

OPEN

Structural genomics applied to the rust fungus *Melampsora larici-populina* reveals two candidate effector proteins adopting cystine knot and NTF2-like protein folds

Karine de Guillen¹, Cécile Lorrain², Pascale Tsan³, Philippe Barthe¹, Benjamin Petre², Natalya Saveleva², Nicolas Rouhier², Sébastien Duplessis², André Padilla¹ & Arnaud Hecker^{2*}

Rust fungi are plant pathogens that secrete an arsenal of effector proteins interfering with plant functions and promoting parasitic infection. Effectors are often species-specific, evolve rapidly, and display low sequence similarities with known proteins. How rust fungal effectors function in host cells remains elusive, and biochemical and structural approaches have been scarcely used to tackle this question. In this study, we produced recombinant proteins of eleven candidate effectors of the leaf rust fungus *Melampsora larici-populina* in *Escherichia coli*. We successfully purified and solved the three-dimensional structure of two proteins, MLP124266 and MLP124017, using NMR spectroscopy. Although both MLP124266 and MLP124017 show no sequence similarity with known proteins, they exhibit structural similarities to knottins, which are disulfide-rich small proteins characterized by intricate disulfide bridges, and to nuclear transport factor 2-like proteins, which are molecular containers involved in a wide range of functions, respectively. Interestingly, such structural folds have not been reported so far in pathogen effectors, indicating that MLP124266 and MLP124017 may bear novel functions related to pathogenicity. Our findings show that sequence-unrelated effectors can adopt folds similar to known proteins, and encourage the use of biochemical and structural approaches to functionally characterize effector candidates.

To infect their host, filamentous pathogens secrete effector proteins that interfere with plant physiology and immunity to promote parasitic growth¹. Although progresses have been made in the past decade, how effectors act in host cells remains a central question in molecular plant pathology. Effectors of filamentous pathogens are secreted and either stay in the apoplast or penetrate inside the cell through specialized infection structures such as haustoria². Effectors are detected by the host plant by two layers of immune receptors at the cell surface or inside the cell, which trigger plant defence response³.

To evade recognition by the host immune system, pathogen effector genes evolve rapidly, notably through the diversification of the amino acid sequence of the encoded proteins⁴. Such diversification impairs the identification of amino acid motifs or sequences similar to known proteins, which could give insights on effector functions inside the host cell⁵. Several superfamilies of effector proteins, such as the fungal MAX or the oomycete WY-domain families, have members showing similar fold but divergent primary sequences^{6,7}, the fold being conserved probably due to the strong link between protein structure and function⁸. Research efforts have been set in this direction and applied in order to determine the structure of several effector proteins^{6,9–11}.

Rust fungi (Pucciniales, Basidiomycetes) constitute the largest group of obligate biotrophic pathogens, that collectively infect almost all plant families, causing serious damages to cultures^{12,13}. In the past decade, genomics and transcriptomics pushed forward rust fungal effector biology research, unravelling hundreds to thousands of candidate effectors with common but not exclusive features¹⁴ such as: protein size, presence of a predicted

¹Centre de Biochimie Structurale (CBS), INSERM U1054, CNRS UMR 5048, Univ Montpellier, F-34090, Montpellier, France. ²Université de Lorraine, INRA, IAM, F-54000, Nancy, France. ³Université de Lorraine, CNRS, CRM2, F-54000, Nancy, France. *email: arnaud.hecker@univ-lorraine.fr

Protein	Protein length (without the signal peptide)	Molecular mass (kDa)	Cysteine residues	Localization in planta ^a	Plant interactors ^a	Avr homologues ^a
MLP124478-(HIS) ₆	70	8.1	6	Nucleolus	Ribosomal proteins	—
MLP124530-(HIS) ₆	96	12	10	Nuclear and cytosolic bodies	—	Avr123
MLP124111-(HIS) ₆	113	13.2	10	Chloroplasts and aggregates	Coproporphyrinogen-III oxidase	—
MLP124561-(HIS) ₆	110	13.7	3	Nucleus and cytosol	—	—
MLP37347-(HIS) ₆	128	15.9	2	Periphery of haustoria ^b , and plasmodesmata	Glutamate decarboxylase	AvrL567
MLP109567-(HIS) ₆	137	16.7	2	Nucleus	—	—
MLP124017-(HIS) ₆	150	18.5	1	Nucleus and cytosol	Topless/Topless-related proteins (TRPs)	—
MLP107772-(HIS) ₆	149	16.2	8	Chloroplasts and mitochondria	—	—
MLP124202-(HIS) ₆	398	45.4	1	Endomembranes	—	AvrM
MLP124266-(HIS) ₆	69	6.1	8	Nucleus and cytosol	—	AvrP4
MLP124499-(HIS) ₆	54	6	4	Nucleus and cytosol	—	—

Table 1. Features of the eleven *M. larici-populina* candidate effector proteins investigated in this study. ^aAs described in Petre *et al.*²² and/or Germain *et al.*²⁴. ^bImmunolocalization performed on infected poplar leaves by Hacquard *et al.*¹⁶.

secretion signal, absence of functional information, richness in cysteines, transcriptional regulation during infection, and/or presence of signatures of rapid evolution. Such features have been used as a basis for identifying candidate effectors^{15–18}. Due to the difficulty to genetically manipulate rust fungi and their host plants, only a handful of rust effectors have been reported so far¹⁹. Apart from their avirulence properties (i.e. recognition by plant immune receptors inside the cell), the functions of these rust effectors remain unknown or need to be clarified¹³. Effectoromic pipelines based on heterologous systems have been recently established to get insights about the plant cellular and molecular targets of candidate effectors, and thus to prioritize candidate effectors for further research^{20–24}. But so far, only one study has set up a small-scale effort using production of candidate effectors in bacterial system to unravel their structure and function²⁵. The identification of plant targets of effectors associated with structure/function analyses of recombinant effectors can reveal how they interact with plant partners and how co-evolution with the host plants promotes the diversification of surface-exposed amino acids^{1,11,26–29}. The avirulence proteins AvrL567, AvrM, and AvrP of the flax rust fungus *Melampsora lini* are the three effector structures described in rust fungi so far^{10,11,30}.

The poplar rust fungus *Melampsora larici-populina* is the causal agent of the poplar leaf rust disease. It causes important damages in poplar plantations across Europe³¹. It is also a model pathosystem to study tree-pathogen interaction. As such, recent research efforts have identified and initiated the characterization of *M. larici-populina* candidate effectors using transcriptomics and functional screens in heterologous plant systems such as *Nicotiana benthamiana* and *Arabidopsis thaliana*^{21,22,24,32}. These studies highlighted that *M. larici-populina* candidate effectors target multiple cell compartments and plant proteins; similar effectoromic screens set with other rust fungi have drawn the same conclusions²⁰.

In this study, we combined biochemical and structural approaches to explore further *M. larici-populina* candidate effector proteins. To this end, we used *Escherichia coli* as an heterologous system to express eleven candidate effectors that were previously described to target particular cell compartments and/or to interact with specific plant proteins and/or that are homologues of known rust avirulence effectors^{21,22,24,32}. Among the eleven selected proteins, three were successfully produced and purified from *E. coli* as recombinant proteins. We could determine the nuclear magnetic resonance (NMR) structures of two of them, highlighting structural similarities with Knottins and with Nuclear Transport Factor 2-like proteins.

Results

Selection of *M. larici-populina* candidate effector proteins. We selected 11 candidate effector proteins of *M. larici-populina* (Table 1), out of a catalogue of 24 candidate effectors previously screened *in planta* for their subcellular localization and plant protein partners²². Notably, we retained proteins showing (i) a specific and informative subcellular localization, such as nucleus (MLP109567), nucleolus (MLP124478), nuclear bodies (MLP124530), chloroplasts and mitochondria (MLP107772, aka CTP1), chloroplasts and aggregates (MLP124111), endomembranes (MLP124202), and plasmodesmata (MLP37347), (ii) specific plant protein partners (MLP124017, MLP37347, MLP124448, MLP124111), (iii) similarities with *M. lini* Avr effectors (MLP124530, MLP37347, MLP124202, MLP124266), or iv) proteins belonging to large families of small-secreted proteins (MLP124499, MLP124561).

Successful production and purification of three candidate effectors in *E. coli*. To investigate the structural properties of the 11 selected candidate effectors (Fig. 1), we first aimed at obtaining the recombinant proteins. To this end, the sequence encoding mature proteins (i.e. without signal peptide) were cloned into pET-26b (for *Mlp124111*, *Mlp124478*, *Mlp124530*, *Mlp124561*, *Mlp37347*, *Mlp109567*, *Mlp107772*, *Mlp124202*, *Mlp124266*, and *Mlp124499*) or pET-28a (for *Mlp124017*) expression vectors, in order to incorporate a C-terminal 6-histidine tag (Table S1). Small-scale expression assays achieved into *E. coli* Rosetta2 (DE3) pLys strain indicated that nine out of the eleven proteins accumulated using a standard induction protocol (i.e.

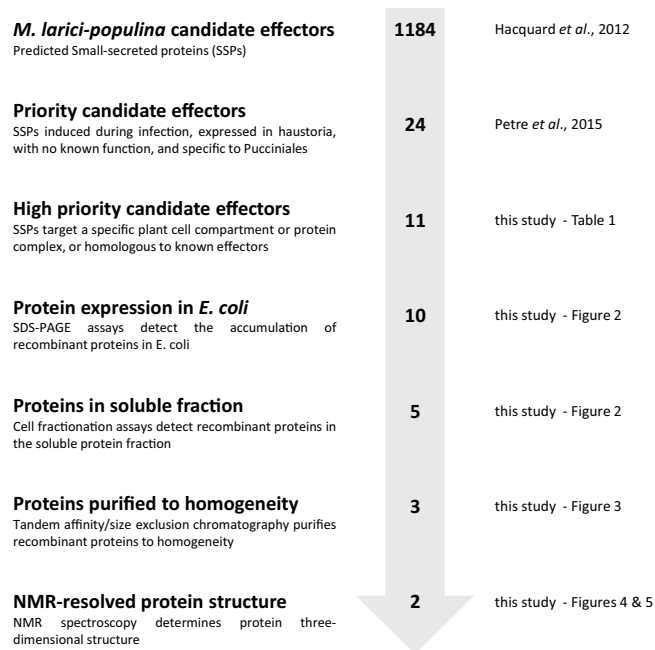


Figure 1. Overview of the effectoromics pipeline. A total of eleven *M. larici-populina* candidate effectors were selected from the previous study of Petre *et al.*²² (i.e. particular localization and/or specific plant interactors and/or homologies to *M. lini* Avr effectors). Effector candidates were expressed in *E. coli* SoluBL21 (DE3) pRARE2, Rosetta2 (DE3) pLysS or RosettaGami2 (DE3) pLysS strains. Soluble recombinant proteins were purified and their structure solved by NMR spectroscopy.

addition of 100 μ M IPTG in mid-exponential growth phase and further growing for 3 to 4 hours at 37 °C). Among those, five (MLP124111, MLP124561, MLP37347, MLP107772 and MLP124202) accumulated in the insoluble protein fraction, and one (MLP109567) was not expressed (Fig. 2), despite the use of other *E. coli* expression strains (SoluBL21 (DE3), Origami2 (DE3) pLysS, Rosetta-Gami2 (DE3) pLysS) and modification of the culture conditions (induction time, temperature, and osmolarity). Among the five remaining soluble proteins we choose MLP124017, MLP124266, and MLP124499, the most stable along the purification procedure, for further analyses. We thus purified the His-tagged recombinant proteins in native conditions using a two-step protocol including immobilized-metal affinity chromatography (IMAC), then size exclusion chromatography (Fig. 3). The purified proteins, yielding respectively 50 mg/L (cell culture), 0.5 mg/L, and 0.5 mg/L for MLP124017, MLP124266 and MLP124499, respectively, eluted in size exclusion chromatography as a single peak corresponding to an estimated apparent molecular mass compatible with a monomeric organization data not shown.

MLP124266 is a thermostable protein that exhibits a cystine knot. From a previous study, we reported that the Mlp124266 and Mlp124499 genes are strongly expressed and induced during poplar leaf colonization by *M. larici-populina*, and belong to large multigene families of 13 and 31 members, respectively, in *M. larici-populina*¹⁶. Mlp124266 and Mlp124499 encode mature proteins of 69 and 50 amino acids, respectively (Fig. S1). MLP124266 has an N-terminal part enriched in charged residues and a C-terminal region that possesses six conserved cysteines predicted to form a cystine knot structure (Fig. S1A). This typical protein organization is shared by all members of the family as well as by alleles of *M. lini* AvrP4^{33,34}. In MLP124499, several acidic residues are found in the N-terminal part whereas the C-terminal part contains three conserved cysteines (Fig. S1B). Prediction programs indicate that all members of both protein families exhibit highly conserved N-terminal signal peptides for protein secretion. Following the production and the purification of both MLP124266 and MLP124499, we undertook a structural characterization of each recombinant protein using a NMR spectroscopy approach.

Standard homonuclear 2D experiments and ¹⁵N-edited TOCSY-HSQC and NOESY-HSQC experiments carried out on MLP124499 allowed the assignment of ¹H and ¹⁵N resonances except for the four N-terminal residues (Fig. S2A). However, several minor peaks were observed, especially for Ala₁₄-Glu₁₆, Gly₂₅-Gln₂₆, Glu₃₀, Asp₄₉ residues, suggesting the presence of multiple forms or conformations. Changing the temperature and the ionic strength, or adding dithiothreitol failed to improve the quality of the MLP124499 NMR spectra. Consequently, very few experimental restraints could be derived and structure calculations led to very ill-defined models.

For MLP124266, the assignment of ¹H, ¹⁵N and ¹³C resonances has been obtained for all residues except for the five N-terminal amino acids (Fig. S2B) and its 3D structure could further be modelled by the NMR derived constraints (Table S2; Figs. 4 and S3). This analysis showed that the Cys₃₆-Leu₆₉ C-terminal region exhibits a typical cystine knot structure involving three disulfide bonds (Cys₃₉-Cys₅₅, Cys₄₄-Cys₅₈, and Cys₅₀-Cys₆₄), a β -sheet composed of anti-parallel strands between Thr₄₂-Cys₄₄, Gly₅₇-Ser₅₉ and Val₆₃-Val₆₅, and a short α -helix formed by the Gln₄₉-Ala₅₂ segment. In contrast, the Met₁-Asp₃₅ N-terminal region displays large structural disorder, as

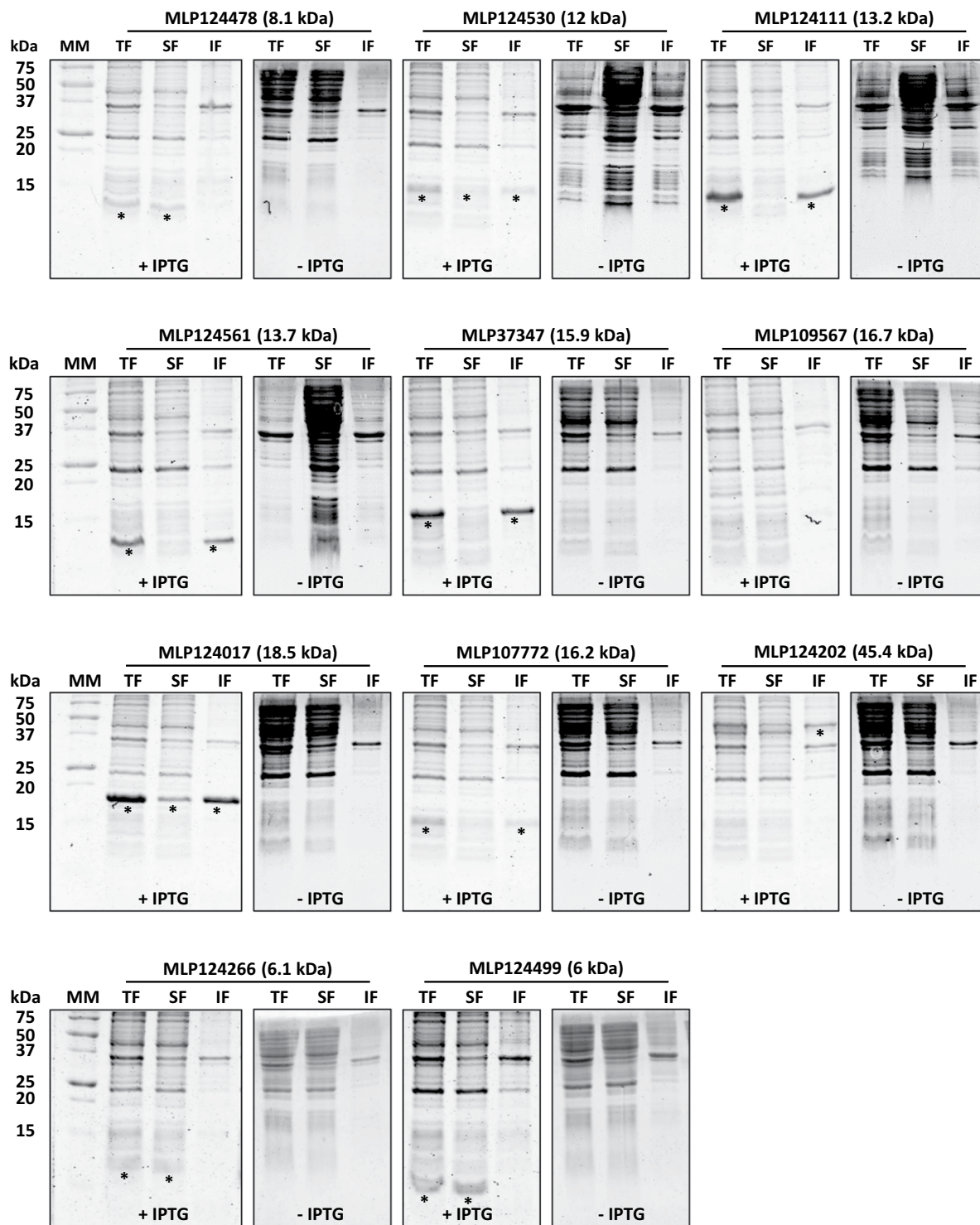


Figure 2. Small-scale expression test of selected candidate effector proteins carried out in *Escherichia coli* Rosetta2 (DE3) pLysS expression strain. Coomassie blue-stained sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of total (TF), soluble (SF) and (IF) insoluble protein fractions of *E. coli* Rosetta2 pLysS expression strain grown in presence (+) or in absence (–) of 0.1 mM isopropyl β -D-thiogalactopyranoside (IPTG). Asterisks indicate the expected migration of overexpressed proteins. MM: molecular mass marker.

shown by the superposition of the 20 NMR models (Fig. S3). Very few NOE correlations were indeed observed for residues 1 to 35. A few sequential and medium-range NOE correlations characteristic of transient helical conformations can however be noticed (Fig. S3) and explain the presence of short secondary structures in some of these 20 models. Indeed, residues 8–17 and 28–30 exhibit helical structures in 30 to 50% and around 25% of

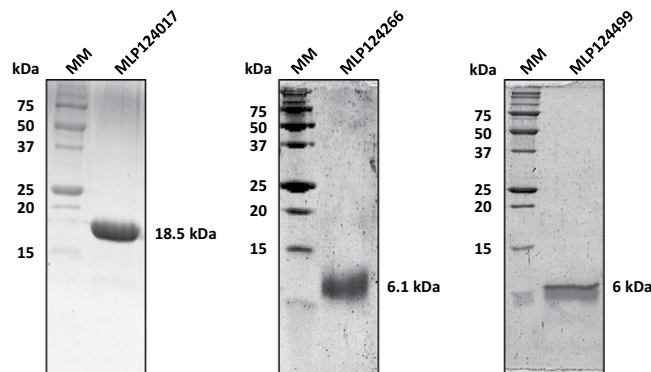


Figure 3. Candidate effectors purified as soluble proteins. Ten micrograms of recombinant MLP124017-(His)₆; MLP124266-(His)₆; MLP124499-(His)₆ have been separated by SDS-PAGE (17%). Molecular mass corresponding to each purified protein is indicated. MM: molecular mass marker.

the models, respectively. Backbone dynamic properties of MLP124266 have been investigated by ¹⁵N relaxation measurements. Heteronuclear ¹H-¹⁵N NOE values showed a contrasted profile with low values for N-terminal residues (indicative of a flexible structure) and high values for C-terminal residues (indicative of a rigid structure). Indeed, amino acids Asp₆ to Gly₃₈ and Cys₃₉ to Leu₆₉ presented heteronuclear NOE averaged values of 0.26 ± 0.05 and 0.66 ± 0.10 , respectively. Several secondary structure prediction softwares predict a helix between residues 11 and 18 (data not shown) and a Consurf analysis shows that the proportion of conserved amino acids in the N-terminal region is much higher than in the C-terminal part (Fig. S4). Interestingly 6 out of the 8 cysteines are gathered in the C-terminal region between Cys₃₉ and Cys₆₉, following a spacing (Cys-X₂₋₇-Cys-X₃₋₁₀-Cys-X₀₋₇-Cys-X₁₋₁₇-Cys-X₄₋₁₉-Cys) typical of cystine knot structures, (i.e., three intricate disulfide bridges that confer very high stability to proteins; Fig. 4)³⁵. Hence, it is likely that rigidity originates from the structure formed by these cysteines that are highly conserved in the protein family, as indicated by the Consurf analysis (Figs. S1 and S4). Thus, we sought to determine whether these disulfides are formed and whether they influence the stability and/or the oligomerization state of the protein by covalent bonds. A single peak corresponding to the theoretical mass of MLP124266 monomer was obtained by mass spectrometry (data not shown). The titration of free thiol groups in an untreated recombinant MLP124266 gave an averaged value of 1 mole SH per mole of protein. Considering the presence of 8 cysteines in the protein, these results are consistent with the existence of three intramolecular disulfide bridges (Fig. S3). The thermostability of recombinant protein was estimated by heating the protein for 10 min at 95 °C. The observation that the protein remained in solution (*i.e.* no precipitation was observed) indicates that it is thermosoluble. In order to investigate the role of the disulfides for such property, we should compare the results obtained with an oxidized and a reduced protein. However, as assessed by thiol titration experiments, we failed to obtain a complete reduction of these disulfides despite extensive incubation of the protein at high temperatures, in denaturing and reducing environments. Altogether, these results indicate that recombinant MLP124266 is properly folded by *E. coli*, and that the disulfide bridges, which are partially resistant to reduction, confer a high rigidity and stability to the protein.

MLP124017 is part of the nuclear transport factor 2-like protein superfamily. MLP124017 is a small-secreted protein (167 amino acids with its signal peptide; 150 amino acids in its mature form, with a molecular mass of 18 kDa) of unknown function, highly expressed during infection of poplar leaves by *M. larici-populina*¹⁶. MLP124017 shares sequence similarity with neither other *M. larici-populina* nor other rust fungal proteins. In a previous study, we demonstrated the nucleocytoplasmic localization of MLP124107 in *N. benthamiana* and its interaction with poplar TOPLESS-related 4 protein²². To further investigate MLP124017 structure and to get insights into its function, we first attempted to solve its 3D structure by crystallization coupled to X-ray diffraction. We were unable to obtain exploitable diffracting crystals despite the use of different versions (untagged or N- or C-terminal His-tagged) of MLP124017 protein and therefore switched to NMR. The recombinant ¹⁵N and ¹³C-labelled MLP124017 protein was used for structure determination by two- and three-dimensional NMR experiments (Table S3). The assigned ¹H, ¹⁵N-HSQC spectra were well dispersed but the peaks for residues from the N-terminal 1–14 and 86–95 segments were missing (Fig. S5). From preliminary structures, the production of a truncated recombinant protein for the first eight N-terminal residues that could mask residues 86–95 did not improve the data. The solution structure of MLP124017 was determined based on 1727 NOE-derived distance restraints, 214 dihedral angle restraints and 102 hydrogen bond restraints. All proline residues have been determined to be in a *trans*-conformation according to the ¹³Cβ chemical shift at 32.21; 32.46; 31.02 and 32.40 ppm for Pro₃₆; Pro₅₁; Pro₅₄ and Pro₁₄₆ respectively. The best conformers with the lowest energies, which exhibited no obvious NOE violations and no dihedral violations >5° were selected for final analysis. The Ramachandran plot produced shows that 99.6% of the residues are in favoured regions (Table S4). MLP124017 structure is composed of a α + β barrel with seven β-strands forming one mixed β-sheet, four β-hairpins, four β-bulges, and four α-helices (Fig. 5A). Residues 1–14 and 150–151 having missing assignments are not defined in the final models. This arrangement of secondary structure produces a cone-shaped fold for the protein, which generates a distinctive hydrophobic cavity (Fig. 5B, Fig. 5C).

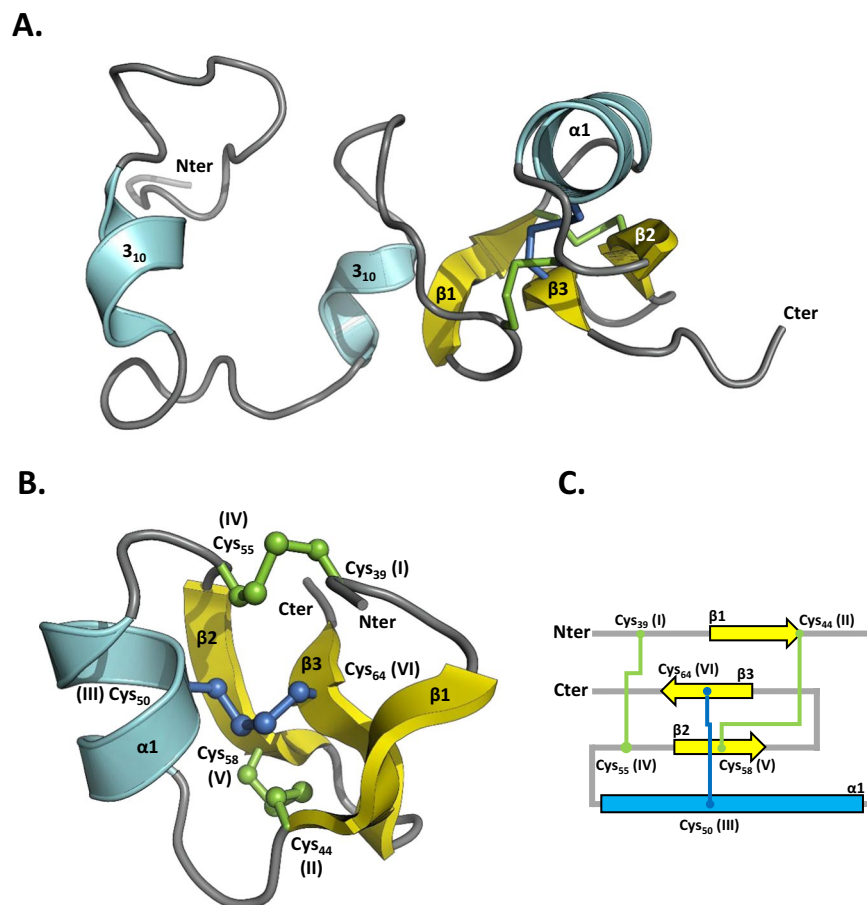


Figure 4. NMR spectroscopy solution structure of MLP124266. **(A)** The structure of MLP124266 is represented as cartoon and consists of one α -helix in cyan and a mixed β -sheet composed by three β -strands coloured in yellow. **(B)** The C-terminal small disulfide-rich domain is related to the structural family of knottins, which contain at least 3 disulfide bridges and 6 cysteines and implies a disulfide between cysteines III and VI going through disulfides I-IV and II-V. In MLP124266, a disulfide bridge between Cys₅₀ (III) and Cys₆₄ (VI), in blue, goes through disulfide bridges between Cys₃₉ (I) and Cys₅₅ (IV) as well as Cys₄₄ (II) and Cys₅₈ (V), in green. **(C)** Schematic representation of disulfide bridge connectivity. Disulfide bridge Cys₅₀ (III)-Cys₆₄ (VI) is represented in blue and disulfide bridges between Cys₃₉ (I)-Cys₅₅ (IV) and Cys₄₄ (II)-Cys₅₈ (V) are represented in green. Helix α (residues 47 to 53) is coloured in cyan.

To identify potential structural homologs of MLP124017, we performed structural similarity searches using the Dali server³⁶. Queries identified SBAL_0622 (PDB code 3BLZ) and SPO1084 (PDB code 3FKA) as the closest structural homologs with the highest Z-score of 6.0 and 6.3, respectively, and a RMSD of 4.1 and 3.5 Å, respectively (Fig. S6). These two proteins, which are from the bacteria *Shewanella baltica* and *Ruegeria pomeroyi*, have no known function, but share a common Nuclear Transport Factor 2-like (NTF2) fold. The NTF2 superfamily comprises a large group of proteins that share a common fold and that are widespread in both prokaryotes and eukaryotes³⁷. Taken together these results show that although MLP124017 do not share sequence similarities or domain with other proteins in sequence databases, its structure is similar to proteins of the NTF2 folding superfamily.

Discussion

In this study, we have set up a small-throughput effectormics pipeline based on recombinant protein production and structural characterization to get insights on 11 candidate effectors of the poplar leaf rust fungus *M. larici-populina*.

Although the production of recombinant proteins in *E. coli* is a valuable approach to perform biophysical and biochemical analyses of candidate effectors^{5,25}, we have faced issues for the production of soluble small-cysteine rich proteins in this system. Indeed, among the eleven candidate effectors screened for expression, only three were found in the soluble protein fraction and stable enough to tolerate the purification procedure. Although we tested different *E. coli* strains and protein expression induction protocols, the other candidate effectors were either not expressed or expressed as inclusion bodies. It is possible to purify recombinant proteins from inclusion bodies by using denaturing extraction conditions and further refolding proteins³⁸. However, this approach is not recommended for structural analysis as the refolding of the proteins *in vitro* may alter folding. Another limit of prokaryote systems to produce eukaryote proteins is the lack of post-translational modifications such as methylations.

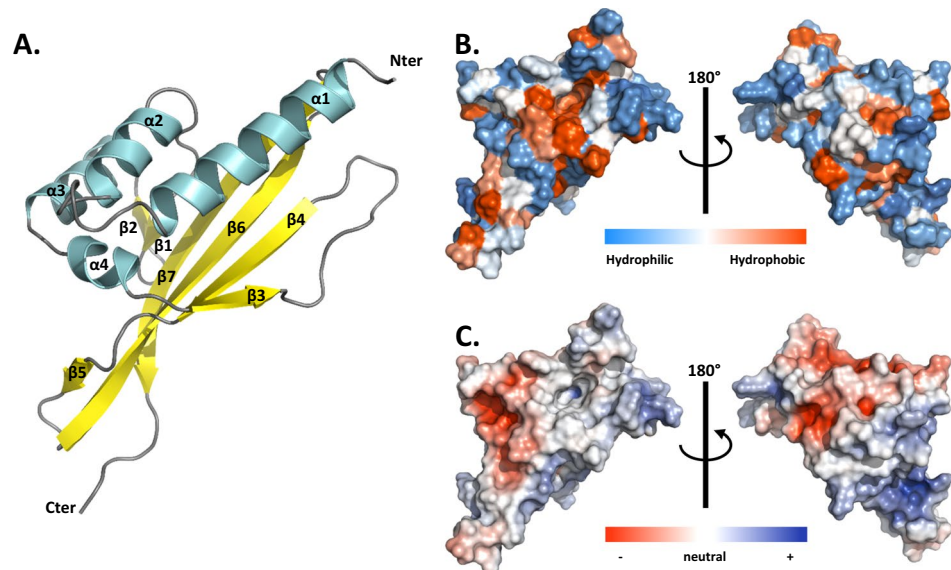


Figure 5. NMR spectroscopy solution structure of MLP124017. (A) The structure of MLP124017 is represented as cartoon and consists of four α -helices in cyan (α 1: residues 16–32; α 2: 41–48; α 3: 65–73; α 4: 80–82) and a mixed β -sheet composed by seven β -strands coloured in yellow (β 1: residues 55–58; β 2: 61–63; β 3: 84–86; β 4: 98–104; β 5: 108–109; β 6: 115–129; β 7: 132–144). (B) Front and rear views on the surface of MLP124017 illustrating the surface hydrophobic potential. The hydrophobic and hydrophilic patches are shown in red and in blue respectively. (C) Front and rear views on the surface of MLP124017 illustrating the surface electrostatic Coulomb potential at pH 7.0 using APBS plugin from Pymol 2.3 software with a contour of -10 kT/e to 10 kT/e. The positive-charge and negative-charge densities are coloured in blue and red respectively.

An alternative to this is to use the yeast *Pichia pastoris*, which has proven useful for several fungal effectors such as *Leptosphaeria maculans* AvrLm4–7 or *Cladosporium fulvum* Avr2 and Avr4^{39,40}. We have tried this system to produce the candidate effector (MLP107772), but without success (data not shown). Nevertheless, this system may be useful and deserves to be considered as an alternative to assay other rust effectors for which we were not able to obtain production in *E. coli*.

Out of the three effectors successfully purified as recombinant proteins, two were structurally characterized by NMR spectroscopy. MLP124266 is a homolog of the *M. lini* AvrP4 effector protein⁴¹, and we showed that it exhibits a cystine knot (or knottin) structural motif commonly encountered in small disulfide-rich proteins. MLP124017 is an orphan protein in *M. larici-populina* with no known ortholog in Pucciniales. MLP124017 physically associates with the poplar TOPLESS-related protein 4 (TRP4)²², and we showed that it exhibits a fold similar to two bacterial proteins that belong to the Nuclear-Transport factor 2-like protein superfamily.

We showed that MLP124266 possesses two distinct regions with contrasted structural properties. The C-terminal part is rigidified by a cystine knot motif whereas the N-terminal part is globally flexible. The knottin folded proteins display a variety of functions such as venoms and spider toxins^{42,43} but also antimicrobial properties such as the cyclotides⁴⁴. Some are also found to interact with protease inhibitors found in plants, insects and plant parasites⁴⁵. The three disulfide bridges within the C-terminal part of MLP124266 confer its rigidity and probably contribute to the high protein stability³⁵. MLP124266 presents a β -sheet structure typical of knottins, but interestingly it also has an additional helix between β 2 and β 3 strands. To our knowledge, the presence of such a helix in knottins has been reported in cyclotides only, and more precisely in bracelet cyclotides containing six or seven residues in the loop between Cys(III) and Cys(IV)⁴⁶. This loop often contains an alanine, which favours the formation of the helix as well as a highly conserved glycine allowing its connection to the cystine knot⁴⁷. Interestingly, the loop in MLP124266 has such residues, i.e. Ala₅₂ and Gly₅₄. Ala and Gly, but consists of four residues only. In *Viola odorata* cycloviolacin O2 (cO2), the additional helix is located in a hydrophobic loop that interacts with the membrane-mimicking micelles⁴⁸. Therefore, it might help disrupting membranes and thus contribute to the cytotoxicity activity of cO2 and play a role in plant defence. In MLP124266, the helical turn is not particularly hydrophobic (Fig. S4B) and may not have these properties. To our knowledge, MLP124266 is the first fungal protein to present a knottin-like structure⁴⁹. It would be interesting to collect structural data from other potential fungal knottins to find out whether the additional helix is always present and to clarify its role.

The intrinsic disorder of the N-terminus of MLP124266 also deserve to be pointed out. This region, approximately extending up to residue 35 and thus representing half of the primary sequence, globally presents a high flexibility, as demonstrated by the NMR dynamic results. Nevertheless, a few residues exhibit a propensity to form helical structures, which may support a biochemical role that remains to be elucidated. Interestingly, other effectors possess a predicted disordered N-terminal region²⁹. For instance, *M. lini* effectors AvrL567 and AvrM have predicted disordered N-terminal regions that are susceptible to protease degradation^{11,50}. Flexible folds are known to be adaptable linkers that favour the ability to bind partners⁵¹. As the N-terminus of many cytoplasmic effectors is anticipated to mediate protein entry into host cells², it is tempting to speculate that this flexible part may bind a target important for cell entry.

The structure of MLP124017 solved by NMR spectroscopy showed a fold similar to members of the NTF2-like superfamily. The NTF2-like superfamily is a group of protein domains sharing a common fold, but showing no sequence similarity. MLP124017 is structurally similar to two bacterial proteins, despite the lack of sequence similarity. The structures of these two bacterial proteins consist of a β -sheet surrounding a binding pocket and α -helices acting as a lid⁵². The NTF2 family regroups catalytic and non-catalytic proteins that contain cone-like structured proteins with a cavity that often acts as a molecular container involved in a wide range of cellular functions³⁷. Interestingly, the cone-shaped structure of MLP124017 is widespread across both prokaryotes and eukaryotes. The first proteins of the NTF2 family were reported to play a role in the transport of molecules from the cytoplasm to the nucleus. Arabidopsis NTF2 protein is required to import nuclear proteins via the recognition of a nuclear localization signal (NLS). This protein also plays a role in the nuclear import of the small-GTPase Ran-GDP that is a central protein in various signal transduction pathways (e.g. mitotic spindle formation, nuclear envelope assembly, or responses to biotic stresses)^{53–56}. In bacteria, some NTF2-like proteins play a role in bacterial conjugation as part of the type IV secretion system⁵⁷, whereas non-catalytic NTF2-like domains act as immunity proteins⁵⁸. In fungi, *Saccharomyces cerevisiae* NTF2 mutant is defective for nuclear import⁵⁹. Although NTF2 folded proteins are widespread across kingdoms, very few is known about their role. A recent study presented that the silencing of NTF2 in wheat decreased the resistance against avirulent isolates of the wheat stripe rust fungus *P. striiformis* f. sp. *tritici*⁶⁰. Since MLP124017 has been shown to interact with TOPLESS and TOPLESS-related proteins²², it is tempting to speculate that the cavity formed by the β -sheet could be involved in the association with these plant partners.

Although MLP124266 and MLP124017 show no primary sequence similarity to known proteins, they adopt a three dimensional fold similar to knottins and NTF2 family members, respectively. Thus, knowing the structure of both candidate effectors allowed us to classify them as members of large superfamilies of proteins. The concept of structural families whose members share no, or very limited, primary sequence homology emerges in effector biology⁶¹. This concept promises to revolution the way we predict and categorize effector proteins^{6,7}. For instance, members of the MAX effector family share a common β -sandwich fold, but show no primary sequence similarity⁶. Similarly, members of the WY superfamily of RXLR effectors in oomycetes share a common three- to four-helix bundle²⁹. Such features are now used to search and categorize fungal and oomycete effector proteins into structural superfamilies^{1,6}. To our knowledge, MLP124266 and MLP124017 are the first effector proteins to adopt knottins and NTF2 folds. Whether other effector proteins adopt similar folds remains to be identified to determine if knottins and NTF2 folds define structural superfamilies in fungi.

Experimental Procedures

Sequence analyses and names. Alignment and phylogenetic analyses were performed on the phylogeny website (www.phylogeny.fr) with default parameters. Alignments were corrected and edited manually, and phylogenetic trees were generated with FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>). Physical and chemical parameters of proteins were estimated using protparam tool (<http://web.expasy.org/protparam/>). Common names and JGI ID of genes described in this study are as follow: *Mlp124266*, *Mlp124499*, *Mlp124111*, *Mlp124478*, *Mlp124530*, *Mlp124561*, *Mlp37347*, *Mlp109567*, *Mlp124017*, *Mlp107772*, and *Mlp124202*. The mapping of the family-wide conservation pattern of amino acids onto MLP124266 structure was performed with Consurf (<http://consurf.tau.ac.il/2016/>).

Cloning of selected effector encoding sequences. Open reading frames coding for the mature forms (i.e. devoid of the sequence encoding N-terminal secretion peptide) of MLP124266 and MLP124499 of *M. larici-populina* isolate 98AG31 were ordered as synthetic genes cloned in pBSK(+) vectors (Genecust). Coding sequence of the mature forms of the nine other candidate effectors (*Mlp124111*, *Mlp124478*, *Mlp124530*, *Mlp124561*, *Mlp37347*, *Mlp109567*, *Mlp124017*, *Mlp107772*, and *Mlp124202*) were amplified by polymerase chain reaction (PCR) using cDNAs from leaves of the poplar hybrid Beaupré infected by *M. larici-populina* (isolate 98AG31) and further cloned into pICSL01005 vector as described previously²². The sequences encoding the mature form of each effector were subsequently cloned by PCR in either pET26b or pET28a vector between *NdeI* and *XhoI* (or *NotI*) or *NcoI* and *XhoI* restriction sites, respectively, using primers shown in Table S1.

Expression and purification of recombinant proteins in *Escherichia coli*. Expression of recombinant proteins was performed at 37°C using the *E. coli* SoluBL21 (DE3) pRARE2 (Amsbio Abington, UK), Rosetta2 (DE3) pLysS, Origami2 (DE3) pLysS or RosettaGami2 (DE3) pLysS strains (Novagen) containing the adequate pET expression vector coding for the selected candidate effector (Table S1) in LB medium supplemented with kanamycin (50 μ g/ml) and chloramphenicol (34 μ g/ml). When the cell culture reached an OD_{600nm} of 0.7, protein expression was induced with 0.1 mM isopropyl- β -D-1-thiogalactopyranoside (IPTG) and cells were grown for a further 4h. To improve the solubility of some recombinant candidate effectors, other protocols were used as follows. First, we added 0.5% (v/v) of ethanol in the medium when culture reached an OD_{600nm} of 0.7. The cells were cooled to 4°C for 3h, recombinant protein expression induced with 0.1 mM IPTG and cells further grown for 18h at 20°C. We also tested a combination of an osmotic and a thermal shock⁶². When the culture reached an OD_{600nm} of 0.5–0.6, 500 mM NaCl and 2 mM of betaine were added to the culture medium and the culture incubated at 47°C for 1 hour under stirring. Cells were then cooled to 20°C and the expression of recombinant proteins induced with 0.1 mM IPTG. After induction, cells were harvested by centrifugation, suspended in a 30 mM Tris-HCl pH 8.0 and 200 mM NaCl lysis buffer, and stored at –20°C. Cell lysis was completed by sonication (three times for 1 min with intervals of 1 min). The cell extract was then centrifuged at 35 000 g for 25 min at 4°C to remove cellular debris and aggregated proteins. After the addition of 10 mM imidazole, soluble fraction containing C-terminal His-tagged recombinant proteins were then purified by gravity-flow chromatography on a nickel nitrilotriacetate (Ni-NTA) agarose resin (Qiagen). After a washing step with lysis buffer supplemented with 20 mM imidazole, the proteins were eluted using lysis buffer containing 250 mM imidazole. The fractions

of interest were pooled, concentrated by ultrafiltration then injected onto a gel filtration HiLoad 16/600 Superdex 75 prep grade (GE Healthcare) column connected to an ÄKTA Purifier™ (GE Healthcare) and eluted with lysis buffer. The fractions containing recombinant MLP124017 were pooled, concentrated, and stored at -20°C as such, whereas for MLP124266 and MLP124499 fractions were pooled, dialyzed against 30 mM Tris-HCl, 1 mM EDTA (TE) pH 8.0 buffer, and stored at 4°C until further use.

For the NMR spectroscopy analyses, MLP124266-(His6) and MLP124499-(His6) recombinant proteins were ^{15}N -labelled in M9 minimal synthetic medium containing $^{15}\text{NH}_4\text{Cl}$ (1 g/L). MLP124017 was single ^{15}N or double ^{15}N and ^{13}C labelled in M9 minimal medium containing 1 g/l NH_4Cl (^{15}N) and 2 g/l glucose (^{13}C) supplemented with 2,5% thiamine (m/v), 1 mg/ml biotin, 50 mM FeCl_3 , 10 mM MnCl_2 , 10 mM ZnSO_4 , 2 mM CoCl_2 , 2 mM NiCl_2 , 2 mM NaSeO_3 , and 2 mM H_3BO_3 . After purification as described above, labelled MLP124266, MLP124499, and MLP124017 were dialyzed against a 50 mM phosphate pH 6.0 or a 20 mM phosphate pH 6.8 buffer supplemented with 200 mM NaCl. The homogeneity of purified proteins was checked by SDS-PAGE and protein concentration determined by measuring the absorbance at 280 nm and using theoretical molar absorption coefficients of $500\text{M}^{-1}\cdot\text{cm}^{-1}$, $3\,105\text{M}^{-1}\cdot\text{cm}^{-1}$, $29\,450\text{M}^{-1}\cdot\text{cm}^{-1}$ deduced from the amino acid sequences of mature MLP124266, MLP124499, and MLP124017 proteins respectively. For MLP124266, protein concentration was also verified using a colorimetric assay (BC assay, Interchim).

Protein sample preparation for NMR spectroscopy. Uniformly labelled ^{15}N MLP124017 (1 mM in 20 mM phosphate pH 6.8, 200 mM NaCl) was supplemented with 5 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) in D_2O as a lock/reference. For the D_2O experiments, the sample was lyophilized and dissolved in 200 μL D_2O . For the 3D heteronuclear experiments, a $^{13}\text{C}/^{15}\text{N}$ labelled sample was diluted at a final concentration of 0.6 mM in 200 μL of the previous phosphate buffer supplemented with 10% D_2O and 0.5 mM DSS as a reference. MLP124266 and MLP124499 samples were dissolved in 50 mM phosphate pH 6.0 buffer with 10% D_2O and 0.02% sodium azide. The concentration of unlabelled MLP124266 and MLP124499 was 0.85 and 0.65 mM, and the one for uniformly ^{15}N labelled MLP124266 and MLP124499 was 1.8 and 0.135 mM, respectively.

Nuclear magnetic resonance spectroscopy. For MLP124017, spectra were acquired on 800 and 700 MHz Avance Bruker spectrometers equipped with triple-resonance (^1H , ^{15}N , ^{13}C) z-gradient cryo-probe at 298 K. Experiments were recorded using the TOPSPIN pulse sequence library (v. 2.1) (Table S2). All spectra are referenced to the internal reference DSS for the ^1H dimension and indirectly referenced for the ^{15}N and ^{13}C dimensions⁶³. Sequential assignment was performed using 3D ^{15}N -NOESY-HSQC, ^{15}N -TOCSY-HSQC, HNCACO, HNCACO, HNCA, HNCOCACB, and HNCACB. Side chain ^1H assignments were carried out using combined analysis with 3D ^{15}N -NOESY-HSQC, ^{15}N -TOCSY-HSQC, and 2D NOESY and TOCSY with D_2O samples. A series of three HSQC spectra was performed after lyophilisation and dilution of the first sample in D_2O to determine amide protons in slow exchange (Table S2).

For MLP124266 and MLP124499, NMR spectra were acquired on a Bruker DRX 600 MHz spectrometer equipped with a TCI cryoprobe. For MLP124266 and MLP124499, COSY, TOCSY (mixing time of 60 ms) and NOESY (mixing time of 150 ms) experiments were run at 298 K, respectively. For MLP124266, HNHA, HNHB, R1 and R2 ^{15}N relaxation rates, ^1H - ^{15}N heteronuclear NOE, HNCA (with 24 (^{15}N) \times 28 (^{13}C) complex points and 192 transients per increment) standard experiments were recorded. Spectra were processed using Topspin® 3.0 software (Bruker) and analysed with NmrView⁶⁴, CcpNmr⁶⁵ and ARIA2⁶⁶.

Structure calculation. For MLP124017 structure calculation, NOE peaks identified in 3D ^{15}N -NOESY-HSQC and 2D NOESY experiments were automatically assigned during structure calculations performed by the program CYANA 2.1⁶⁷. The ^{15}N , H_N , ^{13}C , $^{13}\text{C}\alpha$, $\text{H}\alpha$, and $^{13}\text{C}\beta$ chemical shifts were converted into φ/ψ dihedral angle constraints using TALOS+ (v. 1.2)⁶⁸. Hydrogen bond constraints were determined according to $^1\text{H}/^2\text{H}$ exchange experiments of backbone amide protons (H_N). Each hydrogen bond was forced using following constraints: 1.8–2.0 Å for $\text{H}_\text{N}\cdots\text{O}$ distance and 2.7–3.0 Å for $\text{N}_\text{H}\cdots\text{O}$ distance. Final structure calculations were performed with CYANA (v. 2.1) using all distance and angle restraints (Table S3). 600 structures were calculated with CYANA 2.1, of which the 20 conformers with the lowest target function were refined by CNS (v. 1.2) using the refinement in water of RECOORD⁶⁹ and validated using PROCHECK⁷⁰.

For MLP124266 structure calculation, 783 NOE peaks identified in 3D ^{15}N -NOESY-HSQC or 2D NOESY spectra, 75 ϕ/ψ dihedral angles generated by DANGLE⁷¹ and 10 hydrogen bonds were input as unambiguous restraints in ARIA2. Covalent disulfide bonds between Cys₃₉-Cys₅₅, Cys₄₄-Cys₅₈ and Cys₅₀-Cys₆₄ were also introduced. Among the 400 structures generated by ARIA2, 20 models of lowest energy were refined in water (Table S2).

NMR assignment and structure coordinates have been deposited in the Biological Magnetic Resonance Data Bank (BMRB code 34423 and 34298) and in the RCSB Protein Data Bank (PDB code 6SGO and 6H0I), respectively.

Data availability

The data that support the findings of this study are openly available in the Biological Magnetic Resonance Data Bank (<http://www.bmrb.wisc.edu/>) and in the RCSB Protein Data Bank (<http://www.rcsb.org/>).

Received: 4 September 2019; Accepted: 31 October 2019;

Published online: 02 December 2019

References

1. Win, J., Chaparro-Garcia, A., Belhaj, K. *et al.* Effector Biology of Plant-Associated Organisms: Concepts and Perspectives. *Cold Spring Harb Symp Quant Biol.* 77, 235–247 (2012).
2. Petre, B. & Kamoun, S. How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biol.* 12, e1001801 (2014).

3. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
4. Persoons, A. *et al.* The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Mol. Ecol.* **26**, 1902–1918 (2017).
5. Franceschetti, M. *et al.* Effectors of Filamentous Plant Pathogens: Commonalities amid Diversity. *Microbiol. Mol. Biol. Rev. MMBR* **81** (2017).
6. de Guillen, K. *et al.* Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi. *PLoS Pathog.* **11**, e1005228 (2015).
7. Joe, W. *et al.* Sequence Divergent RXLR Effectors Share a Structural Fold Conserved across Plant Pathogenic Oomycete Species. *PLoS Pathogens* **8**(1), e1002400 (2012).
8. Illergård, K., Ardell, D. H. & Elofsson, A. Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. *Proteins* **77**, 499–508 (2009).
9. Wirthmueller, L., Maqbool, A. & Banfield, M. J. On the front line: structural insights into plant-pathogen interactions. *Nat. Rev. Microbiol.* **11**, 761–776 (2013).
10. Ve, T. *et al.* Structures of the flax-rust effector AvrM reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity. *Proc. Natl. Acad. Sci. USA* **110**, 17594–17599 (2013).
11. Wang, C.-I. A. *et al.* Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *Plant Cell* **19**, 2898–2912 (2007).
12. Dean, R. *et al.* The Top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* **13**, 414–430 (2012).
13. Lorrain, C., Gonçalves dos Santos, K. C., Germain, H., Hecker, A. & Duplessis, S. Advances in understanding obligate biotrophy in rust fungi. *New Phytol.* **222**, 1190–1206 (2019).
14. Duplessis, S., Bakkeren, G. & Hamelin, R. Advancing Knowledge on Biology of Rust Fungi Through Genomics. in *Advances in Botanical Research* vol. 70 173–209 (Elsevier, 2014).
15. Saunders, D. G. O. *et al.* Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS One* **7**, e29847 (2012).
16. Hacquard, S. *et al.* A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Mol. Plant-Microbe Interact. MPMI* **25**, 279–293 (2012).
17. Nemri, A. *et al.* The genome sequence and effector complement of the flax rust pathogen *Melampsora lini*. *Front. Plant Sci.* **5**, 98 (2014).
18. Cantu, D. *et al.* Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorially expressed secreted proteins as candidate effectors. *BMC Genomics* **14**, 270 (2013).
19. Petre, B., Joly, D.L. & Duplessis, S. Effector proteins of rust fungi. *Frontiers in Plant Science* **5**, 416, <https://doi.org/10.3389/fpls.2014.00416> (2014).
20. Lorrain, C., Petre, B. & Duplessis, S. Show me the way: rust effector targets in heterologous plant systems. *Curr. Opin. Microbiol.* **46**, 19–25 (2018).
21. Petre, B. *et al.* Heterologous Expression Screens in *Nicotiana benthamiana* Identify a Candidate Effector of the Wheat Yellow Rust Pathogen that Associates with Processing Bodies. *PLOS ONE* **11**, e0149035 (2016).
22. Petre, B. *et al.* Candidate Effector Proteins of the Rust Pathogen *Melampsora larici-populina* Target Diverse Plant Cell Compartments. *Mol. Plant. Microbe Interact.* **28**, 689–700 (2015).
23. Qi, M. *et al.* Suppression or Activation of Immune Responses by Predicted Secreted Proteins of the Soybean Rust Pathogen *Phakopsora pachyrhizi*. *Mol. Plant-Microbe Interact. MPMI* **31**, 163–174 (2018).
24. Germain, H. *et al.* Infection assays in *Arabidopsis* reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens. *Mol. Plant Pathol.* **19**, 191–200 (2018).
25. Zhang, X. *et al.* Production of small cysteine-rich effector proteins in *Escherichia coli* for structural and functional studies. *Mol. Plant Pathol.* **18**, 141–151 (2017).
26. Yaeno, T. *et al.* Phosphatidylinositol monophosphate-binding interface in the oomycete RXLR effector AVR3a is required for its stability in host cells to modulate plant immunity. *Proc. Natl. Acad. Sci. USA* **108**, 14682–14687 (2011).
27. Chou, S. *et al.* Hyaloperonospora arabidopsidis ATR1 effector is a repeat protein with distributed recognition surfaces. *Proc. Natl. Acad. Sci. USA* **108**, 13323–13328 (2011).
28. Leonelli, L. *et al.* Structural elucidation and functional characterization of the *Hyaloperonospora arabidopsidis* effector protein ATR13. *PLoS Pathog.* **7**, e1002428 (2011).
29. Boutemy, L. S. *et al.* Structures of *Phytophthora* RXLR effector proteins: a conserved but adaptable fold underpins functional diversity. *J. Biol. Chem.* **286**, 35834–35842 (2011).
30. Zhang, X. *et al.* Crystal structure of the *Melampsora lini* effector AvrP reveals insights into a possible nuclear function and recognition by the flax disease resistance protein P. *Mol. Plant Pathol.* **19**, 1196–1209 (2018).
31. Pinon, J. & Frey, P. Interactions between poplar clones and *Melampsora* populations and their implications for breeding for durable resistance. In *Rust diseases of willow and poplar* (eds Pei, M. H. & McCracken, A. R.) 139–154 (CABI), <https://doi.org/10.1079/9780851999999.0139> (2005).
32. Gaouar, O., Morency, M.-J., Letanneur, C., Séguin, A. & Germain, H. The 124202 candidate effector of *Melampsora larici-populina* interacts with membranes in *Nicotiana* and *Arabidopsis*. *Can. J. Plant Pathol.* **38**, 197–208 (2016).
33. Barrett, L. G. *et al.* Diversity and Evolution of Effector Loci in Natural Populations of the Plant Pathogen *Melampsora lini*. *Mol. Biol. Evol.* **26**, 2499–2513 (2009).
34. Persoons, A. *et al.* Patterns of genomic variation in the poplar rust fungus *Melampsora larici-populina* identify pathogenesis-related factors. *Front. Plant Sci.* **5** (2014).
35. Daly, N. L., Gruber, C. W., Göransson, U. & Craik, D. J. Cystine Knot Folding in Cyclotides. In *Folding of Disulfide Proteins* 43–61 (Chang, R. J. Y. & Ventura, S., 2011).
36. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–549 (2010).
37. Eberhardt, R. Y. *et al.* Filling out the structural map of the NTF2-like superfamily. *BMC Bioinformatics* **14**, 327 (2013).
38. Palmer, I. & Wingfield, P. T. Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from *Escherichia coli*. In *Current Protocols in Protein Science* (eds Coligan, J. E., Dunn, B. M., Speicher, D. W. & Wingfield, P. T.) 6.3.1–6.3.20 (John Wiley & Sons, Inc.), <https://doi.org/10.1002/0471140864.ps0603s70> (2012).
39. Blondeau, K. *et al.* Crystal structure of the effector AvrLm4-7 of *Leptosphaeria maculans* reveals insights into its translocation into plant cells and recognition by resistance proteins. *Plant J. Cell Mol. Biol.* **83**, 610–624 (2015).
40. Rooney, H. C. E. *et al.* *Cladosporium* Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. *Science* **308**, 1783–1786 (2005).
41. Catanzariti, A.-M., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A. & Ellis, J. G. Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* **18**, 243–256 (2006).
42. Lee, S.-Y. & MacKinnon, R. A membrane-access mechanism of ion channel inhibition by voltage sensor toxins from spider venom. *Nature* **430**, 232–235 (2004).
43. Garcia, M. L. Ion channels: gate expectations. *Nature* **430**, 153–155 (2004).
44. Tam, J. P., Lu, Y. A., Yang, J. L. & Chiu, K. W. An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine knot disulfides. *Proc. Natl. Acad. Sci. USA* **96**, 8913–8918 (1999).

45. Kim, J.-Y. *et al.* Protease inhibitors from plants with antimicrobial activity. *Int. J. Mol. Sci.* **10**, 2860–2872 (2009).
46. Chen, B. *et al.* Isolation and characterization of novel cyclotides from *Viola hederacea*: solution structure and anti-HIV activity of vhl-1, a leaf-specific expressed cyclotide. *J. Biol. Chem.* **280**, 22395–22405 (2005).
47. Rosengren, K. J., Daly, N. L., Plan, M. R., Waine, C. & Craik, D. J. Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. *J. Biol. Chem.* **278**, 8606–8616 (2003).
48. Wang, C. K., Colgrave, M. L., Ireland, D. C., Kaas, Q. & Craik, D. J. Despite a conserved cystine knot motif, different cyclotides have different membrane binding modes. *Biophys. J.* **97**, 1471–1481 (2009).
49. Postic, G., Gracy, J., Périn, C., Chiche, L. & Gelly, J.-C. KNOTTIN: the database of inhibitor cystine knot scaffold after 10 years, toward a systematic structure modeling. *Nucleic Acids Res.* **46**, D454–D458 (2018).
50. Catanzariti, A.-M. *et al.* The AvrM effector from flax rust has a structured C-terminal domain and interacts directly with the M resistance protein. *Mol. Plant-Microbe Interact. MPMI* **23**, 49–57 (2010).
51. Chouard, T. Structural biology: Breaking the protein rules. *Nature* **471**, 151–153 (2011).
52. Marcos, E. *et al.* Principles for designing proteins with cavities formed by curved β sheets. *Science* **355**, 201–206 (2017).
53. Avis, J. M. & Clarke, P. R. Ran, a GTPase involved in nuclear processes: its regulators and effectors. *J. Cell Sci.* **109**(Pt 10), 2423–2427 (1996).
54. Carazo-Salas, R. E., Gruss, O. J., Mattaj, I. W. & Karsenti, E. Ran-GTP coordinates regulation of microtubule nucleation and dynamics during mitotic-spindle assembly. *Nat. Cell Biol.* **3**, 228–234 (2001).
55. Zhang, C. & Clarke, P. R. Chromatin-independent nuclear envelope assembly induced by Ran GTPase in *Xenopus* egg extracts. *Science* **288**, 1429–1432 (2000).
56. Hetzer, M., Bilbao-Cortés, D., Walther, T. C., Gruss, O. J. & Mattaj, I. W. GTP hydrolysis by Ran is required for nuclear envelope assembly. *Mol. Cell* **5**, 1013–1024 (2000).
57. Goessweiner-Mohr, N., Arends, K., Keller, W. & Grohmann, E. Conjugative type IV secretion systems in Gram-positive bacteria. *Plasmid* **70**, 289–302 (2013).
58. Zhang, D., De Souza, R.F., Anantharaman, V., Iyer, L.M. & Aravind, L. Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol. Direct* **14**, <https://doi.org/10.1186/1745-6150-7-18> (2012).
59. Corbett, A. H. & Silver, P. A. The NTF2 Gene Encodes an Essential, Highly Conserved Protein That Functions in Nuclear Transport in Vivo. *J. Biol. Chem.* **271**, 18477–18484 (1996).
60. Zhang, Q. *et al.* TaNTF2, a contributor for wheat resistance to the stripe rust pathogen. *Plant Physiol. Biochem. PPB* **123**, 260–267 (2018).
61. Bialas, A. *et al.* Lessons in Effector and NLR Biology of Plant-Microbe Systems. *Mol. Plant-Microbe Interact. MPMI* **31**, 34–45 (2018).
62. Oganessian, N., Ankoudinova, I., Kim, S.-H. & Kim, R. Effect of osmotic stress and heat shock in recombinant protein overexpression and crystallization. *Protein Expr. Purif.* **52**, 280–285 (2007).
63. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* **5**, (1995).
64. Johnson, B. A. & Blevins, R. A. NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
65. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
66. Rieping, W. *et al.* ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinforma. Oxf. Engl.* **23**, 381–382 (2007).
67. Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol. Clifton NJ* **278**, 353–378 (2004).
68. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
69. Nederveen, A. J. *et al.* RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59**, 662–672 (2005).
70. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
71. Cheung, M.-S., Maguire, M. L., Stevens, T. J. & Broadhurst, R. W. DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J. Magn. Reson. San Diego Calif* **1997** **202**, 223–233 (2010).

Acknowledgements

The authors would like to thank Prof. David L. Joly (Université de Moncton, Nouveau Brunswick, Canada) and Prof. Hugo Germain (Université du Québec à Trois-Rivières, Québec, Canada) for early discussions about candidate effector selection and continuous collaborations and discussions on the topic through the years. The authors would also like to acknowledge the French EFFECTOME network financed by INRA Plant Health and Environment and Ecology of Forest, Grasslands and Freshwater Divisions, as a fruitful platform for sharing ideas and creating collaborations. Access to the Bruker DRX 600 of the UMS2008 IBSLor Biophysics and Structural Biology Core Facility (Université de Lorraine-CNRS-INSERM) was appreciated. The UMR1136 is supported by a grant overseen by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (ANR-11-LABX-0002-01, Lab of Excellence ARBRE). This work was supported by the French Infrastructure for Integrated Structural Biology (ANR-10-INBS-0005). The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

K.G., C.L., P.T., P.B., B.P., N.S., N.R., S.D., A.P. and A.H. conceived, designed experiments and analysed the data. All the authors wrote and revised the manuscript.

Competing Interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-53816-9>.

Correspondence and requests for materials should be addressed to A.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019