Insights from molecular modeling and dynamics simulation of pathogen resistance (R) protein from brinjal

Dipty Shrivastava[1], Vikrant Nain[2], Shakti Sahi[2], Anju Verma[3], Priyanka Sharma[1], Prakash Chand Sharma[4], Polumetla Ananda Kumar[1]*

[1]National Research Centre on Plant Biotechnology, IARI, New Delhi -110012, India; [2]School of Biotechnology, Gautam Buddha University, Greater Noida- 201308, India; [3]School of Biological Sciences, University of Missouri, Kansas City MO- 64110, USA; [4]School of Biotechnology, GGS Indraprastha University, Delhi-110403, India; Polumetla Ananda Kumar: Email-polumetla@hotmail.com; *Corresponding author.

Abstract:
Resistance (R) protein recognizes molecular signature of pathogen infection and activates downstream hypersensitive response signalling in plants. R protein works as a molecular switch for pathogen defence signalling and represent one of the largest plant gene family. Hence, understanding molecular structure and function of R proteins has been of paramount importance for plant biologists. The present study is aimed at predicting structure of R proteins signalling domains (CC-NBS) by creating a homology model, refining and optimising the model by molecular dynamics simulation and comparing ADP and ATP binding. Based on sequence similarity with proteins of known structures, CC-NBS domains were initially modelled using CED-4 (cell death abnormality protein) and APAF-1 (apoptotic protease activating factor) as multiple templates. The final CC-NBS structural model was built and optimized by molecular dynamic simulation for 5 nanoseconds (ns). Docking of ADP and ATP at active site shows that both ligand bind specifically with same residues and with minor difference (1 Kcal/mol) in binding energy. Sharing of binding site by ADP and ATP and low difference in their binding site makes CC-NBS suitable for working as molecular switch. Furthermore, structural superimposition elucidate that CC-NBS and CARD (caspase recruitment domains) domain of CED-4 have low RMSD value of 0.9 A[0]. Availability of 3D structural model for both CC and NBS domains will help in getting deeper insight in these pathogen defence genes.

Background:

Plants are constantly attacked throughout their life cycle by a range of phytopathogens that includes viruses, mycoplasma, bacteria, fungi, nematodes, protozoa and other parasites. In the first line of defence, presence of an invading pathogen is detected by host immune system through pathogen–associated molecular patterns (PAMPs), such as bacterial flagellin, lipopolysaccharides and fungal-oomycete cellulose-binding elicitor proteins [1]. However, some pathogens have developed a way to evade PAMPs mediated pathogen defence mechanisms. For instance, *Agrobaterium tumefaciens* circumvents recognition by *Arabidopsis* flagellin receptor by domain modification and *Pseudomonas syringae* effectors proteins by suppression of basal immune system in *Arabidopsis* [2, 3]. In the second line of defence 'gene for gene resistance' gets activated which is governed by resistance (R) genes that recognize pathogen race specific effecter (Avr) proteins [4]. The R gene family is one of the largest known plants gene families that mediate elicitor recognition and activate downstream signalling response leading to disease resistance by localized cell death (hypersensitive response) [5]. R proteins monitor the integrity of PAMPs mediated basal immune system and any perturbation of these components by effector proteins could trigger activation of R proteins. Pathogen recognition elicits nucleotide-dependent conformational changes that might induce oligomerisation, thereby providing a scaffold for activation of downstream signalling components [6].

Most of the R proteins contain C-terminal leucine-rich repeats (LRRs), a central nucleotide-binding site (NBS) and a variable amino-terminal domain. The LRRs are mainly involved in recognition, where as the amino-terminal domain determines signalling specificity. The NBS domain that presumably functions as a molecular switch consists of two sub-domains: nucleotide binding (NB) and ARC (common in APAF-1 in human, R protein in plants and CED-4 in *Caenorhabditis elegans*). NB-ARC is a member of the P-loop NTPase domain superfamily that is characterized by conserved nucleotide phosphate-binding motif, also referred to as the Walker A motif, and the Walker B motif. The Walker A and B motifs bind the beta-gamma phosphate moiety of the bound nucleotide (typically ATP or GTP) and the Mg2+ cation, respectively.

Molecular structure of plant R protein homologoues in human APAF-1 and *C. elegans* CED-4 have been resolved by X-ray crystallography [7, 8]. In plants 3D structure of NBS domain with conserved P loop NTPase motif has been modelled for tomato, *Arabidopsis* and flex [9], however such information for coil-coil (CC) domain is lacking. Moreover the previously predicted 3D models are static in nature that do not reflect true and most stable confirmation as observed in dynamic conditions.

In the present work, sequence for CC-NBS domain of R protein was cloned from brinjal (*Solanum melongena*) and a structural model was developed by homology modeling that was further optimized and refined by molecular dynamic simulations for 5 ns. The refined structural model was used for docking with ADP and ATP. Docking studies confirmed specific but differential affinity of ADP and ATP binding at active site.

Methodology:
Target protein:
The brinjal (*Solanum melongena L.*) R protein sequence, for which only one sequence coding for CC-NBS domain is available in database was

Bioinformation

Volume 5

open access

www.bioinformation.net

Issue 8

Hypothesis

retrieved from NCBI GenBank (accession number EU573702) and used for modeling.

Template identification:
For identification of a suitable template and homology modeling different web servers, 3D-Jury [10], ESyPred3D [11], I-TASSER [12], MODEB (http://salilab.org/modweb), PHYRE [13], (PS)2 [14] and SWISS-MODEL [15].

Homology modelling:
Sequence alignment provided by I-TASSER was manually improved and used as an input for homology modeling by HOMMER (http://protein.cribi.unipd.it/homer/) and Swiss-Model [15]. The 3-D model was energy minimized by SPDBV [15] and quality evaluation was done on the basis of Ramachandran's plot and energetics using VERIFY-3D by PROCHECK [16]. Protein 3D structure were aligned and RMSD calculated by TM-Align [17].

Molecular Dynamics Simulations:
The model was further refined and optimized by 5 ns molecular dynamics (MD) simulation. MD simulations were done using Amber 10 with the AMBERff99 forcefield and the SPC/E water model [18]. The predicted model was solvated in a truncated octahedron periodic boxThe system was then energy minimised using a steepest gradient method for 300 steps followed by conjugate gradient method. Initial velocities for each atom were assigned from a Boltzmann distribution at 298K followed by a constant volume MD simulation. A 5 ns MD simulation was run in periodic water box to establish the equilibrium behaviour of protein. All covalent bonds containing hydrogen were fixed at equilibrium lengths using the SHAKE algorithm. A 1 femtoseconds integration time step was used and configurations were collected every 1 picoseconds for subsequent analysis. A real space non-bonded cut off and particle-mesh Ewald summation method was used to compute long range electrostatic energy and force corrections. MD trajectories were analysed using ptraj module of AMBER visualized using VMD [19].

Prediction of protein function:
Protein cavities were identified by SiteHound programme [20]. The residues involved in ligand interaction were identified by PDBsum [21]. Docking of ligand to the protein structure was done using GLIDE [22].

Docking of ADP and ATP:
The geometry of ADP and ATP was optimized by molecular mechanics using IMPACT in a dynamic environment using standard TIP4P water model. The energy minimization was done using Optimized Potentials for Liquid Simulations 2005 force field; using Polak-Ribier conjugate gradient and Truncated Newton conjugate gradient algorithms. The convergence threshold used was rms gradient of 0.01.

Docking of the ATP and ADP ligand with CC-NBS was carried out using extra precision (XP) method called GLIDE (Grid-based Ligand Docking with Energetics). The ligands were prepared for docking using LIGPREP. The receptor grid generation for docking was done using the centroid of selected active site residues as well as blind docking. The different conformations of the compounds were docked flexibly and 1000 poses per compound were obtained. The analysis of the poses, complexes and the binding affinities between the receptor and ligands were analyzed using Schrodinger's suite.

Discussion:
Template identification and sequence alignments:
PDB-BLAST using found human APAF-1 as the closest match and aligned a segment of 112 amino acids (187-299) that share a sequence identity of 32%. Similarly all five homology modeling programs selected either chain A of human apoptotic protease-activating factor 1 (APAF-1) (PDB ID 1Z6T) or *C. elegans* CED-4 chain B (PDB ID 2A5Y, involved in programmed cell death) as a template [12, 13] (**Table 1 see**

Supplementary material). However, none of these programmes were able to align *S. melongena* CC-NBS sequence with either APAF-1 or CED-4 templates completely. Most of these programs except I-TASSER and PHYRE could match only the conserved P-loop containing 'nucleoside triphosphate hydrolase' region, but not the entire CC-NBS sequence. This was quite possible because of very low sequence similarity between the two sequences.

Structural modelling:
Initial CC-NBS structure was modelled with I-TASSER using APAF-1 and CED-4 as multiple templates. Ramachandran plot analysis of I-TASSSER model showed that only 80 percent residues were present in most favoured regions and 5 percent residues in disallowed regions (**Table 2 see Supplementary material**). Although I-TASSER failed to develop a satisfactory model but the predicted model guided sequence alignment of *S. melongena* CC-NBS with CED-4 chain B became an valuable input for further 3D structure modeling with HOMMER and SwissModel (**Figure 1**).

Quality of homology model:
The final structure was subsequently checked by VERIFY-3D graph. The compatibility score above zero in the VERIFY-3D graph corresponded to acceptable side chain environments. All residues appeared to be reasonable and therefore the structure of S. melongena CC-NBS can be considered relibale. Validation of the model was further carried out using Ramachandran's plot calculations computed with PROCHECK program. The phi and psi distributions of the Ramachandran's plot of non-glycine, non-proline residues are summarized in **Table 2 (see Supplementary material)**. Ramachandran's plot analysis further shows that 99.9% main chain bond length and 95.6 bond angles are within limit with overall PROCHECK G factor of -0.01.

Optimisation of the model by MD simulation:
The MD trajectories obtained were analysed to study the behaviour of the structure. **Figure 2A** shows energy plot with respect to time for the complete 5 ns simulation. Comparison of the initial model obtained from homology modeling with the frames from MD trajectories shows that there are pronounced structural fluctuations with root mean square deviations of 2.5-3.0Å (**Figure 2B**). The frames were then taken from the production phase and analysed further. RMSD for backbone atoms was >1.2 Å revealing that structural fluctuations are not highly pronounced and consistent over frames obtained from production phase (**Figure 2C**). Overlapping the different conformations from the latter half of the trajectories shows that the molecule is relatively rigid, an indicative of structure stability. The average structure from the production phase frames was taken as the final model. The structural superimposition of Ca trace of CED-4 template and *S. melongena* CC-NBS shows RMSD of 0.9 A° with an identity score of 0.14 and TM score of 0.52 (**Figure 3A**).

Active site prediction:
The SiteHound program revealed 10 possible ligand binding sites (**Figure 3B**) and residues in the site1, Val160, Gly 190, Gly192, Lys193, Thr194, Thr195, Arg295 were found conserved with the active site of CED-4 . Thus, site1 was chosen as the most favourable binding site to dock ADP and ATP and the remaining 9 sites were excluded from the study.

Predicted model functionality:
In the docked CC-NBS ligand ADP and ATP are located in the centre of the active site and stabilized by hydrogen bond interactions **Figure 3 C-D**. In the docked complex H-bond/electrostatic interaction was observed between ligand and residues Asp155, Gly190, Lys193, Thr194 and Arg295. Interestingly, these residues are in the active site of the protein as predicted by the active site detection program. In the docked complex ADP has slightly lower (-83 kcal/mol) energy than ATP (-82 kcal/mol). It indicates that ADP has slightly competitive edge over ATP in binding to the active site.
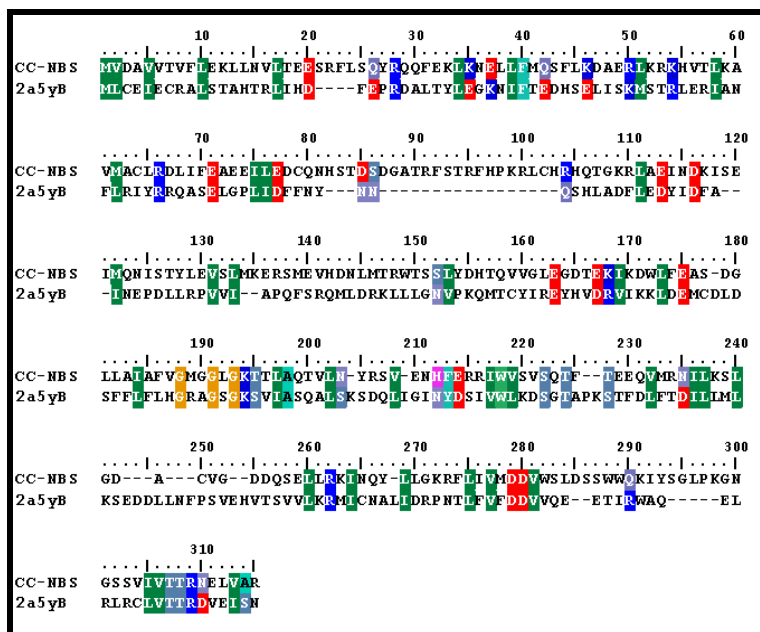
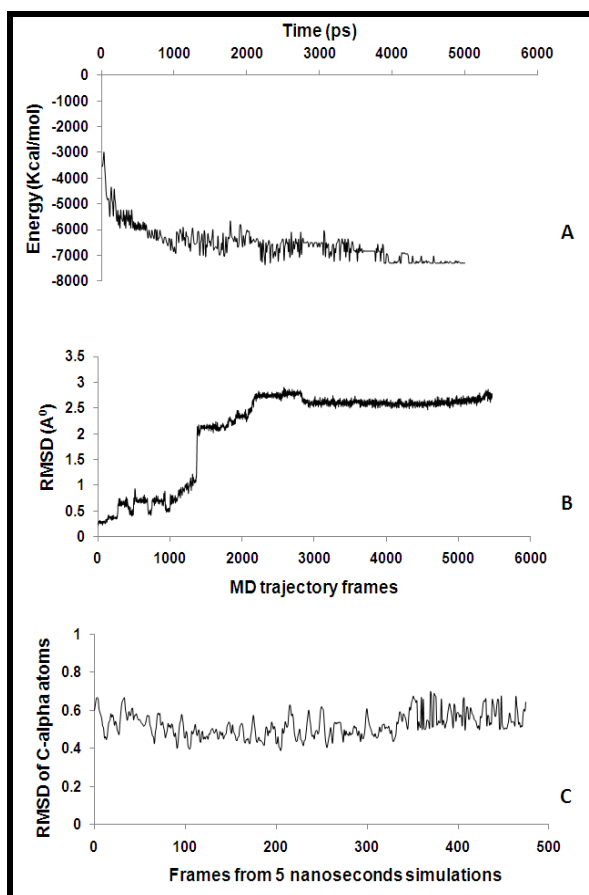**Figure 1:** Alignment of *S. melongena* CC-NBS (query) with CED-4 (template), for homology modeling.

Figure 2: A. Plot of potential energy vs. time in ps for 5 ns simulation. B. Plot of root mean square deviation (RMSD) of Cα vs. frames from MD trajectories. The graph shows the RMSD of backbone atoms (Cα, C, N) during 5ns simulation. C. Plot of RMSD of Cα atoms vs. frames from MD trajectories. The graph represents the variations in Cα atoms for 5ns simulation for frames chosen from production phase.
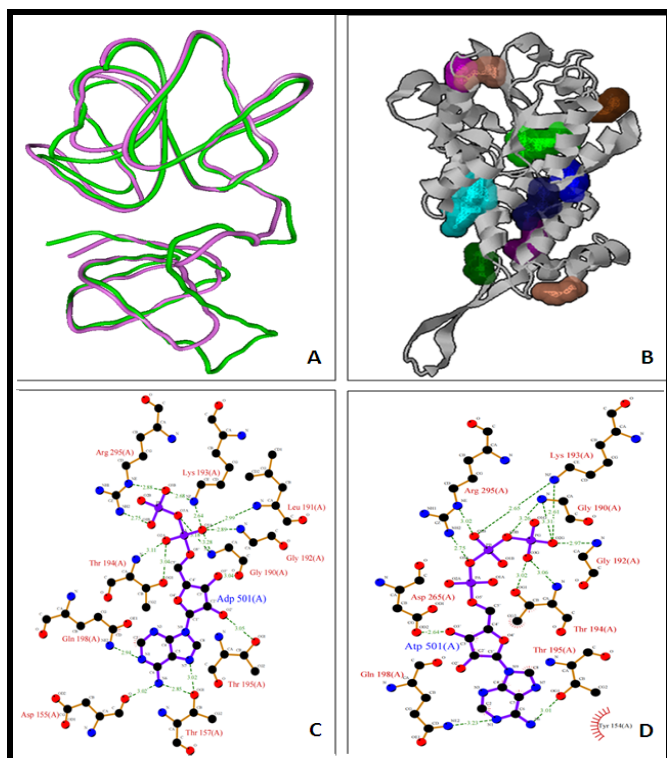
Figure 3: A. Superimposition of main chain alpha trace atoms of *S. melongena* CC-NBS (green) and *C. elegans* apoptosis regulator (magenta); B. Identification of cavities on CC-NBS surface; C-D. Interactions of conserved residues present in active site of *S. melongena* CC-NBS with ADP (C) and ATP (D) using Ligplot.

Conclusion:

The CC-NBS domains have specific binding site for ADP and ATP, with minor difference in their binding affinity that makes it a suitable molecular switch. Presence of slightly lower binding energy of ADP (-84 kcal/mol) in comparison to ATP (-83 kcal/mol) may have a special significance in the context of their binding functioning as a molecular switch. In the ground state there will be more ADP binding because of its lower binding energy. However, in the activated state ADP has to be swapped with ATP, for downstream signalling [5, 6]. Had there been large difference between ADP and ATP binding energies it would have become hard to exchange the two molecules. These studies confirm that CC-NBS domains are perfectly suited for working as molecular switch. Availability of 3D structure of CC-NBS domains will further help us in understanding evolution, structure and function of these proteins and help in designing novel proteins for pathogen resistance management.

References:

[1]   C Zipfel. *Curr Opin Plant Biol* **14**: 14 (2009) [PMID: 19608450]
[2]   L Gomez-Gomez *et al. Plant J* **18**: 277 (1999) [PMID: 10377993]
[3]   T Nurnberger *et al. Immunol Rev* **198**: 249 (2004) [PMID: 15199967]
[4]   EA Van der Biezen & JD Jones. *Trends Biochem Sci* **23**: 454 (1998) [PMID: 9868361]
[5]   P Moffett *et al. Embo J* **21**: 4511 (2002) [PMID: 12198153]

[6]   FL Takken *et al. Curr Opin Plant Biol* **9**: 383 (2006) [PMID: 16713729]
[7]   N Yan *et al. Nature* **437**: 831 (2005) [PMID: 16208361]
[8]   SJ Riedl *et al. Nature* **434**: 926 (2005) [PMID: 15829969]
[9]   R Chattopadhyaya & A Pal. J Biomol Struct Dyn **25**: 357 (2008) [PMID: 18092830]
[10] K Ginalski *et al. Bioinformatics* **19**: 1015 (2003) [PMID: 12761065]
[11] C Lambert *et al. Bioinformatics* **18**: 1250 (2002) [PMID: 12217917]
[12] Y Zhang. *BMC Bioinformatics* **9**: 40 (2008) [PMID: 18215316]
[13] LA Kelley & MJ Sternberg. *Nat Protoc* **4**: 363 (2009) [PMID: 19247286]
[14] CC Chen *et al. Nucleic Acids Res* **34**: W152 (2006) [PMID: 16844981]
[15] N Guex & MC Peitsch. *Electrophoresis* **18**: 2714 (1997) [PMID: 9504803]
[16] RA Laskowski *et al. Journal of Applied Crystallography* **26**: 283 (1993)
[17] Y Zhang & J Skolnick. *Nucleic Acids Res* **33**: 2302 (2005) [PMID: 15849316]
[18] DA Case *et al. J Comput Chem* **26**: 1668 (2005) [PMID: 16200636]
[19] W Humphrey *et al. J Mol Graph* **14**: 33-8, 27 (1996) [PMID: 8744570]
[20] M Hernandez *et al. Nucleic Acids Res* **37**:W413 (2009) [PMID: 19398430]
[21] RA Laskowski. *Nucleic Acids Res* **37**: D355 (2009) [PMID: 18996896]
[22] RA Friesner *et al. J Med Chem* **47**:1739 (2004) [PMID: 15027865]

Supplementary material:

Table 1: Templates selected for homology modeling of brinjal CC-NBS domains.

Template	Phyre	Modweb	(PS)^2	3D Jury	Swiss-Model	I-Tasser
1	1Z6T-A (12)	1Z6T-A (32)	2A5Y-B (12)	2A5Y-B (11)	none	1Z6TA (15)
2	2A5Y-B (14)	-	1Z6T-A (14)	2A5Y-B (10)	-	2A5YA (14)
3	1QVR-B (12)	-	2A5Y-C (15)	2A5Y-B (1)	-	2A5YB (14)

Values in the parenthesis show percentage identity with the template in the aligned region.

Table 2: Ramachandran plot calculations of non proline non glycine residues on 3D model of CC-NBS.

	I-TASEER		HOMMER		SPDBV	
	Number	%	Number	%	Number	%
Residue in most favoured regions	199	70.6	213	85.2	233	82.6
Residue in the additionally allowed zones	55	19.5	32	12.8	42	14.9
Residue in the generously regions	14	5	4	1.6	5	1.8
Residue in disallowed regions	14	5	1	0.4	2	0.7
Total non-glycine and non-proline residues	282	100	250	100	282	100
	ER* 2, gly15, proline 2, total 301.		ER2, gly 13, proline 0, total 265.		ER 2, gly 15, proline 2, total 301.	

*ER: end residues, gly: glycine