J Biochem Tech (2014) 5(2):674-678
ISSN: 0974-2328

Heterologous expression and in-silico characterization of pathogenesis related protein1 (CsPR1) gene from *Camellia sinensis*

Niraj Agarwala*, Priyadarshni Bhorali, Neelakshi Bhattacharyya, Raju Bharalee, Tirthankar Bandyopadhyay, Bornali Gohain , Sushmita Gupta, Manab Deka, Sudripta Das

Received: 03 September 2013 / Received in revised form: 05 October 2013, Accepted: 09 October 2013 Published online: 03 December 2013

Abstract

Pathogenesis related protein1 gene induced after pathogen infection in plants have been frequently used as marker gene for systemic acquired resistance. We have carried out isolation, annotation and expression of *CsPR1*, a potential disease resistance gene. The full length cDNA consist of 671 bp in length containing 162 amino acids with a signal peptide of 22 amino acids and 17.92 kDa predicted molecular weight. Recombinant *CsPR1* was successfully expressed in *BL21(DE3)pLysS* cells using pET 43.1 EK LIC vector system and was purified. Three dimensional models were generated using Phyre2 and I-TASSER and built a compact structure consisting beta sheets surrounded by alpha helixes. The models were validated by MolProbity and RAMPAGE servers. Validation of modelled structures based on Ramachandran plot, revealed I-TASSER produce better quality and reliable 3D model. Purified recombinant *CsPR1* and in silico generated 3D models from this study provide foundation for comprehensive functional and structural characterization of *CsPR1* protein.

Keywords: Heterologous expression, *Camellia sinensis*, *CsPR1*, Phyre2, I-TASSER.

Niraj Agarwala*, Neelakshi Bhattacharyya

Department of Biotechnology, TRA, Assam, India.

*Email: niru.niraj.s@gmail.com

Priyadarshni Bhorali

Assam Agricultural University, Assam, India

Raju Bharalee

The Energy and Resources Institute, Guwahati, Assam, India

Tirthankar Bandyopadhya

National Institute of Plant Genome Research, New Delhi, India

Bornali Gohain

National Bureau of Plant Genetic Resources, New Delhi, India.

Sushmita Gupta

Institute of Advance studies in Science and Technology, Assam, India.

Manab Deka

Department of Biological Sciences, Gauhati University, Assam, India.

Sudripta Das

Institute of Bioresources and Sustainable Development, Takyelpat, Manipur, India.

Introduction

Pathogenesis related proteins are proteins encoded and induced in host plants in response to presence and colonization with pathogens. Pathogenesis related proteins are thermo-stable, low molecular weight and highly resistant to proteases (Van Loon and Van strein, 1999) comprise of 17 families (van Loon et al., 2006). PR proteins found all parts of plants, but most abundant in leaves. PR proteins known to play direct role in disease resistance by exhibiting high antimicrobial activity (Tonon et al., 2002; Anand et al., 2004). PR gene belonging to different PR protein families were used for development of transgenic plant and results in development of potato, rice and *Brassica spp* plants with high level of tolerance or resistance to pathogens (Takahashi et al., 2005L; Datta et al., 2001; Sarowar et al., 2005).

PR-1 is a dominant group of PRs induced by pathogens or SA, and is commonly used as a marker for SAR. Numerous researchers have attempted to assess the function of PR-1 proteins in plants and reported their antifungal activity. The structure of tomato PR-1b (*P14a*) was solved by nuclear magnetic resonance and found to represent a unique molecular architecture. Tomato PR1b have four β strands arranged in antiparallel with four α helices forming a compact structure stabilized by hydrophobic interactions and multiple hydrogen bonds making it more stable and insensitive to proteases (Fernandez et al., 1997). PR1 have

inhibitory effect on *Phytophthora infestans* and *Uromyces fabae* in tomato and broad bean (Niderman et al., 1995; Rauscheret al., 1999) Constitutive expression of PR1a gene in transgenic tobacco confers resistance to *Peronospora tabacina* and *Phytophthora parasitica* var. *nicotianae* (Alexander et al., 1993). Silencing of PR1b in barley confers susceptibility by facilitating *Blumeria graminis* f. sp. *hordei* infection (Schultheiss et al., 2003). Pathogenesis related proteins are now very popular for engineering disease resistance in important plants.

Tea is the oldest and most popular beverage obtained from tender shoots of *Camellia sinensis*. Tea is propagated in various climatic conditions and are challenged by spectrum of biotic and abiotic stresses. Biotic Stress in plants leads to activation of a number of defense and resistance genes. Pathogenesis related protein1 (*CsPR1*) gene emerged as most promising genes from a comprehensive study in tea in response to blister blight disease in tea (Unpublished) considered for heterologous expression.

Plant Protein isolation or purification is costly and time consuming process. Heterologous expression allows production of proteins in a different and simpler organism for easier and higher production .The gene or the DNA portion of interest can be transferred to different system for the synthesis of encoded protein. Different biochemical and biophysical properties (Ligand binding, cell surface characteristics etc.) can be studied by heterologous expression of important proteins. Antimicrobial proteins from plants expressed in heterologous systems (Kovaleva et al., 2009. 2011) for functional characterization. Various computational tools are also now available to understand physico chemical and structural properties of proteins. Homology modeling has proved very useful in the prediction of the 3D structure and function of proteins. In silico approaches are amenable to high throughput and provide solution for time consuming, laborious and expensive experimental processes.

Here we have carried out heterologous expression and in silico analysis of *C.sinensis* Pathogenesis related protein1 (Cs*PR1*).

Materials and Methods

PCR primer designing and amplification

The source sequence accession number of *CsPR1* is KF527571. PCR primers were designed following the recommendation of manufacturer EK LIC kit manual from (NOVAGEN). 5'GAC GAC GAC AAG ATG 3' sequence was incorporated at 5' end of insert specific sequence of forward primers. In case of reverse primer 5' GA GGA GAA GCC CGG TTA 3'sequence was added at the 5'end. Hot start phusion polymerase (Finnzymes) was used in PCR reactions for high fidelity. PCR condition for amplification i.e initial denaturation for 2 minutes at 98°c ,then 35 cycles of 30s at 98°c,1 min at 60°c, 1 min at 72°c and final denaturation at 72°c for 10 minutes. PCR products were run on 1.5%agarose gel and resulting bands excised out and gel extracted using Hielute gel purification kit (Himedia).

Sense primer 5'
GACGACGACAAGATGGTGGTGTGTAACCTTTCATTAGCTT3'
Antisense primer:
5'GAGGAGAAGCCCGGTTAATAAGGAC GTTGTCCAATATAGT 3'

T4 DNA polymerase treatment and annealing in pET 43.1 EK LIC vector

T4 DNA polymerase treatment of the purified fragment was done according to the protocol to make vector compatible overhangs, but incubation time was increased from 30 minutes to 1 hour, for higher efficiency .T4 DNA polymerase treated insert incubated with pET 43.1 EK LIC vector in presence of EDTA for 1 hour.

Transformation and Induction of protein expression

Annealing reaction was directly transformed in NovaBlue GigaSingles™ competent Cells (Cat. No.71127) by heat shock method and selected on LB ampicillin 50µg/ml. Colony PCR was done to confirm the successful transformation of gene. Plasmids extracted and again transformed in expression host strain *BL21 (DE3) pLysS* by heat shock method. Starter culture of *E.coli BL21 (DE3) pLysS* containing plasmid construct with the gene was grown in 2ml LB broth supplemented with 50µg/ml of ampicillin and 34µg/ml of chloramphenicol at 37°c for 12 hours. Fresh LB media (25ml x 4 flask) was inoculated with 500 µl of starter culture up to OD 0.5 and Isopropyl β-D-thiogalactopyranoside (IPTG) was added to final concentration 0.2, 0.5,1 mM in three flask with a control flask without IPTG and grown at 25°c shaking at 225rpm for 12 hours, but aliquots collected at every 4 hours i.e after 4,8,12 hours, three time points.

Purification of recombinant protein

Two ml culture after 4 hours of induction was taken in microcentrifuge tubes and pelleted for protein purification. *CsPR1* protein was purified using Dynabeads® His-Tag Isolation & Pulldown (Invitrogen, cat.no.101.03D) following manufacturer's instructions. Freeze thaw method was used for lysis of bacterial cells, where 4-5 cycles of cells frozen in liquid nitrogen and thawed on ice were repeated. T7 lysozyme expressed by *BL21 (DE3) pLysS* cells. Freezing thawing cause damage to the inner membrane and T7 lysozyme leaks out to the cytoplasm and digest the outer wall of bacterial cells .Proteins were eluted in 50µl of His elution buffer instead of 100µl mentioned in protocol.

I*n-silico analysis and homology modelling*

CsPR1 was selected for In-silico analysis and homology modeling. *CsPR1* protein was aligned with homologues from other plants using Clustal Omega at EBI server (http://www. ebi.ac.uk/ Tools/msa/clustalo/) and coloured view of the con-served amino acids in the aligned sequences analyzed using sequence manipulation suite (http://www.bioinformatics.org /sms2/color_align_prop.html).

Physico chemical characterization

For physico-chemical characterization, Expasy's ProtParam server (Gasteiger et al.,2005) (http://us.expasy.org/tools /protparam.html) was used for the calculation of theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and GRAVY(Kyte and Doolottle.,1982).

Prediction of protein structure

The protein structure were predicted with Phyre2 protein fold recognition server (Kelley and startnberg., 2009) (http://www.sbg.bio.ic.ac.uk /phyre2/html/page.cgi?id=index) and I-TASSER (Zhang 2008, Roy et al.,2010,2012) (http://zhanglab. ccmb.med. umich.edu/I-TASSER/). The obtained structure was validated using MolProbity (Chen et al., 2010, Davis et al., 2007) (http://molprobity.biochem.duke.edu/) and RAMAGE (http:// mordred.bioc.cam.ac.uk/~rapper/rampage.php). Transmembrane protein helix probability curve was analyzed by using SOSUI server (Hirokawa et al., 1998) (http://bp.nuap. ngoau.ac.jp/sosui/ sosui_submit.html). SOPMA (Self-Optimized Prediction Method with Alignment) was used to calculate the secondary structure features of *CsPR1* (http://npsa-pbil.ibcp.fr/cgi-bin/secpred _ sopma.pl). CYS_REC used to locate the position of cysteine residues and SS bonding status in

cysteines and locating disulphide bridges. To identify conserved motifs in *CsPR1*, protein sequences analyzed by MEME (Multiple Em for Motif Elicitation)(http://meme.sdsc.edu/meme/cgi-bin/meme.cgi),using default parameters except occurrences of a single motif distributed among the sequences was set to any repetitions.

categorized the protein as membrane protein. CYS_REC was used to identify cysteine residues and disulphide bonds and results can be found in table no III.

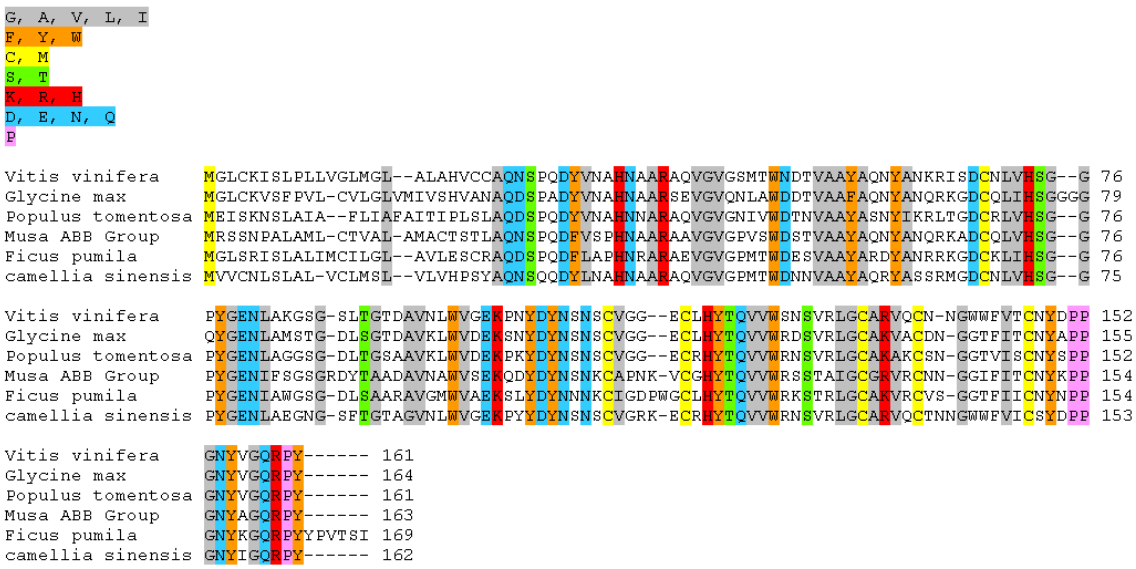

Figure I: Colour aligned *CsPR1* protein on the basis of conserved biochemical properties of the residues

Results and discussion

The cDNA sequence encoding *CsPR1* 671 bp in length, containing 49 bp 5'untranslated region, 489 bp open reading frame (orf) and 136 bp 3'untranslated region. *CsPR1* consist of 162 amino acid and molecular weight of 17.92 kda. Molecular weight of *CsPR1* with tag is 17.92+60Kda=77.92 kda Parameters computed using ExPaSy's ProtParam tool was represented in Table (I). The calculated isoelectric point (pI) will be useful because at pI, solubility is least and mobility in an electro focusing system is zero. Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of protein is zero. At pI, proteins are stable and compact (Sahay et al., 2010).The computed pI value of *CsPR1* 8.1, which shows the basic nature of the protein. The computed isoelectric point (pI) will be useful for developing buffer system for purification by isoelectric focusing method. Extinction co-efficient at 280nm in water was in the range of 45380-45880 in case of *CsPR1*. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. The instability index of *CsPR1* 32.26. So, we can consider *CsPR1* as a stable protein. The aliphatic index (AI) which is defined as the relative volume of a protein occupied by aliphatic side chains (A, V, I and L) is regarded as a positive factor for the increase of thermal stability of globular proteins. Aliphatic index is 69.75 indicate that *CsPR1* protein stable at diverse range of temperatures. The GRAVY (Grand average hydropathy) value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids, divided by total number of residues in the sequence. Here GRAVY value for *CsPR1* is -0.300.Low GRAVY values indicate that *CsPR1* will interact with water.

Functional analysis of these proteins includes prediction of transmembrane region, identification of important motifs. SOSUI distinguishes between membrane and a soluble protein based on amino acid sequences, and predicts the transmembrane helices. The transmembrane regions and their length predicted in *CsPR1* can be found in table II. SOSUI server identified one transmembrane region rich in hydrophobic ami no acids in *CsPR1*, and thus

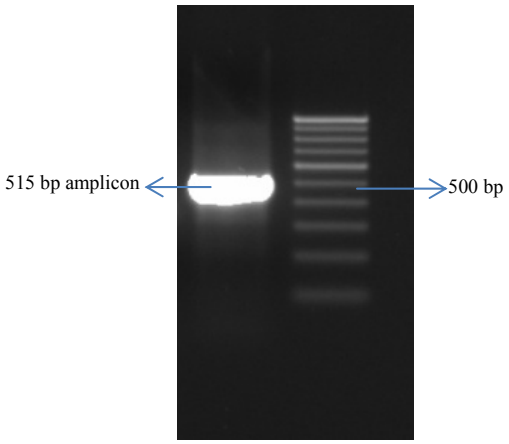

Figure 2: Amplified fragment of *CsPR1* with 100 bp ladder.

Table I: Parameters calculated using ExPaSy's ProtParam tool.

Parameters	***CsPR1***
No of amino acids	162
Molecular weight	17928.1
pI	8.11
+ Residues	9
-Residues	11
Chemical Formula	$C_{784}H_{1180}N_{228}O_{234}S_{12}$
Total atoms	2438
Extinction coefficient ($M^{-1}cm^{-1}$)	45380-45880
Half life(hours)	30 (Mammalian reticulocytes),>20 (Yeast),>10 (E.coli)
Instability index	32.26
Aliphatic index	69.75
GRAVY	-0.300

The secondary structure of *CsPR1* proteins were predicted by SOPMA (Self Optimized Prediction Method with Alignment).The secondary structure shows whether a given amino acid is in helix strand or coils. In case of both the proteins, secondary structures are dominated by random coil, which is followed by alpha helix. Details of the SOPMA prediction given in tabulated form.

Table 2 : SUSUI server prediction of transmembrane helix

SOSUI	N terminal position	Transmembrane helix region	C terminal position	Type of helix	Length
CsPR1	1	MVVCNLSLA LVCLMSLVL VHPS	22	primary	22

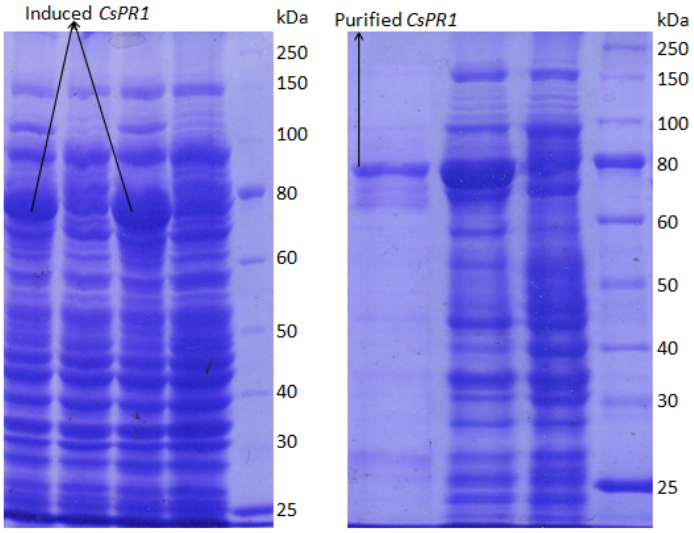

Figure 3: A.CsPR1 Induction using 1mM IPTG (lane1.Induced 2.uninduced 3.induced.4.uninduced 5.protein marker 10-250 kDa). B .*CsPR1* purification (lane 1. *CsPR1* purified 2. Induced *CsPR1* 3. Uninduced control 4.protein marker 10-250 kDa)

Table 3: CYS_REC prediction of cysteine residues

CYS_REC	No of cysteine residues	Position of cysteine residues	Probable pattern pairs
CsPR1	8	4,12,68,111,117,133,138,148	68-138,111-117,133-148

After secondary structure prediction by SOPMA, PROSITE was used to scan *CsPR1* protein on the basis of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles. PROSITE is a database of protein families and domains. It is based on the observation that, while there is a huge number of different proteins, most of them can be grouped, on the basis of similarities in their sequences, into a limited number of families. Proteins or protein domains belonging to a particular family generally share functional attributes and are derived from a common ancestor. In *CsPR1* two CRISP motifs were detected (CRISP1 and CRISP2). CRISP (cysteine-rich secretory proteins) are evolutionary related extracellular proteins from eukaryotes. Rodent sperm-coating glycoprotein (SCP), Plant pathogenesis proteins of the PR-1 family, Mammalian glioma pathogenesis-related protein (GliPR) are some important members of this family. The exact function of these proteins is not yet known.

Table 4: Calculated secondary elements using SOPMA.

Secondary structure (%)	*CsPR1*
Alpha helix	33.95
Extended strand	14.20
Beta turn	3.70
Random coil	48.15

Table 5: Comparison of models generated using Phyre2 and I-TASSER using MolProbity.

MolProbity comparison	*CsPR1* (Phyre2)	*CsPR1* (I-TASSER)
Poor rotamers(%)	9.02	5.26
Ramachandran outliers (%)	11.25	3.75
Ramachandran Favored (%)	79.38	90.00
Ramachandran allowed (%)	88.8	96.2
Cβ deviation>0.25A°	75	10.27
Residue with bad bond (%)	32.1	0.00
Residue with bad angle (%)	36.42	1.12

Conclusion

We have successfully carried out heterologous expression and in *E.coli* strain *BL21(DE3)pLysS* and purification using Dynabeads® His-Tag Isolation & Pulldown. Physicochemical characterization were performed by computing theoretical pI, molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and GRAVY. Disulphide linkages, motifs and profiles were also predicted. Secondary structure analysis revealed that random coils dominated among secondary structure elements in *CsPR1* followed by alpha helix in case of *CsPR1* .The modeling of the three dimensional structure of the protein was performed by 3D homology modeling programs(Phyre2 and I-TASSER) and validated using MolProbity and RAMPAGE. Purified *CsPR1* proteins can be used for fungal bioassay studies and functional characterization and in silico developed models will provide a good foundation for analysis of detailed structural features of *CsPR1* protein.

Acknowledgements

The authors gratefully acknowledged the funding and support from Department of Biotechnology, Govt. of India.

References

Alexander D, Goodman RM., Gut-Rella M et al (1993) Increased tolerance to two oomycete pathogens in transgenic tobacco expressing pathogen-related protein 1a.Proc. Natl. Acad. Sci. USA 90:7327-7331.

Anand A, Lei ZT, Summer LW, Mysore KS et al (2004) Apoplastic extracts from a transgenic wheat line exhibiting lesion-mimic phenotype have multiple pathogenesis-related proteins that are antifungal. Plant-Microbe Interact 17:1306-1317.

Chen VB, Arendall WB, Headd JJ, Keedy DA et al (2010) MolProbity: all-atom structure validation for macromolecular crystallography Acta Cryst D66:12-21.

Datta K, Velazhahan R, Oliva N et al (1999) Over-expression of the cloned rice thaumatin-like protein (PR-5) gene in transgenic rice plants enhances environmental friendly resistance to Rhizoctonia solani causing sheath blight disease. Theor. Appl. Genetics 98:1138-1145.

Davis IW, Leaver-Fay A, Chen VB et al (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35 :W375-W383.

Fernandez C, Szyperski T, Bruyere T et al (1997) NMR solution structure of the pathogenesis-related protein p14a. J Mol Biol, 266:576-593.

Gasteiger E, Hoogland C, Gattiker A et al (2005) Protein Identification and Analysis Tools on the ExPASy Server;(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press.571-607.

Hirokawa T, Boon-Chieng S, and Mitaku S (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics 14: 378-9.

Kelley LA and Sternberg MJE (2009) Protein structure prediction on the web: a case study using the Phyre server, Nature Protocols 4:363 – 371.

Kovaleva V, Kiyamova R, Cramer R et al (2009) Purification and molecular cloning of antimicrobial peptides from scots pine seedlings. Peptides,30(12):2136-2143.

Kovaleva V, Krynytskyy H, Gout I, Gout R (2010) Recombinant expression, affinity purification and functional Characterization of scots pine defensin 1. Appl Microbiol Biotechnol., 89(4):1093-1101

Niderman T, Genetet I, Bruyere T et al (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.

Rauscher M, Adam, AL, Wirtz S et al (1999) PR1 protein inhibits the differentiation of rust infection hyphae in leaves of acquired resistant broad bean. Plant J 19:625-633.

Roy A, Kucukural A, Zhang Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols 5: 725-738.

Roy A, Yang J, Zhang Y. (2012) COFACTOR: An accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Research, 40:W471-W477.

Sahay A. Shakya M(2010) In silico analysis and homology modeling of antioxidant proteins of spinach. J Proteomics Bioinform 3:148-154.

Sarowar S, Kim YJ, Kim EN et al (2005) Overexpression of a pepper basic pathogenesis-related protein 1 gene in tobacco plants enhances resistance to heavy metal and pathogen stresses. Plant Cell Reports 24:216-224.

Schultheiss H, Dechert C, Kiraly L et al (2003) Functional assessment of the pathogenesis-related protein PR-1b in barley. Plant Sci 165:1275-1280.

Stothard P (2000) The Sequence Manipulation Suite: Java Script programs for analyzing and formatting protein and DNA sequences. Biotechniques 28:1102-1104.

Takahashi W, Fujimori M, Miura Y, Komatsu T, Nishizawa Y, Hibi T, Takamizo T (2005) Increased resistance to crown rust disease in transgenic Italian ryegrass (Lolium multiflorum Lam.) expressing the rice chitinase gene. Plant Cell Rep 23:811-818.

Tonon, C, Guevara G, Oliva,C, Daleo G (2002) Isolation of a potato acidic 39 kDa β-1,3-glucanase with antifungal activity against *Phytophthora infestans* and analysis of its expression in potato cultivars differing in their degrees of field resistance. J. Phytopathol 150: 189-195.

van Loon LC, Rep M, Pieterse CMJ (2006) Significance of inducible defense-related proteins in infected plants. Annu Rev Phytopathol 44:135–162

Van Loon, LC, Van Strien EA (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiol. Mol. Plant Pathol 55: 85-97.

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics:9:40