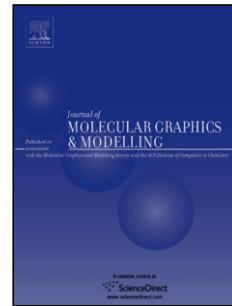


Accepted Manuscript

Title: Molecular recognition of avirulence protein (avrxa5) by eukaryotic transcription factor xa5 of rice (*Oryza sativa L.*): Insights from molecular dynamics simulations

Author: Budheswar Dehury Jitendra Maharana Bikash
Ranjan Sahoo Jagajjit Sahu Priyabrata Sen Mahendra Kumar
Modi Madhumita Barooah



PII: S1093-3263(15)00015-7
DOI: <http://dx.doi.org/doi:10.1016/j.jmgm.2015.01.005>
Reference: JMG 6506

To appear in: *Journal of Molecular Graphics and Modelling*

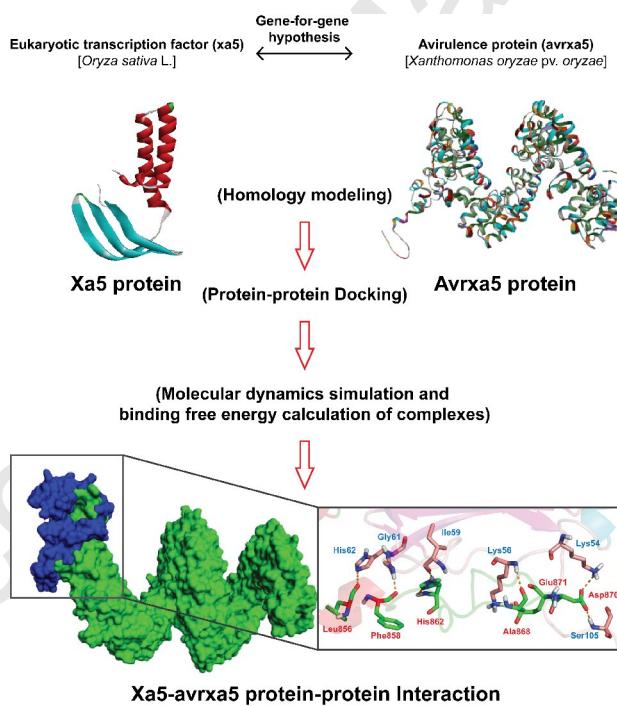
Received date: 17-7-2014
Revised date: 6-11-2014
Accepted date: 9-1-2015

Please cite this article as: B. Dehury, J. Maharana, B.R. Sahoo, J. Sahu, P. Sen, M.K. Modi, M. Barooah, Molecular recognition of avirulence protein (avrxa5) by eukaryotic transcription factor xa5 of rice (*Oryza sativa L.*): Insights from molecular dynamics simulations, *Journal of Molecular Graphics and Modelling* (2015), <http://dx.doi.org/10.1016/j.jmgm.2015.01.005>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Graphical Abstract**Molecular recognition of avirulence protein (avrxa5) by eukaryotic transcription factor****xa5 of rice (*Oryza sativa* L.): Insights from molecular dynamics simulations**Budheswar Dehury^{1,2}, Jitendra Maharana³, Bikash Ranjan Sahoo³, Jagajjit Sahu¹,Priyabrata Sen¹, Mahendra K Modi¹ and Madhumita Barooah^{1*}

For the first time we report a novel mechanism that governs the recognition of avirulence protein avrxa5 from pathogen XOO by eukaryotic transcription factor xa5 of rice. The C-terminal end of xa5 is deeply buried inside the cavity of helical region of C-terminal end of avrxa5 and reinforced by H-bonds and electrostatic interactions contributing towards the molecular recognition process. Further binding free energy analysis showed polar solvation energy contributed more than threefold of the non-polar solvation terms towards molecular recognition.



Highlights

- Highlights gene-for-gene hypothesis involving xa5-avrxa5 of Rice-Xoo.
- In this study we have employed molecular docking, MD simulation and MM/PBSA methods.
- MD simulation revealed key residues involved in molecular recognition of avrxa5.
- MM/PBSA Binding energy calculation emphasized van der Waals term is the driving force for molecular recognition.

Molecular recognition of avirulence protein (*avrxa5*) by eukaryotic transcription factor

xa5 of rice (*Oryza sativa* L.): Insights from molecular dynamics simulations

**Budheswar Dehury^{1,2,*}, Jitendra Maharana³, Bikash Ranjan Sahoo³, Jagajjit Sahu¹, Priyabrata Sen¹,
Mahendra Kumar Modi¹, Madhumita Barooah^{1*}**

¹Department of Agricultural Biotechnology, Assam Agricultural University,

Jorhat-785013, Assam, India

²Department of Life Science and Bioinformatics, Assam University, Silchar-788011, Assam, India

³Department of Bioinformatics, Orissa University of Agriculture and Technology, Bhubaneswar-751003,
Odisha, India

*Corresponding Author

Madhumita Barooah

Email: m17barooah@yahoo.co.in

Tel: +91-(0376)-2340095 (O)

Fax: +91 0376 2340001/2340101

Budheswar Dehury

Email: budheswar.dehury@gmail.com

Abstract

The avirulence gene *avrxa5* of bacterial blight pathogen *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) recognized by the resistant rice lines having corresponding resistance (*xa5*) gene in a gene-for-gene manner. We used a combinatorial approach involving protein-protein docking, molecular dynamics (MD) simulations, and binding free energy calculations to gain novel insights into the gene-for gene mechanism that governs the direct-direct interaction of *R-Avr* protein,. From the best three binding poses predicted by molecular docking, MD simulations were performed to explore the dynamic binding processes of *xa5* and *avrxa5*. Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) techniques were employed to calculate the binding free energy and to uncover the thriving force behind the molecular recognition of *avrxa5* by eukaryotic transcription factor *xa5*. Binding free energy analysis revealed van der Waals term as the most constructive component that favors the *xa5* and *avrxa5* interaction. In addition, hydrogen bonds and essential electrostatic interactions analysis highlighted amino acid residues Lys54/Asp870, Lys56/Ala868, Lys56/Ala866, Lys56/Glu871, Ile59/His862, Gly61/Phe858, His62/Arg841, His62/Leu856, Ser101/Ala872 and Ser105/Asp870 plays pivotal role for the energetically stability of the R-Avr complex. Insights gained from the present study are expected to unveil the molecular mechanisms that define the transcriptional activator mediated transcriptome modification in host plants.

Keywords

Gene-for-gene; molecular dynamics; hypersensitive response (HR); MM/PBSA; structural biology; binding free energy

1. Introduction

Unlike higher animals, plants have well developed multi-faceted defense system to protect against plant pathogens which includes hypersensitive response (HR), expression of pathogenesis-related (PR) proteins, secretion of certain antimicrobial compounds etc. More often plant pathogen effectors encoded by Avirulence (Avr) genes benefit the pathogen by promoting colonization, and benefit plants that have a matching resistance (R) gene thereby constitutes a signal which triggers resistance is termed as gene-for-gene hypothesis [1]. The avirulence proteins encoded by their respective genes (Avr) are actually the effector molecules involved in pathogenecity which are specifically expressed or strongly over-expressed during plant-pathogen interaction. Also, these proteins show the presence of secretion signals and translocation signals which rapidly recognize their corresponding plant resistance genes. Furthermore, Avr genes in naturally-occurring field isolates of pathogens provides a vital resource for the deployment of R genes in crop plants [2].

In direct-direct interaction, studies at molecular level have shown a matched pairs of R genes from plant and corresponding Avr gene from pathogen confirms the gene-for-gene hypothesis where both the gene products interact with each other in receptor- ligand manner. Although the effector molecules encoded by Avr genes plays a key role in the facilitation of disease development process but they are always under natural selection. In certain races of pathogen they develop virulence factor due to genetic modification through diverse mechanisms such as deletion [3], point mutation and frame-shift of Avr genes [4]. The bacterial blight disease caused by the race specific bacterial pathogen *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) is one of the most devastating diseases of rice across the globe. To date 35 resistance genes for bacterial blight have been identified in rice, and the full-length DNA sequences of nine genes (i.e., *Xa1*, *Xa3*, *Xa21*, *xa5*, *xa13*, *xa24*, *Xa21D*, *Xa26*, and *Xa27*) have been isolated and integrated in diverse crop improvement programs. In comparison to number resistance genes cloned from rice, the corresponding counterpart avirulence genes from their respective pathogens are few in number. But recently Zou and co workers [5] has suggested that avirulence gene *avrxa5*, of *Xoo*, the causal agent of bacterial blight in rice, interacts with resistance gene *xa5* from rice plant in a gene-for-gene manner, where, *xa5* an allele of a gene located in chromosome 5 encodes the gamma subunit of general transcription factor TFIIA (TFIIA γ or *Xa5*) [6-8].

The *xa5* in the resistance allele differs in one amino acid from the susceptible one resulted in a substitution from Valine (V) to Glutamic acid (E) in the 39 position. Furthermore it has been reported that *xa5* and *Xa5* could differ in their affinity for transcriptional activators and repressors of defense related genes where *Avrxa5* would be a transcriptional activator of defense related genes and bind with greater affinity to *xa5* than to *Xa5*, which would prefer to bind with the repressors [9]. Although several gene-for-gene interactions have been reported by

several researchers, the interaction between *Xa27* and *avrXa27* gene fully understood till date. The cloned *Xa27* resistance differs from its susceptible allele only within the putative promoter region. The expression host resistance gene (*Xa27*) induced by the corresponding avirulence gene *AvrXa27*, one of several transcription activator-like (TAL) effectors delivered by pathogenic *Xoo* [10]. Furthermore the other TAL effectors (*AvrXa7* and *AvrXa10*) with avirulence function for which the cognate resistance genes have not been cloned yet. The TAL effectors are sequence-unrelated proteins and less conserved share several distinct structural features which includes a N terminal secretion and translocation signals (STS), a central repeat region (RR) consisting of 5-25 repeats of typically 34 amino acids, a leucine zipper region (LZ), three tandem nuclear localization signals (NLS), and an acidic activation domain (AAD) [11].

Similarly the function of *Xoo* TAL effectors PthXo7 and PthXo6 induce expression of TFIIA γ 1 (a parologue on chromosome 1 of the TFIIA gamma subunit), and the TFX1 transcription factor gene, respectively. These evidences support the role of TAL effectors in alteration host gene expression by targeting genes encoding components of the host transcription machinery. The effector molecules which are secreted in to the host plant cell through a conserved type III secretion system (T3SS) encoded by *hrp* genes (for the resistance-associated hypersensitivity reaction and for pathogenesis) and a total of 25 different effector proteins are injected in to the cell by the pathogenic *Xanthomonas* spp. It has also been reported that effector proteins with enzymatic functions play a vital role in the interaction of *Xanthomonas* spp. with the host plants. Contrastingly the unique type III effectors imitate the host (plant transcriptional) activators and manipulate the host (plant) transcriptome [12] by employing a different strategy.

In this scenario, the effector proteins with TAL activity are considered as a precious weapon to understand the recognition of Avr effector by the corresponding resistance gene of the host plant. Zou et al. isolated and cloned the first *avr* gene (*avrxa5*) that corresponds to recessive *R* gene (*xa5*) from the rice pathogen *Xoo*. The sequencing of *avrxa5* gene has revealed that *avrxa5* is highly similar to the members of AvrBs3/pthA and encodes a protein of 1238 amino acid residues with a conserved C-terminal region containing three nuclear localization signals and a transcription activation domain. It has 19.5, 34-amino-acid direct repeats, but the 13th amino acid is missing in the fifth and ninth repetitive units. Also the domain swapping of the repetitive regions between *avrxa5* and *avrXa7* changed the avirulence specificity of the genes in *xa5* and *Xa7* rice lines, respectively [5].

Unraveling the mechanism of plant-pathogen recognition and signal transduction leading to the induction of disease resistance responses has been the key factor for exploiting the innate immunity in plants. In gene-for-gene interactions, avirulence gene products from pathogens generate signals (ligands) and resistance genes of

host plant encode cognate receptors wherein defense response is activated either through direct R-Avr protein interaction or indirectly through detection of changes in their host protein target. We have previously reported that the disease resistance gene *xa5* function as a eukaryotic transcription factor which governs a novel pathway responsible for bacterial blight resistance in rice. This resistance mechanism achieved through the subunit of the basal transcription machinery acts as a polymorphic component of TAL (transcription activator-like) effector recognition as reported in our earlier studies [13].

Most interestingly, *avrxa5* is a member of the AvrBs3 family, also known as TAL effector, is the corresponding avirulence gene of the eukaryotic transcription factor *xa5*, which upon recognition triggers plant defense reactions often culminating in the hypersensitive response (HR) and restricting proliferation of pathogen. Recognition of the corresponding TAL effector Avrxa5 from *X. oryzae* pv. *oryzae* (Xoo), which forms the basis of interaction with the *xa5* protein is still poorly understood due to the lack of any experimental crystal/NMR structures of *xa5* protein and corresponding avirulence protein *avrxa5* of Xoo. In continuation of our earlier studies, an attempt was made to uncover the novel mechanism that govern direct-direct interaction immune receptor of rice and avirulence protein of Xoo, through an integrated computational study that combines molecular docking, molecular dynamics (MD) simulations, and binding free energy calculations using Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) technique. Altogether, this study demonstrates structural and dynamic properties of resistance protein in association with *avrxa5* for the first time. Structural information on plant immune receptors and cognate ligands of pathogens gained from this study is expected to provide a new dimension in unraveling the plant-pathogen interaction at the molecular level.

2. Materials and methods

2.1 Delineation of domain boundary and phylogenetic analysis *avrxa5*

The full length avirulence gene *avrxa5* of bacterial blight pathogen *Xoo* reported earlier by Zou et al. was obtained from the GenBank database of NCBI (GenBank Accession Number: ACM44927). The open reading frame (ORF) consisted of 2700 bp nucleotides which encoded 1238 amino acids was further used for domain architecture study using SMART [14]. SMART search revealed *avrxa5* comprised of conserved carboxy-terminal end having three nuclear localization signals and a transcription activation domain. Rapid Automatic Detection and Alignment of Repeats (RADAR) tool of EBI was used to identify the complex repeat architecture (functional and structural units) of *avrxa5*. The signal peptides were predicted using SignalP v.4.1 [15].

Homologous sequences producing significant alignment (cut-off>90% identity, high query % of query coverage) with *avrxa5* protein from BLAST search [16] against non-redundant (nr) database of NCBI were screened and submitted to Clustal omega server [17] for multiple sequence alignment with default parameters.

Finally the set of aligned sequences were used for establishment of phylogenetic relationships in MEGA v.6.0 [18]. An un-rooted tree was constructed using Neighbor-Joining method [19] incorporating Jukes-Cantor model of substitution model. Finally to estimate the reliability and check the confidence level of each node of tree, bootstrap analysis was performed with an iteration of 1000.

2.2 Template identification and comparative modeling

The mature *avrxa5* consisted of 844 (201-1044 position) amino acid residues with a predicted molecular mass of 88.77 kDa, was used for comparative modeling. The template search was carried out using DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST) [20] program against protein data bank (PDB) which predicts the remote protein homologs and more sensitive and produces more accurate result than general BLAST search. Moreover fold recognition servers *i.e.*, GeneSilico Metaserver (<https://genesilico.pl/meta2/>) [21], and Pcons.net (<http://pcons.net/>) [22] were also employed in template identification process. Along with DELTA-BLAST, fold recognition servers suggested three putative templates as listed in Table S1. In this study, we have employed both single-template and multi-template approach for model construction. For single template approach, crystal structure of Tal effector Pthxo1 [23] (PDB ID: 3UGM A chain) at 3.0 Å resolution was used as the best template (identity of ~79% and positives of 81%) for 3-D modeling as identified by Modeller function. Although both Tal effector Dhax3 (PDB ID: 3V6T) and DNA-Binding domain of Dhax3 (PDB ID: 3V6P) [24] showed a higher level of identity (89.0%) than 3UGM in Modeller, the query coverage (61% and 56%) and alignment score were found to be very low for both the proteins. Both single-template and multi-template approach aid in generation of 200 models by Modeller which was ranked accordingly their low discrete optimized protein energy (DOPE) scores. Models with lowest DOPE score from both the approaches were taken together and superposed with the templates on their C_α atoms. Finally, the model with the multi-template approach which showed least RMSD was considered as best model and subjected to energy minimization and further refinement through MD simulation. In this work, the homology modeled protein of *xa5* was obtained from our earlier report [13].

2.3 Model validation

We used PROCHECK [25], ERRAT [26], Verify-3D [27], ProSA [28], MolProbity [29], ProQ [30] and VADAR [31] tools in model validation process. PROCHECK analyses the Ramachandran plot quality, peptide bond planarity, non-bonded interactions, main chain hydrogen bond energy, C_α chiralities and overall G factor. The non-bonded interactions between different atom types were checked by ERRAT, whereas, Verify-3D assessed the compatibility of the atomic model with its own amino acid sequence. The Protein Structure Analysis (ProSA) tool which is used in the refinement and validation of modeled structure was employed in the

recognition of errors of the theoretical model of protein. In addition, MolProbity structure-validation web server was used to evaluate the modeled protein quality at both global and local level which uses optimized hydrogen placement and all-atom contact analysis, covalent-geometry and torsion-angle criteria to evaluate the protein models.

2.4 Molecular dynamics simulations

The preliminary modeled proteins obtained from the comparative modeling (xa5 and avrxa5) were first refined by energy minimization (EM) in vacuum and subsequently the energy minimized system was subjected to MD simulation in aqueous solution using GROMOS96 53A6 field in GROMACS 4.6 [32]. MD simulations of the both the models were carried in SPC216 water models in separate cubic boxes with a minimum distances of 15 Å between the protein surfaces and box edges. To neutralize both the systems, a physiological ionic strength (0.15 M) of counter ions was added. The final MD simulation of 20 ns (nano second) was carried out for each system. In the refinement process, the structural stability of the both the systems was investigated. A model can be acceptable, if its overall structure is reserved after EM in vacuum (*i.e.*, intrinsic stability) and molecular dynamics simulation in aqueous solution. The water molecules were constrained using the SETTLE algorithm, whereas the covalent bonds within the protein were constrained using the SHAKE algorithm. The particle mesh Ewald method was used to treat electrostatic interactions with a cut-off value of 10 Å and space grid size of 1.2 Å and a cut-off distance of van der Waals interactions was 12 Å. The coordinates of both the systems were saved at every 2 ps and used for further studies. System stability and dynamics parameters including root mean square deviation (RMSD), radius of gyration (rg), energy and root mean square fluctuations (RMSFs) were analyzed from the resulting trajectories of both systems using XMGRACE and VMD. Time-dependent secondary structure analyses of both the models were executed by visual molecular dynamics (VMD 1.9.1). All the MD simulation work of the current work was performed using GROMACS running on a HPC server with 56 nodes in CENTOS environment. The average structures obtained from the stable portion of trajectory in case was extracted which was considered to represent the structure of xa5 and avrxa5 model. The bad contacts between side chains atoms of both the models obtained after MD were optimized by performing energy minimization using GROMACS.

2.5 Protein-protein docking

Generally proteins do not work in isolation; they accomplish their biological function by interacting with one or more molecules to form biological assemblies. Now days, characterization of protein structural assemblies formed of interacting partners is key aspect to understand the molecular interactions, unraveling the molecular basis for different chemical processes in the healthy or diseased cell [33]. Moreover, the two predominant

experimental techniques *i.e.*, X-ray Crystallography (XRD) and Nuclear Magnetic Resonance (NMR) to elucidate the biologically-active structure of a protein assembly is time consuming and labor-intensive and are often limited by the size of the molecular assembly [34]. Due to the biological importance and ubiquity of protein-protein assemblies and current limitations of experimental techniques, computational approaches are emerging to complement wet laboratory efforts in elucidating structures of protein assemblies. When the number of protein units is limited to two, protein-protein docking plays the vital role in predicting the biologically-active or native structure upon docking of the units onto each other is known as. Lack of experimental evidence possesses a great constraint to identify possible interacting surfaces in both the proteins, so in the absence of experimental evidences accurate protein-protein docking and predicting the critical residues involved in recognition mechanism is of prime importance in our study. To address this key issues, we used high ranked protein-protein docking tool Cluspro (top ranked docking software according to CAPRI, 2012) [35] for docking using the both the interacting partners (*i.e.*, xa5 and avrxa5). Based on the gene-for-gene hypothesis, avirulence (avrxa5) gene products from pathogens generate signals (ligands) was considered as ligand and resistance gene (xa5) of host plant encode cognate receptors wherein defense response is activated either through direct R-Avr protein interaction, we employed the same technique in our protein-protein docking methodologies. The resultant 20 docking poses generated from Cluspro were screened to obtain the best three which were further subjected to final rescoring and refinement using HADDOck [34]. Finally, three best docking poses were screened based on the results of HADDOck and subsequently subjected to energy minimization to optimize the close contacts between side chain atoms followed by 20 ns MD simulation.

2.6 MD simulation of the xa5-avrxa5 complexes

The best three xa5-avrxa5 complex structures obtained from molecular docking were used as starting structure for MD simulation studies using GROMOS96 43A1 force field included in GROMACS software package. The large three protein-protein complexes were solvated in three separate cubic boxes SPC216 water molecules box having dimensions of $15 \times 15 \times 15 \text{ \AA}$. A physiological ionic strength of 0.15M counter ions was added to neutralize the systems (the atomic composition of the simulation systems is provided in Table 2). The stepwise protocol applied for MD simulation studies in this work was adopted from our earlier work [13]. All hydrogen-heavy atom bonds were constrained by the SHAKE method, and simulations were performed with a 2 fs time step. Once the system equilibration was achieved, the production MD simulations of 20 ns were performed for all xa5-avrxa5 complexes using NVT ensemble without any positional constrains. The system stability and energetic parameters of all the complexes were analyzed using XMGRACE and VMD. The intermolecular

interactions were visualized using Pymol 1.3 (<http://www.pymol.org/>), Discovery Studio visualizer 3.5, and Ligplot⁺ 1.4.4 (Dimplot) [36] tools.

2.7 Normal mode analysis of the protein-protein complexes

Principal component analysis (PCA) can provide a brief picture of motions, which extracts the highly correlated fluctuations from the MD trajectories by applying the dimensionality reduction method. This method is based on the calculation and diagonalization of the covariance matrix. The eigenvectors (also called the principal modes) of the matrix represent the directions of the concerted motions and the eigenvalues indicate the magnitude of the motions along the direction. To identify the most prominent structural motions during the MD simulation in all the three best complexes, PCA was performed using *g_covar* and *g_anaeig* of the Gromacs suite. The backbone atoms (N, C, Ca, and O) were considered for this analysis. Distinct PCA calculations were performed on the MD trajectories of the three complexes of xa5-avrxa5. Among the generated eigenvectors, the most dominant number was considered, and 100 numbers of frames were created. The top eigenvector was analyzed in ProDy [37] to identify the most eminent dynamic regions. The normal mode analysis was performed using anisotropic network model (ANM) to render the elastic network modes of the R-Avr (protein-protein) complexes.

2.8 Calculation of electrostatic potential

Electrostatic surface potentials around the representative coordinates of xa5 and avrxa5 were calculated by solving nonlinear Poisson Boltzmann (PB) equation using Adaptive Poisson Boltzmann Solver (ABPS) software [38]. The calculation was performed with a grid spacing of 0.4 Å, temperature of 296 K, and salt concentration of 0.15 M. The dielectric constants were set to $\epsilon=2$ and $\epsilon=78$ for protein and solvent, respectively.

2.9 MM/PBSA binding energy calculation

Free energy calculations have proven useful for a number of topics in computational biology, including drug design, receptor-ligand interaction and protein structure determination. Several methods are available to calculate free energies, such as free energy perturbation, replica exchange free energy perturbation, and thermodynamic integration. These methods, although theoretically rigorous, are computationally demanding and become prohibitively expensive as system size increases. In this study, the binding free energy ($\Delta G_{\text{binding}}$) for each system was calculated by the MM/PBSA [39] methodology as shown in eqn. (1) that integrates molecular mechanics and the continuum solvent model. The binding free energy of the three complexes was calculated using MM/PBSA approach that combines internal energy, solvation energy based on electrostatic and nonpolar contributions, and the entropy [40, 41]. A total of 500 structures were generated from the MD trajectory for each MD simulation systems and the binding free energy was calculated using the GMXAPBS tool [42]. The GMXABPS tool combined both GROMACS and APBS by running the shell script files on the trajectory (trr),

topology (tpr) and index (ndx) files for each structure generated from the MD trajectories. The binding free energy ($\Delta G_{\text{binding}}$) was estimated for the three *xa5*-avr*xa5* complexes using following equation:

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{receptor}} + G_{\text{ligand}}) \quad (1)$$

The computation of free energy terms of complex, receptor, and ligand is performed as follows

$$\langle G \rangle = \langle E_{\text{MM}} \rangle + \langle G_{\text{sol}} \rangle - T \langle S_{\text{MM}} \rangle \quad (2)$$

Where E_{MM} stand for the molecular mechanics interaction energy which can be defined as

$$E_{\text{MM}} = E_{\text{int}} + E_{\text{coul}} + E_{\text{vdW}} \quad (3)$$

E_{int} represents bond, angle, and torsion angle energies, E_{coul} indicate electrostatic energy, and E_{vdW} stands van der Waals energy. The solvation free energy term, G_{sol} is the summation of polar and nonpolar contributions which were calculated using APBS program [38] and the polar term (G_{polar}) was calculated by solving non-linearized PB equation.

$$G_{\text{sol}} = G_{\text{polar}} + G_{\text{nonpolar}} \quad (4)$$

The parameters for APBS calculation were as follows: grid spacing was set to an upper limit of 0.5 Å; the temperature was set to 296 K; and the salt concentration was 0.15 M. The nonpolar contribution, G_{nonpolar} is computed as:

$$G_{\text{nonpolar}} = \gamma \text{SASA} + \beta \quad (5)$$

Where $\gamma = 0.0227 \text{ kJ mol}^{-1} \text{ Å}^{-2}$ and $\beta = 0 \text{ kJ mol}^{-1} \text{ 255}$ [43]. The dielectric boundary was defined with 1.4 Å of probe radius. The standard errors were calculated using following equation:

$$\text{Standard error (SE)} = \sigma/\sqrt{N} \quad (6)$$

Where σ is the standard deviation and N is the number of structures used in the calculation. The schematic flow of the present work is listed in Fig. S1.

3. Results and discussion

3.1 Sequence analysis and reconstruction of phylogeny

Repeat search using RADAR revealed that *avrxa5* is comprised of 22 unique repeats with a combination of 32 amino acids where the 21st and 22nd repeats are partial ones (Fig. S2). BLAST search of *Xoo* avirulence protein against non-redundant (nr) database revealed that *avrxa5* is a very closely related member AvrBs3 Tal effector family. Significant aligned sequences with *avrxa5* are notably belongs to Tal effector of pathogen *Xoo*. A total of 19 homologous sequences (cut-off: identity of >90% and E value of 0) of *avrxa5* were aligned in Clustal Omega for phylogeny analysis. Phylogenetic tree inferred using NJ method distinctly showed *avrxa5* shares the same node with members of AvrBs3 family of Tal effectors with bootstrap value of more 90% which signifies *xa5* is a key member of TAL effector family. Although the tree topology showed dichotomy with two different

clusters having strong bootstrap value within their nodes, Tal effector of *Xoo* (*avrxa5*) share common evolutionary origin and may have evolved from a common ancestor as evidenced from Fig. S3.

3.2 Comparative modeling of *xa5* and *avrxa5*

In this study, we adopted the modeled *xa5* from our earlier studies [13]. For 3-D modeling of *avrxa5*, we employed multi-template approach as it is recommended (when suitable templates are available) to avoid biasness towards one protein or one set of side chain conformations. DELTA-BLAST and fold recognition servers identified three putative templates *viz.*, PDB ID: 3UGM, 3V6T and 3V6P (Tal effectors of human) as listed in Table S2 for model building of *avrxa5*. Based on the target-template alignment, Modeller facilitated in the development 20 predicted models. The model with the lowest DOPE score was considered for loop refinement using Lopper algorithm implemented in DS3.5. After a round of energy minimization using CHARMM force field in DS3.5, the refined model was evaluated and compared with the closest template to test the accuracy of the modeled protein.

3.3 Model quality assessment and evaluation

The accuracy of dihedral angles (Φ/Ψ) of *avrxa5* model was measured using Ramachandran plot in Procheck. The refined model showed good percent of residues in most favored regions (662 amino acids, 91.9%), additional allowed regions (48 amino acids, 6.8%), and generously allowed regions (9 amino acids, 1.1%). But none of the residue fell in the disallowed region of Ramachandran plot reflecting acceptability of the model. Furthermore, the quality of the model was well supported by the overall quality (G factors) within the acceptable range -0.08 (acceptable values of G-factor in Procheck are between 0 and -0.5 with the best model displaying values close to zero). Comparison of Ramachandran plot statistics of the modeled *avrxa5* with its closest structural homologue revealed that both the proteins fall almost in the same range. The packing quality of each residue as assessed by the Verify-3D program represents the profile obtained with respect to the residues. Compatibility of the model residues with their environment is assessed by a score function. Residues with a score over 0.2 should be considered reliable. Score for our model *avrxa5* maximally lied above 0.2 which corresponds to acceptable side chain environment. ERRAT score provides accuracy of the non-bonded atomic contacts, and *avrxa5* model had a score greater than the acceptable value (50%) [26]. The Z score of the theoretical model calculated by ProSA are in agreement to that of the PDB structures of similar sizes [28]. ProQ analysis indicated that the quality of *avrxa5* was extremely good [30]. Analysis of the bond length and bond angle of the predicted through MolProbity revealed that none of the residues contains bad side chains or main chain conflicts (**Table 1**). The degree of structure similarity was measured to investigate how well the modeled structure matches the X-ray data of template (PDB ID: 3UGM), the prepared model and template were

superimposed on their backbone atoms using MATRAS web server [44]. The RMSD values of the backbone atoms and C α atoms the model with respect to the template reported to be 0.78 Å and 0.86 Å respectively reflects quality of the model. Moreover, model validation through structural evaluation servers including Vadar and MolProbity signifies the overall quality was good and in some cases better than the closest homolog (**Table 1**). The 3-D structure of both the modeled structures is listed in **Fig. 1**.

3.4 Molecular dynamics simulation of *avrxa5*

The modeled *xa5* and *avrxa5* protein was subjected to MD simulations to estimate the stability of both the systems. The 20ns molecular dynamics for both these proteins *avrxa5* showed linear deviation after till 20ns whereas the showed much flexibility than *xa5* protein. In contrast to *avrxa5*, both RMSD and radius of gyration (Fig. 2a, b) were found to be stable within this time interval (15ns -20ns) of MD simulation in *xa5* protein. Initially, the RMSD of *xa5* increased rapidly in the first 10 ns and gradually became stable at 15 ns Fig. 2c. The average gyration radii of *xa5* and *avrxa5* were found to be 14 Å and 39 Å respectively (Fig. 2b, d) which indicates *xa5* protein remained compact within the trajectory, but, *avrxa5* found be to more flexible during 20 ns MD. Although no significant change was observed in case of *xa5* protein but notable changes are observed in case of *avrxa5* protein during MD simulation (Fig. S4). The small helices in the C-terminal end mostly fluctuated during 10-15 ns which reported stable later during 15-20 ns. Root mean square fluctuations of C α atoms of both the system revealed that, *avrxa5* protein is much more flexible in aqueous solution during 20 ns than *xa5* (data not shown). Moreover, the potential energy of the both the systems was assessed as a function of time in order to measure the stability of the system. Both potential energy as well as total energy of *xa5* was energetically more stable as compared to *avrxa5*. The average changes are plotted and snapshots are extracted to see the structural changes during MD simulations at different time scale. Finally structural variation of both the models before and after refinement at different time scale was extracted (*i.e.*, 0, 5, 10, 15 and 20 ns) and plotted as shown in Fig. 3. In case of *xa5*, transition of a less stable small 3₁₀ helix to turn at the N-terminal helical region was observed, whereas, the dominated β -sheet C-terminal end remained intact reflecting the overall stability of the protein. But in case of *avrxa5*, notable deviations were observed in the flexible loop regions (coil/turn switching and transition of small helices to less turns) during 10-15 ns and remained stable after 15 ns where the helical repeats remains intact at their respective positions without any deviation from the original modeled protein. All of above characteristics conveyed after a 20 ns MD simulation, indicating the stability of both the model in aqueous solution. These average structures further were minimized in DS 3.5 for removing clashes and bad contacts and subsequently used for protein-protein docking. .

3.5 Protein-protein docking

The pathogen (Xoo) *avrxa5* (avirulence) gene, interacts with the corresponding host (rice plant) resistance gene (*xa5*) in a gene-for-gene manner and the specificity of *avrxa5* towards *xa5* was confirmed from the domain swapping experiment conducted by Zou et al. [5]. The *avrxa5* is a close structural homolog of TAL effector PthXo of Xoo which mediates recognition of target DNA is believed to follow same mode of recognition as that of PthXo. This recognition in turn might be triggering the manipulation host (rice) resistance gene expression which corresponds to the earlier work reports of Kay et al. [45]; Romer et al. [46] and Nissan et al. [47], where TAL effector of AvrBs3 family, from *Xanthomonas* spp. mimic eukaryotic transcription factors in turn induces plant gene expression. As such the recognition of *avrxa5* (a member of AvrBs3 family) by the host resistance gene (*xa5*) might be the probable basis of *xa5-avrxa5* (R-Avr) interaction and might be the solitary basis of transcriptional gene regulation in plant-pathogen interaction involving R-Avr gene in rice. To evaluate the direct-direct interaction (gene-for-gene) hypothesis, we docked the disease resistance protein with its corresponding avirulence (*avrxa5*) protein and the affinity of the projected binding surfaces on both the partners was verified using MD simulation. Protein-protein interaction study was performed using Cluspro: an automated, fast rigid-body docking and discrimination algorithm that 1) rapidly filters docked conformations; 2) ranks the conformations using clustering of computed pairwise RMSD values. It is also the first fully automated web-based program for docking proteins and was one of the top performers at CAPRI (Critical Assessment of Predicted Interactions) rounds 1-12, the community-wide experiment devoted to protein docking [48]. The multistage protocol implemented in Cluspro ranked each cluster by its number of members, where we selected three best clusters based on the ClusPro score of the center of the cluster and the lowest ClusPro score found in the cluster as shown in Table 2 (a). The best three ranked clusters are analyzed using DimPlot to identify the direct contacts between *xa5* and *avrxa5* proteins. Since HADDOCK is an information-driven docking approach, in order to drive the docking, the identified residues which form direct contact with the ligand (hydrogen and hydrophobic contacts) were further supplied as restraints for filtering through HADDOCK. Among the different clusters generated from HADDOCK, the best three complexes were screened based on the highest HADDOCK scores (*i.e.*, -179.2+-10.4, -152.1+-4.7 and -152.1+-7.0) (Table 2 (b)). Further these three best complexes were refined through molecular dynamics simulations using GROMACS.

3.6 MD simulation of protein-protein complex

The lack of experimental evidence on the direct-direct interaction of both the proteins, has guided us to delineate the probable mode of binding and mode of recognition between the corresponding partners through MD simulation of the best three complexes (*xa5-avrxa5*). Both the predicted models of *xa5* of rice and *avrxa5* of Xoo differed in their distribution of charged residues on their surfaces as evidenced from the electrostatic

surface potential calculated using APBS software (Fig. S5). The xa5 surface contains an extensive acidic patch spanning the N-terminal helical region whereas C-terminal region mostly dominated by basic amino acids. In contrast, the avrxa5 surface consists of a larger negative patch with a small positive patch. The net charge of xa5 is found to be -9, whereas that of avrxa5 is +2, where this difference in the distribution of charged residues in the two proteins is correlated to their total charge. To better understand the most probable mode of binding of corresponding partners, the affinity of the predicted binding surfaces on xa5 and avrxa5 was verified through MD simulations. To gauge the dynamics stability of each complex from each trajectory, RMSD, Rg and H-bond was taken into account, where complex III shows a linear RMSD, Rg (Fig. 4) and H-bond (Fig. 5) as compared to the rest two complexes (*i.e.*, complex I and II). Moreover, MD simulations reveals that xa5 binds with greater affinity to avrxa5 in complex II having an average of 9 intermolecular hydrogen bonds (H-bonds), which is significantly higher than complex I and III (Fig. 5). As compared to complex I and II, the average number of H-bonds and the total number of contacts within 5 Å of the interacting surfaces are prominent in complex II whereas, the H-bonds and non-bonded contacts between the xa5 and avrxa5 in complex I and II gradually decrease during simulation (Fig. 5). The decrease in ionic interactions might be due to the presence of discrete basic amino acids around avrxa5 that might have dominated its electronegativity, in turn aid in weakening interaction with the positive patch on xa5. Complex III presents a well conserved binding pocket where dominated strand region of xa5 that binds to C-terminal helical region of avrxa5 through a strong network of H-bond. Most importantly retention of three H-bonds *i.e.*, Lys56/Ala866, Lys56/Glu871, and Gly61/Phe85 throughout the simulation process signifies their importance in stabilizing the protein–protein complex and also in recognition (Table 3). In contrary, the average number of H-bonds in complex III is comparatively higher than others but as the interactions are in-consistent which further favors that complex II is more stable than others. Based on intrinsic dynamics stability and H-bond along with other ionic interactions, complex II is considered the most probable binding modes of xa5 and avrxa5 (Table 3), which has been studied in detail in the later section of the manuscript.

3.7 Molecular interaction of xa5 and avrxa5

To understand the behavior of molecular interaction before and after MD simulation, a comparative analysis was performed to map the minute changes. Microscopic observation in complex II showed that Lys56, Ser58, His62, Glu79 and Glu85 of xa5 forms strong H-bond with Phe869, Tyr867, Pro865, Gln860, Ile831 and Tyr867 of avrxa5 initially were broken during MD simulation. These losses of H-bond are can be compensated with through hydrophobic interactions and other ionic interactions. Importantly, N and NZ atom Lys56 (xa5) which interacts of oxygen (O) and OL atom of Ala866 and Glu871 (avrxa5) continuously formed strong hydrogen

bond signifying the importance of this residue in xa5-avrxa5 interaction. Similarly, Nitrogen atom Gly61 interacts with Oxygen atom Phe858 through strong H-bonds. In addition, several additional hydrophobic contacts were noticed in the xa5-avrxa5 protein-protein after MD simulations, but most of the contacts involving both the partners (Lys54/Phe869, Lys56/Phe858, Lys60/Glu857, Glu79/Arg833, Gln90/leu830, and Gly92/Ile831) were retained during MD simulation reflect their importance in protein-protein interaction. Apart from the H-bond and hydrophobic interactions, ionic contacts are found to prominent involving *i.e.*, Lys54/Asp870, Lys56/Glu871, His62/Glu857, Glu79/Arg833 and Glu79/Arg837 (xa5/avrxa5) with an intermolecular distance of 6 Å suggesting their probable role in R-Avr interaction (Table 3). Most interestingly, two crucial intermolecular contacts were observed (*i.e.*, salt bridges) between xa5 and its corresponding effector are seen, where the first one includes NZ atom Lys54 with OD2 of Asp870 whereas the second one involves NZ atom Lys56 with the OE2 of Glu871 with an atomic distance of 3.5 Å. As evidenced from Fig. 6 the helical C-terminal residues located at the curved in face of avrxa5 recognize the β-strand dominated region adjoins the strand dominated region of xa5 through hydrogen bond and hydrophobic interactions. In addition, almost 75% the C-terminal end of xa5 is deeply buried inside the cavity of helical region of C-terminal end and reinforced by H-bonds and electrostatic interactions. Our studies also show that some crucial residues viz., Lys54, Lys56, Ile59, Gly61, His62, Ser101 and Ser105 of xa5 in rice and Asp870, Ala868, Ala866, Glu871, His862, Phe858, Arg841, Leu856, Ala872 and Asp870 of avrxa5 in BB pathogen *Xoo* play a pivotal role in stability of docked complex. Lys54, Lys56, Gly61, His62 and Glu79 of xa5 were indispensable for tight anchoring with the repeats at C-terminal of avrxa5. The proposed mechanism of R-Avr interaction in rice and *Xoo* has been displayed in Fig. 7. Further to justify the stability and to analyze various fluctuations in polar and non-polar contacts in the III complexes, an attempt was made to correlate the differences between hydrophobic-hydrophobic, hydrophilic-hydrophilic, hydrophobic-hydrophilic and acidic-basic contacts of the interacting residues both in xa5 and avrxa5 before MD and after dynamics simulation (Fig. S6). Among the three complexes, complex-II showed less fluctuation in different contact types (*i.e.*, hydrophobic-hydrophilic and hydrophilic-hydrophilic) before and after MD, which perfectly correlate with the stability of the complex. Although, at present no experimental evidence exists regarding the key residues that play vital role xa5-avrxa5 protein-protein interaction, structural biology insights has uncovered key insights into R-Avr direct-direct interaction in rice and BB pathogen *Xoo*. With the advancements in protein expression technologies and in the instrumentation for biophysical characterization and crystallization of recombinant proteins, as well synchrotrons for X-ray diffraction from crystals and high-field solution nuclear magnetic resonance (NMR) can benefit structural biology to which can shed more light into diverse areas of biology including plant-pathogen interactions in near future.

3.8 Normal modes and essential dynamics

The function of proteins usually conveyed by the internal motion and elasticity which aid in proper folding, substrate/ligand recognition, specific aggregations, conformational adaptability etc. Normal mode (NMA) and essential dynamics (EDA) analysis are one of the foremost simulation techniques used to explore the above mentioned characteristics of a protein by identifying the essential degrees of freedom. To understand this in R-Avr complex, we investigated the complex motions and elasticity derived from the MD simulation by ProDy program. PCA analysis of complex-I revealed a significant motion is associated at the curled regions of the chain B (avrxa5). However, in chain B, the motions are comparatively less. The relaxed “S” shaped regions of the chain A (xa5) moved upwards and downwards at the lower and upper half arc of letter “S”, respectively. This resulted in forming a much compact structure with the convex tips facing each other. The N-terminal region of chain A also presented a prominent motion driving inwards. The PCA analysis showed at the end of MD simulation, the stretched chain A formed well refined compact conformation. To further investigate the dynamics of these regions, we calculated the elasticity of the complex using ANM. The ANM analysis of backbone also showed a substantial backbone motion at these regions. The peak regions at the convex sites in both halves shared an eminent elasticity pulling each other in opposite direction. The connecting region of both “C” shaped halves presented a long deep shaped and closely placed conformation (Fig. 8). In chain B, the motions were not so prominent, and showed a firm binding at the C-terminus during the MD simulation. In contrast to complex I, both the complex-II and III showed a minute deviation and displayed more distorted motion (Fig. 8).

3.9 Calculation of binding free energy by MM/PBSA

The binding free (BE) energies can be calculated from MD simulations in a variety of ways such as thermodynamic integration (TI) and one-step perturbation (OSP) approaches, which makes use of the free energy perturbation (FEP). Free energy calculations using MM-GBSA and MM-PBSA provide a comprehensive understanding of molecular recognition in protein-ligand and protein-protein interactions [49]. The MM/PBSA method considers a number of MD snapshots to calculate the binding free energy of protein-protein and protein-ligand complexes. In order to propose a more accurate interaction model of xa5 and avrxa5, the binding free energies were estimated for the three xa5-avrxa5 complexes. MM-PBSA calculations are performed on multiple structures collected from MD simulations. Typically, snapshots are selected at regular intervals from the 20 ns trajectory. A total of 500 complexes were selected for MM/PBSA calculations for each complex. The obtained binding free energies were decomposed into electrostatic, van der Waals, polar, and nonpolar solvation free energies (Table 4). The overall binding free energies of complexes I, II, and III are -263.6, -290.5, -354.253

kJ/mol, respectively. It should be noted that experimental binding affinity for xa5 and avrxa5 has not been reported till date for comparison. Although, Complex-III showed highest binding affinity but keeping the dynamics stability of R-Avr complexes, complex II showed a reasonable binding affinity which is largely governed by the nonpolar terms. Breakdown of the binding free energy into electrostatic, van der Waals, polar, and nonpolar solvation energies reveals that van der Waals energy (ΔG_{vdW}) is the major favorable contributor to protein-protein interaction. Altogether nonpolar solvation energies ($\Delta G_{nonpolar}$) make important contribution to binding. Furthermore, the van der Waals interaction energies dominate over nonpolar solvation terms. Overall binding energy analysis shows the van der Walls and nonpolar solvation terms have important contribution to the xa5 and avrxa5 binding. The polar solvation energy contributed more than threefold of the non-polar solvation terms towards molecular recognition. Energy components of the MM-PBSA calculation are insightful as they endow with insight into the free energy of binding (Table 4), but raises an important question is whether these components can be correlated with the experimentally determined thermodynamic parameters in near future.

Structural biology have revolutionized our understanding of plant-pathogen interactions in recent years, including: the identification of protein functions that were not apparent from sequences alone, the visualization of molecular interfaces of relevance to pathogen virulence and to plant immunity. The accessibility of these protein structures has not only provided the direct insights into molecular function but also enabled us to design new experiments to assess functions that could not or else be envisioned. To some extent where structures do not offer a direct link to functions, it is important to explore the biological context of protein activity. Despite the advances made, the impact of structural biology on plant-pathogen interactions is only just beginning to be felt. Looking forward, there are considerable challenges ahead, especially related to the increasing complexity of samples under study, as the greatest insights will come from the structures of multi-domain proteins or multi-protein complexes, some of which will be ligand-bound. Overcoming these challenges will require the integration of OMICS science including bioinformatics, genomics, cell biology and biochemical approaches in both model plant species and crops. Large-scale screens for protein-protein interactions are building the foundations of protein interactomes in plant cells. Biochemical approaches that include structural biology are expected to substantiate the biological relevance of these interactions and to highlight protein interfaces and activities that are crucial for function. As of now, the comparative modeling, molecular docking and MD simulations are used predominantly to study the key molecular insights in to the interactions between key molecules of innate immune system in plants. These computational methods have provided atomic insights into the phenomena of molecular recognition in the absence of experimental evidences. Although these methods

mostly rely on sophisticated algorithms and scoring functions, suitable care with better equipped high-throughput experiments with scalable hardware has significantly increased the accuracy of such predictions. It is expected that the translation of our findings into crop species will have a meaningful effect on achieving more durable resistance to pathogens in the field which is expected to contribute towards the meeting the challenges in global food security.

4. Conclusion

Unraveling the mechanism of plant-pathogen recognition and signal transduction leading to the induction of disease resistance responses has been the key factor for exploiting the innate immunity in plants. Recognition of the Avrxa5 (a member of Avr/Bs3 family of TAL effector) protein from bacterial blight pathogen *Xoo*, which forms the basis of interaction with the corresponding resistance protein xa5 is still poorly understood due to the lack of any experimental structures. In continuation of our earlier studies, we now predict the probable mode of binding of xa5 and avrxa5 through protein-protein docking, long-term MD simulations and MM/PBSA free energy calculations. Molecular docking and MD simulation reveal that C-terminal β -strand dominated region of xa5 binds to deep crevices of helical region of avrxa5 at the C-terminal end. Moreover, our study shows Lys54, Lys56, Gly61, His62 and Glu79 of xa5 are indispensable for tight anchoring with the repeats at C-terminal of avrxa5. Free energy calculation through QM/MM-PBSA approach are insightful as they provide insight into the free energy of binding where van der Walls and nonpolar solvation terms contributed significantly xa5 and avrxa5 binding. These results represent a first step towards a throughout characterization of the R-Avr recognition process. Site directed mutagenesis along with binding energy of mutants with the experimental free energy can aid in predicting the critical residues involved in interactions in both the partners. This is the first ever study of recognition of TAL effector by a eukaryotic transcription factor in mediating disease resistance. In combination with OMICS science, our result is expected to provide a new dimension in unraveling the plant-pathogen interaction at the molecular level.

Acknowledgment

We sincerely acknowledge support and help from Department of Agricultural Biotechnology at Assam Agricultural University, Jorhat for conducting research studies.

Conflict of interests

The authors declare that no conflict of interests exists.

References

1. H.H. Flor, Current Status of the Gene-For-Gene Concept, *Annu. Rev. Phytopathol.*, 9 (1971) 275-296.
2. E.H. Stukenbrock, B.A. McDonald, Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions, *Mol Plant Microbe Interact.*, 22 (2009) 371-80.
3. S. Schurch, C.C. Linde, W. Knogge, L.F. Jackson, B.A. McDonald, Molecular population genetic analysis differentiates two virulence mechanisms of the fungal avirulence gene NIP1, *Mol Plant Microbe Interact.*, 17 (2004) 1114-25.
4. M.H.A.J. Joosten, A.J. Coijnsen, P.J.G.M. de Wit, Host resistance to a fungal tomato pathogen lost by a single base-pair change in an avirulence gene, *Nature*, 367 (1994) 384-386.
5. H. Zou, W. Zhao, X. Zhang, Y. Han, L. Zou, G. Chen, Identification of an avirulence gene, *avrxa5*, from the rice pathogen *Xanthomonas oryzae* pv. *oryzae*, *Sci. China Life Sci.*, 53 (2010) 1440-49.
6. A. Iyer, S. McCouch, The rice disease resistance gene xa5 encodes a novel form of disease resistance, *Mol Plant Microbe Interact.*, 17 (2004) 1348-1354.
7. A.S. Iyer-Pascuzzi, S.R. McCouch, Recessive resistance genes and the *Oryza sativa*-*Xanthomonas oryzae* pv. *oryzae* pathosystem, *Mol. Plant Microbe Interact.*, 20 (2007) 731-739.
8. A.S. Iyer-Pascuzzi, H. Jiang, L. Huang, S.R. McCouch, Genetic and functional characterization of the rice bacterial blight disease resistance gene xa5, *Phytopathology*, 98 (2008) 289-295.
9. G.H. Jiang, Z.H. Xia, Y.L. Zhou, J. Wan, D.Y. Li, R.S. Chen, W.X. Zhai, L.H. Zhu, Testifying the rice bacterial blight resistance gene xa5 by genetic complementation and further analyzing xa5 (Xa5) in comparison with its homolog TFIIAGamma1, *Mol. Genet. Genomics*, 275 (2006) 354-366.
10. K. Gu, B. Yang, D. Tian, L. Wu, D. Wang, C. Sreekala, F. Yang, Z. Chu, G.L. Wang, F.F. White, Z. Yin, R gene expression induced by a type-III effector triggers disease resistance in rice, *Nature*, 435 (2005) 1122-1125.
11. D.O. Niño-Liu, P.C. Ronald, A.J. Bogdanove, *Xanthomonas oryzae* pathovars: model pathogens of a model crop, *Mol. Plant Pathol.*, 7 (2006) 303-24.
12. S. Kay, U. Bonas, How *Xanthomonas* type III effectors manipulate the host plant, *Curr. Opin. Microbiol.*, 12 (2009) 37-43.
13. B. Dehury, M. Sahu, K. Sarma, J. Sahu, P. Sen, M.K. Modi, G.D. Sharma, M.D. Choudhury, M. Barooah, Molecular phylogeny, homology modeling, and molecular dynamics simulation of race-specific bacterial blight disease resistance protein (xa5) of rice: a comparative agriproteomics approach, *OMICS*, 17 (2013) 423-38.
14. I. Letunic, T. Doerks, P. Bork, SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.*, 40 (2012) D302-305.
15. T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nat. Methods*, 8 (2011) 785-786.
16. S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25 (1997) 3389-3402.
17. F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol.*, 7 (2011) 539.

18. K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0, *Mol. Biol. Evol.*, 30 (2013) 2725-29.
19. N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4 (1987) 406-25.
20. G.M. Boratyn, A.A. Schäffer, R. Agarwala, S.F. Altschul, D.J. Lipman, T.L. Madden, Domain enhanced lookup time accelerated BLAST. *Biol. Direct.*, 7 (2012) 12.
21. M.A. Kurowski, J.M. Bujnicki, GeneSilico protein structure prediction meta-server, *Nucleic Acids Res.*, 31 (2003) 3305-7.
22. B. Wallner, A. Elofsson, Can correct protein models be identified? *Protein Sci.*, 12 (2003) 1073-86.
23. A.N. Mak, P. Bradley, R.A. Cernadas, A.J. Bogdanove, B.L. Stoddard, The crystal structure of tal effector pthx01 bound to its DNA target, *Science*, 335 (2012) 716-719.
24. D. Deng, C. Yan, X. Pan, M. Mahfouz, J. Wang, J.K. Zhu, Y. Shi, N. Yan, Structural basis for sequence-specific recognition of DNA by TAL effectors, *Science*, 335 (2012) 720-723.
25. R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK – a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.*, 26 (1993) 283-291.
26. C. Colovos, T.O. Yeates, Verification of protein structures: Patterns of non-bonded atomic interactions. *Protein Sci.*, 2 (1993) 1511-1519.
27. D. Eisenberg, R. Luthy, J.U. Bowie, VERIFY3D: assessment of protein models with three-dimensional profiles, *Methods Enzymol.*, 277 (1997) 396-404.
28. M. Wiederstein, M.J. Sippl, ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Res.*, 35 (2007) W407-410.
29. V.B. Chen, W.B. Arendall 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr. D Biol. Crystallogr.*, 66 (2010) 12-21.
30. B. Wallner, A. Elofsson, Pcons5: combining consensus, structural evaluation and fold recognition scores, *Bioinformatics*, 21 (2005) 4248-54.
31. L. Willard, A. Ranjan, H. Zhang, H. Monzavi, R.F. Boyko, B.D. Sykes, D.S. Wishart, VADAR: a web server for quantitative evaluation of protein structure quality, *Nucleic Acids Res.*, 31 (2003) 3316-19.
32. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, *Bioinformatics*, 29 (2013) 845-54.
33. C.J. Camacho, S.Vajda, Protein-protein association kinetics and protein docking, *Curr. Opin. Struct. Biol.*, 12 (2002) 36-40.
34. C. Dominguez, R. Boelens, A.M. Bonvin, HADDOCK: a protein-protein docking approach based on biochemical or biophysical information, *J. Am. Chem. Soc.*, 125 (2003) 1731-1737.
35. D. Kozakov, D. Beglov, T. Bohnuud, S.E. Mottarella, B. Xia, D.R. Hall, S. Vajda, How good is automated protein docking? *Proteins*, 81 (2013) 2159-2166.
36. Laskowski R.A., M.B. Swindells, LigPlot⁺: multiple ligand-protein interaction diagrams for drug discovery, *J. Chem. Inf. Model.*, 51 (2011) 2778-86.
37. A. Bakan, L.M. Meireles, I. Bahar, ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics*, 27 (2011) 1575-1577.

38. N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc. Natl. Acad. Sci. U.S.A.*, 98 (2001) 10037-10041.
39. I. Massova, P.A. Kollman, Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies, *J. Am. Chem. Soc.*, 121 (1999) 8133-43.
40. P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham 3rd, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, *Acc. Chem. Res.*; 33 (2000) 889-97.
41. I. Massova, P.A. Kollman, Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding, *Perspect. Drug Discov. Des.*, 18 (2000) 113-135.
42. D. Spiliotopoulos, A. Spitaleri, G. Musco, Exploring PHD fingers and H3K4me0 interactions with molecular dynamics simulations and binding free energy calculations: AIRE-PHD1, a comparative study, *PLoS One*, 7 (2012) e46902.
43. S.P. Brown, S.W. Muchmore, Large-scale application of high-throughput molecular mechanics with Poisson-Boltzmann surface area for routine physics-based scoring of protein-ligand complexes, *J. Med. Chem.*, 52 (2009) 3159-3165.
44. T. Kawabata, MATRAS: A program for protein 3D structure comparison, *Nucleic Acids Res.*, 31 (2003) 3367-69.
45. S. Kay, S. Hahn, E. Marois, G. Hause, U. Bonas, A bacterial effector acts as a plant transcription factor and induces a cell size regulator, *Science*, 318 (2007) 648-651.
46. P. Romer, S. Hahn, T. Jordan, T. Strauß, U. Bonas, T. Lahaye, Plant-pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene, *Science*, 318 (2007) 645-648.
47. G. Nissan, S. Manulis-Sasson, D. Weinthal, H. Mor, G. Sessa, I. Barash, The type III effectors HsvG and HsvB of gall forming *Pantoea agglomerans* determine host specificity and function as transcriptional activators. *Mol. Microbiol.*, 61 (2006) 1118-31.
48. S. Vajda, D. Kozakov, Convergence and combination of methods in protein-protein docking, *Curr. Opin. Struct. Biol.*, 19 (2009) 164-70.
49. B. Wang, L. Li, T.D. Hurley, S.O. Meroueh, Molecular recognition in a diverse set of protein-ligand interactions studied with molecular dynamics simulations and end-point free energy calculations, *J. Chem. Inf. Model.*, 53 (2013) 2659-70.

Figures legends**Fig. 1**

The overall structural features of modeled xa5 of rice **a** and avrxa5 of BB pathogen Xoo **b**.

Fig. 2

Stability parameters for the models of R (xa5) and Avr (Avrxa5) protein as a function of simulation time. **a** Backbone RMSD of xa5 and avrxa5 during 20 ns; **b** Radius of gyration (R_g) of modeled xa5 and avrxa5 protein during 20 ns. Red color line depicts xa5 protein whereas Brown color represents avrxa5 model.

Fig. 3

Snapshots obtained for modeled R **a** and Avr **b** protein during MD simulation at different time scale.

Fig. 4

The stability parameters for the docked complexes (R-Avr) of xa5 and avrxa5 proteins. **a** Backbone RMSD with respect to time (20ns). **b** R_g , during 20 ns molecular dynamics simulations.

Fig. 5

Analysis of total number of intermolecular H-bonds formed between xa5 and avrxa5 in three different docking complexes. **a** Complex I, **b** Complex II, and **c** Complex III. Complex II shows highest stability in terms of average number of H-bonds, signifying it as the most reliable interaction model.

Fig. 6

Intermolecular interactions of xa5 and avrxa5 in three different complexes. **a** Interaction of xa5 with avrxa5 in complex I; **b** Molecular interaction R-Avr protein in complex-II; **c** Interaction of xa5 with the active site amino acids of avrxa5 protein. The hydrophobic amino acids are represented in balls and are colored according to their charges. The H-bond forming amino acids donor and acceptor marked in different colors and displayed as sticks.

Fig. 7

Proposed representation of interaction of xa5 with avrxa5 (i.e., R-Avr molecular interaction) in rice and BB pathogen Xoo. **a** The donor (xa5) and acceptor (avrxa5) amino acids involved H-bonding is shown as sticks with different colors and the H-bond has been displayed with red dotted lines. **b** Hydrophobic contacts with their atomic distance has been displayed and marked in different in red and blue circles.

Fig. 8

Essential dynamics analysis of R-Avr (xa5-avrxa5) protein-protein complexes. **a** Elasticity analysis of xa5-avrxa5 complexes using anisotropic network model in ProDy. **b** Principal component analysis (PCA) of xa5-avrxa5 complexes model from the top eigenvector derived from the 20 ns MD trajectory using ProDy. The first and last frame is shown as ribbon conformation.

Table 1 Model Validation Report of Avrxa5 and Xa5

Servers		Avrxa5	Xa5
Procheck	Most favored regions (%)	92.1 %	90.9%
	Additionally allowed Regions (%)	6.8 %	9.1%
	Generously allowed Regions (%)	1.1 %	0.0%
	Disalloweed regions (%)	0.0 %	0.0%
	Overall G-factor	-0.05	-0.03
Verify3D	Averaged 3D-1D Score > 0.2	86.04%	43.93%
ERRAT	Overall Quality	93.373	93.684
ProSA	Z-Score	-15.95	-2.67
ProQ	LG score	7.306	1.646
	MaxSub	0.730	0.195
MolProbity	Residues with bad bonds (%)	-0.33 %	-0.33 %
(Z-Score)	Residues with bad angles (%)	1.19 %	-0.69 %
	Ramachandran plot outliers	0.76 %	-0.11 %
Vadar	Standard deviation of χ_1 pooled	-0.25	3.00
	Mean H-bond energy	1.32	-1.71
	Generously allowed Ω angles (%)	-1.60	-1.60
	Packing defects (%)	0.01	-2.18
	Percentage of 95% buried residues	0.44	-2.50

Table 2 Docking scores of xa5 and avrxa5 obtained from Cluspro server and HADDOCK

(A) Scores from Cluspro server

Complexes	No. of complexes in Cluster	Centre Score	Lowest energy Score
Complex-I	155	-958.2	-1286.6
Complex-II	141	-1023.0	-1170.1
Complex-III	121	-1036.1	-1195.1

(B) Docking scores and energy components derived from HADDOCK

Scores	Complex-I	Complex-II	Complex-III
HADDOCK score	-179.2 +/- 10.4	-152.1 +/- 4.7	-152.1 +/- 7.0
Van der Waals energy	-67.2 +/- 4.5	-65.6 +/- 10.1	-60.8 +/- 10.5
Electrostatic energy	-263.6 +/- 32.2	-223.3 +/- 44.9	-195.1 +/- 22.3
Desolvation energy	-71.2 +/- 8.2	-67.1 +/- 3.7	-53.6 +/- 8.6
Restraints violation energy	119.5 +/- 21.59	253.1 +/- 28.39	13.2 +/- 18.23
Buried Surface Area	2168.8 +/- 88.8	1855.3 +/- 145.1	1941.5 +/- 124.4
Z-Score	-2.1	-2.0	-1.7

Table 3

Molecular interaction of Xa5-Avrxa5 before and after 20 ns MD simulations

(A) Complex-I

Before MD Simulation			After MD Simulation		
Hydrogen Bond					
Xa5	Avrxa5	length	Xa5	Avrxa5	length
Gln40:NE2	Glu829:OE1	3.05	Glu47:OE1	Arg832:NE	2.63
Lys43:HZ1	Leu830:O	2.75	Glu47:OE1	Arg832:NH2	2.74
Glu47:OE1	Arg832:NE	3.28	Glu47:OE2	Arg832:NH1	3.06
Glu47:OE1	Arg832:NH2	3.05	Asn51:ND2	Arg832:NE	3.29
Lys54:NZ	Gln875:O	2.91			
Lys54:NZ	Cys861:O	2.90			
Ile59:N	Asp870:OD1	2.77			
Lys60:NZ	Asp870:OD2	2.69			
Lys102:NZ	Pro865:O	2.69			
Hydrophobic Interaction within 5 Angstroms					
Ala2	Leu830		Val53	Ala872	
Ile59	Phe869		Leu63	Phe869	
Leu63	Phe869		Tyr66	Ala868	
Trp73	Tyr876		Trp73	Tyr867	
Phe75	Phe869		Trp73	Ala868	
Leu77	Phe869		Phe75	Phe869	
Ala98	Tyr867		Leu77	Phe869	
Leu104	Val855		Leu104	Phe858	
Ionic Interaction within 6 Angstroms					
Glu47	Arg832		Glu47	Arg832	
Glu50	His862		Lys60	Glu871	
Lys60	Asp870				
Aromatic Interaction within 4.5 and 7 Angstroms					
Trp73	Tyr867	6.41	Phe75	Phe869	6.11
Phe75	Phe869	5.14			
Cation-Pi Interaction within 6 Angstroms					
Lys54	Phe876	4.56	Lys54	Phe876	5.99
Lys102	Tyr867	5.04			

(B) Complex-II

Before MD Simulation		After MD Simulation	
Hydrogen Bond			

Xa5	Avrxa5	length	Xa5	Avrxa5	length
Lys56:N	Phe869:O	3.05	Lys54:NZ	Asp870:OD2	2.66
Lys56:N	Ala866:O	2.75	Lys56:N	Ala868:O	2.73
Lys56:NZ	Glu871:OE1	2.55	Lys56:NZ	Glu871:OE1	3.26
Lys56:O	Tyr867:OH	2.79	Ile59:O	His862:NE2	2.73
Ser58:OG	Pro865:O	2.80	Gly61:HN	Phe858:OD1	2.92
Gly61:N	Phe858:O	2.78	His62:NE2	Leu856:O	3.02
His62:NE2	Gln860:NE2	3.04	Ser105:HN	Asp870:OD1	2.66
Glu79:OE1	Ile831:O	2.89			
Glu85:N	Tyr867:OH	2.94			
Hydrophobic Interaction within 5 Angstroms					
Val57	Tyr867		Ala98	Ala872	
Leu63	Phe876				
Tyr66	Phe876				
Phe75	Phe876				
Ionic Interaction within 6 Angstroms					
Lys54	Asp870		Lys54	Asp870	
Lys56	Glu871		Lys56	Glu871	
His62	Glu857		His62	Glu857	
Glu79	Arg832		Glu79	Arg833	
Glu79	Arg833		Glu79	Arg837	
Aromatic Interaction within 4.5 and 7 Angstroms					
-	-	-	-	-	-
Cation-Pi Interaction within 6 Angstroms					
Lys60	Phe859	5.60	Lys60	Phe859	4.21

(C) Complex-III

Before MD Simulation			After MD Simulation		
Hydrogen Bond					
Xa5	Avrxa5	length	Xa5	Avrxa5	length
Tyr7:OH	His864:NE2	3.10	Tyr7:OH	His864:O	3.34
Glu22:OE2	Thr528:OG1	2.63	Glu47:O	Arg845:HH12	2.64
Thr28:OG1	Gly526:O	2.74	Leu49:O	His862:HE2	3.11
Lys54:NZ	His862:ND1	2.84	Glu50:OE1	His864:HN	3.38

Lys54:NZ	Phe858:O	2.75	Glu50:OE2	Ser863:OG	2.66
Lys56:NZ	Tyr867:OH	2.78	Gln52:HE22	Gln851:O	2.96
Lys102:NZ	Ala872:O	2.75	Leu104:O	Ser863:HN	3.04
Lys102:NZ	Glu871:OE1	2.60			
Lys102:NZ	Glu871:OE2	3.24			
Hydrophobic Interaction with in 5 Angstroms					
Ile59	Phe869		Phe4	Ala866	
Leu63	Phe869		Leu6	Ala866	
Trp73	Ala868		Leu6	Ala868	
Phe75	Ala868		Ile59	Ala868	
Phe75	Phe869		Leu63	Phe869	
Leu77	Phe869		Tyr66	Phe869	
Ala98	Ala868		Trp73	Phe869	
			Phe75	Ala868	
			Phe75	Phe869	
			Leu77	Phe869	
			Ile96	Ala868	
			Ala98	Ala866	
			Leu104	Pro865	
Ionic Interaction with in 6 Angstroms					
Glu50	His864		Glu47	Arg837	
Lys102	Glu871		Glu47	Arg841	
			Glu47	Arg845	
			Glu50	Arg837	
			Glu50	Arg841	
			Glu50	His864	
			Lys102	Glu871	
Aromatic Interaction with in 4.5 and 7 Angstroms					
Phe75	Phe869	5.72	Tyr66	Phe869	5.38
			Trp73	Phe869	4.87
			Phe75	Phe869	4.87
Cation-Pi Interaction with in 6 Angstroms					
Lys56	Tyr867	5.23	Lys54	Phe858	5.20

*Conserved interacting residues before and after MD simulation has been highlighted bold font

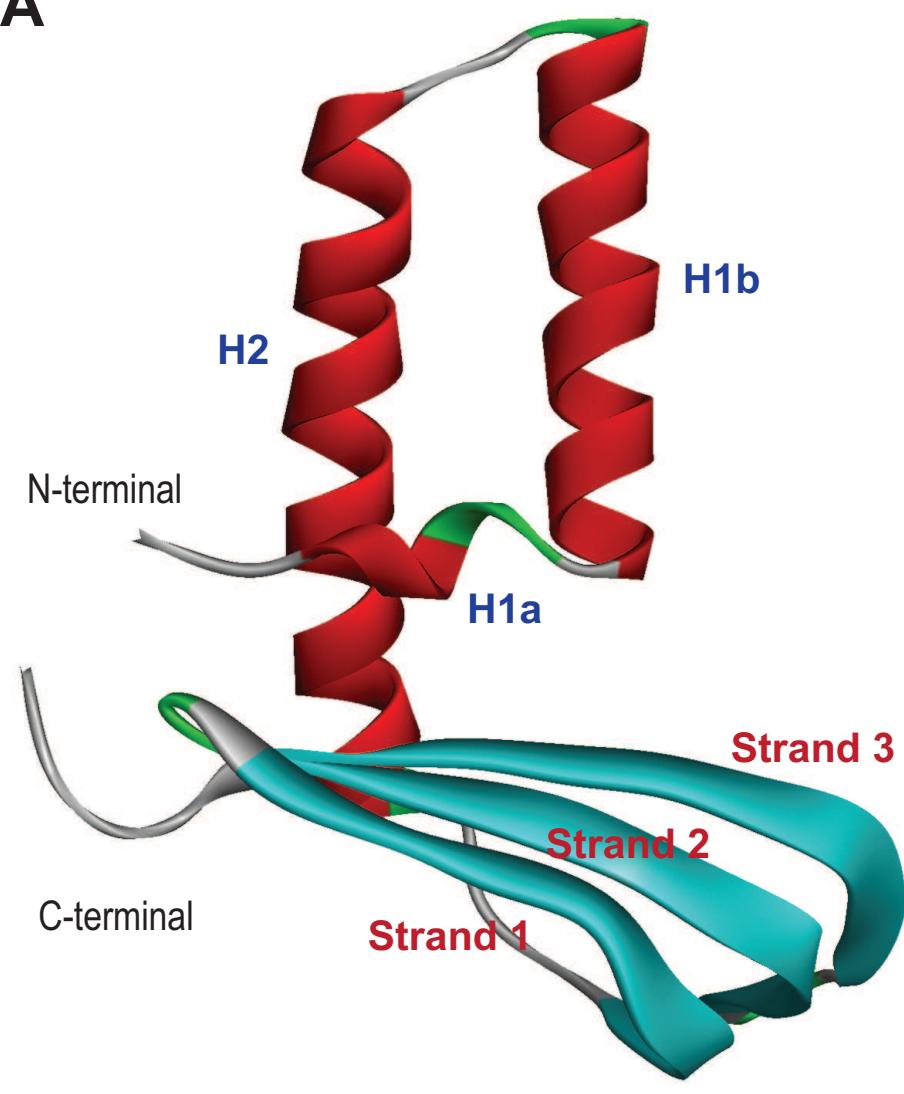
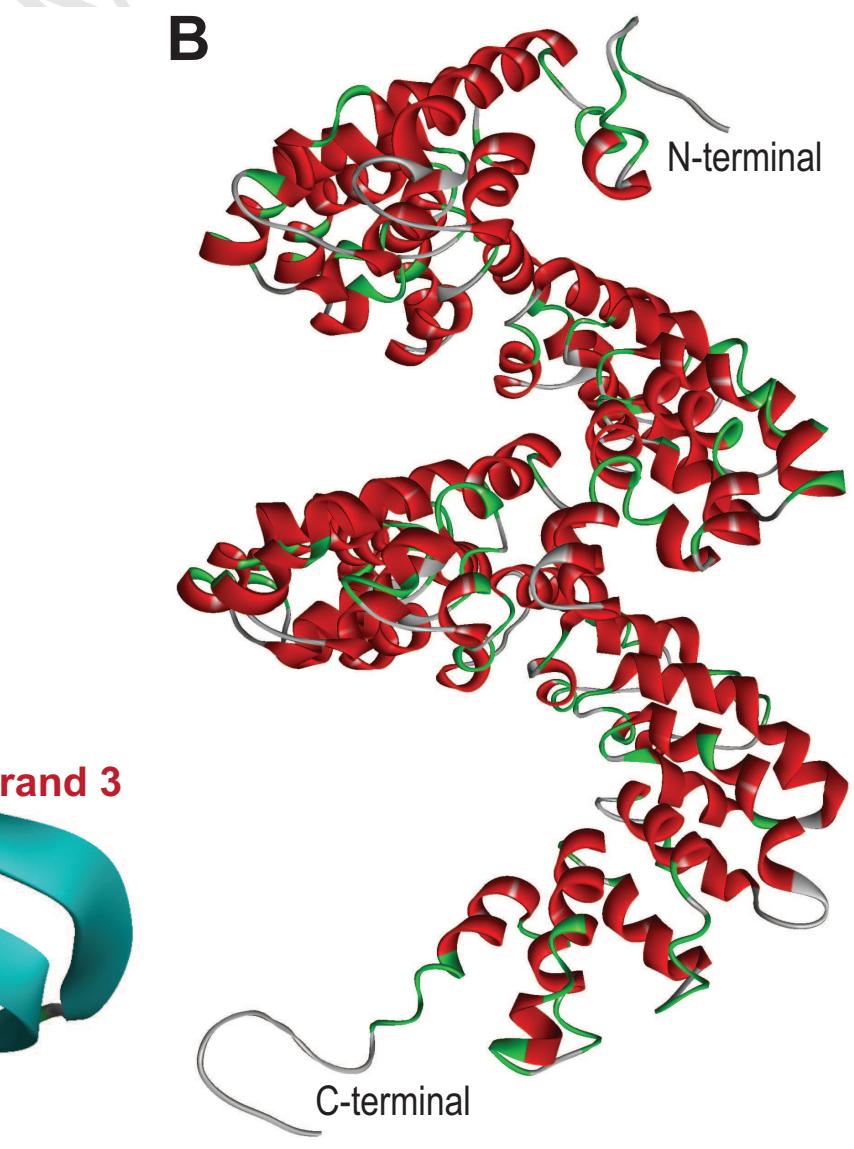
Table 4

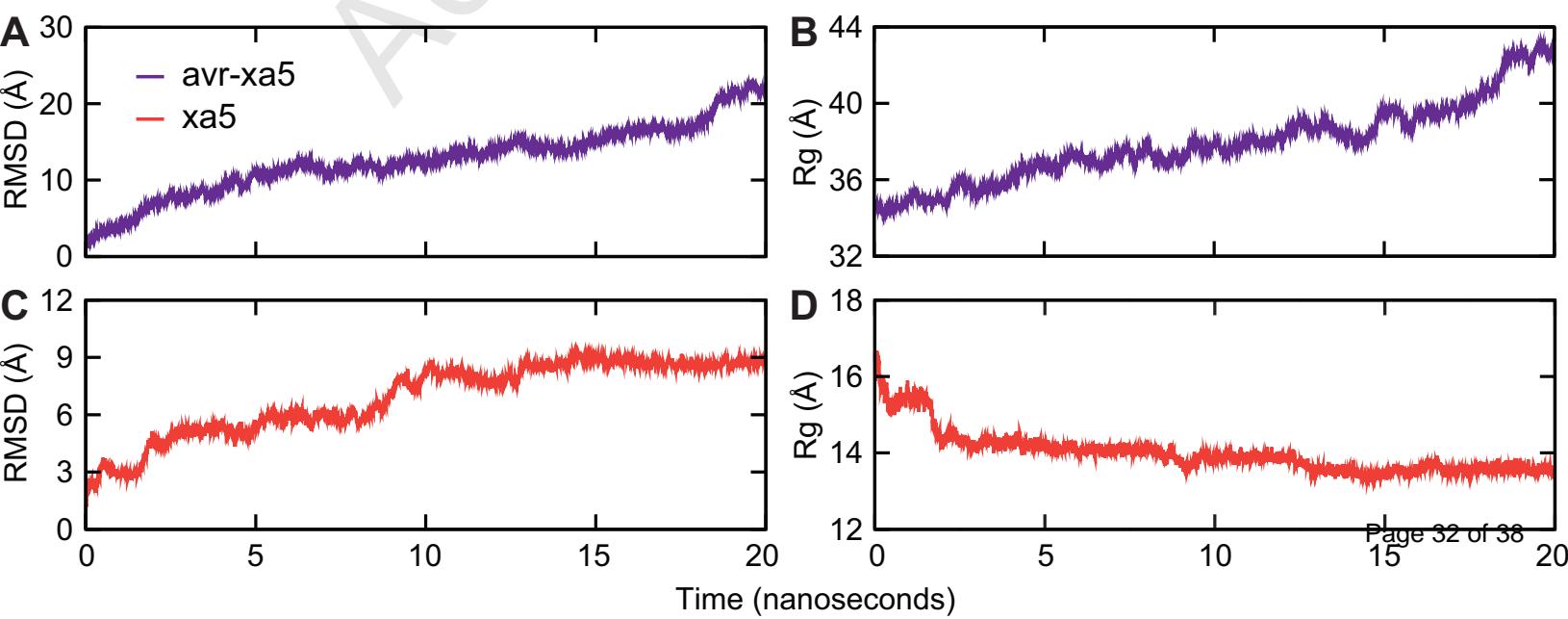
Binding energy (BE) calculation of Xa5-Avrxa5 complexes (complex-I, complex-II and complex-III)

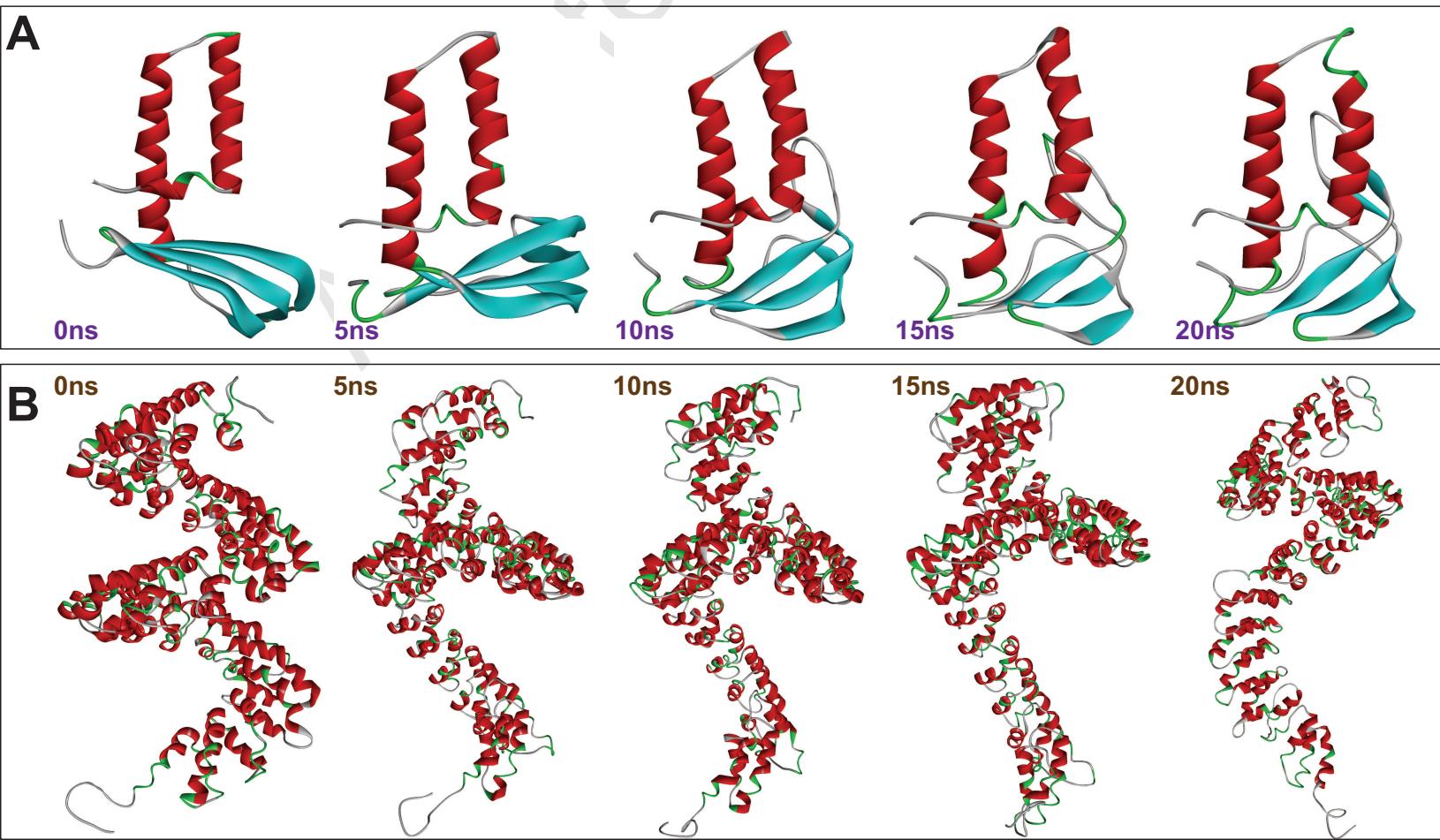
Energy Term	Complex I	Complex II	Complex III
ΔG_{bind}	-263.6 +/- 84.9	-290.5 +/- 80.3	-354.253 +/- 65.9
ΔG_{coul}	402.4 +/- 12.1	500.7 +/- 10.9	2361.0 +/- 22.6
ΔG_{ps}	-45.7 +/- 11.6	-171.6 +/- 11.8	-2071.2 +/- 23.5
ΔG_{polar}	356.7	329.1	289.8
ΔG_{vdW}	-565.7 +/- 85.1	-571.6 +/- 80.4	-603.6 +/- 66.1
ΔG_{nps}	-54.6 +/- 1.4	-48.0 +/- 1.1	-40.4 +/- 2.6
$\Delta G_{nonpolar}$	-620.3	-619.5	-664.0

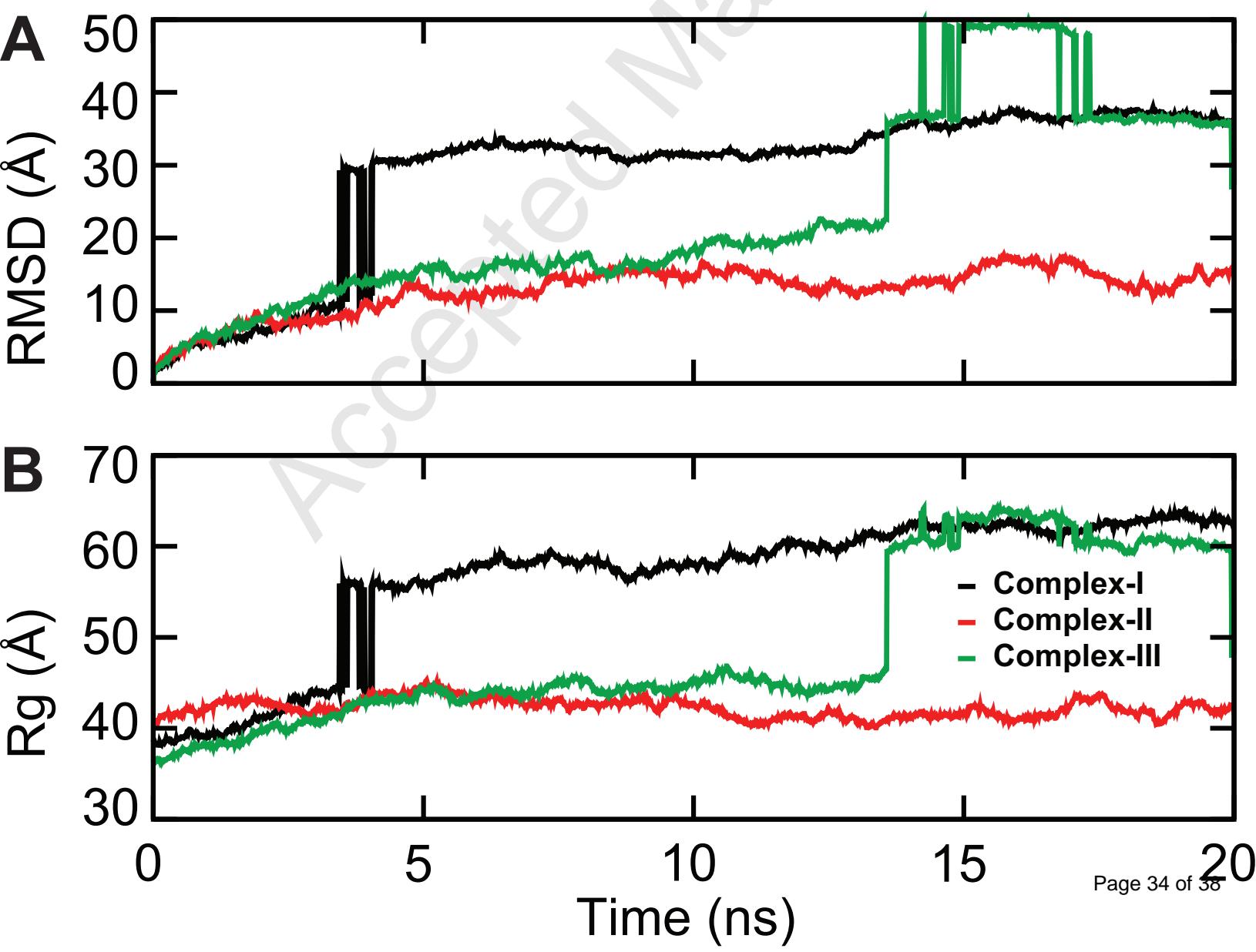
 ΔG_{bind} = Binding free energy. ΔG_{coul} = Electrostatic energy. ΔG_{ps} = Polar solvation energy. ΔG_{polar} = Polar term ($\Delta G_{coul} + \Delta G_{ps}$). ΔG_{vdW} = van der Waals energy. ΔG_{nps} = Nonpolar solvation energy. $\Delta G_{nonpolar}$ = Nonpolar term ($\Delta G_{vdW} + \Delta G_{nps}$).

(+/- Numbers in parenthesis indicate standard errors)

A**B**







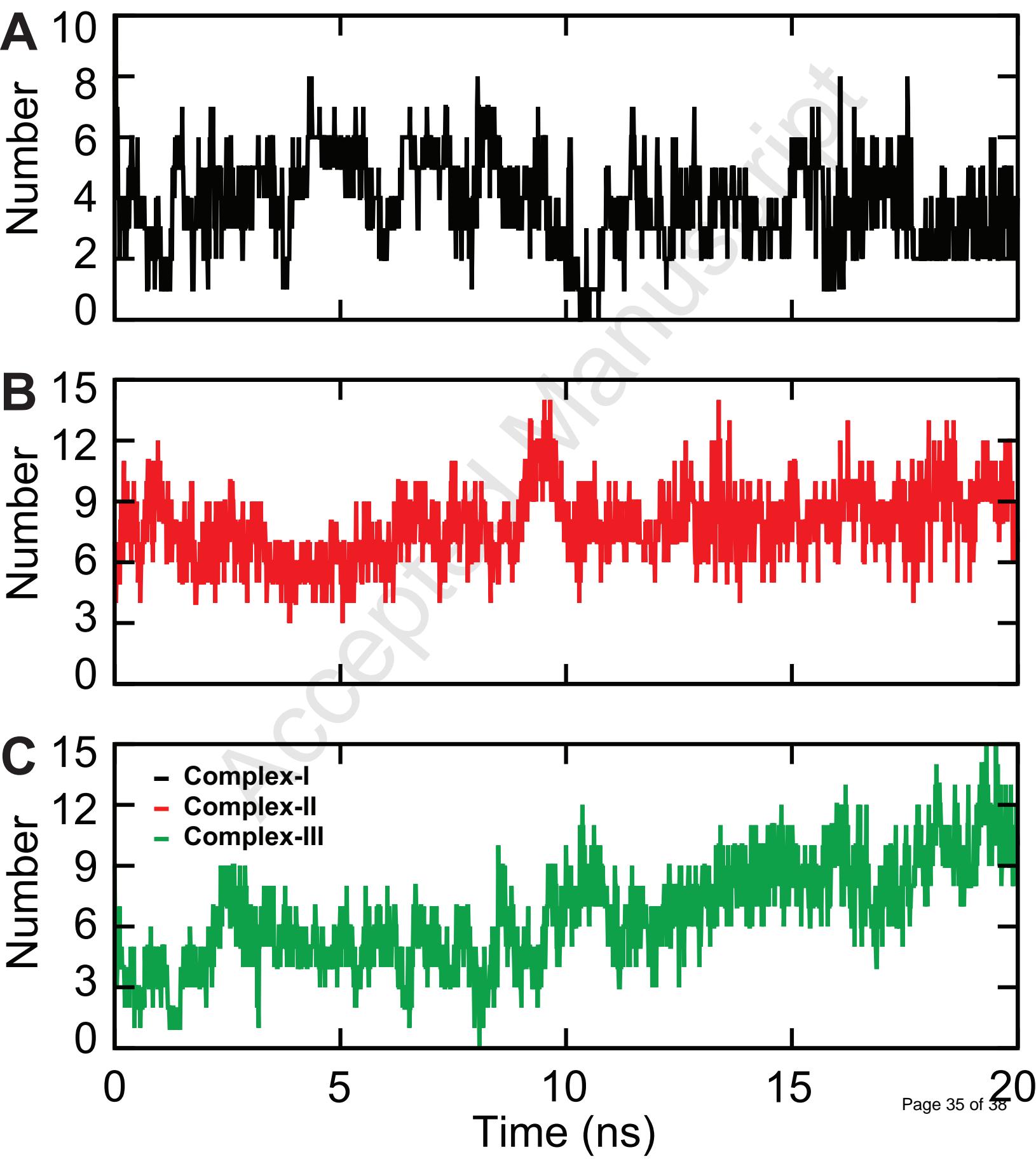


Figure 6