The structure and function of effector proteins expressed by *Austropuccinia psidii* in the early stages of myrtle rust infection.

Tyler Johns[1], Grant R Smith[2], Rebekah Frampton[2], Falk Kalamorz[2], Michael Currie[1], Jenna Gilkes[3], Claudia Meisrimler[1], Renwick C.J. Dobson[1,4]

[1] Biomolecular Interaction Centre, School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand.

[2] Plant and Food Research, Plant & Food Research Gerald St, Lincoln 7608, New Zealand

[3] Callaghan Innovation Protein Science and Engineering Team, University of Canterbury, Christchurch 8140, New Zealand

[4] Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Victoria 3010, Australia.

Corresponding author:

Dr Claudia Meisrimler and Prof. Renwick Dobson; Biomolecular Interaction Centre and School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand.

Email: Claudia Meisrimler <claudia.meisrimler@canterbury.ac.nz> or Renwick Dobson <renwick.dobson@canterbury.ac.nz>

Contributions:

TJ: Devised/conducted all experiments, interpreted the data, designed the images, and wrote the manuscript

GRS, RH, FK: Conceived the project and collaborated with TJ and RCJD.

MC: Trained TJ on analytical machines and helped interpret data.

JG: Trained TJ on purification machines, helped interpret data and provided comments on early manuscript draft.

CM: Provided feedback on results and commented on the early manuscript draft (advisory role).

RCJD: Conceived the project, helped with data interpretation and provided comments on early manuscript draft (advisory role).

This research was funded by the Ministry of Business, Innovation and Employment (Ngā Rākau Taketake – Myrtle Rust and Kauri Dieback Research, C09X1817)

Abstract

Myrtle rust disease, caused by the fungal rust pathogen *Austropuccinia psidii,* has been known to infect a wide range of Myrtaceae hosts. In Australia, the disease has caused the population of Myrtaceae plants to decline, and threatens to do the same to beloved New Zealand natives. Despite the large host range of *A. psidii,* many aspects about the infection process are still unknown. It is hypothesised that like many other pathogens, effector proteins play a significant role in establishing a successful infection in the host. Here we report the identification of five novel *A. psidii* effector proteins that are predicted to be essential in the establishment of myrtle rust infection. These effector proteins are the first to be studied from *A. psidii.* Structural features were identified on *A. psidii* effector candidates using bioinformatics and analytical methods. Bioinformatics show that two of these proteins are similar to fungal cellulase and xyloglucanase enzymes. One effector protein, AP1260 was successfully expressed and purified by recombinant expression in *E. coli* cells. AP1260 is a monodisperse protein as shown by analytical ultracentrifugation experiments and contains secondary structures as shown by circular dichroism. Emphasis was put on obtaining X-ray crystallography data on AP1260 to probe its biochemical function in myrtle rust infection, however further crystallography screening is required. Understanding the structure and mode of action of these effectors is crucial in order to provide insight into the establishment and progression of myrtle rust. Therefore it is crucial to further probe these effectors as they are potential targets for a disease management strategy for myrtle rust disease.

Introduction

Myrtle rust Disease

Myrtle rust disease is caused by the devastating plant pathogen *Austropuccinia psidii,* a biotrophic fungal pathogen of the Sphaerophragmiaceae family, within the order of the Pucciniales *(Carnegie and Pegg 2018; Soewarto et al. 2021). A. psidii* infects plants within the *Myrtaceae* plant family, where it impacts plant health and can lead eventually to plant death (Carnegie et al. 2016; Carnegie and Cooper 2011). Infection is characterised by yellow urediniospores that are formed on newly emerging stems, leaves, fruits and flowers. There are a few biotypes of *A. psidii* shown to establish in different regions, however, the *A. psidii* 'pandemic biotype' is currently associated with the myrtle rust disease in Oceania and Asia- pacific regions (Glen et al. 2007; Roux et al. 2013; Tobias et al. 2021). The unique ability of the *A. psidii* pandemic biotype to infect a broad range within Myrtaceae species has made myrtle rust a serious conservation and commercial issue for environments with Myrtaceae plants. Myrtle rust originated in South America and has since then spread, leading to an increased range of new host Myrtaceae species (Soewarto et al. 2021; Toome-Heller et al. 2020; Stewart et al. 2018). Myrtaceae plants have more than 5,500 species which are diverse and widespread across the globe (Soewarto et al. 2021). Since 2018, myrtle rust has spread across 33 countries, infecting 445 Myrtaceae species, with a majority of the infections identified as the pandemic biotype strain (Carnegie and Pegg 2018; Chock 2020). The pandemic biotype was detected in Australia in 2010, and due to poor management strategies and awareness, it is now widespread across a variety of Myrtaceae species, such as Eucalyptus, in Australia (Tobias et al. 2021; Carnegie and Pegg 2018). Over the last decade myrtle rust has endangered many native Myrtaceae in Australia and caused the further decline of many more species (Fensham et al. 2020). In 2017 myrtle rust was discovered in New Zealand, further increasing the range of the biotype due to the large amount of native Myrtaceae species (Toome-Heller et al. 2020). Currently, 17 native New Zealand taxa and 13 exotic taxa have been infected, including beloved plants such as pōhutukawa, mānuka, and ramarama (Toome-Heller et al. 2020).

The biotrophic lifestyle of A. psidii.

A. psidii belongs to the order *Pucciniales*; also known as "rust fungi" (Aime, Toome, and McLaughlin 2014). *Puccinia graminis f. sp. Tritici, Puccinia striiformis f. sp. tritici* and *Puccinia triticina* are all common rust fungi that are known to establish on wheat crops (Lorrain et al. 2019). Collectively, wheat rust fungi have caused an annual loss of wheat production estimated to be

$4.3-5 billion (USD) globally (Figueroa, Hammond-Kosack, and Solomon 2018). As a result of these losses, rust fungi are put under careful surveillance due to the devastating effects they can have on crops (Lorrain et al. 2019).

The high success of rust fungi as plant pathogens associates to the obligate biotrophic lifestyle and host specific infection mechanisms (Tobias et al. 2016). A biotrophic lifestyle involves infecting and colonising a living host and absorbing nutrients while maintaining host viability (Pandaranayaka et al. 2019; Nazarov et al. 2020). They form specialized structures called a haustorium, which extracts nutrients from the host and secretes virulence proteins that aid in host cell invasion (Laluk and Mengiste 2010; Pandaranayaka et al. 2019; Tobias et al. 2016). Some rust fungi utilise multiple hosts to complete their lifecycle, however, it is believed that *Puccinia Psidii* (later renamed to *A. psidii)* is an autoecious species: completing its lifecycle on one host (Glen et al. 2007; Beenken 2017). *A. psidii* goes through five different spore stages in which all of them re-infect the same plant host to achieve the next stage of its life cycle. The repeated infection cycle on single plant host causes defoliation, lower photosynthetic rates (indirectly impacting plant growth) and eventual plant host death (Chock 2020).

Myrtle rust effector proteins

Successful infection by *A. psidii* urediniospores is dependent on landing on a suitable host, formation of a haustorium and the secretion of proteins that facilitate the biotrophic lifestyle (Glen et al. 2007). The proteins secreted by the haustorium are effector proteins, which are proteins or molecules that facilitate host cell entry, alter host cell processes and structures, but eventually also trigger host cell defence mechanisms (Selin et al. 2016). Effector proteins are utilised by many pathogens across all domains of life and have shown a diverse range of structures and functions (Langin, Gouguet, and Üstün 2020; Liu et al. 2019). In fungal plant pathogens, effectors are delivered to either the apoplast or cytoplasm depending on their function (Djamei et al. 2011). Effectors that are secreted to the apoplast typically target surface receptors or other extracellular targets to facilitate pathogen entry. Cytoplasmic effectors target specific intracellular targets such as transcription factors, cellular machinery, and metabolites to begin colonising the host cell and avoid plant defence mechanisms (Selin et al. 2016; Duplessis et al. 2012). To combat this, plants have a generalized plant immunity known as pattern triggered immunity. Pattern triggered immunity uses pathogen recognition receptors, which recognise pathogenic patterns and triggers signalling and immunity responses (Selin et

al. 2016; Pruitt, Gust, and Nürnberger 2021). The pathogen, in response to the host's immune response, utilizes effector proteins that act to suppress the plant immune system (Andersen et al. 2018; Selin et al. 2016). Since effector proteins can effectively inhibit general plant immunity, the plant has a second tier immune system called effector-triggered immunity (Pruitt, Gust, and Nürnberger 2021). Effector triggered immunity allows the plant to recognize effector proteins or effector- target complexes in a pathogen-specific manner. The infection mechanisms (in the case of the pathogen) and specific immune responses (in the case of the host plant) are due to the co-evolution and life history of pathogen and host. The long evolutionary "arms race" between plant species and native pathogen species has led to greater specialization of effector proteins and host immune response (Kodaman et al. 2014). *A. psidii* has a large host range of plants that have never interacted with this pathogen before (Kodaman et al. 2014; Bufford et al. 2020). Without a specific plant immunity that has co-evolved with the pathogen, these new host plant species are losing a one-sided battle against the new alien pathogen (van Baarlen et al. 2007; Bufford et al. 2020). Effector proteins play an important part in the infection process of plant pathogens, therefore further study into *A. psidii* effector proteins would reveal the underlying pathogenic mechanisms that make it such a widespread successful pathogen (*Lorrain* et al. 2019). This would be a first step towards planning a targeted disease management strategy against myrtle rust disease.

A. psidii effector proteins candidates.

In this study we describe the discovery of five novel *A. psidii* genes that are expressed during the infection of Mānuka plants. Transcriptomic studies have been accomplished by collaborators at Plant and Food Research to identify potential effector genes involved in the establishment of *A. psidii* on Mānuka .The experiments showed that certain pathogenic genes are differential expressed over the course of infection (24 to 48 hours). This finding is consistent with other rust fungi gene expression (Lorrain et al. 2019). The following five candidate *A. psidii* genes were *shown* to express during the first 24 hours of *A. psidii* infection and are hypothesized to play an important role in the early establishing infection of Myrtle rust disease. Early analysis of APSIP001 and APSIP005 genes showed homogly to fungal enzymes. AP1260, AP5292 and AP10948 were also predicted to be effector protein candidates. All of these proteins are novel and are the first *A. psidii* proteins to be studied.

We identified the structural features of these *A. psidii* candidate proteins using bioinformatics and analytical methods. The identification of effector proteins in rust fungi and analysis of their

structural aspects has led to advances in understanding their role in rust associated diseases (Lorrain et al. 2019). Similarly, this study aims to investigate the structure of these five *A. psidii* protein candidates to better understand the mode of *A. psidii* pathogenicity. There are currently no *A. psidii* proteins submitted to the Protein Databank (PDB), which has led to a large gap in knowledge about the involvement of effector proteins in myrtle rust establishment. Obtaining X-ray crystallography structures was a major goal for this project. The high resolution structures determined by X-ray crystallography would allow further probing of the predicted active site, identification of key structural residues, and the possible prediction of the host binding targets of these effectors. Should these effector candidates be identified as important for the establishment of myrtle rust disease, they can be used as targets for a disease management strategy.

Results and Discussion

Bioinformatics show little confidence when predicting the A. psidii effector proteins but predicts APSIP001 and APSIP005 as fungal enzymes.

We began this work by comparing the effector sequences with those in various databases to provide a clue as to their role in infection. The amino acid sequences of both APSIP001 and APSIP005 were run through BLASTP. Similarities to many proteins with sequence similarity above 40 % were identified, but these were largely to "hypothetical proteins" from the *Puccinia* genus. These hypothetical proteins are all submitted from genome wide sequencing experiments of *Puccinia* organisms and experimental data on their structure or function is minimal.

Instead, to find sequence similar proteins submitted to the Protein Databank (PDB), SWISS-MODEL was used to find proteins that could be used as structural templates for APSIP001 and APSIP005 predicted models. SWISS-MODEL identified proteins with sequence similarity of >27 % to the APSIP001 amino acid sequence. The sequence similar proteins used to model APSIP001 were from cellulase and endo-β-1,4-glucanase enzymes from a variety of bacterium and fungi. Endo-beta-1,4-glucanase is cellulase class protein, involved in the internal hydrolysis of the cellulase sugar (Linton 2020). Cellulose is a major structural component in plant cell walls (Eckardt 2008). Multiple Sequence alignment (MSA) of APSIP001 with fungal and bacterial cellulase proteins (from SWISSMODEL) showed some highly conserved sequence regions to the other cellulase enzymes (Supplementary Figure 1a). However, the coverage is not complete

throughout the entire APSIP001 sequence, which produced an incomplete model of the APSIP001 enzyme by SWISS-MODEL. The predicted APSIP001 model adopts a TIM barrel like-fold, with a repeating helix-loop-strand motif and a central beta-barrel core (Vijayabaskar and Vishveshwara 2012) (Figure 1a). This fold is consistent with other cellulase enzymes submitted to the Protein Databank. Interestingly the conserved cellulase-like structure only covered the amino acids 123-465, leaving sizable amounts of amino acid residues on either side not included in the model. These regions spanning 1-123 (on the N-terminus) and 465-489 (on the C-terminus) may be disordered or involved in signaling for enzyme secretion or translocation (Sonah, Deshmukh, and Bélanger 2016).

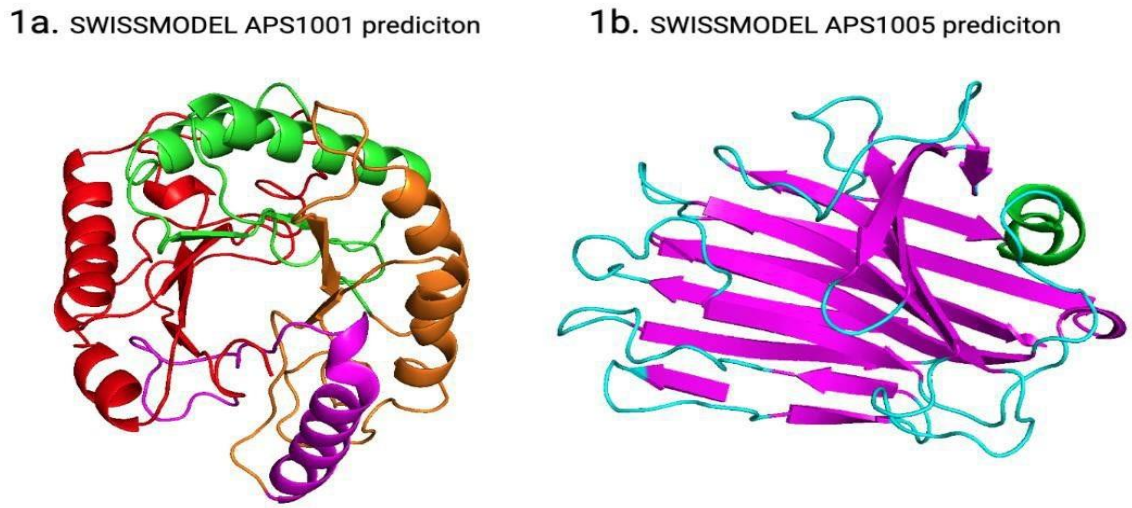


Figure. 1 | Highest GMQE model for APSIP001 and APSIP005 model generated by SWISSMODEL (modelled in Pymol). **1a)** APSIP001 predicted structural model based on the crystal structure of Cel5A (EG2) from *Hypocrea jecorina (PDB: 3QR3)*. **1b)** APSIP005 predicted model based on a Family 12 Xyloglucanase from *Aspergillus niveus* (PDB ID: 4NPR).1). Structural features are differentially coloured.

SWISS-MODEL identified proteins with sequence similarity of >26% to the APSIP001 amino acid sequence. The sequence similar proteins used to model APSIP005 were from xyloglucanase proteins and xyloglucan-specific endoglucanases. Xyloglucanase is an enzyme that facilitates the breakdown of the hemicellulose sugar xyloglucan (Rashmi and Siddalingamurthy 2018). Like cellulose, xyloglucans are also major components in plant cell walls (Eckardt 2008). MSA of APSIP005 compared with xyloglucanase proteins identified by SWISS-MODEL showed broad coverage and many conserved parts of the APSIP005 sequence to other xyloglucanases' (Supplementary Figure 1b). Unlike APSIP001, APSIP005 had excellent coverage across the entire sequence predicted by SWISS-MODEL, and a full predicted model was produced.

BLASTP and SWISSMODEL could not find any substantial sequence identity or homology to any known proteins for all three candidate effector proteins (AP1260, AP5292, AP10948). We hypothesize that this is due to the novelty of the *A. psidii* proteins. There are conserved sequence motifs that are conserved across fungal species and have been found in rust fungi effector proteins (Zhao et al. 2020). However, a limited number of *Puccinia* proteins have been characterized and even less have been structurally resolved and submitted to the PDB. Hence, it difficult to make a confident prediction on the functions and the structure using bioinformatic software relying on homology modelling. These proteins from *A. psidii* are the first to be experimentally characterized and are predicted to be highly species specific. This highlights the need for X-ray crystallography structures of *A. psidii* effector proteins.

AlphaFold2 shows secondary structures of A. psidii effector proteins.

AlphaFold2 is a recently developed protein prediction software developed by Google DeepMind artificial intelligence and is touted as been the most accurate of the available protein structure modelling tools. The predicted structures of APSIP001 and APSIP005 is very similar to the SWISS-MODEL predictions. APSIP001 is still shown as a TIM barrel structure, with very high folding confidence in this region (Figure 2a). APSIP005 retains the conserved xyloglucanase folded structure also (Figure 2b). The AlphaFold2 model for APSIP001 covered the entire sequence, in comparison to the incomplete SWISS-MODEL structure. The C-terminal and N-terminal APSIP001 disordered regions are highlighted more clearly in the AlphaFold2 model, however there is low confidence associated with the folding in these regions (Figure 2a) (Supplementary Figure 2a). AP1260, AP5292 and AP10948 structures were also predicted by AlphaFold2 (Figure 3). The lack of homology or sequence identity has led to some parts of the structures with low confidence in secondary folds (Supplementary Figure 2). AP10948 is shown as almost completely unfolded. However, AP1260 and AP5292 do show moderate confidence in some secondary folded regions predicted by AlphaFold2. These predicted structures are the first described for any *A. psidii* effector proteins.

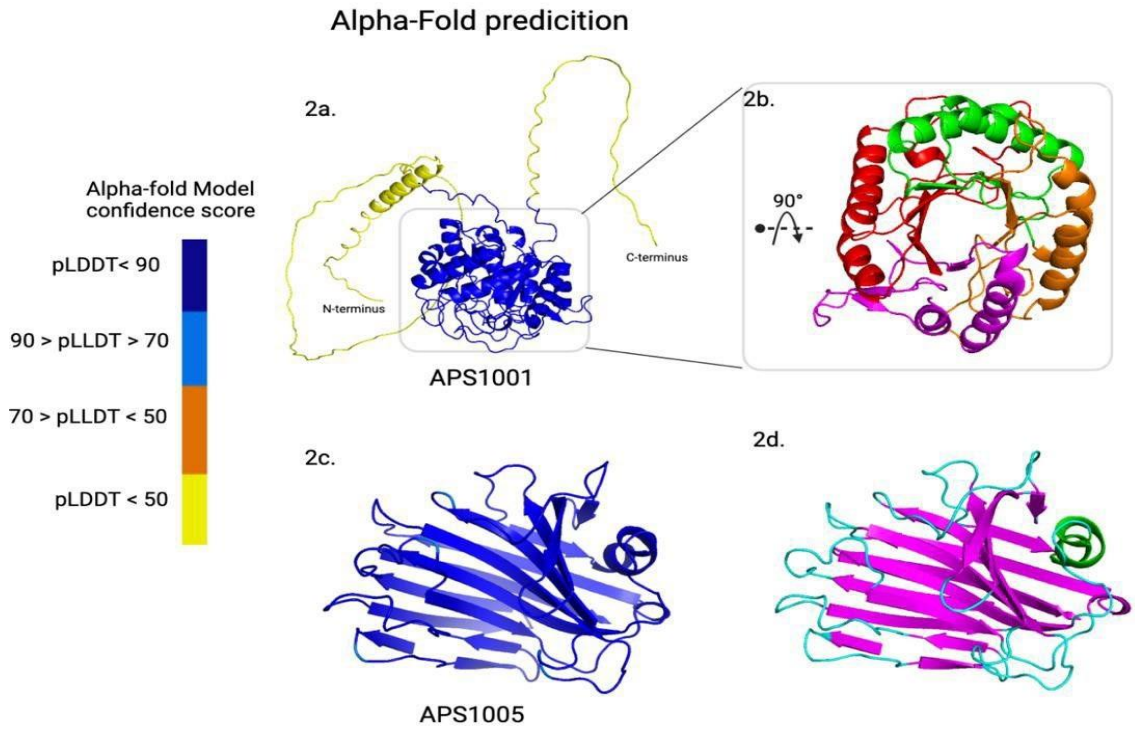


Figure 2 | AlphaFold predicted models of APSIP001 and APSIP005. The predicted local distance difference test (pLDDT) is tool used by AlphaFold2 that evalutes the distance differences between every atom in a model and gives a score (between 1-100), depending on the confidence of stereochemical plausibility(Mariani et al. 2013). Very high confidence score is shown as dark blue (pLLDT > 90), moderate confidence is shown as light blue (90 > pLLDT > 70), low confidence is shown as orange (70 > pLLDT > 50) and very low confidence is shown as yellow (pLLDT < 50). **2a)** The APSIP001 predicted AlphaFold2 model which shows the conserved TIM barrel structure confidently predicted and similar to the SWISS-MODEL prediction. The previously un-modelled N- and C-terminal disordered regions are shown in yellow (very low confidence). **2b)** The conserved TIM barrel structure is identified without disordered regions. The helix-loop-strand-helix-loop-strand motif is highlighted by the variation in colour. **2c)** The AlphaFold prediction of APSIP005. The entire conserved Xyloglucanase structure is confidently predicted and similar in structure to the SWISS-MODEL prediction. **2d)** The Helix, loops, and strands in APSIP005 are highlighted by the variation in colour.

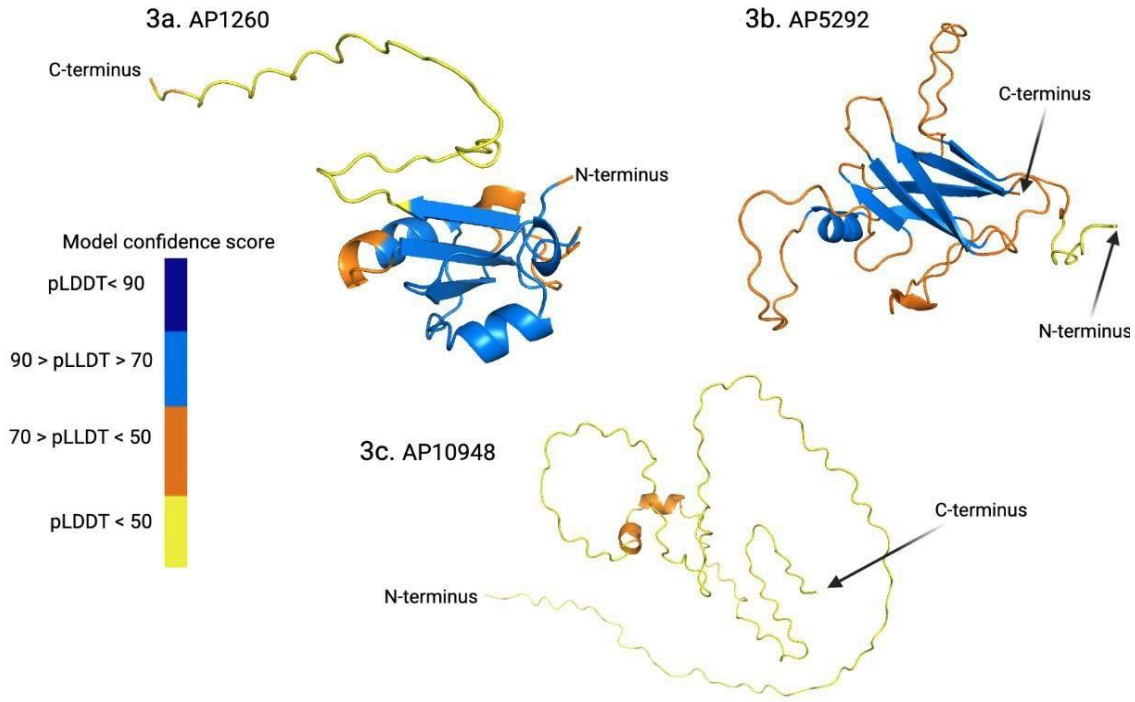


Figure 3 | The AlphaFold predicted structures of the effector proteins: AP1260, AP5292 and AP10948. Model confidence per residue between 1-100 (pLLDT) is colour coded. Very High confidence is shown as dark blue (pLLDT > 90), moderate confidence is shown as light blue (90 > pLLDT > 70), low confidence is shown as orange (70 > pLLDT > 50) and very low confidence is shown as yellow (pLLDT < 50). **3a)** AlphaFold predicted structure of the AP1260 effector protein demonstrating the model has moderate confidence in the secondary folds, which are shown in light blue. **3b)** AlphaFold predicted structure of the AP5292 effector protein. **3c)** AlphaFold predicted structure of the AP10948 effector protein. A majority of the predicted structure is shown as unfolded.

The active site of APSIP001 and APSIP005 was predicted using homology structures.

The more complete APSIP001 and APSIP005 AlphaFold2 structures were used to identify predicted active sites when modelled with a substrate using Pymol (Figure 4). APSIP001 and APSIP005 were superimposed with protein structures from PDB that had significant sequence similarity (identified using SWISS-MODEL) and crystalized with a substrate in the active site. APSIP001 was superimposed with an endo-1,4-glucanase enzyme with celltriose substrate, identified from functional metagenomics of Antarctic soil (PDB ID 5LJF) (Collet et al. 2021). APSIP001 is shown with the sugar substrate celltriose (cyan) crystalized in the predicted active site (Figure 4a). Amino acid residues Asn156, Gly146 and Asp144 were identified as direct polar contacts (within 5 Å) to the celltriose substrate (magenta). APSIP005 is shown with another cellotetrose sugar substrate (beta-D-glucopyranose-(1-4)- beta-D-glucopyranose-(1-4)-beta-D-

glucopyranose-(1-4)-beta-D-glucopyranose) (Figure 4b). APSIP005 was superimposed with an endoglucanase enzyme from *Aspergillus aculeatus* with cellotetrose substrate. There are many polar contacts sites from APSIP005 directly associated with this substrate. Arg140, Glu128, Asp106, Glu215, Trp26 and Asn24 were all identified as direct polar contacts (within 5 Å) to the cellotetrose substrate. The conclusions made about the structure and active site of APSIP001 and APSIP005 are just predictions, but they do provide a hypothesis that could be tested through mutagenesis. X-ray crystallography structures need to be obtained for valid conclusions to be made.

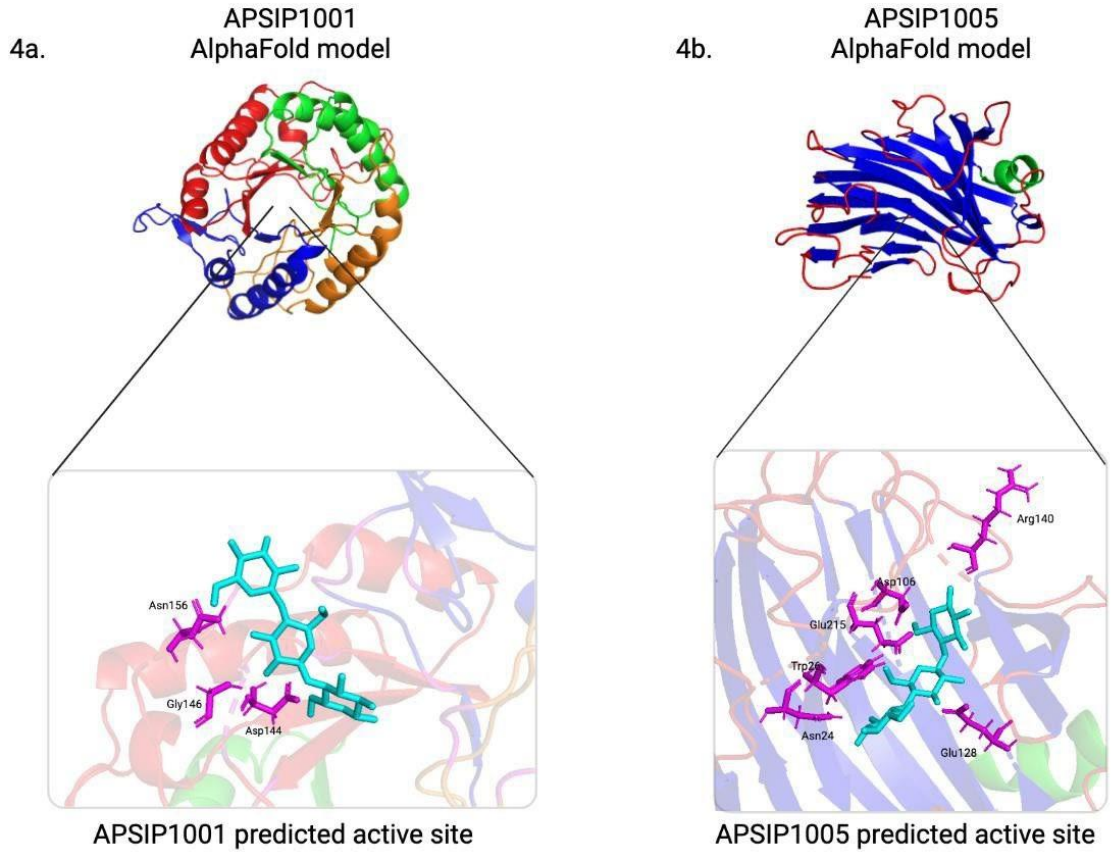


Figure 4. **|** The predicted active site of APSIP001 and APSIP005 . **4a)** The cartoon structure shows the predicted model of APS1001 as determined by AlphaFold2. The active site is enlarged and zoomed in the bottom of the diagram. The cellulase substrate celltriose is shown in cyan and the amino acid residues from APSIP001 with close polar contacts to the substrate are shown in magenta. The APSIP001 model was superimposed over the crystal structure of the endo-1,4-glucanase RBcel1 E135A with celltriose (PDB ID: 5LJF). Asn156, Gly146 and Asp144 were identified as close polar contacts (within 5 Å) to the substrate. **4b)** The top of the diagram shows the APSIP005 structure predicted by Alpha-fold. The active site is zoomed in and shown at the bottom of the diagram. The xyloglucanase substrate, cellotetrose beta-D-glucopyranose-(1-4)- beta-D-glucopyranose-(1-

4)- beta-D- glucopyranose-(1-4)- beta-D-glucopyranose, is shown in cyan and the amino acid residues from APSIP005 with close polar contacts are shown in magenta. The APSIP005 model was superimposed over the crystal structure of FI-CMCase from *Aspergillus aculeatus* F-50 in complex with cellotetrose (PDB ID: 5GM4). Arg140, Glu128, Asp106, Glu215, Trp26 and Asn24 were identified as close polar contacts (within 5 Å) to the substrate.

APSIP001 contains a predicted signal peptide and disordered region

Fungal proteins have been shown to contain a signal peptide that is needed for secretion (Sonah, Deshmukh, and Bélanger 2016). The large, disordered regions, shown in yellow, on the APSIP001 AlphaFold2 model may be a signal peptide (Figure 2a). The bioinformatics software SignalP was used to identify any signal peptides on the N- and C-terminal regions of APSIP001. However, no signal peptide was identified. In fungal proteins there is little homology among signal peptides, and therefore they can escape identification by bioinformatics software (Sonah, Deshmukh, and Bélanger 2016). However, SigCleave identified signal cleavage regions on the N-terminus of the APSIP001. Multiple protein signal cleavage sequences were identified (at residues 105-117, 92-104 and 94-106). All 3 of these cleavage regions are 13 residues long. Although, no signal peptide sequence was recognised by SignalP, the Signal cleave regions may be involved in protein processing, signalling and secretion. Furthermore, the protein disorder prediction server (PrDOS) identified a disordered region at the C-terminus (amino acids 444-489). The predicted disordered region is 45 residues long. Nevertheless, the function of the disordered region of this peptide is unknown.

***The putative A. psidii* cellulase and xyloglucanase proteins were insoluble**

Each protein was expressed in *E. coli* BL21(DE3) strain. SDS-PAGE analysis of the whole cell lysate of APSIP001 and AP1005 showed that both proteins expressed similarly to the predicted molecular weight (Table 1). APSIP001 and APSIP1005 were found to express but were insoluble (Supplementary Figure 3). It was hypothesized that this could be due to the improper folding of the protein. Therefore, to improve the folding conditions and in turn improve the solubility of the proteins, the chaperone plasmid set from TakaraBio was used. The chaperone plasmid system contains several plasmids that encode various chaperone proteins that aid in protein folding. The chaperone plasmids pGro7, pKJE7 and pTf16 were transformed into *E. coli* BL21(DE3) cells harboring the plasmid for the APSIP001 and APSIP005. The chaperone plasmids

encode the chaperone proteins GroEL, GroES, dnaK, dnaJ, grpE and tig, all of which can help with protein folding. Multiple attempts to solubilize APSIP001 and APSIP005 with chaperone proteins were unsuccessful. We hypothesis that the prokaryotic expression led to improper folding of these *A. psidii* proteins, perhaps because of a lack of necessary post-translational modifications, which also lead to insolubility. The *A. psidii* proteins are natively expressed in a fungal/eukaryotic system where the proteins are subjected to post-translational modification and codon bias, which is essential for proper folding (Khow and Suntrarachun 2012). APSIP001 and APSIP005 were expressed in a prokaryotic *E. coli* expression system, which lacks the post-translational machinery and codon bias that may be needed for production.

AP1260 was successfully expressed and purified in E. coli.

AP1260, AP5292, and AP10948 were all successfully expressed in *E. coli*. BL21(DE3) (Supplementary Figure 4). The molecular weights of each of the three effector proteins was slightly larger on the SDS-PAGE than predicted by ProtParam (Table 1). AP1260 was solubilized in PBS buffer and continued to purification by nickel affinity and size exclusion chromatography. Despite AP5292 and AP10948 expression, they were found to be insoluble. We hypothesize that the insolubility of AP5292 and AP10948 and is due to similar expression incompatibility as described for APSIP001 and APSIP005.

PROTEIN	**MOLECULAR WEIGHT (DA)**		**EXTINCTION CO-EFFICIENT**
	ProtParam	SDS page estimation	
APSIP001	52693	55000	1.125
APSIP005	25847	30000	2.173
AP1260	13212	18000	1.371
AP5292	17169	20000	2.604
AP10948	18209	20000	0.636

Table 1 | The calculated parameters of the *A. psidii* proteins by ExPASy ProtParam and their respected molecular weight observed on SDS-PAGE.

The expression of AP1260 (with 6 x His-tag) was scaled to 4L and purified using affinity purification. The two-step purification was successful and validated by SDS-PAGE (Figure 5a). The band at ~18 kDa shows AP1260 over expression (lane 2). The ~18 kDa band persists after IMAC purification using a His-trap column (lane 3). The AP1260 protein was further purified to

homogeneity using SEC (lane 4). A faint band persisted after both purification steps at ~37 kDa. We predicted that it was most likely an *E. coli* contaminant that elutes at the same time as AP1260, as opposed to a higher oligomeric species of AP1260 (Figure 4b). The size exclusion chromatography UV trace showed the AP1260 elution peak being split into twin elution peaks, suggesting that there is the presence of two different AP1260 oligomeric species (Figure 5b). If there is oligomerization, it is not likely due to cystine di- sulfide bonds, which are common for higher oligomerization states (Briggs et al. 2011; Christis, Lubsen, and Braakman 2008). AP1260 only contains two cystines, and they are predicted to form intra-molecular di-sulfide bonds by AlphaFold2 (Figure 6). It has been shown that protein oligomerization state can be concentration dependent (Bhattacharya et al. 2014; Merten, Schultz, and Klug 2012). The AP1260 concentration was higher than >20 mg/mL when loaded onto the SEC column and could have therefore weakly associated with itself during the purification.

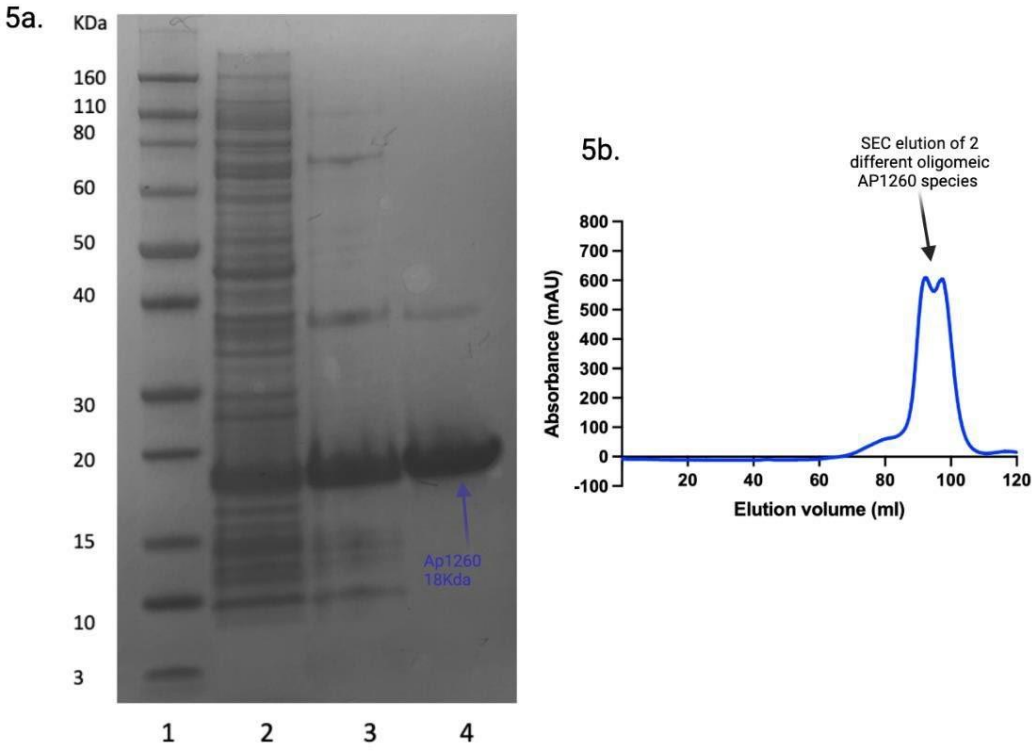


Figure 5 | 5a) SDS-gel of AP1260 protein after expression and purification steps. Lane 1 is loaded with Novex® Sharp pre-stained ladder. Lane 2 is loaded with cell-free lysate after induction with IPTG. Lane 3 is loaded with the pooled fractions after IMAC. Lane 4 is loaded with a protein fraction after SEC. **5b)** A schematic of the SEC-UV trace showing a species eluted at different elution volumes. Two species are eluted close together at ~90ml of buffer.

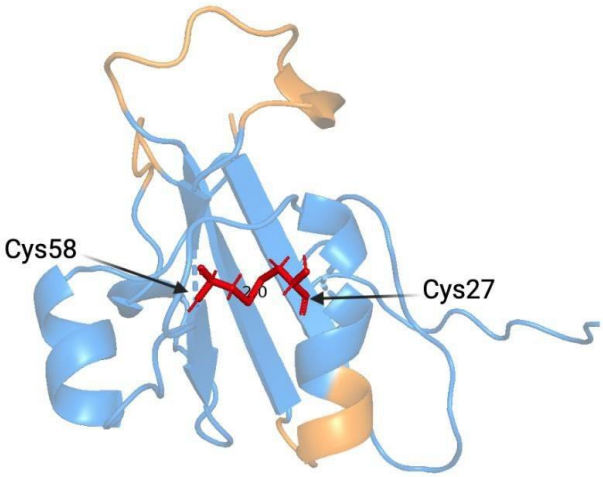


Figure 6 | The AP1260 Alpha-Fold predicted model in cartoon, with cystines (red) shown as sticks. Alpha-Fold predicted that Cys58 andCys27 would form intramolecular di-sulfide bonds.

Sedimentation velocity experiments show that AP1260 is monomeric

To determine the oligomeric state and quaternary structure of the AP1260 protein, analytical ultra-centrifugation (AUC) sedimentation velocity experiments were conducted. These experiments aimed to determine the quaternary structure of the protein and to identify any heterogeneity of the AP1260 protein in solution (as suggested by the SEC results). The AP1260 protein concentration was analyzed at 0.9 mg/mL for the sedimentation velocity experiment (as compared to 20 mg/mL for the SEC experiment). SEDFIT was used to fit the data to a continuous sedimentation model [c(s)] (Figure 7a).

Continuous sedimentations showed two peaks. The larger peak had a sedimentation value (S) of 1.6 and contained >90 % of the overall species. When fitted with a continuous mass model it showed the molecular weight of this species as 14.4 kDa (+/- 0.98) (Figure 7c). This molecular weight is consistent with the calculated monomeric size of AP1260. The smaller species had an c(s) value of 3.1 and contained <4.5 % of the overall species. When fit with the c(m) model the molecular weight of this species was 37.3 kDa (+/-2.1). This is most likely the *E. coli* contaminant identified in SDS-page analysis, which was around ~37 kDa in size.

The frictional ratio provides an estimate on the shape of the protein, typically a globular protein has a frictional ratio of 1.2. The frictional ratio obtained for AP1260 was 1.3, which suggests that the protein is not entirely globular and asymmetrical in one dimension of the protein. We can

be confident that at 0.9 mg/mL AP1260 concentration most of the protein exists as a monomer and does not frequently form high order oligomerization states. The best interpretation of the SEC and AUC data is that the protein can weakly self-associate to form high oligomers at concentrations >1 mg/mL.

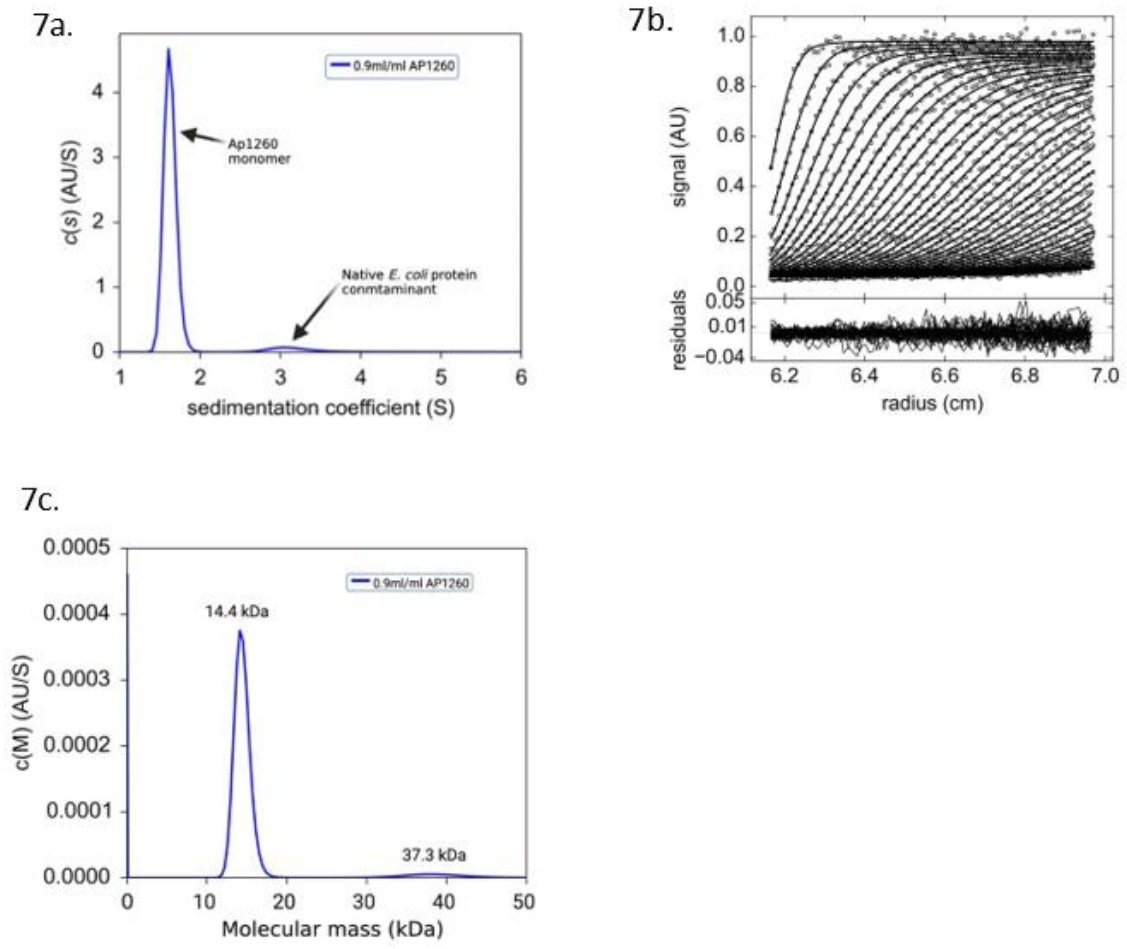


Figure 7| 7a) c(s) distribution plot for AP1260 at 0.9 mg/mL concentration. A well defined peak is shown at S=1.6 and a smaller peak at S=3.1 from an *E. coli* protein contaminant. **7b)** Residual plot of sedimentation velocity**. 7c)** c(m) distribution plot for AP1260 at 0.9 mg/mL concentration. The larger species (monomeric AP1260) has a molecular mass of 14.4 kDa and the smaller peak (*E. coli* contaminant) has a molecular mass of 37.3 kDa.

Circular Dichroism experiments showed that AP1260 has secondary folded structures.

Circular dichroism was used to identify if the AP1260 structure contained secondary structures and how this compared to the secondary folds in the AP1260 AlphaFold2 model. Circular dichroism spectra were obtained at 0.1 mg/mL AP1260 concentration (Figure 8).

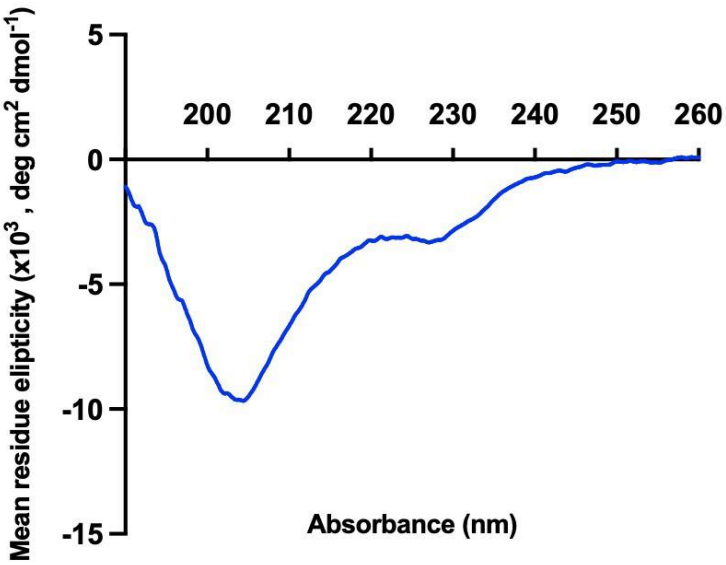


Figure 8 | The CD spectra of AP1260 at 0.1 mg/mL protein concentration.

The CD spectra showed that the AP1260 protein does contain folded secondary structures. The programme Bestsel was used to analysis the CD data and predict the secondary fold structures (Table 2) (Micsonai et al. 2018). 21.2% was interpreted as anti-parallel sheet, 5.5% as a parallel sheet fold, and 11.9% was interpreted as a helix fold and 15% as a turn. 46.5% of the structure could not be reliably determined by Bestsel (Micsonai et al. 2018). In comparison, the AP1260 AlphaFold2 model had 17.5% predicted helix fold and 15.5% predicated beta-sheet. Although CD identified some of the secondary structural characteristics there is not agreement with AlphaFold2 prediction. Both Bestsel and AlphaFold2 are prediction tools, and to reliably identify structural characteristics on the AP1260 protein structure X-ray crystallography is needed.

SECONDARY FOLD	PERCENTAGE OF OVERALL STRUCTURE (%)
REGULAR HELIX	4.5
DISTORTED HELIX	7.4
TURN	15
RIGHT TWISTED	15
RELAXED	6.2
PARALLEL	5.5
OTHER	46.4

Table 2 | The percentages of secondary structural types of the AP1260 protein by interpretation of the CD spectra using BeStSel (Micsonai et al. 2018). The secondary structure types are derived from the DSSP hydrogen bond estimation algorithm (Kabsch and Sander 1983). A regular helix is described as common correctly folded alpha helical structure. A distorted helix differs slightly from a regular helix in that the ends are typically sistroid into a non-helical bonding. A turn is described as the polypeptide chain folding back on itself or changing direction to form a more folded structure. Right twisted and relaxed β- sheets are types of anti-parallel β-sheets where all the strands run in opposite directions. This differs from Parallel β-sheets which has strands that run in the dame direction as each other. “Other” are described as uncharacterized folds or loops in the structure.

X-ray Crystallography

The first attempt at protein crystallization of AP1260 was conducted at 20 mg/mL in a Shotgun (SGN1) screen. The protein immediately precipitated, indicating that the protein concentration was too high. Crystal screen trials were attempted at 10 mg/mL in Shotgun, Morpheus II, PACT, JCSG and PGA Eco screening conditions. Successful crystals were achieved at 10 mg/mL AP1260 concentration in the crystal screen JCSG. The JCSG conditions were 0.2 M magnesium chloride hexahydrate, 0.1 M Tris, 10 % w/v PEG, pH7.0 as well as a 0.02 M calcium chloride dihydrate, 0.1 M sodium acetate, 30 % v/v MPD. Both produced crystals that were morphologically distinct from each other and at different abundances (Figure 9a and 9b). To determine if the crystals formed were protein crystals and not salt- derived crystallization, they were cryo-frozen and sent to the Australian Synchrotron. X-ray crystallization diffraction data was received from both crystal types and did not show indication of protein diffraction. We conclude that the crystal formation of both JCSG crystal screens were salt crystals (due to the lack of protein diffraction). We predict

that the protein can form crystals as suggested by the secondary folds identified by CD data. However, there may be elements of the protein are highly disordered which is consistent with the 46.5 % un-characterized folds in the circular dichroism data and disordered regions shown in the AlphaFold2 models. These disordered regions may inhibit the crystallization of AP1260 into a regularly array (Oldfield et al. 2013).

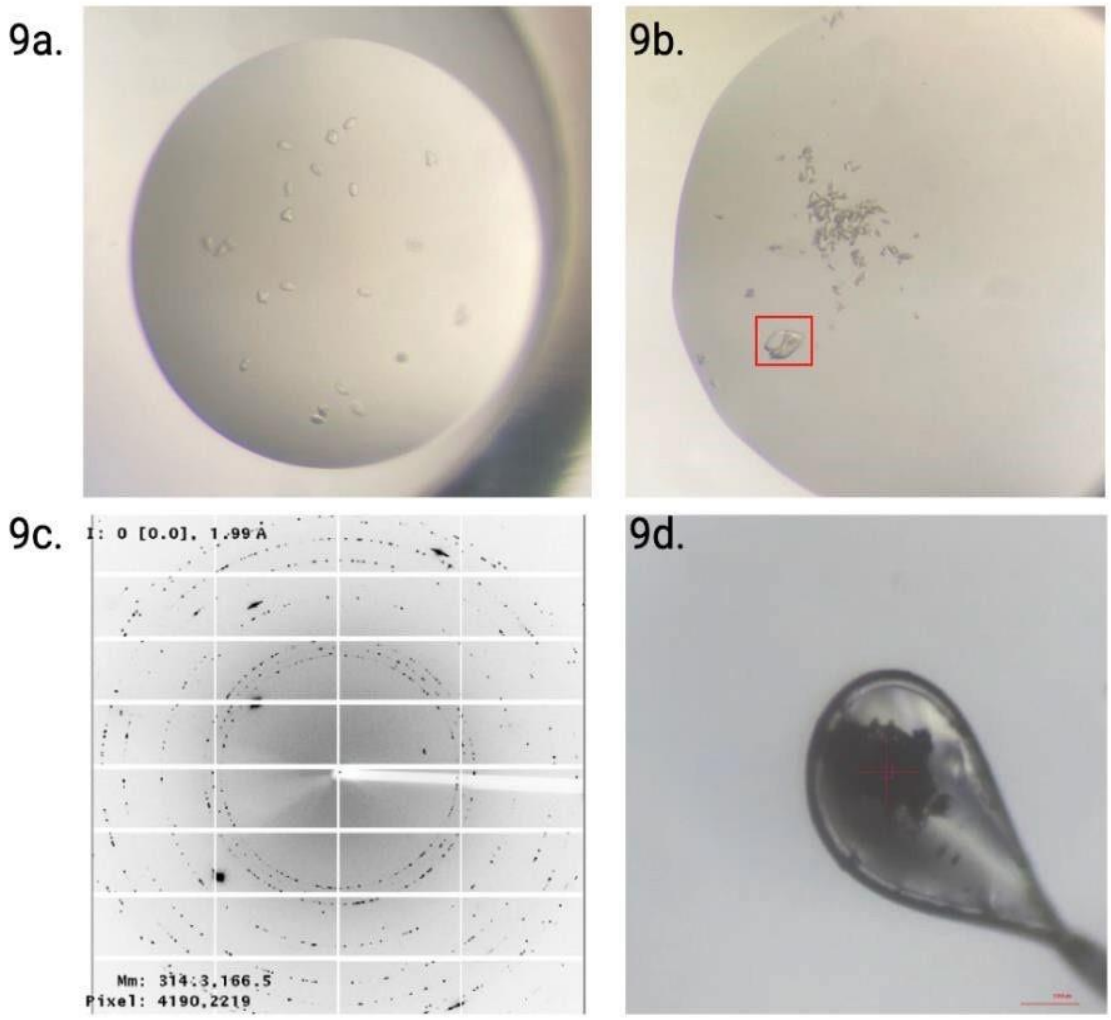


Figure 9 | 9a) The crystal screen JCSG B10 with 10 mg/mL AP1260 protein. Oval shaped crystals can be observed. **9b)** The crystal screen JCSG A4 with 10 mg/mL AP1260 protein. A mass of small "needle-like" crystals and a larger crystal (in red square) can be observed. The larger crystal was send to the Australian synchrotron for X-ray crystallography. **9c)** Diffraction data obtained for the crystal produced in the JSCG A4 condition. No observable protein diffraction pattern was observed. **9d)** The JCSG A4 crystal bound to the loop. The crystal is covered in ice from the cryo-protectant.

Conclusion

This study is the first report of structural results from five *A. psidii* proteins, putatively involved in the establishment of myrtle rust disease. Bioinformatic analysis of APSIP001 and APSIP005 were determined to have high sequence conservation for both proteins, with similarity to cellulase and xyloglucanase respectively. The putative *A. psidii* effector proteins: AP1260, AP5292 and AP10948 are much more species specific and therefore little information could be derived using bioinformatics platforms. Consequently, to further understand the structure and function of these effectors in disease, the research aimed to experimentally characterize the AP1260 effector-like protein. Results demonstrate that AP1260 exists as a monomer and has secondary folded regions as determined by analytical ultracentrifugation and circular dichroism. Furthermore, the research is the first of its kind to probe the structure of an *A. psidii* effector by X-ray crystallography.

The results of this work have begun to experimentally characterize one A. *psidii* effector protein, but further work is needed to solubilise and subsequently characterize the remaining four *A. psidii* effector protein candidates. Future work will aim to deduce the role that these effector proteins play in the establishment of myrtle rust and whether they can be targeted for a disease management strategy.

Materials and Methods

Bioinformatic analysis

A series of widely available bioinformatic platforms were used to study the DNA and amino acid sequences of all five *A. pisidii* proteins. ExPASy ProtParam (Gasteiger et al. 2005) was used calculate the molecular weight and extinction co-efficient of the five proteins. BLAST and SWISSMODEL was used to assess any homology in sequence or structure to other solved protein structures (Altschul et al. 1990; Waterhouse et al. 2018). Clustal Omega (Sievers et al. 2011) was used to align sequences that were identified from SWISS-MODEL. The multiple sequence alignment was visualized using ESPript 3 (Robert and Gouet 2014). AlphaFold2 (ColabFold) was used to predict the protein structures (Jumper et al. 2021). SignalP was used to assess any signal regions on the sequence of the proteins (Almagro Armenteros et al. 2019). SigCleave was used to find any signal cleavage regions on the proteins (von Heijne 1986). PrDOS was used to identify any disordered regions in the amino acid sequences (Ishida and Kinoshita 2007). All structural models were visualized using the modelling tool PyMOL (Schrödinger 2020).

Transformation

Each of the five A. psidii coding sequences were inserted into a pET-24a plasmid containing kanamycin resistance from the biotechnology company Genscript. A polyhistidine tag (6 x His-tag) was added the C-terminus of each coding sequence to aid in nickel affinity protein purification. Recombinant expression in the pET-24 plasmid is controlled by the T7 promoter and a l*ac* operon. Protein expression can be induced by concentrations of the synthetic inducer, isopropyl β- d-1-thiogalactopyranoside (IPTG). The cloning process of *A. psidii* genes was done by GeneScript. Each construct was transformed into competent BL21(DE3) *E. coli.*

An aliquot of BL21(DE3) was defrosted from -80 °C to room temperature. Approximately 100 ng of DNA was added to the cells and left on ice for approximately 30 minutes. To facilitate the uptake of the DNA the cells were heat shocked at 42 °C for 10 seconds, followed by incubation on ice for 5 minutes. Sterile Luria-Bertani (LB) broth was added to the cell culture and incubated at 37° C in shaking incubator (180 revolution per minute) for 1 hour and 30 minutes. The cells were concentrated and further and put onto agar plates containing kanamycin (1x concentration) to select for transformed individuals. Plates were left to grow overnight at 37 °C.

Chaperone proteins

The chaperone plasmid set was used from Takara Bio Inc. The pGro7, pKJE7, and pTf16 plasmids were used and transformed into the BL21(DE3) cells already containing the APSIP001 and APSIP005 plasmid. Transformation procedures were followed as described in the Takara Chaperone plasmid set product manual. Successful transformants were selected for using chloramphenicol and kanamycin plates (1x concentration), to ensure that both plasmids were functional after transformation.

Expression

All incubation temperatures for recombinant proteins expression were conducted at 37°C at 180 revolutions per minute unless stated otherwise. Starter cultures were made using a 10ml LB volume with kanamycin (1x concertation) and inoculated with a single colony of transformed BL21(DE3) *E. coli*. The culture was left to grow in a shaking incubator overnight. This starter culture was used to inoculate fresh sterile LB with kanamycin (1x concentration) at a 1/100 dilution. The culture was left to grow in a shaking incubator for ~3 hours till the optical density at 600nm (OD_{600}) reached 0.4-0.6. Sterile LB was used as a blank for measuring the OD_{600}. Once the OD600 was within this range, IPTG was added to a final concertation of 1mM in the culture. The culture was left to incubate at 20 °C, 180 rpm for 16 hours overnight.

For chaperone plasmid cells, the expression was done under the same conditions as previously described. However, at OD600 = 0.2 the chaperone proteins were induced with a final concentration of 0.5 mg/mL arabinose in the culture. Cultures were then left to incubate until OD600 of 0.4 – 0.6 and induced pET-24a plasmid expression with final concentration of 1 mM IPTG.

Purification

Following protein expression, the culture volume was centrifuged for 20 minutes at 4°C, 6000 x g in a Sorvall™ RC 6 Plus Centrifuge (Thermo Scientific). The cell pellet was kept frozen at -20 °C. The cell pellet was resuspended in PBS buffer (Table 3) at approximately 3 mL/g of pellet weight. Resuspended cells were lysed by sonication using an Ultra-Sonicator (Mielscher) for 15 minutes at 60 % amplitude. Resuspended cells were kept on ice during sonication. The cell free supernatant was collected and hard-spun at 12000 x g for 10 minutes in a Sorvall™ RC 6 Plus Centrifuge (Thermo Scientific).

Purification was done using a two-step purification method using immobilized metal affinity chromatography (IMAC) followed by size exclusion chromatography (SEC) as a second purification step.

BUFFER	COMPOSITION
PBS	137 mM NaCl, 2.7 mM KCl, 10mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4
TRIS	50 mM Tris (trisaminomethane), 150 mM NaCl, pH 7.4
HEPES	20 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, 150 mM NaCl, pH 7.4
SODIUM PHOSPHATE	26 mM NaH_2PO_4, 77mM Na_2HPO_4, pH 7.4

Table 3 | Solubility trial buffers and their composition

Immobilized metal affinity chromatography

Immobilized metal affinity chromatography (IMAC) was performed using a 5 mL HisTrap FF column (5mL) (Cytiva, formerly GE healthcare Life Sciences) loaded onto a ÄKTA protein purification system (GE Healthcare Life Sciences). The HisTrap column was pre-loaded with PBS buffer before loading cell free lysate. Cell free lysate was loaded onto the column at 3 mL/min. The column was washed with PBS wash buffer (PBS + 40 mM imidazole, pH 7.4) for five column volumes (~25 mL). The bound his-tagged protein was eluted with PBS elution buffer (PBS + 400 mM imidazole, pH 7.4); with a high concentration of imidazole for five column volumes. Fractions of eluted protein were collected in 1 mL volumes, identified using SDS-PAGE and were pooled together.

Size-exclusion Chromatography

Pooled fractions from IMAC were concentrated using a 10 K Pall centrifugal concentrator to a final volume that was < 2 mL. The protein sample was loaded onto a 120 mL HiLoad 16/60 Superdex column (GE Healthcare). The column, loaded with the protein sample, was eluted with 120 mL of PBS binding buffer. Absorbance at 280 nm identified protein fractions that corresponded to the protein of interest and were analysed further using SDS-PAGE. Protein fractions were further concentrated using a 10 K pall centrifugal concentrator to the concentration needed for further experiments.

SDS-PAGE

Protein samples were prepared for SDS-PAGE by taking 20 µl of protein sample and 20 µL of

SDS-loading dye containing 50 % (v/v) lithium dodecyl sulfate (LDS) 4 x buffer (Novex®), 20 % (v/v) mercaptoethanol and 30 % (v/v) MilliQ water. Samples were heated to 90 °C for 10 minutes. Protein samples were loaded into lanes in an Invitrogen Bolt™ precast polyacrylamide gel (4-12 % Bis-Tris) and submerged in 1x MES buffer. Novex® Sharp pre- stained ladder was also loaded on to the Bolt gels as a molecular weight marker. The gel was run for 22 minutes at 200 V. After running the gel, it was stained using SimplyBlue™ SafeStain (Invitrogen).

Nanodrop

Protein concentrations were quantified using a Nanodrop™ 1000 Spectrophotometer (Thermo Fisher Scientific). The extinction co-efficient for each of the candidate proteins were calculated using the ProtParam tool from ExPASy. The Beer-lambert law was used to determine the protein concertation using the extinction co-efficient.

Analytical Ultracentrifugation (AUC)

AUC experiments were performed using a Beckman Coulter XL-1 centrifugal system and Beckman Coulter AN50Ti rotor with quartz window cells. The sediment velocity experiments were performed at 50,000 rpm at 200 scans per minute. The different protein concentrations (0.1 mg/mL, 0.31 mg/mL and 0.92 mg/mL were analyzed in a PBS (pH 7.4) buffer. The radial absorbance data was assessed at the wavelength of 280 nm at 20 °C.

Using SedFit (Brown and Schuck 2006), the continuous mass [c(s)] and continuous molecular mass [c(m)] distributions were modelled at 0.9 mg/mL AP1260 protein concentration. Radial absorbance data at 0.1 mg/mL and 0.3 mg/mL did not produce plots that could be interpreted. Conclusions could still be made from 0.9 mg/mL alone. The raw radial absorbance data was refined by removing the plots with high noise ratio using SEDNTERP (Laue 1996). Buffer density and viscosities and were calculated using SEDNTERP for the PBS buffer. Final plots were produced using GUSSI (Brautigam 2015).

Circular dichroism

Proteins were analyzed using a Jasco J-815 CD spectrometer. The protein concentrations were tested at 0.1 mg/mL in a potassium phosphate buffer (0.0348 M K2HPO4, 0.0152 M KH2PO4, pH 7.4). High salt concentration buffers can increase the absorbance, which will affect the CD spectra, therefore the low salinity potassium phosphate buffer was used. Buffer exchange was done using a concentrator to swap the PBS buffer to a lower salinity potassium phosphate buffer. The wavelength was measured between 190 nm to 260 nm. Any results that exceeded a high voltage tension of 700 were removed to reduce the signal to noise ratio. The experiment was done at 20 °C with a quartz pathlength of 2 mm. The raw data was converted to into a mean residue ellipticity using the conversion factor from Equation 1 and 2. GraphPad (Prism) was used to plot the data and produce the graph.

Equation 1:

$$Mean\ residue\ concentration\ = \frac{Number\ of\ backbone\ amino\ acids}{Protein\ molecular\ weight}\ x\ [\ sample\ ](mg/mL)$$

Equation 2:

$$Conversion\ factor\ = \frac{1}{(MRC\ x\ curvette\ path\ length\ (mm)}$$

X-ray crystallography.

Ap1260 protein samples were concentrated using a Pall centrifugal concentrator to 20 mg/mL and 10 mg/mL protein concentration for crystal screening. The crystallization was performed using the sitting drop vapor diffusion method into 96 well crystallization trays. The Shotgun (SG1™) crystal screen was tested at 20 mg/mL as an initial concentration screen. The second crystal screen was tested using Shotgun (SG1™), Morpheus®, PACT premier™, JCSG-plus™, PGA™ crystallization screens at a protein concentration of 10 mg/mL The drops were layed using a Mosquito® Crystal robot at 400 nL reservoir volume and 400 nL protein sample volumes at 20 °C. Successful crystal screens that produced crystal candidates were mounted onto Cryo loops and were flash frozen in liquid nitrogen. Crystal candidates were sent to the Australia Synchrotron to be shot at the MX2 beamline. The protein crystals were mounted onto the beamline in cold nitrogen for collection of crystal diffraction data.

Acknowledgments

Many thanks to my supervisor Ren Dobson, whose constant enthusiasm and support kept me motivated throughout the year. Working with you and the Dobson lab group this year has been amazing and very rewarding.

Special thanks to my co-supervisor and mentor Claudia Meisrimler. Although we worked on very different time schedules, I very am grateful for all the help and guidance you still managed to give me.

Thanks to the team at Plant and Food Research, whose research conceived this project and allowed me to work on it. Special thanks to Grant Smith for taking the time to check-in with me and the project.

I'd like to thank Jenna Gilkes for her moral support and guidance this year. I appreciate all the time and effort you took to help me with my work.

I would like to acknowledge all the members in Dobson lab group, especially Michael Currie for his knowledge and time spent on my project. You all made me feel very welcome and help was always there when needed. Also, special thanks to the Meisrimler lab group for showing me the ropes early on and welcoming me into your group.

And finally, I would like to thank my fellow honours students, Sarah, Tessa, and Abigail. Doing both course work and a full-time project was very challenging, but it was made much easier doing it alongside you guys.

References

Aime, M. Catherine, Merje Toome, and David J. McLaughlin. 2014. '10 Pucciniomycotina.' in David J. McLaughlin and Joseph W. Spatafora (eds.), *Systematics and Evolution: Part A* (Springer Berlin Heidelberg: Berlin, Heidelberg).

Almagro Armenteros, José Juan, Konstantinos D. Tsirigos, Casper Kaae Sønderby, Thomas Nordahl Petersen, Ole Winther, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. 2019. 'SignalP 5.0 improves signal peptide predictions using deep neural networks', *Nature Biotechnology*, 37: 420-23.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. 'Basic local alignment search tool', *J Mol Biol*, 215: 403-10.

Andersen, Ethan J., Shaukat Ali, Emmanuel Byamukama, Yang Yen, and Madhav P. Nepal. 2018. 'Disease Resistance Mechanisms in Plants', *Genes*, 9: 339.

Beenken, Ludwig. 2017. 'Austropuccinia : a new genus name for the myrtle rust Puccinia psidii placed within the redefined family Sphaerophragmiaceae (Pucciniales)', *Phytotaxa*, 297: 53-61.

Bhattacharya, Arpan, Roopali Prajapati, Surajit Chatterjee, and Tushar Kanti Mukherjee. 2014. 'Concentration-Dependent Reversible Self-Oligomerization of Serum Albumins through Intermolecular β-Sheet Formation', *Langmuir*, 30: 14894-904.

Brautigam, C. A. 2015. 'Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data', *Methods in enzymology*, 562: 109-33.

Briggs, David B., Rebecca M. Giron, Pamela R. Malinowski, Martha Nuñez, and Tsu-Shuen Tsao. 2011. 'Role of redox environment on the oligomerization of higher molecular weight adiponectin', *BMC Biochemistry*, 12: 24.

Brown, Patrick H., and Peter Schuck. 2006. 'Macromolecular Size-and-Shape Distributions by Sedimentation Velocity Analytical Ultracentrifugation', *Biophysical Journal*, 90: 4651-61.

Bufford, Jennifer L., Philip E. Hulme, Benjamin A. Sikes, Jerry A. Cooper, Peter R. Johnston, and Richard P. Duncan. 2020. 'Novel interactions between alien pathogens and native plants increase plant–pathogen network connectance and decrease specialization', *Journal of Ecology*, 108: 750-60.

Carnegie, Angus J., and Kevin Cooper. 2011. 'Emergency response to the incursion of an exotic myrtaceous rust in Australia', *Australasian Plant Pathology*, 40: 346.

Carnegie, Angus J., Amrit Kathuria, Geoff S. Pegg, Peter Entwistle, Matthew Nagel, and Fiona R. Giblin. 2016. 'Impact of the invasive rust Puccinia psidii (myrtle rust) on native Myrtaceae in natural ecosystems in Australia', *Biological Invasions*, 18: 127-44.

Carnegie, Angus J., and Geoff S. Pegg. 2018. 'Lessons from the Incursion of Myrtle Rust in Australia', *Annual Review of Phytopathology*, 56: 457-78.

Chock, M. K. 2020. 'The global threat of Myrtle rust (AUSTROPUCCINIA psidii): Future prospects for control and breeding resistance in susceptible hosts', *Crop Protection*, 136: 105176.

Christis, Chantal, Nicolette H. Lubsen, and Ineke Braakman. 2008. 'Protein folding includes oligomerization – examples from the endoplasmic reticulum and cytosol', *The FEBS Journal*, 275: 4700-27.

Collet, L., C. Vander Wauven, Y. Oudjama, M. Galleni, and R. Dutoit. 2021. 'Glycoside hydrolase family 5: structural snapshots highlighting the involvement of two conserved residues in catalysis', *Acta Crystallogr D Struct Biol*, 77: 205-16.

Djamei, Armin, Kerstin Schipper, Franziska Rabe, Anupama Ghosh, Volker Vincon, Jörg Kahnt, Sonia Osorio, Takayuki Tohge, Alisdair R Fernie, and Ivo Feussner. 2011. 'Metabolic priming by a secreted fungal effector', *Nature*, 478: 395-98.

Duplessis, S, DL Joly, PN Dodds, F Martin, and S Kamoun. 2012. "Effectors in plant-microbe interactions." In.: John Wiley & Sons.

Eckardt, Nancy A. 2008. 'Role of xyloglucan in primary cell walls', *The Plant cell*, 20: 1421-22.

Fensham, Roderick J., Angus J. Carnegie, Boris Laffineur, Robert O. Makinson, Geoff S. Pegg, and Jarrah Wills. 2020. 'Imminent Extinction of Australian Myrtaceae by Fungal Disease', *Trends in Ecology & Evolution*, 35: 554-57.

Figueroa, Melania, Kim E Hammond-Kosack, and Peter S Solomon. 2018. 'A review of wheat diseases—a field perspective', *Molecular plant pathology*, 19: 1523-36.

Gasteiger, Elisabeth, Christine Hoogland, Alexandre Gattiker, S'everine Duvaud, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch. 2005. 'Protein Identification and Analysis Tools on the ExPASy Server.' in John M. Walker (ed.), *The Proteomics Protocols Handbook* (Humana Press: Totowa, NJ).

Glen, M., A. C. Alfenas, E. A. V. Zauza, M. J. Wingfield, and C. Mohammed. 2007. 'Puccinia psidii: a threat to the Australian environment and economy —a review', *Australasian Plant Pathology*, 36: 1-16.

Ishida, T., and K. Kinoshita. 2007. 'PrDOS: prediction of disordered protein regions from amino acid sequence', *Nucleic Acids Res*, 35: W460-4.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. 2021. 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596: 583-89.

Kabsch, Wolfgang, and Christian Sander. 1983. 'Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features', *Biopolymers*, 22: 2577-637.

Khow, Orawan, and Sunutcha Suntrarachun. 2012. 'Strategies for production of active eukaryotic proteins in bacterial expression system', *Asian Pacific journal of tropical biomedicine*, 2: 159-62.

Kodaman, Nuri, Rafal S. Sobota, Robertino Mera, Barbara G. Schneider, and Scott M. Williams. 2014. 'Disrupted human–pathogen co-evolution: a model for disease', *Frontiers in Genetics*, 5.

Laluk, Kristin, and Tesfaye Mengiste. 2010. 'Necrotroph attacks on plants: wanton destruction or covert extortion?', *The arabidopsis book*, 8: e0136-e36.

Langin, Gautier, Paul Gouguet, and Suayib Üstün. 2020. 'Microbial Effector Proteins – A Journey through the Proteolytic Landscape', *Trends in Microbiology*, 28: 523-35.

Laue, Thomas M. 1996. 'Analytical Ultracentrifugation', *Current Protocols in Protein Science*, 4: 7.5.1-7.5.9.

Linton, S. M. 2020. 'Review: The structure and function of cellulase (endo-β-1,4-glucanase) and hemicellulase (β-1,3-glucanase and endo-β-1,4-mannase) enzymes in invertebrates that consume materials ranging from microbes, algae to leaf litter', *Comp Biochem Physiol B Biochem Mol Biol*, 240: 110354.

Liu, Liping, Le Xu, Qie Jia, Rui Pan, Ralf Oelmüller, Wenying Zhang, and Chu Wu. 2019. 'Arms race: diverse effector proteins with conserved motifs', *Plant Signaling & Behavior*, 14: 1557008.

Lorrain, Cécile, Karen Cristine Gonçalves dos Santos, Hugo Germain, Arnaud Hecker, and Sébastien Duplessis. 2019. 'Advances in understanding obligate biotrophy in rust fungi', *New Phytologist*, 222: 1190-206.

Mariani, Valerio, Marco Biasini, Alessandro Barbato, and Torsten Schwede. 2013. 'lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests', *Bioinformatics*, 29: 2722-28.

Merten, Jacqueline A., Kathryn M. Schultz, and Candice S. Klug. 2012. 'Concentration-dependent oligomerization and oligomeric arrangement of LptA', *Protein science : a publication of the Protein Society*, 21: 211-18.

Micsonai, András, Frank Wien, Éva Bulyáki, Judit Kun, Éva Moussong, Young-Ho Lee, Yuji Goto, Matthieu Réfrégiers, and József Kardos. 2018. 'BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra', *Nucleic Acids Research*, 46: W315-W22.

Nazarov, P. A., D. N. Baleev, M. I. Ivanova, L. M. Sokolova, and M. V. Karakozova. 2020. 'Infectious Plant Diseases: Etiology, Current Status, Problems and Prospects in Plant Protection', *Acta naturae*, 12: 46-59.

Oldfield, Christopher J., Bin Xue, Ya-Yue Van, Eldon L. Ulrich, John L. Markley, A. Keith Dunker, and Vladimir N. Uversky. 2013. 'Utilization of protein intrinsic disorder knowledge in structural proteomics', *Biochimica et biophysica acta*, 1834: 487-98.

Pandaranayaka, Eswari P. J., Omer Frenkel, Yigal Elad, Dov Prusky, and Arye Harel. 2019. 'Network analysis exposes core functions in major lifestyles of fungal and oomycete plant pathogens', *BMC Genomics*, 20: 1020.

Pruitt, Rory N., Andrea A. Gust, and Thorsten Nürnberger. 2021. 'Plant immunity unified', *Nature Plants*, 7: 382-83.

Rashmi, R., and K. R. Siddalingamurthy. 2018. 'Microbial xyloglucanases: a comprehensive review', *Biocatalysis and Biotransformation*, 36: 280-95.

Robert, Xavier, and Patrice Gouet. 2014. 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic Acids Research*, 42: W320-W24.

Roux, Jolanda, Izette Greyling, Teresa A. Coutinho, Marcel Verleur, and Michael J. Wingfeld. 2013. 'The Myrtle rust pathogen, Puccinia psidii, discovered in Africa', *IMA Fungus*, 4: 155-59.

Schrödinger, L., & DeLano, W. . 2020. ' PyMOL'.

Selin, Carrie, Teresa R. de Kievit, Mark F. Belmonte, and W. G. Dilantha Fernando. 2016. 'Elucidating the Role of Effectors in Plant-Fungal Interactions: Progress and Challenges', *Frontiers in Microbiology*, 7.

Sievers, Fabian, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D Thompson, and Desmond G Higgins. 2011. 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular Systems Biology*, 7: 539.

Soewarto, Julia, Chanatda Somchit, Esna du Plessis, Irene Barnes, Ginna M. Granados, Michael J. Wingfield, Louise Shuey, Michael Bartlett, Stuart Fraser, Peter Scott, Elizabeth Miller, Nick Waipara, Roanne Sutherland, and Beccy Ganley. 2021. 'Susceptibility of native New Zealand Myrtaceae to the South African strain of Austropuccinia psidii: A biosecurity threat', *Plant Pathology*, 70: 667-75.

Sonah, Humira, Rupesh K. Deshmukh, and Richard R. Bélanger. 2016. 'Computational Prediction of Effector Proteins in Fungi: Opportunities and Challenges', *Frontiers in plant science*, 7.

Stewart, J. E., A. L. Ross-Davis, R. N. Graça, A. C. Alfenas, T. L. Peever, J. W. Hanna, J. Y. Uchida, R. D. Hauff, C. Y. Kadooka, M.-S. Kim, P. G. Cannon, S. Namba, S. Simeto, C. Pérez, M. B. Rayamajhi, D. J. Lodge, M. Arguedas, R. Medel-Ortiz, M. A. López- Ramirez, P. Tennant, M. Glen, P. S. Machado, A. R. McTaggart, A. J. Carnegie, and N. Klopfenstein. 2018. 'Genetic diversity of the myrtle rust pathogen (Austropuccinia psidii) in the Americas and Hawaii: Global implications for invasive threat assessments', *Forest Pathology*, 48: e12378.

Tobias, Peri A., David I. Guest, Carsten Külheim, Ji-Fan Hsieh, and Robert F. Park. 2016. 'A curious case of resistance to a new encounter pathogen: myrtle rust in Australia', *Molecular plant pathology*, 17: 783-88.

Tobias, Peri A., Benjamin Schwessinger, Cecilia H. Deng, Chen Wu, Chongmei Dong, Jana Sperschneider, Ashley Jones, Zhenyan Lou, Peng Zhang, Karanjeet Sandhu, Grant R. Smith, Josquin Tibbits, David Chagné, and Robert F. Park. 2021. 'Austropuccinia psidii, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements', *G3 Genes|Genomes|Genetics*, 11.

Toome-Heller, M., W. W. H. Ho, R. J. Ganley, C. E. A. Elliott, B. Quinn, H. G. Pearson, and B. J. R. Alexander. 2020. 'Chasing myrtle rust in New Zealand: host range and distribution over the first year after invasion', *Australasian Plant Pathology*, 49: 221-30.

van Baarlen, Peter, Alex van Belkum, Richard C. Summerbell, Pedro W. Crous, and Bart P. H. J. Thomma. 2007. 'Molecular mechanisms of pathogenicity: how do pathogenic microorganisms develop cross-kingdom host jumps?', *FEMS Microbiology Reviews*, 31: 239-77.

Vijayabaskar, M. S., and Saraswathi Vishveshwara. 2012. 'Insights into the Fold Organization of TIM Barrel from Interaction Energy Based Structure Networks', *PLOS Computational Biology*, 8: e1002505.

von Heijne, G. 1986. 'A new method for predicting signal sequence cleavage sites', *Nucleic Acids Research*, 14: 4683-90.

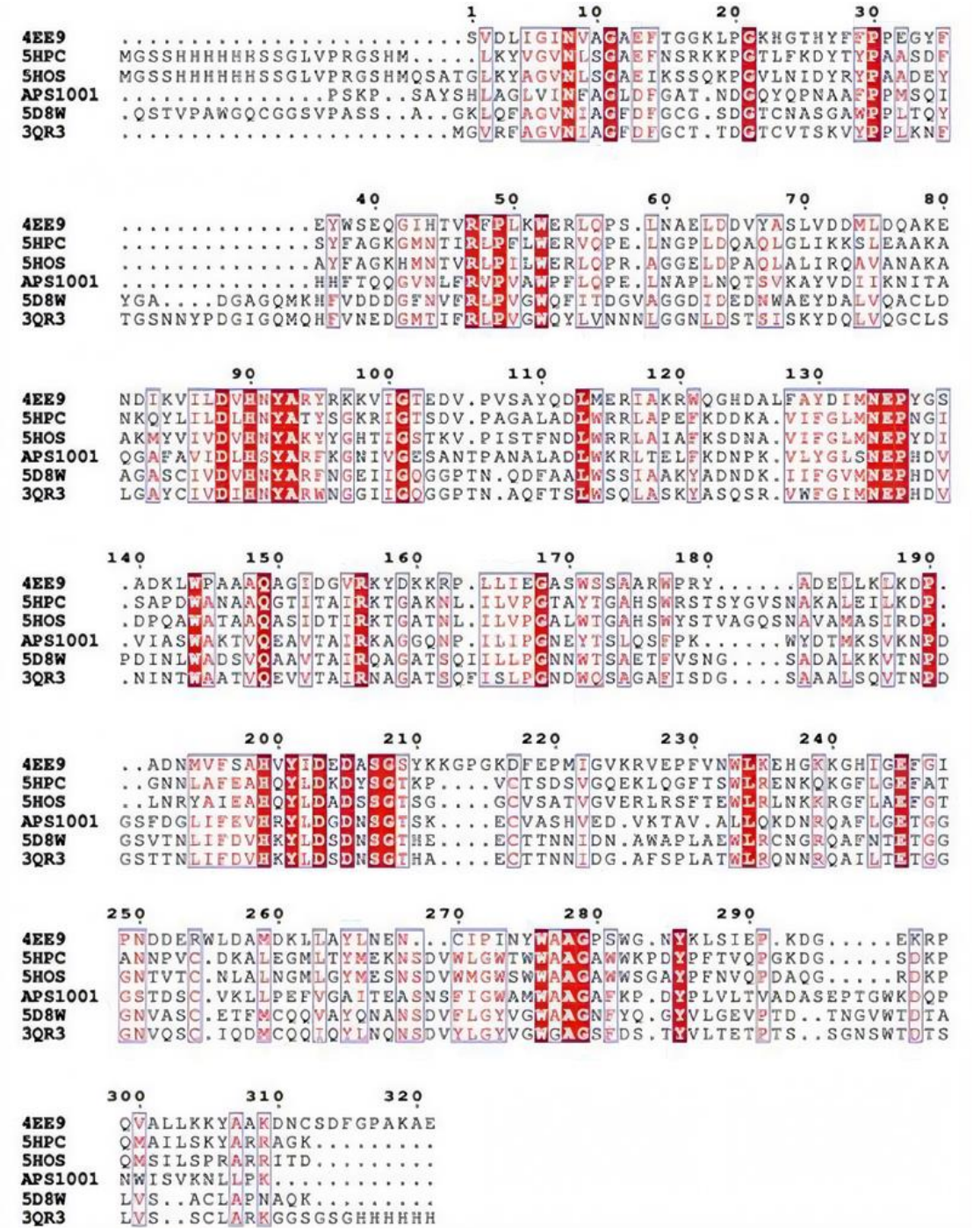
Waterhouse, Andrew, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, Florian T Heer, Tjaart A P de Beer, Christine Rempfer, Lorenza Bordoli, Rosalba Lepore, and Torsten Schwede. 2018. 'SWISS-MODEL: homology modelling of protein structures and complexes', *Nucleic Acids Research*, 46: W296- W303.

Zhao, Shuqing, Xiaofeng Shang, Weishuai Bi, Xiumei Yu, Daqun Liu, Zhensheng Kang, Xiaojie Wang, and Xiaodong Wang. 2020. 'Genome-Wide Identification of Effector Candidates With Conserved Motifs From the Wheat Leaf Rust Fungus Puccinia triticina', *Frontiers in Microbiology*, 11.

Supplementary Information

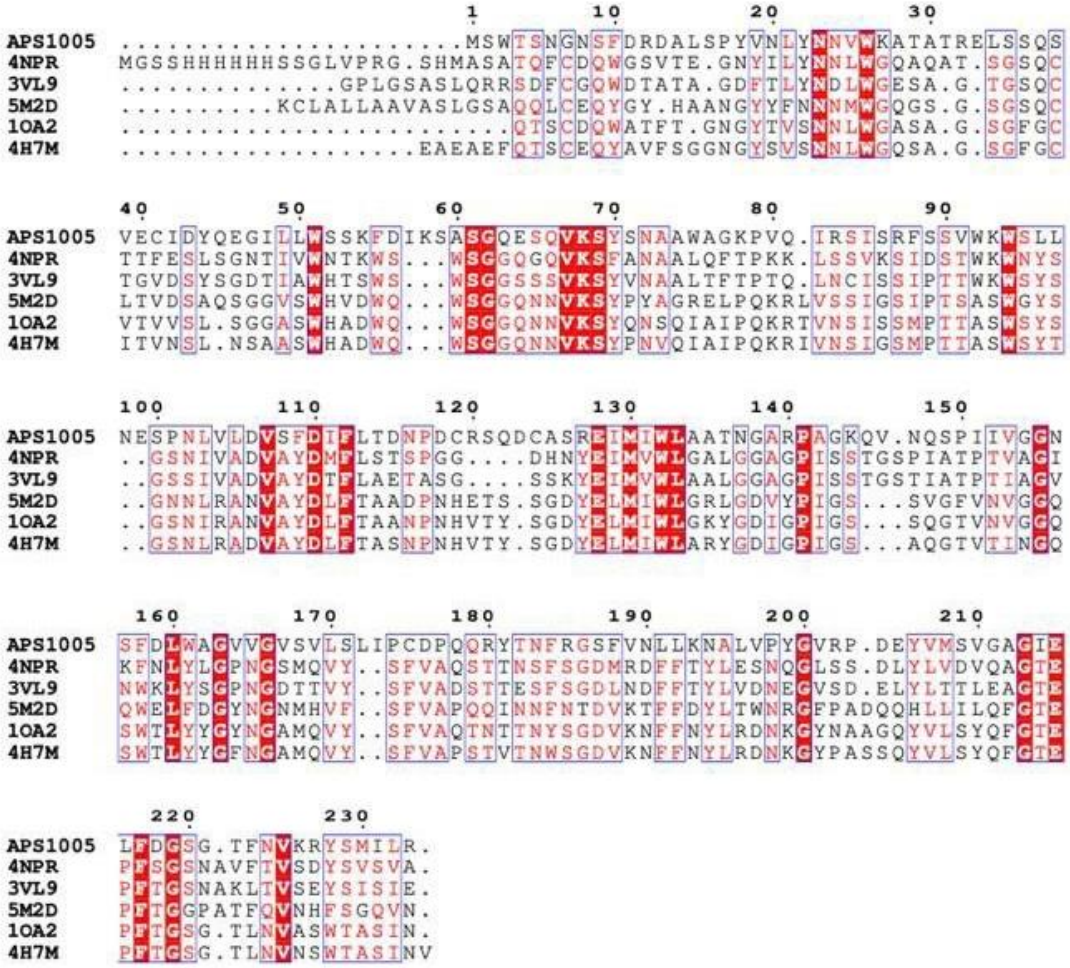
Supplementary Figure 1 | Multiple Sequence alignment of APSI001 and APSI005 similar proteins. Red lettering indicates group similarity, red highlighted box indicates sequence identical ersidues , and a blue outline box indicates global similarity.

S1a)



S1a) Alignment of APSI001 with sequence similar proteins: PDB ID 4EE9 form an uncultured bacterium (27.3 % identity), PDB ID 5HPC from *Xanthomonas citri* pv. Citri (33.55 % identity), PDB ID 5HOS from *Xanthomonas axonopodis* pv. *citri* (29.2 % identify), PDB ID 5D8W from *Ganoderma lucidum* (38.03 % identity), and PDB ID 3QR3 from *Trichoderma reesei* (36.7 % identity).

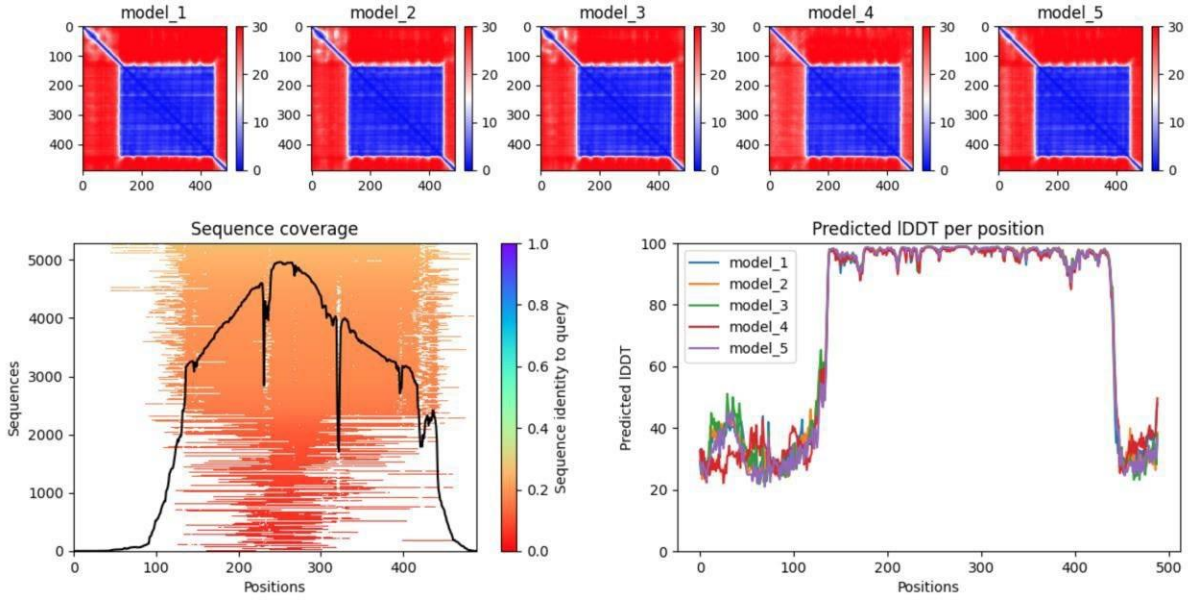
S1b)



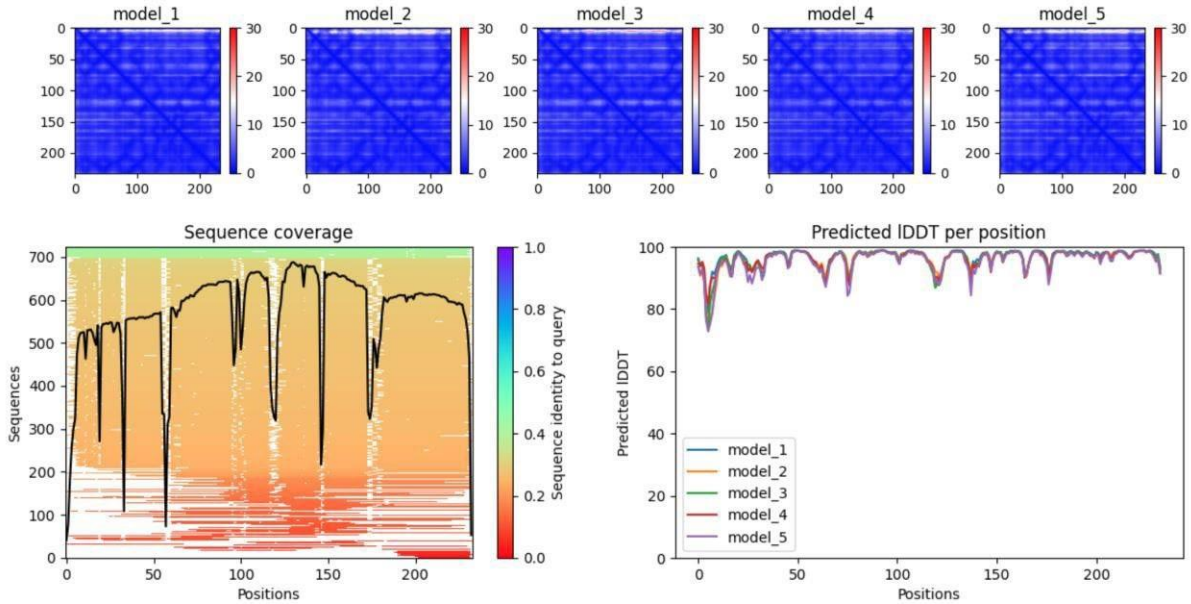
S1b) Alignment of APSI005 with sequence similar proteins: PDB ID 4NPR from *Aspergillus niveus* (31.1 % identity), PDB ID 3VL9 from *Aspergillus aculeatus (*35.7 % identity), PDB ID 5M2D from *Acremonium chrysogenum* (31.2 % identity), PDB ID 1OA2 from *Trichoderma reesei* (31.1 % identity), and PDB ID 4H7M from *Trichoderma harzianum* (26.64 % *identity*).

Supplementary Figure 2 | Sequence **coverage** and predicted local distance difference test (pLDDT) per residue for each of the five *A. psidii* AlphaFold2 predicted models **S2a)** APSIP001, **S2b)** APSIP005, **S2c)** AP1260, **S2d)** AP5292, **S2e)** AP10948

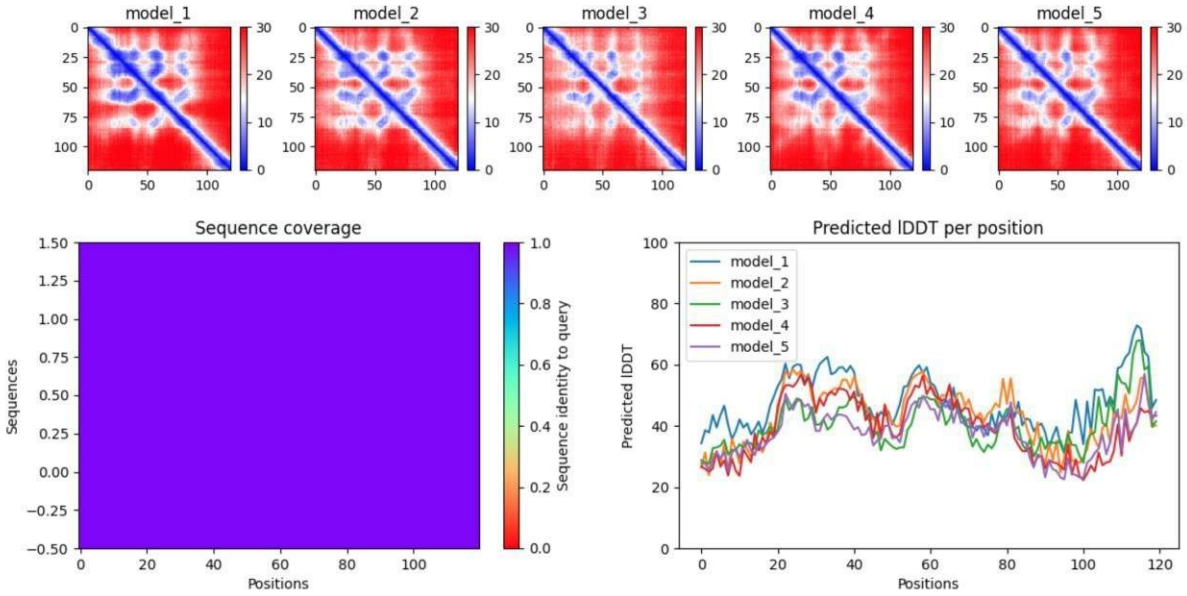
S2a. APSIP001



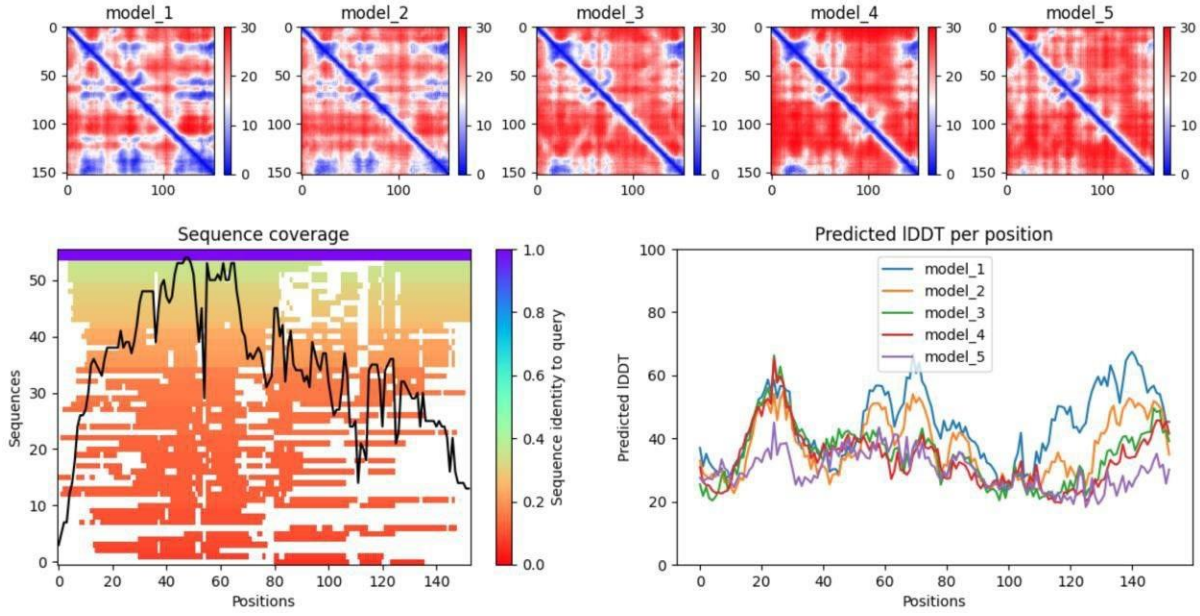
S2b. APSIP005



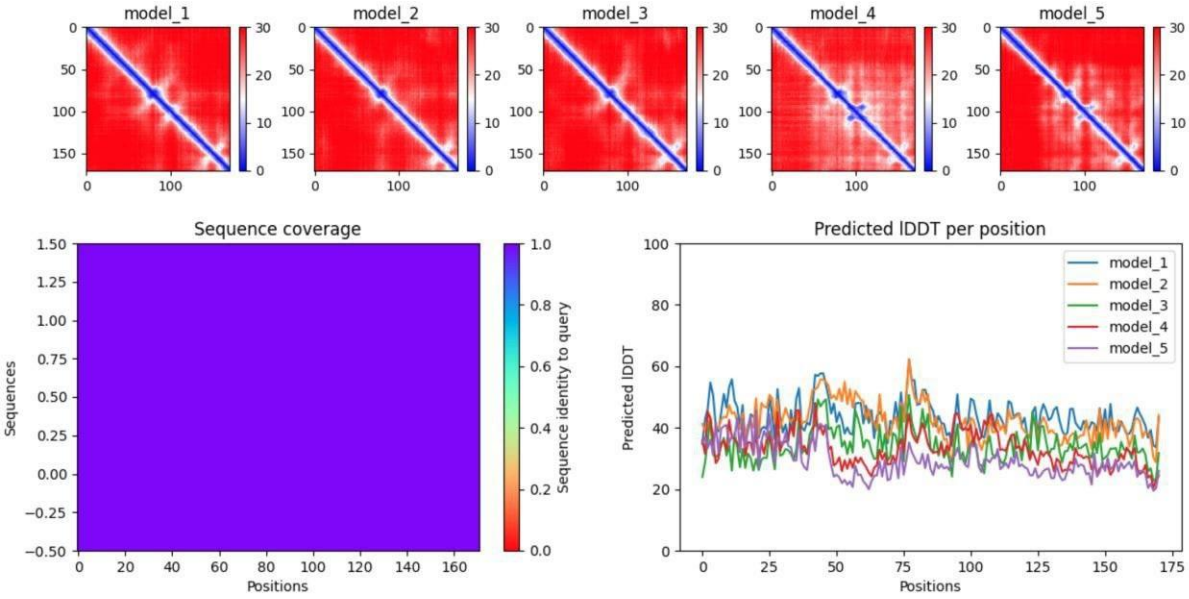
S2c. AP1260



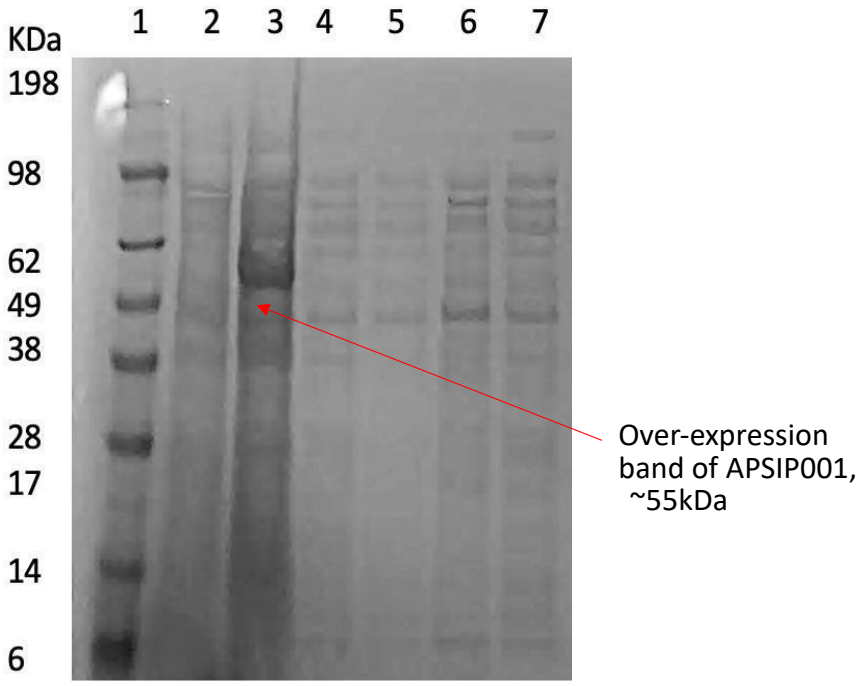
S2d. AP5292



S2e. AP10948


model_1
model_2
model_3
model_4
model_5
Sequence coverage
Sequences
Positions
Sequence identity to query
Predicted lDDT per position
Predicted lDDT
Positions
model_1
model_2
model_3
model_4
model_5

Supplementary Figure 3 | SDS-gel of APSI001 solubility trails. Lane 1 is SeeBlue pre-stained protein standard. Lane 2 shows un-induced APSIP001 expression. Lane 3 shows induced APSP001 expression. The expression of APSIP001 can be seen as around 55kDa. Lane 4 in solubility is Tris buffer. Lane 5 is solubility in PBS buffer. Lane 6 is solubility in HEPES buffer. Lane 7 in solubility in sodium phosphate buffer. Lanes 4-7 shows no solubility of APSIP001 protein.



Supplementary Figure 4| Lane 1: Novex® Sharp pre-stained ladder. Lane 2: Un-induced AP1260. Lane 3: Whole cell lysate AP1260. Lane 4: Soluble AP1260 PBS. Lane 5: un-induced Ap5292. Lane 6: Whole cell lysate AP5292. Lane 7: Soluble AP5292 in PBS. Lane 8: Empty. Lane 9: Whole cell lysate AP10948, Lane 10. Soluble AP10948 in PBS. AP1260 showed soluble expression in lane 4 around ~18 kDa. AP5292 showed expression (~19 kDa) but the degree of solubility is unclear (lane 6). AP10948 showed expression in lane 9 (~19 KDa), but no observable expression in lane 10. The SDS-Gel was run in MOPS buffer (x1) at the wrong voltage, which explains the bunching of the bands towards the end of the gel.