PLOS PATHOGENS

The structural landscape and diversity of *Pyricularia oryzae* MAX effectors revisited

Mounia Lahfa[1], **Philippe Barthe**[1], **Karine de Guillen**[1], **Stella Cesari**[2], **Mouna Raji**[1], **Thomas Kroj**[2], **Marie Le Naour—Vernet**[2], **François Hoh**[1], **Pierre Gladieux**[2], **Christian Roumestand**[1], **Jérôme Gracy**[1], **Nathalie Declerck**[1], **André Padilla**[1]*

1 Centre de Biologie Structurale, Univ Montpellier, CNRS UMR 5048, INSERM U 1054, Montpellier, France,
2 PHIM Plant Health Institute, Univ Montpellier, INRAE, CIRAD, Institut Agro, IRD, Montpellier, France

* andre.padilla@cbs.cnrs.fr

Abstract

Magnaporthe AVRs and ToxB-like (MAX) effectors constitute a family of secreted virulence proteins in the fungus *Pyricularia oryzae (syn. Magnaporthe oryzae)*, which causes blast disease on numerous cereals and grasses. In spite of high sequence divergence, MAX effectors share a common fold characterized by a ß-sandwich core stabilized by a conserved disulfide bond.

In this study, we investigated the structural landscape and diversity within the MAX effector repertoire of *P. oryzae*. Combining experimental protein structure determination and *in silico* structure modeling we validated the presence of the conserved MAX effector core domain in 77 out of 94 groups of orthologs (OG) identified in a previous population genomic study. Four novel MAX effector structures determined by NMR were in remarkably good agreement with AlphaFold2 (AF2) predictions. Based on the comparison of the AF2-generated 3D models we propose a classification of the MAX effectors superfamily in 20 structural groups that vary in the canonical MAX fold, disulfide bond patterns, and additional secondary structures in N- and C-terminal extensions. About one-third of the MAX family members remain singletons, without strong structural relationship to other MAX effectors. Analysis of the surface properties of the AF2 MAX models also highlights the high variability within the MAX family at the structural level, potentially reflecting the wide diversity of their virulence functions and host targets.

Author summary

MAX effectors are a family of virulence proteins from the plant pathogenic fungus *Pyricularia (syn. Magnaporthe) oryzae* that share a similar 3D structure despite very low amino-acid sequence identity. Characterizing the function and evolution of these proteins requires a detailed understanding of their structural diversity. With this in mind, we have determined the NMR structures of four new MAX effectors and shown a near-perfect match with the corresponding AlphaFold2 (AF2) models. We then applied a prediction pipeline based on similarity searches with structural modeling using the AF2 software to predict MAX effectors in a collection of 120 *P. oryzae* genomes. The resulting models and

experimental structures revealed that the MAX core while preserved is highly permissive to secondary structure variations and may coexists with extensive structural diversity in terms of structured N- or C-terminal extensions permitting their classification. For a subset of AF2 models, we have also analyzed the physico-chemical properties of the core domain surfaces, adding another, more functional perspective, notably surface electrostatics and stickiness. This work constitutes a major step in understanding the relationships among MAX effectors by analyzing their structural landscape and cataloguing specific physico-chemical properties. It also provides valuable insights for guiding research into the putative targets of these effectors in infected plant hosts.

Introduction

Fungal plant pathogens secrete small proteins, called effectors, which promote disease by targeting cellular processes in the host plant. There are hundreds of predicted effectors in the genomes of plant pathogenic fungi that are usually identified by their secretion signal and other characteristic features such as cysteine enrichment [1–3]. Some effectors are of particular interest since they constitute avirulence (AVR) factors that are detected by plant immune systems, and this recognition renders crops resistant to severe diseases. Most fungal effectors show no amino-acid sequence homology to known proteins or protein domains. Thus, their biological function cannot be predicted by systematic *in silico* analysis (such as domain searches), but must be experimentally elucidated on a case-by-case basis. Although the similarity between effectors is low, the availability of many highly accurate structural models provides a more precise source of information for inferring the evolution and functional classification of effectors. To this aim, we need new structure-based analysis tools for better phylogeny inference, improved prediction of protein-protein interface prediction and relevant functional classification.

Recently, a combination of primary sequence pattern searches and structural modeling resulted in a major breakthrough in effector biology by revealing that fungal effector repertoires are actually dominated by a limited number of families sharing common structures despite extensive sequence variability [4–6]. One such family are the MAX (*Magnaporthe* AVRs and ToxB-like) effectors we identified in *Pyricularia oryzae* (synonym: *Magnaporthe oryzae*), the causal agent of blast disease in rice, wheat, and other cereals or grasses [7]. This pathogenic fungus is both a major threat to global food security [8] and a prime experimental model in plant pathology [9,10]. By solving the solution structure of two *P. oryzae* effectors, AVR1-CO39 and AVR-Pia, we discovered strong structural similarities between these sequence-unrelated effectors as well as with the ToxB effector from the wheat infecting fungus *Pyrenophora tritici-repentis* [7].

MAX effectors are specific to plant pathogenic ascomycete fungi, and they have undergone a major expansion in *P. oryzae*. Analysis of 120 isolates of *P. oryzae* identified ~7800 putative MAX effectors that were grouped in 94 groups of orthologs (OGs) [11]. Individual isolates have 58 to 78 MAX effectors, corresponding to 5 to 10% of their effector repertoire. This high number suggests that MAX effectors have a critical role in the virulence of the blast fungus. This idea is further supported by the fact that MAX effectors are massively and specifically expressed during the early stages of plant infection and targeted by the plant immune system [7,10,11]. Indeed, nearly half of the cloned AVR genes of *P. oryzae* correspond to MAX effectors [7,12–15]. Analysis of the recognition of the MAX effectors AVR-Pia, AVR1-CO39, and AVR-Pik by the rice immune receptors RGA5 and Pik-1 suggests that they target small heavy

metal-associated domain proteins (sHMAs), which show similarity to copper chaperones [16]. Another MAX effector, AvrPiz-t, targets four different host proteins involved in different cellular processes [17,18]. In comparison with other secreted proteins, MAX effectors show high presence/absence polymorphism and important sequence variability that is maintained by balancing selection [11]. Analysis of the MAX effectors AVR1-CO39, AVR-Pia and AVR-Pik indicates that non-synonymous polymorphisms frequently co-localize with residues interacting with immune receptors and, presumably, also with their host target proteins [11].

To better understand the function and evolution of *P. oryzae* MAX effectors, a systematic and robust analysis of their three-dimensional structure, especially outside the MAX core domain, is still needed. Indeed, in addition to the core, many MAX effectors possess N- and C-terminal extensions that could have critical roles, for instance, by establishing specific protein-protein interactions. Examining these extensions in more detail, as well as other non-conserved structural features, may thus provide insights into the mechanism by which MAX effectors acquire new virulence capabilities, and allow a more comprehensive classification within the MAX family.

In the present study, we combined experimental and computational approaches to finely characterize the structural diversity of MAX effectors. We undertook structural studies of several MAX candidates and solved four new structures by Nuclear Magnetic Resonance (NMR). Comparison of these new experimental MAX structures with corresponding 3D models generated by template-based or *ab initio* approaches revealed the reliability of the MAX predictions. The highest accuracy was achieved with AlphaFold2 (AF2), which predicted the structure of MAX effectors, including non-conserved side-chains in terminal extensions that were not previously observed. We therefore revisited with the use of AF2 the structural landscape of *P. oryzae* MAX effectors, and validated the presence of a MAX core in 77 of the 94 previously defined MAX OGs [11]. Structural alignment of the AF2 models allowed us to refine the structural consensus and to explore the variability within the MAX family, including deviations from the canonical fold, disulfide bond pattern variations, additional secondary structures within N- and C-terminal extensions as well as variations in surface properties, such as stickiness and electrostatics of the core domain.

This work represents the most extensive structural analysis of a fungal effector family of a plant pathogen to date. It also provides valuable knowledge for analyses aimed at elucidating the function of MAX effectors, notably through the prediction of interaction sites within the MAX fold that could contribute to targeting host proteins during infection.

Results

The structure of MoToxB presents the canonical MAX fold

The MAX effector orthogroup OG33 of *P. oryzae* has high protein sequence similarity with the effector ToxB from *P. tritici-repentis* [11]. We determined the structure of its representative in the Br58 isolate (S2 Table) by X-ray crystallography using molecular replacement at 1.38Å resolution (S1 Supplementary Methods). The structure confirmed high structural similarity with ToxB from *P. tritici-repentis* and therefore this MAX effector was renamed MoToxB. The structure was also similar to the experimentally determined structures of five other *P. oryzae* MAX effectors that share less than 13% sequence identity with MoToxB (S1 Fig). Like other MAX effectors, MoToxB is structured as a 6-stranded ß-sandwich (ß1 to ß6) of two triple-stranded antiparallel ß-sheets with a ß6ß1ß2-ß3ß4ß5 topology. A disulfide bond that is conserved in almost all other MAX effectors forms a bridge between ß1 and the loop connecting ß4 and ß5. The two cysteines forming this bond are the only residues that are highly conserved in MAX effectors. In MoToxB a second disulfide bond connects ß2 and ß6 (Fig 1).
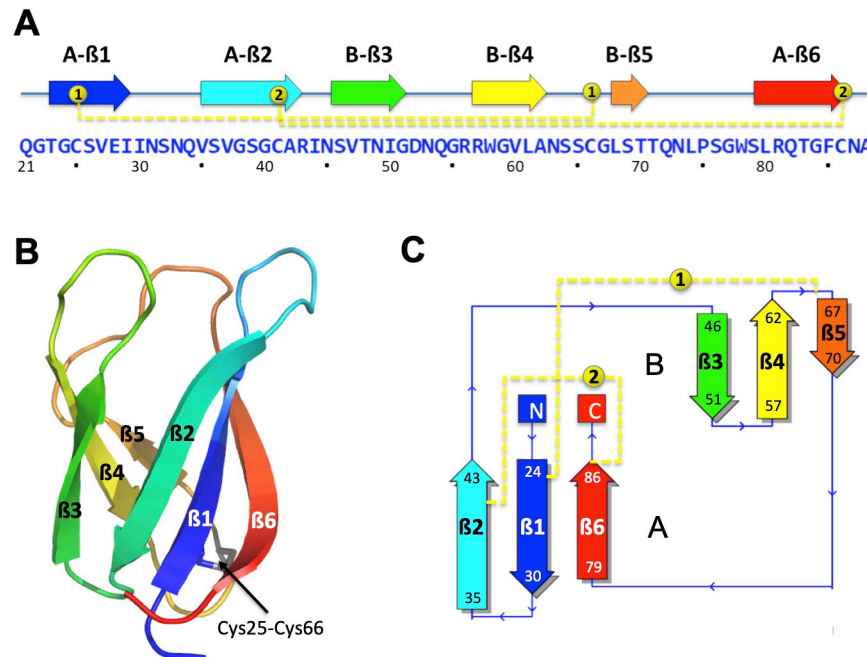
Fig 1. Structure of the M. oryzae ToxB (MoToxB) MAX effector. Primary and secondary structure of MoToxB showing the triple-stranded beta-sandwich forming the conserved MAX core with the two beta-sheets labeled by A and B, strands indicated by arrows and two disulfide bonds in yellow dotted lines. Disulfide bond SS "1" is almost strictly conserved in MAX effectors. (B) Cartoon representation of MoToxB crystal structure (PDB 6R5J) in rainbow color and the conserved disulfide bond "1" shown by sticks. (C) MoToxB topology diagram drawn by PDBsum and colored using the same color scheme as in A and B.

https://doi.org/10.1371/journal.ppat.1012176.g001

NMR structures validate template-based modeling of MAX effectors

In our previous analysis of the MAX effector repertoire in *P. oryzae*, we used a combination of Hidden Markov Model (HMM) pattern searches and hybrid multiple Template Modeling (TM) for predicting the 3D structure of the conserved MAX core of each representative sequences of the 94 MAX effector OGs (S2 Table) defined in that study (OGs are provided in S1 Table) [11]. The reliability of the 3D models (referred as TM-pred models) was evaluated by the TM-pred score, which is an estimate of the TM-score that would be observed in a structural alignment of the TM-pred model with the corresponding experimentally resolved structure using TM-align. For these analyses eight experimental structures of MAX effectors were used as templates for homology modeling, and as a training data set for the TM-pred scoring function [11]. To improve the accuracy of template-query alignments that is crucial for obtaining accurate predicted models as indicated in [19] we have implemented a bidirectional dynamic programming algorithm described in [11] exploring many sub-optimal alignments which where assessed using complementary structural scoring methods. Almost 90% of the OG proteins were modeled at high confidence as MAX structures (TM-pred score > 0.6). Only three TM-pred models, those of OG22, OG77 and OG85, had a TM-pred score below 0.5 and were suspected not to be MAX effectors (S1 Fig).

To deepen insight into MAX effectors and to assess the validity of the predictions, we attempted to resolve the experimental structures of 10 new MAX effector candidates showing high expression during the biotrophic stage of infection (S3 Table) [11]. Using NMR spectroscopy, we successfully determined the structure of four of them (OG28, OG47, OG60 and
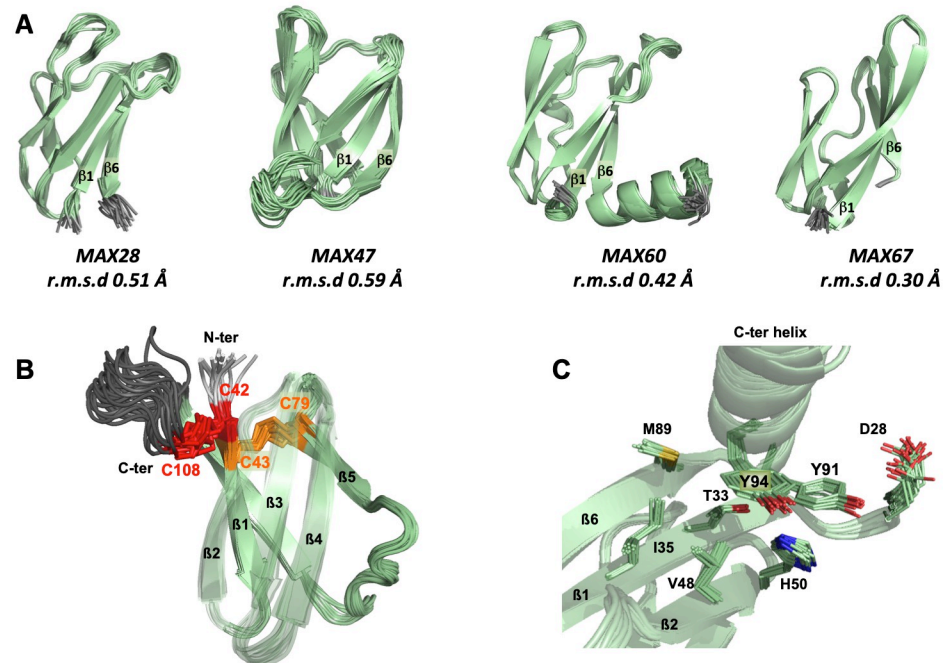
Fig 2. NMR structures of four MAX effectors. (A) NMR structures of MAX28, MAX47, MAX60 and MAX67 showing the superimposition and the r.m.s.d. of their 20 best conformers. The N- and C-terminal unstructured extensions before ß1 and after ß6, respectively, are not shown, except for the C-ter helix of MAX60. (B) View of the two disulfides bonds, C42-C108 (red) and C43-C79 (orange) for NMR structure of MAX47. The loop between the end of ß6 and the C-terminus is colored in dark grey. The ß2, ß3 and ß4 strands are transparent. (C) Local environment of the two residues Y91 and Y94 in the C-terminal helix of MAX60.

OG67), and confirmed that they all have a MAX core fold (referred to as "MAX" instead of "OG clusters" from hereon). The 20 best-refined conformers obtained for each of these effectors were superimposed (Fig 2A), and the high quality of the NMR structures was supported by the low root mean square deviations (r.m.s.d). The complete structural statistics are given in S4 Table.

MAX28, MAX47 and MAX67 displayed the characteristic MAX ß-sandwich fold stapled through the conserved disulfide bridge linking ß1 and the ß4-ß5 loop (Fig 1). The two cysteines forming this bond are the only residues that are highly conserved in MAX effectors. A particular feature of MAX67 was the exceptional length of the ß1 and ß2 strands (10 a.a.), which were longer than those in all other determined MAX effector structures. MAX60 diverged from the canonical MAX fold by the replacement of the ß5 strand by a helical turn, preventing the corresponding residues from forming a regular ß-sheet with ß4 and ß3.

In addition to the central MAX core, MAX28, MAX47 and MAX60 possess remarkable N- and/or C-terminal extensions. For MAX47, the 23 residue-long sequence extending before the ß1 strand was enriched in serine residues and was not resolved in the NMR structure. The ß1 strand started with two consecutive cysteine residues, which formed disulfide bonds that were well defined in the NMR structure (Fig 2B). The first cysteine made a disulfide bond with the last C-terminal cysteine residue (C42-C108). The second cysteine formed the disulfide bond with the cysteine in the ß4-ß5 loop (C43-C79), which is present, as already mentioned, in nearly all canonical MAX effectors, and named SS "1" disulfide bond in the following.

MAX60 has a C-terminal extension, which forms a well-defined α-helix that is attached to the structural core by hydrophobic contacts established by the aromatic rings of two tyrosine

residues (Y91 and Y94). Nuclear Overhauser Effects (NOEs) in the NMR experiments revealed close contacts between tyrosine Y91 and residues D28 to T33 and H50, and between tyrosine Y94 and residues T33, I35, V48 and M89 (Fig 2C).

The resonances of the 42 residue-long C-terminal extension of MAX28 that contains lysine-repeated motifs (KxxxK) were not assigned in the NMR spectra. This is consistent with the prediction of this part of the protein as being unstructured.

The four new NMR structures of MAX effectors were superimposed using TM-align with the corresponding TM-pred models that we previously generated by template homology modeling (Fig 3A). The quality of the models was evaluated by the root-mean-square deviation (r.m.s.d) calculated between the observed and predicted structures and the TM-scores given by TM-align (a value of 1 meaning a perfect match). Comparison of the superimposed back-bones showed that the overall MAX fold as well as the relative orientation of the two ß-sheets forming the central ß-sandwich were all well predicted. The prediction was particularly good for MAX28 whose MAX domain of the TM-pred model precisely matched the experimental structure (r.m.s.d. = 2.11 Å), even for the loops joining the ß-strands. MAX28 was also the effector with the highest estimated TM-pred score (0.75), in remarkably good agreement with the true TM-score (0.74) of the TM-pred model aligned to the NMR structure. Structural predictions of the MAX core were also very good for MAX67, except for the long strands ß1 and ß2 in the NMR structure that were not as long in the TM-pred model. This deviation explains the rather low TM-pred value for TM-pred_MAX67. The models of MAX47 and MAX60 showed also poor definition of certain ß-strands and exhibited strong divergence in connecting loops when compared to the NMR structures. For MAX47, the limited reliability of the TM-pred model was reflected by the low TM-pred score (0.63).

AlphaFold2 reliably predicts MAX effectors core and extensions

To determine the accuracy of AlphaFold2 (AF2) [20] for the prediction of MAX effector structures, we used it to model MAX28, MAX47, MAX60 and MAX67 before the release of their NMR structures in the PDB. We used three different implementations of AF2: 1) ColabFold with MMseqs2 and PDB templates, 2) AlphaFold with Jackhmmer and PDB templates, and 3) ColabFold with *custom* MSA without PDBs (see Materials and methods). For each sequence query the predictive quality of the top-ranked models was assessed based on the predicted local distance difference test (pLDDT) [20,21]. The pLDDT score, scaling from 0 to 100, is a residue-level accuracy score computed by AF2 that provides an estimate of the confidence of each residue's predicted position in the protein structure. We considered the average pLDDT score for the overall protein, or only for residues within the predicted MAX core domain, hereafter, called the MAX pLDDT score. The MAX pLDDT score was used to estimate the prediction confidence on the core structure of each AlphaFold model. This score is optimized during the AlphaFold training phase to predict the per-residue accuracy of the structure. It was shown to reliably estimate the average backbone deviation between the backbones of the predicted model and native structure [20]. For each MAX effector, the AF2 model having the highest MAX pLDDT score was selected (among 15) and referred as its AF_MAX model (Fig 3).

The MAX core domain was predicted with high confidence in AF_MAX47, AF_MAX60 and AF_MAX67 according to their high MAX pLDDT scores, close to or exceeding 90 (Fig 3B and S5 Table). The best models were obtained in all three cases from the Jackhmmer AF2 implementation. A lower confidence score (74.1) was retrieved for the best AF2 model of MAX28, generated with the *Custom* MSA implementation. Nevertheless, AF_MAX28 was very close to the experimental structure of MAX28 according to the average r.m.s.d. value (1.42 Å) calculated from superimposing the MAX core backbone atoms of the NMR
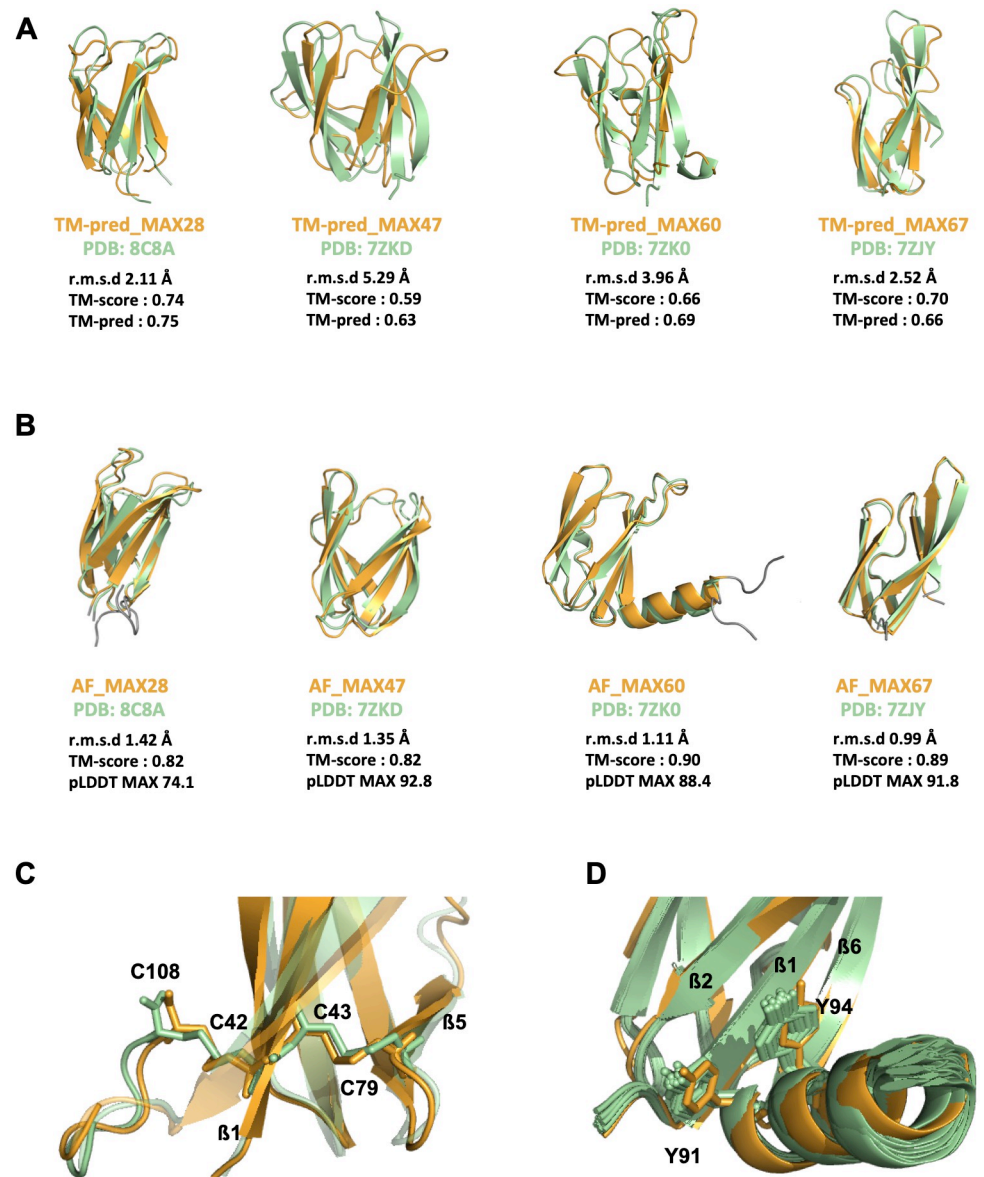
Fig 3. Comparison of newly determined NMR structures of MAX effectors with their TM-pred or AF predicted models. Superimposition of the four MAX effectors determined in this study by NMR shown by the best model (in green) and of their corresponding 3D models (in orange) predicted by hybrid multiple template modeling (TM-pred models shown in A) or AlphaFold2 (AF2 models shown in B). Metrics used for the quantitative assessment of the similarities between the predicted 3D models and their respective experimental structures are indicated: the root mean square deviation (r.m.s.d; the lowest, the best), the template modeling score (TM-score from TM-align, a value of 1 corresponding to a perfect match), TM-pred score (a predictive estimate of the TM-score). The N- and C-terminal boundaries were set according to the TM-pred models and did not include extensions determined in the NMR structures. (B) The r.m.s.d. between backbone heavy atoms of the superimposed NMR structure and AF2 models is given for the MAX domain only (S5 Table), as well as the predicted local distance difference test (pLDDT) that was used to estimate the reliability of the AF2 predictions in the MAX core (MAX pLDDT score, S5 Table). (C) View of the two disulfides bonds of MAX47, C42-C108 and C43-C79, as observed in the best NMR conformer and in the predicted AF_MAX47 model. (D) Position of the two tyrosyl side-chains of Y91 and Y94 in the C-terminal helix of MAX60 in the 20 NMR conformers and in the AF2 predicted model.

https://doi.org/10.1371/journal.ppat.1012176.g003

conformers. Indeed, for all four MAX effectors, the MAX domain of the best AF2 model displayed side-chain rotamers almost identical to those in the experimental structures and that were within the uncertainty of the NMR approach, i.e. density of NMR-derived constraints (Fig 3B).

AlphaFold2 modeling also succeeded in predicting details within the core domains. The cysteine residue side-chains forming the conserved SS1 disulfide bond were well defined in all four AF_MAX effector models. The same was true for those forming the additional disulfide bond, bridging the ß1 and the C-terminal extension, in the structure of MAX47 (Fig 3C). Another example of consistency between experimental structures and AF2 models was the remarkably well defined position and orientation of the C-terminal helix of MAX60, including the two tyrosine residues whose aromatic side chains stacked over the ß1-ß2-ß6 ß-sheet of the MAX core (Fig 3D). For MAX28, both AF2 model and NMR structure were consistent in predicting unstructured N- and C-terminal extensions. Moderate deviations from the experimental structure were only observed for residues in the C-terminus of MAX67 (S2 Fig).

AlphaFold2 validates 77 out of 94 MAX OGs

Given the high quality of the AF2-generated models of MAX effectors, we applied the same AF modeling strategy to all the remaining 90 *P. oryzae* MAX OG representatives. The models were visualized to check the presence of the characteristic MAX core by inspecting the ß-strand topology and the presence of the conserved SS "1" disulfide bond (Pymol. v.1.6; Delano 2002). OG proteins showing significant topological deviations from this canonical MAX fold were discarded. The presence of the short ß5 strand was not used as a filtering criterion. Among the 80 AF2 models that matched the canonical MAX structure (Fig 4 and S5 Table), 57 had MAX pLDDT scores greater than 80 and 20 had MAX pLDDT scores ranging from 60 to 80. Only three OG proteins, OG26, OG73 and OG94, exhibiting a central core compatible with a MAX fold had a MAX pLDDT score below 60 and were not kept in our final selection of 77 validated MAX structures. About one-third of the selected models were generated with the ColabFold implementation using a *Custom* MSA constraining the alignment of the predicted ß1-ß4 strands and of the conserved cysteine residues in the SS "1" disulfide bond (S6 Table). An overview of the general structural features characterizing the 77 validated MAX effector AF models is given in S7 Table.

Among the 14 OG proteins that were not predicted to fold with the MAX topology, three (OG22, OG77 and OG85) were previously flagged as suspects based on their low TM-pred score, and five (OG04, OG59, OG65, OG68 and OG81) had a TM-pred score below 0.6 (S1 Fig). Other OG proteins such as OG51 and OG54 that exhibited a TM-pred score above 0.6 compatible with a MAX structure displayed significant distortions from the canonical MAX fold when modeled by AlphaFold2. For OG51, two models computed with ColabFold MMseqs2 and the AlphaFold Jackhmmer implementations gave very similar models (backbone r.m.s.d. of 1.77Å) with overall pLDDT scores of 52.6 and 61.9, respectively. However, the C-terminal ß-strand of OG51 had a parallel orientation relative to the first ß1 strand that was not compatible with the MAX topology. The best OG54 model had a pLDDT score of 59.4 but deviated from the MAX topology by the absence of the C-terminal ß6 strand, which was not accurately modeled.

The 14 OG clusters that gave inconclusive AF2 models were submitted to three other protein structure prediction web-servers, RaptorX [23] (http://raptorx6.uchicago.edu), Rosetta-Fold [24] (https://robetta.bakerlab.org) and ESMFold [25] (https://esmatlas.com). None of the computed models displayed the canonical MAX fold, with consistently low prediction scores,
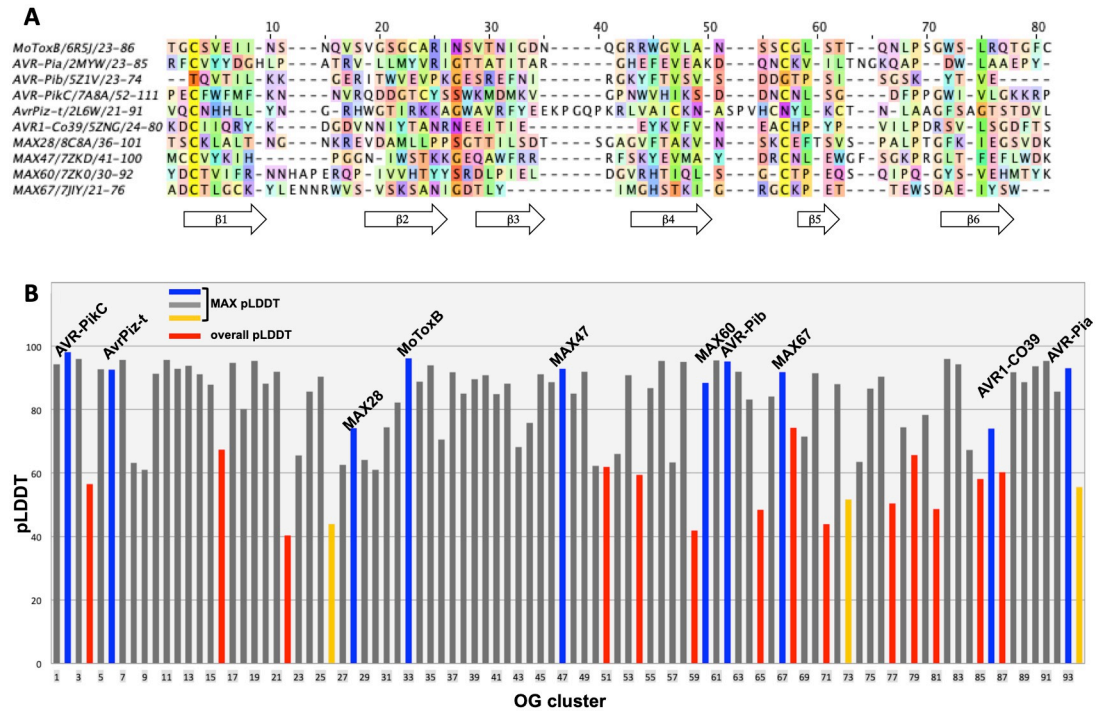
Fig 4. pLDDT scores of AF2 models for known and predicted MAX effectors. (A) Structural alignments of experimentally determined MAX effector structures using MoToxB structure for reference. Residues are colored according to the Taylor scheme [22](Taylor, 1997) and conservation (above 15% threshold) is used as a shading factor. (B) pLDDT or MAX pLDDT scores of the best AF2 models of the 94 OG representatives. OG representatives, whose models did not match the canonical MAX fold, have red bars showing the best overall pLDDT score. Bars of OG representatives with canonical MAX folds indicate the best MAX pLDDT score and are in blue for OGs with experimentally determined structures, in grey if the MAX pLDDT score was higher than 60 and in orange when it was below. Full data is available in S5 Table.

https://doi.org/10.1371/journal.ppat.1012176.g004

confirming the challenging nature of modeling these OG cluster sequences that may deviate from the MAX topology.

Variations around the canonical MAX fold

From the analysis of the 77 AF models we could define more precisely the consensus structural elements that constitute the conserved MAX core. The average size of the ß strands and the connecting loops forming the canonical MAX fold given in S7 Table showed that ß1 and ß2 were usually of similar size and associated together with ß6 to form the longest anti-parallel ß sheet, while ß3 and ß4 strands were generally shorter (Fig 5A and S7 Table). Most variation occurred at the short, 2 to 8 residues long ß5 strand, which is preceded by a loop with variable length that can count up to 25 residues (MAX29). The ß5 strand can be associated with a short helix (MAX15, MAX60), totally absent (MAX78 and MAX83) or replaced by a helix (MAX20).

The number of disulfide bonds stabilizing the MAX protein can also greatly vary, from none in MAX61 and MAX62 (AVR-Pib) up to three in MAX46 or four in MAX52 (S7 Table). A unique member of the MAX family was MAX52, whose AF model consisted of two MAX core domains arranged in tandem and designated MAX52A and MAX52B in S7 Table, each having two disulfide bonds. Besides the conserved disulfide bond (SS "1"), which is a hallmark of the MAX domain, three types of additional disulfide bonds (SS "2", "3" and "4") were found in the experimental and AF2 MAX structures (Fig 5A and 5B). SS "4", joining the N-terminus
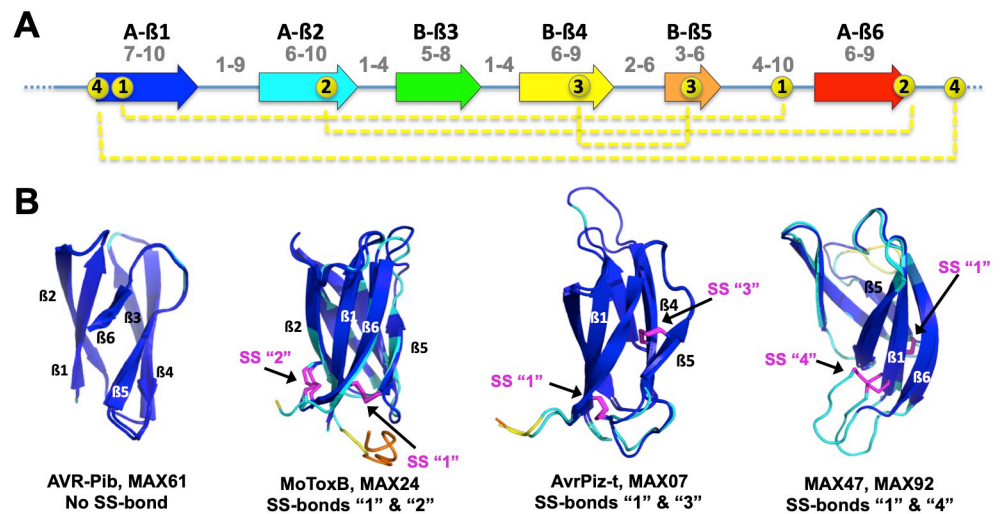
Fig 5. MAX domain structural features. (A) Average size range (indicated in grey) of ß strands (arrows) and connecting loops forming the central ß sandwich of the canonical MAX core, and of the N-ter and C-ter extensions. Average ranges are calculated in S7b Table. The 4 different types of disulfide bonds observed in MAX structures and models are indicated by dotted yellow lines. (B) Four different sets of structural models illustrating the variability of disulfide bond patterns. The AF2 models are colored according to their pLDDT score, by blue for high accuracy (>90), cyan for backbone at good accuracy (> 70), yellow for low confidence (> 50 and < 70) and orange for disordered (< 50). The disulfide bonds are shown in magenta, except for the AVR-Pib structure, which does not have a disulfide bond.

https://doi.org/10.1371/journal.ppat.1012176.g005

of ß1 to a C-terminal cysteine, is well defined in both MAX47 and MAX92. It was not present in any of the previously determined MAX 3D structures that could serve as template and was validated by our NMR structure of MAX47.

N- and C-terminal extensions

Over two-thirds of the 77 AF_MAX models had peptide segments with 15 or more residues extending at one or both ends of the central MAX domain (S7 Table). The length of these extensions can vary among sequences belonging to the same OG cluster, especially for C-terminal extensions (e.g. OG01, OG02 or OG15 clusters in S1a Table). C-terminal extensions were also more numerous and usually longer than N-terminal extensions. They were often modeled by AF2 with well-defined secondary structures, such as additional ß strands extending the ß2ß1ß6 sheet by one or two strands (e.g. MAX08, MAX12, MAX25), or a terminal helix as observed in the model and solution structure of MAX60. In many cases, terminal extensions appeared as unstructured regions that could not be modeled with high confidence by AF2. Long intrinsic disordered regions (IDRs) of more than 30 a.a. [26–28] may have diverse function in bacterial [29,30] and fungal effectors [31,32] and we therefore searched for IDR signatures in MAX effector sequences using ESpritz prediction software [33]. Long IDRs were predicted for ten MAX effectors and were unstructured in six AF_MAX models: in MAX15 (118 a.a.), MAX27 (36 a.a.) and MAX43 (43 a.a) as N-ter extensions, and in MAX28 (42 a.a.), MAX53 (38 a.a.) and MAX78 (43 a.a.) as C-ter extensions. It thus appears that long IDRs are a rare feature among *P. oryzae* MAX effectors, present in less than 8% of all modeled structures. The NMR solution structure of MAX28 validated the unstructured nature of its C-terminal extension.
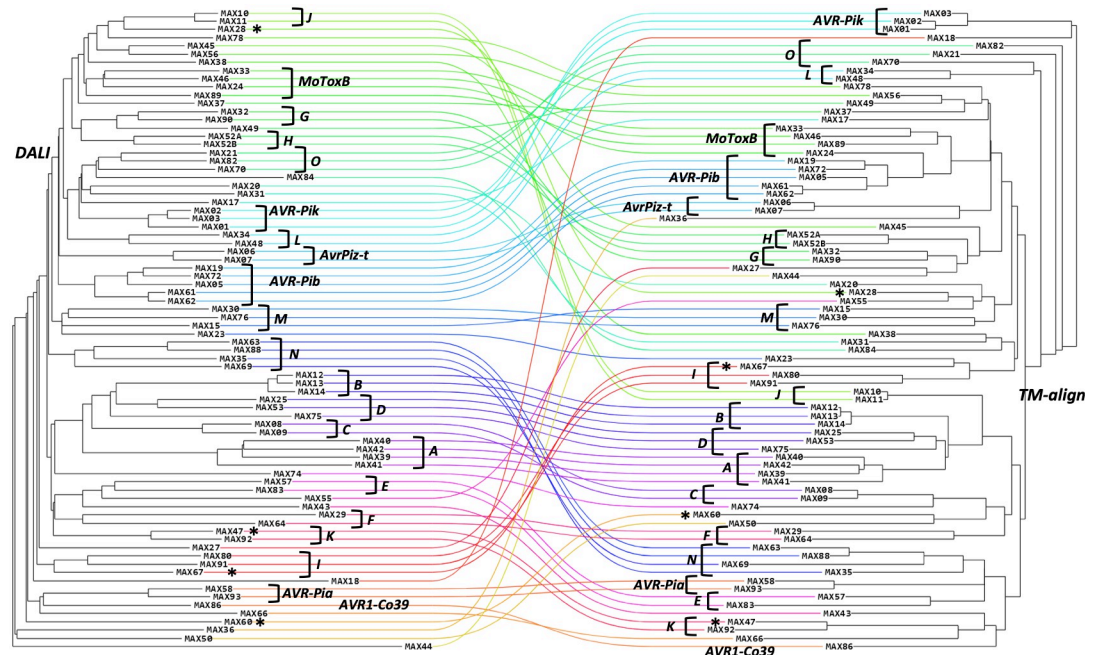
Fig 6. Comparison of the structural similarity trees of MAX effectors based on the Dali Z-score (left) and TM-align TM-score (right) of their superimposed AF models. Unstructured N- and C-terminal regions were removed from the AF_MAX models prior to the analysis. A line of a specific color connects each AF_MAX model in the Dali and TM-align trees. The MAX effectors with an experimental 3D structure are indicated by their name, or a star for the four novel MAX structures. The structural groups to which the MAX effectors were assigned are indicated by brackets and illustrated in Figs 7 and S3.

https://doi.org/10.1371/journal.ppat.1012176.g006

Clustering of MAX effectors in 20 structural groups

Hierarchical clustering of the selected 77 MAX effector models was performed with two protein structure alignment software, Dali [34] and TM-align [35,36] which use different criteria for similarity scoring of superimposed structures. The Dali Z-score relies on secondary structure pairing and is a good estimate of topological conservation while the TM-score is computed for the whole alignment and weights paired residues with low r.m.s.d. more strongly than those that are more distant. When analyzed independently, the structural alignment trees retrieved from these two clustering approaches did not reveal clear sub-families of MAX structures, as shown by the lack of long internal branches in both trees (Fig 6). The comparison between the Dali-based and the TM-align-based trees shows that the MAX effector family is so diverse that even the structural-based trees don't agree on some of the most ancient evolutive relationships between subfamilies. However, comparing evolutive trees obtained by different approaches help to delineate with good confidence consensus subgroups of effectors sharing similar folds. These reliable subgroups are indicated by capital letters or effector names in Fig 6.

In this representation, any bundle of lines of similar colors highlights a possible structural similarity between the models that was common to both clustering methods and that could define a group of MAX models. Each group of at least two models was visually inspected for additional secondary structures that could add to the MAX core, the disulfide bond pattern as well as for specific structural features that fall outside the average ranges reported in Fig 5A. Using this dual clustering method, we defined 15 groups of MAX models sharing common structural features, in addition to the 5 groups of well-established MAX effectors (AVR-Pia,

AVR-Pib, AVR-Pik, AvrPiz-t and MoToxB) (Fig 7). The characteristics of the 20 MAX structural groups are summarized in Fig 7 and illustrated with more details in S3 Fig.

Together, these 20 MAX structural groups comprised about two-thirds of the 77 MAX effectors. Groups A to G gather MAX models possessing major structured elements in addition to the canonical MAX core: groups A and B contain models with 2 extra strands, groups C and D contain models with 1 extra strand, and groups E to G contain models with C-terminal helical extensions after ß6. Group H consists of the MAX52 tandem domains connected by a structured linker. In contrast, groups I to O correspond to plain MAX structures, without other decoration but presenting variations of the MAX fold specific to each group. In addition, the relative orientation and twist of the two β-sheets forming the MAX core β-sandwich can slightly vary, defining different sub-classes of the conserved MAX fold. Some of these groups (I, M, N and O) gather MAX effectors that exhibit strong structural homology that could not be detected at the sequence level, as shown by the comparison of structure *vs* sequence alignment-based similarity trees (S4 Fig). Finally, a total of 23 MAX effectors are singleton, for which no close structural neighbor could be found within our set of 77 AF MAX models. The majority of them (14) consisted of a simple MAX core with unstructured N- and/or C-terminal extensions. Among them was AVR1-CO39 (MAX86). The remaining 9 singletons display diverse structured extensions (S5 Fig), further extending the large structural landscape observed for the MAX effector family in *P. oryza*e.

MAX domains exhibit highly variegated surface properties

The comparison of the molecular surfaces of homologous proteins can highlight common or specific features related to their function. However, size differences or structural elements adding to their common fold can hamper such analysis. We therefore performed a detailed comparative analysis of the surface properties of the *bona fide* MAX effectors by focusing on the MAX core domains extracted from 49 AF_MAX models in which no structured regions interacted with the central core (S7a Table). For this subset of MAX domains, we computed SURF-MAP [37] 2D projections of their molecular envelop, and compared the distribution of the following surface features: exposed secondary structures, electrostatic potential, stickiness and amino-acid polymorphism within the OG cluster to which belongs each representative MAX model (Fig 8).

The electrostatic potential maps (Fig 8D and S2 File) revealed that MAX domain surfaces were more often positively charged than negatively charged or neutral, and that the molecular surfaces can appear entirely positive (e.g. MAX06 (AvrPiz-t), MAX23, MAX34, MAX47) or negative (e.g. MAX10, MAX38, MAX78, MAX86 (AVR1-CO39)), or present intense electrostatic patches (e.g. MAX02 (AVR-PikC), MAX58, MAX62 (AVR-Pib), MAX80). It is well established that positively charged regions in proteins are important for interaction with negatively charged macromolecules, such as nucleic acids and lipopolysaccharides [38], whereas negatively charged protein surfaces can be involved in membrane attachment or DNA mimicking functions [39–41]. In MAX47, we noticed that its unstructured N-terminal extension is rich in aspartic residues, suggesting that it could make transient interactions with the positively charged MAX core in the absence of its cellular target. In MAX62 (AVR-Pib), a surface loop region formed a strong positive patch (Fig 8C), which has been shown to be essential for the avirulence function of AVR-Pib and its nuclear localization in host cells [35]. Interestingly, a similar positive patch was visible on the surface of its structural homolog MAX61 as well as in MAX05 and MAX72 belonging to the same structural group (Fig 7), suggesting that these effectors may also rely on this positive surface loop for their function. Similarly, AvrPiz-t displayed a positively charged surface mostly formed by lysine residues that are required for
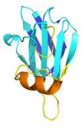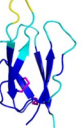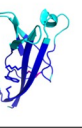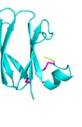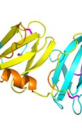
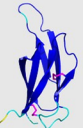| MAX group	representative structure	MAX ID	SS(*)	Structural specificities and extensions	MAX group	representative structure	MAX ID	SS(*)	Structural specificities and extensions
A		*MAX39, MAX40, **MAX41**, MAX42, MAX109*	1	additional ß7ß8 strands flanking one side of the ß2ß1ß6 sheet and C-terminal extensions after ß8 that are in helical conformations	I		***MAX67**, MAX80, MAX91, MAX103*	1	long ß1 strand paired with long ß2 or ß6 strands
B		***MAX12**, MAX13, MAX14*	1	additional ß7ß8 strands flanking one side of the ß2ß1ß6 sheet	J		*MAX10, **MAX11***	1	presence of a short helix in the connecting loop between ß5 and ß6
C		***MAX08**, MAX09*	1	presence of the ß7 additional strand running antiparallel to the ß6	K		***MAX47**, MAX92*	1, 4	disulfide bond "4" joining the N-terminus of ß1 and the C-terminal extension
D		***MAX25**, MAX53, MAX75*	1	presence of the ß7 additional strand running antiparallel to the ß6	L		***MAX34**, MAX48*	1, 3	poorly folded ß2 strand and anti-parallel ß4ß5 strands stabilized by a disulfide bond
E		*MAX57, **MAX83***	1	long ß1 and ß6 strands, C-terminal helical extensions within the MAX core	M		*MAX15, **MAX30**, MAX76, MAX101, MAX111*	1	long ß1 and ß2 strands and their connecting loops
F		*MAX29, **MAX64***	1	C-terminal helical extensions that are stabilized by local residue-specific interactions (disulfide bond or hydrophobic residues)	N		*MAX35, **MAX63**, MAX69, MAX88, MAX102, MAX108*	1	unusually long loop making the connection between strands ß2 and ß3
G		*MAX32, **MAX90***	1, 3	long loop joining the ß1 and ß2 strands and long N- and C-terminal extensions with putative helices of variable length and orientation	O		***MAX21**, MAX82, MAX70, MAX110*	1	Similar ß3ß4ß5 sheet
H		***MAX52A**, **MAX52B***	1, 3	two homologous MAX domains with conserved di-SS pattern	AVR-Pib		*MAX61, **MAX62**, MAX05, MAX19, MAX72*	[1]	MAX61 and MAX62 do not have the conserved SS1
AVR-Pia		*MAX58, **MAX93***	1	ß2 strand interaction with HMA (**)	AvrPiz-t		***MAX06**, MAX07*	1, 3	very similar structures with the main difference for the ß3ß4 connecting loop
AVR-Pik		*MAX01, **MAX02**, MAX03*	1	ß3 strand and N-terminal extension interacting with HMA (**)	MoToxB		*MAX24, **MAX33**, MAX46, MAX89*	1, 2	Structuraly similar to ToxB with disulfide bond "2"

Fig 7. Structural groups identified in P. oryzae MAX effector family. The AF models are colored by their pLDDT scores (see legend of Fig 5). The MAX groups with major structural variations (addition of secondary structures) are listed in the left-hand panels, including the MAX domain duplication of MAX52. The remaining groups, from I to O that do not display additional secondary structural elements but whose MAX core domains have strong structural similarities according to DALI and TM-align (Fig 6) are shown in the right-hand panels. The five groups of well-established MAX effectors (AVR-Pia, AVR-Pib, AVR-Pik, AvrPiz-t and MoToxB) are

shown at the bottom and highlighted in grey color. MAX effectors with ID number above 100 were identified in previous modeling studies in the P. oryzae strain 70–15 as reported in supplemental S8 Table. (*) Type of disulfide bond as defined in Fig 5. (**) from crystallographic structures of complexes.

https://doi.org/10.1371/journal.ppat.1012176.g007

AvrPiz-t avirulence and virulence functions in rice [18]. Inversely, while the surface of MAX86 (AVR1-CO39) is strongly negative, that of MAX93 (AVR-Pia) is neutral, yet both AVR1-CO39 and AVR-Pia interact similarly with the heavy-metal associated (HMA) domain of the rice immune receptor RGA5 [42], respectively (S7 Fig).

Wide divergence was also observed in the surface stickiness of the MAX domains. Surface stickiness is mostly related to surface hydrophobicity and often reflects the propensity of amino acids to be involved in molecular interfaces [43]. Hydrophobic interactions were found to dominate the formation of MAX effector complexes with plant HMA domain binding proteins, as shown by molecular docking [44]. For AVR-Pia, the HMA binding site correlates well with the presence of a large hydrophobic surface patch (Figs 8C, S7B and S7F) and a very sticky ß2 strand, also present in MAX31 (Fig 8E). In AVR1-CO39 as in AVR-Pia, this strand is forming the main interface with the HMA domain of NLR immune receptors (S7A and S7B Fig), yet the hydrophobic patch comprising ß2 appears rather limited in AVR1-CO39 (Figs 8E, S7A and S7E). For the AVR-Pik group (MAX01, MAX02 and MAX03) surface stickiness was high in ß1 and sticky patches were also observed in strands ß3, ß4 and ß6. In all these effectors, the ß3 stickiness could serve in an interaction with strand ß4 of HMA domains, as observed in complexes of different MAX02 effectors with the HMA domain from Pikp-1, Pikh (S7C and S7D Fig, respectively) or from the rice protein OsHIPP19 targeted by AVR-PikF [12–14]. Other MAX effectors (e.g. MAX67, MAX72 and MAX78) possessed a highly sticky ß3 strand that could also associate with the ß-strand of an HMA domain or other type of protein domain. In the crystal structures of the AVR-Pik effectors, the anti-parallel ß1-ß6 strands of the MAX core make hydrophobic contacts with residues in their N-terminal extension which adopts a conserved extended conformation and considerably expands the binding interface with the HMA [12–14]. In MAX structural groups A to D (Fig 7, not included in the present subset of Fig 8), a sticky ß6 strand was often associated with an anti-parallel ß7 extending the MAX core ß-sheet. Similarly, the highly hydrophobic ß6 strand present in MAX33 (MoToxB), MAX45 and MAX50 could interact with target proteins through an antiparallel ß-strand arrangement.

Altogether, this analysis highlights the very variegated surface properties exhibited by the P. oryzae MAX effectors and the difficulties in identifying specific features that could be associated to a common function or interacting partners. In spite of sharing a common fold, these sequence diverse proteins retain extensive diversity at the structural level. On the other hand, sequence conservation was high inside each OG cluster with average conservation scores in strands and loops close to the maximum conservation score of 9 (S7c Table). Only three clusters displayed low conservation scores in ß2, ß3 and ß6 strands for MAX47, in ß2 and the loop joining ß4 to ß5 for MAX63, and in strands ß4 and ß5 for MAX70, which resulted in all cases in increased polymorphism on their surfaces (S6 Fig and S2 File).

Discussion

AF2 outcompetes other strategies for the prediction of MAX effector structures

In this study, we combined experimental structure determination and *in silico* modeling to elucidate the commonalities and variability of the three-dimensional structures in the MAX
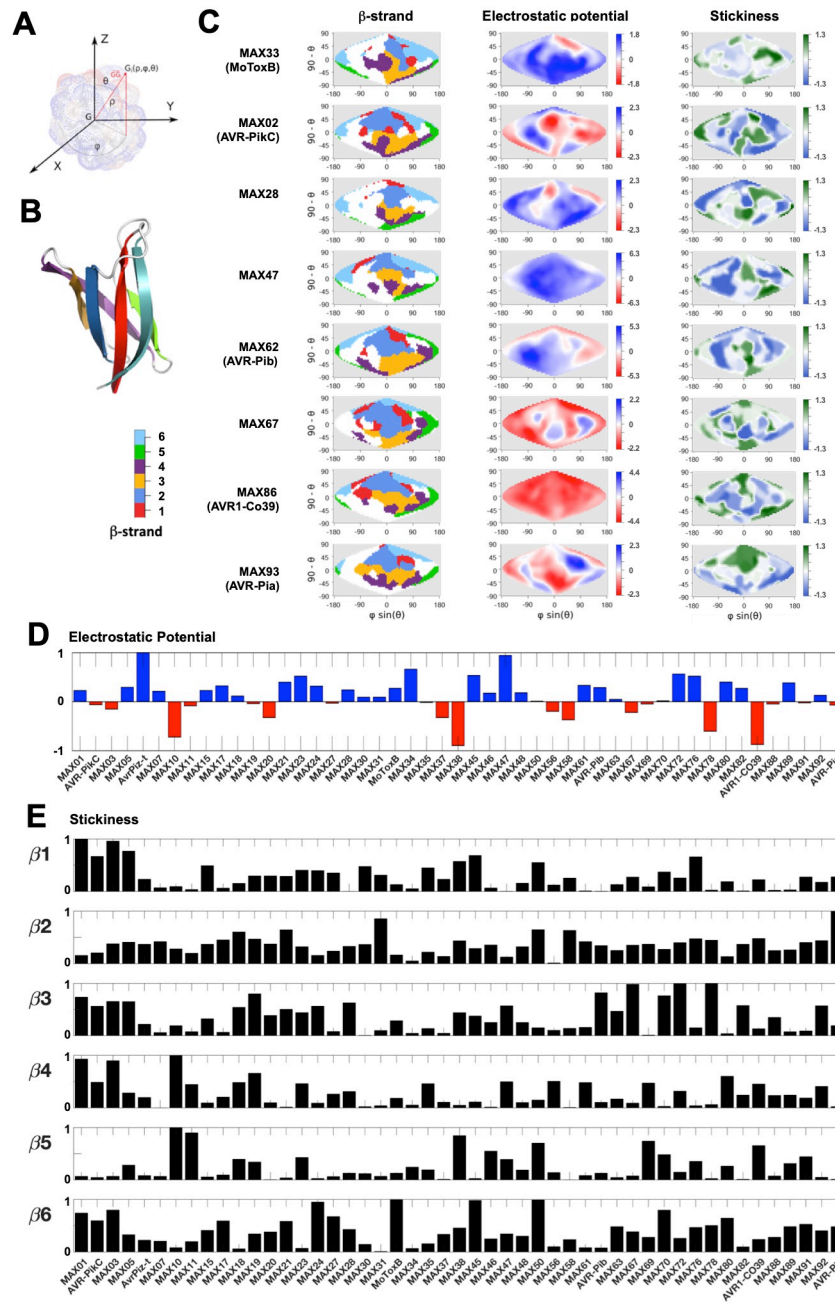
Fig 8. Surface properties of the MAX core domains. Surface properties of MAX core domains computed and represented using SURFMAP. A) Schematic of calculation of the spherical coordinates (from Schweke et al., 2022 [37]). The coordinates of each surface particle G_i is expressed in spherical coordinates (ρ, φ, θ), where ρ represents the distance of the particle G_i to the center of mass G of the protein, φ is the angle between the X axis and the projected vector $\overrightarrow{GGi}$ in the plan ($\overrightarrow{GX}$, $\overrightarrow{GY}$), while θ is the angle between the vector $\overrightarrow{GGi}$ and the Z axis. B) MoToxB structure showing the 6 ß-strands of the MAX core with the color code used for the surface representation in panel C. C) 2D maps of exposed ß-strands, electrostatic potential and stickiness of the molecular surfaces calculated by SURFMAP for MAX domains extracted from AF2 models of MAX effectors with known structure. N- and C-terminal extensions were discarded and the MAX core 3D models were all superimposed to AF_MAX33 (MoToxB) giving a reference frame for the Sanson-Flamsteed 2D projection computed surfaces. The ß-strand maps use the color code given in panel B. The ß1-exposed surface is mostly discontinuous around the ß2 surface that is located at the northern pole of the projection. The continuous ß3 surface lies below the ß2 surface and the discontinuous ß4 surface is found at the bottom of the projection. The ß6 surfaces are found to the west and east of the ß1-ß2 surface areas whereas ß5 surfaces are found on the most eastern areas. The electrostatic potential maps are scaled in the indicated kT/e units and the

stickiness (related to hydrophobicity) scale is that defined by Levy E. D., 2010 [43]. D) Comparison of overall surface electrostatic potentials of MAX core domains, summed over the entire molecular surface and normalized by the highest absolute value calculated for the subset of 49 MAX effector models (S2 File). E) Relative surface stickiness of MAX core ß-strands. Stickiness values were summed for residues forming each of the 6 ß-strands of the MAX core domains and normalized by the highest value calculated for each strand in the subset of 49 MAX core models.

https://doi.org/10.1371/journal.ppat.1012176.g008

effector family of *Pyricularia oryzae*. Using X-ray crystallography we solved the structure of MoToxB and by NMR spectroscopy, we determined the structures of 4 new MAX effectors, MAX28 MAX47, MAX60 and MAX67, bringing to 10 the number of the experimental MAX effector structures from *P. oryzae*. This extended reference set enabled us to evaluate the accuracy with which different modeling techniques can predict the structure of MAX effectors. In a previous work we had used a combination HMM pattern searches and hybrid multiple template modeling to predict the core of MAX effector structures and to select best models according to their TM-pred score [11]. For the selected TM-pred models of MAX28, MAX47, MAX60 and MAX67 the overall structures of the MAX domain were properly modeled, and the observed deviations with their experimentally determined counterparts (TM score) were well predicted by the TM-pred scoring function (Fig 3A). However, the TM-pred models, although limited to the conserved MAX core, deviated substantially from the NMR structures (r.m.s.d ranging from 2.1 to 5.3 Å), highlighting the limits of template-based homology modeling [45], and the need of alternative strategies for the reliable modeling of MAX effectors.

In the present work, we re-explore the structural diversity of the MAX effector family using AlphaFold2 and found that it predicted with very high accuracy the new experimental structures. This was not only the case in the MAX effector core but also in structured regions outside the core. The average r.m.s.d values for superimposed backbone atoms of experimental and modeled structures were between 0,99 and 1,42 Å. Side chain rotamers and disulfide bond conformations were nearly identical, and interactions between secondary structure elements were predicted with high precision. Interestingly, this was true even in the case of MAX28 whose AF2 model had a MAX pLDDT score of only 74. These findings highlighted that AF2 predicts MAX effector structures in a highly reliable and precise manner even at relatively low pLDDT scores. Similar observations were made in other recent studies where experimentally resolved structures of fungal effectors were compared with AF2 predictions, as for example in the case of LARS [46], FOLD [47] and RALPH [48] effectors. Comparison with a published study testing *ab initio* approaches using Rosetta or the two web servers Robetta and QUARK to model MAX effectors with already known structures [49] confirmed that AF2 outcompetes other strategies for the prediction of MAX effector structures.

The repertoire of P. oryzae MAX effectors is still incomplete

Three previous studies [4,5,10] aiming at identifying effector families in *P. oryzae* have used deep-learning methods for systematically modeling the effector candidates from the reference isolate 70–15. Only 11 MAX effectors were detected in the earlier study using TrRosetta [4], 26 and 32 in the studies using AF2 [5,10], respectively. In the present work, a total of 38 validated MAX effectors were found in the *P. oryzae strain 70–15* secretome, among which 18 were not detected in any of the three earlier studies (S8a Table). This high number of false negatives highlights difficulties in model filtering methods based only on structural similarity scores, in contrast to the method used here that relies on HMM-based pattern searches prior to select AF2 models having MAX topology. This method is nevertheless not optimal either since further analysis of rejected MAX OG candidates revealed 12 false negatives, among which 10 had been accurately predicted with a MAX fold in at least one of the three earlier studies and 2

could be modeled with good confidence by the ColabFold *custom* MSA implementation of AlphaFold (S8c Table). This is probably due to the filtering criteria applied for pre-selecting MAX orthogroups (OGs) by the HMM-based searches (protein length < 300 a.a, presence of the conserved SS"1" bond, ratio (>10%) of predicted MAX sequences per OG cluster) [11]. This might also explain why, besides the already known AVR-Pib and its very close homolog MAX61, no other MAX effectors lacking the conserved SS"1" bridge were identified in our study.

Taken together, the four systematic structural modeling studies that have been performed to date on 120 *P. oryzae* genomes have allowed the identification of a total 89 *bona fide* MAX effectors (S8 Table). It is however expected that other MAX effectors have so far escaped detection and that the repertoire of *P. oryzae* MAX effectors is still incomplete. For instance, the recently determined structure of the PWL2 effector from *P. oryzae* [50] revealed the presence of a canonical MAX fold but was not pre-selected as a MAX candidate by our fold-informed protein alignment procedure probably because it contains no cysteine residue and shows no predicted homology with any of the eight experimentally-determined MAX effector templates used in the TM-pred score training data set (S1 Fig). The performance of the HMM-based filtering procedure should improve as more experimental structures are determined and used as training templates.

Commonalities and specificities of MAX structures

The alignment of the 77 structures of MAX OGs we validated by NMR, X-ray crystallography or AF2 modeling allowed the characteristics of the well-preserved structural core to be determined with high precision (S7 Table). It revealed for instance the mean length and variance of the conserved secondary structure elements and allowed the classification of possible cysteine bonding patterns. Only in exceptional cases, canonical structural features were missing or replaced. These cases mostly concerned cysteine bond SS "1" that was lacking in two MAX OGs or beta strand 5 that was replaced by an alpha helix or absent in five MAX OGs. These general features characterizing the MAX core are expected to remain valid as more MAX family members are discovered in *P. oryzae*, as evidenced upon addition of 12 MAX effectors from strain 70–15 to our original set (S7 Table). Comparison of the MAX effector structures also showed that, beyond the well-conserved core, the MAX effector family harbors important structural diversity. Major distinctive features are additional structured regions in the N- or C-termini extending the MAX core by one or two additional strands or forming helical extensions. These regions presumably act in protein-protein interactions or contribute to overall functionality of the effectors. Notably, our study uncovered a domain duplication event within one of the MAX effector clusters (MAX52). Other dual-domain effectors have been described in recent studies, i.e. the Fol dual-domain effectors (FOLD) [47] and effectors predicted from *Puccinia graminis* [5]. The discovery of dual-domain effectors, including the domain duplication found in one of the MAX effector cluster, adds to our understanding of the diversity and complexity of fungal effector proteins. These dual-domain effectors likely have evolved to possess multiple functional domains that contribute to their virulence or interaction with host plants. Unstructured extensions could also play an important functional role since long intrinsically disordered regions (IDRs) that lack a stable 3D structure and exhibit conformational flexibility are known to often interact with multiple binding partners and fulfill various functions [51]. However, long unstructured regions are observed in only a few MAX effectors (S7 Table) and IDRs are difficult to precisely predict from the sequence [52].

In addition to contributing to effector function, N- and C-terminal extensions of MAX effectors may have critical roles in protein folding. This hypothesis is supported by studies

where we used high Hydrostatic Pressure (HP) NMR [53,54] to analyze the folding/unfolding of AVR-Pia, AVR-Pib [55] and MAX60 [56]. While the MAX effector core of AVR-Pia and AVR-Pib folded similarly around a ß3ß4 intermediate, MAX60 had an early folding intermediate formed by ß1, ß6 and the C-terminal helix, a specific extension of this MAX effector. Mutants lacking this helix were not sufficiently stable to be purified. These findings show how additional sequences outside the core can have profound impacts on the folding of MAX effectors.

Structural classification of MAX effectors

In order to better apprehend the structural landscape of the MAX effector family in *P. oryzae*, we attempted the classification of their AlphaFold MAX models. Hierarchical classification of protein structures is not a trivial task and depends greatly on the criteria and metrics applied for superimposing and evaluating distances between 3D structures. Here we retrieved consensus clusters of MAX structures by comparing the similarity trees generated by two different structural alignment software Dali and TM-align (Fig 6). Using this dual clustering method we distinguished 20 subfamilies comprising at least two members that displayed common structural variations of the MAX core (Fig 7). Among the 77 MAX models on which this classification was established, 23 could not be classified and remained singleton because of their lack of strong structural homology with any other MAX model in our data set (S7 Table). However, the number of these singletons as well as the contours of the structural groups defined in Fig 7 are bound to change as more MAX effectors are being modeled and used to build more refined structural similarity trees. Indeed, the arborescence of the trees shown in Fig 6 was slightly modified when the 12 supplementary MAX effectors from *P. Oryzae* strain 70–15 were added to our original data set, resulting in the incorporation of some singletons into redefined structural groups (S8 Fig). This was the case for instance for MAX18, MAX28 and MAX55 that could be added to group F, J, and E, respectively.

MAX effectors of Venturia inaequalis are distinct from P. oryzae MAX effectors

V. inaequalis is an ascomycete fungus, in the *Venturiaceae* family, responsible for apple scab disease. Although it is only very distantly related to *P. oryzae*, V. inaequalis has also an extended MAX effector family as revealed by systematic modeling of its effector repertoire [57]. However, none of the *V. inaequalis* MAX effectors fitted into any of the *P. oryzae* MAX subfamilies. Indeed, all *V. inaequalis* MAX effectors present three conserved disulfide bonds, of which one is the canonical SS "1" bond characteristic of MAX effectors of *P. oryzae* and other fungi, while the remaining two were not found in MAX effectors of any other species. Moreover, MAX-like effectors of *V. inaequalis* usually possess a C-terminal helical extension connected to the MAX core domain *via* a specific disulfide bond. This defines these newly discovered MAX-like effectors from *V. inaequalis* as a distinct subfamily with unique sequence and structure features [57]. *V. inaequalis* colonizes the leaf surface by growing below the cuticle, and releases effectors in this sub-cuticular host environment without penetrating the underlying epidermal cells. Due to this specific life style, the function and host targets of the *V. inaequalis* MAX effectors are presumably fundamentally different from those of *P. oryzae*.

Conclusion and perspectives

Structural information tremendously extends the insight, which can be obtained from primary sequence, and expands our understanding of biological processes or evolution to the atomic level. However, corresponding analyses are far from being trivial especially in rapidly evolving

protein families with high sequence diversity, as it is the case of fungal effector proteins. Our study shows that HMM pattern searches associated with AF2 structure modeling provide a robust method for establishing in a comprehensive manner effector families in fungi. In the case of the MAX effector family, the repertoire of representative members identified in *P. oryzae* has increased from 11 to 89 in just a few years thanks to the advance of deep-learning modeling approaches. Most certainly, the current development of increasingly efficient tools for searching databases and predicting 3D structures will further broaden the family portrait of MAX effectors.

The combination of structural modeling and population genomics provides exciting perspectives for accelerated and deepened investigation of the function and molecular evolution of fungal effector proteins [11]. Ongoing and rapid improvements in *in silico* protein-protein interaction analysis, such as improved prediction of the structures of protein complexes [58] or screening of interacting proteins [59], are opening a new era in this field. Recently, a fast method for directly inferring phylogenies from the structural comparison of predicted 3D models was published [60]. It was shown to improve the accuracy of inferred phylogenies when compared to sequence-based similarity trees and is a promising approach for better analyzing the evolutionary relationships between distantly related effector subfamilies. However, for both of these research areas experimental structure determination remains critical, since large parts of fungal effectoromes can still not be modeled with good confidence, while models of protein complexes generally provide limited insight into the details of the binding interface.

Materials and methods

Experimental structures

MAX28 protein expression and purification. Protein expression and purification experimental details for MAX47, MAX60 and MAX67 are available in Lahfa et al., 2022 [61]. We followed essentially the same protocol for producing the ^{15}N-labelled sample of MAX28. However the protein precipitated once the His$_6$-tag was cleaved, therefore we did not remove it to keep the protein soluble. Uniformly labeled ^{15}N MAX28 was expressed in E. coli BL21 (DE3) cells (Invitrogen, Thermo Fisher Scientific, Waltham, USA) from a homemade plasmid pDB-his-CCDB-3C (courtesy of Frederic Allemand, CBS Montpellier, France). Protein expression was carried out in ^{15}NH$_4$Cl (1 g/l) enriched M9 medium. Cells were grown at 37 ˚C until reaching an OD600 = 0.8 and then, expression proceeded overnight at 30 ˚C after induction by addition of 0.3 mM IPTG. Cells were harvested by centrifugation, re-suspended in denaturing buffer (50 mM Tris, 300 mM NaCl, 1 mM DTT (dithiothreitol), 8 M urea, pH 8) and lysed by ultra-sonication. The supernatant containing the unfolded protein was applied to a HisTrap HP 5 ml affinity column (Cytiva, Freiburg im Breisgau, Germany). The His$_6$-tagged protein was eluted in 50 mM Tris, 300 mM NaCl, 1 mM DTT, 8 M urea, pH 8 with an imidazole gradient up to 500 mM. At this step, MAX28 was directly dialyzed against 10 mM Na Phosphate, 2 mM DTT, 150 mM NaCl, pH 6.8 buffer in order to remove imidazole and urea, allowing the refolding of the protein. The MAX28 samples were then concentrated using Amicon Ultra Centrifugal Filter Devices (MW cutoff 3000 Da), (Merck Millipore, Burlington, USA) prior to size exclusion chromatography (SEC) using HiLoad 16/600 Superdex 75 pg column (Cytiva). Fractions containing protein were pooled, concentrated to 0.4 mM and stored at −20˚C. All NMR experiments were carried out at 27˚C on a Bruker Avance III 800 MHz or Bruker Avance III 700 MHz spectrometer, both equipped with 5 mm z-gradient TCI cryoprobe. NMR samples consisted on approximately 0.4 mM ^{15}N-labeled protein dissolved in 10 mM Na-Phosphate buffer (pH 6.8) and 150 mM NaCl with 5% D$_2$O for the lock.

NMR structure determination of MAX28, MAX47, MAX60 and MAX67. [1]H chemical shifts were directly referenced to the methyl resonance of DSS, while ^{15}N chemical shifts were referenced indirectly to the absolute ^{15}N/^{1}H frequency ratio. All NMR spectra were processed with Topspin 3.6 (Bruker) and analyzed with Cindy 2.1 (Padilla, www.cbs.cnrs.fr). Assignments for MAX28, MAX47, MAX60 and MAX67 have been deposited to and are available from the BMRB data bank under the accession entry 34782, 34731, 34730 and 34729, respectively.

The NMR structures were determined from the NMR constraints listed in S4 Table that were obtained as follow. NOE cross-peaks identified on 3D [^{1}H, ^{15}N] NOESY-HSQC (mixing time 150 ms) were assigned through automated NMR structure calculations with CYANA*3* [62,63]. Hydrogen bond restraints were derived using standard criteria on the basis of the amide [1]H / [2]H exchange experiments and NOE data. When identified, the hydrogen bond was enforced using the following restraints: ranges of 1.8–2.0 Å for d(N-H,O), and 2.7–3.0 Å for d (N,O). Dihedral restraints were obtained from TALOS-N [64] analysis of backbone atom chemical shifts for MAX47, MAX60 and MAX67. For the final list of restraints, distance values redundant with covalent geometry were eliminated and disulfide bonds that were consistent with short distances between cysteine residues were added.

A total of 200 three-dimensional structures were generated using the torsion angle dynamics protocol of CYANA*3* from NOEs, hydrogen bonds and disulfide bond restraints (S4 Table). The 20 best structures (based on the final target penalty function values) were minimized with CNS 1.2 according to the RECOORD procedure [65] and analyzed with PRO-CHECK [66]. The rmsds were calculated with MOLMOL [67]. All statistics are given in S4 Table.

The structure coordinates have been deposited at the Protein Data Bank under the following accession codes: MAX28 (PDB_8C8A), MAX47 (PDB_7ZKD), MAX60 (PDB_7ZK0), MAX67 (PDB_7ZJY).

Modeling by AlphaFold

For all orthogroup (OG) clusters of MAX candidates selected in [11] we further filtered out redundant sequences using CD-HIT v4.3 [68] (S1 Table). The representative sequence of each OG cluster was determined to be the sequence sharing the highest sequence identity with a consensus sequence derived from the OG cluster sequence alignment by MAFFT v7.402 [69] (S2 Table). For each OG representative sequence we computed three AF models differing by the way of building multiple sequence alignment (MSA). The MMseqs2 MSA was obtained from the MMseqs2 [70,71] web server as implemented in the ColabFold alpha release (22th july 2021) that used AlphaFold 2.0 version [72]. We also used the version of AlphaFold 2.2.0 that builds MSAs by Jackhmmer on uniclust, mgnify and uniref90 databases. These two implementations used PDB templates. Finally, we replaced the MMseqs2 MSA by a *Custom* MSA that was build from Muscle_v3.8.31 [73] by inserting (-profile option) the query sequence on top of a previously computed MSA, termed ß1ß4_MSA. The ß1ß4_MSA was build from a Muscle alignment of the OG sequences (S1 Table) by filtering out those having the two flanking cysteine residues in the ß1 and in the loop between ß4 and ß5 strands not correctly aligned to the 8 3D template sequences. The ß1ß4_MSA was further processed by truncating the sequences by eliminating residues (-2 included) before and (+2 included) after the first and last aligned cysteine residues, respectively, and filtering out for redundant sequences by CD-HIT, giving a total size of 247 aligned sequences (S6 Table). For each query, the consistency of the *Custom* MSA was determined by checking the correct alignment of the cysteine residues in the query and in the appended ß1ß4_MSA. When consistent the *Custom* MSA was

converted to a3m format by the reformat.pl script [74] and directly used as input in the Colab-Fold implementation of AlphaFold 2.0 calculations that was setup without the use of PDB templates. *Custom* MSAs could not be built for OG61 and OG62 from absence of cysteine residues in their primary sequences and were not consistent for OG15, OG27, OG71, OG81, OG85 and OG92. For each query the quality of the 15 generated models was assessed by the pLDDT overall score [75]. The correct MAX topology was verified by visual inspection (Pymol v.1.6 Delano 2002). For models having the MAX topology, a MAX pLDDT score that was an average score of residues in the MAX core domain (including residues from ß1 to ß6) was calculated. For each query, the best AF2 MAX model was selected when the MAX pLDDT score was above 60. Sequence information, 3D visualization and PDB files downloading of the complete collection of validated MAX AlphaFold models are accessible online via the webpage https://pat.cbs.cnrs.fr/max and provide the 77 AF2 models in S1 File. The TM-pred models and associated TM-pred scores for the 94 representative sequences of the orthogroup clusters selected in our previous study [11] can also be accessed through the online table https://pat.cbs.cnrs.fr/magmax/model.

Structural alignment and clustering

A standalone implementation of DaliLite.v5 [76] was used for this work. For the all-to-all clustering by Dali we first discarded *unstructured* stretches in each model. For this, the *structured* domain of each model was defined by taking the STRIDE [77] output, and filtering for the first residue in the first and last residue in the last secondary structure (helix or strand), respectively (S7 Table). The model of MAX52 was split in two chains A and B each containing a MAX domain. All these *structured* domains were used for clustering with Dali Z-scores excluding *de-facto* unresolved protein regions without loosing important structural information.

TM-align scoring and side-by-side plot with Dali Z-score tree

The distance between each pair of AF2 models that were used for Dali clustering was estimated by the TM-score obtained from TM-align after pairwise model superposition. A classification tree was then inferred from these pairwise distances using FastME v2 [78]. Finally, Fig 6 was obtained by joining identical models in the FastME tree and in the Dali tree, respectively, by a line of the same color.

Surface properties of MAX core domains

A subset of 49 MAX effector AF2 models, each consisting of a MAX core domain and optional N- and/or C- *unstructured* extensions (S7 Table) was defined by discarding AF2 models having N- and/or C-terminal *structured* extensions (listed in the groups A to H in Figs 7 and S4). All MAX core domains of the 49 AF2 models were superimposed to the reference MoToxB structure with their ß1 strand vertically aligned to the Z Cartesian axe giving a reference frame for the Sanson-Flamsteed 2D projection computed using SURFMAP [37]. Their surface properties including stickiness and electrostatics (APBS) [79] were computed by SURFMAP and are given in S2 File. The temperature factor column of the PDB files was used to encode the color of the exposed surface of the six ß-strands, from 1 to 6, respectively. The sum of the surface stickiness positive values of each individual ß-strand was computed by filtering SURFMAP surface stickiness output and are reported in Fig 8E. The amino-acid conservation scores given for each OG cluster to which belongs each representative MAX model were used to color encode the surface from white for high conservation score of 9, light blue colors for intermediate conservation scores (from 8 to 6), sky-blue for low conservation score of 5 and darker blue colors indicating highly polymorphic positions with conservation scores of 4 and below.

Supporting information

S1 Fig. Template-based modeling of MAX effector sequences.
(PDF)

S2 Fig. NMR structures and AF models.
(PDF)

S3 Fig. Groups of AlphaFold models.
(PDF)

S4 Fig. Comparison of the similarity trees of MAX effectors based on the Dali Z-score (left) and sequence alignment (right) of the AF MAX models.
(PDF)

S5 Fig. Singletons MAX effectors with structured extensions.
(PDF)

S6 Fig. Amino-acid polymorphism mapped on the surface of AF MAX models.
(PDF)

S7 Fig. Comparison of crystal structures and interaction surfaces of MAX-HMA complexes.
(PDF)

S8 Fig. Dali Z-score clustering Heatmap.
(PDF)

S1 Table. List of Pyricularia (syn. Magnaporthe) oryzae entries in the 94 MAX effector orthogroup (OG) clusters.
(XLSX)

S2 Table. Representative sequences of the 94 MAX effector orthogroup clusters.
(XLSX)

S3 Table. Report on structural studies of 10 putative MAX effectors (OG proteins).
(XLSX)

S4 Table. Refinement statistics of MAX effector NMR structures and AF model superimposition.
(PDF)

S5 Table. Summary of AlphaFold2 modeling statistics and list of validated MAX effectors.
(XLSX)

S6 Table. ß1ß4 MSA.
(XLSX)

S7 Table. (a) Structural features of 89 AF-modeled structures of validated MAX effector, (b) statistics and (c) conservation scores.
(XLSX)

S8 Table. Comparison of *P. oryzae* MAX effector candidates in the present (a) and previous modeling studies (b and c).
(XLSX)

S1 Supplementary Methods. Crystallographic structure determination of MoToxB.
(PDF)

S1 File. Zipped folder of AF MAX models.
(ZIP)

S2 File. Zipped folder of AF MAX model surfaces.
(ZIP)

Acknowledgments

Author Contributions

Conceptualization: Stella Cesari, Thomas Kroj, André Padilla.

Data curation: Mounia Lahfa, Philippe Barthe, Pierre Gladieux, Nathalie Declerck.

Formal analysis: Philippe Barthe, Karine de Guillen, Mouna Raji, Marie Le Naour—Vernet.

Funding acquisition: Stella Cesari, Pierre Gladieux, André Padilla.

Investigation: Karine de Guillen, Mouna Raji, Marie Le Naour—Vernet, François Hoh, Pierre Gladieux.

Project administration: Stella Cesari.

Resources: Pierre Gladieux.

Software: Philippe Barthe, Jérôme Gracy.

Supervision: Stella Cesari, Nathalie Declerck, André Padilla.

Validation: Stella Cesari.

Writing – original draft: Mounia Lahfa, Christian Roumestand, Nathalie Declerck, André Padilla.

Writing – review & editing: Stella Cesari, Thomas Kroj, Pierre Gladieux, Nathalie Declerck, André Padilla.

References

1. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal Effectors and Plant Susceptibility. Annu Rev Plant Biol. 2015; 66: 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623 PMID: 25923844

2. Sperschneider J, Williams AH, Hane JK, Singh KB, Taylor JM. Evaluation of Secretion Prediction Highlights Differing Approaches Needed for Oomycete and Fungal Effectors. Front Plant Sci. 2015; 6. https://doi.org/10.3389/fpls.2015.01168 PMID: 26779196

3. Sperschneider J, Dodds PN. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. Mol Plant Microbe Interact. 2022; 35: 146–156. https://doi.org/10.1094/MPMI-08-21-0201-R PMID: 34698534

4. Seong K, Krasileva KV. Computational Structural Genomics Unravels Common Folds and Novel Families in the Secretome of Fungal Phytopathogen *Magnaporthe oryzae*. Mol Plant Microbe Interact. 2021; 34: 1267–1280. https://doi.org/10.1094/MPMI-03-21-0071-R PMID: 34415195

5. Seong K, Krasileva KV. Prediction of effector protein structures from fungal phytopathogens enables evolutionary analyses. Nat Microbiol. 2023; 8: 174–187. https://doi.org/10.1038/s41564-022-01287-6 PMID: 36604508

6. Derbyshire MC, Raffaele S. Surface frustration re-patterning underlies the structural landscape and evolvability of fungal orphan candidate effectors. Genomics; 2023 Jan. https://doi.org/10.1038/s41467-023-40949-9 PMID: 37640704

7. de Guillen K, Ortiz-Vallejo D, Gracy J, Fournier E, Kroj T, Padilla A. Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi. PLOS Pathog. 2015; 11: e1005228. https://doi.org/10.1371/journal.ppat.1005228 PMID: 26506000

8. Fernandez J, Orth K. Rise of a Cereal Killer: The Biology of Magnaporthe oryzae Biotrophic Growth. Trends Microbiol. 2018. https://doi.org/10.1016/j.tim.2017.12.007 PMID: 29395728

9. Fernandez J. The Phantom Menace: latest findings on effector biology in the rice blast fungus. aBIO-TECH. 2023; 4: 140–154. https://doi.org/10.1007/s42994-023-00099-4 PMID: 37581025

10. Yan X, Tang B, Ryder LS, MacLean D, Were VM, Eseola Bisola A, et al. The transcriptional landscape of plant infection by the rice blast fungus Magnaporthe oryzae reveals distinct families of temporally co-regulated and structurally conserved effectors. Plant Cell. 2023; 35: 1360–1385. https://doi.org/10.1093/plcell/koad036 PMID: 36808541

11. Le Naour—Vernet M, Charriat F, Gracy J, Cros-Arteil S, Ravel S, Veillet F, et al. Adaptive evolution in virulence effectors of the rice blast fungus Pyricularia oryzae. Seidl MF, editor. PLOS Pathog. 2023; 19: e1011294. https://doi.org/10.1371/journal.ppat.1011294 PMID: 37695773

12. Maqbool A, Saitoh H, Franceschetti M, Stevenson CEM, Uemura A, Kanzaki H, et al. Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. eLife. 2015; 4: e08709. https://doi.org/10.7554/eLife.08709 PMID: 26304198

13. De la Concepcion JC, Franceschetti M, Maqbool A, Saitoh H, Terauchi R, Kamoun S, et al. Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen. Nat Plants. 2018. https://doi.org/10.1038/s41477-018-0194-x PMID: 29988155

14. Maidment JHR, Franceschetti M, Maqbool A, Saitoh H, Jantasuriyarat C, Kamoun S, et al. Multiple variants of the fungal effector AVR-Pik bind the HMA domain of the rice protein OsHIPP19, providing a foundation to engineer plant defense. J Biol Chem. 2021; 296: 100371. https://doi.org/10.1016/j.jbc.2021.100371 PMID: 33548226

15. Zhang S, Wang L, Wu W, He L, Yang X, Pan Q. Function and evolution of Magnaporthe oryzae avirulence gene AvrPib responding to the rice blast resistance gene Pib. Sci Rep. 2015; 5: 11642. https://doi.org/10.1038/srep11642 PMID: 26109439

16. Oikawa K, Fujisaki K, Shimizu M, Takeda T, Saitoh H, Hirabuchi A, et al. The blast pathogen effector AVR-Pik binds and stabilizes rice heavy metal-associated (HMA) proteins to co-opt their function in immunity. Plant Biology; 2020 Dec. https://doi.org/10.1101/2020.12.01.406389

17. Park CH, Shirsekar G, Bellizzi M, Chen S, Songkumarn P, Xie X, et al. The E3 Ligase APIP10 Connects the Effector AvrPiz-t to the NLR Receptor Piz-t in Rice. Dinesh-Kumar SP, editor. PLOS Pathog. 2016; 12: e1005529. https://doi.org/10.1371/journal.ppat.1005529 PMID: 27031246

18. Bai P, Park C, Shirsekar G, Songkumarn P, Bellizzi M, Wang G. Role of lysine residues of the *Magnaporthe oryzae* effector AvrPiz-t in effector- and PAMP-triggered immunity. Mol Plant Pathol. 2019; 20: 599–608. https://doi.org/10.1111/mpp.12779 PMID: 30548752

19. Rozano L, Jones DAB, Hane JK, Mancera RL. Template-Based Modelling of the Structure of Fungal Effector Proteins. Mol Biotechnol. 2023. https://doi.org/10.1007/s12033-023-00703-4 PMID: 36940017

20. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596: 583–589. https://doi.org/10.1038/s41586-021-03819-2 PMID: 34265844

21. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature. 2020; 577: 706–710. https://doi.org/10.1038/s41586-019-1923-7 PMID: 31942072

22. Taylor WR. Residual colours: a proposal for aminochromography. Protein Eng Des Sel. 1997; 10: 743–746. https://doi.org/10.1093/protein/10.7.743 PMID: 9342138

23. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, et al. Template-based protein structure modeling using the RaptorX web server. Nat Protoc. 2012; 7: 1511–1522. https://doi.org/10.1038/nprot.2012.085 PMID: 22814390

24. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021; 373: 871–876. https://doi.org/10.1126/science.abj8754 PMID: 34282049

25. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science. 2023; 379: 1123–1130. https://doi.org/10.1126/science.ade2574 PMID: 36927031

26. Van Der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of Intrinsically Disordered Regions and Proteins. Chem Rev. 2014; 114: 6589–6631. https://doi.org/10.1021/cr400525m PMID: 24773235

27. Liu Y, Chen S, Wang X, Liu B. Identification of Intrinsically Disordered Proteins and Regions by Length-Dependent Predictors Based on Conditional Random Fields. Mol Ther—Nucleic Acids. 2019; 17: 396–404. https://doi.org/10.1016/j.omtn.2019.06.004 PMID: 31307006

28. Monzon AM, Necci M, Quaglia F, Walsh I, Zanotti G, Piovesan D, et al. Experimentally Determined Long Intrinsically Disordered Protein Regions Are Now Abundant in the Protein Data Bank. Int J Mol Sci. 2020; 21: 4496. https://doi.org/10.3390/ijms21124496 PMID: 32599863

29. Marín M, Uversky VN, Ott T. Intrinsic Disorder in Pathogen Effectors: Protein Flexibility as an Evolutionary Hallmark in a Molecular Arms Race. Plant Cell. 2013; 25: 3153–3157. https://doi.org/10.1105/tpc.113.116319 PMID: 24038649

30. Sun H, Zhu X, Li C, Ma Z, Han X, Luo Y, et al. Xanthomonas effector XopR hijacks host actin cytoskeleton via complex coacervation. Nat Commun. 2021; 12: 4064. https://doi.org/10.1038/s41467-021-24375-3 PMID: 34210966

31. Aparicio Chacón MV, Van Dingenen J, Goormachtig S. Characterization of Arbuscular Mycorrhizal Effector Proteins. Int J Mol Sci. 2023; 24: 9125. https://doi.org/10.3390/ijms24119125 PMID: 37298075

32. Jaswal R, Rajarammohan S, Dubey H, Kiran K, Rawal H, Sonah H, et al. Intrinsically Disordered Kiwellin Protein-Like Effectors Target Plant Chloroplasts and are Extensively Present in Rust Fungi. Mol Biotechnol. 2023. https://doi.org/10.1007/s12033-023-00717-y PMID: 37000361

33. Walsh I, Martin AJM, Di Domenico T, Tosatto SCE. ESpritz: accurate and fast prediction of protein disorder. Bioinformatics. 2012; 28: 503–509. https://doi.org/10.1093/bioinformatics/btr682 PMID: 22190692

34. Holm L. DALI and the persistence of protein shape. Protein Sci. 2020; 29: 128–140. https://doi.org/10.1002/pro.3749 PMID: 31606894

35. Xu J, Zhang Y. How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics. 2010; 26: 889–895. https://doi.org/10.1093/bioinformatics/btq066 PMID: 20164152

36. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005; 33: 2302–2309. https://doi.org/10.1093/nar/gki524 PMID: 15849316

37. Schweke H, Mucchielli M-H, Chevrollier N, Gosset S, Lopes A. SURFMAP: A Software for Mapping in Two Dimensions Protein Surface Features. J Chem Inf Model. 2022; 62: 1595–1601. https://doi.org/10.1021/acs.jcim.1c01269 PMID: 35349266

38. Giraldo MC, Valent B. Filamentous plant pathogen effectors in action. Nat Rev Microbiol. 2013; 11: 800–814. https://doi.org/10.1038/nrmicro3119 PMID: 24129511

39. Wang H, Chou C, Hsu K, Lee C, Wang AH -J. New paradigm of functional regulation by DNA mimic proteins: Recent updates. IUBMB Life. 2019; 71: 539–548. https://doi.org/10.1002/iub.1992 PMID: 30578665

40. Dennehy R, Duggan N, Dignam S, McCormack S, Dillon E, Molony J, et al. Protein with negative surface charge distribution, Bnr1, shows characteristics of a DNA-mimic protein and may be involved in the adaptation of *Burkholderia cenocepacia*. MicrobiologyOpen. 2022; 11. https://doi.org/10.1002/mbo3.1264 PMID: 35212475

41. Goldenberg NM, Steinberg BE. Surface Charge: A Key Determinant of Protein Localization and Function. Cancer Res. 2010; 70: 1277–1280. https://doi.org/10.1158/0008-5472.CAN-09-2905 PMID: 20124473

42. Guo L, Zhang Y, Ma M, Liu Q, Zhang Y, Peng Y, et al. Crystallization of the rice immune receptor RGA5A_S with the rice blast fungus effector AVR1-CO39 prepared *via* mixture and tandem strategies. Acta Crystallogr Sect F Struct Biol Commun. 2018; 74: 262–267. https://doi.org/10.1107/S2053230X18003618 PMID: 29633975

43. Levy ED. A Simple Definition of Structural Regions in Proteins and Its Use in Analyzing Interface Evolution. J Mol Biol. 2010; 403: 660–670. https://doi.org/10.1016/j.jmb.2010.09.028 PMID: 20868694

44. Rozano L, Hane JK, Mancera RL. The Molecular Docking of MAX Fungal Effectors with Plant HMA Domain-Binding Proteins. Int J Mol Sci. 2023; 24: 15239. https://doi.org/10.3390/ijms242015239 PMID: 37894919

45. Fiser A. Template-Based Protein Structure Modeling. In: Fenyö D, editor. Computational Biology. Totowa, NJ: Humana Press; 2010. pp. 73–94. https://doi.org/10.1007/978-1-60761-842-3_6 PMID: 20835794

46. Lazar N, Mesarich CH, Petit-Houdenot Y, Talbi N, Li De La Sierra-Gallay I, Zélie E, et al. A new family of structurally conserved fungal effectors displays epistatic interactions with plant resistance proteins. Kobe B, editor. PLOS Pathog. 2022; 18: e1010664. https://doi.org/10.1371/journal.ppat.1010664 PMID: 35793393

47. Yu DS, Outram MA, Smith A, McCombe CL, Khambalkar PB, Rima SA, et al. The structural repertoire of Fusarium oxysporum f. sp. lycopersici effectors revealed by experimental and computational studies. elife; 2023 Jul. https://doi.org/10.7554/eLife.89280.1

48. Cao Y, Kümmel F, Logemann E, Gebauer JM, Lawson AW, Yu D, et al. Structural polymorphisms within a common powdery mildew effector scaffold as a driver of coevolution with cereal immune receptors. Proc Natl Acad Sci. 2023; 120: e2307604120. https://doi.org/10.1073/pnas.2307604120 PMID: 37523523

49. Rozano L, Mukuka YM, Hane JK, Mancera RL. Ab Initio Modelling of the Structure of ToxA-like and MAX Fungal Effector Proteins. Int J Mol Sci. 2023; 24: 6262. https://doi.org/10.3390/ijms24076262 PMID: 37047233

50. Zdrzałek R, Xi Y, Langner T, Bentham AR, Petit-Houdenot Y, De La Concepcion JC, et al. Bioengineering a plant NLR immune receptor with a robust binding interface towards a conserved fungal pathogen effector. Plant Biology; 2024 Jan. https://doi.org/10.1101/2024.01.20.576400

51. Bondos SE, Dunker AK, Uversky VN. On the roles of intrinsically disordered proteins and regions in cell communication and signaling. Cell Commun Signal. 2021; 19: 88, s12964-021-00774–3. https://doi.org/10.1186/s12964-021-00774-3 PMID: 34461937

52. Zhao B, Kurgan L. Deep learning in prediction of intrinsic disorder in proteins. Comput Struct Biotechnol J. 2022; 20: 1286–1294. https://doi.org/10.1016/j.csbj.2022.03.003 PMID: 35356546

53. Dubois C, Herrada I, Barthe P, Roumestand C. Combining High-Pressure Perturbation with NMR Spectroscopy for a Structural and Dynamical Characterization of Protein Folding Pathways. Molecules. 2020; 25: 5551. https://doi.org/10.3390/molecules25235551 PMID: 33256081

54. Roche J, Dellarole M, Royer CA, Roumestand C. Exploring the Protein Folding Pathway with High-Pressure NMR: Steady-State and Kinetics Studies. In: Akasaka K, Matsuki H, editors. High Pressure Bioscience. Dordrecht: Springer Netherlands; 2015. pp. 261–278. http://link.springer.com/10.1007/978-94-017-9918-8_13.

55. Dubois C, Lahfa M, Pissarra J, de Guillen K, Barthe P, Kroj T, et al. Combining High-Pressure NMR and Geometrical Sampling to Obtain a Full Topological Description of Protein Folding Landscapes: Application to the Folding of Two MAX Effectors from Magnaporthe oryzae. Int J Mol Sci. 2022; 23: 5461. https://doi.org/10.3390/ijms23105461 PMID: 35628267

56. Lahfa M, Mouhand A, de Guillen K, Barthe P, Kroj T, Roumestand C. Does a Similar 3D Structure Mean a Similar Folding Pathway? The Presence of a C-Terminal α-Helical Extension in the 3D Structure of MAX60 Drastically Changes the Folding Pathway Described for Other MAX-Effectors from Magnaporthe oryzae. Molecules. 2023; 28: 6068. https://doi.org/10.3390/molecules28166068 PMID: 37630320

57. Rocafort M, Bowen JK, Hassing B, Cox MP, McGreal B, de la Rosa S, et al. The Venturia inaequalis effector repertoire is dominated by expanded families with predicted structural similarity, but unrelated sequence, to avirulence proteins from other plant-pathogenic fungi. BMC Biol. 2022; 20: 246. https://doi.org/10.1186/s12915-022-01442-9 PMID: 36329441

58. Bryant P, Pozzati G, Elofsson A. Improved prediction of protein-protein interactions using AlphaFold2. Nat Commun. 2022; 13: 1265. https://doi.org/10.1038/s41467-022-28865-w PMID: 35273146

59. Yu D, Chojnowski G, Rosenthal M, Kosinski J. AlphaPulldown—a python package for protein–protein interaction screens using AlphaFold-Multimer. Cowen L, editor. Bioinformatics. 2023; 39: btac749. https://doi.org/10.1093/bioinformatics/btac749 PMID: 36413069

60. Puente-Lelievre C, Malik AJ, Douglas J, Ascher D, Baker M, Allison J, et al. Tertiary-interaction characters enable fast, model-based structural phylogenetics beyond the twilight zone. Evolutionary Biology; 2023 Dec. https://doi.org/10.1101/2023.12.12.571181

61. Lahfa M, Padilla A, De Guillen K, Pissarra J, Raji M, Cesari S, et al. 1H, 13C, 15 N backbone and side-chain NMR assignments for three MAX effectors from Magnaporthe oryzae. Biomol NMR Assign. 2022; 16: 305–309. https://doi.org/10.1007/s12104-022-10095-2 PMID: 35657473

62. Güntert P. Automated NMR structure calculation with CYANA. Methods Mol Biol Clifton NJ. 2004; 278: 353–378. https://doi.org/10.1385/1-59259-809-9:353 PMID: 15318003

63. Güntert P, Buchner L. Combined automated NOE assignment and structure calculation with CYANA. J Biomol NMR. 2015. https://doi.org/10.1007/s10858-015-9924-9 PMID: 25801209

64. Shen Y, Bax A. Protein Structural Information Derived from NMR Chemical Shift with the Neural Network Program TALOS-N. In: Cartwright H, editor. Artificial Neural Networks. New York, NY: Springer New York; 2015. pp. 17–32. https://doi.org/10.1007/978-1-4939-2239-0_2 PMID: 25502373

65. Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CAEM, Nabuurs SB, et al. RECOORD: A recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagRes-Bank. Proteins Struct Funct Bioinforma. 2005; 59: 662–672. https://doi.org/10.1002/prot.20408 PMID: 15822098

66. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK—a program to check the stereo-chemical quality of protein structures. J Appl Crystallogr. 1993; 26: 283–291. https://doi.org/10.1107/S0021889892009944

67. Koradi R, Billeter M, Wüthrich K. MOLMOL: A program for display and analysis of macromolecular structures. J Mol Graph. 1996; 14: 51–55. https://doi.org/10.1016/0263-7855(96)00009-4 PMID: 8744573

68. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012; 28: 3150–3152. https://doi.org/10.1093/bioinformatics/bts565 PMID: 23060610

69. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013; 30: 772–780. https://doi.org/10.1093/molbev/mst010 PMID: 23329690

70. Steinegger M, Soding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017; 35: 1026–1028. https://doi.org/10.1038/nbt.3988 PMID: 29035372

71. Mirdita M, Steinegger M, Söding J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. Hancock J, editor. Bioinformatics. 2019; 35: 2856–2858. https://doi.org/10.1093/bioinformatics/bty1057 PMID: 30615063

72. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022; 19: 679–682. https://doi.org/10.1038/s41592-022-01488-1 PMID: 35637307

73. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32: 1792–1797. https://doi.org/10.1093/nar/gkh340 PMID: 15034147

74. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods. 2011; 9: 173–175. https://doi.org/10.1038/nmeth.1818 PMID: 22198341

75. Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics. 2013; 29: 2722–2728. https://doi.org/10.1093/bioinformatics/btt473 PMID: 23986568

76. Holm L. Using Dali for Protein Structure Comparison. In: Gáspári Z, editor. Structural Bioinformatics. New York, NY: Springer US; 2020. pp. 29–42. https://doi.org/10.1007/978-1-0716-0270-6_3 PMID: 32006276

77. Heinig M, Frishman D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res. 2004; 32: W500–W502. https://doi.org/10.1093/nar/gkh429 PMID: 15215436

78. Lefort V, Desper R, Gascuel O. FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. Mol Biol Evol. 2015; 32: 2798–2800. https://doi.org/10.1093/molbev/msv150 PMID: 26130081

79. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. Protein Sci. 2018; 27: 112–128. https://doi.org/10.1002/pro.3280 PMID: 28836357