

Rothamsted Repository Download

A - Papers appearing in refereed journals

Bresso, E., Leroux, V., Urban, M., Hammond-Kosack, K. E., Maigret, B. and Martins, N. F. 2016. Structure-based virtual screening of hypothetical inhibitors of the enzyme longiborneal synthase - a potential target to reduce Fusarium head blight disease. *Journal Of Molecular Modeling*. 22, p. 163.

The publisher's version can be accessed at:

- <https://dx.doi.org/10.1007/s00894-016-3021-1>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/8v246/structure-based-virtual-screening-of-hypothetical-inhibitors-of-the-enzyme-longiborneal-synthase-a-potential-target-to-reduce-fusarium-head-blight-disease>.

© 21 June 2016, Springer.

Structure-based virtual screening of hypothetical inhibitors of the enzyme longiborneol synthase—a potential target to reduce *Fusarium* head blight disease

E. Bresso¹ · V. Leroux² · M. Urban³ · K. E. Hammond-Kosack³ · B. Maigret² · N. F. Martins¹

Received: 17 December 2015 / Accepted: 27 May 2016 / Published online: 21 June 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract *Fusarium* head blight (FHB) is one of the most destructive diseases of wheat and other cereals worldwide. During infection, the *Fusarium* fungi produce mycotoxins that represent a high risk to human and animal health. Developing small-molecule inhibitors to specifically reduce mycotoxin levels would be highly beneficial since current treatments unspecifically target the *Fusarium* pathogen. Culmorin possesses a well-known important synergistically virulence role among mycotoxins, and longiborneol synthase appears to be a key enzyme for its synthesis, thus making longiborneol synthase a particularly interesting target. This study aims to discover potent and less toxic agrochemicals against FHB. These compounds would hamper culmorin synthesis by inhibiting longiborneol synthase. In order to select starting molecules for further investigation, we have conducted a structure-based virtual screening investigation. A longiborneol synthase structural model is first built using homology modeling, followed by molecular dynamics simulations that provided the required input for a protein–ligand ensemble docking procedure. From this strategy, the three most interesting compounds (hits) were selected among the 25 top-ranked docked

compounds from a library of 15,000 drug-like compounds. These putative inhibitors of longiborneol synthase provide a sound starting point for further studies involving molecular modeling coupled to biochemical experiments. This process could eventually lead to the development of novel approaches to reduce mycotoxin contamination in harvested grain.

Keywords *Fusarium* mycotoxins · Culmorin · Inhibitors · Homology modeling · Molecular dynamics · Ensemble docking

Introduction

Fusarium head blight (FHB), caused by *Fusarium graminearum* and 16 other *Fusarium* species, is one of the most important wheat diseases in the world [1, 2]. Wheat infection during flowering (anthesis) results in grain contamination by mycotoxins. The latter is a well-known phenomenon observed before or during silage storage, reducing quality and yield while constituting an obvious risk factor for human and animal health [3]. Current management of this problem relies on multiple fungicide applications in combination with semi-resistant cultivars. This increases production costs, further fuels public concerns over abuse of chemicals such as pesticides in agriculture, and raises the chances that pathogen resistance will eventually occur and spread [4, 5]. The discovery of new compounds that could block mycotoxin production with exclusivity and no environmental hazard would certainly provide a welcome alternative to control the pathogen.

Among the mycotoxins produced by *Fusarium* [6], culmorin has been found in relatively high level in contaminated wheat grains [7, 8]. Several studies on contaminated grains established a clear link between culmorin (and various hydroxy-culmorin) levels and the trichothecene

Electronic supplementary material The online version of this article (doi:10.1007/s00894-016-3021-1) contains supplementary material, which is available to authorized users.

✉ N. F. Martins
natalia.martins@embrapa.br

¹ EMBRAPA Genetic Resources and Biotechnology Parque Estação Biológica - PqEB, Av. W5 Norte, Brasília, DF, Brazil

² CNRS, LORIA, Campus Scientifique, BP 239, 54506 Vandœuvre-lès-Nancy, France

³ Department of Plant Biology and Crop Science, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK

deoxynivalenol (DON) mycotoxin [9, 10]. Moreover, a growing body of evidence indicates that culmorin actually enhances DON toxicity [10, 11]. Culmorin is produced by the biotransformation of farnesyl diphosphate through a complex pathway. A key element of this pathway is longiborneol synthase, which produces longiborneol (Fig. 1a), which differs from culmorin by the lack of a single hydroxyl group (Fig. 1b), but was also recently identified as being required to complete culmorin biosynthesis in possible association with an oxygenase [12].

This work describes initial efforts to identify potential longiborneol synthase inhibitors, aiming to develop a new suite of agrochemicals for crop protection with reduced risks for human and animal toxic contamination. For this purpose, virtual screening represents a valuable *in silico* approach as it allows screening large chemical libraries in order to detect a limited number of compounds applicable for experimental testing, at a sensible cost [13, 14].

Since no experimentally derived 3D structure exists for longiborneol synthase, we first built a stable homology model that was further subjected to molecular dynamics (MD) simulations. This resulted in the selection of a conformational ensemble encompassing longiborneol synthase flexibility in its most stable structural state. This data was then used to perform ligand-to-protein docking calculations [15]. The present study presents the docking simulation for 15,000 compounds that allowed the selection of three putative longiborneol synthase inhibitors to be submitted to experimental validation.

Materials and methods

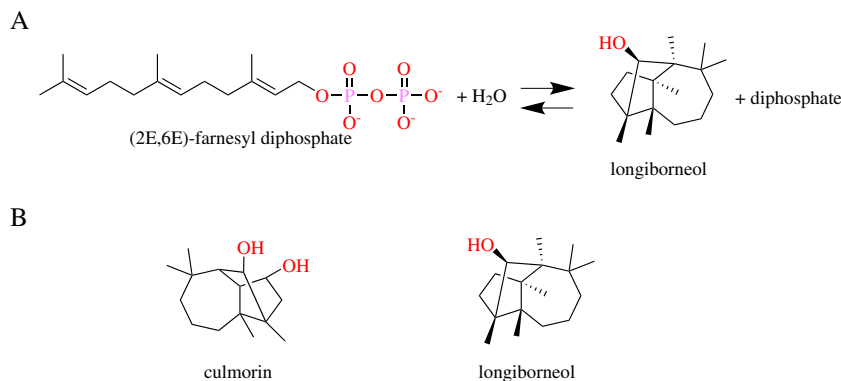
Homology modeling

The *F. graminearum* longiborneol synthase coded by the locus FGSG_10397 (ACY69978.1) [12] is described in the Uniprot database as being encoded by the *CML1* gene (see <http://www.uniprot.org/uniprot/D1M8S2>) and belongs to the fungal sesquiterpene synthase superfamily of enzymes. To identify

possible templates for building a three dimensional (3D) model of this protein, it was first necessary to search protein databases in order to identify proteins that could be suitable templates for obtaining a robust and convincing homology 3D model. Our selection was based on close relatednesses in both structure and function within all members of the family. We focused mostly on the isoprenoid domain characterized by a well-conserved structural organization that is mostly α -helical with a core bundle of antiparallel α -helices. The superposition of 33 structures of this family shows clearly the conservation of this structural organization within the whole family (<http://www.cathdb.info/version/latest/superfamily/1.10.600.10>). We checked carefully that this structural signature had been conserved during our homology modeling process.

Once the protein class was identified and the possible orthologs detected, we searched for similar structures at the Protein Data Bank (PDB), [16] to look for the most suitable template for homology modeling. This step is crucial as the quality of any homology model depends strongly on the percentage of similarity, identity and sequence coverage. When low sequence identities are detected, as in this work (less than 20 %), the modeling task is not trivial and requires a suitable strategy to improve the quality of the proposed models. The use of multiple templates has been shown to generally increase model quality over a single template [17]. Despite recent improvements [18], this approach still needs to use rules to combine information from all the templates. In our case, we chose to use several well-known homology web servers and compared their results. Hence, a comparative dendrogram showing the relationship between the proposed templates and our target protein was established using the phylogeny.fr web server [19, 20], in which MUSCLE 3.8.31 was used for multiple sequence alignment with default parameters (find diagonals option disabled, a maximum number of iterations: 16, no duration limitation and no more than 200 sequences) [21]. To draw the dendrogram, based on the protein alignment, we used the phylogeny.fr web server set PhyML3.1/3.0 aLRT with default parameters (Substitution model: WAG for proteins, aLRT test: SH-like, number of substitution rate categories: 4, gamma parameter: estimated, proportion of invariable

Fig. 1 a Longiborneol biosynthesis pathway. b Longiborneol and culmorin chemical structures



sites: estimated and the transition/transversion ratio: 4; the tree branch support was evaluated by an approximate likelihood-ratio test) [22].

The closest template was retained to perform the homology modeling task using the MODELLER program with its default settings [23] (see the MODELLER manual and tutorial). Additionally, the automatic loop refinement method available in MODELLER was used. The crude longiborneol synthase 3D₀ model was achieved via this route. The model was validated by the server SAVES [24].

Molecular dynamics

Starting from the homology built 3D₀ model obtained from the previous step, we refined the model and investigated its behavior in a physiological medium. The modeled protein was

embedded in a box of 100 Å × 100 Å × 100 Å with TIP3P explicit water molecules [25]. Next, Na⁺ ions were added to ensure electrostatic neutrality. The NAMD program version 2.6 was employed in conjunction with the CHARMM27 force field [26, 27] in order to simulate the ensemble of the 94,678-atom system. The initial state for dynamics was generated from the model after 64,000 steps of conjugate gradient minimization. The conformational behavior of the protein, water and counter ions system was obtained by running 100 ns of MD. The simulations were carried out in the isobaric-isothermal ensemble, maintaining pressure and temperature at 1 atm and 300 K, respectively, using Langevin dynamics (damping parameter of 1 ps⁻¹) and piston approaches. The shake algorithm was used during the simulation. The equations of motion were integrated with a 1 fs time step, using the r-RESPA algorithm [28] electrostatic forces at a slower 2 fs

Table 1 Protein databank (PDB) templates proposed by homology servers for longiborneol synthase [50–56]

Server	PDB ID	Classification	Name	Organism
Phyre2	1K1Y	Transferase	4-Alpha-glucanotransferase	<i>Thermococcus litoralis</i>
	4OKM	Transferase	Selinadiene synthase	<i>Streptomyces pristinaespiralis</i>
	3LG5	Lyase	Epi-isozizaene synthase	<i>Streptomyces coelicolor</i>
	4ZQ8	Transferase	Terpene synthase	<i>Streptomyces lydicus</i>
	3V1V	Lyase	2-Methylisoborneol synthase	<i>Streptomyces coelicolor</i>
	1PS1	Synthase	Pentalene synthase	<i>Streptomyces exfoliatus</i>
SWISSModel	1JFA	Lyase	Trichodiene synthase	<i>Fusarium sporotrichioides</i>
	2Q9Y	Lyase	Trichodiene synthase	<i>F. sporotrichioides</i>
	1N1Z	Isomerase	(+)-Bornyl diphosphate synthase	<i>Salvia officinalis</i>
ROBETTA	3 KB9	Lyase	Epi-isozizaene synthase	<i>S. coelicolor</i>
i-Tasser	1JFA	Lyase	Trichodiene synthase	<i>F. sporotrichioides</i>
	1YYQ	Lyase	Y305F trichodiene synthase	<i>F. sporotrichioides</i>
	1YJ4	Lyase	Trichodiene synthase	<i>F. sporotrichioides</i>
	1DI1	Lyase	Aristolochene synthase	<i>Penicillium roqueforti</i>
Falcon	1PS1	Synthase	Pentalene synthase	<i>S. exfoliatus</i>
	3IOP	Transferase	PDK-1	<i>Homo sapiens</i>
	3 KB9	Lyase	Epi-isozizaene synthase	<i>S. coelicolor</i>
	3RUX	Ligase	Biotin-protein ligase BirA	<i>Mycobacterium tuberculosis</i>
	4MC3	Lyase	Hedycaryol synthase	<i>Kitasatospora setae</i>
Only belonging to the same family	1YYQ	Lyase	Y305F trichodiene synthase	<i>F. sporotrichioides</i>
	3 KB9	Lyase	Epi-isozizaene synthase	<i>S. coelicolor</i>
	4OKZ	Transferase	Selinadiene synthase	<i>Streptomyces pristinaespiralis</i>
	1PS1	Synthase	Pentalene synthase	<i>S. exfoliatus</i>
	5DW7	Lyase	Geosmin synthase	<i>S. coelicolor</i>
	4ZQ8	Transferase	Terpene synthase	<i>S. lydicus</i>
	4KWD	Lyase	Aristolochene synthase	<i>Aspergillus terreus</i>
	3V1V	Lyase	2-Methylisoborneol synthase	<i>S. coelicolor</i>
	1DI1	Lyase	Aristolochene synthase	<i>Penicillium roqueforti</i>
	1JFA	Lyase	Trichodiene synthase	<i>F. sporotrichioides</i>
RaptorX	1JFA	Lyase	Trichodiene synthase	<i>F. sporotrichioides</i>
Our proposal	1JFA	Lyase	Trichodiene synthase	<i>F. sporotrichioides</i>

frequency. Long-range interactions were treated using the particle mesh Ewald approach [29], with an 11 Å cut-off. The calculation of forces and motion equations was repeated to generate a trajectory corresponding to a simulation time of 100 ns in which a conformation was recorded every 1 ps, generating a trajectory of 100,000 conformations. The conservation of the secondary structure elements along the MD trajectories was checked using the Timeline plugin in VMD [30] (data not shown).

Virtual screening

We used the VMS-G software platform [31] to perform virtual screening (VS). The docking program used within the platform was GOLD [32], which has been recognized as among the best docking software [33]. The ensemble docking option was selected in GOLD, in order to perform the calculations with all longiborneol synthase stable conformers detected via MD simulation. The use of such conformational ensembles

was considered as an improved strategy in structure-based docking procedures [15].

The choice of the chemical compounds to be screened is also critical. To ensure the VS campaign, two strategies can be applied: “brute force”, using millions of compounds, or “data-driven” using a chosen limited set. The first strategy requires high computing power but, nevertheless, the bottleneck is the analysis of the enormous number of outputs. The second approach performs a molecular mining with all the possible chemical information available around the biological target and its neighbors. This strategy results in an optimized reduced collection of compounds that will minimize false positives while maximizing candidate hit rate. We used this latter approach by searching the literature and patents for all compounds already proposed as inhibitors of the synthase superfamily. We clustered the compounds using ChemMine [34] in order to extract the most representative compounds and to analyze their chemical diversity. Meanwhile, we analyzed the diversity of the chemical libraries [35] proposed by several

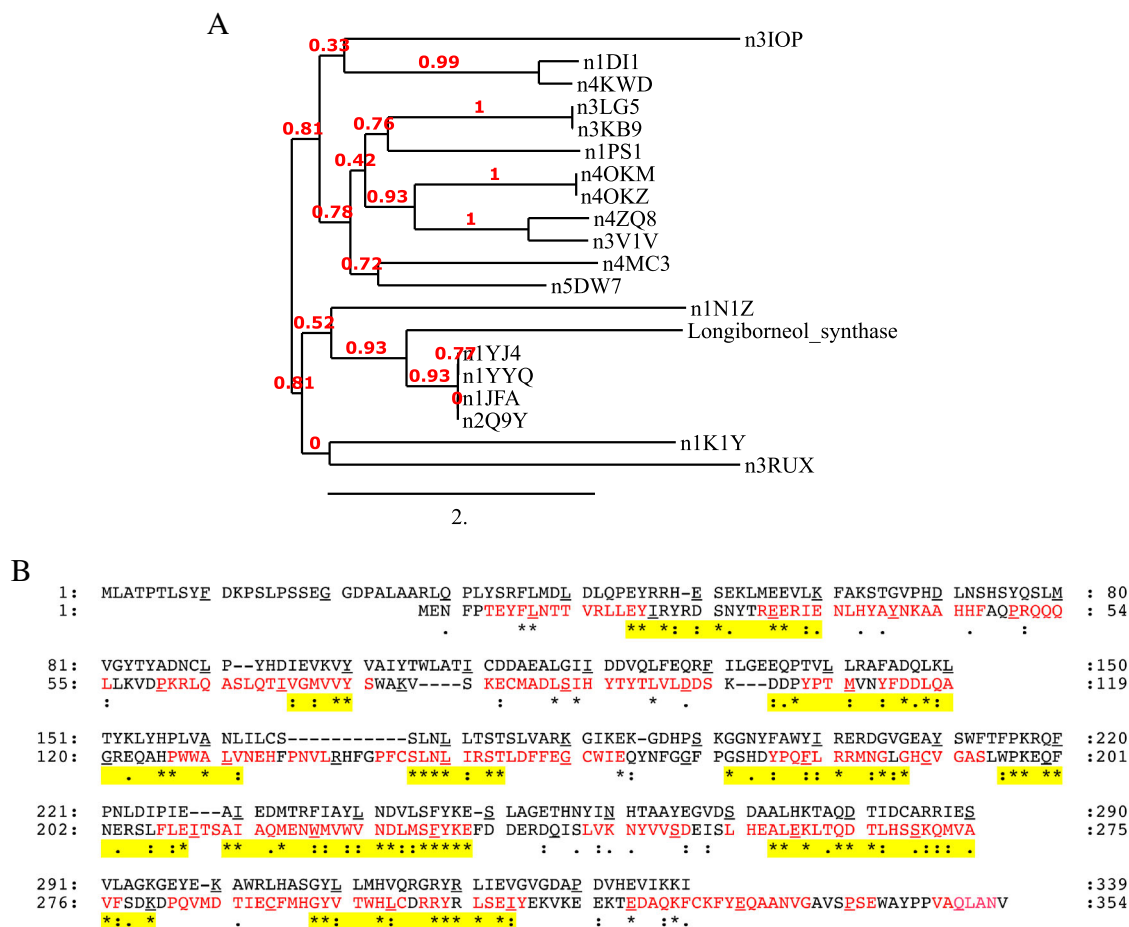


Fig. 2 **a** Dendrogram showing the longiborneol synthase related proteins in the PDB database. Numbers in red correspond to branch support values according to likelihood-ratio test [22]. **b** Protein sequence alignment between the longiborneol synthase (UniProt sequence D1M8S2, top line), and the PDB 1JFA template. Asterisks Strictly conserved residues,

colons conservative substitutions, periods semi-conservative substitutions (defined as PAM250 score > 1.5 and > 0.5 respectively). Segments with significant similarity are highlighted in yellow. Red letters 1JFA helices as assigned by the define secondary structure of proteins (DSSP) algorithm

chemical providers using the ChemicalAssistant software [36, 37]. Through diversity information, we obtained the most suitable library for the VS experiment in order to find putative longiborneol synthase inhibitors [38].

In the present work, the selected chemical library was the one providing the largest diversity of compounds (from the Otava supplier) [39]. Our database contained 15,000 molecules, each one being docked within the longiborneol synthase binding site. We also included in our compound working set all the molecules found in the PDB as known synthase ligands, so, in total, 15,119 compounds were included in the initial docking experiment. The 3D structures of all molecules were obtained using CORINA software [40]. The compound protonation states and atom names were corrected when needed for pH 7 and compatibility with the GOLD program, respectively, using a homemade script. The binding site was previously characterized using the LIGSITE program [41] after performing a structural alignment of all the protein conformers resulted from MD sampling. For each docking, 50 starting ligand conformers were used in GOLD software.

In the docking campaign, all the conformers for longiborneol synthase MD were superimposed to estimate the pocket to be used by GOLD. Pockets were detected using METAPOCKET [42]. Changes in pocket volume and surface were monitored during MD simulations using MDPocket [43]. The docking region was obtained by taking an average of the individual centers in which a sphere of 15 Å was defined for the binding. Docking results were ranked using GoldScore, which has been optimized for the prediction of ligand binding positions, and considered H-bonding energy,

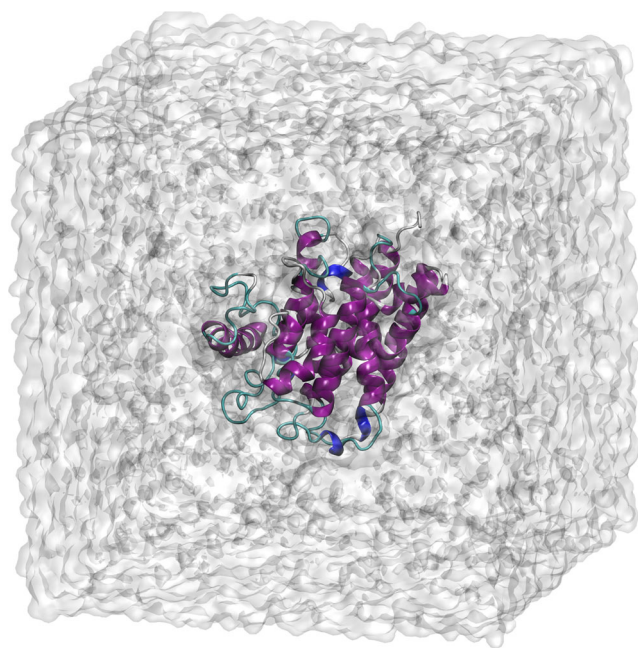


Fig. 3 The longiborneol synthase 3D₁ model in water box of 100 Å × 100 Å × 100 Å with TIP3P explicit water molecules

van der Waals energy, metal interaction and ligand torsion strain energy [32].

In order to avoid possible toxicity, the candidate compounds were surveyed using predictors such as PAINS-remover [44] and the ProTox web service [45]. The structural similarity of the selected compounds was evaluated by the Tanimoto index [31, 38, 46]. The scores were obtained by

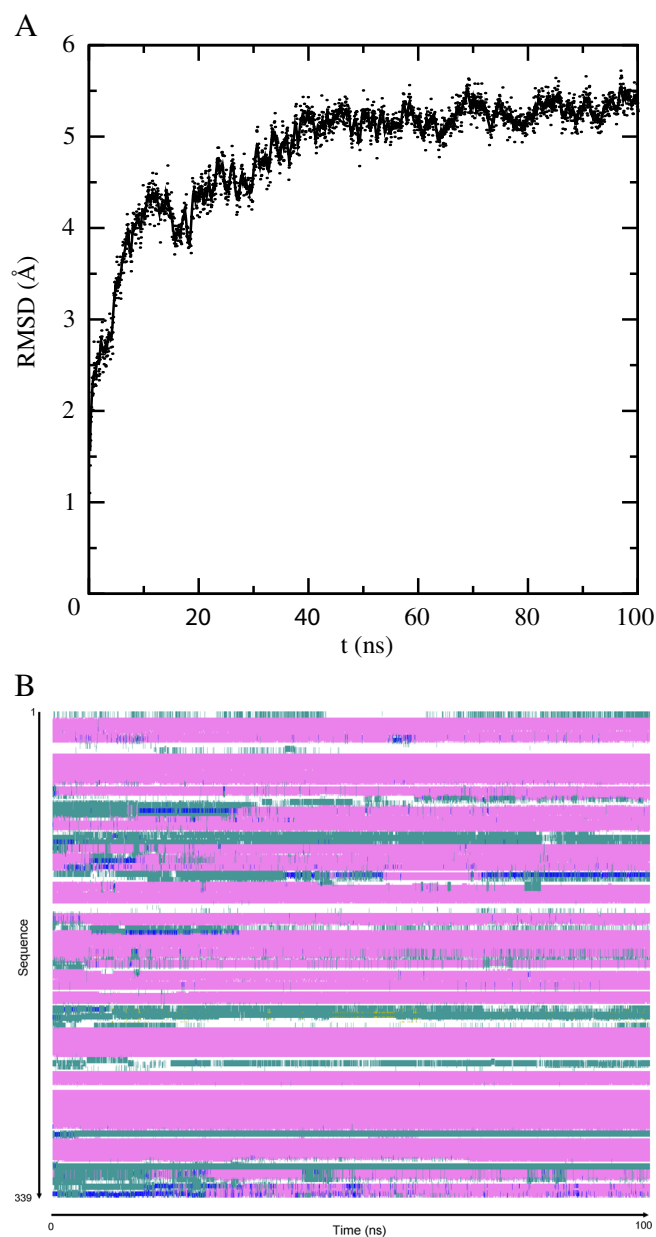


Fig. 4a, b Longiborneol synthase 3D stability in a 100 ns molecular dynamics (MD) simulation. **a** Evolution of the longiborneol synthase backbone atom RMSD during the MD trajectory, with the $t=0$ starting homology model taken as reference. This plot was prepared using 2000 data points for the whole MD trajectory, with 1 point each 50 ps. Data points represent root mean square deviation (RMSD) values for each conformational comparison. The plotted line is a running average spanning 0.5 ns. **b** Secondary structure stability during the 100-ns MD. Purple α -helices, yellow β -sheets, blue 3_{10} -helices, green turns, white coils

Open Babel using the FP2 fingerprint [47]. All calculations were performed using the GRID5K cluster network [48] on which all the batch runs were submitted using the massive spreading facility available within VSM-G [49].

Results

Starting 3D model

Table 1 summarizes the PDB codes of structural templates proposed by the homology modeling. The corresponding cladogram is presented in Fig. 2a, showing the relationship between the proteins. It appears that the closest templates are trichodiene synthase 1YJ4, 1YYQ, 2Q9Y and 1JFA PDB, the first two being a single Y305F point mutation of the native protein, the third a protein–ligand complex, and the last the structure of the free enzyme [57–59].

X-rays studies have shown that terpenoid synthases present an important conformational change to cap the active site upon metal binding [59]. It has been shown [60] that the sequence of events catalyzed by this enzyme family starts with binding of the substrate to the open form of the protein in the absence of metal ions, followed by the coordination of the ions and active site closure. To avoid such a bias in our search for inhibitors, which should bind to the open form of the enzyme, we consequently chose the free form of the trichodiene synthase structure from *Fusarium sporotrichoides* (1JFA PDB code) as the closest template to start the homology modeling process. The sequence alignment used as input to the MODELLER program is shown in Fig. 2b. This image

Fig. 5 RMSD map showing the two conformational families found during the MD simulation

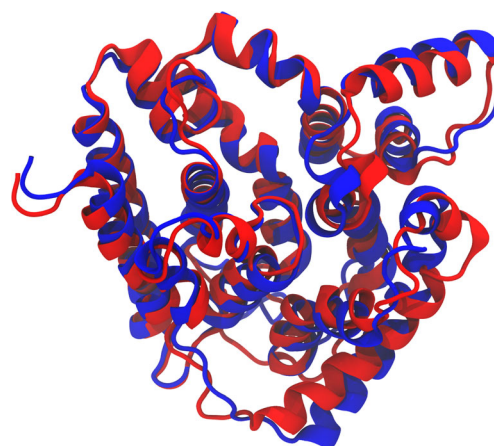
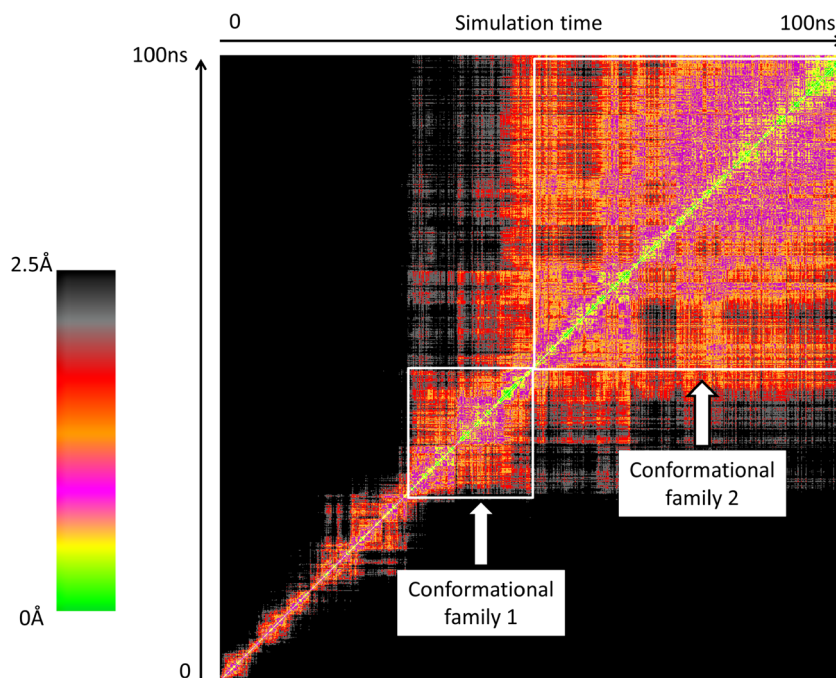


Fig. 6 Superposition of longiborneol synthase representative conformers at 100 ns MD simulation at 40 ns (in red) and 85 ns (in blue)

shows that the sequence identity between the two proteins is quite weak, especially concerning the first domain containing one of the Mg^{2+} ion binding sites.

According to the literature, three Mg^{2+} binding sites are found in the trichodiene synthase family, two corresponding to the DDXXD motif (Mg^{2+}_A and Mg^{2+}_C) and the second to the NDXXSXXXE motif (Mg^{2+}_B) [61]. In the longiborneol synthase protein, only the second metal-binding $N_{241}DXXSXXXE$ motif positioned at helix H is conserved from the proposed alignment, in good agreement with the contribution of these residues in a consensus sequence identified in all terpene synthase sequences. The first aspartate-rich $D_{100}DXXD$ motif found on helix D of the trichodiene synthase corresponds to a $E_{127}XXXXXXEE$ sequence in our alignment. Additionally, the position of the three Asp acidic

side chains of the D₁₀₀DXXD motif at the trichodiene synthase structure could be related to the position of the three Glu_{127,134,135} ones in our model. Nevertheless, this difference in the Mg ions binding groups would be critical for the precise mechanism of longiborneol production. A large loop called H-1 α has been shown to be important, participating in the lead covering the active site after ligand and Mg²⁺ ions binding [62]. In our model, this loop looks disordered, as in the unligated enzyme. Moreover, the Arg₃₀₄ residue shown to be important for the loop closure in the synthase [57] is conserved at longiborneol synthase as Arg₃₁₈.

The 3D₀ model protein obtained belongs to the all- α -protein family type, with a core formed by a bundle of eight helices arranged in two parallel layers and surrounding the ligand binding site. Concerning these helices, our 3D₀ model is very similar to the models proposed by the I-Tasser, RaptorX, Swiss Model, HHPred and Robetta servers. The

protein shape is also observed for the PDB templates considered by all the servers we used in Table 1. Most of the differences between all the proposed models concern the loops connecting the helices as well as the N- and C-terminals. Such a structured consensus at a 3D arrangement of the secondary structure pieces corroborated the folding conservation and allowed subsequent validation of the model in the MD protocol.

3D model conformational behavior

The 3D₁ model obtained after refinement of the homology model is represented in Fig. 3. The conformational behavior of the protein, as represented by the RMSD variations and the stability of areas of secondary structures, is depicted in Fig. 4a and b. It appears from the RMSD plot that the protein reached a very stable backbone conformation after 40 ns simulation,

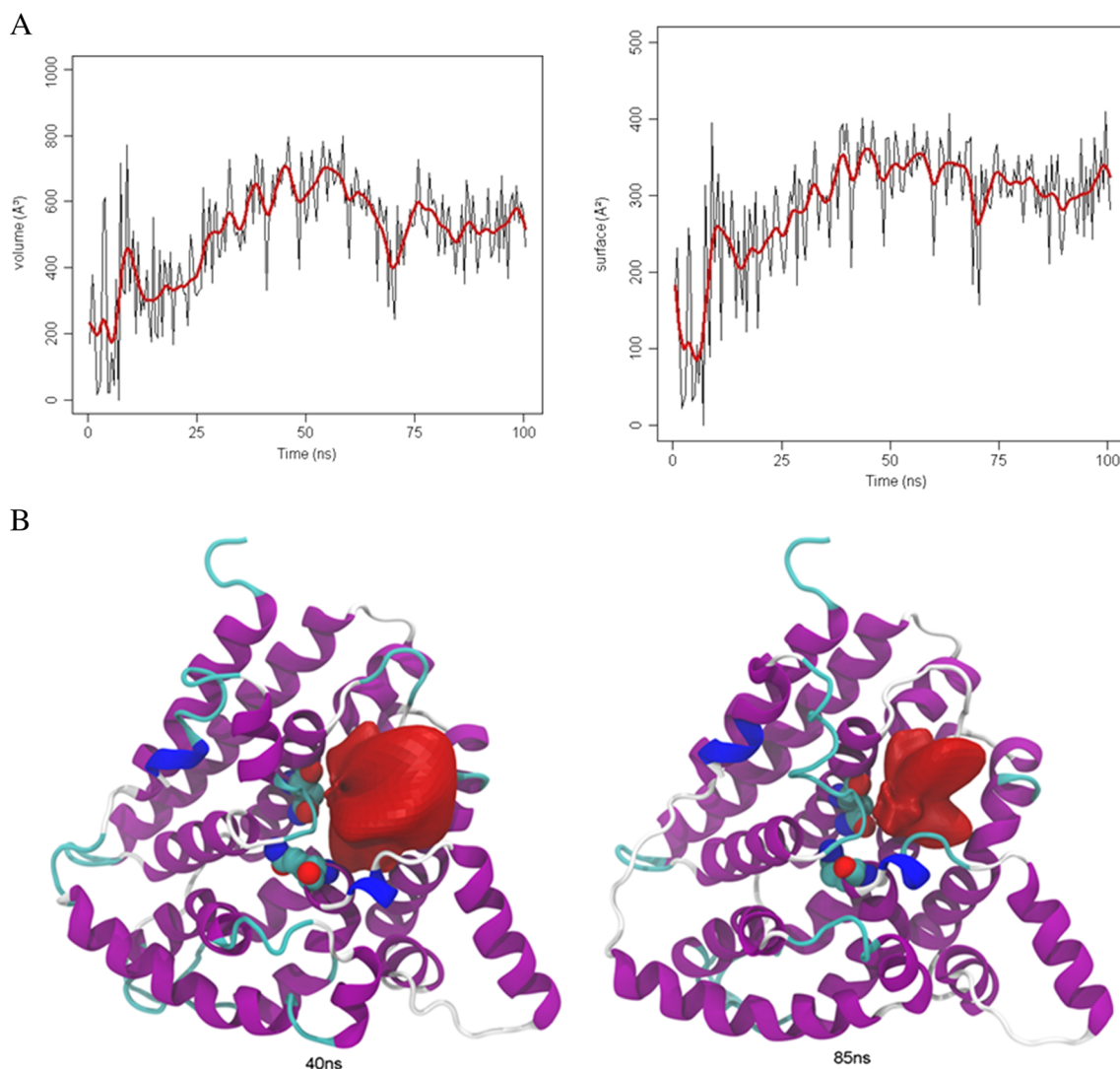


Fig. 7 **a** Pocket evolution during the 100 ns MD. *Grey lines* Volume/surface for each conformation, *red lines* running average spanning 0.5 ns. **b** Binding pocket shapes (*red*) between the two representative conformers

at 40 ns and 85 ns, respectively. The protein structures are presented as helices cartoons (*purple*) and the active site residues are depicted as CPK drawings

presenting stable helices packing and organization. Analysis of the RMSD 2D map (Fig. 5) reinforced the view that the system is stable after 40 ns until the end of the simulation, and that it is organized mostly into two major conformational families. Examples by conformers at 40 ns and 85 ns, respectively, are presented in Fig. 6.

Analysis of the active site pocket organization and properties (Fig. 7a) revealed that the binding pocket presented a large opening during the first 40 ns of the simulation to reach the first plateau, and then decreases a little. Such pocket size changes, illustrated in Fig. 7b, revealed the pocket shapes for the conformations at 40 ns and 85 ns. Despite the change in pocket shape, all residues important for catalytic activity were stable during MD, and in a structural position for proper ligand capture and a catalytic role. So, the relative positions of the acidic residues side chains Glu_{127,134,135} proposed for binding Mg²⁺_A and ions Mg²⁺_C, as well as residues Asp₂₄₂ and Glu₂₄₉, expected to bind to Mg²⁺ prior to the catalysis, were kept in positions similar to those observed in the trichodiene synthase template. The relative positioning of residues Tyr_{309,319} showed a stable structural behavior during

MD simulation, similar to that observed for Tyr_{295,305} in trichodiene synthase.

The salt bridges contributing to binding pocket stability, such as those between Asp₂₄₂ and Arg₂₀₁, Asp₂₀₄ and Lys₁₈₃, Glu₁₂₇ and Arg₃₁₈, were kept either during the trajectory or formed after a few ns simulation. Additionally, some others, such as Glu₁₃₄ with Arg₃₂₀, Glu₁₃₄ with Lys₂₄₈, were obtained later, when the 40 ns plateau was reached, or even later such as Glu₁₂₇ with Lys₁₈₀ at 80 ns (Fig. 8).

Selection of the conformational ensemble for the VS campaign

Within the two major families described above, low RMSD fluctuations were observed. Therefore, we considered that only the two conformers at 40 ns and 85 ns, respectively, should be considered as representative of these families for inclusion in the conformational ensemble for the docking step. Nevertheless, in order to explore the largest conformational space of the protein, and to avoid some important transient conformational states, less stable MD conformational states,

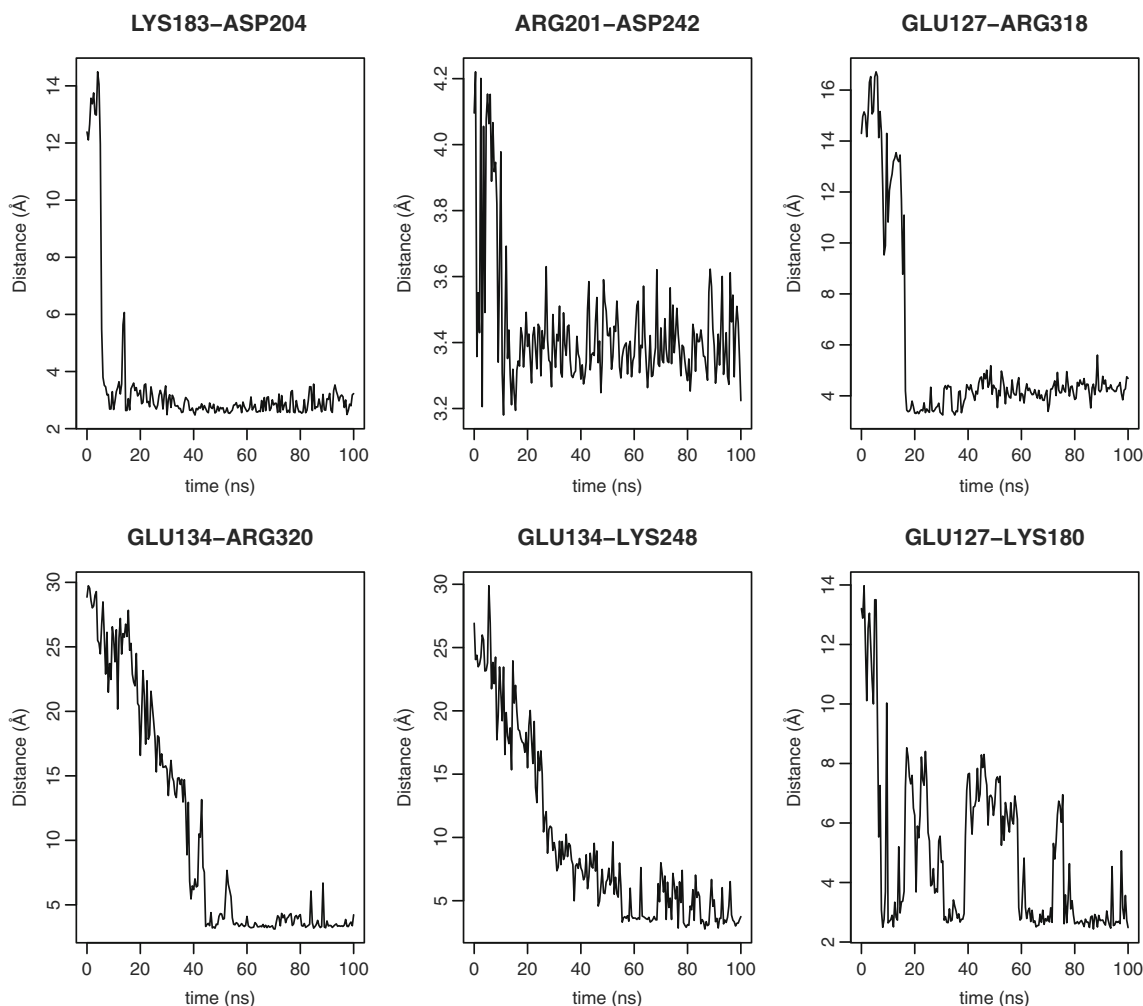


Fig. 8 Salt bridge detection at the binding site among the residues

Table 2 GOLD scores for selected compounds

Rank	Name	GOLD score	Time (ns)	Toxicity
1	2Q80A00	99.1	5	+
2	2J1PA00	98.0	5	+
3	D7119982499 ^a	95.2	17	
4	D7117231002 ^a	94.4	5	
5	D1306769 ^a	94.2	85	
6	2E95B00	94.2	12	+
7	D7211760169	93.8	14	+
8	D0107300006	93.8	40	+
9	D7217180008	93.6	17	+
10	D7011450002	92.9	5	+
11	2ZCRA00	92.3	5	+
12	D7119988305	92.1	5	+
13	D1676975 ^a	91.7	17	
		90.5	85	
14	D7117171303 ^a	91.7	40	
15	D7013940869	91.1	5	++
16	D0105850677 ^a	90.7	5	
17	2E8VA00	90.6	5	+
18	D1668794 ^a	90.6	5	
19	D7018704985 ^a	90.3	5	
20	D2196100	90.3	5	+
21	D0109280061	90.2	40	+
22	D7119988123 ^a	90.1	40	
23	D1686156	90.0	6.8	+
24	D7213330029	90.0	5	+

^a Final selected compounds

such as those appearing at 5 ns, 6.8 ns, 8.3 ns, 10.5 ns, 12 ns, 14 ns and 17 ns, were also added to the ensemble docking run (Fig. S1, Fig. S2). This ensured that good coverage of the conformational space occurred before the stabilization plateau was found after 40 ns. The two stable areas are highlighted in Fig. 7b.

VS results

From the ensemble docking campaign using these two main conformers, and the supplementary conformers as defined above, we selected only the top 100 compounds from the complete GOLD score list for further analysis. After inspection of the protein–ligand interactions involving each molecule, compound similarities (using Tanimoto criteria), and toxicity, only 25 compounds were listed as candidates in Table 2. Among these, nine (from the supplier Otava) can be strongly considered as putative hits for possible biological testing due to their chemical diversity (Tanimoto scores presented in Table 3). Additionally, the chemical formulas of these molecules are shown in Fig. 9. Among the 100 candidates, some already known as synthase inhibitors were also found in the top 25 list, such as geranylgeranyl diphosphate (2Q80A00, 2J1PA00, 2E8VA00), 1-hydroxy-2-[3'-(naphthalene-2-sulfonylamino)-biphenyl-3-yl]ethylidene-1,1-bisphosphonic acid (2E95B00) and tripotassium (1R)-4-(4'-butylbiphenyl-4-yl)-1-phosphonatobutane-1-sulfonate (2ZCRA00).

The docking of the three highest score compounds within the protein binding sites are presented in Fig. 10. The molecular volumes of these compounds fit within the pocket shape, and in the volume was depicted in Fig. 7b. Looking at the different protein–ligand interactions found, it appears that a large variety of binding possibilities occur depending on the protein conformation: for example, compound D7119982499 shares the same Lys₁₈₃ interaction with compound D7117231002, and the Arg₃₁₈ interaction with compound D1306769, while specific interactions are found for each of these compounds (Asp₂₀₄, Arg₂₀₁, Lys₂₄₈ for compound D7119982499; Asn₂₅₇ for compound D7117231002; Ser₂₄₅, His₂₅₆ for compound D1306769). A detailed study of the protein–ligand interactions found with the nine proposed compounds is summarized in Table 4, and showed that three residues are especially conserved in interaction with all ligands: Lys₁₈₀, Asp₂₀₄ and Arg₃₁₈ as shown in Fig. S3.

Table 3 Tanimoto scores (range 0 to 1) between the ten candidate compounds

Rank	D7119982499	D7117231002	D1306769	D1676975	D7117171303	D0105850677	D1668794	D7018704985	D7119988123
D7119982499	–	0.22	0.23	0.21	0.01	0.40	0.24	0.22	0.51
D7117231002		–	0.19	0.51	0.22	0.21	0.31	0.25	0.21
D1306769			–	0.20	0.11	0.24	0.20	0.32	0.23
D1676975				–	0.21	0.21	0.44	0.27	0.21
D7117171303					–	0.11	0.16	0.14	0.08
D0105850677						–	0.23	0.21	0.38
D1668794							–	0.25	0.24
D7018704985								–	0.22
D7119988123									–

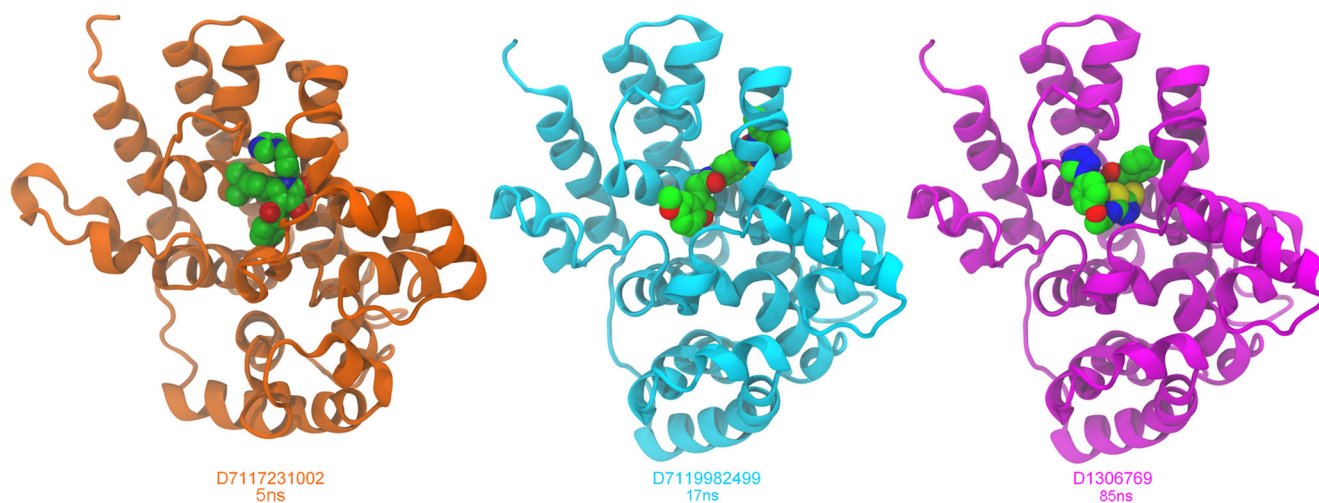


Fig. 9 Position of the three highest GOLD score selected compounds D7119982499, D7117231002 and D1306769 within their binding pocket, showing interactions found between the protein and the ligands

Fig. 10 Chemical formula of candidate compounds

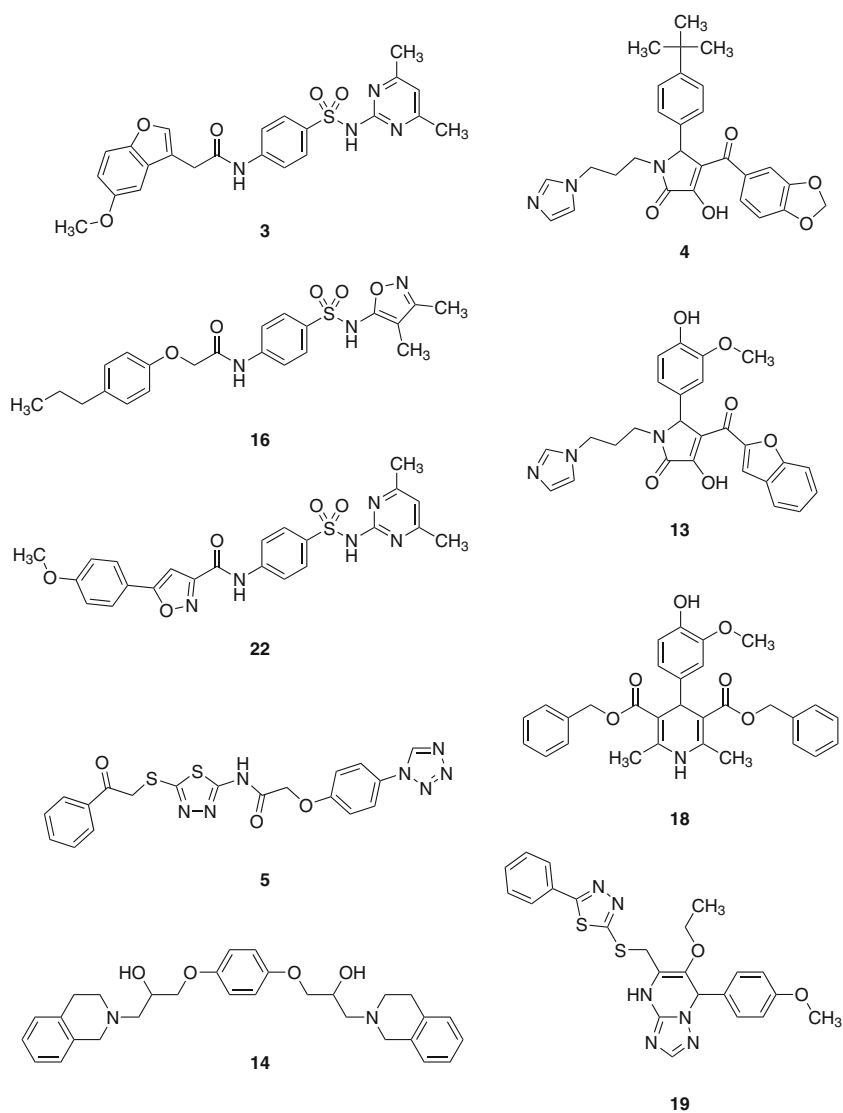


Table 4 The protein residue–ligand interactions revealed by the docking calculations. Hydrophobic interactions are represented by + and H-bonds by ++. All possible interactions are shown in Fig. S3

Compounds	Arg201	Arg318	Asn257	Asp204	Gln136	Gly133	His256	Lys180	Lys183	Lys248	Ser245	Thr255	Tyr309	Tyr319
D7119982499	++	++	–	++	–	–	–	+	++	++	+	–	–	+
D7117231002	+	+	++	+	–	–	–	+	++	–	+	+	+	+
D1306769	+	++	+	+	–	–	++	+	–	+	++	–	–	+
D1676975	–	++	+	+	++	–	–	++	+	++	+	–	–	+
D7117171303	+	+	–	+	–	++	+	+	–	–	–	–	–	–
D0105850677	++	+	+	+	–	–	–	+	–	–	–	++	–	+
D1668794	++	++	+	+	–	–	–	+	–	–	+	+	+	+
D1676975	–	++	+	+	++	–	+	++	+	++	+	–	–	+
D7018704985	+	++	–	+	–	–	–	+	+	–	–	–	++	–
D7119988123	–	++	–	+	+	+	–	++	–	–	–	–	+	++

Data mining with these nine compounds revealed that three of them were already involved in biological high throughput screening (HTS) experiments concerning a large diversity of biological targets outside the fungal kingdom [33, 34]. The chemicals D7119982499, D0105850677 and D1668794 ranked 3, 16 and 18, respectively, in Table 2 (the corresponding PubChem CID numbers are 16013960, 1107619 and 1351488, respectively). Moreover, compound D7117171303, (ranked 14, PubChem CID number 3151540), compound D0105850677 (now assigned PubChem CID number 991039) and compound D1668794 have already been patented [63, 64].

Conclusions

We have identified the first non-phosphonate-like inhibitors of *Fusarium graminearum* longiborneol synthase thanks to a strategy involving homology modeling, MD simulations, structure-based VS using ensemble docking, and data mining. Our approach also benefits from the use of existing compounds coming from a chemical libraries provider as our strategy avoids the time-consuming chemical synthesis steps usually required in classical drug design.

Previous similar approaches, aimed at other synthases, were successfully validated [65, 66]. Thus, we can reasonably expect that the chemicals proposed here would lead to a success story like the one we recently published [38]. In fact, very few inhibitors of sesquiterpene synthases are presently considered as lead compounds against *Fusarium* head blight, leading to an opportunity for the proposal of new inhibitors as shown by recent studies [67, 68].

Unfortunately, until now, no compound has been proposed to inhibit longiborneol synthase although sequences coding for this protein are found in the several *Fusarium* species. Therefore, this protein target could potentially be suitable for

controlling all such complexes globally. Consequently, the chemicals proposed here, after validation, may constitute innovative hits and will be the seeds for the development of new leads to help reduce the worldwide crop disease threat to grain producers/consumers and to global food security.

Acknowledgments This work is funded by a National Council for Scientific and Technological Development (CNPq) grant 400432/2012-9. E.B. is supported by a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) postdoctoral fellowship (#51/2013). M.U. and K.H.K. receive support from Biotechnology and Biological Sciences Research Council, UK, Institute Strategy Grants 20:20® wheat (BB/J/00426X/1).

References

- Driehuis F, Oude Elferink SJ (2000) The impact of the quality of silage on animal health and food safety: a review. *Vet Q* 22:212–216. doi:10.1080/01652176.2000.9695061
- Boenisch MJ, Schäfer W (2011) *Fusarium graminearum* forms mycotoxin producing infection structures on wheat. *BMC Plant Biol* 11:110. doi:10.1186/1471-2229-11-110
- Shimshoni JA, Cuneah O, Sulyok M et al (2013) Mycotoxins in corn and wheat silage in Israel. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* 30:1614–1625. doi:10.1080/19440049.2013.802840
- Becher R, Hettwer U, Karlovsky P et al (2010) Adaptation of *Fusarium graminearum* to tebuconazole yielded descendants diverging for levels of fitness, fungicide resistance, virulence, and mycotoxin production. *Phytopathology* 100:444–453. doi:10.1094/PHYTO-100-5-0444
- Cools HJ, Hammond-Kosack KE (2013) Exploitation of genomics in fungicide research: current status and future perspectives. *Mol Plant Pathol* 14:197–210. doi:10.1111/mpp.12001
- Sweeney MJ, Dobson AD (1998) Mycotoxin production by *Aspergillus*, *Fusarium* and *Penicillium* species. *Int J Food Microbiol* 43:141–158
- Langseth W, Ghebremeskel M, Kosiak B et al (2001) Production of culmorin compounds and other secondary metabolites by *Fusarium culmorum* and *F. graminearum* strains isolated from Norwegian cereals. *Mycopathologia* 152:23–34. doi:10.1023/A:1011964306510

8. Pedersen PB, David Miller J (1999) The fungal metabolite culmorin and related compounds. *Nat Toxins* 7:305–309. doi:10.1002/1522-7189(199911/12)7:6<305::AID-NT72>3.0.CO;2-G
9. Scarpino V, Reyneri A, Sulyok M et al (2015) Effect of fungicide application to control *Fusarium* head blight and 20 *Fusarium* and *Alternaria* mycotoxins in winter wheat (*Triticum aestivum* L.). *World Mycotoxin J* 8:499–510. doi:10.3920/WMJ2014.1814
10. Ghebremeskel M, Langseth W (2001) The occurrence of culmorin and hydroxy-culmorins in cereals. *Mycopathologia* 152:103–108. doi:10.1023/A:1012479823193
11. McCormick SP (2014) Genetic control of *Fusarium* mycotoxins to enhance food safety, Research project #421046, annual report
12. McCormick SP, Alexander NJ, Harris LJ (2010) CLM1 of *Fusarium graminearum* encodes a longiborneol synthase required for culmorin production. *Appl Environ Microbiol* 76:136–141. doi:10.1128/AEM.02017-09
13. Lavecchia A, Di Giovanni C (2013) Virtual screening strategies in drug discovery: a critical review. *Curr Med Chem* 20:2839–2860. doi:10.2174/09298673113209990001
14. Ma XH, Zhu F, Liu X et al (2012) Virtual screening methods as tools for drug lead discovery from large chemical libraries. *Curr Med Chem* 19:5562–5571. doi:10.2174/092986712803833245
15. Korb O, Olsson TSG, Bowden SJ et al (2012) Potential and limitations of ensemble docking. *J Chem Inf Model* 52:1262–1274. doi:10.1021/ci2005934
16. Bernstein FC, Koetzle TF, Williams GJ et al (1977) The protein data bank. A computer-based archival file for macromolecular structures. *Eur J Biochem* 80:319–324. doi:10.1016/S0022-2836(77)80200-3
17. Fernandez-Fuentes N, Rai BK, Madrid-Aliste CJ et al (2007) Comparative protein structure modeling by combining multiple templates and optimizing sequence-to-structure alignments. *Bioinformatics* 23:2558–2565. doi:10.1093/bioinformatics/btm377
18. Meier A, Söding J (2015) Automatic prediction of protein 3D structures by probabilistic multi-template homology modeling. *PLoS Comput Biol* 11, e1004343. doi:10.1371/journal.pcbi.1004343
19. Dereeper A, Guignon V, Blanc G et al (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36:465–469. doi:10.1093/nar/gkn180
20. Dereeper A, Audic S, Claverie J-M, Blanc G (2010) BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol Biol* 10:8. doi:10.1186/1471-2148-10-8
21. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. doi:10.1093/nar/gkh340
22. Guindon S, Dufayard J-F, Lefort V, Anisimova M (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321
23. Eswar N, Eramian D, Webb B et al (2008) Protein structure modeling with MODELLER. *Methods Mol Biol* 426:145–159. doi:10.1007/978-1-60327-058-8_8
24. SAVES. <http://services.mbi.ucla.edu/SAVES/>
25. Jorgensen WL, Chandrasekhar J, Madura JD et al (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926. doi:10.1063/1.445869
26. MacKerell AD, Bashford D, Bellott M et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616. doi:10.1021/jp973084f
27. Phillips JC, Braun R, Wang W et al (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802. doi:10.1002/jcc.20289
28. Humphreys DD, Friesner RA, Berne BJ (1994) A multiple-time-step molecular dynamics algorithm for macromolecules. *J Phys Chem* 98:6885–6892. doi:10.1021/j100078a035
29. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089. doi:10.1063/1.464397
30. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38. doi:10.1016/0263-7855(96)00018-5
31. Beautrait A, Leroux V, Chavent M et al (2008) Multiple-step virtual screening using VSM-G: overview and validation of fast geometrical matching enrichment. *J Mol Model* 14:135–148. doi:10.1007/s00894-007-0257-9
32. Verdonk ML, Cole JC, Hartshorn MJ et al (2003) Improved protein-ligand docking using GOLD. *Proteins Struct Funct Genet* 52:609–623. doi:10.1002/prot.10465
33. Liebeschuetz JW, Cole JC, Korb O (2012) Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *J Comput Aided Mol Des* 26:737–748. doi:10.1007/s10822-012-9551-4
34. Backman TWH, Cao Y, Girke T (2011) ChemMine tools: an online service for analyzing and clustering small molecules. *Nucleic Acids Res* 39:1–6. doi:10.1093/nar/gkr320
35. Akella LB, DeCaprio D (2010) Cheminformatics approaches to analyze diversity in compound screening libraries. *Curr Opin Chem Biol* 14:325–330. doi:10.1016/j.cbpa.2010.03.017
36. Monge A, Arrault A, Marot C, Morin-Allory L (2006) Managing, profiling and analyzing a library of 2.6 million compounds gathered from 32 chemical providers. *Mol Divers* 10:389–403. doi:10.1007/s11030-006-9033-5
37. Le Guilloux V, Arrault A, Colliandre L et al (2012) Mining collections of compounds with screening assistant 2. *J Cheminform* 4:1–16. doi:10.1186/1758-2946-4-20
38. Abadio AKR, Kioshima ES, Leroux V et al (2015) Identification of New antifungal compounds targeting thioredoxin reductase of *Paracoccidioides* genus. *PLoS One* 10, e0142926. doi:10.1371/journal.pone.0142926
39. Otava Chemicals. <http://www.otavachemicals.com>
40. Gasteiger J, Rudolph C, Sadowski J (1990) Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput Methodol* 3:537–547. doi:10.1016/0898-5529(90)90156-3
41. Huang B, Schroeder M (2006) LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct Biol* 6:19. doi:10.1186/1472-6807-6-19
42. Zhang Z, Li Y, Lin B et al (2011) Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics* 27:2083–2088. doi:10.1093/bioinformatics/btr331
43. Schmidtke P, Bidon-chanal A, Luque FJ, Barril X (2011) MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* 27:3276–3285. doi:10.1093/bioinformatics/btr550
44. Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* 53:2719–2740. doi:10.1021/jm901137j
45. Drwal MN, Banerjee P, Dunkel M et al (2014) ProTox: a web server for the in silico prediction of rodent oral toxicity. *Nucleic Acids Res* 42:W53–W58. doi:10.1093/nar/gku401
46. Bajusz D, Rácz A, Héberger K (2015) Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminform* 7:20. doi:10.1186/s13321-015-0069-3
47. O’Boyle NM, Banck M, James CA et al (2011) Open Babel: an open chemical toolbox. *J Cheminform* 3:33. doi:10.1186/1758-2946-3-33
48. Cappello F, Caron E, Dayde M, et al. (2005) Grid’5000: a large scale and highly reconfigurable Grid experimental testbed. In: *Proceedings of IEEE/ACM International Workshop on Grid Computing*. pp 99–106

49. Maigret B, Ghemtio L (2010) Efficiency of a hierarchical protocol for high throughput structure-based virtual screening on GRID5000 cluster grid. *Open Access Bioinforma* 41. doi: [10.2147/OAB.S7272](https://doi.org/10.2147/OAB.S7272)
50. Kelley LA, Mezulis S, Yates CM et al (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845–858. doi: [10.1038/nprot.2015.053](https://doi.org/10.1038/nprot.2015.053)
51. Biasini M, Bienert S, Waterhouse A et al (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42:252–258. doi: [10.1093/nar/gku340](https://doi.org/10.1093/nar/gku340)
52. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res*. doi: [10.1093/nar/gkh468](https://doi.org/10.1093/nar/gkh468)
53. Yang J, Zhang Y (2015) I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* 43: W174–W181. doi: [10.1093/nar/gkv342](https://doi.org/10.1093/nar/gkv342)
54. Wang C, Zhang H, Zheng W-M et al (2015) FALCON@home: a high-throughput protein structure prediction server based on remote homologue recognition. *Bioinformatics*. doi: [10.1093/bioinformatics/btv581](https://doi.org/10.1093/bioinformatics/btv581)
55. Hildebrand A, Remmert M, Biegert A, Söding J (2009) Fast and accurate automatic structure prediction with HHpred. *Proteins Struct Funct Bioinforma* 77:128–132. doi: [10.1002/prot.22499](https://doi.org/10.1002/prot.22499)
56. Källberg M, Wang H, Wang S et al (2012) Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7: 1511–1522. doi: [10.1038/nprot.2012.085](https://doi.org/10.1038/nprot.2012.085)
57. Vedula LS, Cane DE, Christianson DW (2005) Role of arginine-304 in the diphosphate-triggered active site closure mechanism of trichodiene synthase. *Biochemistry* 44:12719–12727. doi: [10.1021/bi0510476](https://doi.org/10.1021/bi0510476)
58. Vedula LS, Zhao Y, Coates RM et al (2007) Exploring biosynthetic diversity with trichodiene synthase. *Arch Biochem Biophys* 466: 260–266. doi: [10.1016/j.abb.2007.06.016](https://doi.org/10.1016/j.abb.2007.06.016)
59. Rynkiewicz MJ, Cane DE, Christianson DW (2001) Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. *Proc Natl Acad Sci USA* 98:13543–13548. doi: [10.1073/pnas.231313098](https://doi.org/10.1073/pnas.231313098)
60. Miller DJ, Allemann RK (2012) Sesquiterpene synthases: passive catalysts or active players? *Nat Prod Rep* 29:60–71. doi: [10.1039/c1np00060h](https://doi.org/10.1039/c1np00060h)
61. Gao Y, Honzatko RB, Peters RJ (2012) Terpenoid synthase structures: a so far incomplete view of complex catalysis. *Nat Prod Rep* 29:1153. doi: [10.1039/c2np20059g](https://doi.org/10.1039/c2np20059g)
62. López-Gallego F, Wawrzyn GT, Schmidt-Dannert C (2010) Selectivity of fungal sesquiterpene synthases: Role of the active site's H-1 α loop in catalysis. *Appl Environ Microbiol* 76:7723–7733. doi: [10.1128/AEM.01811-10](https://doi.org/10.1128/AEM.01811-10)
63. Ladner RD, Neamati N (2011) INHIBITORS OF dUTPase. US Patent 2011/0212467 A1
64. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH (2016) PubChem substance and compound databases. *Nucleic Acids Res* 44(D1):D1202–D1213
65. Durrant JD, Cao R, Gorge AA et al (2011) Non-bisphosphonate inhibitors of isoprenoid biosynthesis identified via computer-aided drug design. *Chem Biol Drug Des* 78:323–332. doi: [10.1111/j.1747-0285.2011.01164.x](https://doi.org/10.1111/j.1747-0285.2011.01164.x)
66. Pani G, Scherm B, Azara E et al (2014) Natural and natural-like phenolic inhibitors of type B trichothecene in vitro production by the wheat (*Triticum* sp.) pathogen *Fusarium culmorum*. *J Agric Food Chem* 62:4969–4978. doi: [10.1021/jf500647h](https://doi.org/10.1021/jf500647h)
67. Johnson ET, Evans KO, Dowd PF (2015) Antifungal activity of a synthetic cationic peptide against the plant pathogens *Colletotrichum graminicola* and three *Fusarium* species. *Plant Pathol J* 31:316–321. doi: [10.5423/PPJ.NT.04.2015.0061](https://doi.org/10.5423/PPJ.NT.04.2015.0061)
68. Crespo-Sempere A, Estiarte N, Marin S et al (2015) Targeting *Fusarium graminearum* control via polyamine enzyme inhibitors and polyamine analogs. *Food Microbiol* 49:95–103. doi: [10.1016/j.fm.2015.01.020](https://doi.org/10.1016/j.fm.2015.01.020)