Effectors of Filamentous Plant Pathogens: Commonalities amid Diversity

Marina Franceschetti,a Abbas Maqbool,a Maximiliano J. Jiménez-Dalmaroni,a Helen G. Pennington,b Sophien Kamoun,b Mark J. Banfielda

Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich, United Kingdoma; The Sainsbury Laboratory, Norwich Research Park, Norwich, United Kingdomb

SUMMARY

Fungi and oomycetes are filamentous microorganisms that include a diversity of highly developed pathogens of plants. These are sophisticated modulators of plant processes that secrete an arsenal of effector proteins to target multiple host cell compartments and enable parasitic infection. Genome sequencing revealed complex catalogues of effectors of filamentous pathogens, with some species harboring hundreds of effector genes. Although a large fraction of these effector genes encode secreted proteins with weak or no sequence similarity to known proteins, structural studies have revealed unexpected similarities amid the diversity. This article reviews progress in our understanding of effector structure and function in light of these new insights. We conclude that there is emerging evidence for multiple pathways of evolution of effectors of filamentous plant pathogens but that some families have probably expanded from a common ancestor by duplication and diversification. Conserved folds, such as the oomycete WY and the fungal MAX domains, are not predictive of the precise function of the effectors but serve as a chassis to support protein structural integrity while providing enough plasticity for the effectors to bind different host proteins and evolve unrelated activities inside host cells. Further effector evolution and diversification arise via short linear motifs, domain integration and duplications, and oligomerization.
INTRODUCTION

Filamentous pathogens (fungi and oomycetes) are the causative agents of some of the world’s most notorious plant diseases. Left unchecked, they can devastate crop harvests, destroy managed and wild forests, affect the supply of ornamental plants, and disturb natural ecosystems (1–3). Perhaps the most famous plant disease outbreak was caused by the oomycete Phytophthora infestans, which spread to Europe and triggered the 19th-century Irish potato famine (4). This pathogen remains relevant in agriculture today, infecting potato and tomato crops throughout the world (5). Diseases caused by fungal pathogens, such as rice and wheat blast and wheat stem and stripe rust, are of immediate concern for global food security (1, 6, 7). Major factors in the ability of these filamentous microbes to cause disease on their hosts are effectors, pathogen-encoded proteins that are secreted to either the apoplast or specialized biotrophic interfaces (both are spaces outside plant cells) or are translocated inside host cells (8–11).

Effectors act to modulate host cell physiology to promote susceptibility to pathogens. In turn, plants have evolved cell surface and intracellular receptors to detect the presence of pathogen signatures and mount an immune response to restrict the progression of disease. Cell surface receptors typically recognize microbe-associated molecular patterns (MAMPs), derived from abundant structural components of microbes’ cell walls, or secreted proteins that function as virulence effectors. Intracellular receptors respond to the presence of translocated effectors and/or their activity on host cell targets. These intracellular receptors are nucleotide-binding domain- and leucine-rich repeat-containing (NLR) proteins that mediate innate immunity to pathogens in both plants and animals (recently reviewed in reference 12).

One of the defining features of effector proteins, be they of bacterial or filamentous pathogen origin, is the lack of clear sequence similarity to proteins of known function. This is thought to be the consequence of evolutionary pressure that drives the rapid diversification of effector activities in host cells to optimize function and/or avoid recognition by the innate immune system. The frequent difficulty in recognizing common motifs that indicate the function or activity of effectors may be due to few of them having enzymatic activity or the absence of known domains for direct interaction with host factors. In addition, many effectors are small proteins of ≤15 kDa, and thus, their rapid diversification would result in a loss of sequence similarity. With a few notable exceptions (the RXLR motif of effectors in some oomycetes being the most prominent), this sequence diversity has meant that it is challenging to confidently produce catalogues of effectors from filamentous plant pathogen genomes despite many of these now being available. In some cases, bioinformatic approaches have been useful in predicting and classifying candidate effectors from filamentous plant pathogens (13–23) (Table 1). However, it can be challenging to pick the most relevant proteins to select for further investigation from these lists. These bioinformatic approaches use some of the commonalities identified among effectors from different organisms, such as genomic context, the presence of a secretion signal, the absence of predicted transmembrane domains, expression patterns, and the lack of similarity to known protein domains. Recent advances in the computational prediction of effectors have employed machine-learning approaches, which are proving useful for prioritizing effectors for further study (24). There are also examples of effectors of filamentous plant pathogens that share common sequence motifs with known enzymes, enzyme inhibitors, sugar-binding proteins, and toxins, with some being shown to possess such activities.

It is well established that protein structure is more conserved than amino acid sequence, and in many cases, this is due to the evolutionary relationship between structure and function (25). The fact that structural conservation can be a powerful method for the functional annotation of proteins is a fundamental concept that has driven the development of structure determination as a tool to understand the effector
biology of both mammalian and plant pathogens (26, 27). In particular, this has been important where the lack of sequence similarity to known functional proteins has prevented the prediction of the molecular mechanism.

In this review, we focus on recent advances that highlight commonalities shared by effectors of filamentous plant pathogens, focusing on functional similarities with known proteins, on effectors that cluster into large structurally common but sequence-divergent families comprising novel folds, or on those that share structural similarity with proteins of known function. It is timely to review progress in this area in light of new insights. We conclude that there is emerging evidence for multiple pathways of evolution of effectors of filamentous plant pathogens, including that some families appear to have evolved from a common ancestor by duplication and diversification in the pathogen.

### EFFECTORS OF FILAMENTOUS PLANT PATHOGENS THAT ENCODE ENZYMES AND PROTEASE INHIBITORS

Structural studies of a number of bacterial plant-pathogenic type III secreted effectors (T3SEs) have revealed similarity with proteins of known function, which suggested both how these proteins act and experiments to test mechanisms (28–31). Remarkably, many of these proteins appear to be enzymes with the potential to catalyze a wide variety of different reactions, such as E3 ligation, ADP ribosylation, and proteolysis. In several cases, specific enzymatic activities have been demonstrated for these proteins (32). In contrast, a number of effectors of filamentous plant pathogens have been predicted to have enzymatic activity, but only a few have had such activities confirmed experimentally. To date, there are no structures of enzymes of effectors of filamentous plant pathogens, so these predictions typically rely primarily on sequence comparisons.

<table>
<thead>
<tr>
<th>Effector class</th>
<th>Hyphal pathogen</th>
<th>Example(s)</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chorismate mutases</td>
<td>Ustilago maydis</td>
<td>Cmu1</td>
<td>45</td>
</tr>
<tr>
<td>Lipase effector</td>
<td>Fusarium graminearum</td>
<td>FGL1</td>
<td>112</td>
</tr>
<tr>
<td>Enzyme inhibitors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protease inhibitors</td>
<td>Cladosporium fulvum</td>
<td>Avr2</td>
<td>41</td>
</tr>
<tr>
<td>Cystatin-like protease inhibitor domains</td>
<td>Phytophthora infestans</td>
<td>EPIC1, EPIC2B</td>
<td>42</td>
</tr>
<tr>
<td>Chitinase inhibitor</td>
<td>Cladosporium fulvum</td>
<td>Avr4</td>
<td>56</td>
</tr>
<tr>
<td>Proteases and peptidases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteases</td>
<td>Zymoseptoria tritici (Mycosphaerella graminicola)</td>
<td>Astacin (peptidase family M12A), serine carboxypeptidase S28</td>
<td>34</td>
</tr>
<tr>
<td>Secreted peptidases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serine protease</td>
<td>Fusarium oxysporum f. sp. lycopersici</td>
<td>Sep1</td>
<td>35</td>
</tr>
<tr>
<td>Alkaline serine protease alp1</td>
<td>Sclerotinia sclerotiorum</td>
<td>Peptidase inhibitor I9</td>
<td>23</td>
</tr>
<tr>
<td>Metalloproteases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zinc metalloprotease</td>
<td>Magnaporthe oryzae</td>
<td>AVR-Pita (AVR2-YAMO)</td>
<td>36, 114</td>
</tr>
<tr>
<td>Deuterolysin metalloprotease</td>
<td>Sclerotinia sclerotiorum</td>
<td>Deuterolysin metalloprotease (M35) family (PF02102) homolog of M. oryzae AVR-Pita</td>
<td>23</td>
</tr>
<tr>
<td>Metalloprotease</td>
<td>Fusarium oxysporum f. sp. lycopersici</td>
<td>Mep1</td>
<td>35</td>
</tr>
<tr>
<td>Nudix hydrolases</td>
<td>Phytophthora sojae</td>
<td>Avr3b</td>
<td>46</td>
</tr>
<tr>
<td>Colletotrichum truncatum</td>
<td>CNUDIX</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>Melampsora lini</td>
<td>AvrM14</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>Crinklers</td>
<td>Kinase activity</td>
<td>Phytophthora infestans</td>
<td>CRN8</td>
</tr>
</tbody>
</table>
Proteases and Protease Inhibitors

Analyses of fungal genomes, including those of *Zymoseptoria tritici* (33), *Colletotrichum* sp. (34), and *Sclerotinia sclerotiorum* (23), identified families of secreted proteases whose expression pattern supports a putative role as effectors, to promote the colonization and growth of the pathogen. *Fusarium oxysporum* f. sp. *lycopersicum* secretes a serine protease, Sep1, and a metalloprotease, Mep1, that act synergistically to cleave host chitinases, preventing their activity in degrading fungal cell walls (35). A double mutant of Sep1 and Mep1 showed reduced disease on tomato, highlighting the importance of these proteins for full virulence.

The rice blast fungus *Magnaporthe oryzae* produces AVR-Pita, an effector with features typical of zinc metalloproteases, including conserved residues known to mediate zinc coordination and catalysis in homologues from other organisms (9, 36). However, to date, actual protease activity for AVR-Pita has not been demonstrated.

A remarkable case is the glucanase inhibitor proteins (GIPs), which are proteins secreted by *Phytophthora* spp. to inhibit the degradation of pathogen β-1,3/1,6-glucans and the release of defense-eliciting oligosaccharides by host β-1,3-endoglucanases (37, 38). GIPs share significant sequence similarity with trypsin serine proteases but are predicted to be proteolytically nonfunctional because they carry mutated catalytic residues.

Interestingly, filamentous plant pathogens also secrete protease inhibitors, which act on host pathogenesis-related proteases to prevent their activities. Examples include EPI1 and EPI10 of *P. infestans*, which carry multiple domains with similarity to the Kazal family of serine protease inhibitors (39, 40). In addition, the Avr2 effector of the fungal pathogen *Cladosporium fulvum* (41) and the *P. infestans* effectors EPIC1 and EPIC2 (42) are unrelated in sequence but have convergently evolved to target the same host proteases (43, 44). The oomycete EPIC family of protease inhibitor effectors has similarity to the widespread cystatin domain (42), whereas *C. fulvum* Avr2 is a small cysteine-rich protein without any notable sequence similarity to other proteins (41).

Fungal Cmu1, an Enzyme Interfering with Metabolic Flux

The maize smut fungus *Ustilago maydis* translocates a chorismate mutase, Cmu1, into plant cells. Cmu1 appears to benefit the pathogen by redirecting the metabolic flux of chorismate away from the biosynthesis of salicylic acid, suppressing the accumulation of this defense-related hormone during infection. Intriguingly, there is evidence to suggest that Cmu1 can move out of infected cells into neighboring cells, where the enzyme’s activity can “prime” the host tissue for infection (45).

Translocated Oomycete Effectors Include Enzymes

Oomycete plant pathogens encode putative enzymes in their effector repertoires. *Phytophthora* species have ~300 to 550 RXLR-type effectors that rarely have sequence similarity to known enzyme folds. However, *P. infestans* and *Phytophthora sojae* contain a sequence signature suggestive of Nudix hydrolase (phosphorylase) activity. The *P. sojae* effector Avr3b has been shown to possess ADP-ribose/NADH pyrophosphorylase activity when expressed and epitope purified from plant tissue (46). Furthermore, the virulence activity of Avr3b was dependent on the conserved Nudix motif. Interestingly, the activity of Avr3b as a Nudix hydrolase is dependent on its modification by plant cyclophilins; when produced in *Escherichia coli*, the protein is not active (47). Recently, a putative Nudix hydrolase effector (AvrM14) was identified in the flax rust fungus *Melampsora lini* (48), but catalytic activity for this protein has yet to be shown.

In addition to RXLR effectors, *Phytophthora* species also contain hundreds of “Crinkler” effectors (CRNs) (13, 16, 49). CRNs are modular proteins, some of which induce cell death upon expression in plant cells (13, 16). One C-terminal CRN domain has significant sequence similarity to protein Ser/Thr kinases of the RD (arginine-aspartate) class. Indeed, *P. infestans* CRN8 was shown to be an active kinase present in an autophosphorylated state in plant cells (50). In planta expression of CRN8 enhanced the growth
of P. infestans, and this required the intact RD motif, suggesting that the enzymatic activity of this kinase is relevant for virulence.

**EFFECTORS OF FILAMENTOUS PLANT PATHOGENS CAN SHARE FOLDS WITH FUNCTIONALLY SIMILAR PROTEINS**

**Chitin-Binding LysM Effectors**

Chitin is a major component of fungal cell walls, and the detection of this homopolymer in the apoplast is used by plants as a strategy for initiating immune responses (51). Plants detect chitin-derived oligosaccharides via cell surface receptors that contain extracellular lysine motif (LysM) domains. Plant LysM domains comprise ~50 amino acids and adopt a βαββ structural fold (52, 53) (Fig. 1). To protect themselves from detection by the plant immune system, fungi use LysM effectors to sequester chitin oligomers in the apoplast, outcompeting binding by host receptor domains. The crystal structure of Cladosporium fulvum Ecp6 confirmed that this protein contained 3 modular LysM domains (54) (Fig. 1 and Table 2). In a strategy to deliver high-affinity ligand interactions, two of the Ecp6 LysM domains (LysM1 and LysM3) dimerize to “sandwich” a chitin oligomer in a groove via multiple hydrogen bonds and hydrophobic interactions (Fig. 1A). To date, this ligand-induced LysM dimerization to increase binding affinity is unique to Ecp6 and highlights the propensity of pathogen effectors to adapt protein folds to acquire new activities (51). Interestingly, the ligand-binding capability of the LysM2 domain of Ecp6 was also shown to interfere with chitin-triggered immunity in planta, but the underlying mechanistic basis remains unclear (55).

Multidomain LysM effectors are also found in other fungal plant pathogens, including the wheat pathogen Zymoseptoria tritici and the rice blast pathogen Magnaporthe oryzae, suggesting that they represent a widespread mechanism for the suppression of detection by the plant immune system. However, unlike Ecp6, Z. tritici LysM effectors protect fungal hyphae against hydrolysis by host chitinases, although the mechanism by which they achieve this is not understood (55).
<table>
<thead>
<tr>
<th>Protein</th>
<th>Origin</th>
<th>Targeted process</th>
<th>Immune receptor(s)</th>
<th>Fold</th>
<th>RMSD (Å) (no. of residues in overlay)</th>
<th>Sequence identity (%)</th>
<th>PDB accession no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avr3a11</td>
<td><em>P. capsici</em></td>
<td>Unknown</td>
<td>WY</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>3ZRB</td>
<td>74</td>
</tr>
<tr>
<td>Avr3a4</td>
<td><em>P. capsici</em></td>
<td>Unknown</td>
<td>WY</td>
<td>1.26</td>
<td>(42)</td>
<td>79.0</td>
<td>2LC2</td>
<td>77</td>
</tr>
<tr>
<td>PexRD2</td>
<td><em>P. infestans</em></td>
<td>MAPKKKε-mediated immune signaling</td>
<td>WY</td>
<td>1.41</td>
<td>(40)</td>
<td>27.8</td>
<td>3ZRG</td>
<td>74</td>
</tr>
<tr>
<td>PexRD54</td>
<td><em>P. infestans</em></td>
<td>Autophagy</td>
<td>WY</td>
<td>1.73</td>
<td>(41)</td>
<td>20.0</td>
<td>5L75</td>
<td>78</td>
</tr>
<tr>
<td>ATR1</td>
<td><em>H. arabidopsis</em></td>
<td>Unknown</td>
<td>RPP1</td>
<td>2.37</td>
<td>(42)</td>
<td>23.7</td>
<td>3RMR</td>
<td>76</td>
</tr>
<tr>
<td>AvrLS67-D</td>
<td><em>M. lini</em></td>
<td>Unknown</td>
<td>L6</td>
<td>2.74</td>
<td>(82)</td>
<td>22.2</td>
<td>2QVT</td>
<td>105</td>
</tr>
<tr>
<td>AvrLS67-A</td>
<td><em>M. lini</em></td>
<td>Unknown</td>
<td>L5 and L6</td>
<td>2.58</td>
<td>(81)</td>
<td>19.7</td>
<td>2OPC</td>
<td>105</td>
</tr>
<tr>
<td>avrM</td>
<td><em>M. lini</em></td>
<td>Unknown</td>
<td>WY-like</td>
<td>ND</td>
<td>26.1</td>
<td>4BJM</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>AvrM-A</td>
<td><em>M. lini</em></td>
<td>Unknown</td>
<td>M</td>
<td>ND</td>
<td>23.9</td>
<td>4BJN</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>AVR-PikD (in complex)</td>
<td><em>M. oryzae</em></td>
<td>Unknown</td>
<td>Pik1/Pik2</td>
<td>ND</td>
<td>2.24 (52)</td>
<td>16.4</td>
<td>2MYW</td>
<td>80</td>
</tr>
<tr>
<td>Avr1-CO39</td>
<td><em>M. oryzae</em></td>
<td>Unknown</td>
<td>RGAS/RGA4</td>
<td>1.36</td>
<td>(55)</td>
<td>17.2</td>
<td>2MYV</td>
<td>80</td>
</tr>
<tr>
<td>AVR-PHa</td>
<td><em>M. oryzae</em></td>
<td>Unknown</td>
<td>RGAS/RGA4</td>
<td>2.24</td>
<td>(52)</td>
<td>16.4</td>
<td>2MYW</td>
<td>80</td>
</tr>
<tr>
<td>AVR-Pizt</td>
<td><em>M. oryzae</em></td>
<td>E3 ligase-mediated immunity</td>
<td>Piz-t</td>
<td>2.33</td>
<td>(58)</td>
<td>15.6</td>
<td>2LW6</td>
<td>84</td>
</tr>
<tr>
<td>Avr4</td>
<td><em>P. fuligina</em></td>
<td>Chitin-mediated immunity/fungally derived chitin perception</td>
<td>CF-4</td>
<td>1.98</td>
<td>(52)</td>
<td>22.2</td>
<td>4Z4A</td>
<td>61</td>
</tr>
<tr>
<td>Ecp6</td>
<td><em>C. fulvum</em></td>
<td>Chitin-mediated immunity/fungally derived chitin perception</td>
<td>LysM1</td>
<td>0.8</td>
<td>(45)</td>
<td>35.9</td>
<td>4B8V</td>
<td>54</td>
</tr>
<tr>
<td>Ecp6</td>
<td><em>C. fulvum</em></td>
<td></td>
<td>LysM2</td>
<td>1.17</td>
<td>(43)</td>
<td>37.1</td>
<td>4B8V</td>
<td>54</td>
</tr>
<tr>
<td>Ecp6</td>
<td><em>C. fulvum</em></td>
<td></td>
<td>LysM3</td>
<td>1.51</td>
<td>(45)</td>
<td>20.8</td>
<td>4B8V</td>
<td>54</td>
</tr>
<tr>
<td>AvrLM4-7</td>
<td><em>L. maculans</em></td>
<td>Production of plant hormones and hydrogen peroxide/plant hormone-mediated immunity</td>
<td>Rlm4 and Rlm7</td>
<td>Unique</td>
<td>ND</td>
<td>ND</td>
<td>4FPR</td>
<td>110</td>
</tr>
<tr>
<td>ToxA</td>
<td><em>P. tritic-repentis</em></td>
<td>Photosynthesis</td>
<td>Tsn1ab</td>
<td>ToxA-like</td>
<td>ND</td>
<td>ND</td>
<td>1ZLE</td>
<td>103</td>
</tr>
<tr>
<td>ToxB</td>
<td><em>P. tritic-repentis</em></td>
<td>Photosynthesis</td>
<td>MAX</td>
<td>2.25</td>
<td>(58)</td>
<td>25.4</td>
<td>2MM0</td>
<td>81</td>
</tr>
<tr>
<td>toxb</td>
<td><em>P. tritic-repentis</em></td>
<td>Inactive allele</td>
<td>MAX</td>
<td>2.33</td>
<td>(57)</td>
<td>19.7</td>
<td>2MM2</td>
<td>81</td>
</tr>
<tr>
<td>NLP</td>
<td><em>P. aphanidermatum</em></td>
<td>Plasma membrane integrity</td>
<td>Actinoporin-like</td>
<td>2.34</td>
<td>(68)</td>
<td>21.9</td>
<td>3GNZ</td>
<td>64</td>
</tr>
<tr>
<td>NLP</td>
<td><em>M. penniscola</em></td>
<td>Plasma membrane integrity</td>
<td>Actinoporin-like</td>
<td>2.24</td>
<td>(68)</td>
<td>19.3</td>
<td>3ST1</td>
<td>70</td>
</tr>
</tbody>
</table>

*Template proteins used for comparison are Avr3a11 (WY and WY-like), AVR-PikD (MAX), tachycitin (CBM14-like), MoCVNH3 (LysM), ToxA (ToxA-like), and sticholysin II (actinoporin-like). RMSD, root mean square deviation; ND, not determined (either to avoid comparison with self or because the comparison is not meaningful).

*bTsn1 is a susceptibility factor.

MAPKKKε, mitogen-activated protein kinase kinase kinase ε.
CBM14-Like Avr4 Effectors

In a second strategy to evade chitin-mediated recognition by the plant immune system, fungi can secrete effector proteins that bind to chitin in their cell wall and prevent the action of host chitinases in generating chito-oligosaccharide fragments. The *Cladosporium fulvum* effector Avr4 was predicted to adopt a carbohydrate-binding module family 14 (CBM14)-like structure, based on its disulfide bond pattern, and in *vitro*, Avr4 protects chitin from hydrolysis by plant chitinases (56, 57). CBM14 proteins are defined as having chitin-binding activity, with one being characterized as having antimicrobial properties (58). The structure of the CBM14 member tachycitin, from the horseshoe crab *Tachypleus tridentatus*, revealed a distorted β-sandwich fold flanked by short loops and turns, stabilized by disulfide bonds (59). Tachycitin was described as sharing some structural similarity to a domain found in the plant chitin-binding protein hevein (60).

Avr4 homologues are found in a number of plant-pathogenic fungal species. Recently, the crystal structure of Avr4 from the tomato pathogen *Pseudocercospora fuligena* confirmed that the Avr4 family of effectors adopts the CBM14-like fold (Fig. 2), and this enabled the investigation of structure-function relationships in chitin binding by these proteins (61). As predicted for tachycitin, the chitin-binding site of Avr4 is located between two β-strands and the connecting β-hairpin and is mediated by aromatic amino acids and adjacent polar residues (Fig. 2).

The evolutionary dynamics of CBM14 family proteins are complex (62). While chitin binding is a critical feature of this fold for fungal defense against the plant immune system, it is clear that other functions can be attributed to the wider family given that CBM14 proteins occur in nonpathogenic species and were previously shown to have antimicrobial properties.

NLPs

NLPs (necrosis- and ethylene-inducing peptide 1-like proteins) are a large family of secreted proteins found in plant-associated fungi, oomycetes, and bacteria. NLPs were initially characterized by their ability to induce necrotic cell death in dicotyledonous plants (63), which is thought to be dependent on toxin-induced host cell damage (64). However, it is now well established that not all NLPs share this activity (65, 66). Despite this, both cytotoxic and noncytotoxic NLPs can trigger cell surface-dependent immune responses in plant cells, and this activity has been localized to a 24-amino-acid peptide (67, 68) recognized by a receptor complex comprising RLP23/SOBIR-1/BAK1 (69). Clues to the mechanism of the cytolytic activity of NLPs came from the crystal structures of

---

**FIG 2** CBM14 family structure of *P. fuligena* Avr4. The structures comprise an alpha helix (yellow) and five beta strands (green). The residues predicted to be involved in the interaction with chitin are shown in blue.
NLPs from *Pythium aphanidermatum* and *Moniliophthora perniciosa* (Fig. 3), which showed that this family of proteins shares a fold with the actinoporin pore-forming toxin stichoysin (64, 70). However, there is no experimental evidence for pore-forming activity by NLPs, and their toxicity may be the result of the NLP-induced release of membrane damage factors that are then sensed by the plant (68). Interestingly, the 24-amino-acid peptide, which acts as a MAMP for the activation of plant immunity, is largely buried within the core of the intact structure, with only a small number of residues being displayed on the surface (67). This suggests that the protein is probably unfolded and/or digested for recognition by the receptor.

**THE THREE-DIMENSIONAL STRUCTURES OF EFFECTORS OF FILAMENTOUS PLANT PATHOGENS SHOW CONSERVED FOLDS WITHIN FAMILIES**

**Oomycete Effectors and the WY Fold**

The RXLR class of host-translocated oomycete effector proteins is defined by the presence of a conserved N-terminal RXLR motif and a diverse C-terminal domain that exerts effector activity inside the host cell (16, 71, 72). Analysis of the sequences of the RXLR repertoires of *Phytophthora sojae* and *Phytophthora ramorum* identified conserved motifs, which were named “W” (Trp), “Y” (Tyr), and “L” (Leu), after the single-letter amino acid code for a highly conserved residue in each sequence (73). Protein structural analysis subsequently revealed that the amino acids at the conserved W and Y positions were buried in the hydrophobic core of a three-α-helical bundle and stacked against one another in an energetically favorable interaction (74) (Fig. 4). Intriguingly, except for the *Hyaloperonospora arabidopsidis* effector ATR13 (75), all of the structures of oomycete RXLR effectors that have been determined to date adopt the “WY domain” fold. Nonetheless, these proteins display significant primary sequence differences. They also show diverse structural adaptations, including N- and C-terminal extensions, loop regions, and domain duplication, that give rise to very different overall structures (74, 76–78) (Fig. 4). Hidden Markov model (HMM) sequence searches, based on the knowledge of the WY domain structure, predicted that nearly half of the RXLR effector complement of *Phytophthora* species would adopt this fold (74).

The structure of the *P. infestans* effector PexRD2 is comprised of five α-helices, three of which contribute to the WY domain three-α-helical bundle (Fig. 4A). The additional helices (present between two helices of the core WY domain) are instrumental in forming an extensive homodimeric interface in the PexRD2 structure, consistent with the observation that PexRD2 self-associates *in planta*. The structures of *Phytophthora capsici* AVR3a4 and AVR3a11 comprise monomeric four-helical bundles (Fig. 4B), with an N-terminal helical extension to the WY domain fold (74). It is possible that the
The HMM-based sequence searches mentioned above revealed that these effectors could also comprise tandemly repeated WY domains encoded by a single gene. The first crystal structure of a tandem WY domain effector was that of ATR1 from *Hyaloperonospora arabidopsidis* (76) (Fig. 4C). In ATR1, two WY domains (each with an N-terminal helical extension) are connected through an additional helix, which acts as a linker. Recently, the crystal structure of PexRD54 revealed how five WY domains can pack together in a stable structure with diverse domain-domain interactions (78) (Fig. 4D). Within each of these tandem WY domain structures, the individual domains can be overlaid with high confidence despite limited sequence identity (76, 78). Interestingly, PexRD54 employs a short linear motif known as the ATG8-interacting motif (AIM) to engage a host protein and to exert its virulence activity (79). The AIM is presented at the C terminus of PexRD54 and is linked to the last WY domain via a short helix (coral) at the C-terminal end prior to the ATG8-interacting motif (AIM) (not shown, as it was disordered in the crystals). All structure figures were prepared with CCP4mg (111).

N-terminal helix is important for maintaining the stability of monomeric, single-WY-domain proteins, although this has not been explicitly tested.

The WY domain fold serves as a chassis for the evolution of novel functions in oomycete effectors while maintaining their structural integrity. The fold presents a flexible platform that supports effector evolution and diversification via the acquisition of short linear motifs, domain duplications, and dimerization. Thus, the WY domain structure is not predictive of the precise function of the effectors but appears to provide enough plasticity for the effectors to bind different host proteins and evolve unrelated activities inside host cells.

**MAX Effectors of *Magnaporthe***

Recently, a new family of effectors of filamentous plant pathogens has been described, which also shares a conserved common structure but displays a diverse...
protein sequence. The Magnaporthe Avrs and ToxB-like (MAX) family was defined following structural work on effectors from the fungal pathogen *M. oryzae*, the causal agent of rice blast disease (80). Despite typically sharing less than 25% sequence identity, each member of this family that has had a structure determined (80–84) shares a characteristic six-stranded β-sandwich fold (Fig. 5). This fold is stabilized by at least one disulfide bond, generally with Cys residues present in β1 and in, or immediately before, β5. In most cases, one of the β-sheets is formed by strands β1, β2, and β6, and the second is formed by strands β3, β4, and β5. The length and orientation of the different structural elements are variable, in particular for strand β5 and for the various connecting loops, giving rise to proteins with distinct shapes and surface properties (80). In addition, the *M. oryzae* effector AVR-PikD contains an N-terminal extension to the six-stranded β-sandwich structure (Fig. 5A), and this region contains polymorphic residues that contribute to the evasion of recognition by the plant innate immune system (82, 85). Interestingly, the *M. oryzae* effectors AVR-Pik, AVR-Pia, and AVR1-CO39 all bind to heavy metal-associated (HMA) domains that have been integrated in intracellular plant immune receptors (NLRs) throughout evolution. This suggests that the conserved MAX effector family fold is well suited to interact with such domains and may suggest a putative virulence target in host cells for these effectors.

Intriguingly, the MAX effector family includes ToxB, a proteinaceous toxin from the fungus *Pyrenophora tritici-repentis* (86). This toxin shares the common three-dimensional structure of MAX effectors (Fig. 5E and F), but its mode of action is unclear, and no interacting partner has been identified. However, the N-terminal region of ToxB has been shown to be essential for activity, while both the central and C-terminal parts are required for full activity (87), suggesting that the conserved structure is important

---

**FIG 5** The structures of MAX effectors reveal the shared β-sandwich fold. The conserved β-strands are shown in a cartoon representation for each protein, with residues contributing to disulfide bridges shown as sticks (in yellow) and loops in gray. Shown are AVR-PikD (A), AVR1-CO39 (B), AVR-Pia (C), AVR-Pizt (D), ToxB (E), and toxb (F), with amino (N) and carboxyl (C) termini labeled.
for function. A naturally occurring nontoxic version of ToxB (toxb) shares 78% sequence identity with the active protein. These proteins share essentially the same structure, although toxb may overall be less stable than ToxB (81).

PSI-BLAST followed by HMM-based profile searches revealed that the majority of MAX effectors are found in Magnaporthe species (80). However, a small number of hits were detected in other fungal species such as Colletotrichum (80). Thus, the discovery of the MAX effectors enables a more robust prediction of candidate effectors in these fungal pathogens.

RALPH Effectors of Powdery Mildew

Nearly 500 candidate effectors of the barley powdery mildew fungus Blumeria graminis f. sp. hordei were predicted from the genome sequence using bioinformatic tools by searching for genes with characteristics of effectors, particularly those encoding small secreted proteins. Many of these candidate effectors have been shown to be expressed during infection (88–90).

To further characterize B. graminis candidate effectors, their sequences were subjected to structural annotation using protein fold recognition methods. A subset of these candidate effectors are predicted to have structural similarities with ribonucleases and were named RALPHs (RNase-like proteins expressed in haustoria) (91). Although confirmation that RALPHs adopt RNase-like folds awaits the determination of an experimentally derived structure, it is intriguing that many B. graminis effectors may share a structural scaffold with each other, a feature common in other families of effectors of filamentous plant pathogens. In another parallel with the MAX effectors, RALPHs have been predicted to contain a disulfide bond, with Cys residues being largely conserved toward both the N terminus (contained within a "YxC" motif) and the C terminus of the proteins.

Recently, data have emerged showing that RALPH effectors function as both virulence and avirulence determinants in B. graminis-barley and -wheat interactions. Using host-induced gene silencing, five RALPHs were shown to be involved in the formation of haustoria (92, 93). AVR_A, and AVR_A13 were shown to be required for disease resistance in barley mediated by the powdery mildew resistance loci Mla1 and Mla13, respectively (94), and AvrPm2 was recently cloned as the cognate effector of the wheat Pm2 gene (95). Furthermore, the B. graminis f. sp. tritici effector SvrPm3a1/f1 (formerly Bcg1avr) has been shown to suppress avirulence triggered by the interaction of effector AvrPm3a2/f2 (svrPm3a1/f1, formerly Bcg1avr) with its receptor Pm3a/f (96, 97). As with other host-translocated effectors, the ability of RALPHs to activate plant immune responses may help explain the strong diversifying selection seen in these proteins.

STRUCTURES OF OTHER NOTABLE EFFECTORS OF FILAMENTOUS PLANT PATHOGENS

Flax Rust Effectors Show Divergent Structures

Melampsora lini causes rust disease on crop plants such as flax and linseed. Genomic analyses of M. lini predicted that this fungus has a large repertoire of putative effector proteins (22). Unlike oomycete RXLR and CRN effectors, but similar to effectors from other fungal species, no widely conserved sequence-based motifs have been identified for flax rust effectors thus far. To date, six M. lini effector proteins have been validated experimentally, based on their avirulence activity (AvrL567, AvrM, AvrP4, AvrP123, AvrL2, and AvrM14) (48, 98–101). These effectors trigger specific immune responses mediated by NLRs in the host cell. AvrL567, AvrM, and their cognate NLRs exhibit polymorphisms giving rise to allelic variants of the effector and receptor with specific recognition profiles (98, 102). For example, AvrL567-A is recognized by the NLRs L5 and L6, whereas AvrL567-D is recognized by L6 but not L5.

Crystal structures of the AvrL567 alleles AvrL567-D and AvrL567-A revealed that the two proteins share the same architecture, adopting a β-sandwich fold comprising seven antiparallel β-strands (Fig. 6A). Interestingly, the structures share some homology with ToxA (103), a host-selective toxin of Pyrenophora tritici-repentis which induces cell
death in sensitive wheat cultivars. ToxA was described as having a distant relationship with mammalian fibronectin proteins, and an Arg-Glu-Asp (RGD) motif was found in a loop region of the protein that may mediate interactions with plant cell integrin-like receptors (103). This motif was subsequently shown to be required for protein internalization (104), although the precise mechanism remains unclear. AvrL567 lacks the RGD motif, implying that it is internalized by a different mechanism. Both AvrL567-D and -A display two positively charged patches on the protein surface and have been shown to bind nucleic acid in vitro (105). However, the biological relevance of nucleic acid binding remains unknown. Structure-led mutagenesis revealed that multiple contacts mediate the interaction between AvrL567 alleles and their cognate receptors (105).

Crystal structures of C-terminal domains of two allelic variants of AvrM (AvrM-A and avrM) revealed an L-shaped /H9251-helical fold comprising two helical repeats (106) (Fig. 6B). The structural repeat, another example of modularity in effectors of filamentous plant pathogens, was not evident from sequence analysis and was revealed only after the structure was determined.

AvrLm4-7, a Lone Effector Structure with a Novel Fold

AvrLm4-7 is a Cys-rich protein that is recognized by oilseed rape cultivars harboring Rlm4 and Rlm7 resistances (107). The loss of AvrLm4-7 in the pathogen strongly impacts pathogen fitness (108, 109). The crystal structure of AvrLm4-7 does not share significant homology with other structures in the Protein Data Bank, and as such, it has proven challenging to infer putative protein function (110). The crystal structure identified the positions of the four disulfide bonds in the protein, which, as for other effectors, are probably involved in stabilizing the structure. In addition, a strongly positive patch was identified on the protein surface, which may represent a functionally relevant surface of the protein, although it has not been possible to show that this region binds a negatively charged ligand. A single amino acid polymorphism that perturbs the recognition of the effector by Rlm4 is located on a loop of the protein, exposed to the surface. It is therefore unlikely that this polymorphism affects the overall structure of the protein, but it may be important for a specific recognition site.

CONCLUSION

The high complexity of the secretomes of filamentous plant pathogens points to a multitude of independent evolutionary pathways to generate effector proteins that target a diversity of host molecules and processes. However, despite this extraordinary sequence diversity, it is now evident that some conserved protein folds, such as the WY
and MAX domains, define widespread families of effector proteins that occur across different plant pathogen taxa. There are both practical and theoretical implications of this finding. Structure-guided sequence similarity searches enable more precise and sensitive annotation of effector catalogues, notably of fungal effectors, which have proven more difficult to annotate than their oomycete counterparts. This should enable prioritization of effectors for further study, thus accelerating their functional characterization.

In addition, the conserved structures provide a framework to unravel how the rapid evolution of effector proteins has resulted in new host targeting activities and prioritization of effectors for further study, thus accelerating their functional characterization.

Acknowledgments

M.J.B. is supported by the BBSRC (UK) (relevant grants J004553 and M02198), the ERC (proposals 294608 [acronym NGRB] and SEP-210218966 [acronym ImmunityByPairDesign]), and the John Innes Foundation. S.K. is funded by the Biotechnology and Biological Sciences Research Council, the European Research Council (NGRB), and the Gatsby Charitable Foundation.

We thank Amey Redkar for discussions.

REFERENCES

Franceschetti et al. Microbiology and Molecular Biology Reviews


Downloaded from http://mmbr.asm.org/ on March 30, 2017 by NBI Library


