher substrate affinity and catalytic activity. These distinctive features, which have also been observed in *Plasmodium falciparum* protease; *Neurospora crassa* [38] and, may be associated with increased stabilization and hyperthermophilic adaptation. Closely packed aromatic interactions have been proposed to increase the  $\Delta G$  of unfolding, thereby increasing thermal stability [34, 37].

### 3.4. Catalytic domain and calcium binding region

The superposed structure has a common core structure with large variation in the loop outside the core region. Protein motif identification confirmed sequence conservation of the catalytic triad residue at aspartic acid (AYVV**DSG**INVNH), serine (**G**TSMATPHIVG), and histidine (**H**GTHVAGTIGG), residues can be used as a signature/motif specific to proteases and each playing an essential role in the cleaving process of the peptide bond. i) the role of Aspartic acid is believed to bring the His into the correct orientation by its carboxyl group hydrogen bonding with the His, thus facilitating the nucleophilic attack; ii) the Serine uses its hydroxyl group (-OH) as the primary nucleophile, which is able to attack the carbonyl carbon (C=O) of the scissile peptide bond of the substrate; iii) the Histidine plays a dual role as the proton donor and acceptor at different steps in the reaction, i.e., a pair of electrons on the His nitrogen has the ability to accept the hydrogen from the serine hydroxyl group, thus coordinating the attack of the peptide bond. Among the sequences analyzed, the highly conserved amino acid Gly<sup>163</sup>, and Gly<sup>203</sup> had the occupancy percentage of 50% and 60%, respectively, which have been previously described [33]. The surface electrostatic potentials around the catalytic site were very similar to template model (Fig. 4C, D), with large patches of electropositive and electro neutral regions around the catalytic site of this modeled protease suggests it favors a negatively charged substrate. The largely electro-neutral regions possibly relax the stringency of the substrate binding, allowing for a number of different protein substrates. Likewise, the largely electronegative catalytic site of the *N. crassa* protease suggests it favors a positively charged substrate [38]. AsP had a higher proportion of polar residues (55%) and basic amino acid (15%), which indicates it, could favor a more hydrophilic environment. Many proteases in the peptide S8 family contain one or more Ca<sup>2+</sup>-binding sites and binding of calcium cations enhances the

thermal stability of the proteases which in turn increase their resistance against proteolysis, either by itself or by other proteases [37]. The binding site residue prediction from FunFOLD 1.0 online server (<http://www.reading.ac.uk/bioinf/IntFOLD/>) was shown in Fig. 4E, with the binding site residues were K84-A90 (Lys<sup>84</sup>, Ile<sup>85</sup>, Lys<sup>86</sup>, Asp<sup>87</sup>, Phe<sup>88</sup>, Ala<sup>89</sup>, Ala<sup>90</sup>), K224-A250 (Lys<sup>244</sup>, Gly<sup>245</sup>, Arg<sup>246</sup>, Thr<sup>247</sup>, Lys<sup>248</sup>, Lys<sup>249</sup>, and Ala<sup>250</sup>) situated in the helix-7 and helix-6 region and the predicted ligand calcium molecules highlighted in blue. The FunFOLD 1.0 methods accurately predicted all of the binding site residues (Fig. 1C and Fig. 4E), but also over-predicted three residues (6, 7 residues). The binding site residue prediction had an MCC score = 0.9012 and BDT score = 0.7744, with the protein predicted to bind to a centroid ligand being calcium.

### **Concluding remarks**

We presented the 3D molecular structural design of serine protease from *Alternaria solani* (AsP) by homology modeling method. AsP is made up of 403 amino acid residues with a molecular weight of 42.1kDa (pI 6.51) and its molecular formula was C<sub>1859</sub>H<sub>2930</sub>N<sub>516</sub>O<sub>595</sub>S<sub>4</sub>. At the core, the modeled structure is a  $\alpha/\beta$  hydrolase fold, which typically consists of 16-stranded  $\beta$ -sheet flanked by ten  $\alpha$ -helices inter-connected with loop region. Biological metal binding region is situated near the 6<sup>th</sup>-helix and His<sup>193</sup> residue is responsible for metal binding site. The substrate specificity of the alkaline serine protease was controlled by the structural arrangement of the two mobile loop regions. Also, calcium binding regions usually have five coordinates with distorted tetrahedral geometry. The calcium ion in the first Ca<sup>2+</sup> binding site (K84-A90) is coordinated by the carboxyl groups of Lys<sup>84</sup>, Ile<sup>85</sup>, Lys<sup>86</sup>, Asp<sup>87</sup>, Phe<sup>88</sup>, Ala<sup>89</sup>, Ala<sup>90</sup> and the calcium ion in the second Ca<sup>2+</sup> site (K224-A250) is coordinated by the carbonyl oxygen atom of Lys<sup>244</sup>, Gly<sup>245</sup>,

Arg<sup>246</sup>, Thr<sup>247</sup>, Lys<sup>248</sup>, Lys<sup>249</sup>, and Ala<sup>250</sup>. The AsP model had sixteen  $\beta$ -sheets and ten  $\alpha$ -helices, with Ser<sup>350</sup> (G347-G357), Asp<sup>158</sup> (A<sup>158</sup>-H<sup>169</sup>) and His<sup>193</sup> (H193-G203) in separate turn/coil structures. These results might further allow studying substrate specificity and other properties contributing as a potential benefit for functional analysis of this protease enzyme and provide a new perspective in the understanding and identification of antifungal protease inhibitor designing.

### **Conflict of Interest**

All authors declare that there are no conflicts of interest.

### **References**

- [1] Datar, V.V., Mayee, C.D., 1981. Assessment of loss in tomato yield due to early blight. *Indian Phytopathology*, **34**, 191–195.
- [2] Kumar, V., Haldar, S., Pandey, K.K., Singh, A.K., Singh, P.C., 2008. Cultural, morphological, pathogenic and molecular variability amongst tomato isolates of *Alternaria solani* in India. *World J. Microbiol Biotechnol.*, **24**, 1003–1009.
- [3] Kikot, G.E., Hours, R.A., Alconada, T.M., 2009. Contribution of cell wall degrading enzymes to pathogenesis of *Fusarium graminearum*: A review. *J. Basic Microbiol.*, **49(3)**, 231–241.
- [4] van Esse, H.P., Van't Klooster, J.W., Bolton, M.D., Yadeta, K.A., et al. 2008. The *Cladosporium fulvum* virulence protein Avr2 inhibits host proteases required for basal defense. *Plant Cell*, **20(7)**, 1948–1963.

- [5] Lawrence, C.B., Mitchell, T.K., Craven, K.D., Cho, Y., et al. 2008. At death's door: *Alternaria* pathogenicity mechanisms. *Plant Pathol. J.*, **24 (2)**, 101–111.
- [6] de Vries, R.P., Visser, J., 2001. *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microbiology and Molecular Biology Reviews*, **65(4)**, 497–522.
- [7] ten Have, A., Espino, J.J., Ester Dekkers, E., Steven, C., et al. 2010. *Botrytis cinerea* aspartic proteinase family. *Fungal Genetics and Biology*, **47**, 53–65.
- [8] Eshel, D., Miyara, I., Ailinnig, T., Dinoor, A., Prusky, D., 2002. pH regulates endoglucanase expression and virulence of *Alternaria alternata* in persimmon fruits. *Mol. Plant-Microbe Interact.*, **15**, 774–779.
- [9] Murphy, J.M., Walton, J.D., 1996. Three extracellular proteases from *Cochliobolus carbonum*: cloning and targeted disruption of ALP1. *Mol. Plant Microbe Interact.*, **9**, 290–297.
- [10] Paris, R., Lamattina, L., 1999. *Phytophthora infestans* secretes extracellular proteases with necrosis inducing activity on potato. *Euro. J. Plant Pathol.*, **105**, 753–760.
- [11] Chandrasekaran, M., Sathiyabama, M., 2013. Production, partial purification and characterization of protease from a phytopathogenic fungi *Alternaria solani* (Ell. and Mart.) Sorauer. *Journal of Basic Microbiology*, *In press*.
- [12] Dubovenko, A.G., Dunaevsky, Y.A., Belozersky, M.A., Oppert, B., 2010. Trypsin-like proteins of the fungi as possible markers of pathogenicity. *Fungal biology*, **114**, 151–159.

- [13] Rawlings, N.D., Morton, F.R., Kok, C.Y., Kong, J., Barrett, A.J., 2008. MEROPS: the peptidase database. *Nucl. Acids Res.*, **36**, D320–D325.
- [14] Huang, X.W., Zhao, N.H., Zhang, K.Q. 2004. Extracellular enzymes serving as virulence factors in nematophagous fungi involved in infection of the host. *Res. Microbiol.* **155**, 811–816.
- [15] St. Leger, R.J., Joshi, L., Roberts, D.W., 1997. Adaptation of proteases and carbohydrases of saprophytic, phytopathogenic and entomopathogenic fungi to the requirements of their ecological niches. *Microbiology*, **143**, 1983–1992.
- [16] Sreedhar, L., Kobayashi, D.Y., Bunting, T.E., Hillman, B.I., Belanger, F.C., 1999. Fungal proteinase expression in the interaction of the plant pathogen *Magnaporthe poae* with its host. *Gene*, **235**, 121–129.
- [17] Batish, S., Hunter, A., Ashby, A.M., Johnstone, K., 2003. Purification and biochemical characterisation of Psp1, an extracellular protease produced by the oilseed rape pathogen *Pyrenopeziza brassicae*. *Physiol. Mol. Plant Pathol.*, **62**, 13–20.
- [18] Han, C.U., Lee, C.H., Jang, K.S., Choi, G.J. et al. Identification of rice genes induced in a rice blast-resistant mutant. *Mol. Cells.* **17**, 462–468.
- [19] Green, P.R., Oliver, J.D., Strickland, L.C., Toerner, D.R., et al. 1993. Purification, crystallization and preliminary X-ray investigation of aqualysin I, a heat-stable serine protease. *Acta Crystallogr. Sect. D*, **49**, 349.

- [20] Barnett, B.L., Green, P.R., Strickland, L.C., Oliver, J.D., 2012. Aqualysin 1: the crystal structure of a serine protease from an extreme thermophile, *Thermus quaticus* YT-1, NCBI.
- [21] Guo, Y. et al. 2009. Methylation-state-specific recognition of histones by the MBT repeats protein L3MBTL2. *Nucl. Acids Res.* **37(7)**, 2204–2210.
- [22] Eisenberg, D., Lüthy, R., Bowie, J.U., 1997. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.*, **277**, 396–404.
- [23] Suyama, M., Matsuo, Y., Nishikawa, K., 1997. Comparison of protein structures using 3D profile alignment. *J. Mol. Evol.*, **44**, S163–73.
- [24] Gundampati, R.K., Chikati, R., Kumari, M., Sharma, A., et al. 2012. Molecular docking and dynamics simulations of *A. niger* RNase from *A. niger* ATCC26550: for potential prevention of human cancer. *J. Mol. Model.*, **18(2)**, 653–62.
- [25] Willard, L., Ranjan, A., Zhang, H., Monzavi, H., et al. 2003. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucl. Acids Res.*, **31(13)**, 3316-3319.
- [26] Monod, M.S., Capoccia, B., Lechenne, C., Zaugg, M., et al. 2002. Secreted proteases from pathogenic fungi. *Int. J. Med. Microbiol.*, 292, 405–419.
- [27] Anandan, D., Marmer, W.N., Dudley, R.L., 2007. Isolation, characterization of culture parameters for production of an alkaline protease isolated from *Aspergillus tamarii*. *J. Ind. Microbiol. Biotechnol.*, **34**, 339–347.

- [28] Wiltgen, M., Tilz, G.P., 2009. Homology modelling: a review about the method on hand of the diabetic antigen GAD 65 structure prediction. *Wien Med Wochenschr.*, **159**, 112–25.
- [29] Morya, V., Yadav, S., Kim, E-K., Yadav, D., 2012. *In silico* characterization of alkaline proteases from different species of *Aspergillus*. *Appl. Biochem. Biotechnol.*, **166**, 243–257.
- [30] Kherraz, K., Kherraz, K., Kameli, A., 2012. Homology modeling of ferredoxin-nitrite reductase from *Arabidopsis thaliana*. *Bioinformation*, **6 (3)**, 115–119.
- [31] Sharma, P., Bhattacharya, A. 2012. 3Dstructure prediction of aquaporin-2, virtual screening and *in silico* docking studies of gold and silver derivatives, used as potent inhibitors. *Int. J. Pharam Pharm Sci.*, **4**, 574–579.
- [32] Ji, X., Wang, W., Zheng, Y., Hao, J., Sun, M. 2012. Homology modeling and molecular dynamics simulation studies of a marine alkaline protease. *Bioinform Biol. Insights*, **6**, 255–263.
- [33] Krem, M.M., Cera, D.E. 2001. Molecular markers of serine protease evolution. *EMBO J.*, **20**, 3036–3045.
- [34] Liang, L., Liu,S., Yang, J., Meng, Z. et al. (2011). Comparison of homology models and crystal structures of cuticle-degrading proteases from nematophagous fungi: structural basis of nematicidal activity. *FASEB J.* **25**, 000–000.
- [35] Yang, J.K., Tian, B.Y., Liang, L.M., Zhang, K.Q. (2007). Extracellular enzymes and the pathogenesis of nematophagous fungi. *Appl. Microbiol. Biotechnol.*, **75**, 21–31.

- [36] Yousef, G.M., Kopolovic, A.D., Elliott, M.B., Eleftherios, P. et al. (2003) Genomic overview of serine proteases. *Biochemical and Biophysical Research Communications*. **305**, 28–36.
- [37] Liu, S.Q., Meng, Z.H., Fu, Y.X., Zhang, K.Q. (2010). Insights derived from molecular dynamics simulation into the molecular motions of serine protease proteinase K. *J. Mol Model.*, **16**, 17–28.
- [38] Aparna Laskar, Rodger, E.J., Aniruddha Chatterjee, Chhabinath Mandal. (2012). Modeling and structural analysis of PA clan serine proteases. *BMC Research Notes*, **5**, 256.

## Figure Legends

**Fig. 1.** (A) Super-position 3D-molecular structure of AsP protease from *A. solani* predicated using *Thermus aquaticus* (4DZT) as a template. Structure super-position of AsP (pink-ribbon model) with template (green-ribbon model). The five regions (N-terminus, C-Terminus, L-loop and  $\alpha$ -Helix,  $\beta$ -sheet) exhibiting relatively large local conformational differences in the ribbon model are labeled. (B) 3D-molecular architecture of AsP. The  $\alpha$ -helix,  $\beta$ -sheet and coil/loop are indicated pink, yellow and cyan color respectively. The catalytic triad aspartic acid (AYVVDSGINVNH), serine (GTSMATPHIVG), Histidine (HGTHVAGTIGG), residues in the serine protease and can be used as a signature specific to that category of proteases marked as a box. His<sup>193</sup> residue involved in metal binding region marked as a star. (C) Structure based sequence alignment of serine protease (AsP). The residue numbering underneath the sequence alignment corresponds to ten amphipathic  $\alpha$ -helical domains are indicated by dark black line and sixteen  $\beta$ -strand marked as a blue arrow line (A). Residues forming the strong calcium binding site and those participating in the formation of the active binding sites (S1-S4) are highlighted in light blue and yellow color respectively.

**Fig. 2.** (A) Ramachandran plot analysis by using RAMPAGE software. (B) Graph showed AsP hydrophobicity and amphipathy. Red color denotes amphipathicity and blue color denotes hydrophobicity.

**Fig. 3.** Serine proteases (see the Table 1) were calculated using Neighbor-Joining (NJ) method and the phylogenetic tree was constructed using the MEGA 4.0 program .

**Fig. 4.** (A) Possible ligand binding site identification of AsP. (B) Surface model. S1, S2, S3, and S4 highlighted in red color residue for binding site and His<sup>193</sup> residue highlighted in green for metal binding site. Comparison of the electrostatic surface potentials of 4DZT (C), AsP (D). The

positive surface potential is colored blue and the negative surface potential is colored red. The approximate locations of the substrate-binding sites/pockets, S1–S3 are labeled on the surfaces. (E) Structure was highlighted in helix-7 and helix-6 regions for Calcium binding sites of Ca 1 and Ca 2. The calcium ion is coordinated by the carboxyl groups of Lys<sup>84</sup>, Ile<sup>85</sup>, Lys<sup>86</sup>, Asp<sup>87</sup>, Phe<sup>88</sup>, Ala<sup>89</sup>, Ala<sup>90</sup> (K84-A90) for Ca 1 binding site. The calcium ion is coordinated by the carbonyl oxygen atom of Lys<sup>244</sup>, Gly<sup>245</sup>, Arg<sup>246</sup>, Thr<sup>247</sup>, Lys<sup>248</sup>, Lys<sup>249</sup>, and Ala<sup>250</sup> (K244–A250), for Ca 2 binding site (blue) also displayed to highlight in the ribbon model.

**Table Legends:**

**Table 1** Sequence identity and list of serine protease family.

Fig.1

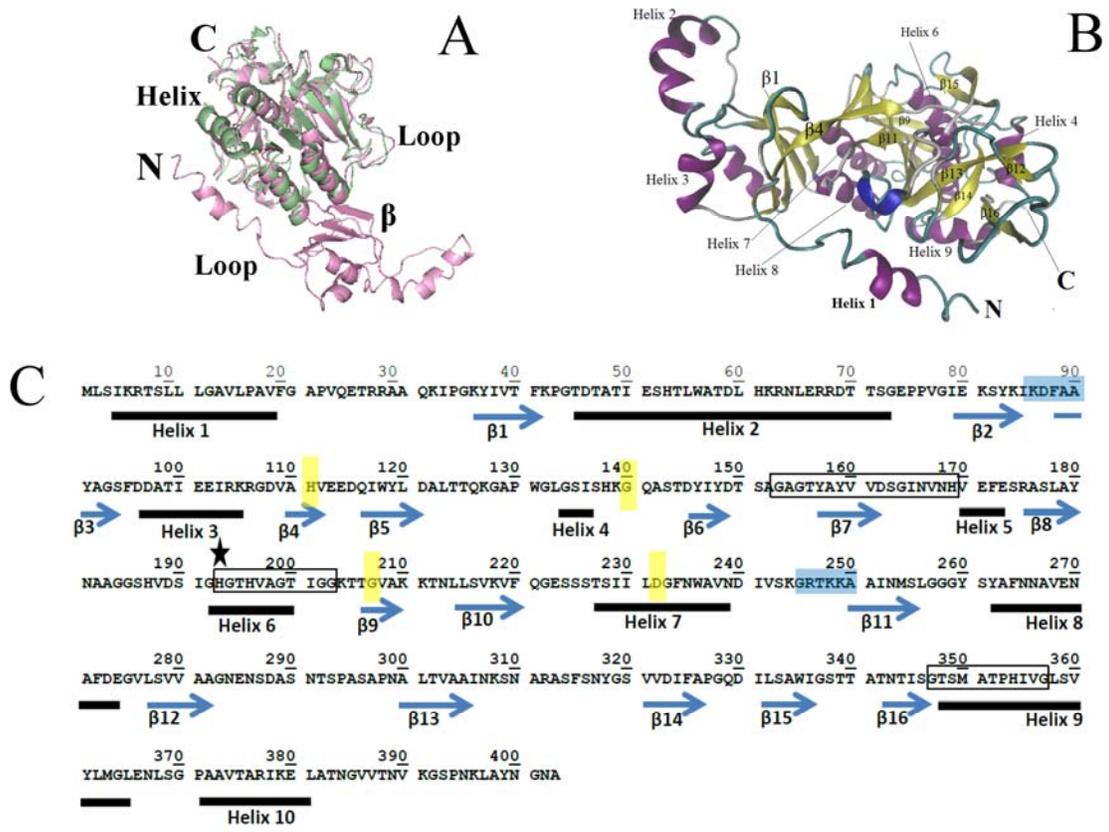


Fig.2

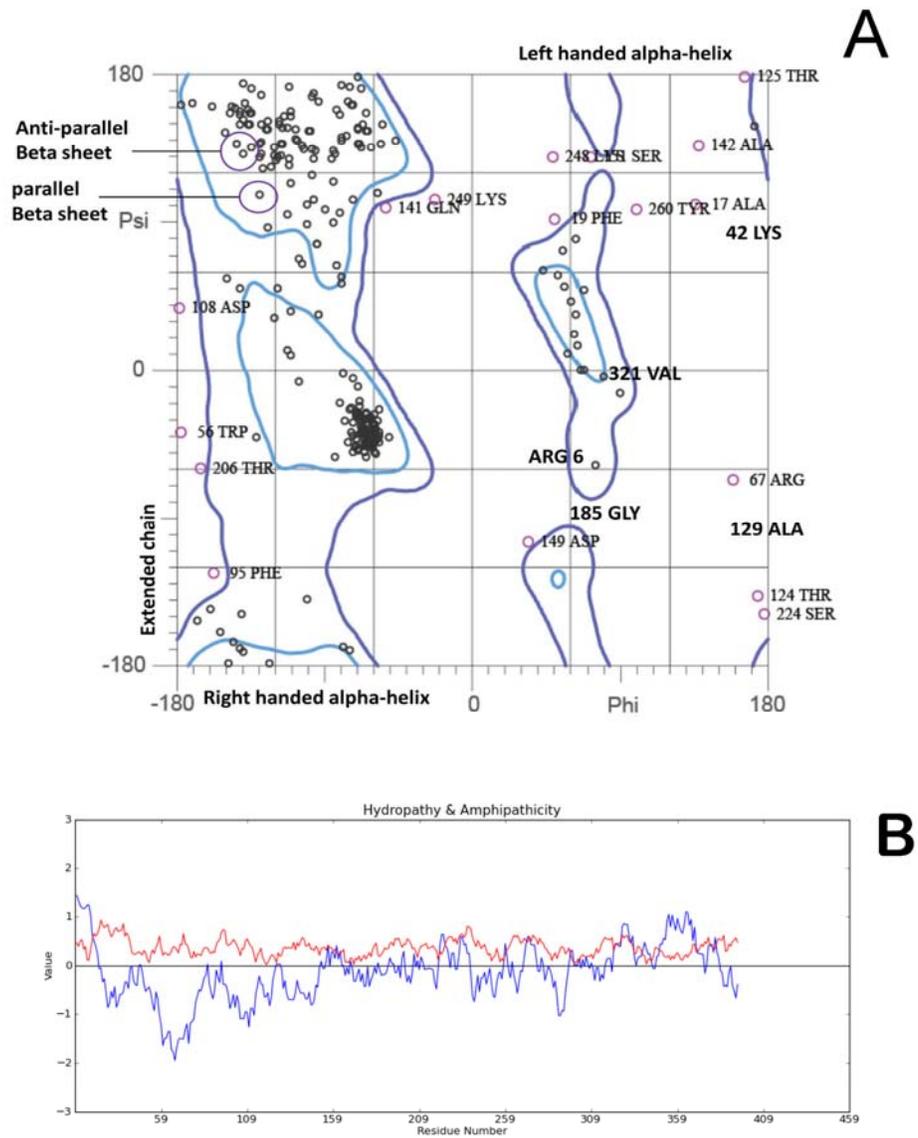


Fig.3

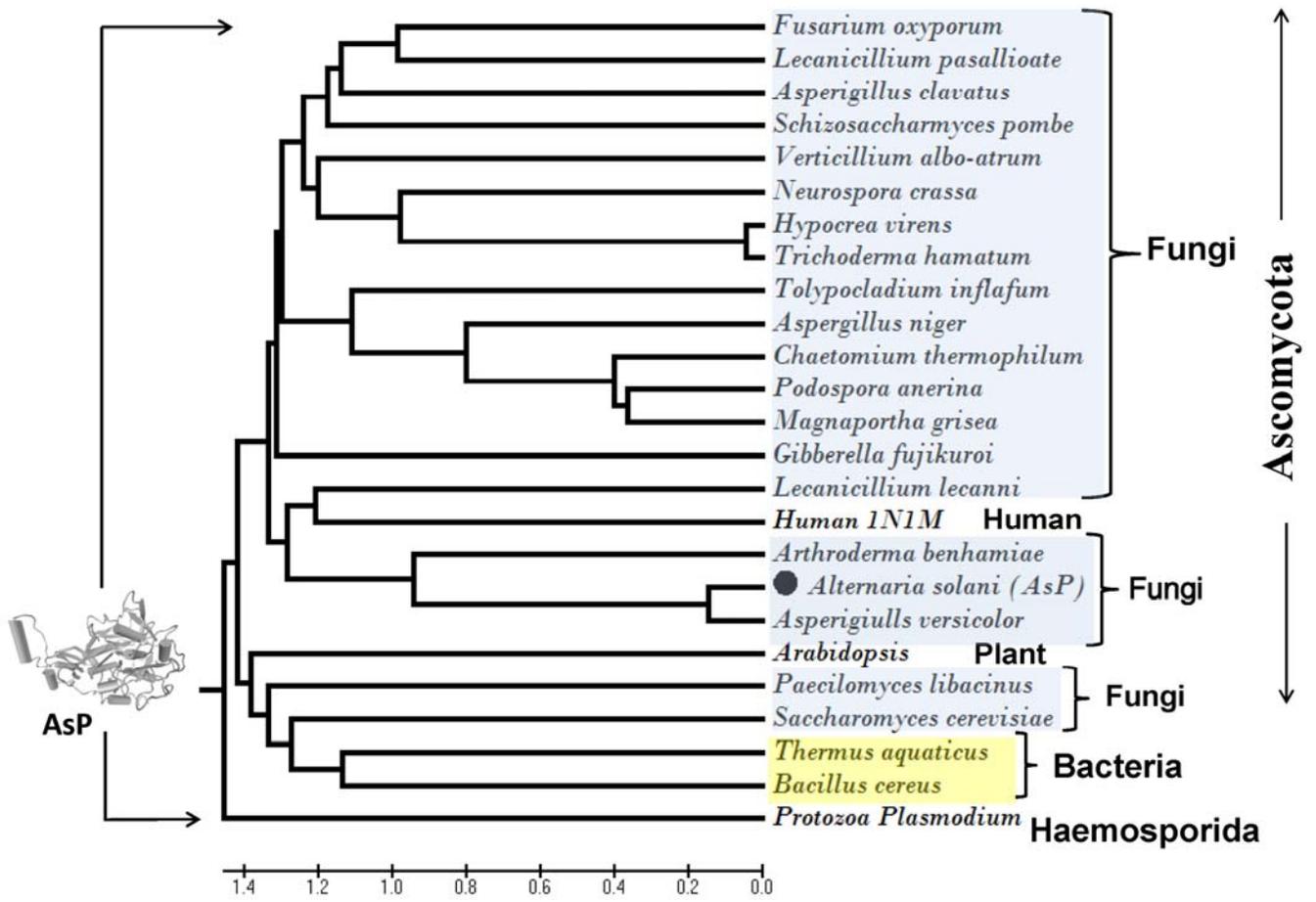
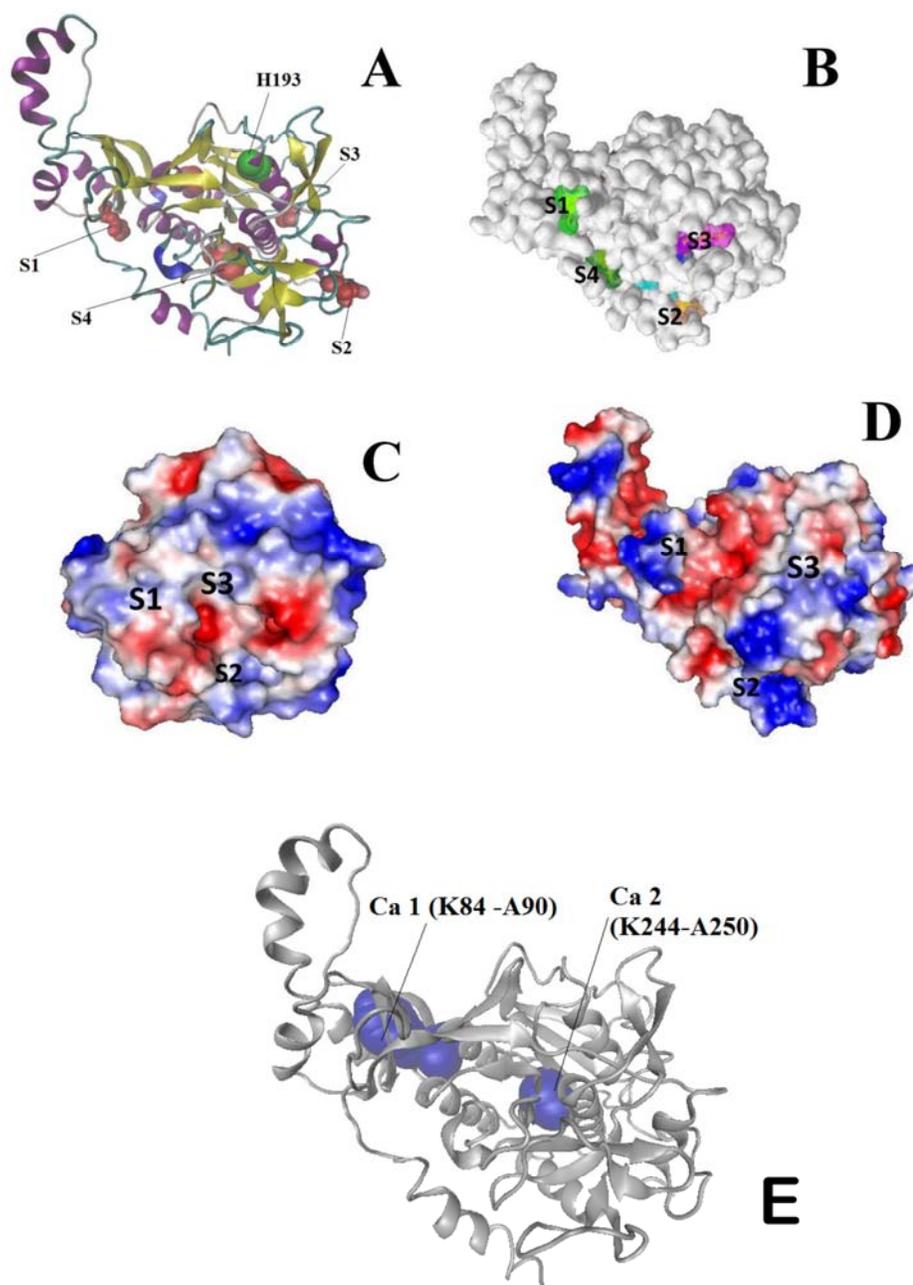


Fig.4



**Table 1**

<b>Species Name</b>	<b>Accession No.</b>	<b>Amino acid length</b>	<b>MW</b>	<b>Seq. Homology Identity</b>
<i>Saccharomyces cerevisiae</i>	X71622	597	67.7kDa	15%
<i>Chaetomium thermophilum</i>	ABK96987	532	57.4kDa	40%
<i>Podospora anserina</i>	AF047689	531	57.0kDa	41%
<i>Fusarium oxysporum</i>	AB110909	397	40.8kDa	41%
<i>Aspergillus niger</i>	M96758	533	56.9kDa	42%
<i>Lecanicillium lecanii</i>	AF059582	400	41.6kDa	42%
<i>Schizosaccharmyces pombe</i>	D14063	467	49.3kDa	43%
<i>Lecanicillium pasallioate</i>	AY870806	382	39.6kDa	45%
<i>Arthroderma benhamiae</i>	XP_003012235	424	45.9kDa	48%
<i>Neurospora crassa</i>	MER028360	299	45.5kDa	48%
<i>Tolypocladium inflafum</i>	AF467982	425	44.1kDa	49%
<i>Paecilomyces libacinus</i>	L29262	367	37.5kDa	52%
<i>Thermus aquaticus</i>	4DZT	513	53.9kDa	54%
<i>Asperigillus clavatus</i>	EAW07090	252	27.5kDa	55%
<i>Gibberella fujikuroi</i>	AY902380	329	34.6kDa	56%
<i>Magnaportha grisea</i>	AB070268	536	57.1kDa	56%
<i>Bacillus cereus</i>	YP_086726	391	41.8kDa	57%
<i>Hypocrea virens</i>	AY242844	403	42.3kDa	59%
<i>Trichoderma hamatum</i>	M87516	403	42.3kDa	59%
<i>Verticillium albo-atrum</i>	EEY20295	556	61.5kDa	67%
<i>Asperigiulls versicolor</i>	ADE74975	403	42.1kDa	78%
<i>Alternaria solani</i> (AsP)	PM0078786*	403	41.6kDa	*
<i>Homo sapiens</i>	MER000401	766	88.2kDa	29%
<i>Arabidopsis thaliana</i>	MER045469	228	24.4kDa	67%

\* Present study