Structures of *Phytophthora* RXLR Effector Proteins
A CONSERVED BUT ADAPTABLE FOLD UNDERPINS FUNCTIONAL DIVERSITY[*][S]

Laurence S. Boutemy[‡], Stuart R. F. King[‡], Joe Win[§], Richard K. Hughes[‡], Thomas A. Clarke[¶],
Tharin M. A. Blumenschein[¶], Sophien Kamoun[§1], and Mark J. Banfield[‡2]

From the [‡]Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom, the [§]Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, United Kingdom, and the [¶]Centre for Molecular and Structural Biochemistry, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom

Phytopathogens deliver effector proteins inside host plant cells to promote infection. These proteins can also be sensed by the plant immune system, leading to restriction of pathogen growth. Effector genes can display signatures of positive selection and rapid evolution, presumably a consequence of their co-evolutionary arms race with plants. The molecular mechanisms underlying how effectors evolve to gain new virulence functions and/or evade the plant immune system are poorly understood. Here, we report the crystal structures of the effector domains from two oomycete RXLR proteins, *Phytophthora capsici* AVR3a11 and *Phytophthora infestans* PexRD2. Despite sharing <20% sequence identity in their effector domains, they display a conserved core α-helical fold. Bioinformatic analyses suggest that the core fold occurs in ~44% of annotated *Phytophthora* RXLR effectors, both as a single domain and in tandem repeats of up to 11 units. Functionally important and polymorphic residues map to the surface of the structures, and PexRD2, but not AVR3a11, oligomerizes *in planta*. We conclude that the core α-helical fold enables functional adaptation of these fast evolving effectors through (i) insertion/deletions in loop regions between α-helices, (ii) extensions to the N and C termini, (iii) amino acid replacements in surface residues, (iv) tandem domain duplications, and (v) oligomerization. We hypothesize that the molecular stability provided by this core fold, combined with considerable potential for plasticity, underlies the evolution of effectors that maintain their virulence activities while evading recognition by the plant immune system.

Many phytopathogens secrete and/or inject "effector" proteins inside host cells to modulate cellular processes (1). Filamentous plant pathogens (oomycetes and fungi) secrete a complex arsenal of effector proteins, with hundreds of effector genes identified in the genomes of several species (2–5). Effec-tor genes tend to reside in dynamic regions of the pathogen genomes, frequently exhibit high levels of sequence polymorphisms, and display signatures of rapid evolution, including high rates of copy number variation, presence/absence polymorphisms, and positive selection. The primary adaptive function of effectors is to promote virulence, mainly by suppressing plant immunity, and some effectors are required for full virulence (6–9). However, plants have evolved mechanisms to recognize effector proteins, resulting in effector-triggered immunity that restricts pathogen growth. This results in selective pressure on the effector to evade host recognition while maintaining its virulence activity (10). How effectors evolve to enable parasites to adapt to their hosts is not well understood. Specifically, the underlying molecular mechanisms by which effectors evolve to gain new virulence functions, adapt to their host targets, and/or evade the plant innate immune system are unclear.

Phytophthora is a genus of the oomycetes (water molds). This group of eukaryotic microorganisms includes plant pathogens that are responsible for some of the most devastating diseases of plants and have serious impacts on global crop production, on forestry, and in ornamental settings (3, 4). *Phytophthora infestans*, the causative agent of potato and tomato late blight, triggered the Irish potato famine in the 1840s. *Phytophthora sojae* causes soybean root rot, and *Phytophthora capsici* causes pepper and cucurbit blight. These destructive parasites continue to be a major problem for modern agriculture, with significant economic and environmental effects realized through crop losses and the methods used to control these diseases. In recent years, the sudden oak death pathogen *Phytophthora ramorum* has emerged as a significant threat to both private and commercial forests.

The genome sequences and effector repertoires of several economically important *Phytophthora* species are available, including *P. infestans*, *P. capsici*, and *P. sojae* (3, 5) (see the *P. capsici* Genome Project Web site). These parasites secrete modular effectors that are translocated into host cells (12, 13). One class of *Phytophthora* effectors is the RXLR family that also occurs in downy mildew oomycetes, such as *Hyaloperonospora arabidopsidis* (2). RXLR effectors are defined by a secretion signal peptide followed by a conserved N-terminal domain defined by the RXLR (Arg-Xaa-Leu-Arg) consensus sequence. The RXLR domain is required for translocation inside plant cells (13) but is dispensable for the biochemical activity of the effectors when expressed directly inside host cells (14). Indeed,

the effector activity appears to reside in the C-terminal regions of RXLR effectors (the "effector domain") (14).

The effector domains of RXLR proteins display extensive sequence diversity and typically share little sequence similarity with proteins with known activities, making functional annotation and prediction of tertiary folds from sequence difficult. Despite this diversity, there are recognizable sequence relationships within the C-terminal regions of *Phytophthora* RXLR effectors, and some repeating sequence motifs termed "W", "Y", and "L" have been described (15). These analyses indicate that many, but probably not all, RXLR effector genes share a common ancestor (3, 15). How RXLR effectors encode adaptability while retaining their viability at the molecular level is unknown. Further, the virulence activities of individual RXLR effectors are only just beginning to emerge, and little is known about the biochemical mechanisms that underlie these activities (6).

Several RXLR effectors, including *P. infestans* AVR3a, AVR4, AVRblb1, and AVRblb2 and *H. arabidopsidis* ATR1 and ATR13, were first identified based on their avirulence activity (their ability to activate effector-triggered immunity in plants that carry a corresponding resistance (R) protein (14) (supplemental Fig. 1). AVR3a is a member of a large subfamily of RXLR effectors with homologues in *P. sojae* (3) and *P. capsici* (see the *P. capsici* Genome Project Web site). In *P. infestans*, AVR3a is found in either of two major allelic forms, AVR3a^{KI} or AVR3a^{EM}, with K/E defining the amino acid residue at position 80 (Lys or Glu) and I/M the residue at 103 (Ile or Met) (16). These different forms trigger different responses *in planta*. AVR3a^{KI}, but not AVR3a^{EM}, activate effector-triggered immunity in the presence of the potato resistance protein R3a (16). In plants that do not carry R3a, both AVR3a forms bind and stabilize the E3 ubiquitin ligase CMPG1, leading to suppression of INF1-elicited plant immunity and programmed cell death (PCD)[3] (17, 18). Further, the identity of the residue at the C terminus of AVR3a is critical for INF1 suppression activity but not for R3a recognition (17). Gain-of-function mutagenesis in AVR3a^{EM} identified a "hot spot" between residues Ser^{123} and Lys^{130} that enhances recognition by R3a and strengthens PCD suppression in response to INF1 (17).

Other oomycete RXLR effectors have been validated based on their ability to modulate defense responses in plants (supplemental Fig. 1). One such effector, *P. infestans* PexRD2, was identified in a functional screen to promote cell death activity in several susceptible and resistant plants (19, 20). The promotion of cell death induced by PexRD2 could reflect its virulence activity (19), but the exact mechanism by which this effector contributes to pathogenesis remains unclear.

The purpose of this study was to investigate the three-dimensional structures of RXLR effectors from *Phytophthora*. Our aim was to decipher the nature of the relationships between these effectors at the molecular level and determine how these might provide information on their remarkable plasticity and accelerated evolution. Here we report the crystal structures of the effector domains from two oomycete RXLR proteins: (i)

P. capsici AVR3a11, a homologue of *P. infestans* AVR3a (6, 16–18) and AVR1b from *P. sojae* (21, 22) and (ii) *P. infestans* PexRD2. These two effectors do not share significant sequence identity in their effector domains but, remarkably, display a similar α-helical fold in their structure. Structure-informed bioinformatic analyses revealed that this fold defines a domain observed in about 44% of annotated *Phytophthora* RXLR effectors. We discovered that many functionally important residues and those that are polymorphic map to the surface of the proteins. Our findings led us to propose a model for the structural basis of plastic evolution in RXLR effectors. We hypothesize that the core α-helical fold (termed the "WY-domain") can adapt through (i) insertion/deletions in loop regions between α-helices, (ii) extensions to the N and C termini, (iii) amino acid replacements in surface residues, (iv) tandem domain duplications, and (v) oligomerization. We propose that the core fold provides both a degree of molecular stability and plasticity that enables development/maintenance of effector virulence activities while allowing evasion of recognition by the plant innate immune system during rapid "arms race" co-evolution.

EXPERIMENTAL PROCEDURES

Construct Design—We used RONN (23) to guide construct design of AVR3a11 and PexRD2 for protein expression. This predicted the N-terminal regions of the effectors up to and including the RXLR signal sequence were likely to be disordered.

Protein Production—Constructs of AVR3a11 and PexRD2 were designed to limit inclusion of any predicted disordered regions, and synthetic genes were ordered (optimized for expression in *Escherichia coli*) and then subcloned into expression vectors (for details, see supplemental material). A second AVR3a11 construct (Thr^{70}–Val^{132}) was produced by PCR and cloned as detailed in the supplemental material. All constructs were verified by sequencing. Details of protein expression, which followed well established procedures, are given in the supplemental material. Proteins were purified by Ni^{2+}-immobilized metal affinity chromatography and gel filtration chromatography (see supplemental material). Protein molecular masses were verified using mass spectrometry (see supplemental material).

Crystallization and Data Collection—Initial crystals of AVR3a11 (the Thr^{70}–Val^{132} construct at 19 mg/ml) and PexRD2 (12 mg/ml) were grown at 20 °C in sitting drop plates set up using a robotic crystallization system and then optimized in 24-well plate hanging drop experiments. Details of precipitant and cryoprotectant solutions are given in the supplemental material. Data used to solve the structure of AVR3a11 were collected using a Rigaku RU-H3RHB generator/Mar345 detector, with high resolution data collected on beamline I04 of the Diamond Light Source, UK. All data for PexRD2 were collected on I02 (Diamond Light Source, UK). Data collection statistics are given in Table 1.

Data Processing and Structure Solution, Refinement/Rebuilding, and Validation—X-ray diffraction data were processed with iMOSFLM (24) and scaled with SCALA (25), as implemented in the CCP4 suite (26). For phasing, AVR3a11 crystals were immersed in cryoprotectant (see supplemental material),

---

[3] The abbreviations used are: PCD, programmed cell death; SAD, single wavelength anomalous dispersion; SIRAS, single isomorphous replacement method with anomalous scattering.

TABLE 1

X-ray data collection, refinement statistics, and model analysis

	AVR3a11 (native)	AVR3a11 (iodide soak)	AVR3a11 (native, DLS)	PexRD2 (bromide peak)	PexRD2 (high resolution)
Data collection[a]					
Instrumentation	Home source	Home source	DLS–I04	DLS–I02	DLS–I02
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters (Å)	$a = 21.21$	$a = 21.29$	$a = 21.30$	$a = 44.80$	$a = 44.45$
	$b = 25.24$	$b = 25.29$	$b = 25.33$	$b = 52.36$	$b = 52.92$
	$c = 90.22$	$c = 90.33$	$c = 90.70$	$c = 53.32$	$c = 53.73$
Wavelength (Å)	1.54	1.54	0.73	0.90	
Resolution range (Å)	16.80–1.80 (1.90–1.80)	13.20–1.90 (2.00–1.90)	18.14–0.90 (0.95–0.90)	44.80–1.90 (2.00–1.90)	37.70–1.75 (1.81–1.75)
R_{merge} (%)[b]	4.1 (9.5)	4.8 (11.3)	6.3 (37.2)	6.5 (43.0)	7.9 (42.5)
$I/\sigma(I)$[b]	51.0 (25.5)	47.2 (24.1)	14.0 (4.3)	25.4 (6.7)	26.3 (9.9)
Completeness (%): all data, anomalous half-sets[b]	99.5 (99.7)	99.7 (100), 99.9 (100)	97.5 (98.0)	99.0 (98.7), 99.3 (99.1)	99.9 (100)
Redundancy: all data, anomalous half-sets[b]	15.0 (13.8)	15.9 (15.7), 9.0 (8.6)	4.7 (5.4)	12.9 (13.3), 7.0 (7.0)	15.8 (16.3)
No. of sites (SAD/SIRAS)		8/4		11	
FOM (SOLVE/RESOLVE)		0.44		0.34	
Refinement[c]					
Resolution range (Å)			17.41–0.90 (0.92–0.90)		37.70–1.75 (1.80–1.75)
R_{work} (%)[b]			12.6 (19.6)		18.9 (23.9)
R_{free} (%)[b]			14.9 (20.9)		24.6 (27.1)
No. of non-hydrogen atoms (Protein, Ligand, Water)			526, 12, 91		967, 7, 125
Root mean square deviation bonds (Å)			0.017		0.015
Root mean square deviation angles (degrees)			2.37		1.45
ESU (based on ML) (Å)			0.012		0.079
Ramachandran favored (%)[d]			98.4		99.2
Ramachandran outliers (%)[d]			0.0		0.0

[a] Values as reported by Scala.
[b] Values in parentheses are for the high resolution shell (as given).
[c] Values as reported by Refmac5.
[d] Values as reported by Molprobity.

supplemented with 100 mM potassium iodide for 30 s before freezing. The AutoSol wizard of PHENIX (27) was used to solve the structure of AVR3a11 (to 1.9 Å resolution), with a combined SAD/SIRAS approach using the anomalous signal from iodide. PHENIX was also used to produce an initial model, which was then refined against the high-resolution data. PexRD2 was co-crystallized with ammonium bromide and anomalous scattering from bound bromide ions was used to solve the structure. Bromide sites were initially identified with SHELXC/D (28), using a SAD approach, and data were collected at a wavelength of 0.90 Å. These positions were then used by PHASER-EP (29), as implemented in the CCP4 suite, to calculate initial phases. These phases were modified with PAR-ROT, and an initial model was built using BUCANEER (30). This model was then refined against the high resolution data. Final models were produced through iterative rounds of refinement using REFMAC5 (26) and rebuilding with COOT (31). Structure validation used COOT and MOLPROBITY (32). Protein structure figures have been prepared with CCP4mg (33). Refinement and validation statistics are given in Table 1. The coordinates and structure factor amplitudes have been submitted to the Protein Data bank with accession codes 3ZR8 (AVR3a11) and 3ZRG (PexRD2).

Analytical Ultracentrifugation—Sedimentation equilibrium experiments were performed using a Beckman XLA-I analytical ultracentrifuge equipped with an An-50 Ti 8 place rotor and absorbance/interference optics. For AVR3a11, samples at 0.3 and 0.45 mg/ml were monitored using absorbance optics, and equilibrium scans were obtained at speeds of 38,000 rpm. Interference patterns were collected at equilibrium for PexRD2 with

samples at 3 and 6 mg/ml and at a speed of 37,000 rpm. Data were fitted to either a single species (AVR3a11) or dimer/tetramer (PexRD2) models using ULTRASCAN (34). For PexRD2, the dimeric form predominates in solution, and there was no evidence for a monomer.

Bioinformatic Analysis—MEME searches allowed for motif lengths up to 100 amino acids (for additional search parameters see the supplemental material). We constructed HMM models of the four motifs that span the WYL sequences and used this to search the effector data base. From this we assembled an updated HMM model describing the WY-domain (49 amino acids). For inclusion of sequences in this updated model, we chose a cut-off of $e < 0.12$ from the initial search, which corresponds to the position of *P. infestans* AVR3a in the list (for sequences used, see supplemental Table 1). This new HMM model was used to re-search the RXLR effector and non-RXLR proteome databases from *Phytophthora* and *H. arabidopsidis* and the non-RXLR secretome of *Phytophthora*. Hits that scored greater than an HMM score of 0.0 were considered putative WY-domain-containing proteins. This cut-off was chosen based on the distribution of WY-domain sequences and our estimate of the false positive discovery rate in the non-RXLR secretome (see Fig. 4c and supplemental material).

Sequence and Structural Alignments—Sequence alignments were prepared using ClustalW2 (35). Structure-based sequence alignments were produced by hand. Structural alignments between AVR3a11/PexRD2 were prepared with SSM (36), producing a root mean square deviation based on C_α positions, as reported.
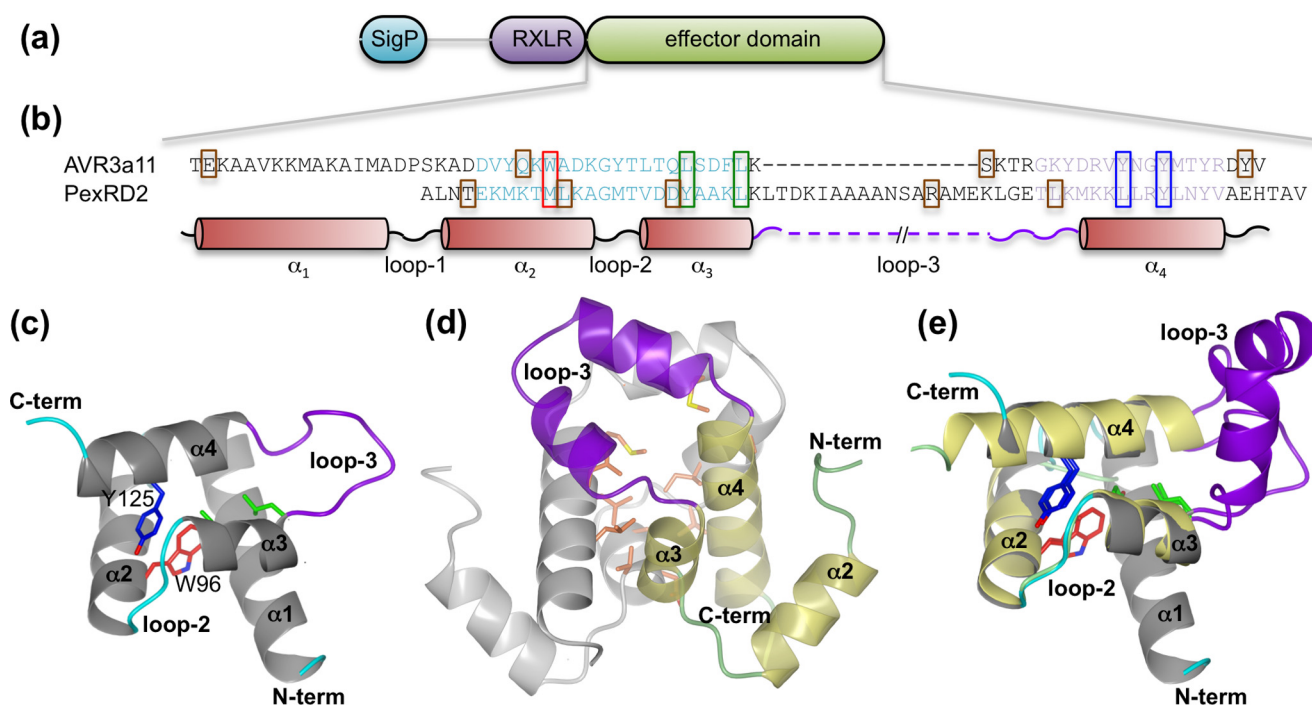
FIGURE 1. **RXLR effectors AVR3a11 and PexRD2 adopt a structurally conserved but adaptable fold.** *a*, oomycete RXLR effectors are modular proteins comprising a secretion signal (*cyan*), RXLR translocation motif (*purple*), and an effector domain (*green*). *b*, structural alignment and secondary structure elements of the effector domains of AVR3a11 and PexRD2. W-motifs (*cyan*) and Y-motifs (*lilac*) are *colored*, with key residues (as discussed under "Results and Discussion") *boxed*. *c*, the structure of AVR3a11 is a monomeric four-helix bundle with a hydrophobic core. Carbon atoms of key residues in the W- and Y-motifs are *colored* as *boxed* in *b*; loop-3 is shown in *purple*. *d*, PexRD2 is a dimer with a hydrophobic interface, including residues Val[73], Asp[74], Ala[77], Thr[83], Ile[86], Ala[90], Met[96], Gly[100], Met[105], Leu[108], Leu[109], and Leu[112] (shown for one monomer only). α-Helices are labeled to correspond to equivalent positions in AVR3a11. *e*, the AVR3a11/PexRD2-monomer overlay generated using SSM, showing the conserved fold. Protein structures are *colored* as in *c* and *d* with key residues of the W- and Y-motifs *colored* as in *b*.
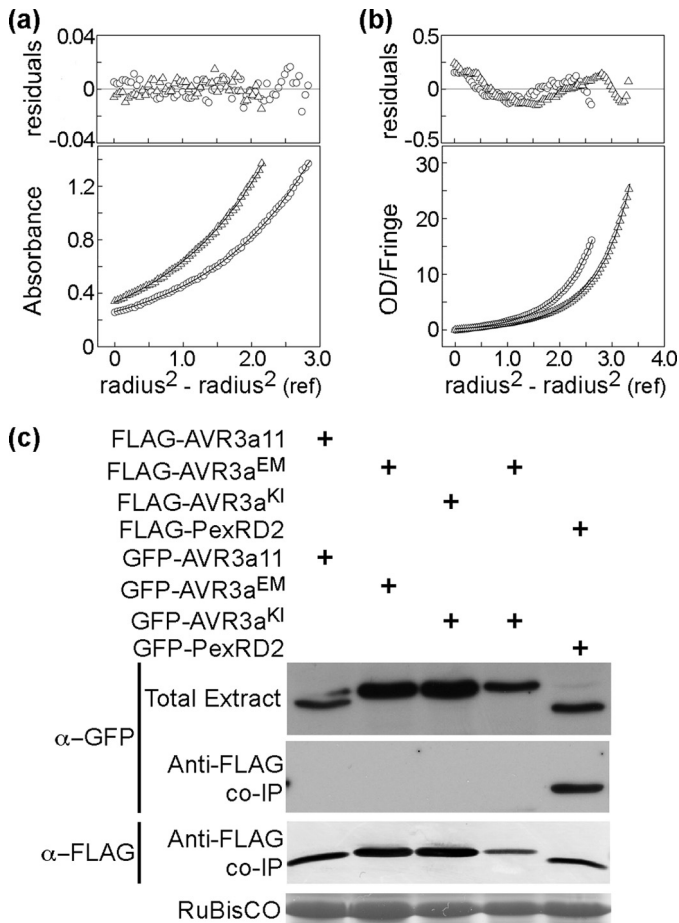
In Planta Co-immunoprecipitations—To determine whether AVR3a11, AVR3a[KI], AVR3a[EM], and PexRD2 exist as oligomers *in planta*, co-immunoprecipitation experiments were performed using tagged effectors expressed in *Nicotiana benthamiana* leaves via agroinfiltration. Two different tagged constructs were made for each effector: (i) FLAG epitope-tagged constructs were cloned into pJL-TRBO (37), and (ii) GFP-epitope effector constructs were generated in pK7WGF2 (38). Details of the cloning are given in the supplemental material. Co-immunoprecipitations were performed with anti-FLAG M2 affinity gel (Sigma) using total protein extracts harvested from leaves 3 days postinoculation. Interacting proteins were detected by Western blot, probing with anti-GFP primary antibodies (Invitrogen), anti-rabbit secondary antibodies conjugated to horseradish peroxidase (Sigma), and a chemiluminescent substrate (Thermo Scientific). Additional details are given in the supplemental material.

RESULTS AND DISCUSSION

To address structure/function relationships in oomycete RXLR effectors, we established an *E. coli*-based expression screen aimed at producing proteins that were amenable to structure determination. This screen identified the *P. capsici* effector AVR3a11 (a homologue of AVR3a from *P. infestans*) and *P. infestans* effector PexRD2, as highly expressed and soluble. The purified proteins were crystallized, and their structures were determined to 0.9 and 1.75 Å, respectively (Table 1).

The Structure of AVR3a11 Is a Monomeric Four-helix Bundle—The structure of AVR3a11 comprises Thr[70]–Val[132] (63 residues) of the native sequence. This construct lacks the N-terminal signal sequence and RXLR translocation motif, which were predicted to be disordered (Figs. 1, *a* and *b*, and supplemental Fig. 2). An earlier construct comprising residues Gly[63]–Val[132] was recalcitrant to crystallization but amenable to NMR analysis. Preliminary structure calculations using NMR data suggested that residues Gly[63]–Lys[69] were not well structured (see supplemental material). The effector domain of AVR3a11 adopts a four-helix bundle fold, stabilized by a hydrophobic core, with the helices connected by loop regions (Fig. 1*c*). The crystal structure suggests the effector domain of AVR3a11 is a monomer. This oligomeric state was also observed in solution, as shown by the retention volume of the protein on a size exclusion column during purification (data not shown) and analytical ultracentrifugation (7350 ± 0.03 Da, Fig. 2*a* (theoretical mass = 7474 Da)). Further, Avr3a11 was shown to be monomeric *in planta* using co-immunoprecipitation (Fig. 2*c* and supplemental Fig. 3). Structure-based database searches (39) using the refined AVR3a11 model revealed low significance hits to other four-helix bundles, including KaiA from *Anabena* sp. PCC7120 (Protein Data Bank code 1R5Q). KaiA is a cyanobacterial circadian clock regulator (40), and although the overall fold of AVR3a11 is loosely related to KaiA, the differences in structure and surface properties do not suggest a role for the AVR3a family as circadian rhythm regulators.
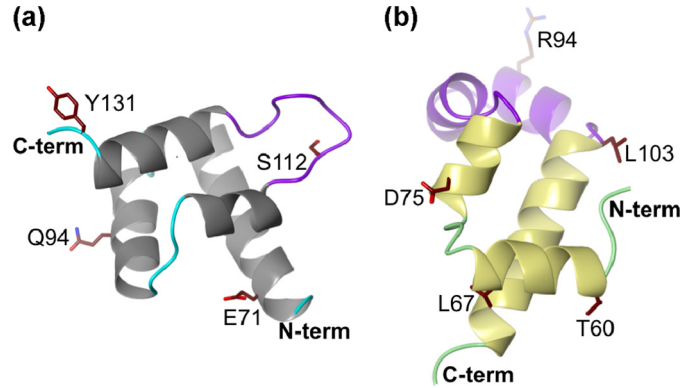
FIGURE 3. **Functionally important and polymorphic residues are presented on the surface of AVR3a11 and PexRD2.** *a*, in AVR3a11, Glu[71] and Gln[94] (equivalent to the Lys/Glu[80] and Ile/Met[103] positions of AVR3a), Ser[112] (equivalent to Ser[132] of AVR3a), and a Tyr residue near the C terminus are located on the protein surface. *b*, in PexRD2, the polymorphic residues between the *P. infestans* and *P. mirabilis* homologues are also surface-displayed. The positions of these residues in the primary sequences of the proteins are shown as *brown boxes* in Fig. 1*b*.

FIGURE 2. **RXLR effectors can adopt different oligomeric states *in vitro* and *in vivo*.** Fits of the analytical ultracentrifugation data (including residuals) for (*a*) AVR3a11 (0.3 mg/ml (*circles*) and 0.45 mg/ml (*triangles*)) (*b*) PexRD2 (3 mg/ml (*triangles*) and 6 mg/ml (*circles*)) to the models described confirmed that monomeric and dimeric species are the prevalent forms of these proteins in solution. *c*, co-immunoprecipitation (*co-IP*) shows that PexRD2 self-associates *in planta*; no oligomerization is apparent for AVR3a11, AVR3a[KI], or AVR3a[EM]. RuBisCO is included as a loading control (Coomassie-stained SDS-polyacrylamide gel).

AVR3a11 belongs to a family of *Phytophthora* effectors that includes *P. infestans* AVR3a and AVR1b from *P. sojae* (supplemental Fig. 4*a*). Pairwise alignments based on the effector domains revealed that AVR3a11 shares 41% sequence identity with AVR3a and 46% sequence identity with AVR1b (spanning residues 70–132, AVR3a11 numbering). The structure of AVR3a11 is therefore representative of this family and provides an opportunity to interpret RXLR effector domain function with respect to structure.

Residues Important for the Function of P. infestans AVR3a Map to the Surface of AVR3a11—Several functionally important residues of the AVR3a11 homologue AVR3a have previously been identified (17). Amino acids 80 and 103 of AVR3a map to Glu[71] and Gln[94] in AVR3a11 (Fig. 3*a* and supplemental Fig. 5) and are positioned at the start of the first α-helix (α$_1$) and midway along α$_2$. Although the C$_α$ atoms of these residues are 16.7 Å apart, the residues locate to the same face of the four-helix bundle, suggesting that this region forms an important surface both for recognition by R3a and PCD suppression activity. The C terminus of the AVR3a family is somewhat divergent (17, 41) and contains indels of up to 3 residues. In the structure

of AVR3a11, the final three residues emerge from α$_4$, and their conformation is not likely to be constrained in solution (Figs. 1*c* and 3*a*). The C terminus of AVR3a, which has a role in PCD suppression activity but not recognition by R3a, is on the opposite face of the four-helix bundle from the K/E and I/M positions. Further, the region comprising residues 123–130 of AVR3a (112–117 in AVR3a11) forms a prominent loop between α$_3$ and α$_4$ ("loop-3"; Figs. 1*c* and 3*a*). The sequence of this loop is variable within the AVR3a family (and includes an up to 4-residue indel). The fact that a single point mutation in this region in AVR3a[EM] (S123C (17)) confers AVR3a[KI] levels of R3a recognition and PCD suppression reveals that this loop is an important determiner of protein activity.

The Structure of AVR3a11 Is Built from Widely Conserved Sequence Motifs—The C-terminal domains of RXLR effectors have been reported to carry conserved sequence motifs termed W, Y, and L (3, 15). The AVR3a family contains a single W- and Y-motif (Figs. 1*b* and 4 and supplemental Fig. 6), with an additional "K" motif to the N terminus (21). The structure of AVR3a11 revealed that these motifs map to discrete secondary structure units that together assemble the four-helix bundle: residues of the K-motif comprise α$_1$ and loop-1; the W-motif comprises α$_2$, loop-2 and α$_3$; and the Y-motif comprises the C terminus of loop-3 and α$_4$ (Fig. 1*b*). Side chains from residues of each helix contribute to the hydrophobic core of the bundle. In AVR3a11, the key residue of the W-motif is Trp[96], which is located at the C-terminal end of α$_2$ and is excluded from bulk solvent. This residue makes favorable edge-to-face stacking interactions with Tyr[125] (42), the most highly conserved residue of the Y-motif, which is positioned toward the end of α$_4$. Tyr[125] also forms a hydrogen bond with the carbonyl oxygen of Ala[101], positioned on loop-2, which maybe important for stabilizing the conformation of this loop. Tyr[122] in AVR3a11 is positioned 3 residues before the conserved Tyr in the Y-motif (this position is frequently a bulky hydrophobic amino acid; see below). It is not immediately apparent from the structure why such high conservation at this position is maintained. Additional conserved residues of the W-motif (Leu[106] and Leu[110] in
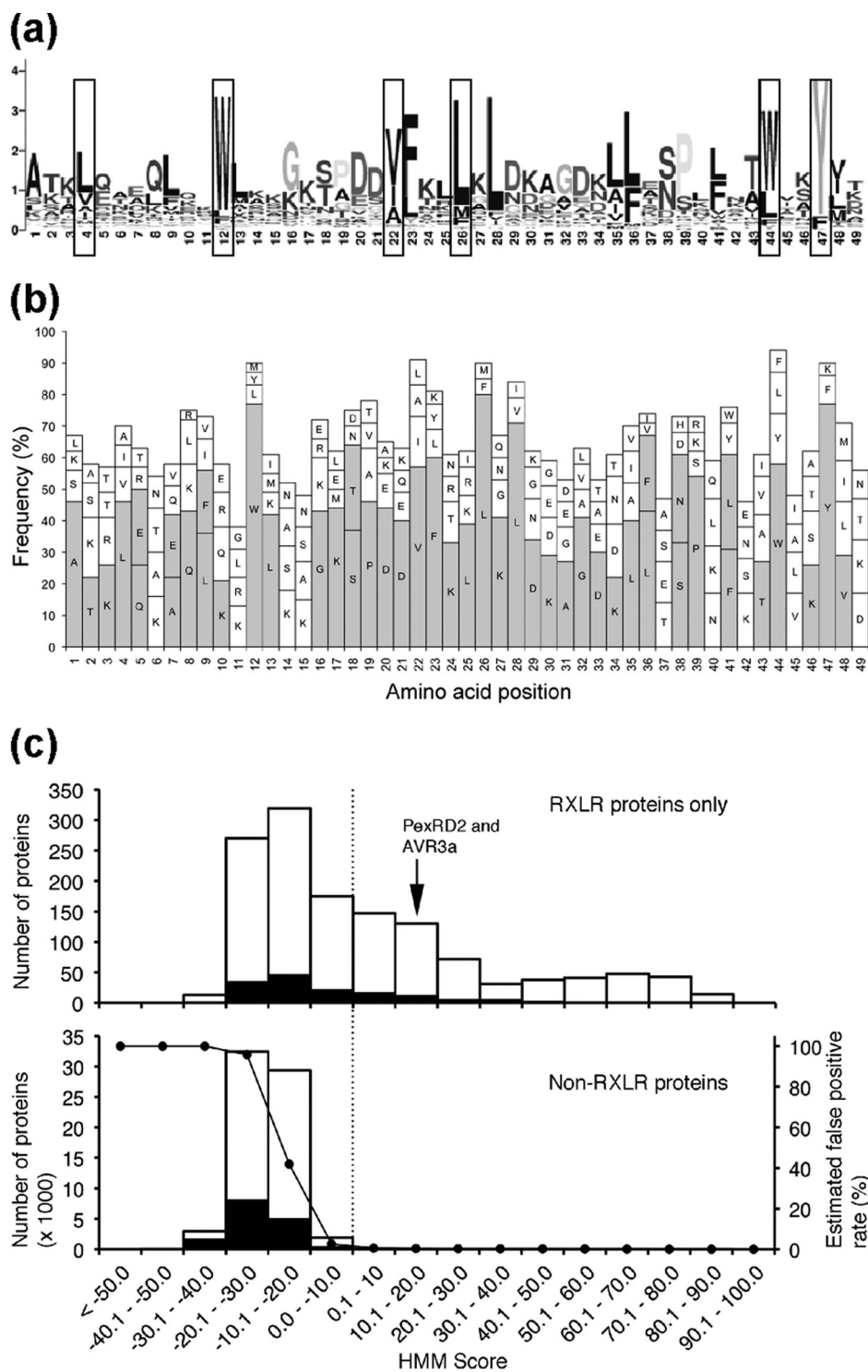
FIGURE 4. **The WY-domain is a conserved unit in *Phytophthora* RXLR effectors.** *a*, the MEME motif spanning the WY-domain identified by searching the RXLR repertoire of *P. infestans*, *P. ramorum*, and *P. sojae*. *Boxed residues*, from *left* to *right* are the Leu of the L-motif, the Trp of the W-motif, the residues from α₃ (AVR3a11) that pack into the hydrophobic core (Val and Leu), and the Tyr of the Y-motif (and the conserved hydrophobic residue two positions before). *b*, graphical representation of the final HMM model used to screen the *Phytophthora* and *H. arabidopsidis* effector repertoires for the presence of the WY-domain. The four most abundant amino acids at each position in the motif are labeled with their *one-letter code*, and those that represented ≥20% are *shaded* in *gray*. *c*, distribution of WY-domains in RXLR effectors (*top*) and the non-RXLR proteome (*bottom*, *stacked bar chart*) following sequence database searches with the WY-domain HMM. The number of proteins in each HMM score bin is shown. The positions of AVR3a and PexRD2 are marked, and the *dashed line* shows the position of the cut-off used to describe WY-domain like sequences. The estimated false positive rate of WY-domain discovery based on the *Phytophthora* non-RXLR secretome is also shown (*bottom*, *line graph*).

AVR3a11) anchor α₃ to the bundle by forming interactions within the hydrophobic core. Mutation of the Trp[96] and Leu[110] equivalents in AVR3a led to reduced protein stability *in planta*

(17); similar mutations in hydrophobic residues of AVR1b also led to loss of function (21), most likely due to protein destabilization. The W- and Y-motifs together appear to form a critical

TABLE 2

Numbers of *Phytophthora* and *H. arabidopsidis* proteins in the RXLR effectorome/non-RXLR proteome that score above/below the threshold described for containing a WY-domain

	RXLR effectors, above threshold	RXLR effectors, below threshold	Non-RXLR proteome, above threshold	Non-RXLR proteome, below threshold
P. infestans	221	342	39	17,553
P. ramorum	181	128	40	15,543
P. sojae	125	210	66	18,787
Total: *Phytophthora*	527 (44%)	680 (56%)	145 (0.3%)	51,883 (97.7%)
H. arabidopsidis	35 (26%)	99 (74%)	91 (0.6%)	14,734 (99.4%)
Total: *Phytophthora* and *H. arabidopsidis*	562	779	236	66,617

folding unit, the WY-domain, that, with the addition of another helix (to the N terminus), can form a four-helix bundle.

PexRD2 Is Also a WY-domain Protein but with Adaptations—The structure of the effector domain of PexRD2 comprises Ala[57]-Ala[120] (64 residues), starting immediately after the RXLR sequence through to the penultimate residue in the construct (Val[121] could not be modeled in the electron density). The structure is composed of five α-helices and crystallizes as a dimer with intimate hydrophobic contacts at the interface (buried surface area = 1830 Å^2, 20% of the surface area of each monomer; PISA CCS score = 0.89 (43)) (Fig. 1*d*). Structure-based database searches (39) did not reveal any significant hits to PexRD2. Given the lack of sequence identity (<20%, supplemental Fig. 4*b*), we were surprised to discover that comparison of the structures of a PexRD2 monomer with AVR3a11 using SSM (36) aligned 37 residues comprising α_2, loop-2, α_3, and α_4 with a root mean square deviation = 1.0 Å (Fig. 1*e*). This region comprises the helices of the WY-domain described for AVR3a11. However, PexRD2 contains no equivalent of α_1 from the AVR3a11 four-helix bundle, and a significant insertion (16 residues) is present between α_3 and α_4 (loop-3). The structurally aligned residues only share 13.5% sequence identity (Fig. 1*b* and supplemental Fig. 4*b*).

PexRD2 Is a Dimer in Solution and Self-associates in Planta—To confirm that the PexRD2 dimer was not a crystallographic artifact, we used analytical ultracentrifugation to show that this oligomeric state predominates in solution (Fig. 2*b*). The best fit of the data was obtained using a dimer/tetramer model with the molecular masses of the species fixed at their theoretical value. The retention volume of PexRD2 on a size exclusion column during purification also supports the dimer conformation (data not shown). Further, we performed *in planta* co-immunoprecipitation experiments with epitope-tagged effectors. These showed that PexRD2, but not AVR3a11 (see above) or AVR3a, self-associates when transiently expressed in *N. benthamiana* leaves via agroinfiltration (Fig. 2*c* and supplemental Fig. 3). Given the evidence from the crystal structure and analytical ultracentrifugation, it is reasonable to assume that PexRD2 is dimerizing *in planta*.

Polymorphic Residues in PexRD2 Are Surface-presented—PexRD2 from all strains of *P. infestans* sequenced to date displays 100% sequence identity in the effector domain (44). However, a PexRD2 homologue in the closely related species *P. mirabilis* PIC99114 contains polymorphisms at five positions (Fig. 3*b* and supplemental Fig. 4*c*) (44). Comparison of these two sequences revealed that positive selection has acted on this gene because the ratio of non-synonymous to synony-

mous substitutions is >1 (dN/dS = 1.27). The PexRD2 structure revealed that these 5 polymorphic residues are presented on the protein surface and are not co-localized (*i.e.* do not define a particular surface region) (Fig. 3*b*). We conclude that polymorphisms in some surface-exposed residues may have contributed to the evolution of PexRD2 without disrupting the WY-domain fold.

The WY-domain Is a Unit Conserved in Many Phytophthora and H. arabidopsidis RXLR Effectors—The unexpected structural similarity we discovered between AVR3a11 and PexRD2 prompted us to re-evaluate the context of the repeated WYL motifs previously reported in *Phytophthora* RXLR effectors (3, 15) and extend this analysis to *H. arabidopsidis*. We used MEME (45) to search for conserved motifs in a data set comprising the *P. infestans*, *P. ramorum*, and *P. sojae* RXLR effector repertoires (1207 proteins, C-terminal effector domain only, downstream of the RXLR). The highest scoring motifs are dominated by overlapping regions incorporating W-, Y-, and L-motifs that can be readily aligned to reveal repeating WYL sequences (supplemental Fig. 6). One newly identified 49-amino acid MEME motif spanned both the W- and Y-motifs and covers the $\alpha_2/\alpha_3/\alpha_4$ WY-domain seen in the structures of AVR3a11 and PexRD2 (Fig. 4*a*). Interestingly, the alignment of overlapping MEME motifs revealed that the conserved Leu of the L-motif is positioned 8 residues to the N terminus of the conserved Trp of the W-motif.

To explore the extent to which the WY-domain is found in RXLR effectors, we constructed HMM models of four MEME motifs (Fig. 4*b*, supplemental Fig. 7, and supplemental Table 1) and rescreened the *Phytophthora* and *H. arabidopsidis* RXLR effector sequences with HMMER (46); we also screened the non-RXLR proteome of *Phytophthora* and *H. arabidopsidis* and the non-RXLR secretome of *Phytophthora* with the WY-domain HMM. This analysis revealed that WY-domains are highly enriched in RXLR effectors compared with the non-RXLR proteome (Fig. 4*c*). Using an appropriate cut-off (see "Experimental Procedures"), we revealed that 527 of 1207 (44%) of *Phytophthora* and 35 of 134 (26%) of *H. arabidopsidis* RXLR effectors contain WY-domain-like sequences (Table 2 and supplemental Table 2). In contrast, in the non-RXLR proteome, only 0.3 and 0.6% (*Phytophthora* and *H. arabidopsidis*, respectively) contain WY-domain-like sequences above the same cut-off (Table 2 and supplemental Table 2). We therefore conclude that our rate of false positive is ≤0.6% and that 562 of 1341 (42%) of oomycete RXLR effectors probably contain a core fold similar to that found in AVR3a11 and PexRD2.

WY-domains can also be found in tandem repeats, and we identified one candidate effector, PITG_23035, with 11 WY-domains. Therefore, many *Phytophthora* RXLR effectors comprise either single or tandem arrays of WY-domains, joined by variable "linker" regions.

A Core Protein Structure for Phytophthora RXLR Effectors That Supports Rapid Adaptive Diversification—Here we present the crystal structures of the effector domains from *Phytophthora* RXLR proteins AVR3a11 and PexRD2. We describe a structural unit, the WY-domain, that bioinformatic analysis suggests is conserved in ~44% of *Phytophthora* RXLR effectors and ~26% of *H. arabidopsidis* effectors. This domain forms a core α-helical fold that tolerates considerable plasticity at both the N and C termini and within the loops between the helices. It also supports amino acid substitutions in surface residues and oligomerization as further mechanisms for modification.

Understanding the mechanisms that underlie protein adaptation under evolutionary pressure is a key question in molecular evolution. Adaptive evolution can drive the development of novel protein functions or the tuning of existing ones (*e.g.* new substrate specificity in an enzyme active site or perturbation of molecular recognition surfaces). Some proteins can also be under selective pressure to lose particular activities (*e.g.* to evade immune system recognition but still need to maintain structural integrity to avoid total loss of function). The evolution of toxicity in Kallikrein-1-like venom proteins involves the acquisition of small insertions in loop regions adjacent to the catalytic cleft (within a conserved core fold), followed by accelerated sequence diversification that increases catalytic efficiency (47). Many small protein domains (~20–60 amino acids) exist as tandem repeats to deliver a structural template that can adapt to provide novel functions. Examples include leucine-rich repeats (48) and the blades of β-propellers. Each of these structural modules encode functional diversity through variation in the number of repeats, mutations in individual surface residues, and larger changes to loop regions. Further, TIM-barrel enzymes are built from a minimal ~20-amino acid α/β unit, usually found in eight repeats, with specificity for the types of reaction catalyzed, and their substrates, encoded by the β/α loops (49). The degree of sequence conservation across the family of known β-propellers and TIM-barrels is below detectable levels. It is generally accepted that all of these proteins have evolved from ancestral precursors through diversification.

Plant pathogen effectors frequently display extreme signatures of positive selection, presumably as a consequence of the co-evolutionary arms race with resistant hosts (3, 11, 44, 50, 51). Our findings provide a molecular framework that explains how RXLR effector proteins tolerate sequence hypervariability. Although alternative hypotheses are possible (*e.g.* the WY-domain is the result of convergence to a fold suited to secretion and/or translocation from *Phytophthora*), we favor the model that the WY-domain is a conserved structural unit found in a large family of oomycete RXLR effectors that share a common evolutionary origin.

Examples from this study that support this include the 16-amino acid insertion in loop-3 observed in PexRD2 compared with AVR3a11, which shows that the WY-fold can incorporate significant insertions without losing integrity (Fig. 1*e*). The

presence of α_1 in AVR3a11 reveals a modification at a terminus, and the structures of AVR3a11 and PexRD2 suggest that additional residues at the C terminus of the WY-fold could easily be accommodated. Known functional and polymorphic residues within the AVR3a and PexRD2 families are positioned on the surface of the proteins where they can influence interaction with host molecules without disrupting the fold. Further, the structure of PexRD2 reveals the WY-domain fold can support oligomerization (which may or may not involve the fold adaptations/elaborations) to deliver new effector structures. Oligomerization in PexRD2 also raises the intriguing possibility that RXLR effectors may permit further functional diversification by forming heterodimers. Future experiments will determine the degree to which homo- and heterodimerization impact the activity of PexRD2.

AVR3a11 and PexRD2 are proteins with only a single WY-domain. However, our computational analyses revealed that *Phytophthora* RXLR effectors can comprise up to 11 WY-domain repeats, greatly enhancing the potential for encoding new structures and oligomeric states. Future studies will address how these repeating WY-domains are organized into higher order structures and how this impacts effector function. We will also use the structures of AVR3a11 and PexRD2 as templates for mutagenesis to understand how these specific effectors modulate immunity inside plant cells (supplemental Fig. 1).

In conclusion, we have defined a large family of oomycete RXLR effectors that contain at least one WY-domain and are structurally related. However, this does not preclude the existence of *Phytophthora* RXLR effectors that contain alternative folds and are phylogenetically unrelated to WY-domain effectors (Fig. 4*c*). In fact, given the diversity of RXLR sequences, it seems likely that other folds do exist, and these have yet to be structurally characterized.

REFERENCES

1. Dodds, P. N., and Rathjen, J. P. (2010) *Nat. Rev. Genet.* **11,** 539–548
2. Baxter, L., Tripathy, S., Ishaque, N., Boot, N., Cabral, A., Kemen, E., Thines, M., Ah-Fong, A., Anderson, R., Badejoko, W., Bittner-Eddy, P., Boore, J. L., Chibucos, M. C., Coates, M., Dehal, P., Delehaunty, K., Dong, S., Downton, P., Dumas, B., Fabro, G., Fronick, C., Fuerstenberg, S. I., Fulton, L., Gaulin, E., Govers, F., Hughes, L., Humphray, S., Jiang, R. H., Judelson, H., Kamoun, S., Kyung, K., Meijer, H., Minx, P., Morris, P., Nelson, J., Phuntumart, V., Qutob, D., Rehmany, A., Rougon-Cardoso, A., Ryden, P., Torto-Alalibo, T., Studholme, D., Wang, Y., Win, J., Wood, J., Clifton, S. W., Rogers, J., Van den Ackerveken, G., Jones, J. D., McDowell, J. M., Beynon, J., and Tyler, B. M. (2010) *Science* **330,** 1549–1551
3. Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I., Bulone, V., Cai, G., Cakir, C., Carrington, J. C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M. A., Fugelstad, J., Gilroy,

E. M., Gnerre, S., Green, P. J., Grenville-Briggs, L. J., Griffith, J., Grünwald, N. J., Horn, K., Horner, N. R., Hu, C. H., Huitema, E., Jeong, D. H., Jones, A. M., Jones, J. D., Jones, R. W., Karlsson, E. K., Kunjeti, S. G., Lamour, K., Liu, Z., Ma, L., Maclean, D., Chibucos, M. C., McDonald, H., McWalters, J., Meijer, H. J., Morgan, W., Morris, P. F., Munro, C. A., O'Neill, K., Ospina-Giraldo, M., Pinzón, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D. C., Schumann, U. D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., van de Vondervoort, P. J., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B. C., van West, P., Ristaino, J., Govers, F., Birch, P. R., Whisson, S. C., Judelson, H. S., and Nusbaum, C. (2009) *Nature* **461,** 393–398

4. Schmidt, S. M., and Panstruga, R. (2011) *Curr. Opin. Plant Biol.* **14,** 392–399

5. Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M., Dorrance, A. E., Dou, D., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., Govers, F., Grunwald, N. J., Huang, W., Ivors, K. L., Jones, R. W., Kamoun, S., Krampis, K., Lamour, K. H., Lee, M. K., McDonald, W. H., Medina, M., Meijer, H. J., Nordberg, E. K., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P. F., Phuntumart, V., Putnam, N. H., Rash, S., Rose, J. K., Sakihama, Y., Salamov, A. A., Savidor, A., Scheuring, C. F., Smith, B. M., Sobral, B. W., Terry, A., Torto-Alalibo, T. A., Win, J., Xu, Z., Zhang, H., Grigoriev, I. V., Rokhsar, D. S., and Boore, J. L. (2006) *Science* **313,** 1261–1266

6. Bos, J. I., Armstrong, M. R., Gilroy, E. M., Boevink, P. C., Hein, I., Taylor, R. M., Zhendong, T., Engelhardt, S., Vetukuri, R. R., Harrower, B., Dixelius, C., Bryan, G., Sadanandom, A., Whisson, S. C., Kamoun, S., and Birch, P. R. (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107,** 9909–9914

7. Kearney, B., and Staskawicz, B. J. (1990) *Nature* **346,** 385–386

8. Kim, M. G., da Cunha, L., McFall, A. J., Belkhadir, Y., DebRoy, S., Dangl, J. L., and Mackey, D. (2005) *Cell* **121,** 749–759

9. Nomura, K., Debroy, S., Lee, Y. H., Pumplin, N., Jones, J., and He, S. Y. (2006) *Science* **313,** 220–223

10. Jones, J. D., and Dangl, J. L. (2006) *Nature* **444,** 323–329

11. Win, J., Morgan, W., Bos, J., Krasileva, K. V., Cano, L. M., Chaparro-Garcia, A., Ammar, R., Staskawicz, B. J., and Kamoun, S. (2007) *Plant Cell* **19,** 2349–2369

12. Schornack, S., van Damme, M., Bozkurt, T. O., Cano, L. M., Smoker, M., Thines, M., Gaulin, E., Kamoun, S., and Huitema, E. (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107,** 17421–17426

13. Whisson, S. C., Boevink, P. C., Moleleki, L., Avrova, A. O., Morales, J. G., Gilroy, E. M., Armstrong, M. R., Grouffaud, S., van West, P., Chapman, S., Hein, I., Toth, I. K., Pritchard, L., and Birch, P. R. (2007) *Nature* **450,** 115–118

14. Schornack, S., Huitema, E., Cano, L. M., Bozkurt, T. O., Oliva, R., Van Damme, M., Schwizer, S., Raffaele, S., Chaparro-Garcia, A., Farrer, R., Segretin, M. E., Bos, J., Haas, B. J., Zody, M. C., Nusbaum, C., Win, J., Thines, M., and Kamoun, S. (2009) *Mol. Plant Pathol.* **10,** 795–803

15. Jiang, R. H., Tripathy, S., Govers, F., and Tyler, B. M. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105,** 4874–4879

16. Armstrong, M. R., Whisson, S. C., Pritchard, L., Bos, J. I., Venter, E., Avrova, A. O., Rehmany, A. P., Böhme, U., Brooks, K., Cherevach, I., Hamlin, N., White, B., Fraser, A., Lord, A., Quail, M. A., Churcher, C., Hall, N., Berriman, M., Huang, S., Kamoun, S., Beynon, J. L., and Birch, P. R. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102,** 7766–7771

17. Bos, J. I., Chaparro-Garcia, A., Quesada-Ocampo, L. M., McSpadden Gardener, B. B., and Kamoun, S. (2009) *Mol. Plant Microbe Interact.* **22,** 269–281

18. Bos, J. I., Kanneganti, T. D., Young, C., Cakir, C., Huitema, E., Win, J., Armstrong, M. R., Birch, P. R., and Kamoun, S. (2006) *Plant J.* **48,** 165–176

19. Oh, S. K., Young, C., Lee, M., Oliva, R., Bozkurt, T. O., Cano, L. M., Win, J., Bos, J. I., Liu, H. Y., van Damme, M., Morgan, W., Choi, D., Van der Vossen, E. A., Vleeshouwers, V. G., and Kamoun, S. (2009) *Plant Cell* **21,** 2928–2947

20. Vleeshouwers, V. G., Rietman, H., Krenek, P., Champouret, N., Young, C., Oh, S. K., Wang, M., Bouwmeester, K., Vosman, B., Visser, R. G., Jacobsen, E., Govers, F., Kamoun, S., and Van der Vossen, E. A. (2008) *PLoS One* **3,** e2875

21. Dou, D., Kale, S. D., Wang, X., Chen, Y., Wang, Q., Wang, X., Jiang, R. H., Arredondo, F. D., Anderson, R. G., Thakur, P. B., McDowell, J. M., Wang, Y., and Tyler, B. M. (2008) *Plant Cell* **20,** 1118–1133

22. Shan, W., Cao, M., Leung, D., and Tyler, B. M. (2004) *Mol. Plant Microbe Interact.* **17,** 394–403

23. Yang, Z. R., Thomson, R., McNeil, P., and Esnouf, R. M. (2005) *Bioinformatics* **21,** 3369–3376

24. Leslie, A. G. (2006) *Acta Crystallogr. D Biol. Crystallogr.* **62,** 48–57

25. Evans, P. (2006) *Acta Crystallogr. D Biol. Crystallogr.* **62,** 72–82

26. Collaborative Computational Project 4 (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763

27. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221

28. Sheldrick, G. M. (2008) *Acta Crystallogr. A* **64,** 112–122

29. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J. Appl. Crystallogr.* **40,** 658–674

30. Cowtan, K. (2006) *Acta Crystallogr. D Biol. Crystallogr.* **62,** 1002–1011

31. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501

32. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) *Nucleic Acids Res.* **35,** W375–W383

33. Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A., and Noble, M. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2288–2294

34. Demeler, B. (2005) *Modern Analytical Ultracentrifugation: Techniques and Methods*, pp. 210–229, Royal Society of Chemistry, London

35. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) *Bioinformatics* **23,** 2947–2948

36. Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2256–2268

37. Lindbo, J. A. (2007) *Plant Physiol.* **145,** 1232–1240

38. Karimi, M., Inzé, D., and Depicker, A. (2002) *Trends Plant Sci.* **7,** 193–195

39. Holm, L., and Rosenström, P. (2010) *Nucleic Acids Res.* **38,** W545–549

40. Markson, J. S., and O'Shea, E. K. (2009) *FEBS Lett.* **583,** 3938–3947

41. Bos, J. I. B. (2007) *Function, Structure and Evolution of the RXLR Effector AVR3a of Phytophthora infestans*, Ph.D. thesis, Ohio State University, Columbus, OH

42. Samanta, U., Pal, D., and Chakrabarti, P. (1999) *Acta Crystallogr. D Biol. Crystallogr.* **55,** 1421–1427

43. Krissinel, E., and Henrick, K. (2007) *J. Mol. Biol.* **372,** 774–797

44. Raffaele, S., Farrer, R. A., Cano, L. M., Studholme, D. J., MacLean, D., Thines, M., Jiang, R. H., Zody, M. C., Kunjeti, S. G., Donofrio, N. M., Meyers, B. C., Nusbaum, C., and Kamoun, S. (2010) *Science* **330,** 1540–1543

45. Bailey, T. L., and Elkan, C. (1994) *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2,** 28–36

46. Eddy, S. R. (1998) *Bioinformatics* **14,** 755–763

47. Aminetzach, Y. T., Srouji, J. R., Kong, C. Y., and Hoekstra, H. E. (2009) *Curr. Biol.* **19,** 1925–1931

48. Kobe, B., and Kajava, A. V. (2001) *Curr. Opin. Struct. Biol.* **11,** 725–732

49. Wierenga, R. K. (2001) *FEBS Lett.* **492,** 193–198

50. Allen, R. L., Bittner-Eddy, P. D., Grenville-Briggs, L. J., Meitz, J. C., Rehmany, A. P., Rose, L. E., and Beynon, J. L. (2004) *Science* **306,** 1957–1960

51. Liu, Z., Bos, J. I., Armstrong, M., Whisson, S. C., da Cunha, L., Torto-Alalibo, T., Win, J., Avrova, A. O., Wright, F., Birch, P. R., and Kamoun, S. (2005) *Mol. Biol. Evol* **22,** 659–672