NBU Journal of
Plant Sciences

Vol. 3, March 2009
ISSN No 0974 0937

Molecular modelling of a pathogenesis related protein from *Solanum tuberosum*

S Thakur, S Sur, D Bose, AK Goyal, T Mishra, R Rai, M Bhattacharya, AK Bothra[1] and A Sen*
Molecular Genetics Laboratory, Department of Botany, University of North Bengal Siliguri 734013, India
[1]Department of Chemistry, Raiganj College, Raiganj

Abstract
The pathogenesis related proteins are an important group of proteins produced in plants in response to infection by phytopathogens. The PR protein of *Solanum tuberosum* form an integral part of the host defence system. The unavailability of the crystal structure of the PR protein of *Solanum tuberosum* prompted us to undertake molecular modeling technique to look into the active sites and infer upon the structure function relationship. The model was built using 1CFE as template. The functionality was studied. The model offers a reliable base to start X-crystallography and NMR based studies.

Keywords: PR proteins, *Solanum tubersosum* molecular modeling, structure prediction, 3D model

The plant system is prone to attack by a number of fungal pathogens. In order to withstand the attack, a number of proteins with putative protective functions are generated as a defense strategy (Datta and Muthukrishnan, 1999). A group of proteins called pathogenesis related proteins (PR) proteins are induced specifically in pathological related situations suggestive of a general role for these proteins in adaptation to biotic stress conditions (Loon and Strein, 1999). These proteins are not pathogen specific rather they are determined by the type of reaction of the host plant (Chakraborty, 2008). Initially five main classes of pathogenesis related proteins (PRps) were typified by both biochemical and molecular-biological techniques in tobacco (Loon and Strein, 1999) however seventeen families of PRps have been recognized till date (Chakraborty, 2008). The infection of potato (*Solanum tuberosum*) leaves with the late blight fungal pathogen *Phytophthora infestans*, or treatment with fungal elicitor, results in huge aggregation of pathogenesis-related (PR) proteins in the extracellular leaf space (Hoegen *et al.*, 2002). These proteins are known to be resistant to digestion by proteolytic enzymes, highlighting the intrinsic stability of these proteins in harsh environments (Fernandez *et al.*, 1997). Although some work related to X-ray crystallography and NMR has been performed on the structures of some PR proteins associated with the golgi complex (Groves *et al.*, 2004), ripe tomato fruits (Ghosh and Chakrabarti, 2005), (Fernandez *et al.*, 1997) and tobacco (Koiwa *et al.*, 1999), almost no work has been executed on the pathogenesis related proteins of *Solanum tuberosum* (potato). Since no crystal structure for the PR proteins of *Solanum tuberosum* are available in the databases, the availability of the raw protein sequence of *Solanum tuberosum* provide a good opportunity to start structure based studies on this protein to gain insights into their functions.

*Corresponding author:
E-mail: senarnab_nbu@hotmail.com

The three dimensional structure of a protein gives information about its function. It is often difficult to ascertain a structure experimentally using X-ray crystallography or NMR. However, computational techniques have become reliable for the creation of 3D structures (Othman *et al.*, 2007). Molecular modeling is a dependable technique that can predict the three-dimensional structure of a protein compared to that obtained at low-resolution by experimental means (Martin-Renom *et al.*, 2000). Three-dimensional structure of any protein is quite useful in making out its functional details (Paramsivasan *et al.*, 2006). In this work, an effort has been made to build the three-dimensional model of a PR protein from *Solanum tuberosum*. The study is expected to infer upon the structure-function relationships of PR protein. It is likely that the homology derived model may serve as a support base for structure based experimental studies.

Materials and Methods

The amino acid sequence of the PR protein of *Solanum tuberosum* bearing EMBL (European Molecular Biology Laboratory) accession numbers AJ250136.1 were obtained from the NCBI (National Center for Biotechnology Information) database (www.ncbi.nlm.nih.gov). The 3D structure of the protein was not available in Protein Data Bank (http://www.rcsb.org/pdb/home/home.do), as a result the work of creating the 3D model of the protein initiated.

In the first step protein structures linked to the target sequence that will be used as template was selected (Centeno et al., 2005). Position specific iterative blast (PSI-BLAST) (Altschul *et al.*, 1997) (http://www.ncbi.nlm.nih.gov/blast/) was carried out against database specification of PDB proteins to detect similarity. Template selection was based upon the nature of the experimental template structure especially, its environmental and functional resemblance as well as phylogenetic similarity. Optimal alignment between the target sequence and template was performed using

```
CLUSTAL W (1.83) multiple sequence alignment

1CFE                    -----------------------QNSPQDYLAVHNDARAQVGVGPMSWDANLASRAQNY
1SOP                    MGLFNISLLLTCLMVLAIFHSCDAQNSPQDYLAVHNDARAQVGVGPMSWDAGLASRAQNY
                                              ***************************.********

1CFE                    ANSRAGDCNLIHSGAGENLAKGGGDFTGRAAVQLWVSERPSYNYATNQCVGGKKCRHYTQ
1SOP                    ANGRTGDCNLIHSGAGENLAKGSGDFTGRAAVQLWVGEKPNYNYGTNQCASGQVCGHYTQ
                        ****.***************.************.*:*.***.****..*: * ****

1CFE                    VVWRNSVRLGCGRARCNNGWWFISCNYDPVGNWIGQRPY
1SOP                    VVWRNSVRLGCGRARCNNGWWFISCNYDPVGNWVGQRPY
                        ************************************:*****
```

Fig. 1: Alignment of the template protein and PR protein from *Solanum tuberosum*. Residues marked with * are conserved.

ClustalW 1.83 (Thompson *et al*., 1994) using default settings. The attained alignment was critically evaluated for consistency. Secondary structures were predicted using HNN (Hierarchical Neural Network method) (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_nn.html).

The crude 3D model of the PR protein was built by MODELLER 9v4 program (Sali and Blundell, 1993). The technique is based upon the satisfaction of the spatial restraints acquired from the alignment (Centeno et al., 2005). The model obtained often contains errors and become crucial when concerned residues are linked with protein function (Centeno *et al*., 2005). To overcome this refinement is indispensable. In this process the constructed model was subjected to constraint energy minimization with a harmonic constraint of 100 kJ/mol/Å^2, using the steepest descent (SD) and conjugate gradient (CG) method to remove any existing bad sectors between the protein atoms. The computations were performed *in vacuo* with the GROMOS96 43B1 factors using the Swiss-Pdb Viewer package (http://expasy.org/spdv/program/spdv37sp5.zip) (Kaplan and Littlejohn, 2001). Hydrogen bonds were ignored.

In order to test the internal quality and reliability, the model was subjected to some evaluations. ProSA (Wiederstein and Sippl, 2007) was performed to judge the accuracy and of the modelled structures and check the 3D models for potential errors. VERIFY3D (Eisenberg *et al*., 1997) was used to authenticate the refined structures. The refined model of PR protein was evaluated for its backbone conformation using Ramachandran plot (Ramachandran et al., 1963). SAVES (Structure analysis and verification server) (http://nihserver.mbi.ucla.edu/SAVS/) was used to verify of the models with ERRAT. Presence of ligand binding pockets in the structures was predicted using CASTp server (Dundas *et al*., 2006). ProFunc (http://www.ebi.ac.uk/thronton-srv/databases/ProFunc) (Laskowski et al., 2005) server was used to identify the functional region in the protein. Deficiency of any data on the site-directed mutagenesis of PR proteins prompted us to carry out site-directed-mutagenesis predictions using the SDM (http://www-cryst.bioc.cam.ac.uk/~sdm/sdm.php) server. The stability changes related with possible mutations were judged by I-Mutant 2.0 (http: //gpcr.biocomp.unibio.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi.).

Intrinsic dynamics of the proteins are vital for garnering information regarding their functions (Yang *et al*., 2006). Intrinsic dynamics studies were carried out using WEBnm (http://www.bioinfo.no/tools/normalmodes) program (Hollup *et al*., 2005) indicating the slowest modes and related deformation energies; ElNemo (http://igs-server.cnrs-mrs.fr/elnemo/index.html.) server (Suhre and Sanejouand, 2004) calculating the normal mode analysis of the protein contributing to the corresponding movement and MolMovDB (http://molmovdb.org/) determining the five lowest frequency modes (Alexandrov *et al*., 2005). Solvent accessibility graphics of the amino-acid residues in the modeled PR protein was studied using ASA-view (http://gibk26.bse.kyutech.ac.jp/~shandar/netasa/asaview/) software (Ahmad *et al*., 2004).

Results and Discussion

The most appropriate template was found to be 1CFE. This is a NMR solution structure of a PR protein, P14a from *Lycopersicon esculentum*. This protein is 135 amino-acids in length. PR protein of *Solanum tuberosum* revealed 90% identity with P14a PR proteins from *Lycopersicon esculentum*. There was a high degree of conservation in the amino-acid residues amongst the two proteins. As anticipated the hydrophilic residues occupied the surface whereas the hydrophobic residues remained within the interior.

Figure 1 shows the alignment of the template and the target protein. The major conserved regions lay between residues 25-51, 53-64, 66-82, 84-96 and 117-154. The conserved nature of the regions are marked by * in the figure. HNN (Hierarchial Neural Network) analysis of the secondary structure illustrated that the alpha helix portion consisted of 39 (24.53%), extended strand 28 (17.61%) and random coil 92 (57.86%) residues. The helix and sheets remain spread all through the protein structure. The modeled structure consists of 1216 atoms and 1250 bonds. It had a molecular formula of $C_{754}H_{1136}N_{226}O_{225}S_{11}$, molecular weight of 17308.1 Da and a molecular volume 10240.1. It had a net partial charge of 3.475.

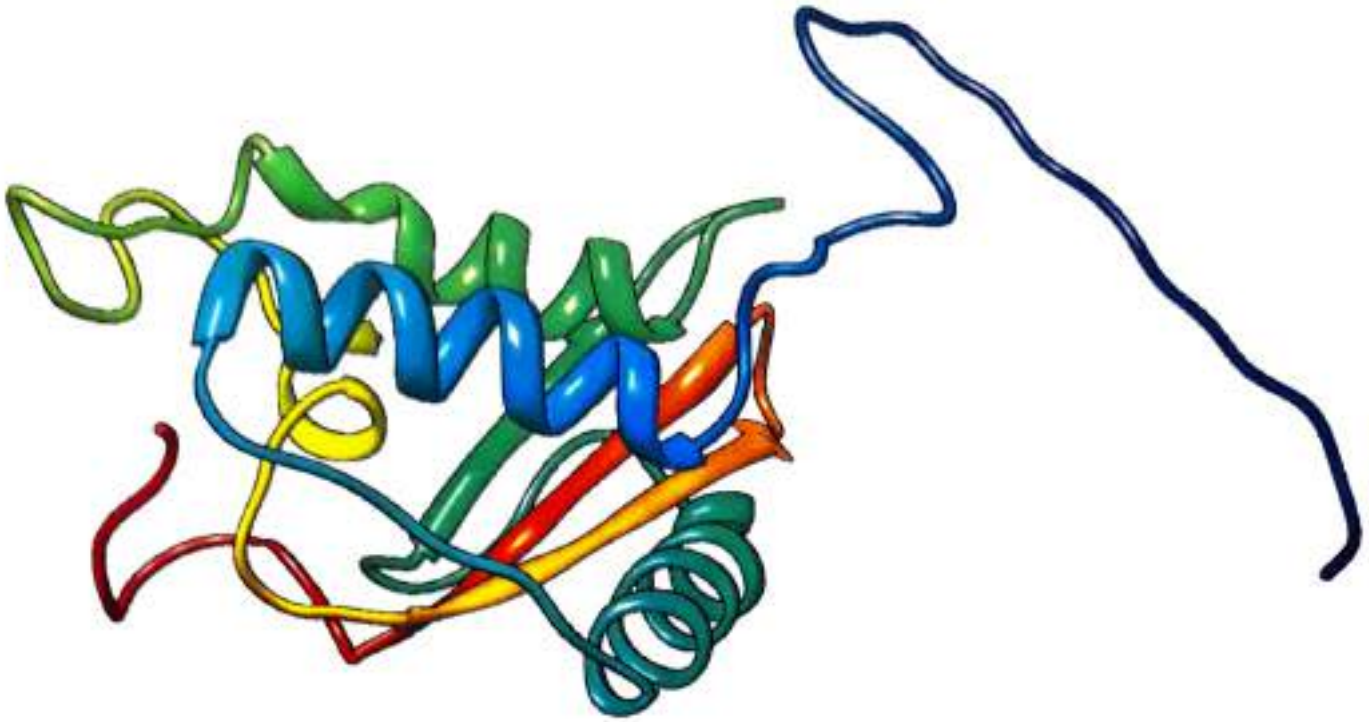

Fig. 2: Three dimensional structure of the PR protein from *Solanum tuberosum*

Figure 2 shows the 3D structure of the modeled protein. In the PR protein of *Lycopersicon esculentum* the following residues Ser 3, Gln 5, His 48, Ser 49, His 903, Arg 100 and Asn 114 are strictly conserved regions of which His 48, Ser 49 and His 93 are functionally important sites. In our modeled protein the functionally important residues i.e., active sites corresponded to histidine in 72 position, serine in 73 position and histidine in 117 position. CASTp revealed the presence of 16 pockets for binding regions with varying area and volume. These pockets play an important role in the protein functionality. ProFunc analysis revealed the presence of 3 nests located in the structure. These nests are structurally crucial motifs forming a concave depression which can serve as a binding site for an atom or a group of atoms. Our modeled protein had 18 matching sequence in the PDB entries. BLAST search of the protein revealed 50 matching sequences in UniProt. No potential helix-turn helix DNA binding motifs were identified. Analysis of the binding sites revealed the presence of clefts and cavities in the surface of the proteins. Further analysis of the modeled protein divulged 8724 significant structural matches.

Root mean square deviation (RMSD) calculations of the backbone demonstrated that the PR protein from *Solanum tuberosum* had a deviation of 0.28 angstroms from that of the template and in the C^{α} residues the deviation was 0.19 angstroms. These results of suggest that the deviations between the template protein and the modeled protein are not significant. It is seen that the histidine residues 73 and 117 in our protein are associated with the functionality of the protein and result

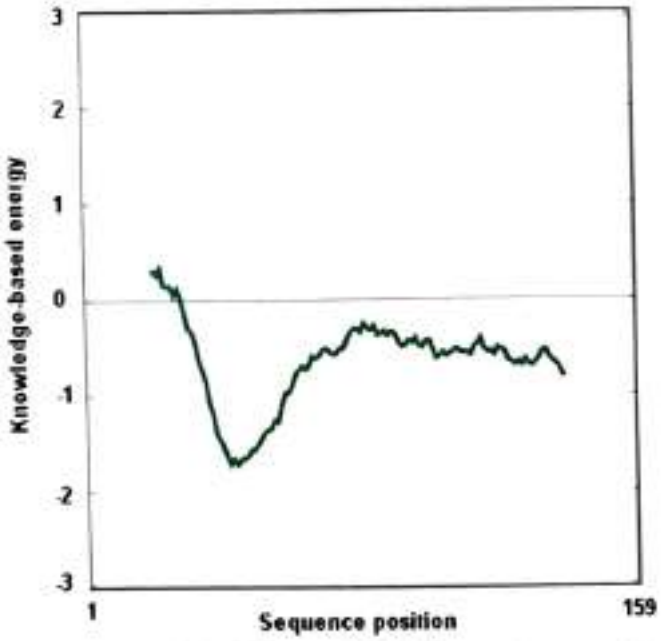

Fig. 3: Energy plot of the protein. Residue energies are averaged over a sliding window are plotted as a function of the central residue in the window.

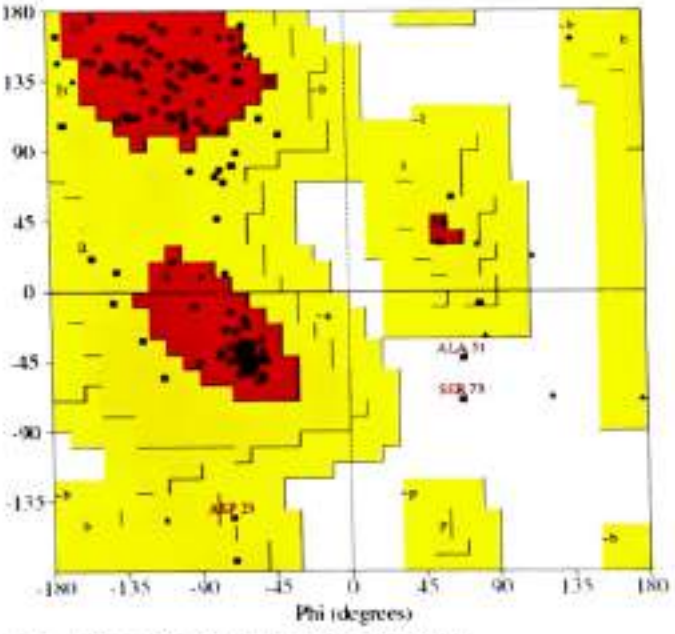

Fig. 4: Ramachandran plot of the PR protein

of the predictions from site-directed mutagenesis demonstrated that, when the residues were replaced with glycine mutations were predicted to be functionally associated. Stability changes were found to be associated with the mutations. The predicted free energy change values (DDG) were found to be -0.97 for position 73 and -0.88 for position 117 suggesting that there is a decrease in stability of the proteins during mutations. Site-directed mutagenesis results confirm that His 73 and His 117 are functionally important residues. This study entail that the residues linked to core functionality are conserved structurally as well as functionally.

The refined model was scrutinized by different programs for assessment of the model quality. Figure 3 demonstrates the result for the monomers of the modeled structures of PR protein from *Solanum tuberosum*. The overall quality score determined by ProSA for our structure is demonstrated in a plot that shows the scores of experimentally determined protein chains available in the Protein Data Bank (PDB). The PR protein of *Solanum tuberosum* had a z score of -4.61. This result points out that the z-score of our model is within the range of scores normally found for proteins of comparable size. Interestingly the energy distribution plot (using a window size of 40 as default) of our modeled protein is below the zero base line. This result of the energy plot suggests that the predicted protein model is quite consistent. The predicted model was confirmed by VERIFY 3D to estimate its correcness. VERIFY 3D revealed that 80.63% of the residues had an averaged 3D-1D>0.2. The plot of average 3D-1D profile score of residues of our predicted protein model signifies that our model reliable. ERRAT evaluation of the protein structure revealed a quality factor of 74.26.Generally the accepted range of a high quality model is <50 (Colovos and Yeates, 1993). In our case, the ERRAT score is well within the range of a high quality model.

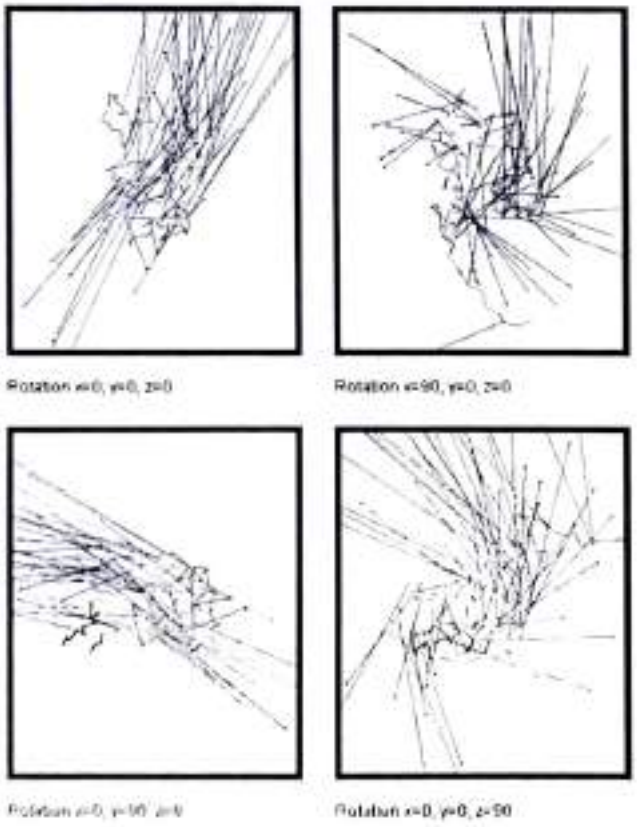

Fig. 5: Vector field representation of the PR protein across different sizes

Figure 4 shows the Ramachandran plot demonstrating the backbone conformations for the modeled protein. On the basis of the analysis of 118 structures having resolution of at least 2.0 angstroms and R factor no greater than 20%, a very good quality model is expected to have more than 90% in the most favored regions of the Ramachandran plot (Rajesh *et al.*, 2007). Ramachandran plot of the PR protein from *Solanum tuberosum* revealed that the numbers of non-glycine and non-proline residues in each of the modeled proteins were 133. Out of this, 110 (82.7%) were in the most favored regions. The allotment of main chain torsion angles phi and psi evidently illustrated that bulk of the amino-acids are in a phi-psi distribution more or less reliable with right handed alpha helices. These results imply that the stereochemical properties and quality of the modeled structures of the in PR protein are quite suitable.

The structural dynamics study of our modeled protein was carried out using the Normal mode analysis (NMA). NMA is a good technique for studying the vibrational and thermal properties of proteins. During normal mode analysis (NMA) the first six modes are associated with global rotation and translation of the system and are ignored (Hollup *et al.*, 2005). Consequently the lowest frequency mode of concern is mode 7 generating the lowest deformation energy. NMA of the PR established that low deformation energies were linked with rigid regions in the protein which have a good possibility of describing domain motions. Normalized atomic displacement analyses were performed for modes 7 to12. These analyses specify the vibrational and thermal properties of the proteins. PR protein from *Solanum tuberosum* had the lowest deformation energy in the seventh mode. B factors from ElNemo analysis were based on the first 100 normal modes. The B-factor analysis of the modeled protein revealed a correlation of 0.106 for 159 C-alpha atoms. Low correlation reveals rigid regions of the protein. A vector field representation of the protein was generated with WEBnm@ server. Figure 5 shows the vector field representation of the protein across different axis. It represents the direction and displacement of the different regions of the protein assesses the possible motion of the protein. Results from ASAVIEW indicated that the accessible residues of the protein were located on the outermost surface of the spiral. These were negatively charged residues and polar uncharged residues whereas most of the hydrophobic residues were confined to the inner rings of the spiral.

Conclusion

The three dimensional structure of the PR protein from *Solanum tuberosum* offer insights into its conformational properties and structure-function relationship. The structures presented here are reliable with their biochemical features. Structural dynamics analysis reveals higher degree of rigidity and refers to

large regions of protein displacement. These results are vital for describing the functioning of the protein. This model offer a base for starting X-ray crystallographic studies.

Acknowledgement

The authors are grateful to the Department of Biotechnology, Government of India, for providing financial help in setting up Bioinformatics Centre, in the Department of Botany, University of North Bengal. ST and AKG is thankful to DBT for providing traineeship and TM is for studentship.

References

Ahmad S, Gromiha M, Fawareh H, Akinori S. 2004. *ASAview*: Database and tools for solvent accessibility representation in proteins. *BMC Bioinform* 5:51

Alexandrov V, Lehnert U, Echols N, Milburn D, Engelman D, Gerstein M. 2005. Normal modes for predicting protein motions: a comprehensive database assessment and associated web tool. *Protein Sci* 14:633-643

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402

Centeno NB, Planas-Iglesias J, Oliva, B 2005. Comparative modelling of protein structure and its impact on microbial cell factories. *Microb Cell Fact* 4:20 Doi: 10.1186/1475-2859-4-20

Chakraborty BN. 2008. Plant defense proteins. *NBU J Plant Sciences* 2: 1-12

Colovos C, Yeates TO. 1993. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 2: 1511–1519.

Datta SK, Muthukrishnan S. 1999. Pathogenesis-related Proteins in Plants. CRC Press, London p. 291

Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang, J. 2006. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 34:116-118

Eisenberg D, Luthy R, Bowie JU 1997. VERIFY3D: Assessment of protein models with three-dimensional profiles. *Metho Enzymol* 277:396-404

Fernandez C, Szyperski T, Bruyere T, Ramage P, Mosinger E, Wuthrich K. 1997. NMR Solution Structure of the Pathogenesis-related Protein P14a. *J Mol Biol* 266:576-593

Ghosh R, Chakrabarti C. 2005. Crystallization and preliminary X-ray diffraction studies of NP24-I, an isoform of a thaumatin-like protein from ripe tomato fruits. *Acta Cryst* F61: 806-807

Groves MR, Kuhn A, Hendricks A, Radke S, Serrano SL, Helms JB, Sinning I. 2004. Crystallization of a Golgi-associated PR-1-related protein (GAPR-1) that localizes to lipid-enriched microdomains. *Acta Cryst* D60: 730-732

Hoegen E, Stromberg A, Pihlgren U, Kommbrink E. 2002. Primary structure and tissue-specific expression of the pathogenesis-related protein PR-1b in potato. *Mol Plant Pathol* 3: 329-345

Hollup SM, Salensminde G, Reuter N. 2005. WEBnm@: a web application for normal mode analysis of proteins. *BMC Bioinforms* 6: 1-8

Kaplan W, Littlejohn TG 2001. Swiss-PDB Viewer (Deep View). *Brief Bioinform* 2: 195-197

Koiwa H, Kato H, Nakatsu T, Oda J, Yamada Y, Sato F. 1999. Crystal structure of tobacco PR-5d protein at 1.8 Å resolution reveals a conserved acidic cleft structure in antifungal thaumatin-like proteins. *J Mol Biol* 286:1137-1145

Laskowski RA, Watson JD, Thronton JM 2005. ProFunc: a server for predicting protein function from structure. *Nucleic Acids Res* 33:89-93

Loon LCV, Strein EAV. 1999. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiol Mol Plant Pathol* 55: 85-97

Martin-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. 2000. Comparative protein structure modeling of genes and genomes. *Ann Rev Biophys Biomol Struc* 29:291-325

Othman R, Wahab HA, Yosof R, Rahman NA. 2007. Analysis of secondary structure predictions of dengue virus type 2 NS2B/NS3 against crystal structure to evaluate the predictive power of the in silico methods. *InSilico Biol* 7: 215-224

Paramsivasan R, Sivaperumal R, Dhnanjeyan KJ, Thenmozhi, V, Tyagi, BK 2006. Prediction of 3-dimensional structure of salivary odorant-binding protein-2 of the mosquito *Culex quinquefasciatus*, the vector of human lymphatic filariasis. *InSilico Biol* 7: 1-6.

Rajesh R, Gunasekaran K, Muthukumaravel S, Balaraman K, Jambulingam P. 2007. In Silico analysis of voltage-gated sodium channel in relation to DDT resistance in vector mosquitoes. *InSilico Biol* 7:413-421

Ramachandran GN, Ramakrishnan C, Sasisekharan, V. 1963. Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7:95-99

Sali A, Blundell, TL 1993. Comparative protein modelling by satisfaction of spatial restraints *J Mol Biol* 234:283-291

Suhre K, Sanejouand YH 2004. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res* 32: w610-614

Thompson JD, Higgins DG, 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 Res* 22:4673-4680

Wiederstein M, Sippl, MJ 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35: 407-410

Yang Lee-We., Rader AJ, Liu X, Jursa CJ, Chen SC, Karimi HA, Bahar I. 2006. oGNM: online computation of structural dynamics using the Gaussian Network Model. *Nucleic Acids Res* 34: w24-w31□