Computational, structural and functional aspects of hypothetical protein of *Aspergillus flavus* Pheromone Receptor Pre-A (PRP-A)

Maneesh Kumar¹, Sindhuprava Rana², Harish Kumar¹, Pratik Kumar¹, Manas Ranjan Dikhit², Rani Mansuri², Jainendra Kumar¹, Ganesh Chandra Sahoo²*

¹Department of Biotechnology, College of Commerce, Arts and Sciences, Magadh University (Bodh Gaya), Patna-800020, Bihar, India.
²BioMedical Informatics Center, Rajendra Memorial Research Institute of Medical Science (ICMR), Agam kuian, Patna-800007, Bihar, India.

**INTRODUCTION**

The genus *Aspergillus* belongs to Ascomycetes encompasses the most common filamentous fungi that can reproduce asexually by forming long conidiospores chains (Ronald Morris *et al.* 1989). *Aspergillus flavus* is generally known for its aflatoxin, a secondary metabolite production, which is highly toxic, mutagenic and carcinogenic to both plants and animals. Aflatoxin contaminates various agricultural products that cause serious health hazards in animals and humans just by inhalation of the fungal spores, having harsh symptoms associated with skin lesion and respiratory problems (Hedayati *et al.* 2007). The biosynthesis pathway of aflatoxin is very much complex and various enzymes are involved that directly or indirectly regulated signals that receive from various receptors (Anderson 1992). Along with aflatoxin biosynthesis in *A. flavus*, the virulence, survival and mating are also regulated by G protein-mediated signaling pathway. Heterotrimeric G protein-mediated signal perception and propagation are conserved from lower eukaryotes to humans. G proteins are a family of heterotrimeric GTPases that exclusively have a huge effect on eukaryotic signal transduction through the coupling of surface receptors to cytoplasmic effector proteins (Dohlman and Thorner 2001; Lengeler *et al.* 2000).
In this filamentous fungus, an unusual mating type gene has been discovered recently. The protein encoded by the gene behaves as pheromone receptor that determines the cell identity. The receptor protein effectively participates in the proliferation of cell and regulates the germination and quorum sensing in heterothallic *Aspergillus flavus* (Coppin et al. 1997; Shi and Glass 2000). The gpr B gene encoding putative GPRCs that is distinctively causes self-fertilization in homothallic fungus *A. nidulans*.

This gpr B is highly similar to *A. fumigates* Pre-A. It can be further analyzed that gpr B (Pre-A) is required for the specialized cell fusion to form a dikaryotic hyphae which is a type of homothallic self-fertilization. In some other fungi, it has been studied that the recognition between nuclei is mediated by the nucleus–limited gene expression of mating type-specific pheromone and receptors (Pöggeler 2002; Debuchy 1999) proofs to be a good target. The analysis of structural features, PRP-A has been taken for our study that responsible for sexual mating in *A. flavus*. Various tools and softwares have been used to understand the natural existence the desired protein. The homology modeled *A. flavus* PRP-A structure was predicted followed by simulation and docking with suitable ligands to see the protein-ligand interaction.

**MATERIALS AND METHODS**

**The Identification of the protein sequence**

To predict the structure and function of the desired protein sequence, various bioinformatics tools and softwares have been used. The primary sequence of the PRP-A in *Aspergillus flavus* Gene ID: AFLA_061620) was taken from NCBI protein database having Acc. No: XP_002378906.1 (Pruitt et al. 2009; Affeldt et al. 2014). This protein sequence has been taken for molecular modeling, computational analyses and to predict the Protein-Ligand interaction with suitable ligands that have the potential to inhibit the protein activities.

**Sequence Analysis and Secondary Structure Prediction**

For a secondary structure analysis, we used GOR4 Server from protein sequence (Garnier et al. 1996). The NCBI Blast was used to compare the query sequence to find its homologues. Conserved domains were determined from BLAST analysis (Table 3). The transmembrane helical regions of Pheromone Receptor Pre-A protein topology prediction and validation were done by using various servers like TMHMM (Krogh et al. 2001), HMMTOP (Tusnady and Simon 1998), TMpred (Claros and von Heijne 1994), MEMSAT (Jones, Taylor, and Thornton 1994) and TopPred (Hofman 1993), that predicted the nature of the query sequence (Sahoo et al. 2013).

**3D Structure Prediction and Model Prediction**

The 3D structure of an *Aspergillus flavus* PRP-A was performed by various online servers like knowledge-based approach (Swiss Model) (Arnold et al. 2006), structure prediction by HMM-HMM comparison (HMpred) (Soding 2005; Remmert et al. 2012), hierarchical method of protein structure and function prediction (I-Tasser) (Zhang 2008), profile-profile matching (PHYRE) (Kelley and Sternberg 2009) and protein structure prediction (Raptor X) (Källberg et al. 2012). Along with all these servers, homology modeling was performed by Modeler of DSv3.5. Based on the DOPE score (Shen and Salì 2006) the best model was selected.

The structural evaluation was carried out by Ramachandran Plot via PROCHECK (Laskowski et al. 1996), Verifield 3D (Bowie, Luthy, and Eisenberg 1991; Luthy, Bowie, and Eisenberg 1992) and ERRAT (Colovos and Yeates 1993) was used to analyze the structural error at each residue of modeled structure. Further validation of the model was done through flexible loop and side chain refinement in DSv3.5. The protein folding energy was evaluated by using ProSA server (Wiederstein and Sippl 2007). The server provided us Z-score that indicates overall model quality.

**Protein Stimulation**

The predicted modeled protein was further stimulated and refined by CHARMM (Karplus 1983) using DSv3.5. CHARMM is a versatile and standard dynamic molecular stimulation program that parameterized the protein atoms. Stimulations were carried out at 300K with 2000 steps of steepest descent minimization techniques, minimization RMS Gradient (0.1), minimization energy change, and implicit solvent model (distance dependent dielectrics), until the RMSD was less than 0.001 kcal mol$^{-1}$ Å$^{-1}$ (Sahoo et al. 2014; Sahoo et al., 2009).

**Active Sites Prediction**

The binding site module has been identified by using DSv3.5. that provides the proper identification and characterization of protein binding/active sites. The entire amino acids of 4JKV_A were selected and allowed Protein Preparation using CHARMM force. The all binding sites are highly active and functional residues were identified and stored for further analysis.

**Docking**

After protein preparation ligand library like Blasticidin, Piperonaline, Piperin, Piperlongumine, Lutein (Xanthophyll), Eriodictyol, Xanthotoxin, Psoralen, Eugenol and Nonyl-aldehyde with their known IC50 value (Holmes, Boston, and Payne 2008; Ansari et al. 2012) was prepared from NBCI PubChem Compound database. Then docking of Protein and Ligands was done by LibDock protocol of DSv3.5 (Rao et al. 2007).

**RESULTS AND DISCUSSION**

In this study, the combined use of both softwares and bioinformatics tool based on the homology analysis of the protein sequence of G-Protein receptor PRP-A with the hypothetical protein 4JKV_A has been retrieved from RCSB PDB tool.
The comparative analysis of transmembrane helices. The computational analysis showed the lowest range and higher range of transmembrane regions in GPRP-A which is summarized in Table 4. The comparative analysis of transmembrane helices prediction programs showed that the lowest range and higher range of transmembrane helices in the first TM is 12-32 residues, 39-61 in second TM, 82-102 in third TM, 123-145 in fourth TM, 167-188 fifth TM, 220-239 in sixth TM and 279-295 seventh TM. This computational analysis showed that there are 7 transmembrane helices in the query sequence that are participated in receptor formation (Table 3).

### Secondary Structure Analysis

The secondary structure analysis of the query protein sequence obtained from GOR4 server shows that random coil was most frequent (52.04 %), followed by alpha helix (13.76%). Extended strand was found to be 34.19% (Table 1) (Neelamathi et al. 2009). The query sequence comparison was evaluated by percentage identity, score and E-value of top five sequences (Table 2).

### Transmembrane helices (TMs) prediction

Five different transmembrane prediction servers like TMHMM, HMMTOP, TMpred, MEMSAT and TopPred were used to predict and validate the position and number of transmembrane regions in G-protein PRP-A which is summarized in Table 4. The comparative analysis of transmembrane helices prediction programs showed that the lowest range and higher range of transmembrane helices in the first TM is 12-32 residues, 39-61 in second TM, 82-102 in third TM, 123-145 in fourth TM, 167-188 fifth TM, 220-239 in sixth TM and 279-295 seventh TM. This computational analysis showed that there are 7 transmembrane helices in the query sequence that are participated in receptor formation (Table 3).

### 3D Structure prediction using homology modeling approaches

3D structure analysis enables to understand the structure, functions, and localization of the receptor protein and their interaction with antifungal ligands. The most common and appropriate prediction method is homology modeling that gives a proper idea about the protein. In the absence of the 3D structure of pheromone receptor Pre-A, we prompted for homology modeling. Suitable template protein was selected on the basis of the sequence similarity with the query sequence that were searched through various online servers and also with inbuilt modeler in DSv3.5. The homology model of the hypothetical protein of PRP-A has shown Fig 1. The figure showed with labeled as sequence alpha (α), beta (β) and flexible loops (FL). All the models were compared and validated by DOPE scores of DSv3.5 (Fig 4). The most suitable template PDB ID: 4JKV_A that retrieved from the HMpred server has been taken with lowest DOPE value (Fig 3) of -61153.003706 as the best-modeled structure which chosen for our further validation.
The model structure was proved by Verify 3D that showed 86% value (Fig 2). The model validation PROCHECK tool was used to determine Ramachandran plot (Fig 4) to assure the quality model.

The result of Ramachandran plot showed 93.3% of residue in favored regions, 6.2% of residues in additional allowed regions, 0.5% generously allowed regions and disallowed regions favored 0% represents that it is reliable and good quality model. The Z-score indicated the overall model quality. The Z-score -7.09 (Fig 5) of input protein model was obtained from ProSA. The reliability of the modeled protein was also checked by using ERRAT that showed 93.072 overall model quality (Fig 6).
**Fig. 3:** Comparative Analysis of 3D Models of Pheromone Pre-B from Different Servers and Software (DSv3.5).

**Fig. 4:** Ramachandran Plot by PROCHECK.
Protein-Ligand Interaction

After detecting the active binding sites of the model PRP-A protein, we tried to analyze the specific substrate ligands that were effectively docked with the 3D model. There are eleven different binding sites were detected by using receptor cavities protocol of DSv3.5. The highest LibDock score has been calculated as 140.104 with Blasticidin S at binding site 1 (the position value of the site 1: -28.787, 22.2787, 20.1339, 19.6) of the model protein. Blasticidin S alone gave 7 different posed at site 1 during dock. It is an effective selective nucleoside antibiotic that acts both eukaryotic and prokaryotic cells. It is an antibiotic, which is isolated from *Streptomyces griseochromogenes* that inhibit translation by altering termination step in both prokaryotic and eukaryotic cells (Takeuchi et al. 1958; Yamaguchi and TANAKA 1966). It shows quick action and causes cell death even at low concentration. It also showed efficient binding respectively that might be the next potential docking values, but it failed to dock with other binding sites within the model protein. Along with Blasticidin S ligand, there are some others ligands which are also perfectly bound at this site 1 as shown in Table 4 along with their LibDock Score. They are Pipernonaline, Piperin, Piperlongumine, Lutein, Eriodictyol, Xanthotoxin, Psoralen, Eugenol and Nonaldehyde. The model protein with ligand Blasticidin S binding was shown in Fig 7.

The figure gives the hypothetical 3D representation of subcellular localization of the model PRP-A protein along with inserted ligand at the outer membrane region of that plasma membrane. The groove contains some positively charged side, negatively charged side and aromatic side chains that interact directly with corresponding charges of the Ligand. The proper Protein-Ligand interaction is shown in Fig8. The PRP-A (gpr B) is just homologous to STE 3 GPCR of *Saccharomyces cerevisiae* pheromone receptor that shares several motifs mainly 7TM. Here, the frequently involved amino acids of model protein that form hydrogen bonds with the ligand are Tyr 394 and Arg 40. The group i.e.: OH of Tyr394 interacts simultaneously with :H33 and :H34 of Blasticidin S, and :HH1 of Arg 400 effectively interact with :O2 of Blasticidin S (Fig 9). Our docking result suggests that the model protein binds close to the active site with similar binding energy.
Fig. 7: 3D representation of subcellular localization of the model protein in Plasma Membrane.

Fig. 8: Protein-Ligand Interaction at Site 1.
CONCLUSION

The main objective of this work was to understand the structural and functional features of PRP-A, a mating type GPCR protein found in heterothallic filamentous Aspergillus flavus, which elicit self-fertilization in the presence of their opposite partner. The sequence analysis and structural analysis of the GPCR protein, PRP-A suggests that the modeled protein is having good geometry and acceptable 3D-profile. The Protein-Ligand interaction was performed using DSv3.5. Compounds like Blasticidin S, Pipernonaline, Piperin, Piperlongumine, and Lutein exhibited high binding activities with the receptor protein.

ACKNOWLEDGEMENTS

I acknowledge ICMR to provide me the computational facility for carrying this work in RMRIMS.

Financial support and sponsorship: Nil.

Conflict of Interests: There are no conflicts of interest.

REFERENCES


