BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access

www.bioinformation.net

Hypothesis

Volume 8(5)

Prediction of the three-dimensional structure of serine/threonine protein kinase pto of *Solanum lycopersicum* by homology modelling

Salam Pradeep Singh[1]*, Sarangthem Vivek[2], Rajib Lochan Bezbaruah[1], Madhumita Barooah[2]

[1]Bioinformatics Infrastructure Facility, Biotechnology Division, North-East Institute of Science & Technology (CSIR), Jorhat- 785006, Assam; [2]Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat, Assam; Salam Pradeep – Email: salampradeep@gmail.com; *Corresponding author

Received January 03, 2012; Accepted January 07, 2012; Published March 17, 2012

Abstract:
The resistant gene Pto of *Solanum lycopersicum* interacts with the *avr Pto* gene product of the bacterial pathogen *Pseudomonas syringae* pv tomato to launch a cascade of molecular events that triggers the hypersensitive disease-resistance response in tamato. The paper describes attempts to predict the structure of Pto encoding a serine/threonine protein kinase to understand the mechanism and function. A three-dimensional model based on the crystal structure of effect protein *Avr ptob* complexed with Kinase Pto and bacterial effector protein *Avrpto* was generated using Modeller9v7. We adopted different modelling approaches for our study, Intialy, we generated a model based on a single template protein and then a model based on multiple templates. The models generated through these approaches were further assessed with ANOLEA energy assessment, Ram Page server and PROCHECK for stereochemistry and geometry check. Comparative analysis suggested that the model generated was better than the templates. This study paves the way for generating computer molecular models for proteins whose crystal structures are not available and which would aid in studying protein-protein interactions.

Keywords: Pto, Modeller9v7, ANOLEA

Background:
Pto encodes a cytoplasmic serine/threonine protein kinase **[1]** that interacts with the avr *Pto* gene product of the bacterial pathogen *Pseudomonas syringae* pv tomato **[2].** The interaction appears to launch a cascade of molecular events that triggers the hypersensitive disease-resistance response **[3].** These experiments provided the first molecular confirmation of Flor's (1956) gene-for-gene hypothesis that predicted a host resistance (R) gene encodes a receptor that recognizes a ligand encoded or produced by the corresponding Avr gene. Protein kinases are a group of enzymes with possess a catalytic subunit which transfers the gamma phosphate from nucleotide triphosphates (often ATP) to one or more amino acid residues in a protein substrate side chain, resulting in a conformational change affecting protein function. The enzymes classified into two broad classes based on substrate specificity are serine/threonine specific and tyrosine specific **[4].** Protein kinase also play a major role in a multitude of cellular processes, including division, proliferation, apoptosis, and differentiation **[5].** The catalytic subunits of protein kinase are highly conserved, and several structures have been generated **[6, 7].**

Methodology:
The amino acid sequence of serine/threonine protein kinase *Pto* of *Solanum lycopersicum* (311 amino acids).was retrieved (Accession No: AAB47421) from the NCBI Genbank database (http:www.ncbi.nlm.gov). A BLAST **[8]** search (PDB-BLAST) **[9]**

was performed with the amino acid sequence of serine/threonine protein kinase *pto*. The PDB-BLAST resulted with two best entries: Crystal Structure of effect protein AvrptoB complexed with Kinase *Pto* (PDB ID: 3HGK (Chain A)) **[10]** and bacterial effector protein Avrpto (PDB ID: 2QKW (Chain B)) **[11].** These two proteins have a common sequence identity of 83% (269/321) with serine/threonine protein kinase *Pto*. Additionally, these two proteins have a crystal structure resolution of 3.30 Angstrom (3HGK) and 3.20 Angstrom (2QKW) respectively to make them excellent reference templates for performing homology modelling. The present study, we approached three different modelling approaches using the Modeller 9v7 **[12]** software for modelling the three-dimensional structures of serine/threonine protein kinase *Pto*. First we aligned the serine/threonine protein kinase *Pto* with the sequence of the template protein 3HGK (Accession No: 3HGKA) using the ALIGN2D command of Modeller 9v7 and built a model. In next phase we aligned the serine/threonine protein kinase *Pto* with the sequence of the template protein 2QKW (Accession No: 2QKWB) using Modeller 9v7 to build another model. And finally, we created a model using multiple templates of Chain A of 3HGK and Chain B of 2QKW. The rough models generated were further refined using loop.py script in Modeller9v7.

Discussion:

Sequence analysis, alignment and Model generation

Initially a structural analysis of the two template structures 3HGK and 2QKW were performed using ICM Molsoft Browser www.molsoft.com/icm_browser.html which revealed that crystal structure of Chain A of 3HGK consisted with only 288 amino acid residues (LYS31-GLU312) whereas crystal structure of Chain B of 2QKW consisted of 292 amino acid residues (PRO30-ILE321). Therefore, we manually retrieved out the sequence form MET01 up to PHE29 of our target sequence and sequences form PRO25-ILE311 (a total of 287 residues) to perform our homology modelling and generated the best fitting structure. For convenience sake, PRO25 was designated as the first residue i.e. PRO01.

Using 3HGK as a template, we created an alignment (using the ALIGN2D command of Modeller9v7) between the sequence of serine/threonine protein kinase *Pto* (pto) and the sequence of Chain A of 3HGK and generated five different structures in 185.80 seconds using the Modeller 9v7 program. Out of the 5 structures generated the best structure was chosen based on the evaluation of molecular Probability Density Function, DOPE Score **[13]** and GA341 **[14].** The best structure gives a molecular Probability Density Function 2520.30786, DOPE score of -29699.56836 and GA341 score of 1.0. In the next phase, 2QKW was used as a template to create another alignment between the sequence of serine/threonine protein kinase *Pto* (pto) and the sequence of Chain B of 2QKW), we again generated another five 5 structures and choose the best structure with molecular Probability Density Function 2494.73193, DOPE score -30093.53711 and GA341 score 1.0 Finally using both templates 3HGK and 2QKW we created an alignment between serine/threonine protein kinase *Pto*, 3HGK and 2QKW. Using this alignment we generated another five structures and choose the best structure with molecular Probability Density Function 9821.59473.

***Model refinement and assessment*:**

The best structure generated using 3HGK as a template was evaluated with ANOLEA energy assessment using the ANOLEA server **[15, 16]** showing a total non-local energy of -742. The structure was further assessed for Ramachandran Plot with the RAM page server **[17, 18]** showing 250 residues in the favoured regions, 27 residues in the allotted regions and 8 residues outside the outlier region. The ANOLEA energy assessment revealed that the energy of few loop regions were high with positive values, therefore, a loop refinement was performed **[19]** that generated a structure with a total non-local energy of -902 with ANOLEA assessment. It showed 261 residues in the favoured regions, 22 residues in the allotted regions and 2 residues outside the outlier region in the Ramachandran plot. ANOLEA energy assessment **[15, 16]** of the best structure generated by using 2QKW was evaluated which showed a total non-local energy of -871. The structure was further assessed for Ramachandran Plot with the RAM page server **[17, 18]** showing 255 residues in the favoured regions, 22 residues in the allotted regions and 8 residues outside the outlier region. Further loop refinement for the loop regions where the energies were high was performed to generate a structure with a total non-local energy of -1110 with ANOLEA assessment and showing 271 residues in the favoured regions, 13 residues in the allotted regions and 1 residues outside the outlier region in the Ramachandran plot.

Similarly, the best structure generated using 3HGK and 2QKW as template was evaluated with ANOLEA energy assessment **[15, 16]** showing a total non-local energy of -886. The structure was further assessed for Ramachandran Plot with the RAM page server **[17, 18]** showing 255 residues in the favoured regions, 20 residues in the allotted regions and 10 residues outside the outlier region. As the ANOLEA energy assessment displayed few loop regions with high energy, we performed loop refinement for the loop regions where the energies were high and generated a structure with a total non-local energy of -1043 with ANOLEA assessment and showing 257residues in the favoured regions 19 residues in the allotted regions and 9 residues outside the outlier region in the Ramachandran plot. A detailed comparative statics on ANOLEA energy assessment and Ramachandran analysis of the three models generated and their Z-score **[20]** based on the templates 3HGK, 2QKW and 3HGK-2QKW in comparison to the templates strucures 3HGK and 2QKW is shown on **Table 1 (see supplementary material**)

Structural Comparison:

The assessment of homology models' accuracy is straightforward when the experimental structure is known. In our study we used the most common method of comparing two protein structures i.e. the root-mean-square deviation (RMSD) metric which measure the mean distance between the corresponding atoms of two superimposed structures. We superimposed the backbone of the built proteins and computed their RMSD - 3HGK based model over the template 3HGK (RMSD 1.33), 2QKW based model over the template 2QKW (RMSD 2.17) and 3HGK-2QKW based model over template 3HGK and 2QKW (RMSD 1.72) using ICM Molsoft Browser www.molsoft.com/icm_browser.html **(Figure 1).**

Additionally we performed a comparative analysis on the secondary structure of the three built models and found out the

difference in the secondary structures. There is a major similarity between all the helix.

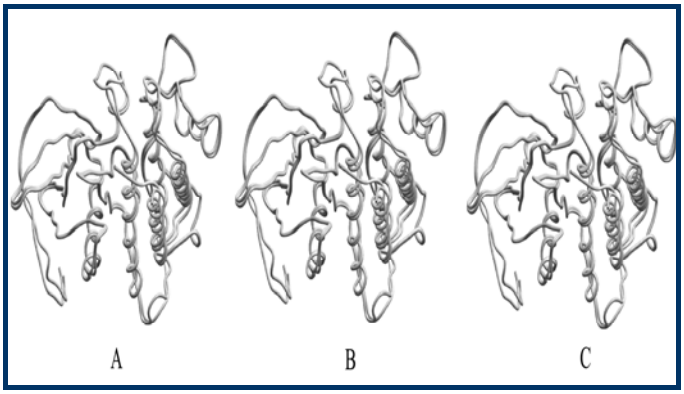

Figure 1: Superimposition of the backbone structures of the three predicted proteins after refinement and computed their RMSD using ICM Molsoft Browser: **(A)** 3HGK based model; RMSD: 1.33 A (1132 atoms superimposed) over the template protein 3HGK; **(B)** 2QKW based model; RMSD: 2.17 (1148 atoms superimposed) over the template protein 2QKW; **(C)** 3HGK-2QKW based model; RMSD 1.72 A (1148 atoms superimposed) over the two templates 3HGk and 2QKW

Conclusions:

The three-dimesnsional model of serine/threonine kinase protein has been generated based on the homology of Crystal Structure of effect protein *Avr ptob* complexed with Kinase Pto (PDB ID: 3HGK (Chain A)) and bacterial effector protein Avrpto (PDB ID: 2QKW (Chain B)). The model generated by using the template 2QKW proved to be the best model generated as compared to the other templates based on the result in model assessment **Table 1 (see supplementary material).** The model contains a compact and a topology common to the template protein. These models have a very similar topology compared to its corresponding template 2QKW. The secondary structural analysis also revealed that the model has a common helix and sheets regions. Based on the promotif **[21]** analysis, both the template 2QKW and the model built by 2QKW as a template consist of 10.5 percent Strands, 30% alpha helix and 2% 310 Helix. The present study would aid in studying protein-protein interactions in future.

Protein structure accession number:

The predicted 3D structures of serine/threonine kinase protein domains were submitted to the Protein Model Database (PMDB) **[22]** and assigned the PMDB ID PM0077821.

Acknowledgement:

S.P.S thanks Director, NEIST, Jorhat for providing Project Assistant Level-II. The authors also thank DBT, Govt of India & CSIR for providing Bioinformatics Infrastructure Facility to carry out this work.

Competing interests:

The authors declare that they have no competing interests

References:

[1] Martin GB *et al. Science.* 1993 **262**: 1432 [PMID: 7902614]
[2] Scofield SR *et al.* Science. 1996 **274**: 2063 [PMID: 8953034]
[3] Tang X *et al.* Science. 1996 **274**: 2060 [PMID: 8953033]
[4] Quinn AM *et al. Science.* 1988 **241**: 42 [PMID: 3291115]
[5] Manning G *et al. Trends Biochem Sci.* 2002 **27**: 514 [PMID: 12368087]
[6] Stout TJ *et al. Curr Pharm Des.* 2004 **10**: 1069 [PMID: 15078142]
[7] Li B *et al. Comb Chem High Throughput Screen.* 2004 **7**: 453 [PMID: 15320712]
[8] Altschul SF *et al. J Mol Biol.* 1990 **215**: 403 [PMID: 2231712]
[9] Sussman JL *et al. Acta Crystallogr D Biol Crystallogr.* **54**: 1078 [PMID: 10089483]
[10] Dong J *et al. Plant Cell.* 2009 **21**: 1846 [PMID: 19509331]
[11] Xing W *et al. Nature.* 2007 **449**: 243 [PMID: 17694048]
[12] Sali A & Blundell TL, *J Mol Biol* 1993 **234**: 779 [PMID: 8254673]
[13] Shen MY & Sali A, *Protein Sci.* 2006 **15**: 2507 [PMID: 17075131]
[14] Melo F *et al. Protein Sci.* 2002 **11**: 430 [PMID: 11790853]
[15] Melo F & Feytmans E, *J Mol Biol.* 1998 **277**: 1141 [PMID: 9571028]
[16] Melo F *et al. Proc Int Conf Intell Syst Mol Biol.* 1997 **5**: 187 [PMID: 9322034]
[17] Ramachandran GN *et al. J Mol Biol.* 1963 **7**: 95 [PMID: 13990617]
[18] Lovell SC *et al. Proteins.* 2003 **50**: 437 [PMID: 12557186]
[19] Fiser A *et al. Protein Sci.* 2000 **9**: 1753 [PMID: 11045621]
[20] Melo F *et al. Protein Sci.* 2002 **11**: 430 [PMID: 11790853]
[21] Hutchinson EG & Thornton JM, *Protein Sci.* 1996 **5**: 212 [PMID: 8745398]
[22] Castrignano T *et al. Nucleic Acids* Res. 2006 **34**: D306 [PMID: 16381873]

Edited by P Kangueane

Citation: **Singh** *et al.* Bioinformation 8(5): 212-215 (2012)

BIOINFORMATION

open access

Supplementary material:

Table1: Comparison of the built proteins

SN	Model Name	Non local energy	High Energy Amino acids	Z-Score	Residues in favoured regions	Residues in allotted regions	Residues in outlier region
1	Pto-3HGKA	902	47	2.79	261 (91.6%)	22 (7.7%	2 (0.7%)
2	Pto-2QKWB	-1110	28	2.23	271 (95.1%)	13 (4.6%)	1 (0.4%)
3	Pto-3HGKA-1QKWB	-1043	47	2.46	257 (90.2%)	19 (6.7%)	9 (3.2%)
4	3HGKA	-502	37	2.60	215 (76.2%)	50 (17.7%)	17 (6.0%)
5	2QKWB	-1390	37	1.60	240 (83.9%)	33(11.5%)	13 (4.5%)