

12-2018

Identification and functional characterization of effectors from an anther smut fungus, *Microbotryum lychnidis-dioicae*.

Venkata Swathi Kuppireddy
University of Louisville

Follow this and additional works at: <https://ir.library.louisville.edu/etd>

 Part of the [Molecular Genetics Commons](#)

Recommended Citation

Kuppireddy, Venkata Swathi, "Identification and functional characterization of effectors from an anther smut fungus, *Microbotryum lychnidis-dioicae*." (2018). *Electronic Theses and Dissertations*. Paper 3078.
<https://doi.org/10.18297/etd/3078>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

IDENTIFICATION AND FUNCTIONAL CHARACTERIZATION OF EFFECTORS
FROM AN ANTHR SMUT FUNGUS, *MICROBOTRYUM LYCHNIDIS-DIOICAE*

By

Venkata Swathi Kuppireddy
M.Sc., Biology, University of Louisville, 2016

A Dissertation
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville in
Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Biology

Department of Biology,
Division of Molecular, Cellular and Developmental Biology,
University of Louisville
Louisville, Kentucky

December 2018

Copyright 2018 by Venkata Swathi Kuppireddy

© All rights reserved

IDENTIFICATION AND FUNCTIONAL CHARACTERIZATION OF EFFECTORS
FROM AN ANTHR SMUT FUNGUS, *MICROBOTRYUM LYCHNIDIS-DIOICAE*

By

Venkata Swathi Kuppireddy
M.Sc., Biology, University of Louisville, 2016

A Dissertation Approved on

November 14, 2018

By the following Dissertation Committee:

Dr. Michael H. Perlin, Principal Advisor

Dr. David Schultz

Dr. Brian Wattenberg

Dr. Carolyn Klinge

Dr. Paul Himes

DEDICATION

This work is dedicated to my beloved family

My parents, brother

&

my loving husband

Thank you for helping to give me the life that I love today

ACKNOWLEDGMENTS

I am highly indebted to my advisor, Dr. Michael Perlin for accepting me into his lab and allowing me to learn all that I learned today. He is extremely patient in answering all the questions and being very supportive in this incredible journey.

I thank my committee members Dr. David Schultz, Dr. Brian Wattenberg, Dr. Carolyn Klinge and Dr. Paul Himes for their valuable comments and inputs in directing my research and encouraging me towards my goal.

I also thank Dr. Michelle Barati for her patience and for all the hours that she sat with me for confocal microscopy assistance and western blots and Dr. Liang Bao for technical support with the SEM. I am grateful to the suggestions of Dr. Ronald Atlas for helping me to develop my teaching abilities, Dr. Mark Running for his valuable inputs and suggestions in my research, and Dr. Gary Cobbs for his help with statistical analysis.

I want to extend my heartfelt gratitude to all the undergraduates Brittany Carman, Catarina Cahill, Laura Irons, Matthew Carruthers, Joseph Paul Ham, Ankur Patel, Patrick Soucy and Medha Desai for their incredible support. I also want to thank my close comrades Parul Singh, Richa Singhal, Sanaya Bamji, Ming Chang Tsai, Joseph Paul Ham, Sunitha Khanal, Lalu Vijaya Krishna Pillai for all the smiles that made this journey less stressful. I would like to offer special thanks to my dear friend Susan Toh, who, although no longer with us, continues to remain as an inspiration, forever. Finally, I like

to thank fellow Goat lab members William Beckerson, Margaret Wallen and Hector Mendoza, Roxanne Leiter for their companionship throughout these years.

I like to appreciate and thank the support of the Biology department Ms. Doris Meadows, Judi Felli, Terri Norris, and Charice Johnson and the Professors Dr. Ronald Fell and Dr. Perri Eason in supporting my graduate career.

Finally, I would like to take this opportunity to express my gratitude to my parents, brother, my husband and my daughter for their unconditional love and support in this arduous journey.

ABSTRACT

IDENTIFICATION AND FUNCTIONAL CHARACTERIZATION OF EFFECTORS FROM AN ANTHRACNOSIS SMUT FUNGUS, *MICROBOTRYUM LYCHNIDIS-DIOICAE*

Venkata Swathi Kuppireddy

November 14, 2018

Microbotryum lychnidis-dioicae causes anther smut fungus in its host plant, *Silene latifolia*. The goal of this work is to identify and characterize the virulence determinants of this pathogen to better understand the molecular basis behind this host-pathogen interaction. This work studied for the first time the key effectors in the mechanism of infection by this fungal species. Using, yeast two-hybrid screens, I have identified the host plant interaction partners for the effector, MVLG_01732. A second effector MVLG_05720, interacts with other fungal proteins that appear to facilitate the fungal establishment and colonization during the infection. Our findings indicate that a third effector, MVLG_04106, could serve as a transcriptional regulator to promote infection. To further characterize the role of the effector, MVLG_01732, I have conducted heterologous expression studies in *A. thaliana* followed by infection assays with the pathogen *Pseudomonas syringae*. Our results indicate that this effector has a role in the early bolting of flowers in *A. thaliana*, this finding provides an important clue about the

role of MVLG_01732 in advancing host plant floral development, which is known to occur in infected *S. latifolia* flowers. Here, we have identified a key molecular link between a fungal effector and the developmental change it triggers in the host plant. Infection assays reveal that this effector might play a role in promoting pathogen growth.

I have also examined the response to this effector by the host by expressing the His-tagged effector in *S. latifolia* in an experiment designed to model the mechanism of infection in the native habitat. This is the first time that this approach of delivering the candidate effector protein has been carried out *in planta* and aimed to provide information about the previously unidentified interacting partners from the host. Overall, this dissertation body aimed to increase the number of available genetic tools to study *M. lychnidis-dioicae* and will serve as a valuable resource for future investigators along with furthering our understanding of the infection mechanism.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
ABSTRACT.....	vi
LIST OF TABLES	x
LIST OF FIGURES	xi
CHAPTER I AN OVERVIEW OF THE EFFECTORS.....	1
CHAPTER II IDENTIFICATION AND INITIAL CHARACTERIZATION OF THE EFFECTORS OF AN ANTHOR SMUT FUNGUS AND POTENTIAL HOST TARGET PROTEINS	42
Overview	42
Introduction	43
Results.....	45
Discussion	57
Materials and Methods	66
Conclusions	76
CHAPTER III FUNCTIONAL CHARACTERIZATION OF THE <i>MICROBOTRYUM</i> <i>LYCHNIDIS-DIOICAE</i> EFFECTORS IN A HETEROLOGOUS HOST MODEL	77
Overview	77
Introduction	78
Materials and Methods	80
Results.....	91
Discussion	103
CHAPTER IV EXPRESSION OF MVLG_01732 HIS TAGGED EFFECTOR IN <i>SILENE LATIFOLIA</i> TO IDENTIFY NATIVE HOST INTERACTORS	107
Introduction	107
Materials and methods	109
Results.....	116
Discussion	126

CHAPTER V CONCLUSIONS AND FUTURE DIRECTIONS	129
REFERENCES	132
APPENDIX.....	147
CURRICULUM VITAE.....	312

LIST OF TABLES

Table 1- 1: Bacterial pathogens that utilize the T3SS.....	3
Table 1- 2: Bacterial effectors and their domains/motifs with roles in virulence.....	5
Table 1- 3: Distribution of RXLR motif proteins in oomycete pathogens along with their host dependency and pathogenicity strategy.....	7
Table 1- 4: List of HST virulence effectors produced by various necrotrophic fungi.....	12
Table 1- 5: <i>Microbotryum</i> pathogen species and their hosts	35
Table 2- 1: Candidate SSPs chosen for further analyses.	51
Table 3- 1: Primers used for expressing the <i>M. lychnidis-dioicae</i> genes in <i>A. thaliana</i>	82
Table 3- 2: Primers used for qRT analysis.....	88
Table 3- 3: RXLR-like variants in the <i>M. lychnidis-dioicae</i> effector proteins	99
Table 4- 1: Primers used to generate MVLG_01732-6x His pPzP-MV-Hyg construct..	111
Table 4- 2: Primers used for sequencing MVLG_01732-6x His pPzP-MV-Hyg.....	114
Table 4- 3: Number of <i>Microbotryum</i> colonies obtained for each transformation experiment.....	120
Table 4- 4: Number of plants infected along with the dose of infection	123

LIST OF FIGURES

Figure 1- 1: Mechanism of entry of different biotrophic fungi.	24
Figure 2- 1: Evaluating intrinsic disorder propensity of protein effectors.....	47
Figure 2- 2: Results of secretion trap experiment with four <i>M. lychnidis-dioicae</i> predicted small secreted proteins (SSP) effectors.....	50
Figure 2- 3: Autoactivation of three reporter genes by MVLG_04106 on QDO/X- α -gal + 3-AT (5 mM) plates.	53
Figure 2- 4: Yeast two-hybrid spot test results.	56
Figure 2- 5: Computational framework for prediction of secretome for <i>M. lychnidis-dioicae</i> and selection of candidate effectors for further analyses.	68
Figure 3- 1: Schematic representation of the T-DNA construct.	91
Figure 3- 2: PCR on <i>A. thaliana</i> lines expressing MVLG_01732 and mCherry.....	92
Figure 3- 3: Phenotypic abnormalities of the transgenic plants expressing MVLG_1732 Δ SP-mCherry.	94
Figure 3- 4: Analysis of MVLG_01732 Δ SP-mCherry transcript levels.....	95
Figure 3- 5: Western blot of protein extractions from the transgenic lines.	96
Figure 3- 6: Localization studies in the 14-day old seedlings.	97
Figure 3- 7: Subcellular localization in protoplasts.	98
Figure 3- 8: Morphological differences in the anthers of WT (<i>Col-0</i>) and MVLG_01732 Δ SP-mCherry expressing plants.....	100
Figure 3- 9: SEM analysis of the pollen in the WT (<i>Col-0</i>), mCherry, and MVLG_01732 Δ SP-mCherry lines.....	101
Figure 3- 10: MVLG_01732 promote the growth of <i>Pst</i> DC3000 growth.	102

Figure 4- 1: Schematic representation of the MVLG_01732-6xHis-T-DNA construct. .	117
Figure 4- 2: PCR on eight <i>A. tumefaciens</i> transformants bearing MVLG_01732-6xHis pZP-Mv-Hyg.....	118
Figure 4- 3: Restriction digestion of plasmids from the <i>E. coli</i> re-transformed with a positive construct from <i>Agrobacterium</i>	119
Figure 4- 4: Electrophoresis of MVLG_01732-6xHis PCR products.	121
Figure 4- 5: Electrophoresis of HYG PCR products.	122
Figure 4- 6: Infected <i>Silene latifolia</i> plant expressing MVLG_01732-6xHis tagged protein.	124
Figure 4- 7: Western blot analysis and Ponceau S staining of proteins.....	125

CHAPTER I

AN OVERVIEW OF THE EFFECTORS

During the arms race between a pathogen and its host, the host produces its defense response to the attack and the pathogen, to counteract defense responses, may utilize a wide array of strategies, one of which involves secreted proteins, called effectors. Effectors are secreted proteins with diverse structural and functional characteristics that are unique to the pathogen and promote pathogenicity. They are known to change the physiology of their hosts to support pathogen growth. Some of them possess enzymatic activity and interact with host proteins to evade the recognition by the host while others suppress the host defense, thereby enhancing pathogen virulence. In bacterial pathogens of mammalian and plant cells, the effectors are usually secreted into the cell using a type III secretion system, a type IV secretion system or a type VI secretion system (Depluvere, Devos et al. 2016). Fungal plant pathogens also use a repertoire of secreted effector proteins (Petre and Kamoun 2014). Such effector proteins are often secreted via conventional endoplasmic reticulum-Golgi apparatus route with their N-terminal secretion signal. They are usually expressed only after contact with their host as their expression is tightly correlated with the different infection stages. Because of the coevolutionary arms race between the host and the pathogen, these effector proteins are also predicted to evolve very rapidly, resulting in high specificity for their hosts (Rouxel, Grandaubert et al. 2011). When recognized by a corresponding host resistance protein (R

protein) the effector proteins are referred to as avirulence proteins (Avr proteins) and those that are not recognized are known as virulence proteins (avr) (Ellis, Catanzariti et al. 2006). According to the gene-for-gene model, the recognition of an avirulence protein by a host resistance protein results in hypersensitive response triggering localized cell death and preventing further parasitic growth (Greenberg and Yao 2004). Effectors and their mode of translocation have been studied in greater detail in prokaryotes (see Table 1-1). In the oomycetes, a fungus-like lineage of eukaryotes, effectors were classified into two categories: apoplastic effectors that reside in the apoplast, the space outside the plasma membrane in a plant cell and the second category were taken up into the host plant cells by a conserved motif (RXLR) and is discussed briefly below. (Brefort, Doehlemann et al. 2009).

Bacterial effectors

Many bacterial effectors mimic host proteins in order to subvert the host systems for the benefit of the pathogen. Effectors have been found to have some of the following enzymatic activities: Small ubiquitin-like modifier (SUMO), cysteine proteases, ubiquitin-like protease, E3 ubiquitin ligase, tyrosine phosphatases, and ADP-ribosyl transferases. The Type III secretion system (T3SS) is a widely used mechanism to secrete effectors by the bacterial pathogens that cause a wide range of diseases in plants, animals, and humans. Extensive research on bacterial effectors dependent on T3SS has been carried out with species such as *Chlamydia*, *Salmonella*, *Shigella*, *Yersinia*, *Pseudomonas*, *Xanthomonas* and on *Escherichia coli* (Table 1-2). The effectors with the conserved domains/motifs that were shown to have a role in virulence are found in Table 1-3 (Dean 2011).

Phyto-bacterial effectors can suppress the ability of the host plant to detect the microbial-associated molecular patterns (MAMP) found on such pathogens. The common MAMPs from bacteria include bacterial flagellin, lipopolysaccharides, the elongation factor Tu, cold shock proteins, and hairpin proteins that represent one of the classes of T3SS-secreted proteins. Upon recognition of these MAMPs by the plant cell, there follows a MAMP-triggered immune response which can prevent infection before the bacteria gains access to the host plant (Tampakaki, Hatziloukas et al. 2009).

Table 1- 1: Bacterial pathogens that utilize the T3SS

Host	Bacterial species	Disease caused	Repertoires of effectors proven to be secreted or translocated
Plant host	<i>Pseudomonas syringae</i> , numerous pathovars	Range of plant diseases, e.g. tomato speck	HopK1, HopY1, HopAS1, HopU1, HopF2, HopH1, HopC1, HopAT1, HopG1, HopD1, HopQ1, HopR1, HopAM1, HopN1, HopM1, AvrE, AvrB3, HopB1, HopX1, HopZ3, HopAb2, AvrPto, HopE1, HopV1, HopAQ1, HopG1, HopI1, HopA1, HopX1, HopO1, HopT1, AvrRpt2, AvrA, HopW1, HopD1, HopQ1, AvrD1, AvrB2, HopAR1 (see Cunnac, 2009 for nomenclature)
	<i>Xanthomonas</i> spp.	Wide range of plant diseases, e.g. rice bacterial blight and citrus canker	AvrBs1, AvrBs2, AvrBs3, AvrRxo1, AvrRxv, AvrXccC, AvrXv3, Ecf, HpaA, XopJ, XopX, XopB, XopC, XopD, XopE, XopF, XopN, XopO, XopP, XopQ

	<i>Ralstonia solanacearum</i>	Plant wilt on many host species	GALA1-7, SKWP1-6, HLK1-3, RipB, PopW, PopP, PopC, RipT, AvrA, PopB, PopA, RipA (many others predicted)
	<i>Erwinia amylovora</i>	Causes fire blight on a range of plant species	DspE, HrpN, HrpW, HopPtoC, AvrRpt2, EopB
	<i>Rhizobium</i> spp.	Symbiont; forms nodules on legumes	NopL, NopP, NopJ, NopM, NopT, NopB, NopN
	<i>Pantoea</i> spp.	Bacterial wilt on corn and maize	WtsE, PthG, HsvG, HsvB
Animal host	<i>Pseudomonas aeruginosa</i>	Opportunist pathogenic. Can cause pneumonia	ExoU, ExoY, ExoS, ExoT
	<i>Escherichia coli</i> (EPEC and EHEC), <i>Citrobacter rodentium</i>	Diarrhoea (EPEC) or haemorrhagic colitis (EHEC). Cattle commensal (EHEC)	Tir, Map, EspF, EspB, EspZ, EspH, EspG, NleA, NleB, NleC, NleD, NleE, NleF, NleG, NleH, NleK, NleL, EspJ, EspK, EspL, EspM, EspY, EspX, EspO, EspW
	<i>Salmonella enterica</i> serovars	Gastroenteritis; typhoid fever	AvrA, SipA, SipB, SipC, SipD, SopA, SopB, SopE, SopE2, SptP, SlrP, SopD, SspH1, SteA, SteB, GogB, PipB, SifA, SifB, SopD, SpiC, SseF, SseG, SseI, SseJ, SseK, SspH, SteC, SpvB, SpvC
	<i>Shigella</i> spp.	Bacillary dysentery; shigellosis	IpaA, IpaB, IpaC, IpaD, IpaJ, IpgD, IpgB, IcsB, OspC, OspD, OspZ, OspB, OspF, VirA, OspE, OspG, IpaH family
	<i>Yersinia</i> spp.	Bubonic plague (<i>pestis</i>); Gastrointestinal disease (<i>enterocolitica</i>)	YopE, YopH, YopP/J, YopE, YopM, YopT, YpkA/YopO
	<i>Yersinia enterocolitica</i> biovar 1B	Severe gastrointestinal disease	YspA, YspL, YspP, YspF, YspE, YspI, YspK, YspM

	<i>Photorhabdus</i> spp.	Opportunistic pathogen (<i>asymbiotica</i>); insect pathogen (<i>luminescens</i>)	LopT
	<i>Chlamydia</i> spp.	Obligate intracellular parasites, sexually transmitted disease; can cause blindness	CADD, CT847, tarp, IncA, IncG, CT229, CT813, Cpn0585, Cpn0909, Cpn1020
	<i>Burkholderia</i> spp.	Melioidosis (<i>B. pseudomallei</i>); glanders (<i>B. mallei</i>)	CHBP, BopE
	<i>Vibrio</i> spp.	Gastroenteritis, wound infections (<i>parahaemolyticus</i>); secretory diarrhoea (<i>cholerae</i>)	VopL, VopA, VPA450, VopT, VopF, VopS, VopQ
	<i>Bordetella</i> spp.	Whooping cough	BopC/BteA, BopN
	<i>Aeromonas</i> spp.	Opportunistic pathogen; fish/humans	AexT, AopB

EPEC, enteropathogenic *E. coli*; EHEC, enterohaemorrhagic *E. coli*. With permission from (Dean 2011)

Table 1- 2: Bacterial effectors and their domains/motifs with roles in virulence

Bacterial effector	Domain or motif with a virulence role
SipA	DEVD – caspase 3 cleavage site
SseI	Cys178 – unknown function
SseL	Cysteine protease site
YopM	Leucine-rich repeats (6–15)
ExoS	ADP-ribosyl transferase domain

HopF2Pto	Myristoylation motif
AvrPto	Ser149 phosphorylation
WtsE and AvrE1	WxxxE motif
SifA	CAAX box
GALA effectors	F-box
SseJ	Lipase domain – SHD triad
SpvB	ADP-ribosyl transferase domain
NleA	PDZ domain
YpkA	Kinase domain
YopE	GAP domain

With permission from (Dean 2011)

Oomycete effectors

Some of the oomycete pathogens encode effectors with a signature amino acid motif RXLR, that aid in the translocation of effectors into plant cells to promote virulence. Analysis of oomycete genomes revealed that RXLR genes are abundant in *Phytophthora* genera and in downy mildew pathogen, *Hyaloperenospora arabidopsidis* (Table 1-3). Furthermore, variants of RXLR such as GKLR or QXLR are also known to exist that are designated as RXLR-like effectors. However, other lineages in oomycetes possess few or no RXLR genes. For example, the biotrophic white blister pathogens in the *Albugo* genus contain few putative effectors with RXLR genes and the necrotrophic pathogens in the *Pythium* genus and the animal pathogens in the *Saprolegnia* possess no RXLR genes (Table 1-3). Another interesting feature about oomycete effectors is that the corresponding R gene in the host plant encodes nucleotide binding, leucine-rich repeat

(NLR) proteins. Examination of the literature suggests that RXLR effectors primarily work through suppression of the plant immune response by manipulating several hormonal signaling pathways, such as salicylic acid and jasmonic acid pathways (Anderson, Deb et al. 2015).

Table 1- 3: Distribution of RXLR motif proteins in oomycete pathogens along with their host dependency and pathogenicity strategy

Oomycete Pathogen	Host dependency & Pathogen Strategy	Number of RXLR Proteins
<i>Albugo candida</i>	Obligate Biotroph	26
<i>Albugo laibachii</i>	Obligate Biotroph	48
<i>Phytophthora sojae</i>	Facultative Hemi-biotroph	374
<i>Phytophthora ramorum</i>	Facultative Hemi-biotroph	396
<i>Phytophthora infestans</i>	Facultative Hemi-biotroph	563
<i>Hyaloperonospora arabidopsidis</i>	Obligate Biotroph	134
<i>Pythium species</i>	Facultative Necrotroph	0

Modified from (Anderson et al., 2015)

Fungal pathogens and effectors

The lifestyles of pathogenic fungi are highly varied based on how they interact with their host plants. Fungal lifestyles for phytopathogens may be classified as biotrophic or necrotrophic. Biotrophy is further subdivided into two classes- obligate and hemibiotrophy. Obligate biotrophs are usually nonculturable in lab conditions and are entirely dependent on the living host plant for growth and reproduction and obtain nutrients from a living tissue whereas hemibiotrophs have an initial biotrophic phase

followed by a necrotrophic phase and are culturable *in vitro*. Necrotrophic fungi upon infecting the host, kill the host tissue and extract the nutrients from the dead host cells (Selin, de Kievit et al. 2016). They also differ in the types of infection structures, called appressoria, that they develop during the infection. The hemibiotrophs like *Magnaporthe oryzae* and *Colletotrichum* species develop a dome-shaped appressorium and use turgor pressure to enter mechanically into the host cells. They initially develop bulged biotrophic invasive hyphae that later change into thin necrotrophic hyphae. However, necrotrophic fungi like *Botrytis cinerea* and *Sclerotinia sclerotiorum* develop unnoticeable appressoria and utilize plant cell wall degrading enzymes (PCWDEs) to penetrate the plant cells. They grow sub-cuticularly and secrete toxic components, killing the host epidermal cells. Biotrophic pathogens, unlike necrotrophs, must stay “under the radar” by overcoming the basal and pathogen-associated molecular patterns (PAMP) triggered defense mechanisms elicited by their host plants. They interfere with various metabolic activities in their hosts to achieve and maintain biotrophy. Some of them develop haustoria, a feeding structure, and use a combination of turgor pressure and PCWDEs to rupture the cell wall without killing the host cell. Once they enter the host tissue, their hyphae grow either intracellularly or intercellularly. For the pathogens with intracellular hyphal growth, the haustoria is surrounded tightly by a host plasma membrane, called the extrahaustorial membrane, although the composition of this membrane differs from the common plant plasma membrane proteins (Lo Presti, Lanver et al. 2015). One of the major differences between the biotrophs and necrotrophic pathogens arises in the targeting of particular defense signaling pathways in their host. Necrotrophs induce salicylic acid dependent cell death responses whereas biotrophs

promote different pathways like jasmonic acid pathways or ethylene responses during compatible interaction with their hosts. There are several ways that a biotroph could sneak into the host without drawing attention. Some of these strategies include: a) modification of the fungal cell walls to escape host recognition, b) sequestering the fragmented fungal cell wall components that trigger PAMP-triggered immune responses from the host, c) secreting apoplastic effectors that could neutralize the activity of antimicrobial enzymes of the host, d) secreting and translocating intracellular effectors into the host cell that interfere with several intracellular defense pathways of the host (Brefort, Doehlemann et al. 2009).

Effector proteins in the fungal repertoire are defined as small secreted proteins containing <300 amino acids that are highly upregulated during infection and bear a signal peptide at their N-terminus region that helps in their secretion. Most are cysteine-rich with disulfide bonds stabilizing their tertiary structures, so they can function in the harsh physiological conditions in the apoplastic compartment of the plant cell (Lo Presti, Lanver et al. 2015). Unlike oomycetes, with a conserved domain, no domain is yet identified as conserved for most of the fungal effectors with the exception of powdery mildews. Effector proteins show signatures of positive diversifying selection because of the co-evolutionary race between hosts and pathogens. The roles of effectors and their mechanism of action differ in biotrophs vs necrotrophs, as the former requires a living host for its survival. The mode of entry of these fungal effectors is still not clearly understood. Some of them are secreted by the haustorial surface into the extracellular space, between haustoria and the host plasma membrane, while others attach to a receptor and get internalized. Although all the effectors are secreted to the apoplast, only some of

them remain there, *e.g.*, cell wall degrading enzymes, while others must be translocated into the host cell.

Plant resistance to pathogens

In the case of phytopathogen systems, the defense system used by plants is different from the one used by mammals or other invertebrates. In general, plants contain two lines of defense. The primary immune strategy by a plant when exposed to a microbe or a pathogen is called MAMP/PAMP-triggered immunity (PTI) by which a basal response is initiated. In the case of fungi, the best known PAMP is chitin, a major structural component in the fungal cell walls, while in bacteria it is flagellin. These conserved molecular patterns that are usually in the extracellular space are recognized by the membrane-associated PAMP recognition receptors (PRR) in plant cells. There are two types of PRRs depending on the presence of the intracellular kinase domain. They are receptor-like kinase proteins (RLKs) and receptor-like proteins (RLPs). Both have the extracellular domain for binding to PAMPs but only RLKs contain an intracellular signal transducing cytoplasmic kinase domain which RLPs lack (Pandey, Rajendran et al. 2016). The recognition of PAMPs by PRRs leads to the activation of signaling cascades like Mitogen-activated protein kinases (MAPK) that lead to the production of reactive oxygen species, accumulation of callose and formation of papilla restricting the cell wall breach created by pathogens, eventually preventing further colonization of the host (Gohre and Robatzek 2008).

Successful pathogens evade their hosts by secreting effectors resulting in effector-triggered susceptibility (ETS). However, plants also evolved to encode resistance R proteins that recognize the effectors, resulting in effector-triggered immunity (ETI),

activating acute disease resistance response with signs of hyper-response and cell death (de Jonge, Bolton et al. 2011). The pathogen's countermove for this strategy is to either modify their arsenal of effectors keeping their virulence or lose and replace the old arsenal of effectors with a new one that can suppress ETI (Jones and Dangl 2006). The majority of R proteins in the cytoplasm consist of a nucleotide binding site (NBS) connected to a region of leucine-rich repeats or LRRs (NBS-LRRs). The second group of R proteins has an extracellular LRR (eLRR) domain and a short transmembrane (TM) domain. The first group of NBS-LRR is further divided to two subclasses based on their N-terminal domains that either contain a coiled-coil region (CC) or Tol/ interleukin-1 receptor (TIR) motif (Stergiopoulos and de Wit 2009).

Necrotrophic fungal effectors

Necrotrophic fungi produce several toxins and cell wall hydrolyzing enzymes for colonization and to induce cell death in their host plants. Some of the broad range fungal necrotrophs include *Botrytis cinerea*, *Alternaria brassicicola*, *Plectosphaerella cucumerina*, and *Sclerotinia sclerotiorum* (Pandey, Rajendran et al. 2016). Their arsenal includes polyketide toxins, secondary metabolites, ROS, nonribosomal peptide toxins, necrosis- and ethylene-inducing peptide 1 (Nep1) proteins (Lo Presti, Lanver et al. 2015). Some of the above components are discussed in detail elsewhere (Qutob, Kemmerling et al. 2006, Stergiopoulos, Collemare et al. 2013). Some host-specific necrotrophic fungi, produce host-selective toxins (HST) as virulence effectors that interact with the cognate dominant receptor gene in the host plant, in turn making the host susceptible to infection (Table 1-4) (Pandey, Rajendran et al. 2016). This is in contrast to the classical gene-for-

gene interaction in which a dominant gene confers disease resistance rather than susceptibility (Ali and Bakkeren 2011).

Table 1- 4: List of HST virulence effectors produced by various necrotrophic fungi

Necrotrophic pathogen	Host	HST	Receptor/target cell organelle
<i>Alternaria alternata</i> f.sp. <i>lycopersici</i>	Tomato	AAL toxin	Endoplasmic reticulum
<i>Alternaria mali</i>	Apple	AM toxin	Chloroplast
<i>Alternaria kikuchiana</i>	Japanese pear	AK toxin	Plasma membrane
<i>Alternaria fragariae</i>	Strawberry	AF toxin	Plasma membrane
<i>Alternaria citri</i>	Lemon	ACT toxin	Plasma membrane/plasmodesmata
<i>Alternaria alternata</i>	Rough lemon	ACR toxin	Mitochondria
<i>Alternaria brassicae</i>	<i>Brassica</i> crops	Destruxin B	Mitochondria, chloroplast
<i>Cochliobolus heterostrophus</i>	Corn	T toxin	URF13 protein located in mitochondria
<i>Cochliobolus carbonum</i>	Corn	HC toxin	Histone deacetylases
<i>Cochliobolus victoriae</i>	Oat	Victorin	Plasma membrane
<i>Pyrenophora tritici-repentis</i>	Wheat	Ptr-Tox A	Chloroplast, ToxA bp1
		Ptr- Tox B	
<i>Periconia circinata</i>	Sorghum	Peritoxin (PC toxin)	Plasma membrane
<i>Phyllosticta maydis</i>	Corn	PM toxin	URF 13 protein of mitochondria
<i>Rhizoctonia solani</i>	Rice	RS toxin	unknown
<i>Parastagonospora nodorum</i>	Wheat	SnToxA	ToxABP1, chloroplast

With permission from (Pandey, Rajendran et al. 2016)

Biotrophic fungal effectors

Biotrophic fungi colonize the living tissue of their host plants and depend on them as their source of nutrients. One of the conserved features of biotrophs that grow intracellularly or that form haustoria is that they are enclosed by plant plasma membrane known as an extra- haustorial membrane. This membrane is continuous with the plant plasma membrane but contains a set of distinct transmembrane proteins. It is also known that these haustorial structures not only help in the nutrient exchange but also are the sites of effector delivery (O'Connell and Panstruga 2006). The mechanism of the entry of some of the biotrophic fungi is shown in Figure 1. In this chapter, I focused on the roles of effectors in various fungal biotrophic pathogens and these are discussed in detail.

Cladosporium fulvum: *C. fulvum* is an ascomycete and an asexual, non-obligate biotrophic pathogen that causes leaf mold disease in tomato. Upon stomatal entry (Figure 1-1), it does not produce any feeding structures like haustoria, but lives in the apoplast and depends on the host's sugars and amino acids. During the incompatible interaction, mesophyll cells of the host plant recognize the fungus and induce a hypersensitive response that blocks fungal proliferation (Stergiopoulos and de Wit 2009). This interaction follows the gene-for-gene concept whereby for every avirulence gene in *C. fulvum*, there is a corresponding tomato resistance gene that mediates the recognition of the fungal pathogen. The *C. fulvum* resistance genes (*Cf*) of the host plant encodes leucine-rich repeat (LRR) receptor-like proteins (RLPs). They lack a domain for signaling and in turn interact with other partners called receptor-like kinases (RLKs) suggesting that the defense response is mediated through such interactions (Faulkner and Robatzek 2012). These LRR-RLPs interact with the respective avirulence protein and

carry out race specific immunity to *C. fulvum*, which is characterized by the hypersensitive response and cell death. The effectors with the well-characterized functions are discussed below.

Avr2: This is one of the fungal effectors, along with Avr4, with a proven role in virulence. It is shown to bind to and inhibit the tomato apoplastic cysteine protease Rcr3. The fungal strains expressing Avr2 protein trigger a hypersensitive-response in tomato lines carrying the corresponding *Cf-2* resistance gene. Heterologous expression of *Avr2* in tomato and *Arabidopsis* enhanced disease susceptibility when challenged with other fungal pathogens like *Botrytis cinerea* and *Verticillium dahliae*. Gene silencing of *Avr2* compromises the virulence in tomato (van Esse, van't Klooster et al. 2008).

Avr4 and Avr9: Avr4 protein has a chitin-binding motif, binds to chitin, and subsequently protects the fungal cell walls from the action of host plant chitinases (van den Burg, Harrison et al. 2006). These chitin binding motifs are recognized by LRRs of RLP proteins like Cf-4 and Cf-9, are plasma membrane destined receptors, delivered by the secretory pathway, and interact constitutively with another LRR-RLK, Suppressor of BIR1 (SOBIR1), that is constitutively endocytosed. This triggers further interaction with another RLK, called BR1-associated kinase 1/ Somatic embryogenesis receptor kinase 3 (BAK1/SERK3), which is known to recycle between the plasma membrane and endosomes. Upon secretion of Avr4 effector to the apoplast by *C. fulvum*, and recognition by Cf-4, the interaction of the latter occurs with BAK1/SERK3. This heterocomplex is recognized and triggered to vacuolar degradation by the endocytotic pathway. This recognition is required for the hypersensitive response and resistance of tomato strains against *C. fulvum*. Gene silencing of the respective genes, *Avr4*, and *Avr9*

in *N. benthamiana* and *Solanum lycopersicum*, suppresses the hypersensitive response and allows fungal proliferation. Live cell imaging of transiently and stably expressed fluorescent protein fusions in *N. benthamiana* suggest the plasma membrane as localization site for Cf-4 (Postma, Liebrand et al. 2016).

Avr9: The founding member of the Cf proteins is Cf9. It is an LRR-RLP that mediates resistance to *C. fulvum* strains producing Avr9. SERK1 and BAK1/SERK3 are recruited to the Cf-9 receptor in an Avr9 responsive manner. The requirement of BAK1/SERK 3 by both Cf-4/9 receptors suggests that they both activate similar kinds of downstream signaling mechanisms within the host plant (Postma, Liebrand et al. 2016).

Ecp6: This effector, like Avr4, binds to chitin. Presence of chitin on fungal cell walls triggers PAMP mediated recognition in plants. Co-precipitation studies with chitin suggested that it is a chitin-binding lectin with the LysM domain, that inhibits the activation of chitin-triggered immunity in host plants. It scavenges the small chitin fragments that are released by the fungal cell walls during the action of plant chitinases, to prevent the elicitation of host immune responses (de Jonge, Peter van Esse et al. 2010). Gene silencing experiments of Ecp6 showed reduced virulence in *C. fulvum*, suggesting that preventing chitin recognition is an important strategy in *C. fulvum* infection (Bolton, van Esse et al. 2008).

Blumeria graminis: Powdery mildews are a large group of ascomycete fungi. *B. graminis* f. sp. *hordei* (*Bgh*) is a barley powdery mildew fungus, an obligate biotroph that infects barley species. During the infection process, it develops a highly specialized structure called a haustorium, commonly seen only in obligate biotrophic rust fungi and some (hemi)-biotrophic oomycetes. It is known that haustorial surface is a major site of

effector delivery and nutrient acquisition by plant pathogens (Stergiopoulos and de Wit 2009). However, the lack of a reliable transformation method for powdery mildew fungus and its inability to grow in culture hampers the progress in the validation of effectors. Host-induced gene silencing (HIGS) is a technology to study the roles of genes *in vivo* where a reporter gene construct together with an RNA interference (RNAi) construct targeting fungal genes are transiently expressed in single leaf epidermal cells of barley. The host leaves were infected with *Bgh* and later stained for the reporter gene activity. The percentage of epidermal cells expressing the reporter gene with haustoria is used as an index to measure fungal proliferation (Zhang, Pedersen et al. 2012). More than 25 *Avr* genes and 85 barley *R* genes (*Ml*) have been identified. These include *Mlk* genes and 28 homologous genes that were mapped to *Mla* (for mildew A) locus from which 6 are highly related to CC-NBS-LRR proteins (Ridout, Skamnioti et al. 2006). The genome is about 120 Mb, due to a large amount of retrotransposon-derived repetitive DNA (Spanu, Abbott et al. 2010). The genome encodes about 248 candidate secreted effector proteins (CSEPs) with around 200 CSEPs that share the common motif Y/F/WxC, where the first amino acid could be any of the three aromatic amino acids and the last one being always cysteine. Moreover, the effectors with this motif are also mostly shown not to have cysteine residues except for a single conserved cysteine residue at the C-terminal end, which likely forms a disulfide bond with the N-terminal cysteine residue in the Y/F/WxC motif (Godfrey, Bohlenius et al. 2010). Interestingly, unlike other fungal *Avr* proteins, the mildew effector proteins lack an N-terminal secretory signal and have been suggested to rely on some other, as yet unknown, mechanism for secretion. The conservation of the Y/F/WxC motif across most of the effectors also suggests its possible role in effector

delivery as observed for oomycete RxLR motif. However, their cytoplasmic expression in transient transformation assays suggests their location of interaction as cytoplasmic in the host cell (Ellis, Dodds et al. 2007).

BEC4 (*Blumeria* effector candidate) was shown to interact with ADP-ribosylation factor (ARF)-GTPase-activating protein (GAP) (ARF-GAP), a regulator of intracellular vesicle trafficking, in yeast two-hybrid (Y2H) screens. Bimolecular fluorescence complementation analysis (BiFC), when components were transiently expressed in barley epidermal cells, showed co-localization of the interactor with the Golgi marker, in line with its function, as ARF regulates vesicle budding at the Golgi apparatus and also in the cytoplasm, a potential site of interaction with the BEC4 effector. However, co-expression of both the effector and interactor could not be detected via this method. Knockout mutant lines of the orthologous ARF in *Arabidopsis* showed an increased susceptibility and haustorium formation with non-adapted powdery mildew pathogen, *Erysiphe pisi* and an enhanced resistance to the adapted *Arabidopsis* downy mildew oomycete pathogen, *Hyaloperonospora arabidopsidis* compared to control lines, indicating the antagonistic action of ARF in defense against adapted vs non-adapted pathogens. This might indicate that the orthologous ARF in *Arabidopsis* is the target for secreted effectors of the downy mildew pathogen, but the mode of action of these downy mildew effectors is opposite to the activity of BEC4. The authors speculate that the ARF-GAP proteins might be common targets for both powdery mildew and downy mildew effectors and BEC4 might interfere with defense-associated host vesicular trafficking. (Schmidt, Kuhn et al. 2014).

AVrA10 and AvrK1: These Avr proteins belong to the large EKA family of around 1000 paralogues in *Bgh* (Zhang, Pedersen et al. 2012) and they lack predicted N-terminal signal peptide for secretion. AVrA10 is known to interact with cognate resistance protein Mla10-containing barley varieties. AVrA10 was found to be a paralogue of AvrK1. Transient expression in single host epidermal cells indicates that these elicit host cell death, reduced fungal sporulation by restricting the accessibility and function as effectors in plants lacking Mlk1. The AvrK1 gene is located near retrotransposons and the protein is known to interact with Mlk1 protein (Ridout, Skamnioti et al. 2006). It has been shown by fluorescence microscopy that most of the Mla10 protein is localized in the cytoplasm with minimal presence in the nucleus. Moreover, cell death and infection assays also demonstrated that the functional site of these effectors is the host cytoplasm. In the barley cell nucleus, Mla10 was found to interact with two WRKY transcription factors in the presence of AvrA10, suggesting that these could be the downstream targets for the AvrA10-Mla10 interaction (Shen, Saijo et al. 2007).

BEC1011 and BEC1054 (*Blumeria* effector candidates) are part of a gene superfamily unique to powdery mildew fungi. Structural modeling suggests that they are similar to microbial RNAses and take on the ribonuclease-like fold structure not previously seen as an effector function; thus, they appear to represent a new class of microbial effectors. HIGS- based identification revealed that these proteins act as effectors and contribute to infection in its earliest stages. (Pliego, Nowara et al. 2013).

CSEP0055 (Candidate for secreted effector proteins) has a three-amino acid motif – ‘YxC’, at the N-terminus of the mature protein. Y2H screens and BiFC experiments confirmed its interaction with barley pathogenesis-related protein PR17c. PR17c was

shown to localize to the papillae in the apoplast during infection by *Bgh*. Downregulation and over-expression analysis of PR17c confirmed it to be important for providing resistance to the penetration of mildew fungus in barley epidermal cells. Silencing of CSEP0055 by HIGS has shown to reduce haustorial index (HI) indicating the role of CSEP0055 in fungal proliferation. Time course experiments on the transcript of CSEP0055 shows that this could play a role in facilitating secondary penetration events (Zhang, Pedersen et al. 2012).

CSEP0105 and CSEP0162: Silencing by HIGS showed a significant reduction in the haustorial index suggesting their roles in the proliferation of the fungus. Y2H screens and BiFC experiments confirmed the interaction partners as small heat shock proteins (sHsps), Hsp16.9 and Hsp 17.5, respectively. Small heat shock proteins are known to play important roles in refolding misfolded proteins, including defense-related proteins, thereby preventing irreversible protein aggregation during stress. Co-expression studies show the localization of these CSEPs follows the sHsps in the cytosol which suggests their role in interfering with the chaperone activity of sHsps. Recombinant *E. coli* expressing His-tagged-Hsp16.9 showed chaperone activity by preventing protein aggregation in *E. coli* when subjected to thermal stress, with a higher percent of soluble protein fraction compared to control extract from cells not expressing His-tagged-Hsp16.9. The same experiment, when repeated in the presence of CSEP0105 was shown to reduce the soluble protein fraction compared to control extract not having CSEP0105. This suggests that CSEP0105 reduces the chaperone activity of Hsp16.9. The same experiments could not work on His tagged Hsp 17.5 because of the insolubility of the

recombinant His-tagged fusion protein when expressed in *E. coli* (Ahmed, Pedersen et al. 2015).

Melampsora lini: Rusts are basidiomycete fungi and cause diseases in many plant species. Rust fungi are obligate biotrophs and similar to powdery mildews, produce a specialized feeding structure called haustoria during the infection process that penetrates the plant cell wall and is surrounded by a host cell membrane (shown in Figure 1-1). But rust hyphae mostly proliferate within the leaf cell rather than on the leaf surface. *M. lini* is a flax rust fungus of *Linum usitatissimum* (flax plant). In contrast to other rust fungi, *M. lini* does not need an alternate host to complete its life cycle (Lawrence, Dodds et al. 2007). In its host, *L. usitatissimum*, around 30 resistance proteins were identified from different polymorphic loci; these recognize about 30 corresponding avirulence proteins (Avr) from different *M. lini* strains and most of the Avr proteins are expressed in haustoria. The studies on *M. lini* and its host plant also served as a model and elucidated the interactions between R-Avr genes based on a gene-for-gene resistance model (Flor 1947). Most of these R proteins are members of intracellular TIR-NBS- LRR class (Dodds, Lawrence et al. 2001, Dodds, Lawrence et al. 2001). However, a role in virulence for the Avr genes of *M. lini* has not been demonstrated yet.

AvrM, AvrP4, and AvrP123: All of these are haustorially expressed secreted proteins (HESPs). AvrM has a conserved RxLR like motif and interacts with resistance protein M. It appears to translocate autonomously into host cells, but the transport mechanism has not yet been identified (Gan, Rafiqi et al. 2010). AvrP123 is a cysteine-rich protein with a Kazal serine protease inhibitor signature and interacts with resistance proteins in flax plants, P1, P2, and P3. AvrP4 is a Cys rich protein and interacts with P4. *Agrobacterium-*

mediated transient expression of the full-length effector proteins in the resistant flax plants and non-host transgenic *N. tabaccum* containing the cognate R gene showed R gene-dependent cell death with appropriate specificities (Catanzariti, Dodds et al. 2006).

AvrL567: Y2H screens identified its interacting partner as L6 receptor coded by the L6 resistance gene. Because of the lack of transformation system for rust fungi, *Agrobacterium*-mediated transient expression studies were carried out both in the native host and in *Nicotiana tabaccum*. In both the studies, infiltration with AvrL567 or its resistance gene encoded L6 immune receptor alone, did not show any effects. However, co-expression of both, induced a necrotic response after recognition of AvrL567 by the resistance L6 protein, indicating the resistance gene-mediated cell death. AvrL567 is a HESP with a conserved 'RxLR like' motif, expressed in the haustorium and is found within the host cell. Like AvrM, it appeared to translocate autonomously into host cells with an unknown transport mechanism (Gan, Rafiqi et al. 2010). Furthermore, having an intracellular NBS-LRR class L receptor protein as interacting partner suggests its delivery into the host cell during the infection process (Dodds, Lawrence et al. 2004).

***Puccinia graminis*:** These rust fungi are heteroecious basidiomycetes, requiring two different hosts, wheat, and barberry, to complete its life cycle. It uses the barberry host to complete its sexual reproduction. It is an obligate pathogen that causes stem rust and is macrocyclic (Petersen 1974), having five spore stages. They are basidiospores, pycniospores (spermatia), aeciospores, uredinospores, and teliospores. As in *Blumeria*, genome expansion has occurred in *Puccinia* and genome sequencing of *P. graminis tritici* identified around 200 candidates as effector genes that are expressed in haustoria (Upadhyaya, Mago et al. 2014).

RGDBP is an RGD Binding Protein of 818 aa, that was isolated as a protein bound to RGD (Arginine-glycine-aspartic acid) peptide via affinity chromatography. It has a fibronectin domain that acts as adhesive ligands connecting the cytoskeleton and the extracellular matrix. VPS9 is a vacuolar protein sorting-associated protein 9 homologue of 744 aa, that co-eluted along with RGDBP in affinity chromatography. Y2H screens showed these two proteins interact with a protein encoded by a cognate resistance R gene called *RPG1* (Reaction to *Puccinia graminis* 1), present in the resistant barley cultivars. *RPG1* encodes a receptor-like kinase with two kinase domains pK1 and pK2. The pK2 domain is catalytically active and pK1 is a pseudokinase. Application of purified RGDBP and VPS9 proteins on to barley leaves induced phosphorylation of *RPG1* within 5 minutes and subsequently degradation after inoculation with urediniospores, leading to a hypersensitive response from the host. Both phosphorylation and degradation of *RPG1* are necessary to exhibit this resistance property. The spores that were treated with RGD peptides failed to germinate and could not phosphorylate *RPG1*; this prevented the formation of adhesion pads with host surface and so, inability to cause disease on susceptible barley cultivars. So, the cooperative action of these two effectors is required to exhibit *RPG1*-mediated resistance in barley cultivars (Nirmala, Drader et al. 2011)

PGTAUSPE-10-1 was an effector identified by *Pseudomonas fluorescens* T3SS-mediated delivery that induces genotype specific hypersensitive response (HR) on W3534 line wheat plants carrying the resistance gene *Sr22* (stem rust resistance gene). So, it is suggested that the possible target protein interactor for this effector is *Sr22* although there could be different resistance genes that might have a role in recognizing this effector and it still needs further analysis. It appeared that this effector is active only when delivered

to the cytoplasm, as the experiments driving its delivery to the plasma membrane did not induce HR (Upadhyaya, Mago et al. 2014).

Uromyces fabae: *U. fabae* is the causative agent of rust on *Vicia faba*, commonly known as a broad bean. It is a biotrophic fungus, producing uredospores which germinate and grow in the direction of stomata by using pH gradient until they reach the guard cells (Edwards and Bowling 1986) and form the appressorium. Once it penetrates the host epidermis it forms haustorial mother cell through which it acquires host nutrients and also could potentially secrete effectors (Figure 1-1). During their life cycle, rust fungi form different types of spores *e.g.*, aecio, basidio, pykno, teleuto, and urediniospores. The lattermost is responsible for epidemic infection (Voegelé 2006).

Rtp1: Rust transferred protein 1 (RTP1p) was the first fungal effector protein that was visualized directly in the cytoplasm and nucleus of host plant cells infected by *Uromyces fabae*. Immunofluorescence and electron microscopic studies identified its accumulation in the projections of the extra-haustorial matrix (EHM), an interface between the host cytoplasm and the pathogen haustorium, before getting transferred to the host cytoplasm (Kemen, Kemen et al. 2005). It is an amyloid-like protein and is classified as a structural effector that forms a filamentous structure. It is suggested that the pathogen secretes this effector into the host cytoplasm in late stages of the biotrophic phase, as a way to protect the haustorium from the defense mechanisms of the host plant (Kemen, Kemen et al. 2013).

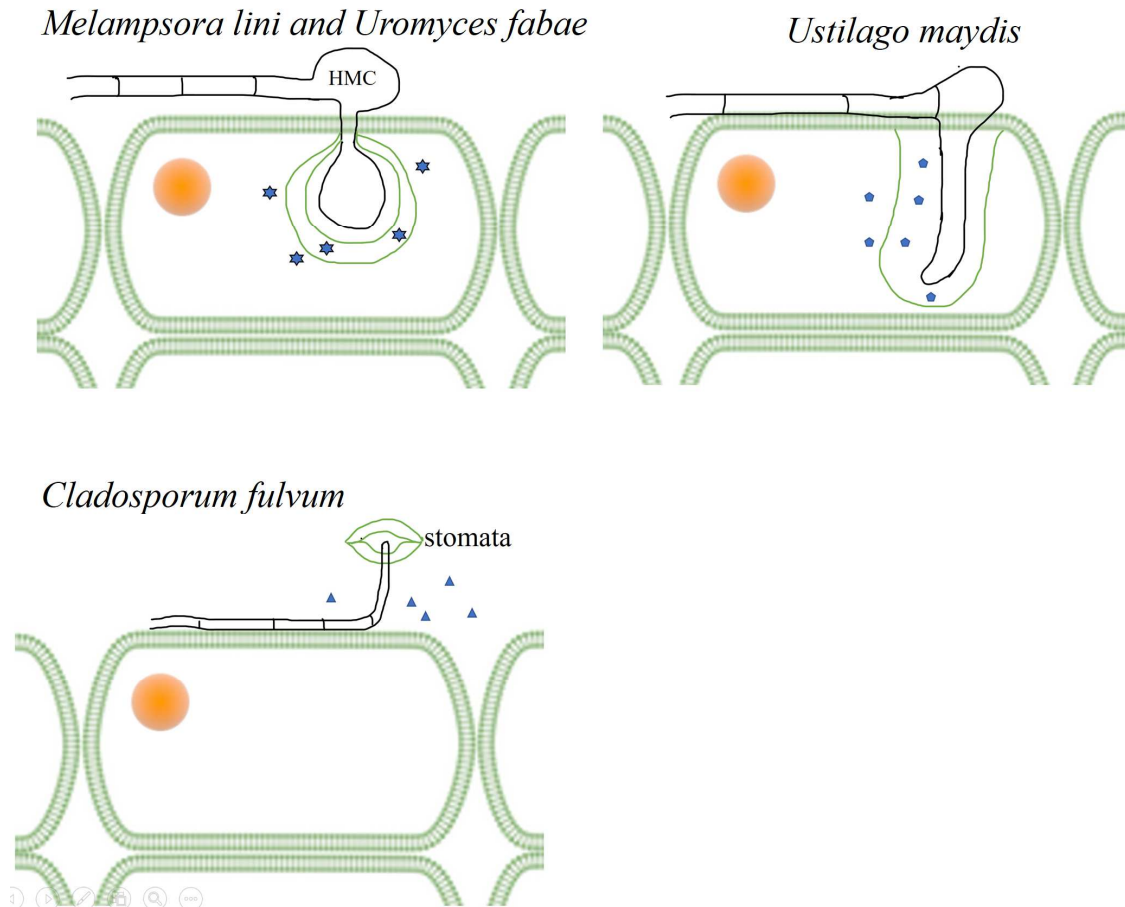


Figure 1- 1: Mechanism of entry of different biotrophic fungi.

M. lini and *U. fabae* form Haustorial mother cell (HMC) that develop into haustoria in the host cells. *U. maydis* hyphae grow intracellularly initially. *C. fulvum* enters through the stomatal opening and grow extracellularly.

***Ustilago maydis*:** Smut fungi are a broad group of basidiomycete fungi with great advantages over rusts and mildews in terms of culturing on defined media and the ease of molecular transformation. *Ustilago maydis*, the causative agent of smut disease in maize is an obligate biotrophic fungal pathogen and is one of the most extensively studied plant pathogens at the molecular level. It initially exhibits intracellular growth (Figure 1-1) and switches to intercellular growth at later stages. It was the first smut fungus to have its

genome sequenced (Kamper, Kahmann et al. 2006). The availability of the full genome sequence and the standard procedures for transformation and gene disruption in *U. maydis* make it an ideal model to understand the molecular mechanisms of an obligate biotroph. In some pathogens, effectors are located in gene-sparse genomic regions with repeat-rich DNA (Raffaele and Kamoun 2012) but in some smut fungi, with a relatively low content of repetitive DNA, these are located as gene clusters that might have originated from gene duplications. Low sequence conservation in these clusters across the smut fungi indicates the rapid evolution within these clusters (Schirawski, Mannhaupt et al. 2010). *U. maydis* genome shows 12 gene clusters encoding around 386 secreted protein effectors (Mueller, Kahmann et al. 2008). Cluster deletion studies reported a change in the virulence phenotype of the pathogen confirming their role in the pathogenicity of *U. maydis* (Kamper, Kahmann et al. 2006). This is one of the unique features of the effectors in *U. maydis* where mutants in single effector genes or gene clusters show a dramatic effect on its biotrophic development, in contrast with effectors in oomycetes exhibiting a lot of functional redundancy.

U. maydis has a tetrapolar mating system with two independently segregating mating type loci *a* and *b* (Puhalla 1968). Locus *a* exists in two alleles and *b* has more than twenty alleles. *U. maydis* has a relatively small genome of 20 Mb, encoding approximately 6,700 genes on 23 chromosomes (Wollenberg and Schirawski 2014). Like the virulence factor, Ecp6 in *C. fulvum*, it was found that *U. maydis* also encodes two proteins with a LysM domain, a carbohydrate-binding module that could sequester the chitin oligosaccharides that elicit the host immune response, although its function needs to be still investigated. Evidence suggests that the plant immune system initially recognizes *U.*

maydis presence and induces a nonspecific PAMP-triggered immune response, but once the pathogen enters the host tissue, the latter establishes a biotrophic environment by suppressing the defense response (Brefort, Doehlemann et al. 2009).

One effector, Cmu1, is a secreted chorismate mutase, a key enzyme at the branching point of the shikimate pathway that produces aromatic amino acids. Analysis of deletion mutants showed a reduction in the formation of large and heavy tumors, a phenotype that could be complemented by the introduction of an HA-tagged Cmu1 wild-type protein to restore the virulent phenotype. Truncated Cmu1-HA lacking secretion signal was unable to complement the virulence phenotype, suggesting that secretion is required for function. Analysis of localization showed that it could be detected inside the fungal hyphae, in the host cytoplasm, and along the biotrophic interface, indicating uptake by the host cell. However, similar experiments with Cmu1-mCherry-HA failed to show the same results suggesting that the addition of the mCherry fluorescent protein could have rendered the protein nonfunctional. By Y2H screens, it was demonstrated that Cmu1 could dimerize and act in conjunction with maize chorismate mutase- ZmCm2 thereby increasing the flow of chorismate from the plastid to the cytosol. This channels the chorismate entry from the salicylic acid pathway to the phenylpropanoid pathway in the cytosol. Salicylic acid is a major plant defense regulator and its suppression likely favors the proliferation of the pathogen (Djamei, Schipper et al. 2011).

Pep1 (protein essential for penetration 1) is a secreted protein that is required for penetration of the *U. maydis* into maize epidermal cells. Pep1 mutants were unable to invade the host cells and not able to establish a compatible interaction. Pep1 mutants are arrested during the penetration of the epidermal cells and induced strong plant defense

response. Pep1 mutants exhibited production of ROS, papilla formation and induction of PR genes that are characteristics of non-host responses in incompatible plant-pathogen interactions. The typical upregulation of jasmonic acid levels is not shown in the Pep1 mutants, which instead, exhibited the higher salicylic acid levels that are a signal for plant defense. The phenotype could be complemented by the ortholog *pep1* from *Ustilago hordei* that is required for penetration in barley. This demonstrates that the Pep1 has a conserved function (possibly among this group of smut fungi) that is not limited to *U. maydis* and maize interaction. Immuno-colocalisation studies with HA-tagged protein demonstrated its presence in the apoplastic space and at sites of cell-to-cell passage (Doehleemann, van der Linde et al. 2009). Although Y2H screens failed to identify the host interactor, biochemical approaches revealed the interactor as POX12, maize peroxidase-12 which was highly induced in plants infected with *pep1* mutants. These class III peroxidases were known to be involved in plant responses to pathogen attack by ROS production, eventually leading to cell death. Pep1 inhibits the apoplastic peroxidase POX12 by scavenging ROS, the first layer of plant defense response, and thus suppressing the plant PAMP-triggered immunity (Hemetsberger, Herrberger et al. 2012).

Pit2 (protein involved in tumors 2): Y2H screen analysis and co-immunoprecipitation studies identified and confirmed the Pit2 interacting partner as maize cysteine protease (CP2). Thus, Pit2 acts as an inhibitor of maize cysteine proteases, an activity essential for *U. maydis* virulence. It shows a concentration-dependent inhibition of CP2 suggesting its role as a competitive inhibitor of the CP2 enzyme. Pit2 has a 14-amino acid conserved motif, which when expressed as a synthetic peptide was still able to inhibit maize cysteine proteases (Mueller, Ziemann et al. 2013). Localization studies using mCherry-

tagged protein showed Pit2 presence mainly at the biotrophic interface with strong accumulations at fungal hyphal tips and in the apoplast of the host cell. Pit2 shows a 35% amino acid sequence identity with its orthologue in *Sporosorium reilianum*, suggesting conserved function between these pathogens. Deletion mutants can infect and proliferate inside the maize plants but fail to maintain biotrophy as they failed to induce tumors and the mutants triggered host defense responses (Doehlemann, Reissmann et al. 2011).

Tin2 (Tumor-inducing 2): One of the hallmarks of *U. maydis* infection is the production of red pigments called anthocyanins. Analysis of Tin2 deletion mutants showed that anthocyanin production was reduced and the ability of the pathogen to reach the vascular tissue was blocked. The compound required for lignin formation, 4-coumaric acid, also contributes to anthocyanin production. Tin2 presence diverts these resources away from lignin production to the anthocyanin pathway (Brefort Thomas, Doehlemann et al. 2009). Thus, the presence of Tin2 interferes with the ability of the host plant cell to lignify the vascular tissues during infection that would have normally blocked fungal access to the vascular tissues. Y2H screens identified its interacting partner as a putative cytoplasmic serine/ threonine kinase, ZmTTK1 (Tin2 targeting kinase 1). Tin2 functions in the cytosol and stabilizes its partner, ZmTTK1, by masking ubiquitin-proteasome degradation motif, which directly increases the synthesis of anthocyanins in the plant. Lack of Tin2 allows resources into the lignin pathway thereby resulting in cell wall fortification in vascular tissue. Although the authors predicted its localization in plant cytoplasm based on the location of anthocyanin production, the mCherry-tagged protein was partially biologically active and could not be seen inside the maize cells (Tanaka, Brefort et al. 2014).

See1 (Seedling efficient effector 1): Another effector, See1, is required during tumor formation in the seedling leaves. Deletion mutants formed only about half of the total number of tumors formed in wild-type infections. Live cell imaging and TEM studies show that the mCherry-tagged truncated protein lacking signal peptide was localized to both the cytoplasm and the nuclei when the construct was transiently expressed in maize leaves. A C-terminal HA-tagged full-length protein upon natural delivery via *U. maydis* hyphae also revealed its translocation from the fungal hyphae to the biotrophic interface and eventually into plant cytoplasm and nucleus. Y2H screens identified the interacting partner of See1 as SGT1, an important factor in the plant host and non-host resistance. The interaction was confirmed by coimmunoprecipitation in *Nicotiana benthamiana* and BiFC *in planta*. See1 interferes with the post-translational modification of SGT1 by inhibiting the MAPK triggered phosphorylation of SGT1 at the Thr-150 residue. This could lead to blocking the host defense signaling and reprogramming the host cell cycle for tumor development (Redkar, Hoser et al. 2015).

Hum3 and RSP1: Some of the secreted proteins in *U. maydis* are involved in the hydrophobic surface interactions of fungal hyphae. RSP 1 (Repetitive and secreted protein) encodes a repellent-like protein with repetitive structure and Hum3 encodes hydrophobin, a protein containing both hydrophobin domain and repetitive regions. Fungal hydrophobins are the secreted proteins that help in the interaction of fungal cell walls with air or with solid surfaces. Repellants serve a similar role although they use a different mechanism and have different biochemical characteristics. However, both of them help in the formation of hydrophobic aerial hyphae in fungi (Kershaw and Talbot 1998). Single mutants of each type did not show any effect on mating or pathogenicity.

However, *rsp1 hum3* double deletion mutants showed normal dikaryotic hyphae but complete loss of pathogenicity *in planta* as shown by growth arrest at early stages after penetration. Infected plants also showed necrotic spots at the infection site suggesting that these mutants are recognized differently by the host compared to wild type. Hence these proteins could be important for *U. maydis* to evade detection by the host surveillance system and allow the proliferation of the fungus (Müller, Schreier et al. 2008).

***Ustilago hordei*:** *U. hordei*, the barley covered smut fungus, is closely related to *U. maydis* although it differs in the aspect that it can infect only at the seed germination stage, which develops later in the meristematic region until sporulation occurs in the seed heads. In contrast, *U. maydis* can infect any above ground plant parts at any plant age. *U. hordei* also differs from *U. maydis* with respect to mating type loci being bipolar in contrast to *U. maydis* with a tetrapolar mating system. Also, *U. hordei* has a much larger genome with more transposable elements (TEs) (Ali, Laurie et al. 2014).

UhAvr1 is an avirulence gene that is in a transposon and repeat-rich region and the presence of transposon activity upstream, in the promoter region, could be responsible for exhibiting virulence on some of the barley cultivars. It has a matching R gene Ruh1 in some barley cultivars and upon recognition, Ruh1 provides complete immunity to the infection by hyphal restriction and necrosis in cells early in the infection. UhAvr1 deletion mutants suggested that it is not crucial for virulence. *U. maydis* and *S. reilianum* both have an orthologous cluster for the UhAvr locus, Cluster 19A. This is the largest effector cluster, with 24 CSEPs encoded for *U. maydis* and 29 CSEPs encoded for *S. reilianum* (Grewal, Rossnagel et al. 2008, Ali, Laurie et al. 2014). Given these findings,

it will be interesting to find the target interactor for the UhAvr1 effector that allows the establishment of biotrophy in the barley host plant.

***Sporisorium reilianum*:** *S. reilianum* is a head smut fungus, that also infects maize and is closely related to *Ustilago maydis* (Begerow, Stoll et al. 2006). It has a tetrapolar mating type system, with two loci, *a* and *b*, like *U. maydis*, although the *a* locus that encodes pheromone and pheromone receptor, has three alleles and *b* locus encoding homeodomain transcription factor has five known alleles (Schirawski, Heinze et al. 2005). It infects maize and sorghum and forms spores inside the sori that eventually replace the inflorescence. These replacements are often accompanied with a complete or partial reversion of floral parts leading to abnormal phyllody formation (Semisi and Ball 1989). Infected plants also show multiple female inflorescences at lateral nodes exhibiting a loss of the apical dominance. The genome of *S. reilianum* shows high synteny with *U. maydis*, although the sequence conservation is low and differs in the gene copy number. In the same line, there are also common effector candidates in both the organisms.

SAD1: Deletion of cluster 19A in *U. maydis* resulted in the inability of the strains to produce anthocyanins and the mutant could not form tumors on the plants (Brefort, Tanaka et al. 2014). Since *S. reilianum* is closely related to *U. maydis*, when similar deletion studies were performed, it led to the identification of a candidate effector called Suppressor of Apical Dominance (SAD1), that suppresses apical dominance in infected maize host plants by inducing more subapical ears. *S. reilianum* infected maize plants produce more female inflorescences and this phenotype is also seen in transgenic *Arabidopsis* plants expressing *SAD1*, which show increased number of secondary rosette-

leaf branches with an additional phenotype of abortion of siliques at early stages of development (Drechsler, Schwinges et al. 2016). Hence, the role of SAD1 in changing the inflorescence branching is suggested. *SADI* deletion studies lead to the loss of apical dominance in infected maize plants and complementation studies with a fluorescently-tagged SAD1 protein in a *SADI* deletion mutant strain restored the original phenotype and showed that the protein is indeed secreted from the fungal hyphae in the infected tissues. Heterologous expression in *Arabidopsis* revealed its subcellular localization in cytoplasm and nucleus. Y2H screens indicated many plant interacting partners and showed that it auto-activated the reporter genes used for Y2H. This latter finding suggested its role as a transcriptional regulator. The authors reported that SAD1 increased the transcript levels of host PIN1, an auxin transporter, and downregulates the branching inhibitor TB1, which could lead to the enhanced branching effect (Ghareeb, Drechsler et al. 2015). Further experiments suggested that SAD1 might be mediating this process through hormone-independent pathways as there were no additional phenotypes detected (Drechsler, Schwinges et al. 2016). However, the exact mechanism of suppression of apical dominance is not yet known and needs further investigation.

Microbotryum violaceum sensu lato: *Microbotryum violaceum sensu lato* is a species complex, and its members are commonly called anther smuts. It was formerly named *Ustilago violacea*. The newly designated genus, *Microbotryum*, currently contains 111 species (V. Robert 2005). These are basidiomycetes and are obligate pathogens of Caryophyllaceae (Pink family) although it can also be found infecting the anthers of many other dicotyledonous plants belonging to Dipsacaceae, Lamiaceae, and Lentibulariaceae (Bauer, Begerow et al. 2006, Kemler, Göker et al. 2006). Individual

species display very high host specificity, infecting different species of plants from a single genus. These fungi sporulate in the flowers of the infected flowers and sterilize the host plant (Baker 1947).

The Lifecycle of *M. violaceum*:

The life cycle of this pathogen starts when diploid teliospores from an infected flower land on a healthy flower through wind dispersal or by pollinators (Jennersten 1983). The diploid teliospores then germinate and undergo meiosis to produce yeast-like haploid sporidia that reproduce by budding. Sexual conjugation takes place between sporidia of opposite mating type (a1 and a2), under suitable conditions, such as low nutrients and cool temperatures. In flowers, where the nutrient availability is high, conjugation takes place only when flowers wither, but on cotyledons, with low nutrient availability, it occurs shortly after teliospore germination. Conjugation results in the formation of an infectious dikaryotic hypha from one of the mated cells, that is stabilized by host cues, allowing the fungus to produce an appressorium and penetrate the host tissue. It is believed that the pathogen never invades the host through stomata and the penetration is not mediated by turgor pressure, but instead it might be mediated by lytic enzymes (Schäfer, Kemler et al. 2010). The fungus grows exclusively intercellularly and overwinters in the meristematic tissue in perennial plants; infection becomes systemic in the following year, producing diseased flowers, in which the pollen has been replaced with fungal spores, thus rendering the male plants sterile. As mentioned above, it is commonly referred to as the “anther smut” (Giraud, Yockteng et al. 2008) and thus, “castrates” its host. Travel from the infection site to the shoot meristems is believed to be mediated by xylem vessels since the fungus was detected in these vessels (Schäfer,

Kemler et al. 2010). Karyogamy occurs in the dikaryotic hyphae resulting in the formation of diploid spores, thus completing the life cycle. The fungal life cycle thus exhibits both a saprobic haploid phase and a parasitic dikaryotic/diploid phase. The disease also aborts the development of female organs in female host plants. Moreover, the female plants develop immature male reproductive anthers, making this one of the most interesting cases of parasitic modification of host floral organs. Linnaeus was the first to notice the smut-induced anthers in the female host plants (Uchida, Matsunaga et al. 2003). Since the infection is propagated via pollination by insects, it may be thought of as a sexually transmitted disease in plants (Antonovics 2005).

One well-characterized *Microbotryum* species, *Microbotryum lychnidis-dioicae*, was the first fungus to have been identified with heteromorphic sex chromosomes. The fungus has 14 chromosomes; of these, the a_1 mating-type chromosome is approximately 3.3 Mbp and that for a_2 is 4.0 Mbp. Electrophoretic karyotypes revealed that the mating type chromosomes are the largest in the genome with only one autosome with the same size range (Hood 2002). These chromosomes determine the mating type compatibility through premating pheromones/receptor factors for recognition or post-mating homeodomain proteins for compatibility (Billiard, Lopez-Villavicencio et al. 2011). It was identified that there is a higher degeneration rate in the non-recombining regions of the a_1 and a_2 chromosomes when compared to autosomes. Those regions that constitute up to 90% of the mating type chromosomes also showed the accumulation of high proportions of TE elements, with gene losses in the mating-type chromosomes (Fontanillas, Hood et al. 2015). Studies indicated that a_1 mating-type chromosome encodes 614 predicted genes, while the a_2 mating-type chromosome encodes 683, with 305 shared predicted genes

among the two chromosomes. The genes that are responsible for encoding transcriptional factors that regulate mating and fungal hyphae are present only on the a₂ chromosome (Badouin, Hood et al. 2015). This is in line with the early research findings that reported that the conjugating tube for mating is initiated mainly from this mating type (Day 1976).

Table 1- 5: *Microbotryum* pathogen species and their hosts

<i>Microbotryum</i> species	Host plant
<i>M. lagerheimii sensu lato</i>	<i>Atocion rupestre</i>
<i>M. shykoffianum</i>	<i>Dianthus pavonius</i>
<i>M. dianthorum</i>	<i>Dianthus seguieri</i>
<i>M. carthusianorum</i> <i>M. shykoffianum</i>	<i>Dianthus carthusianorum</i>
<i>M. superbum</i>	<i>Dianthus gratianopolitanus</i>
<i>M. dianthorum</i> <i>M. superbum</i>	<i>Dianthus monspessulanus</i>
<i>M. shykoffianum</i>	<i>Dianthus neglectus</i>
<i>M. carthusianorum</i>	<i>Dianthus superbus</i>
<i>M. shykoffianum</i>	<i>Dianthus sylvestris</i>
<i>M. coronariae</i>	<i>Lychnis flos-cuculi</i>
<i>M. silene-acaulis</i>	<i>Silene acaulis</i>
<i>M. aff. violaceum</i>	<i>Silene caroliniana</i>
<i>M. chloranthae-verrucosum</i>	<i>Silene chlorantha</i>
<i>M. silene-dioicae</i>	<i>Silene dioicae</i>
<i>M. violaceo-verrucosum</i>	<i>Silene italica</i>
<i>M. lychnidis-dioicae</i>	<i>Silene latifolia</i>
N/A	<i>Silene lemonii</i>
<i>M. bardanense</i>	<i>Silene moorcroftiana</i>
N/A	<i>Silene notarisii</i>

<i>M. violaceum sensu stricto</i>	<i>Silene nutans</i>
<i>M. saponariae</i>	<i>Silene ocymoides</i>
<i>M. saponariae</i>	<i>Silene officinalis</i>
<i>M. majus</i>	<i>Silene ottites</i>
<i>M. violaceum sensu lato</i>	<i>Silene paradoxa</i>
<i>M. aff. violaceum</i>	<i>Silene rupestris</i>
<i>M. silenes-saxifraga</i>	<i>Silene saxifraga</i>
<i>M. lagerheimii</i>	<i>Silene vulgaris</i>
<i>M. silenes-inflatae</i>	
<i>M. intermedium</i>	<i>Salvia pratensis</i> (Mint)
<i>M. saponariae</i>	<i>Saponaria ocymoides</i>
<i>M. scabiosae</i>	<i>Knautia arvensis</i>
<i>M. lagerheimii sensu stricto</i>	<i>Viscaria alpina</i>

N/A indicates no species designation is determined. Species designation is given referring to (Le Gac, Hood et al. 2007, Lutz, Piątek et al. 2008, Denchev, Giraud et al. 2009, Gibson, Petit et al. 2013, Fortuna, Snirc et al. 2016)

Silene latifolia

Silene latifolia, white campion, is the best-studied model for sexual dimorphism in plants. The plant genus *Silene*, was studied by Darwin, Mendel and several other scientists in ecology, evolutionary, and developmental studies (Bernasconi, Antonovics et al. 2009). *S. latifolia* has been extensively studied to address many questions related to sexual and mating systems, the evolution of sex chromosomes, epigenetics, speciation, disease ecology and evolution, and biological invasions. It is a dioecious plant, with male and female flowers on different plants. The Y chromosome of *S. latifolia* is recently evolved

and provides a useful tool to study the evolutionary processes leading to the loss of functional genes from its Y chromosome as compared to more distantly evolved Y chromosome, like that of humans. Its genome is very large and is highly enriched with repetitive content (Bernasconi, Antonovics et al. 2009). A study on 40,000 herbarium specimens collected from natural populations revealed that the *Microbotryum* fungi are mainly confined to the perennials rather than annuals. The life cycle of anther smuts could be responsible for their absence in annuals in nature (Gibson, Petit et al. 2013). Male plants (XY) produce many flowers that are short-lived, while females (XX) produce few flowers that remain accessible for several days on the plant (Kaltz and Shykoff 2001). Moreover, infected flowers have been known to last longer than the uninfected flowers, thus leading to a prediction that the fungus may promote the strengthening of the flower base and filaments of the stamens in infected flowers. This is particularly noticeable when the flowers show a crumpled morphology as they age, but have a well maintained center portion and infected stamens (Uchida, Matsunaga et al. 2003). The mechanism of how the female flowers produce stamens after infection remains unknown. The male flowers have 10 mature stamens and the female flowers have five styles. This is believed to be caused by the three sex determination factors on the Y-chromosome- Gynoecium suppressing factor (GSF), Stamen promoting factor (SPF) and male fertility factor for maturation of anthers in the male plants (Westergaard 1958). The Y-chromosome seems to carry genes for suppression of female development and in infected female plants the fungus partially substitutes for the Y-chromosomal genes that are absent in the female plants although this replacement is not complete as evident by the retention of calyx in the infected female plants (Uchida, Matsunaga et al. 2003).

Recent host shifts were shown by *Microbotryum* from *Silene latifolia*, white campion to *S. vulgaris*, bladder campion. Since the disease has no agricultural or economic impact, it serves as an ideal model to study both wild populations and the possible effects of inbreeding in agricultural crops. Studying this model system may also provide knowledge about future diseases caused by host shifts, possibly affecting humans, like HIV/AIDS (Antonovics, Hood et al. 2002).

Since *Microbotryum-Silene* has been established as a model system for studying host shifts (Antonovics, Hood et al. 2002), epidemiology of diseases (Antonovics J 1997) and for the evolution of sexually transmitted diseases (Nunn, Gittleman et al. 2000), understanding the molecular mechanisms behind pathogenicity is crucial for understanding host shifts and how the pathogens evolve in wild populations.

The genome sequence of *M. lychnidis-dioicae* and transcriptomes from its interaction with the host *S. latifolia* have been produced (Perlin, Amselem et al. 2015, Toh, Chen et al. 2017). The 25.2 Mbp genome of an a₁ mating type strain was sequenced using 454 technology and revealed that 14% of the genome consists of repetitive sequences with the accumulation of transposon elements (TEs) in mating-type chromosomes. The work further identified more than 300 genes linked with the a₁ mating type. The analysis of the genome sequence highlighted the expansion of secretory lipases and significant induction in the transporters that may be required for the intake of necessary components from the host cell. Further analysis also indicated the presence of carbohydrate-active enzymes (CAZymes) that might be involved in host cell degradation, but also the retention of enzymes that break down components of pollen tubes in flowers, in line with its infection location. The study predicted a total of 7,364 protein-coding genes, with 279 secreted

proteins (SPs). Among the 279, seventy-one SPs were smaller than 250 amino acids, hereafter called small secreted proteins, SSPs. Forty-six SSPs were unique to *Microbotryum* with no sequence similarity with any other known proteins and 19 of the SSPs were significantly upregulated during infection (Perlin, Amselem et al. 2015). Thus, such SSPs have the characteristics of fungal effectors and are therefore worthy of further investigation. These results have provided the opportunity for further study of the putative effector proteins at the molecular level and investigation of their role(s) in pathogenesis.

Overview of the dissertation

To shed light on the mechanism of infection of *Microbotryum* on *Silene* host plants, I attempted to identify and characterize for the first time the effectors of this obligate pathogen. In this dissertation, I provide the first study to examine the function of a subset of such candidate proteins, *i.e.*, the putative effectors that are involved in the pathogenicity of this group of fungi. The goal of this study was to deepen the knowledge that is currently available on the host-pathogen interactions of *Microbotryum* and its host based on a study of the genes that are expressed during the infection stage of the life cycle. The research on identification and initial characterization of the effectors has been published (Kuppireddy, Uversky et al. 2017) and is presented in Chapter 2 of this dissertation. I also showed experimentally that these effectors are secreted using an *in vitro* experimental model called yeast secretion trap. To identify the potential host interactors, I utilized yeast two-hybrid screens, and this led to the identification of three host targets for two *M. lychnidis-dioicae* effectors. The identification and characterization

of the effectors and their host targets will permit investigation into the molecular basis for the evolution of this fascinating fungal pathogen and its host adaptation.

Molecular understanding of effectors and their interacting host partners is essential for unraveling plant-pathogen interactions. To further characterize the function and subcellular localization of the effectors and their role in the pathogenicity, I expressed them in a heterologous system. To our knowledge, this is also the first study to express a transgene of *Microbotryum lychnidis-dioicae* (or any other *Microbotryum* species) in *Arabidopsis* plants to understand the phenotypic anomalies caused by the effector and also, its subcellular localization. I generated the stable transgenic *Arabidopsis* lines, and the results are reported in Chapter 3.

I have also overexpressed the fungal effectors via mutant *M. lychnidis-dioicae* strains in the native host plant, *Silene latifolia*, to identify novel interactors if any, and to confirm the current known interactors *in planta* by using Co-IP and mass spectrometry analysis (MALDI-TOF); this work is documented in Chapter 4. The aim of this study was to use the *Microbotryum-Silene* model system to investigate and functionally characterize the biological roles of the selected *M. lychnidis-dioicae* effector candidates in plants.

In the final chapter, Chapter 5, the overall conclusions from this dissertation are provided. In summary, the studies presented provide novel insights into the mechanism of infection of *Microbotryum* in its host plants by studying its effectors, their subcellular localization, and their targets in the host cell, allowing better understanding of the mechanisms used to remodel the host defense mechanism for its own benefit. There is a need to characterize the rest of the effector proteins and I hope that this dissertation provides the experimental

strategies that could be expanded in the future to provide mechanistic insights into the interplay of this biotrophic pathogen with its host.

CHAPTER II
IDENTIFICATION AND INITIAL CHARACTERIZATION OF THE EFFECTORS OF
AN ANTHER SMUT FUNGUS AND POTENTIAL HOST TARGET PROTEINS

Overview

Plant pathogenic fungi often display high levels of host specificity and biotrophic fungi; in particular, must manipulate their hosts to avoid detection and complete their obligate pathogenic lifecycles. One important strategy of such fungi is the secretion of small proteins that serve as effectors in this process. *Microbotryum violaceum* is a species complex whose members infect members of the Caryophyllaceae; *M. lychnidis-dioicae*, a parasite on *Silene latifolia*, is one of the best studied interactions. We are interested in identifying and characterizing effectors of the fungus and possible corresponding host targets. *In silico* analysis of the *M. lychnidis-dioicae* genome and transcriptomes allowed us to predict a pool of small secreted proteins (SSPs) with the hallmarks of effectors, including a lack of conserved protein family (PFAM) domains and also localized regions of disorder. Putative SSPs were tested for secretion using a yeast secretion trap method.

¹ Kuppireddy, V. S., V. N. Uversky, S. S. Toh, M. C. Tsai, W. C. Beckerson, C. Cahill, B. Carman and M. H. Perlin (2017). "Identification and Initial Characterization of the Effectors of an Anther Smut Fungus and Potential Host Target Proteins." *Int J Mol Sci* **18**(11). doi: 10.3390/ijms18112489.

We then used yeast two-hybrid analyses for candidate-secreted effectors to probe a cDNA library from a range of growth conditions of the fungus, including infected plants. Roughly 50 SSPs were identified by in silico analysis. Of these, 4 were studied further and shown to be secreted, as well as examined for potential host interactors. One of the putative effectors, MVLG_01732, was found to interact with orthologues of the *Arabidopsis thaliana* calcium-dependent lipid binding protein (AtCLB) and with cellulose synthase interactive protein 1 (CS11). The identification of a pool of putative effectors provides a resource for functional characterization of fungal proteins that mediate the delicate interaction between pathogen and host. The candidate targets of effectors, *e.g.*, AtCLB, involved in pollen germination suggest tantalizing insights that could drive future studies.

Introduction

During fungal infection of plants, a number of fungi secrete small proteins that serve to manipulate host responses and downstream events in host development during infection. Often such proteins allow biotrophic fungi to evade host defenses, but they can also redirect development so as to specifically benefit the fungus. Such proteins have been termed “effectors”, and many share common characteristics among different fungi [1,2]. For instance, for oomycete pathogens, such as *Phytophthora* species, or rust species (*e.g.*, *Melampsora lini*), small secreted proteins (SSPs) are secreted from specialized structures called haustoria that penetrate host plant cells to draw nutrients from their hosts. Such fungal effectors are SSPs that bear an N-terminal signal peptide; the effectors are usually unique to the pathogen. Most effectors are cysteine-rich and share no sequence similarity with other known proteins, thus revealing the specialized arsenal that each pathogen

possesses and, most likely, their association with the specificity of the pathogen for its host. Some effectors are translocated directly from the infection structures, *i.e.*, haustoria or appressoria, into plant cells, while others interact with host cell receptors and get internalized into the cell [1]. Some studies suggest that secreted proteins can act as structural effectors that could accumulate at the host/pathogen interface and stabilize the fungal filaments [2]. However, the mechanism of how these effectors work in the entry into the plant cell or in the proliferation of the fungus inside the host has yet to be fully elucidated. *Microbotryum lychnidis-dioicae* is an obligate biotrophic basidiomycete smut fungus and is a member of the *Microbotryum violaceum* species complex that infects members of the Caryophyllaceae family. *M. lychnidis-dioicae* infects the dioecious host plant, *Silene latifolia*. The fungal life cycle begins when the fungal spores are disseminated by wind or pollinator species and land on a suitable host. The diploid teliospores then undergo meiosis to produce yeast-like haploid sporidia that reproduce by budding. Conjugation takes place between sporidia of opposite mating type, under suitable conditions, such as low nutrients and cool temperatures. Conjugation results in the formation of an infectious dikaryotic hypha that is stabilized by host cues, allowing the fungus to produce an appressorium and penetrate the host tissue. The fungus overwinters in the meristematic tissue; infection becomes systemic in the following year, producing diseased flowers, in which the pollen has been replaced with fungal spores, thus rendering the male plants sterile. It is thus commonly referred to as the “anther smut” [3]. Karyogamy occurs in the dikaryotic hyphae resulting in the formation of diploid spores, thus completing the life cycle. The fungal life cycle thus exhibits both a saprobic haploid phase and a parasitic dikaryotic/diploid phase. The disease also aborts

the development of female organs in female host plants. Moreover, the female plants develop immature male reproductive anthers, making this one of the most interesting cases of parasitic modification of host floral organs. Linnaeus was the first to notice the smut-induced anthers in the female host plants [4]. Since pollination drives disease transmission, anther smut is considered as a plant sexually transmitted disease (STD) [5].

Recently, the genome sequence and transcriptomes of *M. lychnidis-dioicae* and its interaction with the host *S. latifolia* have been produced [6]. However, there have been no experimental data provided to explain how this fungus can divert the host resources for its own propagation and survival. Here, we provide the first study to examine the function of the candidate proteins, *i.e.*, the putative effectors that might be involved in the pathogenicity of this group of fungi.

Results

***In Silico* Analyses to Identify Potential Effectors**

To provide a conservative estimate of proteins secreted by *Microbotryum lychnidis-dioicae*, several bioinformatic tools were employed, and only those proteins that passed all measures used were retained in the list of predicted secreted proteins (Table S1, Appendix). Out of 7364 proteins, 279 were identified to have a signal peptide; from this group, 71 predicted proteins were smaller than 250 amino acids (hereafter referred to as small secreted proteins, SSPs). Of these, 46 appeared to be unique to *M. lychnidis-dioicae* or to the *Microbotryum* complex, and 60 lacked identifiable PFAM domains. Among the SSPs, 19 were also significantly upregulated during plant infection, suggesting that these may play a role during those stages of the fungal lifecycle and in pathogenicity [7].

Intrinsic Disorder in Predicted Small Secreted Proteins (SSPs)

Intrinsic disorder is known to play an important role in protein-protein interactions [8–14]; intrinsically disordered proteins (IDPs), hybrid proteins containing ordered domains, and intrinsically disordered protein regions (IDPRs) are common among pathogenic microbes [15] and play a number of roles in pathogen-host interactions [16,17].

Accordingly, we analyzed the overall intrinsic disorder predisposition of the 49 predicted secreted proteins from *M. lychnidis-dioicae* upregulated during infection, using a set of established disorder predictors from the PONDR family (PONDR[®] VSL2 [18], PONDR[®] VLXT [19], PONDR[®] VL3 [20], and PONDR[®] FIT [21]). We also used the ANCHOR algorithm [22,23] to evaluate the presence of the disorder-based protein-protein interaction sites, molecular recognition features (MoRFs), *i.e.*, regions that might undergo the binding-induced disorder-to-order transition. Results of these analyses are summarized in Supplementary Materials, Table S2, Appendix. These results draw a picture of an impressive prevalence of intrinsic disorder in the *M. lychnidis-dioicae* SSPs. In fact, all putative effectors have regions of intrinsic disorder, and many of the effectors are very disordered. In particular, 11 effectors (22.4%) can be classified as mostly disordered, since they have >50% disordered residues; 19 effectors (38.8%) are highly disordered, possessing between 30 and 50% of disordered residues; 18 effectors (36.7%) are moderately disordered, since they have between 10 and 30% disordered residues; and just one protein (2.1%) has less than 10% disordered residues and therefore is mostly ordered. These values for disorder content are very high even for a eukaryotic organism and are rather atypical for groups of proteins that are not specifically selected for disorder. Furthermore, many effectors have disorder-based binding sites or MoRFs (*i.e.*,

sites that are disordered in the unbound state and undergo disorder-to-order transition at interaction with the binding partners). Finally, several effectors have more than one MoRF, suggesting that they can be engaged in interaction with multiple partners or, being engaged in interaction with one partner, utilize multivalent “wrapping around”-type binding mode. It is likely that the exceptionally high disorder levels and the presence of MoRFs can simplify interactions of these pathogenic effectors with host proteins or play some other role in regulation of the SSP functionality.

In line with the hypothesis that intrinsic disorder can be of functional importance for the SSPs from *M. lychnidis-dioicae*, Figure 2-1 represents in-depth analysis of the intrinsic disorder predisposition of four putative effector proteins that were up-regulated during infection and were shown to have important functions (Sections 2.3 and 2.4 for the detailed functional characterization of these proteins).

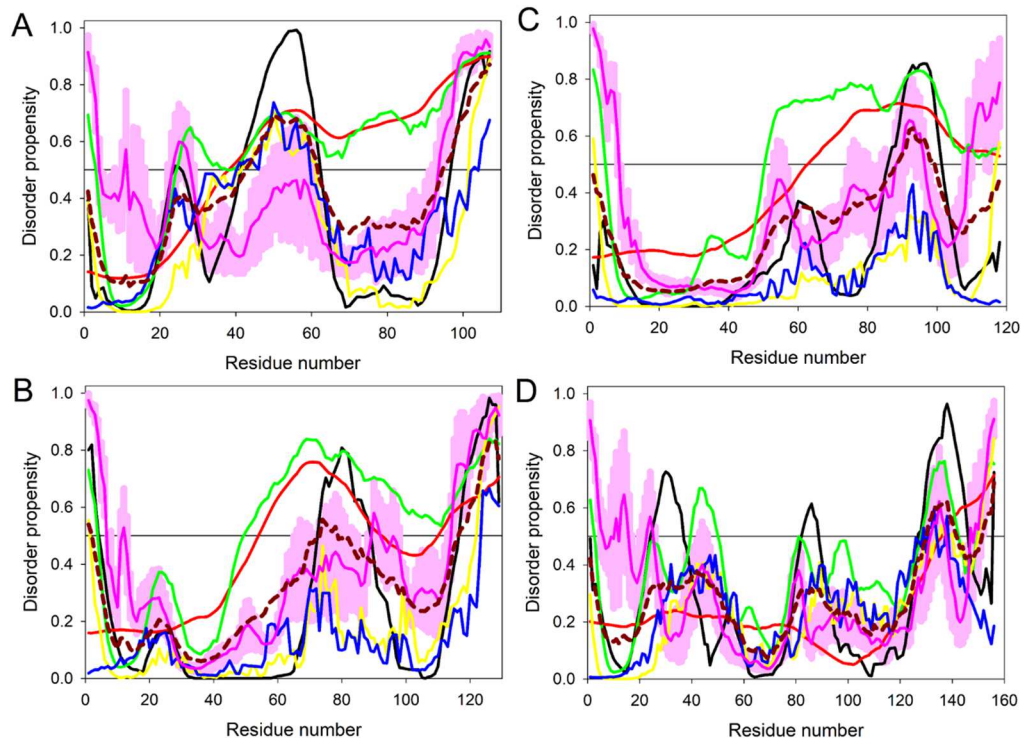


Figure 2- 1: Evaluating intrinsic disorder propensity of protein effectors

(A) MVLG_04106; (B) MVLG_05720; (C) MVLG_06175 and (D) MVLG_01732 by a series of per-residue disorder predictors. Disorder profiles generated by PONDR[®] VLXT, PONDR[®] VL3, PONDR[®] VSL2, IUPred_short, IUPred_long, and PONDR[®] FIT, are shown by black, red, green, yellow, blue, and pink lines, respectively. Dark red dashed line shows the mean disorder propensity calculated by averaging disorder profiles of individual predictors. Light pink shadow around the PONDR[®] FIT shows error distribution. In these analyses, the predicted intrinsic disorder scores above 0.5 are considered to correspond to the disordered residues/regions, whereas regions with the disorder scores between 0.2 and 0.5 are considered flexible.

The corresponding disorder profiles were generated by the overlay of the outputs of six commonly used disorder predictors, PONDR[®] VSL2 [18], PONDR[®] VLXT [19], PONDR[®] VL3 [20], PONDR[®] FIT [21], as well as IUPred_short and IUPred_long [24]. Furthermore, for each of these four proteins, mean per-residue disorder probability was calculated by averaging disorder profiles generated by the individual predictors. The use of consensus for evaluation of intrinsic disorder is motivated by empirical observations that this approach usually increases the predictive performance compared to using a single predictor [25–27]. Figure 1 clearly shows that these four proteins are characterized by high levels of predicted disorder that range (as per the outputs of PONDR[®] VSL2 analysis) from 22.4% in MVLG_01732 to 61.0% in MVLG_06175, to 64.3% in MVLG_05720, and to 79.4% in MVLG_04106. According to the PONDR[®] VSL2-based analysis, there are four IDPRs in MVLG_01732 (residues 1–2, 40–50, 128–142 and 150–156) and three IDPRs in MVLG_04106 (residues 1–3, 25–37, and 39–107), whereas

MVLG_06175 and MVLG_05720 have two IDPRs each (residues 1–4 and 51–118 and residues 1–3 and 50–129, respectively). Furthermore, according to the ANCHOR analysis, each of these four SSPs might have at least one MoRF (residues 145–153 in MVLG_01732, residues 113–118 in MVLG_06175, residues 7–12 in MVLG_05720, and residues 3–12 in MVLG_04106). The presence of MoRFs in these proteins was also analyzed by MoRF_{CHiBi}, which is a new computational approach for fast and accurate prediction of MoRFs in protein sequences. This analysis showed that although there is no MoRF_{CHiBi}-identified MoRF in MVLG_01732, this protein has two regions with some potential to act as MoRFs (residues 29–39 and 139–156). Similarly, there are no MoRF_{CHiBi}-identified MoRFs in MVLG_05720, which, however, have four regions with some potential to act as MoRFs (residues 1–14, 66–76, 97–104, and 120–129). On the other hand, MVLG_06175 has two MoRFs (residues 97–111 and 113–118), and almost the entire chain of MVLG_04106 can act as disorder-based binding region, since this protein has two MoRFs, residues 1–70 and 87–104, that cover almost 83% of its sequence.

Yeast Secretion Trap to Verify the Secretory Nature of Predicted Effectors

We used Yeast Secretion Trap (YST) [28], a molecular genetic approach, to confirm the secretory nature of a small subset of the SSP putative effector proteins (MVLG_01732, MVLG_04106, MVLG_05720, and MVLG_06175; Table 2-1), each of which was also up-regulated during infection. Three of these proteins were also Cys-rich (MVLG_04106, MVLG_05720, and MVLG_06175), another hallmark of effectors in a number of fungal species [29]. YST employs a mutant strain of yeast, SEY 6210, that has a deletion in the *SUC2* locus encoding the enzyme, invertase. Invertase catalyzes hydrolysis of the

disaccharide, sucrose, to glucose and fructose, so that the yeast cell can then take up glucose and metabolize this sugar. Thus, the SEY 6210 mutant yeast strain is normally unable to grow on media where sucrose is the sole carbon source. The method uses a vector, pYSTO-0, bearing the coding region of Suc2 invertase without its signal peptide and its start codon. The protein of interest can be cloned as a translational fusion protein with the invertase driven by a constitutive promoter from *ADHI*. If the protein of interest is secreted, this will result in the reconstituted functional activity of the invertase and enable the yeast cells to grow on sucrose medium. All four predicted effectors from *M. lychnidis-dioicae* examined experimentally with the yeast secretion trap assay indeed appeared to be secreted, since the signal peptide of each allowed Suc2p to be secreted and thus provide for growth of the yeast SEY 6210 mutant on sucrose medium (Figure 2-2). In contrast, SEY 6210 cells transformed with the vector only were unable to grow on such media.

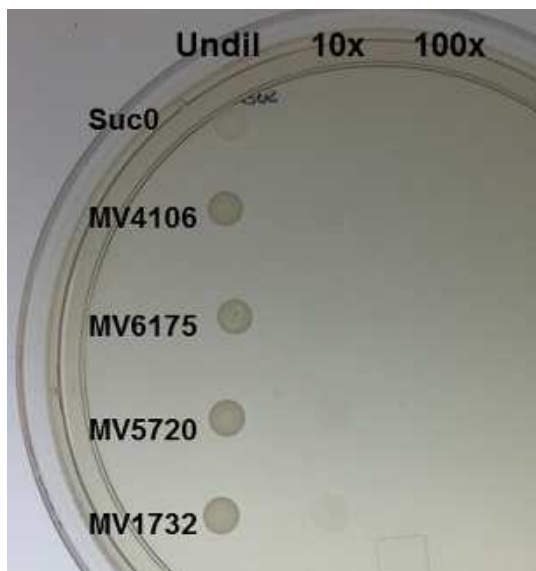


Figure 2- 2: Results of secretion trap experiment with four *M. lychnidis-dioicae* predicted small secreted proteins (SSP) effectors.

Suc0, yeast cells transformed with vector alone on sucrose, leu drop-out medium. Undil, undiluted; 10× and 100× dilutions.

Table 2- 1: Candidate SSPs chosen for further analyses.

Predicted Protein	Expression ^a	Size (Amino Acids)	no. of Cys	Function
MVLG_01732	144 rsem vs. 0	156	1	Candidate effector
MVLG_04106	86 rsem vs. 0	107	6	Candidate effector
MVLG_05720	1164 rsem vs. 0	129	12	Candidate effector
MVLG_06175	127 rsem vs. 0	118	10	Candidate effector

^a rsem normalized counts for infected male *S. latifolia* vs. expression in YPD or nutrient-limited agar [7].

Yeast Two-Hybrid Experiment

Our goal was to determine the function of these fungal proteins that are predicted, and now confirmed, to be secreted, as well as being highly expressed, during infection. We employed yeast two-hybrid genetic screening to identify the possible host interactors for these fungal proteins. As mentioned above, we chose a small subset of the SSPs that were also found to be induced in expression *in planta*.

MVLG_04106 Autoactivates the Reporter Genes in Yeast Two-Hybrid Assay

We expressed MVLG_04106 lacking its signal peptide as a fusion protein to Gal4BD in the bait vector (pGBKT7-MVLG_04106ΔSP) and tested its activity in expressing the

reporter genes. It was found that the yeast strain transformed with this construct activated all three of the reporter genes-*HIS3*, *ADE2*, and *MEL1*, when mated with the opposite mating strain containing only the control prey vector (Figure 2-3). This indicates the cells' ability to grow on media lacking the essential nutrients histidine and adenine because of the activation of the enzymes aminoimidazole ribonucleotide carboxylase 2 (*ADE2*) and imidazole glycerol phosphate dehydratase 3 (*HIS3*). Moreover, the cells were also able to express α -galactosidase, the gene product of the Melibiase 1 (*MEL1*) reporter gene that enables the yeast cells to turn blue-green in the presence of the chromogenic substrate X- α -gal. This was unexpected, so we generated the reciprocal set of constructs to further investigate possible transcriptional activation by MVLG_04106. In this case, a fusion protein was generated with MVLG_04106 and Gal4AD in the prey vector to test if the reporter genes could again be activated. Surprisingly, in this case the reporter genes were not activated. This suggests that MVLG_04106 could activate the transcription of the reporter genes only when attached to the corresponding DNA binding domain for those genes (*i.e.*, Gal4BD). One possibility is that this fungal protein acts as a transcription factor in modulating the host gene expression during infection. In line with the known fact that transcription factors are typically characterized by high levels of intrinsic disorder [30–32], MVLG_04106 was predicted to possess 79.4% disordered residues (see Figure 1A) and is shown to contain long disorder-based interaction regions. The predicted protein contains 107 amino acid residues and is cysteine rich, with approximately 5% Cys residues. Further domain analysis using PROSITE did not yield any information, but prediction of post translational modification sites indicated proteolytic cleavage at residue D35, which could allow the mature protein to function as

a transcriptional regulator [33] (Supplemental Table S2, Appendix). Structural modelling using Swiss-Model yielded chorismite mutase for residues 30–65, for which there was 22.22% similarity in the 3-dimensional structure. When we compared the amino acid sequence of MVLG_04106 with predicted proteins of *M. silenes-dioicae* [34], there was 99.07% identity with the corresponding orthologue, whereas that for the *M. violaceum sensu lato* species [35], only had 63.04% identity.

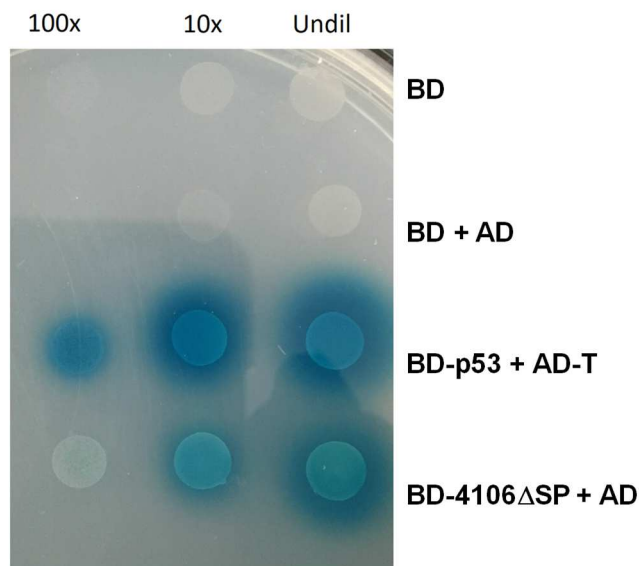


Figure 2- 3: Autoactivation of three reporter genes by MVLG_04106 on QDO/X- α -gal + 3-AT (5 mM) plates.

Undil, undiluted; 10 \times and 100 \times dilutions. QDO (Quadruple drop out media), 3-AT (3-Amino-1,2,4-triazole), BD (DNA binding domain in pGBKT7 vector), AD (Activation domain in pGADT7 vector), BD-p53 (pGBKT7-53 positive control plasmid), AD-T (pGADT7-T positive control plasmid), and BD-4106 Δ SP (MVLG_04106 lacking signal peptide).

MVLG_05720 Fungal Protein Interacts with Fungal Proteins

Yeast two-hybrid screening with MVLG_05720 yielded 614 colonies after the initial stringent selection on QDO medium with 5 mM 3AT, along with screening for α -galactosidase expression on X- α -gal (as blue-green colonies). Further selection on 50 mM 3AT to reduce leaky *HIS* selection yielded 129 colonies for examination via sequence analysis. Of the 129 sequenced clones, we recovered only fungal interactors: 99 of the clones represented MVLG_07305, 27 of the clones were found to be MVLG_04206, and 3 of the clones matched MVLG_04267. Figure 1B illustrates that there are 64.3% disordered residues in MVLG_05720, and this protein has several MoRFs. It contains 129 amino acids and is highly cysteine rich with roughly 9% Cys residues. When the amino acid sequence of MVLG_05720 was compared to the genomes of *M. silenes-dioicae* and *M. violaceum sensu lato*, the corresponding orthologues showed 96.9% identity and 85.93% identity, respectively.

MVLG_06175 Interacts with a Host Protein and a Fungal Protein

Yeast two-hybrid screening with MVLG_06175 initially yielded 1000 colonies after the stringent selection on QDO/X- α -gal + 3AT (5 mM) medium. Further selection to reduce leaky *HIS* selection yielded 201 colonies for examination via sequence analysis. Of the 39 sequenced clones that we recovered, 4 of them were full length clones that encode CASPL2C1, 1 was a fungal protein encoded by MVLG_06379, and the rest of the sequenced clones were for the fungal protein encoded by MVLG_07305 mentioned above. The *S. latifolia* genomic region matching the sequence for CASPL2C1 is found on contig m.88187 (GenBank: FMHP01040264.1) in NCBI for the *Silene latifolia* genome assembly (taxid:37657). It also corresponded to c93454_g1 RNA detected in RNA-Seq

experiments [6,7]. According to Figure 1C, 61.0% of residues in MVLG_06175 are predicted to be intrinsically disordered and this protein can be engaged in disorder-based protein-protein interactions. It contains 118 amino acids and is Cys-rich (roughly 8% Cys residues). When the amino acid sequence of MVLG_6175 was compared to the genomes of *M. silenes-dioicae* and *M. violaceum sensu lato*, there was 95.76%, but only 59.83% identity, respectively, with the corresponding orthologues.

MVLG_01732 Interacts with Host Proteins

Yeast two-hybrid screening with MVLG_01732 yielded 401 colonies after the initial stringent selection on QDO/X- α -gal + 3AT (5 mM) medium. Further selection to reduce leaky *HIS* selection yielded 65 colonies for examination via sequence analysis. From the 65 sequenced clones, yeast two-hybrid screening of MVLG_01732 revealed interesting host plant interactors. One of the interactors, represented by 52 clones, was found from blastp searches of the *Arabidopsis thaliana* genome (TAIR; <https://www.arabidopsis.org/>) as an orthologue of the AT3G61050.2 gene, which encodes a calcium-dependent lipid binding protein (AtCLB). AtCLB has both coiled coil regions and C2 domain similar to synaptotagmins, and synaptotagmins were also identified as hits in blastp searches of the NCBI database. Synaptotagmins are class of proteins with an N terminal transmembrane and two cytoplasmic C₂ domains (Figure S1, Appendix). The *S. latifolia* genomic region matching the AtCLB sequence is found on contig m.108787 (GenBank: FMHP01019528.1) in the *S. latifolia* genome assembly. It also corresponded to c85332_g3 RNA detected in RNA-Seq experiments [6,7].

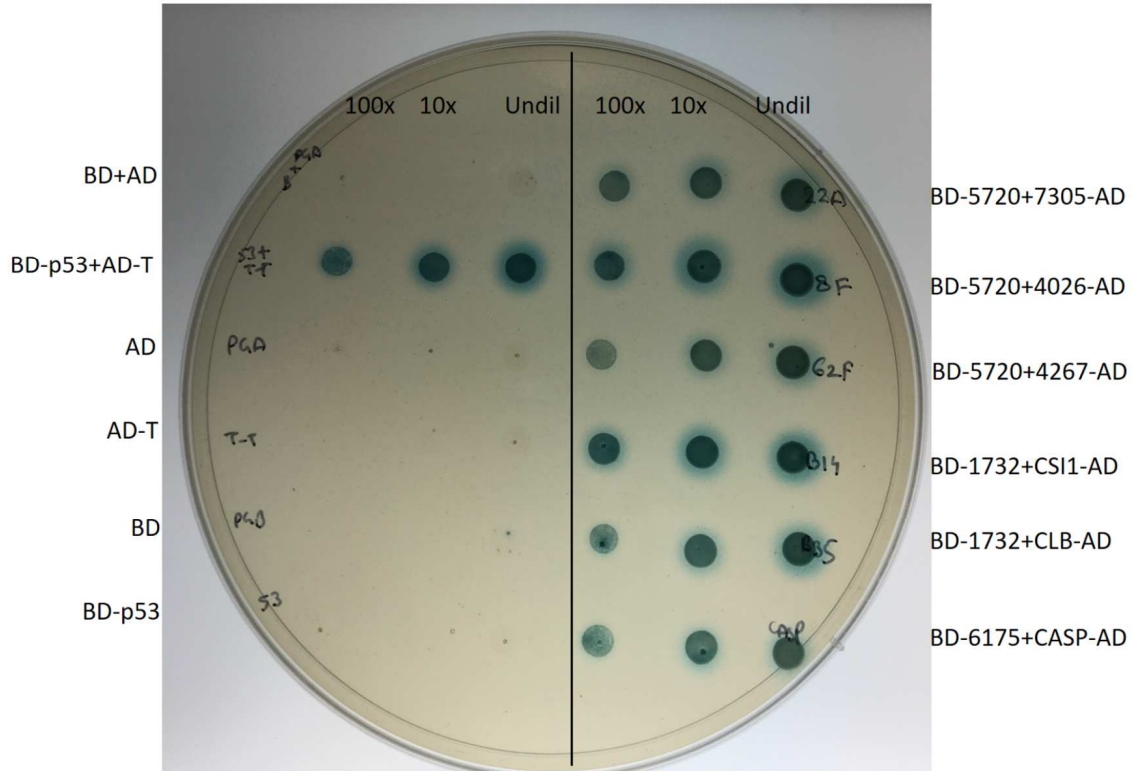


Figure 2- 4: Yeast two-hybrid spot test results.

Yeast two-hybrid spot test results for all four proteins and positive and negative controls on quadruple drop-out medium (QDO)/X- α -gal + 3AT (5 mM) plates. For the spot test, each strain bearing the plasmid was grown in 3 ml of appropriate drop out liquid media at 30 °C and shaken at 280 rpm for 2 days. To reconfirm the interaction, 10 ml of each culture (washed and resuspended in 0.9% w/v NaCl) was mixed and spotted on QDO plates containing X- α -gal at the indicated dilutions and incubated at 30 °C for 3–5 days. Undil, undiluted; 10 \times and 100 \times dilutions.

The other interactor identified by yeast two-hybrid was cellulose synthase Interactive protein 1 (CSI1; Figure S2, Appendix), represented by 13 clones. The *S. latifolia* genomic region matching this sequence is found on contig m.23209 (GenBank:

FMHP01009449.1) in the *S. latifolia* genome assembly. It also corresponded to c93789_g3 RNA detected in RNA-Seq experiments [6,7]. Figure 2-1 (D) shows that with 22.4% disordered residues, MVLG_01732 is the least disordered protein analyzed in this study. However, despite relatively low disorder content, MVLG_01732 contains MoRFs and, therefore, is expected to use intrinsic disorder for protein-protein interactions. The protein is 156 amino acids long and is not rich in Cys residues. When the amino acid sequence of MVLG_1732 was compared to the genome of *M. silenes-dioicae*, a 94.23% identity match was found in the corresponding orthologue, whereas only a 48.99% identity match was observed for the orthologue from *M. violaceum sensu lato*.

Discussion

In this study, we were able to predict from *in silico* analyses a conservative estimate of the secretome of *M. lychnidis-dioicae*. Furthermore, among this group, we identified candidate effectors as SSPs that were also highly expressed during plant infection. For the four putative effectors examined in greater detail in this study, amino acid sequence comparisons between *M. lychnidis-dioicae* and *M. silenes-dioicae* [34] revealed that these two organisms share close similarity in their predicted SSPs. In contrast, comparisons of most of the orthologues identified in *M. violaceum sensu lato* [35], the species that infects *Silene paradoxa*, had significantly lower amino acid similarities to those of the other species. These findings suggest that the latter organism has diverged substantially from the other two species, a finding supported by the phylogenetic relationships of the three respective fungal species [36] and the lack of cross-infectivity for *M. violaceum sensu lato* on either *S. latifolia* or *S. dioicae*; similarly, neither *M. lychnidis-dioicae* nor *M. silenes-dioicae* have been found to infect *S. paradoxa*. Of note, all SSPs were shown to

contain IDPRs, and the vast majority of these proteins (>61.0%) were classified as mostly or highly disordered. Many SSPs were also predicted to have at least one MoRF, with some of the putative effectors possessing multiple MoRFs that can be utilized in promiscuous interactions with the fungal and host proteins. To test some of these predictions, a subset of the SSPs predicted in silico were confirmed to be secreted by YST experiment. We conducted yeast two-hybrid analysis for these SSPs to identify their host interactors and hence to understand their role in the mechanism of the infection (Figure 2-4).

Studies suggest that the proteins that undergo post translational modifications (PTMs) are considered to interact more with other proteins by engaging in more physical contacts and are known to be in the central network pathways more than non-PTM proteins. For example, all but four of the 49 secreted proteins we examined in detail were predicted to be targets for amidation. C-terminal amidation has been shown to be involved in membrane interactions for some proteins. In one case, an antimicrobial peptide, maximin H5, was able to penetrate and lyse erythrocyte membranes when amidated, but the ability to penetrate lipid membranes was severely reduced with deamidated peptide [37]. For the 4 SSPs, we examined in detail by yeast two-hybrid analysis, MVLG_04106 and MVLG_05720, which were predicted to have amidation targets (but, not at the C-terminus), while MVLG_06175 and MVLG_01732 each have a predicted target closer to the C terminus. If amidation plays a similar role for these SSPs as it does for maximin H5, this could indicate that these effectors penetrate host cells as part of their normal function.

MVLG_04106 Could Serve as a Transcriptional Regulator

The finding that MVLG_04106 was able to autoactivate all the reporter genes—*HIS3*, *ADE2*, and *MEL1*—in the yeast two-hybrid screen suggests its role as a transcriptional regulator. However, analysis by structural modelling reveals that a portion of this protein is similar to chorismate mutase, a vital enzyme that catalyzes the conversion of chorismate to prephenate in the shikimate pathway, leading to the production of aromatic amino acids, phenylalanine, and tyrosine, and regulating their balance. Chorismate also serves as a substrate for the production of salicylic acid (SA), which is a major signaling defense molecule in plants. This fungal protein chorismate mutase could deviate the flow of available chorismate for the production of prephenate and hence channel down its availability for SA production. In fact, studies show that *Ustilago maydis*, an obligate biotrophic pathogen that causes corn smut, also secretes an effector called Cmu1, a chorismate mutase taken up by plant cells and spread to adjacent cells causing metabolic priming in the infected cells [38]. However, if MVLG_04106 is a chorismate mutase, this still begs the question of how it autoactivates the reporter genes in yeast two-hybrid assay. Transcriptome analysis revealed that it is highly expressed during infection but not under in vitro conditions, all of which suggests its role in pathogenicity. Thus, further investigation is required to better define its true role during infection.

MVLG_05720 Possibly Regulated by Additional Fungal Proteins

Three fungal proteins were identified as interactors with MVLG_05720: MVLG_07305, MVLG_04026, and MVLG_04267. None of the three were predicted via bioinformatic tools to be secreted. The differential expression data [6,7] indicated that MVLG_07305 was downregulated in late infection stages *in planta* and upregulated in mated conditions

in vitro (while MVLG_05720 was downregulated during mating) [6]. Thus, MVLG_07305 may play some role in mating or in the transition to dikaryotic filaments. The gene is located on the mating-type chromosome, but its expression is similar in both a1 and a2 mating-type strains on either rich or nutrient-limited media [7]. Blastp predicted its function as a putative fimbrial outer membrane usher protein, containing a mannose binding domain. Of note, fimbrial appendages were first observed serendipitously on the haploid cells of an anther smut fungus [39]. They are involved in cell-to-cell communication and adhesion during mating before pathogenesis, as enzymatic and mechanical removal of these structures were shown to delay mating until the regeneration of fimbriae occurred [40,41].

The second fungal interactor, MVLG_04026, followed the same expression pattern as that of MVLG_07305; its predicted function was as a Fibrillin-like protein. Fibrillins are secreted proteins that constitute the backbone of extracellular macromolecular microfibrils [42]. The C terminus of fibrillins can undergo multimerization as a consequence of intermolecular disulfide bonding with itself or other proteins soon after secretion [43]. However, MVLG_04026 was predicted via bioinformatic tools not to be secreted. Its transcription was also upregulated during mating and downregulated during infection.

MVLG_04267 was not found to be differentially expressed under any of the conditions examined. It belongs to the DUF1212 superfamily, a class of membrane proteins with unknown function. Perhaps this protein plays a role in transport of MVLG_05720. If the MVLG_07305 and 4026 proteins are translated during mating and persist during infection, they may interact with MVLG_05720, to sequester it until it is needed for

manipulation of the host. However, the mechanism of their action remains a mystery and requires further investigation.

MVLG_06175 Role in Host Entry During the Infection and in Reproduction

From yeast two-hybrid screening, the protein product of fungal gene *MVLG_06175* interacts with a host CASP-like protein 2C1 orthologue of *Spinacia oleracea* (LOC110788005), transcript mRNA (XM_021992637.1); it also matched CASP-like protein 2C1, AT4G25830.1 of *A. thaliana*. The corresponding transcript from *Silene* expression data (Toh et al. submitted) similarly matched the same *S. oleracea* protein. CASP-like proteins (CASPLs) are homologues of Casparian strip membrane domain proteins (CASPs). With respect to the functions of CASPLs, previous research showed CASPLs might function as protein barriers on the cell membrane of the endodermis and form protein scaffolds for the synthesis of the Casparian strip. Some CASPLs were shown to be expressed in the root endodermis, peripheral root cap, root meristem zone, trichomes, lateral root primordia, young leaves, and the floral organ abscission zone in *Arabidopsis thaliana* [44]. This last role, in floral organs, would be an appropriate target for a fungal effector from an anther smut. Alternatively, since CASPLs are orthologous with MARVEL domain proteins associated with the function of epithelial tight junctions [45, CASPLs might be related to tight junction functions in plant cells as well. Thus, the interaction between *MVLG_06175* and CASPL2C1 of *Silene* could indicate that *Microbotryum* alters the functions of tight junctions to enter the host tissues during infection.

MVLG_06379 was also found as a fungal interactor of *MVLG_06175*. *MVLG_06379* contains a PFAM domain (PF03328.7) for ATP citrate lyase (ACL) beta subunit. The

enzyme converts cytosolic citrate into acetyl-CoA for further fatty acid synthesis, but in the parasitic fungi the transcription and translation of *ACL* appears to be associated with infection and reproduction. *Cryptococcus neoformans* increased transcriptional level of *ACL1* within macrophages. Additionally, mutants lacking *ACL1* showed higher susceptibility to antifungal drugs, a lower survival rate within macrophages, and defects in expression of virulence factors [46].

MVLG_01732 Role in Altering the Vesicular Traffic in the Host and Male Sterility

In blastp analyses searching the *Arabidopsis* genome we found this interactor as the orthologue of the AT3G61050.2 gene. This encodes a calcium-dependent lipid binding protein (AtCLB) that has both coiled coil regions and C2 domain (see Figure S1, Appendix) similar to synaptotagmins. Many coiled coil proteins are involved in regulating gene expression as transcription factors. The motif is present in the nucleotide binding site leucine rich repeat (NBS-LRR) proteins of R genes. *Arabidopsis* encodes 150 NBS-LRR-type proteins and they are either the coiled coil (CC) type or the TIR type [47]. C2 domains in animal cells are involved in signal transduction and vesicle trafficking, but in plant cells they are not well characterized. They could be involved in plant stress signal transduction as positive or negative regulators of stress signaling cascades (Figure S3 for predicted interaction partners of AtCLB, Appendix). AtCLB expression is highly detected in rosette leaves and flowers and low in roots, stems, and cauline leaves. Transcriptome analyses studies on pollen germination and tube growth shows its expression in mature pollen, hydrated pollen, and pollen tube growth, which suggest its role in the development of the male gametophyte [48]. Studies show that AtCLB acts as a DNA-binding protein and binds specifically to the promoter sequence of

Thalional synthase 1 (*THASI*), a key enzyme in the synthesis of the triterpenoid, thalional. AtCLB negatively regulates *THASI* transcription, as part of a response involved in drought and stress tolerance. AtCLB becomes localized on the nuclear membrane and can bind to ceramides, a sphingolipid present in cellular membranes that acts as a second messenger in cell signaling, cell differentiation, and apoptosis [45]. Since it is a membrane protein, its activation by membrane lipid ceramide could result in a proteolytic cleavage and translocate it to nucleus to activate transcription of a different set of genes [49]. Interestingly, analysis of the amino acid sequence showed that there are two proteolytic cleavages at positions 28 (after the transmembrane region (1–22)) and 383 (close to where the C2 domain (264–361) ends and the coiled coil region (390–417) begins). Although this is purely speculative at this point, if the MVLG_01732 effector were to become intracellular, upon Ca²⁺ triggering due to conformational changes, the AtCLB protein could interact with the effector at the coiled coil region to mediate AtCLB activation by the membrane lipid ceramide resulting in proteolytic cleavage after the TM to yield mature protein and translocation to the nucleus to regulate transcription of target genes.

In blastp analyses against the NCBI database, the same host interactor for MVLG_01732 matched a portion of the C₂ domain and mostly the C-terminal coiled coil region of synaptotagmin-5 of *Beta vulgaris* subsp. *vulgaris* (LOC104905441), transcript variant X2, mRNA (XM_010693990.2); the corresponding transcript from *Silene* expression data (Toh et al. submitted) similarly matched the same *B. vulgaris* protein. Synaptotagmins are a family of membrane proteins concentrated on secreted vesicles, including synaptic vesicles. They are composed of a short uncleaved *N*-terminal signal peptide that overlaps

a transmembrane (TM) domain, a synaptotagmin-like mitochondrial and lipid-binding protein (SMP) domain, and two tandem cytosolic calcium binding domains (C₂A and C₂B) at the C-terminus required to bind to phospholipids or different ligands in response to calcium signals [50]. Ca²⁺ plays an important role as a second messenger in response to variety of stimuli like cold, drought, salt, oxidative, and biotic stress. Ca²⁺ binding confers two roles in membrane targeting process. One is to provide a bridge between C₂ domain and anionic phospholipids, and the second is to induce intra or inter domain conformational changes, which further triggers membrane protein interactions.

The second host interactor for MVLG_01732 matched the C₂ domain of Cellulose synthase interactive protein 1 (CSI1) of *Spinach oleracea* (Accession number: XP_021846375) and *Beta vulgaris* (Accession no: XP_010680591), orthologues of *Arabidopsis* AT2G22125.1 gene. Again, the C₂ domain (Figures S1 and S2) is a Ca²⁺ binding motif originally identified in Protein Kinase C [51]. However, not all C₂ domains are regulated by Ca²⁺, with some functioning in a Ca²⁺-independent manner and others having mainly a structural role. C₂ domains interact with cellular membranes and mediate key intracellular processes like insulin secretion and neurotransmitter release in eukaryotic cells. It binds to a multitude of different ligands and substrates that include Ca²⁺, inositol polyphosphates, intracellular proteins, and phospholipids.

Mutant analyses found that CSI physically interacts with microtubules and plays a crucial role in anther dehiscence. This is interesting because the events leading to anther dehiscence are coordinated with pollen differentiation, flower development, and opening for successful pollination. CSI 1 disruption mutants exhibited complete sterility and defective anther dehiscence, with crumpled pollen and defective pollen release from the

anther. Moreover, such mutants had morphological changes in the epidermal and endothelial cell length and width necessary for anther maturation, indicating the reason for defective dehiscence may be due to unstable microtubules. CSI mutants also exhibited altered sensitivity to exogenous Ca^{2+} levels, which indicates that there is Ca^{2+} -mediated regulation in microtubule stability and anther dehiscence [52]. Of note, CSI 1 mutants also exhibited decreased number of ovules per gynoecium but were viable, indicating an additional effect of CSI in early gynoecial development.

We hypothesize that the fungal effector MVLG_01732 modulates the function of CSII by interacting with the C_2 domain (suggested by our yeast two-hybrid results), thereby altering the stability of microtubules, resulting in delayed anther development and dehiscence. This could provide the fungus an opportunity to hijack anther development, replacing the pollen grains with its teliospores. Studies show that calcium binding proteins and calcium dependent signaling are involved in both the development of embryo sacs and during the development of pollen [53]. In both the host interactors we identified, the MVLG_01732 effector binding could modulate C_2 domains and their interaction with Ca^{2+} , triggering several signaling pathways for the benefit of the fungus.

In sum, *in silico* analyses predicted a number of fungal small secreted proteins that could serve as effectors to modulate the plant host. For a subset of these, we identified the host interactors that are candidates for targets of these effectors. Recognizing that, even with appropriate controls, yeast two-hybrid analysis can give false positive results, we are planning to further verify the predicted interactions in future experiments using co-immunoprecipitation from infected plants. There is also a need to characterize the function of these interactors and their roles in the plant. Future experiments to express the

fungal effectors in transgenic plants might recapitulate phenotypes observed during infection. Although the natural host, *S. latifolia*, currently lacks a transformation system, heterologous plant systems like *A. thaliana* are amenable for such experiments and should help in the characterization of these fungal proteins. Moreover, expressing the fungal proteins with a fluorescent marker like GFP or mCherry would help determine the localization of these fungal proteins inside the host. Pull down assays can also be conducted to identify additional host proteins, if any, that may have been missed by yeast two-hybrid analysis. This experimental model could then be expanded in the future to provide mechanistic insights into the interplay of this biotrophic pathogen with its host.

Materials and Methods

Plant and Fungal Growth

Silene latifolia seeds that were used in this study were originally collected from a field population in Clover Hollow near Mountain Lake Biological Station, Virginia. Sterilized seeds were plated on sterile 0.3% phytagar (Life Technologies/Thermo Fisher, Waltham, MA, USA), one half strength Murashige and Skoog salts (Sigma Aldrich, St. Louis, MO, USA), and 0.05% MES (2-(*N*-morpholino) ethanesulphonic acid) buffer (Sigma-Aldrich, St. Louis, MO, USA). Seeds were kept at 4 °C for 5 days to encourage germination and then were transferred to a 20 °C growth chamber with 13 h of fluorescent light. Humidity was kept high initially by using dome covers and flood trays and was gradually decreased to lower levels. Seedlings were transplanted to bigger pots for the emerging new roots to provide hydration requirement when the volume of soil was not sufficient. Plants were grown in Sunshine MVP professional growing mix (Sun gro Horticulture Canada Ltd, cat no: 02392868, Agawam, MA, USA) and were watered every other day with 100-ppm

fertilizer (Peters Professional 15-16-17 Peat-Lite Special, Formula no: S12893, JR Peters, Inc. Allentown, PA, USA) [7].

Fungal strains of *M. lychnidis-dioicae*, p1A1 and p1A2, were axenically grown separately on nutrient rich media (yeast peptone dextrose media (YPD); 1% yeast extract, 10% dextrose, 2% peptone, and 2% agar) at 28 °C for 5 days and nutrient-free water agar media for 2 days (2% water agar).

Plant infection employed haploid *M. lychnidis-dioicae* p1A1 and p1A2 cells that were grown on nutrient rich media (YPD; 1% yeast extract, 10% dextrose, 2% peptone, and 2% agar) at 28 °C; these were harvested and adjusted to a concentration of 1×10^9 cells/mL in equal proportion before being spotted onto nutrient free media (2% agar). The cells were resuspended to a concentration of 1×10^6 in distilled water. Then, 5 μ L of this was dropped onto the 12-day old *S. latifolia* seedlings [7].

***In Silico* Analyses**

Prediction of Small Secreted Proteins (SSPs)

Prediction of the secretome used a pipeline of software packages (TargetP1.1, SignalP3.0, SignalP4.0 (<http://www.cbs.dtu.dk/services/SignalP/>), TMHMM2.0, PredGPI, Phobius, NucPred, Prosite, and WoLF PSORT) to provide a stringent determination of likely secretion [7] (Table S1, Appendix and Figure 2-5).

Computational framework for prediction of the secretome of *M. lychnidis-dioicae*

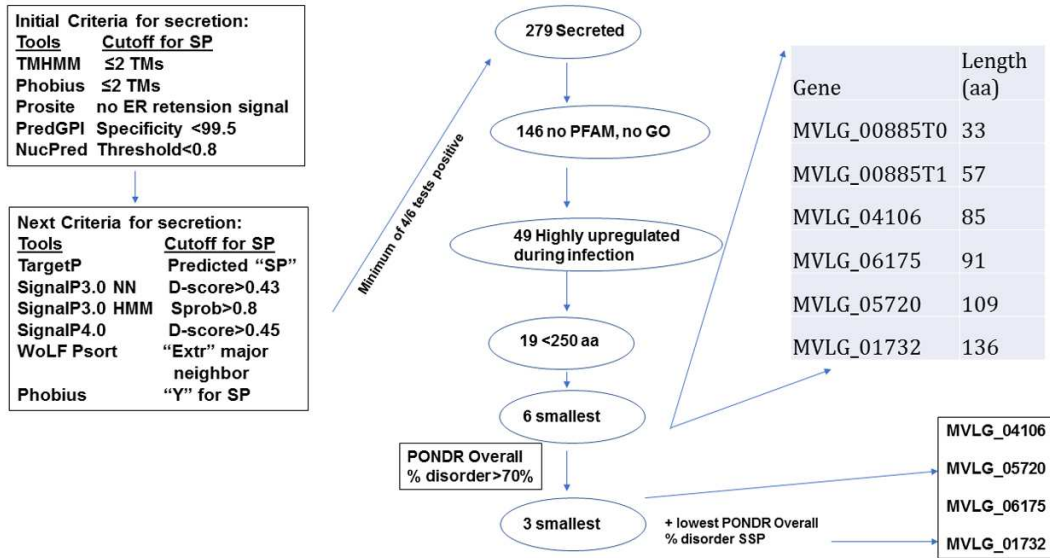


Figure 2- 5: Computational framework for prediction of secretome for *M. lychnidis-dioicae* and selection of candidate effectors for further analyses.

Detailed description of tools and cut-off criteria for secretome prediction and prediction of disorder are provided in Supplementary Methods. Numbers tally for proteins at each stage of secretome prediction are provided in tab 3 of Table S1 (Appendix). TM, transmembrane domain; ER, endoplasmic reticulum; SP, secreted protein; PFAM, protein family; aa, amino acid; GO, gene ontology; MVLG designations refer to specific *Microbotryum lychnidis-dioicae* proteins

Prediction of Intrinsic Disorder

In order to analyze the residue level of disorder propensity of 49 putative effector proteins, four intrinsic disorder predictors were used: POND[®] VSL2 [18], POND[®] VLXT [19], POND[®] VL3 [20], and POND[®] FIT [21]. While evaluating the intrinsic disorder predisposition of four SSPs targeted for functional analysis (MVLG_01732,

MVLG_04106, MVLG_05720, and MVLG_06175), in addition to the members of the PONDR family, IUPred_short and IUPred_long were used [24].

Molecular recognition features (MoRFs) are short segments with increased order propensity located within longer disordered regions. MoRFs bind to globular protein domains and undergo disorder-to-order transition. These disorder-based binding sites are categorized into three types: α -MoRFs (form α -helices upon binding), β -MoRFs (form β -strands), and ι -MoRFs (form irregular structures). For all 49 predicted secreted proteins whose transcription was upregulated during infection, the ANCHOR algorithm (<http://anchor.enzim.hu/>) was used to predict such protein binding regions that are disordered in isolation but can undergo disorder-to-order transition upon binding [22]. This computational tool finds segments within disorder regions that cannot form stable intra-chain interactions to fold on their own, but are likely to gain stabilizing energy by interacting with a globular protein partner [22]. Furthermore, the presence of MoRFs in MVLG_01732, MVLG_04106, MVLG_05720, and MVLG_06175 was further evaluated by another computational tool, MoRF_{chibi} [54].

Additional Bioinformatic Analyses

Alignment of nucleotide and/or amino acid sequences to find regions of similarity between such biological sequences employed Basic Local Alignment Search Tool (BLAST; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Further domain analysis for prediction of post translational modification sites used ModPred [33]. Structural modelling of predicted proteins utilized Swiss-Model [55,56]. Results of these analyses are found in Table S2, Appendix. Additional analysis methods are provided in Supplementary Materials, Appendix and associated references [57,58,59,60].

Yeast Secretion Trap (YST) Experiment

For each candidate effector, validation of secretion employed a yeast-based secretion trap method [28]. Putative secretion signals for each fungal gene were cloned into the pYSTO-0 vector. In such analyses, if the putative signal peptide from a protein provides for secretion of the Suc2p invertase, *S. cerevisiae* cells will be able to grow on sucrose as a sole carbon source; inability to promote growth would indicate that the fungal protein of interest is not normally secreted.

The signal peptide sequence of each fungal protein was determined by Signal P software and amplified by PCR. Standard PCR cycle was used with initial denaturation set at 94 °C for 4 min and 35 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s, with a final extension time of 5 min at 72 °C. The product was held at 4 °C at the end of the cycle.

The PCR products were separated by gel electrophoresis through 1.8% agarose (Agarose LE; USB Corp., Cleveland, OH, USA). The fragments were excised from the gel and purified using the Zymo Gel DNA recovery kit (Orange, CA, USA). The purified fragments were subjected to restriction digestion with *EcoRI* and *NotI* enzymes. The digested fragment was purified and cloned into the pYST-0 vector to obtain a translational fusion with the invertase expressed from the *ADHI* promoter and transformed into *Escherichia coli* DH5 α cells. Cells were plated on LB plates with ampicillin (100 mg L⁻¹) and incubated at 37 °C overnight. *E. coli* strain, DH5 α (Bethesda research Laboratories, Bethesda, MD, USA), was utilized for all cloning purposes. *E. coli* strains were grown at 37 °C in Circle Grow media (MP Biomedicals, LLC, Solon, OH, USA) and plasmid DNA was isolated from potential clones using the alkaline lysis

procedure [61]. The presence of each signal peptide encoded in-frame with the *SUC2* coding region was confirmed by DNA sequencing at the Nucleic Acids Core Facility (Center for Genetics and Molecular Medicine, University of Louisville, Louisville, KY, USA).

Invertase-deficient (*suc2⁻*) *S. cerevisiae* strain (SEY 6210 (MAT α leu2-3, 112 ura3-52 his- Δ 200 trp1- Δ 901 lys2-801 *suc2⁻* Δ 9 GAL)) [62] cells were transformed with the constructs using the lithium acetate/single-stranded carrier DNA/PEG method [63]. Selection was on Synthetic Dropout medium, with SD/-Leu (Clontech, Mountain View, CA, USA) selection plates containing glucose as the sole carbon source. The dropout medium contained glucose (20 g·L⁻¹), yeast nitrogen base (6.7 g·L⁻¹), dropout mix minus leucine (2 g·L⁻¹), agar (15 g·L⁻¹), and water. The plates were incubated at 30 °C for 6–10 days. The colonies were re-streaked for purification onto SD/-Leu drop out selection plates with sucrose as the sole carbon source to select the positive clones that were able to utilize sucrose by secreting invertase enzyme. Such strains were grown overnight in 3 mL of SD/-Leu broth with sucrose, and 10-fold dilutions were spotted onto SD/-Leu with glucose or sucrose as the carbon source and incubated for 5 days at 30 °C. Clones harboring functional signal peptides with reconstituted invertase activity were able to grow on sucrose as the sole carbon source. Untransformed mutant yeast strain SEY 6210 and the same strain, transformed with empty pYST-0 vector, were used as negative controls. Plasmid DNA was extracted from the positive clones and used to retransform *E. coli*. The constructs were again checked for the presence of signal peptide sequence by DNA sequencing.

RNA Extraction and cDNA Library Construction

RNA for generating the cDNA library was obtained from the axenically grown cultures of p1A1 and p1A2 haploid strains [7] on nutrient rich media for 5 days (YPD) at 28 °C, nutrient-free water agar media (2% water agar) for 2 days, and the fungal infected *Silene latifolia* tissue [7]. This latter set of RNAs was extracted from floral stem (pedicle, and remaining cluster and sepals), floral buds (male and female) at different stages (male: 2–6 mm buds, 8 mm to fully opened smutted flowers; female: 3–6 mm, 7–14 mm, and 15–24 mm). The quality of the RNA was checked by Agilent Bioanalyzer and all the samples indicated highly intact RNA with the RNA integrity scores of at least 7.8. The total samples were pooled equally based on Bioanalyzer quantification to generate a normalized cDNA library. The cDNA library was constructed in a Gal4 based prey vector, pGADT7 (Clontech, Mountain View, CA, USA,), by CD Genomics (Shirley, NY, USA) for yeast two-hybrid screening.

Yeast Two Hybrid Screen

The yeast two-hybrid system allows for an initial screening of possible protein-protein interactions [64,65]. A “bait” protein of interest is expressed from a yeast (*Saccharomyces cerevisiae*) expression vector as a fusion with the Gal4 DNA binding domain (BD). Interactors with bait are identified by screening “prey” expressed from a yeast vector where the fusion is with the Gal4 transcriptional activation domain (AD). pGBKT7 was used as a “bait” vector with the GAL4 DNA-binding domain and pGADT7 was used as a “prey” vector with the GAL4 DNA activation domain. While neither the BD, nor AD alone, can activate transcription of the reporter genes used in this system, if two proteins physically interact (*i.e.*, if prey Y interacts with bait X), then the BD and AD

are brought together, and reporter genes will be expressed. In our studies, the prey proteins were all expressed from normalized cDNA libraries of the different stages of *M. lychnidis-dioicae*, including in association with its host, *S. latifolia*. Initial selection of interactors involves ability to grow on increasingly more stringent auxotrophic media, since the yeast strains have auxotrophic mutations that require them to either be provided with the missing nutrients or to have a functional interaction that activates transcription of reporter genes whose read-out is complementation of the growth defect. Additionally, an α -galactosidase gene serves as a reporter, whereby color change to blue-green occurs via cleavage of 5-bromo-4-chloro-3-indolyl α -D-galactopyranoside (X- α -gal) in the medium. In order to avoid false positives, a number of controls were employed, including comparisons using (1) vectors alone/without bait or prey (*i.e.*, pGBKT7 or pGADT7, respectively); (2) bait in BD vector alone; (3) prey in AD vector alone; (4) re-transformation of yeast strains with identified interactors and bait; (5) repetition of the experiment with vectors, in which the bait has been fused to AD and the identified prey interactor has been fused to the BD, so as to avoid artifacts associated with the particular fusion used originally. The interaction of pGBKT7-53 (containing p53 coding sequence) and pGADT7-T (containing T antigen coding sequence) was used as a positive control. Only those candidate interactors that passed these stringent tests were considered worthy of further investigation.

The coding sequences of each of the effector candidates, lacking signal peptides and stop codon, were PCR amplified using cDNA as template, generated from fungal infected *S. latifolia* floral buds, using the primer pairs described in Table S3, Appendix. The effector candidates tested in this study were MVLG_004106, MVLG_005720, MVLG_06175,

and MVLG_001732, for which sequences are available in the JGI Fungal Genome database [66]. The PCR products were cloned into the pCR 2.1 TOPO entry vector (Invitrogen/Thermo Fisher, Waltham, MA, USA). *Escherichia coli* strains, DH5 α (Bethesda research Laboratories, Bethesda, MD, USA), were utilized for all cloning purposes. Plasmid DNA was isolated and the inserts were digested out of this vector with *EcoRI* and *BamHI*. Purified fragments were subsequently cloned into a pGBKT7 destination vector (Clontech) where transcription of the cloned gene would be driven by an *ADHI* promoter, producing fusion proteins at their N termini with the DNA binding domain of the Gal 4 transcription factor.

S. cerevisiae strain Y187 (Library host strain) (MAT α ,ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4 Δ , met $-$, gal80 Δ , URA3 : : GAL1_{UAS}-GAL1_{TATA}-lacZ) [67], containing the *MEL1/lacZ* reporter gene, was transformed with the prey vector containing the cDNA library using the Frozen-EZ Yeast Transformation II kit (Zymo Research) and selected on SD drop out medium lacking Leucine (SD/-Leu). MELIBIASE1 (*MEL1*) reporter gene encodes α -galactosidase and enables yeast cells to turn blue-green in the presence of the chromogenic substrate, 5-bromo-4-chloro-3-indolyl α -D-galactopyranoside (X- α -gal). Cell density of the library was calculated by titering 10 $^{-4}$, 10 $^{-5}$, 10 $^{-6}$, and 10 $^{-7}$ dilutions on SD/-Leu plates.

The AH109 yeast strain (Mating partner) (MAT α , trp1-901, leu2-3, 112, ura3-52, his3-200, gal4 Δ , gal80 Δ , LYS2::GAL1_{UAS}-GAL1_{TATA}-HIS3, GAL2_{UAS}-GAL2_{TATA}-ADE2, URA3::MEL1_{UAS}-MEL1_{TATA}-lacZ) [68], containing *HIS3*, *ADE2*, and *MEL1/lacZ* reporter genes, was used as the host for the bait constructs. The *HIS3*, *ADE2* reporter gene products enable the cells to biosynthesize required nutrients to grow on plates

lacking histidine and adenine. The three reporter genes are under the control of distinct *GAL4* upstream sequences and promoter elements *GAL1*, *GAL2*, and *MEL1*, respectively, yielding strong and specific responses. In AH109, the entire *HIS3* promoter (including both TATA boxes) was replaced by the entire *GAL1* promoter, leading to tight regulation of the *HIS3* reporter gene in this strain. The bait constructs were transformed into AH109 by the lithium acetate/single-stranded carrier DNA/PEG method [63] and selected on SD drop out medium lacking Trp.

The yeast two-hybrid screening was conducted following the Matchmaker Library Construction and Screening Kits User manual (Clontech). Initial screening was conducted on high stringent quadruple drop out media (QDO) SD/-Ade/-His/-Leu/-Trp plates with X- α -gal and 5 mM 3AT. Subsequently, colonies were restreaked onto QDO/X- α -gal plates with 5 mM 3AT initially, and then on to QDO/x- α -gal plates containing 50 mM 3AT to select strong interactors. 3AT was used to inhibit the leaky expression that reduces the effectiveness of histidine selection, and to inhibit X- α -gal to allow detection of the *MEL1/lacZ* reporter. 3-AT is a competitive inhibitor of the yeast HIS3 protein (His3p), blocking low levels of His3p expression, and thus suppressing background growth on SD medium lacking His. Only the positive blue-green clones (indicating α -galactosidase activity) that survived on the highest 3AT levels were used for further screening. To estimate the mating efficiency and to calculate the total number of screened colonies dilution serials were prepared and 100 μ L of each dilution was spread on SD/-Trp, SD/-leu, and DDO plates. Plasmids were isolated from the surviving colonies and were individually used to transform *E. coli*. The prey plasmids were isolated from *E. coli*, and sequenced and analyzed by BLAST screens against the NCBI database

[69]. The dropout medium contained Glucose ($20 \text{ g}\cdot\text{L}^{-1}$), yeast nitrogen base ($6.7 \text{ g}\cdot\text{L}^{-1}$), appropriate dropout mix ($2 \text{ g}\cdot\text{L}^{-1}$), agar ($15 \text{ g}\cdot\text{L}^{-1}$), and water.

Conclusions

In this paper we identified, for the first time, the interactors of the putative effectors of *M. lychnidis-dioicae*. We believe that the protein product of MVLG_04106 codes for a transcriptional regulator/activator of host responses to allow successful infection. The fungal interactors of MVLG_05720 protein product, MVLG_07305 and MVLG_04026, could potentially sequester this effector until it is required during infection. MVLG_06175 appears to interact with a CASP-like homologue and may be involved in cell-cell junctions. The identification of two host interactors of MVLG_01732-AtCLB and CSI I, which play roles in anther/pollen development and dehiscence, provides exciting targets for future studies, as we hypothesize this effector may be crucial in redirecting anther and pollen development in such a way as to benefit the reproductive program of the fungus. Plant infection studies with knockouts or over expression of these effector genes will further our understanding in characterizing the function of these key players in the infection. This strongly suggests the need to also characterize the remaining candidate effector proteins for a more complete understanding of the mechanisms of infection and development of this fascinating plant parasite.

CHAPTER III
FUNCTIONAL CHARACTERIZATION OF THE *MICROBOTRYUM LYCHNIDIS-DIOICAE* EFFECTORS IN A HETEROLOGOUS HOST MODEL

Overview

The anther smut fungus, *Microbotryum lychnidis-dioicae* infects, *Silene latifolia* from the Caryophyllaceae family. It has served as a popular model to study ecology, evolutionary and host specificity of a pathosystem. However, the molecular mechanism of this host-pathogen interaction is poorly understood. The obligate requirement for biotrophic life style and the lack of proper transformation techniques for both the pathogen and the host species have severely curtailed further progress at the molecular level. In this study, for the first time, we expressed a subset of three effector proteins from this fungus in a heterologous host plant model, *Arabidopsis*, to understand the role they may play in the native host. Among the three effectors studied, transgenic *A. thaliana* lines expressing the MVLG_01732 effector, which localized to the plasma membrane, showed curly leafy symptoms with early bolting. This phenotype is interesting because the *Microbotryum lychnidis-dioicae* infected *S. latifolia* flowers bolt earlier than the uninfected flowers in the natural populations. In addition, expression of MVLG_01732 in *A. thaliana* caused enhanced susceptibility to the infection of the plant pathogenic bacteria, *Pseudomonas syringae*.

Introduction

Microbotryum violaceum is a fungal species complex that infects members of Caryophyllaceae. Within this species complex, some taxa emerged as independent lineages based on host specificity and evolved as separate species. *Microbotryum lychnidis-dioicae* is one such species that infects the host plant, *Silene latifolia*. This basidiomycete fungal infection results in the smutted appearance of flowers, where the pollen in anthers has been replaced by this castrating parasite's spores; hence the common name - anther smut fungus. Recently, the genome sequence and the transcriptome of the fungus were determined, and these revealed insights about the role of small secreted effector proteins that are highly upregulated during the infection (Perlin, Amselem et al. 2015). Out of 7364 predicted proteins, 19 were found to be small (less than 250 amino acids) secreted proteins that are unique to *M. lychnidis-dioicae* and were highly upregulated during plant infection, suggesting these may play a role during those stages of the fungal life cycle and in pathogenicity (Perlin, Amselem et al. 2015). In our previous study, we have confirmed that one such effector, MVLG_01732 is secreted (Kuppireddy, Uversky et al. 2017). Yeast two-hybrid analysis using the cDNA library from different stages of growth conditions of fungus and infected plant revealed that it interacts with *Arabidopsis thaliana* orthologues found in *S. latifolia*, i.e., calcium-dependent lipid binding protein (AtCLB) and with cellulose synthase interactive protein 1 (CSII). Identification of host target proteins of *M. lychnidis-dioicae* effector proteins provided a crucial lead to determine the effector function (Kuppireddy, Uversky et al. 2017). We, therefore, utilized fluorescent tagged fusion proteins to localize subcellular locations of target proteins in order to better elucidate potential functions.

To serve this purpose, we generated mCherry tagged versions of fungal proteins to identify their sites of localization *in planta* through stable expression in the heterologous host, *Arabidopsis thaliana*. Heterologous model systems use the expression of a gene in a host organism other than its native host in the usual settings. These systems are popular because they are amendable to use and because of the lack of tractable experimental approaches to study obligate biotrophic pathogens. *A. thaliana* has been a genetic model organism for many years and it has contributed an important role in genetics and molecular biology, including the study of host-pathogen interactions. For recalcitrant fungi two methods have been used to express effectors in non-native host species. *Agrobacterium*-mediated heterologous expression has been used for stable expression of fungal effectors; alternatively, delivery of effectors through *Pseudomonas* type III secretion system has been used for transient expression of such effectors. Both methods are considered reliable methods for understanding the function of effectors (Hugo, L. et al. 2018).

In this study, we conducted *in planta* assays using *A. thaliana* as a heterogenous nonhost model to express the fungal effector, MVLG_01732. Stable expression in *A. thaliana* expressing MVLG_01732 tagged with mCherry fluorescent protein, allowed determination of the subcellular localization of this effector *in planta*. We also report here the phenotypic effect when this effector is expressed in *A. thaliana*. Further, we investigated the interaction and the effect of MVLG_01732 in altering the growth of the pathogen, *Pseudomonas syringae*, in these expression lines. Our results suggest that MVLG_01732 localizes to the plasma membrane and promotes the growth of *P. syringae* during infection of *A. thaliana*.

Materials and Methods

Plant material and Growth conditions of *Arabidopsis*.

Arabidopsis thaliana ecotype Col-0 was used as wild-type background in this study (kindly provided by Dr. Mark Running, University of Louisville). The plants were grown in Sungro Horticulture propagation mix soil (Premium Horticultural Supply, Louisville, KY, cat no.5232601), in a growth chamber, at 22 °C, with 68% relative humidity (RH), light intensity 120 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and with a 16 h / 8 h day/night cycle.

In vitro culture of *Arabidopsis* was performed by plating the seeds on 1/2x MS (Murashige & Skoog, Phytotechnology Laboratories, Cat No: M524) media, 0.05% MES (2-(*N*-morpholino) ethanesulfonic acid, ThermoFisher, Pittsburgh, Pennsylvania, USA) buffer containing 0.8% agar adjusted to a pH 5.7. For the selection of transgenic plants, MS media with kanamycin (50 $\mu\text{g/ml}$) and 300 $\mu\text{g/ml}$ Cefotaxime (Amresco) was used. Seeds were surface sterilized and cold-treated for 2-3 days and then transferred to a growth chamber for germination at 20-24 °C. Seedlings were further grown for 10 days on plates and then transferred to soil pots filled with Sungro Horticulture propagation mix (Premium Horticultural Supply, Louisville, KY, cat no.5232601).

Generation of the CaMV 35S:: MVLG binary construct.

The effector candidates tested in this study were MVLG_004106, MVLG_005720, and MVLG_001732, for which sequences are available in the JGI Fungal Genome database (Nordberg, Cantor et al. 2014). To obtain transgenic plants expressing the coding sequences of each of the effector candidates either as full-length proteins or as truncated versions lacking the N-terminal signal peptide (ΔSP), each target was PCR amplified

using cDNA as template; the cDNA was generated from fungal infected *S. latifolia* floral buds, using the primer pairs described in Table 3-1. For amplification by PCR, ExTaq polymerase was used. Standard PCR cycle was used with initial denaturation set at 94 °C for 4 min and 35 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 45 s, with a final extension time of 5 min at 72 °C. The product was held at 4 °C at the end of the cycle.

The PCR products were separated by gel electrophoresis through 1.5% agarose (Agarose Unlimited USB Corp., Cleveland, OH, USA). The fragments were excised from the gel and purified using the Zymo Gel DNA recovery kit (Orange, CA, USA). The purified fragments were subjected to restriction digestion with *Bam* HI and *Nde* I enzymes and cloned into pRI binary vector under the control of the CaMV 35S promoter. *Escherichia coli* strains, DH5 α (Bethesda Research Laboratories, Bethesda, MD, USA), were utilized for all cloning purposes. Cells were plated on LB agar plates with kanamycin (50 μ g/ml) and incubated at 37 °C overnight. *E. coli* strains were grown at 37 °C in Circle Grow media (MP Biomedicals, LLC, Solon, OH, USA) and plasmid DNA was isolated from potential clones using the alkaline lysis procedure (Sambrook J 2001). The presence of each gene was confirmed by DNA sequencing at the Nucleic Acids Core Facility (Center for Genetics and Molecular Medicine, University of Louisville, Louisville, KY, USA). The coding region for mCherry was amplified from pMF5-5h vector (kindly provided by Dr. Michael Feldbrugge, Universität Düsseldorf). The primers used for amplification are listed in Table 3-1. The PCR products were separated by gel electrophoresis through 1.2% agarose (Agarose Unlimited USB Corp., Cleveland, OH, USA). The fragments were excised from the gel and purified using the Zymo Gel DNA recovery kit (Orange, CA, USA). The PCR product was cloned into the pCR 2.1 TOPO entry vector

(Invitrogen/Thermo Fisher, Waltham, MA, USA). *Escherichia coli* strains, DH5 α (Bethesda Research Laboratories, Bethesda, MD, USA), were utilized for all cloning purposes. Cells were plated on LB plates with ampicillin (100 μ g/ml) and incubated at 37 °C overnight. *E. coli* strains were grown at 37 °C in Circle Grow media (MP Biomedicals, LLC, Solon, OH, USA) and plasmid DNA was isolated from potential clones using the alkaline lysis procedure (Sambrook J 2001). Plasmids that appeared to be larger than the TOPO vector were checked for the presence of amplicon by PCR. The presence of the mCherry insert was confirmed by DNA sequencing at the Nucleic Acids Core Facility (Center for Genetics and Molecular Medicine, University of Louisville, Louisville, KY, USA). The positive TOPO clone was subjected to digestion using *Eco* R1 and *Bam* HI, the products separated on a 1.2 % agarose gel, and purified out of the gel as described above. The insert is expressed as a fusion protein, with mCherry fused to the C-terminus of the effector. Each construct was confirmed again for the presence of mCherry, the MVLG gene and the junction region of the fusion area and kanamycin gene by DNA sequencing. The primers used for screening and sequencing are found in the following Table 3-1.

Table 3- 1: Primers used for expressing the *M. lychnidis-dioicae* genes in *A. thaliana*

Name	Sequence (5' to 3')
MVLG04106pRINde	5'-GCCCATATGTAGGTCTCACTCGCATCCAC-3'
4106NdeFNoSig	5'-GCC CAT ATG GCT GAC GCG ACC AAA C-3'
MVLG04106pRIBam	5'-GCAGGATCCACAACCTTCGGGCTCGGG-3'
MVLG05720pRINde	5'-GCCCATATGTGCCAGTTCAACATGATGC-3'

5720NSPpRIFNde	5'-GCCCATATGAACCCGTGGCCTCCGTC-3'
MVLG05720pRIBam	5'-GCAGGATCCGTAACCCGAACGACGCATCC-3'
6175SignalNdeF	5'-GCCCATATGATACATCGTCCTCAAGCCAG-3'
6175SignalBamR	5'-GCAGGATCCGAGATTTAGAGGAAAGAACCAAT-3'
1732CompletepRIF	5'-GCCCATATGATCTTTCGCCCCGACTTTCA-3'
1732noSPpRIF	5'-GCCCATATGTTGCAAGAAGCGGGCGATAC-3'
1732SignalBamR	5'-GCAGGATCCGGCGTGGATTTTGCCGGAGA-3'
398SignalNdeF	5'-GCCCATATGCTCCCATCGGACTTCAACG-3'
398SignalBamR	5'-GCAGGATCCGATCGAATGCTTGGGAATGT-3'
KMF40	5'-GCTTGGGTGGAGAGGCTATT-3'
KMR677	5'-CGGGTAGCCAACGCTATGTC-3'
mCherryFBamNew	5'-GGA TCC ATG GTG AGC AAG GGC GAG G-3'
mCherryREcoNew	5'-GAA TTC CAA GAC CGG CAA CAG GAT T-3'
MCherry199R	5'-ACAGGATGTCCCAGGCGAAG-3'

Transformation of *Agrobacterium* using electroporation.

The mCherry: MVLG pRI expression effector was introduced by electroporation into *Agrobacterium tumefaciens* strain EHA105 and used to produce *Arabidopsis* transgenic lines. Competent cells of *Agrobacterium* were made following the Pikaard's Lab protocol (Pikaard). Transformation of competent cells of *Agrobacterium* was performed in a BioRad micropulser electroporator with the voltage of 2.5 kV using the 25 uF capacitor and at 400-ohm settings (Mattanovich, Ruker et al. 1989). The competent cells were thawed on ice and 2 µl of the plasmid was added to the tube. The pre-chilled cuvettes

were placed on ice and the mixture was transferred to the cuvette using pre-chilled sterile tips. The pulse buttons were held for a short time until the time constant read about 4.7 sec. Then, 1ml of LB was added to the cuvette and pippered up and down to mix and transferred to a 2 ml tube. The cells were incubated at 28 °C for 2 hours before plating 150 µl onto LB plates with kanamycin (50 µg/ml). The plates were incubated at 28 °C for 2 days. The plasmid DNA was isolated from the potential clones (Wang 2006) and screened by PCR as an initial screen. One such positive clone for each construct type was selected for retransforming *E. coli*. The construct was again checked for the presence of the insert by PCR and restriction digestion analysis to confirm the true *Agrobacterium* clones to be used for the transformation of plants.

Stable transformation of *Arabidopsis* and *in vitro* selection of transformant seedlings.

Transgenic lines of *Arabidopsis* (Col-0) were generated using a modified floral dip transformation protocol (Zhang, Henriques et al. 2006). An *Agrobacterium* strain harboring MVLG gene-pRI expression vector was inoculated in 5 ml of LB medium supplemented with rifampicin, spectinomycin, and kanamycin at 50 µg/ml and grown to the stationary stage (OD₆₀₀ approximately 2.0) at 28 °C with shaking at 250 rpm for 2 days. Next, 3 ml of this feeder culture was used to inoculate 500 ml liquid LB and was grown for 16-24 hours to an OD₆₀₀ of 1.5-2.0. Then, *Agrobacterium* cells were harvested from the liquid medium by centrifugation at 4 °C for 10 min at 4000 x g and washed with freshly made 5% sucrose. Silwet L-77 (PhytoTechnology Laboratories, Lenexa, Kansas) was added to a concentration of 0.02% (v/v) and mixed well before dipping the plants. The plants were inverted to dip the aerial parts along with rosettes to soak shorter axillary

inflorescences for 10 s with gentle agitation. In case the plant was too short for younger meristems to submerge, the suspension was dropped onto the unopened floral buds with a micropipette to infect them. Infected plants were kept at high humidity under a plastic bag overnight and then transferred to the growth chamber. The inoculation procedure was repeated a second time after one week. As a positive control mCherry cloned into pRI was used. As negative controls, pRI only vector, wild-type, and mock-inoculated plants and *Agrobacterium* strains with no pRI vector are used. The mCherry fluorescence was visualized in leaf cells using confocal microscopy.

Each effector, either with the signal peptide or lacking signal peptide, was expressed as a fusion protein with mCherry, (except for MVLG_04106 Δ SP-mCherry, which, despite a number of attempts, was not successfully generated). As a positive control mCherry cloned into pRI was used. As negative controls, pRI only vector, wild-type uninfected, and mock-inoculated plants with cocultivation media with no bacteria and untransformed *Agrobacterium* strains with no pRI vector are used.

Pooled seeds (T0) produced by infected plants were collected and screened for kanamycin (KM)-resistant seedlings. To select *Arabidopsis* transgenic plants, the healthy green colored plants were picked after 14 days and transferred to water-saturated soil. The T1 plants were grown and the seeds from transgenic lines were disinfected and first plated on sterile half strength ($\frac{1}{2}$ x) MS plates with 0.8% agar and with kanamycin and cefotaxime and then transferred to the soil. T2 Lines carrying only a single T-DNA copy were selected based on the Mendelian segregation of the dominant resistant marker segregation ratio (3:1, resistant/susceptible) on kanamycin-containing medium. Independent homozygous single insertion T3 lines exhibiting 100% resistance were

selected as homozygotes and these were used for phenotypic evaluations, subcellular localizations and bacterial infiltration assays.

Statistical analysis:

The T2 progeny of independent *Arabidopsis* lines, selected using kanamycin, was sown and germinated in KM containing MS media. 14 days after germination, the number of resistant and nonresistant seedlings in the segregational analysis was evaluated using the chi-squared test.

Laser Scanning confocal microscopy.

Fourteen-day-old T3 *Arabidopsis* seedlings grown in Petri dishes containing $\frac{1}{2}$ x MS media with kanamycin were observed using confocal microscopy. Images were acquired using an Olympus Fluoview FV-1000 confocal coupled to an Olympus 1X81 inverted microscope, a PlanApoN 60 \times objective, and FV-10 ASW 2.1 software. A single channel scanning configuration was set up for the acquisition of mCherry (excitation 587nm, emission 610nm) using a 543 nm HeNe laser. Optimal brightness for this channel was configured by determining the setting yielding maximal intensity without saturation. Each of the settings was tested against wild-type plants for the mCherry signal to ensure exclusion of non-specific emission. Scanning was performed at a speed of 2 μ s/pixel to acquire z-stacks of each visual field. Images are presented as either single plane images or stacked images, as indicated in the figure 3-6 and 3-7.

Plant genotyping.

Leaves of 4-week old *Arabidopsis* were harvested and snap frozen in liquid nitrogen. DNA was extracted using Plant DNAzol[®] reagent according to the manufacturer's instructions (Thermo Fisher Scientific, Cat No: 10978021). DNA concentration and purity were assessed using Nanodrop 2000[™] UV-Vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

The genomic DNA extracted from transgenic *A. thaliana* lines was used as a template for PCR. Standard PCR was performed using TaKaRa Hot start ExTaq DNA polymerase enzyme (cat no. RR001A, Takara Bio USA, Inc). Primers amplifying the mCherry insert and the MVLG gene were used and the PCR products were separated in 1.3% agarose gels to screen the transgenic lines.

Gene expression analysis using qRT-PCR of the T-DNA insertion lines.

Total RNA was extracted from *Arabidopsis thaliana* leaves with Trizol reagent (Invitrogen). Turbo DNase treatment (Ambion) was performed. cDNA synthesis was done using oligodT primers and Superscript III cDNA synthesis kit (Invitrogen). The qRT-PCR reaction was performed using the primers presented in Table 3-2. 1x Power SYBR Green (Applied Biosystems) was used as the detector. Cycle threshold values were normalized to the housekeeping gene *UBQ10*. Quantifications of RNA expression levels were performed in an Applied Biosystems Step-One thermocycler using the following PCR conditions: 95 °C for 10 min followed by 95 °C for 15 s and 60 °C for 1 min for 40 cycles. Melting curve analysis was performed at the end of each cycle to ensure the specificity of the reaction. Three different RNA isolations for wild-type and

mCherry plants were carried out and nine different RNA isolations from three independent lines were performed for MVLG_01732 Δ SP-mCherry plants; each cDNA was measured in triplicate. Results were expressed as log₂ fold change relative to the housekeeping gene *UBQ10*. Primers, Gene name, and the size of the Genomic and cDNA amplicons are listed in Table 3-2.

Table 3- 2: Primers used for qRT analysis

Name	Sequence (5' to 3')	Source
Ara geno F qRT (At3g18780)	5'-ACTTTCATCAGCCGTTTTGA-3'	(Czechowski, Stitt et al. 2005)
Ara geno R qRT (At3g18780)	5'-ACGATTGGTTGAATATCATCAG-3'	
UBQ10 F qRT (AT4G05320)	5'-GTTGGAGGATGGCAGAACTC-3'	(Czechowski, Stitt et al. 2005)
UBQ10 R qRT (AT4G05320)	5'-GGAGCCTGAGAACAAGATGAA-3'	
mCherry F qRT	5'-CACTACGACGCTGAGGTCAA-3'	This Study
mCherry R qRT	5'-GTGGGAGGTGATGTCCAAC-3'	
1732NSP F qRT	5'-TGCACAAAATCTAACCTCA-3'	This Study
1732NSP R qRT	5'-GTATCGGGTGATGGAGCAAG-3'	
5720NSP F qRT	5'-GATGCAAATGGCGGACAAGT-3'	This Study
5720NSP R qRT	5'-AGGGTTCCATGAAAGCGTCA-3'	
4106NSP F qRT	5'-CATCACGCATCCAGTAGCCT-3'	This Study
4106NSP R qRT	5'-ATACAAGTCCCCCGTGCTTC-3'	

Protein extraction and Immunoblotting.

Leaves of 3-week old T3 *Arabidopsis* transgenic plants were ground to a fine powder after liquid nitrogen exposure. Proteins were extracted in a buffer containing 0.1M Tris-HCl, pH 8, 0.1% SDS, 2% β -mercaptoethanol, and 1X protease inhibitors (ICN Biomedicals Inc, CA). Samples were incubated on ice for 10 min; then the cell debris was pelleted by centrifuging twice for 10 min at 13,000 rpm at 4 °C. The supernatant was separated and protein concentration was quantified using the Bradford reagent (Hugo, L. et al. 2018). An equal amount of protein (10 μ g) was loaded in each well as determined by Bradford assay. The extracted proteins were separated by gradient polyacrylamide gel (4-12%) in Tris-glycine buffer for 1 hour at 30 V and then electrotransferred to a 0.45 μ m nitrocellulose membrane. Immunoblotting was carried out by incubating the membrane with diluted (1:1000) primary antibody, anti-c-Myc-mouse monoclonal antibodies (Cell signaling #2276) in 5% w/v nonfat dry milk/ Tris-tween-buffer saline (TTBS) buffer at 4 °C with gentle shaking, overnight. After transfer, the nitrocellulose membrane was washed three times with 1x TTBS buffer for 5 min at room temperature. Later, the membrane was incubated with secondary antibody, goat-anti-mouse horseradish peroxidase-linked Antibody (1:2000) in 5% w/v nonfat dry milk/ TTBS buffer for 1 hour at room temperature. The membrane was washed three times for 5 min each with 1x TTBS buffer. The membrane was exposed to chemiluminescence reagent (West Pico) for 1 min and the results were generated on X-ray film using an autoradiography cassette.

Isolation of *Arabidopsis* leaf mesophyll protoplasts.

Four-week-old A. thaliana plant leaves were used for mesophyll protoplast preparation (Yoo, Cho et al. 2007). Multiple plants were pooled for protoplast isolation to generate

2×10^5 protoplasts for each of the sample types. Protoplast samples were visualized for fluorescence using a confocal microscope.

Laser Capture Microdissection (LCM) observations of anthers.

The flowers of *Col-0* and MVLG_01732 Δ SP-mCherry transgenic lines were dissected to separate the anthers from other floral parts and mounted using Phosphate buffer solution. Images of the anthers were acquired with a 5x objective on a Leica LMD6500 microscope.

SEM observations of pollen. The anthers of *Col-0* and MVLG_01732 Δ SP-mCherry transgenic lines were sputter coated with gold for 300 s. The material was examined with Zeiss EVO 40 scanning electron microscope (SEM) and the results were documented with smartSEM software at 5.0KV.

Bacterial Infiltration assay. *Pseudomonas syringae* pv. *tomato* (*pst*) DC3000 strain (generously provided by Dr. Alan Collmer and Morgan Carter, Cornell University) was used for bacterial infiltration assay. Cells were grown on King's B plates (2% proteose peptone, 0.2% potassium phosphate dibasic trihydrate, 1.5% agar, 1.4% (v/v) sterile glycerol, 0.5% (v/v) 1M MgSO₄) containing rifampicin (50 μ g/ml) at 28°C overnight. Cells were diluted to an OD₆₀₀ of 0.001 with 10mM MgCl₂ and leaf inoculated on the abaxial side of a 4-week old *A. thaliana* using a 1 ml syringe. Leaf punches (6mm) of the infected leaves were taken on day 0 and day 3 to quantify the bacterial count. Bacterial infiltration was carried out according to the protocol (Liu, Sun et al. 2015).

Results

Analysis of *Arabidopsis thaliana* transformants.

Each MVLG effector gene was expressed as two different types (see Figure 3-1): 1) MVLG-mCherry_pRI and 2) MVLG Δ SP-mCherry_pRI. Despite repeated attempts, generating MVLG_4106 Δ SP-mCherry_pRI construct was not successful. Although constructs with MVLG_6175_pRI and MVLG_398_pRI were generated, transgenic *Arabidopsis* lines were not yet generated with them and will be used by future researchers to investigate their function and subcellular localization.

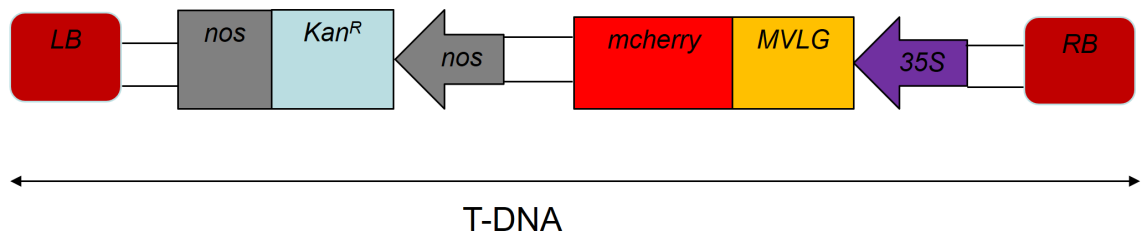


Figure 3- 1: Schematic representation of the T-DNA construct.

LB: left border, nos: nopaline synthase, Kan^R: kanamycin resistance-conferring gene, mCherry: fluorescent tag at the C-terminus of the MVLG effector gene, MVLG: effector gene, 35S: cauliflower mosaic virus promoter, RB: right border

Floral dipping generated 10 primary transformants for MVLG_01732 Δ SP-mCherry construct and 5 primary transformants for the full length fusion construct, MVLG_01732-mCherry. Likewise, MVLG_05720 Δ SP-mCherry floral dip generated 35 primary transformants and its full length fusion construct, MVLG_05720-mCherry generated 16 transformants. Despite multiple attempts generating, MVLG_04106 Δ SP-mCherry

remained not successful but the full-length version of the same effector, MVLG_04106-mCherry generated 14 primary transformants.

Homozygous *Arabidopsis* transformed seedlings from the T3 generation were grown and gDNA was extracted. PCR was carried out to confirm the presence of the gene and the fluorescent marker, mCherry fragment.

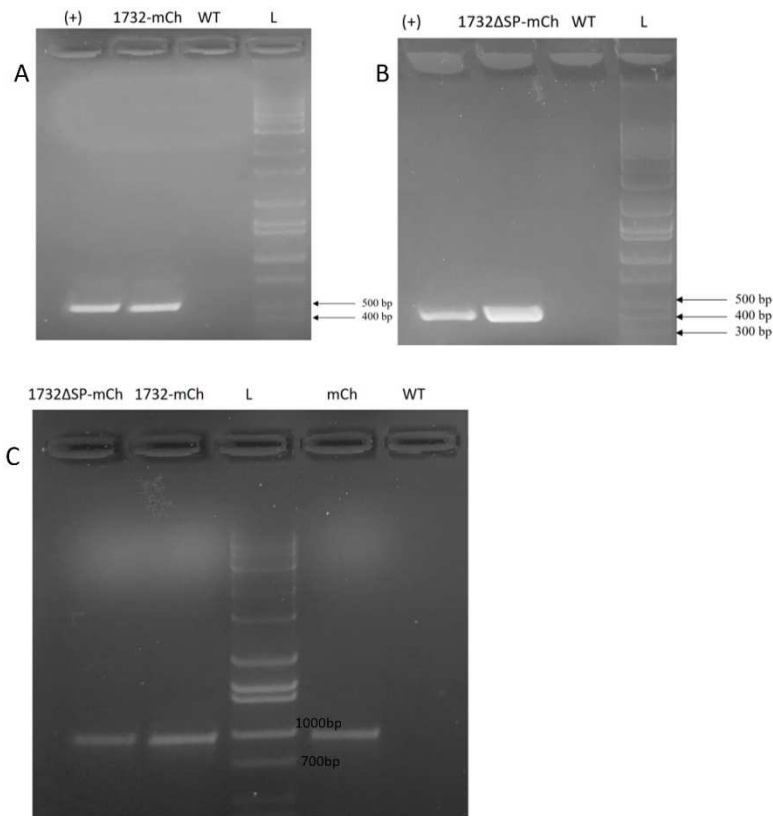
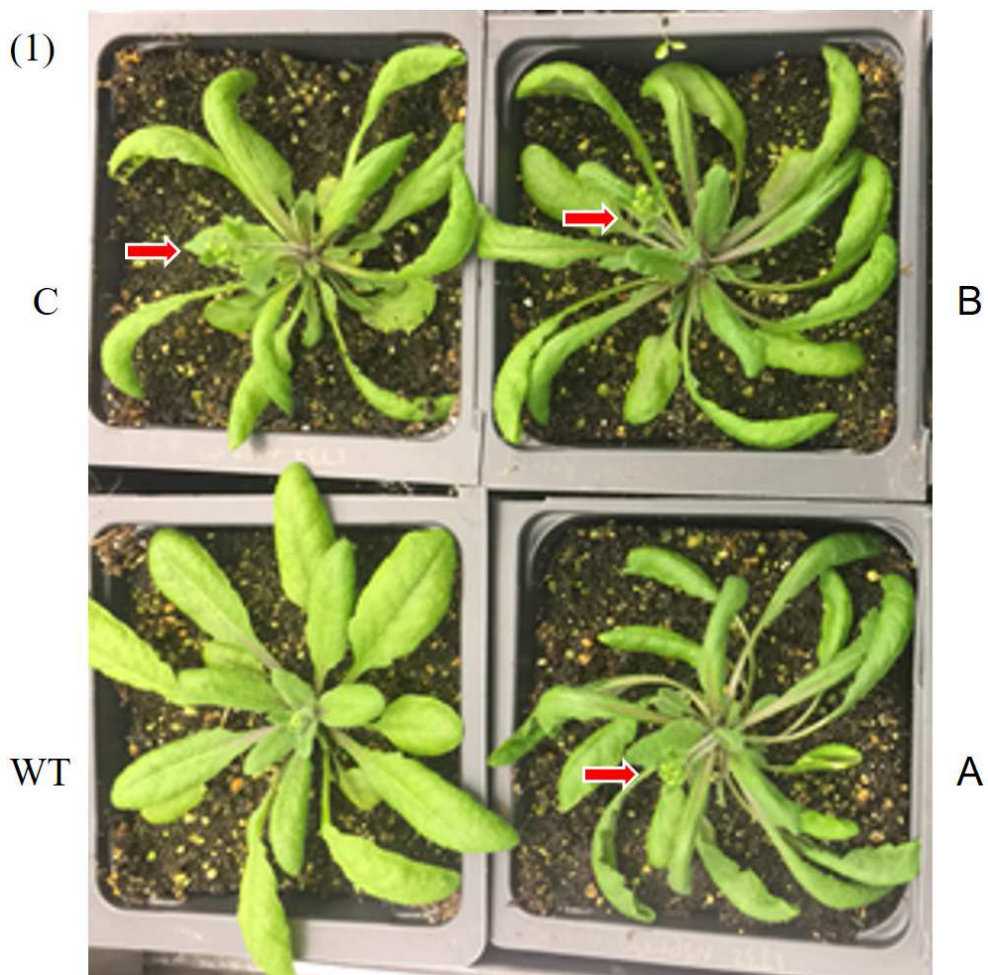


Figure 3- 2: PCR on *A. thaliana* lines expressing MVLG_01732 and mCherry.

A. Transgenic lines expressing MVLG_1732 gene (full length). B. Transgenic lines expressing MVLG_1732ΔSP gene (truncated). C. Transgenic lines expressing the mCherry fluorescent marker. (+): positive *Agrobacterium* clone that was used to transform *A. thaliana* lines. WT: Col-0 lines as negative control, L: DNA ladder, mCh: mCherry expressing lines as a positive control

Homozygous transgenic lines expressing MVLG_01732 have distinct phenotypes compared to wild type and other transgenic lines.

Stable constitutive expression of 1732 Δ SP-mCherry showed a curly leafy phenotype and early bolting in the transgenic *A. thaliana* plants (**Figure 3-3**). The leaves of 1732 Δ SP-mCherry expressing lines are long and narrow and have a pronounced downward curling towards their abaxial side when compared to wild-type leaves that are wide and flat.



(2)

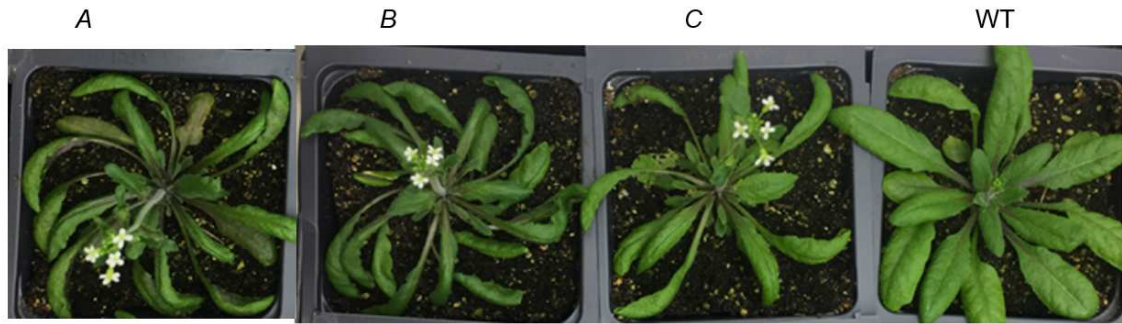


Figure 3- 3: Phenotypic abnormalities of the transgenic plants expressing MVLG_1732ΔSP-mCherry.

(1) A, B and C are MVLG_01732ΔSP-mCherry expressing transgenic lines with narrow leaves curled towards their abaxial side showing the early bolting (arrows indicate the bolts). Col-0 (WT) plant is showing normal leaves and is about to bolt. (2). Open flowers in A, B and C. Col-0 (WT) shows the emerging bud cluster.

Expression of MVLG_01732 in transgenic *A. thaliana*.

To determine the expression levels of the effector gene in the transgenic plants, real-time quantitative PCR was performed. The effector gene was not expressed in the wild-type *Col-0* plants as the control (**Figure 3-4**). The transcript abundance of the effector gene, MVLG_01732ΔSP-mCherry was at least 5.3 log₂-fold higher than that of the wild type.

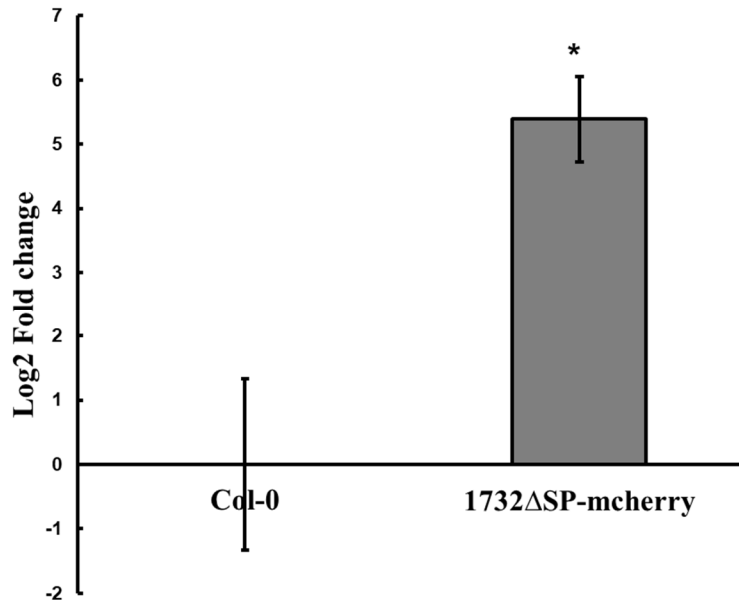


Figure 3- 4: Analysis of MVLG_01732ΔSP-mCherry transcript levels.

Analysis of MVLG_01732ΔSP-mCherry transcript levels in three independently generated transgenic *Arabidopsis* expression lines relative to transcript abundance of the *UBQ10* gene as determined by real-time qRT-PCR. Error bars indicate standard error of the mean of nine biological replicates. An *asterisk* indicates a statistically significant difference (students t-test; $p < 0.01$)

Immunoblot analysis demonstrated expression of the intact tagged protein.

Because of the phenotypic effects of effector gene, MVLG_01732 on the transgenic plant leaf morphology, protein extracts of the leaves from these plants along with those of wild type and mCherry-expressing lines were used as controls and were analyzed by immunoblot to determine the abundance of the intact tagged fusion protein (Figure 3-5). Anti-C-myc western blot revealed a single band signal at the expected size, ~53 kDa

(Figure 3-5, MVLG_1732 Δ SP-mCherry) suggesting that the fusion protein remained intact in the plant cells.

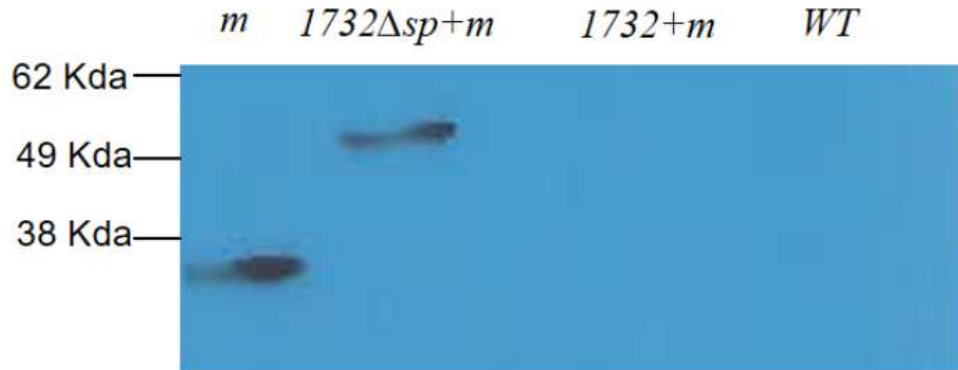


Figure 3- 5: Western blot of protein extractions from the transgenic lines.

The immunoblot was probed with anti-C-myc-antibody. Sizes of proteins in the molecular mass markers (Kda) are shown in the left. m (mCherry), MVLG_1732 Δ SP-mCherry, MVLG_1732-mCherry and WT (*Col-0*).

Subcellular localization of effectors using Floral dip. To gain insight about the function of the effector genes, we used confocal microscopy to detect resulting sub-cellular localization in the leaf epidermal cells of stable *A. thaliana* transgenic lines (**Figure 3-6**). The studies indicated MVLG_01732 Δ SP-mCherry tagged protein is localized to plasma membrane. We further generated the protoplasts from leaf epidermal leaves from these plants to confirm the localization. The protoplasts generated from mCherry plants, showed a dispersed signal in the cytoplasm along with nucleus whereas the protoplasts from MVLG_01732 Δ SP -mCherry plants, the signal was mainly concentrated on the plasma membrane (**Figure 3-7**). However, we could not find a signal for the full length fusion protein, MVLG_01732-mCherry (data not shown).

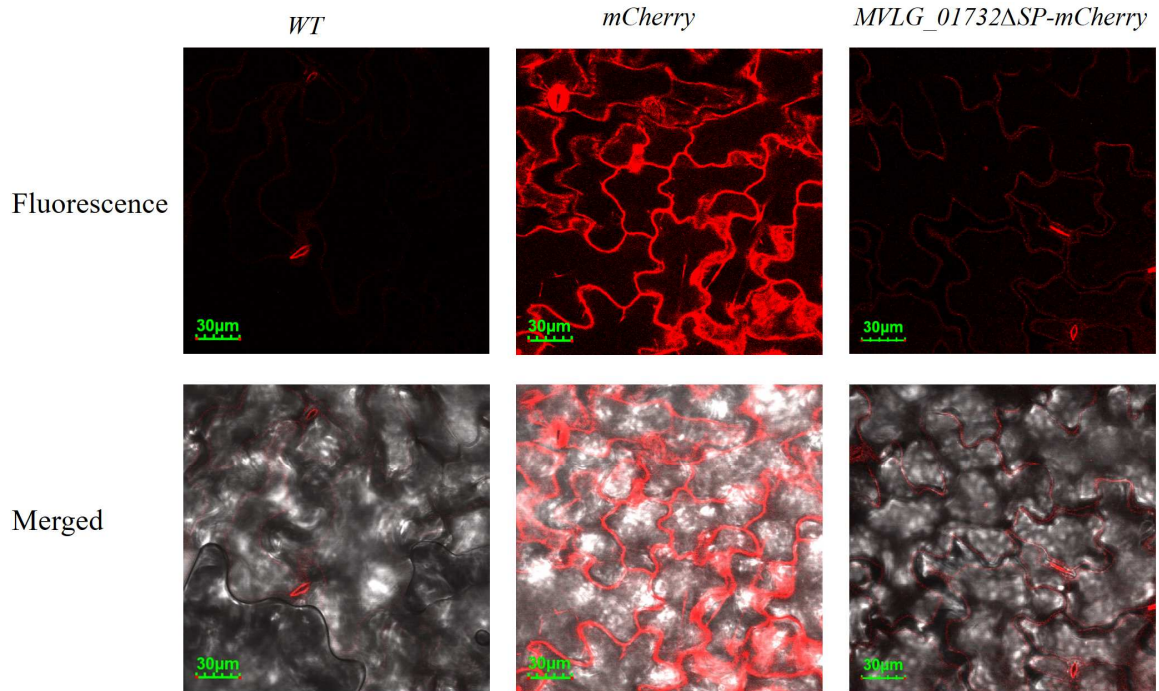


Figure 3- 6: Localization studies in the 14-day old seedlings.

Confocal images of live leaf epidermal cells of 14-day old homozygous stable transgenic plantlets expressing mCherry or MVLG_01732ΔSP-mCherry under the control of the CaMV 35S promoter. Left panel is the Fluorescence view and the right panel is the merged view.

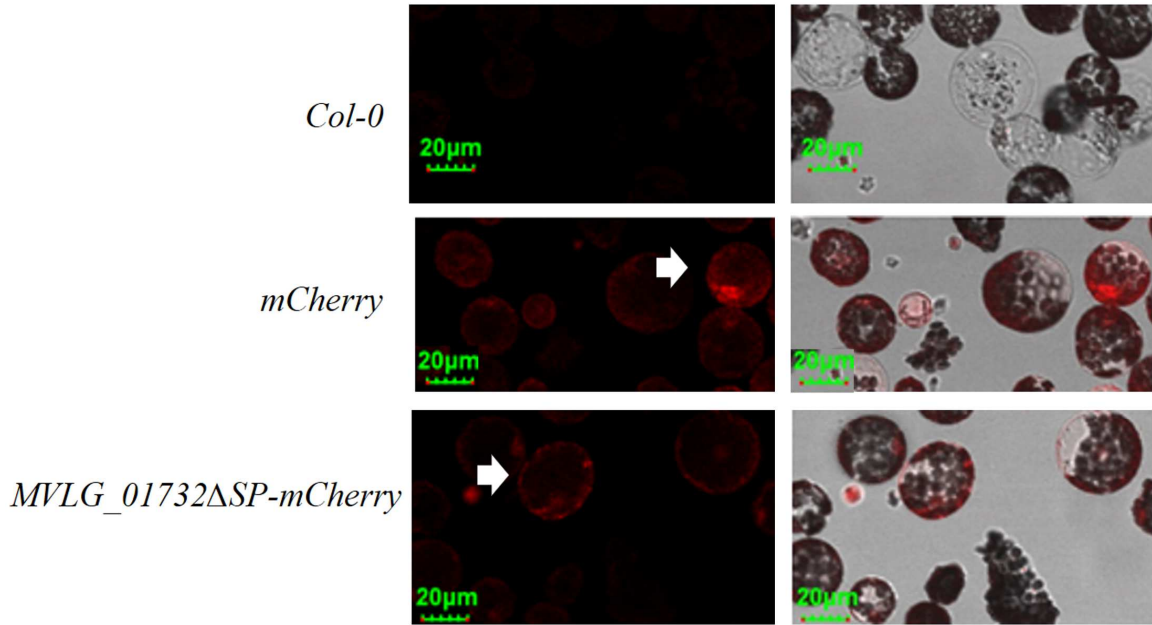


Figure 3- 7: Subcellular localization in protoplasts.

The protoplasts were isolated from the homozygous stable *A. thaliana* transgenic plants of mCherry and MVLG_01732ΔSP-mCherry along with wild type (*Col-0*) plants. Left panel is the Fluorescence view and the right panel is the merged view.

However, during *Microbotryum* infection, for the MVLG_01732 to localize on the plasma membrane of the plant host cell, it must be internalized by the host cell. One study has reported that the effectors bearing N-terminal RXLR motifs or functional variants of such motifs could bind to external phospholipids like phosphatidylinositol-3-phosphate (PI3P), located on lipid rafts could lead to endocytosis thereby allowing the effectors to interact with their host targets (Kale, Gu et al. 2010). Although the exact mechanism of the escape of effectors from the endosomes is currently not known, it is predicted that the moderate affinity of the effectors for PI3P make it possible for the binding on the outer surface and the dissociation inside the cell. Among the three studied

effectors, no obvious RXLR motifs were identified. However, variants of RXLR motifs that could function in the host cell entry were identified and reported in the table 3-3. However, the effectiveness of these RXLR variants in the cell entry need to be tested *in vitro* by experiments like leaf bombardment assays.

Table 3- 3: RXLR-like variants in the *M. lychnidis-dioicae* effector proteins

Effector	RXLR-like motif	Position of motif
4106	RPIT	54
5720	HKLC	55
5720	RDYP	69
5720	KQMQ	76
5720	KCMR	83
5720	HNLT	94
1732	RLIK	32
1732	RKIV	43
1732	KSIP	83

Since the infection target in the native host plant is the anther, we also looked for the morphological differences in the anthers and pollen in the transgenic plants (**Figure 3-8**). The anthers in the MVLG_01732 Δ SP-mCherry lines looked denser than the WT controls. Also, a structural difference in the morphology of anthers in these two lines was observed. Anthers of MVLG_01732 Δ SP-mCherry looked oblong in shape whereas the WT anthers seemed to have a defined heart shape. To better visualize the pollen, we

observed the pollen samples using scanning electron microscopy. Using this more sensitive measure, we did not observe any distinct differences in the pollen of lines expressing MVLG_01732 Δ SP-mCherry compared to wild type or mcherry expressing lines (**Figure 3-9**).

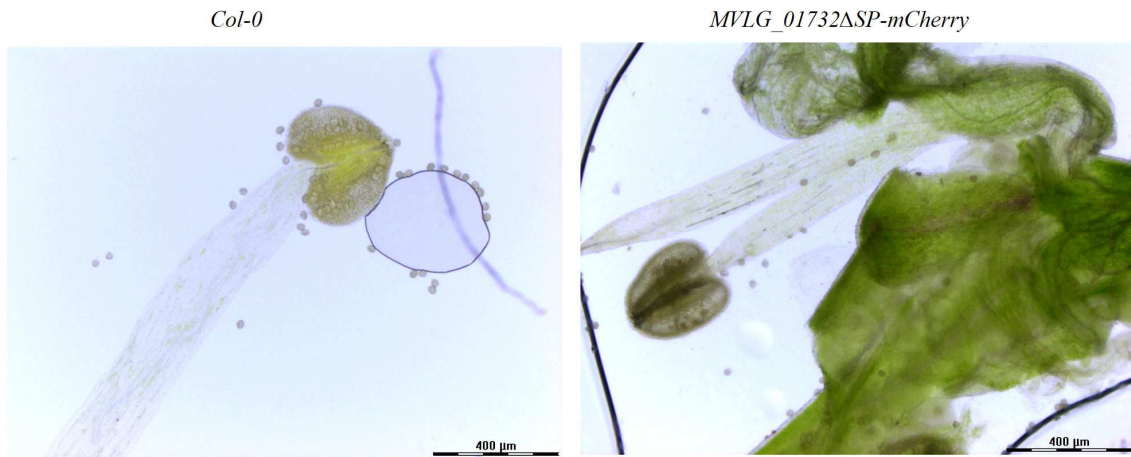
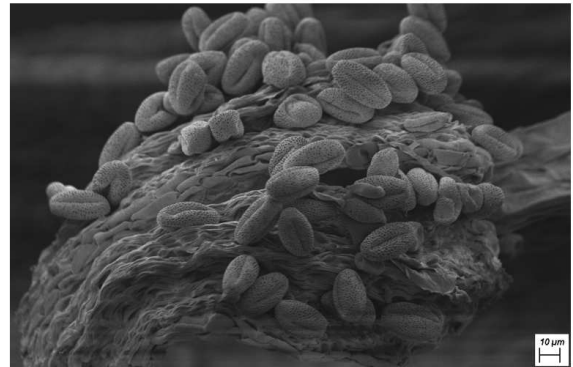
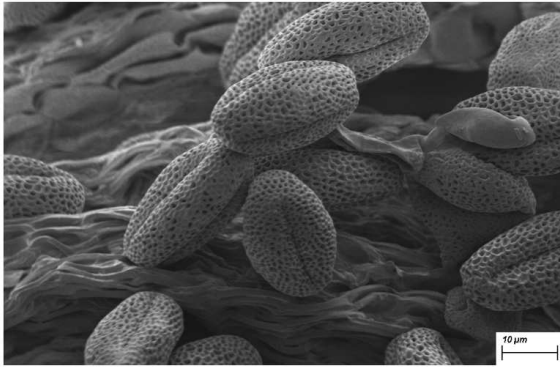


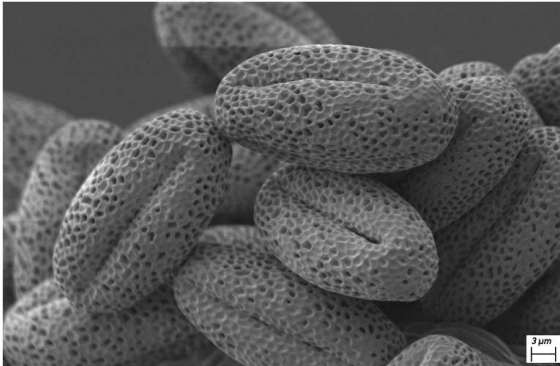
Figure 3- 8: Morphological differences in the anthers of WT (*Col-0*) and MVLG_01732 Δ SP-mCherry expressing plants.

Anthers of the transgenic lines are oblong and denser compared to wild type anthers that are heart-shaped and less dense. Images acquired with a 5x objective on a Leica LMD6500 microscope.

mCherry



MVLG_01732ΔSP-mCherry



Col-0

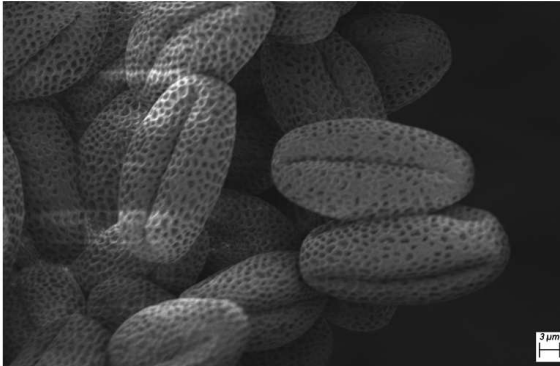


Figure 3- 9: SEM analysis of the pollen in the WT (*Col-0*), *mCherry*, and *MVLG_01732ΔSP-mCherry* lines.

Top panel shows anthers and pollen from *mCherry* transgenic lines. Middle panel shows *MVLG_01732ΔSP-mCherry* transgenic plants with anthers bearing the pollen. Bottom panel shows the pollen from the wild type (*Col-0*) plants. Images acquired with Zeiss EVO 40 scanning electron microscope.

***Pseudomonas* infiltration assay.** To determine if the *M. lychnidis-dioicae* fungal protein, MVLG_01732 has a role in conferring either resistance or susceptibility to the host plants, we performed *Pseudomonas* infection on these lines. Bacterial infiltration assays were conducted on stable *A. thaliana* transgenic lines constitutively expressing the effector genes, as well as on wild type and the line expression only the mCherry. Infection was carried out using wild-type *P. syringae* pv. *tomato* DC3000 and bacterial quantification was carried out. MVLG_01732 Δ SP-mCherry-expressing lines were more susceptible than the wildtype plants and the mCherry expressing plants. However, when compared to mCherry expressing lines, the difference was not statistically significant.

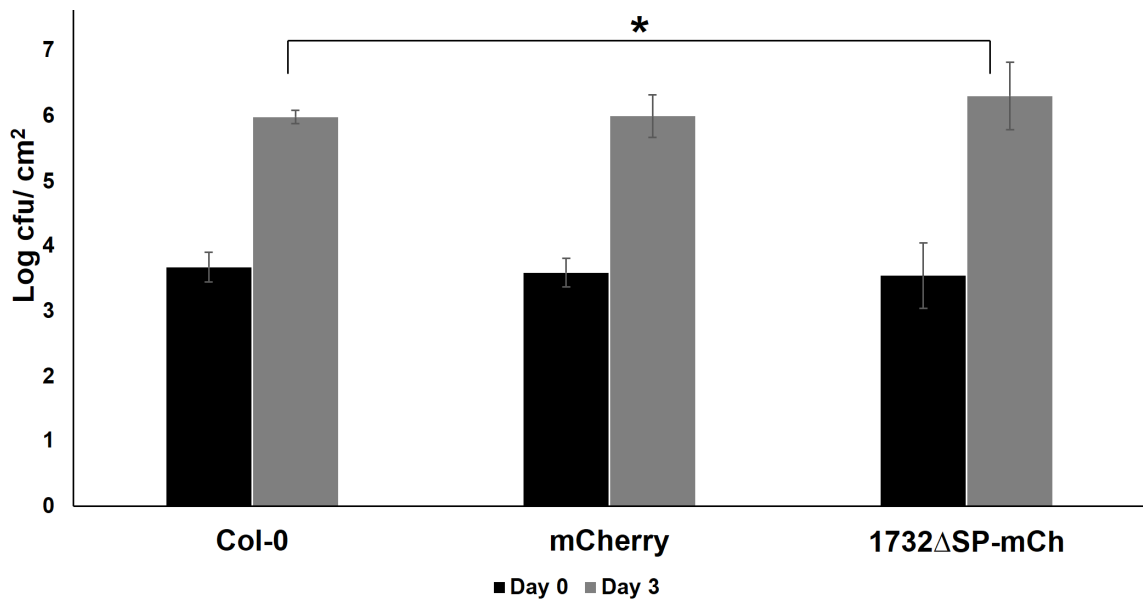


Figure 3- 10: MVLG_01732 promote the growth of *Pst* DC3000 growth.

Pst DC3000 growth in wildtype (*Col-0*), mCherry-only expressing plants, and MVLG_01732 Δ SP-mCherry plants. Bacterial suspension with an OD:0.001 was syringe infiltrated on the abaxial side of the leaves of 4-week old plants and the bacterial growth was quantified on day 0 and day 3. Six replicates were included for each genotype.

Statistical significance was performed by Student's t-test ($P < 0.01$) and an asterisk indicates a statistically significant difference compared with the wild-type plants. cfu, colony forming unit.

Discussion

For many years, *Microbotryum violaceum* species, being obligate biotrophic pathogens, remained a recalcitrant fungus with lack of appropriate genetic tools to study the pathogen (Toh and Perlin 2016). To better understand a host-pathogen interaction, studies at the molecular level remained a crucial step to be taken. Previous studies predicted that *Microbotryum lychnidis-dioicae* secretes 279 proteins, of which 19 bear the hallmarks of effectors and could potentially serve in the virulence and pathogenicity of this fungal pathogen (Perlin, Amselem et al. 2015). As a first step to understand these effectors, we confirmed their secretory nature *in vitro* and identified the potential targets from the host for a subset of these putative effectors (Kuppireddy, Uversky et al. 2017). This allowed us to further our understanding of the mechanism of infection. Although the *Microbotryum-Silene* pathosystem served as an ideal model to study several concepts, e.g., host shifts and plant sexually transmitted diseases, additional experimental approaches to understand the host-pathogen interactions at the molecular level are needed. In the current study, we used *A. thaliana* as a heterologous model to understand the localization and the potential function of the candidate effector, MVLG_01732. For these studies, we generated the stable transgenic *Arabidopsis* lines expressing the full-length mCherry fusion protein and the truncated mCherry fusion protein, lacking the signal peptide to understand the localization and the phenotypic abnormalities caused by the presence of this effector. To assess whether heterologous expression of these effectors

could affect host plant susceptibility to pathogens, we infected the transgenic lines with *Pseudomonas syringae*, *Pst* DC3000, a model pathogen for probing the disease susceptibility (Xin and He 2013) as *A. thaliana* is susceptible to this pathogen.

In this study, stable *Arabidopsis* transgenic lines were generated for the effectors MVLG_01732-mCherry, MVLG_01732 Δ SP-mCherry and with the mCherry construct by itself as a control. For the following effectors- MVLG_05720-mCherry, MVLG_05720 Δ SP-mCherry, MVLG_04106-mCherry, currently the T3 plants are growing and will be a source of investigation in future. We could not generate a MVLG_04106 Δ SP-mCherry construct even after multiple attempts, suggesting that it could be either toxic to the bacterial cells, alternatively this failure could be due to an unresolved cloning issue.

The subcellular localization studies on live leaf epidermal cells showed us that the truncated protein, MVLG_01732 Δ SP-mCherry localized to the plasma membrane (**Figure 3-6 and Figure 3-7**). The localization results are in agreement with respect to its host protein targets, Cellulose synthase interactive protein 1 (CSII) and calcium-dependent lipid binding protein (AtCLB), both membrane binding proteins identified from our previous study (Kuppireddy, Uversky et al. 2017). However, we could not find a signal for the full-length fusion protein containing the fungal protein secretion signal, MVLG_01732-mCherry (**Figure 3-6**). We further generated the protoplasts from the mesophyll leaf cells for better visualization. The mCherry expressing protoplasts showed the cytoplasmic and nuclear localization signal, and protoplasts generated from MVLG_01732-mCherry plants showed plasma membrane localization (**Figure 3-7**).

We conducted Real-time qRT experiments on the transgenic lines and confirmed the detection of the transcript levels in MVLG_01732 Δ SP-mCherry lines (**Figure 3-4**).

Western blot results yielded an intact protein for the truncated fusion protein but not for the full-length protein (**Figure 3-5**). Western blot results of the full-length fusion protein are in agreement with the results of subcellular localization for this protein, indicating why we did not see a fluorescence signal with the full-length fusion protein, suggesting that the translated protein product could be cleaved or degraded and thus rendered nonfunctional.

The *A. thaliana* transgenic lines expressing truncated fusion protein, MVLG_01732 Δ SP-mCherry showed an abnormal curly leafy phenotype toward the abaxial side of the leaf and exhibited early bolting when compared to wild-type and the mCherry expressing plants (**Figure 3-5**). The early bolting is particularly interesting as the fungus mainly resides in the flowers and the infected host plants also exhibit early bolting and flowers remain for prolonged periods when compared to healthy flowers (Kaltz and Shykoff 2001). We have further conducted LCM analysis on the anthers of the transgenic lines and the wild type lines which suggested that the anthers of the MVLG_01732 Δ SP-mCherry lines are oblong in shape and denser compared to wild type anthers that are heart-shaped and less dense (**Figure 3-8**). With this interesting result, we further conducted SEM analysis to see if there were morphological differences in the pollen structure of the MVLG_01732 Δ SP-mCherry lines compared to the wild type and mCherry lines. However, we did not find any obvious difference in the pollen morphology (**Figure 3-9**).

Bacterial infiltration assays with *Pst* DC3000 were conducted to understand if this effector would alter susceptibility of the *A. thaliana* host to this pathogen. The virulence was measured by the quantification of the bacterial count on day 3 (Katagiri, Thilmony et al. 2002). The results suggest that the MVLG_01732 Δ SP-mCherry lines show an increased susceptibility towards the *Pst* DC3000 pathogen compared to the wild type plants. We also compared the MVLG_01732 Δ SP-mCherry lines with the mCherry expressing lines and noticed a similar result, however, although the increase was not statistically significant (**Figure 3-10**).

Further experiments with *Pst* DC3000 mediated effector delivery could be conducted to confirm the results that we obtained through *in planta* effector expressing lines. Bacterial infiltration assays and subcellular localization studies with the rest of the transgenic lines expressing MVLG_05720 and MVLG_04106 generated from this study would be a valuable resource to increase our understanding at the molecular level of the host-pathogen interactions of this fungal pathosystem. co-immunoprecipitation assays could be conducted on these lines by future investigators to detect additional host interactors of these effectors in *A. thaliana*. Overall, the results from the studies described in this chapter would serve as a resource to guide the study the effector genes by using heterologous systems for the fungi that lack proper genetic tools for investigations in native hosts.

CHAPTER IV
EXPRESSION OF MVLG_01732 HIS TAGGED EFFECTOR IN *SILENE LATIFOLIA*
TO IDENTIFY NATIVE HOST INTERACTORS

Introduction

Microbotryum violaceum sensu lato (formerly *Ustilago violacea*) is a species complex, also called the anther smut fungus. It is a basidiomycete and an obligate pathogen of Caryophyllaceae (Pink family) although it can also be found infecting the anthers of many other dicotyledonous plants belonging to Dipsacaceae, Lamiaceae, and Lentibulariaceae (Bauer, Begerow et al. 2006, Kemler, Göker et al. 2006). Recently *Microbotryum lychnidis-dioicae* emerged as an independent lineage based on its host specificity on *Silene latifolia*.

Silene latifolia (white campion) is a dioecious plant and thus male and female flowers emerge from separate plants. It is a well-known model to study sexual dimorphism with XX/XY sex-determining system and evolution of sex chromosomes (Mrackova, Nicolas et al. 2008). The life cycle of the fungus begins when the fungal spores are dispersed by pollinators and land on the healthy host. The diploid teliospores undergo meiosis to give rise to four haploid spores that reproduce by budding to develop into yeast-like sporidia. Conjugation occurs between the sporidia of opposite mating type resulting in the infectious dikaryotic hyphae. The appressorium of dikaryotic hyphae penetrate the epidermis of the host thereby travelling intercellularly (Giraud, Yockteng et al. 2008).

Infected male flowers form smut teliospores instead of pollen in the anthers and the female flowers develop stamens which bear smutted anthers like male flowers (Uchida, Matsunaga et al. 2003). At the early stages of flower development in the uninfected *S. latifolia* flowers, the stamen and pistil primordia can be seen, but at the later stages, mechanisms like cell cycle arrest and cell death are observed in the stamens of female flowers and pistils of male flowers causing the suppression of these flower parts. In contrast, for infected female flowers, the suppression of stamen is released and cell death and cell cycle arrest was not observed (Kawamoto, Yamanaka et al. 2017). The ovules in the female infected flowers are aborted and sterile. The genes that are present on the Y-chromosome of the male flowers of *S. latifolia* are induced upon fungal infection (Uchida, Matsunaga et al. 2003).

Recently, the genome sequence of *M. lychnidis-dioicae* and transcriptomes of its interaction with the host *S. latifolia* have been produced (Perlin, Amselem et al. 2015, Toh, Chen et al. 2017). The study predicted a total of 7,364 protein-coding genes with 279 putative secreted proteins (SPs). Among the 279, 71 were predicted to be smaller than 250 amino acids and categorized as Small Secreted Proteins (SSPs). Of these, 46 were unique to *Microbotryum* with no sequence similarity to any other known proteins; 19 of the SSPs are significantly upregulated during infection (Perlin, Amselem et al. 2015). Thus, the SSPs have the hallmarks of fungal effectors, *i.e.*, proteins secreted by the fungus to manipulate its host. These findings led to further our study of the putative effector proteins at the molecular level and to the investigation of their role in pathogenesis. Our previously published research identified the targets for some of the candidate effectors allowing us to characterize their function (Kuppireddy, Uversky et al.

2017). To this end, we have expressed these effector genes in a non-host heterologous system to identify their sub-cellular localization that could potentially allow us to identify their function with respect to their targets (Chapter 3). To study any pathosystem, availability of tools such as genetic transformation, gene disruption approaches, and protein tagging, are essential techniques for molecular genetic analyses. Over the past 30 years, transformation of *Microbotryum* had limited or no success, but recently a robust and reproducible transformation system was established in our lab by Toh *et al* to overcome this hurdle (Toh, Treves et al. 2016). Following this approach, in this study, we have overexpressed the previously characterized effector gene, MVLG_01732, in *M. lychnidis-dioicae* and then infected the native host *S. latifolia* with the recombinant compatible mating-partners. This would allow us to conduct co-immunoprecipitation and mass spectrometry analysis in future, to identify any new interactors of the MVLG_01732 effector and to confirm the interactors, previously identified via yeast two hybrid analysis. The aim of this study was to use the *M. lychnidis-dioicae* -*S. latifolia* model system to investigate and functionally characterize the biological roles of the selected *M. lychnidis-dioicae* effector gene, MVLG_01732 in its host system. This is also the first *in planta* study to overexpress the effector genes in *M. lychnidis-dioicae* to infect the native host plant species.

Materials and methods

Growth conditions of *Silene latifolia*.

Silene latifolia seeds (generously provided by Dr. Tatiana Giraud, Université du Paris-Sud) were plated onto water agar. The plates were placed at 4 °C for 3 days for stratification allowing for the synchronization of the germination. After germination and

growth, 15-day old seedlings were transferred to 2” square pots filled with Sta-Green Potting Mix plus fertilizer (Lowe’s, item no. 0192430) and covered with domes to retain high humidity. The pots were placed in a growth chamber at 20 °C with 13 h light/11 h dark cycle and were watered every other day. At about 30 days old, the plants were repotted to cone-tainers. After they reached manageable growth, they were finally transferred to 7-inch square pots and were watered with Bloom-booster fertilizer (Miracle Gro, product no.1001921) every other day.

Generation of MVLG_01732-6x His in pPZP-MV-Hyg binary vector. The effector candidate tested in this study was MVLG_001732, for which sequence is available in the JGI Fungal Genome database (Nordberg, Cantor et al. 2014). To obtain the transgenic *M. lychnidis-dioicae* overexpressing this effector, a four-step cloning procedure was employed. First, PCR was used to amplify an upstream 1000 bp predicted promoter region, along with the entire coding sequence of this full-length protein in-frame with a sequence encoding poly-histidine (6x His) tag at C-terminus, using genomic DNA of *M. lychnidis-dioicae* as template; the primer pairs used are described in Table 4-1. For amplification by PCR, Takara Hot-Start ExTaq polymerase (Takara Bio USA, Inc.) was used. Standard PCR cycle was used with initial denaturation set at 94 °C for 4 min and 35 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 3 min, with a final extension time of 7 min at 72 °C. The product was held at 4 °C at the end of the cycle.

The PCR products were separated by gel electrophoresis through 1.3% agarose (Agarose Unlimited, Inc., Alacucha, Florida). The fragments were excised from the gel and purified using the Zymo Gel DNA recovery kit (Orange, CA, USA). The purified PCR fragment was cloned into pCR 2.1 TOPO entry vector (Invitrogen/Thermo Fisher,

Waltham, MA, USA). Plasmids that appeared to be larger than the TOPO vector were checked for the presence of amplicon by PCR. Positive TOPO entry clones are confirmed by PCR and then subjected to restriction digestion with *Bbv*CI and *Asc*I enzymes and cloned into pMF5-5h vector (kindly provided by Dr. Michael Feldbrugge, Universität Düsseldorf) under the control of T3 promoter, as a second step. For the clones that appeared larger than pMF5-5h vector, PCR was carried out to amplify the insert along with NOS terminator using Primers mentioned in Table 4-1. Thirdly, TOPO cloning was carried out again and positive TOPO clones were subjected to restriction digestion using *Pac*I and *Eco*RV. The excised product was purified and finally used to clone into the vector, pPZP-MV-HYG (Toh et al., 2016), which provides hygromycin resistance as a selectable marker for successful *M. lychnidis-dioicae* transformation. *Escherichia coli* strain DH5 α (Bethesda Research Laboratories, Bethesda, MD, USA), was utilized as recipient for all cloning purposes. Cells were plated on LB plates with kanamycin (50 μ g/ml) and incubated at 37 °C overnight. *E. coli* cells were grown at 37 °C in Circle Grow media (MP Biomedicals, LLC, Solon, OH, USA) and plasmid DNA was isolated from potential clones using the alkaline lysis procedure (Sambrook J 2001). The presence of the insert along with 6x His tag was confirmed by DNA sequencing (Eurofins Inc, Louisville, KY, USA).

Table 4- 1: Primers used to generate MVLG_01732-6x His pPzP-MV-Hyg construct

Name	Sequence (5' to 3')
MVLG1732BbvCIF	5'-GCTGAGGAATGAAGGGCGTGAAAGATG -3'
MVLG_01732AscI Rev	5'GGCGCGCCCTAGTGGTGGTGGTGGTGGGCG TGGATTTTGCCGGAGA-3'
1732PacFor	5'-TTAATTAACCTGTCCCCTCAGTTCATGT-3'
1732EcoRVRev	5'- GATATCCTTTCATCGTGCCAGATTT-3'

Transformation of *Agrobacterium* using electroporation.

The MVLG_01732-6x His pPZP -MV-HYG expression construct was introduced by electroporation into *Agrobacterium tumefaciens* strain EHA105 and later used to transform both haploid strains of *M. lychnidis-dioicae*, p1A1 and p1A2 (Toh et al., 2016). Competent cells of *A. tumefaciens* were made by following the Pikaard's Lab protocol (Pikaard). Transformation of competent *A. tumefaciens* cells was performed in a BioRad micropulser electroporator with voltage of 2.5 kV using the 25 uF capacitor and at 400-ohm settings (Mattanovich, Ruker et al. 1989). The competent cells were thawed on ice and 2 µl of the plasmid was added to the tube. The pre-chilled cuvettes were placed on ice and the mixture was transferred to the cuvette by using pre-chilled sterile tips. The bottom and the sides of the cuvette were wiped with a Kimwipe prior to starting the pulse until the time constant read for 4.7 sec. Then, 1ml of LB was added to the cuvette and pipetted up and down to mix the cells, which were then transferred to a 2 ml tube. The cells were incubated at 28 °C for 2 hours before plating 150 µl on to LB plates with kanamycin (50 µg/ml). The plates were incubated at 28 °C for 2 days. The plasmid DNA was isolated from the potential clones (Wang) and screened initially by PCR. One such positive clone for each construct type was selected for retransforming the *E. coli*. The construct was again checked for the presence of the insert by PCR and restriction digestion analysis to confirm the true *Agrobacterium* clone and is used for the transformation of *Microbotryum*.

***Agrobacterium*-mediated transformation of *Microbotryum* strains.**

Agrobacterium-mediated transformation of *Microbotryum* was carried out as published before in the Perlin lab (Toh, Treves et al. 2016). Transformation was carried out for both

compatible mating-partner strains of *M. lychnidis-dioicae*, p1A1 and p1A2, with selection on YPD + HYG/CEF plates (Yeast peptone dextrose solid media plates (YPD plate; 1% yeast extract, 2% peptone, 10% dextrose, 2% agar) containing 150 µg/ml hygromycin (Life Technologies) and 300 µg/ml Cefotaxime (Amresco)). Plates were incubated at 25 °C and the colonies were picked and transferred to YPD + HYG/CEF plates.

Genomic screening of the transformed *Microbotryum* colonies.

DNA extractions for the transformants were carried using the following procedure.

Microbotryum transformants were grown in 4 ml of liquid YPD + HYG/CEF media overnight. Cultures were pelleted by centrifugation at 14000 rpm for 1 min and the supernatants discarded. The pellet was washed with 1 ml of water, pelleted and resuspended in 0.5 ml of lysis buffer (0.5M NaCl, 0.01M EDTA at pH 8.0, 0.2M Tris-Cl at PH 7.5, 1% SDS) and 0.3 g of sterile 0.5 mm glass beads were added and quick spun. To this mixture, 250 µl PCI (25:24:1 v/v phenol:chloroform:isoamyl alcohol) was added and vortexed for 5 min. The cells were centrifuged at 14000 rpm for 3 min. Using wide bore tips, the upper phase was collected to a clean tube. Again 0.25 ml PCI was added and centrifuged as before. The upper phase was collected to a clean 1.6 ml conical tube and 1 ml of ice cold 100% ethanol was added and centrifuged for 5 min. Next, the wash step was repeated with 70% ice cold ethanol twice and the supernatant was discarded. The DNA pellet was vacuum-aspirated and dried on a heat block at 55 °C for 10 min and resuspended in 40 µl TE (pH 8.0). DNA concentration and purity were assessed using Nanodrop 2000™ UV-Vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA)

PCR Screenings. The genomic DNA extracted was used as template for PCR. Standard PCR was performed using Takara Hot-Start ExTaq polymerase enzyme (Takara Bio USA). The transformants were screened for the hygromycin resistance conferring gene, the His-tagged *MVLG_01732* insert and the house keeping gene, *mepA*, the latter used to confirm the identity of the *M. lychnidis-dioicae* cells. PCR products were separated via electrophoresis through 1% agarose gels to screen the transformants for those producing the band of the same size as controls. The PCR fragment was excised and purified. The presence of the overexpressor *MVLG_01732* gene, 6x His tag coding region, along with the 1000 bp upstream promoter was confirmed by DNA sequencing (Eurofins Inc, Louisville, KY, USA). Primers used for this screening and sequencing are found in Table 4-2.

Table 4- 2: Primers used for sequencing MVLG_01732-6x His pPzP-MV-Hyg

Name	Sequence (5' to 3')	Source
HygF Swathi	5'-AAAAGTTCGACAGCGTCTCC-3'	This Study
HygR Swathi	5'-ATTTGTGTACGCCCGACAGT-3'	This Study
1732PacFor	5'-TTAATTAACCTGTCCCCTCAGTTCATGT-3'	This Study
1732EcoRVRev	5'- GATATCCTTTCATCGTGCCAGATTT-3'	This Study
NosSeq	5'-ATCTCATAAATAACGTCATGC-3'	This Study
mepA 5'(4)	CTTTTGCGTAGGAAGAATGC	This Study
mepA 3'(2)	ACGGTGCCGAGGATGATTTGGA	This Study

Infection of *Silene latifolia*.

Freshly grown *M. lychnidis-dioicae* transformant strains of both mating types were scraped from YPD containing hygromycin and cefotaxime (YPD+HYG/CEF) plates. The optical density was adjusted to A_{600} to 1.0 and equal concentrations of mating types are mixed in a microcentrifuge tube. The inoculation treatments were performed on each seed that was plated on nutrient free water agar plates (2% agar). The plates were allowed to dry in the hood and then incubated at 14 °C for 48 hours. After 48 h the plates were transferred to a growth chamber. The infection was repeated after 5 days to increase the success rate of infection. The germinated seedlings after three weeks were planted in pots and watered every other day. Once the vegetative growth was significant, they were transferred to 7-inch pots and used Bloom booster fertilizer once in a week.

Protein Extraction and Western blotting.

Silene latifolia flowers and the buds from different stages were collected and ground to a fine powder under liquid nitrogen. Proteins were extracted in a SDS-based buffer containing 0.1M Tris-HCl, pH 8, 0.1% SDS, 2% β -mercaptoethanol, and 1X protease inhibitors (ICN Biomedicals Inc, CA). In parallel, protein samples were also extracted by NP-40-based lysis buffer (10% Glycerol, 50mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 100mM KCl, 2mM EDTA, 0.1% NP-40, 2mM DTT, 10mM NaF, 0.25mM NaVO₃, 1x protease inhibitor). Samples were sonicated in a water bath for 5 min at room temperature and kept on ice for 5 min. The lysate was loaded in a 1-ml syringe and sheared using a 23G needle. The samples were incubated on ice for 10 min and then the cell debris was pelleted by centrifugation for 30 min at 13000 rpm at 4 °C. The supernatant was separated into a separate tube and protein quantified

using the Bradford reagent (Hugo, L. et al. 2018). Equal amounts of protein (50 µg) were loaded into each well of a gradient polyacrylamide gel (4-12%) in 1x transfer buffer (Nupage) with 20% methanol buffer for 1 hour at 30 V and then electro transferred to a 0.45 µm nitrocellulose membrane. Immunoblotting was carried out by incubating the membrane with diluted (1:1000) primary antibody, mouse-anti-His monoclonal antibodies (Invitrogen #37-2900) in 5% BSA/ Tris-tween-buffer saline (TTBS) buffer at 4 °C, with gentle shaking, overnight. After transfer, the nitrocellulose membrane was washed three times with 1x TTBS buffer for 5 min at room temperature. Later, the membrane was incubated with secondary antibody, goat-anti-mouse horse radish peroxidase-linked Antibody (Santa Cruz #sc-2005) in 5% w/v nonfat dry milk/ TTBS buffer (1:2000) for 1 hour at room temperature. The membrane was washed three times for 5 min each with 1xTTBS buffer. The membrane was exposed to chemiluminescence reagent (west Pico) for 5 sec and the results were generated using an autoradiography cassette.

Results

Cloning into pPZP-MV-Hyg binary plasmid.

Initially, a two-step TOPO cloning procedure was employed for generating MVLG_01732-6x His- pPZP-MV-Hyg clone. First, the PCR product was amplified using primers containing the *Bbv*CI and *Asc*I sites and TOPO cloning was done. The 1689 bp insert was retrieved by using *Bbv*CI and *Asc*I restriction enzymes and cloned into pMF5-5h vector. From this clone, PCR was performed with primers having *Pac*I and *Eco*RV sites to amplify the 2630 bp insert along with the NOS terminator region from the pMF5-5h vector and recloned into the TOPO pCR2.1 vector. This TOPO clone was finally

digested with *PacI* and *EcoRV* and used for pPZP-Mv-Hyg cloning. Bona fide constructs were transformed into *A. tumefaciens*.

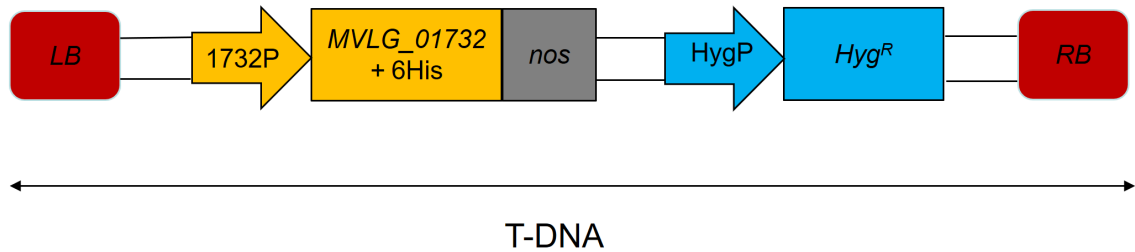


Figure 4- 1: Schematic representation of the MVLG_01732-6xHis-T-DNA construct.

LB: left border, 1732P: native promoter of MVLG_01732, MVLG_01732+6His: Histidine tag at the C-terminus of the MVLG_01732 effector gene, nos: nopaline synthase terminator, HygP: promoter for Hyg gene, Hyg^R: Hygromycin resistance-conferring gene, RB: right border.

Plasmid minipreps from *A. tumefaciens* transformants was carried out and PCR was performed on 8 such plasmid extractions, including bacterial plasmid, MVLG_01732-6xHis pZP-MV-HYG as a positive control. All such transformants that were tested were positive for the construct (**Figure 4-2**).

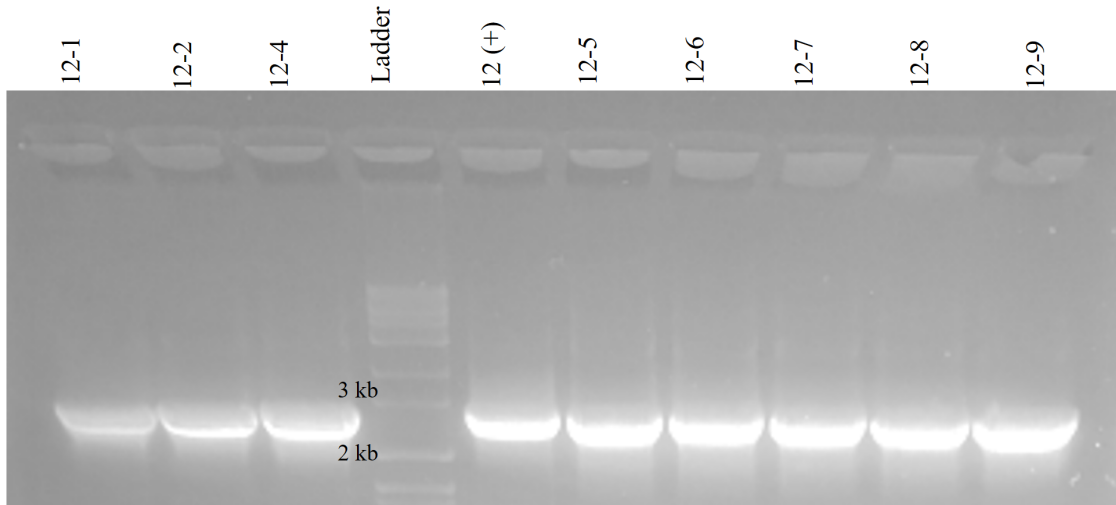


Figure 4- 2: PCR on eight *A. tumefaciens* transformants bearing MVLG_01732-6xHis pZP-Mv-Hyg.

Lanes 12-1 through 12-9 were the *A. tumefaciens* putative transformants. 12 (+) was the positive control from the original *E. coli* clone that was used to transform *A. tumefaciens*. Standard, HiLo linear DNA molecular weight marker (Minnesota Molecular).

As an additional confirmation, one positive *Agrobacterium* clone (12-5) was used to re-transform *E. coli* cells. *PacI* and *EcoRV* digestion of plasmid from *E. coli* released the band of correct size confirming the presence of the MVLG_01732-6xHis insert (**Figure 4-3**).

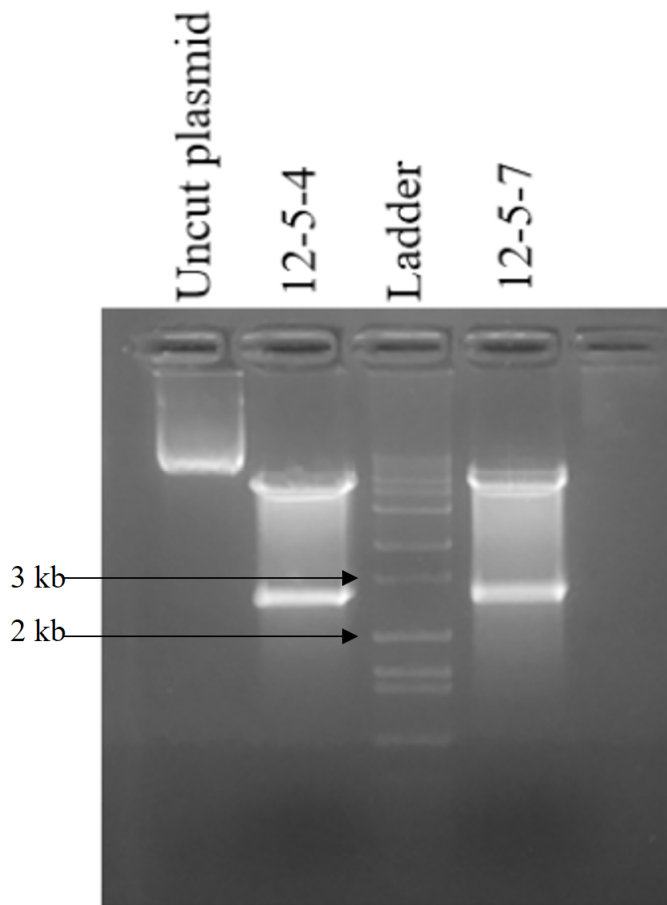


Figure 4- 3: Restriction digestion of plasmids from the *E. coli* re-transformed with a positive construct from *Agrobacterium*.

Lane 1. Uncut pZP-MV-HYG plasmid; lane 2. *E. coli* clone carrying MVLG_01732-6xHis pZP-MV-HYG construct; lane 3. HiLo DNA molecular weight marker; lane 4. Additional *E. coli* clone carrying MVLG_01732-6xHis pPZP-MV-HYG construct

Transformation of *M. lychnidis-dioicae* strains.

Transformation of *Microbotryum* strains, p1A1 and p1A2 was carried out using pPZP-MV-HYG, MVLG_01732-6xHis pPZP-MV-HYG binary plasmids. In the initial experiment, transformation of MVLG_01732-6xHis pPZP-MV-HYG in p1A1 yielded no colonies. A second experiment was performed to generate p1A1 transformants. Among

the two dilutions used, that using equal numbers of *Microbotryum* and *Agrobacterium* cells (M7:A7) was more effective in yielding *Microbotryum* transformants.

Table 4- 3: Number of *Microbotryum* colonies obtained for each transformation experiment

Experiment Number	Category	<i>M. lychnidis-dioicae</i>	<i>A. tumefaciens</i> strain	Colonies as per dilution		Total Number of Transformants
				M6:A7	M7:A7	
1	pPZP-MV-HYG	p1A1	Agro-3	1	1	2
		p1A2	Agro -3	0	3	3
	1732.pPZP-MV-HYG	p1A1	Agro 12-5	0	0	0
		p1A2	Agro 12-5	0	1	1
2	1732.pPZP-MV-HYG	p1A1	Agro 12-5	0	7	7
		p1A2	Agro 12-5	1	4	5

Genomic screening of the transformed *Microbotryum* colonies.

Genomic DNA was extracted from the *Microbotryum* transformants and PCR was used to screen for MVLG_01732-6xHis target (Figure 4-4) and Hyg target (Figure 4-5) (hereafter indicated as 1732p1A1 and 1732p1A2). 5µl of the PCR product was loaded onto the agarose gel in each lane, as indicated in Figure 4-4 and Figure 4-5 to verify the size and the remainder of the PCR sample was purified and sequenced.

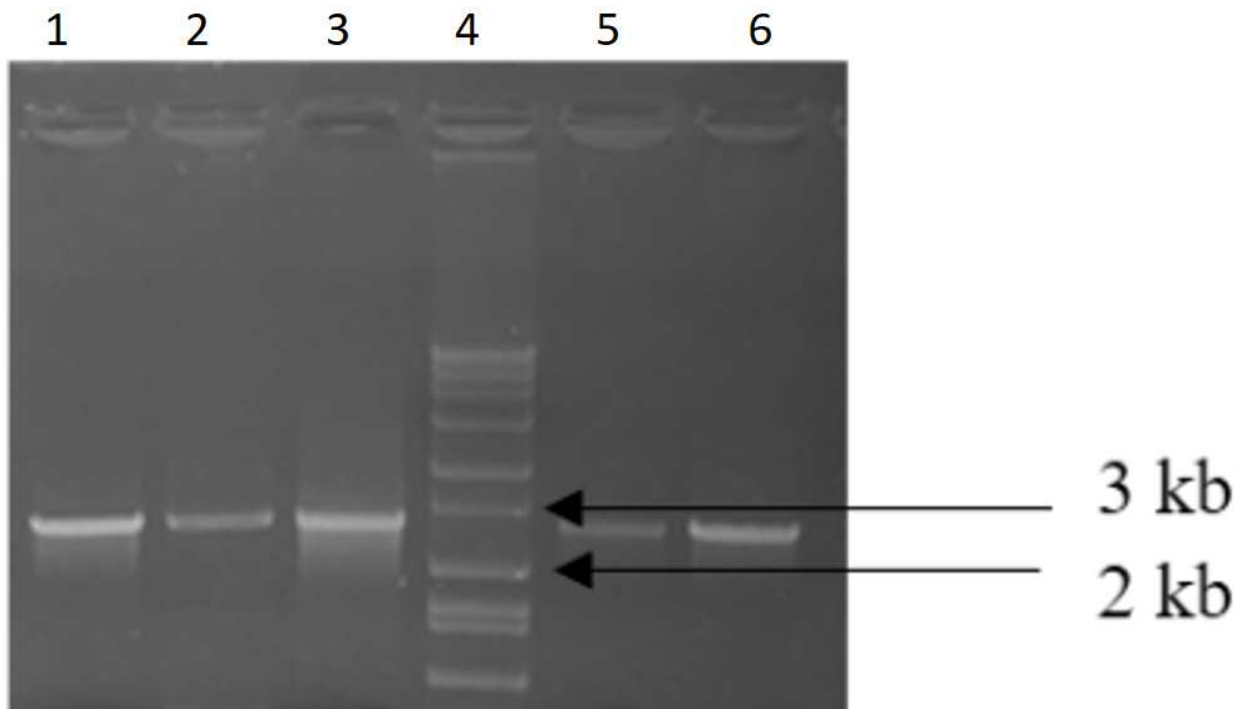


Figure 4- 4: Electrophoresis of MVLG_01732-6xHis PCR products.

Lane 1, MVLG_01732-6xHis p1A1-1; lane 2, MVLG_01732-6xHis p1A2-1; lane 3, MVLG_01732-6xHis pPZP-MV-HYG bacterial plasmid as a positive control for PCR; lane 4, HiLo DNA molecular weight ladder lane 5, MVLG_01732-6xHis p1A1-2; lane 6, MVLG_01732-6xHis p1A2-2.

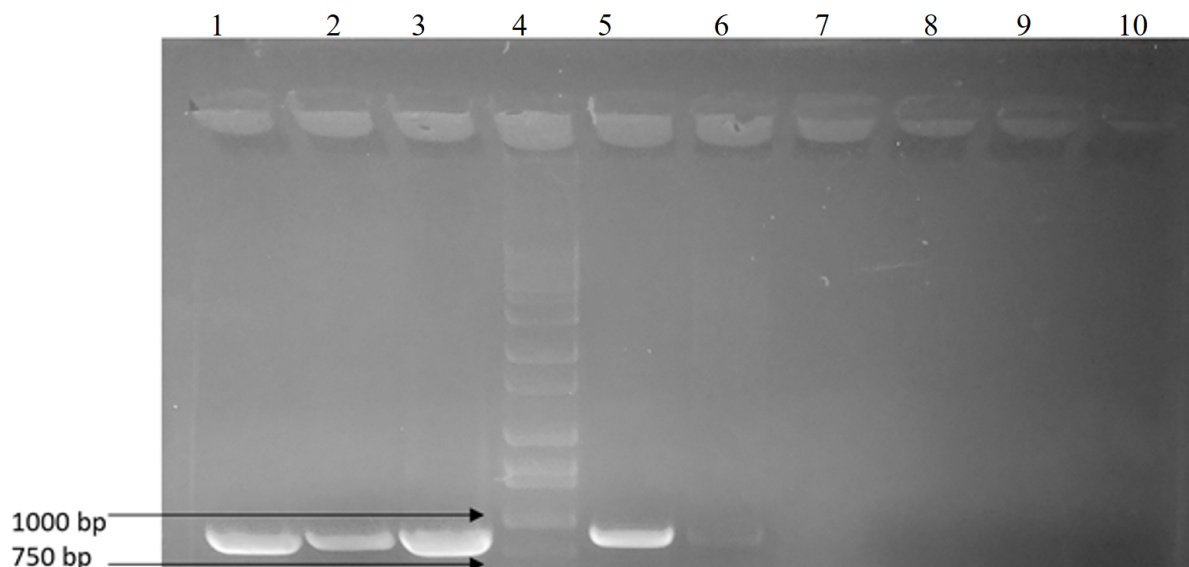


Figure 4- 5: Electrophoresis of HYG PCR products.

Lane 1, MVLG_p1A1-pMV (Vector only); lane 2, MVLG_p1A2-pMV (Vector only); lane 3, pPZP-MV-Hyg bacterial plasmid as a positive control; lane 4, DNA ladder; lane 5, MVLG_01732-6xHis p1A1-1; lane 6, MVLG_01732-6xHis p1A2-1; lane 7, empty lane; lane 8, MVLG_p1A1; lane 9, MVLG_p1A2; lane 10, No DNA control. Lanes 7-10 are negative controls.

Infection of *Silene latifolia*.

Infection of *S. latifolia* was carried out in three different batches. Some of the plants were infected twice to increase the success rate of infection and this is indicated in the following table 4-4. An infected *Silene* plant expressing MVLG_01732-6xHis fungal protein is shown in the **Figure 4-6**. The same plant bearing uninfected and infected flowers can be seen in the picture. The plants were also infected with *Microbotryum* strains carrying pPZP-MV-HYG (vector only) and pPZP-MV-HYG-GFP as control infections.

Table 4- 4: Number of plants infected along with the dose of infection

Batch No	Dose of infection	Strains used	Number of plants sown
1-A	1x infection	p1A1 only x 1732p1A2	3
		pp1A1 x p1A2	3
		pp1A1 -MV-HYG x p1A2-MV-HYG	2
		pp1A1 GFP x p1A2-MV-HYG	3
		Wild Type controls	3
1-B	2x infection	1732p1A1 x 1732p1A2	4
		pp1A1 x p1A2	3
		pp1A1 -MV-HYG x p1A2-MV-HYG	3
		pp1A1 GFP x p1A2-MV-HYG	3
2	2x infection	1732p1A1 x 1732p1A2	15
		p1A1 x p1A2	5
		p1A1 -MV-HYG x p1A2-MV-HYG	6
		p1A1 GFP x p1A2-MV-HYG	4
		Wild Type controls	2
3	2x infection	1732p1A1 x 1732p1A2	30
		p1A1 x p1A2	2
		Wild Type controls	2

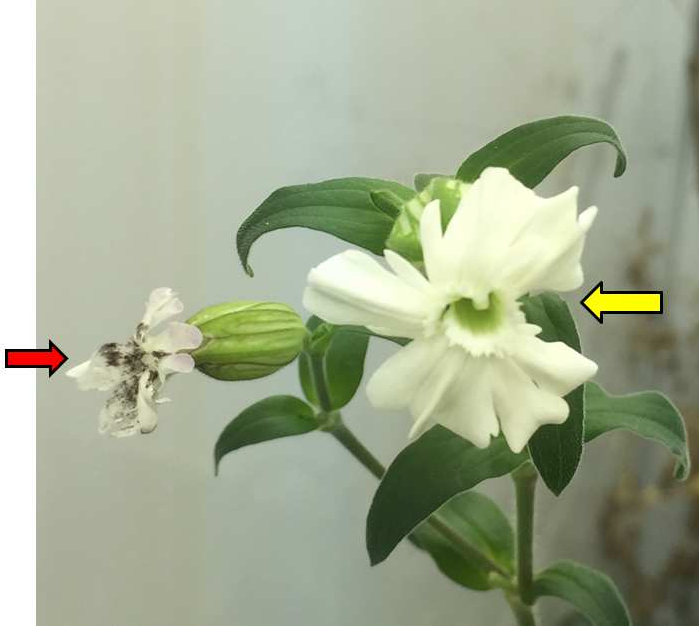


Figure 4- 6: Infected *Silene latifolia* plant expressing MVLG_01732-6xHis tagged protein.

Red arrow indicates the infected flower with the characteristic smutted appearance and the yellow arrow indicates the uninfected healthy flower coming from the same plant.

Western blotting.

Western blot analysis from proteins extracted from the different stages of flower buds from the 1732p1A1 x 1732p1A2, p1A1 x p1A2 and the p1A1-MV -HYG x p1A2-MV -HYG revealed a band with histidine antibody only in 1732p1A1 x 1732p1A2 protein sample extracted by using SDS method at the expected size of 18.15 Kda (Figure 4-7 (A), lane 1, arrow indicates the band). However, the NP-40-based protein extraction method of the same sample did not yield a band of the correct size (Figure 4-7 (A), lane 2). The membranes were stained with Ponceau S initially, to confirm the blotting efficiency as shown in Figure 4-7 (B).

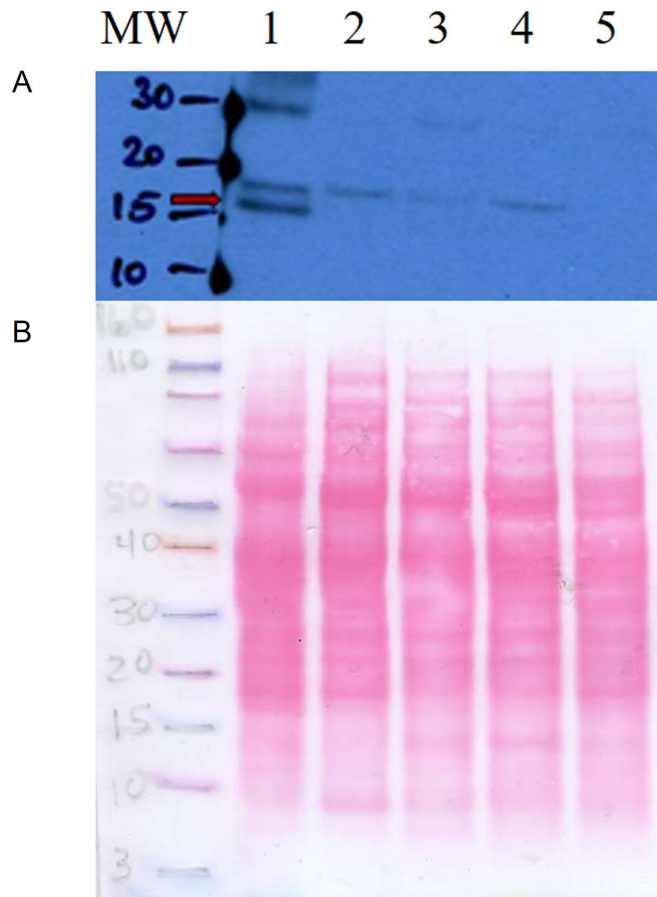


Figure 4- 7: Western blot analysis and Ponceau S staining of proteins

A. Western blots using anti-His antibody. B. Ponceau S staining of proteins after electroblotting of SDS-PAGE gels onto nitrocellulose membranes. Protein standard Molecular weight marker in Kda (Novex®), Lane 1, 1732p1A1 x 1732p1A2 (SDS buffer, arrow indicates the band), Lane 2, 1732p1A1 x 1732p1A2 (NP-40 buffer), Lane 3, p1A1-MV -HYG x p1A2-MV -HYG (SDS buffer), Lane 4, p1A1-MV -HYG x p1A2-MV -HYG (NP-40 buffer), Lane 5, p1A1 x p1A2. Lane 3, 4, and 5 serve as negative controls. Red arrow in Panel A indicates the His-tagged MVLG_01732.

Discussion

Host-pathogen interactions have been intensely studied over the past several decades. Small secreted proteins are effectors that are produced and secreted by pathogens and act as the main components in manipulating the host cell biology by rewiring the host immunity or by supporting pathogen growth and colonization (Sanchez-Vallet, Fouche et al. 2018). *Microbotryum violaceum* is a fungal species complex that has served as a popular model system in ecology, evolutionary biology and host speciation studies (Kemler, Göker et al. 2006, Le Gac, Hood et al. 2007, De Vienne, Refregier et al. 2009, Büker, Petit et al. 2013). It is an obligate parasite that infects more than 100 plant species from Caryophyllaceae family (Thrall, Biere et al. 1993). The best studied species in this group is *Microbotryum lychnidis-dioicae* that infects *Silene latifolia*. Gaining valuable insights from the genomic sequence and initial transcriptomics (Perlin et al., 2015; Toh et al. 2017, 2018), we published the first data on a subset of the candidate effectors from *M. lychnidis-dioicae* that might have a role in the virulence (Chapter 2 of this dissertation) (Kuppireddy, Uversky et al. 2017). We then conducted heterologous expression screens in *A. thaliana* expressing the fungal candidate effectors and observed that the plants expressing MVLG_01732 effector bolted and flowered earlier than the wild type plants, as well as exhibiting leaves that curled towards their abaxial side. Moreover, when challenged with a plant pathogen *Pseudomonas syringae*, the transgenic plants expressing MVLG_01732 effector tended to be more susceptible to the infection compared to the wild type plants (Chapter 3). These observations encouraged us to focus our efforts on this particular effector to characterize its function in its native host. To that end, we generated the His-tagged MVLG_01732 *M. lychnidis-dioicae* overexpression strains for

infection experiments in the native host, *Silene latifolia* to identify potential additional host plant interactors, if any, and to confirm the previously-identified interactors, AtCLB and CSI1. Based on the previous research conducted in Dr. Perlin's lab for generating the *Microbotryum* strain expressing GFP (Toh, Treves et al. 2016), this study further took the next step by expressing the fungal transgene in the infected *S. latifolia* host plants via *M. lychnidis-dioicae* overexpressor strains. To date, this is the first study to extract the fungal transgene protein product from infected *S. latifolia*, introduced through modified overexpressor *M. lychnidis-dioicae* strain (Figure 4-7, A). Considering the limited genetic toolset for this fungus, this serves as a crucial step to further progress on the effectors of this fungus that will ultimately lead to the in-depth understanding of pathogenicity for this fungus.

However, the mechanism whereby fungal protein MVLG_01732 gets internalized into the host plant cell requires further investigation. Specific host targeting cells that are unique to *Microbotryum* could also play a role in the uptake of these effectors into the host cytoplasm. So far, effector translocation and uptake by the host cells remains a mystery for fungal effectors. The unique amino acid sequences of many fungal effectors make it difficult to hypothesize how these effector proteins might end up in the site of action to contribute to the virulence of this pathogen during infection. One study reported that the necrotrophic fungal pathogen, *Pyrenophora tritici-repentis*, secretes a small protein toxin, Tox A, into the apoplast of the cell that can independently cross the plasma membrane from the extracellular space (Manning and Ciuffetti 2005). At this point, we have no evidence that this can occur with MVLG_01732 effector and such a possibility would be interesting to investigate.

Expression of His-tagged MVLG_01732 effector did not yield any observable altered phenotypic effects on the host plant other than the early bolting of the infected flowers (Figure 4-6), it will be interesting to investigate the effect *in planta* of other candidate effectors through this method. Future experiments on these His-tagged effectors expressed in infected plants by Co-IP and mass spectrometry analysis of the host interactors will provide a valuable knowledge on the host targets and will broaden our existing knowledge of the functions of these effectors. Currently, experiments to generate deletion mutants of *Microbotryum* strains for the candidate effectors using CRISPR/CAS system are in progress. We are currently at an exciting stage in unravelling the mysteries behind this fungal infection.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

The infection mechanism of anther smut fungus, *Microbotryum lychnidis-dioicae* on its host plant *Silene latifolia* has remained an unresolved mystery for several centuries. Part of the reason was due to the lack of appropriate genetic tools to manipulate both the pathogen and its host. This dissertation aimed to identify and characterize the key virulence determinants that play a role in the pathogenicity of this fungus. For this study, I selected for study a subset of small secreted proteins (SSPs) that are unique to this species (or even possibly, to this genus) and known to be highly upregulated during the infection. Next, I confirmed that they are secreted *in vitro* by using the yeast secretion trap assay. Identifying the kinds of target host partners of an effector, often helps in deducing the function of the effector. To that end, I identified the host interaction partners for several candidate effectors, with particular emphasis on MVLG_01732, shown to interact with the host AtCLB and CSII orthologues using yeast two hybrid screen approaches. My results also indicate that MVLG_05720 effector interacts with fungal proteins and might play a role in cell-cell communication, mating and the initial colonization of the pathogen. The protein product of MVLG_04106 could function as a transcriptional regulator that could modulate host immune response for the benefit of the pathogen.

Intrinsic disorder among the proteins is known to play a major role in protein-protein interactions and ultimately in the host pathogen interaction. Therefore, we analyzed the intrinsic disorder regions among all the 49 secreted proteins that are upregulated during this infection process. MVLG_04106 is characterized by high levels of predicted disorder, with 79.4%, showing the highest level among the four effectors selected for further investigation in this dissertation (Kuppireddy, Uversky et al. 2017). Moreover, the long-disordered regions (>50 residues) that were found in a number of bacterial effectors also may facilitate their delivery, help to avoid the host immune surveillance, and resemble to the host proteins (Marín, Uversky et al. 2013). The longest disordered region in MVLG_4106 is of 69 residues suggesting that it could act like the bacterial effectors in the mode of translocation. The MoRF analysis by MoRF_{CHIB} predicted that almost the entire chain of MVLG_04106 can act as a disorder-based binding region as this protein has two MoRFs, residues 1–70 and 87–104, that cover almost 83% of its sequence (Kuppireddy, Uversky et al. 2017). Interestingly, disordered regions are also known to be associated with various types of post-translational modification sites (Kurotani, Tokmakov et al. 2014). It is surprising that all three predicted post-translational modifications of MVLG_4106, were from the same predicted disordered regions (Kuppireddy, Uversky et al. 2017).

The biological role of MVLG_04106 remains elusive. Future studies targeting proteomic and metabolomic profiling of wild type and overexpression and pull-down assays in the host plant, *S. latifolia* could offer clues to identify the interacting partners, if any, in the host and thereby, the role of this effector in pathogenicity.

Heterologous expression of the candidate effector, MVLG_01732, in *A. thaliana* provided a key link tying the floral development of the host with this effector. The localization was found to be on the plasma membrane, thereby corroborating its predicted interaction with the target partner *CSII* which is found to have a role in pollen development and anther dehiscence. Our infection studies with the *P. syringae* indicated that the effector also plays a role in making the plant susceptible to the infection.

Finally, I took one last crucial step by functionally characterizing the MVLG_01732 effector *in planta* in the native host by expressing a tagged version of the protein, for both immunodetection and subsequent CoIP. I demonstrated the expression of the protein product *in planta*. Future Co-IP and mass spectrometry results would provide invaluable data on the host target partners and to unify the overall function of this effector.

In the future, the approaches employed in this study coupled with RNAi mediated gene silencing assays such as VIGS and the usage of type III secretion assays to deliver *Microbotryum* effectors to the plant cells could be used to demonstrate the function of the other candidate effectors. These approaches will provide an in-depth understanding of this plant-fungal interaction and also provide general insight about the molecular arsenal employed by the other obligate fungal pathogens.

REFERENCES

- Ahmed, A. A., C. Pedersen, T. Schultz-Larsen, M. Kwaaitaal, H. J. L. Jørgensen and H. Thordal-Christensen (2015). "The Barley Powdery Mildew Candidate Secreted Effector Protein CSEP0105 Inhibits the Chaperone Activity of a Small Heat Shock Protein." *Plant Physiology* **168**(1): 321-333.
- Ali, S. and G. Bakkeren (2011). "Fungal and oomycete effectors – strategies to subdue a host." *Canadian Journal of Plant Pathology* **33**(4): 425-446.
- Ali, S., J. D. Laurie, R. Linning, J. A. Cervantes-Chávez, D. Gaudet and G. Bakkeren (2014). "An Immunity-Triggering Effector from the Barley Smut Fungus *Ustilago hordei* Resides in an Ustilaginaceae-Specific Cluster Bearing Signs of Transposable Element-Assisted Evolution." *PLOS Pathogens* **10**(7): e1004223.
- Anderson, R. G., D. Deb, K. Fedkenheuer and J. M. McDowell (2015). "Recent Progress in RXLR Effector Research." *Molecular Plant-Microbe Interactions* **28**(10): 1063-1072.
- Antonovics, J. (2005). "Plant venereal diseases: insights from a messy metaphor." *New Phytol* **165**(1): 71-80.
- Antonovics, J., M. Hood and J. Partain (2002). "The Ecology and Genetics of a Host Shift: *Microbotryum* as a Model System." *The American Naturalist* **160**(S4): S40-S53.
- Antonovics, J., M. Hood and J. Partain (2002). "The ecology and genetics of a host shift: *microbotryum* as a model system." *Am Nat* **160 Suppl 4**: S40-53.
- Antonovics J, T. P., Jarosz AM (1997). "Genetics and the spatial ecology of species interactions: the *Silene-Ustilago* system."
- Arnold, K. B., L.; Kopp, J.; Schwede, T. (2006). " The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling." *Bioinformatics (Oxford, England)* **22**(2): 195-201.
- Badouin, H., M. E. Hood, J. Gouzy, G. Aguilera, S. Siguenza, M. H. Perlin, C. A. Cuomo, C. Fairhead, A. Branca and T. Giraud (2015). "Chaos of Rearrangements in the

Mating-Type Chromosomes of the Anther-Smut Fungus &Microbotryum lychnidis-dioicae." *Genetics* **200**(4): 1275.

Baker, H. G. H. (1947). "Sex in *Melandrium*." *Nature* **159**(4027).

Bauer, R., D. Begerow, J. P. Sampaio, M. Weib and F. Oberwinkler (2006). "The simple-septate basidiomycetes: a synopsis." *MYCOLOGICAL PROGRESS* **5**(1): 41-66.

Begerow, D., M. Stoll and R. Bauer (2006). "A Phylogenetic Hypothesis of Ustilaginomycotina Based on Multiple Gene Analyses and Morphological Data." *Mycologia* **98**(6): 906-916.

Bendtsen, D. J., H. Nielsen, G. von Heijne and S. Brunak (2004). "Improved prediction of signal peptides: SignalP 3.0." *Journal of Molecular Biology* **340**(4): 783-795.

Berardini, T. Z., L. Reiser, D. Li, Y. Mezheritsky, R. Muller, E. Strait and E. Huala (2015). "The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome." *Genesis* **53**(8): 474-485.

Bernasconi, G., J. Antonovics, A. Biere, D. Charlesworth, L. F. Delph, D. Filatov, T. Giraud, M. E. Hood, G. A. Marais, D. McCauley, J. R. Pannell, J. A. Shykoff, B. Vyskot, L. M. Wolfe and A. Widmer (2009). "Silene as a model system in ecology and evolution." *Heredity (Edinb)* **103**(1): 5-14.

Bernasconi, G., J. Antonovics, A. Biere, D. Charlesworth, L. F. Delph, D. Filatov, T. Giraud, M. E. Hood, G. A. B. Marais, D. McCauley, J. R. Pannell, J. A. Shykoff, B. Vyskot, L. M. Wolfe and A. Widmer (2009). "Silene as a model system in ecology and evolution." *Heredity* **103**: 5.

Biasini, M., S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli and T. Schwede (2014). "SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information." *Nucleic Acids Research* **42**(W1): W252-W258.

Billiard, S., M. Lopez-Villavicencio, B. Devier, M. E. Hood, C. Fairhead and T. Giraud (2011). "Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types." *Biol Rev Camb Philos Soc* **86**(2): 421-442.

Bolton, M. D., H. P. van Esse, J. H. Vossen, R. de Jonge, I. Stergiopoulos, I. J. Stulemeijer, G. C. van den Berg, O. Borrás-Hidalgo, H. L. Dekker, C. G. de Koster, P. J. de Wit, M. H. Joosten and B. P. Thomma (2008). "The novel *Cladosporium fulvum* lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species." *Mol Microbiol* **69**(1): 119-136.

Brefort, T., G. Doehlemann, A. Mendoza-Mendoza, S. Reissmann, A. Djamei and R. Kahmann (2009). "Ustilago maydis as a Pathogen." *Annu Rev Phytopathol* **47**: 423-445.

Brefort, T., S. Tanaka, N. Neidig, G. Doehlemann, V. Vincon and R. Kahmann (2014). "Characterization of the largest effector gene cluster of *Ustilago maydis*." *PLoS Pathog* **10**(7): e1003866.

Brefort Thomas, T., G. Doehlemann, A. Mendoza-Mendoza, S. Reissmann and A. Djamei (2009). "Ustilago maydis as a Pathogen." *Annual review of phytopathology* **47**: 423-445.

Büker, B., E. Petit, D. Begerow and M. E. Hood (2013). "Experimental hybridization and backcrossing reveal forces of reproductive isolation in *Microbotryum*." *BMC Evolutionary Biology* **13**(1): 224.

Catanzariti, A.-M., P. N. Dodds, G. J. Lawrence, M. A. Ayliffe and J. G. Ellis (2006). "Haustorially Expressed Secreted Proteins from Flax Rust Are Highly Enriched for Avirulence Elicitors." *The Plant Cell* **18**(1): 243-256.

Czechowski, T., M. Stitt, T. Altmann, M. K. Udvardi and W.-R. Scheible (2005). "Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in *Arabidopsis*." *Plant Physiology* **139**(1): 5.

Day, A. W. (1976). "Communication through fimbriae during conjugation in a fungus." *Nature* **262**(5569): 583-584.

de Castro, E., C. J. Sigrist, A. Gattiker, V. Bulliard, P. S. Langendijk-Genevaux, E. Gasteiger, A. Bairoch and N. Hulo (2006). "ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins." *Nucleic Acids Res* **34**(Web Server issue): W362-365.

de Jonge, R., M. D. Bolton and B. P. Thomma (2011). "How filamentous pathogens co-opt plants: the ins and outs of fungal effectors." *Curr Opin Plant Biol* **14**(4): 400-406.

de Jonge, R., H. Peter van Esse, A. Kombrink, T. Shinya, Y. Desaki, R. Bours, S. van der Krol, N. Shibuya, M. H. A. J. Joosten and B. P. H. J. Thomma (2010). "Conserved Fungal LysM Effector Ecp6 Prevents Chitin-Triggered Immunity in Plants." *Science* **329**(5994): 953-955.

De Vienne, D. M., G. Refregier, M. E. Hood, A. Guigue, B. Devier, E. Vercken, C. Smadja, A. Deseille and T. Giraud (2009). "Hybrid sterility and inviability in the parasitic fungal species complex *Microbotryum*." *J Evol Biol* **22**(4): 683-698.

Dean, P. (2011). "Functional domains and motifs of bacterial type III effector proteins and their roles in infection." *FEMS Microbiology Reviews* **35**(6): 1100-1125.

Denchev, C. M., T. Giraud and M. E. Hood (2009). "Three new species of anthericolous smut fungi on Caryophyllaceae." *Mycologia Balcanica* **6**(1/2): 79-84.

Depluvere, S., S. Devos and B. Devreese (2016). "The Role of Bacterial Secretion Systems in the Virulence of Gram-Negative Airway Pathogens Associated with Cystic Fibrosis." *Frontiers in Microbiology* **7**: 1336.

Djamei, A., K. Schipper, F. Rabe, A. Ghosh, V. Vincon, J. Kahnt, S. Osorio, T. Tohge, A. R. Fernie, I. Feussner, K. Feussner, P. Meinicke, Y. D. Stierhof, H. Schwarz, B. Macek, M. Mann and R. Kahmann (2011). "Metabolic priming by a secreted fungal effector." *Nature* **478**(7369): 395-398.

Dodds, P. N., G. J. Lawrence, A.-M. Catanzariti, M. A. Ayliffe and J. G. Ellis (2004). "The *Melampsora lini* AvrL567 Avirulence Genes Are Expressed in Haustoria and Their Products Are Recognized inside Plant Cells." *The Plant Cell* **16**(3): 755-768.

Dodds, P. N., G. J. Lawrence and J. G. Ellis (2001). "Contrasting modes of evolution acting on the complex N locus for rust resistance in flax." *Plant J* **27**(5): 439-453.

Dodds, P. N., G. J. Lawrence and J. G. Ellis (2001). "Six amino acid changes confined to the leucine-rich repeat beta-strand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax." *Plant Cell* **13**(1): 163-178.

Doehlemann, G., S. Reissmann, D. Assmann, M. Fleckenstein and R. Kahmann (2011). "Two linked genes encoding a secreted effector and a membrane protein are essential for *Ustilago maydis*-induced tumour formation." *Mol Microbiol* **81**(3): 751-766.

Doehlemann, G., K. van der Linde, D. Abmann, D. Schwammbach, A. Hof, A. Mohanty, D. Jackson and R. Kahmann (2009). "Pep1, a Secreted Effector Protein of *Ustilago maydis*, Is Required for Successful Invasion of Plant Cells." *PLOS Pathogens* **5**(2): e1000290.

Dosztanyi, Z., B. Meszaros and I. Simon (2009). "ANCHOR: web server for predicting protein binding regions in disordered proteins." *Bioinformatics* **25**(20): 2745-2746.

Drechsler, F., P. Schwinges and J. Schirawski (2016). "SUPPRESSOR OF APICAL DOMINANCE1 of *Sporisorium reilianum* changes inflorescence branching at early stages in di- and monocot plants and induces fruit abortion in *Arabidopsis thaliana*." *Plant Signaling & Behavior* **11**(5): e1167300.

- Edwards, M. C. and D. J. F. Bowling (1986). "The growth of rust germ tubes towards stomata in relation to pH gradients." *Physiological and Molecular Plant Pathology* **29**(2): 185-196.
- Ellis, J., A. M. Catanzariti and P. Dodds (2006). "The problem of how fungal and oomycete avirulence proteins enter plant cells." *Trends Plant Sci* **11**(2): 61-63.
- Ellis, J. G., P. N. Dodds and G. J. Lawrence (2007). "The role of secreted proteins in diseases of plants caused by rust, powdery mildew and smut fungi." *Curr Opin Microbiol* **10**(4): 326-331.
- Faulkner, C. and S. Robatzek (2012). "Plants and pathogens: putting infection strategies and defence mechanisms on the map." *Current Opinion in Plant Biology* **15**(6): 699-707.
- Flor, H. H. (1947). "Host-parasite interactions in flax rust: its genetics and other implications." *Phytopathology* **45**: 680-685.
- Fontanillas, E., M. E. Hood, H. Badouin, E. Petit, V. Barbe, J. Gouzy, D. M. de Vienne, G. Aguilera, J. Poulain, P. Wincker, Z. Chen, S. S. Toh, C. A. Cuomo, M. H. Perlin, P. Gladieux and T. Giraud (2015). "Degeneration of the Nonrecombining Regions in the Mating-Type Chromosomes of the Anther-Smut Fungi." *Molecular Biology and Evolution* **32**(4): 928-943.
- Fortuna, T. M., A. Snirc, H. Badouin, J. Gouzy, S. Siguenza, D. Esquerre, S. Le Prieur, J. A. Shykoff and T. Giraud (2016). "Polymorphic Microsatellite Markers for the Tetrapolar Anther-Smut Fungus *Microbotryum saponariae* Based on Genome Sequencing." *PLOS ONE* **11**(11): e0165656.
- Gan, P. H. P., M. Rafiqi, J. G. Ellis, D. A. Jones, A. R. Hardham and P. N. Dodds (2010). "Lipid binding activities of flax rust AvrM and AvrL567 effectors." *Plant Signaling & Behavior* **5**(10): 1272-1275.
- Ghareeb, H., F. Drechsler, C. Lofke, T. Teichmann and J. Schirawski (2015). "SUPPRESSOR OF APICAL DOMINANCE1 of *Sporisorium reilianum* Modulates Inflorescence Branching Architecture in Maize and Arabidopsis." *Plant Physiol* **169**(4): 2789-2804.
- Gibson, A. K., E. Petit, J. Mena-Ali, B. Oxelman and M. E. Hood (2013). "Life-history strategy defends against disease and may select against physiological resistance." *Ecology and Evolution* **3**(6): 1741-1750.

- Giraud, T., R. Yockteng, M. Lopez-Villavicencio, G. Refregier and M. E. Hood (2008). "Mating system of the anther smut fungus *Microbotryum violaceum*: selfing under heterothallism." *Eukaryot Cell* **7**(5): 765-775.
- Giraud, T., R. Yockteng, M. López-Villavicencio, G. Refrégier and M. E. Hood (2008). "Mating System of the Anther Smut Fungus &em>Microbotryum violaceum: Selfing under Heterothallism." *Eukaryotic Cell* **7**(5): 765.
- Godfrey, D., H. Bohlenius, C. Pedersen, Z. Zhang, J. Emmersen and H. Thordal-Christensen (2010). "Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif." *BMC Genomics* **11**: 317.
- Gohre, V. and S. Robatzek (2008). "Breaking the barriers: microbial effector molecules subvert plant immunity." *Annu Rev Phytopathol* **46**: 189-215.
- Greenberg, J. T. and N. Yao (2004). "The role and regulation of programmed cell death in plant-pathogen interactions." *Cell Microbiol* **6**(3): 201-211.
- Grewal, T. S., B. G. Rossnagel, G. Bakkeren and G. J. Scoles (2008). "Identification of resistance genes to barley covered smut and mapping of the *Ruh1* gene using *Ustilago hordei* strains with defined avirulence genes." *CANADIAN JOURNAL OF PLANT PATHOLOGY-REVUE CANADIENNE DE PHYTOPATHOLOGIE* **30**(2): 277-284.
- Hemetsberger, C., C. Herrberger, B. Zechmann, M. Hillmer and G. Doehlemann (2012). "The *Ustilago maydis* Effector *Pep1* Suppresses Plant Immunity by Inhibition of Host Peroxidase Activity." *PLOS Pathogens* **8**(5): e1002684.
- Hood, M. E. (2002). "Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*." *Genetics* **160**(2): 457-461.
- Hugo, G., J. D. L., M. Caroline, P. M. B., L. Claire, S. Donald, M. Marie-Josée, P. Benjamin, D. Sébastien and S. Armand (2018). "Infection assays in *Arabidopsis* reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens." *Molecular Plant Pathology* **19**(1): 191-200.
- Jennersten, O. (1983). "Butterfly Visitors as Vectors of *Ustilago violacea* Spores between Caryophyllaceous Plants." *Oikos* **40**(1): 125-130.
- Jones, J. D. and J. L. Dangl (2006). "The plant immune system." *Nature* **444**(7117): 323-329.
- Kale, S. D., B. Gu, D. G. Capelluto, D. Dou, E. Feldman, A. Rumore, F. D. Arredondo, R. Hanlon, I. Fudal, T. Rouxel, C. B. Lawrence, W. Shan and B. M. Tyler (2010).

"External lipid PI3P mediates entry of eukaryotic pathogen effectors into plant and animal host cells." *Cell* **142**(2): 284-295.

Kaltz, O. and J. A. Shykoff (2001). "Male and female *Silene latifolia* plants differ in per-contact risk of infection by a sexually transmitted disease." *Journal of Ecology* **89**(1): 99-109.

Kamper, J., R. Kahmann, M. Bolker, L. J. Ma, T. Brefort, B. J. Saville, F. Banuett, J. W. Kronstad, S. E. Gold, O. Muller, M. H. Perlin, H. A. Wosten, R. de Vries, J. Ruiz-Herrera, C. G. Reynaga-Pena, K. Snetselaar, M. McCann, J. Perez-Martin, M. Feldbrugge, C. W. Basse, G. Steinberg, J. I. Ibeas, W. Holloman, P. Guzman, M. Farman, J. E. Stajich, R. Sentandreu, J. M. Gonzalez-Prieto, J. C. Kennell, L. Molina, J. Schirawski, A. Mendoza-Mendoza, D. Greilinger, K. Munch, N. Rossel, M. Scherer, M. Vranes, O. Ladendorf, V. Vincon, U. Fuchs, B. Sandrock, S. Meng, E. C. Ho, M. J. Cahill, K. J. Boyce, J. Klose, S. J. Klosterman, H. J. Deelstra, L. Ortiz-Castellanos, W. Li, P. Sanchez-Alonso, P. H. Schreier, I. Hauser-Hahn, M. Vaupel, E. Koopmann, G. Friedrich, H. Voss, T. Schluter, J. Margolis, D. Platt, C. Swimmer, A. Gnirke, F. Chen, V. Vysotskaia, G. Mannhaupt, U. Guldener, M. Munsterkötter, D. Haase, M. Oesterheld, H. W. Mewes, E. W. Mauceli, D. DeCaprio, C. M. Wade, J. Butler, S. Young, D. B. Jaffe, S. Calvo, C. Nusbaum, J. Galagan and B. W. Birren (2006). "Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*." *Nature* **444**(7115): 97-101.

Katagiri, F., R. Thilmony and S. Y. He (2002). "The *Arabidopsis Thaliana*-*Pseudomonas Syringae* Interaction." *The Arabidopsis Book / American Society of Plant Biologists* **1**: e0039.

Kawamoto, H., K. Yamanaka, A. Koizumi, A. Hirata and S. Kawano (2017). "Cell Death and Cell Cycle Arrest of *Silene latifolia* Stamens and Pistils After *Microbotryum lychnidis-dioicae* Infection." *Plant and Cell Physiology* **58**(2): 320-328.

Kemen, E., A. Kemen, A. Ehlers, R. Voegelé and K. Mendgen (2013). "A novel structural effector from rust fungi is capable of fibril formation." *Plant J* **75**(5): 767-780.

Kemen, E., A. C. Kemen, M. Rafiqi, U. Hempel, K. Mendgen, M. Hahn and R. T. Voegelé (2005). "Identification of a Protein from Rust Fungi Transferred from Haustoria into Infected Plant Cells." *Molecular Plant-Microbe Interactions* **18**(11): 1130-1139.

Kemler, M., M. Göker, F. Oberwinkler and D. Begerow (2006). "Implications of molecular characters for the phylogeny of the Microbotryaceae (Basidiomycota: Urediniomycetes)." *BMC Evolutionary Biology* **6**: 35-35.

- Kershaw, M. J. and N. J. Talbot (1998). "Hydrophobins and Repellents: Proteins with Fundamental Roles in Fungal Morphogenesis." *Fungal Genetics and Biology* **23**(1): 18-33.
- Kuppireddy, V. S., V. N. Uversky, S. S. Toh, M. C. Tsai, W. C. Beckerson, C. Cahill, B. Carman and M. H. Perlin (2017). "Identification and Initial Characterization of the Effectors of an Anther Smut Fungus and Potential Host Target Proteins." *Int J Mol Sci* **18**(11).
- Kurotani, A., A. A. Tokmakov, Y. Kuroda, Y. Fukami, K. Shinozaki and T. Sakurai (2014). "Correlations between predicted protein disorder and post-translational modifications in plants." *Bioinformatics* **30**(8): 1095-1103.
- Lawrence, G. J., P. N. Dodds and J. G. Ellis (2007). "Rust of flax and linseed caused by *Melampsora lini*." *Mol Plant Pathol* **8**(4): 349-364.
- Le Gac, M., M. E. Hood, E. Fournier and T. Giraud (2007). "PHYLOGENETIC EVIDENCE OF HOST-SPECIFIC CRYPTIC SPECIES IN THE ANTHR SMUT FUNGUS." *Evolution* **61**(1): 15-26.
- Liu, X., Y. Sun, C. J. Korner, X. Du, M. E. Vollmer and K. M. Pajerowska-Mukhtar (2015). "Bacterial Leaf Infiltration Assay for Fine Characterization of Plant Defense Responses using the *Arabidopsis thaliana*-*Pseudomonas syringae* Pathosystem." *J Vis Exp*(104).
- Lo Presti, L., D. Lanver, G. Schweizer, S. Tanaka, L. Liang, M. Tollot, A. Zuccaro, S. Reissmann and R. Kahmann (2015). "Fungal effectors and plant susceptibility." *Annu Rev Plant Biol* **66**: 513-545.
- Lutz, M., M. Piątek, M. Kemler, A. Chlebicki and F. Oberwinkler (2008). "Anther smuts of Caryophyllaceae: Molecular analyses reveal further new species." *Mycological Research* **112**(11): 1280-1296.
- Manning, V. A. and L. M. Ciuffetti (2005). "Localization of Ptr ToxA Produced by *Pyrenophora tritici-repentis* Reveals Protein Import into Wheat Mesophyll Cells." *The Plant Cell* **17**(11): 3203.
- Marín, M., V. N. Uversky and T. Ott (2013). "Intrinsic Disorder in Pathogen Effectors: Protein Flexibility as an Evolutionary Hallmark in a Molecular Arms Race." *The Plant Cell Online*.

- Mattanovich, D. D., F. Ruker, A. D. Machado, M. Laimer and F. Regner (1989). "Efficient transformation of *Agrobacterium* spp. by electroporation." *Nucleic Acids Research* **17**(16).
- Meszaros, B., I. Simon and Z. Dosztanyi (2009). "Prediction of Protein Binding Regions in Disordered Proteins." *Plos Computational Biology* **5**(5).
- Mrackova, M., M. Nicolas, R. Hobza, I. Negrutiu, F. Moneger, A. Widmer, B. Vyskot and B. Janousek (2008). "Independent origin of sex chromosomes in two species of the genus *Silene*." *Genetics* **179**(2): 1129-1133.
- Mueller, A. N., S. Ziemann, S. Treitschke, D. Assmann and G. Doehlemann (2013). "Compatibility in the *Ustilago maydis*-maize interaction requires inhibition of host cysteine proteases by the fungal effector Pit2." *PLoS Pathog* **9**(2): e1003177.
- Mueller, O., R. Kahmann, G. Aguilar, B. Trejo-Aguilar, A. Wu and R. P. de Vries (2008). "The secretome of the maize pathogen *Ustilago maydis*." *Fungal Genet Biol* **45 Suppl 1**: S63-70.
- Müller, O., P. H. Schreier and J. F. Uhrig (2008). "Identification and characterization of secreted and pathogenesis-related proteins in *Ustilago maydis*." *Molecular Genetics and Genomics* **279**(1): 27-39.
- Nirmala, J., T. Drader, P. K. Lawrence, C. Yin, S. Hulbert, C. M. Steber, B. J. Steffenson, L. J. Szabo, D. von Wettstein and A. Kleinhofs (2011). "Concerted action of two avirulent spore effectors activates Reaction to *Puccinia graminis* 1 (Rpg1)-mediated cereal stem rust resistance." *Proc Natl Acad Sci U S A* **108**(35): 14676-14681.
- Nordberg, H., M. Cantor, S. Dusheyko, S. Hua, A. Poliakov, I. Shabalov, T. Smirnova, I. V. Grigoriev and I. Dubchak (2014). "The genome portal of the Department of Energy Joint Genome Institute: 2014 updates." *Nucleic Acids Res* **42**(Database issue): D26-31.
- Nunn, C. L., J. L. Gittleman and J. Antonovics (2000). "Promiscuity and the primate immune system." *Science* **290**(5494): 1168-1170.
- O'Connell, R. J. and R. Panstruga (2006). "Tete a tete inside a plant cell: establishing compatibility between plants and biotrophic fungi and oomycetes." *New Phytol* **171**(4): 699-718.
- Pandey, D., S. R. C. K. Rajendran, M. Gaur, P. K. Sajeesh and A. Kumar (2016). "Plant Defense Signaling and Responses Against Necrotrophic Fungal Pathogens." *Journal of Plant Growth Regulation* **35**(4): 1159-1174.

Pejaver, V., W. L. Hsu, F. Xin, A. K. Dunker, V. N. Uversky and P. Radivojac (2014). "The structural and functional signatures of proteins that undergo multiple events of post-translational modification." *Protein Sci* **23**(8): 1077-1093.

Peng, K., P. Radivojac, S. Vucetic, A. K. Dunker and Z. Obradovic (2006). "Length-dependent prediction of protein intrinsic disorder." *Bmc Bioinformatics* **7**.

Peng, K., S. Vucetic, P. Radivojac, C. J. Brown, A. K. Dunker and Z. Obradovic (2005). "Optimizing long intrinsic disorder predictors with protein evolutionary information." *J Bioinform Comput Biol* **3**(1): 35-60.

Perlin, M. H., J. Amselem, E. Fontanillas, S. S. Toh, Z. Chen, J. Goldberg, S. Duplessis, B. Henrissat, S. Young, Q. Zeng, G. Aguilera, E. Petit, H. Badouin, J. Andrews, D. Razeq, T. Gabaldon, H. Quesneville, T. Giraud, M. E. Hood, D. J. Schultz and C. A. Cuomo (2015). "Sex and parasites: genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus." *BMC Genomics* **16**: 461.

Perlin, M. H., J. Amselem, E. Fontanillas, S. S. Toh, Z. Chen, J. Goldberg, S. Duplessis, B. Henrissat, S. Young, Q. Zeng, G. Aguilera, E. Petit, H. Badouin, J. Andrews, D. Razeq, T. Gabaldón, H. Quesneville, T. Giraud, M. E. Hood, D. J. Schultz and C. A. Cuomo (2015). "Sex and parasites: genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus." *BMC Genomics* **16**(1): 461.

Petersen, R. H. (1974). "The rust fungus life cycle." *The Botanical Review* **40**(4): 453-513.

Petersen, T. N., S. Brunak, G. von Heijne and H. Nielsen (2011). "SignalP 4.0: discriminating signal peptides from transmembrane regions." *Nat Methods* **8**(10): 785-786.

Petre, B. and S. Kamoun (2014). "How Do Filamentous Pathogens Deliver Effector Proteins into Plant Cells?" *PLOS Biology* **12**(2): e1001801.

Pikaard, C. "Preparation of *Agrobacterium* Competent cells."
<http://www.indiana.edu/~pikweb/PDFs%20and%20protocol%20files%20agrotransform.html>.

Pliego, C., D. Nowara, G. Bonciani, D. M. Gheorghe, R. Xu, P. Surana, E. Whigham, D. Nettleton, A. J. Bogdanove, R. P. Wise, P. Schweizer, L. V. Bindschedler and P. D. Spanu (2013). "Host-induced gene silencing in barley powdery mildew reveals a class of ribonuclease-like effectors." *Mol Plant Microbe Interact* **26**(6): 633-642.

Postma, J., T. W. Liebrand, G. Bi, A. Evrard, R. R. Bye, M. Mbengue, H. Kuhn, M. H. Joosten and S. Robatzek (2016). "Avr4 promotes Cf-4 receptor-like protein association with the BAK1/SERK3 receptor-like kinase to initiate receptor endocytosis and plant immunity." *New Phytol* **210**(2): 627-642.

Puhalla, J. E. (1968). "Compatibility Reactions on Solid Medium and Interstrain Inhibition in *USTILAGO MAYDIS*." *Genetics* **60**(3): 461-474.

Qutob, D., B. Kemmerling, F. Brunner, I. K ufner, S. Engelhardt, A. A. Gust, B. Luberaeki, H. U. Seitz, D. Stahl, T. Rauhut, E. Glawischnig, G. Schween, B. Lacombe, N. Watanabe, E. Lam, R. Schlichting, D. Scheel, K. Nau, G. Dodt, D. Hubert, M. Gijzen and T. N urnberger (2006). "Phytotoxicity and Innate Immune Responses Induced by Nep1-Like Proteins." *The Plant Cell* **18**(12): 3721-3744.

Raffaele, S. and S. Kamoun (2012). "Genome evolution in filamentous plant pathogens: why bigger can be better." *Nat Rev Microbiol* **10**(6): 417-430.

Redkar, A., R. Hoser, L. Schilling, B. Zechmann, M. Krzymowska, V. Walbot and G. Doehlemann (2015). "A Secreted Effector Protein of *Ustilago maydis* Guides Maize Leaf Cells to Form Tumors." *Plant Cell* **27**(4): 1332-1351.

Ridout, C. J., P. Skamnioti, O. Porritt, S. Sacristan, J. D. G. Jones and J. K. M. Brown (2006). "Multiple Avirulence Paralogues in Cereal Powdery Mildew Fungi May Contribute to Parasite Fitness and Defeat of Plant Resistance." *The Plant Cell* **18**(9): 2402-2414.

Romero, P., Z. Obradovic, X. H. Li, E. C. Garner, C. J. Brown and A. K. Dunker (2001). "Sequence complexity of disordered protein." *Proteins-Structure Function and Genetics* **42**(1): 38-48.

Rouxel, T., J. Grandaubert, J. K. Hane, C. Hoede, A. P. van de Wouw, A. Couloux, V. Dominguez, V. Anthouard, P. Bally, S. Bourras, A. J. Cozijnsen, L. M. Ciuffetti, A. Degrave, A. Dilmaghani, L. Duret, I. Fudal, S. B. Goodwin, L. Gout, N. Glaser, J. Linglin, G. H. J. Kema, N. Lapalu, C. B. Lawrence, K. May, M. Meyer, B. Ollivier, J. Poulain, C. L. Schoch, A. Simon, J. W. Spatafora, A. Stachowiak, B. G. Turgeon, B. M. Tyler, D. Vincent, J. Weissenbach, J. Amselem, H. Quesneville, R. P. Oliver, P. Wincker, M.-H. Balesdent and B. J. Howlett (2011). "Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations." *Nature Communications* **2**: 202.

Sambrook J, R. D. (2001). "Molecular cloning : a laboratory manual." Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sanchez-Vallet, A., S. Fouche, I. Fudal, F. E. Hartmann, J. L. Soyer, A. Tellier and D. Croll (2018). "The Genome Biology of Effector Gene Evolution in Filamentous Plant Pathogens." *Annu Rev Phytopathol* **56**: 21-40.

Schäfer, A. M., M. Kemler, R. Bauer and D. Begerow (2010). "The illustrated life cycle of *Microbotryum* on the host plant *Silene latifolia*." *Botany* **88**(10): 875-885.

Schirawski, J., B. Heinze, M. Wagenknecht and R. Kahmann (2005). "Mating type loci of *Sporisorium reilianum*: novel pattern with three a and multiple b specificities." *Eukaryot Cell* **4**(8): 1317-1327.

Schirawski, J., G. Mannhaupt, K. Munch, T. Brefort, K. Schipper, G. Doehlemann, M. Di Stasio, N. Rossel, A. Mendoza-Mendoza, D. Pester, O. Muller, B. Winterberg, E. Meyer, H. Ghareeb, T. Wollenberg, M. Munsterkotter, P. Wong, M. Walter, E. Stukenbrock, U. Guldener and R. Kahmann (2010). "Pathogenicity determinants in smut fungi revealed by genome comparison." *Science* **330**(6010): 1546-1548.

Schmidt, S. M., H. Kuhn, C. Micali, C. Liller, M. Kwaaitaal and R. Panstruga (2014). "Interaction of a *Blumeria graminis* f. sp. *hordei* effector candidate with a barley ARF-GAP suggests that host vesicle trafficking is a fungal pathogenicity target." *Mol Plant Pathol* **15**(6): 535-549.

Selin, C., T. R. de Kievit, M. F. Belmonte and W. G. D. Fernando (2016). "Elucidating the Role of Effectors in Plant-Fungal Interactions: Progress and Challenges." *Frontiers in Microbiology* **7**(600).

Semisi, S. T. and S. F. L. Ball (1989). "Infection of the pearl millet (*Pennisetum americanum*) inflorescence by the downy mildew fungus (*Sclerospora graminicola*)." *Plant Pathology* **38**(4): 571-576.

Shen, Q. H., Y. Saijo, S. Mauch, C. Biskup, S. Bieri, B. Keller, H. Seki, B. Ulker, I. E. Somssich and P. Schulze-Lefert (2007). "Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses." *Science* **315**(5815): 1098-1103.

Sigrist, C. J., E. de Castro, L. Cerutti, B. A. Cuche, N. Hulo, A. Bridge, L. Bougueleret and I. Xenarios (2013). "New and continuing developments at PROSITE." *Nucleic Acids Res* **41**(Database issue): D344-347.

Spanu, P. D., J. C. Abbott, J. Amselem, T. A. Burgis, D. M. Soanes, K. Stuber, E. Ver Loren van Themaat, J. K. Brown, S. A. Butcher, S. J. Gurr, M. H. Lebrun, C. J. Ridout, P. Schulze-Lefert, N. J. Talbot, N. Ahmadinejad, C. Ametz, G. R. Barton, M. Benjdia, P. Bidzinski, L. V. Bindschedler, M. Both, M. T. Brewer, L. Cadle-Davidson, M. M. Cadle-Davidson, J. Collemare, R. Cramer, O. Frenkel, D. Godfrey, J. Harriman, C. Hoede, B. C.

King, S. Klages, J. Kleemann, D. Knoll, P. S. Koti, J. Kreplak, F. J. Lopez-Ruiz, X. Lu, T. Maekawa, S. Mahanil, C. Micali, M. G. Milgroom, G. Montana, S. Noir, R. J. O'Connell, S. Oberhaensli, F. Parlange, C. Pedersen, H. Quesneville, R. Reinhardt, M. Rott, S. Sacristan, S. M. Schmidt, M. Schon, P. Skamnioti, H. Sommer, A. Stephens, H. Takahara, H. Thordal-Christensen, M. Vigouroux, R. Wessling, T. Wicker and R. Panstruga (2010). "Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism." *Science* **330**(6010): 1543-1546.

Stergiopoulos, I., J. Collemare, R. Mehrabi and P. J. De Wit (2013). "Phytotoxic secondary metabolites and peptides produced by plant pathogenic Dothideomycete fungi." *FEMS Microbiol Rev* **37**(1): 67-93.

Stergiopoulos, I. and P. J. de Wit (2009). "Fungal effector proteins." *Annu Rev Phytopathol* **47**: 233-263.

Tampakaki, A. P., E. Hatziloukas and N. J. Panopoulos (2009). *Plant Pathogens, Bacterial A2 - Schaechter, Moselio. Encyclopedia of Microbiology (Third Edition)*. Oxford, Academic Press: 655-677.

Tanaka, S., T. Brefort, N. Neidig, A. Djamei, J. Kahnt, W. Vermerris, S. Koenig, K. Feussner, I. Feussner and R. Kahmann (2014). "A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize." *eLife* **3**: e01355.

Thrall, P. H., A. Biere and J. Antonovics (1993). "Plant Life-History and Disease Susceptibility--The Occurrence of *Ustilago violacea* on Different Species within the Caryophyllaceae." *Journal of Ecology* **81**(3): 489-498.

Toh, S. S., Z. Chen, D. J. Schultz, C. A. Cuomo and M. H. Perlin (2017). "Transcriptional analysis of mating and pre-infection stages of the anther smut, *Microbotryum lychnidis-dioicae*." *Microbiology*.

Toh, S. S. and M. H. Perlin (2016). "Resurgence of Less-Studied Smut Fungi as Models of Phytopathogenesis in the Omics Age." *Phytopathology* **106**(11): 1244-1254.

Toh, S. S., D. S. Treves, M. T. Barati and M. H. Perlin (2016). "Reliable transformation system for *Microbotryum lychnidis-dioicae* informed by genome and transcriptome project." *Arch Microbiol* **198**(8): 813-825.

Toh, S. S., D. S. Treves, M. T. Barati and M. H. Perlin (2016). "Reliable transformation system for *Microbotryum lychnidis-dioicae* informed by genome and transcriptome project." *Archives of Microbiology* **198**(8): 813-825.

- Uchida, W., S. Matsunaga, R. Sugiyama, Y. Kazama and S. Kawano (2003). "Morphological development of anthers induced by the dimorphic smut fungus *Microbotryum violaceum* in female flowers of the dioecious plant *Silene latifolia*." *Planta* **218**(2): 240-248.
- Upadhyaya, N. M., R. Mago, B. J. Staskawicz, M. A. Ayliffe, J. G. Ellis and P. N. Dodds (2014). "A bacterial type III secretion assay for delivery of fungal effector proteins into wheat." *Mol Plant Microbe Interact* **27**(3): 255-264.
- V. Robert, G. S. a. J. S. (2005). "The MycoBank engine and related databases." <http://www.mycobank.org>.
- Van den Burg, H. A., S. J. Harrison, M. H. Joosten, J. Vervoort and P. J. de Wit (2006). "Cladosporium fulvum Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection." *Mol Plant Microbe Interact* **19**(12): 1420-1430.
- Van Esse, H. P., J. W. van't Klooster, M. D. Bolton, K. A. Yadeta, P. van Baarlen, S. Boeren, J. Vervoort, P. J. G. M. de Wit and B. P. H. J. Thomma (2008). "The Cladosporium fulvum Virulence Protein Avr2 Inhibits Host Proteases Required for Basal Defense." *The Plant Cell* **20**(7): 1948-1963.
- Voegelé, R. T. (2006). "Uromyces fabae: development, metabolism, and interactions with its host *Vicia faba*." *FEMS Microbiol Lett* **259**(2): 165-173.
- Wang, K. (2006). "Agrobacterium Protocols." *Methods in molecular Biology*TM **1**.
- Westergaard, M. M. (1958). "The mechanism of sex determination in dioecious flowering plants." *Advances in genetics* **9**: 217-281.
- Wollenberg, T. and J. Schirawski (2014). "Comparative Genomics of Plant Fungal Pathogens: The Ustilago-Sporisorium Paradigm." *PLoS Pathogens* **10**(7): e1004218.
- Xin, X. F. and S. Y. He (2013). "Pseudomonas syringae pv. tomato DC3000: a model pathogen for probing disease susceptibility and hormone signaling in plants." *Annu Rev Phytopathol* **51**: 473-498.
- Xue, B., R. L. Dunbrack, R. W. Williams, A. K. Dunker and V. N. Uversky (2010). "PONDR-FIT: A meta-predictor of intrinsically disordered amino acids." *Biochimica Et Biophysica Acta-Proteins and Proteomics* **1804**(4): 996-1010.
- Yoo, S.-D., Y.-H. Cho and J. Sheen (2007). "Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis." *Nature Protocols* **2**: 1565.

Zhang, W. J., C. Pedersen, M. Kwaaitaal, P. L. Gregersen, S. M. Morch, S. Hanisch, A. Kristensen, A. T. Fuglsang, D. B. Collinge and H. Thordal-Christensen (2012). "Interaction of barley powdery mildew effector candidate CSEP0055 with the defence protein PR17c." *Mol Plant Pathol* **13**(9): 1110-1119.

Zhang, X., R. Henriques, S. S. Lin, Q. W. Niu and N. H. Chua (2006). "Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method." *Nat Protoc* **1**(2): 641-646.

APPENDIX

Supplemental data for Chapter 2

Prediction of secreted proteins

Prediction of secreted proteins was performed using a combination of online tools:

TargetP1.1, SignalP3.0(Bendtsen, Nielsen et al. 2004)(Bendtsen, Nielsen et al. 2004)(Bendtsen, Nielsen et al. 2004)(Bendtsen, Nielsen et al. 2004)(Bendtsen, Nielsen et al. 2004)(Bendtsen, Nielsen et al. 2004)(Bendtsen, Nielsen et al. 2004)(Bendtsen, Nielsen et al. 2004)(Bendtsen, Nielsen et al. 2004)(Bendtsen, Nielsen, von Heijne, & Brunak, 2004)(Bendtsen, Nielsen, von Heijne, & Brunak, 2004)(Bendtsen, Nielsen, von Heijne, & Brunak, 2004)(Bendtsen, Nielsen, von Heijne, & Brunak, 2004)(Bendtsen, Nielsen, von Heijne, & Brunak, 2004)(Bendtsen, Nielsen, von Heijne, & Brunak, 2004), SignalP4.0, TMHMM2.0, PredGPI, Phobius, NucPred, Prosite and WoLF PSORT.

The initial absolute cutoffs for excluding non-secreted proteins are listed in the following table.

Criteria for secretome discovery.

Tools	Cutoff for non-SP	Remain in pool of probable secretome
TMHMM	>2 TMs	≤2 TMs
Phobius	>2TMs	≤2 TMs
Prosite	ER retention signal: 00014	-
PredGPI	Specificity of >99.5% using the General Model	Specificity of <99.5, where 99.0-99.5 is characterized as lowly probably for GPI linkage
NucPred	Threshold>0.8	Threshold<0.8

The following guidelines were then used in determining whether the protein was secreted or not. Out of the 6 tests for indicating SP localization, at least 4 tests should be positive for SP in order to remain in the probable SP pool.

Further guidelines for secretome discovery on the remaining pool of probable secretome.

Tools	Guideline for SP
TargetP	Predicted localization to “SP”
SignalP3.0 NN	D-score>0.43
SignalP3.0 HMM	Sprob>0.8
SignalP4.0	D-score>0.45
WoLF Psort	“Extr” listed as major neighbor
Phobius	“Y” for SP

Where there are ambiguous predictions, TMHMM and Phobius have to agree on the existence of 1-2 TM(s) on the protein to exclude it from the probable secretome pool. In addition, if a protein is predicted with “lowly probable” GPI linkage, but has TM predicted by TMHMM and/or Phobius around the same region, they serve as corroborating evidence for anchorage to the membrane.

Where there were contractions or insufficient evidence for determining secretome status, BLASTp and Pfam domains is used to establish probable orthologs, followed by referencing the UniProtKB and FunSecKB database for confirmation of localization of the orthologs, where available.

Analysis of intrinsic disorder predisposition and presence of functional sites in proteins encoded by 49 MVLG genes

For each protein, we present:

- 1) MVLG ID and disorder content evaluated by PONDR VSL2 (Peng, Radivojac et al. 2006).

- 2) Amino acid sequence. Position of secretion signal peptide found by SignalP4.1 (Petersen, Brunak et al. 2011) are shown by red bold font.
- 3) Results of intrinsic disorder predisposition analysis by four predictors of PONDR family, PONDR[®] VSL2 (Peng, Radivojac et al. 2006), PONDR[®] VLXT (Romero, Obradovic et al. 2001), PONDR[®] VL3 (Peng, Vucetic et al. 2005), and PONDR[®] FIT (Xue, Dunbrack et al. 2010).
- 4) PONDR[®] VSL2 disorder prediction statistics including location of predicted disordered regions.
- 5) Results of evaluation of the presence and localization of disorder-based binding sites conducted using the ANCHOR algorithm (Dosztanyi, Meszaros et al. 2009, Meszaros, Simon et al. 2009).
- 6) Presence of possible post-translational modification sites found by ModPred (Pejaver, Hsu et al. 2014). Prediction results with High Confidence are listed.
- 7) Presence of possible functions found by PROSITE (de Castro, Sigrist et al. 2006, Sigrist, de Castro et al. 2013).
- 8) Molecular models built using SWISS-MODEL [(Biasini, Bienert et al. 2014), (Arnold 2006)].

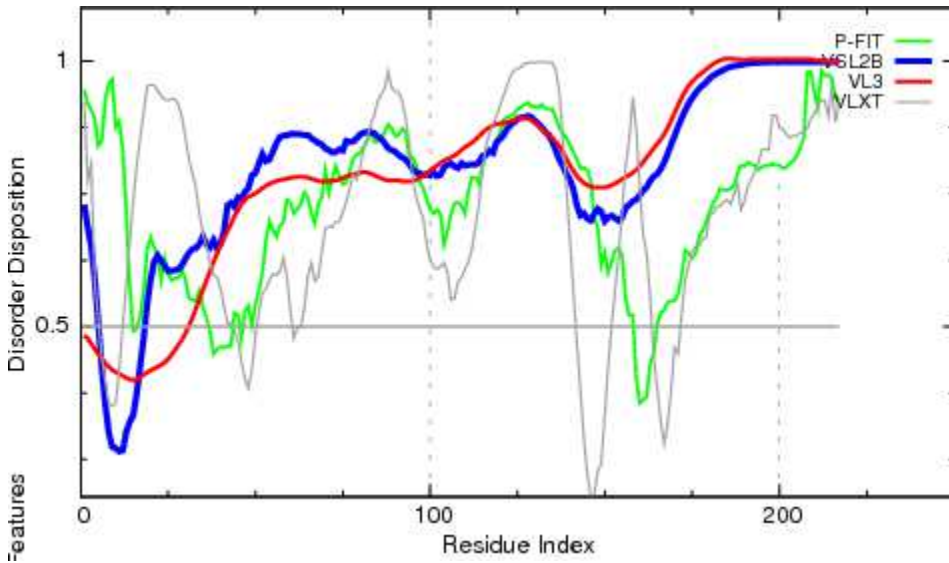
Table S1. List of Predicted Secreted Proteins (see separate Excel File)

Table S2. Mostly disordered proteins (>50% disordered residues by PONDR® VSL2 analysis; percent disordered indicated in parentheses after the protein name)

> MVLG_01284T0 (94.01%)- 217aa

MLMLKSLSVLIVAASAAHALQSPAAASNLERGLTDGTTGLLDNLPGLPGLGD
 LLGGGGKTTTGKVKRELDDFVSDTNSATPVKSPPGTDSALNDVTS DHNVLPQDA
 GLESLIPGLHKRQDEDEVDQDEDDVDSNSGLDAGAEVDADFDLVRRGYRKASE
 VKFKVNEPKPATFVKKSKKHPKTADKEHGKKHDGVHKKPDEEHKKHDEVPKK
 HHPVA

PONDR:



=====PONDR VSL2
 STATISTICS=====

Predicted residues: 217	Number Disordered Regions: 2
Number residues disordered: 204	Longest Disordered Region: 199
Overall percent disordered: 94.01	Average Prediction Score: 0.7951
Predicted disorder segment [1]-[5]	Average Strength= 0.6332
Predicted disorder segment [19]-[217]	Average Strength= 0.8290

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	41	56	16
2	68	78	11
3	91	118	28
4	136	178	43
5	209	217	9
Filtered Regions			
	From	To	Length
1	1	14	14
2	192	192	1

ModPred and PROSITE:

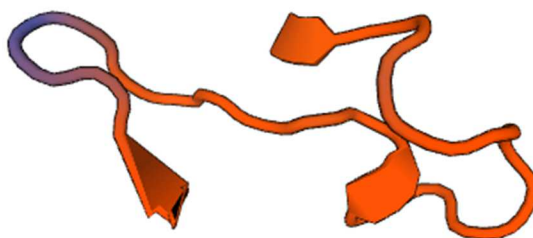
ModPred: Sumoylation (K5), Amidation (G37, F73, P86, D128, F148, E160, E208), ADP-ribosylation (R68), Phosphorylation (T81), Proteolytic cleavage (D133, R152, K162, T172, H192, D193).

PROSITE: No identified domain recognition sites.

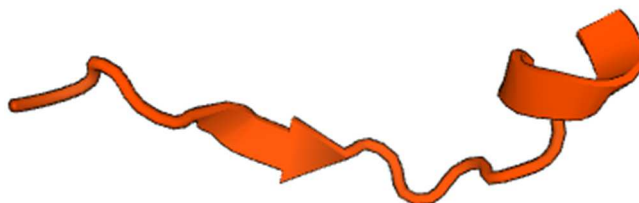
Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
5bs7.1.F	Protein SPT2 homolog	26.83	X-ray, 3.3Å	hetero-oligomer	None

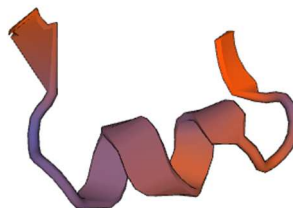
Name	Title	Identity	Method	Oligo State	Ligands
5g1k.1.B	THIOL DISULFIDE INTERCHANGE PROTEIN DSBG	30.30	X-ray, 2.0Å	homo-dimer	None
1ywf.1.A	PHOSPHOTYROSINE PROTEIN PHOSPHATASE PTPB	23.08	X-ray, 1.7Å	monomer	None



Model #1: Residues 43-68 of MVLG_01284T0 with 1ywf.1.A (23.08 % sequence identity) as a template



Model #2: Residues 138-153 of MVLG_01284T0 with 5bs7.1.F (26.83 % sequence identity) as a template

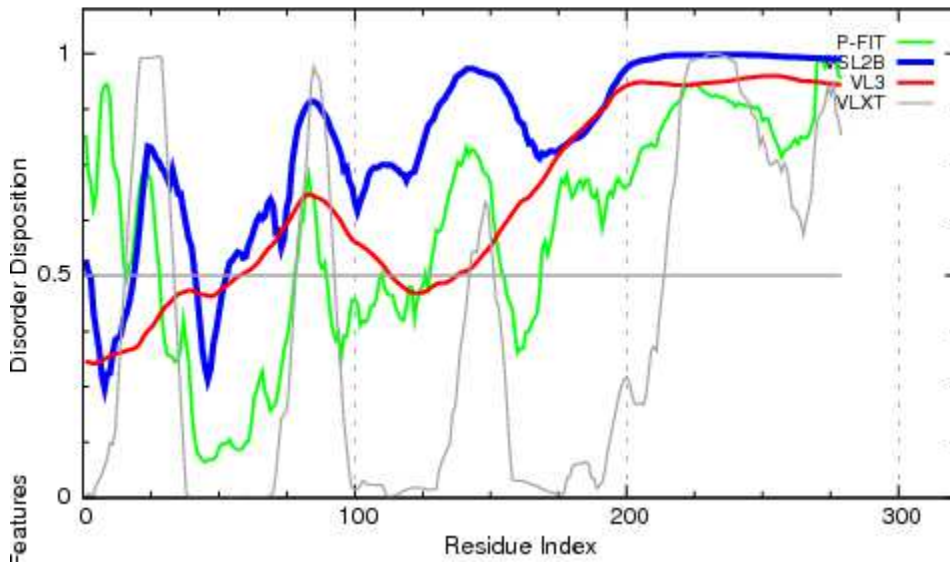


Model #3: Residues 21-35 of MVLG_01284T0 with 5g1k.1.B (30.30 % sequence identity) as a template

> MVLG_00385T0 (90.32%)

MHIHFFT VLSASALLALSHAAPTPGGPSSVHNVSPKYGPDSKCFHYFFDHLDEY
KPKYGCQDYEQYKCSYADAYVKKKAEQDKKEAECKKQDYDFKKHTWHYTS
VKNSFEEEEKKYNALLKQYNLETTRYEEHKKNYEAFKRAREEENRKNEETKKY
CEKVFEVKEYFKPKQSYHIDGHQNSHGGNLKGDGKGFAGKKDEDEGEKEGGH
GGVKNVGGKDGKGRAKDGKDGKLDDEKDEHDGKDGHDGKASKIHGKKDEKC
HDEKKDFSSGGNPKSGW

PONDR:



=====PONDR VSL2
STATISTICS=====

Predicted residues: 279	Number Disordered Regions: 3
Number residues disordered: 252	Longest Disordered Region: 227
Overall percent disordered: 90.32	Average Prediction Score: 0.7983
Predicted disorder segment [1]-[2]	Average Strength= 0.5224
Predicted disorder segment [19]-[41]	Average Strength= 0.6797
Predicted disorder segment [53]-[279]	Average Strength= 0.8639

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	119	125	7
2	158	180	23
3	250	257	8
4	267	279	13

ModPred and PROSITE:

ModPred:Proteolytic cleavage (Y75, K228, D229, D232, D242, D248, K267), Amidation (Q92), Carboxylation (E127, E133, E146, E147, E154, E160, E164, E165), Amidation (R145, Y176, L190), Acetylation (K195, K200, K220, K224), ADP-ribosylation (R226).

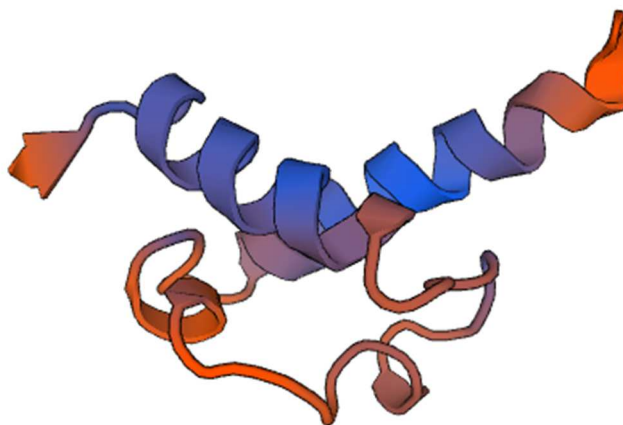
PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
3etz.1.A	Adhesin A	20.63	X-ray, 2.0Å	monomer	None
2dod.1.A	Transcription elongation regulator 1	11.67	NMR	monomer	None



Model #1: Residues 104-169 of MVLG_00385T0 with 3etz.1.A (20.63 % sequence identity) as a template

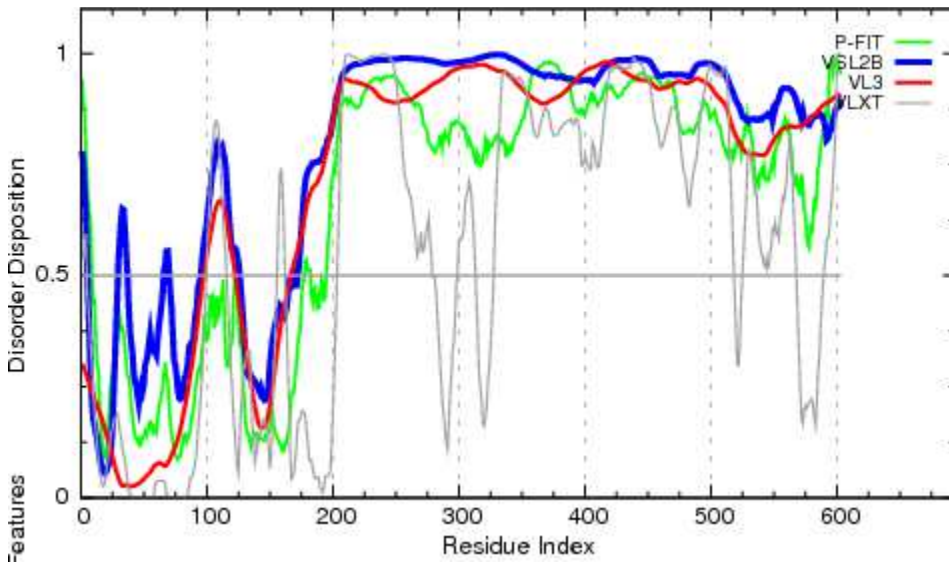


Model #2: Residues 107-168 of MVLG_00385T0 with 2dod.1.A (11.67 % sequence identity) as a template

>MVLG_06764T0 (80.27%)- 603 aa

MKLIRSARGAFLGAIGLSCLAVRLVRAQAGDEGNSYYADDTNSWIKFDGSWK
 EVVSDQFHQYTSMLTSDKGATATFKFIGESFKILGISDQANFSVTVDDGQDADVPS
 VSSDPDKTSALFSQTQLRPGTQHTVILEVISSGNLSFDAVVIGGGPASKAALCPVK
 GHQPCTSVNFQSIPKAHQSKPSTDRGGLFTTSLHAGIKLKLKRPDPDPSTPKDNS
 GDKSGNPLRHPKGAVGRKQKVDNTDKQDEKMGKDDVDQDGGKHQHQNKGK
 APKDQPTNDSADKEYDPGMKIGKTSEHHSKEDPAGADTTKSEKAGLEGHELKPS
 KQNKPDSPKTENPAPDDPKPKETSAPHNQTDGAGGDSGLLEGLLGGAVGSSGS
 NEHESHGGDHREGDSASQTPKSSLIPAIVKLVGSDGGHKPSKSKASSKPTDAG
 TLGNSSDPHKPSSSDHSTTLVGLSRPLGDGTGEEHHSNSKDSGIEQPEPKTLVDIK
 GGGDKHPDDPAADKPPPTPKKLVSIKGGPTDTGLQAGIDLKIVLPGSPATKLPD
 GQDVLGKKNVTGPSKDPGGKDHDQDHPVIVPVVIKFSDSHDDPGSVATVVTVSQE
 PVPKP

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 603	Number Disordered Regions: 6
Number residues disordered: 484	Longest Disordered Region: 432
Overall percent disordered: 80.27	Average Prediction Score: 0.7831
Predicted disorder segment [1]-[5]	Average Strength= 0.6747
Predicted disorder segment [30]-[37]	Average Strength= 0.5884
Predicted disorder segment [66]-[70]	Average Strength= 0.5322
Predicted disorder segment [96]-[126]	Average Strength= 0.6613
Predicted disorder segment [168]-[170]	Average Strength= 0.5116
Predicted disorder segment [172]-[603]	Average Strength= 0.9337

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	141	149	9
2	156	165	10

3	168	176	9
4	189	207	19
5	223	242	20
6	249	271	23
7	278	326	49
8	350	420	71
9	424	440	17
10	446	464	19
11	471	494	24
12	507	555	49
13	568	597	30
Filtered Regions			
	From	To	Length
1	134	138	5

ModPred and PROSITE:

ModPred: Sumoylation (K2), Amidation (I79, K410, T591), Phosphorylation (T214, Y285, S331, T398, T507, S537), Proteolytic cleavage (K231, D253, D393, K410, D547), ADP-ribosylation (R458), Ubiquitination (K494), Hydroxylation (P505, P506).

PROSITE: PROKAR_LIPOPROTEIN (Prokaryotic membrane lipoprotein lipid attachment site profile, 1-19, PROSITE entry PS51257). Signal (1-18), N-palmitoyl cysteine (19), S-diacylglycerol cysteine (19)

In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The

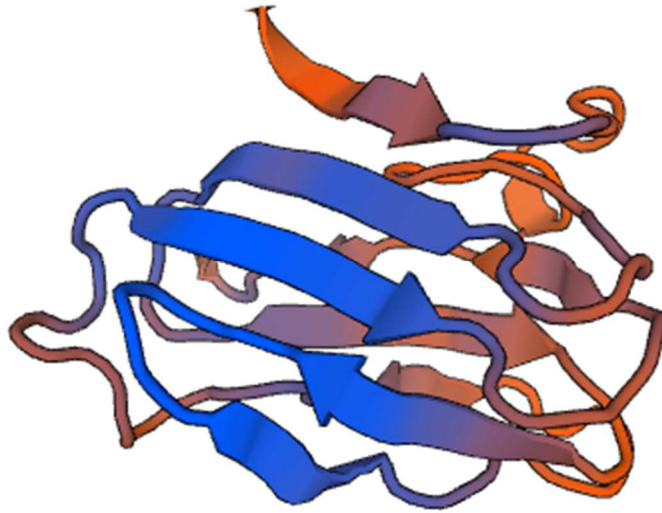
peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached.

PROSITE: TONB_DEPENDENT_REC_1 (TonB-dependent receptor proteins signature 1, 1-101, PROSITE entry PS00430)

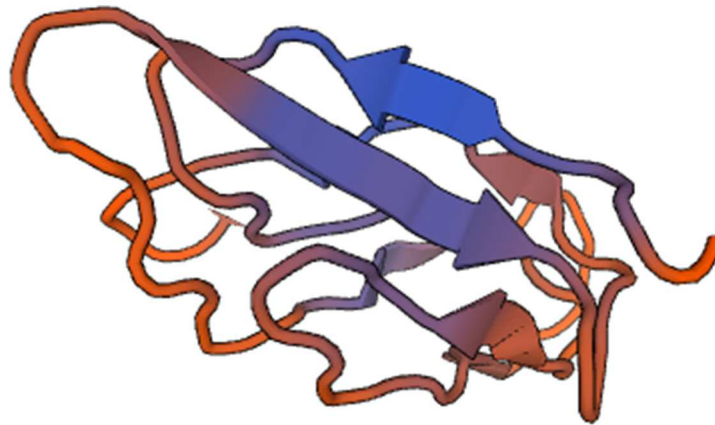
In *Escherichia coli*, the tonB protein interacts with outer membrane receptor proteins that carry out high-affinity binding and energy-dependent uptake of specific substrates into the periplasmic space. These substrates are either poorly permeable through the porin channels or are encountered at very low concentrations. In the absence of tonB these receptors bind their substrates but do not carry out active transport. The tonB protein also interacts with some colicins.

Structural modelling:

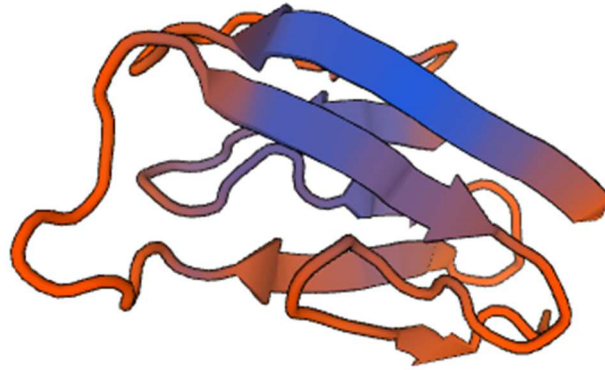
Name	Title	Identity	Method	Oligo State	Ligands
2wab.1.A	ENDOGLUCANASE E	24.04	X-ray, 1.9Å	monomer	None
5x7o.1.A	Glycoside hydrolase family 31 alpha-glucosidase	14.63	X-ray, 2.0Å	homo- dimer	2 x <u>NI</u> , 6 x <u>CA</u> , 9 x <u>MG</u> , 4 x <u>MES</u>
1fv3.1.A	TETANUS TOXIN HEAVY CHAIN	16.25	X-ray, 2.3Å	monomer	None
2vyu.1.A	CHOLINE BINDING PROTEIN F	24.59	X-ray, 2.5Å	monomer	None
3ron.1.A	Type-1Aa cytolytic delta- endotoxin	17.95	X-ray, 2.2Å	monomer	None



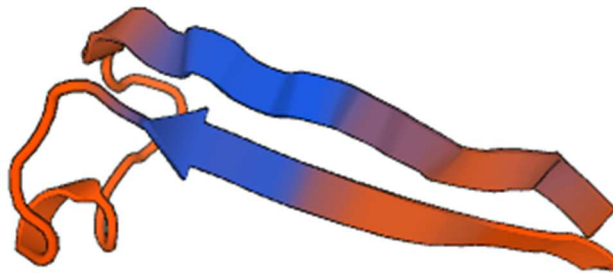
Model #1: Residues 42-152 of MVLG_06764T0 with 2wab.1.A (24.04 % sequence identity) as a template



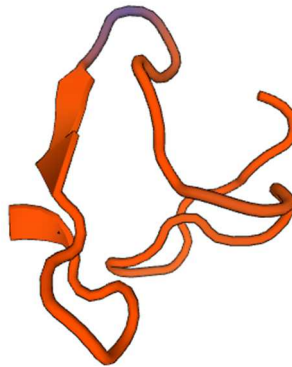
Model #2: Residues 69-154 of MVLG_06764T0 with 5x7o.1.A (19.05 % sequence identity) as a template



Model #3: Residues 71-151 of MVLG_06764T0 with 1fv3.1.A (16.25 % sequence identity) as a template



Model #4: Residues 65-103 of MVLG_06764T0 with 3ron.1.A (17.95 % sequence identity) as a template

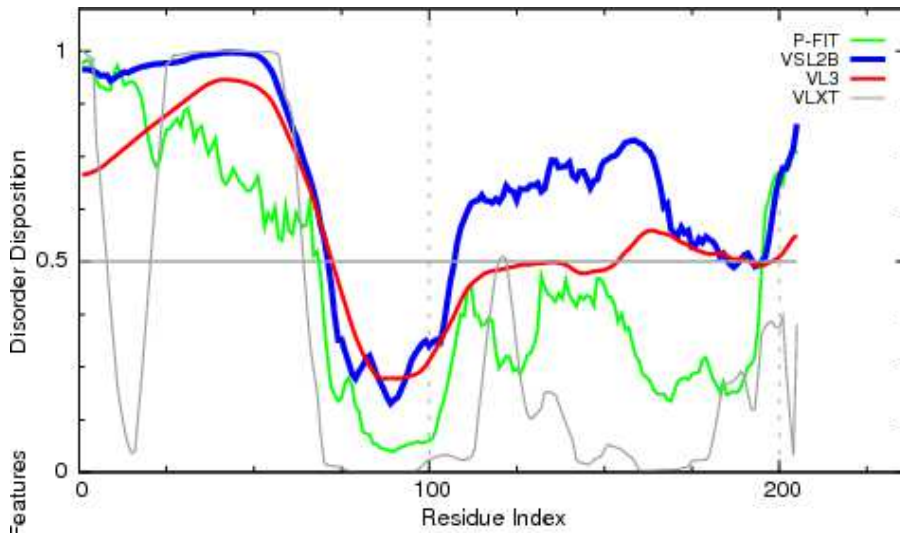


Model #5: Residues 29-61 of MVLG_06764T0 with 2vyu.1.A (24.59 % sequence identity) as a template

>MVLG_00398T0 (80.00%)- 205 aa

MPSMRSLTSAFALSALVAVSSAPT VSSMPSLIERHDSPLSPPSLPPPPSPSVKSSLP
TFSPPPPTYENQTM CINYYFEHLPEYKDL YDCTVYDATKSFYDARYYQKKAIEKL
EEEAQCAKDQADFADRVNQFAQAELAYAAEQKRFEWASKKFEHEKASLETSKK
AFQALINQKVLEASQAMEIKHTCEKVFSEHTNEYIPKHSI

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 205	Number Disordered Regions: 4
Number residues disordered: 164	Longest Disordered Region: 78
Overall percent disordered: 80.00	Average Prediction Score: 0.6783
Predicted disorder segment [1]-[71]	Average Strength= 0.9207
Predicted disorder segment [108]-[185]	Average Strength= 0.6637
Predicted disorder segment [189]-[192]	Average Strength= 0.5114
Predicted disorder segment [195]-[205]	Average Strength= 0.6655

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	7	18	12
Filtered Regions			
	From	To	Length
1	71	79	9
2	86	89	4

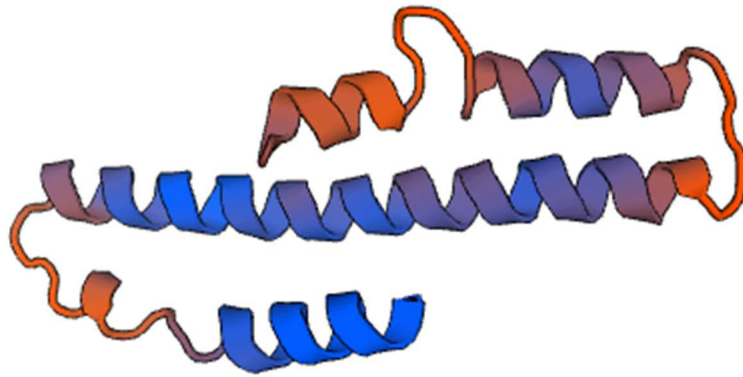
ModPred and PROSITE:

ModPred: Phosphorylation (S37, S40, S49), Hydroxylation (P47), O-linked glycosylation (T58), Amidation (Y199).

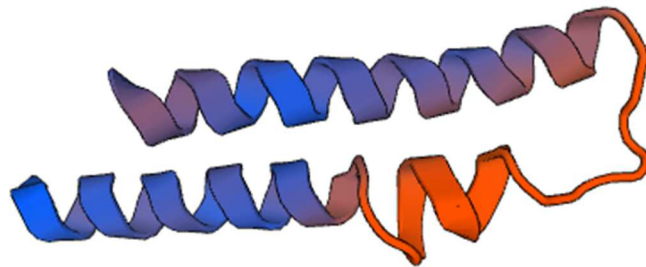
PROSITE: No identified domain recognition sites.

Structural modelling:

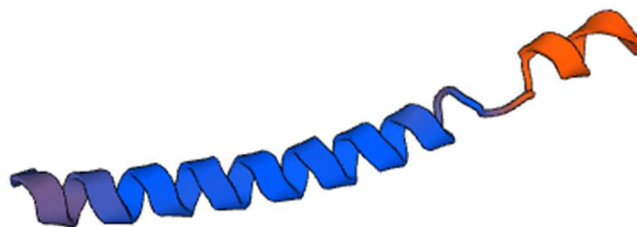
Name	Title	Identity	Method	Oligo State	Ligands
2ap3.1.A	conserved hypothetical protein	20.00	X-ray, 1.6Å	monomer	None
4bne.1.A	PROTEIN KINASE C AND CASEIN KINASE SUBSTRATE IN NEURONS PROTEIN 2	18.33	X-ray, 2.6Å	homo-dimer	None
5cx2.1.B	Coronin	26.83	X-ray, 2.2Å	hetero-oligomer	None



Model #1: Residues 106-199 of MVLG_00398T0 with 2ap3.1.A (19.78 % sequence identity) as a template



Model #2: Residues 125-187 of MVLG_00398T0 with 4bne.1.A (20.00 % sequence identity) as a template

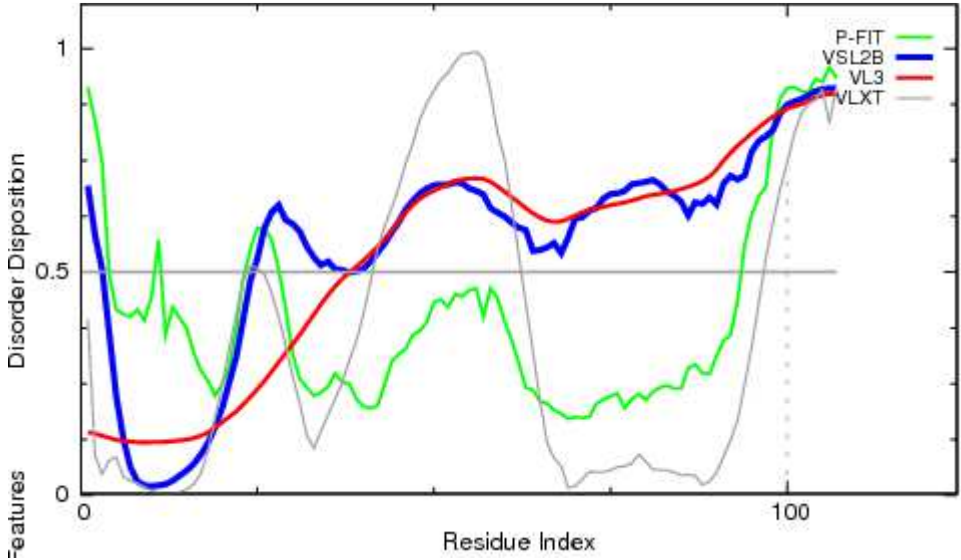


Model #3: Residues 142-182 of MVLG_00398T0 with 5cx2.1.B (26.83 % sequence identity) as a template

> MVLG_04106T0 (79.44%)- 107aa

MKYSLVFVALVVIATRIVSALAADATKQASTSEVDYPYPFPEEHAATVSQGPTR
 PITHPVASTLNLNCKAEKCTTCKGEARGTCIEQCASWMAHQASQPEPEGC

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 107	Number Disordered Regions: 3
Number residues disordered: 85	Longest Disordered Region: 69
Overall percent disordered: 79.44	Average Prediction Score: 0.5594
Predicted disorder segment [1]-[3]	Average Strength= 0.5938
Predicted disorder segment [25]-[37]	Average Strength= 0.5667
Predicted disorder segment [39]-[107]	Average Strength= 0.6827

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			

	From	To	Length
1	3	14	12

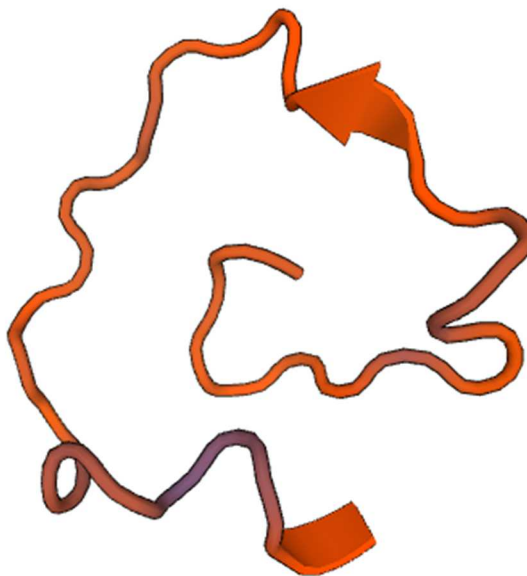
ModPred and PROSITE:

ModPred: Post translational modification sites include proteolytic cleavage (D35), sulfation (Y36), Amidation (S48).

No identified domain recognition sites (PROSITE)

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
1csm.1.A	CHORISMATE MUTASE	22.22	X-ray, 2.2Å	homo-dimer	2 x TRP

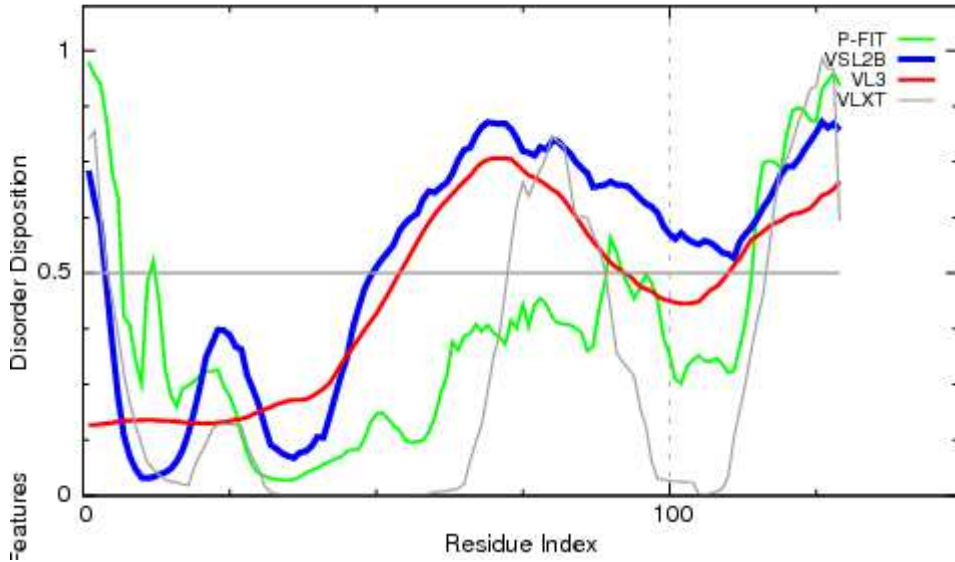


Model: Residues 30-65 of MVLG_04106T0 with 1csm.1.A (22.22% sequence identity) as a template

> MVLG_05720T0 (64.34%)- 129aa

MMRSLIKLLVLFTAVSVALANPWPPSVQDSCNWLKAWCTDCQTSFCGNITSHK
 QHKLCFKTHCESHHPRDYPRPCKQMQMADKCMRSCQWKRSHNLTLSWNPFIN
 HDKCRHCCDMQGGPTEKRMRRSGY

PONDR:



=====PONDR VSL2
 STATISTICS=====

Predicted residues: 129	Number Disordered Regions: 2
Number residues disordered: 83	Longest Disordered Region: 80
Overall percent disordered: 64.34	Average Prediction Score: 0.5194
Predicted disorder segment [1]-[3]	Average Strength= 0.6671
Predicted disorder segment [50]-[129]	Average Strength= 0.6944

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			
	From	To	Length

1	7	12	6
---	---	----	---

ModPred and PROSITE:

ModPred: SUMOylation (k7), Amidation (A14), Hydroxylation (P119), Proteolytic cleavage (R123, R126, S127)

PROSITE: No identified domain recognition sites.

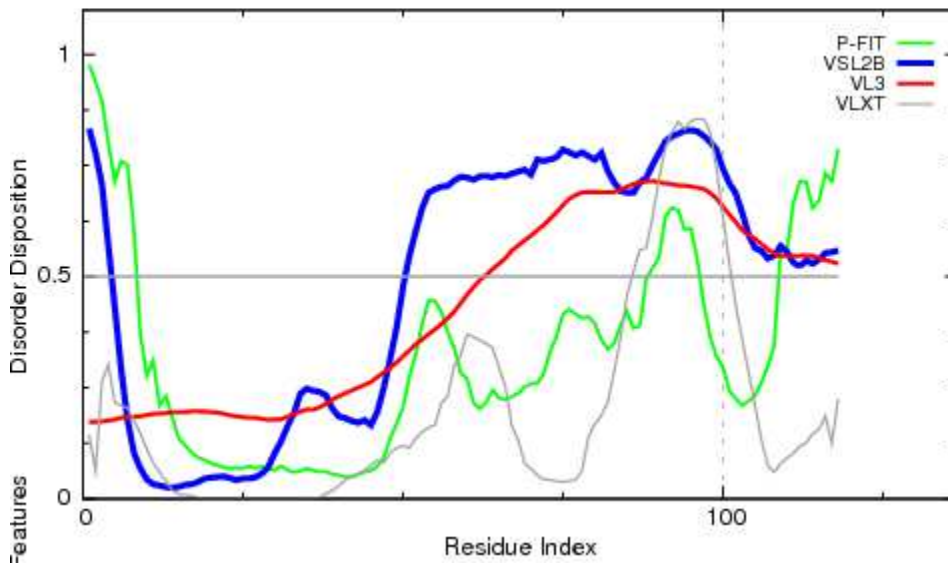
Structural modelling:

No templates were found matching target sequence

> MVLG_06175T0 (61.02%)- 118aa

MWTSSIVQAALLFAVIVLYSSPVVAWAFCPFGKTAEHMAICSSLCRMRCYDPN
 SGTSNSTCRNACTGQYHVSRLNAADQCMQQCDRFTKDKKKQGEGKLEHKRCL
 HKCTDWWFFPLNL

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 118	Number Disordered Regions: 2
Number residues disordered: 72	Longest Disordered Region: 68
Overall percent disordered: 61.02	Average Prediction Score: 0.4819
Predicted disorder segment [1]-[4]	Average Strength= 0.7216
Predicted disorder segment [51]-[118]	Average Strength= 0.6969

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			
	From	To	Length
1	113	118	6

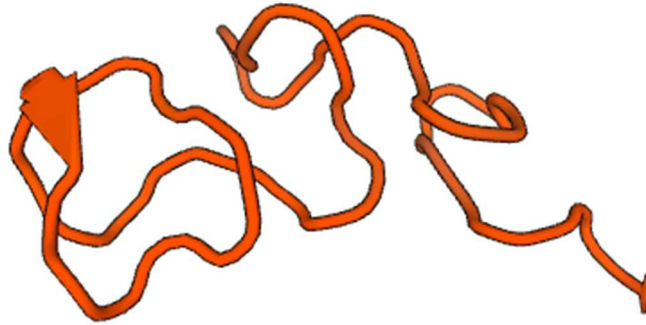
ModPred and PROSITE:

ModPred: Amidation (A10, S59, Y69), GPI anchor amidation (N53).

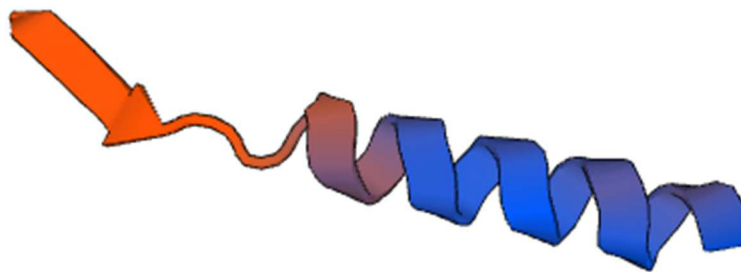
No identified domain recognition sites (PROSITE)

Structural modelling:

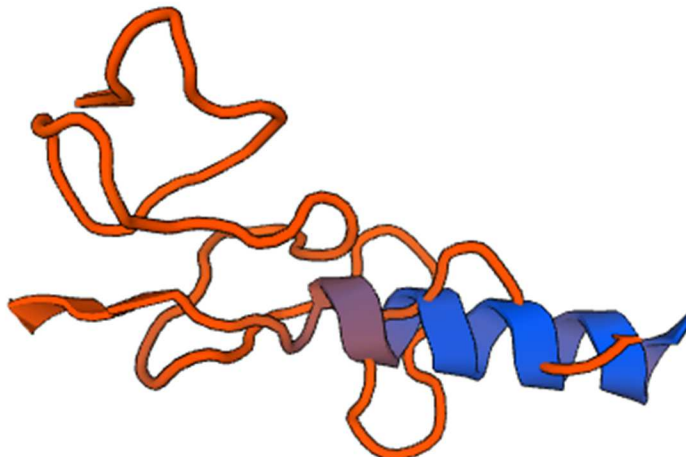
Name	Title	Identity	Method	Oligo State	Ligands
2v79.1.B	DNA REPLICATION PROTEIN DNAD	20.59	X-ray, 2.0Å	homo-tetramer	6 x <u>NA</u> , 22 x <u>CL</u>
2ahx.1.A	Receptor tyrosine-protein kinase erbB-4	27.08	X-ray, 2.4Å	monomer	5 x <u>NAG</u> , 3 x <u>NDG</u>



Model #1: Residues 40-87 of MVLG_06175T0 with 2ahx.1.A (27.08% sequence identity) as a template



Model #2: Residues 66-88 of MVLG_06175T0 with 2v79.1.B (20.59% sequence identity) as a template

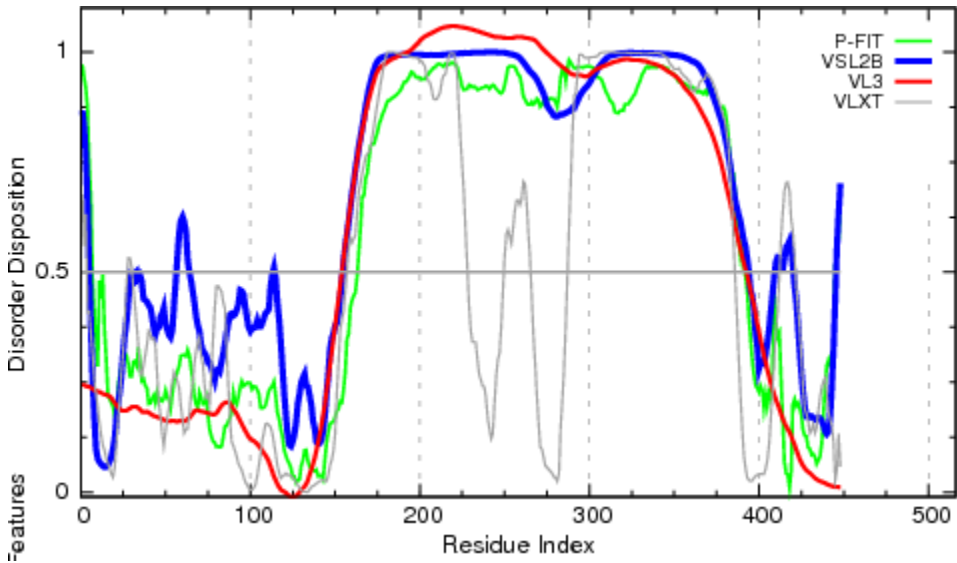


Overlap of models 1 and 2 for MVLG_06175T0

> MVLG_01591T0 (59.82%)- 448 aa

MRSLQLLSVLVTTCLPAAFAEIIDNTDAKKVITESGADISGNVTIQNAEGDYM
HFQRDGTPGNSHVLSLFPQYSTVEVVSRTIYGASGRMHGRFTGVSLSGANKCAA
TQFNSNEGRDYDVVAYGCTFNRNHTGAKMIFNVLPCGNTEDALSLAQKIRGVSK
KEDFKFKKANPKSSPSRKSSGKSGAHRNTPHRPQSDLSSSGQPGRHHVGGYR
GRK R HSGHGRRRRGGHGHHGGHGHEGGNHHGGGGHGHKGGNHHGGGGHGHKGGNHH
GGGHPQHVVRSLSLCTGNSLACQRRRHYLAKRDSRSQMLVSPQGPSPQGPSPQGP
VSPSGTPKQSASGASGGAGSAAGDHGPGPQSTAKKTQDGAVSQQASKDPNPASE
ADKSNSEIADHLRKNLMSGKAQTVCIQGQDHLSDMQTAGLTGKETVGGAGGVP
GLMYDLFDASNDAFWLTMTRVN

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 448

Number residues disordered: 268

Overall percent disordered: 59.82

0.6601

Predicted disorder segment [1]-[5]

Predicted disorder segment [33]-[33]

Predicted disorder segment [57]-[63]

Predicted disorder segment [114]-[114]

Predicted disorder segment [155]-[394]

Predicted disorder segment [410]-[420]

Predicted disorder segment [446]-[448]

Number Disordered Regions: 7

Longest Disordered Region: 240

Average Prediction Score:

Average Strength= 0.7239

Average Strength= 0.5020

Average Strength= 0.5857

Average Strength= 0.5102

Average Strength= 0.9325

Average Strength= 0.5355

Average Strength= 0.6314

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	136	144	9
2	151	174	24
3	181	311	131
4	318	349	32
5	352	367	16
6	378	407	30
7	426	441	16
Filtered Regions			
	From	To	Length
1	122	126	5

ModPred and PROSITE:

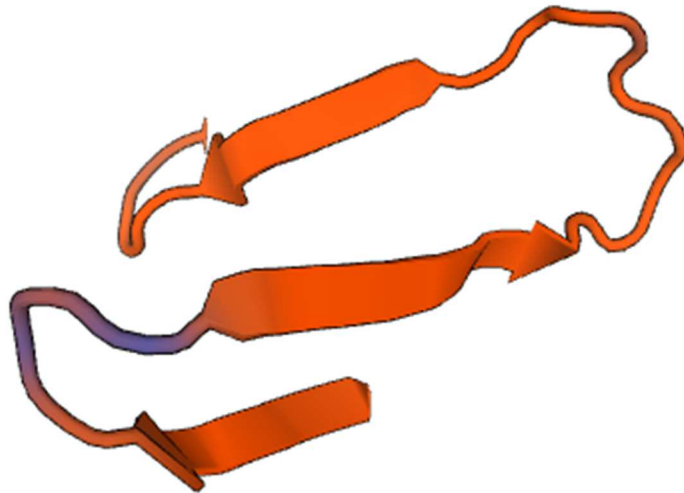
ModPred: Amidation (M53, G59, N104, R290, K354, M429, T445), Proteolytic cleavage (Y119, D120, K180, K184, R189, Q302, M303), Acetylation (K162, K216), Sumoylation (K167), Phosphorylation (S176, S322), Methylation (R225).

PROSITE: No identified domain recognition sites.

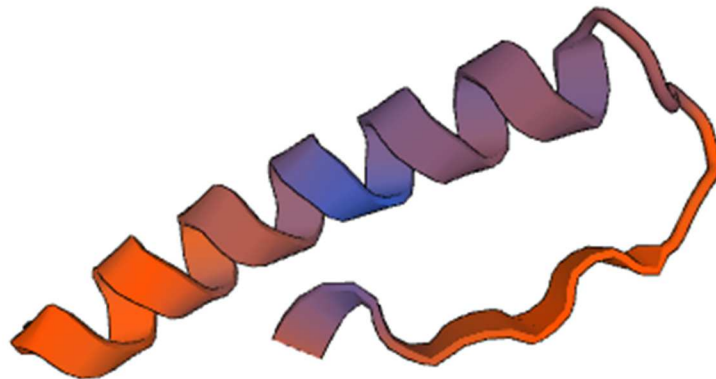
Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
4y3k.1.A	Serpin A12	17.14	X-ray, 2.2Å	monomer	None

Name	Title	Identity	Method	Oligo State	Ligands
4k07.1.A	Amyloidogenic immunoglobulin light chain protein AL-103	5.71	X-ray, 2.8Å	homo-dimer	None



Model #1: Residues 29-63 of MVLG_01591T0 with 4k07.1.A (5.71% sequence identity) as a template

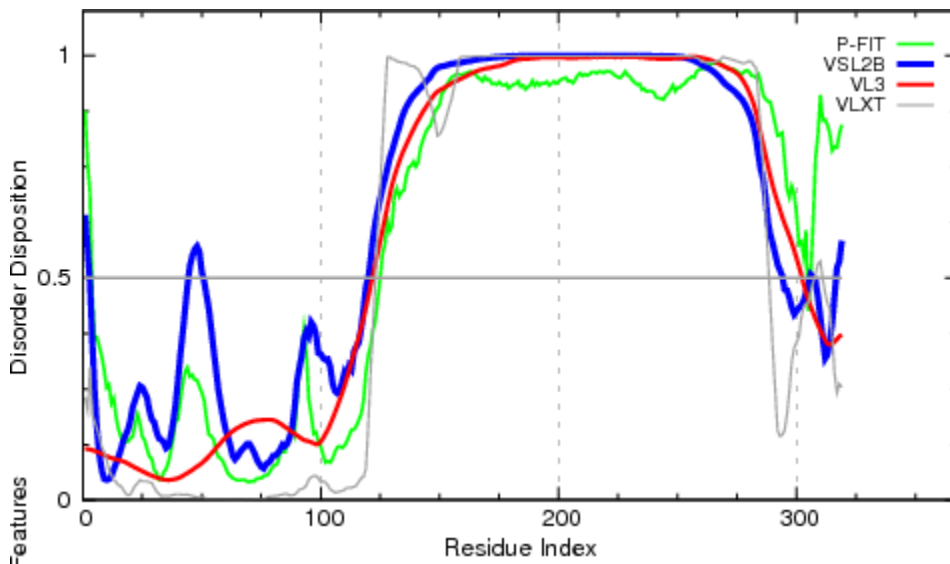


Model #2: Residues 274-308 of MVLG_01591T0 with 4y3k.1.A (17.14% sequence identity) as a template

> MVLG_04168T0 (58.93%)- 319 aa

MHLVTVLPFALVAFLGSTGVQALIKTANATLEALYLVEYDTKAYGTPETGAVK
HFGTNFHQACVAAAIQKIPIMHCRVADIPFNPLIIPQDGTWDNHGPVTDLMDLAM
NNTLFLGNARIVSGPDGKPYAPAYPPPTKPTPETRDADAASHVPDSTGLVPGTTL
PGPSTIPGPGTTPPGPGTTLPGPSTIPGPGTTPPGPGTTLPGPSTIPGPGTTPGPGTTL
PGPSTIPGPGTTPPGPGPTLPGPGPTLVPVSRPTTPLTGGHGRKGRKHRNG
RKGGFKRQVTVDDTVDDFLMNGQVQPPDSVDTLGGIAF

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 319	Number Disordered Regions: 5
Number residues disordered: 188	Longest Disordered Region: 175
Overall percent disordered: 58.93	Average Prediction Score: 0.6365
Predicted disorder segment [1]-[2]	Average Strength= 0.6015
Predicted disorder segment [45]-[50]	Average Strength= 0.5467
Predicted disorder segment [120]-[294]	Average Strength= 0.9317
Predicted disorder segment [306]-[307]	Average Strength= 0.5051
Predicted disorder segment [317]-[319]	Average Strength= 0.5479

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	105	121	17
2	144	171	28
3	180	193	14
4	198	238	41
5	246	273	28
6	277	306	30
7	309	319	11

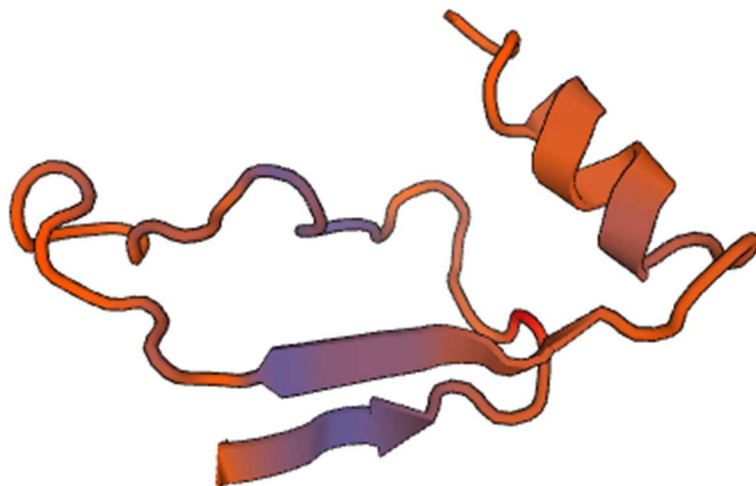
ModPred and PROSITE:

ModPred: Proteolytic cleavage (R118, S121, R143, G156, R277, R280, K281, K285, R286, D311), Hydroxylation (P133, P134, P135, P237, P244), Amidation (P134, P204, P257), Phosphorylation (T139, T175, T195, T240, T263), O-linked glycosylation (S167, S187, S207, T215), ADP-ribosylation (R260).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
1z5z.2.A	Helicase of the snf2/rad54 family	11.11	X-ray, 2.0Å	monomer	None

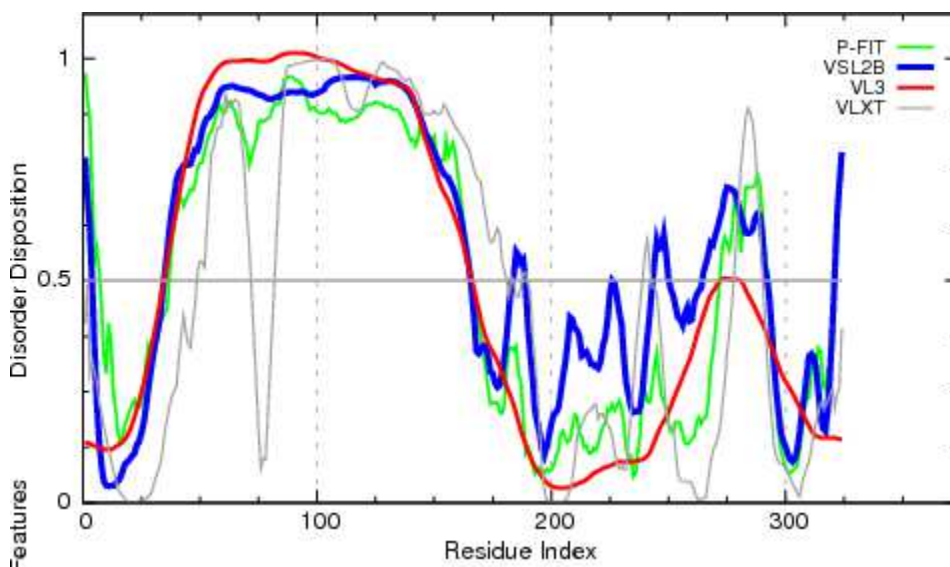


Model: Residues 253-307 of MVLG_04168T0 with 1z5z.2.A (11.11% sequence identity) as a template

> MVLG_02288T0 (55.25%)- 324 aa

MRFLTSQITLCLLVLTSTLAFHVANLFPDLSVDAHRTVPSSNRRHHRCLAHSV
 RHNHHKKCRHSRKTGLKHFDQEKS**AHTHGH**LAHRNRRKKPAIKRLGKKRPVSH
 VDPNHKRPEHDQSNPPTILTGPTLQQPEPSGVHVS**KTPAGKTAPERQIPDGKDQV**
 SEIQALALEEINAFRALHNAPPLQTSPELVQNAV**VVWTSKCHYGHTRGAFTGEYGE**
 IIARTSGSWGNNMSKAIELWTVDEENDFNPRKPQ**TTHTQAVWKSSRLLGCASS**
 DKCNDPADNSTTVTGDDIPPDEHNSVLYICRFLPAGNLNDKD**DVDIIMLKG**FAD

PONDR:



STATISTICS=====

Predicted residues: 324	Number Disordered Regions: 7
Number residues disordered: 179	Longest Disordered Region: 131
Overall percent disordered: 55.25	Average Prediction Score:
0.5690	
Predicted disorder segment [1]-[3]	Average Strength= 0.6941
Predicted disorder segment [35]-[165]	Average Strength= 0.8695
Predicted disorder segment [184]-[188]	Average Strength= 0.5446
Predicted disorder segment [226]-[226]	Average Strength= 0.5008
Predicted disorder segment [244]-[250]	Average Strength= 0.5636
Predicted disorder segment [265]-[292]	Average Strength= 0.6244
Predicted disorder segment [321]-[324]	Average Strength= 0.6675

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	46	112	67
2	122	147	26
3	161	180	20
4	192	205	14
5	232	237	6
6	259	268	10
7	297	305	9
Filtered Regions			
	From	To	Length
1	10	27	18
2	218	218	1
3	318	319	2

ModPred and PROSITE:

ModPred: Amidation (A21), Proteolytic cleavage (D34, H36, R37, H52, R88, R98),
Acetylation (K71, K97).

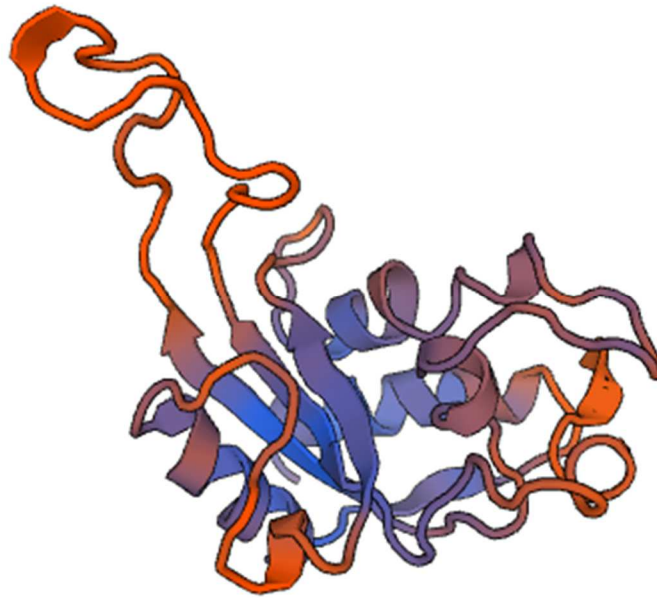
PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
5jys.1.A	Protein PRY1	25.81	X-ray, 1.9Å	monomer	1 x <u>MG</u>
2giz.1.A	Natrin-1	26.62	X-ray, 1.7Å	monomer	None
4kt0.1.I	Photosystem I reaction center subunit XII	30.77	X-ray, 2.8Å	hetero-oligomer	4 x <u>LHG</u> , 3 x <u>SF4</u> , 90 x <u>CLA</u> , 2 x <u>PQN</u> , 2 x <u>LMU</u> , 2 x <u>CL0</u> , 17 x <u>BCR</u> , 1 x <u>LMG</u>



Model #1: Residues 163-321 of MVLG_02288T0 with 2giz.1.A (26.62% sequence identity) as a template



Model #2: Residues 162-320 of MVLG_02288T0 with 5jys.1.A (25.81% sequence identity) as a template

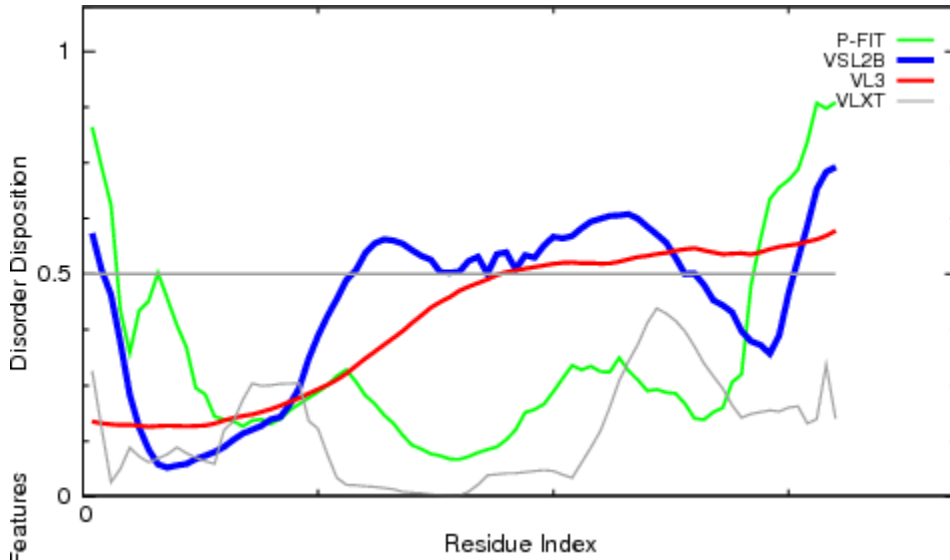


Model #3: Residues 3-26 of MVLG_02288T0 with 4kt0.1.I (30.77% sequence identity) as a template

> MVLG_00566T0 (55.00%)- 80 aa

MRTCSIVFALGTLTLLSLTQVVVAAPKAAADSTDFTKGMSCNSCVKTCNQKHL
ATGSADMEAGTSLVDCMDSCISVYNCES

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 80	Number Disordered Regions: 3
Number residues disordered: 44	Longest Disordered Region: 37
Overall percent disordered: 55.00	Average Prediction Score: 0.4306
Predicted disorder segment [1]-[2]	Average Strength= 0.5504
Predicted disorder segment [29]-[65]	Average Strength= 0.5601
Predicted disorder segment [76]-[80]	Average Strength= 0.6607

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			

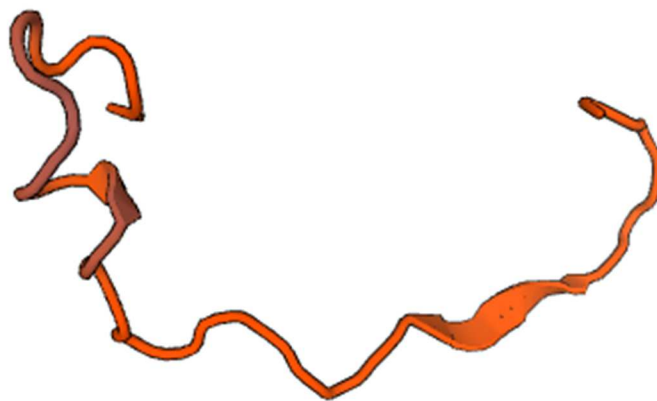
ModPred and PROSITE:

ModPred: Proteolytic cleavage (D33), Carboxylation (E79).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
5d6s.1.A	Epoxyqueuosine reductase	15.15	X-ray, 2.6Å	monomer	2 x <u>SF4</u> , 1 x <u>B12</u>



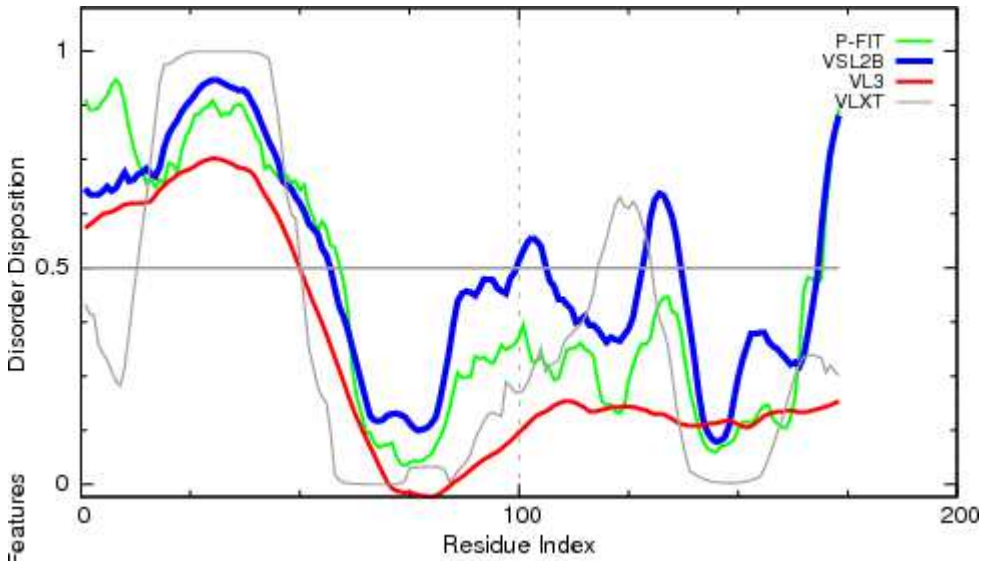
Model: Residues 21-53 of MVLG_00566T0 with 5d6s.1.A (15.15% sequence identity) as a template

Highly disordered proteins (30-50% disordered residues by PONDR® VSL2 analysis)

> MVLG_05122T0 (43.93%)- 173 aa

MLFKVSAALVLAGLSLGASALPSMSTESRAQPSPSSNKSPYGRGTGYIDSPADRK
TTTYKVGDKIHFVYTSAPATYFVDVSLMLANGSQSFQLANRLTGSSMISNDANA
RAYFRMPENLKTIAATELLAASQDEHSGAMKNNNCILAYLIAKETQNGQYGLVGN
LETKQAIAISM

PONDR:



=====PONDR VSL2
 STATISTICS=====

Predicted residues: 173	Number Disordered Regions: 4
Number residues disordered: 76	Longest Disordered Region: 56
Overall percent disordered: 43.93	Average Prediction Score: 0.4965
Predicted disorder segment [1]-[56]	Average Strength= 0.7669
Predicted disorder segment [100]-[106]	Average Strength= 0.5439
Predicted disorder segment [129]-[136]	Average Strength= 0.6229
Predicted disorder segment [169]-[173]	Average Strength= 0.7406

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	142	147	6

Filtered Regions			
	From	To	Length
1	1	16	16
2	75	81	7

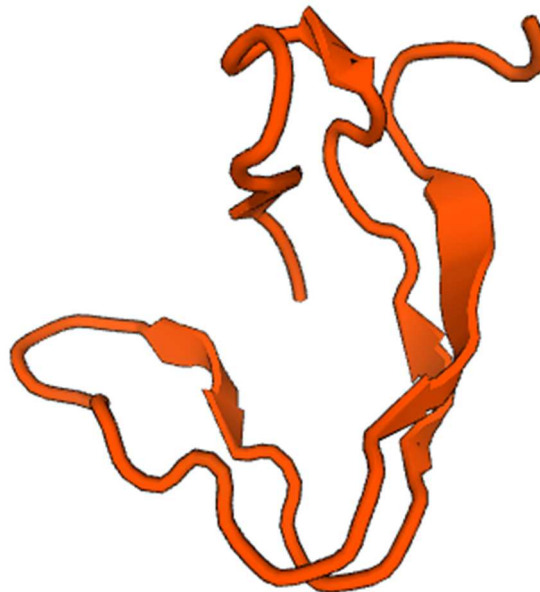
ModPred and PROSITE:

ModPred: Proteolytic cleavage (E27), Phosphorylation (S49), GPI anchor amidation (N104), Amidation (Y157).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
2o2o.1.A 1	SH3-domain kinase-binding protein	15.91	NMR	monomer	None

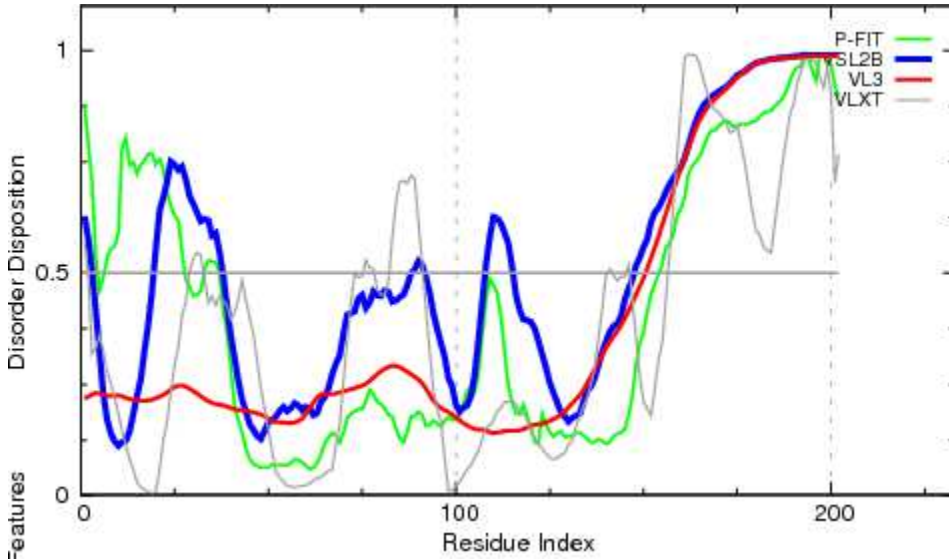


Model: Residues 35-78 of MVLG_05122T0 with 2o2o.1A (15.91 % sequence identity) as a template

> MVLG_03398T0 (43.07%)- 202 aa

MLTLPLLVASLGFCHSVVGSRHKNQDAHQVQGSTGKEPLIDADKMYLIRVNPQ
 ITGLQSACTFLSDRCEKYVKRGPNVKQLDVCSSAGQAVTSTSPYLWASCFETGT
 NEKDGRARDVSFNAFAGSDHAIVFLRGDEVFREVEIDEDLLKTESKKWKPTVNH
 QPQNRSPRQAGHRQINHETTGPKTHSGHNGDRPKRHKQET

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 202	Number Disordered Regions: 5
Number residues disordered: 87	Longest Disordered Region: 55
Overall percent disordered: 43.07	Average Prediction Score: 0.5044
Predicted disorder segment [1]-[3]	Average Strength= 0.5705
Predicted disorder segment [20]-[37]	Average Strength= 0.6548
Predicted disorder segment [89]-[91]	Average Strength= 0.5176
Predicted disorder segment [108]-[115]	Average Strength= 0.5787
Predicted disorder segment [148]-[202]	Average Strength= 0.8722

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	98	103	6

2	120	146	27
3	172	179	8
Filtered Regions			
	From	To	Length
1	6	13	8
2	148	149	2

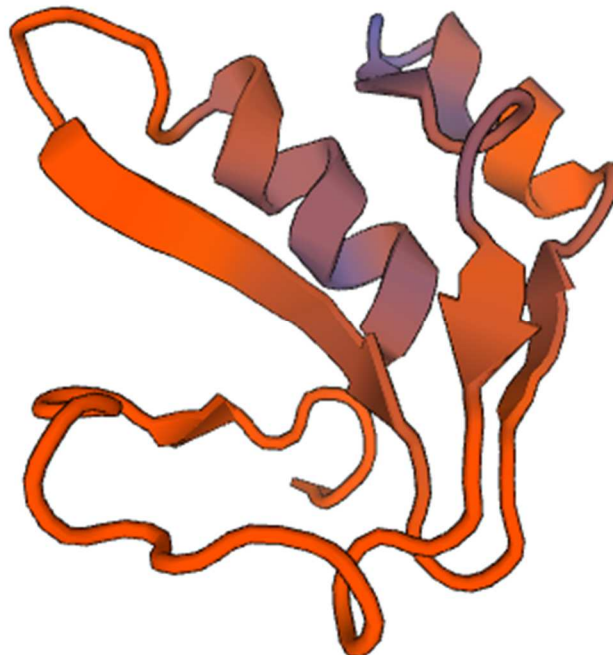
ModPred and PROSITE:

ModPred: Proteolytic cleavage (R116, D117, N121).

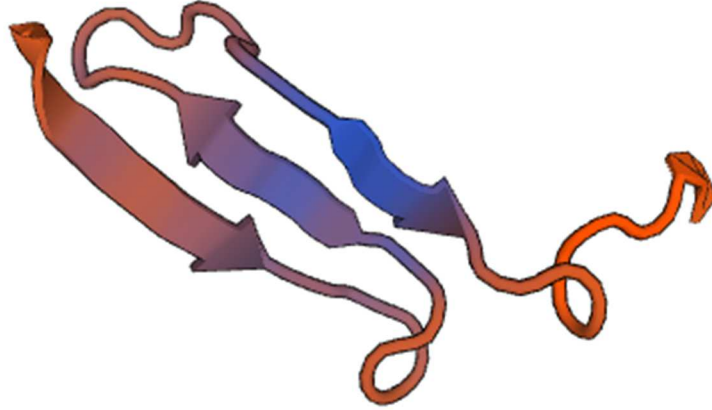
PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
2myn.1.A	Glutaredoxin arsenate reductase	17.91	NMR	monomer	None
1ybx.1.A	Conserved hypothetical protein	19.44	X-ray, 1.8Å	homo-dimer	None



Model #1: Residues 35-78 of MVLG_03398T0 with 2myn.1.A (17.91 % sequence identity) as a template

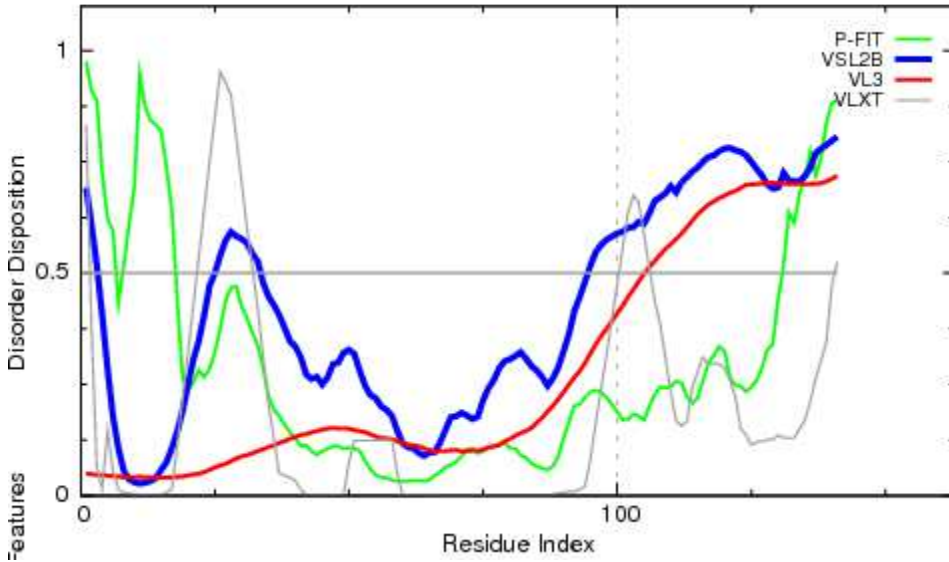


Model #2: Residues 117-152 of MVLG_03398T0 with 1ybx.1.A (19.44 % sequence identity) as a template

> MVLG_05716T0 (41.84%)- 141 aa

MMYSSLFIFAFTVVGAIVNAKMAKVATNSQTTSLGVPVAGVEKFHQPYWKNGT
 AAPAACVAVSQACFECLSKCYQHNNQWGFNGKTDCCYYGQCNNTRETRYKESC
 AIENNAKTCS DGLPKAQQTGGPMLNCKQANGTSLY

PONDR:



=====PONDR VSL2
 STATISTICS=====

Predicted residues: 141	Number Disordered Regions: 3
Number residues disordered: 59	Longest Disordered Region: 47
Overall percent disordered: 41.84	Average Prediction Score: 0.4194
Predicted disorder segment [1]-[3]	Average Strength= 0.6057
Predicted disorder segment [25]-[33]	Average Strength= 0.5559
Predicted disorder segment [95]-[141]	Average Strength= 0.6970

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			

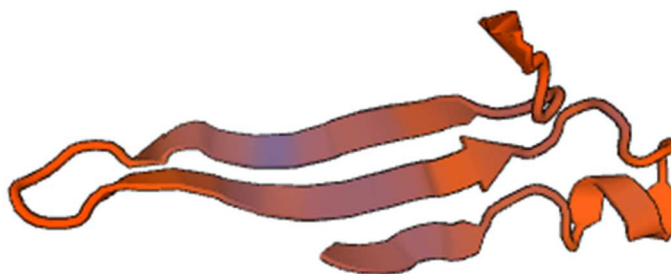
ModPred and PROSITE:

ModPred: Disulphide linkage (C87, C92), Carboxylation (E102).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
1mm9.1.A	Streptavidin	15.56	X-ray, 1.7Å	homo-tetramer	8 x MRD

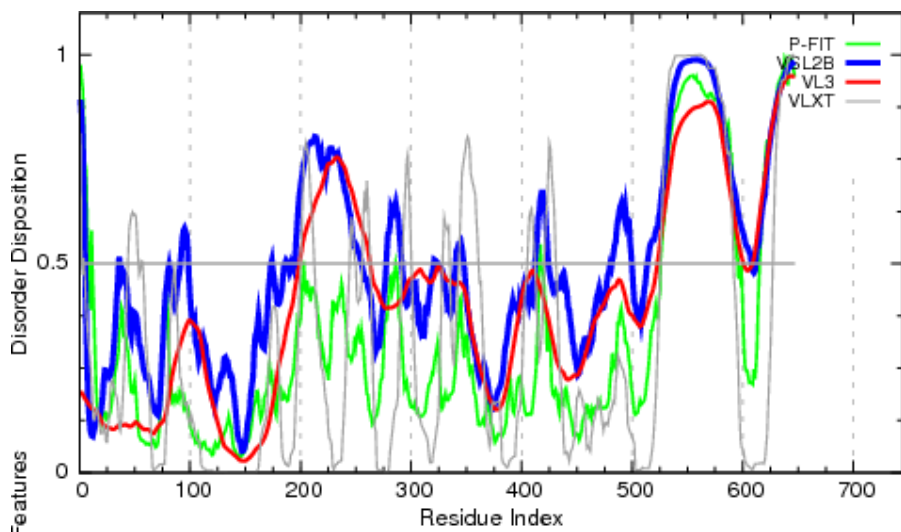


Model: Residues 79-125 of MVLG_05716T0 with 1mm9.1.A (15.56 % sequence identity) as a template

> MVLG_04806T0 (40.65%)—647 aa

MPDLSWFHSALLLSCLGTLALSQPINLTTYHLYDSSHAAQGLVHPDALRGELV
QTPSALLNALGLRDAELFWASPKGTRGTSQVAIMQMETHEKQKVLNMQRFSQ
LLAAVKCKPSQVTIQFVTKAAFEAASQLWSAVNSDRIWHLQLFTSWKGCYTDG
GNLKPFLHTEVDFSEKLAATLTGNETDWKTA AHTFIMSSGEHFDETPPTADGSS
RPLLTRSSIFTKAGKEFAKEFKSVAKSIDKVENKFLKKIRKELHATLANKHKVLRI
AFDKSYTGKTFSTPGKAASLNGSVTCTGCGPTGSLVLHTVIKVTLGEEPTVKLT
MKPQNLGVSLGLAMSAKSDFPESFGIETPLLEQTIPAAGFKIPEIASVGLVASLGY
GISVSNFKGNLNVAKNVSVSIPDGAKLNLLVNPSNDEQSKLGGSWAPVAQSTPM
HISGFGSGEFAVGLGISLSLKMEVLSFEITPAKITLSGPSIGFEFGVSSKECGPYTSV
GSSHDPTYITPKLGFSLVSGSDLNVDSGTWGLGIPIPDGGPNVNSSHGTGGSGG
GGGGTNPYPSSHTNPDQKDSSQPAVKTRSVNKRHNHDPGASKISISASKTYE
QSFRLHSPICFMSGLHQRRPLSNHGRKDEPDGGTRRLSNSSE

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 647	Number Disordered Regions: 14
Number residues disordered: 263	Longest Disordered Region: 94
Overall percent disordered: 40.65	Average Prediction Score: 0.4942
Predicted disorder segment [1]-[5]	Average Strength= 0.7473
Predicted disorder segment [36]-[36]	Average Strength= 0.5165
Predicted disorder segment [92]-[98]	Average Strength= 0.5592
Predicted disorder segment [174]-[174]	Average Strength= 0.5079
Predicted disorder segment [186]-[188]	Average Strength= 0.5043
Predicted disorder segment [192]-[192]	Average Strength= 0.5019
Predicted disorder segment [195]-[255]	Average Strength= 0.6964
Predicted disorder segment [278]-[291]	Average Strength= 0.5958
Predicted disorder segment [321]-[322]	Average Strength= 0.5085
Predicted disorder segment [341]-[347]	Average Strength= 0.5378
Predicted disorder segment [412]-[424]	Average Strength= 0.5981
Predicted disorder segment [482]-[499]	Average Strength= 0.5869
Predicted disorder segment [514]-[607]	Average Strength= 0.8037
Predicted disorder segment [612]-[647]	Average Strength= 0.8180

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	503	531	29

2	573	584	12
3	587	626	40
Filtered Regions			
	From	To	Length
1	268	270	3
2	642	642	1
3	644	647	4

ModPred and PROSITE:

ModPred: Proteolytic cleavage (R49, D570, R579), Amidation (T82, T119, L217, I351, S461, S521, K583), ADP-ribosylation (R215, R579), Acetylation (K233, K244, K248), O-linked glycosylation (S533, S547), Phosphorylation (S643, S645, S646).

PROSITE: No identified domain recognition sites.

Structural modelling:

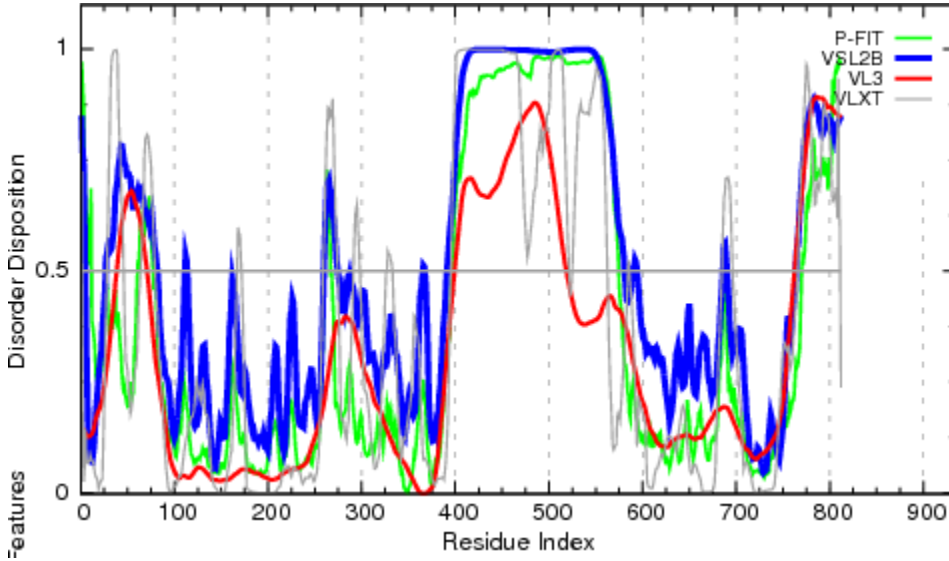
No templates were found matching target sequence.

> MVLG_03092T0 (40.02%)- 812 aa

MKFSNSIIACALLASGWADLARTHGVDAGGGSIGVSLPPTTPSDPSCTGSNEWF
 SPYRSTFEAHERGENLDQLKQPGSIDDSEYLLNPNFNSINNCQTRYYYLDIHETRA
 APDGFEREMFLFNGRINGPLIEANQGDITVVYVHNYLDIGTTVHWHGLAQNGSG
 WADGPLGVTQCPIPPGTTFIYKYTLSRFDQCGTYWYHAHRLAHYS DGLVAPLVI
 HCPNDPLKRGDLYDIDQVVVVRDHYHPLSTRIISALLVNGSFQGSATPSPNAGLI
 NGRGRYNCSFAPEGSVCTDDAPLTEFEFPKGSRVRLRLINPSAHAQFLVSVDEHP
 LNVVEADDTPVWQTTVHRIPINVGQRY SAILNTADNNEGDSFWMRADINTACFG
 ANFTDLNPEVKAIIRIGPASSSPSSSVSSASSENGSNPPSQGASNGSSDSSGDHSS
 GNPSDLQSSNNNGTSGDAPDGSQSDNSGDGNASGDGRGDWGNQEDSDGSG
 DSGDQQGSQRGSRWKRAAGLRKRNGNNNDNNDNNDNNDNNDNNDNNDNNDNND
 NNDNNSGSNTNTNQNLPTSTDWSDAVNGSCHDLAESTLVPRVFPNPPGASISH
 EFRATILTPSGAFGFAANNV SFESFVDDPFLFRVNRGDDIPLGLSASIVLDDKSLA
 HDIVINNANPIDHPFHLHGVMHLIARGAGSVSADNISSVALNLNLPPIRRDTISVT

GNTFAIVRVVADNAGVWAIHCHILPHQVTGLMGVVVIRPDLIRKMEIPQHARDL
 CTLGSSLSSAQQQDPQPNIEPGRRIIRSINPLLPSKDFIRKRVLNND

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 812	Number Disordered Regions: 9
Number residues disordered: 325	Longest Disordered Region: 189
Overall percent disordered: 40.02	Average Prediction Score: 0.4951
Predicted disorder segment [1]-[4]	Average Strength= 0.7432
Predicted disorder segment [27]-[81]	Average Strength= 0.6519
Predicted disorder segment [112]-[113]	Average Strength= 0.5300
Predicted disorder segment [260]-[274]	Average Strength= 0.6220
Predicted disorder segment [366]-[367]	Average Strength= 0.5105
Predicted disorder segment [392]-[580]	Average Strength= 0.9414
Predicted disorder segment [590]-[594]	Average Strength= 0.5136
Predicted disorder segment [686]-[690]	Average Strength= 0.5364
Predicted disorder segment [765]-[812]	Average Strength= 0.8098

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	98	103	6
2	371	403	33

3	440	452	13
4	459	465	7
5	471	488	18
6	498	516	19
7	563	588	26
8	801	809	9
Filtered Regions			
	From	To	Length
1	6	17	12

ModPred and PROSITE:

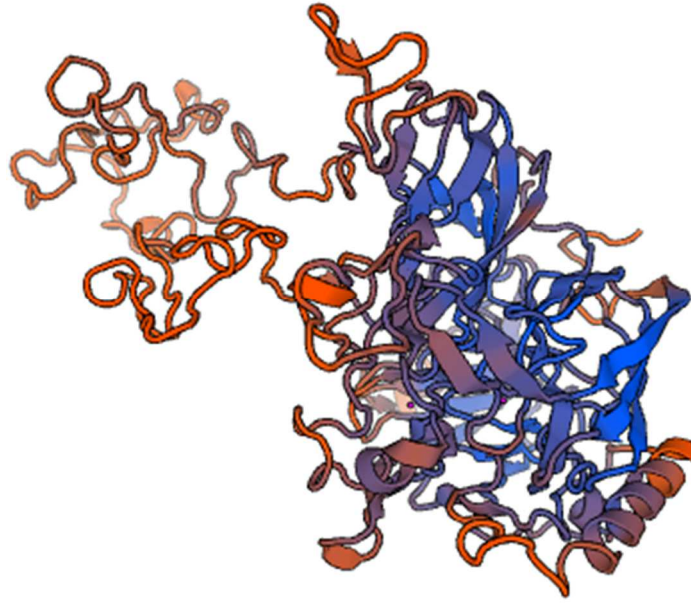
ModPred: Amidation (S4, H346, E579, V586, L797), Proteolytic cleavage (R22, Q235, R509, R791), Sumoylation (K394), Phosphorylation (S405), GPI anchor amidation (N418), ADP-ribosylation (R514)

PROSITE: Multicopper_oxidase1 (725-745, PROSITE entry PS00079),

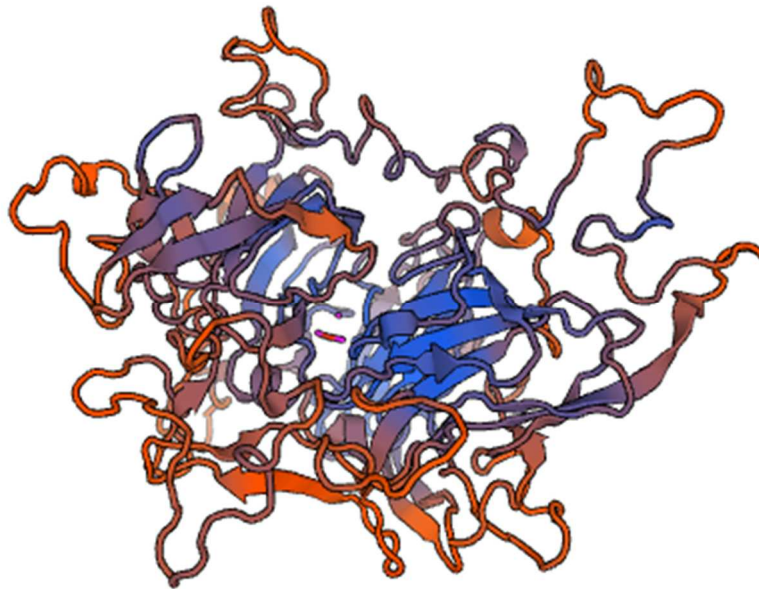
Multicopper_oxidase2 (730-741, PROSITE entry PS00080).

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
3sqr.1.A	laccase	31.29	X-ray, 1.7Å	monomer	5 x <u>NAG</u> , 3 x <u>CU</u> , 1 x <u>MAN</u>
3gyr.1.A	Phenoxazinone synthase	17.43	X-ray, 2.3Å	homo-hexamer	6 x <u>C2O</u> , 18 x <u>CU</u>



Model #1: Residues 83-787 of MVLG_03092T0 with 3sqr.1.A (31.29 % sequence identity) as a template

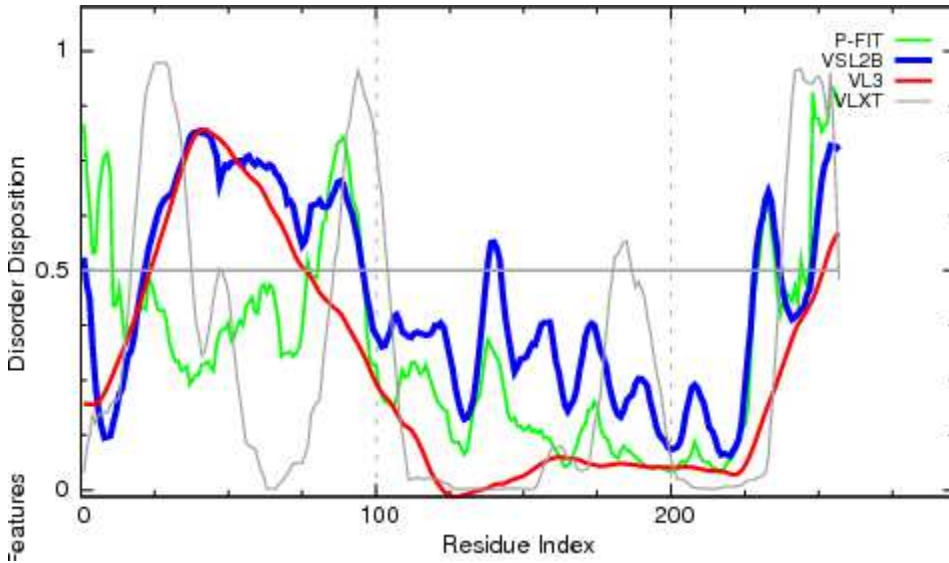


Model #2: Residues 70-756 of MVLG_03092T0 with 3gyr.1.A (17.43 % sequence identity) as a template

> MVLG_01652T0 (38.13%)- 237 aa

MFRTLYRLACLSLVPALLQAAPAPVDSSLGQGVSA LKTRGSQSCRRVAQCTQA
 APRNAVQTCNGGKCGFACKSGYTWKDKKCQAASSGQATSGGTL LAAVSGHMV
 DAQLASNGITGFRAQSN GWNTNAIASWFRTDSIQDSTNGH SWCYNEYDDSLPGF
 APDVSVM LANFGGSNV RAGQAYCGLEAEVVTADGRTVNLIIMDGFDSK WVRTP
 ASIDVIYNAFGL LHGSTTNDKNTVESGVKWR L TGRRDSRYTFNSS

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 257	Number Disordered Regions: 4
Number residues disordered: 98	Longest Disordered Region: 74
Overall percent disordered: 38.13	Average Prediction Score: 0.4307
Predicted disorder segment [22]-[95]	Average Strength= 0.6931
Predicted disorder segment [138]-[142]	Average Strength= 0.5438
Predicted disorder segment [229]-[236]	Average Strength= 0.6128
Predicted disorder segment [248]-[257]	Average Strength= 0.7063

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			

	From	To	Length
1	1	8	8

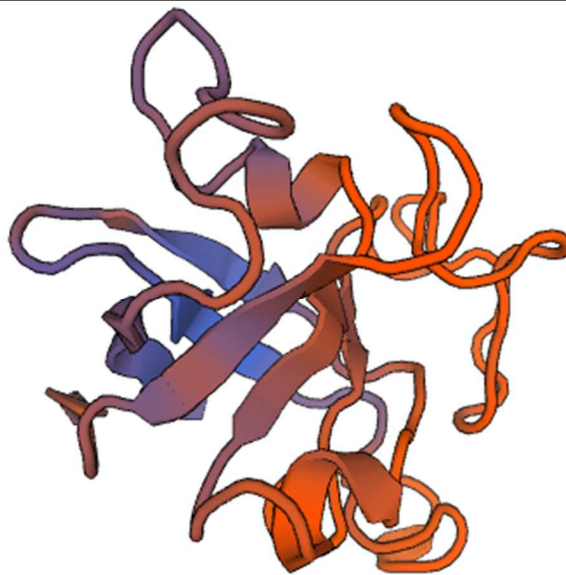
ModPred and PROSITE:

ModPred: Amidation (A9, G32, F222), Pyrrolidone carboxylic acid (Q19), Disulphide linkage (C67, C148), O-linked glycosylation (S92), Proteolytic cleavage (D136).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
2hcz.1.A	Beta-expansin 1a	19.82	X-ray, 2.8Å	monomer	None
1k4r.1.A	MAJOR ENVELOPE PROTEIN E	17.65	EM, 24.0Å	homo- trimer	None



Model #1: Residues 123-244 of MVLG_01652T0 with 2hcz.1.A (19.82 % sequence identity) as a template

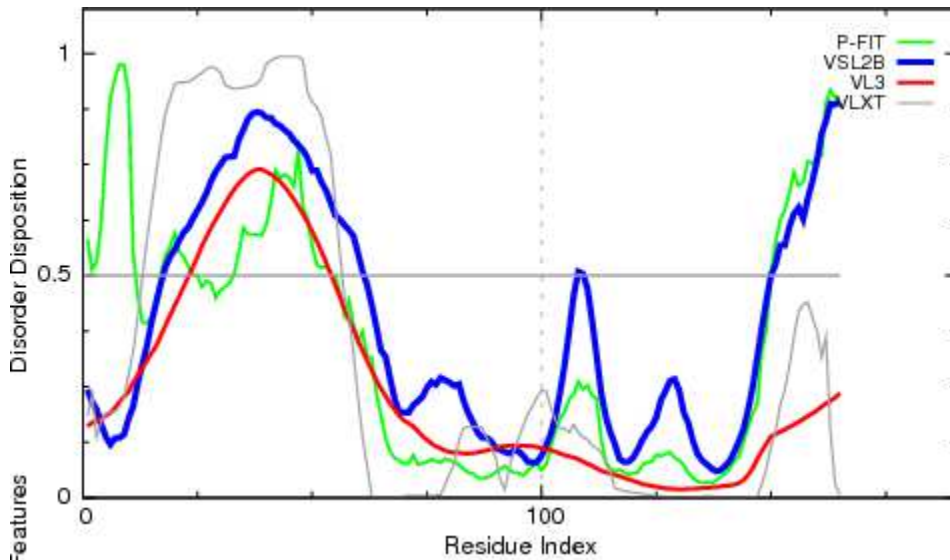


Model #2: Residues 130-191 of MVLG_01652T0 with 1k4r.1.A (17.65 % sequence identity) as a template

> MVLG_02018T0 (37.58%)- 165 aa

MLLWLRVAALALAATGGPVSASPLKALDNALSSLNSTGPASKLTPIPTPPTISL
 LSKSKTYYPGDTVFFKWDRAAPTMQSADLFIAYSGPLATVPICVTRDMLLQPDR
 GSMVLHSAYVIPWKELLGQKRATVEGFIYFVYSPTHYRFDTGVTGGKSDSFTIQH
 R

PONDR:



=====PONDR VSL2
 STATISTICS=====

Predicted residues: 165
 Number residues disordered: 62
 Overall percent disordered: 37.58
 Predicted disorder segment [18]-[61]
 Predicted disorder segment [108]-[109]
 Predicted disorder segment [150]-[165]

Number Disordered Regions: 3
 Longest Disordered Region: 44
 Average Prediction Score: 0.3924
 Average Strength= 0.7153
 Average Strength= 0.5065
 Average Strength= 0.6979

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			
	From	To	Length
1	1	8	8

ModPred and PROSITE:

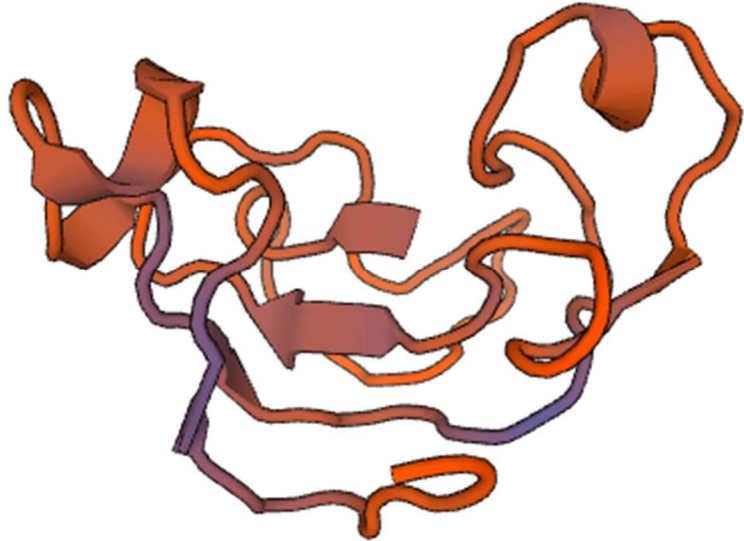
ModPred: O-linked glycosylation (T49), Amidation (V140).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Coverage	Identity	Method	Oligo State	Ligands
2cp5.1.A	Restin	Created with Raphaël 2.2.0	23.44	NMR	monomer	None
1bcp.1.D	PERTUSSIS TOXIN	Created with Raphaël 2.2.0	35.19	X-ray, 2.7Å	hetero-oligomer	1 x <u>ATP</u>
2jgx.1.A	COMPLEMENT FACTOR H	Created with Raphaël 2.2.0	9.80	NMR	monomer	None

Name	Title	Coverage	Identity	Method	Oligo State	Ligands
1tnr.1.A	TUMOR NECROSIS FACTOR BETA	Created with Raphaël 2.2.0	28.57	X-ray, 2.8Å	hetero- oligomer	None



Model #1: Residues 123-244 of MVLG_02018T0 with 2cp5.1.A (23.44 % sequence identity) as a template



Model #2: Residues 123-244 of MVLG_02018T0 with 1bcp.1.D (35.19 % sequence identity) as a template



Model #3: Residues 123-244 of MVLG_02018T0 with 2jgx.1.A (9.80 % sequence identity) as a template

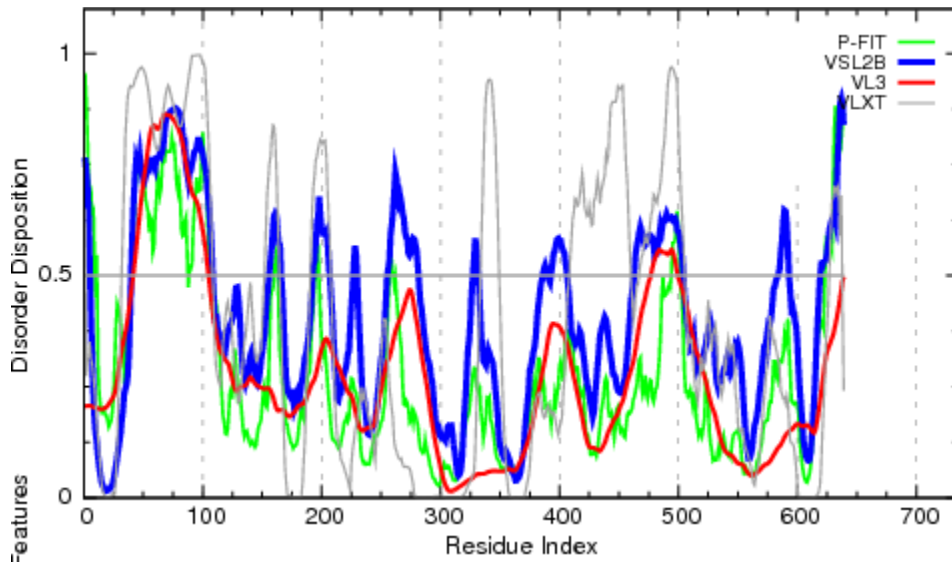


Model #4: Residues 123-244 of MVLG_02018T0 with 1tnr.1.A (28.57 % sequence identity) as a template

> MVLG_01824T0 (34.74%)- 639 aa

MVYTRPCRARAAVFASIVLCLIVVVVPTSvlaLETAIVAPPDGNSSKGPDTETIP
 TLAGADDGETLFQLNSTEMSPEEPLLEEPHWLPLLKRARRRKPPSEDAWSGETTT
 RIIGKLGIGAMQAIQSSDDELLVRHRAAFGAKLRPRILTQDITETSLIGSSSLSDNP
 YLGIGRKCDFCAGGTFLSNGDIISVGGQPSEHTELGKPGFAEDGFTGLRIFQPTSH
 RLLDNPKKVHIQSARWYASVVRVTDGSALIMGGSKKGQYNNDPKVDNPTMEFF
 PSKGPQFYKFLQDALDSNLFPLAFLLSGSGNIFVANHVAMIYDWKHNREHRV
 KGVPGGIVATYPGSGTAVLLPLTIKNNWISEVLICGGVFNTVNLNPGFNVRAD
 PVSDQCARTSFPRGNSMSGWEVEHMLSPRIMGDPVITPDGQVLIVGGAKTGTAG
 YGNAIGMDAAVPNLVPTLYNPDAPRGQRFSEEFPPAKIERMYHSTSLTTEGSVL
 TMGSSPNRILTRLTYKSRFEVELIAPPYMTKKRPAILNYPQQIKYNGRYTLTMSN
 PMGCDNVRVVLIDGGYATHALHMNQRSVELLVTSSNQSTITFQSPHDGTIWP
 PAFLWITVCEGKIPSKGHKIMVGDGSNPPNYKAPF

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 639

Number residues disordered: 222

Overall percent disordered: 34.74
0.4069

- Predicted disorder segment [1]-[5]
- Predicted disorder segment [40]-[108]
- Predicted disorder segment [156]-[165]
- Predicted disorder segment [193]-[206]
- Predicted disorder segment [227]-[229]
- Predicted disorder segment [256]-[281]
- Predicted disorder segment [328]-[331]
- Predicted disorder segment [386]-[390]
- Predicted disorder segment [393]-[406]
- Predicted disorder segment [463]-[504]
- Predicted disorder segment [584]-[593]
- Predicted disorder segment [620]-[639]

Number Disordered Regions: 12
Longest Disordered Region: 69
Average Prediction Score:

- Average Strength= 0.6753
- Average Strength= 0.7657
- Average Strength= 0.5868
- Average Strength= 0.5926
- Average Strength= 0.5466
- Average Strength= 0.6077
- Average Strength= 0.5577
- Average Strength= 0.5129
- Average Strength= 0.5531
- Average Strength= 0.5905
- Average Strength= 0.5737
- Average Strength= 0.6653

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length

1	85	90	6
2	113	121	9
3	173	183	11
4	606	614	9
Filtered Regions			
	From	To	Length
1	14	27	14
2	66	69	4
3	363	366	4
4	427	430	4
5	510	511	2
6	560	560	1

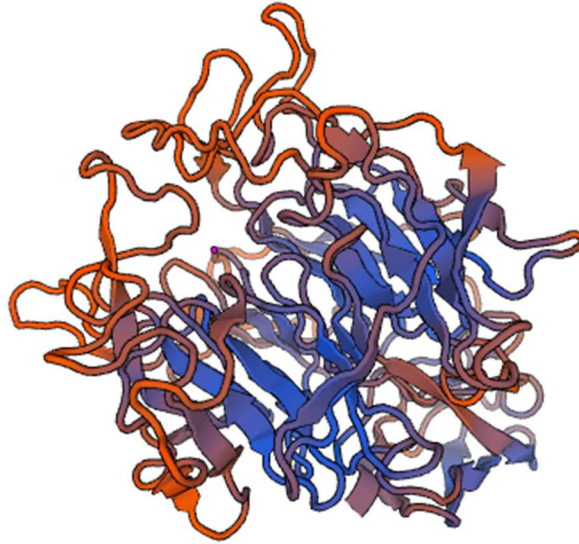
ModPred and PROSITE:

ModPred: Amidation (A12, F276, Y635), Proteolytic cleavage (R92, R95, R136, R329, K331, R464, R467), Ubiquitination (K323), Methylation (R464).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
5lxz.1.A	Secreted protein	21.57	X-ray, 1.5Å	monomer	2 x <u>CU</u>

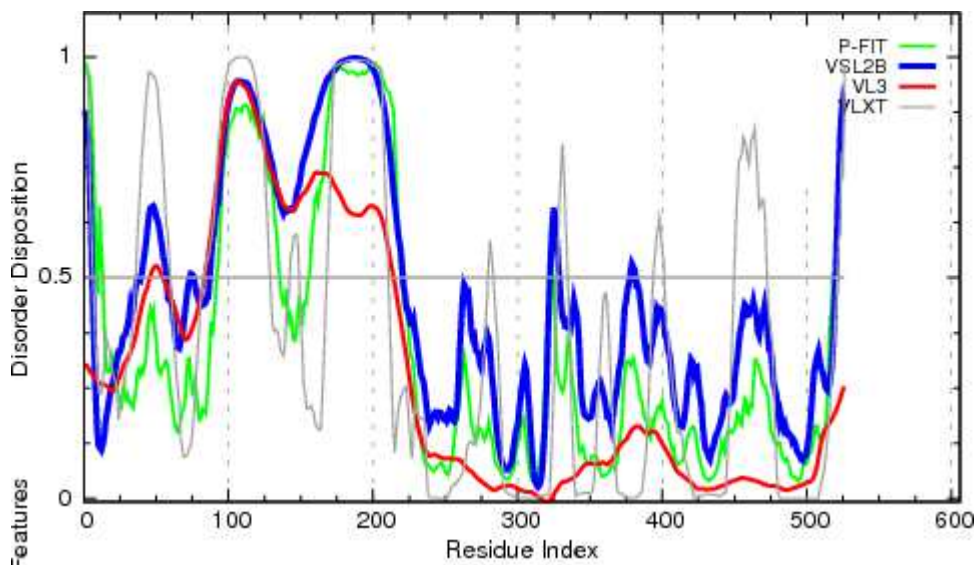


Model: Residues 99-627 of MVLG_01824T0 with [5lxz.1.A](#) (21.57 % sequence identity) as a template

> MVLG_00243T0 (33.33%)- 528 aa

MSRMNTRAVLALVLVASLQALSPITGITFAAGPEQALAVRAPAPAAAATARS
RVQQLGADRYVRLATLHTRDDEHLLLEARLLNHVMIEPGLAIRAADESESEEDTEE
AETDDEAEFEETVEEVIERGNSRVFATSARPILAGMRPPHVIPNPKDPDFGKLT
TPVRTTTTTASAKSTATTPRATTAPSTMTTTTTTTTTTTSTRDAAAASIPTGLG
CFPSNVKSIPTGVNYTATDLASSWWCADSSEYAFIGFSYSVDECQSPSTLLASFTR
MRKQFGARDVRLYGACDATWFNDALVDAAASANLDVYHLIWFDFGDDQRKS
RYSAFVKTMRTNPKAPFVFKNVAIGSEPLYDGVLSATNLVTEIFSMKSKMAPYG
TKATFSEMPYGLQINNGAPSTMAAADFVEGNVLPFFDSQATTGANAWGVVSW
LSYFASLAPGKIIRMTQTGWPSDQSVWKANTPTAVSSILSQASYALLDSKCSWF
NANGGIGWFAHIYSDDSLPGWLLNNGNLKFPFAPKSSC

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 528

Number residues disordered: 176

Overall percent disordered: 33.33
0.4540

Predicted disorder segment [1]-[5]

Predicted disorder segment [40]-[57]

Predicted disorder segment [74]-[76]

Predicted disorder segment [88]-[220]

Predicted disorder segment [323]-[328]

Predicted disorder segment [378]-[381]

Predicted disorder segment [522]-[528]

Number Disordered Regions: 7

Longest Disordered Region: 133

Average Prediction Score:

Average Strength= 0.7728

Average Strength= 0.5946

Average Strength= 0.5057

Average Strength= 0.8449

Average Strength= 0.5981

Average Strength= 0.5212

Average Strength= 0.7649

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	79	96	18
2	116	151	36
3	155	166	12
4	214	221	8

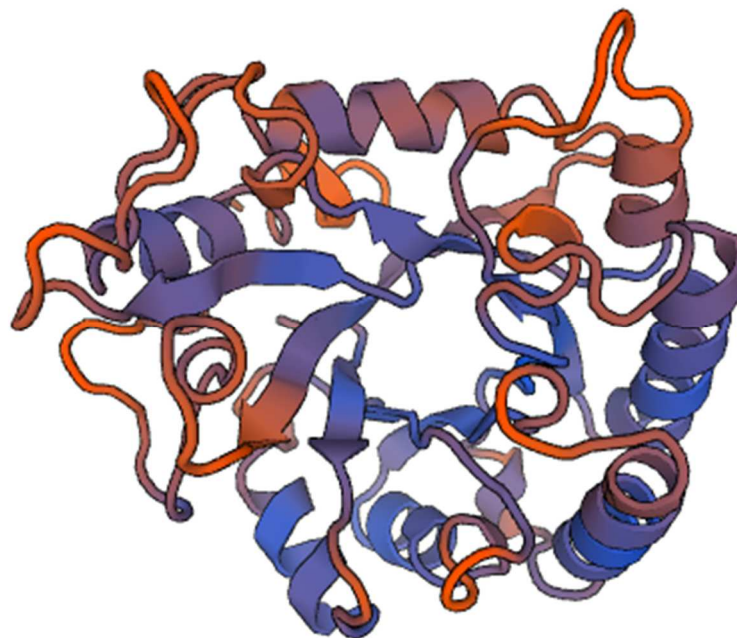
ModPred and PROSITE:

ModPred: Amidation (A8, G359, W454, G496, Y502), Proteolytic cleavage (R62, K160, R283, R286, D322, K326, Y329, S330, D407), Carboxylation (E99, E106, E107, E109, E113, E117, E118, E121, E122), O-linked glycosylation (T169, T170, T171, T178, T180, T182, T186, S190, T196, T197, T198, T199, T200, T201, T202, T203, T204, T205, T206), Hydroxylation (P189), N-linked glycosylation (N397).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
4wtr.1.A	beta-1,3-glucanosyltransferase	20.54	X-ray, 2.3Å	monomer	4 x <u>BGC</u>
3pz9.1.A	Mannan endo-1,4-beta-mannosidase. Glycosyl Hydrolase family 5	16.02	X-ray, 1.4Å	monomer	None



Model #1: Residues 250-528 of MVLG_00243T0 with 4wtr.1.A (20.54 % sequence identity) as a template

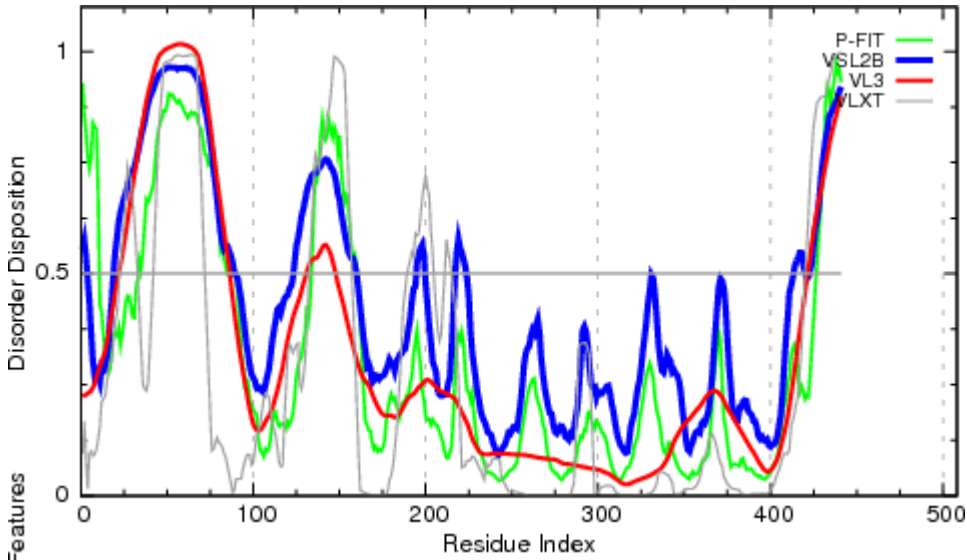


Model #2: Residues 251-458 of MVLG_00243T0 with 3pz9.1.A (16.02 % sequence identity) as a template

> MVLG_02872T0 (33.33%)- 441 aa

MFFRASTVFTSLLLVAPALAEAPSPDKRGFKLELHSRASHNPQHAGKAPKSGH
AQPYAKRLKHGPAHPRRGGAHAHQGHPPFLAATKAGSRQSFANLANVENIDWS
VEVTFGSPQVRVPLFLSMGSSLSSVADQNIKSDAKTRYNPSKSLTARNMTKAQV
DPNTGVTFITYKDKISIGGFVSDQTFVMTSTPNGDPLERVYDSPVPWAGALAL
GRTSKGTPSLSFLENLIRSKVIDNAVCGISLTVEGGALFFGGIDSHSFKGKIVWSPV
ETHYMEGFWTIKTGGWGWKGVATGTAGLLQFAPENTYTYISAILGNKLFAGIK
HHVDSKTQRYLLPCNSNASDTIGFFIHNRMFPVPIPDILFPSDSDPTMCHTALLQ
VTNKHILDDYTVVMGALHMRSFYTILSYEKEHGGPAIGLAESSIKVMGGDPGPG
GHSEK

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 441

Number residues disordered: 147

Overall percent disordered: 33.33
0.4227

Predicted disorder segment [1]-[3]

Predicted disorder segment [19]-[89]

Predicted disorder segment [124]-[159]

Predicted disorder segment [195]-[199]

Predicted disorder segment [218]-[223]

Predicted disorder segment [414]-[420]

Number Disordered Regions: 7

Longest Disordered Region: 71

Average Prediction Score:

Average Strength= 0.5550

Average Strength= 0.8052

Average Strength= 0.6611

Average Strength= 0.5464

Average Strength= 0.5458

Average Strength= 0.5277

Predicted disorder segment [423]-[441]

Average Strength= 0.7575

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	29	38	10
2	50	65	16
3	73	108	36
4	118	128	11
5	169	175	7
Filtered Regions			
	From	To	Length
1	1	19	19

ModPred and PROSITE:

ModPred: Amidation (A16, D139, P271, P362), Proteolytic cleavage (K31, H35, D204),

Acetylation (K59), Ubiquitination (K74), Methylation (K427)

PROSITE: Peptidase_A1 domain (105-422, PROSITE entry PS51767), Disulphide bridge (340-376)

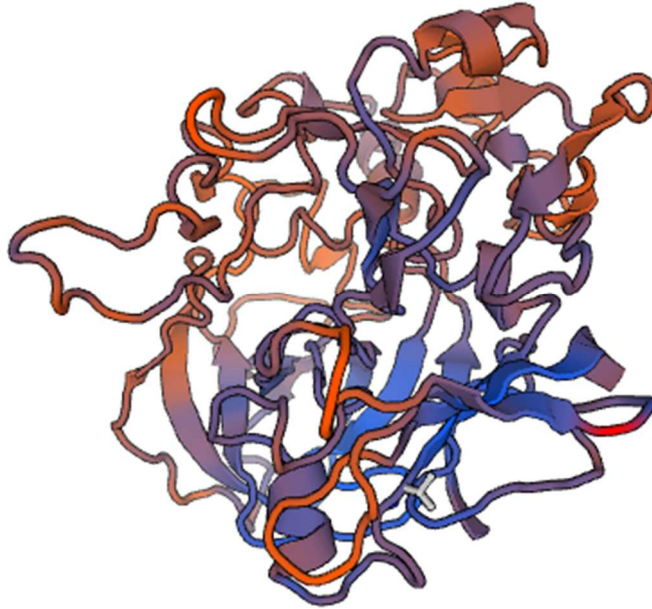
Eukaryotic Aspartyl proteases (Aps) form peptidase family A1.

Known eukaryotic Aps in Fungi:

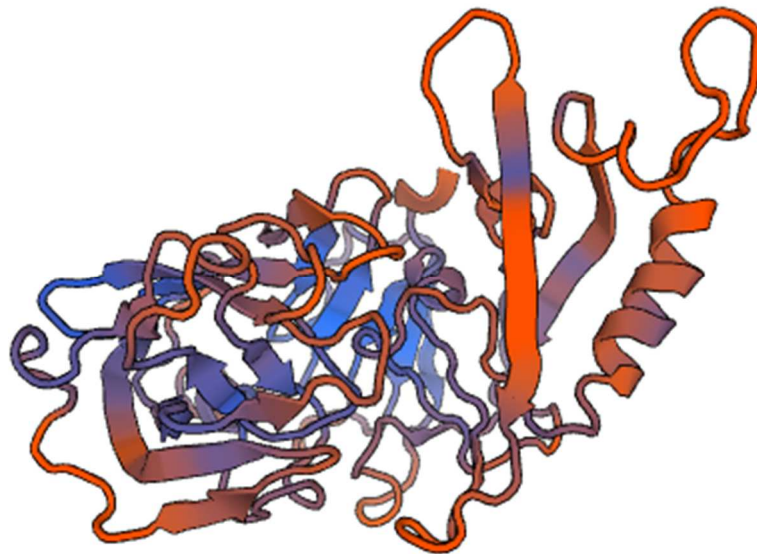
- Fungal proteases such as aspergillopepsin A (EC 3.4.23.18), candidapepsin (EC 3.4.23.24), mucoropepsin (EC 3.4.23.23) (mucor rennin), endothiapepsin (EC 3.4.23.22), polyporopepsin (EC 3.4.23.29), and rhizopuspepsin (EC 3.4.23.21).
- Yeast saccharopepsin (EC 3.4.23.25) (proteinase A) (gene PEP4). PEP4 is implicated in posttranslational regulation of vacuolar hydrolases.
- Yeast barrierpepsin (EC 3.4.23.35) (gene BAR1); a protease that cleaves α -factor and thus acts as an antagonist of the mating pheromone.
- Fission yeast *sxa1* which is involved in degrading or processing the mating pheromones.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
2psg.1.A	PEPSINOGEN	22.58	X-ray, 1.8Å	homo-dimer	None
3zkm.1.A	BETA-SECRETASE 2	24.17	X-ray, 1.8Å	hetero-oligomer	None



Model #1: Residues 29-425 of MVLG_02872T0 with 2psg.1.A (22.58 % sequence identity) as a template

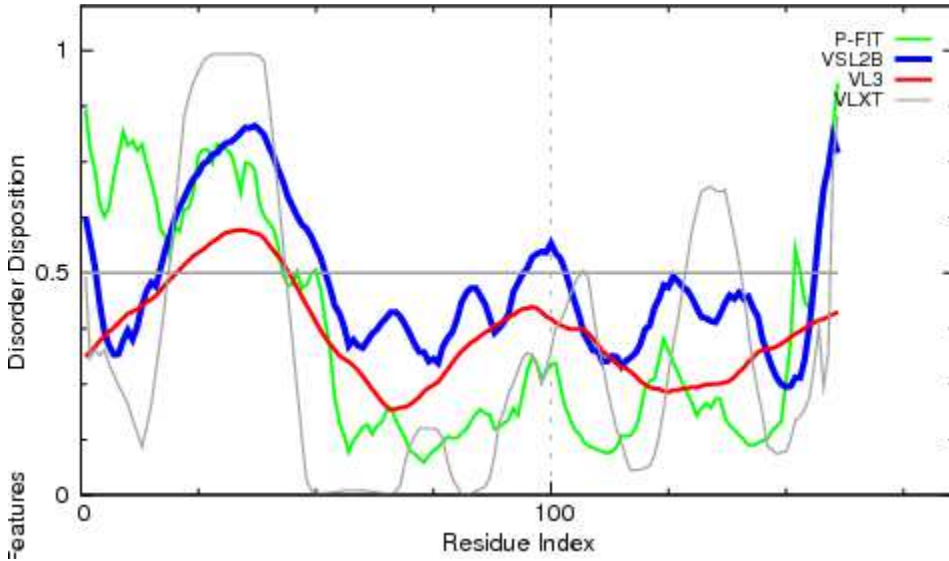


Model #2: Residues 96-428 of MVLG_02872T0 with 3zkm.1.A (24.17 % sequence identity) as a template

> MVLG_05120T0 (32.92%)

MLLKLTFLLALTLVSLGVSASQSNDTVGRAMSPSSTKSTVTSPAYGKALKTD
 EAFSFVYYPAEGDRQDSFFQITKLSLKAVEKDQFPTDFANDLSSAQTEPVSVD
 RLPPLEYFNANTVKTAKTGDSIEAMLEISEQNFQKQVQTINVPLTITLHAHSQ

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 161
 Number residues disordered: 53
 Overall percent disordered: 32.92
 0.4757
 Predicted disorder segment [1]-[3]
 Predicted disorder segment [17]-[52]
 Predicted disorder segment [95]-[103]
 Predicted disorder segment [157]-[161]

Number Disordered Regions: 4
 Longest Disordered Region: 36
 Average Prediction Score:
 Average Strength= 0.5702
 Average Strength= 0.6993
 Average Strength= 0.5370
 Average Strength= 0.7201

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			
	From	To	Length

1	1	13	13
---	---	----	----

ModPred and PROSITE:

ModPred: Amidation (A20, Y115), Proteolytic cleavage (E28, R30, F56, D69, D107),

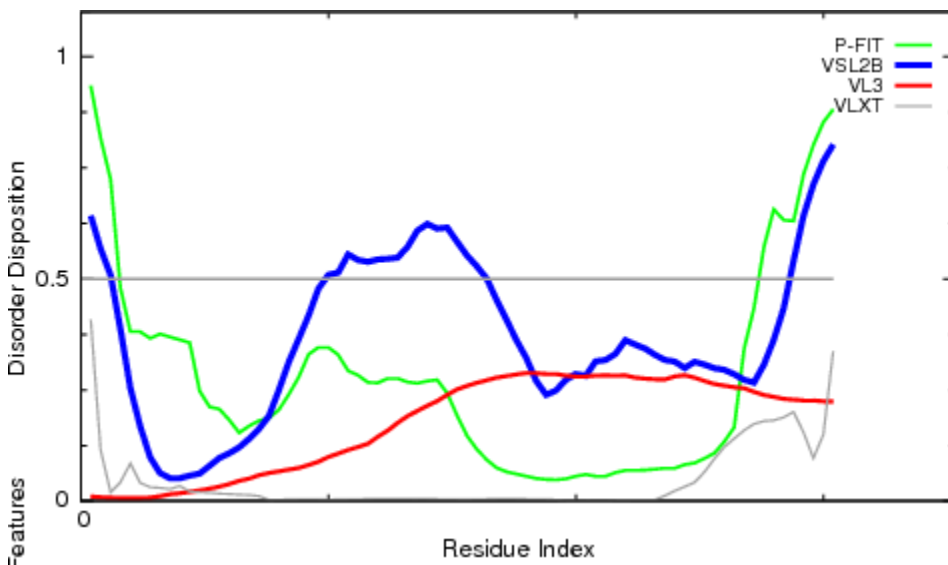
Phosphorylation (S33), ADP-ribosylation (R109).

PROSITE: No identified domain recognition sites.

> MVLG_00885T0 (32.89%)- 76 aa

MRFSLAFFAVPFLVGQVVASVSDWSAKNGSFKCTSNEAGKGGKCMVCVHSNL
 DIFNTSLSQACGNCGEFCTSNVHA

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 76
 Number residues disordered: 25
 Overall percent disordered: 32.89
 0.3704
 Predicted disorder segment [1]-[3]
 Predicted disorder segment [25]-[41]
 Predicted disorder segment [72]-[76]

Number Disordered Regions: 3
 Longest Disordered Region: 17
 Average Prediction Score:
 Average Strength= 0.5726
 Average Strength= 0.5583
 Average Strength= 0.6922

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			

ModPred and PROSITE:

ModPred: Amidation (A6, L52).

PROSITE: No identified domain recognition sites.

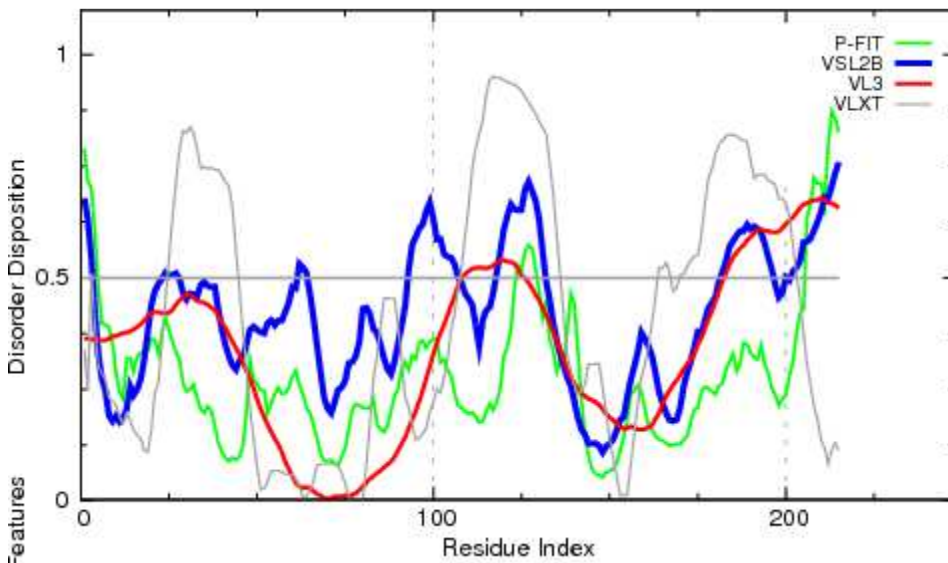
Structural modelling:

No templates were found matching target sequence.

> MVLG_00677T0 (32.03%)- 215 aa

MRASVILPLCLGLLSYCASAPSLEPPFTYQIPSPERIAELAQPYLINPENYSKTV
YYRHEPRDHISYVMYAFQSMTNTSKWASAVVRHTYGHPPKDSNFEPPTTDIAL
SMPHIIPAPIPEGTPEGEGQPVQVEYVRNRNVYRKVFLWHQPRSAVLGLEEKLQL
GTLIDLGNIDPTTQPVMREYSEDMAKELGEGCPVSGAMFERHMKDVQMYQ

PONDR:



STATISTICS=====

Predicted residues: 215	Number Disordered Regions: 8
Number residues disordered: 69	Longest Disordered Region: 15
Overall percent disordered: 32.09	Average Prediction Score:
0.4189	
Predicted disorder segment [1]-[3]	Average Strength= 0.6180
Predicted disorder segment [24]-[27]	Average Strength= 0.5074
Predicted disorder segment [62]-[64]	Average Strength= 0.5218
Predicted disorder segment [93]-[107]	Average Strength= 0.5807
Predicted disorder segment [118]-[131]	Average Strength= 0.6365
Predicted disorder segment [182]-[196]	Average Strength= 0.5799
Predicted disorder segment [200]-[200]	Average Strength= 0.5007
Predicted disorder segment [202]-[215]	Average Strength= 0.6257

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	142	150	9
2	163	171	9
Filtered Regions			
	From	To	Length
1	10	13	4
2	70	74	5

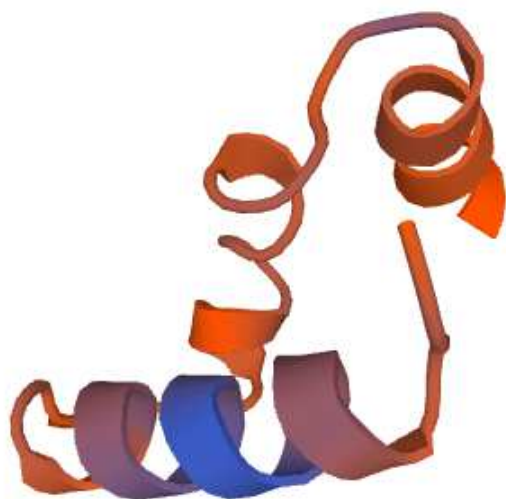
ModPred and PROSITE:

ModPred: Proteolytic cleavage (E135), ubiquitination (K191).

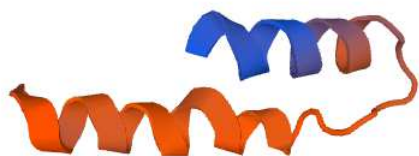
PROSITE: No identified domain recognition sites.

Structural modelling:

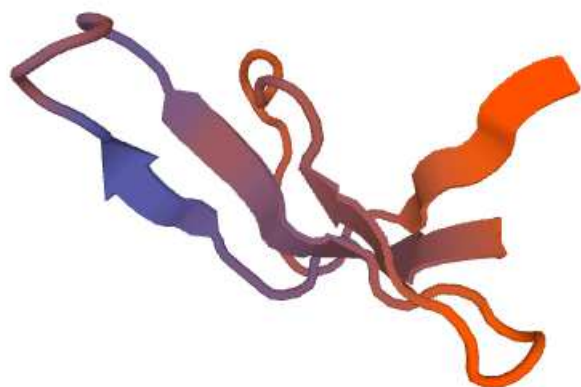
Name	Title	Identity	Method	Oligo State	Ligands
4x3i.1.A	Activity-regulated cytoskeleton-associated protein	27.91	X-ray, 1.8Å	monomer	1 x <u>ALA-THR-ARG-ASN-PHE-SER-GLY</u>
2d9d.1.A	BAG family molecular chaperone regulator 5	23.68	NMR	monomer	None
1r12.1.A	PROTEIN (RIBOSOMAL PROTEIN L2)	20.00	X-ray, 2.3Å	monomer	None
1m6x.1.I	Flp recombinase	33.33	X-ray, 2.8Å	homo-tetramer	None



Model #1: Residues 60-102 of MVLG_00677T0 with 4x3i.1.A (27.91% sequence identity) as a template



Model #2: Residues 155-193 of MVLG_00677T0 with 2d9d.1.A (23.68% sequence identity) as a template



Model #3: Residues 128-172 of MVLG_00677T0 with 1rl2.1.A (20.00% sequence identity) as a template

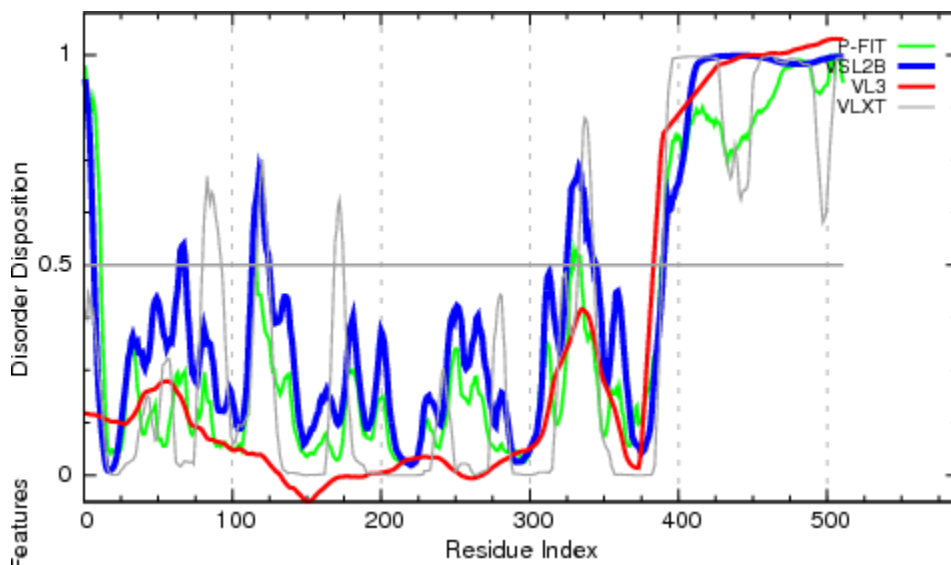


Model #4: Residues 58-99 of MVLG_00677T0 with 1m6x.1.I (33.33% sequence identity) as a template

> MVLG_02682T0 (31.70%)- 511 aa

MVQVHHRPTGIKRLFCWVLVLQLASLSAFAALDTAKREACQVFTNSPDLSKC
PVDTIYVSARDPKAKFKSIQQAINYLKATTANTPATILIGSGVYQEQLVVDGFASI
TLLGQTTSPRSSYAHNTVDINHNRVLQKSDGYQNWLTSTLLVNGCADFKAYNL
NLRQTAPVGIALAVAVMSSSGSFYACAIEGYQDTLFLGPNKTRGYLYGCYVSGV
VDFIYGWATLVVKDSQIMLLGEGTAYVAWRGAETTTSGAYFFGSTFDAAQNSF
GKIYPRTVAVGRAWNDKARIVILDCYLGSMIVPGIFAPWSYNPKDTRLSNEVFFG
EYNSQGPSEAKSKIVDSRTGKVDVDHLLHVLDTKSAAPYYSLSTIFGQDILWID
GNFNVKAVSLASGGVGGSSTPGALAAAPALAAASLPPPAPSAGKGTPLTKQDPKK
PTQDPKKTEHNPTKQKSKHHSHPKRKKKATSTKPQVDSGAGASSNHVSGGRP
NLASLPKKRGHGGSGGHPHHHHHHHHHHG

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 511

Number residues disordered: 162

Overall percent disordered: 31.70
0.4298

Predicted disorder segment [1]-[6]

Predicted disorder segment [65]-[68]

Predicted disorder segment [114]-[124]

Predicted disorder segment [325]-[342]

Predicted disorder segment [344]-[344]

Predicted disorder segment [390]-[511]

Number Disordered Regions: 6

Longest Disordered Region: 122

Average Prediction Score:

Average Strength= 0.8039

Average Strength= 0.5361

Average Strength= 0.6219

Average Strength= 0.6352

Average Strength= 0.5044

Average Strength= 0.9446

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	400	415	16
2	444	511	68
Filtered Regions			
	From	To	Length
1	368	389	22
2	422	426	5

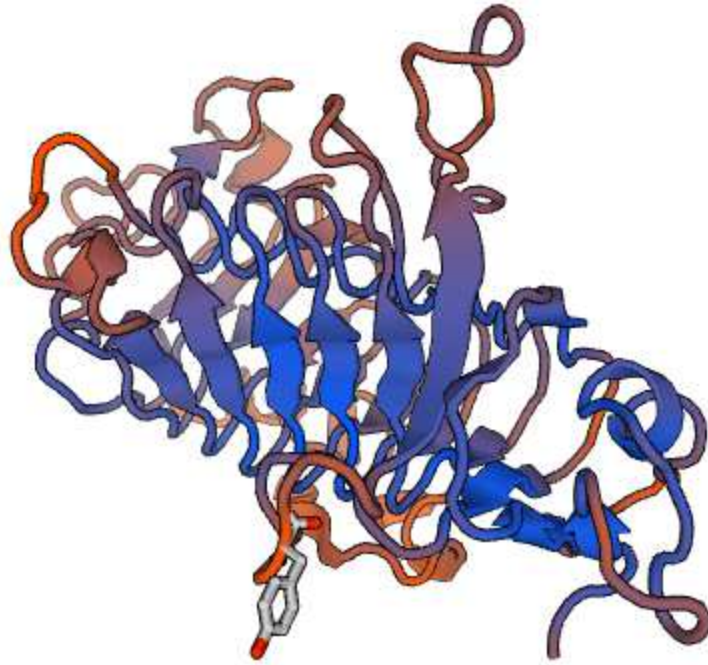
ModPred and PROSITE:

ModPred: Amidation (N45, L111, Y363, Y364), N-linked glycosylation (N151), O-linked glycosylation (S252), Proteolytic cleavage (D340, K421, D469), Hydroxylation (P414, P415), ADP-ribosylation (R483), Methylation (K491).

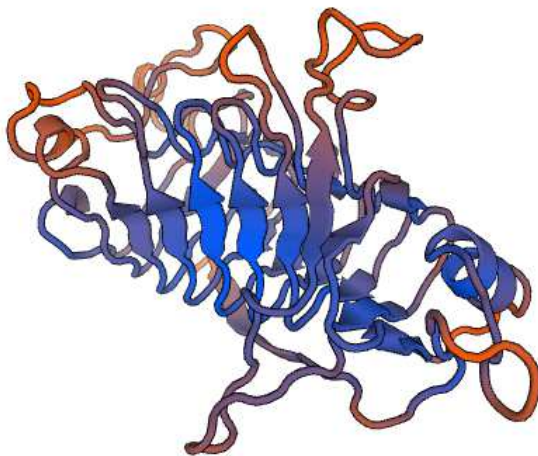
PROSITE: No identified domain recognition sites.

Structural modelling:

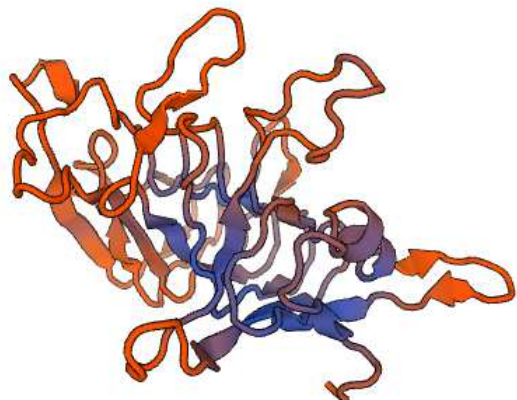
Name	Title	Identity	Method	Oligo State	Ligands
1xg2.1.A	Pectinesterase 1	21.12	X-ray, 1.9Å	hetero-oligomer	None
4pmh.1.A	Pectinesterase	22.18	X-ray, 1.8Å	monomer	None
4pew.1.A	Putative secreted protein	17.53	X-ray, 1.5Å	monomer	1 x <u>MG</u>
4xr6.1.A	Tail spike protein	12.24	X-ray, 1.8Å	homo-trimer	6 x <u>GLC</u> , 3 x <u>GLA</u> , 3 x <u>RAM</u> , 6 x <u>NAG</u> , 3 x <u>NDG</u>
4xop.1.A	Tail spike protein	10.34	X-ray, 1.6Å	homo-trimer	6 x <u>GLC</u> , 3 x <u>GLA</u> , 3 x <u>RAM</u> , 6 x <u>NAG</u> , 3 x <u>NDG</u>



Model #1: Residues 54-387 of MVLG_02682T0 with 1xg2.1.A (21.12% sequence identity) as a template



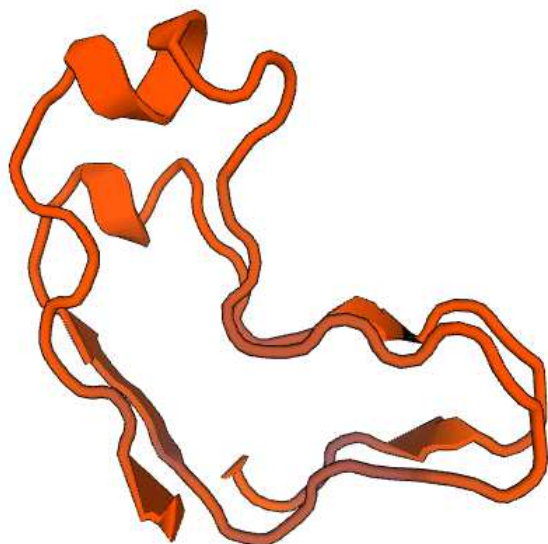
Model #2: Residues 41-332 of MVLG_02682T0 with 4pmh.1.A (22.18% sequence identity) as a template



Model #3: Residues 52-305 of MVLG_02682T0 with 4pew.1.A (17.53% sequence identity) as a template



Model #4: Residues 71-284 of MVLG_02682T0 with 4xr6.1.A (12.24% sequence identity) as a template

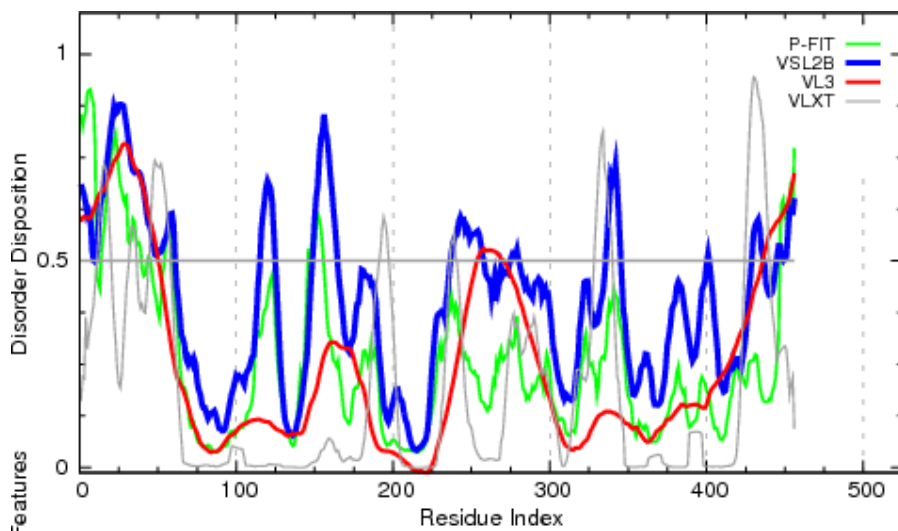


Model #5: Residues 153-215 of MVLG_02682T0 with 4xop.1.A (10.34% sequence identity) as a template

> MVLG_03747T0 (31.36%)- 456 aa

MKILAALPLSLAALAGAHKHSSGHSSHRYRHHRASGVLQASSGTTTCIVDESG
VGQDSTPKIMDAFTKCQKNAKIVLNGNYLVKSLLYTPMLYNVEIELTGTLTYSD
DIAYWSKPTTDTHGDGSYELYYQNVTFFFLQGEKIWLHGSPTSKTСКАЕКQSTF
NGNGQKWWDQFVKDKKAGNLHGIESTEYARPILLTIGNAKNVRVEYINFLNGPF
WNIFITHSKQVTMSNINIDAVSKSDSLPYNTDGVDTYNSDDVTLDFNVNNADD
CVSLKPNSTNVEVGRVNCNGSHGISVGS LGQYVDSYDIVENVYIHDISMSNAQA
GARIKAWPDRNGTAKDAGGGSGYVKNITFQNFVNKNVDEPLLITSCYMNSNEY
CTKFPSKMTVSDVHYINVTGTSSGKYKDVVALLDCSKECTGITAIGTHLSLPTPST
PPVYNCHNV DSEKQLDFHCTEL

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 456	Number Disordered Regions: 10
Number residues disordered: 143	Longest Disordered Region: 51
Overall percent disordered: 31.36	Average Prediction Score:
0.3970	
Predicted disorder segment [1]-[9]	Average Strength= 0.6124
Predicted disorder segment [11]-[61]	Average Strength= 0.6787
Predicted disorder segment [116]-[124]	Average Strength= 0.6396
Predicted disorder segment [149]-[164]	Average Strength= 0.6972
Predicted disorder segment [237]-[257]	Average Strength= 0.5645
Predicted disorder segment [275]-[280]	Average Strength= 0.5112
Predicted disorder segment [336]-[345]	Average Strength= 0.6419
Predicted disorder segment [401]-[401]	Average Strength= 0.5162
Predicted disorder segment [429]-[436]	Average Strength= 0.5581
Predicted disorder segment [445]-[456]	Average Strength= 0.5574

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	307	314	8
Filtered Regions			
	From	To	Length
1	1	13	13
2	136	139	4

3	213	213	1
4	215	218	4

ModPred and PROSITE:

ModPred: Proteolytic cleavage (R34, K340), Ubiquitination (K340), Amidation (K415).

PROSITE: POLYGALACTURONASE (Polygalacturonase active site, 286-299,

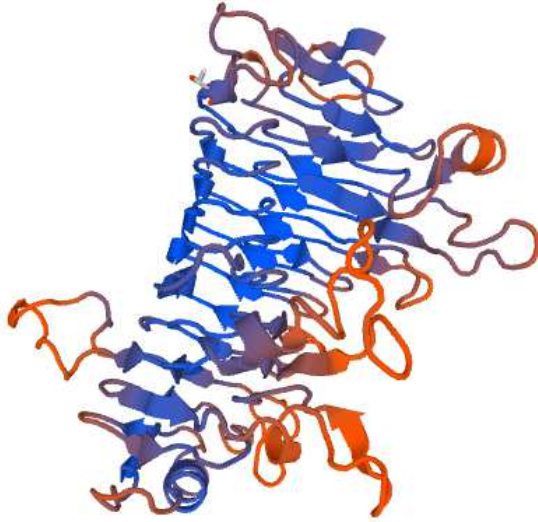
PROSITE entry PS00502)

Polygalacturonase (EC 3.2.1.15) (PG) (pectinase) catalyzes the random hydrolysis of 1,4- α -D-galactosiduronic linkages in pectate and other galacturonans. In plant bacterial pathogens such as *Erwinia carotovora* or *Pseudomonas solanacearum* and fungal pathogens such as *Aspergillus niger*, polygalacturonase is involved in maceration and soft-rotting of plant tissue.

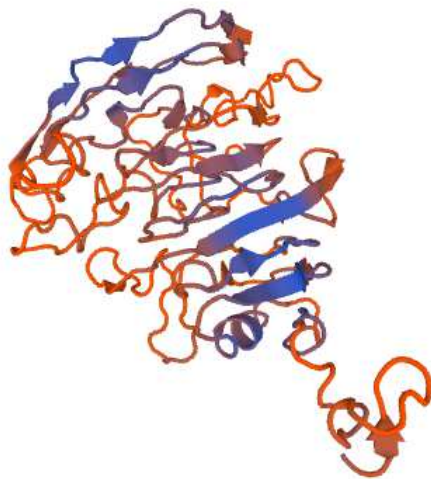
Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
4c2l.1.A	ENDO-XYLOGALACTURONAN HYDROLASE A	37.40	X-ray, 1.8Å	monomer	1 x <u>NAG</u> , 1 x <u>NAG</u> , 1 x <u>MAN</u>
3lmw.1.A	Iota-carrageenase, CgiA	14.05	X-ray, 2.6Å	monomer	1 x <u>NI</u> , 1 x <u>CA</u>
4xr6.1.A	Tail spike protein	17.69	X-ray, 1.8Å	homo-trimer	6 x <u>GLC</u> , 3 x <u>GLA</u> , 3 x <u>RAM</u> , 6 x <u>NAG</u> , 3 x <u>NDG</u>

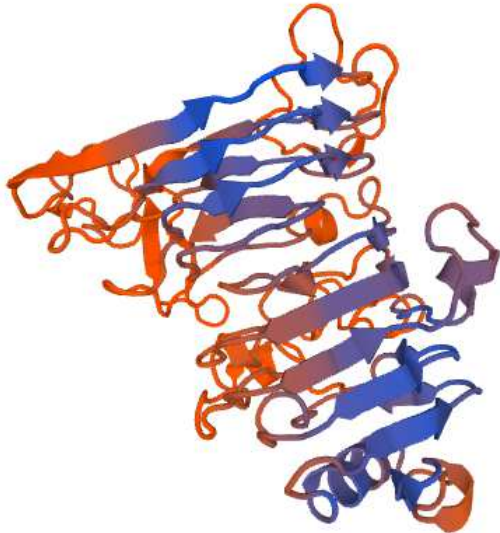
Name	Title	Identity	Method	Oligo State	Ligands
3lmw.1.A	Iota-carrageenase, CgiA	10.56	X-ray, 2.6Å	monomer	1 x <u>NI</u> , 1 x <u>CA</u>
5gai.1.Z	Tail fiber protein	17.82	EM	hetero-oligomer	None



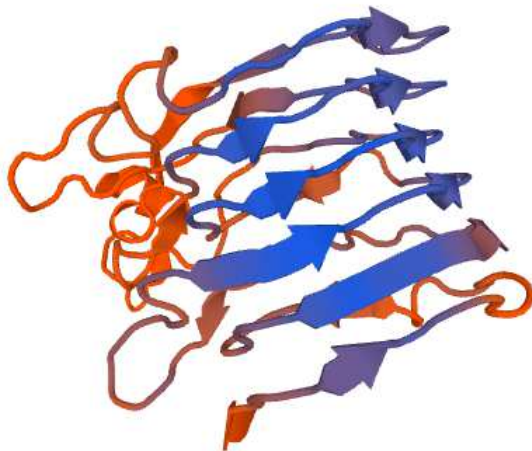
Model #1: Residues 46-455 of MVLG_03747T0 with 4c2l.1.A (37.40% sequence identity) as a template



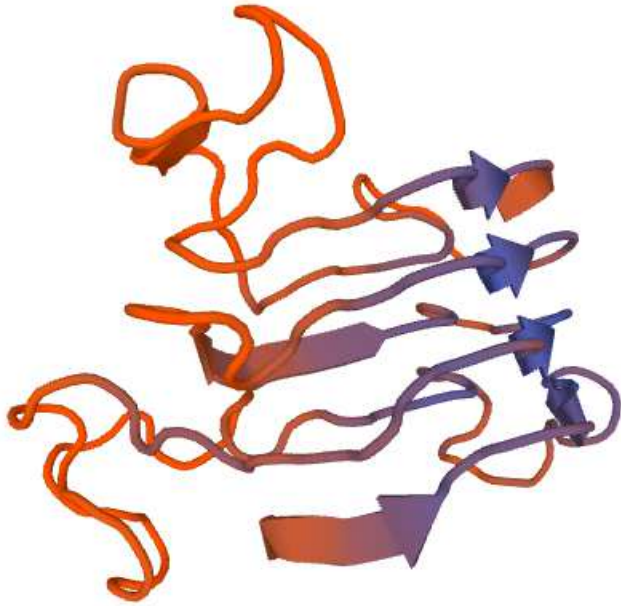
Model #2: Residues 24-401 of MVLG_03747T0 with 4xr6.1.A (17.69% sequence identity) as a template



Model #3: Residues 46-422 of MVLG_03747T0 with 3lmw.1.A (14.05% sequence identity) as a template



Model #4: Residues 199-397 of MVLG_03747T0 with 3lmw.1.A (10.56% sequence identity) as a template

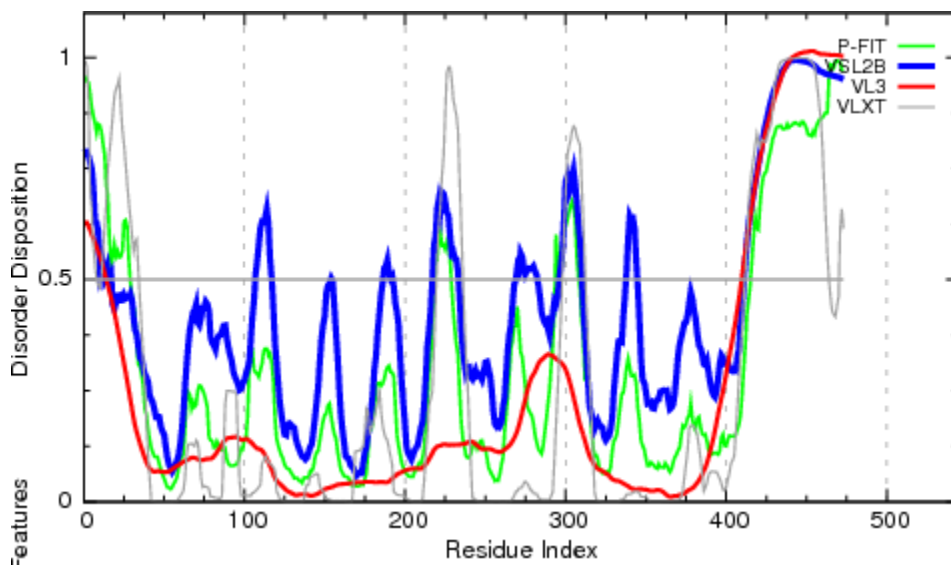


Model #5: Residues 226-362 of MVLG_03747T0 with 5gai.1.Z (17.82% sequence identity) as a template

> MVLG_05737T0 (31.08%)- 473 aa

MMTPLSTLVVAAATLSSLLQVAATQAATTAPMLSSYFPAYTEGATVAWNQTK
LAMYFVDITTKDGF EIGNQPLDGIKKFTSQAYANGAKPMVTLGSWNGSLYFSK
QLSTPEGRTKLASQLQNYLYYKEFKGVDVSWLYPAQQGIGCNTVSPKDTDNFLK
FLKTLRGWLGMGYLISIAAPPGGFLTGNGTEHVVDYSEWATVLDHINVMTYDYT
GPWSSKTGPLSPMHSCASGGGVTA AVKYWTSSGFPAEKIFISIPSYAISFTLKSSTL
EKTYMTDGDGGTFNYSSLIYQSFSSIPKGEAADSNEPTTDGCGVVTANYTGQWH
YTSLIKEGLLAHDGSKGLKGYARYMDGCSQTPFLFNPTNKHFIAYEDAASASIKA
GFARDNGLKGVTVFTSEGFD DTVYDAIVTDLNRPKKELES GGATGKSDTPQAQA
GGKTTKPSTPPPSSKQPQDMPKKASHGAGILGKMNLRAR

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 473

Number residues disordered: 147

Overall percent disordered: 31.08
0.4255

Number Disordered Regions: 11

Longest Disordered Region: 62

Average Prediction Score:

Predicted disorder segment [1]-[12]

Average Strength= 0.6728

Predicted disorder segment [14]-[15]

Average Strength= 0.5140

Predicted disorder segment [17]-[17]

Average Strength= 0.5080

Predicted disorder segment [107]-[117]

Average Strength= 0.5893

Predicted disorder segment [154]-[154]

Average Strength= 0.5021

Predicted disorder segment [187]-[192]

Average Strength= 0.5213

Predicted disorder segment [218]-[233]

Average Strength= 0.6169

Predicted disorder segment [269]-[282]

Average Strength= 0.5254

Predicted disorder segment [296]-[310]

Average Strength= 0.6447

Predicted disorder segment [338]-[344]

Average Strength= 0.5982

Predicted disorder segment [412]-[473]

Average Strength= 0.9104

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	399	411	13
2	455	473	19
Filtered Regions			

	From	To	Length
1	290	290	1
2	393	394	2

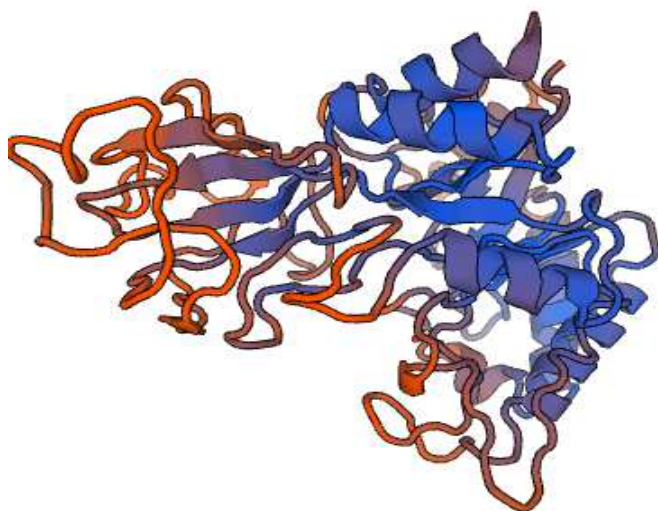
ModPred and PROSITE:

ModPred: Acetylation (K160, K467), Amidation (Y196, S197, V240, Y275, Y370, Y404), Ubiquitination (K220, K467), Proteolytic cleavage (D405, T409, D410, E419), Phosphorylation (T429, T443), Hydroxylation (P446), ADP-ribosylation (R471).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
4txg.1.A	Chitinase	22.56	X-ray, 1.8Å	monomer	11 x <u>CS</u>

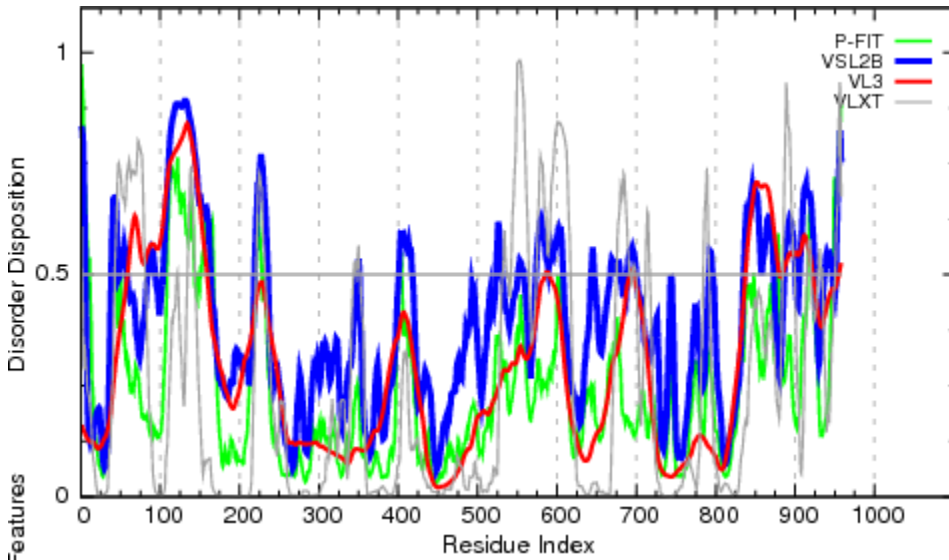


Model #1: Residues 28-413 of MVLG_057377T0 with 4txg.1.A (22.56% sequence identity) as a template

> MVLG_01520T0 (31.00%)- 958 aa

MRVGNSLAVLAAAAAIVPAAFAQKTQVFGRYIIVYPESNKEGKALHARDHLSN
LRAKSGVEPLEVVQEYHMPGVLVGQSVNAPGVTKEQLEKMPGVKAVYPVFDYS
FAAVQQQKEPPSQSFQQHHNHRKVKESGAAPPHLELRDQQFLGQKNMPNTTG
GFSPHRMTSIDILHKRGFFGQGVKSCFIEGHTQYTHPLLKRGCFGSKNCAIQFGA
DLVGTDPNHPQPGDPHADCDVRSTHILGQMVAPENRFDVGAIPQAEIGWYSIF
PCGGGGATGDIIGAFKAADDGCKVISNSLISSVGWNDNDLGPITLNKLAEKGG
VFAVSAWGVSRDEGLFYPAGPATGTEGVGAAYVDLNQYPFAYTLTFENGEATL
PYISVYPIPYDDSFVEVYFLSTSSTDTAATGCDDLPHDTPDLTNRAVVVQRASCGFE
TQMANVRKFGARVLLVVNYPASVGWPAPYFDGIAPSVFPVGMHSDGAKLLE
YYRKNNSGLKLNFKDRTLHPVNADTGGKISFYSSYGPDNSLTTGPTFGV PANQI
AGIRPNGSVGTIDATSSPITNAIATLVLGARKNDNLKPDELRSLLATTAKPISIHPR
ADGELETTTLAGSGLVNALRAVEAQT LVTPFSFKINDTAHVQKEQQLTLKNMG
HASITYTFDSTAAQTKMTYDGGAKQDIVPSSLPTVLQEA EAKVSFDKTSITIEAGQ
TATVKVTITPPQLTAREKDYPVYSGFINIHASNKQEFHVS YFGLAADIVDMPIID
VSTSFASAFRSGLPQGLTYPYLLDNSPNSVKVPTQLTTFDRSIGVGVFIRFAQATR
HVTVDVIAGNSTFKGTLPSHEGRNHRRSLNAADENHLVARRLARQSRADPNQLY
TDVQVLGRIYEKKNQARDDKGPDALVVFKGSMHKDLSMDGQASDLPDGMPY
RVLVRALKTTADPSLEASWESWVSPPVQFKS

PONDR:



STATISTICS=====

Predicted residues: 958	Number Disordered Regions: 21
Number residues disordered: 297	Longest Disordered Region: 60
Overall percent disordered: 31.00	Average Prediction Score:
0.3976	
Predicted disorder segment [1]-[5]	Average Strength= 0.7563
Predicted disorder segment [39]-[47]	Average Strength= 0.6197
Predicted disorder segment [50]-[58]	Average Strength= 0.5426
Predicted disorder segment [85]-[93]	Average Strength= 0.5318
Predicted disorder segment [104]-[163]	Average Strength= 0.7643
Predicted disorder segment [219]-[235]	Average Strength= 0.6649
Predicted disorder segment [349]-[351]	Average Strength= 0.5175
Predicted disorder segment [401]-[417]	Average Strength= 0.5639
Predicted disorder segment [524]-[530]	Average Strength= 0.5664
Predicted disorder segment [551]-[558]	Average Strength= 0.5182
Predicted disorder segment [574]-[607]	Average Strength= 0.5686
Predicted disorder segment [643]-[646]	Average Strength= 0.5319
Predicted disorder segment [668]-[679]	Average Strength= 0.5201
Predicted disorder segment [689]-[699]	Average Strength= 0.5432
Predicted disorder segment [792]-[796]	Average Strength= 0.5282
Predicted disorder segment [835]-[857]	Average Strength= 0.6365
Predicted disorder segment [859]-[873]	Average Strength= 0.5794
Predicted disorder segment [887]-[900]	Average Strength= 0.5821
Predicted disorder segment [906]-[924]	Average Strength= 0.6381
Predicted disorder segment [938]-[946]	Average Strength= 0.5367
Predicted disorder segment [952]-[958]	Average Strength= 0.7087

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	96	110	15
2	140	150	11
3	166	187	22
4	204	214	11
5	565	570	6
6	615	621	7

7	627	634	8
8	875	885	11
9	900	908	9
10	927	933	7
Filtered Regions			
	From	To	Length
1	266	269	4
2	586	588	3
3	812	815	4
4	859	860	2

ModPred and PROSITE:

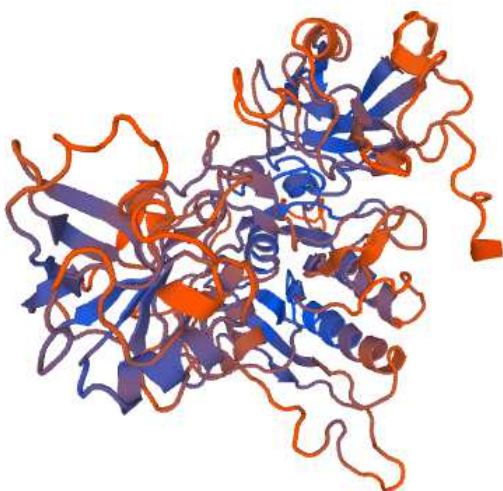
ModPred: Amidation (A12, I259, F392), Proteolytic cleavage(R30, Y31, D216, D254, W332, L340, Y395, R584, Y733, D923, R928), Ubiquitination (K40, K57), ADP-ribosylation (R546), O-linked glycosylation (T716), Phosphorylation (S897).

PROSITE: No identified domain recognition sites.

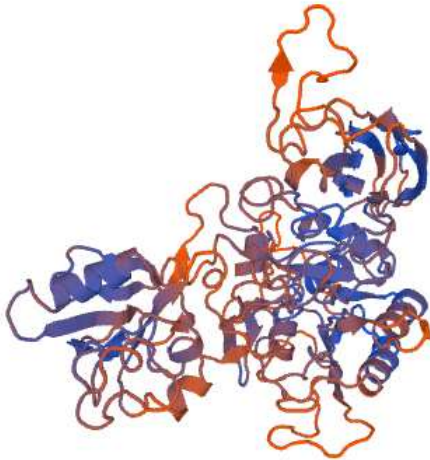
Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
3eif.1.A	C5a peptidase	19.82	X-ray, 1.9Å	monomer	1 x <u>CA</u> , 1 x <u>MLA</u>
4yn3.1.A	Cucumisin	19.24	X-ray, 2.0Å	monomer	1 x <u>MAN</u> , 1 x <u>BMA</u> , 4 x <u>NAG</u> , 1 x <u>FUC</u>

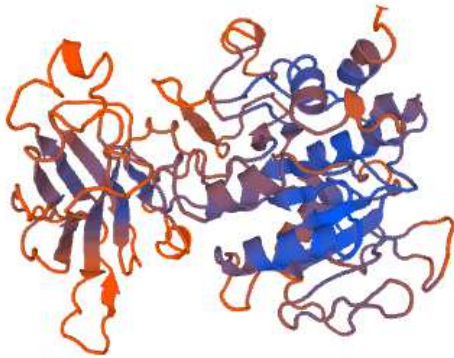
Name	Title	Identity	Method	Oligo State	Ligands
4i0w.1.B	Protease CspB	14.39	X-ray, 1.6Å	hetero-oligomer	None
3lxu.1.A	Tripeptidyl-peptidase 2	15.80	X-ray, 3.1Å	homo-dimer	None
1r6v.1.A	subtilisin-like serine protease	19.59	X-ray, 1.7Å	monomer	1 x <u>CA</u>
1y9z.1.A	alkaline serine protease	20.21	X-ray, 1.4Å	monomer	2 x <u>CA</u> , 1 x <u>PMS</u>
4mzd.1.A	Nisin leader peptide-processing serine protease NisP	18.84	X-ray, 1.1Å	monomer	None



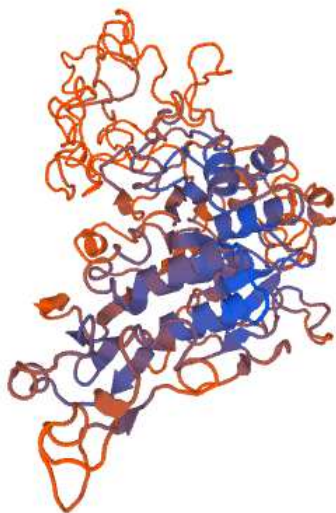
Model #1: Residues 163-766 of MVLG_01520T0 with 3eif.1.A (19.82% sequence identity) as a template



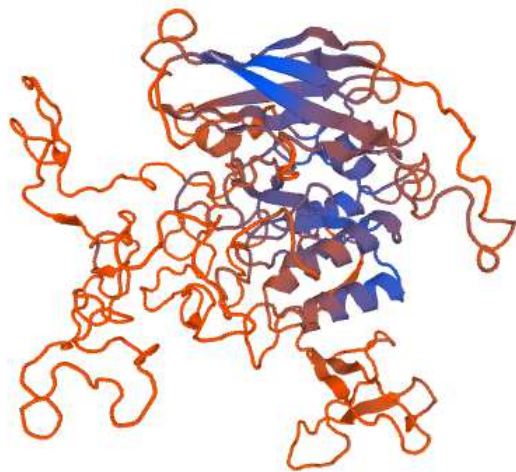
Model #2: Residues 144-754 of MVLG_01520T0 with 4yn3.1.A (19.24% sequence identity) as a template



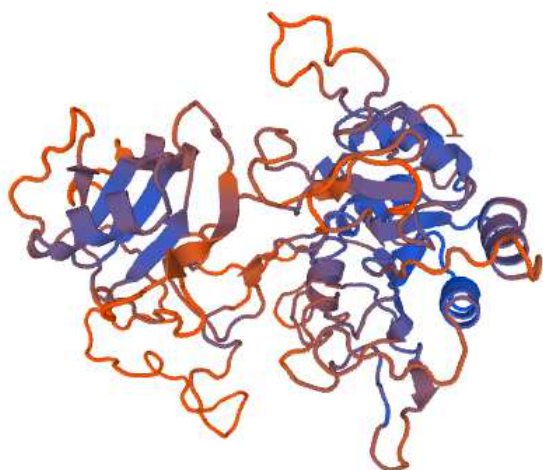
Model #3: Residues 167-629 of MVLG_01520T0 with 4i0w.1.B (14.39% sequence identity) as a template



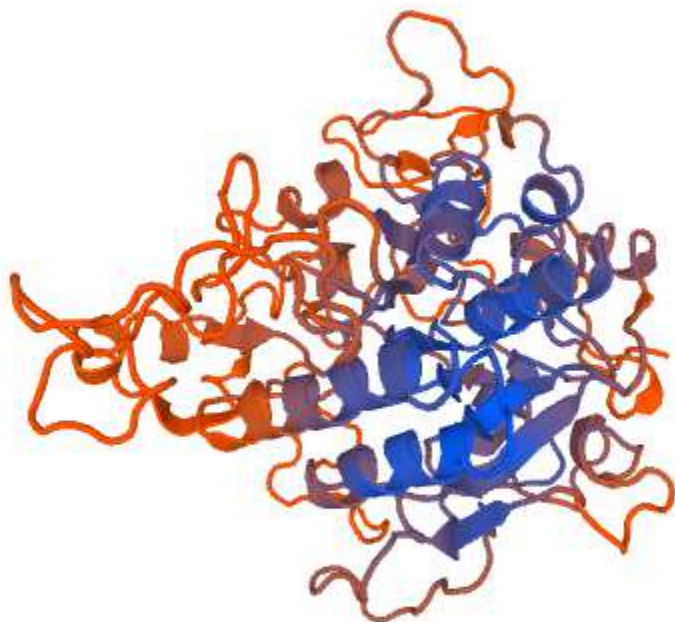
Model #4: Residues 19-626 of MVLG_01520T0 with 1r6v.1.A (19.59% sequence identity) as a template



Model #5: Residues 177-755 of MVLG_01520T0 with 3lxu.1.A (15.80% sequence identity) as a template



Model #6: Residues 168-629 of MVLG_01520T0 with 1y9z.1.A (20.21% sequence identity) as a template

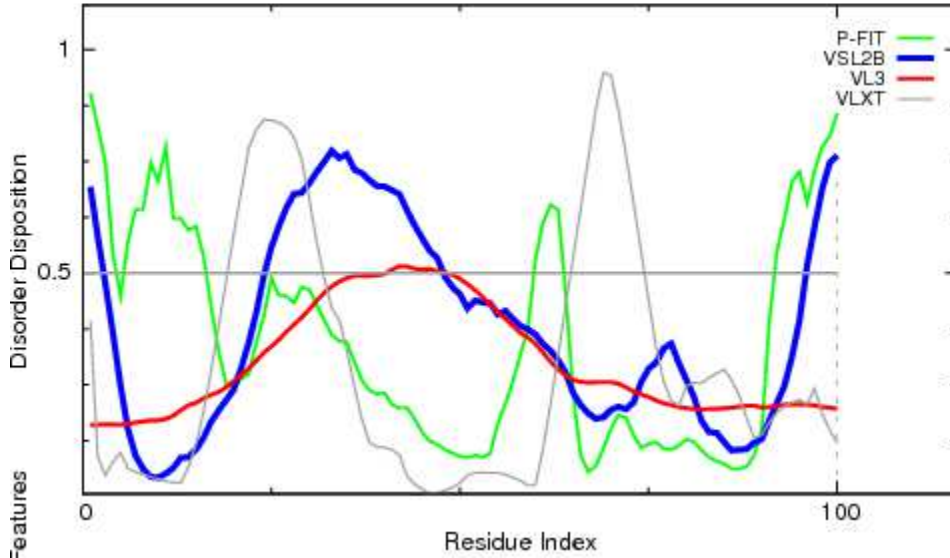


Model #7: Residues 145-627 of MVLG_01520T0 with 4mzd.1.A (18.84% sequence identity) as a template

> MVLG_04107T0 (30.00%)- 100 aa

MKYSLVFVTLVLM AAINVSAIPADLTKPTSTSSEVDKVDHPKKYAPPAVISFISK
ANATVARQTKDCCNYCLKRRRDGVKLNSCYAICLWSSGKWTTKCP

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 100
 Number residues disordered: 30
 Overall percent disordered: 30.00
 0.3768
 Predicted disorder segment [1]-[2]
 Predicted disorder segment [25]-[47]
 Predicted disorder segment [96]-[100]

Number Disordered Regions: 3
 Longest Disordered Region: 23
 Average Prediction Score:
 Average Strength= 0.6372
 Average Strength= 0.6721
 Average Strength= 0.6643

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			

	From	To	Length
1	6	12	7

ModPred and PROSITE:

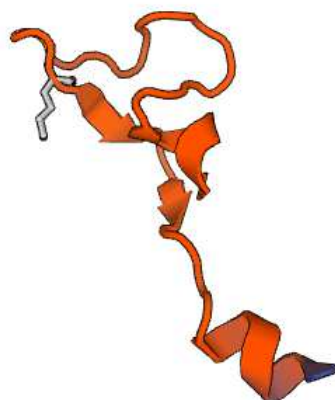
ModPred: Proteolytic cleavage (D36, R76, D77), Disulphide linkage (C67, C71, C99),

Amidation (Y70), Hydroxylation (P100).

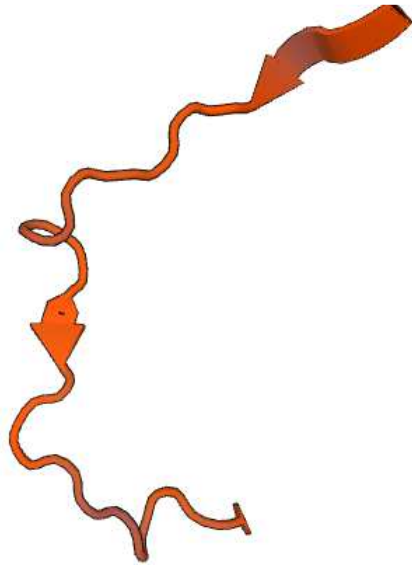
PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
3hwe.1.A	Neutrophil gelatinase-associated lipocalin	15.22	X-ray, 2.8Å	monomer	2 x <u>RKS</u>
5ool.1.2	39S ribosomal protein L36, mitochondrial	16.22	EM	hetero-oligomer	1 x <u>PNS</u> , 12 x <u>MG</u> , 3 x <u>ZN</u>
2yeu.2.B	DR2231	23.53	X-ray, 2.0Å	homo-dimer	2 x <u>GD</u>



Model #1: Residues 16-56 of MVLG_04107T0 with 2yeu.2.B (23.53% sequence identity) as a template



Model #2: Residues 21-46 of MVLG_04107T0 with 3hwe.1.A (15.22% sequence identity) as a template



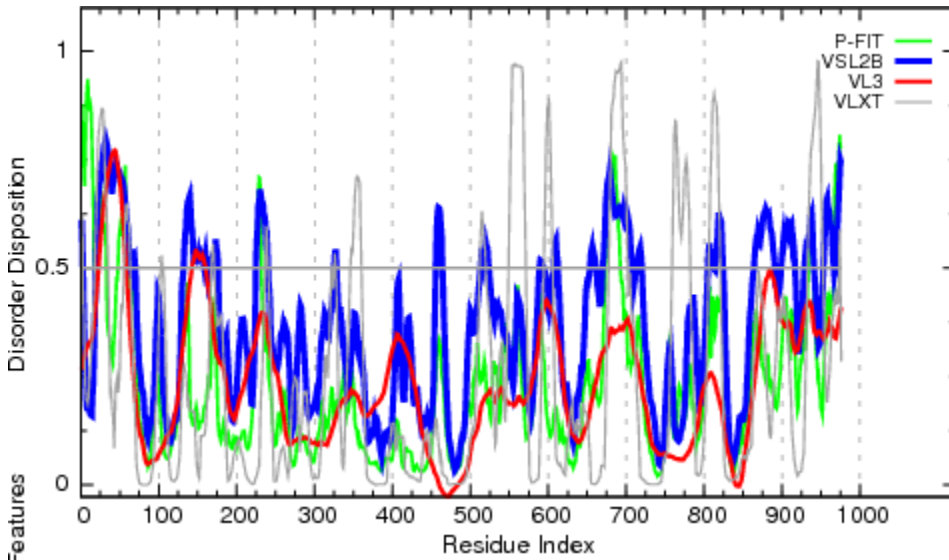
Model #3: Residues 58-80 of MVLG_04107T0 with 5ool.1.2 (16.22% sequence identity) as a template

Moderately disordered proteins (10-30% disordered residues by PONDR® VSL2 analysis)

> MVLG_02763T0 (29.30%)- 976 aa

MLVGKVSLVILWT**ATMALA**SPSRRGNHKSTTGKRSSFSGRYLVKLSTQTDM
 KGSKVPVEQHRQDCNDQIKNLSSDFFGAIKVHNTFNFA YQCSMSVEAMEGFSP
 DLADYLGVTGVDVIKVVVHGRGLSFPMPPPVSKLQPKEGFATPWRKNHHPLLLS
 PSVYRNNSFAPHVESEVHLMHNLGLLGDSNVSVCLVDTGVDYTNRRLLGEGFGK
 GFKIVLGHDFVGNDGKHPGSPYTNCTDHGTHVTGIVGANFDPDFKFSGAAPV
 TLGHYRAFACTGQSTEDTIAAALLRAHADGCKVITLSLGGPSAWEDGLVADAAS
 HVTNQGSLVVSSAGNFGTQGLFYGDVPGELPEVLGTAATDLREYPVGYLLDFVD
 HSFQPIPYFAVYPVKINETLDVCYIPPSITDDPKCNLSTIVLPKGD LKNSLLVLELG
 QCPHSLVAKWAVANKLRVAMVSFKPEDAQSP LNYYSNHFARGIDYFLIVPHSW
 VETLIRYYTASRGKLQVSFAAGKRAPVEALANHESGGNMAFYSSYGPTATLEGF
 GNTLAAPGTNILSTVTVAQGGVGVMSGTSMACPLAAGIAALLFSHRKADNLTPR
 QVKSLMATT AQPVRISQKPKDAFATVVQQGAGIVSAYRAYIAKTLIEPHSIALGD
 LEHFKNSHSITLKNTNKFVITYTLSSTSSQTVTTYDKSASIDINPSGIPRPGIAGAA
 TVAFTPRSLRIPPGQSATFTATFTLPNFSKIDFFRVPVVS GWLLIDSAGDPVPTYRI
 AYAGVAAGLQIMPVLDSTDVASQSYGIKGLRHPFIMVGNSLDPDALPSTAADV L
 SDPNKVTSVSRKDGIFVFLRFAMATPYVQVDLVDANTTFIPTIPSNNNLHLAENN
 SLTGDLRKRPHNPPLFDSVSIVGTAATANELTRDPTDFSHGSGADTSFINFNGMV
 AVKHPDDPATT SVDKGRPYRLLIRARRMNSNPEFSASYDSWLSPPFQFLD

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 976
 Number residues disordered: 286

Number Disordered Regions: 23
 Longest Disordered Region: 44

Overall percent disordered: 29.30
0.3690

Average Prediction Score:

Predicted disorder segment [1]-[2]
 Predicted disorder segment [20]-[63]
 Predicted disorder segment [66]-[69]
 Predicted disorder segment [133]-[162]
 Predicted disorder segment [166]-[170]
 Predicted disorder segment [172]-[176]
 Predicted disorder segment [225]-[240]
 Predicted disorder segment [326]-[328]
 Predicted disorder segment [456]-[464]
 Predicted disorder segment [513]-[523]
 Predicted disorder segment [589]-[591]
 Predicted disorder segment [608]-[612]
 Predicted disorder segment [654]-[660]
 Predicted disorder segment [670]-[707]
 Predicted disorder segment [714]-[720]
 Predicted disorder segment [805]-[812]
 Predicted disorder segment [815]-[824]
 Predicted disorder segment [865]-[891]
 Predicted disorder segment [899]-[916]
 Predicted disorder segment [929]-[941]
 Predicted disorder segment [954]-[966]
 Predicted disorder segment [968]-[968]
 Predicted disorder segment [970]-[976]

Average Strength= 0.5861
 Average Strength= 0.6953
 Average Strength= 0.5239
 Average Strength= 0.5575
 Average Strength= 0.5125
 Average Strength= 0.5355
 Average Strength= 0.6081
 Average Strength= 0.5265
 Average Strength= 0.5943
 Average Strength= 0.5317
 Average Strength= 0.5126
 Average Strength= 0.5391
 Average Strength= 0.5248
 Average Strength= 0.6281
 Average Strength= 0.5272
 Average Strength= 0.5269
 Average Strength= 0.5722
 Average Strength= 0.5827
 Average Strength= 0.5697
 Average Strength= 0.5631
 Average Strength= 0.6000
 Average Strength= 0.5005
 Average Strength= 0.6613

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	889	894	6
2	918	926	9
Filtered Regions			
	From	To	Length
1	1	16	16
2	42	43	2
3	216	219	4

4	832	839	8
5	946	949	4
6	974	976	3

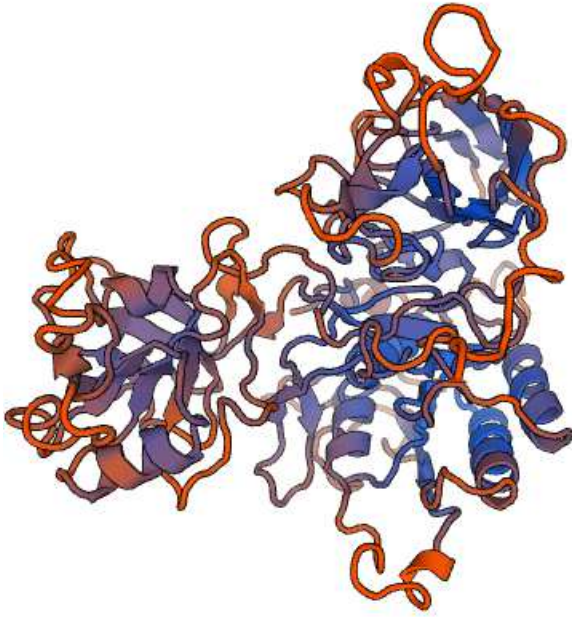
ModPred and PROSITE:

ModPred: Amidation (A19, F131, N392, F652, F709, R953), Proteolytic cleavage (R34, R41, D110, D222, Y527, R715, D781, K828, D888, D965), Acetylation (K213), Ubiquitination (K229, K456), Sumoylation (K259), O-linked glycosylation (S694), ADP-ribosylation (R698)

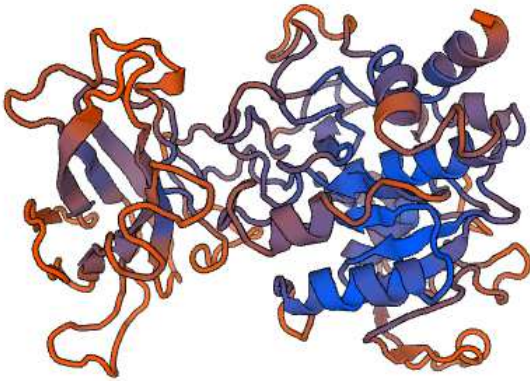
PROSITE: Subtilase_ASP (193-204, PROSITE entry PS00136), Subtilase_HIS (242-252, PROSITE Entry PS00137), Subtilase_SER (566-576, PROSITE PS00138)

Structural modelling:

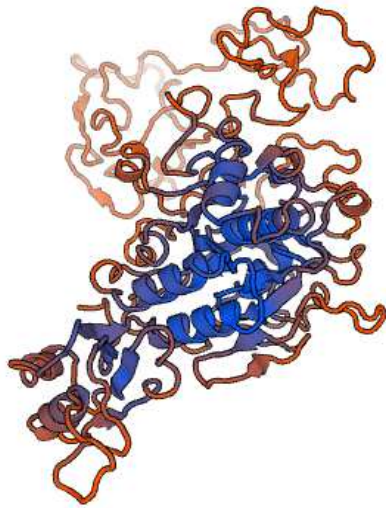
Name	Title	Identity	Method	Oligo State	Ligands
3eif.1.A	C5a peptidase	22.10	X-ray, 1.9Å	monomer	1 x <u>CA</u> , 1 x <u>MLA</u>
4i0w.1.B	Protease CspB	20.42	X-ray, 1.6Å	hetero-oligomer	None
3lxu.1.A	Tripeptidyl-peptidase 2	18.67	X-ray, 3.1Å	homo-dimer	None
1r6v.1.A	subtilisin-like serine protease	22.94	X-ray, 1.7Å	monomer	1 x <u>CA</u>
1y9z.1.A	alkaline serine protease	20.58	X-ray, 1.4Å	monomer	2 x <u>CA</u> , 1 x <u>PMS</u>
4mzd.1.A	Nisin leader peptide-processing serine protease NisP	21.30	X-ray, 1.1Å	monomer	None



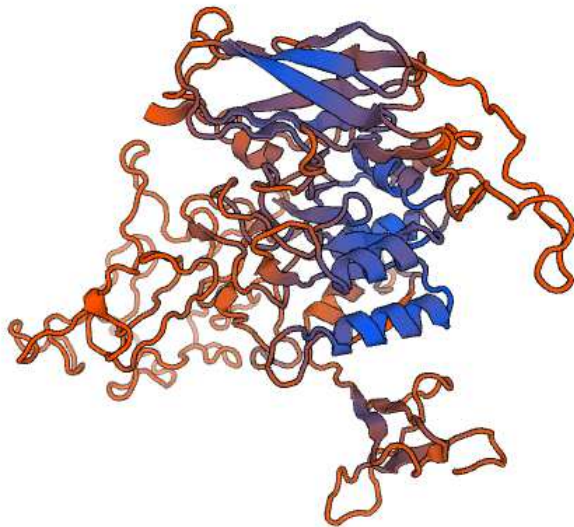
Model #1: Residues 172-780 of MVLG_02763T0 with 3eif.1.A (22.10% sequence identity) as a template



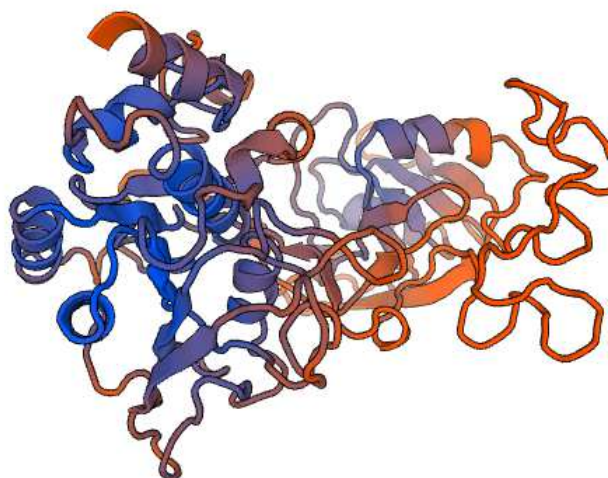
Model #2: Residues 172-639 of MVLG_02763T0 with 4i0w.1.B (20.42% sequence identity) as a template



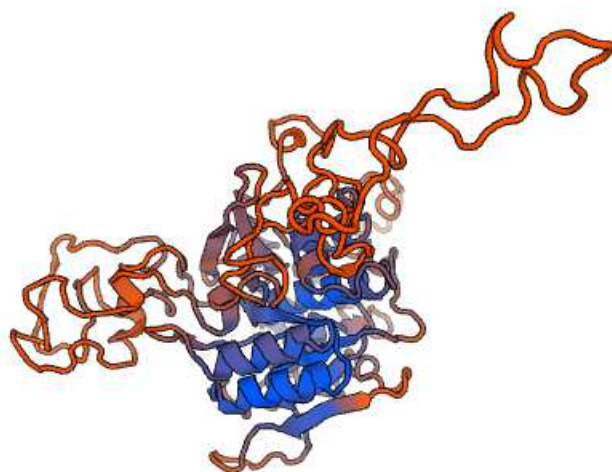
Model #3: Residues 34-637 of MVLG_02763T0 with 1r6v.1.A (22.94% sequence identity) as a template



Model #4: Residues 185-770 of MVLG_02763T0 with 3lxu.1.A (18.67% sequence identity) as a template



Model #5: Residues 177-639 of MVLG_02763T0 with 1y9z.1.A (20.58% sequence identity) as a template

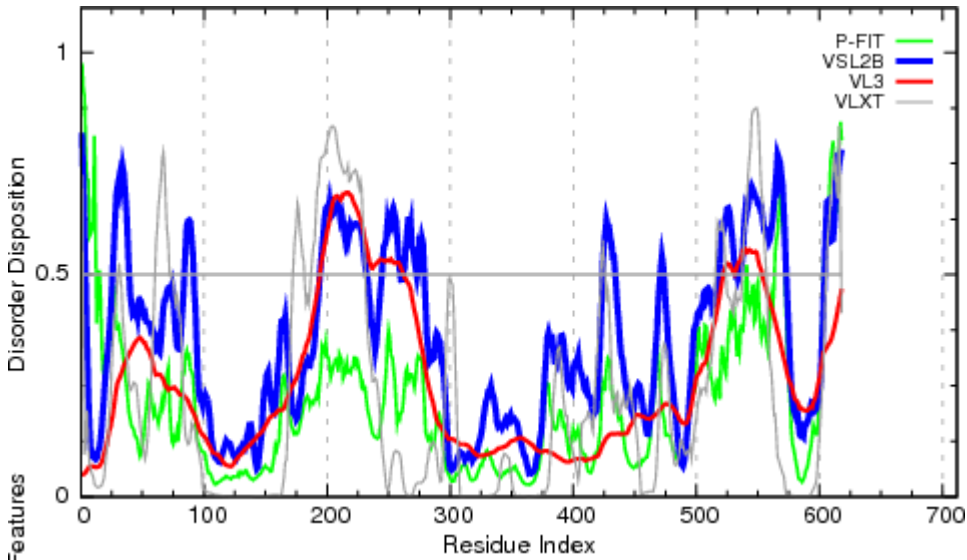


Model #6: Residues 144-638 of MVLG_02763T0 with 4mzd.1.A (21.30% sequence identity) as a template

> MVLG_03994T0 (28.32%)- 618 aa

MRKAFSFFALLYATSWARAQVLRPHPLHDSISSSDQIHPRGLEDYFLDKKRLSA
ALDLLDEEHIMWSGDETPAGDFPVATMHFETHSKEKVLNMQRFSRLISGVSCGV
GRIYITFKSRVAFDYAAQAWDWVHLHPEHVFTLLAHWRDCMNPDGHFIPFHFK
KAVPEASTLTITLEGIEVAWEEAGHTFTLHVGSGLREGEELQNTATKELEVPSHFT
EVAHPLPAPNAGPLLEERFHIGHLNFKLPDPSVSRSKGYSVHLDHKYNGEMVSK
HHLGQNGYEATSHCINCGSSGRIDISFRLRIKWFDIKEMGIYATAFNVGARLQWD
LSLKANTIASLDFGGNIFEFPLPGLGLEIHKIFKLGLIASVGVGIGCRNYTGHEMS
HGIQFRIQDGAEAHIDLVKGIGGNHWRPQVWSAPLHIEGKVKANPAASAGSTV
GFEMELFKTTLAAGLRISAPSAVFLVKLNDANKGPCGQRLHRRSIQVDAILLAYL
GMSGGLNGAFAGSEPIGKRGLIDGGVHTSKNNQTRRERLLDMSWMEPNFTEEEF
REFLESNDVDVSAHAHLNSSKHS�HPRRLGLFANLPIYHHSWPLIKELCIPIGPH
VPLQKRSHPRDLLIGP

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 618

Number residues disordered: 175

Overall percent disordered: 28.32

0.3560

Predicted disorder segment [1]-[3]

Predicted disorder segment [26]-[39]

Predicted disorder segment [85]-[91]

Predicted disorder segment [195]-[232]

Predicted disorder segment [245]-[259]

Predicted disorder segment [262]-[272]

Predicted disorder segment [274]-[279]

Predicted disorder segment [424]-[434]

Number Disordered Regions: 12

Longest Disordered Region: 38

Average Prediction Score:

Average Strength= 0.7133

Average Strength= 0.6514

Average Strength= 0.5943

Average Strength= 0.6078

Average Strength= 0.5904

Average Strength= 0.5630

Average Strength= 0.5489

Average Strength= 0.5695

Predicted disorder segment [471]-[473]
 Predicted disorder segment [519]-[533]
 Predicted disorder segment [536]-[573]
 Predicted disorder segment [605]-[618]

Average Strength= 0.5258
 Average Strength= 0.5848
 Average Strength= 0.6425
 Average Strength= 0.6718

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			
	From	To	Length
1	236	240	5
2	486	492	7
3	579	579	1

ModPred and PROSITE:

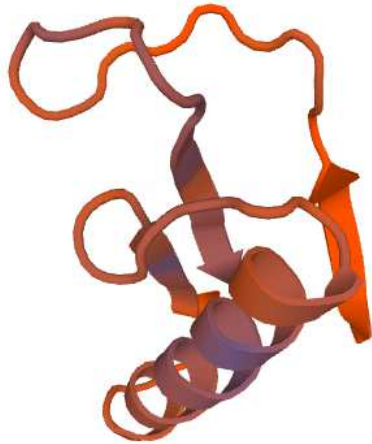
ModPred: Sumoylation (K3), Amidation (A4, L353, Y376, N407, Y491, Y585),

Proteolytic cleavage (D29, R530, R608), Ubiquitination (K263, K271, K426, K568).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
1p1g.1.A	PROTEIN (MACROPHAGE MIGRATION INHIBITORY FACTOR)	10.29	X-ray, 2.5Å	homo-trimer	None
3iq2.1.A	Sorting nexin-7	28.00	X-ray, 1.7Å	monomer	1 x <u>GOL</u> , 6 x <u>SO4</u>



Model #1: Residues 53-121 of MVLG_03994T0 with 1p1g.1.A (10.29% sequence identity) as a template



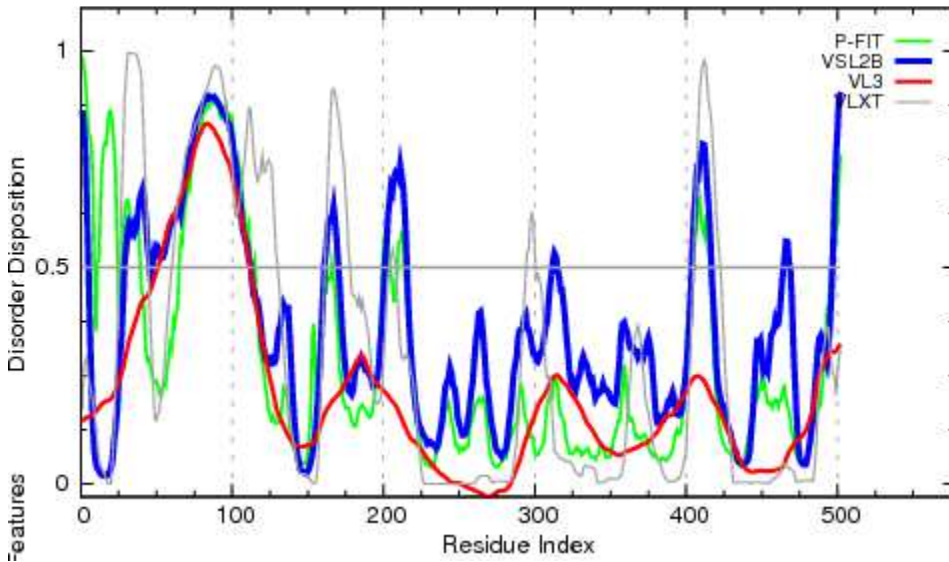
Model #2: Residues 536-556 of MVLG_03994T0 with 3iq2.1.A (28.00% sequence identity) as a template

>MVLG_01159T0 (27.49%)- 502 aa

MHGNYLRLLLSILICQVIAKWKYKYDTGDHRIPRRHRGVDLRGRHHASGVLG
GDFSSNIETANQTIQIDELRQTDTKGENEDLQVSDQEGNGSSFDES DKASNGRIII
SGPPRYVNPAVREKYLRLIDMRSGFTLDEEPFRVVGINIIYWLCNDENVLGVKPG
TPTQKRRIREALAAAVAMGANTVRVGSGLGYADALQPDQHHRAAPRSPAM
DIHDYAIYAAGRYGLKIILPLMDNYDYHGGKYTVLKWLGISAEHNGANFFTD
RAIAFFKSYIEFVLNRKNPYTMRTYGEDPVVSIIEDGNEFGAYKKGSEGYPLAFTD

EIAAQVKKFAPQALFMDGTDGFFNLTAHLQAPGLRSKAVDIVTDHPYPRDIPLLQ
 MQAFLARISGKAFILGEMDWVPSAPSRNPPSRLVEPSLSAYLNVLD RYPNIGVLA
 WSLFVHTDDCRDWVRHHDGYEMYYPLPQDTAEKQANVLT LVQWFYARTGRE
 VPSSLPYQTCPQEEF

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 502	Number Disordered Regions: 9
Number residues disordered: 138	Longest Disordered Region: 65
Overall percent disordered: 27.49	Average Prediction Score:
0.3579	
Predicted disorder segment [1]-[5]	Average Strength= 0.7244
Predicted disorder segment [29]-[45]	Average Strength= 0.5999
Predicted disorder segment [48]-[112]	Average Strength= 0.7189
Predicted disorder segment [161]-[170]	Average Strength= 0.5798
Predicted disorder segment [201]-[215]	Average Strength= 0.6589
Predicted disorder segment [312]-[314]	Average Strength= 0.5251
Predicted disorder segment [404]-[416]	Average Strength= 0.6889
Predicted disorder segment [465]-[468]	Average Strength= 0.5394
Predicted disorder segment [497]-[502]	Average Strength= 0.7601

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length

None			
Filtered Regions			
	From	To	Length
1	9	23	15
2	55	55	1
3	107	111	5
4	144	153	10

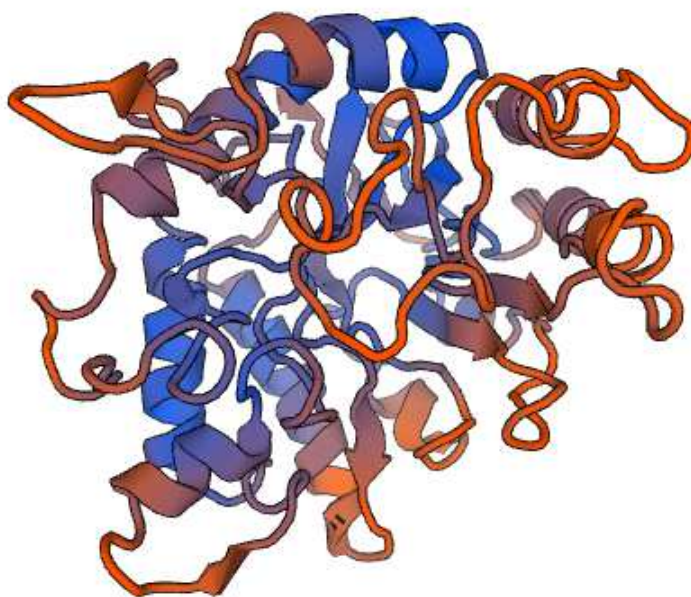
ModPred and PROSITE:

ModPred: Amidation (A8, V55, L723, Y752, D770, D955, Y1022), Proteolytic cleavage (R62, K160, R283, R286, D322, K326, Y329, S330, D407, R562, R565, D568, T897, P931), Carboxylation (E99, E106, E107, E109, E113, E117, E118, E121, E122), O-linked glycosylation (T169, T170, T171T178, T180, T182, T186, S190, T192, T196, T197, T198, T199, T200, T201, T202, T203, T204, T205, T206), Hydroxylation (P189), Nlinked glycosylation (N397), Phosphorylation (T692), Disulphide linkage (C1025).

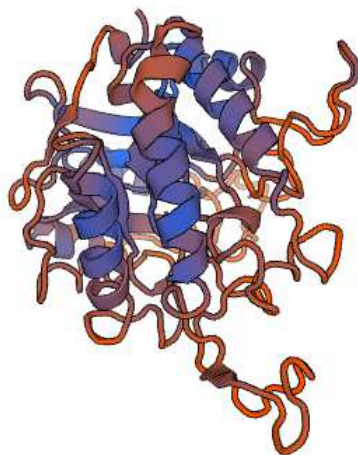
PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
1uuq.1.A	MANNOSYL- OLIGOSACCHARIDE GLUCOSIDASE	16.91	X-ray, 1.5Å	monomer	None
4xzw.1.A	endo-glucanase chimera C10	16.92	X-ray, 1.5Å	monomer	2 x <u>O4B</u> , 1 x <u>CA</u>



Model #1: Residues 122-485 of MVLG_01159T0 with 1uuq.1.A (16.91% sequence identity) as a template

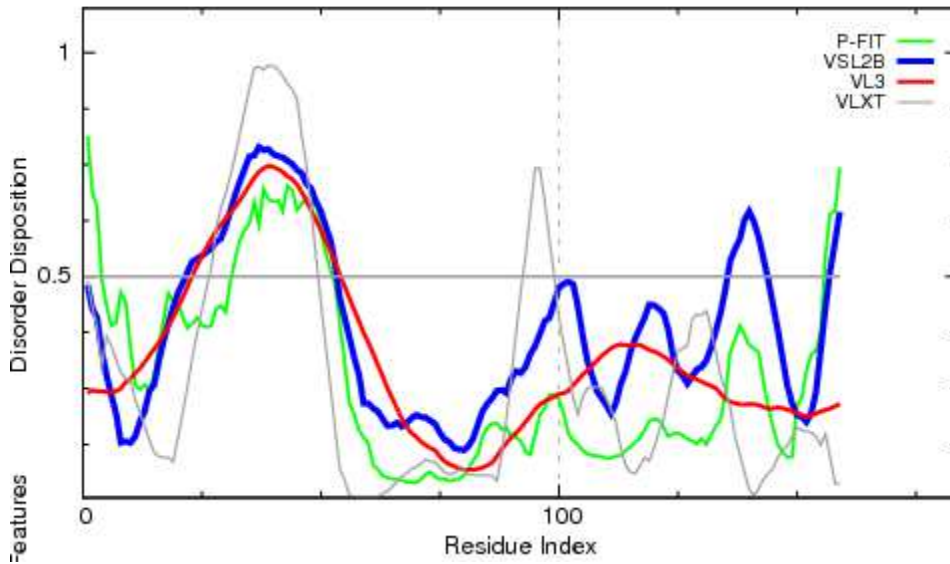


Model #2: Residues 116-445 of MVLG_01159T0 with 4xzw.1.A (16.92% sequence identity) as a template

>MVLG_07010T0 (27.04%)- 159 aa

MLFKLPVVLAMALLTLGASASERFTVTSLRRRDKPGDYPQDRGSVKSPAEGQQ
LKVGTLFPFRNPISVGDLDVTLDEVEVFLKIKSLNYSRRLVTNLMSPGGNKPIVQN
FIVMHPKGSIVKRGTIMPGTIEVFEEQQNGTKANGNGKYFLNQAVGVTFQF

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 159

Number residues disordered: 43

Overall percent disordered: 27.04

0.3847

Predicted disorder segment [22]-[53]

Predicted disorder segment [136]-[143]

Predicted disorder segment [157]-[159]

Number Disordered Regions: 3

Longest Disordered Region: 32

Average Prediction Score:

Average Strength= 0.6694

Average Strength= 0.5874

Average Strength= 0.5780

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length

None

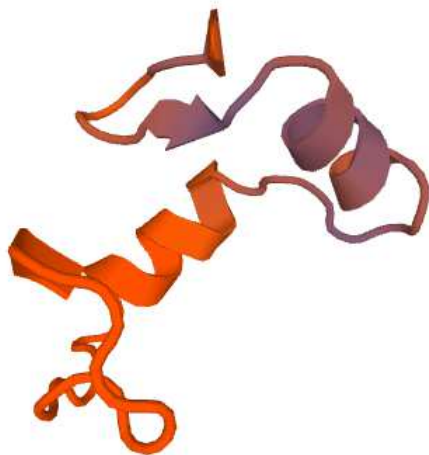
ModPred and PROSITE:

ModPred: Proteolytic cleavage (R23, T27, S28, R30, R32, R42, K46), ADP-ribosylation (R122), Acetylation (K146), Amidation (Y147).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
5i4q.1.A	Contact-dependent inhibitor A	22.45	X-ray, 2.3Å	hetero-oligomer	2 x <u>SO4</u> , 2 x <u>CL</u>



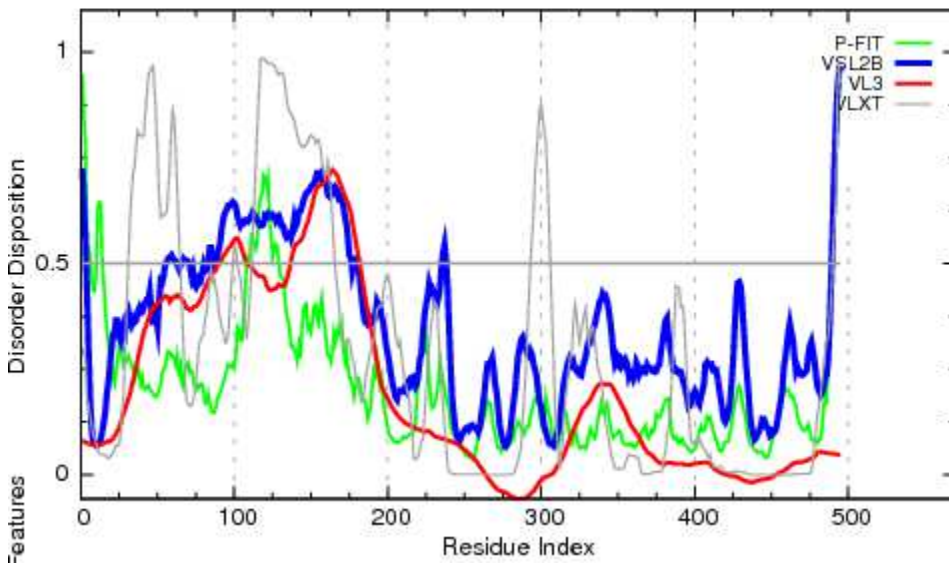
Model #1: Residues 81-129 of MVLG_07010T0 with 5i4q.1.A (22.45% sequence identity) as a template

> MVLG_01192T0 (23.84%)- 495 aa

MLRSYLLLVSILVHDAASWYEAASRLDDTVQPWNRSCPRRVRAQAELLRGV
WRDSNGDRISFDLRKRNLQDFDFVEPAPDFLGLPPLGTAVSSARSARNFAVRPR
ATGGQGTDYLTAPINSSASAVDTIQPEASIIIDGPPASFQSGSYLEVDPSTGLT

LDGEPFRPVGPRCLCNDESLSCLPRGYTDKSRIREALAAAVAMGANTIRINSCGIS
 TGFPQAVQPSLHTYGTDEQLDIHDYVIYAAGEYGLKVILPLTDNYDYYHGGKYT
 FLRWLNVPDNDAGAQQFTDRQVRRAFKRYIKFLLTRVNQYNGLAYGEDPTIAIIE
 DGNEFGAYMGKEGFPPLSFTEDIAKYVKSLAPQALLMDGTDGFYNYTTKAVAP
 GVTSPYVDIVTDHAYPRNIALLRQVDIAHSNGKVFLIGEMDWTNNGGADFGA
 YLNLLYNYKSVGMAWSLFTHTPCSSYVIHDDAYSIIYPNGGQHLTLILLQSS
 GAEGF

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 495
 Number residues disordered: 118
 Overall percent disordered: 23.84
 0.3486
 Predicted disorder segment [1]-[3]
 Predicted disorder segment [56]-[62]
 Predicted disorder segment [66]-[70]
 Predicted disorder segment [83]-[86]
 Predicted disorder segment [89]-[176]
 Predicted disorder segment [179]-[180]
 Predicted disorder segment [236]-[238]
 Predicted disorder segment [490]-[495]

Number Disordered Regions: 8
 Longest Disordered Region: 88
 Average Prediction Score:
 Average Strength= 0.6444
 Average Strength= 0.5107
 Average Strength= 0.5060
 Average Strength= 0.5228
 Average Strength= 0.6229
 Average Strength= 0.5117
 Average Strength= 0.5279
 Average Strength= 0.8299

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			
	From	To	Length
1	136	140	5

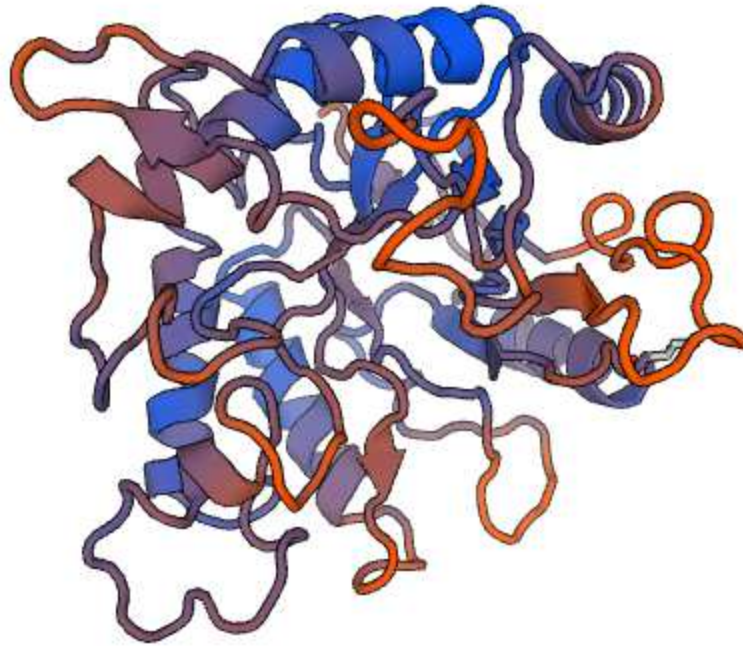
ModPred and PROSITE:

ModPred: Amidation (A16, M450, Y471, Y475, Q488), Pyrrolidone carboxylic acid (Q45), Proteolytic cleavage (R97, W452), Methylation (K256, K301).

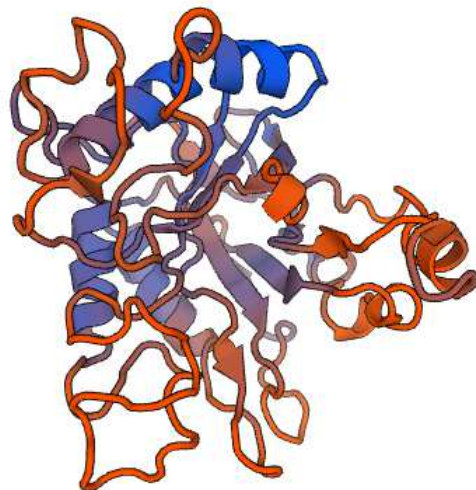
PROSITE: No identified domain recognition sites.

Structural modelling:

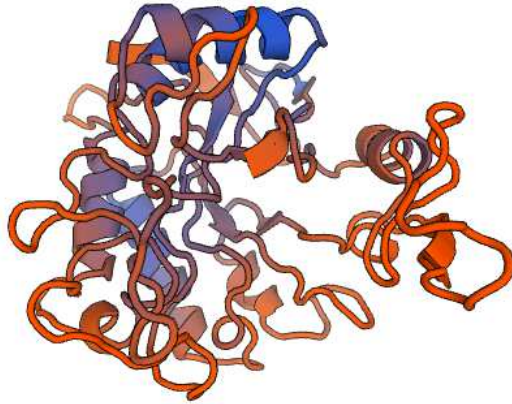
Name	Title	Identity	Method	Oligo State	Ligands
1rh9.1.A	endo-beta-mannanase	24.92	X-ray, 1.5Å	monomer	None
2zun.1.A	458aa long hypothetical endo-1,4-beta-glucanase	17.99	X-ray, 2.0Å	monomer	2 x <u>CBI</u>
4cu6.1.A	BETA-GALACTOSIDASE	19.29	X-ray, 2.7Å	monomer	None



Model #1: Residues 153-491 of MVLG_01192T0 with 1rh9.1.A (24.92% sequence identity) as a template



Model #2: Residues 151-457 of MVLG_01192T0 with 2zun.1.A (17.99% sequence identity) as a template

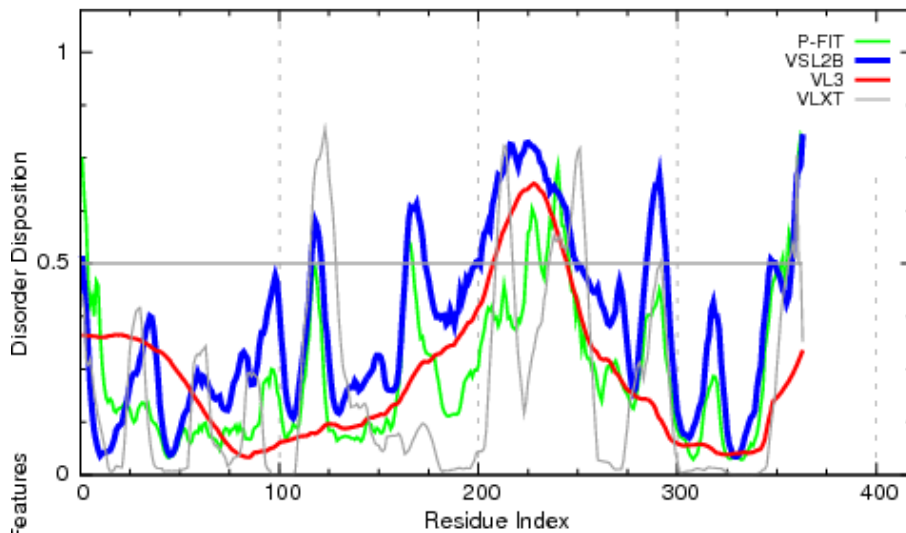


Model #3: Residues 152-457 of MVLG_01192T0 with 4cu6.1.A (19.29% sequence identity) as a template

> MVLG_05108T0 (23.69%)- 363 aa

MLCKLSLVLILASSFWVALATPPAACTIVSSTDIPKVQKCKVITIMAFIMPAGQT
 LMLDVQAGTTINQLGDIIFEHRGPWRGPLMSIYGDSITYNGNNKKLYCNGQMYW
 DGMGVTGTTKPGPALLITGTVSDLIHNSPLNAVVEANGKTLLSNIFVNNTD
 GDRMGHNTDGFNVVQKTRDLTISGCTVINQDDCISITSGQGITISQNTCKNGHGI
 SIGSIKSNEHVSQVTISQNHVENSQQGYRIKTYSGATRGSVDNITFHGNTGNGLTH
 YGVVVEQDYTESGPKGPSFATNGVLISNIRFVGPITTLSMAGDKAQKVYVLCGV
 NSCIGDWDWSSLKFTRGGSLGSITRAPIRGLTA

PONDR:



STATISTICS=====

Predicted residues: 363	Number Disordered Regions: 7
Number residues disordered: 86	Longest Disordered Region: 48
Overall percent disordered: 23.69	Average Prediction Score:
0.3541	
Predicted disorder segment [116]-[121]	Average Strength= 0.5558
Predicted disorder segment [164]-[173]	Average Strength= 0.5941
Predicted disorder segment [199]-[200]	Average Strength= 0.5054
Predicted disorder segment [202]-[249]	Average Strength= 0.6874
Predicted disorder segment [285]-[294]	Average Strength= 0.6248
Predicted disorder segment [347]-[349]	Average Strength= 0.5044
Predicted disorder segment [358]-[363]	Average Strength= 0.6849

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			
	From	To	Length
1	327	329	3

ModPred and PROSITE:

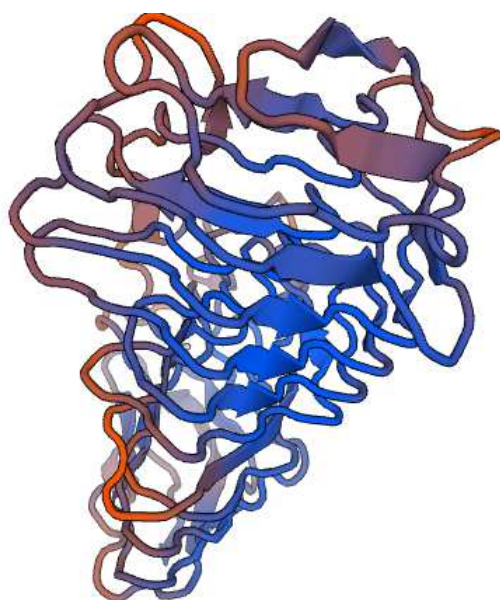
ModPred: Disulphide linkage (C103), Amidation (Y248, Y325).

PROSITE: No identified domain recognition sites.

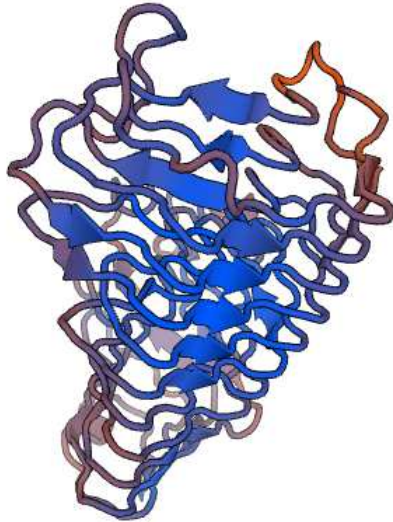
Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
1czf.1.A	POLYGALACTURONASE II	40.48	X-ray, 1.7Å	monomer	3 x <u>ZN</u> , 1 x <u>NAG</u>
1k5c.1.A	ENDOPOLYGALACTURONAS E	41.72	X-ray, 1.0Å	monomer	2 x <u>NAG</u>

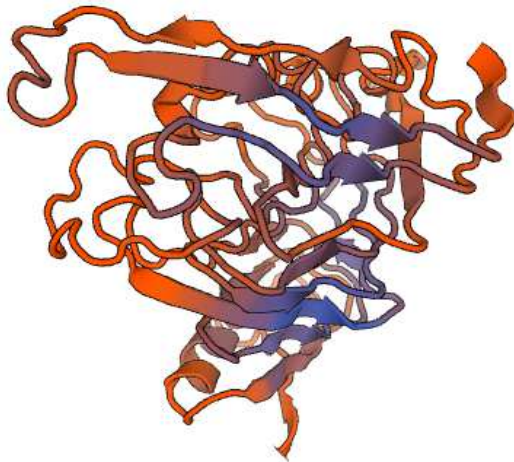
Name	Title	Identity	Method	Oligo State	Ligands
3lmw.1.A	Iota-carrageenase, CgiA	10.14	X-ray, 2.6Å	monomer	1 x <u>NI</u> , 1 x <u>CA</u>
4xqi.1.A	Tail spike protein	13.82	X-ray, 1.8Å	homo-trimer	3 x <u>GLC</u> , 3 x <u>GLA</u> , 3 x <u>RAM</u> , 3 x <u>NAG</u> , 3 x <u>NDG</u>
5gai.1.Y	Tail fiber protein	17.27	EM	hetero-oligomer	None
4xor.1.A	Tail spike protein	12.07	X-ray, 1.5Å	homo-trimer	3 x <u>GLC</u> , 3 x <u>GLA</u> , 3 x <u>RAM</u> , 6 x <u>NAG</u> , 3 x <u>NDG</u>



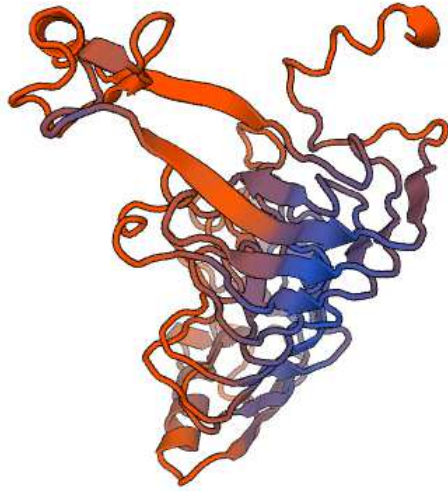
Model #1: Residues 26-360 of MVLG_05108T0 with 1k5c.1.A (41.72% sequence identity) as a template



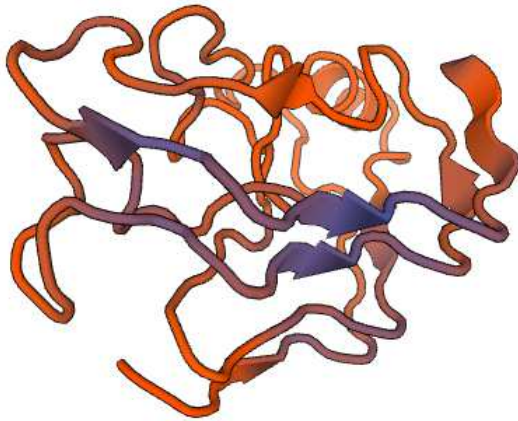
Model #2: Residues 26-347 of MVLG_05108T0 with [1czf.1.A](#) (40.48% sequence identity) as a template



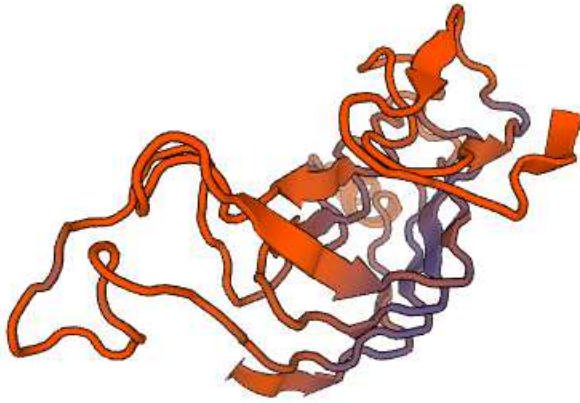
Model #3: Residues 13-314 of MVLG_05108T0 with 4xqi.1.A (13.82% sequence identity) as a template



Model #4: Residues 14-330 of MVLG_05108T0 with 3lmw.1.A (10.14% sequence identity) as a template



Model #5: Residues 142-274 of MVLG_05108T0 with 4xor.1.A (12.07% sequence identity) as a template

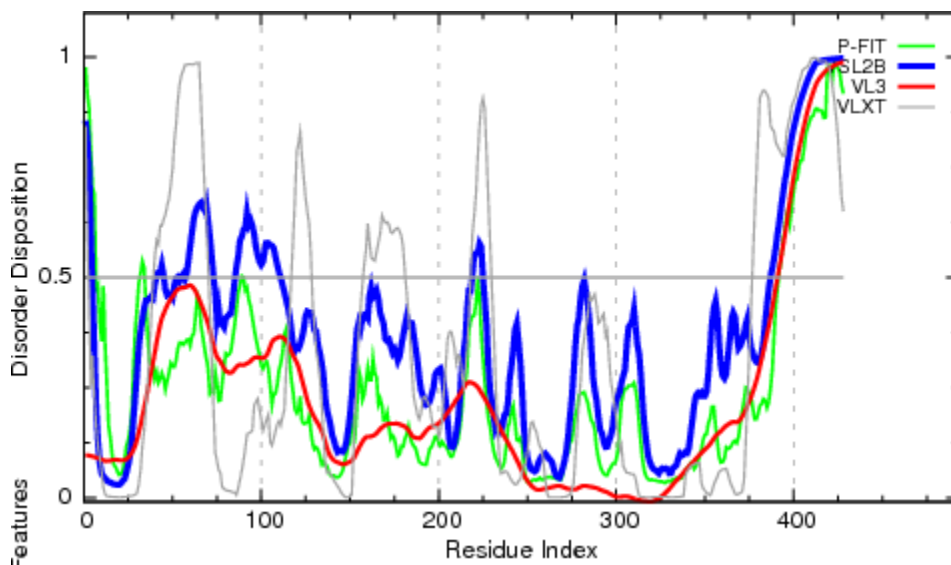


Model #6: Residues 151-313 of MVLG_05108T0 with 5gai.1.Y (17.27% sequence identity) as a template

> MVLG_04797T0 (23.60%)- 428 aa

MPKAAIFNVATVRVLVTFVTLVVFVSPLAAADAVDHPKLRVKGGLGRAGAPA
EVIASAAASVTGSNSKSYPRGYSAVIAFGASYMDNAHKRSKKYATSFRDQQEYP
FSDRGRYTNGPVAVEYMKPSTNPALRPFQIDPPVLFDFAYGGSVIKNNLTGTAG
PHNIPDLGREIKQYLEQLDDEIIDPGRGRVLHVIHTGTNPISQMWLHALTANITHA
KTRRSIGKQVTQMAKYIRYLATHDSL RDNVVAADYLIVGLPPLGIVPNLYFNIA
AFPNHTAAQRDAALEYAGELVDL FNVELEAFTSSLKAYVKPGSRILYYDLANLF
KTIYRFPRIYGITAPVTQACWSSSTRVLCKDPEHHL YIDTLHPTTSAHKIWASRMN
RLV NKVARQADAKTLETTVEDDPTTDDHPSSKTDDEPSSSPGDLRC

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 428
 Number residues disordered: 101
 Overall percent disordered: 23.60
 0.3703

Number Disordered Regions: 8
 Longest Disordered Region: 42
 Average Prediction Score:

- Predicted disorder segment [1]-[4]
- Predicted disorder segment [41]-[45]
- Predicted disorder segment [53]-[53]
- Predicted disorder segment [56]-[56]
- Predicted disorder segment [58]-[73]
- Predicted disorder segment [86]-[110]
- Predicted disorder segment [219]-[225]
- Predicted disorder segment [387]-[428]

- Average Strength= 0.7574
- Average Strength= 0.5111
- Average Strength= 0.5008
- Average Strength= 0.5064
- Average Strength= 0.6098
- Average Strength= 0.5725
- Average Strength= 0.5442
- Average Strength= 0.8670

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	140	150	11
2	383	389	7
Filtered Regions			
	From	To	Length

1	424	428	5
---	-----	-----	---

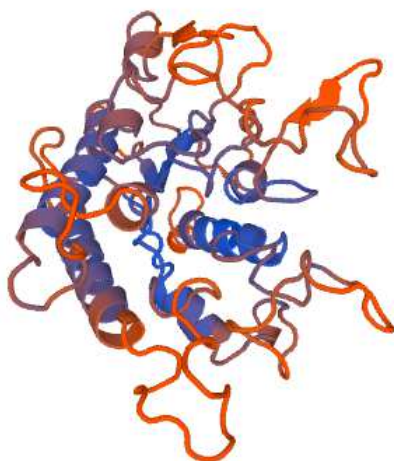
ModPred and PROSITE:

ModPred: Ubiquitination (K3), Amidation (A4, S26, A373, A389, K414), Methylation (K68), Proteolytic cleavage (R100), Phosphorylation (S422).

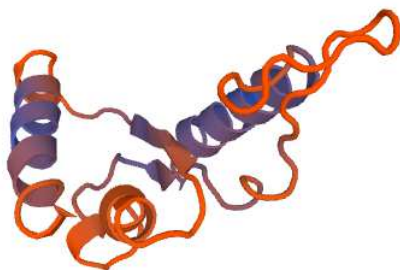
PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
3kvn.1.A	Esterase estA	21.58	X-ray, 2.5Å	monomer	None
1zmb.1.A	Acetylxylan esterase related enzyme	9.52	X-ray, 2.6Å	homo-dimer	None



Model #1: Residues 71-399 of MVLG_04797T0 with 3kvn.1.A (21.58% sequence identity) as a template

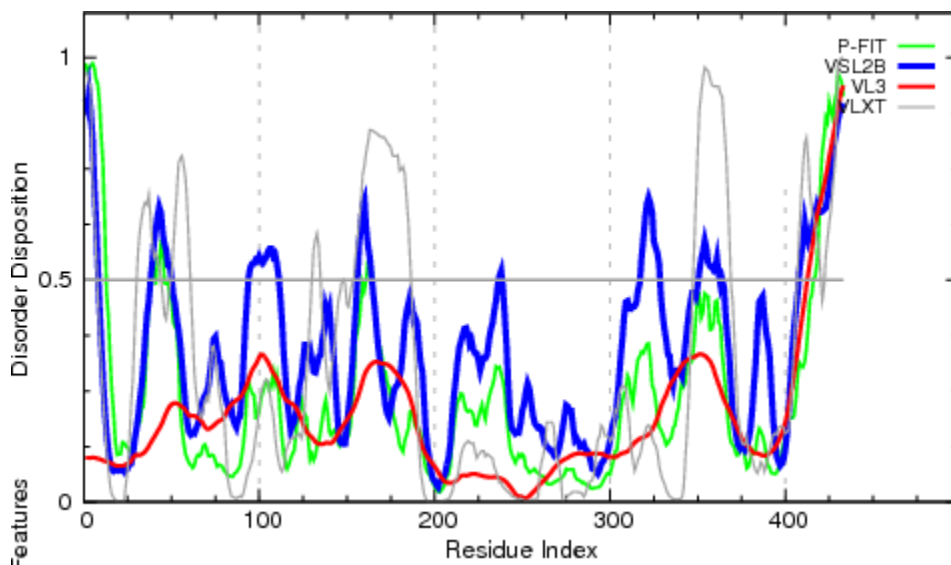


Model #2: Residues 293-389 of MVLG_04797T0 with 1zmb.1.A (9.52% sequence identity) as a template

> MVLG_03707T0 (22.86%)- 433 aa

MLSHSTSRFLRSWGAFLCLPWNPLALVLAATIPTSHSLLTNSGLDSRTLKGV
RERLDLVARDVWVSGTQTEAYLELDQPQLTVFNPYVFNPFKSSRDRVILESSF
PNSSNRIVLNWLERLGPDDEQFAVIKGGGAAGDPASLGYAWMIAQATTTDEGTQE
RLERMIEAEVEWLEKVPRTMDGAISHRKEVTQLWSDFIYMVPPFLAARGIATSN
HSLLESYRQIKLYRSHLQDTSTHLWRHVRYGTWEDPSLWATGNAWAAAGITR
VLATLTNSFHTAMYWEEIRDALWANEIVEAGFARVKKDGLLPNHLDDPYDFSD
SASSALLASTVFRLHQLGMIRKPSTTLAKAEKIRSKINDKIDPKTGWLRGCVNPLS
WYQRDQSPAQAFVILLEAAWRDSRMISERNSIASKEKVFGGQGTHTRRRDR

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 433	Number Disordered Regions: 9
Number residues disordered: 99	Longest Disordered Region: 26
Overall percent disordered: 22.86	Average Prediction Score:
0.3405	
Predicted disorder segment [1]-[8]	Average Strength= 0.7855
Predicted disorder segment [39]-[50]	Average Strength= 0.5846
Predicted disorder segment [94]-[111]	Average Strength= 0.5440
Predicted disorder segment [156]-[164]	Average Strength= 0.5907
Predicted disorder segment [237]-[238]	Average Strength= 0.5152
Predicted disorder segment [318]-[328]	Average Strength= 0.6108
Predicted disorder segment [351]-[362]	Average Strength= 0.5418
Predicted disorder segment [364]-[364]	Average Strength= 0.5008
Predicted disorder segment [408]-[433]	Average Strength= 0.6937

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	393	401	9
Filtered Regions			
	From	To	Length
1	148	148	1
2	375	377	3

ModPred and PROSITE:

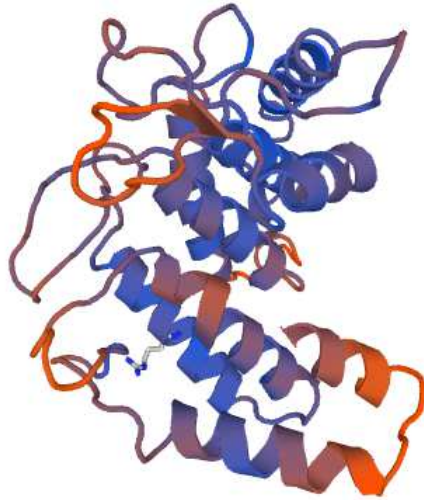
ModPred: Proteolytic cleavage (R9, R12, D321, R431, R433), Sumoylation (K178),

Sulfation (Y320).

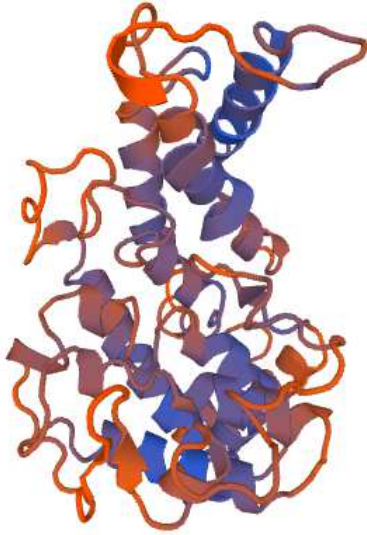
PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
4mmi.1.A	Heparinase III protein	10.34	X-ray, 2.4Å	monomer	2 x <u>CA</u>
3qwt.1.A	Putative GH105 family protein	24.18	X-ray, 2.2Å	homo-octamer	None



Model #1: Residues 150-402 of MVLG_03707T0 with 3qwt.1.A (24.18% sequence identity) as a template

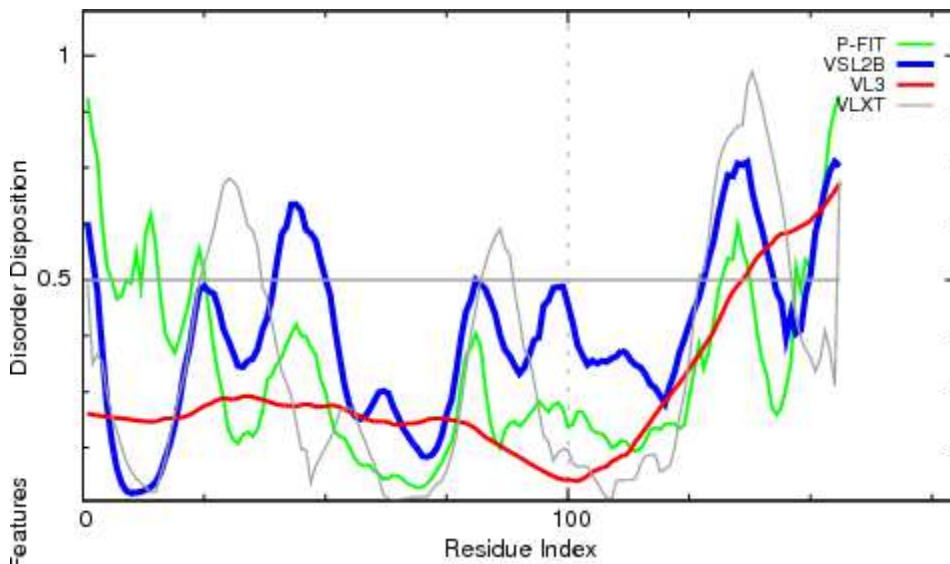


Model #2: Residues 45-378 of MVLG_03707T0 with 4mmi.1.A (10.34% sequence identity) as a template

> MVLG_01732T0 (22.44%)- 156 aa

MLLKLITLIVALLVLNVSALQEAGDTKAEFRLIKRAAAQKSNLTQPTE~~NASFL~~
 LHHPIPFELNYDPKIVYAVDVELLSEHDKSIPLAVQMAGDHTGAISTTFSTPYFAE
 NSVKYRNVTLRVTEWSLPSHNTAPKKKSS~~TIDRKIVCRNFSGKIHA~~

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 156	Number Disordered Regions: 4
Number residues disordered: 35	Longest Disordered Region: 15
Overall percent disordered: 22.44	Average Prediction Score:
0.3764	
Predicted disorder segment [1]-[2]	Average Strength= 0.5863
Predicted disorder segment [40]-[50]	Average Strength= 0.5999
Predicted disorder segment [128]-[142]	Average Strength= 0.6503
Predicted disorder segment [150]-[156]	Average Strength= 0.6757

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	145	153	9

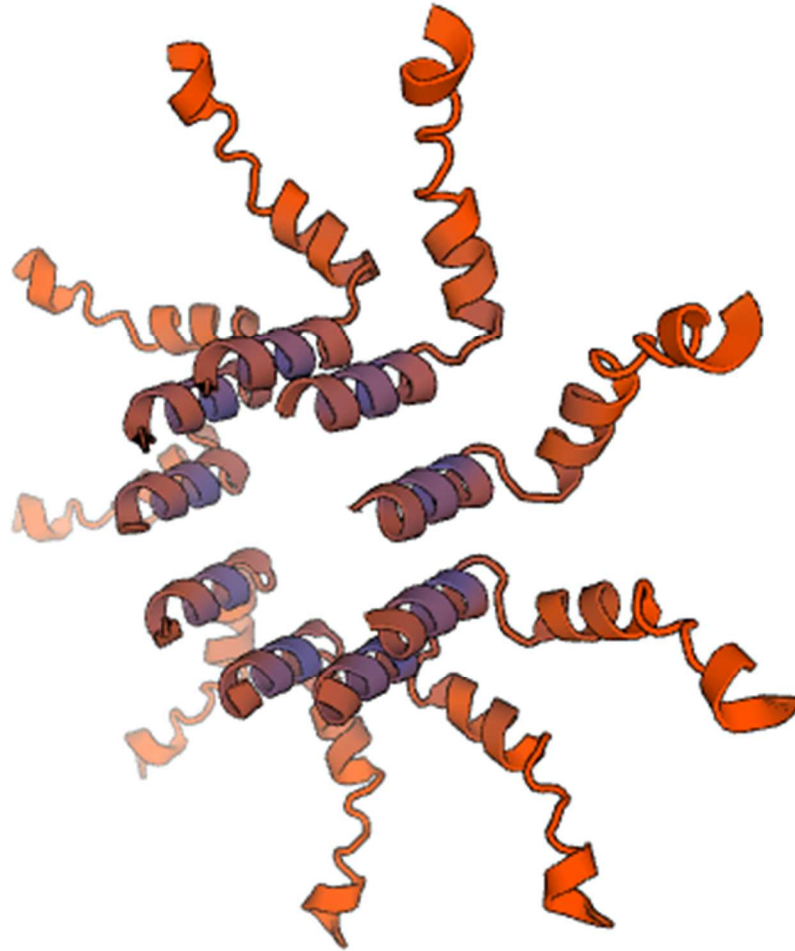
ModPred and PROSITE:

ModPred: Amidation (Y115).

PROSITE: No identified domain recognition sites.

Structural modelling:

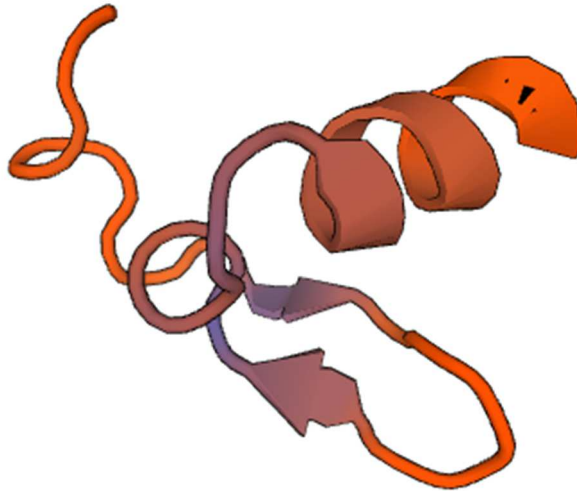
Name	Title	Identity	Method	Oligo State	Ligands
2lzt.1.A	Sec-independent protein translocase protein TatA	20.59	NMR	monomer	None
5cwb.1.A	Designed helical repeat protein	28.21	X-ray, 1.5Å	monomer	None
5bwd.1.A	benzylsuccinate synthase alpha chain	17.65	X-ray, 2.0Å	hetero-oligomer	1 x FUM
2jn8.1.A	Putative cytoplasmic protein	39.29	NMR	monomer	None



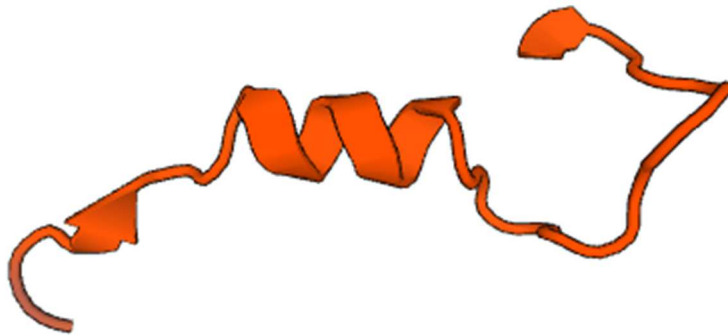
Model #1: Residues 4-37 of MVLG_01732T0 with 2lzs.1.E (20.59 % sequence identity) as a template



Model #2: Residues 18-51 of MVLG_01732T0 with 5cwb.1.A (28.21 % sequence identity) as a template



Model #3: Residues 110-143 of MVLG_01732T0 with 5bwd.1.A (17.65 % sequence identity) as a template



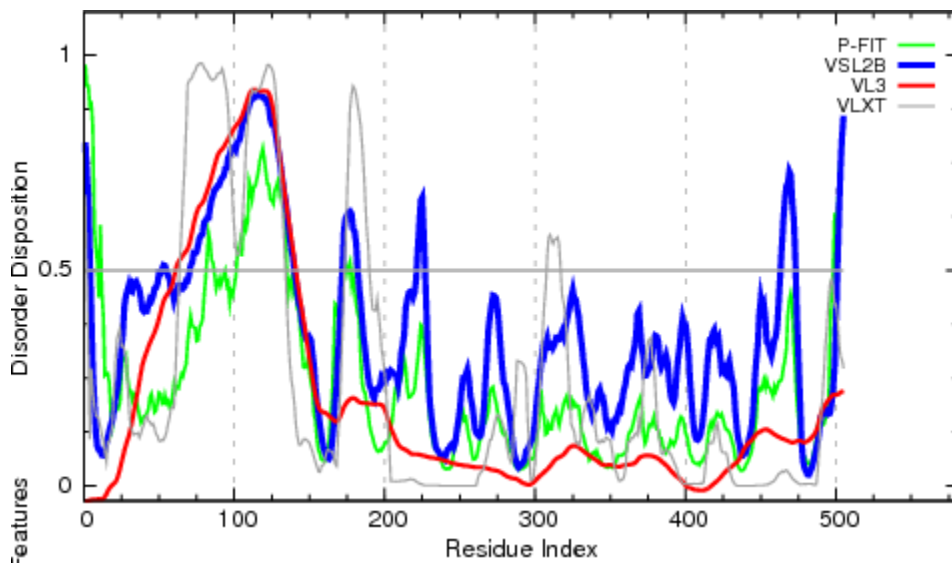
Model #4: Residues 81-110 of MVLG_01732T0 with 2jn8.1.A (36.67 % sequence identity) as a template

> MVLG_01191T0 (21.19%)- 505 aa

MRYSRQLWFGLGLLSQLVTAVTPEVYNNGDCTSSRRFRGLAHTNALMRHSAD
TFDAPRGVPLLALSAPNQMSKIVGLGKRSPAVPAGVSVSVRANVEVGGQSLQP
CADD CDHPQPPPSEEPIVISGPGTLPRSDSYLTLDDARTGLLLDDEPFRPVGINIYW
LCNDENIEGRPKGYPTDKTRVREALAAAVAMGANTVRIGSCGTSLGFHDAIQPD
LHHYADDDGMDIHDIWAAGRYDLKVILTLTDNYDYYHGGKYTILRWLGEPT
DDAGARFFADERPIQVYLRYAKWVLGRVNRYNNAIYGEDPTVSIHETGNELGAY
MGKEGYPPLNWTDRVAQRIKQLAPLALVMDGTDGIYNWSTKATAPGLLSPHIDI
VTDHPYPRDINLFRTQAQLAKSANKVFLLGEMNWLPTGATNANLSDYLEVLDK

YPSVGVLVWSLFTHDSQCSEYVLHNSYSIYYPDGPNTPEEKQNIWSLVQWFYR
 VTDRAVPAVLPVQACPQEVF

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 505	Number Disordered Regions: 7
Number residues disordered: 107	Longest Disordered Region: 69
Overall percent disordered: 21.19	Average Prediction Score:
0.3548	
Predicted disorder segment [1]-[4]	Average Strength= 0.6816
Predicted disorder segment [52]-[54]	Average Strength= 0.5082
Predicted disorder segment [71]-[139]	Average Strength= 0.7406
Predicted disorder segment [172]-[182]	Average Strength= 0.5923
Predicted disorder segment [222]-[226]	Average Strength= 0.6133
Predicted disorder segment [464]-[473]	Average Strength= 0.6476
Predicted disorder segment [501]-[505]	Average Strength= 0.7122

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	88	94	7
2	103	108	6
3	136	144	9

4	157	166	10
Filtered Regions			
	From	To	Length
1	61	65	5
2	146	148	3

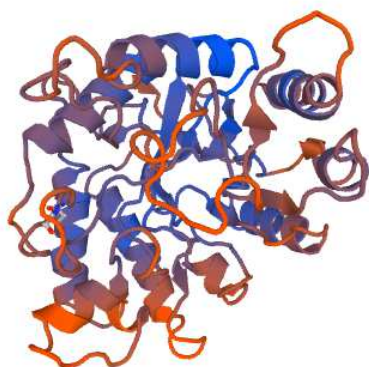
ModPred and PROSITE:

ModPred: Amidation (A20, P82, D337, P492), Proteolytic cleavage (R36, R38, R58, R276), ADP-ribosylation (R80, R133), O-linked glycosylation (S126).

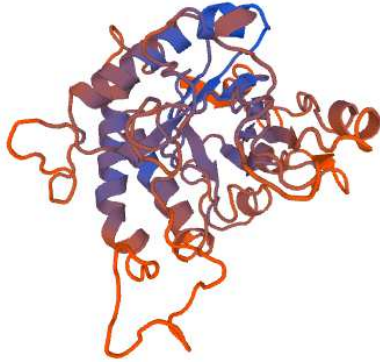
PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
1uuq.1.A	MANNOSYL- OLIGOSACCHARIDE GLUCOSIDASE	19.59	X-ray, 1.5Å	monomer	None
5byw.1.A	Endoglucanase H	28.21	X-ray, 1.5Å	monomer	None



Model #1: Residues 135-488 of MVLG_01191T0 with 1uuq.1.A (19.59 % sequence identity) as a template

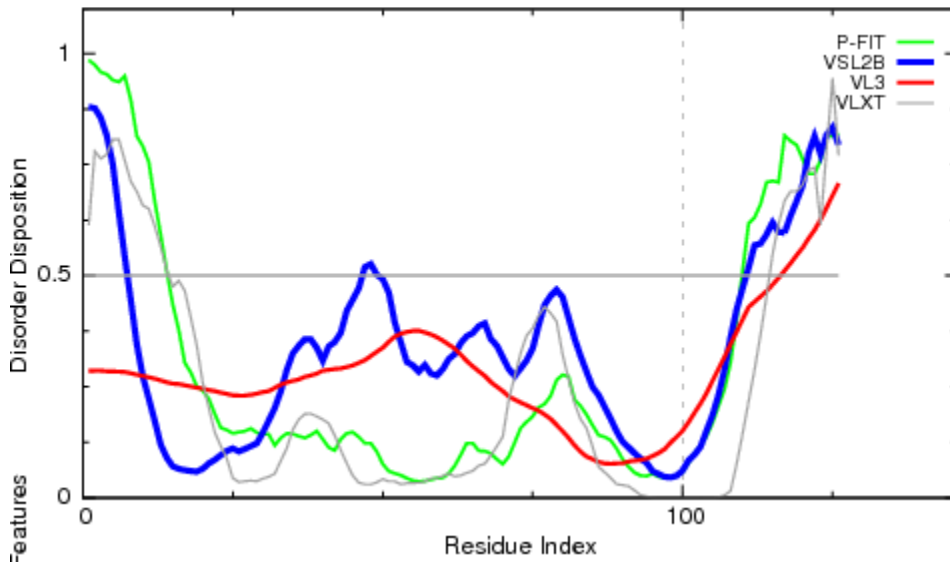


Model #2: Residues 137-467 of MVLG_01191T0 with 5byw.1.A (15.33 % sequence identity) as a template

> MVLG_05525T0 (20.63%)- 126 aa

MCRSSNMARPGMLILALITTVAPIAIALTHVENACAKEAVLHNDLSDGAKCKS
ITDLGCVCSESTGDAFLRSLGDYVKDGRGRCREQYFANIQAYACAYCLFKDLVP
PQSCGKASVSTVPTESDVQ

PONDR:



=====**PONDR VSL2**=====
STATISTICS=====

Predicted residues: 126
 Number residues disordered: 26
 Overall percent disordered: 20.63
 0.3398
 Predicted disorder segment [1]-[7]
 Predicted disorder segment [47]-[49]
 Predicted disorder segment [111]-[126]

Number Disordered Regions: 3
 Longest Disordered Region: 16
 Average Prediction Score:
 Average Strength= 0.7662
 Average Strength= 0.5157
 Average Strength= 0.6798

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			

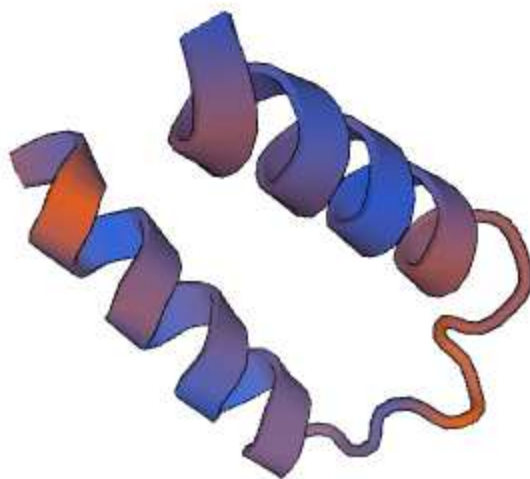
ModPred and PROSITE:

ModPred: GPI anchor amidation (N6).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
5hiu.1.A	GTPase activator-like protein	10.53	X-ray, 2.5Å	monomer	None

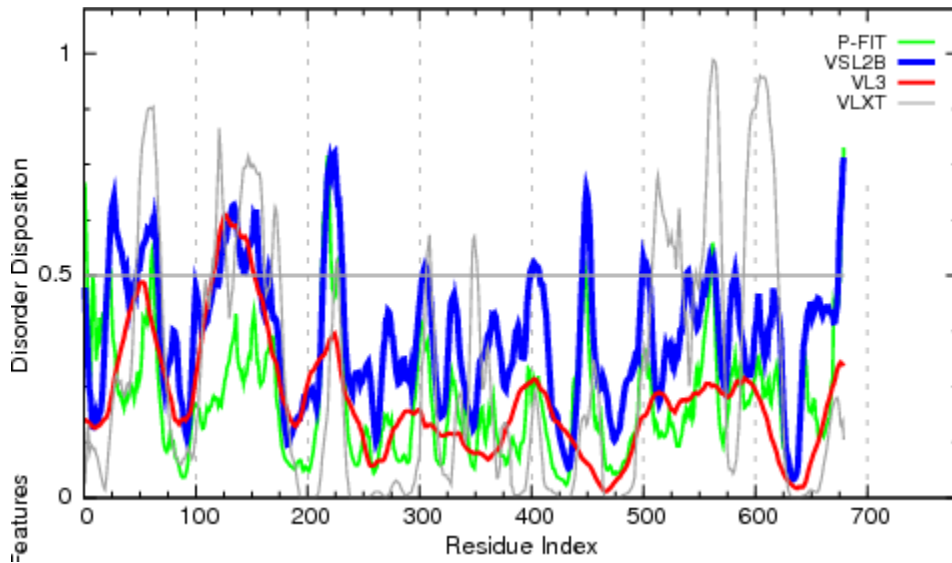


Model #1: Residues 64-101 of MVLG_05525T0 with 5hiu.1.A (10.53 % sequence identity) as a template

> MVLG_02331T0 (19.73%)- 679 aa

MWACFSASLVTVTCAALASAHDPFSRSTIRHSRRYSLLPDTYILETECPDASILA
EASPRTSGANMDSLVTFLAAFP DARPRHVYDSALFCGLSVELGADSPHTQLLRI
QRLKSVSPVRSIQLGVRTAGSSDTPFQAEASLNVVPRTEARGESDYLSHVMLEV
DKMHMGLFGSRETLACVIDSGIDLMHPLLGNCGFGSNCKVVTGYDFVGGDDGH
SPKRSPQTSCSDHGTHIAGILAADKFKAFGFSGVAPNASLGVYRVFSCCKGAASSD
TFLKAMLMAADDGCRVLSLSFGKALGWDQDDGDDPFRKVVSR LATRGVFIAAA
SGNDASQGLMFAQTPADLAGILAVGSVEPVAAPRGFKLSFEHNRYP SMTYLALR
PVNHSQTFQIH FHSIRRAKDTSCDPLLPRSSNFTNSVVVLQKGACGTKLIHFFVRH
GARVVIAHDNGDPEQAQNWRRTAYAVHSQEGLEWLLKWPTS AVHTLLDHYLD
SLGDLQVNF RSKDPVPQDELIDRVAGGLVSEYTEFGPAATLDTLAAHVSAPGSSI
LSTFPLNKGGYGVASGTSMATPMAAGVATLLISHRKDDHLTPAQIRSLMITTAGP
VATKLNSVHPLTTVMQQGGGLVSAH RAYHAMTLIWPYALALYDTPRHVKDHV
VTLTNTHKSVVTYSFN SVPSQTLAMY NKV

PONDR:



=====PONDR VSL2
STATISTICS=====

Predicted residues: 679	Number Disordered Regions: 13
Number residues disordered: 134	Longest Disordered Region: 36
Overall percent disordered: 19.73	Average Prediction Score:
0.3656	
Predicted disorder segment [22]-[38]	Average Strength= 0.5801
Predicted disorder segment [48]-[66]	Average Strength= 0.5721
Predicted disorder segment [122]-[157]	Average Strength= 0.5821
Predicted disorder segment [165]-[165]	Average Strength= 0.5088
Predicted disorder segment [214]-[231]	Average Strength= 0.6758
Predicted disorder segment [303]-[307]	Average Strength= 0.5114
Predicted disorder segment [399]-[407]	Average Strength= 0.5154
Predicted disorder segment [446]-[453]	Average Strength= 0.6192
Predicted disorder segment [499]-[504]	Average Strength= 0.5225
Predicted disorder segment [539]-[540]	Average Strength= 0.5076
Predicted disorder segment [558]-[564]	Average Strength= 0.5195
Predicted disorder segment [582]-[582]	Average Strength= 0.5082
Predicted disorder segment [675]-[679]	Average Strength= 0.6825

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length

1	88	95	8
Filtered Regions			
	From	To	Length
1	429	432	4

ModPred and PROSITE:

ModPred: Proteolytic cleavage (Y35, Y210, D211, R310), O-linked glycosylation (S62, S227), Phosphorylation (S117), N-linked glycosylation (N197), Amidation (S664)

PROSITE: Subtilase_ASP (181-192, PROSITE entry PS00136), Subtilase_HIS (231-241, PROSITE Entry PS00137), Subtilase_SER (559-569, PROSITE entry PS00138)

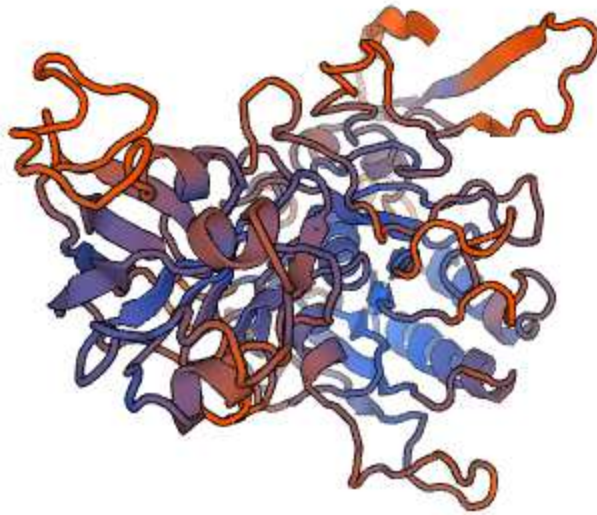
Subtilases are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases but which evolved by independent convergent evolution. Subtilase family currently includes the following proteases in Fungi:

- Alkaline extracellular protease (AEP) from *Yarrowia lipolytica* (gene xpr2).
- Alkaline proteinase from *Cephalosporium acremonium* (gene alp).
- Cerevisin (EC 3.4.21.48) (vacuolar protease B) from yeast (gene PRB1).
- Cuticle-degrading protease (pr1) from *Metarhizium anisopliae*.
- KEX-1 protease from *Kluyveromyces lactis*.
- Kexin (EC 3.4.21.61) from yeast (gene KEX-2).
- Oryzin (EC 3.4.21.63) (alkaline proteinase) from *Aspergillus* (gene alp).
- Proteinase K (EC 3.4.21.64) from *Tritirachium album* (gene proK).
- Proteinase R from *Tritirachium album* (gene proR).

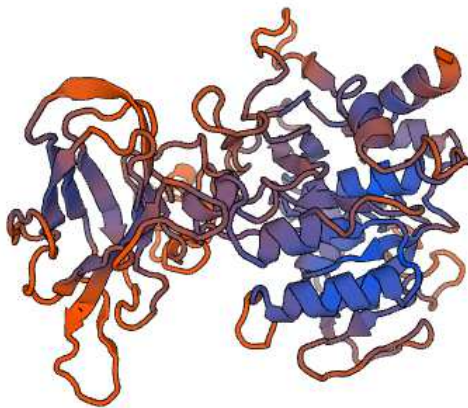
- Proteinase T from *Tritirachium album* (gene proT).
- Subtilisin-like protease III from yeast (gene YSP3).
- Thermomycolin (EC 3.4.21.65) from *Malbranchea sulfurea*.

Structural modelling:

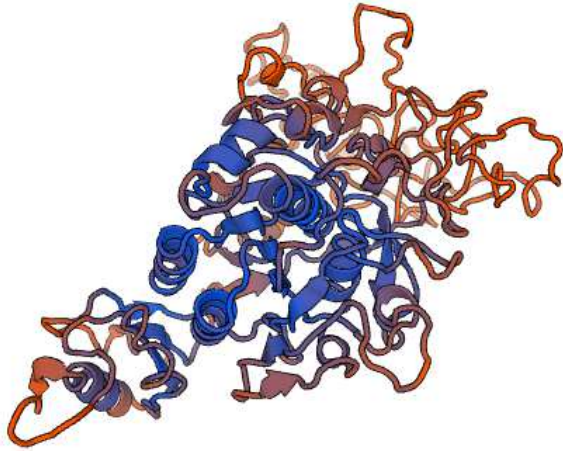
Name	Title	Identity	Method	Oligo State	Ligands
3eif.1.A	C5a peptidase	23.48	X-ray, 1.9Å	monomer	1 x CA , 1 x MLA
4i0w.1.B	Protease CspB	18.69	X-ray, 1.6Å	hetero-oligomer	None
1r6v.1.A	subtilisin-like serine protease	24.16	X-ray, 1.7Å	monomer	1 x CA
1y9z.1.A	alkaline serine protease	23.16	X-ray, 1.4Å	monomer	2 x CA , 1 x PMS
3i74.1.A	Subtilisin-like protease	32.94	X-ray, 2.6Å	homo-dimer	1 x NAG-NAG , 1 x NAG-FUC , 2 x ACE-PHE-GLU-LYS-ALV-0QE



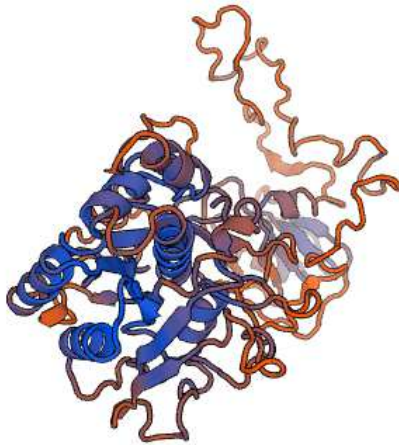
Model #1: Residues 160-667 of MVLG_02331T0 with 3eif.1.A (23.48 % sequence identity) as a template



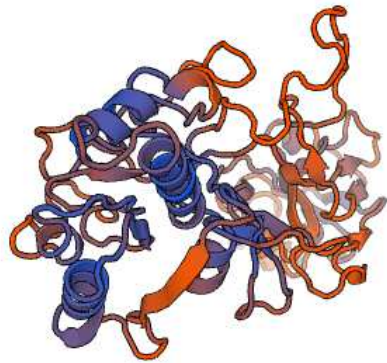
Model #2: Residues 160-632 of MVLG_02331T0 with 4i0w.1.B (18.69% sequence identity) as a template



Model #3: Residues 34-630 of MVLG_02331T0 with 1r6v.1.A (24.16% sequence identity) as a template



Model #4: Residues 163-633 of MVLG_02331T0 with 1y9z.1.A (23.16% sequence identity) as a template

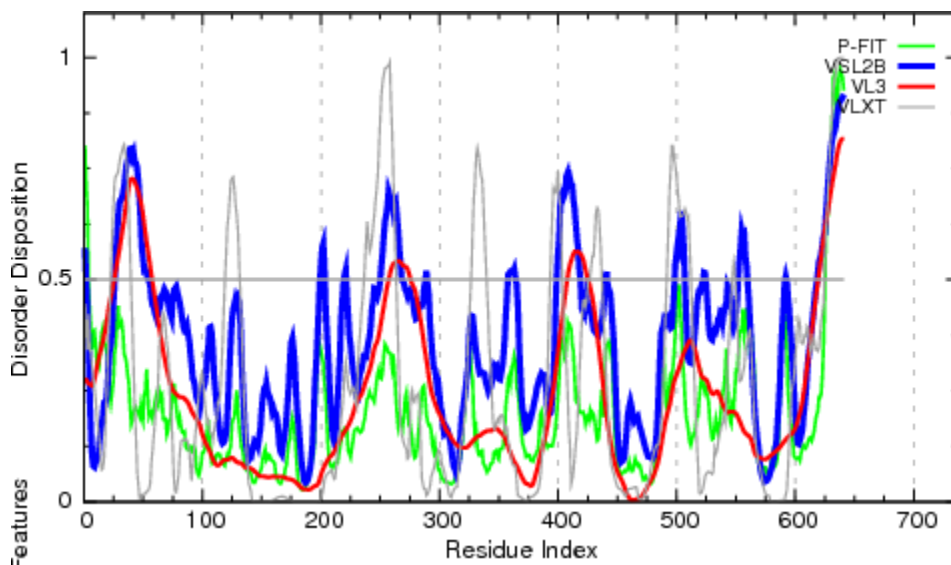


Model #5: Residues 231-625 of MVLG_02331T0 with 3i74.1.A (32.94% sequence identity) as a template

> MVLG_02184T0 (19.34%)- 641 aa

MLIGDLLLPCAFALGRLVLASPLPLSEFETELHRHKTSPGHPEKGSVFHSAYK
PLWEAHRGQQLKLVTYDHEDYRLSSDFEITDIPTERNYYFDVSEVTAAPDG
VTRKMFLVNARVNGELIEANEGDTIKLHVRNWLRVGTGIHFHGIPQAHVNYFDG
PVGVVTCPIASKSEFTFSFKLVNVCCTYFWHGHRSTQSDGINGPVVHCRNDT
LKKGADFDREQVVMVTDNYHELSSVIMEKLRSSAGVYGSTSTPTPKSGLIQGRG
DFDCKNRTNILKGHSCKKQSIYSEIAPAGSLRLRFINAGMHAFWRISVDEHEM
KLIEVDDTPIDAVGMPRIPINAGQRFSAVLDRSDKAGSSFWMRSAATQCFRAP
LNGFNPETLAIVRVDPYASTSSSSGQQKFPTSKPFTHDLVELCEDPATSILRPRV
AENTADTADQVDYFNATYILAPEGGRFYMNGISFEAYAYDPLLFRAIRGESIANR
SATVIMSETTGEGGRARVHDIIVNPNPGGGAHPFHLHGPRSYIVGGGDGTISKETW
ATMTPMTQNPTRRDVFTVPPNSYIVIRIVADLVGVHAFHCHVSPHTSVGMAGAL
VVRPDLIRQIQLPQESIDMCKASHYDSGFSEETPESARRR

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 641

Number residues disordered: 124

Overall percent disordered: 19.34
0.3630

Number Disordered Regions: 13

Longest Disordered Region: 28

Average Prediction Score:

Predicted disorder segment [27]-[54]

Average Strength= 0.6757

Predicted disorder segment [200]-[203]

Average Strength= 0.5562

Predicted disorder segment [220]-[221]

Average Strength= 0.5163

Predicted disorder segment [251]-[268]

Average Strength= 0.6264

Predicted disorder segment [289]-[289]

Average Strength= 0.5192

Predicted disorder segment [358]-[364]

Average Strength= 0.5156

Predicted disorder segment [399]-[418]

Average Strength= 0.6553

Predicted disorder segment [441]-[441]

Average Strength= 0.5179

Predicted disorder segment [499]-[506]

Average Strength= 0.5902

Predicted disorder segment [523]-[524]

Average Strength= 0.5158

Predicted disorder segment [552]-[559]

Average Strength= 0.5641

Predicted disorder segment [592]-[592]

Average Strength= 0.5069

Predicted disorder segment [619]-[641]

Average Strength= 0.7449

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			

	From	To	Length
1	5	18	14
2	474	477	4
3	479	479	1
4	570	586	17

ModPred and PROSITE:

ModPred: Amidation (A14, F316, S405, R507, I571), Proteolytic cleavage (Q200, R394, R560, D561, D627, F630, R641), O-linked glycosylation (S256), Phosphorylation (T259), ADP-ribosylation (R270)

PROSITE: Multicopper oxidase2 (586-597, PROSITE entry PS00080)

Multicopper oxidases [1,2] are enzymes that possess three spectroscopically different copper centers. These centers are called: type 1 (or blue), type 2 (or normal) and type 3 (or coupled binuclear). Consensus pattern: H-C-H-x(3)-H-x(3)-[AG]-[LM]

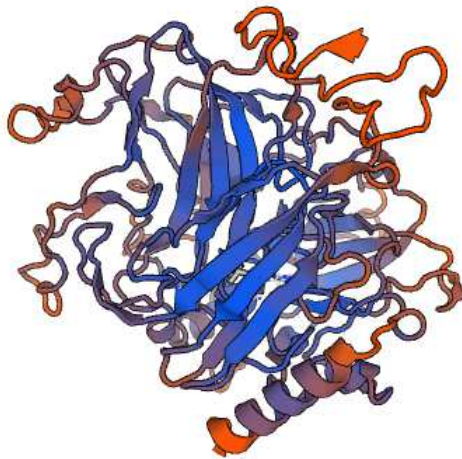
The first 2 H's are copper type 3 binding residues; The C, the third H, and L or M are copper type 1 ligands. The enzymes that belong to this family are:

- Laccase (EC 1.10.3.2) (urishiol oxidase), an enzyme found in fungi and plants, which oxidizes many different types of phenols and diamines.
- In addition to the above enzyme there are a number of proteins which, on the basis of sequence similarities, can be said to belong to this family. These proteins are:

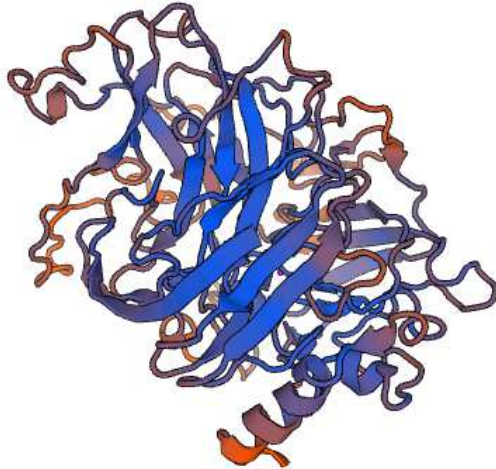
- Copper resistance protein A (copA) from a plasmid in *Pseudomonas syringae*. This protein seems to be involved in the resistance of the microbial host to copper.
- Yeast FET3, which is required for ferrous iron uptake.
- Yeast hypothetical protein YFL041w and SpAC1F7.08, the fission yeast homolog.

Structural modelling:

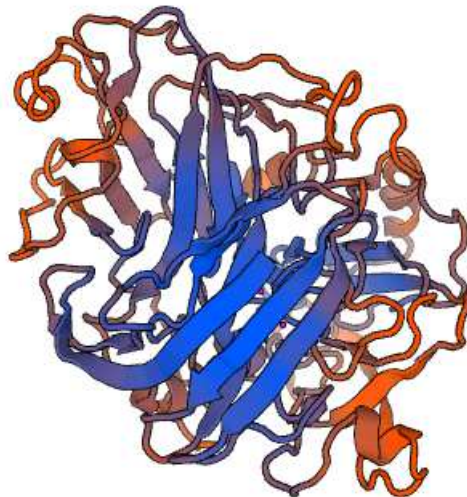
Name	Title	Identity	Method	Oligo State	Ligands
3v9e.1.A	Laccase	27.48	X-ray, 1.7Å	monomer	5 x <u>NAG</u> , 3 x <u>CU</u> , 1 x <u>MAN</u>
1aso.1.A	ASCORBATE OXIDASE	27.23	X-ray, 2.2Å	homo- dimer	2 x <u>NAG</u> , 9 x <u>CU</u> , 2 x <u>OH</u>
5mew.1.A	Laccase 2	28.06	X-ray, 1.3Å	monomer	4 x <u>NAG</u> , 4 x <u>CU</u> , 2 x <u>OXY</u>



Model #1: Residues 50-629 of MVLG_02184T0 with 3v9e.1.A (27.48% sequence identity) as a template



Model #2: Residues 95-631 of MVLG_02184T0 with 5mew.1.A (28.06% sequence identity) as a template

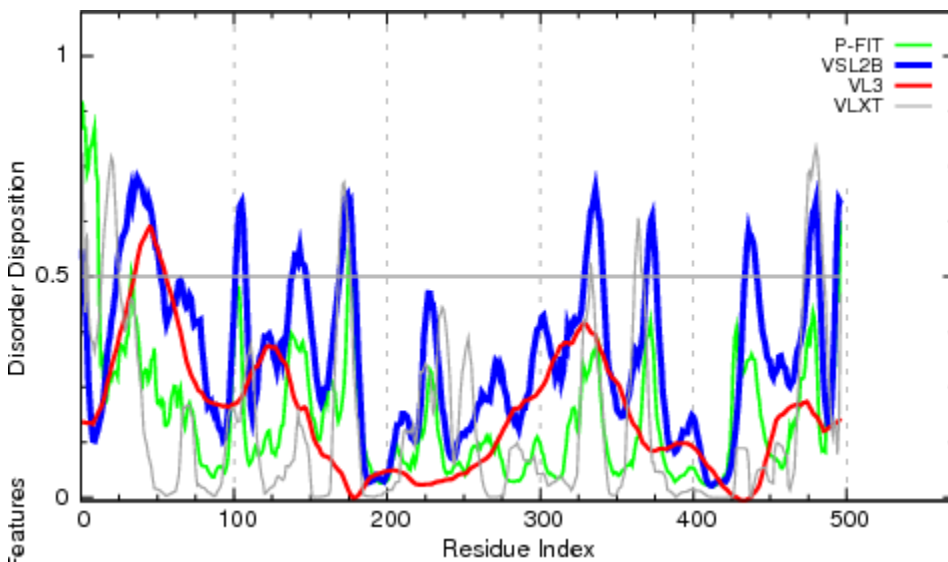


Model #3: Residues 94-598 of MVLG_02184T0 with 1aso.1.A (27.23% sequence identity) as a template

> MVLG_07060T0 (18.95%)- 496 aa

MMWSTLLVPITAALAATAVHAATNHAAVGAHSSLDHNKGSQNVAKGDISKG
NTFQVISHPDFPNHKLRIKESQLCGDKEKIYSGFLDIAEHTHLFFAFAESRDKPDE
DSVLLWLNGGPGCSSMAGFLENGPCLVTNNGGNSSTFNPYSWNSNANMIFLDSP
VKVGFSNARKPVDTSRKTAEDIYAFMQLFYQVFPRFAMLDIFILAGESYAGMYIP
QVASVIVQKNKLVDGASSNTIYVPLVSMIAIGNGFVEIVSALSAEVDFACGKGVH
KAIYNSSTCDALYPQIPICRSVATCRQNLTRQNCQQAELDCFVLGAPFDNTGLN
PYDVTKKCDRSPSKDGPLCYKEASWLPIYLNRPDIRAKLGVHAKAKPFEECSDSV
HTAFLLSGDWVVNTPAVLSDLLEAGIKLLLYVGVNDFICNYLGVNRNWTAMKW
SGQDQYSKAPFHEFRMPNGTVVGLTKSYGPLTYLEVKDAGHMVPRDKPDEALE
MIKTWIRGDQF

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 496
Number residues disordered: 94
Overall percent disordered: 18.95
0.3250

Predicted disorder segment [24]-[52]
Predicted disorder segment [101]-[107]
Predicted disorder segment [138]-[147]
Predicted disorder segment [170]-[178]
Predicted disorder segment [329]-[340]
Predicted disorder segment [370]-[375]
Predicted disorder segment [434]-[441]

Number Disordered Regions: 9
Longest Disordered Region: 29
Average Prediction Score:

Average Strength= 0.6341
Average Strength= 0.6034
Average Strength= 0.5319
Average Strength= 0.6156
Average Strength= 0.6186
Average Strength= 0.5888
Average Strength= 0.5633

Predicted disorder segment [475]-[482]
Predicted disorder segment [493]-[496]

Average Strength= 0.6040
Average Strength= 0.6175

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			
	From	To	Length
1	1	3	3
2	6	8	3
3	94	96	3

ModPred and PROSITE:

ModPred: Proteolytic cleavage (H31, D108, S109, L111, W113, K362, H473),

Amidation (Y465)

PROSITE: CARBOXYPEPT_SER_SER (Serine carboxypeptidases, serine active site, 204-211, PROSITE entry PS00131), CARBOXYPEPT_SER_HIS (Serine carboxypeptidases, histidine active site, 463-480, PROSITE entry PS00560)

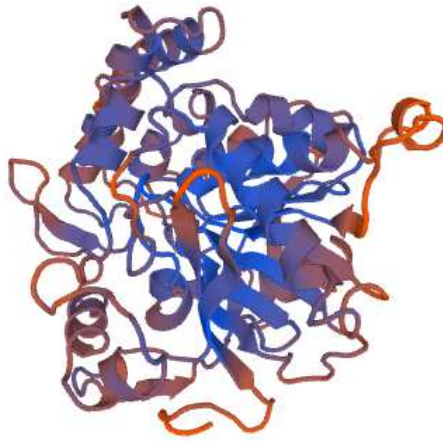
All known carboxypeptidases are either metallo carboxypeptidases or serine carboxypeptidases (EC 3.4.16.5 and EC 3.4.16.6). The catalytic activity of the serine carboxypeptidases, is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which is itself hydrogen-bonded to a serine.

Fungal Proteins known to be serine carboxypeptidases are:

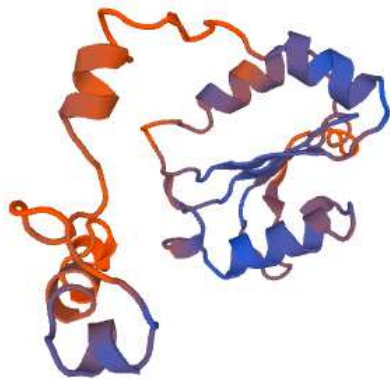
- Yeast carboxypeptidase Y (YSCY) (gene PRC1), a vacuolar protease involved in degrading small peptides.
- Yeast KEX1 protease, involved in killer toxin and α -factor precursor processing.
- Fission yeast *sxa2*, a probable carboxypeptidase involved in degrading or processing mating pheromones.
- *Penicillium janthinellum* carboxypeptidase S1
- *Aspergillus niger* carboxypeptidase pepF.
- *Aspergillus satoii* carboxypeptidase cpdS.
- Yeast hypothetical protein YBR139w.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
1ysc.1.A	SERINE CARBOXYPEPTIDASE	37.35	X-ray, 2.8Å	monomer	3 x <u>NDG</u>
1gxs.1.B	P-(S)- HYDROXYMANDELONITRILE LYASE CHAIN B	23.97	X-ray, 2.3Å	hetero- oligomer	2 x <u>BEZ</u> , 2 x <u>NAG</u> , 2 x <u>DKA</u>



Model #1: Residues 72-495 of MVLG_07060T0 with 1ysc.1.A (37.35% sequence identity) as a template



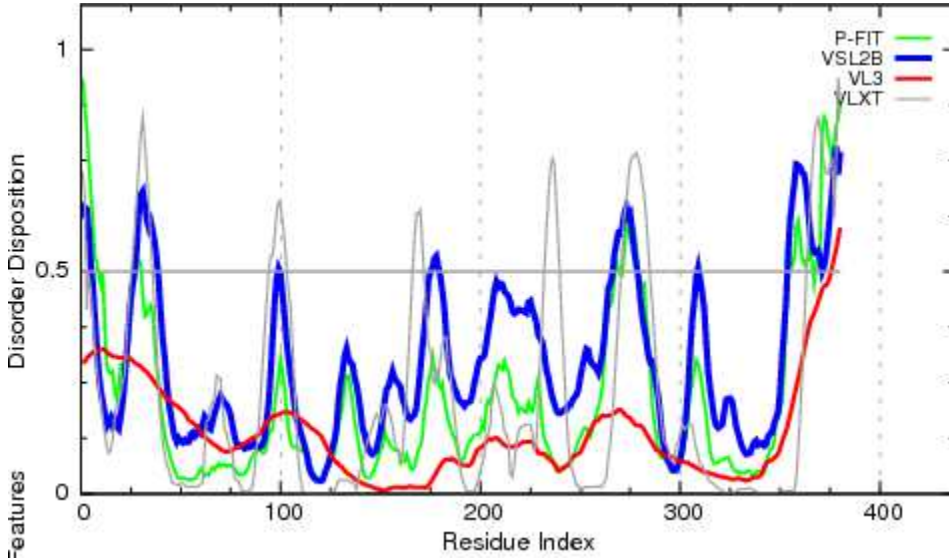
Model #2: Residues 341-496 of MVLG_07060T0 with 1gxs.1.B (23.97% sequence identity) as a template

> MVLG_01474T0 (16.32%)- 380 aa

MLSKLRNVSVAALLFAGLAIAAPAPVSNSSEARHGKQNLLTPKVMII SMFAP
ERAVWIKPMKLVHNVSVVGLSPLYPYVACNNEYDVCIMTTGEAEINAAAS MMA
LALSPLFCLQHTYFLIAGIGGVNPYAGTLGSAAFARFAVQVALEYELDARQIPSN

WTTGYWMQNTAGPGQLSATKDLYGTELFEVNTNLLAKAYSAAKGVTLNDSTT
 AQAYRQKFDYAPANQPPQVILGDVATSDVYYAGTLLSESGNYTALLTNGTGKY
 TTTAQEDNATLESMVRATKAGLLDYARVILRTCSDFDRAPPKVTAYDAFFAN
 QGGFELALQNLVIAGKPVVDMILKDWSTFKNGVQPQSKGNGSYYGDDLGLTRS
 GPALA

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 380	Number Disordered Regions: 7
Number residues disordered: 62	Longest Disordered Region: 27
Overall percent disordered: 16.32	Average Prediction Score:
0.2960	
Predicted disorder segment [1]-[5]	Average Strength= 0.6035
Predicted disorder segment [27]-[38]	Average Strength= 0.6006
Predicted disorder segment [99]-[99]	Average Strength= 0.5068
Predicted disorder segment [176]-[179]	Average Strength= 0.5212
Predicted disorder segment [267]-[278]	Average Strength= 0.5886
Predicted disorder segment [309]-[309]	Average Strength= 0.5085
Predicted disorder segment [354]-[380]	Average Strength= 0.6394

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			

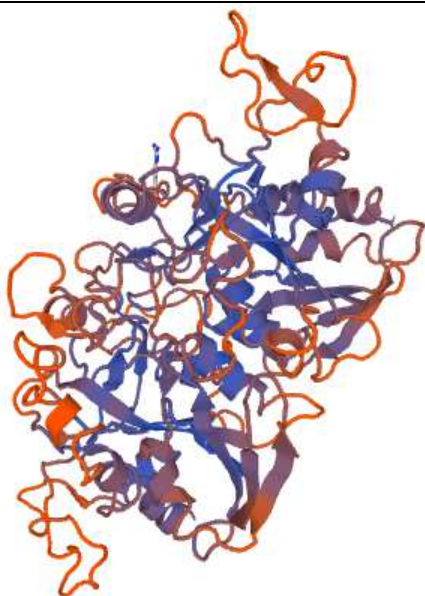
ModPred and PROSITE:

ModPred: GPI anchor amidation (N29), Amidation (A227, Q232, Y366), Proteolytic cleavage (R374).

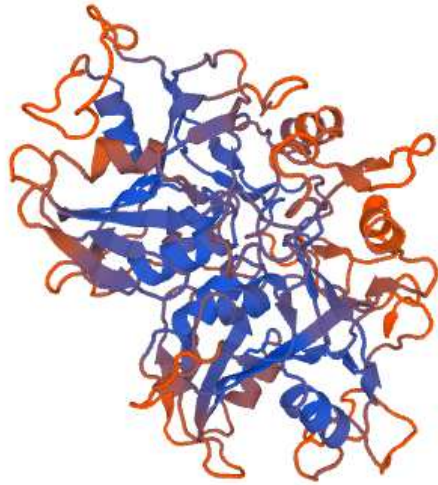
PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
1zos.1.A	5'-methylthioadenosine / S-adenosylhomocysteine nucleosidase	18.78	X-ray, 1.6Å	homo-dimer	2 x <u>MTM</u>
4g89.1.B	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	17.65	X-ray, 2.1Å	homo-dimer	1 x <u>SAH</u> , 1 x <u>ADE</u>



Model #1: Residues 45-349 of MVLG_01474T0 with 1zos.1.A (18.78% sequence identity) as a template

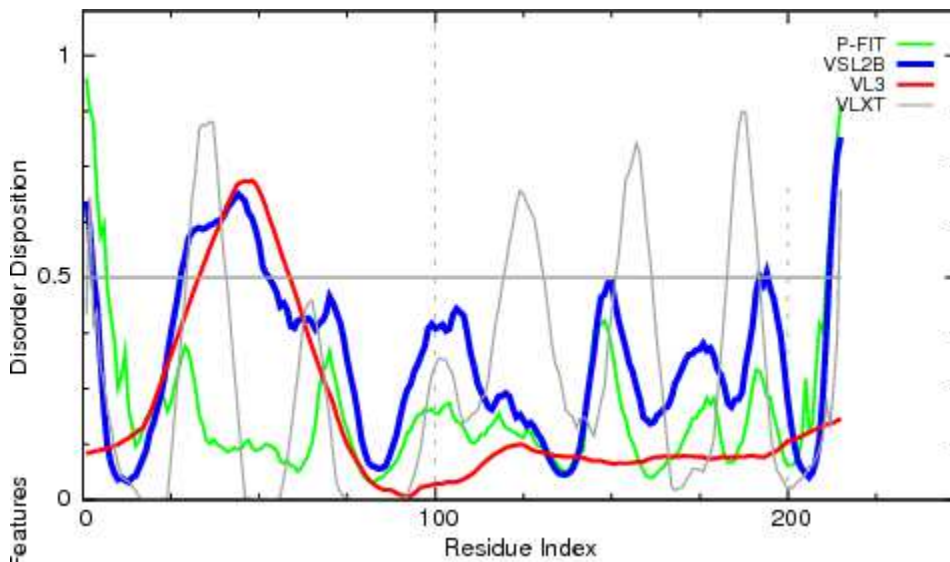


Model #2: Residues 44-311 of MVLG_01474T0 with 4g89.1.B (17.65% sequence identity) as a template

> **MVLG_00784T0 (15.81%)- 215 aa**

MLVSFKPFGLVWLCMSV FVSATFGRVTKDPKTPIVKGPKAPVVKEPKTPLVFQ
 ECKKYTYSKDIGKLFKRKGVEGKATFFINGYNYGCIYNKENVNALRARYHEGHA
 DLVNLSSVQIVKQVELLET AVERILGVRLGMFIAPYDSIDEKAAKVIRDKGYKIVR
 WSLDSGDTTFYLGRPQSSVNLIRKWIKKASGKSGIGLFDEVGFIFLTCHGSE

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 215	Number Disordered Regions: 5
Number residues disordered: 34	Longest Disordered Region: 25
Overall percent disordered: 15.81	Average Prediction Score:
0.3172	
Predicted disorder segment [1]-[3]	Average Strength= 0.5966
Predicted disorder segment [28]-[52]	Average Strength= 0.6149
Predicted disorder segment [192]-[192]	Average Strength= 0.5020
Predicted disorder segment [194]-[194]	Average Strength= 0.5114
Predicted disorder segment [212]-[215]	Average Strength= 0.7071

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			

ModPred and PROSITE:

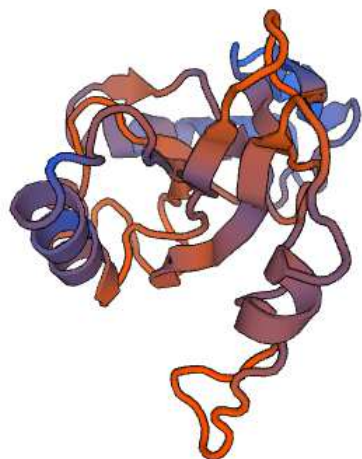
ModPred: Sumoylation (K6), Proteolytic cleavage (T22, F23, R25), Acetylation (K62, K66), Methylation (K69), Amidation (I89, V95, Y102), Ubiquitination (K92), Pupylation (K152).

PROSITE: No identified domain recognition sites.

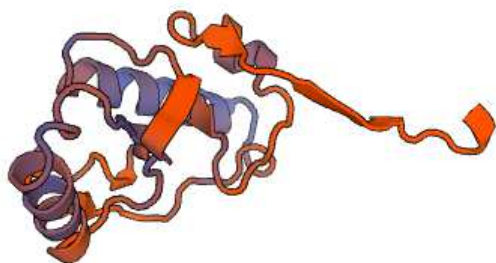
Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
2iw0.1.A	CHITIN DEACETYLASE	26.88	X-ray, 1.8Å	monomer	1 x <u>ZN</u>
3qbu.1.A	Putative uncharacterized protein	16.81	X-ray, 2.6Å	homo-tetramer	4 x <u>ZN</u>

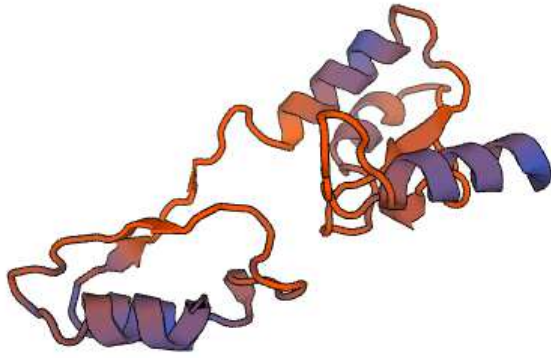
Name	Title	Identity	Method	Oligo State	Ligands
3l6u.1.A	ABC-TYPE SUGAR TRANSPORT SYSTEM PERIPLASMIC COMPONENT	17.39	X-ray, 1.9Å	homo-dimer	None
2eqo.1.A	TNF receptor-associated factor 3-interacting protein 1	13.21	NMR	monomer	None



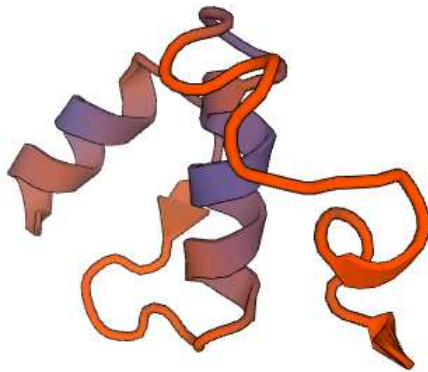
Model #1: Residues 37-203 of MVLG_00784T0 with 2iw0.1.A (26.88% sequence identity) as a template



Model #2: Residues 48-173 of MVLG_00784T0 with 3qbu.1.A (16.81% sequence identity) as a template



Model #3: Residues 61-191 of MVLG_00784T0 with 3l6u.1.A (17.39% sequence identity) as a template

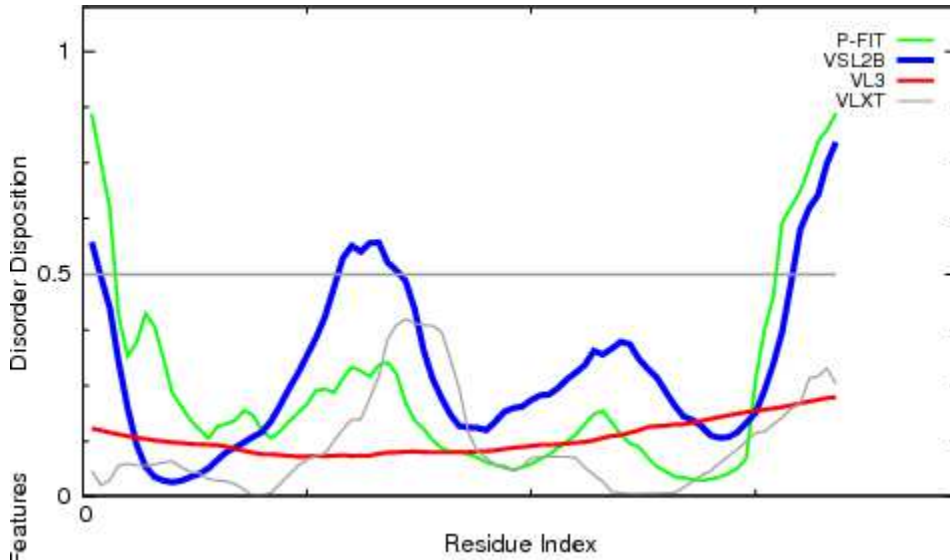


Model #4: Residues 121-173 of MVLG_00784T0 with 2eqo.1.A (13.21% sequence identity) as a template

> MVLG_06541T0 (15.48%)- 84 aa

MRFSLIVLATLLVGFAAAAPVLQSDIFKDTQKADRAQNLVHRLMVRVSTVYSQ
CTHNCEDEYARYKIGPYELIACKKGCQSDAI

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 84
Number residues disordered: 13
Overall percent disordered: 15.48
0.2829
Predicted disorder segment [29]-[35]
Predicted disorder segment [80]-[84]

Number Disordered Regions: 2
Longest Disordered Region: 7
Average Prediction Score:
Average Strength= 0.5466
Average Strength= 0.6944

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			

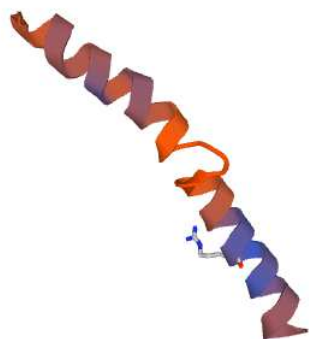
ModPred and PROSITE:

ModPred: Amidation (Y70).

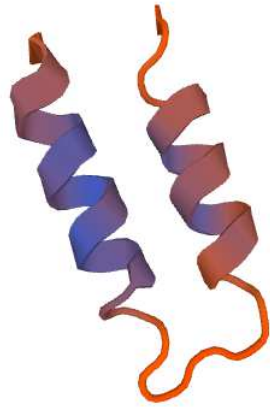
PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
2m7g.1.A	Geopilin domain 1 protein	21.62	NMR	monomer	None
2lqx.1.A	Trypsin inhibitor BWI-2c	32.35	NMR	monomer	None
2l03.1.A	Ly-6/neurotoxin-like protein 1	22.58	NMR	monomer	None



Model #1: Residues 6-44 of MVLG_06541T0 with 2m7g.1.A (21.62% sequence identity) as a template



Model #2: Residues 29-64 of MVLG_06541T0 with 2lqx.1.A (32.35% sequence identity) as a template



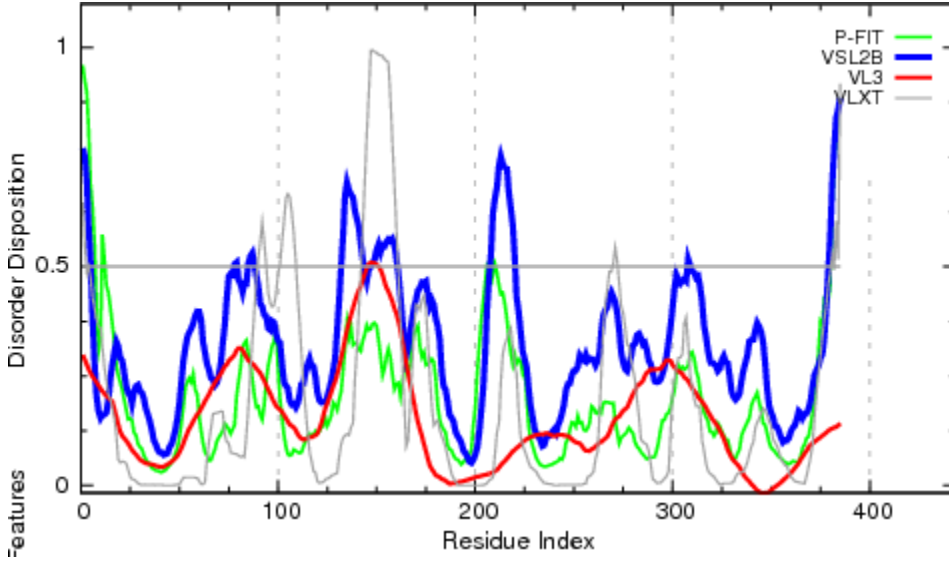
Model #3: Residues 48-82 of MVLG_06541T0 with 2l03.1.A (22.58% sequence identity) as a template

> MVLG_01897T0 (14.81%)- 385 aa

MLLSTLLASLLVAVSAKKGLLAAYYPASQVDA AIDWNVTDIGYYMAAVTAK
 NGLAFPAGKPGLADFVLR AHAHKKKAVLSIGGPEGSQYFSSLVRTETARAKFVE
 QILEVGRKYNTDGV DISWQFPTVHG NPKNEIDPKDSANLLKLLKDLRRSRPK EW
 LSAAVSPNGIFAPSGTTTLSNYQDFAEVVDAFNVMAYHYVGAWNEWTGPD SPS
 HQCGTGRSVTTDIERFIKAGFVANKIMLGIP SFGKAFTLHNNTLR TSVVYGDQQV
 PKKYQIRIRQRYETYNGADNTFVKLKAQGILKGEDGLTAGRGYKRHYDHCSRTP

FLFNPGTKSFITYLDARSASYRAQIAVQQEYLGVFVSSIGLDNLAFNAIDKALKTP
IDGFATD

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 385	Number Disordered Regions: 8
Number residues disordered: 57	Longest Disordered Region: 13
Overall percent disordered: 14.81	Average Prediction Score:
0.3237	
Predicted disorder segment [1]-[5]	Average Strength= 0.6648
Predicted disorder segment [78]-[79]	Average Strength= 0.5072
Predicted disorder segment [85]-[88]	Average Strength= 0.5170
Predicted disorder segment [132]-[143]	Average Strength= 0.6103
Predicted disorder segment [148]-[160]	Average Strength= 0.5425
Predicted disorder segment [208]-[220]	Average Strength= 0.6601
Predicted disorder segment [308]-[309]	Average Strength= 0.5091
Predicted disorder segment [380]-[385]	Average Strength= 0.7677

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length

None			
Filtered Regions			
	From	To	Length
1	194	198	5

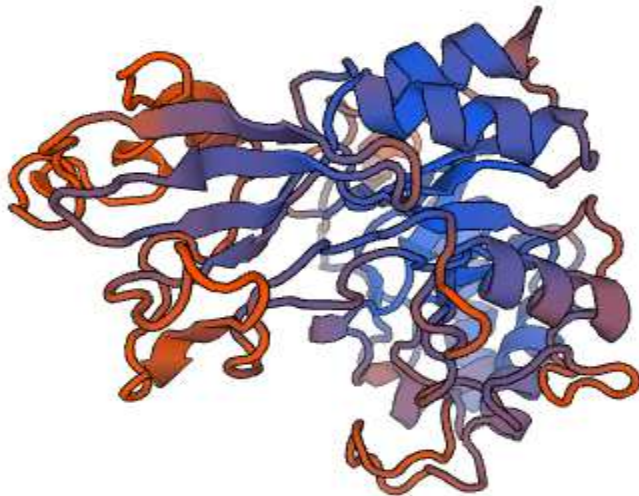
ModPred and PROSITE:

ModPred: Amidation (A9, P209, Y283, S340, Y353), Ubiquitination (K52), Proteolytic cleavage (D66, R154, R220), Acetylation (K248, K300), Pupylation (K376).

PROSITE: No identified domain recognition sites.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
4txg.1.A	Chitinase	21.02	X-ray, 1.8Å	monomer	11 x CS

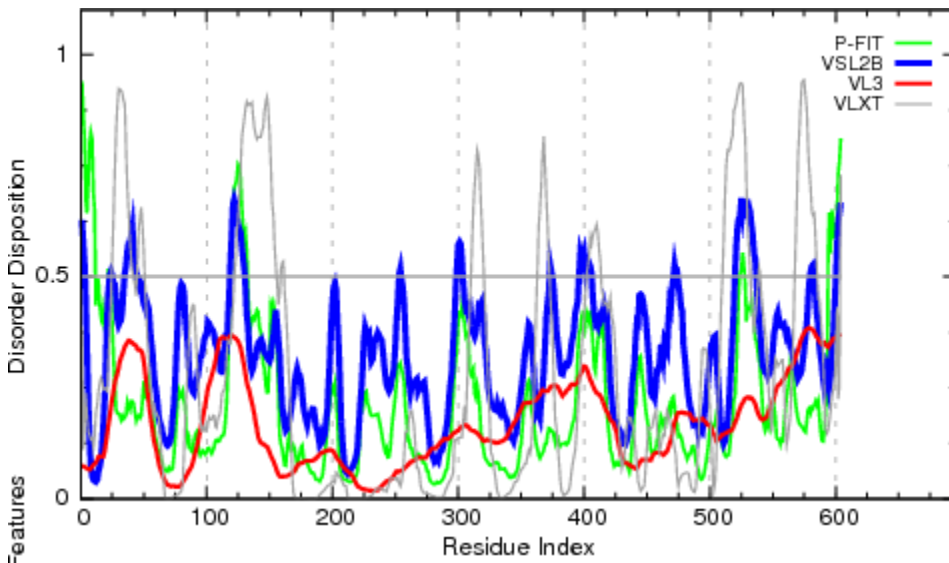


Model #1: Residues 17-377 of MVLG_01897T0 with 4txg.1.A (21.02% sequence identity) as a template

> MVLG_06973T0 (12.58%)- 604 aa

MWALGGIAGSVLWICAVNGQGSSYAPRRVRCPTDGPLVKSTGSPLAGNQFLES
REATYQAARWNKVLEPLYLKYLGNQDTGYSTAQIATIVKHEPRIGTACSGGGL
RASLYCAGTLSALDSRSRSHAAPVLQLSAYMTGLSGGSWAITSLATSNLGPTSIY
DIVLGKNGAPGWKLDLNLNFPNIKHLIPFNANIIRDLHEKNRAHFSVTLIDYWGR
LGHHFLPGTTRASFFSQLAPNDNGLLFDAINSTSKFKEFEMPYPITVTTTSRVRP
WDQFKVVHDYIPAINTVFEISPYSFGSFDPSLSAHIPTIYMGSYVEQTQTGVART
CVNGFDSASFIMGCSAGLFTAIESMLQPDMKTFRRLLSLIHRVSKEEKLDILTSK
VPNTFYGYNSGLMGSRRFESAENKNLYLTDGGMNGENIPLAPLLVKARRLDTIFA
IDASQDTKMSWPNGVSLHRTWERINRTANGYSDFPPVPSKPYDFLMGGLTRRPV
FFGCNVKDARVDKPGNYPILYLPNAPVPHSGYSTNTKTSQMEYSISDTEAFLNT
VQANAMKGYPPGDAVVDREYKTALKCATVDRARQRGNMARSACIQMQRYCWPPL
KA

PONDR:



=====PONDR VSL2

STATISTICS=====

Predicted residues: 604
Number residues disordered: 76
Overall percent disordered: 12.58
0.3269
Predicted disorder segment [1]-[3]
Predicted disorder segment [23]-[25]
Predicted disorder segment [35]-[44]

Number Disordered Regions: 12
Longest Disordered Region: 17
Average Prediction Score:

Average Strength= 0.5765
Average Strength= 0.5060
Average Strength= 0.5536

Predicted disorder segment [119]-[130]	Average Strength= 0.6016
Predicted disorder segment [253]-[254]	Average Strength= 0.5118
Predicted disorder segment [299]-[303]	Average Strength= 0.5561
Predicted disorder segment [373]-[374]	Average Strength= 0.5188
Predicted disorder segment [396]-[406]	Average Strength= 0.5331
Predicted disorder segment [471]-[474]	Average Strength= 0.5104
Predicted disorder segment [520]-[536]	Average Strength= 0.6177
Predicted disorder segment [580]-[582]	Average Strength= 0.5206
Predicted disorder segment [601]-[604]	Average Strength= 0.6225

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
1	491	496	6
Filtered Regions			
	From	To	Length
1	11	15	5
2	70	72	3
3	510	514	5

ModPred and PROSITE:

ModPred: Amidation (K39, E373), Acetylation (K73), Sumoylation (K255)

PROSITE: PLA2C (PLA2c domain profile, 30-604, PROSITE entry PS51210)

The PLA2c domain is the catalytic lipase domain in cytosolic phospholipase A2 (cPLA2) (EC 3.1.1.4) and lysophospholipase or phospholipase B (PLB) (EC 3.1.1.5) of vertebrates and fungi. It catalyzes the carboxylic ester hydrolysis of glycerophospholipids or lysophospholipids. The mammalian cPLA2 group IVA enzymes cleave intracellular phospholipid membranes to produce lipid mediators, which also play a role in inflammatory diseases such as asthma and arthritis. This enzyme contains a N-terminal calcium-binding C2 domain that presents the catalytic domain to the membrane. Fungal

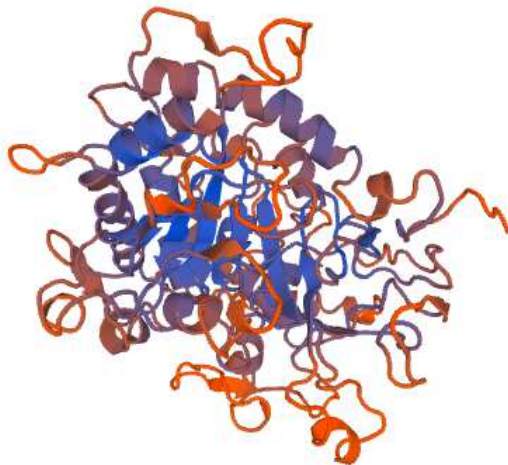
secreted lysophospholipase/ PLB can possess three different enzymatic activities, the hydrolase activity of phospholipase, lysophospholipase and a lysophospholipase transacylase activity.

Some fungal proteins known to contain a PLA2c domain:

- Fungal lysophospholipases/PLB, which are considered to be important for virulence of pathogenic fungi.
- Yeast sporulation-specific protein 1 (SPO1), which is required for meiosis.

Structural modelling:

Name	Title	Identity	Method	Oligo State	Ligands
5iz5.1.A	Cytosolic phospholipase A2 delta	19.26	X-ray, 2.2Å	monomer	None



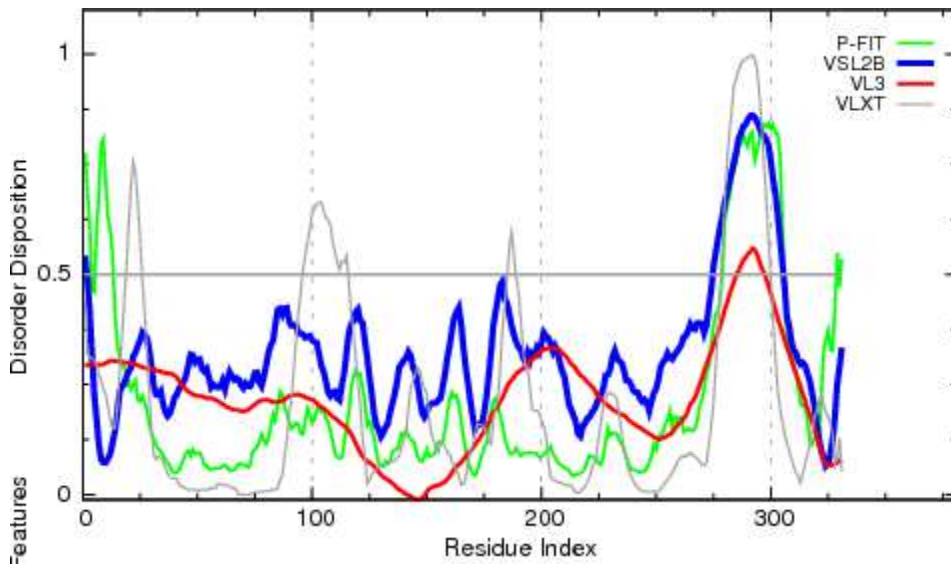
Model #1: Residues 31-577 of MVLG_06973T0 with 5iz5.1.A (19.26% sequence identity) as a template

Mostly ordered proteins (<10% disordered residues by PONDR® VSL2 analysis)

> MVLG_01005T0 (9.67%)- 331 aa

MHYTRLCLLCAALS**NAPILLSA**RPQITMSADLSLECFIATTGLLSSPFARCADAS
GFLAALDAKIGLADALSDWLNNFCKDTCPPDARAKAWSGLEGGCADDELAREIA
LPSVLLGTGHLRSSATAVQCPRFDNFVESNPILNARSTAVVANYDVLKRSACTG
SVSRQSYCFVEFVKDLEEANKRNFTMSVDLLSPTCAELNAVPRSKLCTLCNQML
FEMMVKLLSRPIDRVTLTDHARQACGLAFASLSALQADFLRPVGGQFDQAVAR
YTA-AAIKDLDQHNSTPSTPTTSSAYSSRTLRRSTFALDTVVKAVAPTAATILLY
GWVQ

PONDR:



=====PONDR VSL2
STATISTICS=====

Predicted residues: 331
Number residues disordered: 32
Overall percent disordered: 9.67
0.3224
Predicted disorder segment [275]-[305]

Number Disordered Regions: 1
Longest Disordered Region: 31
Average Prediction Score:
Average Strength= 0.7347

ANCHOR:

Predicted Disordered Binding Regions			
	From	To	Length
None			
Filtered Regions			
	From	To	Length
1	325	331	7

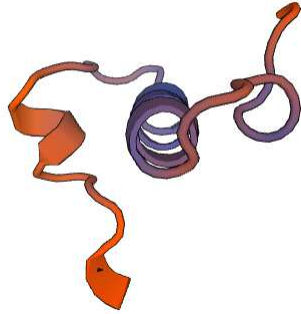
ModPred and PROSITE:

ModPred: Pupylation (K177, K278), ADP-ribosylation (R206), N-linked glycosylation (N284), O-linked glycosylation (S293), Proteolytic cleavage (R304), Amidation (Y327), Pyrrolidone carboxylic acid (Q331).

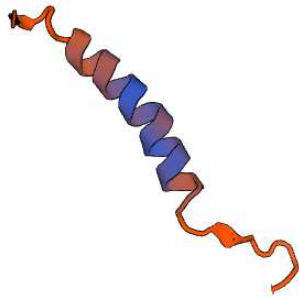
PROSITE: No identified domain recognition sites.

Structural modelling:

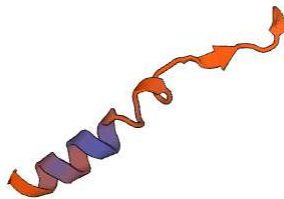
Name	Title	Identity	Method	Oligo State	Ligands
3gxv.1.A	Replicative DNA helicase	22.50	X-ray, 2.2Å	hetero-oligomer	None
5n9j.1.D	Mediator of RNA polymerase II transcription subunit 21	25.71	X-ray, 3.4Å	hetero-oligomer	None
3ajb.1.B	Peroxisomal biogenesis factor 19	22.22	X-ray, 2.5Å	hetero-oligomer	None



Model #1: Residues 202-241 of MVLG_01005T0 with 3gxv.1.A (22.50% sequence identity) as a template



Model #2: Residues 230-264 of MVLG_01005T0 with 5n9j.1.D (25.71% sequence identity) as a template

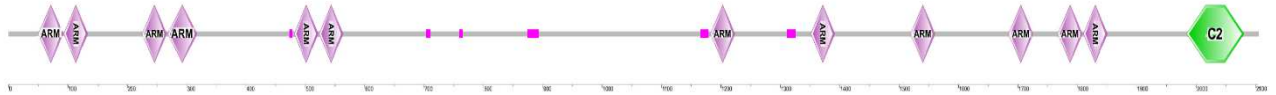


Model #3: Residues 266-291 of MVLG_01005T0 with 3ajb.1.B (22.22% sequence identity) as a template

Table S3. Primers used for PCR and sequencing

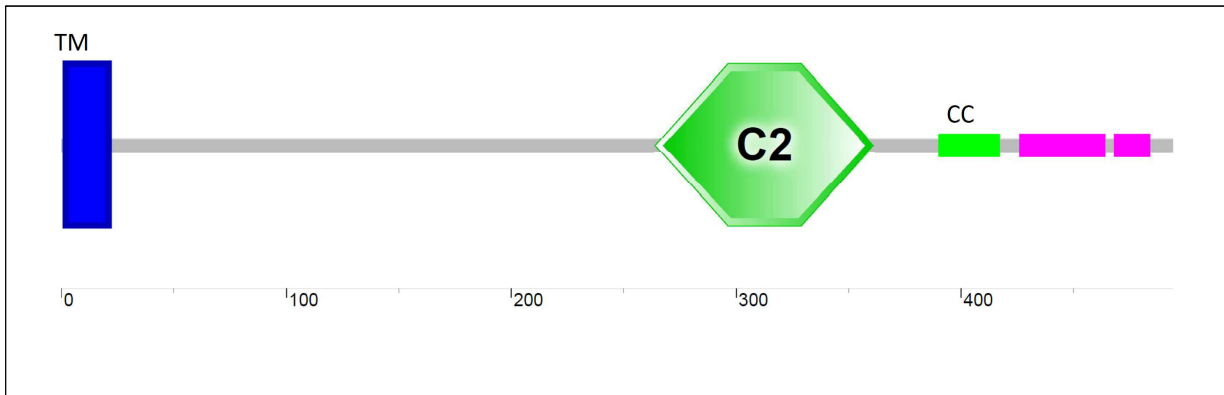
Primer	Sequence (5'→3')	Usage
MVLG_04106F	GCC GAA TTC GCT GAC GCG ACC AAA CAG GCC TCC	Yeast two hybrid
MVLG_04106R	GAC GGA TCC TCA ACA ACC TTC GGG CTC GGG TTG	Yeast two hybrid
MVLG_05720F	GCC GAA TTC AAC CCG TGG CCT CCG TCG GTT CAA	Yeast two hybrid
MVLG_05720R	GAC GGA TCC CTA GTA ACC CGA ACG ACG CAT CCT	Yeast two hybrid
MVLG_06175F	GCC GAA TTC TTT TGT CCC TTT GGA AAA ACG GCG	Yeast two hybrid
MVLG_06175R	GAC GGA TCC TTA GAG ATT TAG AGG AAA GAA CCA	Yeast two hybrid
MVLG_01732F	GCC GAA TTC TTG CAA GAA GCG GGC GAT ACC AAG	Yeast two hybrid
MVLG_01732R	GAC GGA TCC CTA GGC GTG GAT TTT GCC GGA GAA	Yeast two hybrid
T7 Sequencing primer	AATACGACTCACTATAGGGCG	Yeast two hybrid
MVLG_04106F	GCC GAA TTC ATG AAG TAC TCG CTC GTC TTT GTC	Yeast secretion trap
MVLG_04106R	GCG GCC GCC GGC GAG AGC CGA GAC GAT GCG CGT G	Yeast secretion trap
MVLG_05720F	GCC GAA TTC ATG ATG CGT TCC CTC ATC AAG TTG	Yeast secretion trap
MVLG_05720R	GCG GCC GCC CGC AAG AGC CAC ACT GAC GGC GGT G	Yeast secretion trap
MVLG_06175F	GCC GAA TTC ATG TGG ACC TCT TCG ATC GTC CAA	Yeast secretion trap
MVLG_06175R	GCG GCC GCC AGC CCA CGC CAC GAC AGG GCT CGA G	Yeast secretion trap
MVLG_01732F	GCC GAA TTC ATG CTG TTA AAG CTT ACC ATC ACC	Yeast secretion trap
MVLG_01732R	GCG GCC GCC TGC CGA AAC ATT GAG GAC GAG TAA G	Yeast secretion trap
Sequencing primer	TCCTCGTCATTGTTCTCGTTCC	Yeast secretion trap

Figure S1. Structure of Protein: Domain structures of the Cellulose Synthase Interactive 1 protein from *Silene latifolia* drawn based on SMART searches (<http://smart.embl-heidelberg.de/>). The length of the protein and the positions of ARM and C₂ domains are scaled below.



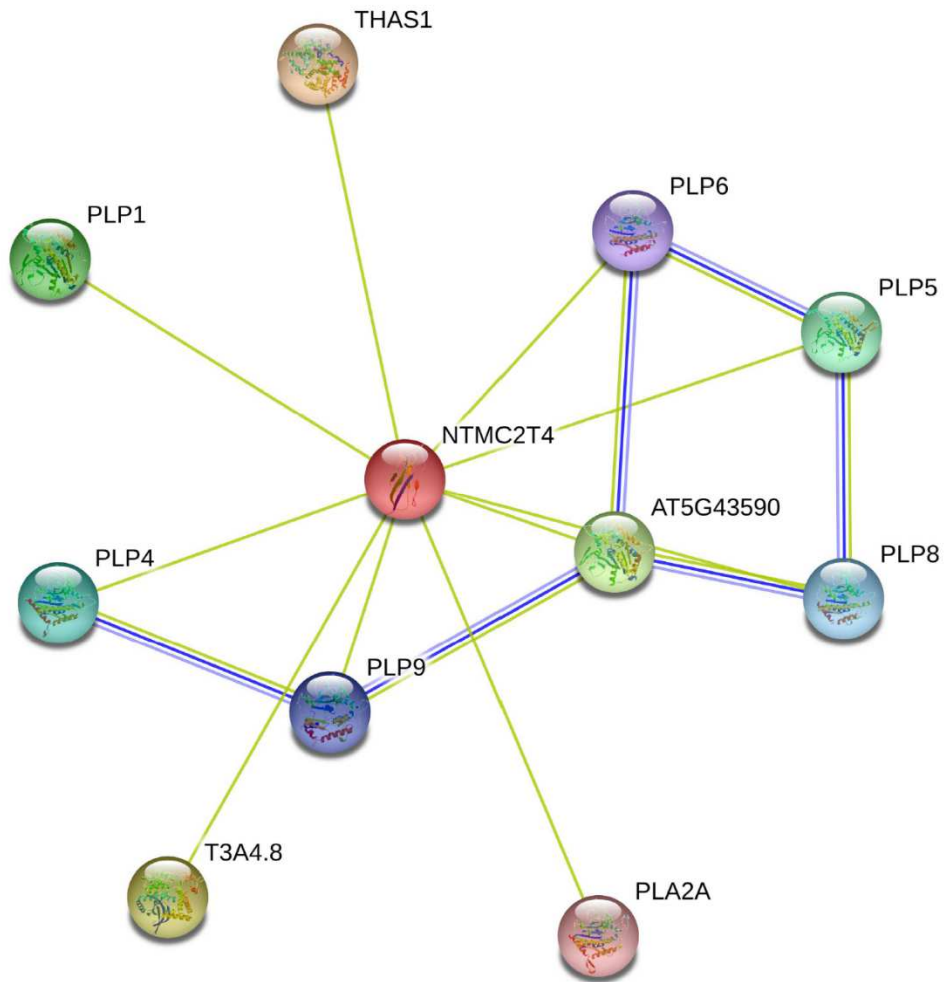
Arm repeats (50-92, 93-133, 225-266, 268-317, 481-521, 523-563, 1182-1223, 1351-1391, 1519-1560, 1684-1725, 1767-1808, 1810-1850), C₂ domain (1984-2080)

Figure S2. Structure of Protein: Domain structures of AtCLB protein from *Silene latifolia* drawn based on SMART searches (<http://smart.embl-heidelberg.de/>). The length of the protein and the positions of TM, C₂ domains are scaled below.



TM- Transmembrane region (2-21), C₂ domain (264-361), CC-Coiled coil (390-417), The regions highlighted in pink are low compositional complexity regions.

Figure S3. Predicted functional partners using STRING for the predicted MVLG_01732 host target, AtCLB, *Arabidopsis thaliana* gene At3G61050.1, network and interactions with other proteins.



Edges:

<p>Edges represent protein-protein associations</p> <p><i>associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding each other.</i></p>	<p>Known Interactions</p> <p> from curated databases</p> <p> experimentally determined</p>	<p>Predicted Interactions</p> <p> gene neighborhood</p> <p> gene fusions</p> <p> gene co-occurrence</p>	<p>Others</p> <p> textmining</p> <p> co-expression</p> <p> protein homology</p>
--	---	--	--

Your Input:

NTMC2T4 calcium-dependent lipid-binding domain-containing protein (510 aa)

Predicted Functional Partners:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
	THAS1	thalianol synthase 1; Converts oxidosqualene to thalianol (766 aa)								0.829
	T3A4.8	CDPK-related kinase 3; May play a role in signal transduction pathways that involve calcium as a second messenger (By si...								0.596
	AT5G43590	acyl transferase/acyl hydrolase/lysophospholipase-like protein (401 aa)								0.569
	PLP1	patatin-like protein (414 aa)								0.569
	PLP5	PATATIN-like protein 5 (435 aa)								0.569
	PLP4	PATATIN-like protein 4 (428 aa)								0.569
	PLP8	PATATIN-like protein 8 (526 aa)								0.569
	PLP9	PATATIN-like protein 9 (384 aa)								0.569
	PLP6	PATATIN-like protein 6 (499 aa)								0.569
	PLA2A	phospholipase A 2A (407 aa)								0.569

Functional enrichments in the network

Biological Process (GO)			
pathway ID	pathway description	count in gene set	false discovery rate
GO:0016042	lipid catabolic process	8	5.46e-12
GO:0006629	lipid metabolic process	9	1.22e-09
GO:0006952	defense response	7	3.16e-05
GO:0044238	primary metabolic process	10	0.00354
GO:0044699	single-organism process	10	0.0037
GO:0071704	organic substance metabolic process	10	0.00408
GO:0050896	response to stimulus	8	0.0107
GO:0010311	lateral root formation	2	0.0249
GO:0048513	organ development	4	0.0306
GO:0010102	lateral root morphogenesis	2	0.0373

(less ...)

Molecular Function (GO)			
pathway ID	pathway description	count in gene set	false discovery rate
GO:0047372	acylglycerol lipase activity	3	1.21e-06
GO:0016298	lipase activity	4	3.51e-05
GO:0016787	hydrolase activity	8	8.82e-05
GO:0004620	phospholipase activity	3	0.000517
GO:0003824	catalytic activity	10	0.000652

▲ node1	node2	node1 accession	node2 accession	node1 annotation	node2 annotation	score
AT5G43590	NTMC2T4	AT5G43590.1	AT3G61050.1	acyl transferase/acyl hydrolase...	calcium-dependent lipid-bindin...	0.569
AT5G43590	PLP6	AT5G43590.1	AT2G39220.1	acyl transferase/acyl hydrolase...	PATATIN-like protein 6	0.407
AT5G43590	PLP8	AT5G43590.1	AT4G29800.2	acyl transferase/acyl hydrolase...	PATATIN-like protein 8	0.430
AT5G43590	PLP9	AT5G43590.1	AT3G63200.1	acyl transferase/acyl hydrolase...	PATATIN-like protein 9	0.421
NTMC2T4	AT5G43590	AT3G61050.1	AT5G43590.1	calcium-dependent lipid-bindin...	acyl transferase/acyl hydrolase...	0.569
NTMC2T4	PLA2A	AT3G61050.1	AT2G26560.1	calcium-dependent lipid-bindin...	phospholipase A 2A	0.569
NTMC2T4	PLP1	AT3G61050.1	AT4G37070.2	calcium-dependent lipid-bindin...	patatin-like protein	0.569
NTMC2T4	PLP4	AT3G61050.1	AT4G37050.1	calcium-dependent lipid-bindin...	PATATIN-like protein 4	0.569
NTMC2T4	PLP5	AT3G61050.1	AT4G37060.2	calcium-dependent lipid-bindin...	PATATIN-like protein 5	0.569
NTMC2T4	PLP6	AT3G61050.1	AT2G39220.1	calcium-dependent lipid-bindin...	PATATIN-like protein 6	0.569
NTMC2T4	PLP8	AT3G61050.1	AT4G29800.2	calcium-dependent lipid-bindin...	PATATIN-like protein 8	0.569
NTMC2T4	PLP9	AT3G61050.1	AT3G63200.1	calcium-dependent lipid-bindin...	PATATIN-like protein 9	0.569
NTMC2T4	T3A4.8	AT3G61050.1	AT2G46700.1	calcium-dependent lipid-bindin...	CDPK-related kinase 3; May pla...	0.596
NTMC2T4	THAS1	AT3G61050.1	AT5G48010.2	calcium-dependent lipid-bindin...	thalianol synthase 1; Converts ...	0.829
PLA2A	NTMC2T4	AT2G26560.1	AT3G61050.1	phospholipase A 2A	calcium-dependent lipid-bindin...	0.569
PLP1	NTMC2T4	AT4G37070.2	AT3G61050.1	patatin-like protein	calcium-dependent lipid-bindin...	0.569
PLP4	NTMC2T4	AT4G37050.1	AT3G61050.1	PATATIN-like protein 4	calcium-dependent lipid-bindin...	0.569
PLP4	PLP9	AT4G37050.1	AT3G63200.1	PATATIN-like protein 4	PATATIN-like protein 9	0.434
PLP5	NTMC2T4	AT4G37060.2	AT3G61050.1	PATATIN-like protein 5	calcium-dependent lipid-bindin...	0.569
PLP5	PLP6	AT4G37060.2	AT2G39220.1	PATATIN-like protein 5	PATATIN-like protein 6	0.418

▲ node1	node2	node1 accession	node2 accession	node1 annotation	node2 annotation	score
PLP5	PLP8	AT4G37060.2	AT4G29800.2	PATATIN-like protein 5	PATATIN-like protein 8	0.426
PLP6	AT5G43590	AT2G39220.1	AT5G43590.1	PATATIN-like protein 6	acyl transferase/acyl hydrolase/ly...	0.407
PLP6	NTMC2T4	AT2G39220.1	AT3G61050.1	PATATIN-like protein 6	calcium-dependent lipid-binding d...	0.569
PLP6	PLP5	AT2G39220.1	AT4G37060.2	PATATIN-like protein 6	PATATIN-like protein 5	0.418
PLP8	AT5G43590	AT4G29800.2	AT5G43590.1	PATATIN-like protein 8	acyl transferase/acyl hydrolase/ly...	0.430
PLP8	NTMC2T4	AT4G29800.2	AT3G61050.1	PATATIN-like protein 8	calcium-dependent lipid-binding d...	0.569
PLP8	PLP5	AT4G29800.2	AT4G37060.2	PATATIN-like protein 8	PATATIN-like protein 5	0.426
PLP9	AT5G43590	AT3G63200.1	AT5G43590.1	PATATIN-like protein 9	acyl transferase/acyl hydrolase/ly...	0.421
PLP9	NTMC2T4	AT3G63200.1	AT3G61050.1	PATATIN-like protein 9	calcium-dependent lipid-binding d...	0.569
PLP9	PLP4	AT3G63200.1	AT4G37050.1	PATATIN-like protein 9	PATATIN-like protein 4	0.434
T3A4.8	NTMC2T4	AT2G46700.1	AT3G61050.1	CDPK-related kinase 3; May play a...	calcium-dependent lipid-binding d...	0.596
THAS1	NTMC2T4	AT5G48010.2	AT3G61050.1	thalianol synthase 1; Converts oxi...	calcium-dependent lipid-binding d...	0.829

NTMC2T4- FUNCTIONS IN: lipid binding; INVOLVED IN: biological_process
unknown; LOCATED IN: plasma membrane; EXPRESSED IN: male gametophyte,
cultured cell, callus, pollen tube; EXPRESSED DURING: M germinated pollen stage;
CONTAINS InterPro DOMAIN/s: C2 membrane targeting protein (InterPro:IPR018029),
C2 calcium/lipid-binding domain, CaLB (InterPro:IPR008973), C2 region
(InterPro:IPR020477), C2 calcium-dependent membrane targeting (InterPro:IPR000008);
BEST Arabidopsis thaliana protein match is: Calcium-dependent lipid-binding (CaLB

domain) family protein (TAIR:AT3G61030.1). /db_xref="Araport:[AT3G61050](#)" ,
/db_xref="GeneID:[825277](#)", /db_xref="TAIR:[AT3G61050](#)" (Berardini, Reiser et al.
2015).

CURRICULUM VITAE

NAME: Venkata Swathi Kuppireddy

ADDRESS: Department of Biology
University of Louisville
Louisville, KY 40208

EMAIL: mail2swathi.k@gmail.com

EDUCATION

PhD, Molecular, Cellular and Developmental Biology 2013- 2018

University of Louisville, KY (3.87/4.0)

MS, Molecular, Cellular and Developmental Biology 2013-2016

University of Louisville, KY

Bachelor of Technology, Biotechnology 2005-2009

Jawaharlal Nehru Technological University, India

PUBLICATIONS

- **Kuppireddy, V. S.**, Uversky, V. N., Toh, S. S., Tsai, M. C., Beckerson, W. C., Cahill, C., Carman, B., and Perlin, M. H. (2017) Identification and Initial Characterization of the Effectors of an Anther Smut Fungus and Potential Host Target Proteins. *International journal of molecular sciences* **18**. doi: 10.3390/ijms18112489.
- Article published in Special Issue reprint Book _ Plant Microbe Interaction 2017, <https://www.mdpi.com/books/pdfview/book/855>, October, 2018.

ORAL PRESENTATIONS

- Identifying effectors for a biotrophic fungal plant pathogen, Bochum, Germany, 2016.

- Characterizing host target proteins for a biotrophic fungal plant pathogen, 102nd Kentucky Academy of Sciences meeting, KY, 2016.
- Identification of host targets for effectors from the anther smut fungus, *Microbotryum lychnidis-dioicae*, 29th Fungal Genetics Conference, Asilomar, CA, 2017.
- A Tale of Smut, Sex Change and STDs, 3-MT Thesis Competition, UofL, KY, 2017.
- A Tale of Smut, Sex Change and STDs, 3-MT Thesis Regional Competition, Fayetteville, AR, 2018.

HONORS/ AWARDS/ GRANTS

- **Dean's Citation Award**, UofL, 2018
- **Graduate Student Spotlight** (Spotlighted in the 'The Advocate' spring 2018 newsletter), UofL, 2018
- **Bill Furnish Award for Excellence in Teaching of Biology**, awarded to an 'outstanding Instructor', UofL, 2018
- **1st place** in Oral Presentation at 3-Minute Thesis Graduate Oral Competition at UofL, 2017
- **People's Choice Award** in Oral Presentation at 3-Minute Thesis Graduate Oral Competition at UofL, 2017
- **1st place** in an Oral presentation at the 102nd Annual Kentucky Academy of Science meeting, 2016
- **Finalist** in 3-MT Thesis Regional Competition, Fayetteville, AR, 2018
- **Dissertation Completion Award**, UofL, KY, 2018
- **GSC Research Grant**, UofL, KY, 2017
- **GSC Travel Grant**, UofL, KY, 2017
- **29th Fungal Genetics Conference Travel Grant**, Asilomar, CA, 2017
- **Ustilago/ Smut convergence Travel Award**, Asilomar, CA, 2017
- **Arts and Sciences research and creative grant** award, UofL, KY, 2016
- **GNAS Intramural grant**, UofL, KY, 2016

TEACHING EXPERIENCE

Graduate Teaching Assistant for Biology 258- Introduction to Microbiology lab,
Department of Biology, UofL, 2013-2018.

VOLUNTEER EXPERIENCE

Louisville Regional Science Fair Judge for Molecular & Cellular Biology division, 2018.