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

* Corresponding author

E-mail: andre.padilla@cbs.cnrs.fr

Keywords: *Pyricularia oryzae*, MAX effectors, NMR, protein structure, AlphaFold

Short title: Structural diversity in MAX effectors

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.

60 # Introduction

61 Fungal plant pathogens secrete small proteins, called effectors, which promote
62 disease by targeting cellular processes in the host plant. There are hundreds of
63 predicted effectors in the genomes of plant pathogenic fungi that are usually
64 identified by their secretion signal and other characteristic features such as cysteine
65 enrichment [1–3]. Some effectors are of particular interest since they constitute
66 avirulence (AVR) factors that are detected by plant immune systems and render
67 crops resistant to severe diseases. Fungal effectors usually show no amino-acid
68 sequence homology to known proteins or protein domains, and consequently their
69 biological function cannot be inferred from systematic *in silico* analysis (such as
70 domain searches) but must be elucidated on a case-by-case basis. As a huge and
71 growing number of putative effectors is being discovered, deciphering the functions
72 and adaptive evolution of fungal effectors requires robust predictive tools for
73 analyzing protein sequences and prioritizing candidates.

74 Recently, a combination of primary sequence pattern searches and structural
75 modeling resulted in a major breakthrough in effector biology by revealing that fungal
76 effector repertoires are actually dominated by a limited number of families sharing
77 common structures despite extensive sequence variability [4–6]. One such family are
78 the MAX (*Magnaporthe* AVRs and ToxB-like) effectors we identified in *Pyricularia*
79 *oryzae* (synonym: *Magnaporthe oryzae*), the causal agent of blast disease in rice,
80 wheat, and other cereals or grasses [7]. This pathogenic fungus is both a major
81 threat to global food security [8] and a prime experimental model in plant pathology
82 [9,10]. By solving the solution structure of two *P. oryzae* effectors, AVR1-CO39 and
83 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

109 MAX effectors possess N- and C-terminal extensions that could have critical roles,
110 for instance, by establishing specific protein-protein interactions. Examining in more
111 details these extensions, as well as other non-conserved structural features, may
112 thus provide insights into the mechanism by which MAX effectors acquire new
113 virulence capabilities, and allow a more comprehensive classification within the MAX
114 family.

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

131 This work represents the most extensive structural analysis of a fungal effector
132 family of a plant pathogen to date. It also provides valuable knowledge for analyses
133 aimed at elucidating the function of MAX effectors, notably through the prediction of

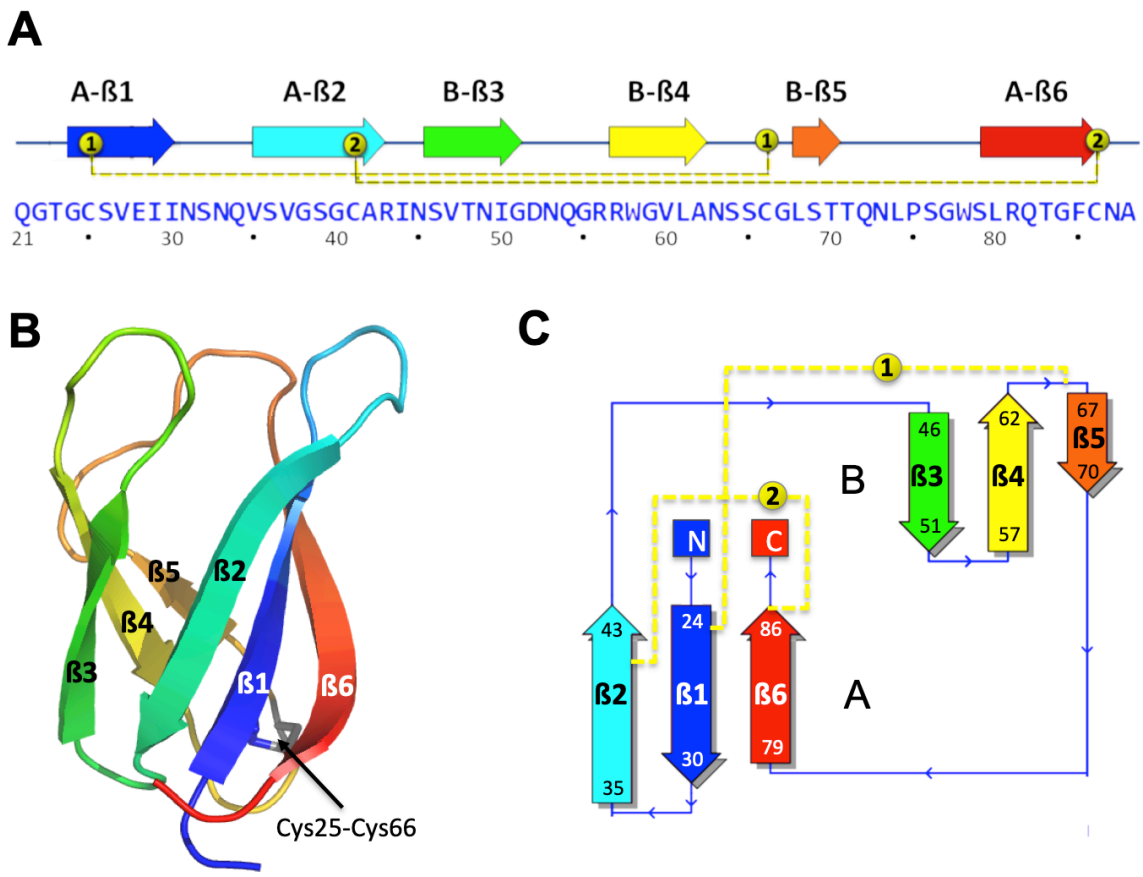
134 interaction sites within the MAX fold that could contribute to targeting host proteins
135 during infection.

136

137 # Results

138 ## *The structure of MoToxB presents the canonical MAX fold*

139 The MAX effector orthogroup OG33 of *P. oryzae* has high protein sequence
140 similarity with the effector ToxB from *P. tritici-repentis* suggesting that the
141 corresponding genes are orthologs [7,11]. We determined the structure of its
142 representative in the Br58 isolate (S2 Table) by X-ray crystallography using
143 molecular replacement at 1.38Å resolution (Supplementary Materials and Methods).
144 The structure confirmed high structural similarity with ToxB from *P. tritici-repentis* and
145 therefore this MAX effector was renamed MoToxB. The structure was also similar to
146 the experimentally determined structures of five other *P. oryzae* MAX effectors that
147 share less than 13% sequence identity with MoToxB (S1 Fig). Like other MAX
148 effectors, MoToxB is structured as a 6-stranded ß-sandwich (ß1 to ß6) of two triple-
149 stranded antiparallel ß-sheets with a ß6ß1ß2-ß3ß4ß5 topology. A disulfide bond that
150 is conserved in almost all other MAX effectors forms a bridge between ß1 and the
151 loop connecting ß4 and ß5. The two cysteines forming this bond are the only
152 residues that are highly conserved in MAX effectors. In MoToxB a second disulfide
153 bond connects ß2 and ß6 (Fig 1).



154

155 ***Fig 1. Structure of the M. oryzae ToxB (MoToxB) MAX effector***

156 *Primary and secondary structure of MoToxB showing the triple-stranded beta-sandwich*
157 *forming the conserved MAX core with the two beta-sheets labeled by A and B, strands*
158 *indicated by arrows and two disulfide bonds in yellow dotted lines. Disulfide bond SS "1" is*
159 *almost strictly conserved in MAX effectors. (B) Cartoon representation of MoToxB crystal*
160 *structure (PDB 6R5J) in rainbow color and the conserved disulfide bond "1" shown by sticks.*
161 *(C) MoToxB topology diagram drawn by PDBsum and colored using the same color scheme*
162 *as in A and B.*

163

164 ## *NMR structures validate template-based modeling of MAX effectors*

165 In our previous analysis of the MAX effector repertoire in *P. oryzae*, we used a
166 combination of Hidden Markov Model (HMM) pattern searches and hybrid multiple
167 Template Modeling (TM) for predicting the 3D structure of the conserved MAX core
168 of each representative sequences of the 94 MAX effector OGs (S2 Table) defined in

169 that study (OGs are provided in S1 Table) [11]. The reliability of the 3D models
170 (referred as TM-pred models) was evaluated by the TM-pred score, which is an
171 estimate of the TM-score that would be observed in a structural alignment of the TM-
172 pred model with the corresponding experimentally resolved structure using TM-align.
173 For these analyses eight experimental structures of MAX effectors were used as
174 templates for homology modeling, and as a training data set for the TM-pred scoring
175 function [11]. Almost 90% of the OG proteins were modeled at high confidence as
176 MAX structures (TM-pred score > 0.6). Only three TM-pred models, those of OG22,
177 OG77 and OG85, had a TM-pred score below 0.5 and were suspected to be no MAX
178 effectors (S1 Fig). All 94 TM-pred models and associated TM-pred scores can be
179 downloaded from the online table https://pat.cbs.cnrs.fr/magmax/model/ (see also
180 Materials and Methods).

181 To deepen insight into MAX effectors and to assess the validity of the
182 predictions, we attempted to resolve the experimental structures of 10 new MAX
183 effector candidates with high expression during the biotrophic stage of infection (S3
184 Table) [11]. Using NMR spectroscopy, we successfully determined the structure of
185 four of them (OG28, OG47, OG60 and OG67), and confirmed that they all have a
186 MAX core fold (referred to as “MAX” instead of “OG clusters” from hereon). The 20
187 best-refined conformers obtained for each of these effectors were superimposed (Fig
188 2A), and the high quality of the NMR structures was supported by the low root mean
189 square deviations (r.m.s.d). The complete structural statistics are given in S4 Table.
190

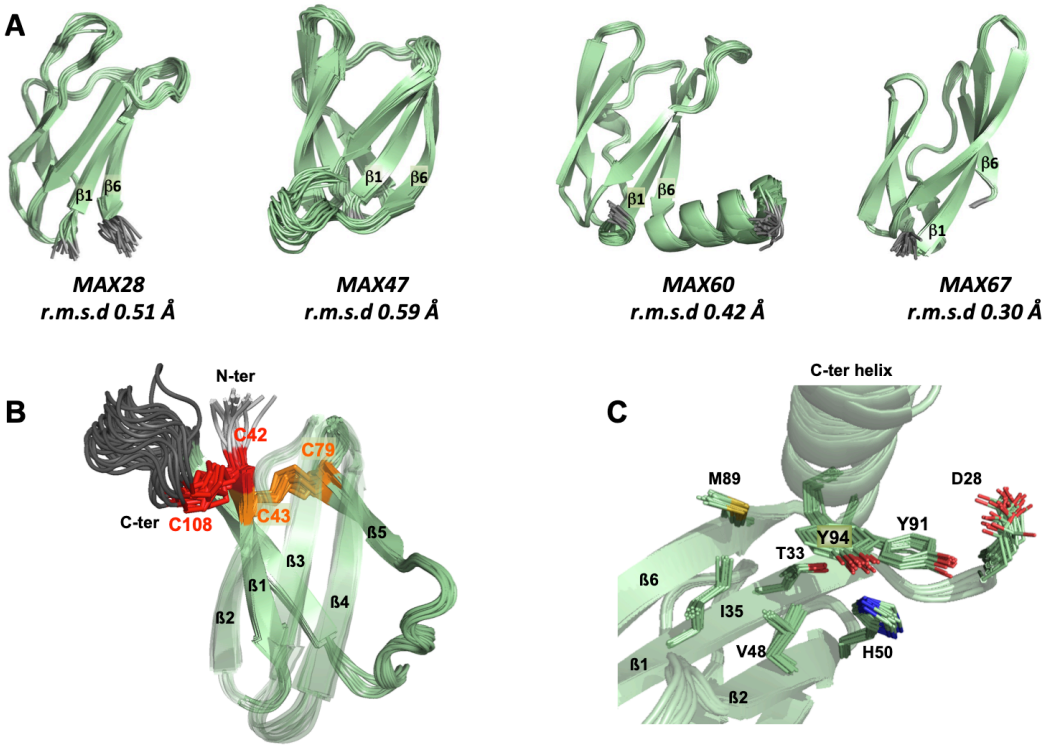


Fig 2. NMR structures of four MAX effectors. *(A) NMR structures of MAX28, MAX47, MAX60 and MAX67 showing the superimposition and the r.m.s.d. of their 20 best conformers. The N- and C-terminal unstructured extensions before ß1 and after ß6, respectively, are not shown, except for the C-ter helix of 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.*

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

207 of the ß5 strand by a helical turn, preventing the corresponding residues from forming
208 a regular ß-sheet with ß4 and ß3.

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

218 MAX60 has a C-terminal extension, which forms a well-defined α-helix that is
219 attached to the structural core by hydrophobic contacts established by the aromatic
220 rings of two tyrosine residues (Y91 and Y94). Nuclear Overhauser Effects (NOEs) in
221 the NMR experiments revealed close contacts between tyrosine Y91 and residues
222 D28 to T33 and H50, and between tyrosine Y94 and residues T33, I35, V48 and M89
223 (Fig 2C).

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

227 The four new NMR structures of MAX effectors were superimposed using TM-
228 align with the corresponding TM-pred models that we previously generated by
229 template homology modeling (Fig 3A). The quality of the models was evaluated by
230 the root-mean-square deviation (r.m.s.d) calculated between the observed and
231 predicted structures and the TM-scores given by TM-align (a value of 1 meaning a

232 perfect match). Comparison of the superimposed backbones showed that the overall
233 MAX fold as well as the relative orientation of the two ß-sheets forming the central ß-
234 sandwich were all well predicted. The prediction was particularly good for MAX28
235 whose MAX domain of the TM-pred model precisely matched the experimental
236 structure (r.m.s.d.=2.11 Å), even for the loops joining the ß-strands. MAX28 was also
237 the effector with the highest estimated TM-pred score (0.75), in remarkably good
238 agreement with the true TM-score (0.74) of the TM-pred model aligned to the NMR
239 structure. Structural predictions of the MAX core were also very good for MAX67,
240 except for the long strands ß1 and ß2 in the NMR structure that could not be
241 accurately modeled. This deviation explains the rather low TM-pred value for TM-
242 pred_MAX67. The models of MAX47 and MAX60 showed also poor definition of
243 certain ß-strands and exhibited strong divergence in connecting loops when
244 compared to the NMR structures. For MAX47, the limited reliability of the TM-pred
245 model was reflected by the low TM-pred score (0.63).
246

247 ***AlphaFold2 reliably predicts MAX effectors core and extensions.***

248 The recent release of Artificial Intelligence (AI)-based protein structure
249 modeling methods have demonstrated the high accuracy with which deep-learning
250 softwares such as AlphaFold2 (AF2) can predict the 3D structure of most structured
251 proteins [19]. To determine the accuracy of AF2 for the prediction of MAX effector
252 structures, we used it to model MAX28, MAX47, MAX60 and MAX67 before the
253 release of their NMR structures in the PDB. We used three different implementations
254 of AF2 to generate the multiple sequence alignment (MSA) from which protein 3D
255 models are predicted: the MMseqs2 and Jackhmmer implementations, which use
256 MSAs generated automatically online or locally and that can include PDB templates,

257 as well as the *Custom* MSA implementation that uses MSAs provided by the user
258 and that was parameterized to not use PDB templates (see Materials and Methods).
259 For each sequence query, the three MSAs were independently submitted to AF2,
260 and the predictive quality of the top-ranked models was assessed thanks to the
261 predicted local distance difference test (pLDDT) [19,20]. The pLDDT score, scaling
262 from 0 to 100, is a residue-level accuracy score computed by AF2 that provides an
263 estimate of the confidence of each residue's predicted position in the protein
264 structure. We considered the average pLDDT score for the overall protein, or only for
265 residues within the predicted MAX core domain, hereafter, called the MAX pLDDT
266 score. The MAX pLDDT score allows better filtering of the AF models, since it
267 provides a quantitative measure of prediction quality focusing on the MAX core
268 domain. For each MAX effector, the AF2 model having the highest MAX pLDDT
269 score was selected and referred as its AF_MAX model (Fig 3).

270 The MAX core domain was predicted with high confidence in AF_MAX47,
271 AF_MAX60 and AF_MAX67 according to their high MAX pLDDT scores, close to or
272 exceeding 90 (Fig 3B and S5 Table). The best models were obtained in all three
273 cases from the Jackhmmer AF2 implementation. A lower confidence score (74.1)
274 was retrieved for the best AF2 model of MAX28, generated with the *Custom* MSA
275 implementation. Nevertheless, AF_MAX28 was very close to the experimental
276 structure of MAX28 according to the average r.m.s.d. value (1.42 Å) calculated from
277 superimposing the MAX core backbone atoms of the NMR conformers. Indeed, for all
278 four MAX effectors, the MAX domain of the best AF2 model displayed side-chain
279 rotamers almost identical to those in the experimental structures and that were within
280 the uncertainty of the NMR approach, i.e. density of NMR-derived constraints (Fig
281 3B).

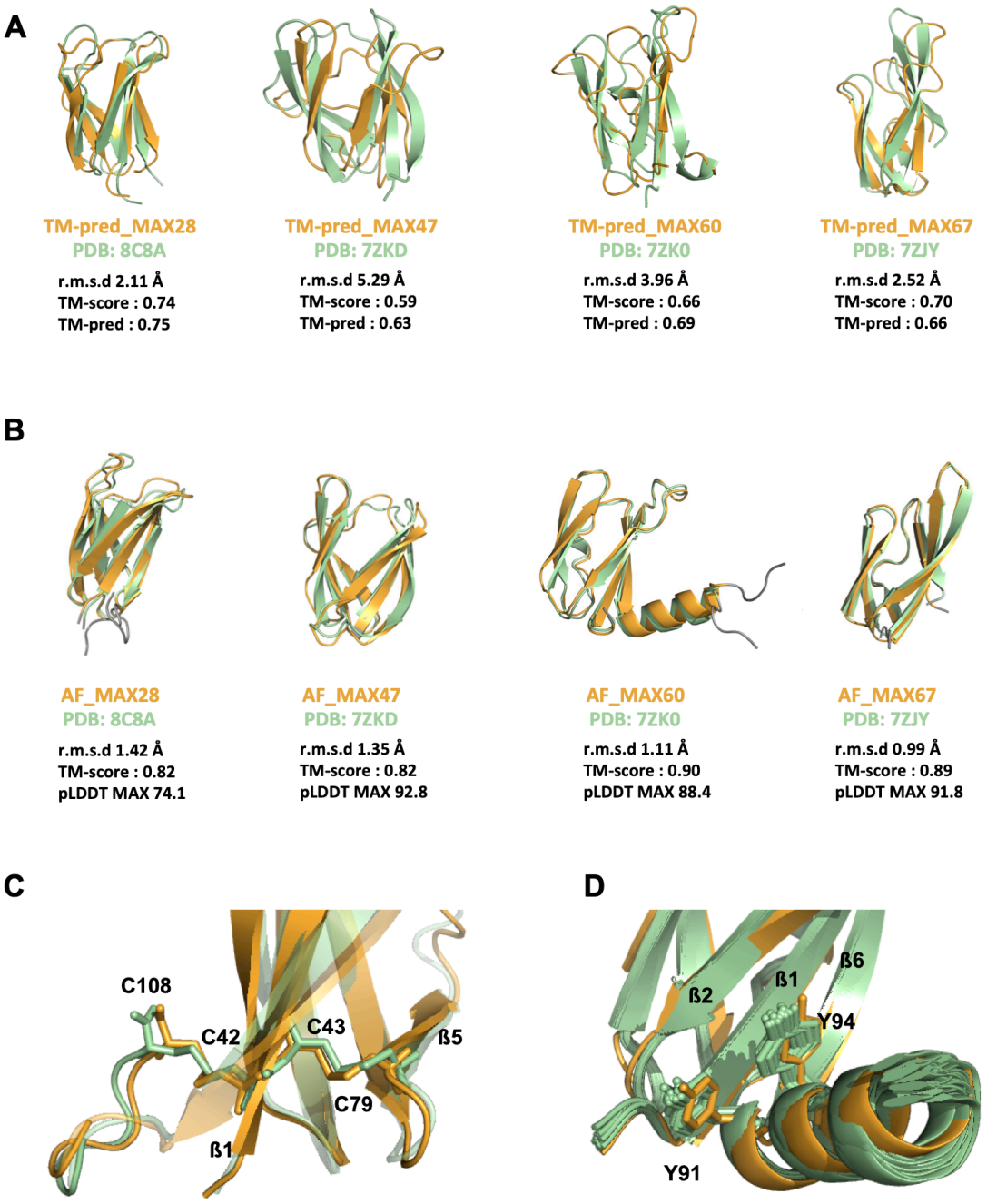


Fig 3. Comparison of newly determined NMR structures of MAX effectors with their TM-pred or AF predicted models. *Superimposition of the four MAX effectors determined in this study by NMR shown by the best model (in green) and of their corresponding 3D models (in orange) predicted by hybrid multiple 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*

292 *corresponding to a perfect match), TM-pred score (a predictive estimate of the TM-score).*
293 *The N- and C-terminal boundaries were set according to the TM-pred models and did not*
294 *include extensions determined in the NMR structures. (B) The r.m.s.d. between backbone*
295 *heavy atoms of the superimposed NMR structure and AF2 models is given for the MAX*
296 *domain only (S5 Table), as well as the predicted local distance difference test (pLDDT) that*
297 *was used to estimate the reliability of the AF2 predictions in the MAX core (MAX pLDDT*
298 *score, S5 Table). (C) View of the two disulfides bonds of MAX47, C42-C108 and C43-C79,*
299 *as observed in the best NMR conformer and in the predicted AF_MAX47 model. (D) Position*
300 *of the two tyrosyl side-chains of Y91 and Y94 in the C-terminal helix of MAX60 in the 20*
301 *NMR conformers and in the AF2 predicted model.*
302
303 AlphaFold2 modeling also succeeded in predicting details within the core
304 domains. The cysteine residue side-chains forming the conserved SS1 disulfide bond
305 were well defined in all four AF_MAX effector models. The same was true for those
306 forming the additional disulfide bond, bridging the ß1 and the C-terminal extension, in
307 the structure of MAX47 (Fig 3C). Another example of consistency between
308 experimental structures and AF2 models was the remarkably well defined position
309 and orientation of the C-terminal helix of MAX60, including the two tyrosine residues
310 whose aromatic side chains stacked over the ß1-ß2-ß6 ß-sheet of the MAX core (Fig
311 3D). For MAX28, both AF2 model and NMR structure were consistent in predicting
312 unstructured N- and C-terminal extensions. Moderate deviations from the
313 experimental structure were only observed for residues in the C-terminus of MAX67
314 (S2 Fig).
315

316 ## *AlphaFold2 validates 77 out of 94 MAX OGs*

317 Given the high quality of the AF2-generated models of MAX effectors, we
318 applied the same AF modeling strategy to all other *P. oryzae* MAX OG
319 representatives. The models were visualized to check the presence of the
320 characteristic MAX core by inspecting the ß-strand topology and the presence of the

321 conserved SS “1” disulfide bond (Pymol. v.1.6; Delano 2002). OG proteins showing
322 significant topological deviations from this canonical MAX fold were discarded. The
323 presence of the short ß5 strand was not used as a filtering criterion. Among the 80
324 AF2 models that matched the canonical MAX structure (Fig 4, S5 Table), 57 had
325 MAX pLDDT scores greater than 80 and 20 had MAX pLDDT scores ranging from 60
326 to 80. Only three OG proteins, OG26, OG73 and OG94, exhibiting a central core
327 compatible with a MAX fold had a MAX pLDDT score below 60 and were not kept in
328 our final selection of 77 validated MAX structures. About 1/3rd of the selected models
329 was generated with the AF2 implementation using a *Custom* MSA constraining the
330 alignment of the predicted ß1-ß4 strands and of the conserved cysteine residues in
331 the SS “1” disulfide bond (S6 Table). An overview of the general structural features
332 characterizing the 77 validated MAX effector AF models is given in S7 Table.

333 Among the 14 OG proteins that were not predicted to fold with the MAX
334 topology, three (OG22, OG77 and OG85) were previously flagged as suspects based
335 on their low TM-pred score, and five (OG04, OG59, OG65, OG68 and OG81) had a
336 TM-pred score below 0.6 (S1 Fig). Other OG proteins such as OG51 and OG54 that
337 exhibited a TM-pred score above 0.6 compatible with a MAX structure displayed
338 significant distortions from the canonical MAX fold when modeled by AlphaFold2. For
339 OG51, two models computed with MMseqs2 and the AF2 Jackhmmer
340 implementation gave very similar models (backbone r.m.s.d. of 1.77Å) with pLDDT
341 overall scores of 52.6 and 61.9, respectively. However, the C-terminal ß-strand of
342 OG51 had a parallel orientation relative to the first ß1 strand that was not compatible
343 with the MAX topology. The best OG54 model had a pLDDT score of 59.4 but
344 deviated from the MAX topology by the absence of the C-terminal ß6 strand, which
345 was not accurately modeled.

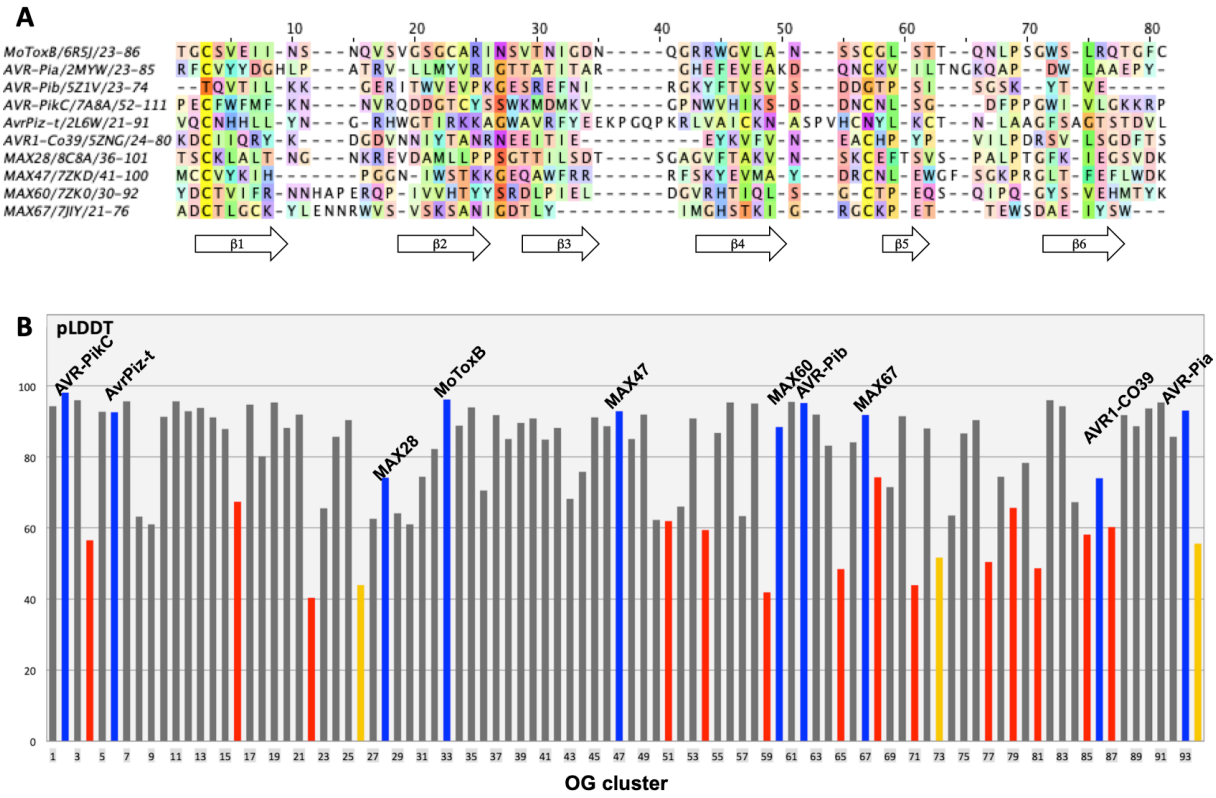


Fig 4. pLDDT scores of AF2 models for known and predicted MAX effectors.

(A) Structural alignments of experimentally determined MAX effector structures using MoToxB structure for reference. Residues are colored according to the Taylor scheme [21](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.

The 14 OG clusters that gave inconclusive AF2 models were submitted to two other protein structure prediction web-servers, RaptorX [22] and RosettaFold [23]. None of the computed models displayed the canonical MAX fold, with consistently low prediction scores, confirming the challenging nature of modeling these OG cluster sequences that may deviate from the MAX topology.

364 ***Variations around the canonical MAX fold***

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

375 The number of disulfide bonds stabilizing the MAX protein can also greatly
376 vary, from none in MAX61 and MAX62 (AVR-Pib) up to three in MAX46 or four in
377 MAX52 (S7 Table). A unique member of the MAX family was MAX52, whose AF
378 model consisted of two MAX core domains arranged in tandem and designated
379 MAX52A and MAX52B in S7 Table, each having two disulfide bonds. Besides the
380 conserved disulfide bond (SS “1”), which is a hallmark of the MAX domain, three
381 types of additional disulfide bonds (SS “2”, “3” and “4”) were found in the
382 experimental and AF2 MAX structures (Figs 5A and B). SS “4”, joining the N-
383 terminus of ß1 to a C-terminal cysteine, is well defined in both MAX47 and MAX92. It
384 was not present in any of the previously determined MAX 3D structures that could
385 serve as template and was validated by our NMR structure of MAX47.

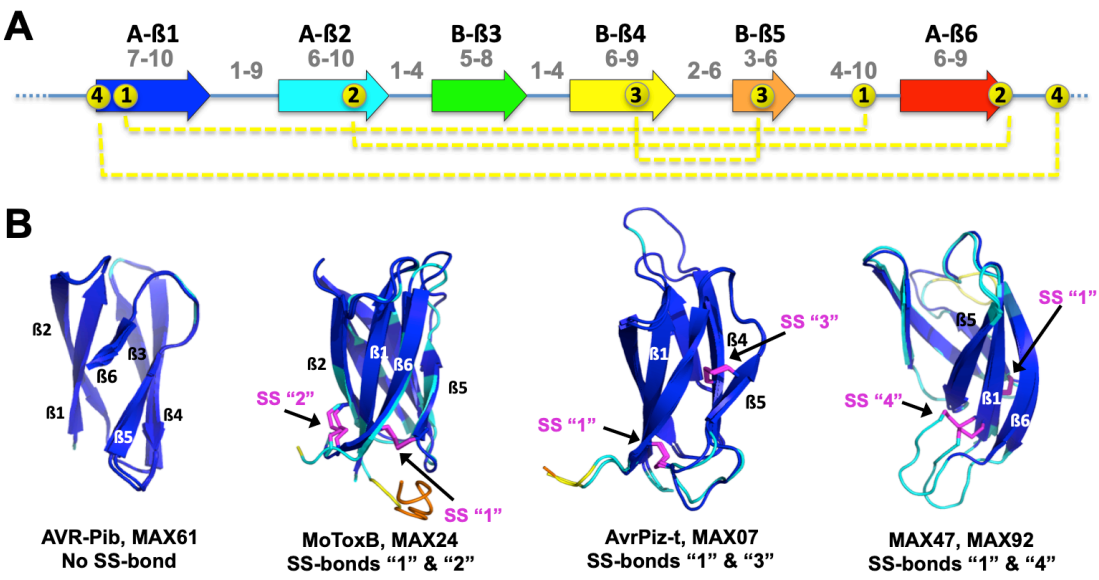


Fig 5. MAX domain structural features. *(A) Average size range (indicated in grey) of ß strands (arrows) and connecting loops forming the central ß sandwich of the canonical MAX core, and of the N-ter and C-ter extensions. Average ranges are calculated in Table S7b. The 4 different types of disulfide 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.*

N- and C-terminal extensions

Over two-third 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

406 the model and solution structure of MAX60. In many cases, terminal extensions
407 appeared as unstructured regions that could not be modeled with high confidence by
408 AF2. Long intrinsic disordered regions (IDRs) of more than 30 a.a. [24–26] may have
409 diverse function in bacterial [27,28] and fungal effectors [29,30] and we therefore
410 searched for IDR signatures in MAX effector sequences using ESpritz prediction
411 software [31]. Long IDRs were predicted for ten MAX effectors and were unstructured
412 in six AF_MAX models: in MAX15 (118 a.a.), MAX27 (36 a.a.) and MAX43 (43 a.a)
413 as N-ter extensions, and in MAX28 (42 a.a.), MAX53 (38 a.a.) and MAX78 (43 a.a.)
414 as C-ter extensions. It thus appears that long IDRs are a rare feature among *P.*
415 *oryzae* MAX effectors, present in less than 8% of all modeled structures. The NMR
416 solution structure of MAX28 validated the unstructured nature of its C-terminal
417 extension.
418

419 ***Two-thirds of the MAX effectors grouped in 20 structural classes***

420 Here we performed hierarchical clustering of the selected 77 MAX effector
421 models with two protein structure alignment software, Dali [32] and TM-align [33,34]
422 which use different criteria for similarity scoring of superimposed structures. The Dali
423 Z-score relies on secondary structure pairing and is a good estimate of topological
424 conservation while the TM-score is computed for the whole alignment and weights
425 paired residues with low r.m.s.d. more strongly than those that are more distant.
426 When analyzed independently, the structural alignment trees retrieved from these
427 two clustering approaches did not reveal clear sub-families of MAX structures, as
428 shown by the lack of long internal branches in both trees (Fig 6). To facilitate the
429 comparison of the trees, we differently colored the lines connecting each MAX model
430 in both trees (see Materials and Methods).

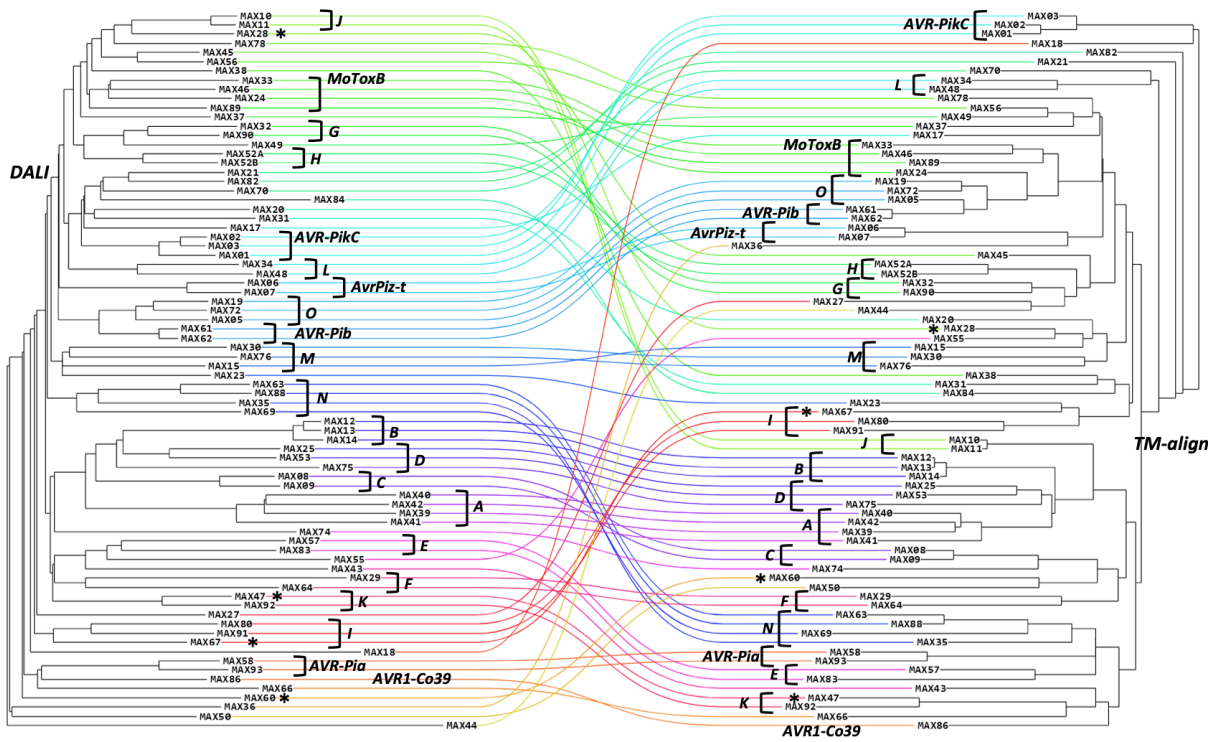
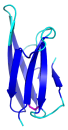
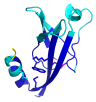
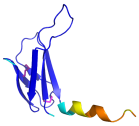


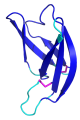
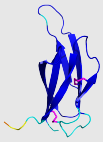
Fig 6. Comparison of the structural similarity trees of MAX effectors based on the Dali Z-score (left) and TM-align TM-score (right) of their superimposed AF models. *Unstructured N- and C-terminal regions were removed from the AF_MAX models prior to the analysis. A line of a specific color connects 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 Table 1 and S3 Fig.*

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 statistical average values 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-PikC, AvrPiz-t and MoToxB) (Table 1). Together, these 20 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. The characteristics of the 20 MAX structural groups are summarized in Table 1 and illustrated with more details in S3 Fig. Only 6 groups had a disulfide bond pattern with a specific SS bond adding to the conserved SS ”1”, and only the AVR-Pib group had no SS bond.

462 ***Table 1. Structural groups identified in P. oryzae MAX effector family***

Group	representative structure	MAX Id	SS(*)	Structural Specificities and Extensions
A		*MAX39, MAX40,* ***MAX41,*** *MAX42, MGG_16939T0 MGG_18062T0*	*1*	additional ß7ß8 strands flanking one side of the ß2ß1ß6 sheet and C-terminal extensions after ß8 that are in helical conformations
B		***MAX12,*** *MAX13, MAX14 MGG_16619T0*	*1*	additional ß7ß8 strands flanking one side of the ß2ß1ß6 sheet
C		***MAX08,*** *MAX09*	*1*	presence of the ß7 additional strand running antiparallel to the ß6
D		***MAX25,*** *MAX53, MAX75 MGG_02635T0 MGG_15207T0 MGG_11304T0*	*1*	presence of the ß7 additional strand running antiparallel to the ß6
E		*MAX57,* ***MAX83***	1	long ß1 and ß6 strands,C-terminal helical extensions that are stabilized by hydrophobic interactions
F		*MAX29,* ***MAX64*** *MGG_17255T0*	*1*	C-terminal helical extensions that are stabilized by local residue-specific interactions (disulfide bond or hydrophobic residues)
G		*MAX32,* ***MAX90*** *MGG_08469T0*	*1, 3*	long loop joining the ß1 and ß2 strands and helical C-terminal extensions that are not strictly conserved in their lengths and orientations
H		***MAX52***	*1, 3*	two MAX domains
AVR-Pia		*MAX58,* ***MAX93***	*1*	ß2 strand interaction with HMA (**)
AVR-PikC		*MAX01,* ***MAX02****, MAX03 MGG_15972T0*	*1*	ß3 strand and N-terminal extension interacting with HMA (**)

Group	representative structure	MAX Id	SS(*)	Structural Specificities and Extensions
I		***MAX67,*** *MAX80, MAX91 MGG_16175T0*	*1*	long ß1 strand paired with long ß2 or ß6 strands
J		*MAX10,* ***MAX11***	*1*	presence of a short helix in the connecting loop between ß5 and ß6
K		***MAX47,*** *MAX92*	*1, 4*	disulfide bond "4" joining the N-terminus of ß1 and the C-terminal extension
L		***MAX34,*** *MAX48 MGG_15625T0 MGG_08992T0*	*1, 3*	malformed ß2 strand and anti-parallel ß4ß5 strands stabilized by a disulfide bond
M		*MAX15,* ***MAX30****, MAX76 MGG_10282T0*	*1*	longer ß1 and ß2 strands and their connecting loops
N		***MAX63,*** *MAX88, MAX35, MAX69 MGG_02207T0*	*1*	unusually long loop making the connection between strands ß2 and ß3
O		*MAX05,* ***MAX19****, MAX72 MGG_12426T0 MGG_14793T0 MGG_18060T0*	*1*	Structuraly similar to AVR-Pib.
AVR-Pib		*MAX61,* ***MAX62***		Structuraly similar to group O models.
AvrPiz-t		***MAX06,*** *MAX07 MGG_18041T0*	*1, 3*	very similar structures with the main difference for the ß3ß4 connecting loop
MoToxB		*MAX24,* ***MAX33****, MAX46, MAX89 MGG_17132T0*	*1, 2*	Structuraly similar to ToxB with disulfide bond "2"

463 *The AF models are colored by their pLDDT scores (see legend of Fig 5). The groups with*
464 *major structural variations (addition of secondary structures) are listed in the left-hand*
465 *panels, including the MAX domain duplication of MAX52. The remaining groups, from I to O*
466 *that do not have additional secondary structural elements but displayed specific structural*
467 *variations of the MAX core domain itself are shown in the right-hand panels. To complete the*
468 *overview, the five groups of well-established MAX effectors (AVR-Pia, AVR-Pib, AVR-PikC,*

AvrPiz-t and MoToxB) are shown at the bottom and highlighted in grey color. The MAX Id column gives the AF model identifiers of the MAX effector list reported in the present study (representative MAX model indicated in bold) and the corresponding MAX effectors reported by Seong & Krasileva, 2021 [4]; Seong and Krasileva, 2023 [5]; Yan and Talbot, 2023 [10] in the P. oryzae strain 70-15 are referred to by their MGG identifier (S8 Table).

(*) Type of disulfide bond as defined in Fig 5

(**) from crystallographic structures of complexes

One-third of the MAX effectors are singletons

Singletons MAX effectors had no common structural features with other MAX effectors and constituted one-third of the AF_MAX effector models (26/77). The majority of them (17) consisted of a simple MAX core with unstructured N- and/or C-terminal extensions. Among them was AVR1-CO39 (MAX86). The remaining 9 singletons had diverse structured extensions (S4 Fig). In MAX49, N-terminal and C-terminal helices formed a helix bundle structure, not observed in other MAX effector models. MAX36, MAX43, MAX66 and MAX74 had an additional ß7 strand, which aligned anti-parallel to ß6, as observed for group D MAX effectors. Additional C-terminal helices were predicted for MAX44, MAX84, MAX55 and MAX60, in the last two cases with high confidence.

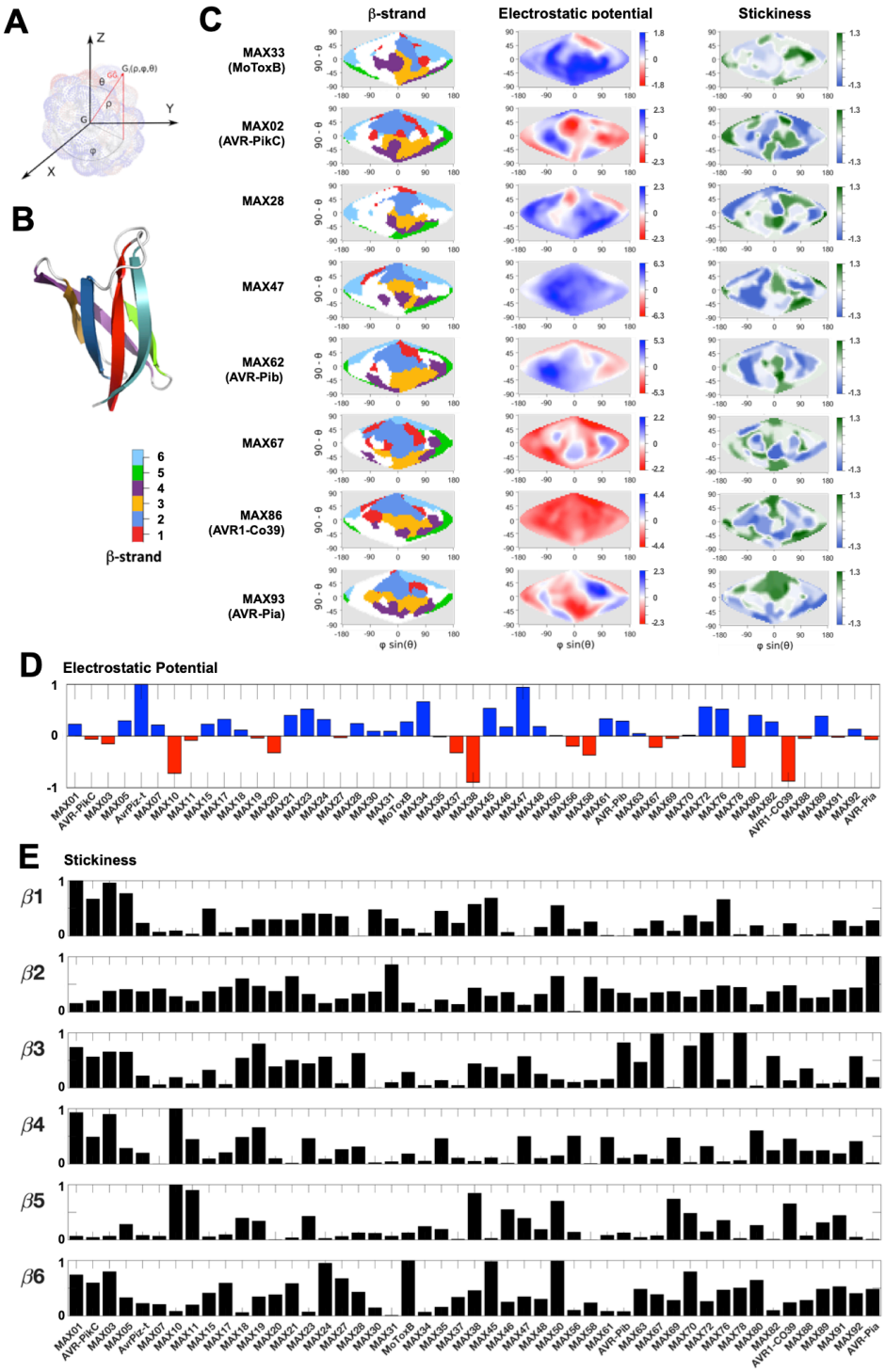
MAX domains exhibit highly variegated surface properties

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

496 central core (S7a Table). For this subset of MAX domains, we computed SURFMAP
497 [35] 2D projections of their molecular envelop, and compared the distribution of the
498 following surface features: exposed secondary structures, electrostatic potential,
499 stickiness and amino-acid polymorphism within the OG cluster to which belongs each
500 representative MAX model (Fig 7).

501 The electrostatic potential maps (Fig 7D and S2 files) revealed that MAX
502 domain surfaces were more often positively charged than negatively charged or
503 neutral, and that the molecular surfaces can appear entirely positive (e.g. MAX06
504 /AvrPiz-t, MAX23, MAX34, MAX47) or negative (e.g. MAX10, MAX38, MAX78,
505 MAX86/AVR1-CO39), or present intense electrostatic patches (e.g. MAX02/AVR-Pik,
506 MAX58, MAX62/AVR-Pib, MAX80). It is well established that positively charged
507 regions in proteins are important for interaction with negatively charged
508 macromolecules, such as nucleic acids and lipopolysaccharides [36], whereas
509 negatively charged protein surfaces can be involved in membrane attachment or
510 DNA mimicking functions [37–39]. In MAX47, we noticed that its unstructured N-
511 terminal extension is rich in aspartic residues, suggesting that it could make transient
512 interactions with the positively charged MAX core in the absence of its cellular target.
513 In MAX62 (AVR-Pib), a surface loop region formed a strong positive patch (Fig 7C),
514 which has been shown to be essential for the avirulence function of AVR-Pib and its
515 nuclear localization in host cells [35]. Interestingly, a similar positive patch was visible
516 on the surface of its structural homolog MAX61 belonging to the same structural
517 group, as well as in MAX05 and MAX72 belonging to group O (Table1), suggesting
518 that these effectors may also rely on this positive surface loop for their function.
519 Similarly, AvrPiz-t displayed a positively charged surface mostly formed by lysine
520 residues that are required for AvrPiz-t avirulence and virulence functions in rice [18].

521 Inversely, while the surface of MAX86 (AVR1-CO39) is strongly negative, that of
522 MAX93 (AVR-Pia) is neutral, yet both AVR1-CO39 and AVR-Pia interact similarly
523 through their ß2 strand with the HMA domain of the rice immune receptors RGA5
524 [40], respectively.



525

526 ***Fig 7. 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). 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 $\overrightarrow{GG_i}$ in the plan ($\overrightarrow{GX}$, $\overrightarrow{GY}$), while θ is the angle between the vector $\overrightarrow{GG_i}$ 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 [41]. *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 files). 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.*

Wide divergence was also observed in the surface stickiness of the MAX domains. Surface stickiness is mostly related to surface hydrophobicity and reflects the propensity of amino acids to be involved in molecular interfaces [41]. For AVR-Pia, the HMA binding site correlated well with the presence of a large hydrophobic surface patch (Fig 7C and panel C in S5 Fig) and a very sticky ß2 strand, also present in MAX31 (Fig 7E). This was however not the case for AVR1-CO39 whose surface hydrophobicity was limited except in strand ß5. This strand is particularly variable in length and surface properties, exhibiting high stickiness in only a few MAX effectors (e.g. MAX10, MAX11 and MAX38) where it could contribute to the binding site of host target proteins. 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 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 (S5 Fig). In MAX structural groups A to D (Table 1, not included in the present subset of Fig 7), 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. 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 Files).

585 # Discussion

586 *AF2 outcompetes other strategies for the prediction of MAX effector structures*

587 In this study, we combined experimental structure determination and *in silico*
588 modeling to elucidate the commonalities and variability of the three-dimensional
589 structures in the MAX effector family of *Pyricularia oryzae*. Using X-ray
590 crystallography we solved the structure of MoToxB and by NMR spectroscopy, we
591 determined the structures of 4 new MAX effectors, MAX28 MAX47, MAX60 and
592 MAX67, bringing to 10 the number of the experimental MAX effector structures from
593 *P. oryzae*. This extended reference set, enabled us to evaluate the accuracy with
594 which different modeling techniques can predict the structure of MAX effectors. In a
595 previous work we had used a combination HMM pattern searches and hybrid multiple
596 template modeling to predict the core of MAX effector structures and selecting best
597 models according to their TM-pred score [11]. For the selected TM-pred models of
598 MAX28, MAX47, MAX60 and MAX67 the overall structures of the MAX domain were
599 properly modeled, and the observed deviations with their experimentally determined
600 counterparts (TM score) were well predicted by the TM-pred scoring function (Fig 3A).
601 However, the TM-pred models, although limited to the conserved MAX core, deviated
602 substantially from the NMR structures (r.m.s.d ranging from 2.1 to 5.3 Å), highlighting
603 the limits of template-based homology modeling [42], and the need of alternative
604 strategies for the reliable modeling of MAX effectors.

605 Contrary to this, AF2 predicted with very high accuracy the new experimental
606 structures. This was not only the case in the MAX effector core but also in structured
607 regions outside the core. The average r.m.s.d values for superimposed backbone
608 atoms of experimental and modeled structures were between 0,99 and 1,42 Å. Side
609 chain rotamers and disulfide bond conformations were nearly identical, and

610 interactions between secondary structure elements were predicted with high
611 precision. Interestingly, this was true even in the case of MAX28 whose AF2 model
612 had a MAX pLDDT score of only 74. These findings highlighted that AF2 predicts
613 MAX effector structures in a highly reliable and precise manner even at relatively low
614 pLDDT scores. Similar observations were made in other recent studies where
615 experimentally resolved structures of fungal effectors were compared with AF2
616 predictions, as for example in the case of LARS [43], FOLD [44] and RALPH [45]
617 effectors. Comparison with a published study testing *ab initio* approaches using
618 Rosetta or the two web servers Robetta and QUARK to model MAX effectors with
619 already known structures [46] confirmed that AF2 outcompetes other strategies for
620 the prediction of MAX effector structures.
621
622 *Comparison with modeling of MAX effectors from P. oryzae strain 70-15*
623 Three previous studies [4,5,10] aiming at identifying effector families in *P.*
624 *oryzae* have used deep-learning methods such as TrRosetta or AF2 for
625 systematically modeling the effector candidates from the reference isolate 70-15. In
626 these three studies a total of either 11, 26 and 32 MAX effectors were detected,
627 respectively. Ten effector candidates were identified as MAX effectors by all three
628 studies, and another set of 22 MAX effectors were identified in only one or two (S8
629 Table). The high number of false negatives (18) in these studies that missed a large
630 part of the MAX effectors of the 70-15 isolate highlights difficulties in model selection
631 based on pTM-score and in detecting structural similarities using methods like CATH
632 [47] and SCOPe [48], which were employed in these studies. All MAX effectors that
633 were reported in these previous TrRosetta and AlphaFold studies were also identified
634 in our analysis (Table 1 and S8 Table) with the exception of OG54 that we did not

635 retain as a MAX effector. MAX members of the groups C, E and H were not found in
636 the previous reports and members of groups J and K are absent from the genome of
637 70-15 for MAX. Taken together, these comparisons reveal that the combination of
638 HMM-based pattern searches with systematic AF2 modeling is the most
639 comprehensive approach for establishing the composition of fungal effector families,
640 while structure modeling alone results in numerous false negatives, and
641 underestimates the size of effector families.
642
643 *Classification of MAX effector AF2 models predicted in 120 P. oryzae genomes*
644 The alignment of the 77 *bona fide* MAX effector models from AF2, NMR or X-
645 ray crystallography permitted to determine with high precision the characteristics of
646 the well-conserved structural core. It revealed for instance the mean length and
647 variance of the conserved secondary structure elements and allowed the
648 classification of possible cysteine bonding patterns. Only in exceptional cases,
649 canonical structural features were missing or replaced. These cases mostly
650 concerned cysteine bond SS “1” that was lacking in two MAX OGs or beta strand 5
651 that was replaced by an alpha helix or absent in five MAX OGs.
652 Comparison of the 77 MAX effector structures also showed that, beyond the
653 well-conserved core, the MAX effector family harbors important structural diversity.
654 Clustering of the structures distinguished 20 subfamilies comprising 51 orthogroups
655 as well as 26 singletons that could not be classified. Major distinctive features are
656 additional structured regions in the C-terminus characteristic of certain subfamilies.
657 Groups A to D have one or two additional strands and groups E to G possess helical
658 extensions. These regions presumably act in protein-protein interactions or contribute
659 to overall functionality of the effectors. Notably, our study uncovered a domain

660 duplication event within one of the MAX effector clusters (group H). Other dual-
661 domain effectors have been described in recent studies, i.e. the Fol dual-domain
662 effectors (FOLD) [44] and effectors predicted from *Puccinia graminis* [5]. The
663 discovery of dual-domain effectors, including the domain duplication found in one of
664 the MAX effector cluster, adds to our understanding of the diversity and complexity of
665 fungal effector proteins. These dual-domain effectors likely have evolved to possess
666 multiple functional domains that contribute to their virulence or interaction with host
667 plants.

668 In addition to contributing to effector function, N- and C-terminal extensions of
669 MAX effectors may have critical roles in protein folding. This hypothesis is supported
670 by studies, where we used HP-NMR (high Hydrostatic Pressure NMR) [49,50] to
671 analyze the folding/unfolding of AVR-Pia, AVR-Pib [51] and MAX60 [52]. While the
672 MAX effector core of AVR-Pia and AVR-Pib folded similarly around a ß3ß4
673 intermediate, MAX60 had an early folding intermediate formed by ß1, ß6 and the C-
674 terminal helix, a specific extension of this MAX effector. Mutants lacking this helix
675 were not sufficiently stable to be purified. These findings show how additional
676 sequences outside the core can have profound impacts on the folding of MAX
677 effectors.

678 Beside structured extensions, long intrinsically disordered regions (IDRs) were
679 only observed in six MAX effectors. IDR regions lack a stable 3D structure and
680 exhibit conformational flexibility, allowing them to interact with multiple binding
681 partners and fulfill various functions [53] but are difficult to precisely predict from the
682 sequence [54]. Modest structural variations were detected within the MAX core of the
683 seven subgroups I to O and distinguished these groups from each other. The
684 classification provided in Table 1 can be further improved by incorporating surface

685 properties. Thus, Avr-Pib and O groups could be combined, since they are
686 structurally similar and members of the O group have electrostatic surfaces related to
687 AVR-Pib.
688
689 *MAX effectors of Venturia inaequalis are distinct from P. oryzae MAX effectors*
690 *V. inaequalis* is an ascomycete fungus, in the *Venturiaceae* family, responsible
691 for apple scab disease. Although it is only very distantly related to *P. oryzae*, V.
692 *inaequalis* has also an extended MAX effector family as revealed by systematic
693 modeling of its effector repertoire [55]. However, none of the *V. inaequalis* MAX
694 effectors fitted into any of the *P. oryzae* MAX subfamilies. Indeed, all *V. inaequalis*
695 MAX effectors present three conserved disulfide bonds, of which one is the canonical
696 SS "1" bond characteristic of MAX effectors of *P. oryzae* and other fungi, while the
697 remaining two were not found in MAX effectors of any other species. Moreover,
698 MAX-like effectors of *V. inaequalis* usually possess a C-terminal helical extension
699 connected to the MAX core domain *via* a specific disulfide bond. This defines these
700 newly discovered MAX-like effectors from *V. inaequalis* as a distinct subfamily with
701 unique sequence and structure features [55]. *V. inaequalis* colonizes the leaf surface
702 by growing below the cuticle, and releases effectors in this sub-cuticular host
703 environment without penetrating the underlying epidermal cells. Due to this specific
704 life style, the function and host targets of the *V. inaequalis* MAX effectors, are
705 presumably fundamentally different from those of *P. oryzae*.
706
707 *Conclusion and perspectives*
708 Structural information tremendously extends the insight, which can be
709 obtained from primary sequence, and expands our understanding of biological

710 processes or evolution to the atomic level. However, corresponding analyses are far
711 from being trivial especially in rapidly evolving protein families with high sequence
712 diversity, as it is the case of fungal effector proteins. Our study shows that HMM
713 pattern searches associated with AF2 structure modeling provide a solid basement
714 for establishing in a comprehensive manner effector families in fungi. In addition to
715 providing detailed insight into effector family diversity, such studies generate solid
716 bases for the functional investigation of fungal effectors and for the investigation of
717 their evolutionary trajectories. The combination of structural modeling and population
718 genomics provides exciting perspectives for accelerated and deepened investigation
719 of the molecular evolution of fungal effector proteins [11]. Ongoing and rapid
720 improvements in *in silico* protein-protein interaction analysis, such as improved
721 prediction of the structures of protein complexes [56] or screening of interacting
722 proteins [57], are opening a new era in the functional investigation of effectors by
723 generating at high throughput hypothesis for effector function and evolution.
724 However, for both of these research areas experimental structure determination
725 remains critical, since large parts of fungal effectoromes can still not be modeled with
726 good confidence, and models of protein complexes generally lack precision and
727 therefore only provide limited insight into the identity and conformation of the directly
728 interacting residues and the details of the binding interface.
729

Materials and Methods

Experimental Structures

MAX28 Protein expression and purification.

Protein expression and purification experimental details for MAX47, MAX60 and MAX67 are available in [58] Mounia et al. 2022. For MAX28 we followed essentially the same protocol than for the other MAX effectors for producing the ^{15}N-labelled NMR sample but the His$_6$-tag was not cleaved 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 $^{15}NH_4Cl$ (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

755 at 27°C on a Bruker Avance III 800 MHz or Bruker Avance III 700 MHz spectrometer,
756 both equipped with 5 mm z-gradient TCI cryoprobe. NMR samples consisted on
757 approximately 0.4 mM ^{15}N-labeled protein dissolved in 10 mM Na-Phosphate buffer
758 (pH 6.8) and 150 mM NaCl with 5% D_2O for the lock.
759
760 *NMR Structure determination of MAX28, MAX47, MAX60 and MAX67.*
761 ^{1}H chemical shifts were directly referenced to the methyl resonance of DSS, while
762 ^{15}N chemical shifts were referenced indirectly to the absolute ^{15}N/^{1}H frequency ratio.
763 All NMR spectra were processed with Topspin 3.6 (Bruker) and analyzed with Cindy
764 2.1 (Padilla, www.cbs.cnrs.fr). Assignments for MAX28, MAX47, MAX60 and MAX67
765 have been deposited to and are available from the BMRB data bank under the
766 accession entry 34782, 34731, 34730 and 34729, respectively.
767 The NMR structures were determined from the NMR constraints listed in S4 Table
768 that were obtained as follow. NOE cross-peaks identified on 3D [^{1}H, ^{15}N] NOESY-
769 HSQC (mixing time 150 ms) were assigned through automated NMR structure
770 calculations with CYANA*3* [59,60]. Hydrogen bond restraints were derived using
771 standard criteria on the basis of the amide ^{1}H / ^{2}H exchange experiments and NOE
772 data. When identified, the hydrogen bond was enforced using the following restraints:
773 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
774 obtained from TALOS-N [61] analysis of backbone atom chemical shifts for MAX47,
775 MAX60 and MAX67. For the final list of restraints, distance values redundant with
776 covalent geometry were eliminated and disulfide bonds that were consistent with
777 short distances between cysteine residues were added.
778 A total of 200 three-dimensional structures were generated using the torsion angle
779 dynamics protocol of CYANA*3* from NOEs, hydrogen bonds and disulfide bond

780 restraints (S4 Table). The 20 best structures (based on the final target penalty
781 function values) were minimized with CNS 1.2 according to the RECOORD procedure
782 [62] and analyzed with PROCHECK [63]. The rmsds were calculated with MOLMOL
783 [64]. All statistics are given in S4 Table.
784 The structure coordinates have been deposited at the Protein Data Bank under the
785 following accession codes: MAX28 (PDB_8C8A), MAX47 (PDB_7ZKD), MAX60
786 (PDB_7ZK0), MAX67 (PDB_7ZJY),).
787

788 ***TM-pred MAX models Web Table***

789 Homology models of each OG representative sequence relative to each 3D
790 template were built using MODELLER v9.1 [65] with several alternative query-template
791 threading alignments as described in Le Naour—Vernet et al., 2023 [11]. The top-5
792 models were selected according to the TM-pred scoring function and are given in the
793 Web page: https://pat.cbs.cnrs.fr/magmax/model/
794 The Web table has the following columns: Group (OG cluster), Score (composite
795 evaluation score of the best model -best scores are in red, worst are in blue-), Dfire,
796 Goap, Qmean, E1D, E2D, E3D (individual evaluation scores of the best model),
797 Alignment (aligned identifier sequences), Identifier (protein identifier of the orthologs
798 prioritizing *Oryzae* infecting strains). Structural models and alignments are available
799 by clicking on each OG cluster in the first column. A multiple sequence alignment of
800 non-redundant orthologous proteins is displayed at the top of the page, starting with
801 the modeled representative sequence. For each OG cluster given in [11] we further
802 filtered out redundant sequences using CD-HIT [66] (S1 Table). The representative
803 sequence of each OG cluster was determined to be the sequence sharing the
804 highest sequence identity with a consensus sequence derived from the OG cluster

805 sequence alignment by MAFFT [67] (S2 Table). Signal peptide prediction before the
806 cleaved residue (cs.SIGNALP41), mean hydrophobic index (ih.MEAN), sequence
807 conservation (1D.CSRV), alternative alignments and model secondary structures
808 (2d.STRIDE) are displayed below the sequence alignment. Sequence conservation
809 scoring implemented in PAT [68] was calculated according to Sonnhammer and
810 Hollich, 2005 [69] with a Gonnet matrix [70]. At the bottom of the page, the 5 best
811 models can be displayed using different representations and color schemes. Models
812 can also be downloaded in PDB format.
813

814 ***Modeling by AlphaFold.***

815 For each OG representative sequence we computed three AF models differing
816 by the way of building multiple sequence alignment (MSA). The MMseqs2 MSA was
817 obtained from the MMseqs2 [71,72] web server as implemented in the ColabFold
818 version of AlphaFold 2.0 [73]. We also used the version of AlphaFold 2.2.0 that builds
819 MSAs by Jackhmmer on uniclust, mgnify and uniref90 databases. These two
820 implementations use PDB templates. Finally, a *Custom* MSA was build from
821 Muscle_v3.8.31 [74] by inserting (-profile option) the query sequence on top of a
822 previously computed MSA, termed ß1ß4_MSA. The ß1ß4_MSA was build from a
823 Muscle alignment of the OG sequences (S1 Table) by filtering out those having the
824 two flanking cysteine residues in the ß1 and in the loop between ß4 and ß5 strands
825 not correctly aligned to the 8 3D template sequences. The ß1ß4_MSA was further
826 processed by truncating the sequences by eliminating residues (-2 included) before
827 and (+2 included) after the first and last aligned cysteine residues, respectively, and
828 filtering out for redundant sequences by CD-HIT, giving a total size of 247 aligned
829 sequences (S6 Table). For each query, the consistency of the *Custom* MSA was

830 determined by checking the correct alignment of the cysteine residues in the query
831 and in the appended ß1ß4_MSA. When consistent the *Custom* MSA was converted
832 to a3m format by the reformat.pl script [75] and directly used as input in the
833 ColabFold implementation of AlphaFold 2.0 calculations that was setup without the
834 use of PDB templates. *Custom* MSAs could not be built for OG61 and OG62 from
835 absence of cysteine residues in their primary sequences and were not consistent for
836 OG15, OG27, OG71, OG81, OG85 and OG92.

837 The quality of each model was assessed by the pLDDT overall score [76]. The
838 correct MAX topology was verified by visual inspection (Pymol v.1.6 Delano 2002).
839 For models having the MAX topology, a MAX pLDDT score that was an average
840 score of residues in the MAX core domain (including residues from ß1 to ß6) was
841 calculated. For each query, the best AF2 MAX model was selected when the MAX
842 pLDDT score was above 60. The complete set of AF2_generated models for the 77
843 validated MAX effectors is available in S1 Files.

844

845 ***Dali implementation and search***

846 A standalone implementation of DaliLite.v5 [77] was used for this work. For the
847 all-to-all clustering by Dali we first discarded *unstructured* stretches in each model.
848 For this, the *structured* domain of each model was defined by taking the STRIDE [78]
849 output, and filtering for the first residue in the first and last residue in the last
850 secondary structure (helix or strand), respectively (S7 Table). The model of MAX52
851 was split in two chains A and B each containing a MAX domain. All these *structured*
852 domains were used for clustering with Dali Z-scores excluding *de-facto* unresolved
853 protein regions without loosing important structural information.

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

855 The distance between each pair of AF2 models that were used for Dali
856 clustering was estimated by the TM-score obtained from TM-align after pairwise
857 model superposition. A classification tree was then inferred from these pairwise
858 distances using FastME [79]. Finally, Fig 6 was obtained by joining identical models
859 in the FastME tree and in the Dali tree, respectively, by a line of the same color.

860

861 ***Surface properties of MAX core domains.***

862 A subset of 49 MAX effector AF2 models, each consisting of a MAX core
863 domain and optional N- and/or C- *unstructured* extensions (S7 Table) was defined by
864 discarding AF2 models having N- and/or C-terminal *structured* extensions (listed in
865 the groups A to H in Table 1 and in S4 Fig). All MAX core domains of the 49 AF2
866 models were superimposed to the reference MoToxB structure with their ß1 strand
867 vertically aligned to the Z Cartesian axe giving a reference frame for the Sanson-
868 Flamsteed 2D projection computed using SURFMAP [35]. Their surface properties
869 including stickiness and electrostatics (APBS) [80] were computed by SURFMAP and
870 are given in S2 File. The temperature factor column of the PDB files was used to
871 encode the color of the exposed surface of the six ß-strands, from 1 to 6,
872 respectively. The sum of the surface stickiness positive values of each individual ß-
873 strand was computed by filtering SURFMAP surface stickiness output and are reported
874 in Fig 7E. The amino-acid conservation scores given for each OG cluster to which
875 belongs each representative MAX model were used to color encode the surface from
876 white for high conservation score of 9, light blue colors for intermediate conservation
877 scores (from 8 to 6), sky-blue for low conservation score of 5 and darker blue colors
878 indicating highly polymorphic positions with conservation scores of 4 and below.

881

882 **Author Contributions:** M.L., and K.d.G. prepared the ^{15}N-^{13}C-labeled protein
883 samples (sub-cloning, protein expression and purification). P.B. made the NMR
884 resonance assignment of the four proteins in the study conditions. J.G. wrote scripts
885 for protein sequence analysis and homology modeling. C.R. and A.P. supervised the
886 project, conceived experiments and modeling, participated in the interpretation, and
887 M.L., A.P., N.D., C.R., S.C., P.G. wrote the article. A.P. also contributed to the
888 funding acquisition with P.G. and T.K. All authors have read and agreed to the
889 published version of the manuscript.

890

891 **Funding:** This research project was funded by the ANR project MagMAX (ANR-18-
892 CE20-0016-02), the European Research Council (ERC-2019-STG-852482-ii-MAX),
893 and supported by French Infrastructure for Integrated Structural Biology (FRISBI)
894 grant No. ANR-10-INSB-05.

895

896 Acknowledgements

897 We are grateful to Liisa Holm for feedback in implementing Dali. The authors
898 are grateful to J. Maidment (CBS, Montpellier) for careful proof reading and language
899 correction of our manuscript. We are grateful to Matthew Bowler and Didier Nurizzo
900 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France for
901 providing assistance in using beamline ID30A-1. The authors gratefully acknowledge
902 the ESRF for provision of synchrotron radiation facilities via Block Allocation Group
903 beamtime.

904

905

Supporting information

S1 Fig. Template-based modeling of MAX effector sequences

S2 Fig. NMR structures and AF models

S3 Fig. Groups of AlphaFold models

S4 Fig. Singletons MAX effectors with structured extensions

S5 Fig. Hydrophobic residues on the surface of AVR-Pik and AVR-Pia

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

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

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

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

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

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

S6 Table. ß1ß4 MSA

S7 Table. Structural features of 77 AF-modeled structures of validated MAX effector, statistics and conservation scores.

S8 Table. Comparison of our selection of MAX effectors with previous modelisation studies

Supplementary Materials_and_Methods. Crystallographic structure determination of MoToxB

S1 File. Zipped folder of AF MAX models

S2 File. Zipped folder of AF MAX model surfaces

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. doi:10.1146/annurev-arplant-043014-114623
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. doi:10.3389/fpls.2015.01168
3. Sperschneider J, Dodds PN. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. Mol Plant-Microbe Interactions®. 2022;35: 146–156. doi:10.1094/MPMI-08-21-0201-R
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 Interactions®. 2021;34: 1267–1280. doi:10.1094/MPMI-03-21-0071-R
5. Seong K, Krasileva KV. Prediction of effector protein structures from fungal phytopathogens enables evolutionary analyses. Nat Microbiol. 2023;8: 174–187. doi:10.1038/s41564-022-01287-6
6. Derbyshire MC, Raffaele S. Surface frustration re-patterning underlies the structural landscape and evolvability of fungal orphan candidate effectors. Genomics; 2023 Jan. doi:10.1101/2023.01.06.522876
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.
8. Fernandez J, Orth K. Rise of a Cereal Killer: The Biology of Magnaporthe oryzae Biotrophic Growth. Trends Microbiol. 2018 [cited 28 May 2018]. doi:10.1016/j.tim.2017.12.007
9. Fernandez J. The Phantom Menace: latest findings on effector biology in the rice blast fungus. aBIOTECH. 2023;4: 140–154. doi:10.1007/s42994-023-00099-4
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. doi:https://doi.org/10.1093/plcell/koad036
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. doi:10.1371/journal.ppat.1011294
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.
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 [cited 10 Jul 2018]. doi:10.1038/s41477-018-0194-x
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. doi:10.1016/j.jbc.2021.100371
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. doi:10.1038/srep11642

980 16. Oikawa K, Fujisaki K, Shimizu M, Takeda T, Saitoh H, Hirabuchi A, et al. The blast
981 pathogen effector AVR-Pik binds and stabilizes rice heavy metal-associated (HMA)
982 proteins to co-opt their function in immunity. Plant Biology; 2020 Dec.
983 doi:10.1101/2020.12.01.406389
984 17. Park CH, Shirsekar G, Bellizzi M, Chen S, Songkumarn P, Xie X, et al. The E3 Ligase
985 APIP10 Connects the Effector AvrPiz-t to the NLR Receptor Piz-t in Rice. Dinesh-Kumar
986 SP, editor. PLOS Pathog. 2016;12: e1005529. doi:10.1371/journal.ppat.1005529
987 18. Bai P, Park C, Shirsekar G, Songkumarn P, Bellizzi M, Wang G. Role of lysine
988 residues of the *Magnaporthe oryzae* effector AvrPiz-t in effector- and PAMP-triggered
989 immunity. Mol Plant Pathol. 2019;20: 599–608. doi:10.1111/mpp.12779
990 19. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly
991 accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589.
992 doi:10.1038/s41586-021-03819-2
993 20. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved
994 protein structure prediction using potentials from deep learning. Nature. 2020;577:
995 706–710. doi:10.1038/s41586-019-1923-7
996 21. Taylor WR. Residual colours: a proposal for aminochromography. Protein Eng
997 Des Sel. 1997;10: 743–746. doi:10.1093/protein/10.7.743
998 22. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, et al. Template-based protein
999 structure modeling using the RaptorX web server. Nat Protoc. 2012;7: 1511–1522.
1000 doi:10.1038/nprot.2012.085
1001 23. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al.
1002 Accurate prediction of protein structures and interactions using a three-track neural
1003 network. Science. 2021;373: 871–876. doi:10.1126/science.abj8754
1004 24. Van Der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al.
1005 Classification of Intrinsically Disordered Regions and Proteins. Chem Rev. 2014;114:
1006 6589–6631. doi:10.1021/cr400525m
1007 25. Liu Y, Chen S, Wang X, Liu B. Identification of Intrinsically Disordered Proteins
1008 and Regions by Length-Dependent Predictors Based on Conditional Random Fields. Mol
1009 Ther - Nucleic Acids. 2019;17: 396–404. doi:10.1016/j.omtn.2019.06.004
1010 26. Monzon AM, Necci M, Quaglia F, Walsh I, Zanotti G, Piovesan D, et al.
1011 Experimentally Determined Long Intrinsically Disordered Protein Regions Are Now
1012 Abundant in the Protein Data Bank. Int J Mol Sci. 2020;21: 4496.
1013 doi:10.3390/ijms21124496
1014 27. Marín M, Uversky VN, Ott T. Intrinsic Disorder in Pathogen Effectors: Protein
1015 Flexibility as an Evolutionary Hallmark in a Molecular Arms Race. Plant Cell. 2013;25:
1016 3153–3157. doi:10.1105/tpc.113.116319
1017 28. Sun H, Zhu X, Li C, Ma Z, Han X, Luo Y, et al. Xanthomonas effector XopR hijacks
1018 host actin cytoskeleton via complex coacervation. Nat Commun. 2021;12: 4064.
1019 doi:10.1038/s41467-021-24375-3
1020 29. Aparicio Chacón MV, Van Dingenen J, Goormachtig S. Characterization of
1021 Arbuscular Mycorrhizal Effector Proteins. Int J Mol Sci. 2023;24: 9125.
1022 doi:10.3390/ijms24119125
1023 30. Jaswal R, Rajarammohan S, Dubey H, Kiran K, Rawal H, Sonah H, et al. Intrinsically
1024 Disordered Kiwellin Protein-Like Effectors Target Plant Chloroplasts and are
1025 Extensively Present in Rust Fungi. Mol Biotechnol. 2023.
1026 doi:https://doi.org/10.1007/s12033-023-00717-y
1027 31. Walsh I, Martin AJM, Di Domenico T, Tosatto SCE. ESpritz: accurate and fast
1028 prediction of protein disorder. Bioinformatics. 2012;28: 503–509.

1029 doi:10.1093/bioinformatics/btr682
1030 32. Holm L. DALI and the persistence of protein shape. Protein Sci. 2020;29: 128–
1031 140. doi:10.1002/pro.3749
1032 33. Xu J, Zhang Y. How significant is a protein structure similarity with TM-score =
1033 0.5? Bioinformatics. 2010;26: 889–895. doi:10.1093/bioinformatics/btq066
1034 34. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on
1035 the TM-score. Nucleic Acids Res. 2005;33: 2302–2309. doi:10.1093/nar/gki524
1036 35. Schweke H, Mucchielli M-H, Chevrollier N, Gosset S, Lopes A. SURFMAP: A
1037 Software for Mapping in Two Dimensions Protein Surface Features. J Chem Inf Model.
1038 2022;62: 1595–1601. doi:10.1021/acs.jcim.1c01269
1039 36. Giraldo MC, Valent B. Filamentous plant pathogen effectors in action. Nat Rev
1040 Microbiol. 2013;11: 800–814. doi:10.1038/nrmicro3119
1041 37. Wang H, Chou C, Hsu K, Lee C, Wang AH -J. New paradigm of functional regulation
1042 by DNA mimic proteins: Recent updates. IUBMB Life. 2019;71: 539–548.
1043 doi:10.1002/iub.1992
1044 38. Dennehy R, Duggan N, Dignam S, McCormack S, Dillon E, Molony J, et al. Protein
1045 with negative surface charge distribution, Bnr1, shows characteristics of a DNA-mimic
1046 protein and may be involved in the adaptation of *Burkholderia cenocepacia*.
1047 MicrobiologyOpen. 2022;11. doi:10.1002/mbo3.1264
1048 39. Goldenberg NM, Steinberg BE. Surface Charge: A Key Determinant of Protein
1049 Localization and Function. Cancer Res. 2010;70: 1277–1280. doi:10.1158/0008-
1050 5472.CAN-09-2905
1051 40. Guo L, Zhang Y, Ma M, Liu Q, Zhang Y, Peng Y, et al. Crystallization of the rice
1052 immune receptor RGA5A_S with the rice blast fungus effector AVR1-CO39 prepared *via*
1053 mixture and tandem strategies. Acta Crystallogr Sect F Struct Biol Commun. 2018;74:
1054 262–267. doi:10.1107/S2053230X18003618
1055 41. Levy ED. A Simple Definition of Structural Regions in Proteins and Its Use in
1056 Analyzing Interface Evolution. J Mol Biol. 2010;403: 660–670.
1057 doi:10.1016/j.jmb.2010.09.028
1058 42. Fiser A. Template-Based Protein Structure Modeling. In: Fenyö D, editor.
1059 Computational Biology. Totowa, NJ: Humana Press; 2010. pp. 73–94. doi:10.1007/978-
1060 1-60761-842-3_6
1061 43. Lazar N, Mesarich CH, Petit-Houdenot Y, Talbi N, Li De La Sierra-Gallay I, Zélie E,
1062 et al. A new family of structurally conserved fungal effectors displays epistatic
1063 interactions with plant resistance proteins. Kobe B, editor. PLOS Pathog. 2022;18:
1064 e1010664. doi:10.1371/journal.ppat.1010664
1065 44. Yu DS, Outram MA, Smith A, McCombe CL, Khambalkar PB, Rima SA, et al. The
1066 structural repertoire of Fusarium oxysporum f. sp. lycopersici effectors revealed by
1067 experimental and computational studies. elife; 2023 Jul. doi:10.7554/eLife.89280.1
1068 45. Cao Y, Kümmel F, Logemann E, Gebauer JM, Lawson AW, Yu D, et al. Structural
1069 polymorphisms within a common powdery mildew effector scaffold as a driver of
1070 coevolution with cereal immune receptors. Proc Natl Acad Sci. 2023;120: e2307604120.
1071 doi:10.1073/pnas.2307604120
1072 46. Rozano L, Mukuka YM, Hane JK, Mancera RL. Ab Initio Modelling of the Structure
1073 of ToxA-like and MAX Fungal Effector Proteins. Int J Mol Sci. 2023;24: 6262.
1074 doi:10.3390/ijms24076262
1075 47. Sillitoe I, Dawson N, Lewis TE, Das S, Lees JG, Ashford P, et al. CATH: expanding
1076 the horizons of structure-based functional annotations for genome sequences. Nucleic
1077 Acids Res. 2019;47: D280–D284. doi:10.1093/nar/gky1097

1078 48. Fox NK, Brenner SE, Chandonia J-M. SCOPe: Structural Classification of Proteins—
1079 extended, integrating SCOP and ASTRAL data and classification of new structures.
1080 Nucleic Acids Res. 2014;42: D304–D309. doi:10.1093/nar/gkt1240
1081 49. Dubois C, Herrada I, Barthe P, Roumestand C. Combining High-Pressure
1082 Perturbation with NMR Spectroscopy for a Structural and Dynamical Characterization of
1083 Protein Folding Pathways. Molecules. 2020;25: 5551. doi:10.3390/molecules25235551
1084 50. Roche J, Dellarole M, Royer CA, Roumestand C. Exploring the Protein Folding
1085 Pathway with High-Pressure NMR: Steady-State and Kinetics Studies. In: Akasaka K,
1086 Matsuki H, editors. High Pressure Bioscience. Dordrecht: Springer Netherlands; 2015.
1087 pp. 261–278. Available: http://link.springer.com/10.1007/978-94-017-9918-8_13
1088 51. Dubois C, Lahfa M, Pissarra J, de Guillen K, Barthe P, Kroj T, et al. Combining High-
1089 Pressure NMR and Geometrical Sampling to Obtain a Full Topological Description of
1090 Protein Folding Landscapes: Application to the Folding of Two MAX Effectors from
1091 Magnaporthe oryzae. Int J Mol Sci. 2022;23: 5461. doi:10.3390/ijms23105461
1092 52. Lahfa M, Mouhand A, de Guillen K, Barthe P, Kroj T, Roumestand C. Does a Similar
1093 3D Structure Mean a Similar Folding Pathway? The Presence of a C-Terminal α-Helical
1094 Extension in the 3D Structure of MAX60 Drastically Changes the Folding Pathway
1095 Described for Other MAX-Effectors from Magnaporthe oryzae. 2023.
1096 53. Bondos SE, Dunker AK, Uversky VN. On the roles of intrinsically disordered
1097 proteins and regions in cell communication and signaling. Cell Commun Signal. 2021;19:
1098 88, s12964-021-00774–3. doi:10.1186/s12964-021-00774-3
1099 54. Zhao B, Kurgan L. Deep learning in prediction of intrinsic disorder in proteins.
1100 Comput Struct Biotechnol J. 2022;20: 1286–1294. doi:10.1016/j.csbj.2022.03.003
1101 55. Rocafort M, Bowen JK, Hassing B, Cox MP, McGreal B, de la Rosa S, et al. The
1102 Venturia inaequalis effector repertoire is dominated by expanded families with
1103 predicted structural similarity, but unrelated sequence, to avirulence proteins from
1104 other plant-pathogenic fungi. BMC Biol. 2022;20: 246. doi:10.1186/s12915-022-01442-
1105 9
1106 56. Bryant P, Pozzati G, Elofsson A. Improved prediction of protein-protein
1107 interactions using AlphaFold2. Nat Commun. 2022;13: 1265. doi:10.1038/s41467-022-
1108 28865-w
1109 57. Yu D, Chojnowski G, Rosenthal M, Kosinski J. AlphaPulldown—a python package
1110 for protein–protein interaction screens using AlphaFold-Multimer. Cowen L, editor.
1111 Bioinformatics. 2023;39: btac749. doi:10.1093/bioinformatics/btac749
1112 58. Lahfa M, Padilla A, De Guillen K, Pissarra J, Raji M, Cesari S, et al. 1H, 13C, 15 N
1113 backbone and side-chain NMR assignments for three MAX effectors from Magnaporthe
1114 oryzae. Biomol NMR Assign. 2022;16: 305–309. doi:10.1007/s12104-022-10095-2
1115 59. Güntert P. Automated NMR structure calculation with CYANA. Methods Mol Biol
1116 Clifton NJ. 2004;278: 353–378. doi:10.1385/1-59259-809-9:353
1117 60. Güntert P, Buchner L. Combined automated NOE assignment and structure
1118 calculation with CYANA. J Biomol NMR. 2015 [cited 18 Apr 2015]. doi:10.1007/s10858-
1119 015-9924-9
1120 61. Shen Y, Bax A. Protein Structural Information Derived from NMR Chemical Shift
1121 with the Neural Network Program TALOS-N. In: Cartwright H, editor. Artificial Neural
1122 Networks. New York, NY: Springer New York; 2015. pp. 17–32. doi:10.1007/978-1-
1123 4939-2239-0_2
1124 62. Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CAEM, Nabuurs SB, et
1125 al. RECOORD: A recalculated coordinate database of 500+ proteins from the PDB using
1126 restraints from the BioMagResBank. Proteins Struct Funct Bioinforma. 2005;59: 662–

1127 672. doi:10.1002/prot.20408
1128 63. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK—a program to
1129 check the stereochemical quality of protein structures. : 283–291.
1130 64. Koradi R, Billeter M, Wüthrich K. MOLMOL: A program for display and analysis of
1131 macromolecular structures. J Mol Graph. 1996;14: 51–55.
1132 65. Webb B, Sali A. Protein Structure Modeling with MODELLER. Methods Mol Biol.
1133 2020;2199: 239–255. doi:10.1007/978-1-0716-0892-0_14
1134 66. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-
1135 generation sequencing data. Bioinformatics. 2012;28: 3150–3152.
1136 doi:10.1093/bioinformatics/bts565
1137 67. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
1138 Improvements in Performance and Usability. Mol Biol Evol. 2013;30: 772–780.
1139 doi:10.1093/molbev/mst010
1140 68. Gracy J, Chiche L. PAT: a protein analysis toolkit for integrated biocomputing on
1141 the web. Nucleic Acids Res. 2005;33: W65–W71. doi:10.1093/nar/gki455
1142 69. Sonnhammer EL, Hollich V. Scoredist: A simple and robust protein sequence
1143 distance estimator. BMC Bioinformatics. 2005;6: 108. doi:10.1186/1471-2105-6-108
1144 70. Gonnet GH, Cohen MA, Brenner SA. Exhaustive Matching of the Entire Protein
1145 Sequence Database. Science. 1992;256: 1443–1445. doi:10.1126/science.1604319
1146 71. Steinegger M, Soding J. MMseqs2 enables sensitive protein sequence searching
1147 for the analysis of massive data sets. Nat Biotechnol. 2017;35: 1026–1028.
1148 doi:10.1038/nbt.3988
1149 72. Mirdita M, Steinegger M, Söding J. MMseqs2 desktop and local web server app for
1150 fast, interactive sequence searches. Hancock J, editor. Bioinformatics. 2019;35: 2856–
1151 2858. doi:10.1093/bioinformatics/bty1057
1152 73. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold:
1153 making protein folding accessible to all. Nat Methods. 2022;19: 679–682.
1154 doi:10.1038/s41592-022-01488-1
1155 74. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high
1156 throughput. Nucleic Acids Res. 2004;32: 1792–1797. doi:10.1093/nar/gkh340
1157 75. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein
1158 sequence searching by HMM-HMM alignment. Nat Methods. 2011;9: 173–175.
1159 doi:10.1038/nmeth.1818
1160 76. Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score
1161 for comparing protein structures and models using distance difference tests.
1162 Bioinformatics. 2013;29: 2722–2728. doi:10.1093/bioinformatics/btt473
1163 77. Holm L. Using Dali for Protein Structure Comparison. In: Gáspári Z, editor.
1164 Structural Bioinformatics. New York, NY: Springer US; 2020. pp. 29–42.
1165 doi:10.1007/978-1-0716-0270-6_3
1166 78. Heinig M, Frishman D. STRIDE: a web server for secondary structure assignment
1167 from known atomic coordinates of proteins. Nucleic Acids Res. 2004;32: W500–W502.
1168 doi:10.1093/nar/gkh429
1169 79. Lefort V, Desper R, Gascuel O. FastME 2.0: A Comprehensive, Accurate, and Fast
1170 Distance-Based Phylogeny Inference Program: Table 1. Mol Biol Evol. 2015;32: 2798–
1171 2800. doi:10.1093/molbev/msv150
1172 80. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to
1173 the APBS biomolecular solvation software suite. Protein Sci. 2018;27: 112–128.
1174 doi:10.1002/pro.3280
1175