

RESEARCH

Open Access



In-silico study of E169G and F242K double mutations in leucine-rich repeats (LRR) polygalacturonase inhibiting protein (PGIP) of *Gossypium barbadense* and associated defense mechanism against plant pathogens

MURMU Sneha^{1†}, RASHMI Mayank^{1†} , NAGRALE Dipak T.^{2†}, KOUR Tejasman³, SINGH Mahender Kumar⁴, CHAURASIA Anurag⁵, BEHERA Santosh Kumar⁶, SHANKAR Raja⁷, RANJAN Rajiv⁸, JHA Girish Kumar¹, GAWANDE Shailesh P.², HIREMANI Neelakanth S.², PRASAD Y. G.² and KUMAR Sunil^{1*}

Abstract

Background Polygalacturonase inhibiting proteins (PGIPs) play a pivotal role in plant defense against plant pathogens by inhibiting polygalacturonase (PG), an enzyme produced by pathogens to degrade plant cell wall pectin. PGIPs, also known as leucine-rich repeat pathogenesis-related (PR) proteins, activate the host's defense response upon interaction with PG, thereby reinforcing the host defense against plant pathogens attacks. In Egyptian or extra-long staple cotton (*Gossypium barbadense*), the interaction between PGIP and PG is one of the crucial steps in the defense mechanism against major pathogens such as *Xanthomonas citri* pv. *malvacearum* and *Alternaria macrospora*, which are responsible for bacterial leaf blight and leaf spot diseases, respectively.

Results To unravel the molecular mechanisms underlying these PR proteins, we conducted a comprehensive study involving molecular modeling, protein-protein docking, site-specific double mutation (E169G and F242K), and molecular dynamics simulations. Both wild-type and mutated cotton PGIPs were examined in the interaction with the PG enzyme of a bacterial and fungal pathogen. Our findings revealed that changes in conformations of double-mutated residues in the active site of PGIP lead to the inhibition of PG binding. The molecular dynamics simulation studies provide insights into the dynamic behaviour and stability of the PGIP-PG complexes, shedding light on the intricate details of the inhibitory and exhibitory mechanism against the major fungal and bacterial pathogens of *G. barbadense*, respectively.

Conclusions The findings of this study not only enhance our understanding of the molecular interactions between PGs of *Xanthomonas citri* pv. *malvacearum* and *Alternaria macrospora* and PGIP of *G. barbadense* but also present a potential strategy for developing the disease-resistant cotton varieties. By variations in the binding affinities

[†]Murmu S, Rashmi M, and Nagrale DT contributed equally to this work.

*Correspondence:

Kumar Sunil

Sunil.Kumar7@icar.gov.in; skybiotech@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

of PGs through specific mutations in PGIP, this research offers promising avenues for the development of enhanced resistance to cotton plants against bacterial leaf blight and leaf spot diseases.

Keywords Polygalacturonase inhibiting proteins, Polygalacturonase, Plant-pathogen interaction, Protein-protein interaction, Docking, Molecular dynamics simulation

Introduction

Cotton, revered as “The white gold” and “The king of fibers,” stands as a valuable cash crop cultivated across India, China, the United States, Australia, Egypt and other African tropics (Fortucci 2002). In particular, India holds the largest area under cotton cultivation, reaching 13.049 million hectares during the 2022–2023 season and ranks among the world’s top cotton producers as per Indian Council of Agricultural Research–Central Institute for Cotton Research (ICAR-CICR 2022). Notably, *Gossypium barbadense*, a species of cotton known for its extra-long staple fibers (ELS), contributes significantly to the specialty of cotton fabrics and textiles, comprising 5% of the world’s cotton production (Hu et al. 2019). To maximize cotton yield and quality, modern techniques such as genetic and molecular breeding, genetic engineering, and disease management have been extensively explored. However, cotton faces challenges from various phytopathogenic fungi and bacteria, causing foliar and boll rot diseases that significantly vary with agroclimatic zones and weather conditions. For instance, in Brazil, cotton boll rot disease can lead to productivity losses of 20%–30% (Antonio Zancan et al. 2013), while in the USA, bacterial boll rot disease resulted in the significant yield losses of 100 153 to 182 708 bales during 2010–2011 (Goldberg et al. 2009). Furthermore, in South Carolina, USA, upland cotton (*Gossypium hirsutum*) experienced 10%–15% yield losses due to seed and boll rot disease (Hudson 2000; Hollis 2002). In India, diseases like *Alternaria* blight (*Alternaria macrospora*) and bacterial blight (*Xanthomonas axonopodis* pv. *malvacearum*) pose substantial threats, causing annual yield losses of up to 26% and 30%, respectively (Chattannavar et al. 2006). These challenges highlight the urgency to address these challenging bacterial leaf blight and boll rot diseases caused by *Xanthomonas citri* pv. *malvacearum* and *Alternaria macrospora*, which significantly reduce cotton production and productivity. However, in-depth studies on these major diseases are lacking, thus hindering the development of resistance breeding strategies to improve cotton productivity (Shete et al. 2018). Consequently, transgenic approaches incorporating novel genes offer promising solutions to enhance the cotton genome and combat these biotic stress challenges in

cotton production and cultivation in climate change scenarios.

Polygalacturonase (PG) inhibiting proteins (PGIPs) are the crucial components found in plant cell walls, playing a significant role in inhibiting enzymes known as endo-polygalacturonases produced by certain plant pathogens and insects. Structurally, PGIPs are composed of leucine-rich repeats (LRRs) organized into β -sheets, facilitating their interaction with PGs. This interaction triggers defense responses in host plants by the accumulation process of oligogalacturonides (OGs), which in turn induce the synthesis of defense compounds and thereby the expression of defense-related genes. Despite the lack of detailed structural information about the PG-PGIP complex, studies have been emphasized on the importance of PGIPs in plant defense. Furthermore, overexpression of PGIPs genes in host crop plants has shown promise in protecting against insect pests and plant pathogens.

Interestingly, PGs exhibit considerable variability in terms of primary structure, specificity, pH optimum, substrate preference, and mode of action (De Lorenzo et al. 2002). These plant cell wall-degrading enzymes (PCWDEs) include both exo- and endo-forms of PG, pectate lyase, and pectin methylesterase (PME), collectively known as pectinases. Produced by bacterial or fungal pathogens during the infection process, these enzymes plays a vital role in certain plant-pathogen interactions (Rodriguez-Palenzuela et al. 1991; Brown et al. 1992; Somai-Jemmali et al. 2017; Singh et al. 2019). PGs are considered major pathogenic factors in various fungal and bacterial pathogens, including *Botrytis cinerea*, *A. citri*, and *Claviceps purpurea* (Oeser et al. 2002), *Agrobacterium tumefaciens* (Rodriguez-Palenzuela et al. 1991), and *Ralstonia solanacearum* (Huang and Allen 2000; Kumar et al. 2022). Various necrotrophic pathogens harbor multiple PG genes within their genomes (Amselem et al. 2011). PGs play a critical role in virulence, as evidenced by gene-deletion studies conducted in *Botrytis cinerea* (Have et al. 1998) and enzymatic as well as expression studies in *Sclerotinia sclerotiorum* (Kasza et al. 2004; Li et al. 2004). Similarly, triggering rapid activation of host plant defense responses is crucial to combat infection by phytopathogenic bacteria (Ge et al. 2019; Rathinam et al. 2019). Nonetheless, the significance of pectinases has been extensively explored in soft rot diseases caused

by genus *Erwinia* through physiological and genetic studies (Barras et al. 1994; Collmer and Keen 1986). Moreover, in bacterial disease pathogenesis, endo-PG plays a pivotal role in the colonization and virulence of tomato plants by the plant pathogenic bacterium, *Ralstonia solanacearum* (Huang et al. 2000). However, inhibiting PGs with PGIPs may not always be effective against bacterial PGs or other pectinolytic enzymes of microbial or host plant origin (Cervone et al. 1990). Recently, the molecular mechanism of pathogenesis-related (PR) proteins was investigated through molecular modeling, protein-protein docking, directed mutagenesis, and molecular dynamics simulations of PGIP from banana with phytopathogenic bacterium *Erwinia carotovora* PG (ecPG), the causative agent of soft rot and rhizome rot diseases in banana and many cultivated plants (Kumar et al. 2020). Nevertheless, the role of fungal pectinases derived from phytopathogens in host-pathogen interactions remains poorly understood and documented. Furthermore, numerous fungal genes encoding endo-PGs, crucial elements in pathogenesis, have been cloned and characterized from *Aspergillus niger*, *Aspergillus tubingensis*, *Aspergillus parasiticus*, *Cochliobolus carbonum*, *Colletotrichum lindemuthianum*, *Cryphonectria parasitica*, *Fusarium moniliforme*, and *Sclerotinia sclerotiorum* (Shieh et al. 1997).

In various studies, PGIPs derived from different plant species have demonstrated efficacy against PGs secreted by plant pathogens, highlighting their potential in inhibiting wilt pathogens and boll rot disease. Molecular genetic evidence has linked specific polygalacturonase variants to the infection and spread of fungal pathogens in cotton bolls, emphasizing the significance of PGIPs in plant defense. Moreover, the identification of key amino acids in PGIPs and their manipulation through site-directed mutagenesis have shown promise in enhancing resistance to wilt diseases caused by *Verticillium dahliae* in cotton plants. The interaction between PGs and PGIPs serves as a model for understanding the host-pathogen recognition processes, offering insights into plant resistance and immunity against diseases. The interaction between host and pathogen can be investigated using molecular dynamics (MD) simulation studies (Fazil et al. 2012; Choudhary et al. 2022; Kumar et al. 2023). Hence, our study utilizes molecular modeling techniques along with MD simulations to help us to understand how PGIP from *G. barbadense* bind to PG of *A. macrospora* and *X. citri* pv. *malvacearum*, and the resulting effects on physiological, biochemical, and molecular mechanisms. Additionally, through computational methods, we aim to uncover the binding interactions between PGIP and PG, which induce structural alterations in the enzyme, thereby changing its activity.

Materials and methods

Structure prediction

The protein sequences of *G. barbadense* PGIP (gbPGIP) (AAQ19808.1), *A. macrospora* PG (amPG) (ACF19803.1), and *X. citri* pv. *malvacearum* PG (xcPG) (EKQ58646.1) were obtained from the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>) protein database. In the absence of experimental structures for PGIP and PG, we employed multiple structure prediction tools, including Modeller (Webb and Sali 2016), Swiss-model (Schwede et al. 2003), and AlphaFold (Jumper et al. 2021), to generate the models. The crystal structure of *Phaseolus vulgaris* PGIP (PDB: 1OGQ) was used as a template for PGIP modeling, while the crystal structures of *Pectobacterium carotovorum* PG (PDB: 1BHE) and *Aspergillus aculeatus* PG (PDB: 1IA5) served as templates for modeling the PG of *X. citri* pv. *malvacearum* and *A. macrospora*, respectively. These templates are functional homologs as well as the closest sequence of the PGIP and PGs.

Validation of models

The validity of the predicted 3-Dimensional (3D) models was assessed using the SAVES (Structure Analysis and Verification Server; <https://saves.mbi.ucla.edu/>) to detect any deviations from standard parameters, including bond lengths, dihedral structures, and atom-atom distances. Verification of the stereochemical quality of the predicted 3D structure was performed using Verify-3D (Bowie et al. 1991) and ERRAT (Colovos et al. 1993) tools available on the SAVES server. PROCHECK analysis (Laskowski et al. 1993) was employed to evaluate stereochemical characteristics, including bond lengths, bond angles, and Ramachandran plots, as well as main chain and side chain parameters of the 3D models. Additionally, WHATCHECK program (Hooft et al. 1996) was utilized to examine extensive stereochemical parameters of residues within the models.

Active site identification and protein-protein docking

Active site residues were determined through both sequence and structural alignment of the template and modeled proteins, supplemented by information from relevant literatures. Protein-protein interactions (PPIs) between PGIP of *G. barbadense* (gbPGIP) and PG of *A. macrospora* (amPG) and *X. citri* pv. *malvacearum* (xcPG) were analyzed using High Ambiguity Driven biomolecular DOCKing (HADDOCK) software (Dominguez et al. 2003). HADDOCK integrates biochemical and biophysical interaction data, including chemical shift perturbation data from nuclear magnetic resonance (NMR) titration analyses and mutagenesis data, to calculate the intermolecular energy. It generates a set of

complex structures, typically around 1 000, which are then classified based on their electrostatic, van der Waals, and restraint violation energy (RVE) factors. The structures were further evaluated and ranked based on criteria such as pairwise backbone root mean square deviation (RMSD) at the interface, average buried surface area, and average interaction energies. HawkDock server (Weng et al. 2019) was utilized to assess complex clusters based on free binding energy using molecular mechanics/generalized born surface area (MM/GBSA) analysis. Subsequently, mutations were introduced at active site residues, Glu169 and Phe242, replacing them with Gly and Lys, respectively, using the MutaBind2 server (Zhang et al. 2020). HawkDock server was again employed for MM/GBSA analysis of the mutated complexes. MutaBind2 provides 3D structure models for complexes with single or multiple mutations and calculates changes in binding affinity resulting from the mutations.

MD simulations

Performing MD simulations is a powerful approach to investigate the stability, conformational flexibility, and dynamic behavior of molecular systems, such as protein-protein and protein-ligand complex atoms (Singh et al. 2013; Varshney et al. 2023). We carried out MD simulations for the docked conformations of two complexes, i.e., gbPGIP complex with amPG and xcPG.

The aim of these simulations was to assess the stability of docked complexes both before and after mutation. Utilizing the GROMACS-2021 package (Bekker et al. 1993) and the CHARMM27 force field, the simulations were conducted. Initially, the docked structures were positioned within a water box containing ions (Bjelkmar et al. 2010). A dodecahedron box was created to accommodate the docked complexes, with the addition of sufficient chloride ions to neutralize the system. Steric clashes were resolved through energy minimization (EM), employing a steep descent method with a cutoff of 1 000 kJ·mol⁻¹ and 50 000 iterations or reduction steps. The EM approach effectively addressed steric clashes and served as a simulation framework. Subsequently, a two-phase equilibration process ensured the stability of the solvent and ions surrounding the complex. The first phase involved stabilizing the system's temperature through equilibration in the NVT (number of particles, system volume, and temperature) ensemble at 300 K, followed by stabilizing the system's pressure in the second phase through equilibration in the NPT (number of particles, system pressure, and temperature) ensemble, maintaining a constant number of particles, pressure, and temperature. Each equilibration phase required 100 picosecond (ps) to complete. Finally, a production run

was conducted for 100 nanosecond (ns) to collect atomic trajectories.

Trajectories analysis

Molecular dynamics (MD) simulations generate extensive datasets that capture the dynamic behaviour of atoms or molecules over time. Analyzing MD trajectories is crucial for understanding how biomolecules behave and for validating computational models. Key stability parameters such as RMSD, root mean square fluctuation (RMSF), radius of gyration (Rg), and solvent-accessible surface area (SASA) (Connolly 1983) are intrinsic to this analysis. Following the completion of the simulations, we employed the integrated tools within GROMACS to analyze the trajectories of protein-protein interactions. The stability of the protein backbone structure throughout the simulation was assessed by calculating RMSD and RMSF using the rms and rmsf modules, respectively. Additionally, the gyrate and sasa modules were utilized to determine Rg and SASA, providing insights into the protein's shape and size. The Xmgrace program was then used for further data analysis and visualization of the obtained results (Vaught 1996).

Essential dynamics and free energy landscape

The investigation into protein structure dynamics during MD simulations involved the application of principal component analysis (PCA). PCA served to simplify the understanding of complex systems by elucidating collective motions within proteins. The identification of the first principal component (PC1) through RMSF delineated the primary source of fluctuations within the system. Eigenvalues, derived from GROMACS tools, provided an indication of the cumulative fluctuations per atom associated with these collective motions, with a specific focus on C α atoms due to their robustness against statistical noise. Porcupine plots were generated to visualize the collective motion of atoms or residues within the biomolecular system, while free energy landscapes were constructed to analyse the energetic landscape of the system as it explored different conformational states.

To assess the conformational diversity present among protein structures generated during MD simulations, a clustering method was employed. Clustering analysis was performed using the Jarvis-Patrick method (Jain 2010), implemented with the cluster module of GROMACS. The Jarvis-Patrick method is a non-parametric clustering procedure that identifies clusters based on the similarity of structures without requiring a predetermined number of clusters. For this analysis, average structures were determined as the middle snapshot within each cluster, identified through the Jarvis-Patrick clustering procedure. This approach was based on the mutual RMSD of

all snapshots collected from the final 100 ns of the simulations, allowing for a detailed examination of the conformational space sampled by each complex.

Results and discussion

Structure prediction and validation

Using three computational tools, namely Modeller, Swiss-model, and AlphaFold, we analyzed the structural characteristics of PGIPs and PGs. Modeller and Swiss-model employ a comparative modeling approach, leveraging existing experimental structures of similar proteins as templates to predict the structure of the target protein. These tools optimize the alignment between the target sequence and the template structures, facilitating the generation of a 3D-model. On the other hand, AlphaFold utilizes advanced neural network architectures and training techniques, integrating evolutionary, physical, and geometric properties of protein structures to predict their 3D structures. In our analysis, we found that the Swiss-model server provided more accurate models of PGIP and PG compared with Modeller and AlphaFold. The predicted structures of cotton GbPGIP, consisting of 330 amino acids, and PG from *X. citri* pv. *malvacearum* (463 amino acids) and *A. macrospora* (379 amino acids), are depicted in Fig. 1. Further, upon structural assessment, we confirmed the validity of the predicted structures, indicating their suitability for further investigation and functional studies. This comprehensive analysis highlights the effectiveness of computational modeling tools in predicting protein structures, with the Swiss-model demonstrating superior accuracy in this context.

The stereochemical properties of the predicted gbPGIP and PGs were thoroughly investigated. Analysis of the Ramachandran plot revealed that approximately 77.4% of the residues' phi/psi angles in the modelled structure were situated within the most favoured regions, while 22.2% of residues resided in additional allowed regions. For xcPG and amPG, a notable percentage of PG residues, 88.3% and 80.8% respectively, occupied the most favoured areas of the Ramachandran plot (Fig. 1d-f). The PROCHECK G-factor values for all predicted structures fell within the ideal range of -0.4 to 0.5 , indicating their high quality.

Various evaluation methods were employed to assess the overall quality of the models. The ERRAT gives the quality score and Z-score of the PROSA server indicating the consistency of the models. The negative values of Z-score signifying their reliability (Wiederstein et al. 2007). Further validation from the VERIFY 3D server confirmed the quality of the models, as a significant proportion of residues scored above the threshold of 0.1 . The WHATCHECK results supported the accuracy of the models, while VADAR (Volume, Area, Dihedral

Angle Reporter) analysis provided smaller chi-square standard deviations, suggesting a better fit and higher confidence (Willard et al. 2003). Additionally, ProQ utilized MaxSub and LGscore to evaluate the protein model's accuracy, whereas LGscore above 4.0 indicated a very good model (Wallner et al. 2003). Collectively, these validations underscore the high level of structural quality exhibited by the modelled structures. Further details on the quality assessments are provided in Table 1.

Sequence alignment and active site residues

The alignment of PGIP sequences with their respective templates and PG from pathogenic organisms was conducted using ESPript (Robert et al. 2014), and the results are depicted in Fig. 2. The secondary structures such as α -helices, η -helices, β -sheets, and strict β -turns were annotated as α , η , β , and TT, respectively. Conserved residues were highlighted in red boxes, while similar residues were represented by white boxes. Additionally, the alignment accounted for insertions and deletions of residues, particularly in the loop regions.

The PGIP sequence from *G. barbadense* exhibited significant alignment and conservation with the PGIP template from *P. vulgaris* (PDB: 1OGQ), as illustrated in Fig. 2a. However, notable mutations were observed at the active residues Val152 (V152) and Gln224 (Q224) in the crystal structure of *P. vulgaris* PGIP, where *G. barbadense* displayed Glu169 (E169) and Phe242 (F242), respectively. Due to distinct evolutionary histories between the two species, these changes most likely resulted from the evolutionary processes.

Furthermore, the PG sequences from selected plant pathogens, such as *A. macrospora* and *X. citri* pv. *malvacearum*, were aligned with sequences from other crystal structures. Specifically, PG sequences from *Fusarium verticillioides* (PDB ID: 1HG8) and *A. aculeatus* (PDB ID: 1IA5) served as templates for fungal pathogens, while *P. carotovorum* PG (PDB ID: 1BHE) served as a template for bacterial pathogens. Though complete sequence conservation was not observed, active residues were conserved across all five plant pathogens (Fig. 2b). These findings underscore the dynamic nature of protein interactions in plant-pathogen systems and highlight the evolutionary adaptations that occur in response to selective pressures.

Site-directed mutations

Maulik et al. (2013) conducted research on the PGIP of *P. vulgaris*, where they found that site-directed double mutagenesis at residues Val152 and Gln224, replacing them with Gly and Lys, respectively, led to significant structural alterations in the protein's concave structure. This alteration ultimately hindered the binding of PG from the pathogen *F. phyllophilum*. The mutations caused

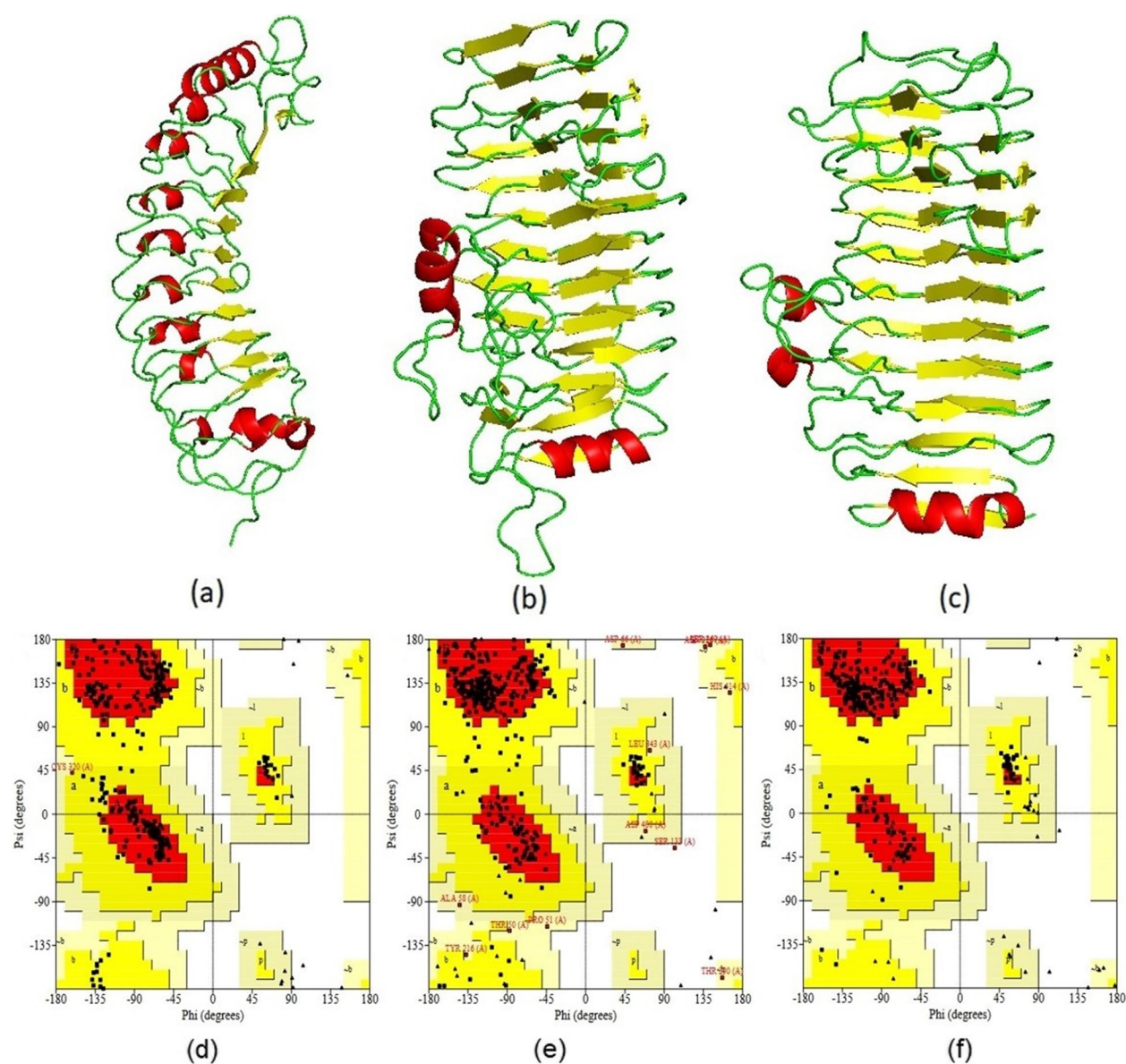


Fig. 1 Predicted structure and the Ramachandran plot of respective structures, PGIP of (a and d) *G. barbadense* and PG of (b and e) *X. citri* pv. *malvacearum*, (c and f) *A. macrospora* using the Swiss-model

Table 1 Quality measures of the predicted structure

Protein	Overall quality factor (ERRAT)	Z-score (ProSA-web)	LGScore (ProQ)	MaxSub (ProQ)	Remark
gbPGIP	86.913	−7.47	10.134	−0.662	Good
amPG	89.058	−7.17	9.250	−0.409	Good
xcPG	86.555	−5.76	8.997	−0.416	Good

subtle changes in the concave face of the PGIP, affecting its functionality. Interestingly, in the PGIP of *G. barbadense*, evolutionary processes have led to mutations at similar positions: Val152 mutated to Glu169 and Gln224 mutated to Phe242 (as depicted in Fig. 3). Inspired by these findings, we conducted a similar study on the PGIP of *G. barbadense*. Utilizing MutaBind2, we performed double mutations to observe the effects of replacing Glu169 and Phe242 with Gly and Lys, respectively.

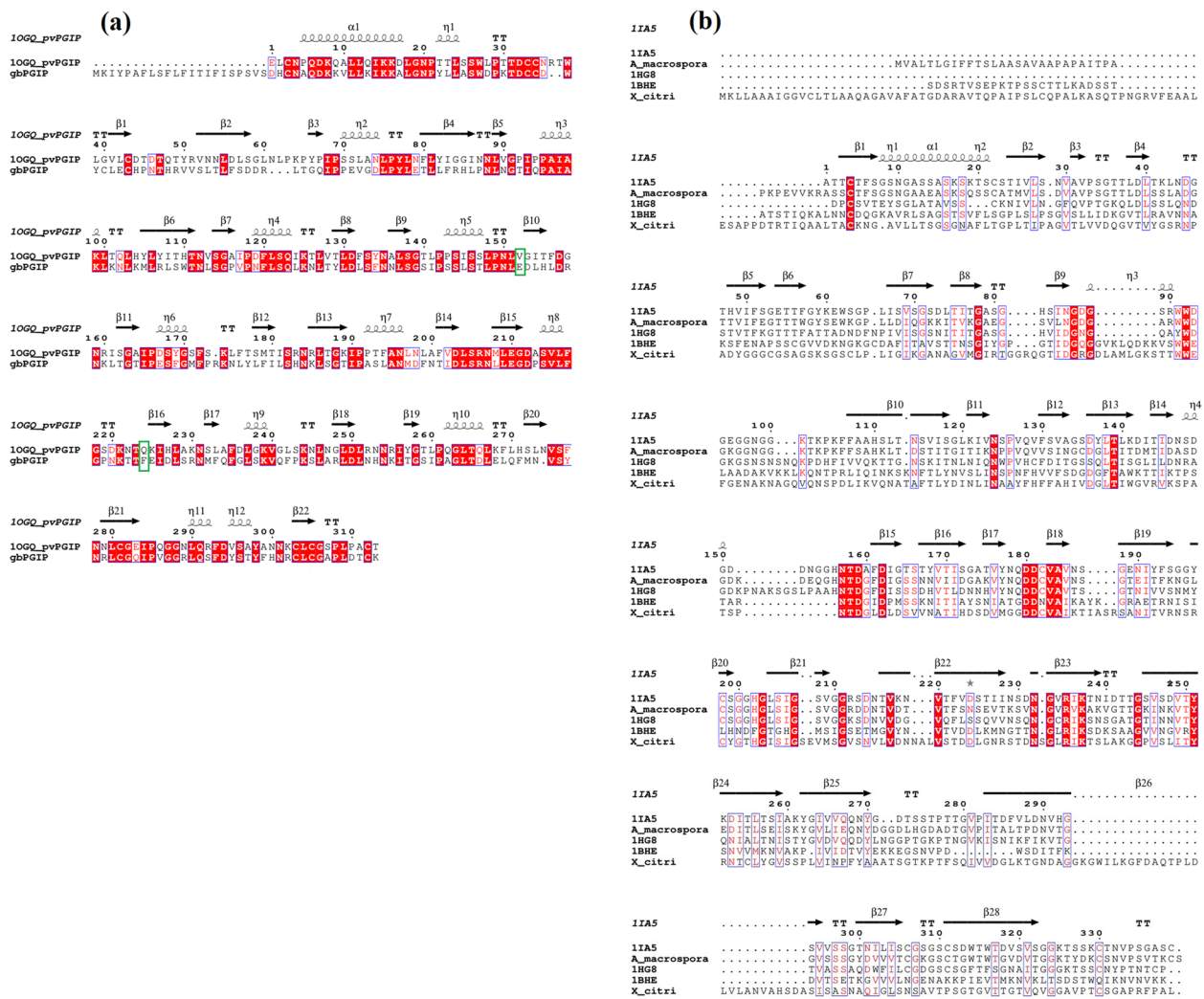


Fig. 2 Sequence alignment of PGIP and PG sequences generated by MultiAlign and image was produced by ESPrInt. **a** PGIP of *P. vulgaris* (PDB: 1OGQ) and *G. barbadense*, **(b)** Crystal structures selected as templates PDB id 1HG8, 1IA5, 1BHE and selected PG of parasites *X. citri* pv. *malvacearum* and *A. macrospora*

Docking

The crucial residues comprising the active sites of PGIPs and PGs were highlighted in red within the sequence alignment and Supplementary file S1. To ascertain the optimal conformation of the gbPGIP receptor proteins with the amPG and xcPG effectors, protein-protein docking simulations were executed using HADDOCK. The structural models of ghPGIP, amPG, and xcPG served as inputs for the docking process. Clusters of generated models were assessed, with preferences given to conformations exhibiting the most negative Z-score, as outlined in Table 2. Each selected cluster consisted of four protein complex models, and the model with the lowest binding energy and most favorable interactions with active site residues was chosen for further analysis using the Hawkdock server. The calculated binding energies (in

kcal·mol⁻¹) for the selected models were -92.1 (gbPGIP-ampG) and -48.1 (gbPGIP-xcPG), indicating their stability. Following this, doubly mutated complex structures of PGIP-PG were generated using the MutaBind2 server and subsequently submitted to the HADDOCK server. In-depth intermolecular analysis uncovered the presence of hydrogen bonds within the complexes. By scrutinizing the active sites and evaluating binding energies, the study elucidates critical insights into the stability and specificity of these interactions.

The gbPGIP-amPG and gbPGIP-xcPG complexes displayed hydrogen bond counts, with 7 and 3, respectively, along with several hydrophobic interactions, as illustrated in Fig. 4a and c. Specifically, the amino acid Glu169 of gbPGIP formed a hydrogen bond with amPG

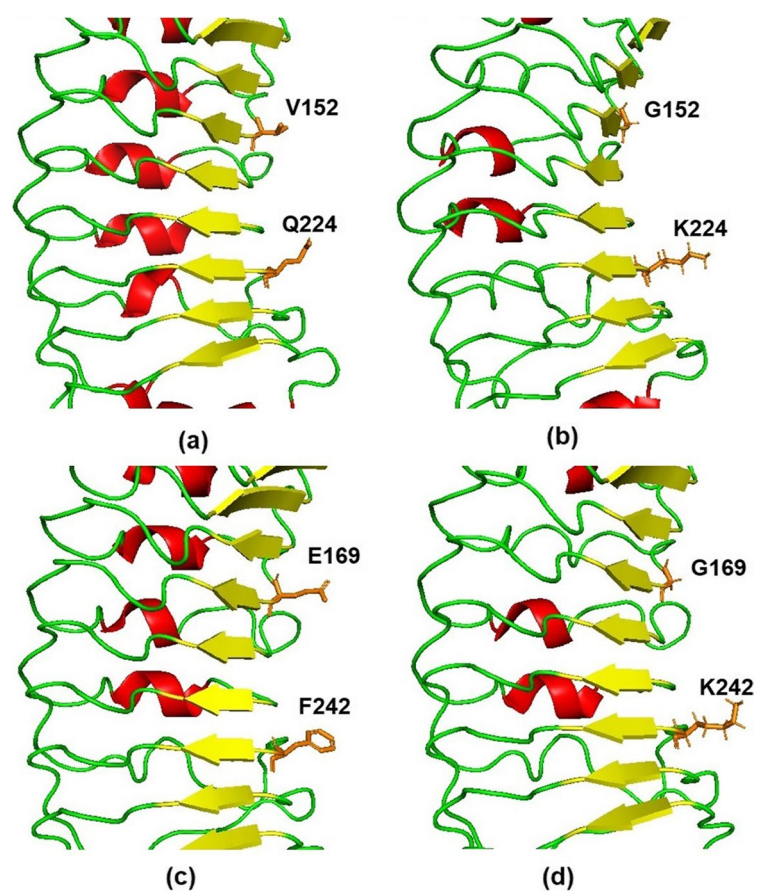


Fig. 3 Double mutated active residues of template (1OGQ) and gbPGIP; **(a)** 1OGQ- before mutation, **(b)** 1OGQ- after mutation, **(c)** gbPGIP- before mutation, and **(d)** gbPGIP- after mutation. V: Valine (Val), Q: Glutamine (Gln), E: Glutamate (Glu), F: Phenylalanine (Phe), G: Glycine(Gly) and K: Lysine (Lys)

Table 2 HADDOCK parameters for PGIP-PG complexes

Protein complexes	Structures	Clusters	Water refinement	Z-score
gbPGIP-amPG	121	12	60%	−1.4
gbPGIP-xcPG	131	11	65%	−1.1

and engaged in a hydrophobic interaction with xcPG. Additionally, the Phe242 residue of gbPGIP participated in hydrophobic interactions with both amPG and xcPG. These findings suggest a nuanced interaction pattern between gbPGIP and its interacting partners, indicating potential differences in binding affinity and specificity between amPG and xcPG.

Figure 4b clearly illustrates that residues Gly169 and Lys242 of mutant gbPGIP do not participate in the interaction with amPG. However, when xcPG interacts with mutant gbPGIP, it disrupts the hydrophobic interaction with Gly169, instead it forms a hydrogen bond

with the Lys242 residues of mutant gbPGIP (Fig. 4d). The green and red dotted lines indicate the hydrogen bonds and hydrophobic interactions, respectively, between amino acids of gbPGIP which interact with amPG or xcPG residues. While the mutation does affect the flexibility and interactions of active residues in the docked complexes, its impact appears to be minimal at this stage. To comprehensively assess the effects of the mutations in gbPGIP on the interaction with amPG and xcPG, molecular dynamics simulations were conducted in the subsequent step. These simulations provide deeper insights into the dynamic behaviour and stability of the protein-protein complexes, shedding light on the conformational changes and energetic profiles underlying the interaction dynamics.

Analysis of trajectories

The docked complexes underwent molecular dynamic simulations lasting 100 ns to observe their behavior and evaluate their structural stability. Various

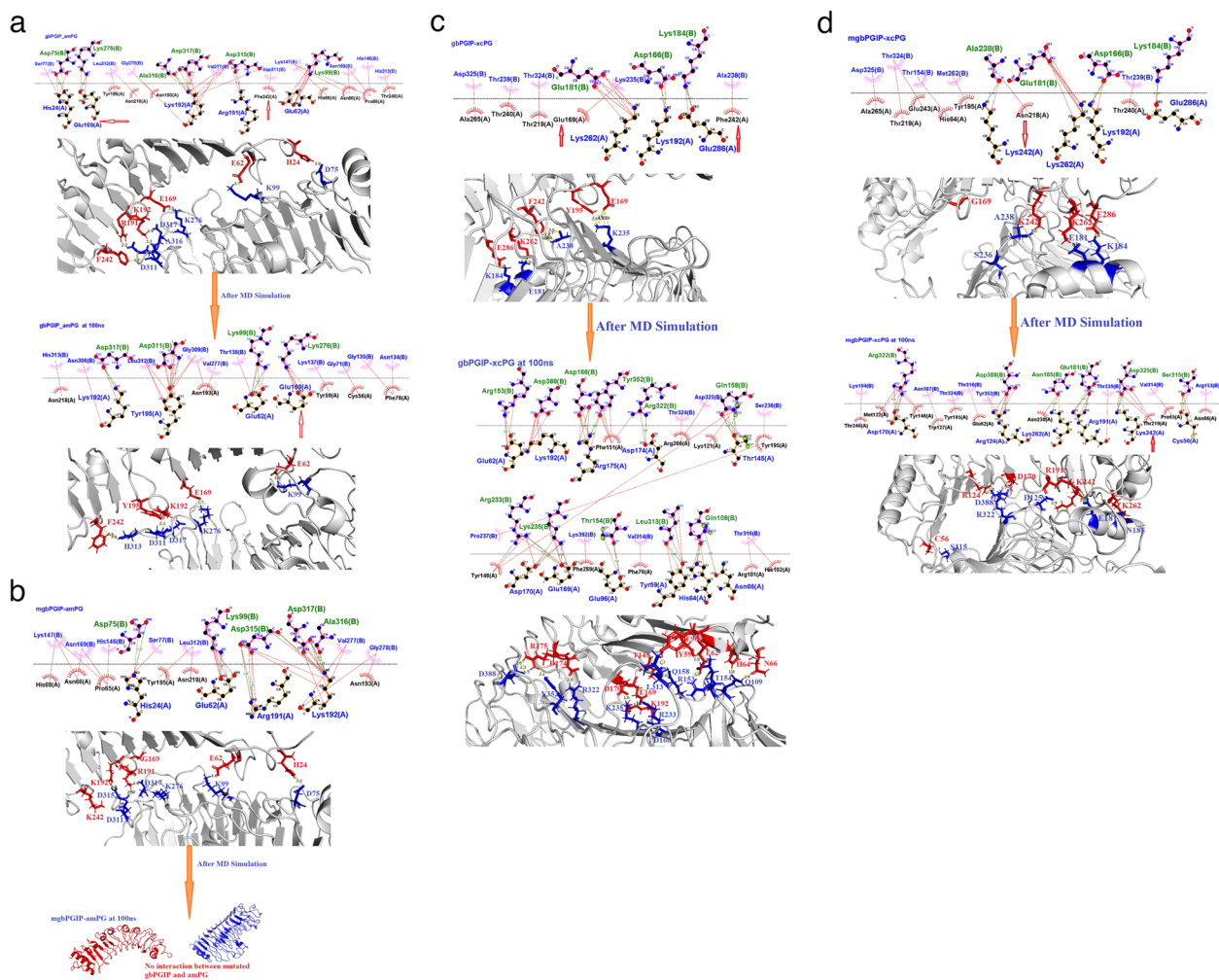


Fig. 4 Inter-molecular analysis of four complexes **a** Inter-molecular analysis of gbPGIP-amPG complex, pre and post MD simulations. **b** Inter-molecular analysis of mutated gbPGIP with amPG complex, pre and post MD simulations. **c** Inter-molecular analysis of gbPGIP-xcPG complex, pre and post MD simulations. **d** Inter-molecular analysis of mutated gbPGIP with xcPG complex, pre and post MD simulations. A is used for PGIP and B for PG. 2-D and 3-D images are generated by LigPlot and PyMol tools, respectively. Green dotted lines represent hydrogen bonds and Red dotted lines illustrate hydrophobic interactions between PGIP and PG

metrics, including RMSD, RMSE, Rg, and SASA, were used to assess the stability of the complexes during the simulations.

RMSD served as a quantitative measure to gauge the stability of the complexes throughout the 100 ns MD simulation, measuring the variation in the positions of the protein's backbone from the beginning to the end of the simulation. Lower RMSD values indicated greater stability in the docked complexes. In the case of the gbPGIP-amPG complex, consistent stability was observed, maintaining an average RMSD of ~ 0.39 nm over the entire 100 ns simulation, as depicted in Fig. 5. Conversely, the mgbPGIP-amPG complex showed a maximum deviation of ~ 4.6 nm. Particularly, this

complex displayed an average deviation of ~ 1.29 nm, with significant fluctuations of ~ 2.4 nm during the time interval of 62–67 ns and ~ 1.89 nm during 74–80 ns. After 80 ns, it consistently exhibited high fluctuation, peaking at ~ 3.9 nm.

The gbPGIP-xcPG complex also demonstrated stable dynamic behaviour with a RMSD of ~ 0.48 nm throughout the simulation. The mgbPGIP-xcPG complex showed an initial RMSD of ~ 0.5 nm up to 50 ns. Subsequently, it slightly decreased to ~ 0.47 nm for the remainder of the simulation (Fig. 5).

The results suggest that the three trajectories gbPGIP-amPG, gbPGIP-xcPG, and mgbPGIP-xcPG demonstrate consistent structural stability throughout the simulation,

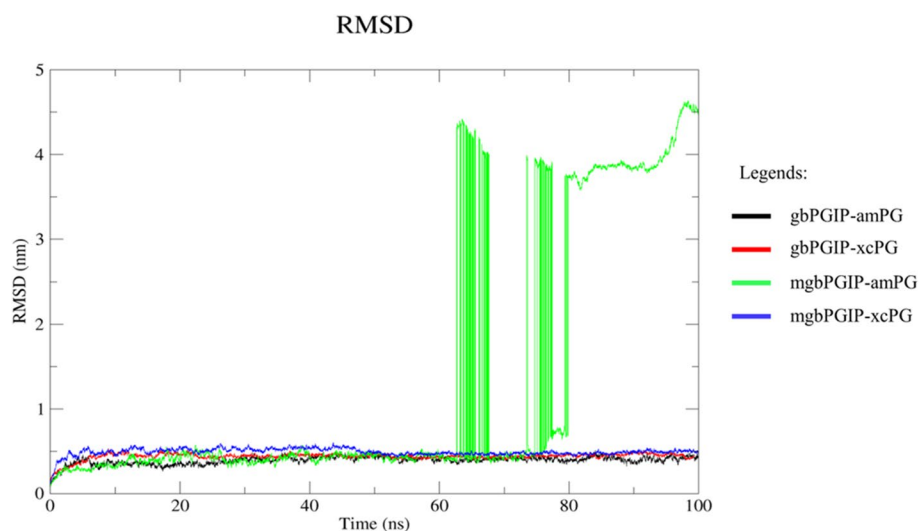


Fig. 5 RMSD of the backbone atoms as a function of time during MD simulations

maintaining RMSD values around 0.5 nm. These structures remain close to their initial conformations, indicating minimal structural changes. The interaction between wild gbPGIP and xcPG which is stabilized by a hydrophobic bond with Lys242 and possibly other interactions with Gly169 as observed in Fig. 4c. This shows that these interactions do not induce substantial structural changes in the PGIP, allowing it to remain stable and presumably functional in inhibiting the polygalacturonase activity of the pathogen. Conversely, the mgbPGIP-amPG trajectory exhibits initial stability similar to the other trajectories but shows significant fluctuations starting around 60 ns which could be due to lack of interaction between Gly169 and Lys242 residues of gbPGIP and amPG as evident in Fig. 4b. This instability culminates in a dramatic increase in RMSD values, reaching up to 5 nm after 80 ns. This behaviour indicates a substantial structural transition or instability in the latter part of the simulation for mgbPGIP-amPG. This was not observed in other systems, suggesting it undergoes critical structural changes. Such instability suggests that the mutation disrupts the ability of PGIP to maintain a stable complex with amPG, likely impairing its inhibitory effectiveness. Consequently, this could make the host more susceptible to infection by *A. macrospora*, highlighting the critical role of specific amino acids in maintaining the structural integrity and function of PGIP. In *X. citri* pv. *malvacearum*, both the wild-type and mutated PGIPs (gbPGIP-xcPG and mgbPGIP-xcPG) show stable interactions with xcPG, suggesting that the structural changes induced by the mutation do not significantly affect the ability of PGIP to interact with xcPG. This indicates that key binding residues for xcPG are not substantially impacted by the mutation,

allowing both the wild-type and mutated PGIPs to effectively inhibit the polygalacturonase activity of the plant pathogen. Thus, the host resistance to *X. citri* pv. *malvacearum* was not compromised by the mutation in PGIP.

Furthermore, a superimposition analysis of the wild-type and mutant PGIP structures was conducted to assess the extent of structural deviation induced by mutations. The RMSD values were calculated for both pre- and post-MD simulation structures. The RMSD of the pre-MD superimposed structures was determined to be ~0.11 nm, indicative of minimal structural variance between the initial conformation of the wild-type and mutant PGIP. However, following MD simulations, the RMSD increased substantially to ~2.82 nm, suggesting a remarkable divergence in the structural alignment between the wild-type and mutant PGIP structures (Fig. S2).

The RMSF serves as a tool to discern regions of proteins that undergo variations by gauging the flexibility of specific residues over time. A lower RMSF score indicates stability within protein complexes, while a higher score implies increased flexibility and instability. To assess the flexibility of individual residues relative to their average position during simulations, we computed the mobility of C α atoms across all systems. Consistent fluctuations were observed across the systems, as illustrated in Fig. 6.

In the gbPGIP-amPG complex, RMSF values for residues ranged from ~0.05 to ~0.29 nm, averaging at ~0.12 nm. The mgbPGIP-amPG complex displayed a wider range of RMSF values, spanning from ~0.05 to ~0.30 nm, with an average of ~0.15 nm. Similarly, in the gbPGIP-xcPG complex, RMSF values ranged from ~0.05 to ~0.33 nm, averaging at ~0.13 nm. Conversely, the

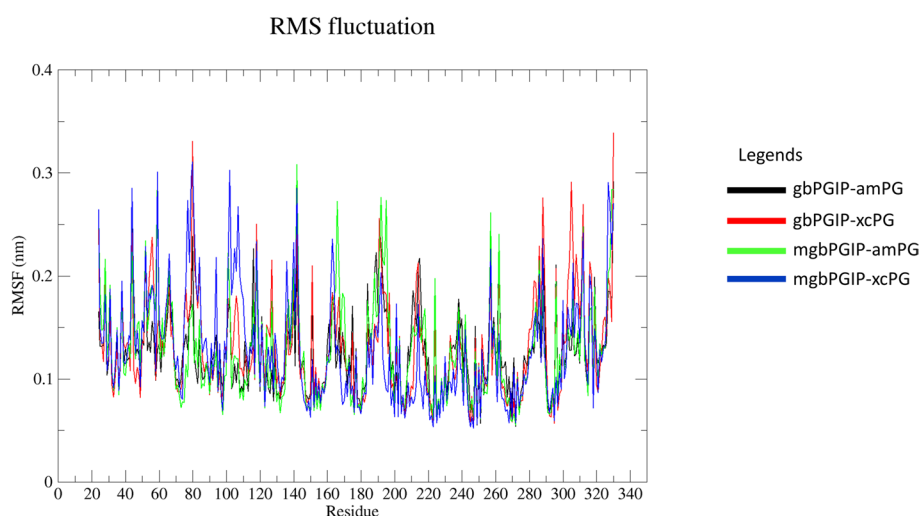


Fig. 6 RMSF of Ca during 100 ns MD simulation

mgbPGIP-xcPG complex exhibited RMSF values ranging from ~ 0.05 to ~ 0.31 nm, with an average of ~ 0.12 nm.

These findings indicate distinct patterns of residue flexibility within the complexes. Particularly, the mgbPGIP-amPG complex showed higher overall RMSF values, suggesting increased flexibility in specific residues compared with the non-mutated gbPGIP-amPG complex. Conversely, the mgbPGIP-xcPG complex displayed similar RMSF values to its non-mutated counterpart, indicating comparable flexibility in residues. These observed variations in RMSF values may denote altered dynamics and structural adaptability induced by mutations in the PGIP.

To further evaluate the structural stability of the docked complexes, additional analyses were conducted using measurements of the Rg and SASA. SASA quantifies the extent of the protein surface exposed to the

solvent, while Rg provides insights into the compactness of the system and the dimensions of protein-protein complexes, thus indicating its crucial role for their proper interactions.

The Rg values for gbPGIP-amPG, gbPGIP-xcPG, mgbPGIP-amPG, and mgbPGIP-xcPG were approximately 2.86, 2.72, 3.48, and 2.76 nm, respectively (as depicted in Fig. 7a). Notably, mgbPGIP-amPG exhibited major fluctuations during the 62–67 ns and 74–80 ns intervals, followed by consistent higher fluctuations after 80 ns, averaging at 5 nm. These significant fluctuations in Rg for the mgbPGIP-amPG complex indicate substantial structural instability. This instability suggests that the mutation disrupts the compactness of the PGIP when interacting with amPG.

For the gbPGIP-amPG complex, the SASA values remained stable with an average of 299 nm^2 , suggesting

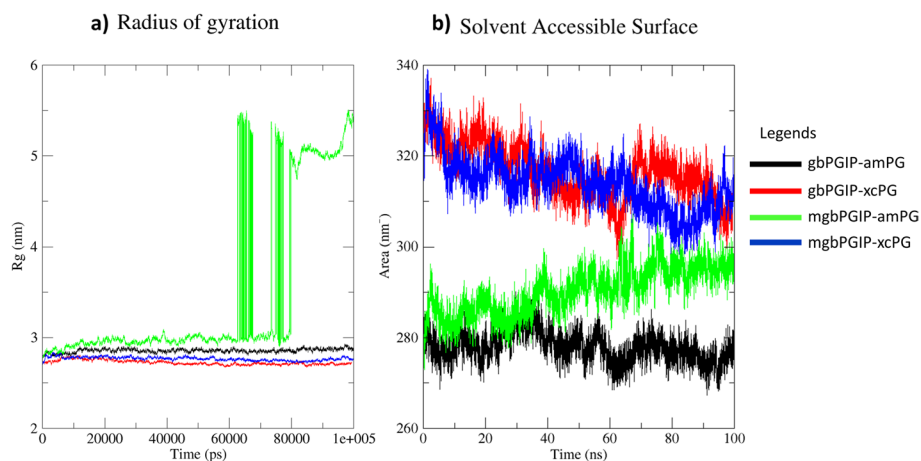


Fig. 7 a Rg profile and b SASA analysis over a time range of 100 ns MD simulation

a consistent and stable exposure to the solvent. For the gbPGIP-xcPG complex, a decreasing trend in total SASA was observed, averaging at 316 nm², which might indicate slight compaction or reduced exposure to the solvent over time, reinforcing the stability of this complex. In contrast, the SASA for the mutated complex mgbPGIP-amPG increased from 267 nm² to 295 nm², highlighting structural changes that lead to greater solvent exposure. This increased SASA further supports the notion of instability and conformational changes observed in the mgbPGIP-amPG complex. For the mgbPGIP-xcPG complex, the SASA changes from 308 nm² to 315 nm² (illustrated in Fig. 7b) suggests minor adjustments in solvent exposure but overall maintains stability, consistent with the stable Rg result for this complex.

Both the wild-type and mutated PGIPs (gbPGIP and mgbPGIP) demonstrate stable interactions with xcPG, maintaining consistent Rg and SASA values. In contrast, mgbPGIP-amPG complex exhibits significant structural instability, as evidenced by fluctuating Rg and increasing SASA values. These structural changes likely impair the ability of mgbPGIP to effectively inhibit amPG, making the host more susceptible to infection by *A. macrospora*.

Analysis of intermolecular hydrogen bonds

Intermolecular hydrogen bonds play a key role in maintaining molecular stability and facilitating recognition processes. To assess the dynamic stability of four complexes (gbPGIP-amPG, gbPGIP-xcPG, mgbPGIP-amPG, and mgbPGIP-xcPG), H-bond analysis was performed during 100 ns MD simulations as shown in Fig. 4a-d and Fig. 8.

Within the gbPGIP-amPG complex, a consistent and average of 4.8 hydrogen bonds was observed throughout the simulation. In the gbPGIP-xcPG complex, an increasing trend in the number of hydrogen bonds between the interacting proteins was noted, with an average of 8.5 over the course of the simulation.

Nevertheless, the mgbPGIP-amPG complex exhibited a decreasing trend in the number of hydrogen bonds among the interacting proteins, with an average of 2.6. After 70 ns, the number of hydrogen bonds continuously decreases and at the end of the simulation process the complex is separated, as represented by the green colour in Fig. 8. The mgbPGIP-xcPG complex displayed a pattern similar to gbPGIP-xcPG, with a slight decrease in the number of hydrogen bonds. The average number of hydrogen bonds for mgbPGIP-xcPG was 8.2. The difference of hydrogen bond interactions in gbPGIP-xcPG (red) and mutated gbPGIP-xcPG complex (blue) reflects higher variation between 80–100 ns. At the end of the simulation, gbPGIP-PG shows more hydrogen interaction and site directed mutations in gbPGIP inhibit the formation of hydrogen bonds and their stability with xcPG.

These results suggest that while the non-mutated complexes (gbPGIP-amPG and gbPGIP-xcPG) maintain stable hydrogen bond interactions, the mutated complexes (mgbPGIP-amPG and mgbPGIP-xcPG) exhibit alterations in the dynamics of intermolecular hydrogen bonding. The decreasing trend in hydrogen bond numbers in mgbPGIP-amPG and the slight decrease in mgbPGIP-xcPG may indicate potential disruptions in the stability of these complexes.

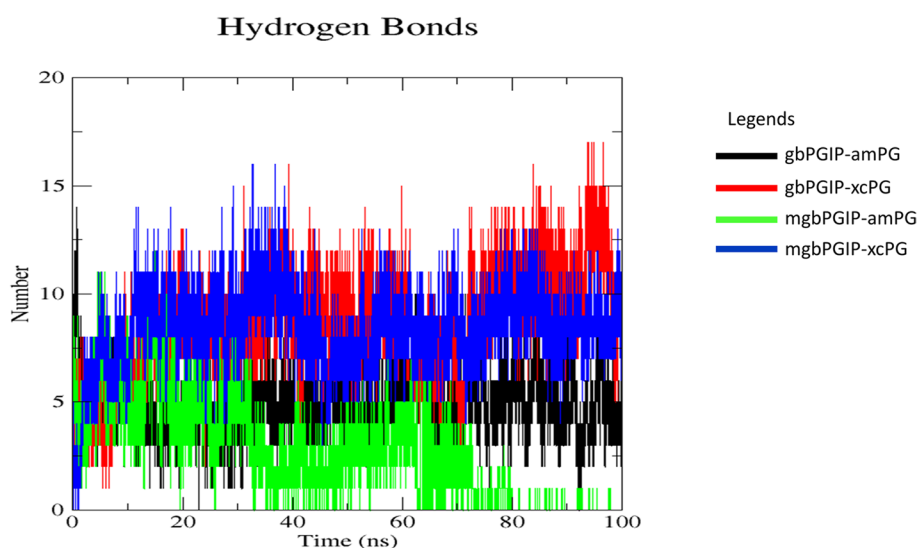


Fig. 8 H-bond analysis during 100 ns MD simulations

Clustering analysis

The cluster analysis of the four complexes (gbPGIP-amPG, gbPGIP-xcPG, mgbPGIP-amPG, and mgbPGIP-xcPG) reveals significant insights into their structural stability and conformational dynamics. The superimposition of the top two representative structures, obtained using the Jarvis-Patrick clustering method, shows that the gbPGIP-amPG (Fig. 9a) and mgbPGIP-amPG (Fig. 9c) complexes exhibit significant deviations between the structures, indicating significant conformational flexibility and instability. This is contrasted by the gbPGIP-xcPG (Fig. 9b) and mgbPGIP-xcPG (Fig. 9d) complexes, where the structures are closely aligned, reflecting structural stability and minimal conformational changes (Fig. 9).

For the gbPGIP-amPG complex, the range of RMSD values was 0.05–4.5 nm, with an average RMSD of 2.1 nm, resulting in 25 clusters. The mgbPGIP-amPG complex showed a similar pattern with a range of RMSD values from 0.05 to 4.6 nm, an average RMSD of 2.2 nm, and 40 clusters. These high RMSD ranges and large number of clusters indicate significant structural deviations and instability for the amPG complexes. Conversely, the gbPGIP-xcPG complex had a much narrower RMSD range of 0.05–0.3 nm, an average RMSD of 0.2 nm, and formed only 4 clusters. Similarly, the mgbPGIP-xcPG complex exhibited an RMSD range of 0.05–0.45 nm, an average RMSD of 0.2 nm, and 7 clusters. The small RMSD ranges and fewer clusters in the xcPG complexes suggest a higher degree of structural stability and minimal conformational changes.

The RMS distribution further supports these observations (Fig. 10). For the gbPGIP-amPG (black) complex, the distribution shows a broad peak, indicating a range of conformations and higher RMS values, which suggests its structural instability. Similarly, the mgbPGIP-amPG (green) complex exhibits multiple peaks, with one significant peak at higher RMS values, reinforcing the idea of considerable conformational flexibility and instability. In contrast, the RMS distributions for the gbPGIP-xcPG (red) and mgbPGIP-xcPG (blue) complexes were narrow and focused at lower RMS values, implying that these complexes maintain similar conformations and exhibit structural stability.

The combined analyses from the superimposed structures and RMS distributions indicate that the gbPGIP-amPG and mgbPGIP-amPG complexes undergo significant conformational changes and are less stable. In contrast, the gbPGIP-xcPG and mgbPGIP-xcPG complexes exhibit minimal conformational changes and greater structural stability. These findings suggest that interactions involving xcPG are more stable, whereas those involving amPG lead to considerable conformational flexibility and instability in the complexes.

Principal component (PC) and Free Energy Landscape (FEL) analysis

In the study of protein dynamics, principal component analysis (PCA) was utilized to delve into the conformational variations of PGIP complexes (both the wild-type and mutant) with PG. This analysis allowed for the

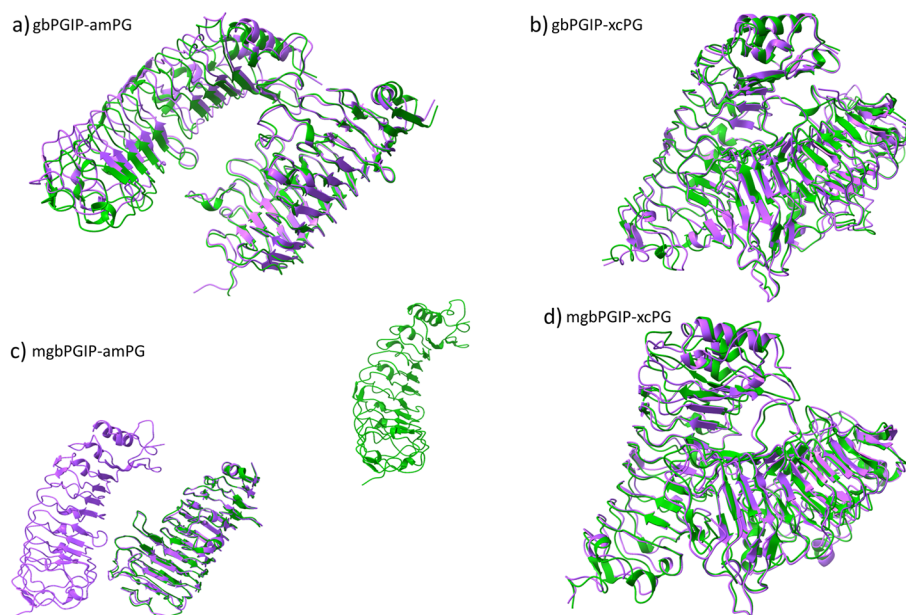


Fig. 9 Superimposed of the representative structures of the top 2 clusters

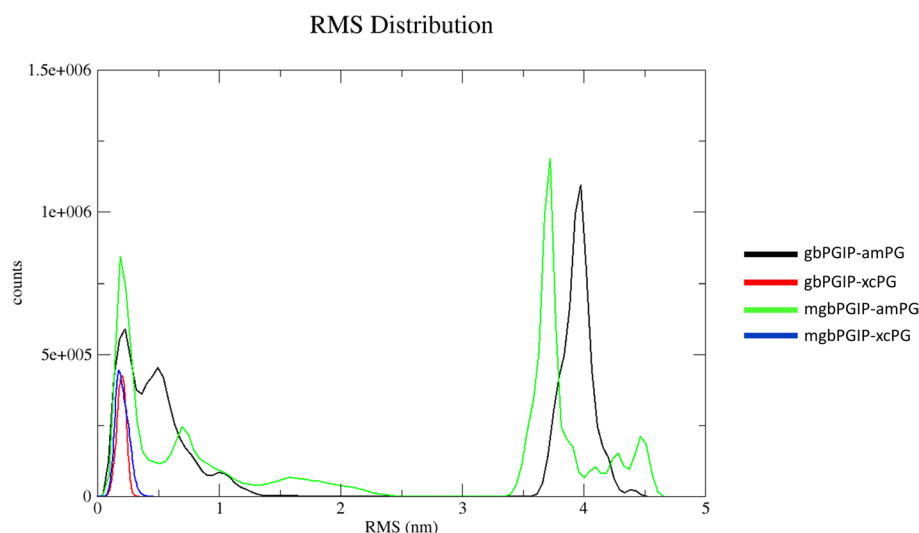


Fig. 10 RMS distribution for conformations of the complexes

examination of their collective motions using the essential dynamics approach. By employing PCA, we were able to visualize how proteins move in their atomic space to execute specific functions. Eigenvalues were extracted from the covariance matrix, while PCs were identified using tools like *gmx anaeig* and *gmx covar*. These eigenvalues provided insights into the atomic contributions to motion, while PCs elucidated the overall direction of motion of the atoms. Our analysis of MD trajectories unveiled distinct behaviors between the wild-type and mutant gbPGIP complexes, particularly when interacting with amPGIP and xcPGIP. Notably, the wild-type gbPGIP demonstrated stability with increased dynamics in structural conformation, whereas the mutant gbPGIP

exhibited instability, especially in the complex with amPG (Fig. 11).

Additionally, porcupine plots were generated to visualize the global motions in both systems, leveraging the top two principal components PC1 and PC2. These plots provided further insights into the collective motions of the protein complexes, enhancing our understanding of their dynamic behaviour.

Porcupine plots were employed to visualize the movement patterns captured by the top PCs obtained from PCA (Fig. 12). These plots are valuable tools for understanding how protein structures move over time during molecular dynamics simulations. Porcupine plots depict the extreme projections of protein structures onto the

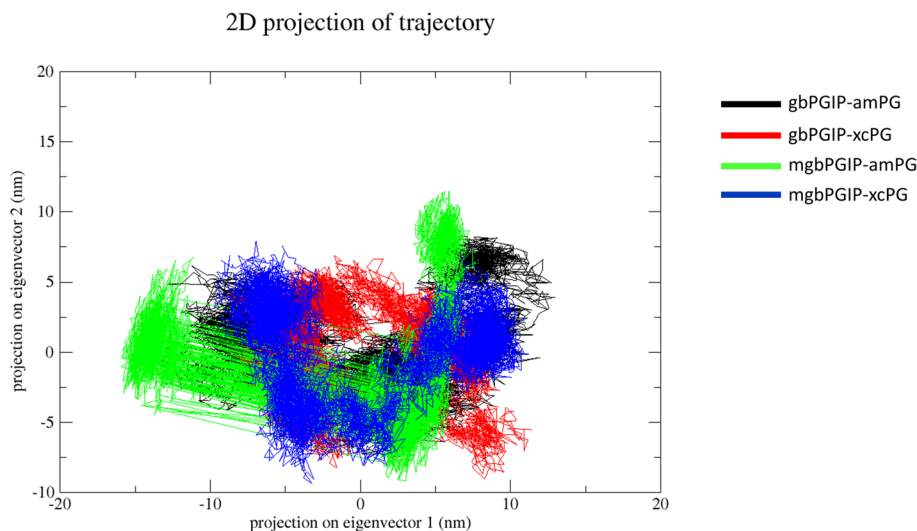


Fig. 11 Principal components of the complexes during 100 ns

PC1, which accounts for the maximum variance in the dataset. The length of the arrows in the plots represents the strength of motion, while the direction indicates the direction of movement. The analysis revealed distinct motion patterns among the protein complexes. For instance, gbPGIP-amPG showed a pronounced inward motion, suggesting a conformational change towards a more compact structure. Conversely, gbPGIP-xcPG and mgbPGIP-xcPG displayed subtle outward motion along PC1, indicating a slight expansion of the structures. These observations were consistent with the RMSF analysis, which identified the terminal ends of the complexes as the regions of greatest flexibility. Interestingly, mgbPGIP-amPG exhibited minimal motion due to lack of interaction.

In the examination of the Gibbs FELs, the first two eigenvectors (EVs) were utilized to investigate deeper into the conformational dynamics of the complex. The FELs of the gbPGIP-amPG, gbPGIP-xcPG, mgbPGIP-amPG, and mgbPGIP-xcPG complexes over a 100 ns time frame were presented in Fig. 13. Within these plots, regions shaded in deeper blue indicated energetically favourable conformational states characterized by lower energy levels, whereas yellow regions denoted less favorable conformations. Particularly, the gbPGIP-xcPG and mgbPGIP-xcPG complexes exhibited stable, energetically favourable states indicating minimal conformational changes. In contrast, the gbPGIP-amPG and mgbPGIP-amPG complexes showed more pronounced

fluctuations and less stable energy states, suggesting significant conformational changes and instability. Further analysis revealed that the presence of amPG influenced both the size and position of the essential space sampled by gbPGIP and mgbPGIP, resulting in unstable configurations and higher energy states.

The FEL analysis complements the RMSD findings observed for the mgbPGIP-amPG trajectory, which reached up to 5 nm after 80 ns, highlighting that specific mutations can have differential impacts on PGIP interactions with different pathogens. The stable interaction with xcPG despite the mutations suggests that the inhibitory function of PGIP against *X. citri* pv. *malvacearum* remains intact. In contrast, the instability observed with amPG indicates that the global structural integrity required for effective inhibition was compromised, emphasizing the critical role of specific amino acids in maintaining both local and global structural stability and function of PGIP.

Discussion

The complex of protein-protein PGIP-PG is explained to be important for the plant-pathogen interaction mechanism at atomic level (Misas-Villamil et al. 2008). The crystal structures of PG of various parasitic species, *P. carotovorum* (1BHE), *A. aculeatus* (1IA5), *E. leycettana* (7E56), *F. verticillioides* (1HG8), *A. thaliana* (7B8B), *C. lupini* (2IQ7), *A. niger* (1CZF), *C. purpureum* (1K5C), *T. maritima* (3JUR), and *Y. enterocolitica* (2UVE), have been

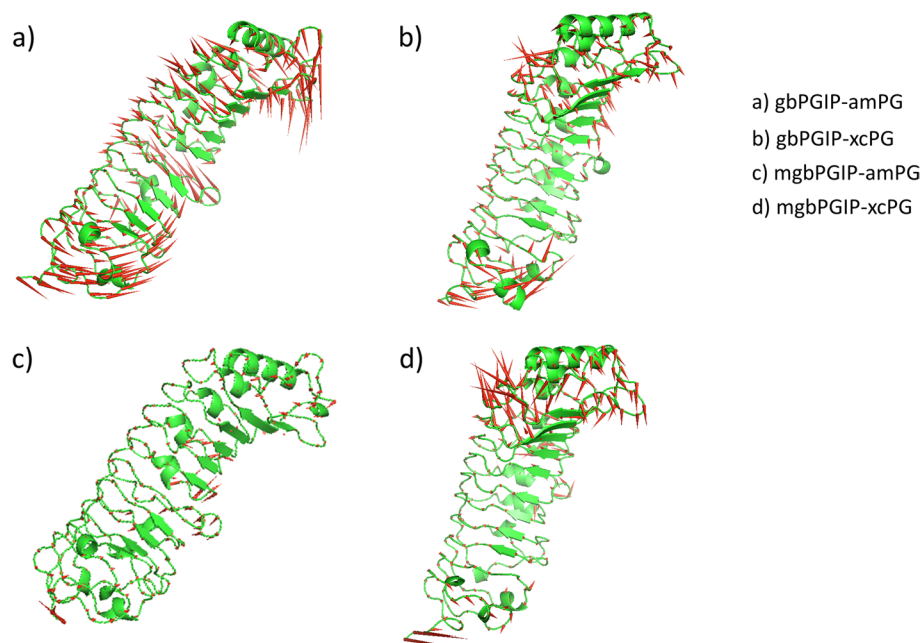


Fig. 12 Porcupine plot of gbPGIP-amPG, gbPGIP-xcPG, mgbPGIP-amPG, and mgbPGIP-xcPG complexes

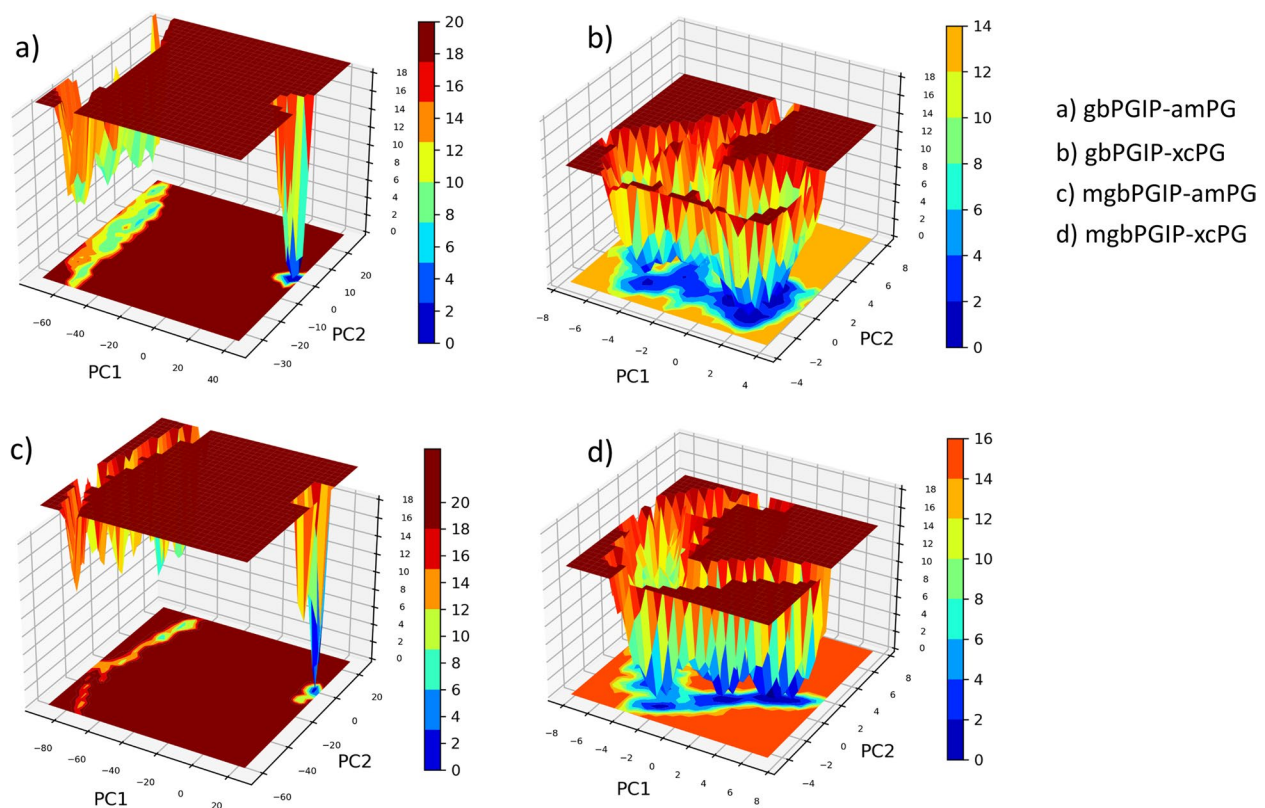


Fig. 13 Free energy landscape of gbPGIP-amPG, gbPGIP-xcPG, mgbPGIP-amPG, and mgbPGIP-xcPG complexes during 100 ns

resolved and submitted RCSB in Protein Data Bank. But, the crystal structure of PGIP is available only for PGIP of *P. vulgaris* in the Protein Data Bank. The study of plant-pathogen interaction (PPI) is instrumental in deciphering the intricate molecular interactions during host-pathogen interactions (Murmu et al. 2024; Kumar et al. 2020). The PPI of PGIP from *G. barbadense* with bacterial pathogen, *X. citri* pv. *malvacearum* and fungal pathogen, *A. macrospora*, and effect of double mutations in the PGIP have not been studied yet in *Gossypium* spp. Here, we employed a combinatorial approach of structural bioinformatics including structural modelling, active site prediction, protein-protein docking, double mutated complex formation, and molecular dynamic simulation to explore the mechanism of the PGIP-PG complex to explain the plant-pathogen interactions. The mode of interaction of fungal amPG with gbPGIP differs from the mode of interaction of bacterial xcPG with gbPGIP. The results of docking studies are evident of hydrogen bonding, gbPGIP-xcPG complexes have higher number of stronger H-bonds and hydrophobic interactions compared to gbPGIP-amPG.

Several studies have reported that 10 LRRs are associated to make a solenoidal shape of the PGIPs and these LRRs are directly involved in various cellular processes by protein-protein interactions. But only nine LRRs are

made of the solenoid structure of PGIP in *G. barbadense* (Fig. S3). However, mutations at selective sites result in changes to the structure, creating a concave form that inhibits interaction with the pathogenic PG. The studies of MD simulation of complexes, before and after mutation, have been performed for the gbPGIP-amPG, mutated gbPGIP-amPG, gbPGIP-xcPG, and mutated gbPGIP-xcPG. The MD simulation results collectively suggested that while the non-mutated complexes exhibit stable structures and interactions, the mutated complexes, particularly mgbPGIP-amPG, undergo alterations in dynamics, leading to potential disruptions in stability. These findings highlight the importance of considering the impact of mutations on the structural and dynamic behaviour of protein complexes, providing valuable insights for further investigations into the functional consequences of these changes. A deep understanding of interaction at the molecular level for cotton (*G. barbadense*) and its two major fungal and bacterial plant pathogens has been provided in the current study and it would be helpful in devising the disease resistant varieties and developing the disease management strategies as well as in improving crop production at the commercial level.

The results of the dynamic behaviour of the complexes provide compelling evidence that different pathogens interact differently with host PGIPs and mutations in PGIPs which could have varying effects on these interactions, ultimately impacting host resistance to pathogen infection. Stable trajectories were observed for interactions between the wild-type and mutated PGIPs with *X. citri* pv. *malvacearum*, suggesting that mutations do not significantly impact the ability of PGIP to inhibit xcPG. These findings aligned with the previous research where mutations did not compromise PGIP ability to inhibit the polygalacturonase activity of *Fusarium phyllophilum* (Maulik et al. 2013). Conversely, a mutated PGIP interacting with *A. macrospora* showed significant structural instability, potentially impairing its inhibitory effectiveness. These results emphasize the critical role of specific amino acids in maintaining the function of PGIP. This observation is consistent with the previous study where mutations in PGIPs were shown to affect their ability to inhibit *Verticillium dahliae* (Liu et al. 2018). These findings highlight the complex interplay between PGIP mutations and pathogen interactions, influencing host resistance to infection.

The observed alterations in protein complex dynamics due to mutations, also present a unique opportunity for crop protection in cotton through genome editing technologies like clustered regularly interspaced short palindromic repeats (CRISPR) -CRISPR associated protein (Cas) (Cardi et al. 2023). This approach aligns well with the growing interest in utilizing genome editing tools for crop improvement and disease resistance (Abdallah et al. 2015). The destabilization observed in mutated complexes, as demonstrated in MD simulations, offers insights into potential vulnerabilities which can be targeted for modification, aiming to optimize plant defence mechanisms. This strategy not only holds promise for accelerating breeding programs but also contributes to sustainable agriculture by developing cotton varieties with improved disease resistance, aligning with the broader goals of precision agriculture and crop protection. Further experimental validations will be significant in confirming the predicted effects of genomic modifications in plant phenotype and performance in real-world agricultural crop production.

Conclusion

The present study provides valuable insights into the molecular interaction between PGIP and PGs, shedding lights on the defense mechanisms employed by host plants *G. barbadense* against major plant pathogens. These findings have significant implications in modern crop production practices, cultivation, and plant breeding programs aimed at improving crop resistance to

diseases caused by plant pathogens like *A. macrospora* and *X. citri* pv. *malvacearum* in *G. barbadense*. By harnessing the knowledge gained from this study, researchers, plant pathologists, and breeders can devise the targeted approaches to enhance the plant defense systems and mitigate the impact of plant diseases on crop production and productivity. These findings open up new avenues for research and hold promise for developing disease-resistant cotton varieties and in further improving the crop protection strategies. By deciphering the intricacies of the cotton-pathogen interaction, these findings contribute to the knowledge advancement in molecular basis of host-pathogen interaction and crop science in particular, also showing a way for the sustainable management of cotton diseases.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42397-024-00203-z>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

We gratefully acknowledge CABIn grant (F. no. Agril. Edn.4-1/2013-A & P), Indian Council of Agricultural Research, Ministry of Agriculture and Farmers' Welfare, Govt. of India and ICAR-Indian Agricultural Statistics Research Institute, New Delhi for providing facilities and support. We acknowledge the Department of Biotechnology, Govt. of India for the BIC project grant (BT/PR40161/ BTIS/137/32/2021).

Authors' contributions

Kumar S and Nagrale DT designed the study; Murmu S, Rashmi M, Kour T, Singh MK, and Behera SK performed the experiments; Kumar S, Murmu S, Rashmi M, Behera SK, and Nagrale DT wrote the manuscript; Chaurasia A, Shankar R, Ranjan R, and Jha GK edited the manuscript and made modifications; Gawande SP, Hiremani NS, and Prasad YG have shared their expertise in giving their views on impact of such mutations and assisted in writing the manuscript.

Funding

CABIn grant (F. no. Agril. Edn.4-1/2013-A & P), Indian Council of Agricultural Research, Ministry of Agriculture and Farmers' Welfare, Govt. of India and Department of Biotechnology, Govt. of India for BIC project grant (BT/PR40161/ BTIS/137/32/2021)

Data availability

The data presented in this study are available in this article and the accompanying Supplementary Materials.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Yes.

Competing interests

The authors declare no conflict of interest.

Author details

¹ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India. ²ICAR-Central Institute for Cotton Research, Nagpur, India. ³SKUAST, Jammu, India. ⁴National Brain Research Center, Manesar, India. ⁵ICAR-Indian Institute of Vegetable Research, Varanasi, India. ⁶National Institute of Pharmaceutical Education and Research, Ahmedabad, India. ⁷ICAR-Indian Institute of Horticultural Research, Bengaluru, India. ⁸Dayalbagh Educational Institute, Agra, India.

Received: 27 June 2024 Accepted: 12 November 2024

Published online: 13 January 2025

References

- Abdallah NA, Prakash CS, McHughen AG. Genome editing for crop improvement: challenges and opportunities. *GM Crops Food*. 2015;6(4):183–205. <https://doi.org/10.1080/21645698.2015.1129937>.
- Amselem J, Cuomo CA, van Kan JA, et al. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis Cinerea*. *PLoS Genet*. 2011;7(8):e1002230. <https://doi.org/10.1371/journal.pgen.1002230>.
- Antonio Zancan WL, Gonzaga L, Silva G. Cotton in Brazil: importance and chemical control of bolls rot. In: Nita M, editor. *Fungicides - showcases of integrated plant disease management from around the world*. InTech; 2013. <https://doi.org/10.5772/53344>. Accessed 20 Apr 2024.
- Barras F, van Gijsegem F, Chatterjee AK. Extracellular enzymes and pathogenesis of soft-rot *Erwinia*. *Annu Rev Phytopathol*. 1994;32(1):201–34. <https://doi.org/10.1146/annurev.py.32.090194.001221>.
- Bekker H, Berendsen HJC, Dijkstra EJ, et al. Gromacs: A parallel computer for molecular dynamics simulations. In: DeGroot RA, Nadrichal J, editors. *Physics computing '92*. Singapore: World Scientific; 1993. p. 252–6.
- Bjelkmar P, Larsson P, Cuendet MA, et al. Implementation of the CHARMM force field in GROMACS: analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J Chem Theory Comput*. 2010;6(2):459–66. <https://doi.org/10.1021/ct900549r>.
- Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*. 1991;253(5016):164–70. <https://doi.org/10.1126/science.1853201>.
- Brown RL, Cleveland TE, Cotty PJ, et al. Spread of *aspergillus flavus* in cotton bolls, decay of intercarpellary membranes, and production of fungal pectinases. *Phytopathology*. 1992;82(4):462–7. <https://doi.org/10.1094/PHYTO-82-462>.
- Cardi T, Murovec J, Bakhsh A, et al. CRISPR/Cas-mediated plant genome editing: outstanding challenges a decade after implementation. *Trends Plant Sci*. 2023;28(10):1144–65. <https://doi.org/10.1016/j.tplants.2023.05.012>.
- Cervone F, De Lorenzo G, Pressey R, et al. Can Phaseolus PGIP inhibit pectic enzymes from microbes and plants? *Phytochemistry*. 1990;29(2):447–9. [https://doi.org/10.1016/0031-9422\(90\)85094-v](https://doi.org/10.1016/0031-9422(90)85094-v).
- Chattannavar SN, Kulkarni S, Khadi BM. Chemical control of *Alternaria blight* of cotton. *J Cotton Res Dev*. 2006;20(1):125–6.
- Choudhary P, Bhowmik A, Chakdar H, et al. Understanding the biological role of PqgB in *Pseudomonas stutzeri* using molecular dynamics simulation approach. *J Biomol Struct Dyn*. 2022;40(9):4237–49. <https://doi.org/10.1080/07391102.2020.1854860>.
- Collmer A, Keen NT. The role of pectic enzymes in plant pathogenesis. *Annu Rev Phytopathol*. 1986;24(1):383–409. <https://doi.org/10.1146/annurev.py.24.090186.002123>.
- Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. 1993;2(9):1511–9. <https://doi.org/10.1002/pro.5560020916>.
- Connolly ML. Solvent-accessible surfaces of proteins and nucleic-acids. *Science*. 1983;221(4612):709–13. <https://doi.org/10.1126/science.6879170>.
- De Lorenzo G, Ferrari S. Polygalacturonase-inhibiting proteins in defense against phytopathogenic fungi. *Curr Opin Plant Biol*. 2002;5(4):295–9. [https://doi.org/10.1016/s1369-5266\(02\)00271-6](https://doi.org/10.1016/s1369-5266(02)00271-6).
- Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*. 2003;125(7):1731–7. <https://doi.org/10.1021/ja026939x>.
- Fazil MHUT, Kumar S, Farmer R, et al. Binding efficiencies of carbohydrate ligands with different genotypes of cholera toxin B: molecular modeling, dynamics and docking simulation studies. *J Mol Model*. 2012;18(1):1–10. <https://doi.org/10.1007/s00894-010-0947-6>.
- Fortucci P. The contributions of cotton to economy and food security in developing countries. *World Cotton Research Conference Proceedings*, Cairo, Egypt, 20–25 Oct 2002. 2022. p. 1–5. https://staging.icac.org/meetings/cgtn_conf/documents/11_fortucci.pdf.
- Ge T, Huang X, Xie R. Recent advances in polygalacturonase gene in fruit tree species. *Plant Physiol J*. 2019;55:1075–88. <https://doi.org/10.13592/j.cnki.ppj.2019.0147>.
- Goldberg S, Koenning JT, Pitts J, et al. 2009 cotton disease loss estimate committee report. *Proceeding of the Beltwide Cotton Conference*. New Orleans, 4–7 Jan 2010; 2009. p. 237–40.
- Have At, Mulder W, Visser J, et al. The endopolygalacturonase gene *Bcpg1* is required for full virulence of *Botrytis cinerea*. *Mol Plant Microbe Interact*. 1998;11(10):1009–16. <https://doi.org/10.1094/mpmi.1998.11.10.1009>.
- Hollis P. Seed rot has lowered South Carolina cotton yields. *Decatur, IL, USA: Farm Progress*; 2002. <https://www.farmprogress.com/cotton/seed-rot-has-lowered-south-carolina-cotton-yields>. Access 3 Apr 2024.
- Hoofst RW, Vriend G, Sander C, et al. Errors in protein structures. *Nature*. 1996;381(6580):272. <https://doi.org/10.1038/381272a0>.
- Hu Y, Chen J, Fang L, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet*. 2019;51:739–48. <https://doi.org/10.1038/s41588-019-0371-5>.
- Huang Q, Allen C. Polygalacturonases are required for rapid colonization and full virulence of *Ralstonia solanacearum* on tomato plants. *Physiol Mol Plant Pathol*. 2000;57(2):77–83. <https://doi.org/10.1006/pmpp.2000.0283>.
- Hudson J. Seed rot hits South Carolina cotton. *Decatur, IL, USA: Farm Progress*; 2000. <https://www.farmprogress.com/cotton/seed-rot-hits-south-carolina-cotton>. Accessed 10 Apr 2024.
- ICAR-CICR, Annual Report. ICAR-Central Institute for Cotton Research, Nagpur, India; 2022. p. 135. Published by Director, ICAR-CICR, Nagpur.
- Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*. 2010;31(8):651–66. <https://doi.org/10.1016/j.patrec.2009.09.011>.
- Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2>.
- Kasza Z, Vagvölgyi C, Fèvre M, et al. Molecular characterization and *in planta* detection of *Sclerotinia Sclerotiorum* endopolygalacturonase genes. *Curr Microbiol*. 2004;48:208–13. <https://doi.org/10.1007/s00284-003-4166-6>.
- Kumar S, Ahmad K, Behera SK, et al. Biocomputational assessment of natural compounds as a potent inhibitor to Quorum sensors in *Ralstonia solanacearum*. *Molecules*. 2022;27(9):3034. <https://doi.org/10.3390/molecules27093034>.
- Kumar S, Behera SK, Gururaj K, et al. In silico mutation of aromatic with aliphatic amino acid residues in *Clostridium perfringens* epsilon toxin (ETX) reduces its binding efficiency to Caprine Myelin and lymphocyte (MAL) protein receptors. *J Biomol Struct Dyn*. 2023;42(5):2257–69. <https://doi.org/10.1080/07391102.2023.2204362>.
- Kumar S, Dehury B, Tandon G, et al. An insight into molecular interaction of PGIP with PG for banana cultivar. *Front Biosci (Landmark Ed)*. 2020;25(2):335–62. <https://doi.org/10.2741/4809>.
- Laskowski RA, MacArthur MW, Moss DS, et al. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 1993;26(2):283–91. <https://doi.org/10.1107/S0021889892009944>.
- Li R, Rimmer R, Buchwaldt L, et al. Interaction of *Sclerotinia sclerotiorum* with *Brassica napus*: cloning and characterization of endo- and exo-polygalacturonases expressed during saprophytic and parasitic modes. *Fungal Genet Biol*. 2004;41(8):754–65. <https://doi.org/10.1016/j.fgb.2004.03.002>.
- Liu N, Sun Y, Wang P, et al. Mutation of key amino acids in the polygalacturonase-inhibiting proteins Ck PGIP 1 and gh PGIP 1 improves resistance to Verticillium wilt in cotton. *Plant J*. 2018;96(3):546–61. <https://doi.org/10.1111/tpj.14048>.
- Maulik A, Basu S. Study of Q224K, V152G double mutation in bean PGIP2, an LRR protein for plant defense—an in silico approach. *Proteins*. 2013;81(5):852–62. <https://doi.org/10.1002/prot.24243>.
- Misas-Villamil JC, Van der Hoorn RAL. Enzyme–inhibitor interactions at the plant–pathogen interface. *Curr Opin Plant Biol*. 2008;11(4):380–8. <https://doi.org/10.1016/j.pbi.2008.04.007>.
- Murmu S, Archak S. In-silico study of protein-protein interactions in wheat blast using docking and molecular dynamics simulation approach. *J Biomol Struct Dynamics*. 2024;42(11):5747–57. <https://doi.org/10.1080/07391102.2023.2228907>.

- Oeser B, Heidrich PM, Müller U, et al. Polygalacturonase is a pathogenicity factor in the *Claviceps purpurea*/rye interaction. *Fungal Genet Biol*. 2002;36(3):176–86. [https://doi.org/10.1016/S1087-1845\(02\)00020-8](https://doi.org/10.1016/S1087-1845(02)00020-8).
- Rathinam M, Mishra P, Vasudevan M, et al. Comparative transcriptome analysis of pigeonpea, *Cajanus cajan* (L.) and one of its wild relatives *Cajanus platycarpus* (Benth.) Maesen. *PLoS One*. 2019;14(7):e0218731. <https://doi.org/10.1371/journal.pone.0218731>.
- Robert M, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;42(W1):W320–4. <https://doi.org/10.1093/nar/gku316>.
- Rodriguez-Palenzuela P, Burr TJ, Collmer A. Polygalacturonase is a virulence factor in *Agrobacterium tumefaciens* biovar 3. *J Bacteriol*. 1991;173(20):6547–52. <https://doi.org/10.1128/jb.173.20.6547-6552.1991>.
- Schwede T, Kopp J, Guex N, et al. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res*. 2003;31(13):3381–5. <https://doi.org/10.1093/nar/gkg520>.
- Shete PP, Kasal YG, Perane RR. Screening of the cotton genotypes against *Ramularia Areola* ATK under field condition. *Plant Archives*. 2018;18(1):734–6.
- Shieh MT, Brown RL, Whitehead MP, Cet al, et al. Molecular genetic evidence for the involvement of a specific polygalacturonase, P2c, in the invasion and spread of *Aspergillus flavus* in cotton bolls. *Appl Environ Microbiol*. 1997;63(9):3548–52. <https://doi.org/10.1128/aem.63.9.3548-3552.1997>.
- Singh S, Sablok G, Farmer R, et al. Molecular dynamic simulation and inhibitor prediction of cysteine synthase structured model as a potential drug target for trichomoniasis. *Biomed Res Int*. 2013;2013(1):390920. <https://doi.org/10.1155/2013/390920>.
- Singh UB, Singh S, Malviya D, et al. Integration of anti-penetrant tricyclazole, signaling molecule salicylic acid and root associated *Pseudomonas fluorescens* enhances suppression of *Bipolaris sorokiniana* in bread wheat (*Triticum aestivum* L.). *J Plant Pathol*. 2019;101:943–54. <https://doi.org/10.1007/s42161-019-00296-5>.
- Somai-Jemmal L, Siah A, Harbaoui K, et al. Correlation of fungal penetration, CWDE activities and defense-related genes with resistance of durum wheat cultivars to *Zymoseptoria tritici*. *Physiol Mol Plant Pathol*. 2017;100:117–25. <https://doi.org/10.1016/j.pmp.2017.08.003>.
- Varshney N, Murmu S, Baral B, et al. Unraveling the Aurora kinase A and Epstein-Barr nuclear antigen 1 axis in Epstein Barr virus associated gastric cancer. *Virology*. 2023;588:109901. <https://doi.org/10.1016/j.virol.2023.109901>.
- Vaught AV. Graphing with Gnuplot and Xmgr. *Linux Journal*. 1996. <https://www.linuxjournal.com/article/1218>. Retrieved 19 June 2009.
- Wallner B, Elofsson A. Can correct protein models be identified? *Protein Sci*. 2003;12(5):1073–86. <https://doi.org/10.1110/ps.0236803>.
- Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics*. 2016;54:5.6.1–37. <https://doi.org/10.1002/cpbi.3>.
- Weng G, Wang E, Wang Z, et al. HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res*. 2019;47(W1):W322–30. <https://doi.org/10.1093/nar/gkz397>.
- Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 2007;35(suppl_2):W407–10. <https://doi.org/10.1093/nar/gkm290>.
- Willard L, Ranjan A, Zhang H, et al. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res*. 2003;31(13):3316–9. <https://doi.org/10.1093/nar/gkg565>.
- Zhang N, Chen Y, Lu H, et al. MutaBind2: Predicting the impacts of single and multiple mutations on protein-protein interactions. *iScience*. 2020;23(3):100939. <https://doi.org/10.1016/j.isci.2020.100939>.