BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 8(17)

Homology Modeling and Functional Characterization of PR-1a Protein of *Hordeum vulgare subsp. Vulgare*

Vahid Aslanzadeh[1]* & Mostafa Ghaderian[2, 3]

[1]Department of Biotechnology, Research Institute of Physiology and Biotechnology, University of Zanjan, Zanjan, Iran; [2]Department of Chemistry, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia; [3]Department of Pharmacy, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia; Vahid Aslanzadeh – Email: vaslanzadeh@gmail.com; *Corresponding author

Received August 18, 2012; Accepted August 20, 2012; Published September 11, 2012

Abstract:
Pathogenesis-related protein 1a of *Hordeum vulgare subsp. Vulgare* (HvPR-1a) is induced by various pathogens and stress related factors. It plays important roles in plant defense system. Since the discovery of HvPR-1a a great deal of research has been focused on its isolation and characterization. However, three dimensional structure of HvPR-1a is still unknown. 3D structure can be used for determining protein function, and identifying novel protein folds and potential targets for regulation. The protein model was developed using MODELLER 9v10. Physicochemical characterization and functional annotation of the model carried out with Expasy's ProtParam server and three different conserved domain finding programs including InterProScan, Proteins Families Database (Pfam), and NCBI Conserved Domains Database (NCBI-CDD). Applying validation programs revealed that the model has good quality and the RMSD value is 0.7. The predicted model submitted in Protein Model Database, PMDB for public use. This model will be used in wide range of studies for functional analysis and improvement activity of the protein.

Key words: HvPR-1a, Homology modeling, Three dimensional structure, MODELLER, Expasy's ProtParam server

Background:
Plants are subjected to wide range of biotic and abiotic stresses during growth and development. They respond to these stresses by activating a series of hypersensitive reactions (HR), which contribute to the neutralization of the stresses and reducing productivity loss. HR response is accompanied with synthesis of secondary metabolites and induction of defense enzymes and proteins for triggering active defense. The defense proteins include pathogenesis-related (PR) proteins that are suggested to be effective in variety of mechanisms. They are involved in inducing plant cell wall reinforcement, suppressing and degrading pathogenic enzymes and proteins, and are associated with development of systemic acquired resistance (SAR) to impede pathogenic attacks and environmental stresses **[1]**. These proteins accumulate in vacuolar compartment, cytoplasm and apoplastic space **[2]**. They are pathogen specific and are regulated at the level of transcription **[3]**. PR proteins were first observed in tobacco and subsequently, many other types of PR proteins have been detected and characterized in tobacco and other plant species. Based on serological properties and sequence homology, seventeen major groups of PRs (PR-1 to PR-17) have been identified **[2]**. PR-1 proteins with 15-17 kDa molecular mass are the dominant group, and have been detected in Arabidopsis, maize, wheat, tomato, tobacco, barley, rice and pepper. They are resistant to proteolysis and adapted to extracellular environment **[4]**. Despite other groups, exact mode of action of the PR-1 proteins is still unknown, but some of recent studies have been assigned antifungal activity to members of this group, which include HvPR-1a. In 1993 Alexander et al. reported that tobacco plants transformed to

constitutively express PR-1 showed increased tolerance to two oomycete pathogens, *Peronospora tabacina* (blue mold) and *Phytophthora parasitica var nicotinae* (black shank). But the transgenic tobacco plants remained susceptible to infection with tobacco mosaic virus (TMV), potato virus Y (PVY), *Cercopspora nicotianae*, and *Pseudomonas syringae pv tabaci* **[5]**. In vitro activity test of tomato PR-1 revealed that PR-1 inhibited zoospore germination of *Phytophthora infestans* and, in vivo, its application decreased the surface area of leaf discs infected with this oomycete **[6]**. PR-1 proteins are strongly conserved proteins and their homologues sequences have been detected in fungi, insects, and vertebrates, including human **[1].**

Barley is a vitally important crop for human consumption. In this study we modeled three dimensional structure of HvPR-1a based on the high quality homology modeling approach to establish a basis for future researches about its biological function and interaction properties. Computational structure prediction methods provide cost-effective and time-effective alternative in absence of experimentally derived structures. This study predicts some of the properties of HvPR-1a and proposes a high quality 3D structure for it. The results will be useful to understand structural features and anti-pathological function of HvPR-1a, and will raise the prospects of its potential use for engineering plants for resistance to phytopathogenic fungi.

Methodology:

Template identification and structure prediction

The amino acid sequences of HvPR-1a protein were retrieved from NCBI protein sequence database **[7]**. Following BLASTP run of the query model, appropriate template to use in building a 3D model was selected based on the high score, lower e-value and maximum sequence identity. Although some techniques can produce 3D models from more than one template, most of them depend on a single template. Accordingly, choosing the appropriate template is a critical stage to certify quality of the final structure. To ensure the high accuracy of the structure, sequence alignment between query and template has shown in **Table 1** (**see supplementary material**). Prediction of the 3D-structure of query protein was performed by popular homology modeling program, MODELLER 9v10 **[8]**. After careful alignment of query and template proteins and removing potential errors, loop refinement was performed using loop optimization method of MODELLER.

Model evaluation

Model evaluation is the most important step in homology modeling to show the acceptable quality of the modeled protein. Hence, we used PROCHECK **[9]** to evaluate the quality of our predicted models by Ramachandran plot assessment. Models subjected to energy minimization with GROMOS 96 43B1 parameters set, implementation of Swiss-PDB Viewer (SPDBV) to rectify unfavorable clashes and improve the quality of stereochemistry **[10]**. The Root Mean Square Deviation (RMSD) values were computed to measure differences between the predicted model and the template protein structures. SPDBV and Chimera programs **[11]** were used to analyze and represent the structure of generated model. Further, models verification was performed with VERIFY3D **[12]**, ERRAT **[13]**, and WHAT IF **[14]** programs. ProSA **[15]** was employed for the analysis of structural errors and making plot of residue scores and calculating Z scores.

Functional characterization

HvPR-1a protein submitted to three different protein annotation databases including InterProScan **[16]**, Proteins Families Database (Pfam) **[16]**, and NCBI Conserved Domains Database (NCBI-CDD) **[17]** for functional characterization and classification. InterProScan classifies sequences, predict the occurrence of functional domains and adds GO terms to protein signature. Pfam is a large collection of annotated protein families that generated using hidden Markov models. NCBI-CDD is a database for fast identification of conserved domains in protein sequences. Protein was further subjected to CYS_REC program(http://linux1.softberry.com/berry.phtml?topic=cys_rec&group=programs&subgroup=propt) for locating SS-bonding states of cysteines and disulphide bridges in proteins, if present.

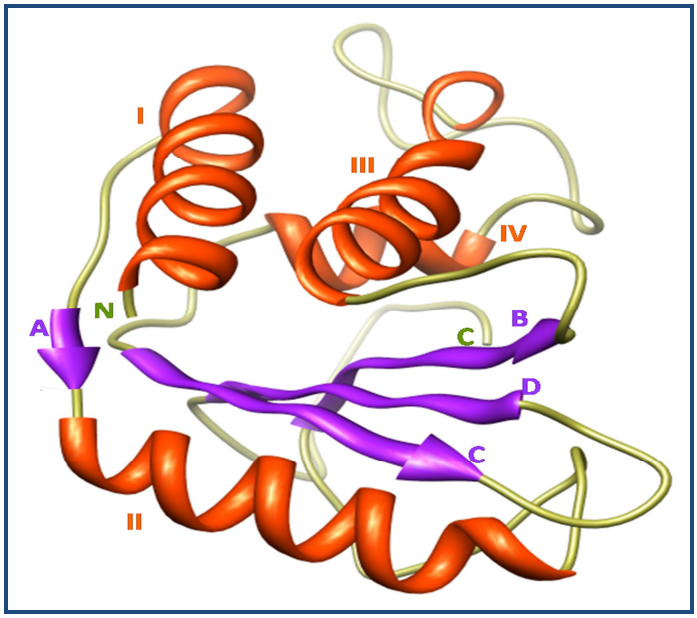

Figure 1: Modeled structure of HvPR-1a as visualized by Chimera 1.6.1. The structure has been formed of four α helices and four β strands. Strand **D** is antiparallel with strand **B** and **C**; and short strand **A** is parallel with strand **B**. Strands **A** and **B** connected to each other with helix **II** that is located under the β sheets. Helices **III** and **I** are parallel, located above the β-sheets and attached as horizontal and right angle related to β-strands. Helix **IV** located between helices **I**, **III** and β-strands. Locating helices in two sides of the beta plane provide a tight and stable folding for protein. **N** shows the N-terminal and **C** shows the C-terminal of the protein. C-terminal is longer because it contains CRISPs family signal motifs which are important for protein function. A 3_{10}-helix located just following the helix **III** compromising the residues 76, 77 and 78.

Discussion:

Assessment of resultant structure with several model evaluation tools demonstrated high quality of the model. Ramachandran plot of the structure showed that 93.9% of residues fell in most favorite regions, and 6.1% fell in additional allowed regions; none of the residues found in the generously allowed regions and disallowed regions. This validates the quality of modeled structure. Overall average G-factor of dihedral angles and main-chain covalent forces was -0.21 that was greater than acceptable cutoff -0.5. The *G*-factor provides a measure of the plausibility of a stereochemical property and a high G-factor means the property corresponds to a high-

probability conformation. Other structure validation programs such as WHAT IF, VERIFY3D and ERRAT were employed for further validation of models. WHAT IF server assigned a Ramachandran Z-score of -1.274 and structural average packing score of -1.138. Both of these scores were within the expected range (between 0.0 to -2.0) for well-refined structures. RMS Z-scores for bond angles and bond lengths as determined by WHAT IF was 1.404 and 0.936 respectively which is very close to 1.0 suggesting high model quality. VERIFY3D found that at least 80% of the amino acids have scored >= 0.2 in the 3D/1D profile. This suggests that atomic model (3D) is compatible with its amino acid sequence (1D).

Overall calculated quality factor of ERRAT was 94.40 for modeled protein, indicated the error values of individual residue are negligible. ProSA calculated Z-score for modeled protein was -6.79 which fell in the range of scores typically found for experimentally determined (X-ray, NMR) structure for native proteins of similar size (137 AA) in current PDB database. The z-score measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations **[14]**. Z-scores outside characteristic range for native proteins indicate erroneous structures. Further, calculated average energy over 40 residues with ProSA validated good quality of the modeled structure.

SPDBV analysis of predicted model

Analysis of the model using SPDBV reveals that none of the residues were found to make clashes in their existing position. HvPR-1a final model shows four α-helices (I to IV) containing the residues 2 to 16, 24 to 39, 64 to 75 and 93 to 99, and four β-strands (A to D) compromising the residues 21-22, 51 to 57, 106 to 112 and 118 to 124 arranged antiparallel between helices I, III and IV and II, respectively **(Figure 1)**. This identifiable α + β topology having helices in above and below the β strands clearly forming a compact folding. This unique 3 layer alpha-beta-alpha fold which first detected in PR-1 proteins reflects the high stability of modeled structure and its resistance to several proteases **[1]**. Estimated RMSD after superimposition structures of modeled and template proteins with SPDBV was 0.7. This indicates high similarity between HvPR-1a and template protein and therefore good quality of modeled HvPR-1a structure.

Functional characterization

In this study, several bioinformatics tools and databases were used for structural prediction and validation of the HvPR-1a. For identifying conserved domains and functional annotation of the HvPR-1a protein we used three different web tools including InterProScan, Pfam, and NCBI-CDD. All of these tools confirmed that HvPR-1a belongs to CAP super family and contains CAP domain. This family includes plant pathogenesis-related protein 1 (PR-1), which accumulates after infections with pathogens with hypersensitive reaction, and may act as an anti-fungal agent or be involved in cell wall loosening. Therefore, they play important roles in plant defence systems. Furthermore, the sequence and structural evidences support that members of this family are involved in Ca2+ chelating function and potentially act as either protease or protease inhibitors. Members are most often secreted and have an extracellular function. This family also includes CRISPs, cysteine-rich secretory proteins, which combine CAP with a C-terminal cysteine rich domain. Further, analysis with PROSITE **[18]**, database of protein domains, families and functional sites, two conserved CRISP family motif in C-terminal of the protein was detected **Table 2** (**see supplementary material**). Cysteine rich proteins form disulfide bridges that are important for their stability and extracellular function. For prediction disulphide bridges in proteins we used CYS_REC program. This program recognized three probable SS bounding patterns of pairs between cysteine residues 41-113, 86-92, and 108-123.

To have an accurate interpretation about function of the model we used Expasy's ProtParam server **[19]** for computing physicochemical properties. The computed parameters were included the molecular weight (MW), theoretical isoelectric point (pI), extinction coefficient (EC), instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY) **Table 3 (see supplementary material**). The results showed that protein have a basic nature and is thermostable and hydrosoluble. These properties are compatible with the extracellular function of the protein.

Biological function of the protein

Antifungal activity: Since the discovery of pathogenesis-related proteins in 1970 the biological function of the PR-1proteins remained unclear. To understand the functions of proteins at a molecular level, determining their three-dimensional structure is a prerequisite. This is the first work in proposing a three dimensional model to HvPR-1a protein. Having this structure will provide an opportunity for researchers to discover its mode of action. It has been reported that PR-1 proteins suppress infections induced by selected pathogens from oomycete mainly by targeting their cell wall. By using cell wall polymer microarrays researchers can design different substrates from cell wall components of these fungi and assign the candidate substrates with high confidence to the protein by employing various and powerful protein docking tools. These achievements will help to develop *in vitro* protein function assessments against candidate substrates to explore biological activity of the protein.

Disturbing Ca2+ channel function: Although the exact antifungal and enzymatic activity of the plant PR-1 proteins remains largely unknown, a PR-1-like protein, helothermine, from the Mexican banded lizard interacted with membrane-channel proteins of target cells, inhibiting the release of Ca2+ **[20]**. A tip-high Ca2+ gradient is observed in growing fungal hyphae, and, low Ca2+ concentrations result in inhibiting hyphal growth and branching **[21]**. High sequential and structural similarity of the helothermine and HvPR-1a increase the probability that PR-1 proteins may be act as antifungal agents by inhibiting the release of Ca2+ in fungi cells. Positioning of the secondary structures elements in helothermine as shown in **Figure 2** is more similar to HvPR-1a. However, more computational and experimental studies are required for assessment Ca2+ channel disturbing function of the HvPR-1a.

Submission of the protein structure in Protein Model Database (PMDB)

The modeled structure of HvPR-1a was successfully deposited in PMDB **[22]** with accession number PM0078336 after successfully passing PMDB stereochemical checks. This *in silico* generated proteins structures database is freely available for

public use and users can access to the model by its accession number.

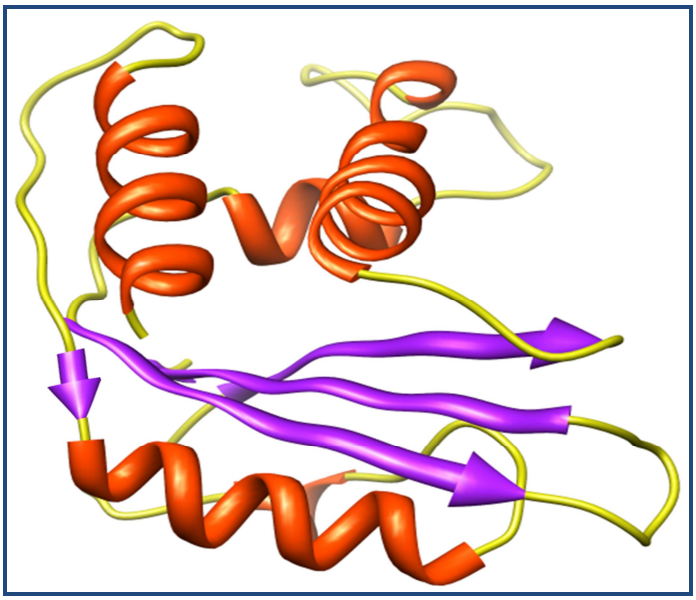

Figure 2: 3D modeled structure of SCP_CRISP region of helothermine protein (GenBank: AAC59730.1). This structure is compromised of four antiparallel β-strands and four α helices located above and below the β plane same as the HvPR-1a protein. A short α helix also located under the β-strands.

Conclusion:

In this study PR-1a protein from *Hordeum vulgare subsp. Vulgare* was modeled based on the comparative modeling approach using MODELLER. After refinement of the model, verification of the modeled structure performed with different programs. Results of these verification tools and low RMSD value indicated that the modeled structure possess good and reasonable structural quality. Physicochemical and functional studies performed for characterization of HvPR-1a in reaching conclusions about the biochemistry and biological function of the modeled protein. The results have shown that this protein is stable and secretory protein and it is play antifungal and Ca2+ cannel blocking function in extracellular environment. However, the significance and precise characterization of HvPR-1a associated with defensive response of plants against infections remain to be determined. High quality modeled structure of HvPR-1a may be exploited in molecular docking and developing wet lab experiments for profound insights into its biological activity and for biotechnological engineering of the protein for improvement its anti-pathogenic response in transgenic plants.

Acknowledgement:

We thank Mostafa Afshari for critical reading of the manuscript.

References

[1] Van Loon LC *et al*. *Annu Rev Phytopathol*. 2006 **44**: 135 [PMID: 16602946]
[2] Stintzi A *et al. Biochimie.* 1993 **75**: 687 [PMID: 8286442]
[3] Van Loon LC & Van Strien EA, *Physiol Mol Plant Pathol.* 1999 **55**: 85
[4] Liu Q & Xue Q, *J Genet.* 2006 **85**: 193 [PMID: 17406093]
[5] Alexander D *et al. Proc Natl Acad Sci U S A.* 1993 **90**: 7327 [PMID: 8346252]
[6] Niderman T *et al. Plant Physiol.* 1995 **108**: 17 [PMCID: PMC157301]
[7] http://www.ncbi.nlm.nih.gov/sites/entrez?db=protein
[8] Sali A & Blundell TL, *J Mol Biol.* 1993 **234**: 779 [PMID: 8254673]
[9] Laskowski RA *et al. J Appl Cryst.* 1993 **26**: 283
[10] Kaplan W & Littlejohn TG, *Brief Bioinform.* 2001 **2**: 195 [PMID: 11465736].
[11] Pettersen EF *et al. J Comput Chem.* 2004 **25**: 1605 [PMID: 15264254]
[12] Eisenberg D *et al. Methods Enzymol.* 1997 **277**: 396 [PMID: 9379925]
[13] Colovos VC & Yeates TO, *Protein Sci.* 1993 **2**: 1511 [PMID: 8401235]
[14] Vriend G, *J Mol Graph.* 1990 **8**: 52 [PMID: 2268628]
[15] Sippl MJ, *Proteins.* 1993 **17**: 355 [PMID: 8108378]
[16] Zdobnov EM & Apweiler R, *Bioinformatics.* 2001 **17**: 847 [PMID: 11590104]
[17] Marchler-Bauer A *et al. Nucleic Acids Res.* 2011 **39**: D225 [PMID: 21109532]
[18] Falquet L *et al. Nucleic Acids Res.* 2002 **30**: 235 [PMID: 11752303]
[19] Wilkins MR *et al. Methods Mol Biol.* 1999 **112**: 531 [PMID: 10027275]
[20] Morrissette J *et al*. *Biophys J.* 1995 **68**: 2280 [PMID: 7647234]
[21] Regalado CM, *Microbiology.* 1998 **144**: 2771 [PMID: 9802018]
[22] Castrignanò T *et al. Nucleic Acids Res.* 2006 **34** :D306 [PMID: 16381873]

Edited by P Kangueane

Citation: Aslanzadeh & Ghaderian, Bioinformation 8(17): 807-811 (2012)

BIOINFORMATION open access

Supplementary material:

Table 1: Sequence alignment between query and template. Query sequence has shown the highest similarity and identity with 1CFE sequence. These observations and belonging 1CFE to PR-1 group of proteins left only 1CFE as the best potential template.

Selected template: Pathogenesis-related protein P14a of Solanum lycopersicum
PDB ID: 1CFE
UniProtKB: P04284
Identity: 55.7%
Similarity: 66.4%
E-value: 4e-48

```
Query    1 QNSPQDYLAVHNDARAQVGVGPMSWDANLASRAQNYANSRAGDCNLIHSG  50
             |||||:.||.|||.||||.:||...|.:.||:|||.|.|||.|.|||
1CFE     1 ---PQDYLSPHNAARAAVGVGAVSWSTKLQAYAQSYANQRIGDCKLQHSG  47

Query   51 A--GENLAKG--GGDFTGRAAVQLWVSERPSYNYATNQCVGGKKCRHYTQ  96
           .  |||:..|  |.|:....||:|||.|:..|:|.:|.|.|||.|.||||
1CFE    48 GPYGENIFWGSAGADWKAADAVKLWVDEKKDYDYGSNTCAGGKVCGHYTQ  97

Query   97 VVWRNSVRLGCGRARC-NNGWWFISCNYDPVGNWIGQRPY  135
           ||||.|..:||.|..| |||..||:|||:|.||.:||:||
1CFE    98 VVWRASTSIGCARVVCNNNGGVFITCNYEPAGNVVGQKPY  137
```

Table 2: Functional conserved motifs found in HvPR-1a using PROSITE (release 20.83)

Motif Family	Motif Found	Position in the sequence
CRISP	GHYTQVVWraS	93 - 103
CRISP	FItCNYePaGNV	120 - 131

Table 3: Physicochemical parameters computed for HvPR-1a. Extinction coefficient (EC) indicates higher concentration of Tyr, Trp, and Cys (13.9%) residues. EC is valuable for purifying protein through spectrophotometer. Instability index (II) indicates higher stability of protein. Aliphatic index (AI) is relatively high but high contents of serine and threonine (13.4%) (hydrogen bond-forming amino acids) increase theromstability of protein. Very low GRAVY value suggests solubility of protein in water.

Sequence length	MW	pI	EC	II	AI	GRAVY
137	14656.2	8.36	40910 to 41258	26.86 is below the threshold value (40)	62.70	-0.383