BMC Bioinformatics

Open Access

GPCRs from *fusarium graminearum* detection, modeling and virtual screening - the search for new routes to control head blight disease

Emmanuel Bresso[1], Roberto Togawa[1], Kim Hammond-Kosack[2], Martin Urban[2], Bernard Maigret[3,1] and Natalia Florencio Martins[1*]

Abstract

Backgound: *Fusarium graminearum* (FG) is one of the major cereal infecting pathogens causing high economic losses worldwide and resulting in adverse effects on human and animal health. Therefore, the development of new fungicides against FG is an important issue to reduce cereal infection and economic impact. In the strategy for developing new fungicides, a critical step is the identification of new targets against which innovative chemicals weapons can be designed. As several G-protein coupled receptors (GPCRs) are implicated in signaling pathways critical for the fungi development and survival, such proteins could be valuable efficient targets to reduce *Fusarium* growth and therefore to prevent food contamination.

Results: In this study, GPCRs were predicted in the FG proteome using a manually curated pipeline dedicated to the identification of GPCRs. Based on several successive filters, the most appropriate GPCR candidate target for developing new fungicides was selected. Searching for new compounds blocking this particular target requires the knowledge of its 3D-structure. As no experimental X-Ray structure of the selected protein was available, a 3D model was built by homology modeling. The model quality and stability was checked by 100 ns of molecular dynamics simulations. Two stable conformations representative of the conformational families of the protein were extracted from the 100 ns simulation and were used for an ensemble docking campaign. The model quality and stability was checked by 100 ns of molecular dynamics simulations previously to the virtual screening step. The virtual screening step comprised the exploration of a chemical library with 11,000 compounds that were docked to the GPCR model. Among these compounds, we selected the ten top-ranked nontoxic molecules proposed to be experimentally tested to validate the *in silico* simulation.

Conclusions: This study provides an integrated process merging genomics, structural bioinformatics and drug design for proposing innovative solutions to a world wide threat to grain producers and consumers.

Keywords: G-protein coupled receptors, *Fusarium graminearum*, Structural bioinformatics, Fusarium head blight

* Correspondence: natalia.martins@embrapa.br
[1]EMBRAPA Genetic Resources and Biotechnology, Brasília, DF 70770-917, Brazil
Full list of author information is available at the end of the article

Background

The ascomycete *Fusarium graminearum* (FG) is a filamentous fungus dwelling on and in a wide range of plant species, on crop debris and within soil. This fungus causes Fusarium head blight (FHB) disease on wheat, barley. FG is also responsible for various corn and rice diseases [1]. FG is a highly destructive pathogen of cereals reduces grain quality rather than grain production. FG causes two main problems: first, seed quality is reduced, and secondly, infection produces mycotoxin-contaminated grains. Among the various sesquiterpenoid trichothecene toxins produced by FG, deoxynivalenol, also known as vomitoxin, is one of the most important [2]. Deoxynivalenol contaminated grains are often considered unfit for animals and/or human consumption leading to considerable economic losses [3, 4].

Fungicide applications are only moderately effective at controlling FHB and often intrinsic resistance problems have been encountered [5, 6]. The identification of new fungicides is urgently required to limit FG development. In the search for new, efficient and selective fungicides able to control the development of the pathogen, the first step is to find relevant targets [7, 8].

G-protein coupled receptors (GPCRs) are the starting point for the control of several signaling pathways and are therefore considered a potentially rich source of innovation as drug targets and for drug design to alleviate many human diseases of genetic and / or biotic origins [9]. GPCRs, which are activated by a large panel of factors ranging from light, small amines to hormones and chemokines, initiate signaling cascades resulting in multiple cell responses. GPCRs constitute a large family of proteins, the signature of which consists of a transmembrane domain embedded within the plasma membrane and possess seven transmembrane helices. Their functions are to detect extracellular signals and to activate intracellular mediated signal transduction pathways and appropriate cellular responses. GPCRs classically transmit a signal via the activation of heterotrimeric G proteins. The sustained stimulation leads to the activation of G protein-coupled receptor kinases and the recruitment of arrestin proteins, which engage alternative signaling pathways [10].

In fungi, GPCRs are known to be implicated in biological processes including vegetative growth, sporulation, stress responses and pathogenicity [11]. GPCRs have been the subject of numerous bioinformatics studies to explore their potential suitability as drug targets [12]. As a result, the entire set of GPCRs encoded by various fungi has been predicted for several fungi including *Saccharomyces cerevisae*, *Schizosaccharomyces pombe*, the saprobes *Aspergillus spp.*, *Neurospora crassa*, and *Trichoderma spp.*, the plant pathogens *Magnaporthe grisea* and *Verticillium spp.*, and the human/animal infecting pathogen *Cryptococcus*

neoformans [13–21]. Ma et al. previously explored the GPCRs repertoire for *Fusarium* species, but only sequence alignments were used for these predictions [22]. As GPCRs are known not to share a high sequence similarity, such predictions may increase the risk of occurrence of false positives [23, 24].

The primary goal of this study was to detect GPCR s in the predicted FG proteome, to select the best candidates for potential use to control this pathogen and to identify potential inhibitors. Several *in silico* predictive filters were used leading to the selection of one of the most relevant GPCR target. Prior to the rational screening of putative active compounds, the three-dimensional structure of this GPCR should be known. As no FG GPCRs' 3D structures are presently available, we used homology modeling and molecular dynamics simulations in order to obtain a convincing model for the selected GPCR candidate. Then, stable conformations of this model were used to identify potential inhibitors using the virtual screening approach [25, 26].

Results and discussion

Identification of putative GPCRs

The 13,321 predicted protein sequences of *Fusarium graminearum* (Version 32) were submitted to the bespoke GPCRpipe, and only nine sequences were identified as putative GPCRs. After checking for the number of their transmembrane helices by TMHMM, HMMtop, and Phobius, only six proteins among the nine were found to contain the necessary 7 transmembrane helix (TM) feature confirmed by all three programs (Table 1). For the three others, as at least two programs predicted 7 TMs, we finally considered all the nine as putative GPCRs. Furthermore, all these predicted GPCRs presented an extracellular N-terminus and an intracellular C-terminus like other known GPCRs, strengthening, therefore, this selection. In the group of nine candidates, seven are in common with the previous annotation by

Table 1 Number of transmembrane helices determined by TMHMM, HMMtop and Phobius

Protein ID	# Helices		
	TMHMM	HMMtop	Phobius
FGSG_01861	7	7	7
FGSG_02655	6	7	7
FGSG_02942	7	7	5
FGSG_03023	7	7	7
FGSG_05006	5	7	7
FGSG_05239	7	7	7
FGSG_05404	7	7	7
FGSG_07270	7	7	7
FGSG_07716	7	7	7

Ma et al. (2010). Among them two (FGSG_02942 and FGSG_05404) are novel GPCR candidates coming from the stringent structural/function prediction (Table 1).

Functional classification and final selection

As fungal GPCRs are associated with different functions, identification of these functions for the nine putative GPCRs was used as a first step to select the best target to inhibit fungal growth. Zheng et al. (2010) proposed a new classification of fungal GPCRs and classified 40 GPCRs from the ascomycete pathogens *Verticillium dahliae* and *Verticillium albo-atrum*, and we used these sequences to build the phylogenetic tree of the nine putative GPCRs (Fig. 1). FGSG_05239, FGSG_07716, FGSG_03023 and FGSG_01861 are present in the same branch as class V GPCRs. FGSG_05006 and FGSG_02942 are grouped into class III GPCRs. FGSG_02655 and FGSG_07270 are both pheromone receptors belonging respectively to class I and class II. The last GPCR, FGSG_05404, is similar to class X. To confirm this phylogeny relationship, a domain composition of each protein was determined using Pfam (Table 2) showing that all extracted domains are coherent with GPCRs functions: STE2 and STE3 are mating type pheromone receptor domains, Git3 is a glucose receptor, 7tm_1 and 7tm_2 are respectively rhodopsin and secretin receptors, dicty_CAR is a cyclic AMP receptor, and Lung_7-TM_R is a known seven transmembrane helix domain.

Table 2 Classes of the retained GPCRs identified in FG and Pfam domains

Protein ID	Conserved Pfam domain	Class [17]
FGSG_02655	STE2 (PF02116)	I. Ste2-like pheromone receptor
FGSG_07270	STE3 (PF02076)	II. Ste3-like pheromone receptor
FGSG_02942	Git3 (PF11710)	III. G protein-coupled glucose receptor regulating Gpa2
FGSG_05006		
FGSG_05239	7tm_1 (PF00001)	V. cAMP receptor like
FGSG_01861	7tm_2 (PF00002)	
FGSG_03023	Dicty_CAR (PF05462)	
FGSG_07716		
FGSG_05404	Lung_7-TM_R (PF06814)	X. PTM1-like GPCR

To identify potential new fungicide targets, an important step is to verify that the identified targets are not present in host organisms (principally wheat and corn) and humans. As shown in Table 3, for eight of the nine identified GPCRs, no similar protein was found in *Homo sapiens*, *Zea mays* or *Triticum* species. For the putative GPCR FGSG_05404 similar was found to a protein occurring in the three different tested species. As a consequence, FGSG_05404 was discarded from any further analyses.

From the eight remaining possible GPCR targets, only FGSG_02655, predicted to code for a class I pheromone receptor, was retained for entering into the molecular
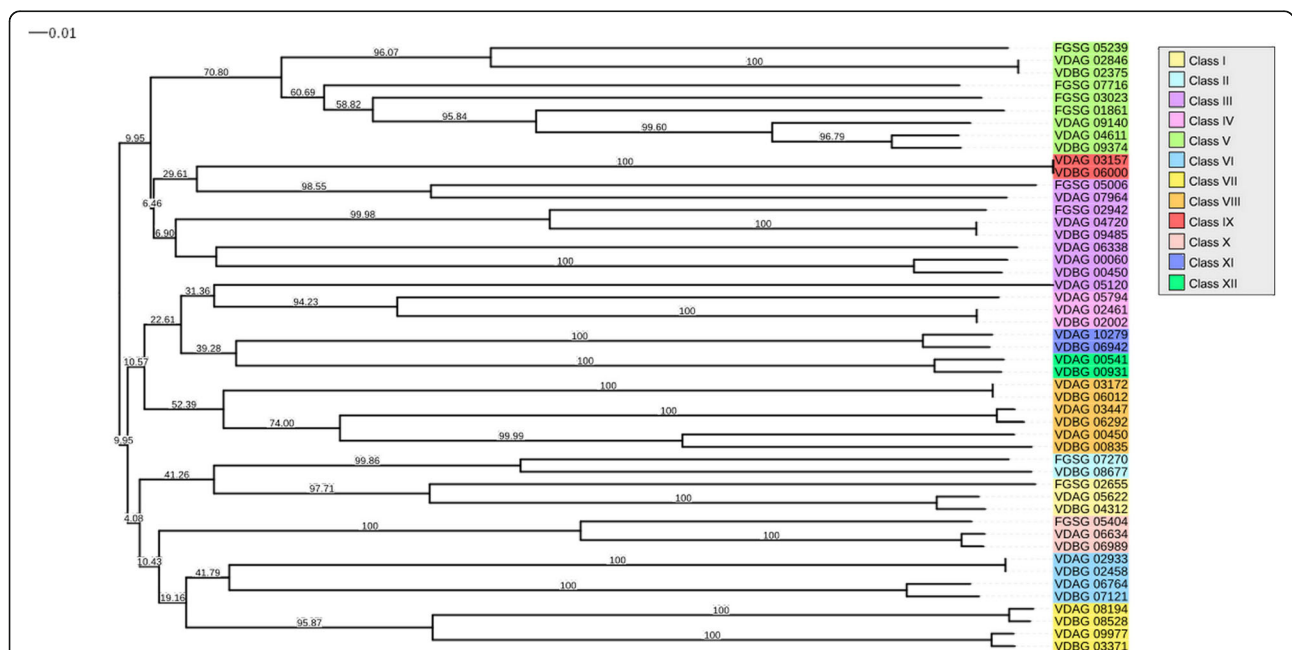

Fig. 1 Phylogeny of nine putative GPCRs identified in *Fusarium graminearum* and 40 GPCRs identified by Zheng et al. in *Verticillium dahliae* and *Verticillium albo-atrum*. The unrooted tree with bootstrap value (10,000 repetitions) shown in every branch was constructed using the neighbor-joining method

Table 3 Similarity with other species

Protein ID	Blastp best results								
	Zea mays			Homo sapiens			Triticum.		
	E-value	Query-Cover	Ident	E-value	Query-Cover	Ident	E-value	Query-Cover	Ident
FGSG_02655	–	–	–	–	–	–	–	–	–
FGSG_07270	–	–	–	–	–	–	–	–	–
FGSG_02942	–	–	–	8e-03	46%	26%	–	–	–
FGSG_05006	–	–	–	5e-02	26%	25%	–	–	–
FGSG_05239	–	–	–	7e-03	41%	24%	–	–	–
FGSG_01861	4e-06	39%	26%	2e-04	41%	24%	–	–	–
FGSG_03023	–	–	–	2e-02	44%	24%	–	–	–
FGSG_07716	6e-05	85%	24%	3e-04	43%	25%	–	–	–
FGSG_05404	6e-33	70%	30%	9e-21	51%	28%	6e-29	73%	27%

Blastp best results: the nine identified GPCR vs Homo sapiens, Zea mays & Triticum (nr database). Only best results with an E-value lower than 1e-01 are displayed

modeling pipeline. This protein contains a mating type pheromone receptor domain (PF02116). In fungi, one of the first steps in sexual reproduction is sexual phero- mone reception by a mating type receptor. As sexual reproduction is a fundamental part of the FG disease cycle, FGSG_02655 is therefore considered a good choice for developing as a fungicide target [27]. Using a reverse genetics approach, FGSG_02655 has already been shown to be required for sexual mating and viru- lence in *F. graminearum.* Single gene deletion strains ex- hibit reduce female fertility and fewer mature perithecia were produced when strains were selfed [28, 29]. In addition, a separate study [30] confirmed the sexual mat- ing defects and also showed the reduced ability of a delta FGSG_02655 strain to cause disease on wheat ears and maize cobs. Therefore, the selection of FGSG_02655 as a candidate fungicide target could be used to restrict *F.*

graminearum growth and development at two distinct phases in its predominately monocyclic disease cycle.

Homology modeling of FGSG_02655

The first step in homology modeling is to identify the most suitable templates to use to build the query 3D model. The sequence similarity between FGSG_02655 and known GPCRs' PDB was found to be very low. Therefore, no structure could be used as a template. In- stead the FGSG_02655 transmembrane helices were pre- dicted using several tools (Table 4). Firstly, sequence similarity was not used to build the models as it is usu- ally done in homology modeling. The 7 TM helices posi- tions were used, aligning the TM sequences predicted for the FGSG_02655 query with those observed in the PDB templates. Moreover, we used supplementary infor- mation concerning the third transmembrane helix: in

Table 4 Transmembrane helix predictions for FGSG_02655

| Prediction method	TM1	TM2	TM3	TM4	TM5	TM6	TM7
DAS	49–63	74–93	126–136	155–174	193–220	241–258	271–285
PRED-TMR	45–63	71–93	156–178	192–214	240–258	269–187	
HMMTOP	46–63	72–93	116–134	155–177	200–219	240–258	269–286
TMHMM	42–64	71–93	113–135	156–178	198–220	241–263	
GPCRHMM	41–63	72–93	114–135	152–171	192–214	232–258	263–283
PredictProtein	44–61	76–95	123–143	155–179	202–219	241–259	269–286
TOPCON	43–62	72–92	120–140	156–176	200–220	240–260	268–298
MINNOU	39–62	71–105	118–147	152–185	192–224	243–257	269–289
SOSUI	117–138	146–168	193–215	225–247	268–290	310–332	340–362
SPLIT	122–138	147–170	198–213	227–251	264–296	312–338	350–368
TMpred	127–146	155–175	198–216	237–256	277–296	322–338	350–368
Model 1	39–61	86–109	129–156	176–197	230–252	275–299	312–333
Model 2	39–61	95–118	129–156	173–195	225–247	263–287	311–337
Model 3	39–61	86–109	129–156	176–197	225–247	263–287	307–328
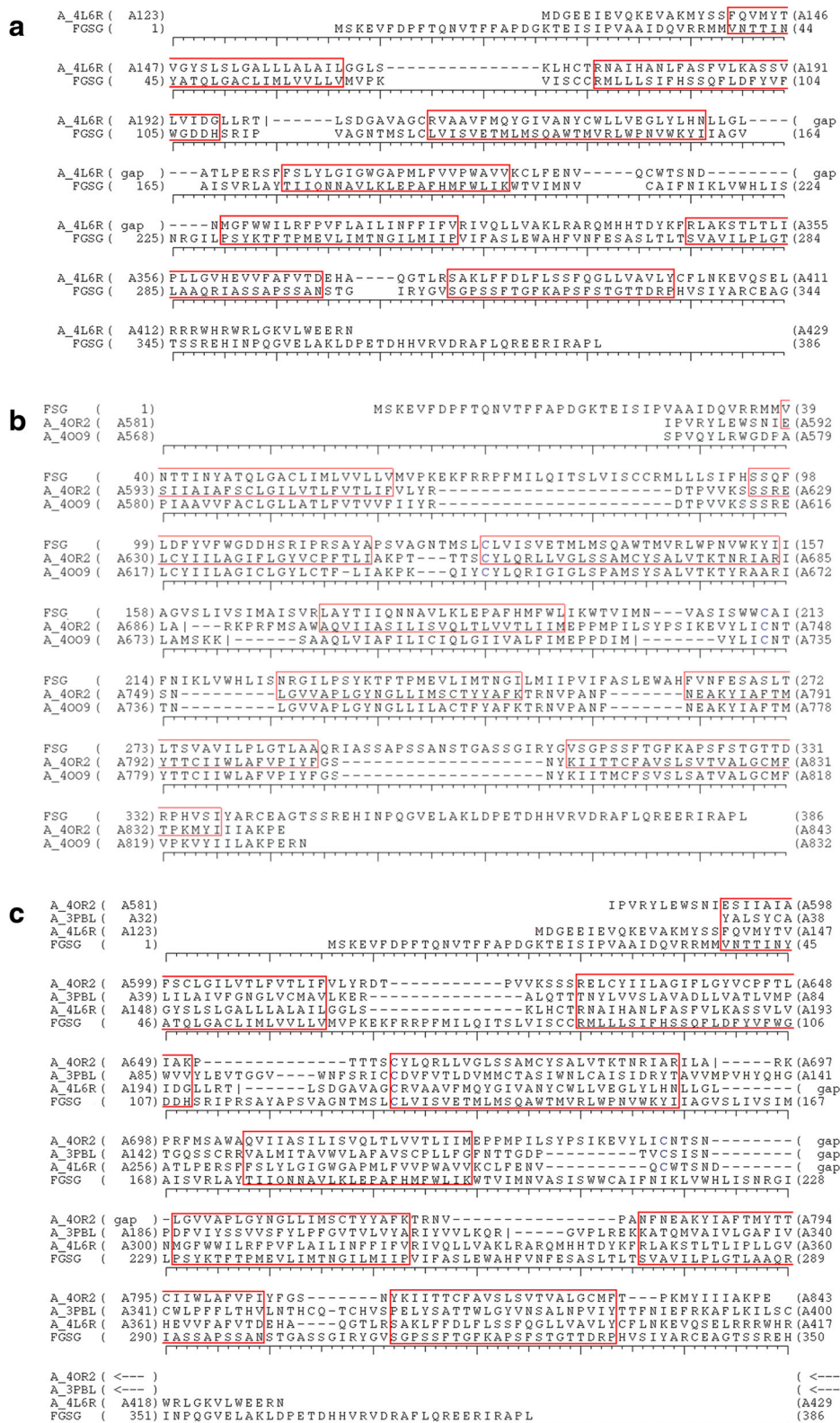
Fig. 2 (See legend on next page.)

(See figure on previous page.)
Fig. 2 a Alignment of FGSG_02655 with a sequence of human glucagon G-protein coupled receptor (4L6R). This alignment was used to build the model 1. **b** Alignment of FGSG_02655 with sequences of human G protein-coupled metabotropic glutamate receptor 1 (4OR2) and human metabotropic glutamate receptor 5 (4OO9). This alignment was used to build the model. **c** Alignment of FSG_02655 with sequences of G protein-coupled metabotropic glutamate receptor 1 (4OR2), the human dopamine D3 receptor (3PBL) and human glucagon G-protein coupled receptor (4L6R). This alignment was used to build the model 3. Red squares correspond to transmembrane helix positions

the structures used as templates as well as in the majority of PDB GPCR structures [31], a cysteine forms a disulfide bond with the second extracellular loop [12]. After analyzing these different transmembrane helix predictions, three models of FGSG_022655 were finally retained. The corresponding TM helices are highlighted in Table 4, and the corresponding sequence alignments are presented in Fig. 2a–c. As quality score determined by DOPE (discrete optimized protein energy) were similar (Table 5) for the 3 selected models, they were submitted to molecular dynamics (MD) simulations to check their stabilities.

Checking the stability of the models by molecular dynamics

The three protein models of FGSG_02655 selected from the previous step were subjected to a ten ns MD simulation to determine their stability as a GPCR. Evaluation of stability was realized by analyzing secondary structure evolution during the dynamics. The main focus of this analysis was the transmembrane helices: if the seven transmembrane helices were not broken during the Molecular Dynamics, then we consider the model as stable. The analysis of secondary structure evolution shows that for two protein models (model 1 and 2), transmembrane helices were not retained during the simulation. In model 2, the third transmembrane helix broke during the equilibration and after 4 ns TM helix 7 was broken. Both helices did not re-form during the MD. Concerning model 1, helix 4 was broken at the beginning of the Molecular Dynamics and did not reform during the simulation. These observations indicate that model 1 and model 2 were not stable and were not considered further.

For model 3 (Fig. 3), all the secondary structures were stable during the original simulation. As this model appeared to be stable, we prolonged the MD simulation until 100 ns.

Table 5 DOPE score of the 3 selected models

Model	DOPE score
1	−34,627.5
2	−34,211.42
3	−34,766.89

Scores were calculated by the assess_dope function of Modeller

Choice of the final model

The timeline analysis of the secondary structures during the 100 ns of MD simulation showed that the third model is stable during the whole trajectory and especially the transmembrane helices (Fig. 4). Then, each frame of the trajectory was aligned to the first one based on the protein backbone. Root-mean-square deviation (RMSD) analysis was performed for both the whole protein and its seven transmembrane helices (Fig. 5). It appears from these simulations that a very stable conformational regime was obtained after 42 ns, mostly due to a rearrangement of the TM helices within the palmitoyloleoylphosphatidylcholine (POPC) bilayer. During the remaining 58 ns, the protein conformation fluctuates between two quite similar conformational families (Fig. 6). Concerning these two conformations, three of their transmembrane helices had the same positions (helices three, five and seven), two were shorter in the first conformation (helices one and six) and two others were shorter in the second conformation (helices two and four) (Fig. 7). For the first family, the distance between the α carbon of the N-terminal residue and the nearest POPC atom was around 28 Å and 20 Å for the C-terminal side. For the second conformation, these same distances were both 13 Å, showing that the first conformation family was the outermost of the membrane. The representative 3D structures of each of these families can be used for performing a structure-based virtual chemical screen using the ensemble docking procedure.

Virtual screening

From the ensemble docking campaign using these two main conformers, we retained only the top 30 compounds from the complete GOLD score list to further analysis. After removing possible toxic molecules and compounds outside the pocket, we retained only 10 molecules (Table 6) for possible biological testing. It should be noted that the majority of the compounds bound the first conformation. The chemical formulas and names of these molecules are shown on Additional file 1: Table S1. The positioning of the 3 best-score retained compounds within the binding site is presented on Fig. 8. Looking at the protein/ligand interaction found, it appears that each ligand had specific interactions with the protein: for example, compound F0514-4158 interacts with Phe_{214} Lys_{217} and Ser_{231}, while the molecule F0514-3978 interacts only with Ser_{292}.
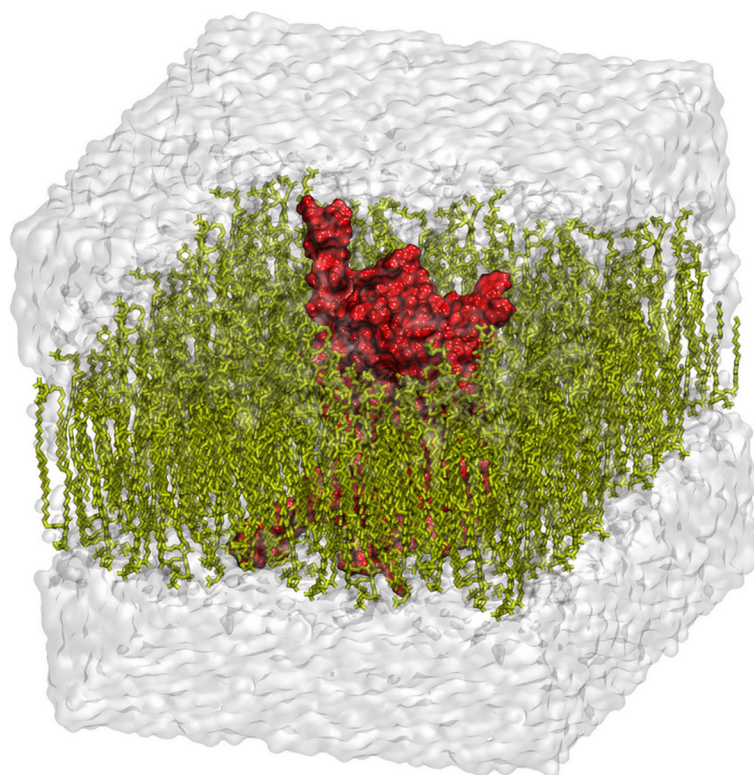
Fig. 3 The appearance of the initial Model 3. The red color indicates the GPCR, the olive green color indicates the membrane lipids and the grey color indicates the water box
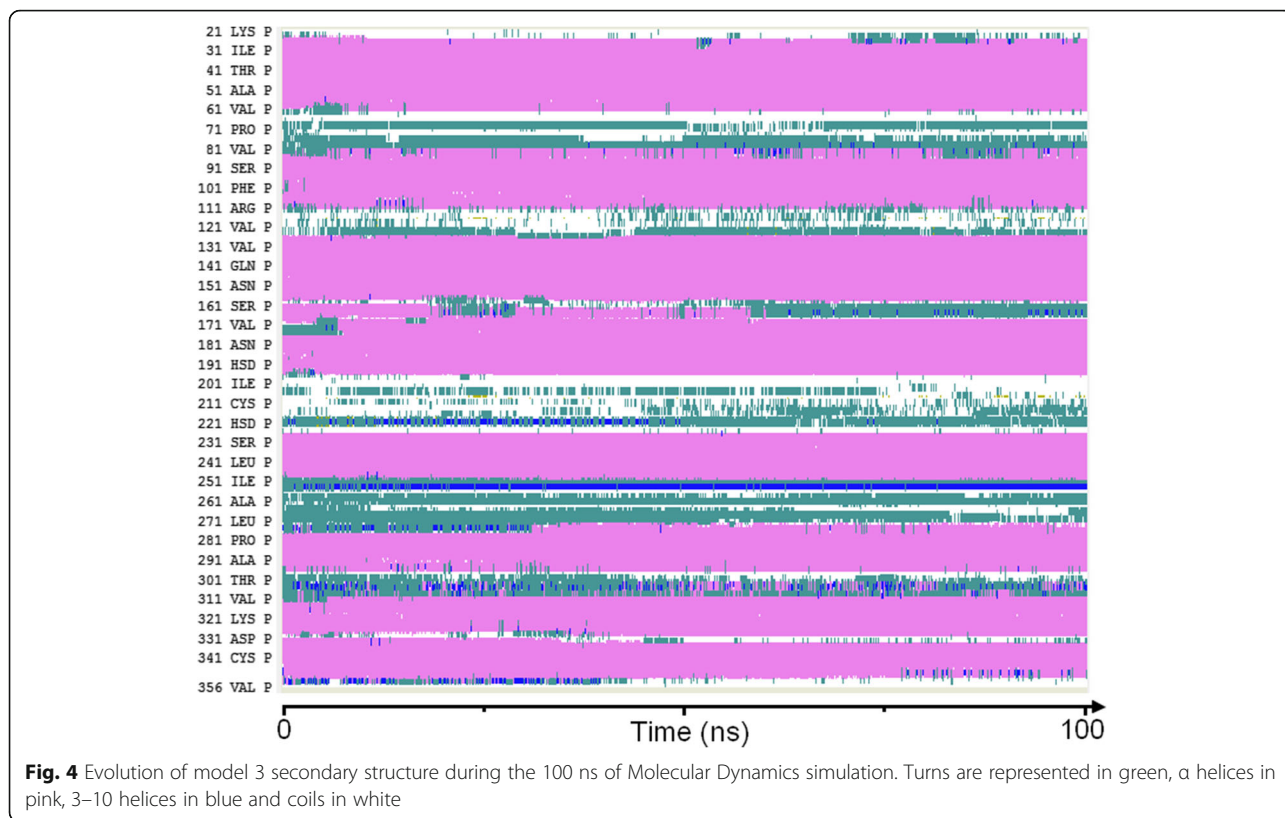
Conclusion

For this study we have used a bespoke analysis pipeline called GPCRpipe followed by a stepwise funnel strategy to identify, select and model one putative GPCR protein that could be of possible use as drug target to design new compounds active against *Fusarium graminearum*. This iterative search procedure is innovative because it has combined the use of genomics and molecular protein modeling approaches. Considering the 117 GPCRs candidates previously predicted in the FG genome sequencing [22] and nine sequences initially obtained by our strategy, we anticipated that a few false positives were retained. Therefore, we used additional filters based on structural and functional criteria to predict the most authentic GPCR candidates. Our selection of the nine resulting FG GPCR candidates was based not only on sequence similarity but on Molecular Dynamic modelling of the 3D protein structure. In addition, literature mining and a phylogeny analysis was used to explore the potential biological processes associated/proposed for each GPCR. For the final sequence retained, namely FGSG_02655, the molecular dynamics simulations proved to be an efficient method to choose between several alignments between the putative FG GPCR query and the template used in the homology modeling process as only one of the three predicted models came

through the extended 100 ns MD stimulation intact. Furthermore, a method developed to model GPCR was recently published [32] and may be interesting to use in future studies.

The receptor conformational flexibility highlighted by the MD simulation on the retained final robust model was used for running next an efficient ensemble docking structure-based virtual screening which provides interesting hypothetical hits to be now proposed for experiments. The whole *in silico* selection funnel presented in this study provide an example of an integrated process merging genomics, structural bioinformatics and drug design and leading to propose valuable and innovative solutions to a world-wide threat to grain producers and consumers.
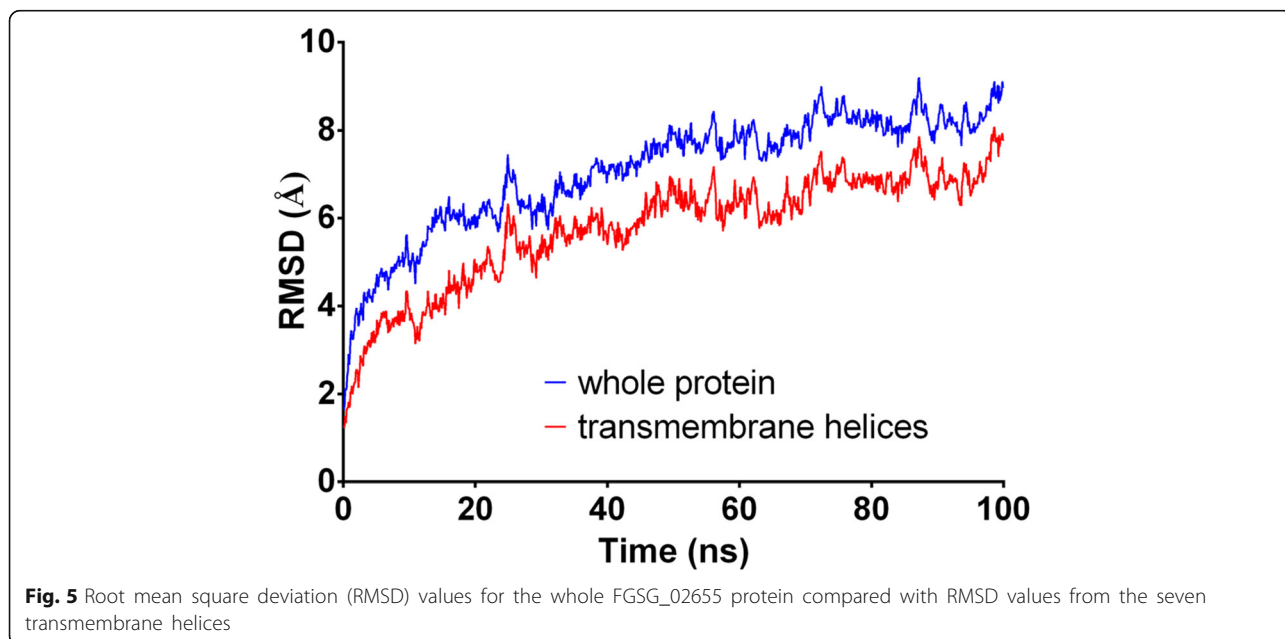
Methods

Our in silico experimental approach was composed of three main steps as illustrated in Fig. 9. The first step was related to the identification of GPCRs in the FG genome using several GPCRs prediction tools. Next, having identified putative GPCRs, these were classified according to their function to select a limited set of possible targets for designing new and innovative compounds against FG. The third step of this funnel consisted of a molecular modeling approach to building the three-dimensional models of

Fig. 4 Evolution of model 3 secondary structure during the 100 ns of Molecular Dynamics simulation. Turns are represented in green, α helices in pink, 3–10 helices in blue and coils in white

these targets. The last step corresponds the virtual screening. Binding pockets were detected in stable models identified in the previous step. Then, a large compound library was used with a docking program to find putative inhibitors. All calculations were performed on a 64 cores computer.

Step 1: GPCR identification in FG proteome.
The *Fusarium graminearum* genome was firstly published in 2007 [33]. The complete proteome of *Fusarium graminearum* PH-1 assembly FG3 (13,321 proteins) was downloaded from the BROAD Institute *Fusarium* comparative database [34].

Fig. 5 Root mean square deviation (RMSD) values for the whole FGSG_02655 protein compared with RMSD values from the seven transmembrane helices
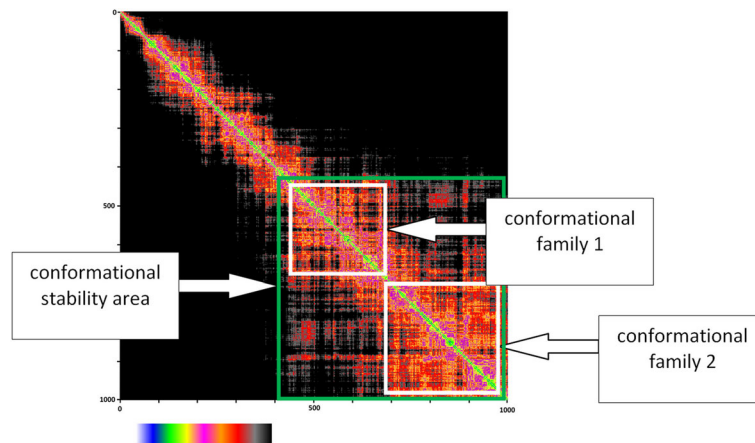
Fig. 6 Root mean square deviation (RMSD) map showing the conformational behavior of the protein during the 100 ns MD. The color scale is given below, the black color corresponding to dissimilar conformations (RMSD > 3.5 Å) and white corresponds to identical conformations (RMSD = 0 Å). The two conformational families 1 and 2 found as stable structural states are highlighted by the white squares
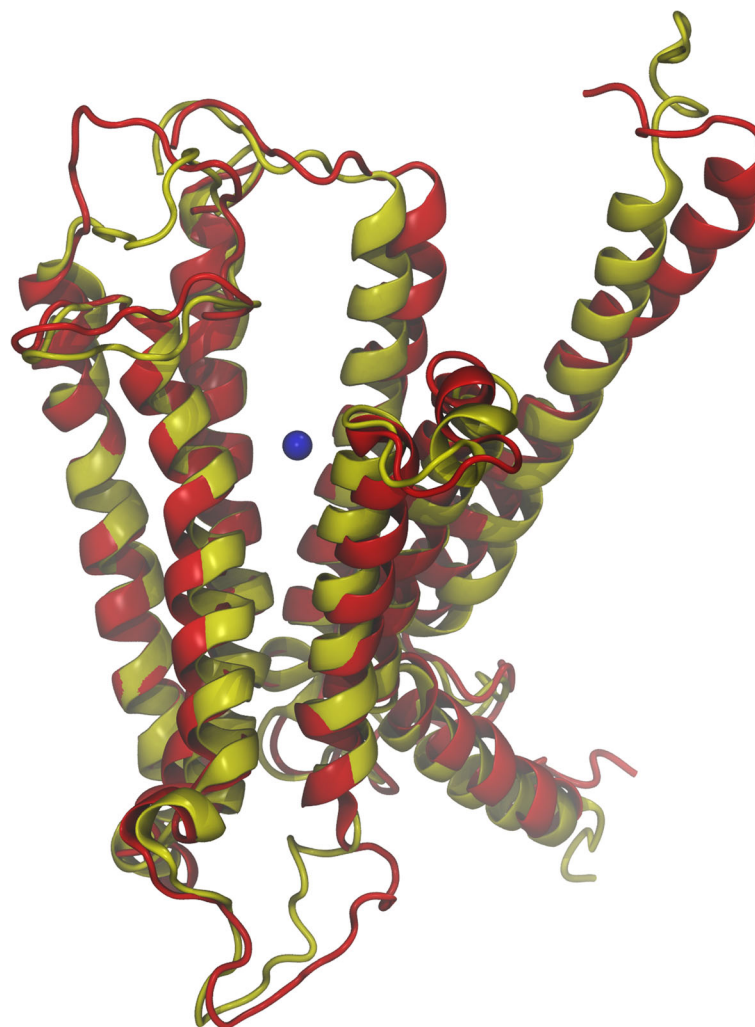

Fig. 7 Structural alignment of the conformational families. The first conformation is in yellow and the second in red. The center of the binding pocket is in blue

Table 6 GOLD scores for the finally selected compounds (in bold)

Rank	Score (PLP)	ID dataset	#Atoms outside	Frame	Toxicity
1	*110.31*	*F1044-0055*	*25*	*525*	
2	*109.35*	*F1044-0055*	*18*	*840*	
3	107.46	**F0514-4158**	10	525	
4	106.61	**F0514-3978**	4	525	
5	*106.59*	*F0514-5375*	*18*	*525*	
6	105.88	**F0514-4003**	8	525	
7	105.7	**F0617-0172**	16	840	
8	105.67	F0520-1906	0	840	+
9	105.66	**F3407-3991**	0	840	
10	105.57	**F0514-4846**	12	525	
11	105.5	**F0514-0510**	10	525	
12	105.35	**F0514-3894**	5	525	
13	105.27	**F0514-5342**	13	525	
14	105.24	**F0514-4074**	0	840	
15	105	F0514-3894	8	525	

Molecules in italics are not selected because of their high number of atoms outside the pocket

FG putative GPCRs were firstly detected using the GPCRpipe program [24]. Two methods in GPCRpipe were used for the detection of GPCR. Based on Hidden Markov Model, the first step was designed for the detection of GPCR. The second one is a library that consists of 39 Pfam profile HMMs (specific to different families of GPCR). We used the GPCRpipe "AND" method, meaning that a GPCR was predicted only if the two methods confirmed the prediction. This choice resulted in a reduced number of GPCRs predictions and limited the number of false positive predictions. The next step, to validate the GPCRpipe predictions, involved using three transmembrane prediction softwares namely HMMtop [35], TMHMM [36] and Phobius [37]. The first two are the best-known trans-membrane prediction methods, and Phobius was reported to perform comparably [37, 38].

Step 2: GPCR characterization, selection, and annotation. The functional classification of the GPCRs identified in *Fusarium graminearum* was realized using a PfamA analysis [39]. In parallel, the phylogenetic tree of these GPCRs and the 40 identified GPCRs in *Verticillium dahliae* and *Verticillium albo-atrum* [17] was built. This tree was built by Mega 6.0 using the Neighbor-Joining method and 10,000 repetitions.

Step 3: 3D Model Building.

Homology modeling

The construction and validation of the various homology models of GPCRs is still a challenge [40] and requires not only the sequence alignments but would also include structure-based alignments. This approach has been proved successful in many studies

[41, 42]. Nowadays, the structure of 20 different class A, two class B, two class C, and one frizzled GPCRs are available in the PDB [43], providing a reasonable set of possible templates to be used.

The choice of the proper template is crucial for ensuring the validity of the homology model. For that, several strategies are proposed, and many questions still remain [44]: for example, do we have to choose a single template, and in this case how to select it, or a set of templates? Such a decision can be difficult as contradictory results were obtained such as the ones claiming that a single well-chosen template is better than a set [45] versus the ones with opposite conclusions [46]. Moreover, it appears clearly from recent studies that the accuracy of the model greatly depends on the phylogenetic tree proximity of the template and the target [47]. Consequently, when considering (i) the remarks above, (ii) the conservation of the 7TM bundles in all GPCRs and the observed deformation of its helices [48], and (iii) the sequence conservation of several motifs [49] we decided to start our homology modeling phase using both phylogenetic data (for selecting the most suitable template) and helix predictions information (to align the TM helices sequences between the template and the target). Next the loops connecting the TM helices were added to the models obtained this way, considering also the constraint of the possible disulfide bridges [50, 51]. To build our models, transmembrane helix positions were determined by 11 transmembrane helix predictors (Additional file 2: Table S2 [23, 35, 52–59]). Then, our models were constructed by aligning transmembrane helices instead of similar amino acids. Moreover, crystal structures of several GPCRs, as well as experimental evidence, have shown the presence of a disulfide bond linking transmembrane helix 3 (TM3) to the second extracellular loop (ECL2) [60] and we used this additional information positioning this cysteine residue at the top of TM3. Finally, the position of the conserved motifs was also used for our selection process [61]. The homology modeling task was performed using the MODELLER program with its default settings [62]. Additionally, the automatic loop refinement method available in MODELLER was used. The DOPE score from MODELLER was used to estimate model quality.

Molecular dynamics (MD)

The next steps using Molecular Dynamics simulations were required to refine the preliminary crude models and then analysis the stability of the GPCR within the membrane [63]. MD is now commonly used to validate homology models, especially in the GPCRs field [64–67].
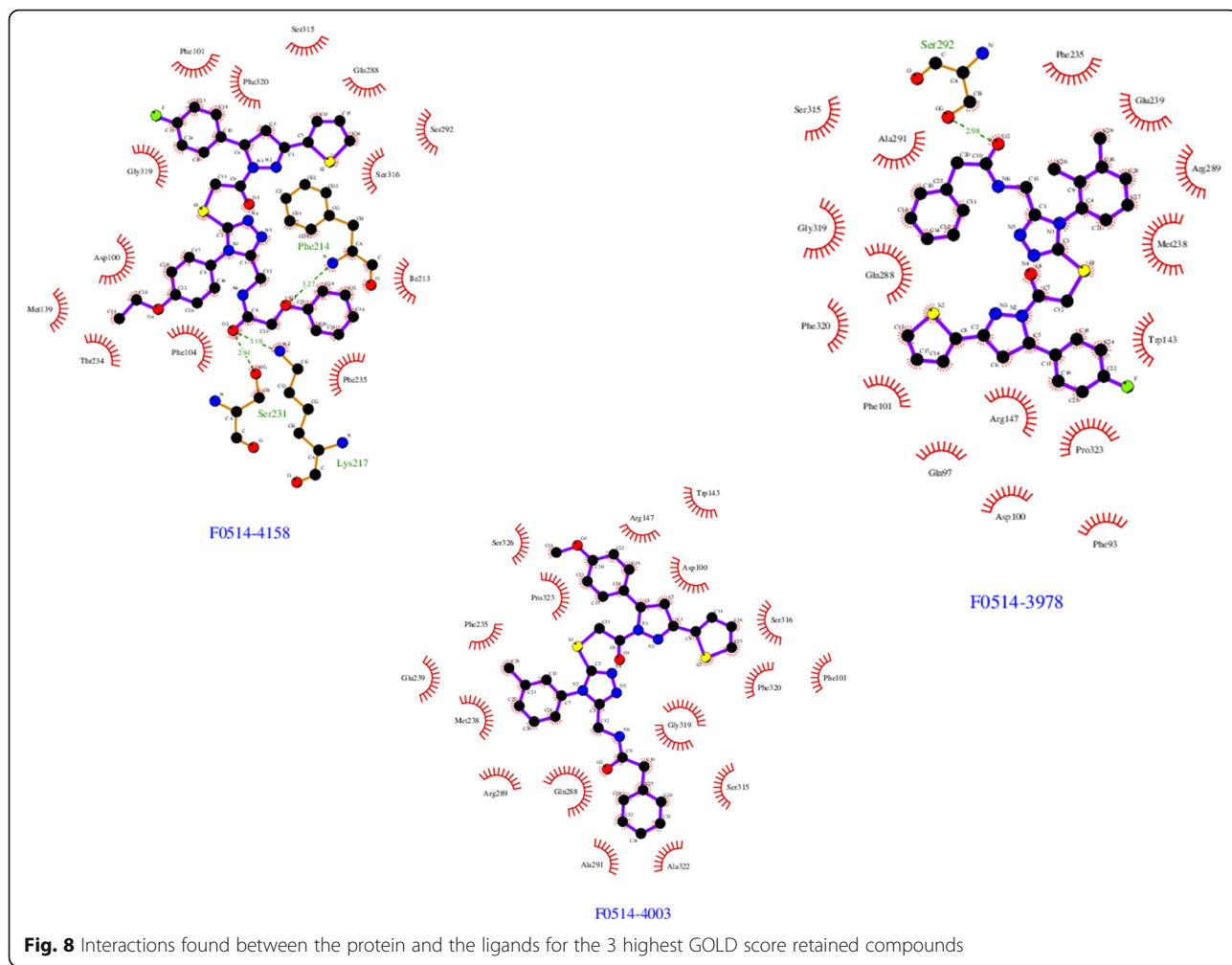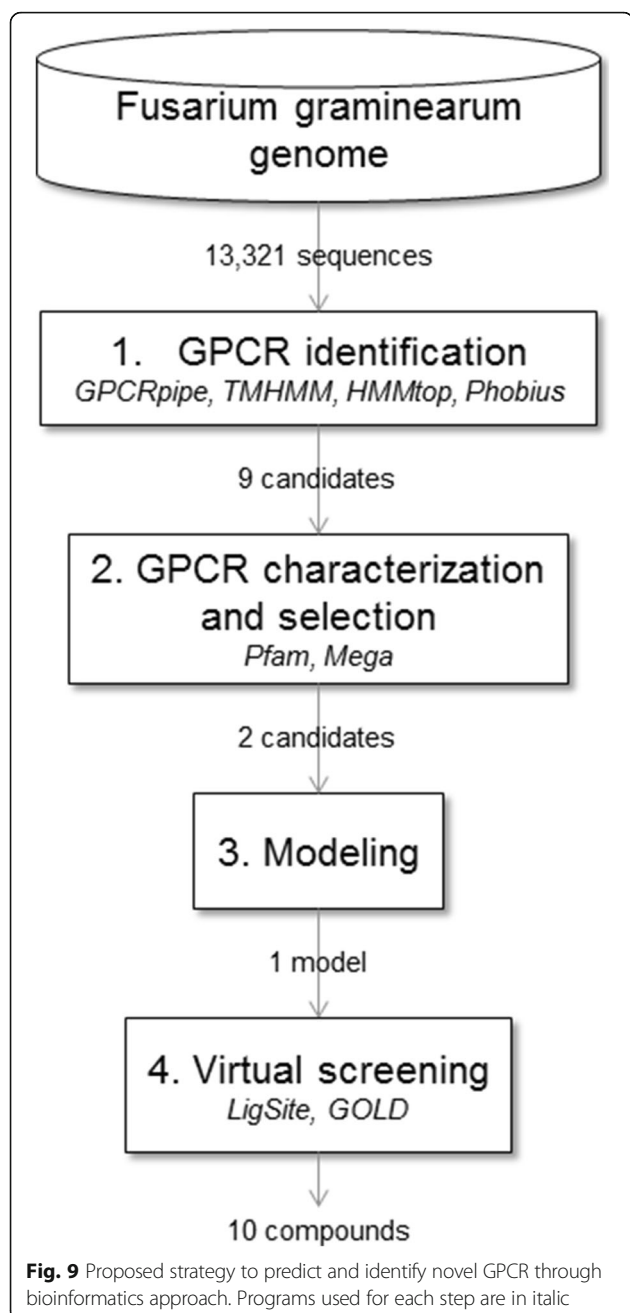
Fig. 8 Interactions found between the protein and the ligands for the 3 highest GOLD score retained compounds

For this purpose, we used molecular dynamics simulation on the receptor homology models that were embedded in a fully hydrated POPC bilayer [68]. No ligand was positioned within the receptor at this level as it has been shown [47] that the presence of a ligand does not change the accuracy of the structure produced. Initially, the receptors models were positioned across the equilibrated bilayer while seeking to match the hydrophobic protein segments with the layer formed by the lipid hydrocarbon tails. Lipids overlapping with the protein complex were deleted, leaving a bilayer consisting of 357 POPC molecules. To ascertain that the cytoplasmic and extracellular loops did not interact, an amount of 35,479 water molecules was added, as well as 10 counterions to make the whole system-neutral, thus making a total number of atoms equal to 159,461. The complete system, represented in Fig. 3 was replicated periodically in the three directions of space, with a repeat distance of ∼ 120 Å.

The MD simulations were carried out in the isobaric-isothermal ensemble, maintaining the pressure and the temperature of 1.0 atm and 300.0 K, respectively, using Langevin dynamics and the Langevin piston approach. The MD program NAMD [63] was employed in conjunction with the CHARMM27 force field [69] to describe the receptor, the lipid bilayer, and the water molecules. Coulomb forces were evaluated with the particle-mesh Ewald method. The equations of motion were integrated with a 1-fs time step, using the r-RESPA algorithm to update short- and long-range contributions at different frequencies. Each system was energy minimized and then equilibrated (3 ns) before recording trajectories. All MD trajectory frames were recorded at 1 ps intervals, for a total of 10 ns simulation. Model stabilities were then checked by analyzing secondary structure evolution during the MD simulation. If at least one transmembrane helix broke, then the model was not considered stable. For the stable models obtained, the

Fig. 9 Proposed strategy to predict and identify novel GPCR through bioinformatics approach. Programs used for each step are in italic

simulations were extended to reach a 100 ns simulation time.

MD simulation analysis

Once the MD simulation was completed, all the frames were aligned by only take into account the protein backbone. RMSDs were calculated by the "RMSD trajectory tool" plugin from Visual Molecular Dynamics (VMD) [70]. RMSD maps were built using a previously developed in-house Tool Command Language (TCL) script.

Step 4: Virtual screening

Ligand library

The chemical libraries used for the virtual screening were the GPCR Targeted Libraries (11,571 compounds) from Life chemicals. These libraries contained compounds for sixteen types of GPCRs.

Pocket detection and analysis

For each stable conformer detected thanks to the MD simulation, the coordinates of the binding pocket center were identified using LigSite [71].

Ensemble docking

For the docking, we used the docking program GOLD [72] which has been considered as one of the best docking software [73]. Because several stable conformations were identified, we used the ensemble docking possibility available in GOLD. The use of such conformational ensembles was considered as an improved strategy in structure-based docking calculations [74]. For each docking, 100 starting ligand conformers were used in GOLD. All target conformers used were aligned in a common reference system and the center of the pocket cavity is an average of the individual centers found in each conformation. A sphere of 15 Å was selected to define the binding region around this center.

Toxicity

In order to remove probable toxic molecules, the side effects of the finally identified compounds were detected using toxicity predictors such as PAINS-remover, Badapple and Protox webservers [75–77].

Additional files

Additional file 1: Table S1. Name and 2D structure of the 10 retained compounds (PDF 108 kb)

Additional file 2: Table S2. Programs used for detecting transmembrane helix positions (PDF 34 kb)

Abbreviations

3D: Tri-dimensional; DOPE: Discrete optimized protein energy; FG: *Fusarium graminearum*; FHB: *Fusarium* head blight; GPCR: G-protein coupled receptor; MD: Molecular dynamics; PDB: Protein databank; RMSD: Root mean square deviation; TCL: Tool command language; TM: Transmembrane; VMD: Visual molecular dynamics

Availability of data and materials
The datasets during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions
EB carried out the GPCR identification, modeling, virtual screening and drafted the manuscript. NFM drafted the manuscript, supervised the project. BM drafted the manuscript, molecular dynamics, implemented virtual screening and designed the strategy and conceived the study. RT performed database management and integration of bioinformatics tools MU performed complementary bioinformatics analysis. KHK conceived of the study. All authors reviewed and approved the final manuscript.

Competing interests
The authors declare no competing financial interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Author details
¹EMBRAPA Genetic Resources and Biotechnology, Brasília, DF 70770-917, Brazil. ²Department of Plant Biology and Crop Science, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK. ³CAPSID Team, LORIA, UMR 7503, CNRS, Lorraine University, Vandœuvre-lès-Nancy 54506, France.

Published: 15 December 2016

References
1. Goswami RS, Kistler HC. Heading for disaster: Fusarium graminearum on cereal crops. Mol Plant Pathol. 2004;5(6):515–25.
2. Escrivá L, Font G, Manyes L. In vivo toxicity studies of fusarium mycotoxins in the last decade: a review. Food Chem Toxicol. 2015;78:185–206.
3. Döll S, Dänicke S. The Fusarium toxins deoxynivalenol (DON) and zearalenone (ZON) in animal feeding. Prev Vet Med. 2011;102:132–45.
4. Pinton P, Nougayrède JP, Del Rio JC, Moreno C, Marin DE, Ferrier L, et al. The food contaminant deoxynivalenol, decreases intestinal barrier permeability and reduces claudin expression. Toxicol Appl Pharmacol. 2009;237:41–8.
5. Yuen GY, Schoneweis SD. Strategies for managing Fusarium head blight and deoxynivalenol accumulation in wheat. Int J Food Microbiol. 2007;119: 126–30.
6. Lucas JA, Hawkins NJ, Fraaije BA. The evolution of fungicide resistance. Adv Appl Microbiol. 2015;90:29–92.
7. Tripathi H, Luqman S, Meena A, Khan F. Genomic identification of potential targets unique to Candida albicans for the discovery of antifungal agents. Curr Drug Targets. 2014;15:136–49.
8. Al Abdallah Q, Fortwendel JR. Exploration of Aspergillus fumigatus Ras pathways for novel antifungal drug targets. Front Microbiol. 2015;6:128.
9. Teixeira A. G Protein-Coupled Receptors: Structure, Signaling, and Physiology. The Yale Journal of Biology and Medicine. 2011;84(3):326-327.
10. Reiter E, Lefkowitz RJ. GRKs and β-arrestins: roles in receptor silencing, trafficking and signaling. Trends Endocrinol Metab. 2006;17(4):159–65.
11. Xue C, Hsueh YP, Heitman J. Magnificent seven: Roles of G protein-coupled receptors in extracellular sensing in fungi. FEMS Microbiol Rev. 2008;32: 1010–32.
12. Van Dijck P. Nutrient sensing G protein-coupled receptors: interesting targets for antifungals? Med Mycol. 2009;47:671–80.
13. Lafon A, Han K-HH, Seo J-AA, Yu J-HH, D'Enfert C. G-protein and cAMP-mediated signaling in aspergilli: A genomic perspective. Fungal Genet Biol. 2006;43:490–502.
14. Kulkarni RD, Thon MR, Pan H, Dean RA. Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus Magnaporthe grisea. Genome Biol. 2005;6:R24.
15. Xue C, Bahn Y-S, Cox GM, Heitman J. G protein-coupled receptor Gpr4 senses amino acids and activates the cAMP-PKA pathway in Cryptococcus neoformans. Mol Biol Cell. 2006;17:667–79.
16. Li L, Wright SJ, Krystofova S, Park G, Borkovich KA. Heterotrimeric G protein signaling in filamentous fungi. Annu Rev Microbiol. 2007;61:423–52.
17. Zheng H, Zhou L, Dou T, Han X, Cai Y, Zhan X, et al. Genome-wide prediction of G protein-coupled receptors in Verticillium spp. Fungal Biol. 2010;114:359–68.
18. Gruber S, Omann M, Zeilinger S. Comparative analysis of the repertoire of G protein-coupled receptors of three species of the fungal genus Trichoderma. BMC Microbiol. 2013;13:108.
19. Affeldt KJ, Carrig J, Amare M, Keller NP. Global survey of canonical Aspergillus flavus G protein-coupled receptors. MBio. 2014;5:e01501–14.
20. Hoffman CS. Except in every detail: Comparing and contrasting G-protein signaling in Saccharomyces cerevisiae and Schizosaccharomyces pombe. Eukaryot Cell. 2005;4:495–503.
21. Ladds G, Goddard A, Davey J. Functional analysis of heterologous GPCR signalling pathways in yeast. Trends Biotechnol. 2005;23(7):367–73.
22. Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature. 2010;464:367–73.
23. Wistrand M, Käll L, Sonnhammer ELL. A general model of G protein-coupled receptor sequences and its application to detect remote homologs. Protein Sci. 2006;15:509–21.
24. Theodoropoulou M, Tsaousis G, Litou Z, Bagos P, Hamodrakas S: GPCRpipe: a pipeline for the detection of G-protein coupled receptors in proteomes. 2013.
25. Abadio AKR, Kioshima ES, Teixeira MM, Martins NF, Maigret B, Felipe MSS. Comparative genomics allowed the identification of drug targets against human fungal pathogens. BMC Genomics. 2011;12:75.
26. Beautrait A, Leroux V, Chavent M, Ghemtio L, Devignes M-DD, Smaïl-Tabbone M, et al. Multiple-Step Virtual Screening Using VSM-G: Overview and Validation of Fast Geometrical Matching Enrichment. J Mol Model. 2008;14:135–48.
27. Trail F. For blighted waves of grain: Fusarium graminearum in the postgenomics era. Plant Physiol. 2009;149(January):103–10.
28. Lee J, Leslie JF, Bowden RL. Expression and function of sex pheromones and receptors in the homothallic ascomycete Gibberella zeae. Eukaryot Cell. 2008;7:1211–21.
29. Kim HK, Lee T, Yun SH. A putative pheromone signaling pathway is dispensable for self-fertility in the homothallic ascomycete Gibberella zeae. Fungal Genet Biol. 2008;45:1188–96.
30. Van Nguyen T. Signal transduction pathways in the fungal wheat pathogen Fusarium graminearum. 2013.
31. Venkatakrishnan a J, Deupi X, Lebon G, Tate CG, Schertler GF, Babu MM. Molecular signatures of G-protein-coupled receptors. Nature. 2013;494:185–94.
32. Zhang J, Yang J, Jang R, Zhang Y. GPCR-I-TASSER: A Hybrid Approach to G Protein-Coupled Receptor Structure Modeling and the Application to the Human Genome. Structure. 2015;23:1538–49.
33. Cuomo CA, Güldener U, Xu J-R, Trail F, Turgeon BG, Di Pietro A, et al. The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization. Science. 2007;317:1400–2.
34. Fusarium Comparative Sequencing Project, Broad Institute of Harvard and MIT [www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative/fusarium-comparativegenome-project]. Accessed Jan 2015.
35. Tusnády GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics. 2001;17:849–50.
36. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567–80.
37. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. Nucleic Acids Res. 2007;35:W429–32.
38. Cuthbertson JM, Doyle DA, Sansom MSP. Transmembrane helix prediction: a comparative evaluation and analysis. Protein Eng Des Sel. 2005;18:295–308.
39. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: The protein families database. Nucleic Acids Res. 2014;42:D222–30.
40. Kufareva I, Katritch V, Stevens RC, Abagyan R. Advances in GPCR modeling evaluated by the GPCR dock 2013 assessment: Meeting new challenges. Structure. 2014;22:1120–39.
41. Du H, Brender JR, Zhang J, Zhang Y. Protein structure prediction provides comparable performance to crystallographic structures in docking-based virtual screening. Methods. 2015;71:77–84.

42. Tang H, Wang XS, Hsieh JH, Tropsha A. Do crystal structures obviate the need for theoretical models of GPCRs for structure-based virtual screening? Proteins Struct Funct Bioinf. 2012;80:1503–21.

43. Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, et al. The Protein Data Bank and the challenge of structural genomics. Nat Struct Biol. 2000;7(Suppl):957–9.

44. Yarnitzky T, Levit A, Niv MY. Homology modeling of G-protein-coupled receptors with X-ray structures on the rise. Curr Opin Drug Discov Devel. 2010;13:317–25.

45. Bera I, Laskar A, Ghoshal N. Exploring the structure of opioid receptors with homology modeling based on single and multiple templates and subsequent docking: A comparative study. J Mol Model. 2011;17(5):1207–21.

46. Sokkar P, Mohandass S, Ramachandran M. Multiple templates-based homology modeling enhances structure quality of AT1 receptor: Validation by molecular dynamics and antagonist docking. J Mol Model. 2011;17:1565–77.

47. Lupala CS, Rasaeifar B, Gomez-Gutierrez P, Perez JJ. 193 Effect of template selection on the construction of atomistic models of GPCRs by homology modeling. J Biomol Struct Dyn. 2015;33 Suppl 1:127–8.

48. van der Kant R, Vriend G. Alpha-bulges in G protein-coupled receptors. Int J Mol Sci. 2014;15:7841–64.

49. Nagarathnam B, Kannan S, Dharnidharka V, Balakrishnan V, Archunan G, Sowdhamini R. Insights from the analysis of conserved motifs and permitted amino acid exchanges in the human, the fly and the worm GPCR clusters. Bioinformation. 2011;7:15–20.

50. Olivella M, Caltabiano G, Cordomí A. The role of Cysteine 6.47 in class A GPCRs. BMC Struct Biol. 2013;13:3.

51. Wang C-IA, Lewis RJ. Emerging opportunities for allosteric modulation of G-protein coupled receptors. Biochem Pharmacol. 2013;85:153–62.

52. Cserzö M, Wallin E, Simon I, von Heijne G, Elofsson A. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. Protein Eng. 1997;10:673–6.

53. Cao B, Porollo A, Adamczak R, Jarrell M, Meller J. Enhanced recognition of protein transmembrane domains with prediction-based structural profiles. Bioinformatics. 2006;22:303–9.

54. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, et al. PredictProtein-an open resource for online prediction of protein structural and functional features. Nucleic Acids Res. 2014;42:W337–43.

55. Pasquier C, Promponas VJ, Palaios GA, Hamodrakas JS, Hamodrakas SJ. A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the SwissProt database: the PRED-TMR algorithm. Protein Eng. 1999;12:381–5.

56. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics. 1998;14:378–9.

57. Juretić D, Zoranić L, Zucić D. Basic charge clusters and predictions of membrane protein topology. J Chem Inf Comput Sci. 2002;42:620–32.

58. Hofmann K, Stoffel W. TMbase-A database of membrane spanning proteins segments. Biol Chem Hoppe Seyler. 1993;374:166.

59. Bernsel A, Viklund H, Hennerdal A, Elofsson A. TOPCONS: Consensus prediction of membrane protein topology. Nucleic Acids Res. 2009;37:W465–8.

60. Mann RJ, Al-Sabah S, de Maturana RL, Sinfield JK, Donnelly D. Functional coupling of Cys-226 and Cys-296 in the glucagon-like peptide-1 (GLP-1) receptor indicates a disulfide bond that is close to the activation pocket. Peptides. 2010;31:2289–93.

61. Nagarathnam B, Sankar K, Dharnidharka V, Balakrishnan V, Archunan G, Sowdhamini R. TM-MOTIF: an alignment viewer to annotate predicted transmembrane helices and conserved motifs in aligned set of sequences. Bioinformation. 2011;7(5):214–21.

62. Eswar N, Eramian D, Webb B, Shen M-Y, Sali A. Protein structure modeling with MODELLER. Methods Mol Biol. 2008;426:145–59.

63. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem. 2005;26:1781–802.

64. Fano A, Ritchie DW, Carrieri A. Modeling the structural basis of human CCR5 chemokine receptor function: from homology model building and molecular dynamics validation to agonist and antagonist docking. J Chem Inf Model. 2006;46:1223–35.

65. Singh R, Sobhia ME. Structure prediction and molecular dynamics simulations of a G-protein coupled receptor: human CCR2 receptor. J Biomol Struct Dyn. 2013;31:694–715.

66. Maganti L, Open Source Drug Discovery Consortium, Ghoshal N. Probing the structure of Mycobacterium tuberculosis MbtA: model validation using molecular dynamics simulations and docking studies. J Biomol Struct Dyn. 2014;32:273–88.

67. Tautermann CS, Seeliger D, Kriegl JM. What can we learn from molecular dynamics simulations for GPCR drug design? Comput Struct Biotechnol J. 2015;13:111–21.

68. Benedetto A, Bingham RJ, Ballone P. Structure and dynamics of POPC bilayers in water solutions of room temperature ionic liquids. J Chem Phys. 2015;142(12):124706.

69. Mackerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998;102:3586–616.

70. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph. 1996;14:33–8.

71. Huang B, Schroeder M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. BMC Struct Biol. 2006;6:19.

72. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins Struct Funct Genet. 2003; 52:609–23.

73. Liebeschuetz JW, Cole JC, Korb O. Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. J Comput Aided Mol Des. 2012;26:737–48.

74. Korb O, Olsson TSG, Bowden SJ, Hall RJ, Verdonk ML, Liebeschuetz JW, et al. Potential and limitations of ensemble docking. J Chem Inf Model. 2012;52: 1262–74.

75. Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem. 2010;53:2719–40.

76. Yang J. The BADAPPLE promiscuity plugin for BARD Evidence-based promiscuity scores. 2013:8–12.

77. Drwal MN, Banerjee P, Dunkel M, Wettig MR, Preissner R. ProTox: A web server for the in silico prediction of rodent oral toxicity. Nucleic Acids Res. 2014;42:W53–8.