OMICS A Journal of Integrative Biology
Volume 17, Number 8, 2013

DOI: 10.1089/omi.2012.0131

Molecular Phylogeny, Homology Modeling, and Molecular Dynamics Simulation of Race-Specific Bacterial Blight Disease Resistance Protein (*xa*5) of Rice: A Comparative Agriproteomics Approach

Budheswar Dehury,[1,2] Mousumi Sahu,[1] Kishore Sarma,[1] Jagajjit Sahu,[1] Priyabrata Sen,[1] Mahendra Kumar Modi,[1] Gauri Dutta Sharma,[2] Manabendra Dutta Choudhury,[2] and Madhumita Barooah[1]

Abstract

Rice (*Oryza sativa* L.), a model plant belonging to the family *Poaceae,* is a staple food for a majority of the people worldwide. Grown in the tropical and subtropical regions of the world, this important cereal crop is under constant and serious threat from both biotic and abiotic stresses. Among the biotic threats, *Xanthomonas oryzae* pv. *oryzae,* causing the damaging bacterial blight disease in rice, is a prominent pathogen. The *xa*5 gene in the host plant rice confers race-specific resistance to this pathogen. This recessive gene belongs to the *Xa* gene family of rice and encodes a gamma subunit of transcription factor IIA (TFIIAγ). In view of the importance of this gene in conferring resistance to the devastating disease, we reconstructed the phylogenetic relationship of this gene, developed a three-dimensional protein model, followed by long-term molecular dynamics simulation studies to gain a better understanding of the evolution, structure, and function of *xa*5. The modeled structure was found to fit well with the small subunit of TFIIA from human, suggesting that it may also act as a small subunit of TFIIA in rice. The model had a stable conformation in response to the atomic flexibility and interaction, when subjected to MD simulation at 20 nano second in aqueous solution. Further structural analysis of *xa*5 indicated that the protein retained its basic transcription factor function, suggesting that it might govern a novel pathway responsible for bacterial blight resistance. Future molecular docking studies of *xa*5 underway with its corresponding avirulence gene is expected to shed more direct light into plant–pathogen interactions at the molecular level and thus pave the way for richer agriproteomic insights.

Introduction

Rice (*Oryza sativa* L.) is one of model plants belonging to the family *Poaceae* and is considered as one of the staple foods grown in tropical and subtropical regions for a majority of the people worldwide. Unfortunately, such an important cereal crop is under serious threat from biotic and abiotic stresses. Among the biotic threats, *Xanthomonas oryzae* is one of the prominent and notorious pathogens that is the sole cause for bacterial blight (BB) disease of rice.

Development of resistant cultivars is considered to be the most effective way to control BB of rice. The exploitation of host resistance is considered as the only reliable method to control the disease (Gnanamanickam et al., 1999). Plants possess highly sophisticated and multifaceted defense mechanisms against pathogens. The defense response includes a number of strategies, including hypersensitive response (HR), production of antimicrobial compounds, lignin formation, and increased production of pathogenesis-related (PR) proteins (Baker et al., 1997; Cutt and Klessig, 1992; Goodman and Novacky, 1994; Levine et al., 1994; Mehdy, 1994). Disease resistance genes are single, dominant resistance (*R*) genes that either directly or indirectly recognize products of pathogen effector (*Avr*) genes commonly used in breeding programs (Flor, 1971; Keen, 1990). The *R*/*Avr* interactions are concisely described as gene-for-gene interaction. This initiates a signal transduction pathway that often, but not always, leads to a rapid and localized plant cell death (Greenberg, 1997). In plant–pathogen interaction systems, resistance is often achieved by plant *R* proteins that recognize corresponding effector proteins from the invading pathogen; whereas lack of recognition results in disease. The effectiveness of *R*-gene mediated resistance in restricting

[1]Agri-Bioinformatics Promotion Programme, Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat, Assam, India.
[2]Department of Life Science & Bioinformatics, Assam University, Jorhat, Assam, India.

bacterial colonization and host phenotypic responses varies considerably, depending upon several factors such as the specificity of *R* gene, the mode of interaction/recognition, and other factors such as environmental conditions (Iyer and McCouch, 2004). The *R* genes known so far mostly fall into five classes (Martin et al., 2003). The majority of them result in resistance only to specific races of pathogens expressing the corresponding *Avr* or race-specific resistance.

In rice, resistance to a range of *X. oryzae* pv. *oryzae* (*Xoo*) strain is mediated by *Xa* genes. Genetics of host resistance to Xoo has been studied extensively and revealed two different types of host resistances [i.e., complete resistance (CR), and partial resistance (PR)]. Until now, more than 35 BB resistance genes (*Xa*) have been identified and categorized into dominant and recessive genes. Out of them, *xa*5 (Iyer and McCouch, 2004; Jiang et al., 2006), *xa*8 (Iyer-Pascuzzi and McCouch, 2007), *xa*13 (Yuan et al., 2009), *xa*24 (Wu et al., 2008), *xa*26 (Cao et al., 2007), and *xa*28 occur naturally and confer race-specific resistance. Although several resistance genes from crop plants and their corresponding avirulence genes from pathogens have been identified and cloned, only the interaction between *Xa*27 and *avrXa*27 (Tian and Yin, 2009) has been fully characterized till date.

Although resistance is conferred by dominant resistance genes, there are certain recessive resistance genes that also confer resistance, of which very few have been characterized. Among the members of the recessive resistance gene family, *xa*5 is a race-specific resistance gene that provides immunity to races of *Xoo* expressing avirulence gene *avrxa*5 (Iyer-Pascuzzi et al., 2008). It is an allele of a gene on chromosome 5 that encodes the gamma subunit of the general transcription factor (TFIIAγ) (Jiang et al. 2006; Iyer and McCouch, 2004). Among the members of recessive genes, *xa*13 of rice (Yuan et al., 2009), *mlo* of barley (Buschges et al., 1997) and *RRS*1-*R* of *Arabidopsis* (Deslandes et al., 2005) provide a wide spectrum of immunity to their respective pathogens in contrast to the *xa*5 gene. Though *mlo*, *RRS*1-R, and *xa*13 provide wide spectrum of immunity against their respective pathogens, they are structurally unlike to *xa5*.

TFIIA is one of a set of general transcription factors required for transcription by RNA polymerase II, where it is composed of two subunits in yeast and three in higher eukaryotes like humans and plants (Orphanides et al., 1996; DeJong and Roeder, 1993; Ranish and Hahn, 1991). TFIIA acts as a basal transcription factor in eukaryotes to initiate the mRNA synthesis, generally forming an assembled complex of transcriptionally competent pre-initiation complex with TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (Orphanides et al., 1996). Besides playing a key role in cell growth, it has the abundance in transcription, including stimulation and stabilization of the interaction between the TATA-box binding protein and the general transcription factor TFIID, promoter selection, gene-specific regulation, and activator-dependent transcription (Hampsey et al., 1998). Like Drosophila (Aoyagi and Wassarman, 2000), rice also contains two copies of TFIIAγ in its genome where one copy corresponds to *xa*5 on chromosome 5 (TFIIAγ5) and the other to chromosome 1 (TFIIAγ1), in contrast to Arabidopsis, which has only one of the genes. The TFIIAγ5 shares 85.8% sequence identity with TFIIAγ1 and lack of three amino acid residues from TFIIAγ1. TFIIAγ1 has the susceptible haplotype and is expressed at lower levels than *xa*5 in adult plant leaves.

At the amino acid level, the TFIIA from yeast, Drosophila, and man show a remarkable conservation in their small subunits of TFIIAc (Li et al., 1999). Sequence analysis has revealed that *xa*5 protein of rice is 50% identical to the Drosophila and human TFIIAγ subunits, and shares 85%–93% identity with *Arabidopsis*, maize, wheat, barley, and sugarcane genes. Most of the plant species bears the TFIIAγ subunit which conforms to the susceptible allele model in rice, with a valine or leucine at position 39. In contrast, yeast contains only single copy of TFIIAγ, which encodes a functional protein with a glutamic acid residue at the position corresponding to the resistant *xa*5 protein. While the yeast TFIIAγ is only 39% identical to *xa*5, it thereby suggests that the singe mutation in this type of TFIIAγ subunits does not affect the essential function of TFIIAγ (Iyer-Pascuzzi et al., 2008).

The scarcity of experimental 3-D structures of rice *xa5* protein is a constraint for understanding the molecular mechanisms of action against the corresponding avirulence protein. The homology modeling approach is a novel way for obtaining structural information about *xa5* when no crystal structure of the protein is available. Unlike other classes of *R* genes, the *xa5* gene encodes a eukaryotic transcription factor TFIIAγ. Understanding of the evolutionary relationship through molecular phylogeny with other eukaryotic transcription factors from different plant species will better facilitate the origin of evolution of xa5 protein. A detailed analysis of molecular phylogeny, comparative modeling, and molecular dynamics (MD) simulation at 20 ns of the recessive *xa*5 disease resistance (*R*) gene, which encodes the gamma subunit of TFIIA ubiquitous protein that confirms resistance against bacterial blight of rice protein, were performed to explore the evolution, structure–function, and dynamics of *xa*5 protein. Further study of the refined 3-dimensional model of rice *xa5* protein, which is stable through out the simulation process, will be useful for gaining insight in to the protein–protein interaction with the corresponding avirulence protein from *Xoo* and will aid in novel understanding of the plant–pathogen interaction at the molecular level.

Materials and Methods

Sequence retrieval and domain analysis

The reviewed amino acid sequence encoded by the rice *xa5* resistance gene (GenBank Accession no: AAV53715) conferring resistance against bacterial blight disease caused by *Xanthomonas oryzae* was retrieved from the GenBank database of NCBI (http://www.ncbi.nlm.nih.gov/). The InterProScan tool (http://www.ebi.ac.uk/Tools/pfa/iprscan/) (Zdobnov and Apweiler, 2001) was used to deduce the protein family, super family, and domain arrangement within the protein. Conserved domains of the *xa*5 protein were explored by using the following databases: Pfam (http://pfam.janelia.org/) (Fin et al., 2010), SMART (http://smart.embl-heidelberg.de/) (Letunic et al., 2012; Schultz et al., 1998), and CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) (Marchler-Bauer et al., 2011). The signal peptide was predicted by SignalP 4.0 server (http://www.cbs.dtu.dk/services/SignalP/) (Petersen et al., 2011).

Molecular evolutionary analysis

To search for homologue sequence of rice *xa*5, BLASTP (Altschul et al., 1990) was carried out against the nonredundant

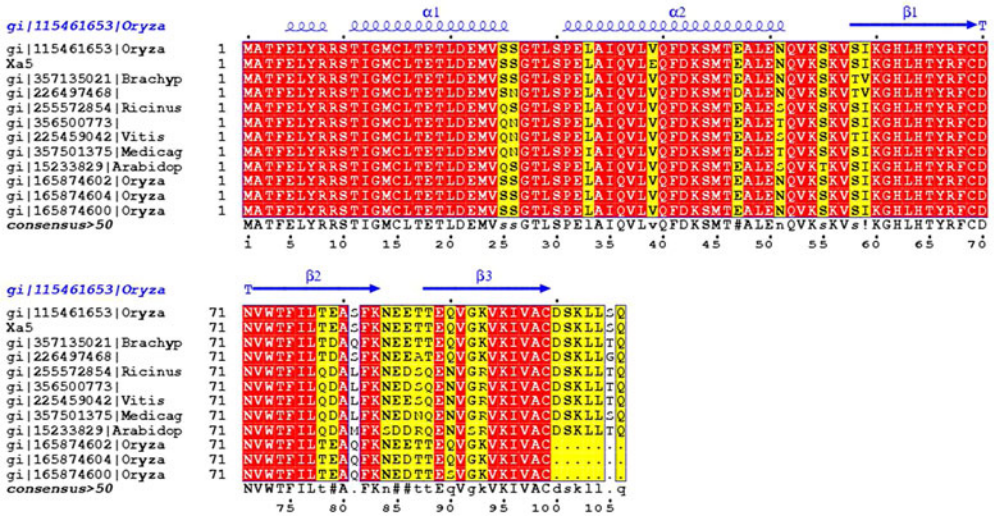

FIG. 1. Multiple sequence alignment of *xa*5 with other transcription factors from different plant species. Image was generated in ClustalX and image was displayed using ESPript. Conserved residues are *highlighted and shaded in boxes*.

(nr) database of NCBI. Sequences selected from sequence similarity search (cut-off 90% identity), including the query sequence, were aligned using ClustalX (Larkin et al., 2007). The molecular evolutionary genetic tree through Neighbor-Joining (NJ) method (Saitou and Nei, 1987) was constructed in MEGA5.0 (Tamura et al., 2011). Poisson correction was used to calculate protein distances using gap opening penalty of 10, gap extension penalty of 0.1, and gap separation distance of 4, employing Blossum weight matrix with no residue specific or hydrophilic penalties (Thompson et al., 1994). The robustness of the constructed phylogenetic tree was tested by bootstrap analysis using 1000 iterations.

Primary structure analysis

To have a broader chemistry on *xa*5, the primary structure of the protein was studied using ProtParam tool (http://expasy.org/cgi-bin/protparam) (Gasteiger et al., 2005) of Expasy Proteomic Server. Various physico-chemical parameters such as molecular weight, isoelectric point, instability index, aliphatic index, and grand average hydropathy (GRAVY) were computed. PSIPRED server (Buchan et al., 2010) was used for the prediction of secondary structure of *xa*5 protein.

Homology modeling of xa5

Template search and target-template alignment. The 106 amino acids residue long *xa*5 protein of rice was subjected to BLASTP analysis against PDB (http://www.rcsb.org/) to identify suitable template for comparative protein structure modeling and furthermore functional prediction. In addition to BLASTP search, I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Roy et al., 2010) was also employed to search for the best template for modeling of *xa*5 protein. To ensure the sensitivity and accuracy for the selection of template for *xa*5, GeneSilico MetaServer (https://genesilico.pl/meta2) (Kurowski and Bujnicki, 2003) was used, which uses a consensus approach to predict the template for model building

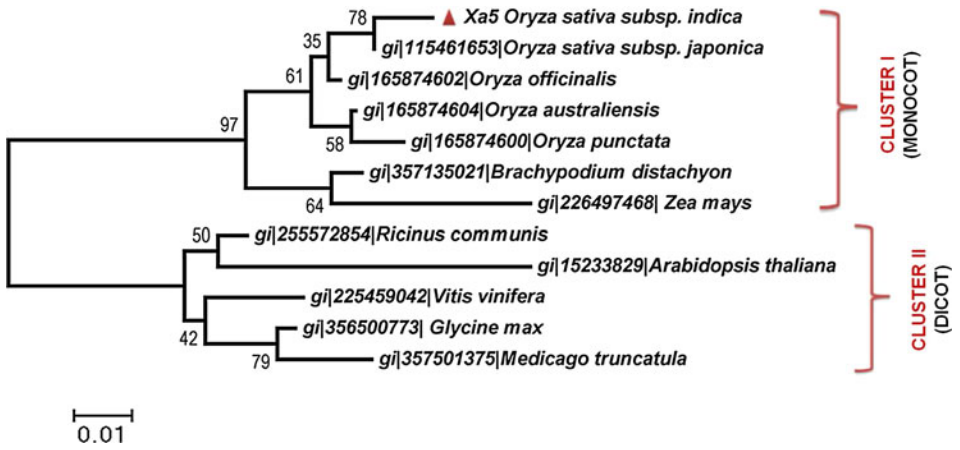

FIG. 2. Neighbor-Joining tree inferred from *xa*5 and its related transcription factor sequences in MEGA5.0. The numbers of nodes represent the percentage of boot strap value obtained from 1000 sampling. Bar, 0.01 shows the substitutions per nucleotide position.

Table 1. Templates Selected for Comparative Model Building of *xa*5 from BLAST Search Against PDB

Templates (PDB with their chain)	*Total score*	*Query coverage (%)*	E *value*	*% of identity*	*Resolution (A°)*
1NVP-D	118	93	9e-35	52	2.1
1RM1-B	98.2	93	9e-27	42	2.5
1YTF-D	97.8	93	1e-26	42	2.5

Table 2. Secondary Structure Comparison of *xa*5 and Template

Target/ Template	*Number of amino acid residues (%)*			
	Turn	*Helix*	*Strand*	*Total number of amino acids*
*xa*5 (target)	31 (29.25)	41 (38.67)	34 (32.08)	106
1NVP-D (template)	36 (33.33)	38 (35.19)	34 (31.48)	108

of protein. All the above prediction methods suggested 1NVP (D-chain) is the most appropriate template for *xa*5 protein model building, having the highest sequence identity, query coverage, and less E-value. As comparative modeling relies on a sequence alignment between target sequence and the template sequence whose structure has been experimentally determined, a target-template alignment was performed using ClustalX. The Easy Sequencing in Postscript 2.2 (ESPript) (http://espript.ibcp.fr/ESPript/ESPript/) (Gouet et al., 2003) server was used to render the target-template alignment result. Based on the target-template alignment, 20 different 3D models of *xa5* were generated by MODELLER9.11 (Sali et al., 1995) that employs extraction of spatial restraints from two sources (i.e., homology derived and CHARMm force field derived) (Brooks et al. 1983). These theoretical structural models of *xa5* were ranked based on their normalized discrete optimized protein energy (DOPE) scores. The model with the lowest DOPE score was considered as the best model and was subjected for further refinement in Discovery Studio (DS) 3.5 (Accelrys Software Inc., Discovery Studio Modeling Environment, Release 3.5, San Diego: Accelrys Software Inc., 2012). The model was refined by CHARMm force field in DS, which provides powerful mechanics and dynamics protocols for studying the energetics and motion of molecules. In the present analysis, CHARMm force field was used throughout the simulation.

Model quality assessment. The quality of *xa*5 model was evaluated by a number of tools to test the internal consistency and reliability of the model. PROCEHCK (Laskowski et al., 1993) analysis, which quantifies the residues in available zones of Ramachandran plot, was used to assess the stereo chemical quality of the model. ERRAT (Colovos and Yeates, 1993) tool, which finds the overall quality factor of the protein, was used to check the statistics of nonbonded interactions between different atom types. Similarly to determine the compatibility of the atomic model (3D) with its own amino acid sequence (1D), the VERIFY-3D (Luthy et al., 1992) program was used. All the above analysis was carried out using Structural Analysis and Verification Server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES/). Furthermore, standard bond lengths and bond angles of the xa5 model was determined using WHAT IF (http://swift.cmbi.ru.nl/whatif/) (Hekkelman et al., 2010) web server. The MolProbity web server (http://molprobity.biochem.duke.edu.) (Chen et al., 2010) was used in the quality validation of the 3D model, which provides details of atomic contact analysis of any steric problems within the molecules, as well as updated dihedral-angle diagnostics. Subsequently, the ProSA (Wiederstein and Sippl, 2007) tool was employed in the refinement and validation of the modeled structure to check the native protein folding energy of the model by comparing the energy of the model with the potential mean force derived from a large set of known protein structures. The pairwise 3-D structural superimposition of the proposed model of *xa*5 protein with its closest structural homologue was carried out using a iPBA web server that computes the root mean square deviation (RMSD) between the Cα-atoms and all atoms of the homology model and template. Finally, to have a better knowledge on the conservedness in the secondary structural topology of the refined model and the template 1NVP-D, the pair-wise structural alignment was performed using MATRAS web server (http://strcomp.protein.osaka-u.ac.jp/matras/) (Kawabata, 2003).

Molecular dynamics (MD) simulations of xa5

To investigate the stability and dynamics of the modeled protein MD simulations were conducted using GROMOS96 43A1 force field (Walter et al., 1999) and the flexible SPC water model in GROMACS 4.5.4 (Groningen Machine for Chemical

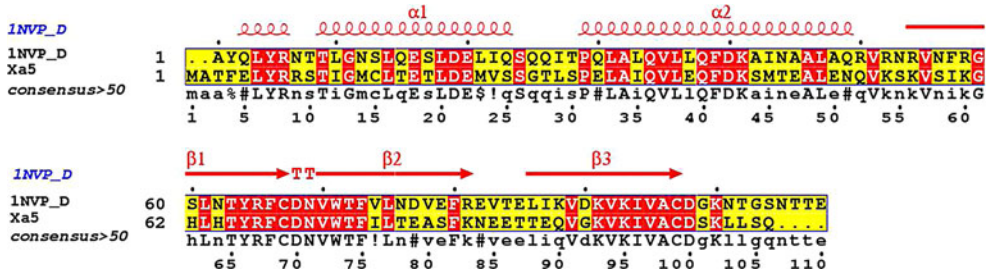

FIG. 3. Pair-wise sequence alignment of the target protein xa5 and D-chain of 1NVP. The secondary structural elements were identified from the 1NVP structure using ESPript. The α-helices, η-helices, β-sheets and strict β-turns are denoted α, η, β, and TT, respectively. Similar amino acids are *highlighted in boxes*, and completely conserved residues are indicated by *white lettering on a red background*.

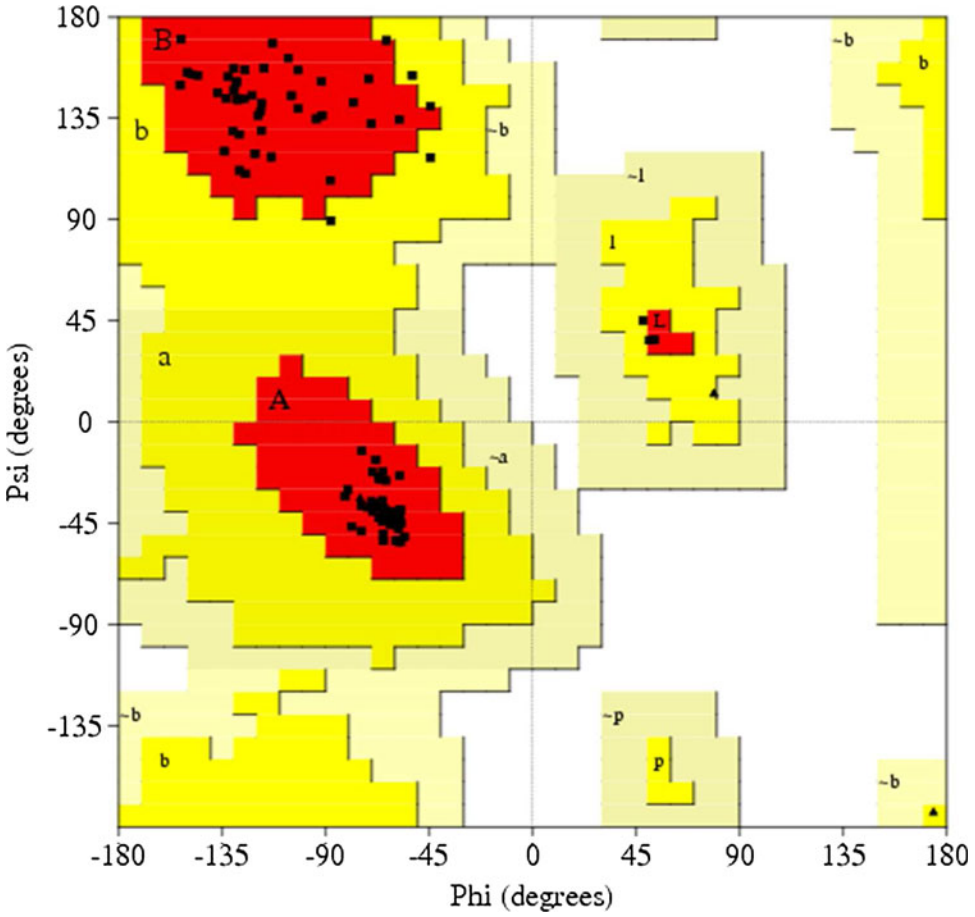

FIG. 4. Ramachandran plot of the *xa*5 model. The plot was generated with the PROCHECK program.

Simulations) (Van et al., 2005) package running on a high performance CentOS6.0 cluster computer. The protonation states of all the ionizable residues in the modeled protein were set to their normal states at pH 7.0. During the MD simulations, all atoms of the xa5 protein was surrounded by a octahedron water box of SPC3 water molecules that extended 9 A° from the protein, and periodic boundary conditions were applied in all directions. The system was solvated with the simple point charge (SPC) of 25416 water molecules. After solvating, the system was neutralized with four Na^+ counter ions. The solvated system was further subjected to energy minimization through 2000 steps of the steepest descent minimization to remove bad van der Waals contacts, which generates a good starting structure for MD simulation. All bonds were constrained using a LINCS algorithm (Hess et al., 1997), whereas the electrostatic interactions were calculated by the particle meshEwald (PME) algorithm (Darden et al., 1999). A cut-off value was set for long-range interactions including 0.9 nm for van der Waals and 1.4 nm for electrostatic interactions using the PME method. The energy-minimized model was subjected to position-restrained MD under isothermal-isobaric condition. Equilibration was carried

Table 3. Comparison of Ramachandran Plot Statistics of *xa*5 Model with its Template 1NVP

Ramachandran plot statistics	*Homology model structure of xa5*		*X-ray crystallographic Structure of template (D-Chain of 1NVP)*	
	Residues	Percentage	Residues	Percentage
Residues in most favored regions	94	94.9	79	85.9
Residues in additionally allowed regions	5	5.1	12	13.0
Residues in generously allowed regions	0	0	1	1.1%
Residues in disallowed regions	0	0	0	0
Number of non-glycine and non-proline	99	100.0	92	100.0%
Number of end residues (excluding Gly and Pro)	2		2	
Number of glycine residues	4		2	
Number of proline residues	1		1	
Overall G factor	0.01			0.30

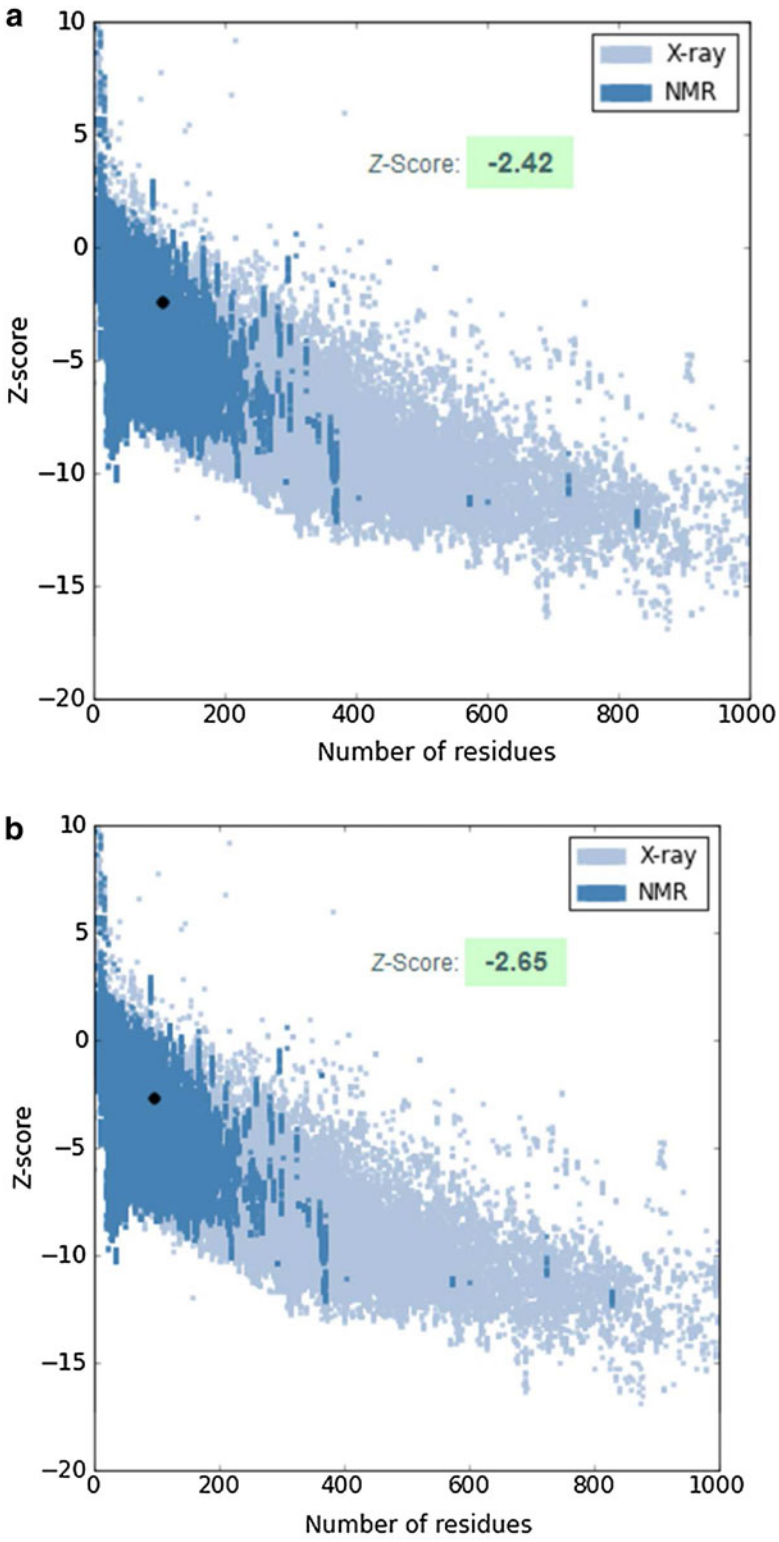

FIG. 5. Protein Structure Analysis (ProSA) of model *xa*5. **(a)** Overall quality of *xa*5 model showing a z-score of −2.42 (Native conformation to its template). **(b)** Overall quality of template (1NVP-D) model showing a z-score of −2.65.

out at 300 K and a pressure of 1 bar for 1, 00, 000 steps of 200 ps. Finally, the system was simulated for 100, 00,000 steps of 20 nano second (ns), maintaining the same temperature and pressure using PME method. Snapshots of the trajectory were taken on every 1 pico second and stored for further analyses. The system stability and differences in the trajectories were analyzed using XMGRACE software. The simulated model was subjected to validation process in SAVES server. Finally the secondary structure of the MD simulated protein model was predicted by STRIDE (http://webclu.bio.wzw.tum.de/

stride/) (Heinig and Frishman, 2004) and DSSP (http://swift.cmbi.ru.nl/gv/dssp/) (Kabsch and Sander, 1983) program which assigns secondary structure elements from atomic coordinates of proteins.

Results and Discussion

Sequence analysis and domain prediction

The published primary complete amino acid sequence of the target protein, *xa5* (106 amino acids) from *Oryza sativa* L., which is thought to provide a complete protection against BB disease of rice, was downloaded from NCBI. SMART search revealed that *xa5* possesses two domains *viz*, N-terminal transcription factor IIA (TFIIA) gamma domain (Ala2-Glu50) and C- terminal transcription factor IIA (TFIIA) gamma domain (Gln52-Ser101). Domain prediction in CDD and Pfam also suggested the same domains and fortify our domain predictions for further analysis. The two discrete domains were linked through a polar hydrophilic residue Asparagine (Asn51). The transcription factor IIA (TFIIA) is one of the factor that forms part of a transcription pre-initiation complex along with RNA polymerase II, the TATA-box-binding protein (TBP) and TBP-associated factors, on the TATA-box sequence upstream of the initiation start site. After initiation, some components of the pre-initiation complex (including TFIIA) remain attached and re-initiate a subsequent round of transcription, after which TFIIA binds to TBP to stabilize TBP binding to the TATA element. The CDD search also revealed that the C-terminal end of the TFIIA gamma subunit, which interacts with the TBP interface (polypeptide binding site) to support TFIIA function in the core promoter, is also conserved in the set of aligned transcription factors from different species. The N-terminal domain contained the alpha-helical domain of the gamma subunit of transcription factor TFIIA, whereas the C-terminal domain possessed the beta-barrel domain (Tan et al. 1996). The results from the above search tools perfectly correlated with that of the result (data not shown) from InterProscan, which showed that N-terminal was helical

a

```
     :       H1                    H2                              E1
SecA : GGGGGSHHHHHHHHHHHHHHHHHTSS HHHHHHHHHHHHHHHHHHHHHHH     EEEEEE
   4 :FELYRRSTIGMCLTETLDEMVSSGTLSPELAIQVLEQFDKSMTEALENQVKSKVSIKGHL:  63
       ***  * *  * * ***   *    * ** *** ****    **   *   *   * *
   3 :YQLYRNTTLGNSLQESLDELIQSQQITPQLALQVLLQFDKAINAALAQRVRNRVNFRGSL:  62
SecB : GGGGSSHHHHHHHHHHHHHHHHTTSS HHHHHHHHHHHHHHHHHHHHHT     EEEEEE
     :       H1                    H2                              E1

     :        E2              E3
SecA :EEEEEETTEEEEEEEEEEEE SS EEEEEEEEEEEE
  64 :HTYRFCDNVWTFILTEASFKNEETTEQVGKVKIVACD: 100
       *********** *    *       * ********
  63 :NTYRFCDNVWTFVLNDVEFREVTELIKVDKVKIVACD:  99
SecB :EEEEEETTEEEEEEEEEEEE SS EEEEEEEEEEEE
     :        E2              E3
```

b

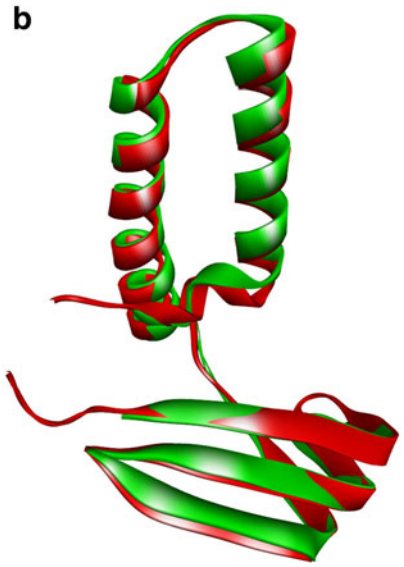

FIG. 6. Three-dimensional structural superimposition of *xa*5 model with template 1NVP. **(a)** A residue-based pairwise alignment where the lines with "SecA" and "SecB" are secondary structures of the modeled protein xa5 and template 1NVP, respectively. These secondary structures are determined by the DSSP program. The DSSP codes for secondary structures are H = alpha helix, B = residue in isolated beta-bridge, E = extended strand, participates in beta ladder, G = 3-helix (3/10 helix), I = 5 helix (pi helix), T = hydrogen bonded turn, S = bend. **(b)** Pair-wise secondary structure alignment of *xa*5 with template D-chain of 1NVP by iPBA web server. The homology modeled protein *xa*5 is colored in *red* and D-chain of the template 1NVP is colored in *green*.

whereas C-terminal formed a beta barrel. SignalP4.0 predicted *xa5* without any signal peptide cleavage sites.

Primary structure analysis of xa5

The ProtParam tool computed a molecular weight of 11.9 KDa for rice *xa5* protein. The 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. The protein *xa5* had a pI of 5.12, indicating its acidic nature (pI < 7.0). The aliphatic index (AI) is defined as the relative volume of a protein occupied by aliphatic side chains such as alanine, valine, isoleucine, and leucine. It is considered as a positive factor for the increase of thermal stability of globular proteins (Ikai, 1980). Aliphatic index of rice *xa5* protein was very high (i.e., 85.47), indicating that *xa5* protein may be stable for a wide range of temperature. The instability index provides an estimate of the stability of protein in a test tube. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable (Guruprasad et al., 1990). It was calculated that the instability index of the protein is about 40.82, which indicated the unstable nature of the protein. The GRAVY indices of *xa5* was very low (-0.170), indicating its high affinity for water.

Molecular evolutionary analysis of xa5

Comparative sequence analysis using BLAST search of *xa*5 against the nr database revealed that the *xa*5 is closely related to members of the small subunit of transcription factors (TFIIA) from various plant species with the maximum evolutionary relation to monocot and dicot plant species. Sequences producing significant alignment with the query protein *xa5* were selected based on their query coverage (90%), percentage (cut off >95%) of identity, and E-value (cut off 0). A total of 12 sequences from different organisms (including *xa*5) were considered for multiple sequence alignment in ClustalX. It is evident from Figure 1 that the transcription factors from different plant species along with *xa5* have been highly conserved throughout the evolutionary process.

Phylogenetic analysis of *xa5* using the Neighbor-Joining method in MEGA 5.0 with the other transcription factors from different plant species revealed a clear demarcation of TFIIAs into two large clusters representing species-specific divergence with strong bootstrap values within their nodes. The final rooted phylogenetic tree forms dichotomy comprising of two clusters *viz*, Cluster I and Cluster II. TFIIA sequences from monocot plants (*Oryza sativa* subsp. *indica*, *Oryza sativa* subsp. *japonica*, *Brachypodium distachyon*, *Zea mays*, *Oryza officianalis*, *Oryza australiensis*, and *Oryza punctata*) formed Cluster I, whereas Cluster II included sequences from dicot plant species (*Ricinus communis*, *Vitis vinifera*, *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*) (Fig. 2). Though all TFIIA sequences from monocots formed a single cluster within one group (Cluster I), TFIIA from *O. sativa* subsp. *japonica* was clustered together with *O. sativa* subsp. *Indica*, indicating that *indica* and *japonica* are highly similar in contrast to other members of the class. Cluster II consisted of the members from the dicot species in which *Arabidopsis* TFIIA clustered with *Ricinus communis*, whereas TFIIA from *Vitis vinifera*, *Glycine max*, and *Medicago truncatula* generated a monophyletic clade. The tree showed a distinct crop-specific clustering of sequences, representing clear crop-specific

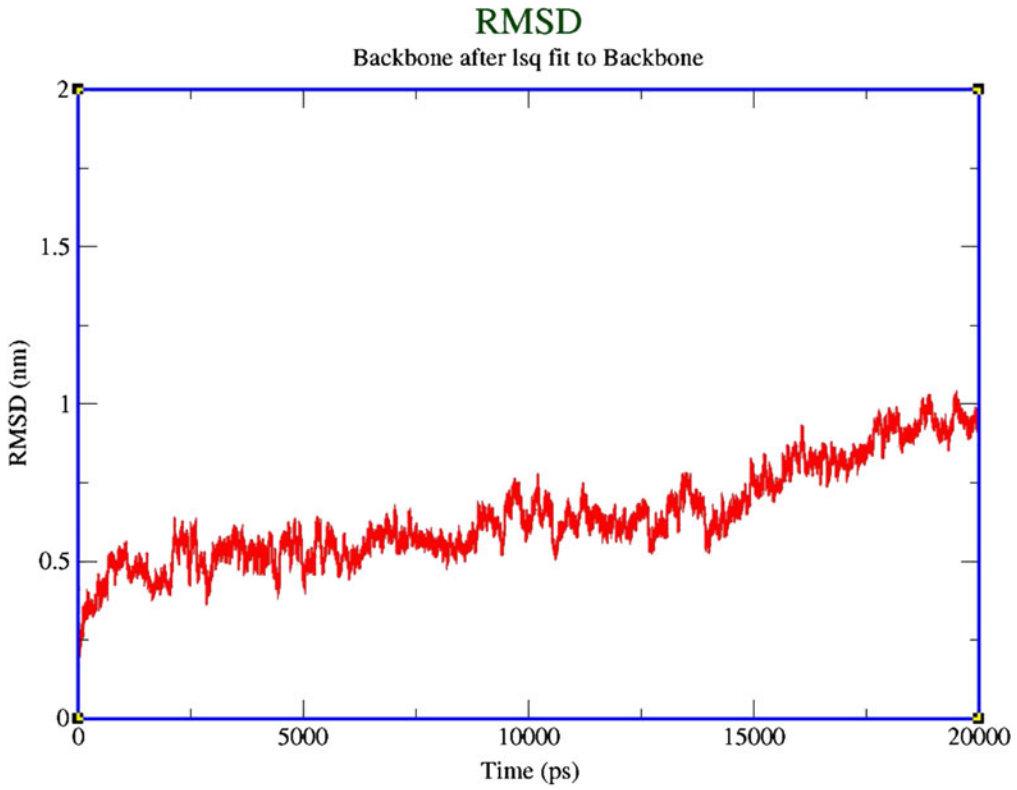

FIG. 7. The RMSD values with respect to simulation time for a 20 ns MD simulation on the *xa*5 model. The *black lines* represent the values for the model *xa*5.

sequence differences. Although TFIIA subunits from monocots and dicots share a high percentage of sequence similarity with *xa*5 of rice, they form a clear distinct crop-specific clustering within the phylogenetic tree, reflecting that they are still evolving and evolved differentially.

Homology modeling of xa5 *protein*

Comparative modeling of protein is considered as one of the most accurate methods for 3D structure prediction, yielding suitable models for a wide spectrum of applications (Bodade et al., 2010). It is usually a method of choice when a clear relationship of homology between the sequence of target protein and at least one known structure is found. This approach would give reasonable results based on the assumption that the tertiary structure of two proteins will be similar if their sequences are related. A high level of sequence identity promises a more reliable alignment between the target sequence and the template structure. BLAST search revealed three putative templates (PDB id: 1NVP, 1RM1, and 1YTF) of high-level identity with the target sequence, as shown in Table 1. These templates are the crystal structures of TFIIA from human and yeast (Bleichenbacher et al., 2003; Tan et al., 1996; Geiger et al., 1996), respectively.

Both I-Tasser and GeneSilico MetaServer suggested that the D-chain of human transcription initiation factor IIA gamma

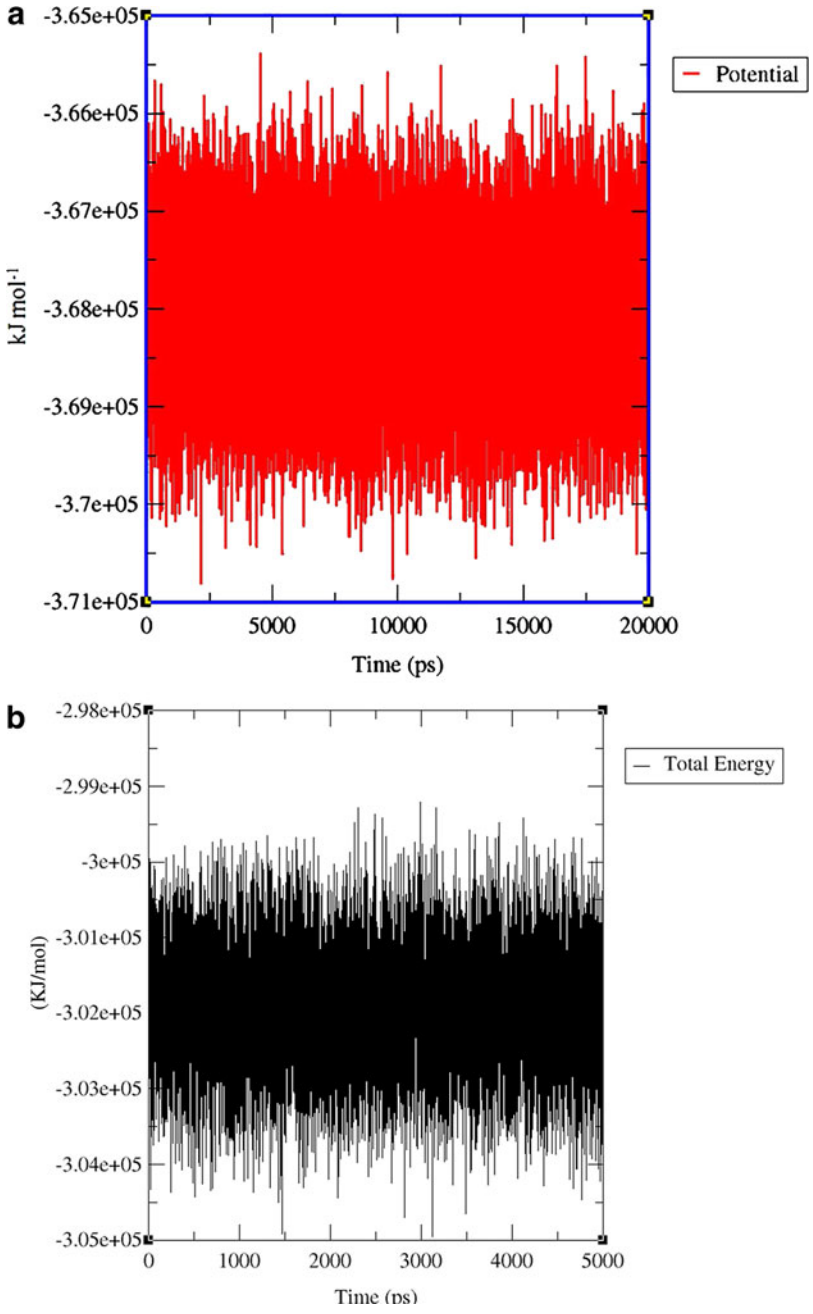

FIG. 8. Calculated energy versus time plot for MD simulations of *xa*5 using GROMACS software. **(a)** Potential energy (kJ/mol) during 20.0 ns trajectory. **(b)** Total energy (kJ/mol) during 20.0 ns trajectory.

subunit (PDB ID: 1NVP), with a resolution of 2.1 Å, as the best template for comparative modeling. The pair-wise sequence alignment of rice *xa*5 and template was generated using ClustalX, and the alignment was displayed in ESPript 2.2 (Fig. 3). Based on the target–template alignment, Modeller 9.11 generated 20 rough models of *xa5*. Out of these 20 different models, the model with the lowest DOPE score was considered to be thermodynamically stable and chosen for further refinement and validation. To assess the conservedness among the secondary structure elements, the secondary structure of the *xa*5 and the template was predicted and compared from their primary sequence. The secondary structure comparison between the target and template showed N-terminal domain was comprised of helices, whereas the C-terminal domain was comprised of beta sheets that shared strong homology across the entire length. The identity of each alpha and beta sheets between target and template is shown in Table 2. The conservation of the secondary structure revealed the reliability of our proposed model predicted by Modeller based on the target–template alignment.

Model assessment and validation

PROCHECK was used first to check the reliability of the backbone of torsion angles ϕ and Ψ of the model, which quantifies the residues fall in the available zones of Ramachandran plot, as shown in Figure 4. Ramachandran plot analysis for the modeled protein of *xa*5 showed that 94.9% residues fell in the most favored regions, 5.1% residues were in additional allowed regions, and no residue was in the generously allowed and disallowed regions. Comparing with the template, the built 3-D model had a similar Ramachandran plot, which signifies the predicted model is reliable in terms of its backbone conformation, as reported in Table 3. The high quality of the structure is further evident by the G-factor of 0.07 computed in PROCHECK. The quality of our model *xa*5 was further supported by a high ERRAT score of 81.522 (a value of ~95% shows high resolution), which indicates acceptable protein environment (Colovos et al., 1993). The VERIFY-3D results of the *xa*5 model showed 64.15% of the amino acids had an average 3D-1D score of >0.2, and 72.64% of the residues showed positive scores (cut-off score was >0), indicating the reliability of the proposed model. The PROVE program was used to measure the average magnitude of the volume irregularities in terms of the Z-score root mean square deviation of the model. The Z-score RMS values of the model and template were 1.594 and 1.866, respectively (a Z-score RMS value of ~1.0 indicates good resolution of structures). WHAT IF server analyzed the coarse packing quality, anomalous bond length, planarity, packing quality, and the collision with symmetry axis, distribution of omega angles, proline puckering, and anomalous bond angles of the model protein, reflecting its acceptance of good quality. The results from MolProbity server revealed 0% of the residues had bad bonds (goal 0%), 0.94% of the residues had bad angles (goal<0.1%), and that 0% of the Cβ deviations were >0.25 Å (goal 0%). This further confirmed the reliability of the *xa*5 model.

The energy profile of the model and the Z-score value (a measure of model quality as it measures the total energy of the structures) were obtained using ProSA program that calculates the interaction energy per residue using a distance-based pair potential. The ProSA analysis of the model *xa*5 (Fig. 5a) achieved a Z score of −2.42 and that of template was −2.65

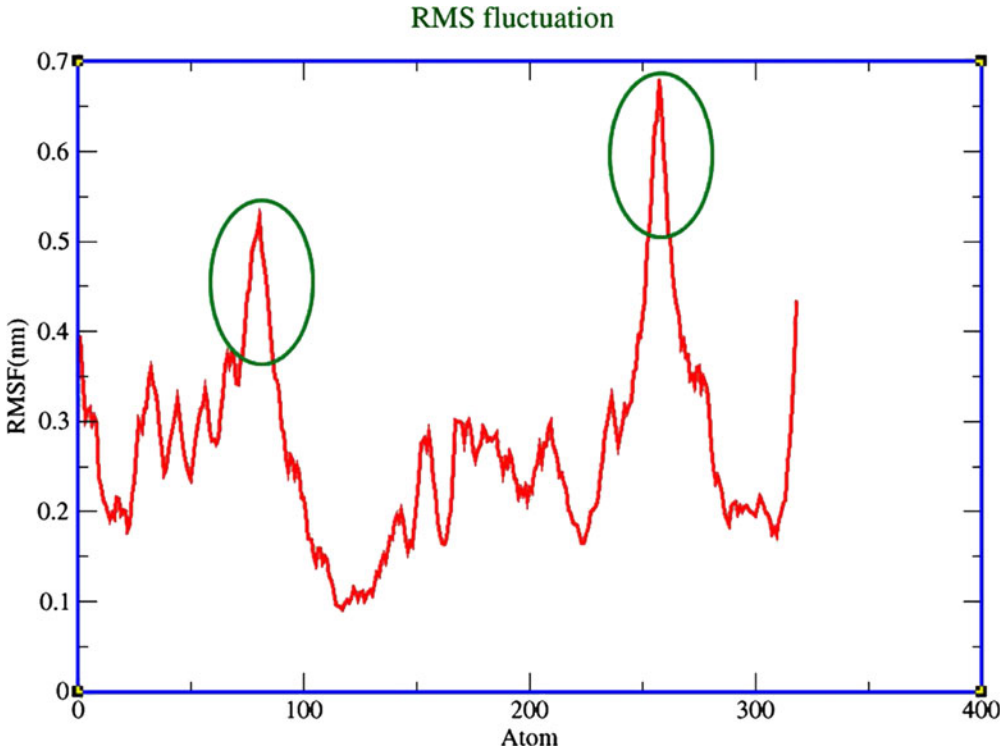

FIG. 9. The RMS fluctuation of xa5 protein backbone is represented with a *red line*. The most fluctuating region (flopping region) of xa5 protein backbone during 10 ns simulation is marked with a *green circle*.

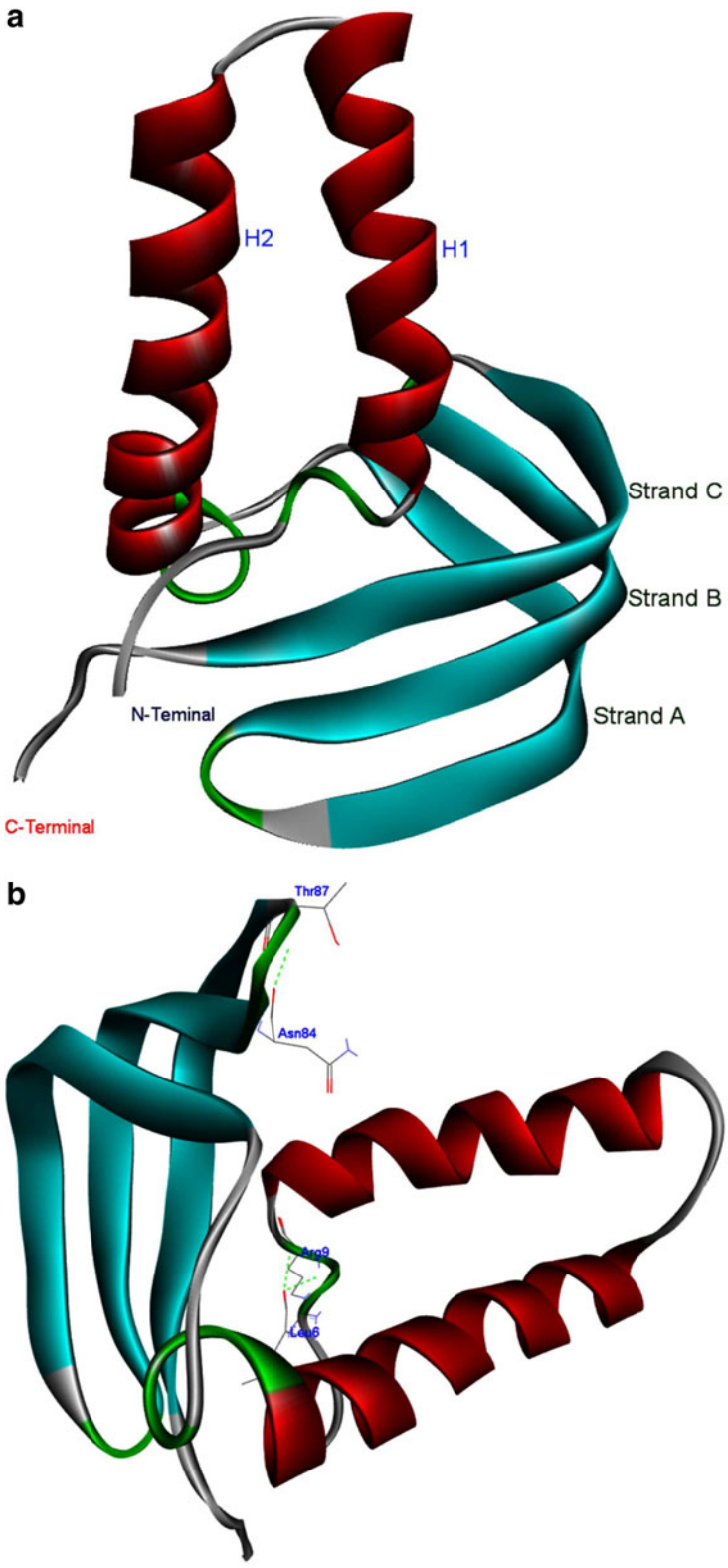

FIG. 10. Homology models of *xa*5 of rice. **(a)** Solid ribbon representation of the model colored by their secondary structure elements. The helices (H1and H2) and the strands (A, B, and C) are labeled. **(b)** Amino acid residues of gamma turn forming hydrogen bond are labeled.

(Fig. 5b), (where the negative PROSA energy reflects reliability of the model) reflecting the good quality of the model.

The quality of the model was also assessed by comparing the predicted structures to the experimentally determined structure by superimposition and atoms RMSD assessment. Consequently, superimposition of the D chain of the template with the homology model was executed by combinatorial extension of polypeptides. The RMS deviation of Cα trace between the modeled structure and template was 0.90 A°, which indicates the generated model is reasonably good and quite similar to template (Fig. 6a). The 3-D alignment of the model and template predicted by MATRAS server showed the key elements of secondary structure are strongly conserved (secondary structure elements identity of 97.9%) in the alignment where there exists a sequence similarity of mere 50.5% (Fig. 6b).

MD simulation

The predicted model of *xa5* from *Oryza sativa* pv. *indica* was subjected to molecular dynamics simulation to determine the stability and MD properties of the model using GROMACS4.5.4 simulation suite. To explore the stability and dynamics properties of the model, a molecular dynamic simulation study for 20 ns was performed for the modeled *xa*5 protein in explicit water. The energy minimization for the solvated protein model showed that maximum force dropped below the defined value of 2000 kJ $mol^{-1}nm^{-1}$ after 479 steps. Positioned restrained molecular dynamics running for 200 ps was fixed for all bond lengths in the system. To understand the stability of the system, the root mean square deviation (RMSD) of Cα backbone atoms of *xa5* model was calculated after the final MD simulation. The calculated RMSD was found to be ~0.9 at 20 ns, which indicates that the system was equilibrated after, at most, 19794 ps with respect to the initial structure. It is evident from Figure 7 that RMSD value of the model increased slowly up to 18911 ps, then slowly decreased up to 19303 ps, then slightly increased up to 19500 ps, and finally the protein remained in the plateau state till the end of equilibration.

Based on intrinsic dynamics, structural stability, and the improved relaxation of the modeled protein, the potential energy and total energy (Fig. 8a and 8b) of the structure was calculated, and the radius of gyrations graph was also generated. The energy and RMSD calculations for *xa5* protein demonstrated that the protein is not very flexible over the timescale of the MD simulations. The *RM8F* of xa5 protein backbone is shown in Figure 9. Extensive study was made on the structural evolution of the *xa*5 protein over the different time scale by comparing four snapshots at 5 ns, 10 ns, 15 ns, and 20 ns. The MD simulated 3-D structure of *xa5* comprises of two domains *viz*, N-terminal TFIIA subunit and C-terminal TFIIA subunit. The N-terminal subunit of the protein is comprised of two α-helices (H1 and H2) linked to the C-terminal subunit, which is composed of three antiparallel β-strands (A, B, and C) as shown in Figure 10a. The protein also contains 2 gamma turns and 11 beta turns in which two β-turns are stabilized by hydrogen bonding (Leu6-Arg9 and Asn84-Thr87) (Fig. 10b). The presence of β-turns in *xa5* protein signifies that they might be playing an important role in protein folding, protein stability, and molecular recognition process. We found that the N-terminal of the protein showed a minute fluctuation at 5 ns where the initial 3_{10}-helix involve the residues: Tyr7, Arg8, and Arg9 disappeared and formed a turn. During 5 to 10 ns interval, the turn transit to 3_{10}-Helix and remained there until 15 ns. After 20 ns, MD simulation these residues (Tyr7, Arg8, and Arg9) transit back to a turn in place of the 3_{10}-helix and remained stable (Fig. 11). The 3_{10} helix is the most common variant of the α-helix motif that contains 10 atoms within the hydrogen bond. These kind of secondary topologies of protein are typically not stable in solvent and very rarely play a part in overall structure of protein. The 3_{10} helix typically possesses three residues per turn, with $\{\phi, \psi\}$ angles around $\{-50°, -25°\}$. In our study, three residues (i.e., Tyr7, Arg8, and Arg9) form a 3_{10} helix at different time-scales during the molecular dynamics process involving bonds between residues i and i+3 instead of i and i+4 as in α-helix. The root mean square fluctuations (RMSF) of the *xa*5 protein backbone atom was studied to characterize the atomic fluctuation during the 20 ns MD simulation; and the most flexible regions (flopping regions) has been highlighted in Figure 9. A detailed comparison of the secondary structure elements using DSSP and STRIDE result revealed that the conformations of *xa5* remained stable after 20 ns MD simulations. Prior to simulation, the N-terminal of the *xa5* protein was made up of three helices (one 3_{10}-helix and two α-helices) and the C-terminal end comprised of three β-strands, which were connected by a turn. Although the 3_{10} helix forms a turn during MD, the other secondary structure elements remained stable throughout the simulation process (Fig. 11), which highlights the stability and reliability of the model.

The effector proteins (avirulence proteins), which are translocated directly in to the plant cell, are specifically recognized by the resistant plants having corresponding resistance genes. This recognition triggers plant defense reaction which in turn aid in HR response, localized cell death, and restricts the proliferation plant pathogens. In the process of conserved type III secretion by the pathogenic *Xanthomonas* spp., 25 numbers of effector proteins injected directly into the plant cell. Effector proteins with specific enzymatic functions play a vital role in plant pathogen interaction (Kim et al., 2006), whereas unique type III effectors mimic plant transcriptional activators and manipulate the host (plant) transcriptome (Kay and Bonas, 2009).

Studies have confirmed that activities of various type III effectors result in a change in the plant gene expression. For example, *Xanthomonas campestris* pv. *vesicatoria* AvrBs3, which is also known as TAL (transcription activator-like) effector, acts as a transcription factor and induces plant gene expression (Kay et al., 2007; Romer et al., 2007). Nissan and co-workers (Nissan et al., 2006) reported that AvrBs3 family type III effectors with TAL activity in *Xanthomonas* spp. mimic the eukaryotic transcription factors. The subunit of basal transcription machinery, which acts as an indispensable component of TAL effector recognition, is another important aspect of plant resistance mechanism. In this scenario, the gamma subunit of eukaryotic transcription factor TFIIA encoded by the rice resistance gene (*xa*5) differs from protein encoded by the susceptible allele (*Xa*5) by only one amino acid (E39V). This gamma subunit is involved in the recruitment of the basal transcription machinery by eukaryotic transcription factors. The cloning and sequencing of the corresponding *Avr* gene (*avrxa*5) of *xa*5 from the rice pathogen *Xanthomonas oryzae* pv. *oryzae* has shown that *avrxa*5 is highly similar to the members of AvrBs3 /pthA family (Zou et al., 2010).

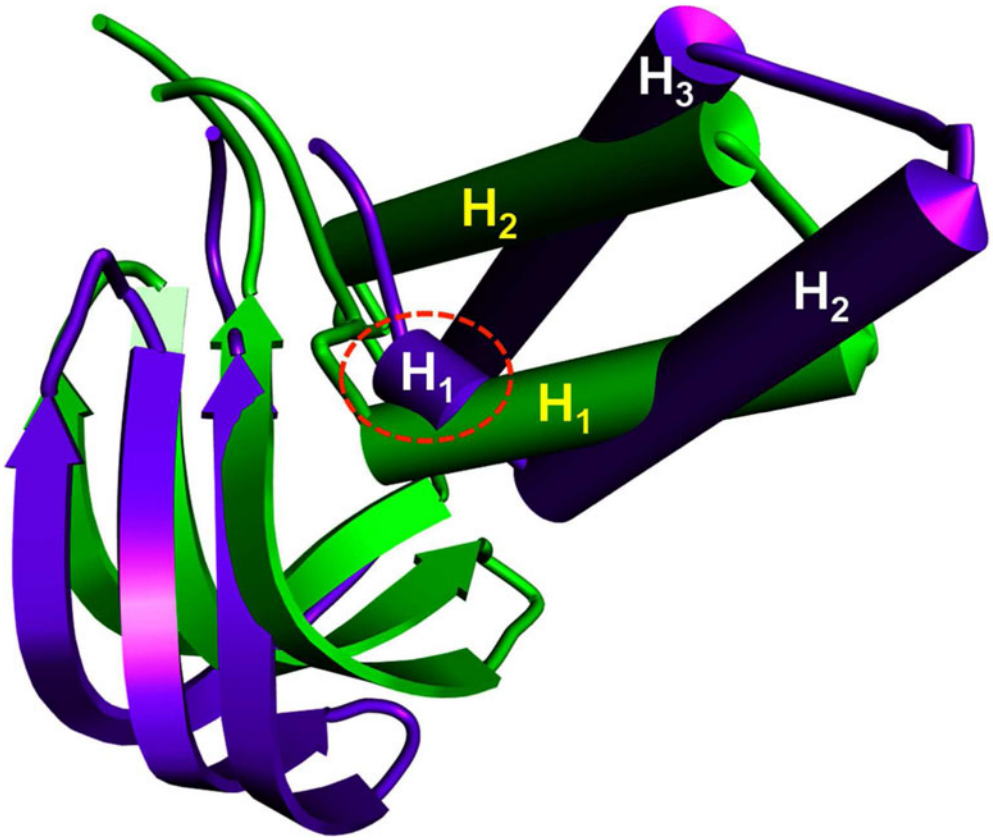

FIG. 11. Structural changes in the *xa*5 model after 20 nano second MD simulation. The change in one 3_{10} helix in to a turn after MD has been highlighted which was there in the initial model and changed to a turn during simulation. The helices of the N-terminal region are numbered with (H1, H2, and H3). The *purple color* shows the secondary structure elements of *xa*5 model before MD simulation, and the *green color* reflects the secondary structure of *xa*5 after MD simulation.

Furthermore, sequence-structural analysis of avrxa5 revealed that it is a structural homologue (sequence identity of 80%) of TAL effector PthXo1 (PDB ID: 3UGM). Since the TAL effector of *AvrBs*3 family, from *Xanthomonas* spp. mimics eukaryotic transcription factors and induces plant gene expression, as such recognition of *avrxa*5 (a member of AvrBs3 family) by the host resistance gene (*xa*5) might be the probable basis of *xa*5-*avrxa*5 (R-Avr) interaction and which may be missing in susceptible genes in *Xa*5 rice lines to promote transcriptional gene regulation. In addition, domain swapping experiment conducted by Zou et al., (2010) confirmed that *avrxa*5 gene possesses avirulence specificity towards *xa*5 gene in rice lines. As the action of *xa*5 depends on the delicate equilibrium between the conformational stability and protein stability, more structural study along with the specificity of *avrxa*5 effector protein toward recessive *xa*5 through molecular docking will provide a deeper insight in to the direct–direct (*R-Avr*) interaction of plant–pathogen at molecular levels. The coordinates of the model is deposited in the Protein Model DataBase (PMDB) (http://mi.caspur.it/PMDB/) and can be accessed using the PMDB ID: PM0077993.

Conclusion

*Xa*5, a recessive gene belonging to the rice *Xa* gene family, encodes the gamma subunit of TFIIA that provides complete resistance against race-specific pathogen *Xanthomonas oryzae* pv. *oryzae*. Although *xa*5 protein represents a small subunit transcription initiation factor of rice, it shares significant sequence similarity with the gamma subunit of TFIIA of human and is also a structurally homolog. Despite both the transcription factors of rice and human being structurally homologous, the mode of behavior in terms of their molecular function is totally different from each other, where the former provides complete resistance against bacterial blight disease, and the latter is responsible for accurate transcription by RNA polymerase II in human. The molecular evolutionary analysis revealed that the transcription factors of monocot and dicot plant species evolved differentially, and *xa*5 most probably has evolved due to gene duplication as the functional divergence is reflected by conferring resistance to rice BB. The theoretical 3-D structure of rice *xa*5 protein was built using TFIIA protein (1NVP-D) to compare the structural variations between them. The modeled structure of *xa*5 fits well with the structure of TFIIA small subunit from human, suggesting that it may also act as a small subunit of TFIIA in rice. The molecular dynamics study confirmed the structural stability of the modelled *xa*5 protein supported by RMSD and energy graphs. However, a minute change in the arrangement of secondary structure elements was observed during the simulation process. The structural analysis of *xa*5 indicates that this protein potentially retains the basic transcription factor function, which in turn may govern the novel pathway responsible for bacterial blight resistance and susceptibility.

Since the corresponding effector *avrxa*5 from XOO, a member of AvrBs3 family, shares significant structural similarity with TAL effector, it is supposed that the *avrxa*5 gene might imitate the eukaryotic transcription factor and induce host plant gene expression of *xa*5. Further studies involving protein–protein (*xa*5–*avrxa*5) docking will provide useful information to understand the gene-for-gene interaction theory of plant-pathogens at the molecular level.

Acknowledgment

The authors thankfully acknowledge the financial support for Agri-Bioinformatics Promotion Program by Bioinformatics Initiative Division, Department of Information Technology (DIT), Ministry of Communications & Information Technology, Government of India, New Delhi, as well as Biotechnology Information System Network (BTISNET), Department of Biotechnology, Government of India.

Author Disclosure Statement

No competing financial interests exist.

References

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. (1990). Basic local alignment search tool. J Mol Biol 215, 403–410.

Aoyagi N, and Wassarman DA. (2000). Genes encoding *Drosophila melanogaster* RNA Polymerase II general transcription factors: Diversity in TFIIA and TFIID components contributes to gene-specific transcriptional regulation. J Cell Biol 150, F45–F49.

Baker B, Zambryski P, Staskawicz B, and Dinesh-Kumar SP. (1997). Signaling in plant-microbe interactions. Science 276, 726–733.

Bleichenbacher M, Tan S, and Richmond TJ. (2003). Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. J Mol Biol 332, 783–793.

Bodade RG, Beedkar SD, Manwar AV, and Khobragade CN. (2010). Homology modeling and docking study of xanthine oxidase of Arthrobacter sp. XL26. Int J Biol Macromol 47, 298–303.

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, and Karplus M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem 4, 187–217.

Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, and Jones DT. (2010). Protein annotation and modelling servers at University College London. Nucleic Acids Res 38, W563–W568.

Buschges R, Hollricher K, Panstruga R, et al. (1997). The barley Mlo gene: A novel control element of plant pathogen resistance. Cell 88, 695–705.

Cao Y, Duan L, Li H, Sun X, Zhao Y, Xu C, Li X, and Wang S. (2007). Functional analysis of Xa3/Xa26 family members in rice resistance to *Xanthomonas oryzae* pv. oryzae. Theor Appl Genet 115, 887–895.

Chen VB, Arendall WB 3rd, Headd JJ, et al. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. Acta Crystallographica D66, 12–21.

Colovos C, and Yeates TO. (1993). Verification of protein structures: Patterns of non-bonded atomic interactions. Protein Sci 2, 1511–1519.

Cutt JR, and Klessig DF. (1992). Pathogenesis-related proteins. In Boller, T. and Meins, F (eds), *Genes Involved in Plant Defense*, Springer-Verlag, New York, NY, pp. 209–243.

Darden T, Perera L, Li L, and Pedersen L. (1999). New tricks for modelers from the crystallography toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations. Structure 7, R55–R60.

Daura X, Mark AE, and Van Gunsteren WF. (1998). Parametrization of aliphatic CHn united atoms of GROMOS96 force field. J Comp Chem 19, 535–547.

DeJong J, and Roeder RG (1993). A single cDNA, hTFIIA/a, encodes both the p35 and p39 subunits of human TFIIA. Genes Dev 7, 2220–2234.

Deslandes L, Olivier J, Theulieres F, et al. (2002). Resistance to *Ralstonia solanacearum* in *Arabidopsis thaliana* is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes. Proc Natl Acad Sci USA 99, 2404–2409.

Finn RD, Mistry J, Tate J, et al. (2010). The Pfam protein families database. Nucleic Acids Res 38, D211–222.

Flor HH. (1971). Current status of the gene-for-gene concept. Ann Rev Phytopathol 9, 275–296.

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, pp 571–607.

Geiger JH, Hahn S, Lee S, and Sigler PB. (1996). Crystal structure of the yeast TFIIA/TBP/DNA complex. Science 272, 830–836.

Gnanamanickam SS, Pryiyadarasani V, Narayanan NN, Vasudevan P, and Kavitha (1999). An overview of bacterial blight disease of rice and strategies for management. Curr Sci 77, 1435–1444.

Goodman RN, and Novacky AJ. (1994). The hypersensitive reaction in plants to pathogens: A resistance phenomenon, APS PRESS, St. Paul, Minnesota, pp. 244.

Gouet P, Robert X, and Courcelle E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 31, 3320–3323.

Greenberg JT. (1997). Programmed cell death in plant-pathogen interactions. Annu Rev Plant Physiol Plant Mol Biol 48, 525–545.

Gu K, Yang B, Tian D, et al. (2005). R gene expression induced by a type-III effector triggers disease resistance in rice. Nature 435, 1122–1125.

Guruprasad K, Reddy BV, and Pandit MW. (1990). Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. Prot Eng 4, 155–164.

Hampsey M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol Mol Biol Rev 62, 465–503.

Heinig M, and Frishman D. (2004). STRIDE: A Web server for secondary structure assignment from known atomic coordinates of proteins. Nucl Acids Res 32, W500–W502.

Hekkelman ML, Beek TT, Pettifer SR, Thorne D, Attwood TK, and Vriend G. (2010). WIWS: A protein structure bioinformatics Web service collection. Nucleic Acids Res 38, W719–723.

Hess B, Bekker H, Berendsen HJC, and Fraaije JGEM. (1997). LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18, 1463–1472.

Ikai AJ. (1980). Thermostability and aliphatic index of globular proteins. J Biochem 88, 1895–1898.

Iyer AS, and McCouch S. (2004). The rice bacterial blight resistance gene *xa*5 encodes a novel form of disease resistance. Mol Plant Microbe Interact 17, 1348–1354.

Iyer-Pascuzzi AS, Jiang H, Huang L, and McCouch SR. (2008). Genetic and functional characterization of the rice bacte-

rial blight disease resistance gene xa5. Phytopathology 98, 289–295.

Iyer-Pascuzzi AS, and McCouch SR. (2007). Recessive resistance genes and the *Oryza sativa-Xanthomonas oryzae pv. oryzae* pathosystem. Mol Plant-Microbe Interact 20, 731—739.

Jiang GH, Xia ZH, Zhou YL, et al. (2006). Testifying the rice bacterial blight resistance gene *xa5* by genetic complementation and further analyzing *xa*5 (Xa5) in comparison with its homolog TFIIAgamma1. Mol Genet Genomics 275, 354–366.

Kabsch W, and Sander C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637.

Kawabata T. (2003). MATRAS: A program for protein 3D structure comparison. Nucleic Acids Res 31, 3367–3369.

Kay S, and Bonas U. (2009). How Xanthomonas type III effectors manipulate the host plant. Curr Opin Microbiol 12, 37–43.

Kay S, Hahn S, Marois E, Hause G, and Bonas U. (2007). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science 318, 648–651.

Keen NT. (1990). Gene-for-gene complementarity in plant-pathogen interactions. Annu Rev Genet 24, 447–463.

Kim JG, Taylor KW, Hotson A, Keegan M, Schmelz EA, and Mudgett MB. (2006). XopD SUMO protease affects host transcription, *Pantoea agglomerans* determine host specificity and function as transcriptional activators. Mol Microbiol 61, 1118–1131.

Kottapalli KR, Kottapalli P, Agrawal GK, Kikuchi S, and Rakwal R. (2007). Recessive bacterial leaf blight resistance in rice: Complexity, challenges and strategy. Biochem Biophys Res Commun 355, 295–301.

Kurowski MA, and Bujnicki JM. (2003). GeneSilico protein structure prediction meta-server. Nucleic Acids Res 31, 3305–3307.

Larkin MA, Blackshields G, Brown NP, et al. (2007). ClustalW and ClustalX version 2.0. Bioinformatics 23, 2947–2948.

Laskowski RA, MacArthur MW, Moss DS, and Thornton JM. (1993). PROCHECK—A program to check the stereochemical quality of protein structures. J App Cryst 26, 283–291.

Letunic I, Doerks T, and Bork P. (2012). SMART 7: Recent updates to the protein domain annotation resource. Nucl Acids Res 40, D302–D305.

Levine A, Tenhaken R, Dixon R, and Lamb C. (1994). H2O2 from the oxidative burst orchestrates the plant hypersensitive disease resistance response. Cell 79, 583–593.

Li YF, Gourierrec JL, Moez T, Kim YJ, Guerineau F, and Zhou DX. (1999). Characterization and functional analysis of Arabidopsis TFIIA reveal that the evolutionarily unconserved region of the large subunit has a transcription activation domain. Plant Mol Biol 39, 515–525.

Luthy R, Bowie JU, and Eisenberg D. (1992). Assessment of protein models with three-dimensional profiles. Nature 356, 83–85.

Marchler-Bauer A, Lu S, Anderson JB, et al. (2011). CDD: A Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39, D225–D229.

Martin GB, Bogdanove AJ, and Sessa G. (2003). Understanding the functions of plant disease resistance proteins. Annu Rev Plant Biol 54, 23–61.

Mehdy MC. (1994). Active oxygen species in plant defense against pathogens. Plant Physiol 105, 467–472.

Nissan G, Manulis-Sasson S, Weinthal D, Mor H, Sessa G, and Barash I. (2006). The type III effectors HsvG and HsvB of gall-forming *Pantoea agglomerans* determine host specificity an dfunction as transcriptional activators. Mol Microbiol 61, 1118–1131.

Orphanides G, Lagrange T, and Reinberg D. (1996). The general initiation factors of RNA polymerase II. Genes Dev. 10, 2657–2683.

Petersen TN, Brunak S, von Heijne G, and Nielsen H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat Methods 8, 785–786.

Rafiqi M, Bernoux M, Ellis JG, and Dodds PN. (2009). In the trenches of plant pathogen recognition: Role of NB-LRR proteins. Semin Cell Dev Biol 20, 1017–1024.

Ranish JA, and Hahn S. (1991). The yeast general transcription factor TFIIA is composed of two polypeptide subunits. J Biol Chem 266, 19320–19327.

Romer P, Hahn S, Jordan T, Strauß T, Bonas U, and Lahaye T. (2007). Plant-pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. Science 318, 645–648.

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

Saitou N, and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4, 406–425.

Sali A, Potterton L, Yuan F, van Vlijmen H, and Karplus M (1995). Evaluation of comparative protein modeling by MODELLER. Proteins 23, 318–326.

Schultz J, Milpetz F, Bork P, and Ponting CP. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. Proc Natl Acad Sci USA 95, 5857–5864.

Song WY, Wang GL, Chen LL, et al. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science 270, 1804–1806.

Sun XL, Cao YL, Yang ZF, et al. (2004). Xa-26, a gene conferring resistance to *Xanthomonas oryzae pv. oryzae* in rice, encoding an LRR receptor kinase-like protein. Plant J 37, 517–527.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28, 2731–2739.

Tan S, Hunziker Y, Sargent DF, and Richmond TJ. (1996). Crystal structure of a yeast TFFIA/TBP/DNA complex. Nature 381, 127–151.

Tian D, and Yin Z. (2009). Constitutive heterologous expression of avrXa27 in rice containing the R gene Xa27 confers enhanced resistance to compatible *Xanthomonas oryzae* strains. Mol Plant Pathol 10, 29–39.

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, and Berendsen HJ. (2005). GROMACS: Fast, flexible, and free. J Comput Chem 26, 1701–1718.

Walter RPS, Philippe HH, Ilario GT, et al. (1999). The GROMOS biomolecular simulation program package. J Phys Chem 103, 3596–3607.

Wang GL, Ruan DL, Song WY, et al. (1998). Xa21D encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell 10, 765–779.

Wiederstein M, and Sippl MJ. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35, W407–W410.

Wu X, Li X, Xu C, and Wang S. (2008). Fine genetic mapping of xa24, a recessive gene for resistance against *Xanthomonas oryzae pv. oryzae* in rice. Theor Appl Genet 118, 185–191.

Yoshimura S, Yamanouchi U, Katayose Y, et al. (1998). Expression of Xa-1, a bacterial blight resistance gene in rice, is induced by bacterial inoculation. Proc Natl Acad Sci USA 95, 1633–1668.

Yuan M, Chu Z, Li X, Xu C, and Wang S. (2009). Pathogen-induced expressional loss of function is the key factor in race-specific bacterial resistance conferred by a recessive R gene xa13 in rice. Plant Cell Physiol 50, 947–955.

Zdobnov EM, and Apweiler R. (2001). InterProScan—An integration platform for the signature-recognition methods in InterPro. Bioinformatics 17, 847–848.

Zou H, Zhao W, Zhang X, Han Y, Zou L, and Chen G. (2010). Identification of an avirulence gene, avrxa5, from the rice pathogen *Xanthomonas oryzae* pv. *oryzae*. Sci China Life Sci 53, 1440–1449.

Address correspondence to:
Dr. Madhumita Barooah
Agri-Bioinformatics Promotion Programme
Department of Agricultural Biotechnology
Assam Agricultural University
Jorhat-785013
Assam
India

E-mail: m17barooah@yahoo.co.in