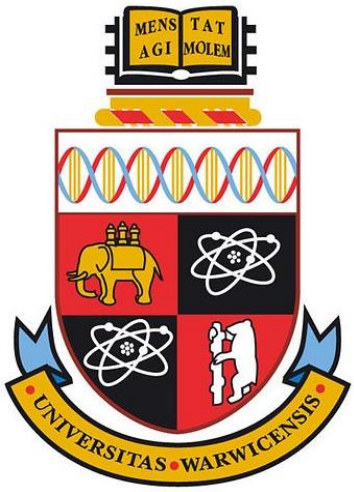


The *Photorhabdus* Virulence Cassette: The Role of Payload Tags in Loading Bacterial Syringes

by

Rhys David Evans

Doctoral Thesis

for the degree of

Doctor of Philosophy

School of Life Sciences

University of Warwick

Supervisors: Prof. Nicholas R. Waterfield and Dr. John James

August 2023

Contents

List of Figures ..... **viii**

List of Tables ..... **xi**

Acknowledgements ..... **xii**

Declaration ..... **xv**

Abstract ..... **xvi**

List of Abbreviations ..... **xvii**

1. Introduction **1**

1.1 *Photorhabdus* **3**

1.1.1 The Bacteria-Nematode Life Cycle 5

1.1.2 *Photorhabdus* as a Pathogen 9

1.1.2.1 Tc toxins 9

1.1.3 The *Photorhabdus* Pangenome 10

1.1.4 *Photorhabdus* in Biotechnology 11

1.2 Contractile Protein Injection Systems **14**

1.2.1 Secretion Systems and Their Effectors 14

1.2.2 Extracellular Contractile Injection Systems (eCIS) 17

1.2.2.1 Anti-feeding Prophage (Afp and AfpX) 22

1.2.2.2 Metamorphosis-Associated Contractile (MAC) and T6SS[iv] 24

1.2.2.3 Antibacterial eCIS 25

1.2.2.4 Tailocins as a Model for Contraction 26

1.2.3 Other CISs: BIS, tCIS, and AlgoCIS 27

1.3 The *Photorhabdus* Virulence Cassette **30**

1.3.1 Encoding the PVC 31
1.3.1.1 Contractile Syringe: Tube and Sheath 32
1.3.1.2 Stabilisation: The Pvc16 Cap 38
1.3.1.3 The Needle: Spike, Tip, and Baseplate 39
1.3.1.4 Cell Binding: Tail Fibres 41
1.3.1.5 Structural PVC Gene Homology 42
1.3.1.6 The Elusive Components: Pvc6, 14, and 15 45
1.3.2 PVC Toxin Payloads 47
1.3.2.1 Pnf 48
1.3.2.2 LopT: a Class II PVC Effector 52
1.3.2.3 PaTox: a Class III PVC Effector 52
1.3.2.4 Cyclomodulins 53
1.3.2.4.1 $RRSP_{Pa}$ 55
1.3.2.4.2 Cif 55
1.3.2.5 Pdp1 / SepC-like 58
1.3.2.6 PVC Units 1-4 59
1.3.3 Protein Delivery Systems in Biomedicine 59
1.3.4 AAA+ ATPase Superfamily 60
1.3.4.1 Classification of AAA+ ATPases 61
1.3.4.2 ATP Hydrolysis Modes of Action 66
1.3.4.3 Pvc15 as an AAA+ ATPase 68
1.4 Aims and Summary 70
2. Materials and Methods 71
2.1 Strains 71
2.1.1 Preparation of Chemically Competent Cells 73
2.1.2 Preparation of Glycerol Stocks 74
2.2 Media and Stocks 74
2.2.1 Culture Media 74
2.2.1.1 Lysogeny Broth (Luria-Bertani: LB) 74
2.2.1.2 SOC 74
2.2.1.3 Antibiotics and Supplements 75

2.3 Cell Culture Techniques 76
2.3.1 Biological Replicates for Cell Culture 76
2.3.2 Mammalian Cell Cultures 76
2.3.2.1 Jurkat T Cells and Raji B Cells 76
2.3.2.2 HEK293T 77
2.3.2.3 THP-1 Monocytes 77
2.3.3 Insect Cell Culture 78
2.3.4 Resazurin Assay: Measuring The Effects of PVCs on Cell Lines 78
2.3.4.1 Resazurin Cell Growth Calibration Curve 79
2.3.4.2 Analysis and Visualisation of Plate Reader Results 80
2.3.5 Transient Transfection 82
2.3.5.1 Measuring Protein Degradation 82
2.3.6 Generating Stable Cell Lines – Transduction 83
2.4 Molecular Techniques – Nucleic Acids 84
2.4.1 Plasmids 85
2.4.2 Polymerase Chain Reaction 91
2.4.2.1 Primers 92
2.4.2.1.1 Primer Design 96
2.4.2.2 Q5 PCR 97
2.4.2.3 *Taq* and Colony PCR 98
2.4.2.4 Overlap Extension PCR 98
2.4.2.5 DNA clean-up 99
2.4.3 Agarose Gel Electrophoresis 99
2.4.3.1 Gel Extraction 100
2.4.4 Cloning Techniques 100
2.4.4.1 Restriction Digestion 100
2.4.4.2 Ligation 100
2.4.4.3 Transformation of *E. coli* 101
2.4.4.4 Plasmid Extraction 102
2.4.5 Sequencing 102
2.4.5.1 Di-deoxy-chain-termination (Sanger) Sequencing 102

2.4.5.2 Oxford Nanopore Whole-Plasmid Sequencing 102
2.5 Molecular Techniques – Proteins 103
2.5.1 SDS-PAGE 103
2.5.1.1 Eukaryote Lysate Preparation 103
2.5.1.2 Staining 104
2.5.1.3 Western blot (WB) 105
2.5.1.3.1 Quantification 105
2.6 Microscopy and Computation 106
2.6.1 Statistical Analysis 106
2.6.2 Sample Preparation for Fluorescence Microscopy 106
2.6.2.1 Image Analysis 107
2.6.3 Transmission Electron Microscopy (TEM) 107
2.6.4 Molecular Visualisation 108
2.6.5 Generating Figures and Schematics 109
2.6.6 Sequence Alignments 109
3. Investigating PVC Effectors and ATPases 110
3.1 Introduction 110
3.1.1 The Bacterial eCIS Effectome 110
3.1.2 Predicting Structures of PVC Effectors and the Pvc15 ATPase 111
3.1.3 PVC Baseplate Carbohydrate Binding Modules 111
3.2 Aims 112
3.3 Methods 112
3.3.1 Glycowork/LectinOracle for *in silico* Protein-Glycan Predictions 112
3.3.2 Using the dbeCIS for Analysing Putative Payload Functions 112
3.3.3 Using AlphaFold Packages to Predict Protein Structure 113
3.3.3.1 pLDDT Calculation from MoLPC mpDockQ Output 114
3.3.3.2 PyMol: Monoplanar Angles of Two Centres of Mass 114
3.4 Results 116
3.4.1 Distribution of Functions of eCIS-Downstream ORFs 116
3.4.1.1 Ubiquity of Leader Sequences in eCIS-downstream ORFs 123

3.4.1.2 Assessment of *Photorhabdus* Leader Sequence Diversity 124
3.4.2 Effector Predictions 129
3.4.2.1 Pnf and Cif Effector Homology 129
3.4.2.2 Comparison of LopT-PVC with LopT-TSS 135
3.4.3 Predicting the Structure and Function of Pvc15 138
3.4.3.1 Inferring Hexameric Assemblies of Pvc15 151
3.4.4 Inferring the Binding Capacity of the PVC*lopT* Chassis 156
3.5 Discussion 163
3.5.1 Implications of Functional Diversity of eCIS-associated ORFs 163
3.5.2 Leader Sequence Homologies 165
3.5.3 Implications of Machine Learning Methodologies 166
3.6 Conclusions 166
4. Modulating PVC*pnf* Payloads by Encoding Native Leader Sequences 168
4.1 Introduction 168
4.2 Aims 168
4.3 Methods 169
4.3.1 PVC Purification 169
4.3.1.1 Expression and Harvesting 169
4.3.1.1.1 IPTG Gradient 169
4.3.1.2 Lysis 170
4.3.1.3 Lysate Preparation 171
4.3.1.3.1 Resin Immunoprecipitation 171
4.3.1.4 Formaldehyde Cross-linking 172
4.3.2 Cloning Heterologous Payloads into a PVC-compatible Vector 173
4.4 Results 179
4.4.1 Leader Sequences as Packaging Signals for Payloads 179
4.4.2 *Serratia entomophila* Afp17 is Toxic to *E. coli* 182
4.4.3 Verifying Injection of Payloads by their Effect on Respiration 183
4.5 Discussion 188

4.5.1 Heterologous Payloads....188
4.5.2 Verifying Injection by Observing Cellular Effects....188
4.5.3 PVC Production and 'Regulatory' Genes....190
5. Uncovering Pvc15-Leader Sequence Interactions....191
5.1 Introduction....191
5.2 Aims....191
5.3 Methods....192
5.3.1 Bacterial Lysate Protein Abundance and Degradation....192
5.3.2 ATPase assay....192
5.4 Results....194
5.4.1 Effects of Leader Truncation on Pnf Payload Stability....198
5.4.2 PVC*pnf* Loading Depends on Pvc15 and the Leader Sequence....208
5.4.3 The Effect of Mutations on the Roles of Pvc15....210
5.4.4 Pvc15 Domain D2 Deletion Reduces Pnf's Optimal Abundance....212
5.4.4.1 Synergistic Interference of Arabinose with IPTG Induction....215
5.4.5 ATPase Activity is Depleted in Pvc15 Mutants....216
5.4.6 ATPase Competence is Required for Payload Loading....218
5.4.7 Additional PVC*pnf* Loading Biological Replicates....220
5.5 Discussion....222
5.5.1 Cif Leader Interactions....223
5.5.2 Interpretation of Protein Abundance Experiments....223
5.5.3 Alternative Interpretations of Pnf LS-Pvc15 Interactions....224
5.5.4 Additional Binding Roles of PVC*pnf* Components....230
6. Conclusions and Future Directions....232
7. Bibliography....235
8. Appendices....249

List of Figures

Figure 1.1 | Maximum-likelihood phylogenetic tree of *Photorhabdus* 4
Figure 1.2 | Life cycle of the *Photorhabdus-Heterorhabditis* EPN complex. 6
Figure 1.3 | *Photorhabdus* species have an open pangenome. 11
Figure 1.4 | Tomography of T6SS in *M. xanthus* spans the cell cytosol. 15
Figure 1.5 | T6SS effectors are recruited to VgrG, PAAR, and Hcp. 16
Figure 1.6 | Genetic and Structural Schematic for PaPVC*pnf* 18
Figure 1.7 | eCIS states change during and after injection. 19
Figure 1.8 | Conserved structure of tailed contractile systems. 21
Figure 1.9 | Anti-feeding prophage (Afp) particles in TEM. 22
Figure 1.10 | The MAC Amorphous Array. 24
Figure 1.11 | T6SS^{iv} form ordered membrane-bound arrays. 25
Figure 1.12 | Gene cassettes of PVC classes I-III. 32
Figure 1.13 | Pvc1 hexamers have complementary electrostatic potentials. 33
Figure 1.14 | Molecular overview of the baseplate-sheath-tube complex. 35
Figure 1.15 | PVC sheath conformations are altered by contraction. 37
Figure 1.16 | Schematic of 'interface 1' of the PVC baseplate. 40
Figure 1.17 | *Photorhabdus* effectors modify Rho-subfamily GTPases. 51
Figure 1.18 | PVC effectors modulate the cell cycle. 54
Figure 1.19 | Cif causes the accumulation of regulators. 57
Figure 1.20 | AAA+ ATPase clades 1-7 conserved supersecondary motifs. 65
Figure 2.1 | Schematic of inputs to the Plate Reader Analysis R script. 81
Figure 3.1 | Predicted functions of lineage Ia and Ib putative eCIS payloads. 120
Figure 3.2 | Ubiquity of leader sequences in lineage Ia putative eCIS payloads. 123
Figure 3.3 | Putative PVC effector LSs show conserved motifs within clustered groups. 128
Figure 3.4 | AlphaFold2 is an accurate predictor of effector structure. 130

Figure 3.5 | Local multiple sequence alignment of Pnf and CNF catalytic domain homologues and Cif homologues. 132
Figure 3.6 | Pnf and CNF catalytic domains form similar protein structures. 133
Figure 3.7 | Cif homologues form similar protein structures and encode LSs. 134
Figure 3.8 | N-terminal disorder predicts LSs through low pLDDT. 135
Figure 3.9 | LopT and YopT local sequence alignments. 136
Figure 3.10 | LopT and YopT form similar protein structures and encode LSs. 137
Figure 3.11 | Local sequence alignment of *Photorhabdus* Pvc15 ATPase homologues. 142
Figure 3.12 | Pvc15 is a Cdc48-like type II ATPase precluded by a single N-domain. 144
Figure 3.13 | Superimposition of Pvc15 D1 and D2 domains reveals tertiary similarities. 146
Figure 3.14 | Pvc15 ATPase motifs are highly conserved in D2. 148
Figure 3.15 | Pvc15 homologues have vast structural similarities. 151
Figure 3.16 | A complete prediction of the Pvc15 hexamer. 154
Figure 3.17 | DALI server results suggests the Pvc15 N-domain may bind nucleic acids. 156
Figure 3.18 | DALI server results suggests Pvc11$_{PAU lopT}$ contains a carbohydrate binding module within the extended domain III. 158
Figure 3.19 | Pvc11$_{lopT}$ domain IIIb is rich in β-sheet supersecondary structures. 161
Figure 3.20 | The *P. luteoviolacea* HI MAC has an effector encoded within the operon. 165
Figure 3.21 | Alignment of LSs of PaTox and ATCC43949 Pnf. 166
Figure 4.1 | Schematic of the pBAD-PVC*pnf* and pVTRa vectors. 173
Figure 4.2 | Workflow schematic of plasmid construction. 177
Figure 4.3 | Western blot of aliquots of the heterologous payload-PVC*pnf* co-purification process. 181
Figure 4.4 | *S. entomophila* Afp17 is toxic when induced in *E. coli*. 183
Figure 4.5 | *Drosophila* S2 cells are sensitive to native PVC*pnf*. 185
Figure 4.6 | Optimal starting cell numbers using resazurin calibration growth curves. 186
Figure 4.7 | PVC*pnf* perturbs respiration of mammalian immune cell lines. 187
Figure 5.1 | Whole-plasmid sequencing confirms the PVC*pnf*Δ*pvc15* vector. 194
Figure 5.2 | Pvc15 is required for loading and functional injection of the cognate payload. 196
Figure 5.3 | Time course of Pnf abundance post-induction. 199
Figure 5.4 | LS C-terminal truncations differentially affect the stability of the payload. 200
Figure 5.5 | Collected data for Pnf stability and degradation experiments. 204
Figure 5.6 | N-terminal Pnf50 truncations enhance Pnf abundance when co-expressed with PVC*pnf*. 205

Figure 5.7 | Pvc15 is the sole PVC*pnf* chaperone protein for its native payload. 207
Figure 5.8 | An intact LS and Pvc15 are required for PVC*pnf*-payload association..... 209
Figure 5.9 | pVTRaDuet vector: simultaneous Pnf-Myc and Pvc15 expression. 211
Figure 5.10 | Pnf abundance is depleted when the Pvc15 D2 domain is deleted. 212
Figure 5.11 | Co-expression of PVC*pnf*Δ*pvc15* does not affect Pvc15's role in Pnf stabilisation. 214
Figure 5.12 | Arabinose enhances IPTG-induced Pnf expression. 215
Figure 5.13 | Standard curve for conversion of OD (650 nm) to nmol of phosphate. 217
Figure 5.14 | D2 is the sole functional ATPase domain with PVC*pnf*-dependent activity. .. 217
Figure 5.15 | PVC*pnf* cargo loading requires ATPase competent Pvc15. 220
Figure 5.16 | Biological replicates for statistical analysis in previous loading figures. 221
Figure 5.17 | Inferred mass action kinetics model of Pnf abundance. 226
Figure 5.18 | RNA secondary structures of wild-type and C-terminal truncated LSs. 229

List of Tables

Table 1.1 | *Photorhabdus-Heterorhabditis* entomopathogenic complexes vary by species. 8
Table 1.2 | Contractile injection system homologues. 43
Table 2.1 | *Escherichia coli* strains used. 72
Table 2.2 | Competent cell preparation solutions. 73
Table 2.3 | Recipes of solutions used for bacterial growth. 74
Table 2.4 | Antibiotics and supplements. 75
Table 2.5 | Plasmid backbones and their derivations used. 85
Table 2.6 | PCR setup and reagents. 91
Table 2.7 | PCR conditions. 92
Table 2.8 | DNA primers used. 93
Table 3.1 | Lineage Ia operon numbers by strain. 121
Table 3.2 | Lineage Ib operon numbers by species. 122
Table 4.1 | Gel loading samples for assaying PVC purification. 172
Table 4.2 | Validations of natural and heterologous PVC*pnf* payloads. 179

Acknowledgements

Firstly, I thank my lab colleagues Amy Godfrey, Elena Carter, Shathviga Manoharan, and Max Addison, as well as my many colleagues in the adjoining labs for making me feel so welcome every day of the last 3 years. I have been privileged to have worked alongside such greatly aspiring scientists and supportive individuals; I wish you all the very best.

I also thank the staff at Nanosyrinx Ltd for their continued collaboration and guidance: Dr. Joe Healey for insights into bioinformatics and experimental design early on in the project, Dr. Alexia Hapeshi, Dr. Sabine Bowman-Grahl, and Dr. Mohsina Bashir for their guidance and assistance for experiments throughout my research, as well as other members of the research team who assisted me in the final stages of experiments.

I thank Saskia Bakker and Ian Hands-Portman for their permission and assistance in the use of the transmission electron microscope within the University of Warwick advanced bioimaging platform facility. I also thank the staff of the Scientific Computing RTP (SCRTP) for their valuable technical support and usage of their high-performance computer cluster. I also thank Lesley Harvey-Smith and the School of Life Sciences tech team; without them, none of this PhD would have been possible.

Thanks also go to my advisory panel: Prof. Yin Chen and Dr. Chrystala Constantinidou for their support throughout the PhD and coming along to my poster events and presentations. They always provided new and

insightful questions and threw their interest into the project to provide great forums of discussion (and to read my rather lengthy reports).

I thank the wonderful staff of the MIBTP: Prof. Vardis Ntoukakis, Kerry Davies, and Nikki Glover. They have been the go-to voices of support and have made the running of the program as smooth as possible. I also thank my MIBTP peers for their engagement, communication, and support throughout the years and even throughout exceptional circumstances.

I also thank everyone involved with the Postgraduate Certificate for Transferable Skills in Science (PGCTSS); the modules have been a great way to keep on track with the PhD and encourage me to look to additional methods of learning, such as ethics and intellectual property. I have also used this unique opportunity to advise the 2023 Warwick international Genetically Engineered Machines (iGEM) competition team and act as a co-supervisor for the World Congress in Undergraduate Research (WCUR).

Next, I thank Dr. Chris Corre, Dr. Corinne Hanlon, Warwick Integrative Synthetic Biology (WISB) centre, and the staff at the iGEM competition for the invaluable opportunity they gave me to get stuck into a research project working as part of a team. The skills I accumulated during the summer of 2018 have been crucial for the work I have done in this PhD and I cannot recommend this unique opportunity for undergraduate students enough. It is only fitting, then, that I thank all of the Warwick iGEM team of 2018.

I acquired additional lab skills from my 3rd-year dissertation project done in the lab of Dr. Munehiro Asally whom I thank for providing a great lab environment to try new ideas and lead an exciting independent project in such a small amount of time.

Huge thanks go to my supervisors Prof. Nick Waterfield and Dr. John James. Both have been brilliant inspirations for what makes a great scientist and each has provided unique ideas and perspectives of how the project can progress. Even in areas outside of their expertise they have played an

important part in pushing my project in the right direction. My 3-month placement with John as part of the MIBTP training year was an extraordinary time of seeing how uniquely engineered hardware devices could be used for optogenetics and the wise insights he provided throughout changed my perspectives of various lab techniques and made me think more critically of how to interpret results of my own as well as others.

I also thank my undergraduate personal tutor Prof. Philip Young who first mentioned to me that a PhD would be "right up my street" and told me to apply for one rather than only pursuing a masters. He gave me the confidence that I needed to believe that making that jump was possible and that the PhD was only as daunting as what you believe it to be. Thank you, Phil.

Profound gratitude also goes to my partner for her continued love and support throughout this research venture and whose encouragement has given me confidence in my abilities each and every day. I also thank my close friends Jack Lawrence, Jonny Whiteside, and Valentin Dospinescu who have made my higher education experience so memorable. Finally, many thanks go to my family for their invaluable support, encouragement, and guidance throughout my full-time education and who made my work possible.

Thank you again to all of those who have provided feedback and support throughout my PhD journey; despite the uncertainty from recent years, you have made the last 3 years some of my favourite (so far).

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by the author.

Some methods for protein purification were done in collaboration with Nanosyrinx Ltd. The finer details of these methods have been omitted due to potential conflicts of interest with their patented works.

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) and University of Warwick funded Midlands Integrative Biosciences Training Partnership (MIBTP) [grant number BB/M01116X/1].

Abstract

The *Photorhabdus* virulence cassette (PVC) is an elegant, multi-protein, contractile nanostructure which injects bioactive polypeptide 'payloads' across eukaryotic cell membranes. N-terminal leader sequences (LSs) from native payloads are necessary for loading and can also associate heterologous proteins into the hollow tube of the PVC ready for injection. Previous works identified an accessory protein encoded in the PVC operon, Pvc15, to be responsible for payload loading via an interaction with the LS. This thesis builds on these findings and concludes that Pvc15 also uses the LS to stabilise payloads and that the Pvc15 hexamer's ATPase activity is necessary for association of the LS-tagged payload to the PVC chassis.

Bioinformatic analyses in this work found that Pvc15 encodes an elusive N-domain and two tandem AAA domains: D1 and D2. *E. coli* carrying plasmids encoding the PVC operon and the LS-tagged payload were used to investigate the effects of mutations in either the LS or Pvc15 using quantitative western blot. Domain D2 was found to stabilise the Pnf payload via the LS. Given that only mutation of D2 was necessary to abolish the ATPase, D2 was identified as the sole functional ATPase domain. In addition, truncations to the LS of less than 50 amino acids conferred payload stability and abolished its capacity to load into the PVC. Overall, this thesis provides insight into the interactions between Pvc15 and the LS. Elucidating the nature of PVC loading may enable the development of PVCs as a therapeutic synthetic biology protein delivery toolkit in the near future.

List of Abbreviations

aa	amino acid
AAA+	ATPase Associated with diverse cellular Activities
ADP	Adenosine DiPhosphate
Amp	Ampicillin
ATCC	American Type Culture Collection
Afp	Anti-feeding prophage
ATP	Adenosine TriPhosphate
BDBH	BiDirectional Best Hits
BLAST	Basic Local Alignment Search Tool
bp	base pairs
Chlor	Chloramphenicol
Cif	Cycle-inhibiting factor
CNF	Cytotoxic Necrosis Factor
COG	Cluster(s) of Orthologous Groups
CPU	Central Processing Unit
Cryo-EM	Cryo-Electron Microscopy
DMSO	Di-Methyl-SulphOxide
DNA	DeoxyriboNucleic Acid
DTT	DiThioThreitol
DUF	Domain of Unknown Function
(e)CIS	(Extracellular) Contractile Injection System
EDTA	Ethylene-Diamine-Tetra-Acetic acid

EPN	EntomoPathogenic Nematode
ER	Endoplasmic Reticulum
Fwd	Forward primer
GDP	Guanosine DiPhosphate
GFP	Green Fluorescent Protein
GFP(1-10)	GFP β-sheets 1-10 (large component)
GFP11	GFP β-sheet 11 (small component)
GPU	Graphics Processing Unit
GTP	Guanosine TriPhosphate
HEPES	4-(2-Hydroxy-Ethyl)-1-Piperazine-Ethane-Sulphonic acid
IJ	Infective Juvenile nematode
IMAC	Immobilised Metal ion Affinity Chromatography
IP	ImmunoPrecipitation
IPTG	IsoPropyl β-D-1-ThioGalactopyranoside
ISM	Initiator-Specific Motif
kDa	kiloDaltons
LacZ	β-galactosidase
LacZα	LacZ small α component
LacZω	LacZ large ω component
LB	Luria Bertani broth
LopT	*Photorhabdus* YopT-like toxin
LS	Leader Sequence
MAC	Metamorphosis Associated Contractile (eCIS)
MARTX	Multifunctional Auto-processing RTX protein
Mcf	'Makes caterpillars floppy' toxin
MCS	Multiple Cloning Site
mRFP	monomeric Red Fluorescent Protein
MSA	Multiple Sequence Alignment

NEB	New England Biolabs
OMCL	Ortho-Markov CLustering algorithm
ORF	Open Reading Frame
PAAR	Proline-Alanine-Alanine Repeat protein
PAE	Predicted Alignment Error
PaTox	*Photorhabdus asymbiotica* Toxin (previously LumT)
PCR	Polymerase Chain Reaction
PDB	Protein DataBank
pLDDT	predicted Local Distance Difference Test
PL1/PL2	Pore Loop 1 / Pore Loop 2
PLTS	Phage-Like protein Translocation System
Pnf	*Photorhabdus* necrosis factor
PVC	*Photorhabdus* Virulence Cassette
PVDF	PolyVinyliDene Fluoride
Rev	Reverse primer
RMSD	Root Mean Square Deviation
RNA	RiboNucleic Acid
RRSP	Ras/Rap1-Specific endoPeptidase domain
Rtx	Repeats-in-toxin toxin
SA-β-Gal	Senescence-Activated β-Galactosidase
SDS-PAGE	Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SOC	Super Optimal media with Catabolite repression
SP	Signal Peptide
T*x*SS	Type *x* Secretion System
Ta	Annealing temperature
TBE	Tris-Borate-EDTA

TEM	Transmission Electron Microscopy
Tm	Melting temperature
TMP	Tape Measure Protein
WB	Western Blot

1. Introduction

In 1862, the Battle of Shiloh amassed around 20,000 casualties in one of the bloodiest battles of the American Civil War. During the cold and dark night some wounded soldiers noticed their open wounds glowing a blueish green. Intriguingly, soldiers who displayed these glowing wounds would have a higher chance of survival and lower rates of infection (Lamoureux, 2018). This phenomenon, termed the 'Angel's Glow', is explained by the presence of a bioluminescent, insect-pathogenic bacterium with a notable capability for antimicrobial production: *Photorhabdus luminescens*. This species cannot ordinarily cause infection at the human body temperature of 37 °C, but hypothermia likely enabled favourable conditions for growth whilst preventing further infection by more virulent pathogens.

141 years later, whole genome sequencing of *Photorhabdus luminescens* revealed a striking number of toxins, haemolysins, adhesins, lipases, proteases, and antibiotic-producing genes (Duchaud *et al.*, 2003). In addition, variants of prophage-like loci were identified across members of the genus, each encoding a different putative proteinaceous effector thought to play a role in insect pathogenicity; these were designated the '*Photorhabdus* Virulence Cassettes' (PVCs) (Ffrench-Constant *et al.*, 2000).

We now know far more about these proteinaceous effectors. The PVC is an elegant, nano-sized, phage-like 'syringe', hypothesised to be altruistically released from *Photorhabdus* after cell death (Becker *et al.*, 2022), which can deliver its loaded bioactive polypeptide payload into eukaryotic cells

independently from the producing cell (Vlisidou *et al.*, 2019). One can think of the PVC as the casing to a missile whereby its payload determines the choice of weapon upon successful delivery to the host's cells during infection. The PVC itself has a remarkable mode of action to deliver these payload proteins: using its tail fibres to anchor itself to the membrane, it contracts to inject its sharp needle, the spike protein, across the membrane and deliver its payload directly into the cytosol. The capacity for PVCs to translocate polypeptides in this way could serve as a tool to deliver active pharmacological or bioengineered polypeptides into patient cells in research and therapeutic applications in the future. However, it is not fully known how payload proteins are selected for association to the lumen of the PVC during nanosyringe assembly, nor the key roles of the few non-structural proteins encoded in the PVC operon.

1.1 *Photorhabdus*

The genus of *Photorhabdus* describes a remarkable genus of insect pathogens which express a panoply of secondary metabolites and insecticidal virulence factors, though some species have also adapted their metabolic pathways and entomopathogenic effectors for mammalian infection (Mulley *et al.*, 2015).

Photorhabdus are Gram-negative rod-shaped γ-proteobacteria from the family of opportunistic human pathogens of *Morganellaceae*. Despite having close genetic phylogeny, *Photorhabdus* occupies a range of environmental niches from soil and sediment to host tissues in animals and plants (Geller *et al.*, 2021). In addition, *Photorhabdus* is the only known bioluminescent terrestrial bacterium; lending to its name meaning 'glowing-rod' via its luciferase (Foucault *et al.*, 2010). Other bioluminescent bacteria are primarily marine-based, such as *V. fischeri* in its symbiosis with the bobtail squid (Miyashiro and Ruby, 2012).

Among the genus of *Photorhabdus* are 20 species, many of which were recently elevated from subspecies of only 4 species (Machado *et al.*, 2018, 2020; Abd-Elgawad, 2021). *Photorhabdus* speciation events have been identified through phylogenetic studies using specific genetic sites (Figure 1.1) (Machado *et al.*, 2018), as well as core and pangenomic methods (Rivera-Ramírez *et al.*, 2022).

Having first been isolated from a leg ulcer in the USA in 1989 (Farmer *et al.*, 1989), *P. asymbiotica* and *P. australis* are thought to be some of the only species capable of human infection at 37 °C, although there are less than 20 cases reported in the literature (Gerrard and Stevens, 2017). Infections typically occur in warm climates via soil (Gerrard *et al.*, 2004; Gerrard and Stevens, 2017). However, the incidence of infections in the USA and Australia is possibly under-reported since *Photorhabdus* is not on the databases of many common

automated clinical diagnosis machines and infections are often misdiagnosed as coliforms (Gerrard *et al.*, 2004; Weissfeld *et al.*, 2005; Mulley *et al.*, 2015). Furthermore, identifying photorhabdosis requires specialist knowledge but can be definitively diagnosed by 16S ribosomal RNA (rRNA) sequencing (Dutta *et al.*, 2018).

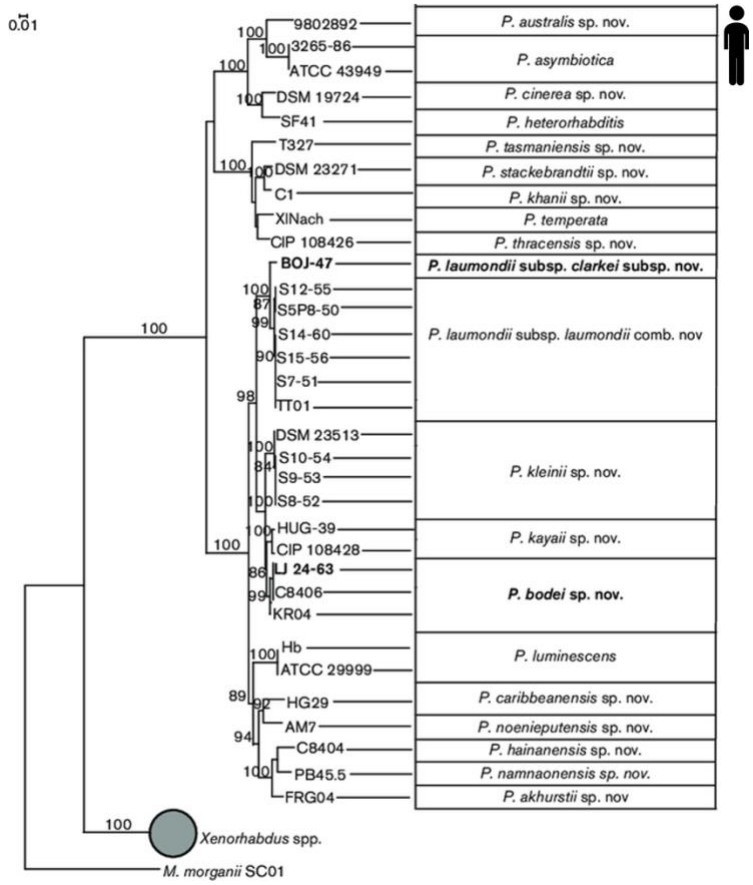


Figure 1.1 | Maximum-likelihood phylogenetic tree of *Photorhabdus*

Adapted from Machado *et al.*, 2018. *P. asymbiotica* and *P. australis* are the primary causative agents of human photorhabdosis in symbiosis with the *Heterorhabditis* nematode; most other strains are exclusive insect pathogens. Scale bar represents nucleotide bases per site across the 4318 nucleotide positions; bootstrap values are based on 100 replications.

Interestingly, the two European strains of *P. asymbiotica*, HIT and JUN, are unable to cause human infection; much like other strains, infection may be limited by their ability to infect the human host since their optimal growth

temperature is maintained at the insect body temperature of 34 °C (Mulley *et al.*, 2015). In a unique case in Texas documented by Dutta and colleagues in 2018, a newborn female was abandoned outside an apartment complex and admitted to hospital with hypothermia at 92.5 °F (~33.6 °C) with multiple insect bites on the extremities. Lesions on the spine were predicted to be an infection by *P. luminescens*, though further analysis of the strain is ongoing. This case may offer insight into how *Photorhabdus* has evolved for infection in humans and ground mammals as dead-end hosts.

1.1.1 The Bacteria-Nematode Life Cycle

Photorhabdus spp. live symbiotically in the anterior intestine of obligate entomopathogenic nematodes (EPNs): *Heterorhabditis* (Silva *et al.*, 2002), which are ubiquitous in soils around the world. The nematodes cannot develop in the absence of their cognate bacterial symbiont (Han and Ehlers, 2000); together, *Photorhabdus* serves as an essential ally to *Heterorhabditis* by quickly killing the insect host before it can mount an effective immune response as well as facilitating the bacteria-nematode life cycle (Forst *et al.*, 1997; Parihar *et al.*, 2022).

Figure 1.2 details a typical *Heterorhabditis-Photorhabdus* life cycle where free-living stage 3 infective juvenile nematodes (**1**) enter insects via the anal, oral, or tracheal routes (Sheets and Aktories, 2017) (**2**). Upon entry to the gut, the nematodes receive a yet-unidentified regurgitation signal in the insect haemolymph (**3**), causing the excretion of bacteria (as few as 50-200 bacteria can quickly kill the insect in between 20-72 hours) (Sheets and Aktories, 2017) (**4**). In addition, the regurgitation signal marks the start of the 'IJ recovery' development pathway where an IJ matures into a self-fertile adult hermaphrodite. Synergistically, the nematode and bacteria convert the organs and tissues of the cadaver into bacterial biomass to be consumed by the IJs, so the hermaphrodite can reproduce to lay around 300 eggs (Clarke, 2017) (**5**).

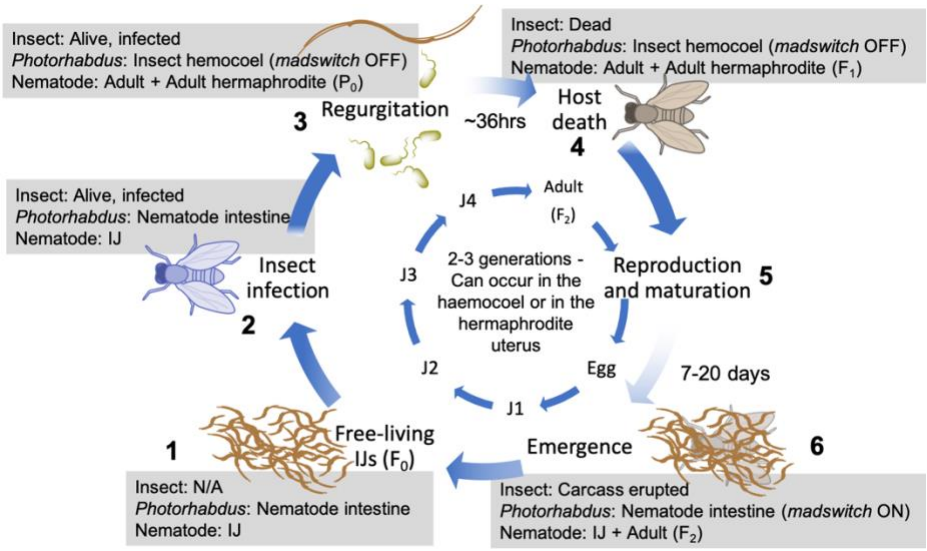


Figure 1.2 | Life cycle of the *Photorhabdus-Heterorhabditis* EPN complex.

This life cycle can complete in less than 24 hours in the case of wax moth larvae.

(1) Infective juvenile (IJ) nematodes living in soil bear *Photorhabdus* in their intestines and actively seek insect hosts to infect.

(2-3) Upon entry into an insect host via the anal, spiracle, or oral routes, contact with the host haemocoel triggers regurgitation of the bacteria.

(3) With the promoter of the *madswitch* in its 'off' orientation, *Photorhabdus* has an invasive, pathogenic (P) form. The regurgitation signal induces nematode maturation to dioecious adults and hermaphrodites (P_0) which produce more adults and some hermaphrodites (F_1).

(4) The bacteria produce a plethora of virulence factors which suppress the insect immune response, killing it typically around 36 hours post-infection.

(5) Exoenzymes from the bacteria bioconvert and liquefy the organs of the host. F_1 adults lay some of their eggs in the biomass to mature, facilitated by the secretion of bacterial 'food signals'. Some eggs are retained inside of the hermaphrodite adults' uteri, where offspring hatch and mature by progressively consuming the parent by endotokia matricida.

(6) After 2-3 generations of nematodes, the bacteria signal that food sources are low. This switches *Photorhabdus* to a mutualistic (M) phenotype: expressing Mad fimbriae to re-associate with IJ intestines and re-establish the EPN complex. Tens to hundreds of thousands of EPN complexes rupture the carcass 10-20 days post-infection.

Once eggs have hatched, juveniles undergo several moults (J1-J4; Figure 1.2) to adulthood. During this time, *Photorhabdus* maintain a promoter known as the *madswitch* in its OFF orientation; this promotes a virulent, secondary metabolite-producing pathogenic phenotype (the 'P' form) and enables IJ maturation via secretion of 'food signals' (Clarke, 2017). Some of the hermaphrodites' eggs are retained inside the uterus and hatch within the adult as *Photorhabdus*-colonised IJs. These feed off of the parent nematode in a process termed endotokia matricida (Ciche *et al.*, 2008); by the time the parent is killed, the offspring have matured into adults. *Photorhabdus* also facilitates maturation of its nematode symbiont; in addition to the food signals, crystal inclusion proteins CipA and CipB are crucial for the development of IJs in *Heterorhabditis* (You *et al.*, 2006).

Once the organs of the cadaver have been fully bioconverted and consumed by the nematode offspring, the bacteria re-associate with their vector nematodes. Inverting the promoter of the *madswitch* downregulates virulence and secondary metabolism whilst upregulating surface-localized Mad fimbriae that facilitate recolonisation of the nematode via adhesion to the maternal nematode intestinal wall and rectal gland cells (the mutualistic 'M' form) (Somvanshi *et al.*, 2010, 2012). Finally, matured nematodes rupture the carcass to be released back into the environment (**6**), completing the life cycle in between 7 and 20 days, and releasing >100,000 IJs from infection of a single IJ (Silva *et al.*, 2002; Clarke, 2017).

More recently, it was found that the colour changes by melanisation of *Galleria mellonella* (wax moth) larvae after infection by different *Heterorhabditis zealandica* strains is explained by their associations with different *Photorhabdus* species (Booysen, Malan and Dicks, 2022). These colour changes represent a potentially useful insight into identifying EPN isolates for biological studies (Table 1.1) (Abd-Elgawad, 2021).

Table 1.1 | *Photorhabdus-Heterorhabditis* entomopathogenic complexes vary by species. *Heterorhabditis* nematodes vector *Photorhabdus* bacteria according to their species. Adapted from Abd-Elgawad, 2021.

***Heterorhabditis* species**	***Photorhabdus* species**
H. atacamensis	*P. khanii* subsp. *guanajuatensis*
H. bacteriophora	*P. caribbeanensis*, *P. kayaii*, *P. khanii*, *P. laumondii* subsp. *clarkei* subsp. *laumondii*, *P. thracensis*
H. bacteriophora, *H. downesi*, *H. megidis*	*P. cinerea*
H. bacteriophora, *H. georgiana*	*P. kleinii*
H. bacteriophora, *H. indica*	*P. luminescens*
H. baujardi	*P. namneonensis*
H. beicherriana	*P. bodei*
H. downesi, *H. megidis*, *H. zealandica*	*P. temperata*
H. gerrardi, *H. indica*	*P. australis* subsp. *Thailandensis* subsp. *australis*
H. indica	*P. aegypti*, *P. akhurstii*, *P. noenieputensis*
H. marelatus, *H. zealandica*	*P. tasmanensis*
H. mexicana	*P. luminescens* subsp. *mexicana*
H. sonorensis	*P. luminescens* subsp. *sonorensis*
H. zealandica	*P. heterorhabditis* subsp. *Aluminescens* subsp. *heterorhabditis*

1.1.2 *Photorhabdus* as a Pathogen

There are some key features for diagnosis of *Photorhabdus* infection in humans: yellow pigmentation and weak bioluminescence in bacterial colonies, and unusual annular (i.e., ring-shaped) haemolysis surrounding the colony on blood agar plates (Weissfeld *et al.*, 2005). Photorhabdosis is most often characterised by soft tissue damage in the form of 'erythematous' nodules or ulcers, and dissemination often leads to multifocal infection. Indeed, bacteraemia was documented in 5 out of the 18 reported cases (Gerrard and Stevens, 2017).

Some *Photorhabdus* species achieve mammalian infection due to their plethora of virulence factors. For example, the Mcf ('makes caterpillars floppy') toxin family (Mcf1 and Mcf2) causes apoptosis in haemocytes in insects but also mammalian epithelial cells (Y. Lee *et al.*, 2019). The Bcl2-homology 3-like (BH3) domain of Mcf1 mimics mitochondrial pro-apoptotic proteins to induce widespread programmed cell death (Dowling *et al.*, 2004; Yang *et al.*, 2006). In insects, tissue damage is so abundant that infecting insects with *E. coli* heterologously expressing Mcf1 results in loss of turgor pressure and host death (Daborn *et al.*, 2002; Rodou, Ankrah and Stathopoulos, 2010).

In addition, the multifunctional auto-processing repeats-in-toxin (Rtx)-like toxins (MARTX) release their modular structures of effector domains upon entering the target cell; Mcf-like domains are found in approximately one third of all MARTX toxins (Y. Lee *et al.*, 2019).

1.1.2.1 Tc toxins

Encoded by pathogenicity islands in the genomes of both *Yersinia* and *Photorhabdus*, the Tc toxins 'inject' a proteinaceous payload, TcC, into the host cell cytosol. The unfolded, toxic ADP-ribosyltransferase domain of TcC is encased in a cocoon structure by TcB and then cleaved from the rest of TcC. The

TcB-TcC complex associates with TcA: an α-helical, homopentameric, channel-like shell. The variable receptor-binding domain of TcA recognises and binds receptors on the cell surface and enables internalisation of the complex. Once endocytosed, the endosome's acidification opens the TcA shell tip to enable its channel to cross through the membrane and the cargo to diffuse into the cytoplasm for refolding and, ultimately, cell killing (Meusch *et al.*, 2014; Roderer *et al.*, 2019).

Such novel protein shuttles with specific compartments for processing, unfolding, and translocating their cargo have potential uses in biomedical research if toxins can be replaced with medically important bioactive polypeptides; notably, the toxins in Tc systems must be of sufficient size (20-35 kDa) and an isoelectric point (pI) of 8.0 or greater (Roderer *et al.*, 2019).

1.1.3 The *Photorhabdus* Pangenome

Bacteria with parasitic or commensal lifestyles can have closed pangenomes, such as *Staphylococcus lugdunensis*, or open pangenomes such as *S. aureus* and *S. epidermidis* (Argemi *et al.*, 2018). Core genomes are made up of genes encoded in all genomes in a dataset, whereas accessory genes are possessed by only some of the genomes analysed. Analysis by Rivera-Ramírez and colleagues (2022) suggests that *Photorhabdus* possesses an open pangenome and that the core genome converges to n = 348 as the number of genomes increases (given 77 genomes in the dataset). As more genomes are added to the dataset, the open pangenome size increases indefinitely since identifying new strains of *Photorhabdus* tends to add novel genes to the accessory or unique genomes (Rivera-Ramírez *et al.*, 2022). This is hardly surprising given *Photorhabdus*' wide range of virulence factors, secondary metabolites, and regions known to have undergone extensive horizontal gene transfer (HGT), so it contains many non-essential (i.e., accessory) genes. For an open pangenome such as *Photorhabdus*,

the smaller core genome is complemented by the comparatively large accessory genome of over 9000 genes (Figure 1.3).

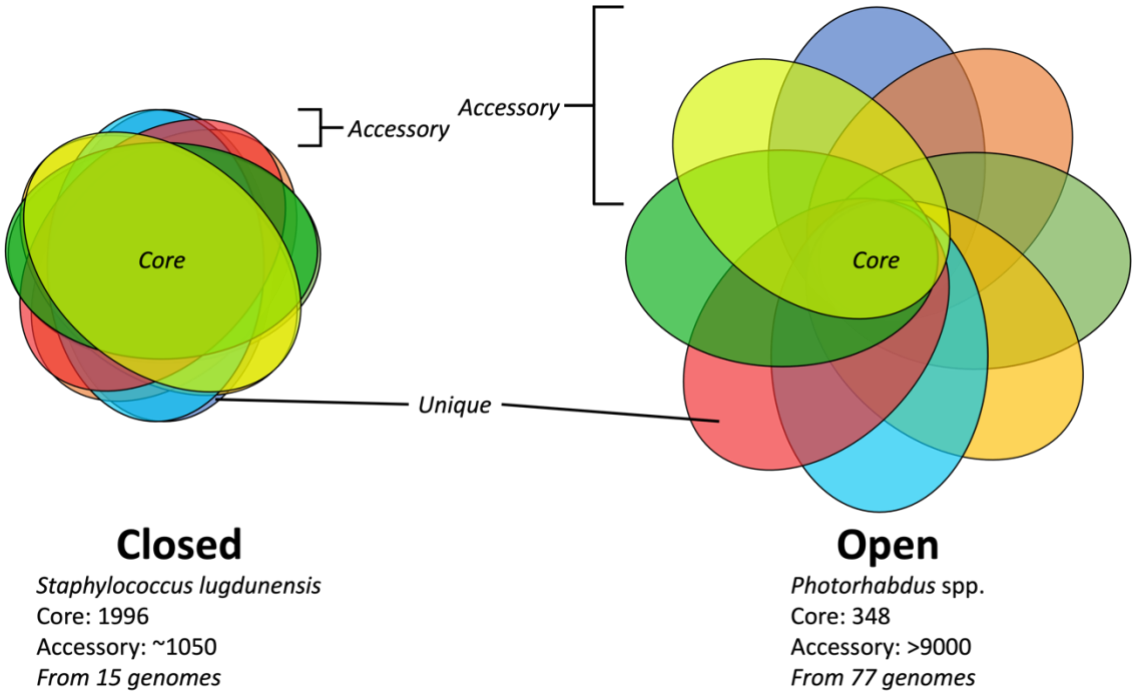


Figure 1.3 | *Photorhabdus* species have an open pangenome.
Data from Argemi *et al.*, (2018) indicates that *Staphylococcus lugdunensis* contains a closed pangenome such that addition of new genomes to the dataset is unlikely to yield a large number of novel genes. On the other hand, *Photorhabdus* possesses an open pangenome whereby its core genome is comparatively smaller than the accessory genomes (Rivera-Ramírez *et al.*, 2022). Addition of strains to the dataset would likely reveal novel genes which would increase the unique/accessory genome sizes. Adapted from McInerney, McNally and O'Connell, 2017.

1.1.4 *Photorhabdus* in Biotechnology

Photorhabdus toxins are effective at killing a range of insect hosts, making it an ideal candidate for biological pest control and integrated pest management (Abd-Elgawad, 2021). These are methods that enable reduction of the use of hazardous chemical pesticides such as organophosphates that have led to

resistances in insects including *Aedes* mosquito populations, for example (da Silva *et al.*, 2020). In fact, safety risks to humans, environmental concerns, outbreaks of secondary pests, and decreases to biodiversity have all led to the removal of some of the most effective and widely used pesticides from the market (Orozco *et al.*, 2016). For the most part, *Bacillus thurigiensis israelensis* (*Bti*) has been used widely as a biolarvicide of mosquitoes and flies and has been the focus of research to improve its effectiveness (Bravo *et al.*, 2011; da Silva *et al.*, 2020). Even still, prolonged culturing of transgenic plant cultivars employing *Bacillus thuringiensis* insecticidal virulence factors has led to subsequent resistance in some insects (Xiao and Wu, 2019). Hence, *Photorhabdus* toxin genes are an attractive alternative for protecting transgenic crops (da Silva *et al.*, 2020; Abd-Elgawad, 2021).

Heterorhabditis-Photorhabdus EPN complexes also have many advantages as biocontrol agents: i) they have efficient and specific insect-seeking abilities, ii) the process yields many hundreds more EPN-bacterial complexes from a single infection, and iii) they provide highly virulent infection towards insects. These factors have facilitated their approval in the US and European pest control markets (Shi and Bode, 2018). For example, Santos Amaral and colleagues (2022) demonstrated the potential of the *H. bacteriophora* HP88-*P. luminescens* EPN complex as a pest control agent of the *Biomphalaria glabrata* amphibious snail by effectively halving the number of snail eggs laid.

Biopesticides comprise 5-6% of the global crop protection market at 3-4 billion USD (Ndolo Obonyo *et al.*, 2019), but new prospects of environmental sustainability, improvements to formulation and delivery technologies, health awareness, and consumer demand is expected to grow this market share to upwards of $8 billion by 2025 (ResearchAndMarkets.com, 2020).

Another exciting aspect of using this unusual bacterium in biotechnology is that of its antimicrobial production. ‘Bio-prospecting’ (i.e., genome mining) identifies new bioactive secondary metabolites to combat antimicrobial

resistance in agricultural and therapeutic settings (Malik and Müller, 2016; Zhou *et al.*, 2019). Additional genomic mining studies have identified clusters involved in β-lactam carbapenem production (Derzelle *et al.*, 2002) and a novel catechol-type siderophore antibiotic termed photobactin (Ciche *et al.*, 2003).

1.2 Contractile Protein Injection Systems

1.2.1 Secretion Systems and Their Effectors

Secretion systems are a well-characterised class of membrane-spanning structures which translocate 'effector' proteins. Whilst the general secretion (Sec) and twin arginine translocation (Tat) systems can translocate proteins across the cytoplasmic membrane, Gram negative bacteria must also translocate extracellular proteins across the periplasmic space (Green and Mecsas, 2016). The secretion systems type I to VI are employed for this purpose and types III, IV, and VI can also be used for translocation across a third membrane of the host cell.

Recruitment of effectors also differs between the secretion systems. The T3SS, for example, recruits effector proteins to the apparatus via N-terminal signal sequences and uses ATP to secrete unfolded effector proteins through the inner tube (Lara-Tejero and Galán, 2019). The T6SS is unique amongst its TXSS siblings since injection relies on contraction of its sheath. The structure is composed of 14 core components, TssA-M and PAAR, and comprises a needle-like structure whose length depends upon the width of the cell in which it is anchored (Santin *et al.*, 2019) (Figure 1.4). There are 6 T6SS subtypes; subtype IV (T6SS^{iv}) from *A. asiaticus* is closely related to MACs and is discussed in chapter 1.2.2.2.

Generally, T6SS effectors can act as periplasmic lytic toxins or cytoplasmic toxins. Since T6SS can target both prokaryotes and eukaryotes, some trans-kingdom effectors target conserved molecules such as DNA or phospholipids, as well as ADP-ribosyltransferases such as Tre (Jurėnas and Journet, 2021). Anti-bacterial effectors are encoded with a corresponding immunity protein to prevent self-intoxication. Notably, not all T6SS payloads are toxins as some proteins can be secreted from the injectisome into the surrounding milieu. For

example, the metal-chelating molecule YezP is used to acquire surrounding zinc ions to combat host immunity (Wang *et al.*, 2015).

Anti-host effectors include VgrG1 encoded in *V. cholerae* which cross-links actin to prevent phagocytosis and alter cell morphology (Dutta *et al.*, 2019). *P. aeruginosa* also encodes many T6SS-associated effectors including phospholipases D (PldA and PldB) and Vgr2b which facilitate bacterial entry into host cells (Monjarás Feria and Valvano, 2020).

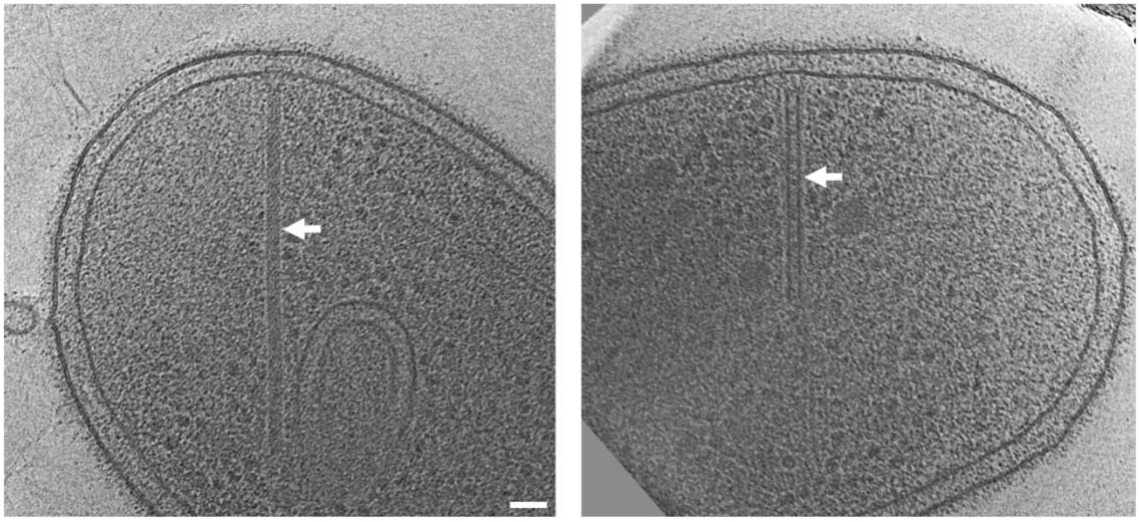


Figure 1.4 | Tomography of T6SS in *M. xanthus* spans the cell cytosol.
The membrane anchored T6SS (white arrows) from *Myxococcus xanthus* is shown to span the cytosol but differs vastly in length depending on its contraction state. Scale bar = 100 nm. Left = extended; right = contracted. Adapted from Chang *et al.*, 2017.

T6SS assembly is initiated by the binding of 10/12 copies of the cytoplasmic outer membrane protein TssJ with the inner membrane proteins TssLM (Nguyen *et al.*, 2018). TssK then links the membrane complex to the hexameric $TssEF_2G(K_3)_2$ baseplate complex which surrounds the $VgrG_3$PAAR spike/tip complex. The tube consisting of Hcp (TssD) hexamers then assembles towards the cytoplasmic end from this complex, surrounded by the polymerised TssBC sheath, and terminates with the TssA cap.

Three structural T6SS components associate with the effectors: i) inside the 40Å Hcp hexamers prior to assembly of the tube (Silverman *et al.*, 2013), ii) the VgrG spike, and iii) the N- or C-terminal extensions of the sharp PAAR tip which can be accompanied by chaperones/adaptors (Figure 1.5). Adaptors bind the same spike protein and, thus, diversify the arsenal of cargo effectors that can be loaded onto the tip of the apparatus. Alternatively, 'specialised' effectors, such as the large rearrangement hot spot (RHS) core-containing proteins, are effector domains covalently fused to the C-terminus of VgrG, PAAR, or Hcp proteins (Alcoforado Diniz, Liu and Coulthurst, 2015). Either way, a PAAR-containing protein tip must be part of the structure for the T6SS to assemble and function (Beauvois *et al.*, 2023).

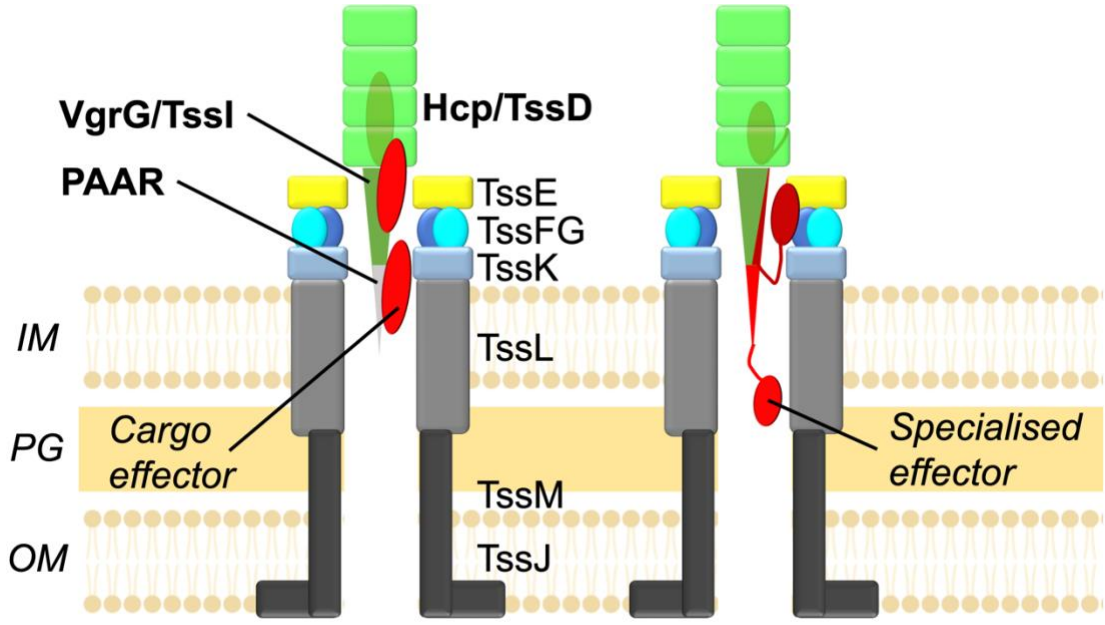


Figure 1.5 | T6SS effectors are recruited to VgrG, PAAR, and Hcp.

The type 6 secretion system (T6SS) spans the inner membrane (IM), peptidoglycan (PG), and outer membrane (OM). Structural components of the membrane complex (TssJLM) and baseplate complex (TssEFG) are connected via TssK and associate with the VgrG spike and PAAR tip. Stacking of the hexameric tube proceeds away from the membrane inside the cell and either associates with cargo effectors, which may require adaptor proteins, or specialised

effectors which have VgrG or PAAR domains covalently attached. Adapted from Coulthurst, 2019.

A phylogenomic study revealed that PAAR domains encode up to 3 interacting motifs (often of the form 'PAAX') within the loops at the base of the structure to form a sharp conical 'tip' with which to penetrate through the membrane of the target cell (Zhang *et al.*, 2021). PAAR proteins can associate with T6SS or eCIS with no obvious sequence differences. PAAR proteins are divided into 8 types (A-H) and 16 subtypes and are encoded in ~23% of bacterial genomes. Thus, they are likely important versatile tools for bacteria to compete in their niche; free-living bacterial 'generalists' often encode multiple copies of the same gene whilst host-dwelling bacteria are less enriched in PAAR proteins, possibly since many of these associated toxins would damage the bacterium's host (Zhang *et al.*, 2021). Moreover, the interplay between PAAR and toxin is vast; a given PAAR subtype can associate with up to 40 toxins, and a given toxin can associate with up to 10 PAAR subtypes (Zhang *et al.*, 2021).

1.2.2 Extracellular Contractile Injection Systems (eCIS)

eCISs are structures found across the microbial kingdoms, including bacteria and archaea, and are documented in the databases dbeCIS (http://www.mgc.ac.cn/dbeCIS/) and eCIStem (http://ecistem.pythonanywhere.com/). Previously termed 'phage-like protein translocation structures', or PLTSs, their architecture resembles that of a nano-sized syringe capable of injecting proteinaceous payloads independently of the cell which produced it. These are differentiated from ordinary CIS since eCIS do not require anchoring to the membrane of the producing cell.

The eCIS that will be the focus of this work is the *Photorhabdus asymbiotica* virulence cassette encoded upstream of the Pnf protein payload (PaPVC*pnf*) shown in Figure 1.6. As will be discussed in more detail in chapter 1.3.1, this structure packages payload proteins into a hollow tube which is 'injected'

across the membrane after contraction of the sheath, similarly to the T6SS (Figure 1.7).

For context, it will be useful to refer to Figure 1.8 as a summary of the structural conservation between some of the CISs mentioned in this chapter. These structures have similarities to the PVC and many of them also deliver proteinaceous payloads into target cells.

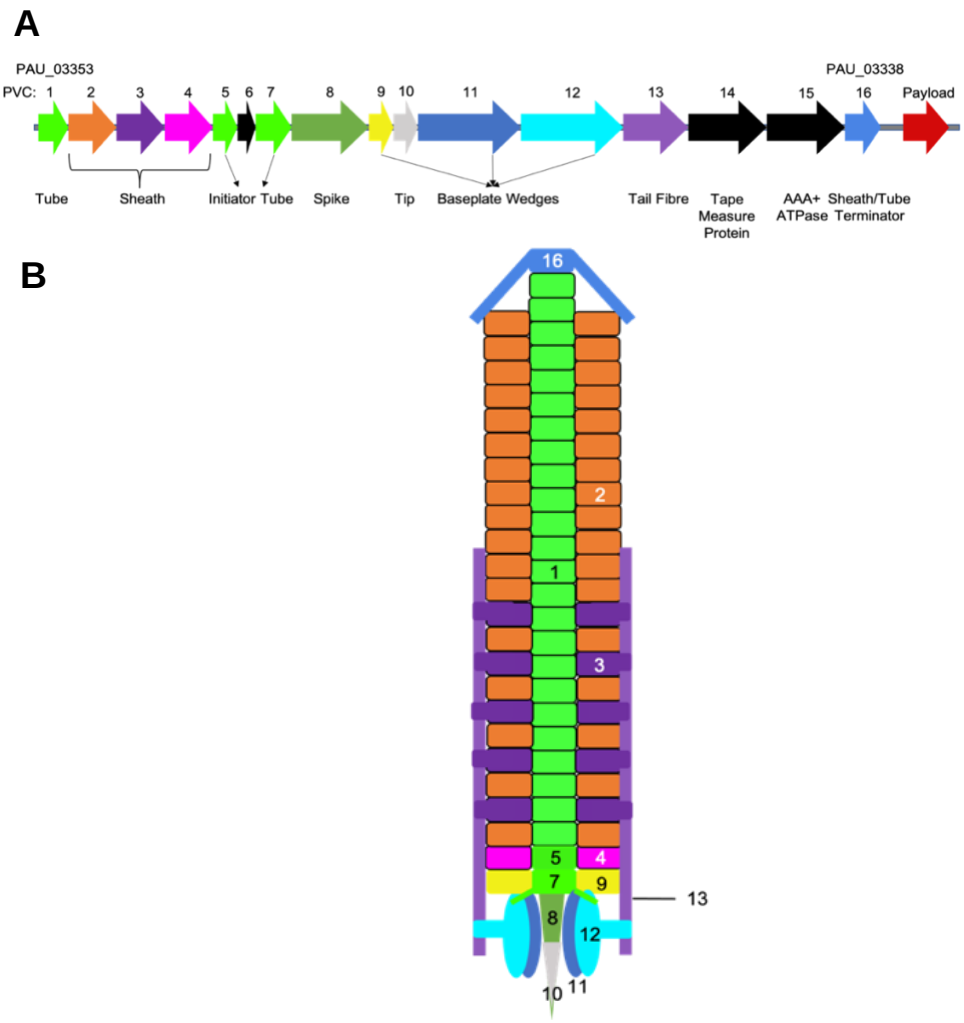


Figure 1.6 | Genetic and Structural Schematic for PaPVC*pnf*.

(A) The depicted operon is based on the well-characterised PVC*pnf* operon from *Photorhabdus asymbiotica* ATCC43949. A simplified schematic of the mature, fully assembled PVC 'syringe' is shown in **(B)**. Pvc6, Pvc14, and Pvc15 are not present in the mature syringe structure and the payload is loaded inside the syringe (not depicted). Adapted from Jiang *et al.*, 2019. Not to scale.

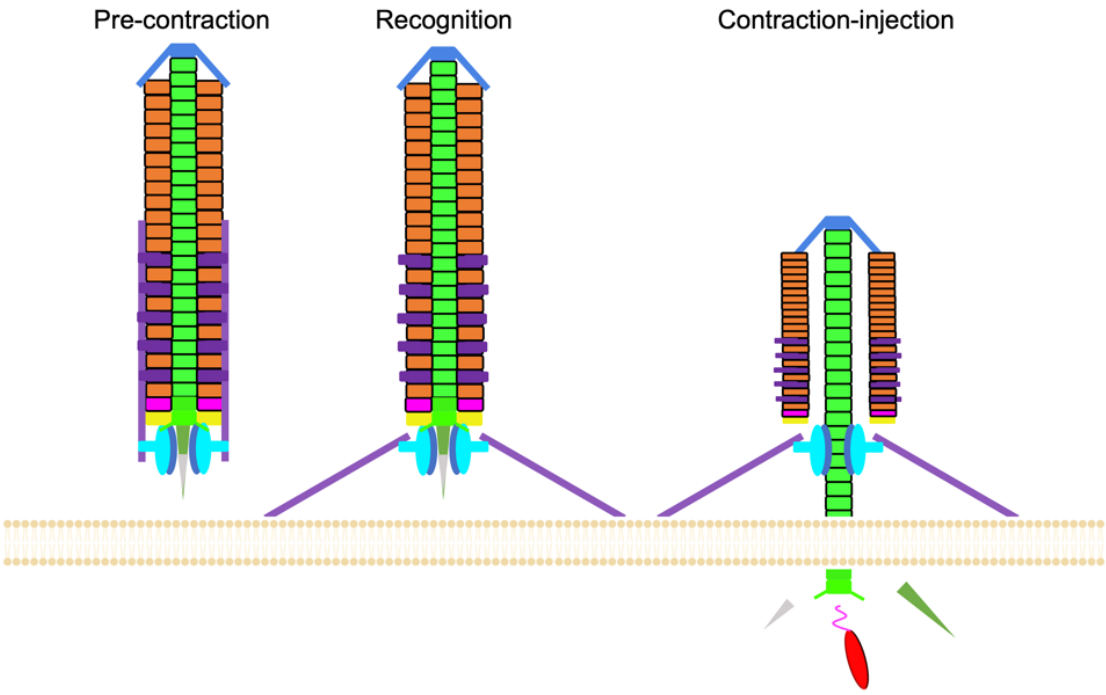


Figure 1.7 | eCIS states change during and after injection.

The mature eCIS is initially in an extended state. Tail fibres binding the target cell membrane causes mechanical contraction of the sheath, pushing the tube through the membrane. Finally, payload proteins in the inner tube diffuse into the cytosol. Adapted from Jiang *et al*., 2019.

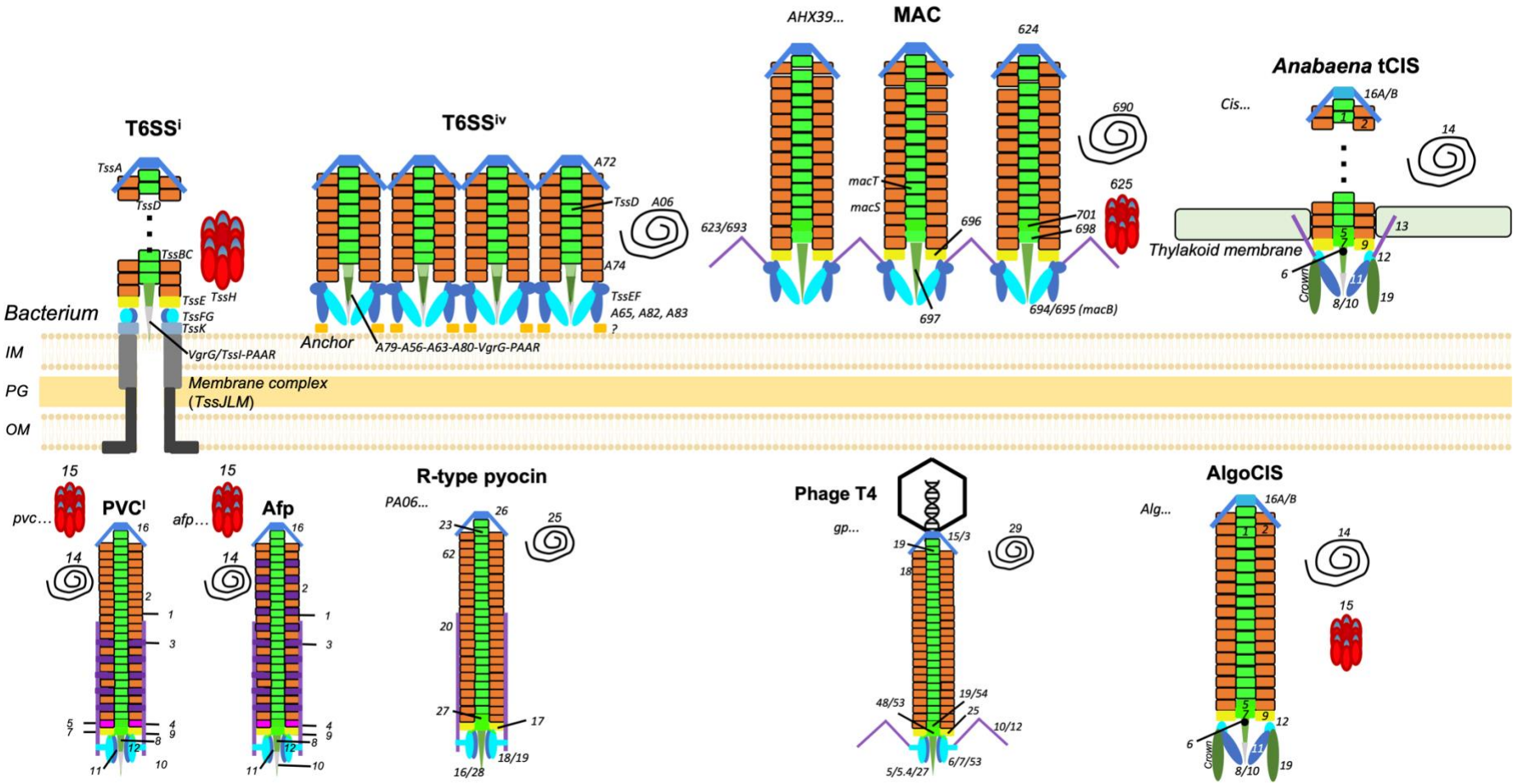

T6SSi
TssA
TssD
TssBC
TssE
TssH
TssFG
TssK
VgrG/TssI-PAAR
Bacterium
IM
PG
OM
Membrane complex
(TssJLM)
T6SSiv
A72
TssD
A06
A74
TssEF
A65, A82, A83
?
Anchor
A79-A56-A63-A80-VgrG-PAAR
AHX39...
MAC
624
690
macT
macS
623/693
696
701
698
625
697
694/695 (macB)
Anabaena tCIS
Cis...
16A/B
14
13
Thylakoid membrane
Crown
8/10
19
15
pvc...
PVCl
afp...
Afp
R-type pyocin
PA06...
Phage T4
gp...
29
15/3
48/53
19/54
10/12
5/5.4/27
6/7/53
AlgoCIS
Alg...
16/28
18/19
17

Figure 1.8 | Conserved structure of tailed contractile systems.

Tailed contractile systems (i.e., CISs + R-type phages) have vastly conserved architecture. Shown eCISs are class I PVC (modelled after PVC*pnf*), Afp (structured almost synonymously with AfpX), R-type pyocin, and AlgoCIS from *Algoriphagus machipongonensis*. Other CISs include subtype I and subtype IV T6SS, MAC, and the thylakoid-associated tCIS from *Anabaena* spp. Phage T4 is also shown; structures possessing tape-measure proteins (e.g., Pvc14 and Afp14) and ATPases (e.g., Pvc15 and Afp15) are shown with these proteins adjacent to the structure. Other structures discussed in this chapter such as BIS and *Sc* CIS do not yet have structural data to display. Payload proteins (not shown) are loaded into the tube lumen of the PVC, Afp, both T6SSs, MAC, and potentially the tCIS and AlgoCIS; phage T4 is loaded with phage genomic DNA.

Afp12 homologues are hotspots for evolutionary re-engineering, as shown by the capacity to link T6SS^{iv} structures together as well as to intercalate the long tail fibres of adjacent MACs.

The tCIS and AlgoCIS both have Cis6/Alg6 protein 'plugs' that insert between the tube and the spike and, in addition, their Cis11/Alg11 baseplate proteins form enclosed spike cages flanked by Cis19/Alg19 'crowns'. Meanwhile, the AlgoCIS appears to lack tail fibres encoded by an Afp13-like gene, akin to class II PVC, PVC*lopT*. On the other hand, the tCIS uses its Cis13 gene to anchor itself to the thylakoid membrane for injection from the stress-induced 'ghost cells'.

1.2.2.1 Anti-feeding Prophage (Afp and AfpX)

Anti-feeding prophage (Afp) from the *Serratia entomophila* plasmid pADAP (amber-disease associated plasmid) is the most similar homologue to the *Photorhabdus* virulence cassette (PVC). Structurally, the Afp represents a nano-sized 'syringe'-like protein complex (Rybakova, Mitra and Hurst, 2014) (Figure 1.9). Unlike *Photorhabdus*' wide host range, Afp has only been associated with cessation of feeding in the larvae of the New Zealand grass grub, *Costelytra zealandica*; ingestion of less than 500 particles is often sufficient to induce this behaviour (Kube and Wendler, 2015).

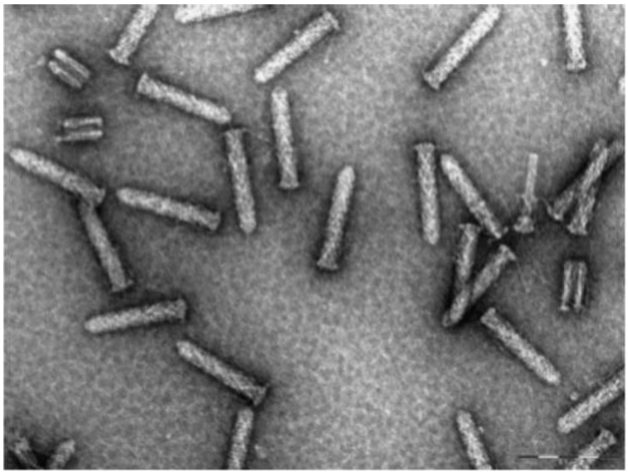


Figure 1.9 | Anti-feeding prophage (Afp) particles in TEM.
Tomography reconstruction of Afp particles. The baseplate is marginally thicker than the sheath and can be seen as short 'skirts' at the ends of the particles. Striations can be seen in the sheaths representing the multiple stacked hexamers of the sheath around the inner tube. Scale bar = 100 nm. Adapted from Rybakova, Mitra, and Hurst, 2014.

Whilst the T4 and λ phages are degraded at pH ranges greater than 10, the Afp is unaffected at high pH. Both the Afp and PVC tube lumens are negatively potentiated; for the Afp, this is probably required for the conductance of its Afp18 effector into target cells after contraction (Desfosses *et al.*, 2019; Zinke, Schröder and Lange, 2022). The *C. zealandica* gut is pH 11, which likely justifies some of these biochemical differences between the Afp

and other eCISs. In addition, both the Afp and PVC may be similarly regulated as the Afp is regulated by AnfA1: an RfaH-like antiterminator protein which binds the operon polarity suppressor (ops) element immediately upstream of Afp1: the first gene in the operon (Hurst *et al.*, 2007). Similar RfaH boxes are found upstream of the *pnf*, *lopT*, and *PaTox* PVC operons.

The Afp16 'cap' is a tail terminator protein (TrP) which terminates elongation of the tube and stabilises the sheath to yield longer (>200 nm) sheathed particles (Rybakova *et al.*, 2013). Genes encoded downstream of Afp16, meanwhile, appear to encode the toxins designated Afp17 and the much larger Afp18, though Afp17 is not found within the mature structure so its role is unclear (Rybakova *et al.*, 2013).

AfpX, or AGR96X Afp, is a close orthologue from *Serratia proteamaculans* which causes much faster killing of its grass grub host than that of *S. entomophila* and is also capable of killing manuka beetle larvae (*Pyronata festiva*) (Hurst *et al.*, 2018). There are some structural differences in the AfpX which has two AfpX16 orthologues (AfpX16a and AfpX16b); the latter has a 34-amino acid deletion which may mediate the greater variability in lengths observed with the AfpX compared with the Afp (Hurst *et al.*, 2018). Additionally, the payload proteins, AfpX17 and AfpX18, are homologous to the *S. entomophila* Afp17 and 18. AfpX17 is more than twice as large as Afp17 and has sequence similarity to a putative ADP-ribosyltransferase from *P. asymbiotica* (PAU_03033), whilst AfpX18 has roughly 80% sequence similarity to the Afp18 in *Yersinia ruckeri* Afp-like eCIS: a known tyrosine-glycosylation toxin (Jank *et al.*, 2015).

Crucially, unlike the PVC, the AFP is i) plasmid-encoded, ii) localised for use against very specific host in New Zealand, and iii) are limited to only one operon per genome. Taken together, a more restricted set of selection

pressures may explain some of the differences that exist between the structures and uses of the PVC and AFP in nature.

1.2.2.2 Metamorphosis-Associated Contractile (MAC) and T6SS^{iv}

The metamorphosis-associated contractile (MAC) in *Pseudoalteromonas luteoviolacea* facilitates the development of its symbiont tubeworm, *Hydroides elegans*. Mif1 is an effector packaged inside of the MAC tube lumen which triggers metamorphosis (Ericson *et al.*, 2019). Intriguingly, MACs expand as an ordered series of hundreds of tethered structures emanating from the site of cell lysis; the array consists of interlocking tail fibres with 6 adjacent structures to form a 'sea mine-like' amorphous array (Shikuma *et al.*, 2014) (Figure 1.10).

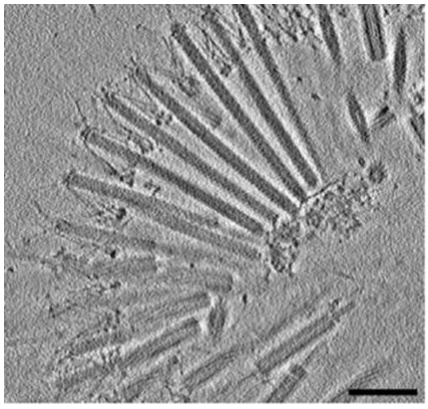


Figure 1.10 | The MAC Amorphous Array.
Electron cryo-tomography of MACs from *Pseudoalteromonas luteoviolacea*. A meshwork of tail-fibres holds this amorphous structure together after release from the producing bacterium. Scale bar represents 50 µm. Adapted from Rocchi *et al*., 2019.

On the other hand, Rocchi and colleagues (2019) identified a MAC that can be loaded with a non-specific endonuclease effector, Pne1, which causes cell death upon injection into host cells.

The subtype IV T6SS (T6SS^{iv}), is also a lineage Ib eCIS. T6SS^{iv} is produced by *Amoebophilus asiaticus* and forms ordered hexagonal arrays while

membrane anchored. In contrast to the MAC, however, these structures are i) less numerous and ii) putatively joined at the baseplates in the inner membrane (Figure 1.11). Authors speculate that the T6SS^{iv} is used to escape the phagosome in the bacterium's amoeba host (Böck *et al.*, 2017; Xu *et al.*, 2022).

The requirement for an intact bacterial membrane blurs the line between whether T6SS^{iv} is a T6SS or an eCIS. For example, compared to a traditional T6SS, the T6SS^{iv} has a much shorter tube (Böck *et al.*, 2017), encodes a TMP, and is not recycled by an ATPase: all features more closely resembling an eCIS, though no effectors have currently been characterised. While traditional T6SS spans the entire bacterial membrane, T6SS^{iv} localises only to the cytoplasmic membrane anchored by an unknown protein (Unni *et al.*, 2022).

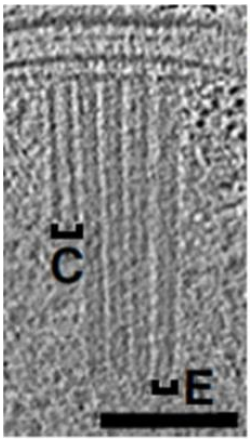


Figure 1.11 | T6SS^{iv} form ordered membrane-bound arrays.
Cryotomography shows T6SS^{iv} forming ordered arrays bound at the inner membrane of *A. asiaticus*. C = contracted, E = extended. Scale bar = 100 nm. Adapted from Böck *et al.*, 2017.

1.2.2.3 Antibacterial eCIS

Despite being widespread in microbes, an analysis by Geller and colleagues (2021) indicated that eCIS distribution is heavily taxonomically biased towards aerobic environmental microbes such as *Chitinophaga*, *Dickeya*, and

Algoriphagus. In addition, they propose that some eCIS can target bacteria, presumably in a manner similar to the binding of R-type pyocins, discussed more in chapter 1.2.2.4. This hypothesis stems from the BLAST hits of the tail fibre gene sequences from 12 eCIS loci which share more homology with phage tail fibres than the other 99 eCIS tail fibres which share homology with eukaryote-targeting viruses such as mammalian Rhesus adenovirus 60, or Bat mastadenovirus B. The putative toxins encoded at these loci, termed eCIS-associated toxins, or EATs, were shown to inhibit growth when heterologously expressed in *E. coli* which was rescued when the corresponding immunity gene was expressed, though the contractile systems themselves were never characterised. This could represent an exciting new class of eCIS which target bacteria in the fight against antimicrobial and phage-resistance which should be explored further.

1.2.2.4 Tailocins as a Model for Contraction

A class of caudate bacteriocins (often referred to as tailocins) are found in *Pseudomonas aeruginosa*. R-type pyocins are used as a bactericidal weapon against closely related species and up to 200 R-type pyocins can be released from a single cell in response to DNA damage (Carim *et al.*, 2021). Tailocins bind strongly to the LPS of sensitive – usually non-sister – cells and puncture the cell membrane through contraction of the sheath which forms a pore for rapid ionic flux and non-lytic cell death (Uratani and Hoshino, 1984; Lee *et al.*, 2016). How self-intoxication is mitigated in *P. aeruginosa* strains is unclear though O-antigen presentation and composition, as well as LPS thickness, alters the sensitivity to pyocins from certain strains (Carim *et al.*, 2021). R-type pyocin, meanwhile, can be retargeted to *E. coli* when engineered with tail fibres from the *E. coli*-specific phage, ΦV10 (Scholl *et al.*, 2009).

These caudate structures are more simplistic than the other eCISs as they do not inject polypeptides. They are also more streamlined than the

bacteriophage since tailocins elicit potent toxicity against competing bacteria without having to inject nucleic acids (Heo *et al.*, 2007). On the other hand, Lee and colleagues (1999) demonstrated an exceptional tailocin containing single-stranded nucleotides.

Fraser and colleagues (2021) used Domain Motion in Atomic Detail (DMAD) modelling to study the contraction energetics of R-type pyocin (Ge *et al.*, 2015; Kudryashev *et al.*, 2015). The authors were able to accurately predict a free-energy profile for the propagation of contraction between each subunit in the structure and estimate the activation energy (E_a) and Gibbs Free energy (ΔG); confirmed experimentally as 160 kcal/mol and -2160 kcal/mol, respectively. The free energy of the R-type pyocin is significantly lower than that predicted for the Afp (-730 kcal/mol) and suggests a greater tendency to fire (Desfosses *et al.*, 2019). Stabilisation of the tail fibres upon membrane binding initiates sheath contraction from the tip to the sheath and through each helix to the distal end.

1.2.3 Other CISs: BIS, tCIS, and AlgoCIS

Many unique CISs have been characterised in recent years. Metadata analysis of eCIS-encoding bacterial isolates indicates depletion of eCIS in most human tissues including the skin, urogenital tract and oral and digestive systems (Geller *et al.*, 2021). However, the gut-commensal *Bacteroidales* family are mildly enriched in eCIS as well as other CIS systems such as the subtype 3 T6SS (T6SS^{iii}) and a unique '*Bacteroidales* injection system' (BIS). Rojas and colleagues (2020) identified that BIS-like gene clusters enable commensal bacteria to target diverse organisms, such as tubeworms and amoebae, which promotes a healthier gut; BIS-like elements are more enriched in healthy gut microbiomes than in those with inflammatory bowel disease.

Whilst many CISs may serve as additional mode of virulence for the producing cell, some serve as a crucial part of its growth. The *Sco* eCIS from *S. coelicolor* appears to maintain cell shape, facilitate stress-induced cell death and modulate the switch to sporulation (Vladimirov *et al.*, 2023).

Other recent discoveries include the 120 nm-long thylakoid-anchored CIS (tCIS) from the multicellular cyanobacterium, *Anabaena* sp. PCC 7120. In this organism, tCIS induces ghost cell formation in stressed cells, exposing the tCIS to the as-yet-unknown target cell (Weiss *et al.*, 2022). Similarities of the tCIS structure to previously discussed CISs are abundant, including a negatively potentiated inner tube lumen with a diameter of 37Å. Interestingly, the Cis6 component caps a cavity of the spike protein closest to the tube lumen, contrasting Afp6 and Pvc6 which are not part of their mature structures. The tCIS also has other structural features: six trimers of Cis19 form a so-called 'crown' with pseudo-C3 symmetry which has short tail fibre-like extrusions, and Cis12 contains extensions required for anchoring to the thylakoid membrane alongside the Cis13 fibres (Weiss *et al.*, 2022).

Yet another recent discovery is the eCIS derived from *Algoriphagus machipongonensis* termed AlgoCIS (Xu *et al.*, 2022). AlgoCIS is closely related to the MAC and Afp and is comprised of a single sheath protein, Alg2 (i.e., there are no '*alg3*' or '*alg4*' encoded in the operon). Similarly to the tCIS Cis6, the Alg6 'plug protein' associates with the tube initiator proteins, Alg5 and Alg7, as well as the Alg8 spike protein. Interestingly, Alg6 is required for mature AlgoCIS assembly; the MAC also shared this requirement for the homologous JF50/12690 gene, in contrast to the PVC which can form mature particles from a ΔPvc6 mutant (Jiang *et al.*, 2019). Additionally, the AlgoCIS operon lacks a canonical '*alg13*', but encodes a putative tail fibre, *alg19*, which shares a general structural similarity to the baseplate TssK from T6SS (Park *et al.*, 2018). The researchers hypothesised that the 'spike cage', an

extension of the Alg11 baseplate, may have conserved sugar binding modules which may be involved in attachment to the cell glycocalyx. The spike cage also appears to be conserved in MACs and T6SS[iv].

1.3 The *Photorhabdus* Virulence Cassette

Of the many virulence factors *Photorhabdus* possesses, the *Photorhabdus* virulence cassette, or PVC, is one of the largest. *P. luminescens* TT01 was first sequenced in 2003 by Duchaud and colleagues which indicated that this strain contained a large repertoire of virulence-associated loci, and encoded many sites with unusual GC content and similarity to phage baseplate and tail regions (Yang *et al.*, 2006). Cosmid libraries were prepared from both *Photorhabdus luminescens* TT01 and *P. asymbiotica* ATCC43949 and transformed into *E. coli* EC100. When expressing a particular cosmid-encoded PVC operon with its cognate putative effector, Pnf, and injecting the lysate into insect larvae, some larvae were killed within 15 minutes (Yang *et al.*, 2006; Waterfield *et al.*, 2008). The supernatants of these bacterial cosmid clones could be fractionated by size-exclusion chromatography (SEC) and visualised with transmission electron microscopy (TEM). This revealed the expression of 150-800 nm-long structures, similar to the antibacterial R-type pyocins which appeared to affect insect cell lines (Michel-Briand and Baysse, 2002; Yang *et al.*, 2006). It was concluded that PVCs must be functional homologues of the anti-feeding prophage (Afp) identified in *Serratia entomophila* (Hurst, Glare and Jackson, 2004) which is itself a homologue of *Myoviridae*-like (i.e sheathed) phages such as T4 (reviewed by Zinke, Schröder and Lange, 2022).

Since the PVC's discovery, many Afp- and PVC-like elements have been identified and were collectively termed phage-like protein translocation structures (PLTSs), though current terminology tends to use the acronym 'eCIS': extracellular contractile injection system. eCIS undergo structural transitions after release from the producing cell: i) pre-contraction (also called the 'extended' state), ii) recognition, where binding elements such as tail fibres bind a host cell membrane, and iii) contraction, where the sheath

twists and compresses the rigid tube through the membrane such that payload protein is released directly into the cytosol (Figure 1.7).

This chapter will cover the structural and genetic components of PVCs, as well as what can be learnt from the analogous structures from other organisms. This chapter also discusses current knowledge of the elusive AAA+ ATPase which may directly interact with the payload (chapter 1.3.4 on page 60). Information about the typical effector proteins that can be loaded in conjunction with the contractile nanomachinery are discussed in chapter 1.3.2 starting on page 47.

1.3.1 Encoding the PVC

Each PVC is encoded by 15-20 'core' genes as well as at least one putative effector protein (Yang *et al.*, 2006). Each *Photorhabdus* species then encodes typically 5-6 PVC loci with different effector proteins and subtle differences in the sequence of some structural genes. Thus, the naming convention of the PVC operons is often associated with the cognate effector that is loaded into the complex as well as the species in which it is found. For example, PaPVC*pnf* is derived from *P. asymbiotica* (Pa) and encodes the Pnf toxin downstream of the operon core genes, though the PaPVC*pnf* locus also encodes the sepC-like gene, Pdp1 (Wang *et al.*, 2021). For PaPVC*pnf*, in particular, the 10.7 MDa PVC complex is composed of 328 polypeptide chains (Jiang *et al.*, 2019). ThePVCs get loaded with an associated effector and are likely released only upon lysis of the producing cell: a form of altruism that primarily benefits the sister cells as well as its nematode symbiont (Sarris *et al.*, 2014; Shikuma *et al.*, 2014).

A schematic of the *P. asymbiotica* ATCC43949 PVC*pnf* operon and 'nanosyringe' are shown in Figure 1.6. This is a 'class I' PVC which possesses the classical 16-gene cassette. The 'class II' PVC, such as PVC*lopT* lacks the tail fibre, *pvc13*, and *pvc3*, whilst the 'class III' PVC, represented by

PVC*patox*, encodes a unique *pvc0* (*PAU02248*) and two more open reading frames (ORFs) with unknown functions (Vlisidou *et al.*, 2019) (Figure 1.12). Work by Jiang and colleagues (2019) used cryo-electron microscopy to study the macromolecular structure of PVCs which has significantly advanced our understanding of how PVCs assemble and function. A similar study conducted for the homologous Afp was conducted the same year (Desfosses *et al.*, 2019).

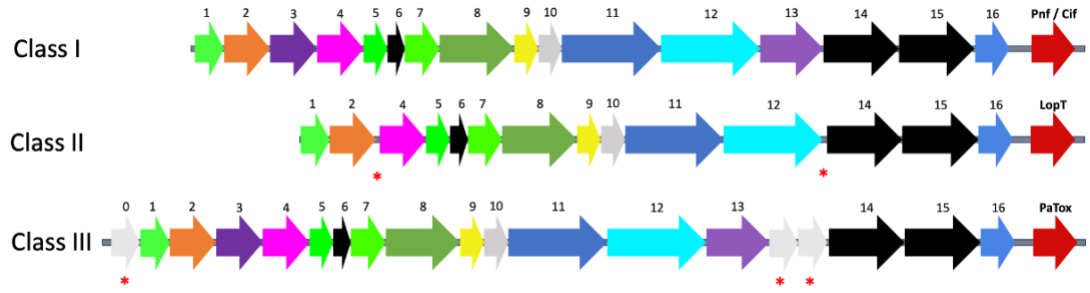


Figure 1.12 | Gene cassettes of PVC classes I-III.
The class I PVC Is represented by *P. asymbiotica* PVC*pnf* and PVC*cif* operons; note that these operons contain subtle differences in certain structural genes as well as the downstream payload gene(s). For example, *pvc11* in PVC*pnf* is far shorter than in PVC*cif*. PVC*lopT* is an example of a class II PVC which lacks the *pvc13* tail fibre gene as well as the *pvc3* sheath protein. PVC*patox* is the only example of class III PVC in *P. asymbiotica* and encodes an upstream *pvc0* open reading frame (ORF) and two further ORFs between *pvc13* and *pvc14* all of unknown function. Adapted from Vlisidou *et al.*, 2019.

1.3.1.1 Contractile Syringe: Tube and Sheath

Looking first at the tube, this is an inner compartment into which the payload is loaded. The tube is composed of 24 stacked hexameric rings which form a continuous inner channel of 4 nm in inner diameter and 8 nm in outer diameter. At the proximal end (i.e., near the spike and tip), hexameric Pvc7 is stacked with the β-barrel ring of the trimeric baseplate

wedge, Pvc8: likely the initial point of assembly of the 'syringe'. The interaction between Pvc7 and Pvc8 accommodates the C3 (spike) to C6 (tube and sheath) symmetry. Meanwhile, Pvc9 encases Pvc7 and links the tube to the other baseplate wedges. Pvc5 is stacked onto Pvc7 and these two proteins, thus, make up the initial structure of the tube. In the structure resolved by Jiang and colleagues (2019), Pvc1 makes up the next 22 layers stacked towards the distal end where it is terminated by the Pvc16 cap. The electrostatic potentials of each ring (negative charge on the proximal side; positive charge on the distal side) likely facilitate the correct assembly and orientation of each ring (Figure 1.13). These three variants, Pvc1, Pvc5, and Pvc7, are structurally similar and manifest an overall negative potential on their inner surface. It is unknown whether the toxin favours a negative charge for expulsion from the tube after contraction, such as with the Afp18 payload in the Afp (Zinke, Schröder and Lange, 2022).

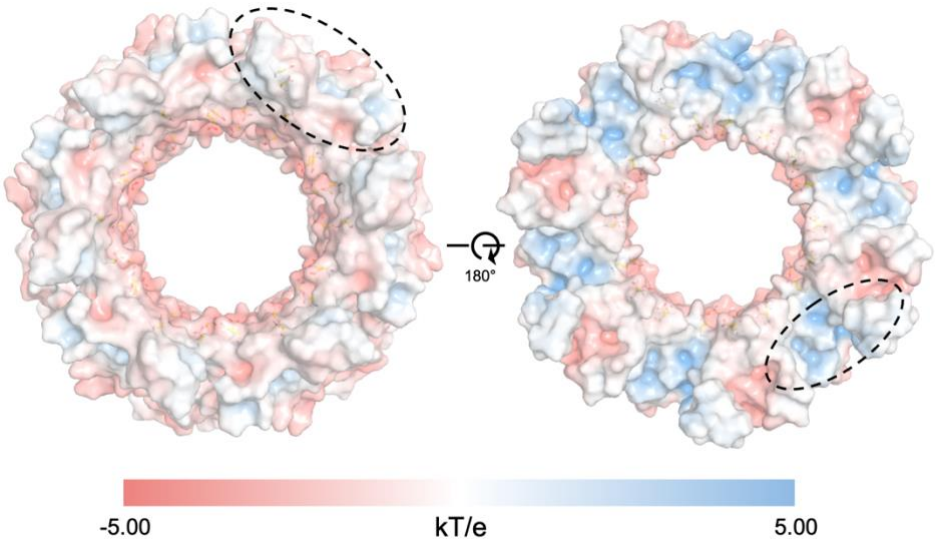


Figure 1.13 | Pvc1 hexamers have complementary electrostatic potentials.

22 Pvc1 hexamers comprise the tube of the PVC by stacking. The top/distal side of Pvc1 (left) has overall negative surface potential whilst the bottom/proximal side (right) has more positively-charged residues. The inner part of the hexamer is also comprised of negatively-charged residues: Glu and Asp residues are represented here by sticks and one

of the six binding sites are circled by dashes. Electrostatic potential was calculated using APBS in the PyMol suite. PDB ID: 6J0B, adapted from Jiang et al., 2019.

Each tube monomer consists of an N-loop on one side and an α-loop on the other, connected by a β-sandwich-like fold. The N-loop, as well as loop β4-β5.1, contributes to the tube-sheath interaction whilst the β2-β3 loop conveys a set of large inter-ring contacts between tube subunits. The α-loop forms a bridge between the sheath Pvc2, the loop β2-β3, and the N-loop of an adjacent monomer. This conserved set of interactions enables the tube-tube interaction to be stronger than the tube-sheath interaction which, in turn, facilitates the tube to 'slide' away from the sheath during contraction.

The sheath, meanwhile, has a comparable yet distinct structure from the tube. As shown in Figure 1.14, its first layer is composed of the sheath initiator complex, Pvc9 which encases Pvc7, and Pvc4 which encases Pvc5 and forms the first layer of the sheath. From here, an alternating sequence of Pvc2 and Pvc3 compose the next 10 layers of the hexamers, followed by another 12 layers of Pvc2 (Jiang *et al,*. 2019).

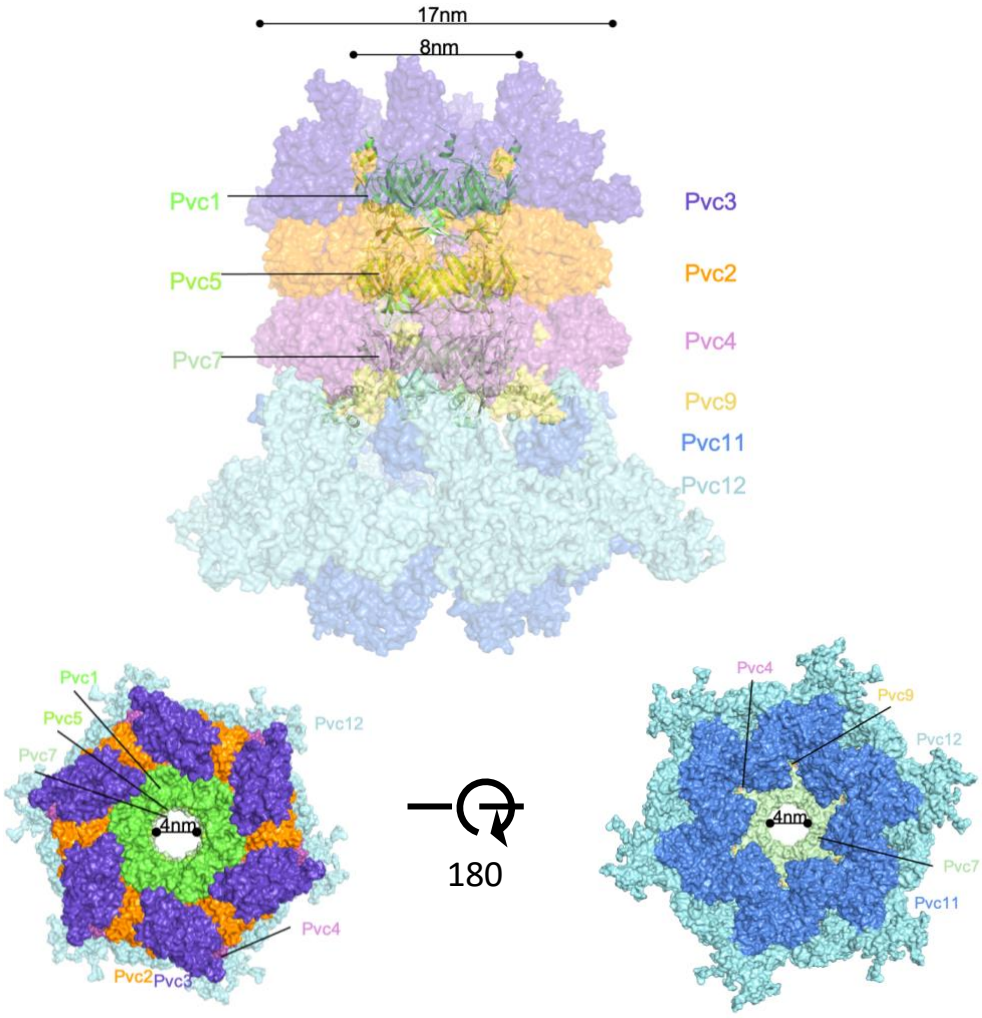


Figure 1.14 | Molecular overview of the baseplate-sheath-tube complex.

(A) A side view of the baseplate-sheath-tube complex in the intact extended PVC. The inner tube (green; Pvc1, Pvc5, and Pvc7) is shown as a cartoon whilst the outer sheath (Pvc2, Pvc3, Pvc4, and Pvc9) and baseplate (Pvc11 and Pvc12) are shown as a surface. The total diameter of the inner tube and outer sheath are 8 nm (80Å) and 17 nm (170Å), respectively. **(B)** Cross-sectional view of the syringe cut above the most proximal Pvc3 hexamer. The distal (left) and proximal (right) ends show a 4 nm (40Å) cavity inside the syringe where the protein payload is loaded. PDB ID: 6J0N, adapted from Jiang et al., 2019.

The ‘sequence’ of the sheath, according to this cryo-EM data, is Pvc4-(Pvc2-Pvc3)$_5$-(Pvc2)$_{12}$ and, like the inner tube, is terminated by the Pvc16 cap. Much like the tube, the three variants (Pvc2, 3, and 4) are structurally similar. Pvc2 most closely resembles the R-type pyocin sheath protein whilst Pvc4 is slightly larger. Pvc3 differs from the other variants by the possession of a protrusion which forms a fibre docking site for binding Pvc13 fibres when the syringe is in its extended state. Upon sheath contraction (i.e payload injection), the sheath’s outer diameter increases from 17 to 22 nm, as shown in Figure 1.15. This conformational transition increases the sheath’s inner diameter to 10 nm and results in separation of the sheath from the tube since the tube’s outer diameter stays constant at 8 nm. In addition, each subunit’s termini are rearranged such that the sheath is vertically compressed to push the spike and tube into the cell surface when contracted; the sheath’s helical rise, or advancement per subunit, decreases from 39.3 Å to 17 Å, and its helical twist increases from 19.9° to 31.4°.

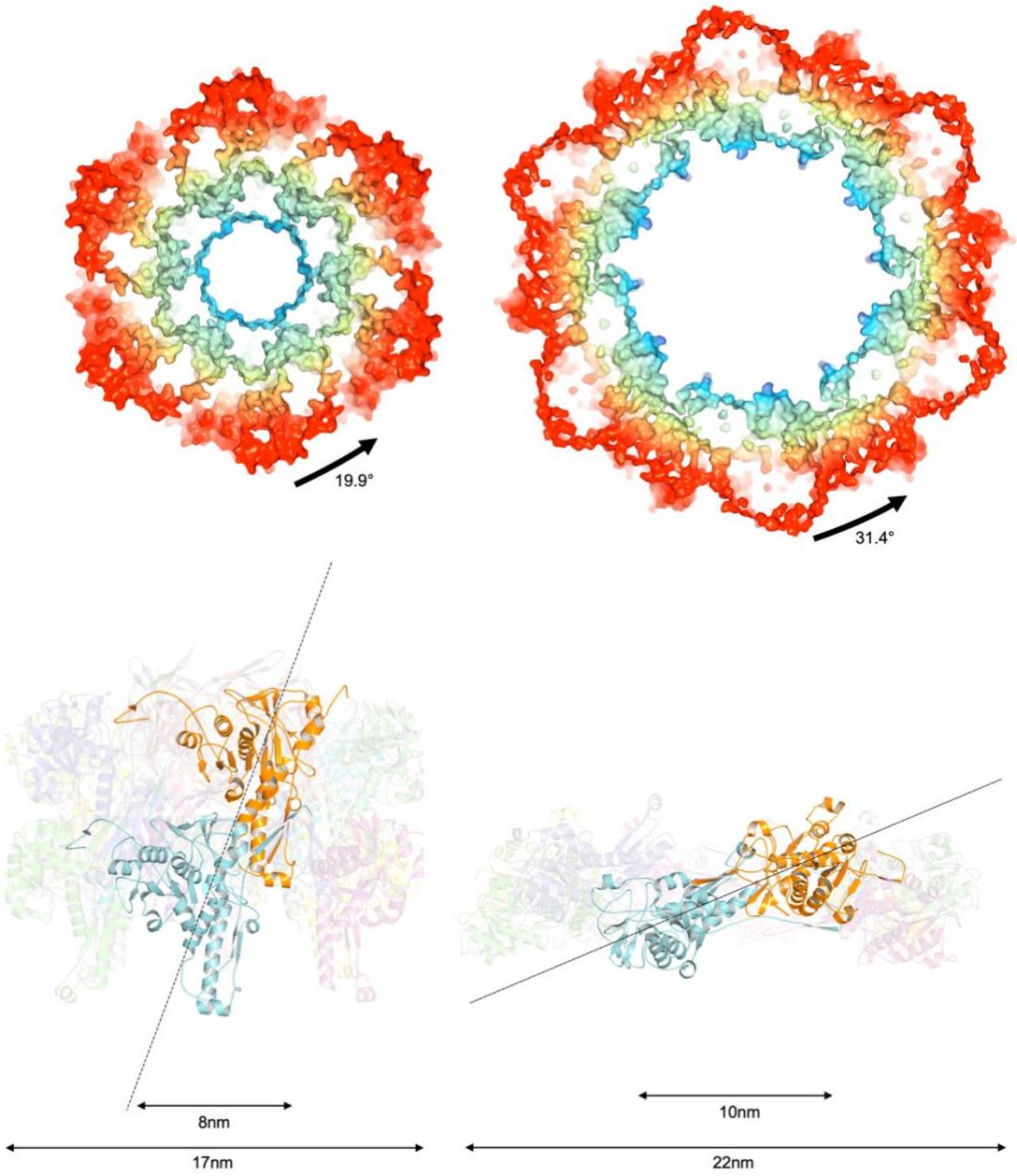


Figure 1.15 | PVC sheath conformations are altered by contraction.

Sheath monomers are coloured by chain in the extended (left) and contracted (right) states with top and side views. Upon contraction, each disc of hexamers gets closer together (the helical rise to the next subunit increases from 39.3Å to 17.0Å), whilst the subunits in a given ridge increase their helical twist from 19.9° to 31.4°. Contraction also causes the sheath's diameters – inner and outer – to increase: 8 nm to 10 nm and 17 nm to 22 nm, respectively. PDB IDs: 6J0B and 6J0C, respectively. The top view is coloured according to the proximity to the centre of the tube.

Though the exact interactions between sheath subunits have not been clarified in the PVC, in another homologous tailed structure called the R-type pyocin, each sheath subunit has two long protrusions extending to the disc below it; one provides a β-strand to the subunit directly below it whilst the other extends to the subunit in the neighbouring ridge (Ge *et al.*, 2015). This interaction, termed the 'β-handshake', is also used to anchor the sheath to the baseplate and cap (Ge *et al.*, 2015). This mechanism is also employed by the Afp (Desfosses *et al.*, 2019) so it is safe to assume a similar structure is used to connect the sheaths of the PVC. The β-handshake is often maintained during contraction and, in the Afp, a rigid body rotation between alternating Afp2-Afp3 sheath subunits of 33° is observed, whilst the sheath itself compresses to half of its original length.

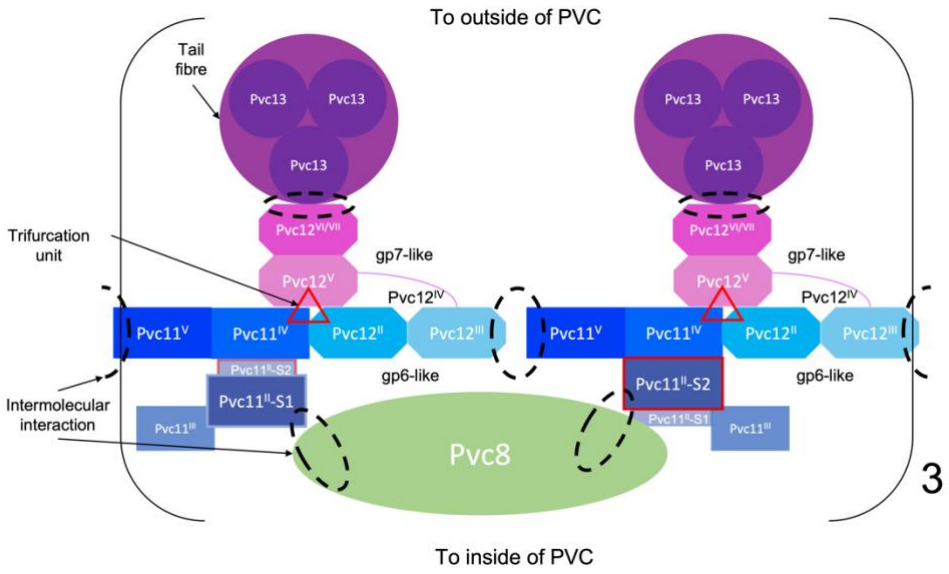
1.3.1.2 Stabilisation: The Pvc16 Cap

The Pvc16 terminator cap appears to be an important stabiliser of the tube-sheath complex. In fact, Δ*pvc16* and Δ*pvc2* mutants produce PVC particles lacking an outer sheath (Jiang *et al,*. 2019). Each Pvc16 monomer is a dumbbell-like shape whereby its N- and C-terminal domains are connected by a single β-sheet middle linker. The N-terminal domain consists primarily of a long loop that interacts with two Pvc1 monomers of the tube, and the N-terminus which interacts with another Pvc1. The middle linker and the C-terminal domain interact with one Pvc2 monomer each. In summary, each Pvc16 monomer stabilises three Pvc1 monomers, and two Pvc2 monomers. These Pvc16 units link to become a hexameric cap of the 80Å inner pore formed by Pvc1, leaving only 7Å at the centre of the Pvc16 hexameric ring (Jiang *et al,*. 2019).

1.3.1.3 The Needle: Spike, Tip, and Baseplate

To form a continuous channel from the inner tube through to the spike and needle tip (Pvc8 and Pvc10, respectively), the baseplate wedge stabilises the complex. A Pvc8 trimer forms the main body of the spike that attaches to the sharp conical Pvc10. The baseplate, meanwhile, has two primary interfaces: the interface of Pvc11 with Pvc8 (Pvc11:Pvc8), and the interface between the Pvc11-Pvc12 heterodimers and that of Pvc9-Pvc7 (Pvc11-Pvc12:Pvc9-Pvc7). In the ring structure encircling Pvc8, composed of 6 copies of Pvc11-Pvc12 dimers (a dodecamer), the 6 copies of Pvc11 domain II (designated here as Pvc11-II or $Pvc11^{II}$) form the inner ring of the baseplate that encircles the Pvc8 spike. As seen in Figure 1.16, this circle is formed by means of a 'trifurcation unit' consisting of domains II and V of Pvc12 with domain IV of Pvc11. From here, Pvc11-V can bind Pvc12-III from a neighbouring trifurcation unit to form one component of the hexameric ring. Meanwhile, Pvc12-VI and VII protrude to bind to the Pvc13 tail fibres. Notably, due to the structural asymmetry of two Pvc11-II domains binding two different regions of the same Pvc8 monomer, the structure of Pvc11-II alternates between two conformations termed 'state I' and 'state II', encoded separately and flanking the Pvc11-III domain in the primary sequence (Figure 1.16). Although the main difference between $Pvc11^{II}$-S1 and $Pvc11^{II}$-S2 is a rigid body rotation, the primary sequences of these domains differ vastly.

A



B

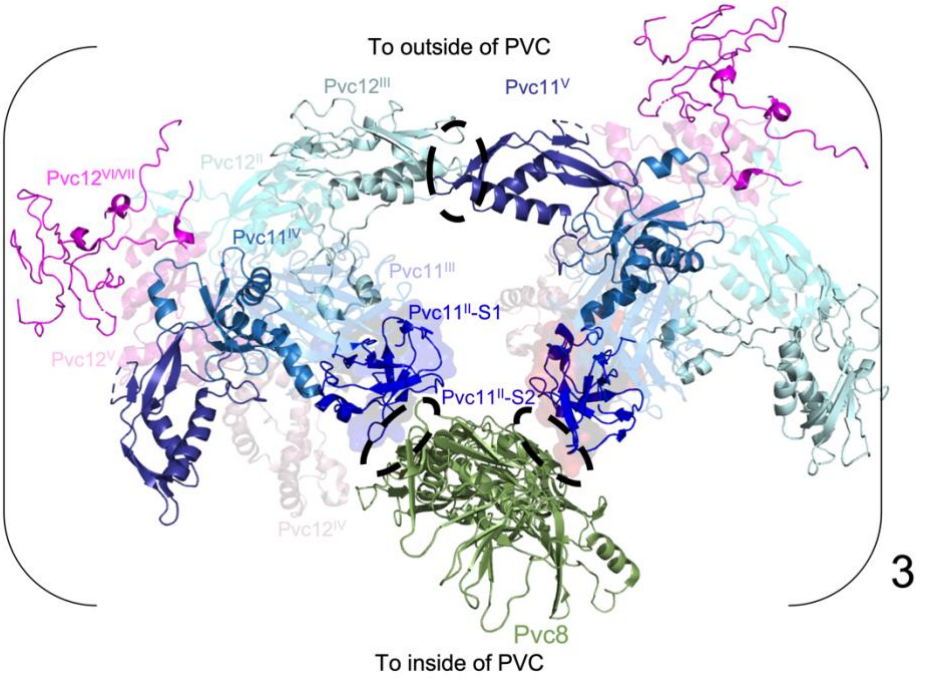


Figure 1.16 | Schematic of 'interface 1' of the PVC baseplate.

Each trifurcation unit is composed of a Pvc11 domain IV (blue), a Pvc12 domain II (gp6-like; cyan), and a Pvc12 domain V (gp7-like: pink). Each unit is connected to neighbouring units via the bond between Pvc11-V and Pvc12-III. Domain VI/VII of Pvc12 coordinates one Pvc13 trimer tail fibre, whilst every neighbouring pair of Pvc11-II coordinates a single Pvc8 monomer. Due to each Pvc11-II binding two different parts of the same Pvc8 monomer,

this structural asymmetry results in two conformations of Pvc11-II: state I (S1) and state II (S2). This structure is repeated for each Pvc8 monomer to form the full baseplate for the syringe (i.e. 3 times). **(B)** The schematic in **(A)** is reconstructed in molecular format using PDB IDs 6J0M and 6J0N from the data of Jiang et al., 2019. As in **(A)**, pink colours represent gp6-like domains and blue regions represent gp7-like domains; the Pvc8 monomer is shown in green. Each schematic is viewed from the proximal side of the syringe looking towards the distal side.

Lastly, the second interface associates the baseplate to the central hub. Pvc9 and the LysM domain from Pvc7 bind and stabilise each Pvc11-Pvc12 dimer. Both of interfaces 1 and 2 are essential for structure assembly: Δ*pvc1*, Δ*pvc3-5*, Δ*pvc7-9*, and Δ*pvc11-12* mutants do not produce intact PVCs (Jiang *et al,*. 2019).

Whilst the PVC appears to have a PAAR domain within Pvc10, it is unlikely to be used for payload binding, as in the case of the T6SS, since PVC payloads are loaded exclusively to the internal tube as demonstrated by ImmunoGOLD-TEM (Vlisidou *et al.*, 2019) and electron density mapping (Wang *et al.*, 2021).

1.3.1.4 Cell Binding: Tail Fibres

Pvc13 tail fibres facilitate binding of the complex to host cell membranes and may determine the target specificity of the PVC. The tail fibres are composed of a trimer of Pvc13 polypeptides with three main components (Kreitz *et al.*, 2023): the N-terminus, made up of 3 helices; the shaft, comprising at least 10 repetitive β-sheets, each 17 amino acids in length, which shares homology with the adenovirus type 2 shaft, and the C-terminal domain which is homologous to gp10/12 from phage T4 and appears to stabilise the 'tube' structure connecting the phage-like fibre tip to the N-terminus. The Pvc13 trimeric fibres dock on both Pvc12 domain VII as well as the Pvc3 docking

sites on layers 3, 5, and 7 which retains the tail fibres closely compacted to the PVC sheath until recognising a target cell when they appear to extend to contact the membrane.

It is not known what the binding target is for the native tail fibres, but Kreitz and colleagues (2023) demonstrated that Sf9 insect cells are one of a few cell lines which display cytotoxicity when injected with native payloads such as Pnf. J774A.1 macrophages were also shown to be susceptible to PVC*pnf* loaded with Pnf or Pdp1, regardless of whether Pvc13 tail fibres were deleted. For other cell lines, Pvc13 can be retargeted by substituting the putative receptor-binding domain, comprising residues 403-477, with an adenoviral binding domain, such as Ad5-knob, or epitope tag binding domains from antibodies such as α-HA, α-FLAG etc. Thus, *Cre* recombinase, encoded as a heterologous PVC payload, results in *loxP*-GFP expression from the injected cells: the first published evidence of a bioengineered PVC (Kreitz *et al.*, 2023).

1.3.1.5 Structural PVC Gene Homology

As mentioned, each gene of the PVC operon shares great sequence, structural, and functional similarity to other caudate (tailed) contractile injection systems (CISs) such as bacteriophage T4, R-type pyocin, and T6SS. The homologues and functions are shown in Table 1.2. Interestingly, the LysM domain of Pvc7 appears to be highly conserved to that of the T4 gp53, the phage P2 gpX, and PA0627 in R-type pyocin, implying that the LysM domain is indispensable for function of the baseplate (albeit fused to a different component depending on the caudate structure). The association of Pvc12 with Pvc11 – a gp6 homologue – is reminiscent of the $(gp6)_2$-gp7 heterotrimer in T4 phage (Taylor, van Raaij and Leiman, 2018).

Table 1.2 | Contractile injection system homologues.

Isofunctional homologues in the *Photorhabdus* virulence cassette (PVC), *Serratia entomophila* anti-feeding prophage (Afp), phage T4, *P. aeruginosa* type 6 secretion system (T6SS), R-type pyocin, and *P. luteoviolacea* strain HI MAC. VipAB are the homologous sheath proteins used in *Vibrio cholerae*.

* = Unless other phage is specified in parentheses. ** = predicted structure (Barret et al., 2011). ° = domains I, III/IV-VII. Adapted from Jiang et al., 2019.

PVC	Afp	Phage T4*	T6SS	R-type pyocin	MAC	Function
Pvc1	Afp1	gp19	TssD/Hcp	PA0623	AHX39701-2.1 (macT)	Tube
Pvc2	Afp2	gp18	TssBC/VipAB	PA0662	AHX39703.1 (macS)	Sheath
Pvc3	Afp3	-	-	-	-	Sheath
Pvc4	Afp4	-	-	-	-	Sheath
Pvc5	Afp5	gp19/gp54	-	-	AHX39701-2.1	Tube initiator
Pvc6	Afp6	-	-	-	-	*Unknown*
Pvc7	Afp7	gp48/gp53; gpX (P2)	TssG/PA5265**	PA0627	AHX39698-9.1	Tube initiator (LysM)
Pvc8	Afp8	gp5/gp27 (spike/hub)	TssI/VgrG	PA0616/PA0628	AHX39697.1	Spike
Pvc9	Afp9	gp25	TssE	PA0617	AHX39696.1	Baseplate
Pvc10	Afp10	gp5.4	PAAR-repeat motif	PA0616	-	Tip
Pvc11	Afp11	gp6	TssF	PA0618	AHX39695.1 (macB)	Baseplate
Pvc12	Afp12	gp6/gp7°	TssF/TssG	PA0619	AHX39694.1	Baseplate
Pvc13	Afp13	gp10/gp12	N/A	PA0620	AHX39693.1, AHX39623.1	Tail fibre
Pvc14	Afp14	gp29	N/A	PA0625	AHX39690.1	Tape measure protein
Pvc15	Afp15	-	TssH/ClpV	-	AHX39625.1	AAA+ ATPase
Pvc16	Afp16	gp15/gp3; gpU/gp3 (λ)	TssA	PA0626	AHX39624.1	Tube-sheath terminator

The spike and tip proteins, Pvc8 and Pvc10, appear to be an example of evolutionary reshuffling; the spike-hub complex is formed by a homotrimer of Pvc8, a gp27-gp5 homologue, and Pvc10, a PAAR (Proline-Alanine-Alanine-Repeat)-containing protein homologue which forms the spike tip and may be stabilised by the presence of a metal ion such as Zn or Fe (Shneider *et al.*, 2013). In T4 phages, three proteins constitute these components: the hub is formed by gp27, the spike by gp5, and the tip by gp5.4. More similarly to the PVC, however, the T6SS adopts a single homotrimeric polypeptide for the spike and hub: VgrG (the Pvc8 / gp5-gp27 orthologue) forms a β-helical prism, and a PAAR protein (Pvc10 orthologue) for the tip (Leiman *et al.*, 2009; Park *et al.*, 2018). Additionally, the T6SS baseplate components are known structural and sequence homologues of T4 phage: TssE is a gp25 homologue, TssF is a gp6 homologue, and TssG is a gp7 homologue. Altogether, there is a common ancestry between each of these components (Chen *et al.*, 2019). Similarly to Pvc8's homology to gp5-gp27, a fusion event may have constituted the homotrimeric structure of the central spike in the T6SS VgrG (Leiman *et al.*, 2009; Barret *et al.*, 2011).

Many – but not all – caudate structures possess binding fibres which are ancestral homologues. For example, the function of Pvc13 is likely an ancestral fusion of gp10/12, from phage T4, and adenovirus fibre, which Jiang and colleagues (2019) speculated to have enabled the recognition of eukaryotic cell membranes. The mechanism of how this binding fibre was adopted in a bacterial operon is not currently agreed upon; adenoviruses primarily infect mammals, though some reptilian adenoviruses also exist (Menéndez-Conejero *et al.*, 2017). Alternatively, adenovirus-like fibres in PVCs could be an example of convergent evolution. Notably, however, it is still unknown what component of the cell membrane that native PVC tail fibres bind. As far as current knowledge suggests, PVCs exclusively inject into eukaryotic cells.

Most of the effectors of the PVC also have homologues likely obtained by horizontal gene transfer (HGT); these are discussed in chapter 1.3.2 starting on page 47.

1.3.1.6 The Elusive Components: Pvc6, 14, and 15

PVC mutants lacking Pvc6, 10, 13, 14, or 15 are still able to produce intact PVCs, though they do not necessarily translocate their payload (Jiang *et al.*, 2019, 2022). Since some of the functions of Pvc6, 14, and 15 have yet to be defined, it is useful to look to the largely homologous structures in *S. entomophila*: the Afp. To this end, Pvc14 could act as a 'tape-measure protein' (TMP): regulating the sizes of PVC particles in much the same vein as that of the TMPs in virtually all non-contractile phages (Pedulla *et al.*, 2003).

In the Afp, the TMP is theorised to stretch along the length of the assembling tube and prevent the premature attachment of the Afp16 cap (sometimes referred to as the tail-length termination protein, or TrP) (Rybakova *et al.*, 2013, 2015). In fact, there appears to be a direct linear relationship to the size of Afp14, whether truncated or elongated, and that of the length of the resulting Afps ($R^2 = 0.92$). The 98 N-terminal amino acids of Afp14 are essential for Afp baseplate assembly, which is analogous to the N-terminus of the T4 phage gp29 TMP (Rybakova *et al.*, 2015; Bhardwaj, Mitra and Hurst, 2021). Each of the ΔPvc14, ΔPvc6, and ΔPvc16 PVCs have abnormal lengths compared with wild-type, so Pvc6 may interact with Pvc14 in a manner which regulates PVC length (Jiang *et al.*, 2019).

The PVC has far greater variability in length than the Afp which may indicate a different set of interactions between the Afp14 and Pvc14 TMPs and their respective distal sheath proteins. Notably, the mature Afp sheath alternates between Afp2 and Afp3 throughout the entire length of the syringe; 77% of purified Afp particles contain an Afp16-Afp2 interaction, whilst the

remaining bind Afp3; the PVC sheath, meanwhile, is made up of Pvc2 exclusively at the distal end (Jiang *et al.*, 2019).

Pvc15 is very highly conserved even between all classes of the PVC. Pvc15 is an ATPase associated with diverse cellular activities (AAA+) which was first speculated to function in loading of the toxin into the complex (Jiang *et al.*, 2019). A homologue is not found in R-type pyocin nor T4 phage, but based on its similarity to Afp15, it has also been hypothesised to be involved in baseplate and sheath assembly (Hurst, Glare and Jackson, 2004; Rybakova *et al.*, 2013; Kube and Wendler, 2015; Hurst *et al.*, 2018). Later experiments confirmed the co-immunoprecipitation of Afp15 with the Afp16 cap protein, which itself co-purifies with the Afp3 sheath protein (Bhardwaj, Mitra and Hurst, 2021). Whilst it may be tempting to speculate a similar role for Pvc15, unlike Afp15 which is stably produced on its own and is essential for particle assembly, Pvc15 is not stably expressed on its own at levels detected by western blot (WB) and ΔPvc15 PVCs are often found intact (Jiang *et al.*, 2019, 2022).

It should also be noted that the T6SS Pvc15 homologue, TssH/ClpV, is associated with recycling of the complex to its smaller molecular components. A function in disassembly would, presumably, not be required for the secreted single-use PVCs. In contrast, Hurst, Glare and Jackson (2004) predicted that Afp15 is used for disassembly of the protein complex (Neuwald *et al.*, 1999).

Overall, PVC nucleation begins with assembly of the baseplate Pvc11-Pvc12 hexamers to the Pvc8-Pvc10 spike-tip complex. Initiation tube components Pvc5 and Pvc7 enable docking of baseplate wedge Pvc9 and the sheath Pvc4 as prerequisites for Pvc1 tube assembly and Pvc13 attachment to the baseplate. Finally, binding of the Pvc16 tube terminator enables sheath polymerisation of Pvc2 and Pvc3 . Crucially, it is unknown at what stage payload association occurs and whether payload is associated with the inner tube before or after stacking. In addition, whilst PVCs are released after cell death, it is unknown if other secretion mechanisms of the structure exist in living cells.

1.3.2 PVC Toxin Payloads

The payloads of the PVC are often – but not always – encoded downstream of the final structural gene in the operon. Patented work from the Waterfield lab first identified that the N-termini of native PVC payloads are sufficient to load polypeptides into the mature nanosyringe. Recent studies have confirmed that the payload's N-terminal region of roughly 50-70 amino acids acts as a leader sequence (LS), or packaging signal also called the 'signal peptide', which interacts with core PVC proteins to facilitate PVC association (Jiang *et al.*, 2022; Kreitz *et al.*, 2023). Encoding these natural LSs N-terminal to unnatural (i.e., heterologous) proteins enables their co-purification with the PVCs.

The PVC is an attractive means with which to deliver proteins into the cytosol as, currently, there have not been any known limitations to the types of heterologous polypeptides that can be loaded. Jiang and colleagues (2022) used the Pnf and Pdp1 N-terminal sequences to load mRFP, Nano- and *Renilla* luciferases, and β-lactamase (BlaM) to observe fluorescence, with their respective substrates, in J774 macrophages incubated with loaded PVCs. The researchers observed loading for payloads ranging from 10 to 132 kDa as well as a range of isoelectric points (4.5-9.9) of proteins from bacteria, eukaryotes, and fungi.

Electron density mapping by Wang and colleagues (2021) showed that Pnf and Pdp1 – proteins of about 40 kDa each – are loaded into the PVC's Pvc1 tube lumen in a 'pea in the pod' manner: many payload proteins are stacked on top of one another throughout the tube lumen. Whether larger proteins are loaded in an alternative manner, or whether a particular number of payload molecules are loaded per Pvc1 ring, is yet to be more thoroughly explored. In addition, 'regulatory' genes found at this operon (PAU_RS16560-70 and PAU_RS24015) are thought to act as important proteins required for the correct assembly of PVC*pnf* (Jiang *et al.*, 2022; Kreitz *et al.*, 2023). However, these genes were found

to be dispensable for mature PVC assembly in the work presented here since mature PVC*pnf* could be obtained from plasmids even in their absence.

In *P. asymbiotica* (Pa) ATCC43949, there are 5 PVC operons, two of which – PVC*lopT* and PVC*cif* – are also found in *P. luminescens* TT01 (Pl). Many of the PVC effectors modify small Rho-subfamily GTPases: molecular switches which depend on whether they are bound with GDP (inactive) or GTP (active). The switch 'on' or 'off' is mediated by guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs) (Hodge and Ridley, 2016). These molecular switches have roles in regulation of the cell cycle, responses to cell surface receptors, and cell differentiation and motility (Nakamura, 2013). The exact roles of the GTPase are often mediated by the combinatorics of the many parallel signalling pathways; how certain PVC effectors dysregulate these processes will be discussed in this subchapter.

1.3.2.1 Pnf

Pnf exhibits homology to the *E. coli* cytotoxic necrosis factor (CNF1, CNF2, and CNF3) (Knust and Schmidt, 2010; Vlisidou *et al.*, 2019), and Cnf1γ in *Yersinia pseudotuberculosis*. Pnf was first characterised by transfection into NIH-3T3 mammalian tissue culture cells by Yang and colleagues (2006) in which it caused rich actin-cabling. Pnf encodes only the catalytic domain of the CNF toxins and has a similar mode of action, as confirmed by Vlisidou and colleagues (2019).

CNF toxins – and Pnf by extension – deamidate RhoA at Gln63, preventing GTP hydrolysis by GAPs to inactivate and sequester RhoA by binding guanine dissociation inhibitors (GDIs) (Figure 1.17). Constitutively activated RhoA causes cell proteins such as mammalian homologue of *Drosophila* diaphanous (mDia) to build actin filaments, whilst Rho-associated coiled-coil forming kinase (ROCK) maintains them by activating LIM-kinase and phosphorylates myosin light chain (MLC) and myosin light chain phosphatase (MLCP)

(O'Connor and Chen, 2013). This combination of interactions leads to the formation and maintenance of bundles of parallel myosin and actin filaments termed actomyosin stress fibres.

Other Rho-family GTPases such as Cdc42, RhoB, and RhoC can be targeted by the other CNF isoforms; Pnf also targets Rac1 for deamidation (Hoffmann *et al.*, 2004; Knust and Schmidt, 2010; Vlisidou *et al.*, 2019). Activation of both potentially conflicting signals from Rac1 and RhoA may dysregulate cell fate, division, and migration. In addition, CNF and Pnf may be involved in cytokine dysregulation in innate immunity since Rac1 induces pro-inflammatory IL-1β secretion (Diabate *et al.*, 2015) whilst RhoA inhibits inflammasome activation via pyrin phosphorylation (Park *et al.*, 2016).

Despite its name, CNF only causes necrosis or lysis in neutrophils *in vivo*, primarily causing apoptosis in epithelial cell lines (Russo *et al.*, 2005). On the other hand, CNF has been shown to promote cell survival by upregulation of Bcl-2, likely to establish a chronic intracellular uropathogenic *E. coli* infection (Fiorentini *et al.*, 1998; Thomas *et al.*, 2001).

Interestingly, CNF1 does not require an injection system for cell entry as it is capable of receptor-mediated endocytosis via its N- and C-termini (Chaoprasid *et al.*, 2021). It is unknown whether Pnf has similar mechanisms of cell entry without the PVC, particularly given that Pnf does not encode N-terminal receptor binding domains (Knust and Schmidt, 2010).

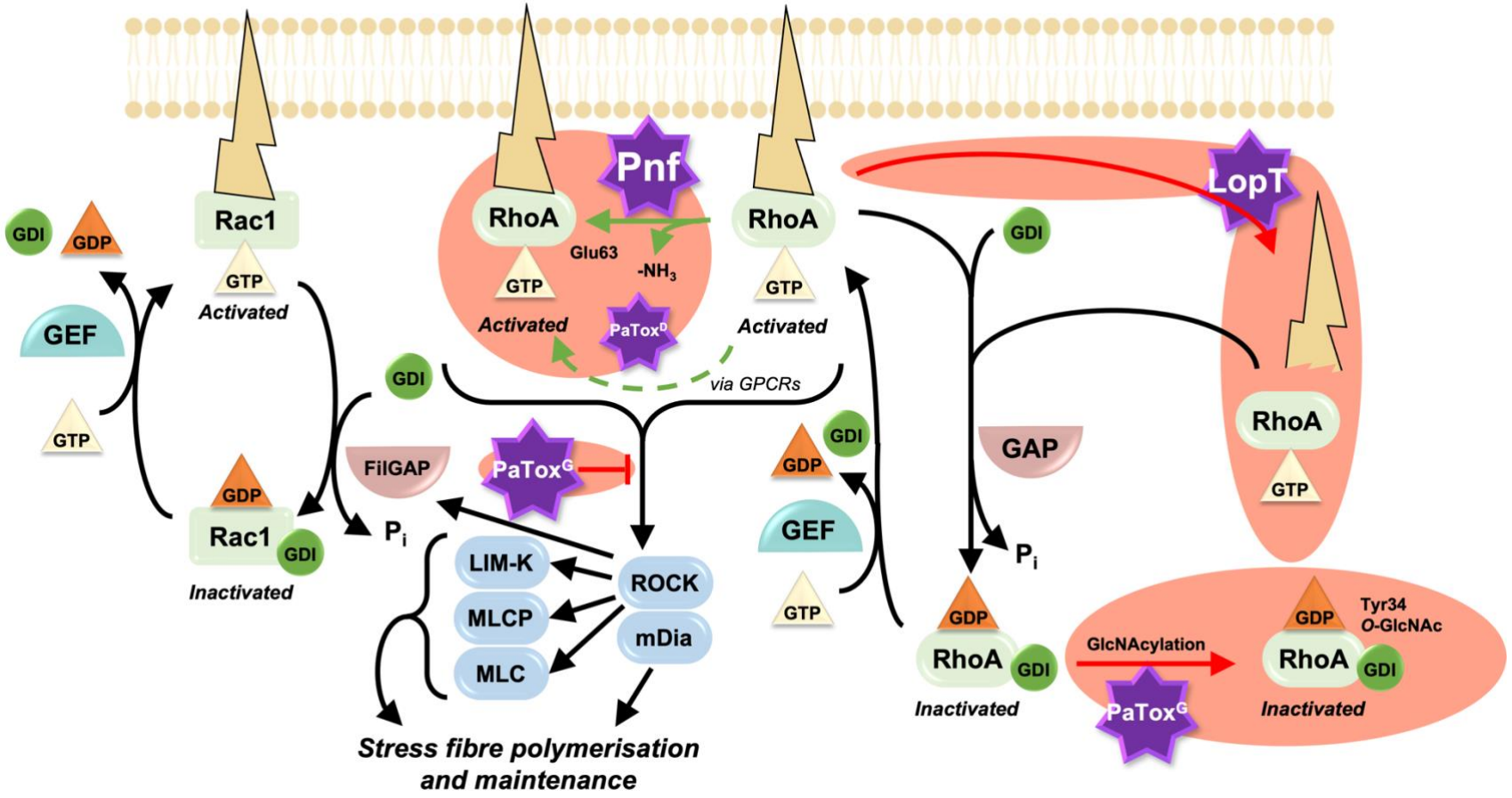

Rac1
GDI
GDP
GTP
Activated
GEF
GTP
GDP
Rac1
GDI
Inactivated
RhoA
Pnf
Glu63
-NH3
GTP
Activated
PaToxD
RhoA
GTP
Activated
via GPCRs
GDI
FilGAP
PaToxG
Pi
LIM-K
MLCP
MLC
ROCK
mDia
Stress fibre polymerisation
and maintenance
GDI
GDP
GEF
GTP
LopT
GAP
Pi
RhoA
GTP
GDP
RhoA
GDI
Inactivated
GlcNAcylation
PaToxG
GDP
Tyr34
O-GlcNAc
RhoA
GDI
Inactivated

Figure 1.17 | *Photorhabdus* effectors modify Rho-subfamily GTPases.

Activated Rho GTPases such as RhoA and Rac1 bind GTP through interaction with GEFs and localise to the plasma membrane. Activated RhoA promotes cell cycling and induces stress fibres via ROCK and mDia; Rac1 signalling is then inhibited by FilGAP. When deactivated, GAPs replace GTP with GDP and GTPases are sequestered by GDIs.

Pnf acts to promote cell cycling whilst inducing stress fibres to inhibit cytokinesis; both Pnf and the deamidase domain of PaTox ($PaTox^{D}$) act by sequestering Rho-family GTPases in their active GTP-bound states. Meanwhile, the glycosyl transferase domain of PaTox ($PaTox^{G}$) sequesters RhoA in an inactive form through UDP-GlcNAcylation of Tyr34; this also prevents interaction of activated RhoA with its downstream effectors and works to dysregulate RhoA control in conjunction with $PaTox^{D}$. LopT sequesters RhoA by displacement of the GTPase from the membrane, causing it to bind cytosolic GDIs. Orange areas indicate effector pathways; green arrows indicate RhoA activation and red arrows indicate RhoA inactivation.

1.3.2.2 LopT: a Class II PVC Effector

Sharing homology to the *Yersinia* outer protein T (YopT) cysteine protease, LopT shares the conserved catalytic C-terminus (Brugirard-Ricaud *et al.*, 2005). YopT enters the target cell, along with its chaperone *sycT*, via the T3SS; LopT enters via the PVC along with its chaperone *slcT*. In the host cell, RhoA, Rac, and Cdc42 are localised to the membrane by prenylation of their C-termini; YopT cleaves the prenylated cysteine and releases the GTPases to the cytosol to be inactivated by cytosolic RhoGDIs (Shao *et al.*, 2002; Garcia-Mata, Boulter and Burridge, 2011; Schaefer, Reinhard and Hordijk, 2014) (Figure 1.17). The inhibition of Rho GTPases causes cell rounding by disruption of the cytoskeleton and prevents phagocytosis of *Yersinia* from macrophages (Aepfelbacher *et al.*, 2003); these effects are retained when heterologously expressing LopT in *Y. pseudotuberculosis* (Brugirard-Ricaud *et al.*, 2005).

Curiously, PVC*lopT* is the only PVC locus known to not encode Pvc13 tail fibres. It is unknown how, if at all, this mechanism of contraction differs to the more classical class I PVCs of the *pnf* and *cif* loci.

1.3.2.3 PaTox: a Class III PVC Effector

Photorhabdus asymbiotica toxin (PaTox) is one of the largest PVC payloads and is encoded downstream of the unique "class III" PVC locus (Vlisidou *et al.*, 2019). Intriguingly, PaTox is capable of entering the cytosol without the PVC through a mechanism involving the acidification of the early endosome (Jank *et al.*, 2013); its dependency on the PVC itself has yet to be explored. Its theoretical size is 334.9 nDa and consists of three known domains: the protease (P), deamidase (D), and glycosyltransferase (G) domains.

The P domain (PaTox^{P}) shares homology to a class of C58-proteases such as the MARTX toxin from *Vibrio vulnificus*, discussed in chapter 1.1.2, and the

Yersinia YopT, discussed in the previous chapter. PaTox^{P} has been shown to enhance cytotoxicity in HeLa cells (Bogdanovic *et al.*, 2019).

Both the G and D domains were characterised in a paper by Jank and colleagues (2013). PaTox^{G} is a class of glycosyltransferases which contains a conserved DxD motif and is a functional homologue of Afp18 from *Yersinia ruckeri* (Jank *et al.*, 2015). Homologues such as TcdA and TcdB from *C. difficile* and TpeL from *C. perfringens* are typical examples of this class of effector which act at serine or threonine residues of small GTPases. PaTox^{G} transfers mono-*O*-linked β-N-acetylglucosamine (O-GlcNAc) to tyrosine residues of host Rho-GTPases: Y32 of Rac and Cdc42, and Y34 of RhoA. This modification prevents binding of GEFs and GAPs to the GTPases, inhibiting interactions with downstream components. PaTox^{G} also acts upon RhoB and C as well as Rac1, 2, and 3. Figure 1.17 depicts PaTox^{G} sequestering the inactive GDP-bound RhoA though it can also target active GTP-bound RhoA (Jank *et al.*, 2013).

Finally, the D domain (PaTox^{D}) shares roughly 20% similarity to the deamidase domain of *Pasteurella multocida* toxin (PMT). In contrast to the G domain, PaTox^{D} deamidates Gln205 in Gα proteins of G protein-coupled receptors (GPCRs) and locks Gα_{q}-family proteins in an active state, thus activating RhoA (Figure 1.17). Constitutive RhoA activation by PaTox^{D} and inactivation by PaTox^{G} presumably leads to cytoskeletal dysregulation, producing the lethal effect in *Galleria mellonella* larvae (Jank *et al.*, 2013)

1.3.2.4 Cyclomodulins

Bacterial effectors capable of modulating the host's cell cycle – whether promoting or inhibiting it – are collectively termed cyclomodulins (Oswald *et al.*, 2005). Since the cell cycle is a universal property of eukaryotes, cyclomodulins are an effective toxin to vertebrate and invertebrate hosts. There are three notable PVC payloads that influence the cell cycle: Pnf, Cif, and RRSP. Both Pnf and RRSP perturb Rho family GTPases which are involved in the cell

cycle. The set of actions of each effector on cell cycling is summarised in Figure 1.18.

As discussed in chapter 1.3.2.1, Pnf causes RhoA to be constitutively activated. As well as stress fibre formation, Pnf subsequently causes exit from G_0 and progression through the G_1 checkpoint to the S phase of the cell cycle (Nougayrède *et al.*, 2005). The undergoing of DNA replication with such vast destabilisation of the cytoskeleton inhibits cytokinesis and subsequently causes multinucleation and aneuploidy (Oswald *et al.*, 2005; Knust and Schmidt, 2010).

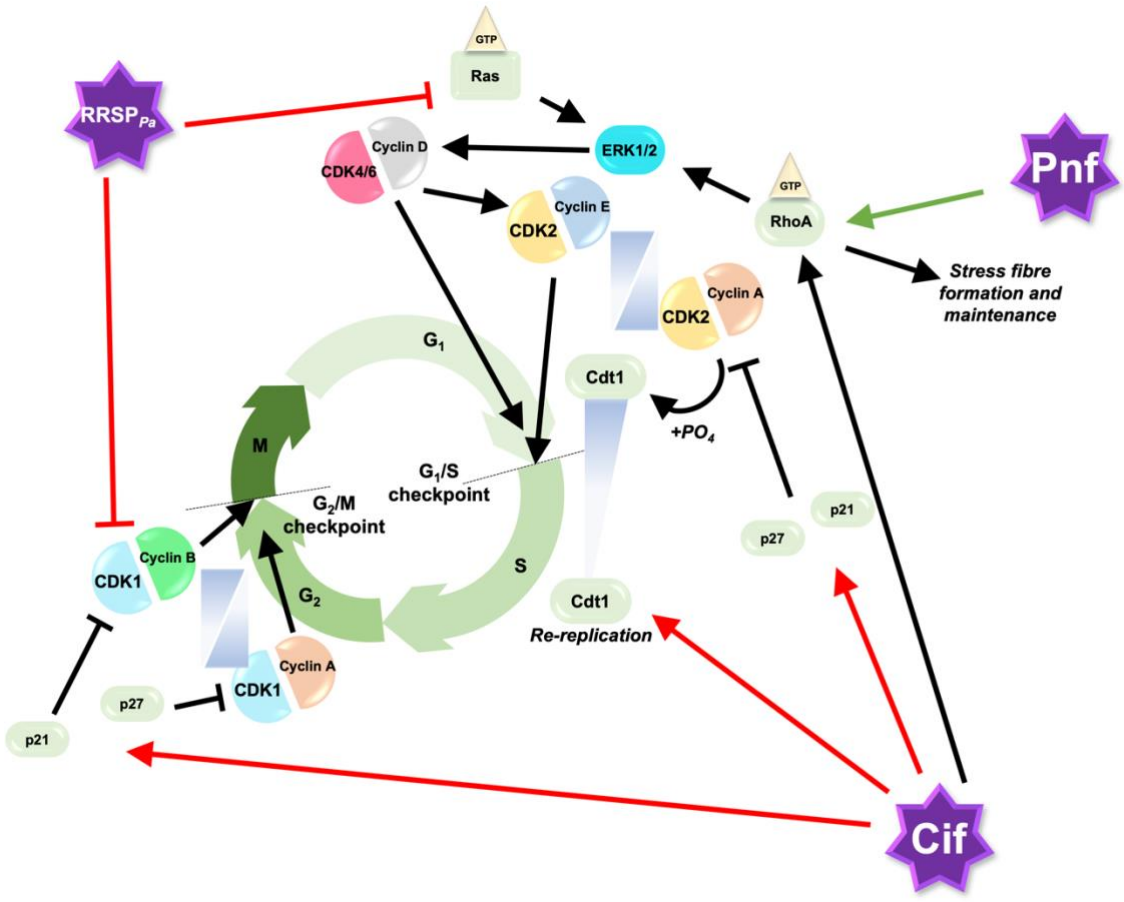


Figure 1.18 | PVC effectors modulate the cell cycle.

During infection, *Photorhabdus* uses cyclomodulins to hijack cellular processes of the host. Green arrows represent induction or promotion of cell cycling whilst red arrows indicate its inhibition. The *Photorhabdus* effectors are indicated as purple stars. Pnf constitutively activates RhoA which stimulates cell cycling but also causes the formation and maintenance of stress fibres, inhibiting cytokinesis and causing multinucleation and aneuploidy. Cif causes the

accumulation of inhibitors of the cell cycle: p21, p27, and Cdt1 but also accumulates RhoA to induce stress fibres. RRSP encoded in *P. asymbiotica* ($RRSP_{Pa}$) cleaves Ras and inhibits cyclin-dependent kinase (CDK) 1 phosphorylation required at the G_2/M checkpoint to stimulate mitosis.

1.3.2.4.1 $RRSP_{Pa}$

Previously termed domain of unknown function (DUF) 5, the Ras/Rap1-specific endopeptidase (RRSP) first identified in *Vibrio vulnificus* isolates ($RRSP_{Vv}$) cleaves within Ras (Rat sarcoma) to prevent ERK1/2 phosphorylation and inhibit cell proliferation (Antic *et al.*, 2015) (Figure 1.18). $RRSP_{Pa}$ is encoded at the *Pa*PVC*lopT* locus and inhibits cyclin-dependent kinase 1 (CDK1) which prevents mitosis and causes apoptosis in HeLa cells (Wang *et al.*, 2020).

1.3.2.4.2 Cif

Cycle-inhibiting factor (Cif) proteins are a family of cysteine proteases: proteins which have a crucial cysteine in the active site. Homologous *cif* sequences found in enteropathogenic and enterohemorrhagic *E. coli*, *Y. pseudotuberculosis*, and *B. pseudomallei* are primarily used by the type 3 secretion machinery and are a subtype of orthologues obtained by horizontal gene transfer, termed xenologues (Jubelin *et al.*, 2009). Cif encodes a highly conserved triad of residues (C109, H165, and Q185) essential for its function, as well as the first 16 N-terminal amino acids which are required for translocation by the T3SS machinery (Taieb *et al.*, 2006); a similar N-terminal chain is found in Cif encoded in the PVC*cif* locus so may also be used for trafficking into the PVC.

Cif results in many cytopathic effects since it inhibits the proteasomal degradation of various signalling molecules by preventing the activation of a specific family of cullin-RING E3 ubiquitin ligases (CRLs) (Figure 1.19). CRLs transfer activated ubiquitin from E2 ubiquitin conjugating enzyme to substrates

for targeting to the proteasome for degradation. CRLs are composed of i) the cullin scaffold protein, ii) a substrate recognition module (SRM), and iii) an activator molecule called NEDD8: the direct target of Cif in the nucleus (Jubelin *et al.*, 2010; Morikawa *et al.*, 2010). Cif itself does not encode a classical nuclear localisation signal, so how it localises to the nucleus to affect NEDD8 is unclear. At roughly 30 kDa in size, however, there is a possibility that Cif enters the nucleus quite simply by slow diffusion (Wang and Brattain, 2007).

Under normal conditions, active NEDD8-cullin-RING-E2 complex ubiquitylates both the substrate and SRM. Then, the COP9 signalosome hydrolyses the bond between NEDD8 and the cullin subunit of the inactive CRL complex. Deneddylation enables CAND1 to bind the cullin-RING complex which is subsequently displaced by a new SRM and reactivated after the binding of NEDD8. Cif hijacks this process by directly binding and deamidating Gln40 of NEDD8, effectively impairing hydrolysis by COP9 and rendering the NEDD8-CRL complex constitutively inactivated, and leading to the accumulation of the target substrates of the CRL (Figure 1.19) (Cui *et al.*, 2010; Crow *et al.*, 2012).

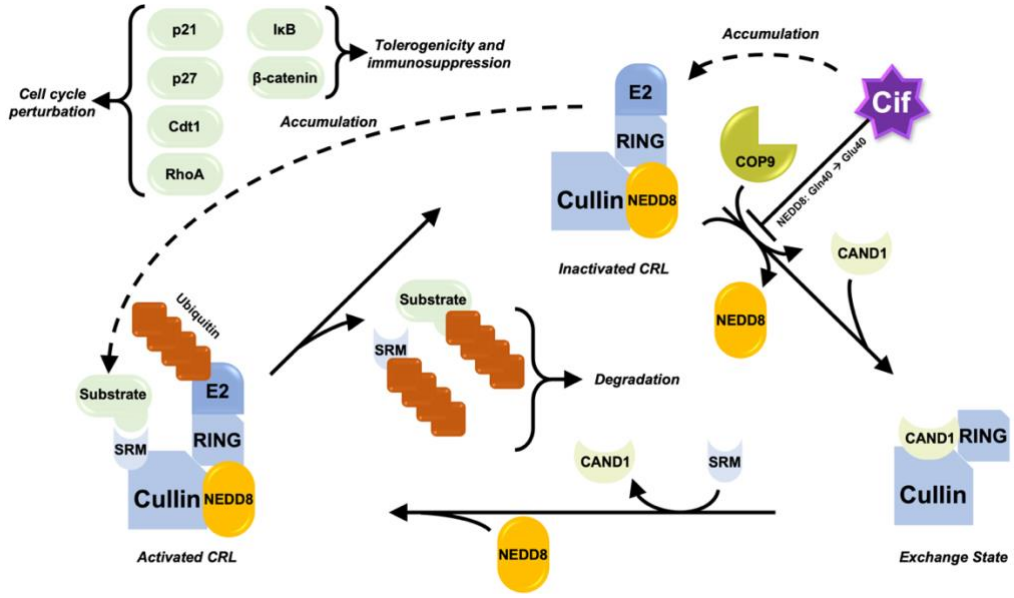


Figure 1.19 | Cif causes the accumulation of regulators.

Cif, encoded in both *P. asymbiotica* and *P. luminescens*, acts by modifying and inactivating cullin-RING E3 ubiquitin ligases (CRL). After trafficking to the nucleus, Cif deamidates Gln40 of the licensing factor NEDD8, preventing hydrolysis of this protein from the inactive CRL complex. Accumulation of inactive CRLs causes accumulation of the CRL's targets, particularly those involved in the regulation of cell cycling. The Cif orthologue in *B. pseudomallei* also deamidates Gln40 of ubiquitin, affecting all classes of E3 ubiquitin ligases and causing accumulation of a wider diversity of target proteins. Adapted from Taieb, Nougayrède, and Oswald 2011.

One such set of accumulating targets are those involved in cell cycling. Most notably, Cif indirectly and irreversibly inhibits cell cycling at the G_1/S and G_2/M checkpoints due to the accumulation of $p21^{waf1/cip1}$ and $p27^{kip1}$ (p21 and p27 hereafter) cyclin-dependent kinase (CDK) inhibitors (CKIs). p27 phosphorylates CDK2-Cyclin E and A which prevents S-phase progression; p21 does the same but also phosphorylates CDK1-Cyclin B to prevent mitotic progression (Figure 1.18) (Taieb, Nougayrède and Oswald, 2011).

Sequestering of active CRLs also accumulates β-catenin, Cdt1, RhoA, and IκB. Most notably, β-catenin induces tolerogenicity in dendritic cells, the survival of regulatory T cells, suppression of effector T cells, and regulation of

macrophage migration and adhesion (Suryawanshi *et al.*, 2020). These functions likely owe to an, overall, immunosuppressive phenotype since IκB accumulation inhibits NF-κB-mediated inflammation (Collier-Hyams *et al.*, 2005; Manicassamy *et al.*, 2010). Similarly to Pnf, stabilisation of activated RhoA induces stress fibres in HeLa cells, and accumulation of the licensing factor Cdt1 leads to multiple rounds of DNA replication, leading to increased ploidy (Tada, 2007).

Additionally, Cif from *B. pseudomallei* deamidates Gln40 of ubiquitin, extending the function of this homologue to disrupt degradation of all cell processes (Cui *et al.*, 2010). Meanwhile, Cif from *Pseudomonas aeruginosa* downregulates surface-level class I major histocompatibility complexes (MHC-I) in human airway epithelial cells (Bomberger *et al.*, 2014). Given that it can also promote ubiquitylation of the cystic fibrosis transmembrane conductance regulator (CFTR) channel, Cif likely reduces mucal clearance to favour the establishment of a respiratory polymicrobial infection (Bomberger *et al.*, 2011).

1.3.2.5 Pdp1 / SepC-like

Encoded just upstream of the *pnf* gene, Pdp1 (previously termed SepC or SepC-like) shares homology to the C-termini of a superfamily of rearrangement hotspot (Rhs) proteins and is a dNTP pyrophosphatase. Wang and colleagues (2020) demonstrated that Pdp1 disrupts the pool of nucleotides in the nucleus and cytosol: an apparently unique mechanism in the context of pathogenic bacterial effectors. Despite its dual-localisation, Pdp1 is cytotoxic in human macrophage and HEK293T cells when contained to the cytosol by mutation of its nuclear localisation signal (NLS). Yang and colleagues (2006) also demonstrated that Pdp1 causes dense actin cabling in the same vein as that of Pnf.

1.3.2.6 PVC Units 1-4

P. luminescens TT01 encodes a tandem of 4 PVC structural operons each with their own effectors. PlPVCu1 encodes a Pdp1-like effector, whilst PlPVCu2 and PlPVCu3 encode Rhs-like toxins. *P. asymbiotica* does not encode units 1-3 but does encode an ancestral homologue of PVCu4, albeit with distinct effectors. PaPVCu4 encodes a photopexin (*phxA*)-like effector as well as two *gogB*-like genes homologous to *Salmonella enterica* serovar Typhimurium and *Escherichia coli*, hinting at its usage in mammalian infection (Crennell *et al.*, 2000). PlPVCu4 encodes *phxA* but also encodes a halovibrin (*hvnA*) homologue at this locus (Ffrench-Constant *et al.*, 2000): a novel NAD^+-glycohydrolase (NADase) whose reaction catalyses the release of free reactive ADP-ribose which has been shown to act as a signalling molecule in the symbiosis of *Vibrio fischeri* with the Bobtail squid (Reich and Schoolnik, 1996; Stabb, Reich and Ruby, 2001). Authors speculate that the halovibrin in *Photorhabdus* could be an ancient mechanism of interaction with its *Heterorhabditis* symbiont (Hapeshi and Waterfield, 2017).

1.3.3 Protein Delivery Systems in Biomedicine

PVCs appear to be capable of loading payloads of wide-ranging sizes; the Cif protein is ~29.5 kDa (Crow *et al.*, 2009) and Pnf is predicted to be ~38.0 kDa. These are smaller payloads compared with other naturally occurring ones; the protease domain alone in PaTox is thought to be around 47 kDa in a larger structure of around 335 kDa (Bogdanovic *et al.*, 2019) and $RRSP_{Pa}$ is thought to be around 61 kDa in size. Nonetheless, in the T6SS the Hcp tube acts as a chaperone for a variety of specialised cargoes, but fusing the large protein β-lactamase to a natural Tse1 payload jams the injection system and abrogates the function of the T6SS (Howard *et al.*, 2021). It is unknown whether the PVC tube plays a similar role, and whether payloads are similarly capable of abrogating PVC function.

The trafficking of bioactive proteins across the eukaryotic cell membrane remains a core challenge for many drug-development companies with interests in biologics or protein therapeutics. Endosomal uptake, for example, presents the unique challenge of preventing protein degradation; hence, delivery directly into the cytosol is more desirable to get payloads to act at the necessary organelles (Y.-W. Lee *et al.*, 2019).

Protein therapeutics avoids some of the inherent issues with nucleic acid delivery by viral vectors, including immunogenicity and toxicity, as well as the heterogeneity in protein expression levels over time from nanoparticle-based delivery methods (Mout *et al.*, 2017; Marschall, 2021). Other methods of protein delivery have been developed in recent years: covalent modifications, such as cholesterol-tagging (Tai, Zhao and Gao, 2020), sometimes lose activity after purification. Cationic lipid-based methods, such as cell culture transfection reagents, as well as peptide-based carriers have been largely successful though conditions must be optimised for each protein payload, especially for the treatment of complex diseases such as cancer and diabetes (Y.-W. Lee *et al.*, 2019).

1.3.4 AAA+ ATPase Superfamily

Pvc15 is conserved in all of the PVC classes and operons and has been shown to be involved in payload loading (Jiang *et al.*, 2022). The ATPases associated with diverse cellular activities (AAA+) represent an ancient, universally encoded superfamily of molecular machines which contain an ATP-binding and hydrolysing domain to exert large macromolecular mechanical changes. As the name suggests, these proteins – often in the form of homohexamers or other homooligomers – are responsible for a myriad of cell processes despite being structurally similar. This diversity is owed to the N- and C-terminal components appended either side of the conserved base and lid supersecondary motifs (Snider, Thibault and Houry, 2008; Miller and Enemark, 2016).

Though often used interchangeably, "AAA" family members are a subset of the AAA+ superfamily which are now referred to as the 'classic' clade of ATPases; care is taken in this thesis to ensure this nomenclature is upheld: "AAA ATPase" is synonymous with 'classic' or clade 3 AAA+ ATPases.

1.3.4.1 Classification of AAA+ ATPases

As illustrated in Figure 1.20, AAA+ ATPases exist in 7 clades based on the structures of the AAA+ domain: a domain functionally akin to a 'battery' compartment which provides the energy for the other N- or C-terminal domains to function, in addition to the mechanical work that it provides. The large β-α-β 'sandwich' subdomain acts as the base for ATP binding whilst the lid domain typically encloses the binding site and mediates subunit interactions.

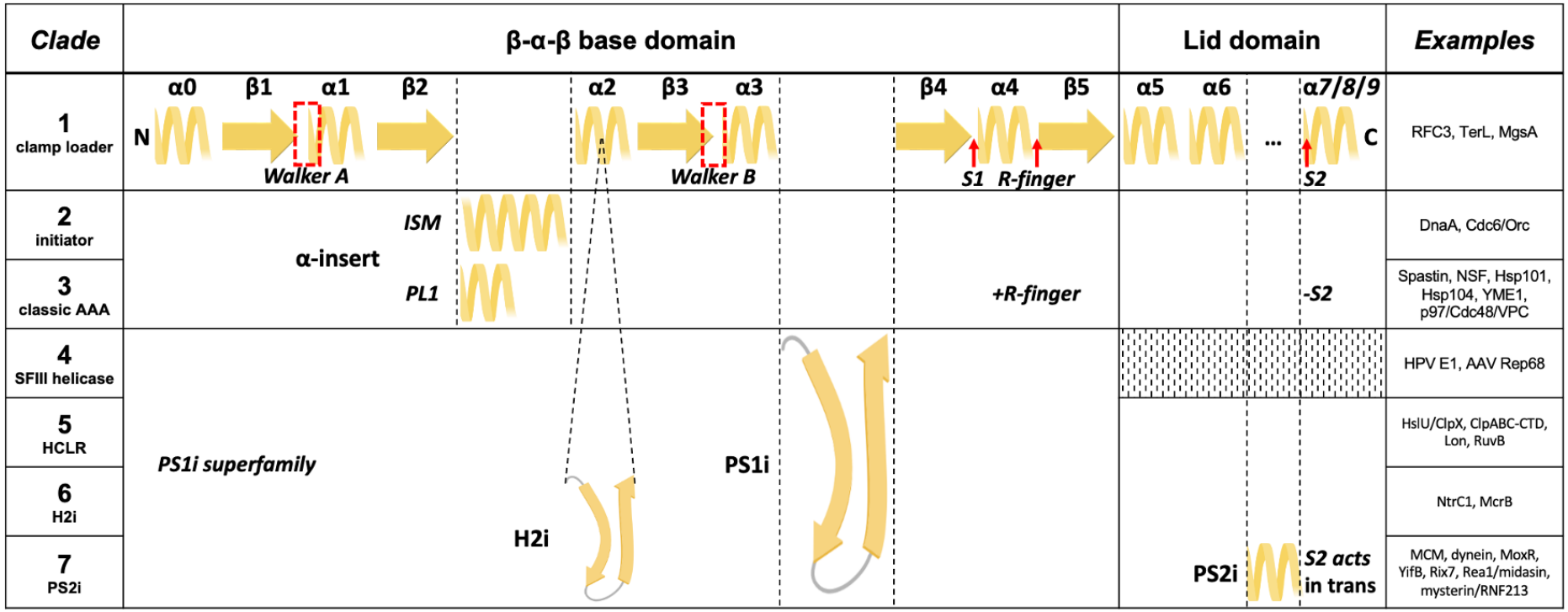
Clade 1 consists of the minimal AAA+ domain and includes members such as RFC and *E. coli* DNA polymerase III subunits γ and δ (Jeruzalmi, O'Donnell and Kuriyan, 2001; Miller and Enemark, 2016). Clades 4-7, meanwhile, are distinguished by the presence of a β-hairpin prior to the sensor 1 motif in the AAA+ domain (i.e. pre-sensor 1 β-hairpin insert; the PS1βh/PS1i superclade) which is used for signalling between subunits during concerted ATP binding and hydrolysis events in the hexamer; alternatively, PS1i can also be positioned for direct interaction with DNA in the case of the papillomavirus E1 helicase and other clade 4 proteins (Enemark and Joshua-Tor, 2008; Shin *et al.*, 2020).

A recent review of the structural and mechanistic differences between clades is covered in-depth by Khan, White and Brunger (2022), the key points of which will be summarised here.

Most AAA+ ATPases form hexamers when undergoing hydrolysis but clade 1, the clamp-loader clade, tend to either work as monomers or form pentamers in the case of the T4 TerL terminase (Hilbert *et al.*, 2015); the MgsA protein can also form a tetramer (Page *et al.*, 2011). Clade 2, the initiator clade, encodes an

additional long α-helix between β2 and α2 (termed the initiator-specific motif; ISM) and, along with some members of clade 5, can form filamentous oligomers around DNA. Examples are DnaA and DnaC in bacteria, and Orc1-5 in eukaryotes. Clade 3, the 'classic' AAA clade, is the most well-studied clade and is characterised by a shorter α-helix insertion (which also includes pore loop 1), an additional R-finger, and substitution of the S2 arginine with alanine. Clades 1, 2, and 4 are restricted to exclusively DNA replication functions, whereas clades 3 and 5 are far more functionally diverse and have can be involved in protein complex disassembly, protein unfolding and degradation, trafficking, disaggregation, and cytoskeletal remodelling. Clade 5 ATPases are also involved in chaperoning, maintaining the morphology of the endoplasmic reticulum (ER), and have DNA helicase and protease roles.

A


Clade
β-α-β base domain
Lid domain
Examples
α0 β1 α1 β2 α2 β3 α3 β4 α4 β5 α5 α6 ... α7/8/9
N
C
Walker A
Walker B
S1
R-finger
S2
1 clamp loader
RFC3, TerL, MgsA
2 initiator
DnaA, Cdc6/Orc
3 classic AAA
Spastin, NSF, Hsp101, Hsp104, YME1, p97/Cdc48/VPC
ISM
α-insert
PL1
+R-finger
-S2
4 SFIII helicase
HPV E1, AAV Rep68
5 HCLR
HslU/ClpX, ClpABC-CTD, Lon, RuvB
6 H2i
NtrC1, McrB
7 PS2i
MCM, dynein, MoxR, YifB, Rix7, Rea1/midasin, mysterin/RNF213
PS1i superfamily
H2i
PS1i
PS2i
S2 acts in trans

B

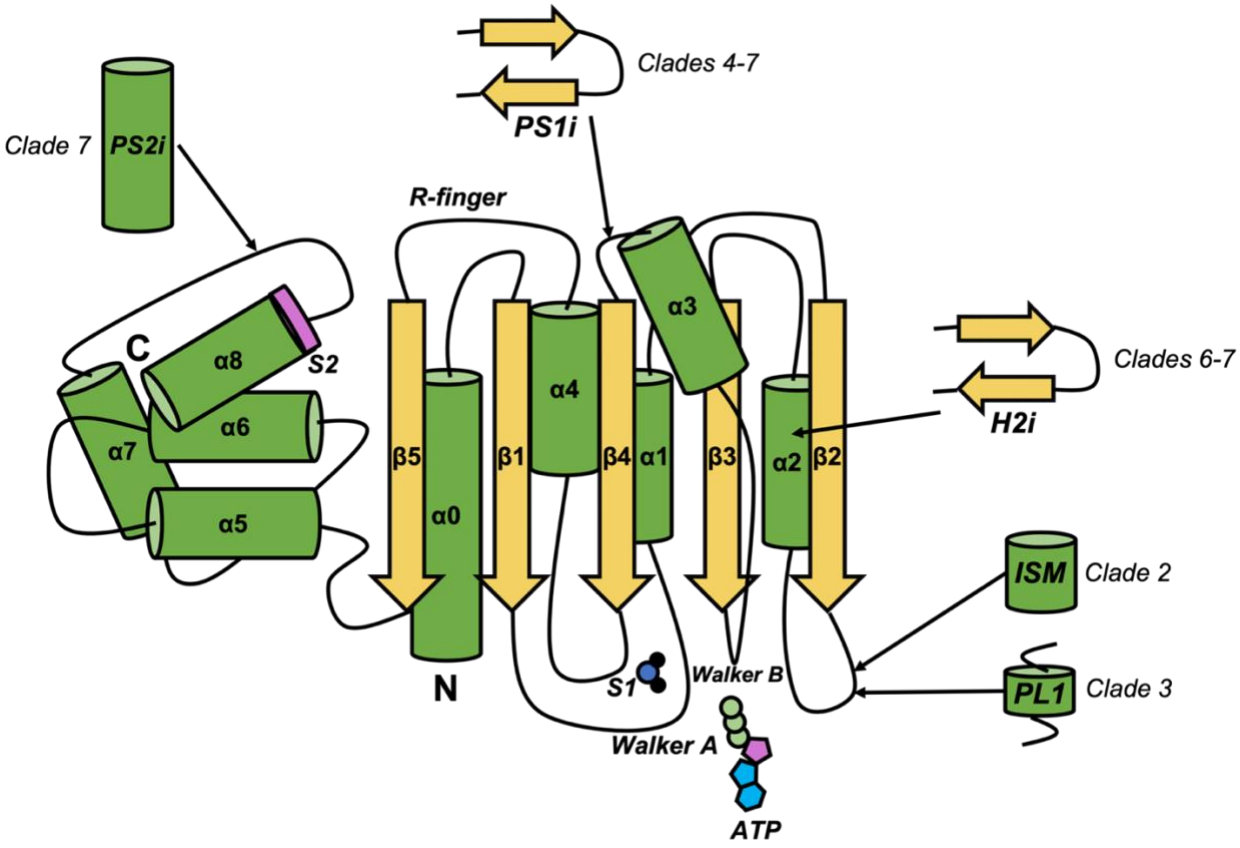

Clades 4-7
PS1i
Clade 7
PS2i
R-finger
α3
C
α8
S2
α4
Clades 6-7
H2i
α7
α6
β5
β1
β4
α1
β3
α2
β2
α5
α0
ISM
Clade 2
PL1
Clade 3
N
S1
Walker B
Walker A
ATP

Figure 1.20 | AAA+ ATPase clades 1-7 conserved supersecondary motifs.

(A) Clade 1 is the archetypal AAA+ ATPase domain consisting of two sub-domains: the β-α-β base domain, sometimes called the 'sandwich' or simply the 'large subdomain', and the lid, C-terminal bundle, or 'small' subdomain which encloses the nucleotide binding site; in clade 4, the latter is replaced by unique N- and C-terminal helix domains. The lid subdomain can consist of 3 (as shown), 4, or 5 helices which may or may not include the PS2i of clade 7. The large subdomain consists of alternating α-helices and β-sheets which form the base for which to bind (at the Walker A site) and catalyse (via the Walker B site) the ATP substrate whilst coordinating a Mg^{2+} ion and the water molecule required for hydrolysis, done by the sensor 1 (S1) residue. A sensor 2 residue found in the lid domain may provide binding energy for the nucleotide in the monomer (*in cis*), or in an adjacent monomer for clade 7 (*in trans*). Finally, the arginine finger (R-finger) acts *in trans* for oligomeric AAA ATPases: communicating the hydrolysis from one monomer to the adjacent one.

(A) and **(B)**: An additional insertion of an α-helix between β2 and α2 is found in clade 2 (called the initiator-specific motif; ISM) and in the classic AAA clade which incorporates pore loop 1 (PL1); Clade 3 also encodes an additional R-finger and substitutes the arginine of S2 with an alanine. Clades 4-7 make up the pre-sensor 1 insert (PS1i) superfamily, which is a β-hairpin loop found between α3 and β4. Both clades 6 and 7 encode an inserted β-hairpin within α2 (H2i); clade 7 encodes an additional α-helix in the lid domain pre-sensor 2 (PS2i).

Clades 4-7 all contain a pre-sensor 1 insert (PS1i) and are, thus, classified as the PS1i superfamily/superclade. Clades 4 and 5 do not contain any additional structural features though clade 4, the 'superfamily III (SFIII) helicase' clade, which is composed of only viral helicases, replaces the canonical lid domain with unique α-helix domains N- and C-terminal of the β-α-β core. Clade 5 is referred to as the HCLR clade, named after each of the primary families: HslU/ClpX, ClpABC-CTD, Lon, and RuvB. Both clades 6 and 7 contain a β-hairpin insert inside of α2 (helix 2 insert; H2i) and clade 7 contains an additional α-helix in the lid domain prior to the sensor 2 motif (pre-sensor 2 insert; PS2i). Clade 6 consists of restriction endonucleases and transcription factors such as NtrC1 and McrB: an example of a notable AAA+ GTPase (Nirwan *et al.*, 2019). Clade 7 is more functionally diverse and includes the mini-chromosome maintenance (MCM) helicase but also the dynein motor family and essential ribosome biogenesis proteins Rix7 and Rea1/midasin from *S. cerevisiae* (Sosnowski *et al.*, 2018; Lo *et al.*, 2019). The dynein family is especially interesting as these proteins form pseudo-hexameric AAA+ rings: they are encoded on a single polypeptide protomer and separated by a single microtubule-binding domain after the 4th AAA+ domain which act to move the complex in concert along microtubules for transporting macromolecular complexes in the cytosol (Liu, Rao and Gennerich, 2020).

Finally, some proteins are outliers to this classification system. Pch2/Trip13 family AAA+ ATPases, involved in spindle assembly, are not classified by any clade since they lack distinct PS1i and pre-α2 helix inserts, yet seem to have diverged from an ancestor of clade 3 and the PS1i superclade (Ye *et al.*, 2015).

1.3.4.2 ATP Hydrolysis Modes of Action

ATPase hydrolysis events can be conducted according to the sequential 'around-the-ring' model of hydrolysis such as with *Yersinia pestis* Lon protease (Shin *et al.*, 2020), or as two trimers of dimers alternating ADP/ATP-bound

states such as with the *Meiothermus taiwanensis* Lon protease (Lin *et al.*, 2016). Alternatively, Clp-family proteases appear to function independently of one another in a so-called stochastic model; these pseudohexameric protomers still function if one AAA+ domain is mutated (this would normally stall an ATPase relying on sequential firing) and amino acid step sizes can be larger than sequentially-firing AAA+ proteins (Fei *et al.*, 2020). Stochastic firing such as this can result in much faster translocation of substrate as residues in pore loops PL1 (found in the β2-α2 loop) and PL2 (found in the β3-α3 loop) can ensure unidirectional movement by Brownian ratchet theory. For example, ClpB pore loops act as a ratchet-like 'pawl' that asynchronously reconfigure throughout the process of ATP binding and hydrolysis to prevent stalling (Mazal *et al.*, 2021). Either way, polypeptides or DNA, such as in the RuvAB complex (Wald *et al.*, 2022), are translocated through the axial pore formed by the hexameric subunits.

The AAA+ domain binds ATP at the conserved Walker A and B motifs and are responsible for ATP binding and hydrolysis, respectively. The Walker A motif, also called the P-loop, consists of a conserved motif: (G/A)xxxxG**K**(T/S); the Walker B motif is less well conserved but is generally of the form (R/K)-xxx-G-xxx-hhhh**DE** where 'h' denotes a hydrophobic residue. The 'DEXD' motif is sometimes referred to as the B-box. The bold letters represent catalytic residues and, in the case of Pvc15, the 'DE' in the Walker B motif is found within the 'DEAD' B-box notably found in some nucleotide helicases (Linder *et al.*, 1989; Linder, 2006). Additional residues may also facilitate ATP hydrolysis: *cis*-acting sensor residues 1 (N, S, T, or D) and 2 (K or R) as well as a *trans*-acting arginine finger from a neighbouring protomer are responsible for stabilising the reaction of hydrolysis of the γ phosphate (Shin *et al.*, 2020).

1.3.4.3 Pvc15 as an AAA+ ATPase

Researchers identified that Pvc15 is essential for loading the native Pnf and Pdp1 payloads into the PVC (Jiang *et al.*, 2022). Point mutations within the Walker A and B motifs (though it is not specified which ones) failed to abolish the unfolding activity of Pvc15 upon fluorescent proteins which had been fused with an N-terminal Pnf LS. This may indicate roles of Pvc15 beyond the ATPase activity itself.

In addition, phylogenetic analysis of Pvc15 by Jiang and colleagues (2022) indicates the closest neighbouring homologue to be HflB from *Y. similis*: an FtsH-type magnesium-coordinating metalloprotease (Sprague and Neubauer, 2014). FtsH and its orthologues are widely encoded from bacteria to eukaryotes FtsH localises to the cytoplasmic side of the inner membrane and has functions in dislocating membrane proteins and degrading regulatory and misfolded proteins (Janska, Kwasniak and Szczepanowska, 2013).

Another example of an abundant eukaryotic AAA+ protein which is highly structurally homologous to Pvc15 is Cdc48 (also known as p97 or VPC) (HHPred probability = 99.8%). Cdc48 makes up 1% of total protein in the eukaryotic cytosol (Wolf and Stolz, 2012). The myriad of roles that Cdc48 plays in almost every compartment of the cell is owed to its capacity to bind up to 6 of over 40 protein cofactors at the N domain, and another 6 at the flexible C-terminal tail; in one paper, Cdc48 is referred to as the 'Swiss army knife' of the cell (Baek *et al.*, 2013). Each monomer of the hexameric structure consists of two tandem AAA+ domains, D1 and D2, with each containing their own Walker A and B motifs. Notably, neither is redundant since mutation of the catalytic glutamate to glutamine in either domain is lethal to the cell, implicating both domains as necessary for essential functions (Esaki and Ogura, 2010). The tight association with different cofactors also tethers Cdc48 to different locales; for example, association with Ubx2 and VIMP tethers Cdc48 to the ER for efficient

ER-associated degradation, whilst ubiquitinated transmembrane proteins can be extracted using Cdc48's unfolding activity (Olszewski *et al.*, 2019).

1.4 Aims and Summary

Recent advances in the study of eCISs in prokaryotes and eukaryotes have indicated a high conservation of AAA+ ATPases at these loci. However, the relationship between N-terminal leader sequences (LSs) and their associated ATPases to that of the mature eCIS has not been fully investigated. In particular, the diversity of functions of putative payloads has not been previously shown, nor the distributions and roles of LSs to their associated eCISs. These points will be addressed regarding the PVC in this thesis in numerous ways:

1. The diversity of eCIS payload functions, how LSs are distributed across species, and predictions of the Pvc15 protein structure will be demonstrated in Chapter 3, starting on page 110.
    i. Alignments of PVC payload LSs are used to classify them into homologue groups in Chapter 3.4.2, page 129.
    ii. Pvc15 structural predictions are used to identify domains and residues for hypothesis testing. An in-depth look at the Pvc15 monomer and hexamer are in Chapter 3.4.3, page 138.
2. In Chapter 4, beginning on page 168, different heterologous payload proteins are verified for loading and injection into cell lines.
3. In Chapter 5, evidence for the Pvc15/LS interaction will be presented. Abundance of payload in cell lysate is investigated using N- and C-terminal truncations of the LS. Then, Pvc15's putative roles are investigated: i) stabilising the native Pnf payload via its LS, ii) PVC loading of functional payload via its LS, and iii) catalysing ATP. Pvc15 can be deleted from the PVC*pnf* operon then reconstituted with Pvc15 mutants in a separate plasmid; this work is done starting on Chapter 5.4.2 on page 208. Altogether, these experiments verify the hypotheses made by the bioinformatic approaches.

2. Materials and Methods

All experimental and bioinformatics methods and materials can be found in detail in this chapter. Methods and their respective materials are organised in order of the experimental workflow where appropriate.

2.1 Strains

All strains used in this project were gifts of the Waterfield lab and supplied from preparations of competent cells used for previous or ongoing experiments. As shown in Table 2.1, DH5α or DH10β were used for general purpose transformations and cloning because of the *recA1* mutation G160E which disables this recombinase to prevent homologous recombination with the chromosome (Bryant, 1988), and the *endA1* mutation to prevent endonuclease activity upon the plasmid (Lin, 1992). For expression, an NEB BL21(DE3)-derived strain, NiCo21, was used due to its high expression of T7 RNA polymerase upon induction with isopropyl β-D-1 thiogalactopyranoside (IPTG) and, thus, potent expression of the desired gene constructs. The strain is also optimised for immobilised metal-ion affinity chromatography (IMAC) protocols due to 6 mutations in the surface histidines of GlmS to alanines, resulting in a loss of affinity to nickel (Ni)-containing resins (Robichon *et al.*, 2011).

Table 2.1 | *Escherichia coli* strains used.

Genotypes, purpose of use, and origin is displayed.

Strain	Genotype	Description	Reference
		Cloning Strains	
DH5α	F- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80*lacZ*ΔM15 Δ*(lacZYA-argF)U169*, *hsd*R17(rK-, mK+), λ-	General purpose cloning with high transformation efficiency, HB1100 derivatised strain from Bethesda Research Laboratories	(Glover, 1985)
		Expression Strains	
NEB 'NiCo21' BL21(DE3)	*can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	DE3 prophage for T7 RNA polymerase enables high expression of constructs under the control of the T7 promoter after addition of IPTG. Also optimised for IMAC purification. Derivatised BL21(DE3) with deleted proteases (OmpT and Lon) to decrease degradation of heterologously expressed proteins.	New England Biolabs

2.1.1 Preparation of Chemically Competent Cells

Competent *E. coli* strains were made using a $CaCl_2$ method with additional salts $MnCl_2$ and NaCl, as shown in Table 2.2. Cells were grown in 8 mL of lysogeny broth (LB; chapter 2.2.1.1) overnight at 180 rpm and 37 °C in a 50 mL Falcon tube to ensure enough aeration during shaking by keeping a ~5:1 air to liquid ratio. This starting culture was then inoculated 1:100 in LB – such as 2 cultures of 0.5 mL into 50 mL – and grown at 37 °C at 180 rpm until reaching an OD_{600} of 0.5 (roughly 2-3 hrs). Subsequent steps were conducted on ice, letting cultures sit first for 20 minutes and pre-cooling the centrifuge to 4 °C. Cells were isolated by centrifugation for 10 minutes at 4500 rpm, and pellets were resuspended in 20 mL of solution I (see Table 2.2), and incubated on ice for a further 20 minutes or more. The centrifugation step was repeated, and the pellet was resuspended in 4 mL of pre-cooled solution II. 1.5 mL Eppendorf centrifuge tubes were used to store the 100 µL aliquots in dry ice whilst being transferred to a -80 °C freezer.

Table 2.2 | Competent cell preparation solutions.
Ingredients for preparing chemically competent *E. coli* and their respective concentrations as used in chapter 2.1.1.

Media	Components	Concentrations
Solution I (pH 5.6-6.0)	Sodium Acetate	10 mM
	$MnCl_2$	50 mM
	NaCl	5 mM
Solution II (pH 5.6-6.0)	Sodium Acetate	10 mM
	Glycerol	5% v/v
	$CaCl_2$	85 mM
	$MnCl_2$	5 mM

2.1.2 Preparation of Glycerol Stocks

For storage of bacterial strains and constructs, 500 μL of overnight culture was added to 500 μL of 50% glycerol (25% glycerol final concentration) and kept on dry ice whilst being transferred to a -80 °C freezer. Cells were harvested from these stocks by slightly melting the top of the ice using the heat from one's finger and scraping an ice crystal onto an agar plate or into liquid broth.

2.2 Media and Stocks

2.2.1 Culture Media

2.2.1.1 Lysogeny Broth (Luria-Bertani: LB)

E. coli was cultured in lysogeny broth (LB) liquid and agar plates, usually at 180 rpm in a shaking incubator for liquid broth or a static incubator for plates.

2.2.1.2 SOC

Super Optimal media with Catabolite repression (SOC) is a high-glucose alternative to LB used primarily for recovery of bacteria, reducing stress after heat shock or electropermeabilisation.

Table 2.3 | Recipes of solutions used for bacterial growth.

* = Glucose is added after autoclaving the remaining ingredients.

Media	Components	Concentrations
Lysogeny Broth (LB)	Bacto-tryptone	10 g/L
	Yeast extract	5 g/L
	NaCl	10 g/L
SOC	Bacto-tryptone	20 g/L
	Yeast extract	5 g/L
	NaCl	0.5 g/L
	KCl	0.186 g/L
	$MgCl_2$	10 mM
	$MgSO_4$	10 mM

Glucose*	20 mM

2.2.1.3 Antibiotics and Supplements

The antibiotics ampicillin (Amp) and chloramphenicol (Chlor) were used to select for cells containing both the pBAD-PVC*pnf* and pVTRa vectors, respectively. Arabinose and IPTG were then used to induce expression of the PVC syringe and epitope-tagged payload respectively. Concentrations of components and supplements can be seen in Table 2.4.

Table 2.4 | Antibiotics and supplements.
A concentration gradient experiment was used to find the optimal concentration of IPTG to use for induction of each PVC-payload combination between 0.1 and 2 mM.

Supplement	Working Concentration	Purpose
Antibiotic Selection		
Ampicillin	100 μg/mL	Selection of pBAD-PVC*pnf* vector
Chloramphenicol	25 μg/mL	Selection of pVTRa vector
Kanamycin	50 μg/mL	Selection of pET28a vector for expression of individual PVC components
Induction		
Arabinose	0.2% w/v	Induction of PVC syringe structural operon from the pBAD-PVC*pnf* vector
IPTG	0.1-2 mM	Induction of leader sequence-linked-epitope-tagged payload from the pVTRa vector

2.3 Cell Culture Techniques

2.3.1 Biological Replicates for Cell Culture

Biological replicates contribute to the sample size, *n*; in cell culture, the line between biological and technical replicates is harder to define (Lazic, Clarke-Williams and Munafò, 2018). In this work, biological replicates were defined using the criteria outlined by Lazic, Clarke-Williams and Munafò (2018) for experimental units (EUs): independent samples that i) have been measured separately, ii) have been treated separately, and iii) have minimal to no interaction/interference with one another.

For bacterial culture, biological replicates of cell lysates were taken as different starter cultures standardised from the same glycerol stock. For eukaryotic cell culture, cells were grown in separate flasks after thawing from the same liquid nitrogen stock and standardised by assaying at the same passage number.

2.3.2 Mammalian Cell Cultures

2.3.2.1 Jurkat T Cells and Raji B Cells

Jurkat T cells and Raji B cells were supplied as a gift from the John James lab: cells that have been used in previous works (James and Vale, 2012; James, 2018). Both Jurkat and Raji T and B cell lines, respectively, were grown in Roswell Park Memorial Institute (RPMI)-1640 advanced media (GIBCO #21870) supplemented with 10% foetal bovine serum (FBS), 10 mM HEPES, 2 mM L-Glutamine, 100 units/mL penicillin, and 100 μg/mL streptomycin. Cells were incubated in 10 mL flasks at 5% CO_2 in a 37 °C incubator.

When obtaining cells from storage at -80 °C, cells stored in growth medium supplemented with 5% v/v DMSO in 1 mL aliquots of 2.5 x 10^6 cells were quickly thawed in a 37 °C water bath. Cells were then spun at 2,000 rpm for 2

minutes and the supernatant was discarded. Cells were resuspended in 10 mL fresh media and left to incubate for 2 days at 5% CO_2 in a 37 °C incubator. Cells were then split every 3 days to a concentration of 2 x 10^5 cells/mL, counted by de-clumping and mixing 1:1 with trypan blue to stain for dead cells then using a haemocytometer.

2.3.2.2 HEK293T

Gifted from the John James lab, HEK293T cells (ATCC, CRL-11268) were grown in 10 mL growth medium in a T75 flask: Dulbecco's modified Eagle Medium (DMEM) supplemented with 10% FBS, 2 mM L-Gln, 100 units/mL penicillin, and 100 μg/mL streptomycin. Upon reaching >80% confluency after growth at 37 °C and 5% CO_2 – usually after around 3 days – media was carefully removed leaving adherent cells intact. 10 mL PBS was carefully added to gently wash cells, then was removed. Cells were trypsinised using 4 mL trypsin-EDTA and left at 37 °C and 5% CO_2 for a few minutes. 6 mL of growth medium was then added and split as normal: cells were counted using a haemocytometer after being mixed 1:1 with trypan blue and the remaining culture was diluted to 0.1 x 10^6 cells per mL.

2.3.2.3 THP-1 Monocytes

THP-1 TIB-202 human monocytes, below passage number 15, were passaged at 0.2 x 10^6 cells per mL in RPMI-1640 + 6 mM Gln + 10% FBS and grown at 37 °C and 5% CO_2 and split every 3 days. Before experimentation, cells were differentiated into THP-1 macrophages (M0) with 10 ng/mL 12-phorbol myristate acetate (PMA) for 72 hours (Starr *et al.*, 2018).

2.3.3 Insect Cell Culture

S2 insect cells derived from late-stage *Drosophila melanogaster* embryos were stored in Schneider's *Drosophila* medium 2 (S2) + 10% FBS + 10% DMSO, were thawed at 28 °C and centrifuged at 1000 rpm for 5 minutes. Supernatant was discarded to remove any remaining DMSO. After resuspension in S2 medium, cells were grown in 5 mL within a T25 flask at 28 °C at a starting cell density of 2-4 x 10^6 cells/mL. Cells were passaged when cell density was between 6-20 x 10^6 cells/mL, ensuring cell density never went below 0.5 x 10^6 cells/mL.

2.3.4 Resazurin Assay: Measuring The Effects of PVCs on Cell Lines

The blue compound, resazurin, is metabolically reduced in the mitochondria to yield the highly fluorescent pink compound, resorufin.

Jurkat and Raji cell lines were diluted to 8 x 10^5 cells/mL and 2 x 10^5 cells/mL for S2 cells (determined by resazurin calibration curves). Cells were seeded at 100 μL in quadruplicate in a black, clear-bottomed 96-well plate. Blank wells were prepared in triplicate for each cell line, and an additional set of wells contained only media. The plate was incubated for 1 h at 5% CO_2 in a 37 °C incubator. 10 μg of prepared PVC samples were added and the plate was incubated for a further 4 h. Finally, 10 μL of resazurin (12.5 mg/mL stock) was added to each well in quick succession (~5 μM final concentration). The plate was incubated for a further 6 h and fluorescence was measured (excitation = 530-570 nm, emission = 580-620 nm).

A follow-up experiment was conducted which aimed to test the dose-dependency of PVC*pnf* on cellular respiration and its effect on the cytoskeleton. 0.0 (PBS), 1.0, 5.0, 10.0, and 20.0 μg of PVCpnf was added into 100 μL wells of S2 cells and resorufin fluorescence was measured 6 h after inoculation and blank-corrected. Cells were grown for a further 12 h and then centrifuged at 200 x g for 5 minutes and washed with 100 μL PBS two times. Samples were then

fixed with 100 μL 4% paraformaldehyde (PFA) in PBS for 20 minutes. Cells were pelleted and washed with PBS two more times with 5 minute washes; cells were permeabilised using 100 μL 0.1% Triton-X-100 in PBS for 15 minutes and cells were then washed a further two times. Next, a solution of 200 μL PBS with 0.5 μL 400X phalloidin-conjugate in DMSO (Alexa Fluor™ 568 Phalloidin; Thermo Scientific, cat. no.: A12380) was added and cells were incubated at room temperature for 1 h. Finally, cells were washed twice more with PBS, then resuspended in PBS and added to wells in a black, clear- and flat-bottomed 96-well plate (Greiner μClear; cat. no.: 655090) for fluorescence microscopy using a 60X objective in a Cytation5 multi-mode imager to observe actin filaments.

2.3.4.1 Resazurin Cell Growth Calibration Curve

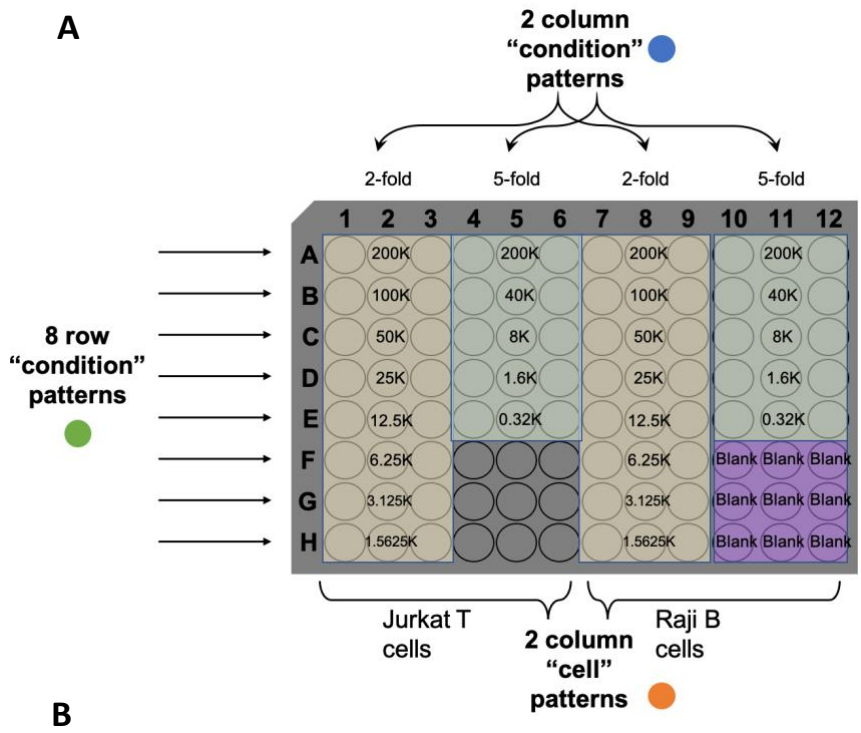
Jurkat and Raji cell concentrations were measured by counting via a haemocytometer, and an equal volume of culture for both cell lines was centrifuged at 1,000 rpm for 4 minutes to pellet the cells. The supernatant was discarded, and cells were resuspended to the appropriate volume to yield 2 x 10^6 cells/mL in culture medium.

In a black, clear-bottomed 96-well plate, 100 μL was added to each well apart from one row. 200 μL of cells was added to the empty row, and each was serially diluted either 2- or 5-fold in triplicate to concentrations of 15.625 x 10^3 and 3.200 x 10^3 cells/mL, respectively. 9 wells were left with only media as blanks. The plate was left to incubate at 5% CO_2 at 37 °C for 4-5 h. After incubation, a resazurin tablet was added to 20 mL of media (12.5 mg/mL), and 10 μL was added to each well in quick succession to a final concentration of ~5 μM; this concentration is thought to be low enough to not cause cytotoxicity in Jurkat T cells according to data by Erikstein and colleagues (2010). Fluorescence was measured every 30 minutes for 24 h at 5% CO_2 and 37 °C (excitation = 530-570 nm, emission = 580-620 nm). The optimal starting cell density was that which could just reach maximal cell density at around 24 h.

2.3.4.2 Analysis and Visualisation of Plate Reader Results

To efficiently proceed from the tabular output of the plate reader to analysis and visualisation of results, a script written in R version 4.0.2 'Taking Off Again' (R Core Team, 2020) utilised ggplot2 (Wickham, 2016, p. 2) and the reshape packages (Wickham, 2007) (see Appendices 10.1 for the full script used for this example). By inputting row and column 'patterns' that the samples follow and assigning these to each condition, the script can create a ggplot2 figure simply from the .csv file provided as output from the plate reader. The schematic Figure 2.1 shows how a 96-well plate can be input into the script, and the subsequent output whose colours and parameters can be changed using the ggplot2() function at the bottom of the script.

As shown in Figure 2.1a, in this example the 2-fold and 5-fold dilutions were laid out symmetrically on the plate between the two cell types, so were assigned as two input variables. To assign these variables, the column numbers – as assigned in the plate reader's output as Well Col – were assigned as separate vectors in the list variable, well.cols.list seen in Figure 2.1b. For instance, the 2-fold values were conducted in columns 1 to 3 and 7 to 9, and the 5-fold dilution was carried out in the other 6 wells. Then, for each starting cell number – the set of conditions – a value, 1 or 2, was assigned in the variable well.cols.groups to specify which vector within well.cols.list applied to that condition i.e 1:3 and 7:9 (1) or 4:6 and 10:12 (2). By assigning the same variables in the other dimensions by row and cell type, data are triangulated to their correct positions and datasets. The script then calculates the mean and standard deviation of each set of repeats. In this way, the user can effectively triangulate the position of each datapoint and specify which experiment the reading belongs to using 2 dimension variables: row and column. The experiment can be read into different figures for different cell types which are iterated over a time course.



B

```
cells <- c("Resazurin Calibration Curve - Jurkat T Cells",
           "Resazurin Calibration Curve - Raji B Cells")  # as displayed in figure title.
# One for each cell type / experiment. Creates different graphs

### Condition patterns
conditions.list <- list(c("200K", "200K ", "100K", "50K", "40K", "25K", "12.5K", "8K",
                          "6.25K", "3.125K", "1.6K", "1.563K", "0.32K"))  # List. What sets of conditions exist?
# as displayed in legend. Ensure all names within vectors are different, and each set is the same length.
# If same conditions in different cell types, enter full set of conditions above and use c(1,1,...) for each cell type below
conditions.groups <- c(1,1) # Vector. Which set of "conditions" corresponds to each cell of "cells"? Length=length(cells)

### Plate patterns
# If only one set of conditions; .corr vectors should be c(1,1,...). Length=length(cells)
(well.rows.list <- strsplit(toupper(letters[1:8]), split=""))  # List. What are the row patterns of the plate? Ensure to use toupper()
# Hint e.g: use letters[1:4] to select wells A:D
well.rows.groups <- list(c(1,1,2,3,2,4,5,3,6,7,4,8,5))  # List. Indices of groups in "well.rows.list", one for each condition.
# i.e which row pattern applies to each condition? Length of vector number i=length(conditions.list[[i]])
well.rows.corr <- c(1,1)  # Vector. Indices of "well.rows.groups" that correspond to each condition, one for each cell. Length=length(cells)

(well.cols.list <- list(c(1:3,7:9),c(4:6,10:12)))  # List. What are the col patterns of the plate?
well.cols.groups <- list(c(1,2,1,1,2,1,1,2,1,1,2,1,2))  # List. Indices of groups in "well.cols.list", one for each condition.
# i.e which col pattern applies to each condition? Length of vector number i=length(conditions.list[[i]])
well.cols.corr <- c(1,1)  # Vector. Indices of "well.cols.groups" that correspond to each condition, one for each cell. Length=length(cells)

(cells.range.list <- list(c(1:6), c(7:12)))  # List. What are the cell patterns of the plate? Ignore other conditions here
# Use nums for cols and letters for rows. Can use both here
cells.range.groups <- c(1,2)  # Vector. Indices of groups in "cells.range.list", one for each cell type.
# i.e which well pattern applies to each cell type / experiment? Length=length(cells)
# Hint: If only one cell type, assign these two variables as well.{rows/cols}.list and .groups
```

Figure 2.1 | Schematic of inputs to the Plate Reader Analysis R script.

By recognising condition-dependent and cell-dependent 'patterns' in the layout of the 96-well plate, variable inputs can be specified in the plate reader analysis R script. **(A)** Representation of how Jurkat T cell and Raji B cell calibration optimisation curves can be generated using the R script variables in **(B)**. Numbered wells represent starting cell concentrations and each filled coloured circle represents one of these three patterns identified in the 96-well plate. Each row of 3 wells can be averaged into a single datum for each time increment in the experiment. Each starting concentration of cells, specified as a 'condition', is assigned a group of rows and columns that describe its position in the plate.

2.3.5 Transient Transfection

HEK293T cells, gifted from the John James lab, were first grown to >80% confluency, trypsinised, and counted in a haemocytometer with 1:1 volume trypan blue. Cells were diluted to 0.5 x 10^6 cells per mL using DMEM, and 2 mL was added into each well of a sterile 6-well plate. Transfection was done using GeneJuice transfection reagent (Sigma-Aldrich cat. no. 70967), a non-liposomal, non-cytotoxic agent for maximal transfection efficiency. For transient transfection of an expression construct, 1 μg of DNA was added to 3 μL of GeneJuice and made up to 20 μL with sterile distilled water. 100 μL of serum-free DMEM was added and left for 20 minutes before being dripped into the appropriate wells. Growth medium was replaced the next day used for analysis or experiment 48 h after transfection.

2.3.5.1 Measuring Protein Degradation

HEK293T cells were transfected with the relevant protein construct in a 1:1 ratio with a constitutive GFP reporter plasmid [pSBtet-GN; Addgene 60501, (Kowarz, Löscher and Marschalek, 2015)] such that transfection could be verified by fluorescence microscopy prior to the experiment. 1.5 x 10^6 cells in 10 mL were added to a 100 mm culture dish (Starlab; cat. no.: CC7682-3394) and GeneJuice reagent volumes were scaled up accordingly; this ensures a homogenous distribution of transfected constructs across all cells in the population used for the subsequent experiment. After verifying transfection, 0.5 x 10^6 cells were added to individual wells of a 12-well plate (each sample in biological triplicate). Cycloheximide was added (50 μg/mL) in reverse order of time points (i.e., the overnight sample was added first) to ensure minimal disturbance of other samples from being taken in and out of the incubator.

At the end of the time course, growth medium from all samples was removed and cells were washed with PBS and lysed using the protocol found ahead in chapter 2.5.1.1 on page 103.

2.3.6 Generating Stable Cell Lines – Transduction

Transductions were conducted using a second generation packaging system: pCMV-dR8.91 (Addgene: #2221, updated: #8455; the packaging plasmid) and pMD2.G (Addgene: #12259; the vesicular stomatitis virus glycoprotein (VSV-G) envelope plasmid) (Table 2.5). When transducing cell lines, the protocol found in chapter 2.3.5 applies for transient transfection of lentivirus plasmids, except the following mass values were used: 600 ng of the transfer plasmid encoding the relevant HIV genes necessary for packaging, 600 ng of pCMV-dR8.91, and 300 ng of pMD2.G were added to the same tube. Then, 4.5 μL of GeneJuice was added and made up to 20 μL with serum-free media.

After transfecting cells with this mix, cells were left for 48 h before collecting the expressed lentivirus. The supernatant was collected, attempting not to disturb the cells adhered at the bottom of the wells, and centrifuged at 8,000 g for 2 minutes. The cells to be transfected were aliquoted into 1 mL for each transduction to be conducted and centrifuged at 1,300 g for 3 minutes. The supernatant was discarded and the pellet was resuspended in the residual. Up to 2 mL of viral supernatant was added to the suspension and made up to 3 mL using growth media and plated in a separate 6-well plate. After 24 h, 2 mL of growth media was added to each well and after another 24 h, transduced cultures were collected in T25 flasks and made up to 10 mL for subsequent passages.

2.4 Molecular Techniques – Nucleic Acids

The workflow of cloning the various aspects of this project can be seen in chapter 2.4.4.

2.4.1 Plasmids

Table 2.5 | Plasmid backbones and their derivations used.

Plasmid backbones were either already in use for similar projects in the lab or were gifted from other labs. IPTG = isopropyl β-D-1 galactopyranoside.

Designation	Components	Purpose	Reference
		Plasmid Backbones	
pBAD24	*araBAD*, *ampR*, *ori* and *f1*	Selected using ampicillin. Ori group = A; ColE1/pBR322	(Guzman *et al.*, 1995)
pVTRa	*lacO*, *catR*, *ori*	IPTG-inducible expression vector with a pTrc99 promoter. Selected using chloramphenicol. Restriction enzyme recognition sites for XbaI, SalI, PstI, and HindIII are contained in the multiple cloning site (MCS). Ori group = C; pSC101.	(Pérez-Martin and de Lorenzo, 1996)
pHR'	*ampR*, *ori*, 5' LTR, HIV-1 ψ, RRE, cPPT, WPRE, 3' LTR	Can be packaged into lentiviral virions for transduction when transfecting with a second generation lentiviral packaging system. Expression construct is inserted upstream of the WPRE and downstream of a strong SFFV promoter. Selected in bacteria using ampicillin. Ori group = A.	(Luigi Naldini *et al.*, 1996a)
pCMV-dR8.91	*ampR*, *ori*, CMV enhancer/promoter, HIV-1 *gag/pol*-cPPT, RRE, *tat*/*rev*-	Second generation lentiviral packaging plasmid. Expresses *gag* (matrix, capsid, nucleocapsid) and *pol* (reverse transcriptase, integrase) for effective virion production. Tat is required for transcriptional activation by binding the 5' LTR prior to packaging. Rev binds the Rev-response element (RRE) to facilitate viral RNA nuclear export.	(Luigi Naldini *et al.*, 1996b)

	NES		
pMD2.G	*ampR*, *ori*, VSV-G *env*	Envelope plasmid for pseudotyping lentiviral virions with the VSV glycoprotein to confer promiscuous tropism.	Addgene plasmid #12259
pD864-LacZ	*ampR*, *ori*, *lacZ*	Template plasmid for obtaining the full β-galactosidase (*lacZ*) sequence for cloning.	Addgene plasmid #109380 (Ceroni *et al.*, 2018)
pET28a	*kanR*, *ori*, *lacO*	IPTG-inducible T7 RNA polymerase promoter. Amp selection; ori group = A.	(Dubendorf and Studier, 1991)
		Derived Expression Vectors	
pBAD-PVC*pnf*	pBAD	Arabinose-inducible expression vector derived from pBAD24. Used for expression of the core/structural PVC proteins.	This work
pBAD-PVC*pnf*Δ*pvc15*	pBAD	Arabinose-inducible expression vector derived from pBAD24. Deletion of *pvc15* and non-coding region between *pvc14* and *pvc15* achieved using overlap extension PCR and insertion using SalI + PstI.	This work
pVTRa-pnf50-Pnf(C190A)-Myc	pVTRa	Standard expression vector used as a positive control for loading and injection of the PVC. C190A ablates deamidase activity. Myc-tagged non-functional Pnf preceded by the 50 amino acid Pnf LS with a short 'VD' linker (MLKYANPQTVAKAQRTKNTAKKPPSSTSFDGHLELSNGENQPYEGHKIRKIVD).	This work

pVTRa-cif50-Pnf(C190A)-Myc	pVTRa	Myc-tagged non-functional Pnf preceded by 50 N-terminal amino acids of the *cif* effector. C190A ablates deamidase activity. Encodes the full-length 50 amino acid Cif LS with a short 'VD' linker (MREYSKEDDCVKEKTNLAESENVEADNYLEMDCLNYLAKLNGMPERKDHSVD).	This work
pVTRa-pnf50-Pnf-Myc	pVTRa	Standard expression vector used as a positive control for loading and injection of the PVC. Myc-tagged Pnf preceded by the 50 amino acid Pnf LS with a short 'VD' linker (MLKYANPQTVAKAQRTKNTAKKPPSSTSFDGHLELSNGENQPYEGHKIRKIVD).	This work
pVTRa-pnf10-Pnf-Myc	pVTRa	Expression vector encoding Myc-tagged Pnf with its LS truncated to the first 10 amino acids and with a short 'VD' linker (MLKYANPQTVVD).	This work
pVTRa-cif50-Pnf-Myc	pVTRa	Myc-tagged Pnf preceded by 50 N-terminal amino acids of the *cif* effector. Encodes Myc-tagged Pnf with the 50 amino acid Cif LS with a short 'VD' linker (MREYSKEDDCVKEKTNLAESENVEADNYLEMDCLNYLAKLNGMPERKDHSVD).	This work
pVTRa-cif10-Pnf-Myc	pVTRa	Myc-tagged Pnf preceded by 50 N-terminal amino acids of the *cif* effector. Encodes Myc-tagged Pnf with the 50 amino acid Cif LS and a short 'VD' linker (MREYSKEDDCVKEKTNLAESENVEADNYLEMDCLNYLAKLNGMPERKDHSVD).	This work
pVTRa-pnfN25-Pnf-Myc	pVTRa	Expression vector encoding Myc-tagged Pnf with its LS truncated to the last 25 amino acids and with a short 'VD' linker (MTSFDGHLELSNGENQPYEGHKIRKIVD).	This work
pVTRa-pnfN40-Pnf-Myc	pVTRa	Expression vector encoding Myc-tagged Pnf with its LS truncated to the last 10 amino acids and with a short 'VD' linker (MPYEGHKIRKIVD).	This work
pVTRaDuet-6xHis-Pvc15-6xHis_pnf50-Pnf(C190A)-	pVTRa	Expression vector encoding N- and C-terminal hexahistidine tagged Pvc15 under the control of the same promoter as Myc-tagged non-functional Pnf preceded by the 50 amino acid Pnf LS	This work

Myc		with a short 'VD' linker (MLKYANPQTVAKAQRTKNTAKKPPSSTSFDGHLELSNGENQPYEGHKIRKIVD).	
pVTRaDuet-6xHis-Pvc15NTD_pnf50-Pnf(C190A)-Myc	pVTRa	Expression vector encoding N-terminal hexahistidine tagged Pvc15 domains N-D1 (residues 1-334; i.e., deletion just before D1 lid subdomain; ΔD2) under the control of the same promoter as Myc-tagged non-functional Pnf preceded by the 50 amino acid Pnf LS with a short 'VD' linker (MLKYANPQTVAKAQRTKNTAKKPPSSTSFDGHLELSNGENQPYEGHKIRKIVD).	This work
pVTRaDuet-Pvc15CTD-6xHis_pnf50-Pnf(C190A)-Myc	pVTRa	Expression vector encoding C-terminal hexahistidine tagged Pvc15 domain D2 (residues 434-698; i.e., ΔN-D1) under the control of the same promoter as Myc-tagged non-functional Pnf preceded by the 50 amino acid Pnf LS with a short 'VD' linker (MLKYANPQTVAKAQRTKNTAKKPPSSTSFDGHLELSNGENQPYEGHKIRKIVD).	This work
pVTRa-pnf50-GFP11x1-Myc	pVTRa	A single unit of GFP11 with a Myc tag at its C-terminus and preceded by 50 N-terminal amino acids of the *pnf* effector to act as a leader sequence for loading. To be co-expressed with pBAD-PVC*pnf* encoding the PVC chassis genes. Reconstitutes functional GFP with GFP(1-10).	This work (Ghosh, Hamilton and Regan, 2000; Cabantous, Terwilliger and Waldo, 2005)
pVTRa-pnf50-LacZα-Myc	pVTRa	The inactive α-complementation subunit of β-galactosidase (LacZα) encoded with a Myc tag at its C-terminus and preceded by 50 N-terminal amino acids of the *pnf* effector. To be co-expressed with pBAD-PVC*pnf* encoding the PVC chassis. Reconstitutes functional β-galactosidase with LacZω association.	This work (Bebenek and Kunkel, 1995)

pVTRa-pnf50-LacZ-Myc	pVTRa	β-galactosidase (LacZ) encoded with a Myc tag at its C-terminus and preceded by 50 N-terminal amino acids of the *pnf* effector.	This work (Kalnins *et al.*, 1983)
pVTRa-pnf50-GFPmut3-Myc	pVTRa	A faster-folding mutant of green fluorescent protein encoded with a Myc tag at its C-terminus and preceded by 50 N-terminal amino acids of the *pnf* effector.	This work (Cormack, Valdivia and Falkow, 1996)
pVTRa-cif50-GFPmut3-HA	pVTRa	A faster-folding mutant of green fluorescent protein encoded with an HA tag at its C-terminus and preceded by 50 N-terminal amino acids of the *cif* effector.	This work (Cormack, Valdivia and Falkow, 1996)
pVTRa-pnf50-Ptemp_PVCu1-Myc	pVTRa	A putative phosphatidyl phosphodiesterase derived from *Photorhabdus temperata* strain DSM 15199 (VY86_19225) encoded with a Myc tag at its C-terminus and preceded by 50 N-terminal amino acids of the *pnf* effector.	This work
pVTRa-pnf50::Sentomo_Afp17::Myc	pVTRa	A putative glycosyl transferase derived from *Serratia entomophila* strain A1MO2 (VY86_19225) encoded with a Myc tag at its C-terminus and preceded by 50 N-terminal amino acids of the *pnf* effector.	This work
pHR′- GFP(1-10)	pHR′	The first 10 β-sheets of GFP encoded in a mammalian expression vector. Reconstitutes functional GFP when bound to GFP11.	This work (Ghosh, Hamilton and Regan, 2000; Cabantous, Terwilliger and Waldo, 2005)
pHR′-mScarlet-GFP11	pHR′	Positive control for transient transfection of GFP(1-10) cell lines to be identified using flow	This work (Ghosh, Hamilton

		cytometry. GFP11 with N-terminal mScarlet fluorescent protein. Red fluorescence indicates the correct transfection of the plasmid whilst blue/green fluorescence indicates the correct expression and association of both GFP-truncated forms.	and Regan, 2000; Cabantous, Terwilliger and Waldo, 2005)
pHR′-LacZω	pHR′	An α-accepting complement of β-galactosidase with ΔM15 mutation (LacZω). Residues 12-42 are removed from the original *lacZ* gene. Encoded in a mammalian expression vector. Reconstitutes functional β-galactosidase when bound with LacZα.	This work (Prentki, 1992)
pHR′-LacZ	pHR′	Positive control for transient transfection of β-galactosidase encoded in a mammalian expression vector. LacZ activity is tested by adding fluorescent probes for LacZ activity, or by using X-gal.	This work (Kalnins *et al.*, 1983)
pHR′-LacZα	pHR′	The α-complement of β-galactosidase encoded in a mammalian expression vector. When transiently transfected with LacZω, this cell line acts as a control for the association of LacZα and LacZω to make functional LacZ.	This work (Prentki, 1992)

2.4.2 Polymerase Chain Reaction

Table 2.6 | PCR setup and reagents.

Quantities of reaction reagents and their concentrations before and after addition to the reaction mix are shown. For overlap extension PCR, any number of fragments can be added to the mix with primers that flank the outermost regions of the fragments used for amplification of the completed fragment, though usually no more than three fragments at a time were used in this work.

Reagent	Conc.	Volume µL	Final conc.
Routine PCR (Q5)			
Q5 Reaction Buffer	5X	10.0	1X
dNTPs	10 mM	1.0	200 µM
Q5 DNA polymerase	2.0 units/µL	0.5	0.02 units/µL
Forward primer	10 µM	2.5	0.5 µM
Reverse primer	10 µM	2.5	0.5 µM
Template DNA	variable	1.0	>0.2 ng/µL
Nuclease-free water	--	to 50	--
Colony PCR (GoTaq Green Master Mix)			
GoTaq Green Master Mix	2X	12.5	1X
Forward primer	10 µM	1.2	~0.5 µM
Reverse primer	10 µM	1.2	~0.5 µM
Template DNA	variable	half colony	variable
Nuclease-free water	--	to 25	--
Overlap Extension PCR (Q5)			
Q5 Reaction Buffer	5X	10.0	1X
dNTPs	10 mM	1.0	200 µM
Q5 DNA polymerase	2.0 units/µL	0.5	0.02 units/µL
Forward primer	10 µM	2.5	0.5 µM
Reverse primer	10 µM	2.5	0.5 µM
Fragment 1	variable	5.0	>0.2 ng/µL
Fragment 2+	variable	5.0	>0.2 ng/µL
Nuclease-free water	--	to 50	--

Table 2.7 | PCR conditions.

Polymerase Chain Reactions (PCRs) were carried out largely according to manufacturer's instructions. Denaturation, annealing, and extension steps are repeated for the corresponding number of cycles. The enzyme used for each reaction is shown in the header in parentheses.

Stage	Time (s)	Temperature (°C)	Cycles
Routine and Overlap Extension PCR (Q5)			
Initial denaturation	60	98	
Denaturation	30	98	
Annealing	15	>60	34
Extension (/kb)	30	72	
Final extension	120	72	
Hold	Infinite	4	
Colony PCR (GoTaq Green Master Mix)			
Initial denaturation	60	98	
Denaturation	30	98	
Annealing	15	>60	30
Extension (/kb)	60	72	
Final extension	120	72	
Hold	Infinite	4	

2.4.2.1 Primers

All primers used in this work were purchased from Integrated DNA Technologies (IDT).

Table 2.8 | DNA primers used.

'Sequence' base pairs in lowercase and 'Name' terms after an underscore ('_') indicate overhang sections which do not bind the template in the first cycle of PCR and, therefore, are not considered in the Tm. Restriction sites are **bolded**.

Name	Sequence (5′ → 3′)	Tm (°C)	Purpose
	Sequencing and Screening		
pVTRa Seq Fwd	GGGAAATGAGCTGTTGAC	54	Sanger sequencing pVTRa and colony PCR forward
pVTRa Seq Rev	GCAGTTCCCTACTCTCGCA	58	Sanger sequencing pVTRa and colony PCR reverse
HA-Tag ColPCR Rev	TGCGTAGTCTGGTACGTCG	58	Colony PCR reverse
	Cloning		
pIDTSmart Fwd	GGCTCGTCCTGAATGATATGC	58	PCR of ordered construct
pIDTSmart Rev	GATCGTGGACCGATCATACG	58	PCR of ordered construct
Linker Fwd	atc**GTCGAC**GGTTCATCTGGAA	56	PCR of construct containing linker (VDGSSG) for cutting with SalI
GFP Fwd_pnf	ccataagattaggaaaatc**GTCGAC**ATGAGTAAAGGAGAAGAACTTTTCAC	55	PCR of GFP with *pnf* overhang for OEPCR or cutting with SalI
GFP Rev_Tag	ATA**CTGCAG**TTTGTATAGTTCATCCATGCCATGTG	57	PCR of GFP with a PstI site-containing overhang
pVTRa Rev_NotI	gtcg**cggccg**cCTGCGTTCTGATTTAATCTGTATCAGG	57	PCR of pVTRa constructs for splicing into pHR vector

cif50 Fwd_BamHI-Kozak	ccg**ggatcc**gccgccaccATGAGAGAATACAGTAAAGAAGATGATTGC	57	PCR of pVTRa constructs for splicing into pHR vector
pnf50 Fwd_BamHI-Kozak	ccg**ggatcc**gccgccaccATGTTAAAATATGCTAATCCTCAGACCG	58	PCR of pVTRa constructs for splicing into pHR vector
SalI-Pnf Fwd_(P)	**GTCGAC**AAAGGATTGAGACAACATC	62	Forward primer for whole-plasmid PCR and blunt-end re-ligation. Used for truncating the leader sequences from the C-terminus
cif40 Rev	atgttgtctcaatcctttt**gtcgac**TAATTTAGCTAAATAATTAAGACAATCCATTTCCAGATAGTTATCTGCC	62	Dual-purpose PCR of the C-terminal-truncated *cif* leader sequence; can be followed by restriction ligation using SalI, or overlap extension PCR for other vectors
cif30 Rev	cgcaagatgttgtctcaatcctttt**gtcgac**TTCCAGATAGTTATCTGCCTCCAC	58	
cif20 Rev	gcaagatgttgtctcaatcctttt**gtcgac**AGATTCTGCTAAATTTGTTTTTTCTTTTACG	58	
cif10 Rev	cgcaagatgttgtctcaatcctttt**gtcgac**GCAATCATCTTCTTTACTGTATTCTCTCATTC	58	
pnf40 Rev	cgcaagatgttgtctcaatcctttt**gtcgac**GTTTTCACCATTTGAAAGTTCAAGGTGC	58	Dual-purpose PCR of the C-terminal-truncated *pnf* leader sequence; can be followed by restriction ligation using SalI, or overlap extension PCR for other vectors
pnf30 Rev	cgcaagatgttgtctcaatcctttt**gtcgac**ATCAAAAGAGGTTGATGATGGCG	58	
pnf20 Rev	cgcaagatgttgtctcaatcctttt**gtcgac**CGCAGTATTTTTAGTACGTTGTGC	58	
pnf10 Rev	cgcaagatgttgtctcaatcctttt**gtcgac**TACGGTCTGAGGATTAGCATATTTTAACAT	58	
BamHI-Kozak Fwd	CCG**GGATCC**ACCGGTCGCCACC	69	PCR of pHR mammalian expression vector
mScarlet OE Rev_LacZα	ggcttggcgtaatcatggTGCTGCCGCTGCCCTTGTACAGCTC	68	PCR of mScarlet to overlap with a LacZα overhang
LacZα OE Fwd_mScarlet	gtacaagggcagcggcagCACCATGATTACGCCAAGCC	58	PCR of LacZα to overlap with an mScarlet overhang
pBAD-PVC Fwd PstI	GTCTGCAGACGCACGATC GTTTGCGTGTTTAACATAAAAAACCTCTCTGTGGTCAATGGTTATAA	57	Overlap extension PVC for deletion of *pvc15* from the pBAD-

pBAD-PVC Rev Δ15	CCACG	57	PVC*pnf* operon by subsequent digestion with PstI and SalI
pBAD-PVC Fwd Δ15	CGTGGTTATAACCATTGACCACAGAGAGGTTTTTTATGTTAAACAC	59	
pBAD-PVC Rev SalI	GCAAAC GATGATGATGATGATGGTCGACG	58	
Pvc15 Fwd 6xHis_OrigRBS_XbaI	ctctagaTTGACCACAATGACTTAGTCTGAGTAAAAAATATGCACCAC CACCACCACCACAATATATCGCCTGTTTTTTATGATTCATTG	58	Initial cloning primers for making a pET28-compatible Pvc15 expression vector by digesting with XbaI + SalI (redundant due to addition into pVTRa; this strategy also enabled the addition of 6xHis tags on both sides of the ORF). Subsequent primers for cloning from pET28a-compatible vector into pVTRa by digesting with KpnI + BamHI.
Pvc15 Rev SalI	cttgtcgacaaatgttaatcgtccgactttagc	56	
pET Fwd KpnI	agaGGTACCTTGACCACAATGACTTAGTCTGAGT	57	
Pvc15 Rev 6xHis-BamHI-RBS	agcggatccTCCtcaatgatgatgatgatgatgaaatgttaatcgtccgactttagc	57	
Pvc15 NTD Rev_STOP-BamHI	agc**ggatcc**TCCtcataccggtatctttttcagccatacc	59	Cloning of mutant Pvc15 strains: NTD (Δ335-698), CTD (Δ1-433), and E555Q done by overlap extension PCR. Each uses flanking sequencing primers for cleavage using the KpnI + BamHI sites for digestion.
Pvc15 CTD Fwd_START-KpnI	atg**ggtacc**atgcaaaatttcggtcaattggcaca	56	
Pvc15 E555Q Fwd	ggtattattctttgatCaagctgacgcactg	57	
Pvc15 E555Q Rev	cagtgcgtcagcttGatcaaagaataatacc	58	

2.4.2.1.1 Primer Design

SnapGene was used to design primers for all cloning experiments. Primers were designed to optimally anneal 18-20 bp with a melting temperature of 58 °C, but always between 55-68 °C, ignoring non-annealing regions. Forward and reverse primers always had 50-65% GC content and were always within 5 °C melting temperature of one another. Each primer had a good mix of all four nucleotides. The 5 bp at the 5' and 3' ends of every primer were treated with the same conditions during design:

- Essential GC clamp
    - 20-80% GC
    - No more than two repeating bases
- Ideal GC clamp
    - 40-60% GC
    - 'GC' or 'CG' at the terminus

For primers with overhangs, only the annealing temperature of the region which would first bind the template was considered for the optimal annealing temperature for the reaction, but it was also ensured that subsequent cycles amplifying the full-length fragment had melting temperatures within 5 °C of each other. For primers with site-directed mutations, there was at least 12 annealing bases either side of the non-binding nucleotide(s) (though no more than 2 consecutive mutageneses at a time).

After designing the primers, NUPACK (http://www.nupack.org/) was used to assess the energetics of the formation of secondary structures at 55 °C. Inputs were usually the default with DNA selected, but with 2 mM Mg^{2+} and a maximum complex size of 2 for the two strand species. Primer secondary structure was acceptable as long as the ends of the primer were shown to be single-stranded and the predicted concentration of forward primer-reverse

primer conjugates was sufficiently lower than the predicted concentrations of each primer on its own (Zadeh *et al.*, 2011).

2.4.2.2 Q5 PCR

To ensure a minimal number of point mutations because of low fidelity polymerisation, Q5 High-fidelity DNA polymerase (New England Biolabs; M0491) was used for most cloning reactions. In a 50 μL reaction, 10 μL of 5X Q5 reaction buffer was added to 33 μL of distilled sterile water kept on ice. 1 μL of 10 mM dNTPs was added along with 2.5 μL of both 10 μM primers to yield a final concentration of 0.5 μM. Around 40 ng of template DNA was used, or 0.5 μL of the stock solution. Finally, 0.5 μL of the provided Q5 DNA polymerase was added to a final concentration of 0.02 units/μL; the remaining volume was made up using nuclease-free water.

Reactions were carried out in a Bio-Rad thermocycler and pre-heated to a lid temperature of 105 °C. PCR setup and reagents are shown in Table 2.6 whilst conditions can be seen in Table 2.7. Initial denaturation was carried out at 98 °C for 30 seconds followed by 32-34 cycles of subsequent reaction; Denaturation: 98 °C for 10 seconds; Annealing: 60-68 °C for 10-15 seconds; Extension: 72 °C for 10 seconds per kilobase. Annealing temperatures for Q5 polymerase are often higher than the traditional calculations for melting temperature (Tm) and optimal annealing temperature (Ta; lower Tm of primer pair + 1 °C, among other methods) given to *Taq*-based polymerases. Hence, annealing temperature was calculated using the provided NEB Tm calculator (http://tmcalculator.neb.com/). Particularly with small fragments, the annealing time was kept low to avoid excessive primer dimer formation and low specificity binding events. Where reaction yields were low or secondary structure was expected even in the presence of 2 mM Mg^{2+} in the Q5 reaction buffer, 5% DMSO was used to reduce secondary structure formation in the DNA fragments during amplification (Jensen, Fukushima and Davis, 2010).

2.4.2.3 *Taq* and Colony PCR

Colony PCR was used to confirm true positive colonies. One primer annealed to the backbone whilst the other annealed to the insert which would simultaneously indicate the presence of the insert in the correct backbone as well as assuring the correct orientation of the insert.

200 μL of HotStart (HS) MyTaq Red DNA Polymerase (Meridian Bioscience; Cat. No.: BIO-21115) was added to 2 mL of 5X MyTaq Red reaction buffer to be stored in 250 μL aliquots; 5.5 μL of this mix was added to 17.1 μL of nuclease-free water with 1.2 μL of each 10 μM primer to make-up a 25 μL reaction. To add the template DNA, half a colony was carefully scraped from the plate using a sterile filter pipette tip and inserted into a PCR tube where it was twisted vigorously.

The PCR reaction included an initial 4-minute denaturation step at 95 °C, followed by 32-34 cycles; Denaturation: 95 °C for 15 seconds; Annealing: 55-58 °C for 20 seconds; Extension: 72 °C for 20 seconds per kilobase. Agarose Gel Electrophoresis was used to confirm the presence of amplified fragments.

2.4.2.4 Overlap Extension PCR

For some cloning workflows, primers were designed to be homologous to the primer annealing region of the other fragment. These reverse complementary primers were used in an initial PCR reaction to amplify each fragment with a homologous region. Both fragments were added to a second reaction which utilised the forward primer of the first fragment and the reverse primer of the second, enabling amplification of only the fragments which had overlapped at the homologous region to create a larger fragment. This was done particularly when cloning small fragments such as epitope tags or leader sequences which are too small to reliably visualise and isolate by gel electrophoresis and extraction.

After the first PCR reaction, the 50 μL 'overlap' reaction consisted of 10 μL of 5X Q5 reaction buffer added to 24.5 μL of sterile water. 5 μL of each of the gel-purified PCR products were added with 2.5 μL of each of the outermost primers. Finally, 0.5 μL of Q5 High-fidelity polymerase was added. The PCR was conducted using the same conditions and cycles as the first reaction.

2.4.2.5 DNA clean-up

It was sometimes necessary to clean DNA samples after gel extraction or PCR to remove residual ethanol, buffers, or enzymes. DNA purification was conducted, as per the manufacturer's instructions, using the GE Healthcare 'illustra GFX' PCR DNA and Gel Purification Kit. DNA was finally eluted into 30 μL of type 6 elution buffer, warmed for 10 minutes in a dry bath at 60 °C prior to elution to ensure better unbinding of DNA from the silica beads of the spin column and, thus, a higher yield.

2.4.3 Agarose Gel Electrophoresis

Agarose gels were made at 0.8-1.2% w/v. Weighed powdered agarose was added to 40 mL Tris-acetate-EDTA (TAE) composed of 40 mM Tris base, 20 mM acetic acid, and 1 mM ethylene-diamine-tetra-acetic acid (EDTA). After heating, an additional 35-40 mL TAE was added at room temperature to hasten cooling. SybrSafe was added and the gel was cast and allowed to polymerise at room temperature for a minimum of 30 minutes. After adding the appropriate volume of loading dye to the samples, agarose gel electrophoresis was conducted at a constant voltage of 90V for 45-60 minutes in TAE buffer. Gels were visualised using an ultra-violet light box or G:BOX-F3 (Syngene).

2.4.3.1 Gel Extraction

Bands were extracted from agarose gels using a sharp scalpel, and gel extraction kit as per the manufacturer's instructions (QIAQuick Gel Extraction Kit or QIAEX II Gel Extraction Kit; Qiagen). Final elution was into 40 μL of elution buffer provided in the kit or distilled sterile water. If possible, fragments <300 bp in size were gel extracted from 1.5% agarose gels since yield was significantly decreased in 2.0% gels. DNA was kept at -20 °C for storage during cloning to prevent the action DNAses.

2.4.4 Cloning Techniques

Cloning workflows were designed via Snapgene (GSL Biotech LLC) including primer design and the visualisation of vectors and sequencing alignment results.

2.4.4.1 Restriction Digestion

Restriction digestion was carried out in a 50 μL reaction mixture. More than 1 μg of vector was added, 5 μL of CutSmart (New England Biolabs), and no more than one tenth of the reaction mixture contained the appropriate restriction enzyme, usually 1.3 μL each or 26 units. The final volume was made up with distilled sterile water. Digestion was conducted at 37 °C for 1-2 h. 6X loading dye was added to a total volume of 60 μL and DNA fragments were separated by agarose gel electrophoresis followed by gel purification according to the manufacturer's instructions (QIAquick Gel Extraction Kit, Qiagen).

2.4.4.2 Ligation

In a 20 μL reaction mixture, 2 μL of 10X T4 DNA ligase buffer and 1 μL of T4 DNA ligase were used (New England Biolabs). Backbone DNA and insert DNA were quantified using a nanophotometer. 80 ng of backbone DNA was added to

a 1.5 mL Eppendorf tube with an appropriate volume of insert DNA which was either incubated at room temperature for 2 h or overnight at 4 °C. The mass and volume of insert to use was derived using the following equations, assuming a molar ratio of 1:3 vector : insert is ideal for stochastic ligation:

$$mass_{insert}(ng) = 3 \times \frac{mass_{vector}\ (ng) \times length_{insert}\ (kb)}{length_{vector}\ (kb)}$$

$$V_{insert}(\mu L) = \frac{mass_{insert}(ng)}{concentration_{insert}\ (ng/\mu L)}$$

For each reaction, a control reaction was conducted which omitted addition of the insert. This helped to deduce the number of self-ligation mutant constructs (false positive colonies) and to compare against the transformation of correctly ligated vectors (construct$^+$ colonies) upon transformation into the chassis bacterium.

2.4.4.3 Transformation of *E. coli*

50 μL of chemically competent *E. coli* were thawed at 4 °C for 30 minutes. 1 μL of purified DNA or 5 μL of reaction mixture containing less than 50 ng of DNA was added and cells were left to sit for a further 10 minutes. The cell reaction mixture was incubated for 5 minutes at 37 °C and placed immediately back on ice for 2 minutes. Pre-heated LB was added (400 μL) and cells were grown at 37 °C for 1 h. The mixtures, including control reactions, were added to LB agar (LBA) plates containing the appropriate antibiotic at working concentration, left to aerate for 5 minutes, then incubated at 37 °C overnight (16-20 h). Given that only successfully transformed colonies should grow on these plates, the number of self-ligation mutant colonies on the control plate were used as a measure of the rate of success for each transformation. The positive likelihood ratio (LR(+)) was, thus, defined as:

$$LR(+) = \frac{number\ of\ construct\text{-}positive\ colonies}{number\ of\ false\ positive\ colonies}$$

Colony PCR was used to confirm successful insertion and correct orientation of fragments before being sequenced by Sanger di-deoxynucleotide sequencing.

2.4.4.4 Plasmid Extraction

A bacterial colony which was positive for the vector and insert as shown by colony PCR was inoculated into 12 mL liquid culture in a 50 mL tube, spun at 200 rpm for 20 h to an OD of ~0.8. Liquid cultures were spun at 4000 g for 6 minutes and supernatant was removed. Before using the plasmid extraction kit, cells were resuspended in the liquid residual by aggressive vortexing to de-clump cells and improve yield. After vortexing, plasmids were extracted as per the manufacturer's instructions (Qiagen; Plasmid Miniprep Kit). For highly concentrated samples, extracted DNA was diluted to 200 ng/μL, measured using a standard nanophotometer.

2.4.5 Sequencing

2.4.5.1 Di-deoxy-chain-termination (Sanger) Sequencing

Validation of cloned constructs was done by the sequencing services of GATC Biotech, Germany. Less than 400 ng of vector was added to an Eppendorf tube with 2.5 μL of 10 μM primer and made up to 10 μL with sterile distilled water. Primers used for amplification were designed for binding at least 200 bp from the site of interest as Sanger sequencing is error-prone especially at the ends of the amplicon.

2.4.5.2 Oxford Nanopore Whole-Plasmid Sequencing

To verify the whole sequence of the >20 kb pBAD-PVC*pnf* plasmid, the sample was sent to Plasmidsaurus (formerly SNPsaurus) in Oregon, USA. Oxford Nanopore sequencing was used to generate long reads to a high depth to attain a consensus sequence for the entire plasmid including the region of interest.

2.5 Molecular Techniques – Proteins

2.5.1 SDS-PAGE

Samples were added to LDS sample buffer (1X final concentration; ThermoFisher Scientific, NP0007) with 50 mM dithiothreitol (DTT). Sodium-Dodecyl-Sulphate (SDS) Polyacrylamide Gel Electrophoresis was often used to estimate the purity of protein samples. Precast gels (Bio-Rad TGX Mini-Protean 4-15% gels with Dual Colour Standard, cat. no.: 4561084DC) were used as per the manufacturer's instructions. SDS-PAGE was conducted in a Mini-Protean gel tank using PAGE buffer (25 mM Tris base, 190 mM glycine, 0.1% SDS w/v) and 150V for 45 minutes.

2.5.1.1 Eukaryote Lysate Preparation

When lysing eukaryotic cells to test for injection of PVC payloads, care was taken to ensure the nucleus was unbroken since the genomic DNA can cause the samples to become overly viscous. For adherent cells, media was aspirated and discarded and an equal volume of PBS was added; cells were gently resuspended with mixing. Suspended cells were spun at 2,500 g for 5 minutes and washed carefully with the same volume of PBS; this was repeated once, then cells were spun again and all supernatant was removed. For every 1 x 10^6 cells, 200 µL of eukaryotic lysis buffer (e-lysis buffer) was added using the recipe below. Cells were transferred to a microcentrifuge tube, where necessary, and shaken gently at 4 °C for 30 minutes. Finally, nuclei were spun at maximum speed in a microcentrifuge (>12,000 rpm) for 15 minutes to isolate them as a pellet. Supernatant was transferred to a separate tube and the nuclei were either discarded or assayed by adding 100 µL RIPA (radio-immunoprecipitation) lysis buffer. 90 µL of sample lysate was added to 45 µL

of 3x LDS sample buffer with 50 mM DTT for running SDS-PAGE as described above (chapter 2.5.1); the remaining volume was stored at -20 °C.

- e-Lysis buffer:
    - 1x TBS (made up with ddH_2O)
    - 1 tablet complete EDTA-free protease inhibitor per 50 mL
    - 1% Triton X-100

2.5.1.2 Staining

Protein gels were stained by adding to a Coomassie brilliant blue R-250 solution (Abcam InstantBlue® Coomassie Protein Stain; cat. no.: ab119211) with agitation (120-150 rpm) overnight or by agitating for 1 h. Stained gels were visualised in a G:BOX-F3 (Syngene) using an infra-red emission wavelength. Alternatively, gels could be visualised after transfer to a polyvinylidene fluoride (PVDF) membrane. Though Pryor, Xu and Hamilton (1992) detailed a method to completely destain PVDF membranes after staining with Coomassie brilliant blue R-250 by adding 50% methanol and 25% acetic acid for 30 minutes, washing with acetic acid in this way tended to reduce the assay sensitivity for subsequent immunodetection. Instead, immunodetection was conducted on the PVDF-blotted membrane first, visualised, then stained with Coomassie brilliant blue R-250 after 3 washes with ddH_2O, similarly to that described by Welinder and Ekblad (2011). The membrane was stained for 5 minutes then destained using a solution of 50% methanol and 7% acetic acid for 5 minutes then rinsed with ddH_2O again a further 3 times (Goldman, Harper and Speicher, 2016). This method was preferred over Ponceau S staining since Coomassie blue can detect protein at a minimum of 50 ng whereas Ponceau S is less sensitive, only detecting proteins at levels of 200 ng or more, and never seemed to give adequate results with a 0.1% Ponceau S solution in 5% sodium acetate even after pre-wetting with 20% methanol.

2.5.1.3 Western blot (WB)

After gel running, the extracted gel was loaded onto a PVDF membrane using a Biorad Transblot Turbo electroblotter in a 7 minute protocol. The Thermo Pierce Fast western blot kit was used according to the manufacturer's instructions. Rabbit anti-FLAG (DYKDDDDK, Cell Signaling, 14793), mouse anti-Myc (Cell Signaling, 2276), mouse anti-HA (Fisher Scientific, 11553060), or rabbit anti-β-RNAP (Abcam, ab191598) monoclonal antibodies were used as the primary antibodies and visualised using horseradish peroxidase (HRP)-bound secondary antibody and adding luminol and peroxide buffer. Bands were visualised using the chemiluminescence protocol in a G:BOX -F3 (Syngene) and using 2x2 binning for more than 300 seconds of exposure.

2.5.1.3.1 Quantification

Samples were normalised based on their OD_{600} for cell lysates or protein concentration for purifications, as determined using the BSA protocol on a spectronanophotometer. For a given band, the mean background value of the blot was subtracted and a corresponding loading control was used for that replicate either in the form of a fixed point band which appeared in all samples or the mean of all points in the corresponding lane in the Coomassie blue stain calculated using FIJI (version 2.14.0; ImageJ version 1.54f; Java version 1.8.0_202). After dividing the mean grey values of each band by its corresponding loading control mean grey value, test bands were relativised to their control samples; at least 3 biological replicates were included for each sample and statistical analysis was done using a Student's T-test.

Both normalisation methods (fixed-point and mean/sum of the replicate) are appropriate to use for medium-intensity bands with a high signal:noise ratio, as discussed by Degasperi and colleagues (2014). Normalisation by sum of the replicate tends not to adjust the sensitivity and specificity of normalised data

compared to the raw data since it reduces variance in the normalised data. In contrast, normalisation by fixed point reduces the proportion of false positive results but increases the proportion of false negatives.

2.6 Microscopy and Computation

2.6.1 Statistical Analysis

In the instance of multiple statistical inference comparisons, p-value adjustments were made only if the multiple comparisons were *arbitrary*: data with a 'global hypothesis' which may be tested using analysis of variance (ANOVA), for example. Statistical inference in this work was always planned at the experimental design stage: keeping arbitrary comparisons to a minimum. In addition, comparisons conducted between a given control and another sample were done using T tests assuming equal variance given similar sample sizes; these are presented as 't(df) = t; p-value = p'.

Overall, these measures ensured that p-value adjustments were rarely necessary to control family-wise error rates or false discovery rates. When required, adjustments were often done conservatively at low sample sizes using the Bonferroni correction method by dividing the critical p-value threshold, α, by the number of arbitrary levels (Dunn, 1961; Rothman, 1990).

2.6.2 Sample Preparation for Fluorescence Microscopy

Fluorescent cells (bacteria or eukaryotic cell lines) were washed twice with PBS by using a centrifuge at 4000 g to isolate cells then resuspending the pellet. Cells were added to a glass slide; a cover slide was then added. A Leica inverted fluorescence microscope was used with fluorescence excitation and emission wavelength filter values as follows: DAPI = 360/40 and 460/50 nm, and FITC = 475/35 and 530/40 nm. Images were captured using a Hamamatsu C11440-22CU. Objective lenses used were 10x (NA = 0.32), 20x (NA = 0.40), or 100x (NA

= 1.25). For the green (FITC) channel, laser intensity was set to 80 with an exposure time of 200 ms. For the blue (DAPI) channel, laser intensity was set to 50 with an exposure time of 150 ms. Phase contrast was done using a PH1 light ring, 38 ms exposure, and a light intensity of 90.

2.6.2.1 Image Analysis

TIFF images were analysed using Fiji (Schindelin *et al.*, 2012). First, phase contrast images were converted to an 8-bit image and a binary mask was applied using the 'auto-threshold' command to obtain objective cell boundaries. The fluorescence images were converted to 32-bit and background was subtracted using a rolling ball radius of 0.5. The binary mask was then multiplied by the processed fluorescence image and the new image was analysed. Mean gray values and standard deviations were specified to be analysed using the 'Analyze Particles…' command and exported into .csv formatted to be statistically analysed using R. Only particles with a circularity of greater than 0.4 and an area of greater than 0.05 were considered when analysing particles for most cell types being image with 10 or 40x objectives.

2.6.3 Transmission Electron Microscopy (TEM)

Purified PVC elutions were imaged by transmission electron microscopy (TEM) using a Jeol 2100 LaB6 TEM and a OneView 16 Mpi camera (Gatan), located in the University of Warwick. 300-mesh TEM carbon support grids were glow discharged for 1 minute. 5 μL of sample was added and blotted off after 1 minute. 5 μL of 2% uranyl acetate was added for 1 minute and blotted off at the side of the grid then lightly from the top of the grid. The grid was air dried for at least 5 minutes before mounting into the TEM dock and imaging at 160 kV.

2.6.4 Molecular Visualisation

In PyMol version 2.3.3 (Schrödinger LLC), the following Python code was used for exporting images as a portable network graphics (.png) file:

```
bg_color black
set antialias
set ray_trace_mode, 1
set gamma, 1.5; set direct, 0.5
set two_sided_lighting, 1; set ray_shadow, 0

set cartoon_fancy_helices, 1
set cartoon_smooth_loops, 1
set cartoon_highlight_color, grey7
set stick_radius, 0.22

set label_size, -0.3

ray 2400,2400
png {file_name}.png, dpi=300
```

These settings were saved to a .py file and run during the PyMol 2.4.2 session using '@script.py'. Some other applications used UCSF ChimeraX 1.5 (Pettersen *et al.*, 2004, 2021):

```
save {file_name}.png transparentBackground true pixelSize 0.02
```

To colour an object by its distance from the centre of mass or geometry (COM and COG), the center_of_mass.py script obtained the coordinates (https://pymolwiki.org/index.php/Center_of_mass). After download:

```
run center_of_mass.py
as surface
com all, object=COG
ramp_new proximityRamp, COG, range=[5,65], color=rainbow
set surface_color, proximityRamp
recolor; rebuild
```

'Range' values (bold) could be changed according to the desired distance to colour between the COG and each residue in Å.

2.6.5 Generating Figures and Schematics

Schematics were designed using Microsoft PowerPoint or BioRender.

2.6.6 Sequence Alignments

Multiple sequence alignments (MSA) were done using Clustal Omega 1.2.4 (Sievers *et al.,* 2011) and visualised using Jalview 2.11.2.7 (Waterhouse *et al.,* 2009).

3. Investigating PVC Effectors and ATPases

3.1 Introduction

3.1.1 The Bacterial eCIS Effectome

Whilst eCIS operons often encode a nearby effector, many of these putative payloads have not been functionally defined or investigated for the abundance of LSs to facilitate loading. A database of eCIS elements known as 'dbeCIS' (http://www.mgc.ac.cn/dbeCIS/) identifies operons which are predicted to encode structural components of extracellular contractile protein injection systems (eCISs) across bacterial and archaeal genomes (Chen *et al.*, 2019). Based on the phylogeny of the conserved *afp11* orthologues, eCIS subtypes are grouped into two lineages with six total subgroups (Ia-b and IIa-d); lineage Ia includes the PVC, Afp, and AfpX, whereas Ib contains the MAC from *S. entomophila* and the T6SS[iv] from *A. asiaticus*.

After identifying a putative effector payload, functions can be defined using homology searches including those that employ secondary structure conservation using Hidden Markov Models (HMMs) which are useful for finding homologous sequences when sequence identity is less than 20% (Eddy, 2004).

3.1.2 Predicting Structures of PVC Effectors and the Pvc15 ATPase

Whilst some PVC payload homologues have already been resolved at the atomic level, the structures of *Photorhabdus*-specific homologues have not been confirmed to contain structural conservation. Moreover, the structure of the Pvc15 AAA+ ATPase has not yet been described.

AlphaFold2 uses multiple sequence alignment (MSA) to identify related proteins which have been experimentally resolved for end-to-end training. Unlike other methods, it uses machine learning between the observed sequence and MSA hits at *each level* of the neural network. AlphaFold2 uses its own measure of prediction accuracy which correlates with the Cα-root mean square deviation (RMSD): the predicted local distance-difference test (pLDDT). In addition, the predicted alignment error (PAE) is useful for assessing the confidence of inter-chain interfaces for homo-oligomers; in contrast to pLDDT, a lower PAE score means higher confidence.

3.1.3 PVC Baseplate Carbohydrate Binding Modules

The AlgoCIS, MAC, and T6SS^{iv} all have two notable features: i) they do not possess canonical tail fibres, and ii) they possess a unique extension in the *afp11* homologue, forming the 'spike cage' (Xu *et al.*, 2022). Alg11 possesses putative carbohydrate binding modules (CBMs) with high sequence similarity to that of the *Thermotoga maritima* TmCBM27 (PDB: 1OF4) mannan-binding module, and chitosan-binding module of *Paenibacillus* spp. (PDB: 4ZXE). Xu and colleagues (2022) hypothesise that these CBMs may enable attachment to target membranes in the absence of tail fibres. Since the class II PVC also lacks tail fibres, this opens up the possibility for Pvc11 of the PVC*lopT* chassis (Pvc11$_{lopT}$) to facilitate attachment via similarly encoded CBMs.

3.2 Aims

i) Explore the functional diversity of putative eCIS payloads/effectors across lineages Ia and Ib.
ii) Define sequence and structural conservation of some of the PVC payloads as well as the Pvc15 AAA+ ATPase.
iii) Identify core domains and residues of Pvc15 which may elucidate its function and inform interesting mutations for *in vitro* experiments.
iv) Build a hypothesis of how Pvc11 from PVC*lopT* may substitute the role of the tail fibre using deep learning to predict protein-glycan interactions.

3.3 Methods

3.3.1 Glycowork/LectinOracle for *in silico* Protein-Glycan Predictions

Assessment of a range of protein-glycan interactions was assessed using LectinOracle via a Jupyter Notebook and using a Google VM via Colaboratory: LectinOracle_Microarray.ipynb (https://colab.research.google.com/drive/1Tk67xxra3W1fc4kJDQ5TLlpWUQaBEgGw). Though this notebook works as described in glycowork, its effectiveness depends heavily on the suitability of the protein being predicted according to the training dataset (Thomès, Burkholz and Bojar, 2021).

3.3.2 Using the dbeCIS for Analysing Putative Payload Functions

It was hypothesised that putative eCIS payloads would be encoded up to 4-5 kb downstream of the final 'core' gene in the operon: usually the 'cap'-equivalent gene. By appending the sequence information of these putative payloads to a custom database, each open reading frame (ORF) could have its function predicted by sequence using BLASTp and domain architecture using HHPred to predict its most likely function.

3.3.3 Using AlphaFold Packages to Predict Protein Structure

For hetero-oligomer prediction, the original AlphaFold Multimer script (github.com/deepmind/alphafold/) has been shown to have an improved accuracy over others' inspired methods (Evans *et al.*, 2021). However, a script which is run using Google Colab – termed ColabFold – is advantageous: i) it enables open-access use of graphics and tensor processing units (GPUs and TPUs, respectively) dependent on usage limits on the Google Colaboratory server (running AlphaFold2 locally requires two terabytes of storage to download each of the necessary protein databases), and ii) open-source code enables derivatives of the authors' script to give more advanced options to users. For the work done in this thesis, a script using mmseqs2 instead of jackhmmer for MSA significantly reduces the runtime (colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) (Mirdita *et al.*, 2021).

ColabFold is optimised for obtaining diverse MSA hits quickly; accurate molecular models can be built from only ~30 aligned sequences and, once a sufficient number of sequences are obtained, ColabFold has stricter stop-point criteria which prevents finding hits with sequences that will not significantly improve the model. The script also outputs all descriptive statistics – such as pLDDT and PAE data and figures – as well as multiple ranked models using a single ensemble seed to compare and find the best predicted model over multiple runs. Proteins and complexes totalling more than ~1400 amino acids could not be run with ColabFold due to GPU limitations. In these cases, LocalColabFold was used instead (github.com/YoshitakaMo/localcolabfold); this script has some advantages over ColabFold: there are no limitations on how long a run can be before timing out and no limitations on GPU usage (if one is available locally), though runs can take 5-10 times longer than ColabFold if structure prediction must be done on a CPU. Finally, for structures larger

than ~1000 amino acids, a local high performance computer cluster based at the Scientific Computing Research Technology Platform at the University of Warwick could be used for running AlphaFold locally using a Quatro RTX 6000 GPU and 16 CPUs per task (though, even this method usually fails to allocate memory effectively for structures greater than ~1800 amino acids).

3.3.3.1 pLDDT Calculation from MoLPC mpDockQ Output

For the hexameric Pvc15, the mpDockQ score which assesses completeness and accuracy of the model was an acceptable 0.233. This score, as well as the constant values assigned by MoLPC, were used to calculate the pLDDT of residues at an interface with another residue ($pLDDT_{IF}$). These constant values are provided by the authors on GitHub (Bryant *et al.*, 2022):

```
L = 0.783
x0 = 289.79
k = 0.061
b = 0.23
```

$$x = x_0 - \frac{\ln\left(\frac{L}{mpDockQ - b} - 1\right)}{k}$$

$$pLDDT_{IF} = \frac{x}{\log(n_{contacts})}$$

```
For the predicted Pvc15 hexamer:
mpDockQ = 0.233029686521374
number of contacts (n_contacts) = 184+240+379+242+438+243 = 1,726
```

$$x = 289.79 - \frac{\ln(258.44 - 1)}{0.061} = 198.793$$

$$pLDDT_{IF} = \frac{198.793}{\log(1726)}$$

$$pLDDT_{IF} = 61.41$$

3.3.3.2 PyMol: Monoplanar Angles of Two Centres of Mass

The angle of rise between two centres of mass in a protein complex (*A* and *B*) was calculated after re-orienting the complex using the 'rotate' command in

PyMol such that the protein was in the plane of the origin's orientation. Then, a straight line was drawn from point A to a point, C, which creates a right-angled triangle: x and z coordinates of A and the y coordinate of B. The 'get_angle' command could be used to find the angle $\widehat{CAB}$. This was all done in a Python script, 'pseudoatom_planar':

```
'''
Creates an object which has the same x and z coordinates as the second
object, but the same y coordinate as the first object. This puts two
atoms in the same plane such that planar and composite angles can be
calculated.
'''
from pymol import cmd
from sys import argv


def pseudoatom_planar(p1, p2, create=True):

    x1, y1, z1 = cmd.get_coords(p1)[0]
    x2, y2, z2 = cmd.get_coords(p2)[0]

    try:
        flat_point = cmd.get_unused_name("flat_point", 0)
    except AttributeError:
        flat_point = "flat_point"

    cmd.pseudoatom(flat_point, pos=[x2, y1, z2])
    cmd.show("spheres", flat_point)

    print("Angle ", flat_point, "•", p1, "•", p2, ": ",
cmd.get_angle(flat_point, p1, p2), "°")

    if (not create):
        cmd.delete(flat_point)

cmd.extend("pseudoatom_planar", pseudoatom_planar)
```

As with any .py file, this can be run using the 'run' command in Pymol followed by providing the two object arguments (pseudoatoms p1 and p2):

```
run pseudoatom_planar.py
pseudoatom_planar p1, p2
```

3.4 Results

3.4.1 Distribution of Functions of eCIS-Downstream ORFs

The conserved eCIS structure is found in many homologous variants spanning the microbial kingdoms. Due to the multi-component complexity of some of these operons, it can be difficult to distinguish structural genes from potential payloads. Thus, it was desirable to analyse a subset of eCISs, such as lineages Ia and Ib, to identify and classify the functions of potential eCIS payloads which are often, though not always, encoded adjacent to the structural operon (Geller *et al.*, 2021).

Using R within RStudio (R Core Team, 2020; RStudio Team, 2020), a script was written to use regular expressions (via the *grepl()* function) to count the occurrences of chosen keywords that best described many of the entries in the database of functions. Of those entries that did not include these keywords, the phrases 't\\dss', 'toxin', or 'effector' were used to infer the ORF as a toxin, designated as 'Other Toxin'; entries with none of these keywords were assigned to the category 'Other'. The distributions of predicted functions in lineage Ia and Ib are shown in Figure 3.1A and B.

Function counts were normalised by the number of strains within each species as well as the number of operons in each genome. For lineage Ib, the counts were also normalised by the number of species in the database. Thus, one can infer the frequency of each function for ORFs encoded up to 5 kb downstream of any given eCIS operon from any given strain.

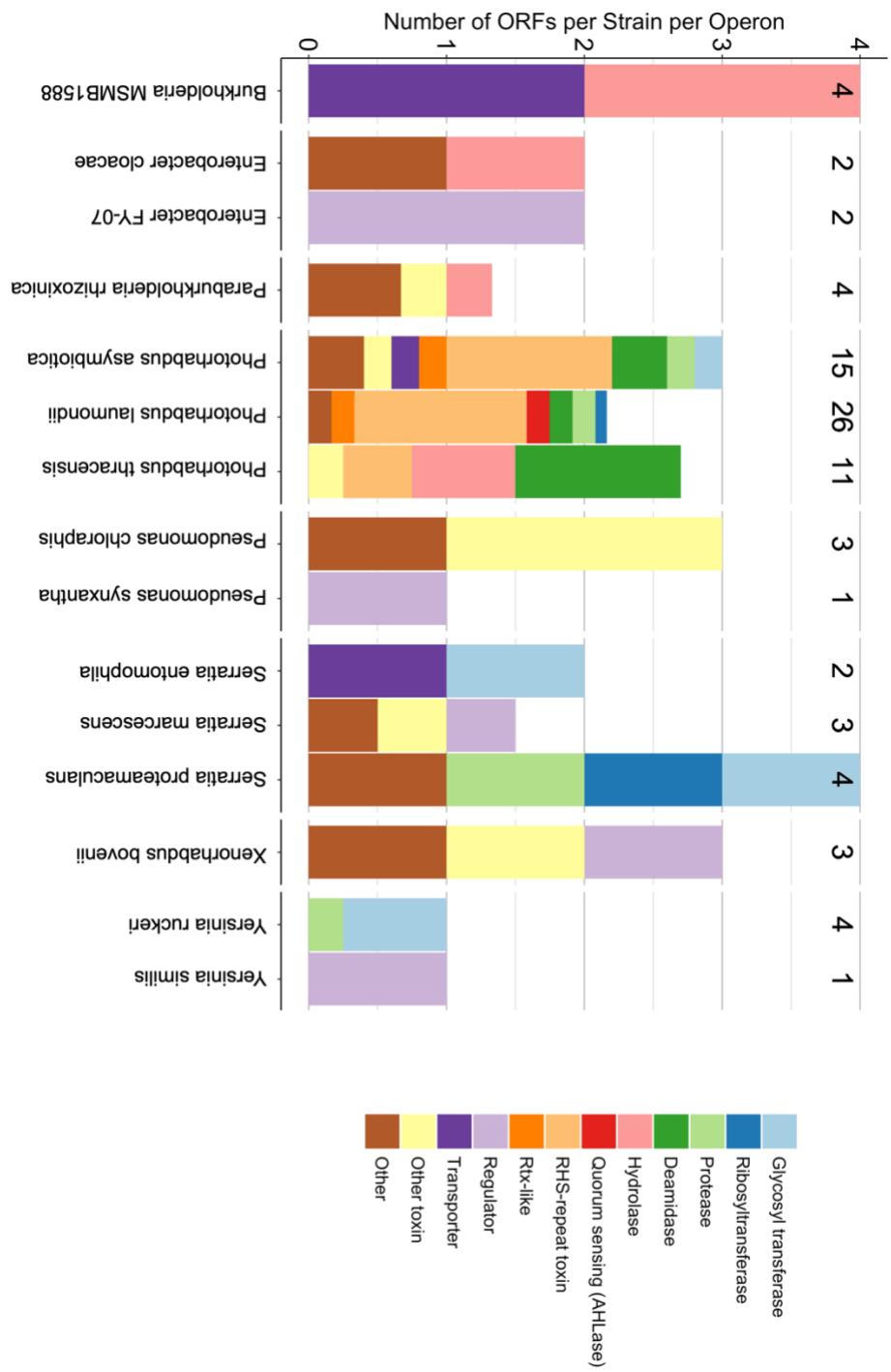
Photorhabdus eCIS-downstream operons are significantly more highly represented in their use of rearrangement hotspot (Rhs) repeat-containing proteins, particularly in *P. asymbiotica* and *P. luminescens*: both species which have been documented to cause mammalian infection. eCIS-downstream ORFs of the exclusively insect-pathogenic *P. temperata*, meanwhile, were more

commonly predicted to have deamidase and hydrolase activities. It should be noted that the dbeCIS database does not include certain strains of *Photorhabdus*, including the HIT and JUN European strains as well as *P. australis*.

The 5 kb cut-off in this methodology reflects the approximate sizes of each ORF within an eCIS operon as well as the distances between them. In lineage Ia, the analysed dataset contained 15 total species across 8 genera encoding a total of 41 core eCIS operons and 88 associated putative effector ORFs (Table 3.1). In lineage Ib, the dataset contained 67 total species across 42 genera encoding a total of 110 core eCIS operons and 240 associated putative effector ORFs (Table 3.2). 24 ORFs distributed across 13 species in lineage Ib (10% of all ORFs in Ib) had either sequence or domain architecture similarity to a highly conserved domain of unknown function (DUF), DUF4157: a previously-termed PVC-metallopeptidase, although its function has never been confirmed. It contains a Zn-binding motif, HExxH, typical of metallopeptidases; authors have speculated its function to be involved in toxin interactions such as for loading or export (Geller *et al.*, 2021). This hypothesis may be reflected in the current work since every instance of DUF4157 was encoded adjacent to another eCIS-downstream ORF with the only exception being *Stigmatella aurantiaca* strain DW4/3-1.

Notably, many eCIS effectors were predicted to have antibacterial functions; a finding which is also in agreement with Geller and colleagues, though their study extended to lineages outside of lineage I.

More details of the script used for analysing the custom database can be found in Appendices chapter 10.1. Appendices chapter 10.3 also shows the AlphaFold2 predictions of each of the ORFs identified in this analysis for finding the frequency of LSs.


A
Number of ORFs per Strain per Operon
0
1
2
3
4
Burkholderia MSMB1588
Enterobacter cloacae
Enterobacter FY-07
Paraburkholderia rhizoxinica
Photorhabdus asymbiotica
Photorhabdus laumondii
Photorhabdus thracensis
Pseudomonas chloraphis
Pseudomonas syrxantha
Serratia entomophila
Serratia marcescens
Serratia proteamaculans
Xenorhabdus bovenii
Yersinia ruckeri
Yersinia similis
4
2
2
4
15
26
11
3
1
2
3
4
3
4
1
Glycosyl transferase
Ribosyltransferase
Protease
Deamidase
Hydrolase
Quorum sensing (AHLase)
RHS-repeat toxin
Rtx-like
Regulator
Transporter
Other toxin
Other

B

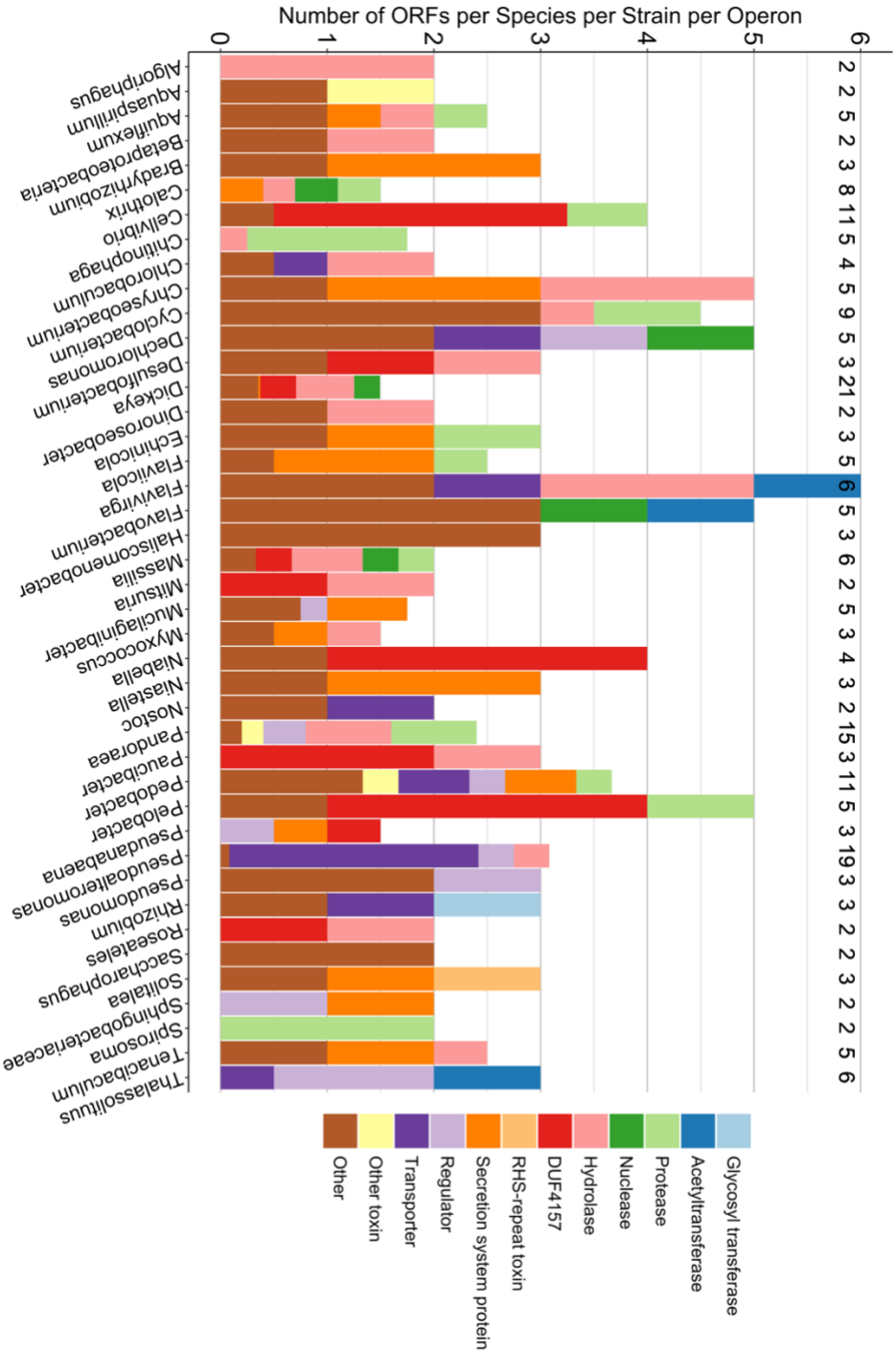

Number of ORFs per Species per Strain per Operon
0
1
2
3
4
5
6
Thalassolituus
Tenacibaculum
Spirosoma
Sphingobacteriaceae
Solitalea
Saccharophagus
Roseateles
Rhizobium
Pseudomonas
Pseudoalteromonas
Pseudanabaena
Pelobacter
Pedobacter
Paucibacter
Pandorea
Nostoc
Niastella
Niabella
Myxococcus
Mucilaginibacter
Mitsuria
Massilia
Haliscomenobacter
Flavobacterium
Flaviviriga
Flavicola
Echinicola
Dinoroseobacter
Dickeya
Desulfobacterium
Dechloromonas
Cyclobacterium
Chryseobacterium
Chlorobaculum
Chitinophaga
Cellvibrio
Calothrix
Bradyrhizobium
Betaproteobacteria
Aquiflexum
Aquaspirillum
Algorphagus
2 2 5 2 3 8 11 5 4 5 9 5 3 2 1 2 3 5 6 5 3 6 2 5 3 4 3 2 15 3 11 5 3 19 3 3 2 2 3 2 2 5 6
Other
Other toxin
Transporter
Regulator
Secretion system protein
RHS-repeat toxin
DUF4157
Hydrolase
Nuclease
Protease
Acetyltransferase
Glycosyl transferase

Figure 3.1 | Predicted functions of lineage Ia and Ib putative eCIS payloads.

The online dbeCIS database was used to identify structural operons of extracellular contractile injection system (eCIS) subtypes belonging to lineage Ia **(A)** and Ib **(B)**, as defined by Chen *et al*., 2019. Open reading frames (ORFs) up to 5 kb downstream of the final structural gene were appended to a custom database to explore the functions of likely payloads. BLASTp and HHPred programs were used to identify sequence homology to predict each ORF's most likely function, and regular expressions were used to count how many times certain keywords appeared using scripts written in R. ORFs which did not fit any of the assigned keywords were classed as either 'Other Toxin', or 'Other' depending on what additional information was found for that entry. Each bar represents a normalised value of the average number of ORFs found within 5 kb of the final amino acid of the final eCIS-structural gene in the operon for any given operon in any given strain or species, where applicable. Numbers presented at the top of each bar represent the total number of ORFs used to generate the bar. The diversity of functions is represented as subsets of each bar; functions were chosen based on the most commonly predicted functions appearing in the database. In **(B)**, 'Protease' and 'Nuclease' are treated as subsets of 'Hydrolase'; they are preferred to the more ambiguous 'Hydrolase' term and are only assigned as such in the absence of more specific information from the homology search. Genera encoding only one eCIS operon with one or less downstream ORFs were excluded from analysis; lineage Ia: *Chania* and *Shewanella*, lineage Ib: *Gynuella*, *Minicystis*, *Sorangium*, *Moorea*, *Geobacter*, *Dyadobacter*, *Draconibacterium*, *Anabaena*, *Aulosira*, *Tolypothrix*, *Fremyella*, *Herbaspirillum*, *Tenderia*, and *Stigmatella*.

Table 3.1 | Lineage Ia operon numbers by strain.

The dataset of lineage Ia includes a total of 15 species across 8 genera. The database initially included *P. laumondii* as a strain of *P. luminescens* and *P. thracensis* as a strain of *P. temperata*. This has been updated to reflect the new taxonomy (Machado *et al*., 2018).

Strain	Operons
Burkholderia sp. MSMB1588	1
Enterobacter sp. FY-07	1
Enterobacter cloacae complex 'Hoffman cluster IV' DSM 16690	1
Paraburkholderia rhizoxinica HKI	3
Photorhabdus asymbiotica ATCC43949	5
Photorhabdus laumondii DPSV002N	6
Photorhabdus luminescens TT01	6
Photorhabdus thracensis DSM 15199	4
Pseudomonas chloraphis 189	1
Pseudomonas synxantha LBUM223	1
Serratia entomophila A1MO2	1
Serratia marcescens BSR3	1
Serratia marcescens SOLR4	1
Serratia proteamaculans AGR96X	1
Xenorhabdus bovienii CS03	1
Yersinia ruckeri Big Creek 74	1
Yersinia ruckeri SC09	1
Yersinia ruckeri QMA0440	1
Yersinia ruckeri YRB	1
Yersinia similis 228	1

Table 3.2 | Lineage Ib operon numbers by species.

The dataset in lineage Ib includes a total of 67 species across 42 genera. Most species possess only 1 strain, though *Pseudoalteromonas luteoviolacea* includes 4 strains.

Species / Strain	Operons	Species / Strain	Operons
***Algoriphagus* sp. PR1**	2	*Mitsuaria* sp. 7	2
***Aquaspirillum* sp. LM1**	2	*Mucilaginibacter mellensis*	3
Aquiflexum balticum	5	*Mucilaginibacter* sp. BJC16-A31	2
Betaproteobacteria bacterium	2	*Myxococcus stipitatus*	1
Bradyrhizobium oligotrophicum	3	*Myxococcus fulvus*	2
***Calothrix* sp. NIES-2098**	2	*Niabella soli*	4
***Calothrix* sp. NIES-2100**	2	*Niastella koreensis*	3
***Calothrix* sp. PCC**	1	*Nostoc linckia*	2
***Calothrix* sp. 336-3**	2	*Pandoraea norimbergensis*	2
***Calothrix* sp. NIES-3974**	1	*Pandoraea oxalativorans*	2
Cellvibrio japonicus	5	*Pandoraea sputorum*	6
***Cellvibrio* sp. PSBB023**	6	*Pandoraea faecigallinarum*	3
Chitinophaga caeni	2	*Pandoraea pulmonicola*	2
Chitinophaga pinensis	3	*Paucibacter* sp. KCTC	3
Chlorobaculum limnaeum	1	*Pedobacter ginsengisoli*	4
Chlorobaculum parvum	3	*Pedobacter steynii*	4
***Chryseobacterium* sp. IHBB**	5	*Pedobacter* sp. PACM	3
Cyclobacterium amurskyense	4	*Pelobacter propionicus*	5
Cyclobacterium marinum	5	*Pseudanabaena* sp. PCC	3
Dechloromonas aromatica	5	*Pseudoalteromonas* sp. 12015MBL	3
Desulfobacterium autotrophicum	3	*Pseudoalteromonas* sp. Bsw20308	3
Dickeya fangzhongdai	4	*Pseudoalteromonas luteoviolacea* S4054249	3
Dickeya solani	9	*Pseudoalteromonas luteoviolacea* S4054	3
Dickeya chrysathemi	1	*Pseudoalteromonas luteoviolacea* S40542	3
Dickeya zeae	5	*Pseudoalteromonas luteoviolacea* HI1	4
Dickeya dianthicola	2	*Pseudomonas knackmussii*	3
Dinoroseobacter shibae	2	*Rhizobium* sp. NXC24	3
Echinicola vietnamensis	3	*Roseateles depolymerans*	2
Flaviicola riflensis	5	*Saccharophagus degradans*	2
Flavivirga eckloniae	6	*Solitalea canadensis*	3
Flavobacterium anhuiense	5	*Sphingobacteriaceae* sp. GW460-11-11-14-LB5	2
Haliscomenobacter hydrossis	3	*Spirosoma pollinicola*	2
***Massilia* sp. B2**	3	*Tenacibaculum jejuense*	4
***Massilia* sp. ZMN-3**	1	*Tenacibaculum* sp. SZ-18	1
Massilia timonae	2	*Thalassolituus oleivorans*	6

3.4.1.1 Ubiquity of Leader Sequences in eCIS-downstream ORFs

To affirm that leader sequences (LSs) are ubiquitous in eCIS, the custom database was also used for structural prediction of lineage Ia ORFs using AlphaFold2 (Jumper *et al.*, 2021). Any long, low-confidence N-terminal extensions were recorded: amino acids 1-50 possessing an average pLDDT < 40. LSs were found in 55.4% of ORFs up to 5 kb downstream of a lineage Ia eCIS operon (Figure 3.2). These leaders are localised almost entirely to the PVC loci in *Photorhabdus* and the Afp loci in *Serratia* and *Yersinia*. Therefore, whilst LSs may be a common feature of Afp-like eCISs, other mechanisms of eCIS loading may exist.

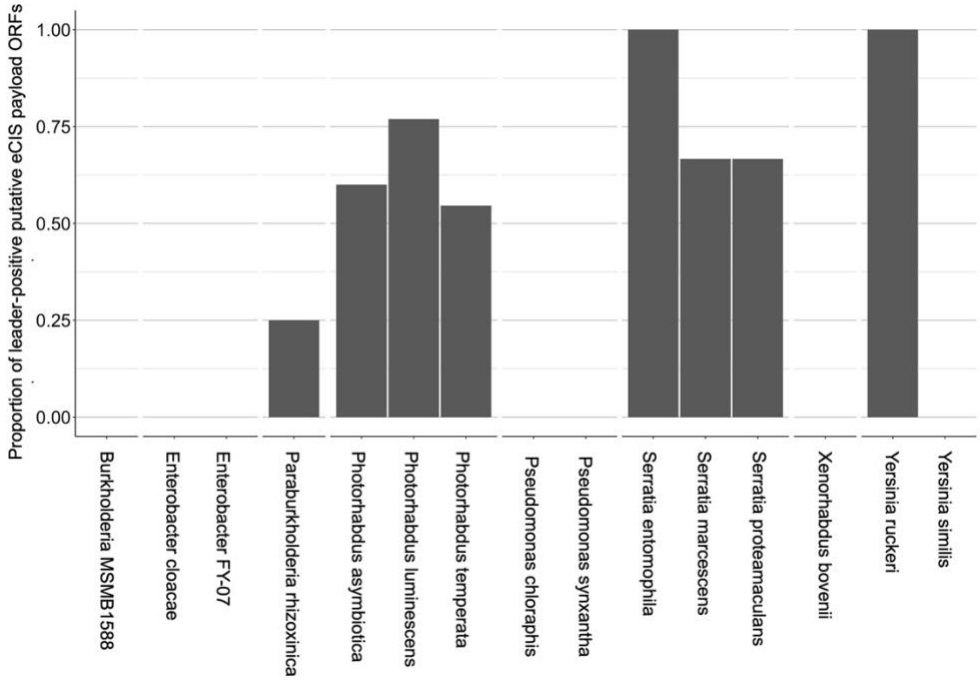


Figure 3.2 | Ubiquity of leader sequences in lineage Ia putative eCIS payloads.
AlphaFold 2.0 was used to predict the presence of low-confidence (i.e highly disordered) N-terminal extensions in ORFs encoded within 5 kb downstream of lineage Ia eCIS loci encoded in the genomes of Gram-negative bacteria. *Photorhabdus* is known to encode the PVC whilst *Serratia* and *Yersinia* are known to encode Afp with an Afp18/AfpX18 ORF downstream. Sample sequences with only one ORF and one operon were excluded (e.g *Chania* and

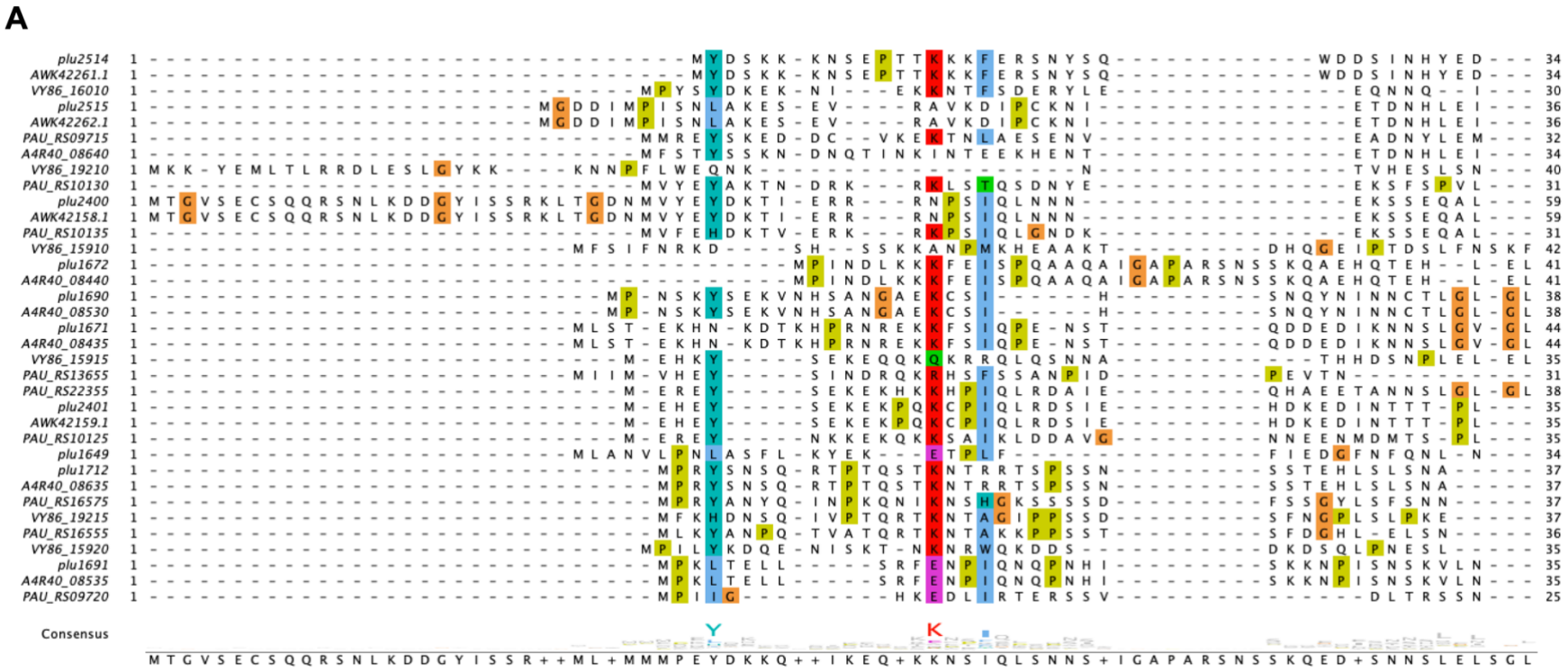
Shewanella). The overall ubiquity of leader sequences in lineage Ia is 55.4% and is heavily biased towards the Afp-homologous structures.

3.4.1.2 Assessment of *Photorhabdus* Leader Sequence Diversity

Sequence alignment of putative PVC effectors, as done by Jiang and colleagues (2022), reveals vaguely conserved residues in the effector LSs (the first ~70 amino acids of most putative effectors). Whilst some putative PVC effector leaders do, indeed, show homology, they are often restricted to short motifs in i) the orthologous gene in a different strain, or ii) other putative effectors downstream of the same PVC operon; the latter of which may reflect the unique interactions between effectors and their respective PVC core components.

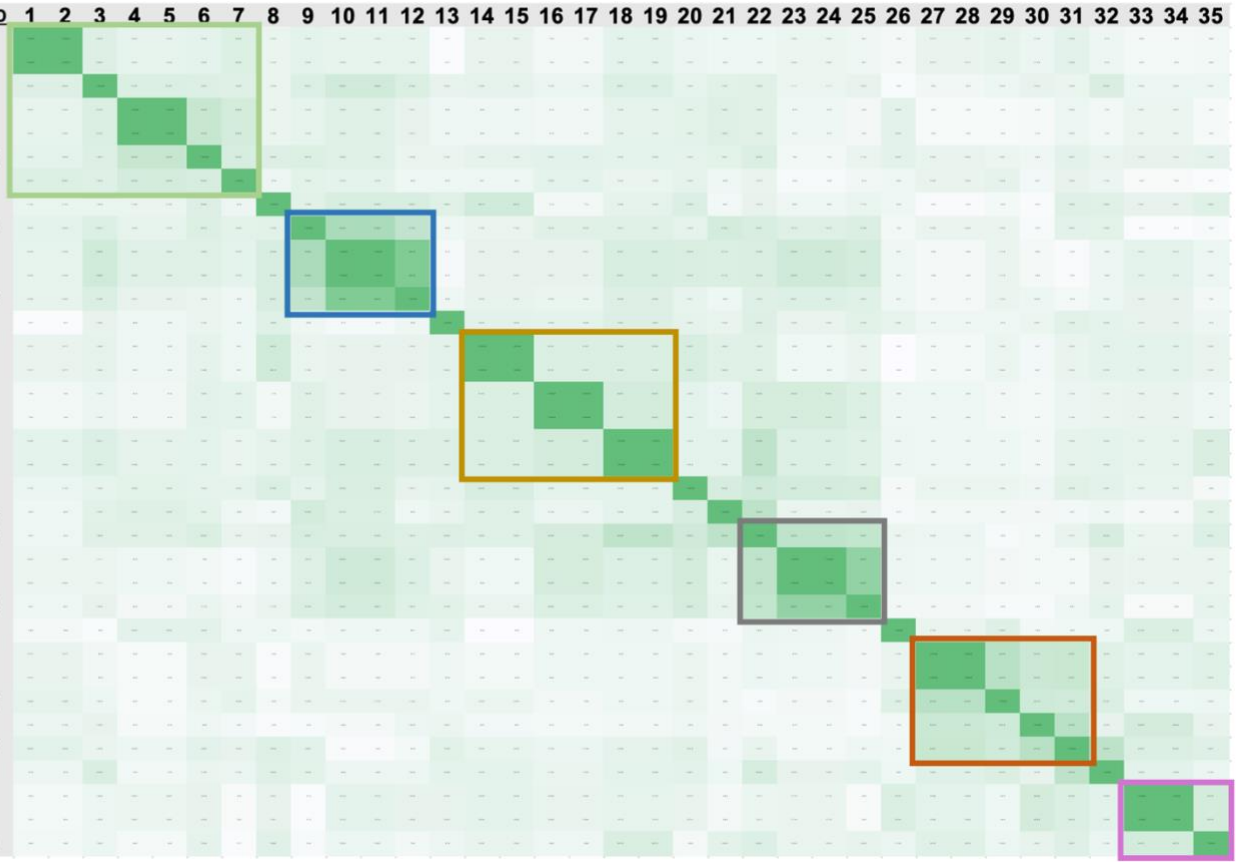
To better describe the 'groups' of PVC effector LSs, local alignment was performed using Clustal Omega on all *Photorhabdus* putative effector LSs (the first 70 amino acids identified using AlphaFold2.0, as will be discussed in the next chapter) using Clustal Omega (Figure 3.3A). A heatmap was generated from the percent identity (%ID) matrix to identify groups of leaders which cluster together (Figure 3.3B). These clusters were aligned again via Clustal Omega to compare LS diversity like-for-like (Figure 3.3C). Using this method, only a few LSs clustered less well: most of these outliers (5 out of 6) occurred downstream of unit 4 PVCs and 4 out of 6 were found in *P. thracensis*.

Overall, 6 general groups of leaders were identified: i) the Cif-Unit3 group, ii) the Rtx group, iii) the *Luminescens* Rhs group (Rhs-*Plum*), iv) the LopT group, v) the Pnf group, and vi) the Tcc-like group. More common instances of sequence conservation were found in these groups compared to all-versus-all alignment with distinct motifs identified in each. The diversity observed between effectors from different PVCs may be indicative of their specificity to load into their cognate PVC, rather than loading into a PVC from *any* operon.


A
plu2514
AWK42261.1
VY86_16010
plu2515
AWK42262.1
PAU_RS09715
A4R40_08640
VY86_19210
PAU_RS10130
plu2400
AWK42158.1
PAU_RS10135
VY86_15910
plu1672
A4R40_08440
plu1690
A4R40_08530
plu1671
A4R40_08435
VY86_15915
PAU_RS13655
PAU_RS22355
plu2401
AWK42159.1
PAU_RS10125
plu1649
plu1712
A4R40_08635
PAU_RS16575
VY86_19215
PAU_RS16555
VY86_15920
plu1691
A4R40_08535
PAU_RS09720
Consensus
M T G V S E C S Q Q R S N L K D D G Y I S S R + + M L + M M M P E Y D K K Q + + I K E Q + K K N S I Q L S N N S + I G A P A R S N S S K Q E D + S N N S L E L S G L

B

Number	Species	Strain	Operon	Gene / Protein ID
1	luminescens	TT01	PVC*cif*	plu2514
2	luminescens	laumondii	PVC*cif*	AWK42261.1
3	temperata	thracensis	PVCu3	VY86_16010
4	luminescens	TT01	PVC*cif*	plu2515
5	luminescens	laumondii	PVC*cif*	AWK42262.1
6	asymbiotica	ATCC43949	PVC*cif*	PAU_RS09715
7	luminescens	laumondii	PVCu3	A4R40_08640
8	temperata	thracensis	PVCu1	VY86_19210
9	asymbiotica	ATCC43949	PVC*lopT*	PAU_RS10130
10	luminescens	TT01	PVC*lopT*	plu2400
11	luminescens	laumondii	PVC*lopT*	AWK42158.1
12	asymbiotica	ATCC43949	PVC*lopT*	PAU_RS10135
13	temperata	thracensis	PVCu4	VY86_15910
14	luminescens	TT01	PVCu3	plu1672
15	luminescens	laumondii	PVCu4	A4R40_08440
16	luminescens	TT01	PVCu2	plu1690
17	luminescens	laumondii	PVCu1	A4R40_08530
18	luminescens	TT01	PVCu3	plu1671
19	luminescens	laumondii	PVCu4	A4R40_08435
20	temperata	thracensis	PVCu4	VY86_15915
21	asymbiotica	ATCC43949	PVCu4	PAU_RS13655
22	asymbiotica	ATCC43949	PVCu4	PAU_RS22355
23	luminescens	TT01	PVC*lopT*	plu2401
24	luminescens	laumondii	PVC*lopT*	AWK42159.1
25	asymbiotica	ATCC43949	PVC*lopT*	PAU_RS10125
26	luminescens	TT01	PVCu4	plu1649
27	luminescens	TT01	PVCu1	plu1712
28	luminescens	laumondii	PVCu3	A4R40_08635
29	asymbiotica	ATCC43949	PVC*pnf*	PAU_RS16575
30	temperata	thracensis	PVCu1	VY86_19215
31	asymbiotica	ATCC43949	PVC*pnf*	PAU_RS16555
32	temperata	thracensis	PVCu4	VY86_15920
33	luminescens	TT01	PVCu2	plu1691
34	luminescens	laumondii	PVCu1	A4R40_08535
35	asymbiotica	ATCC43949	PVC*cif*	PAU_RS09720



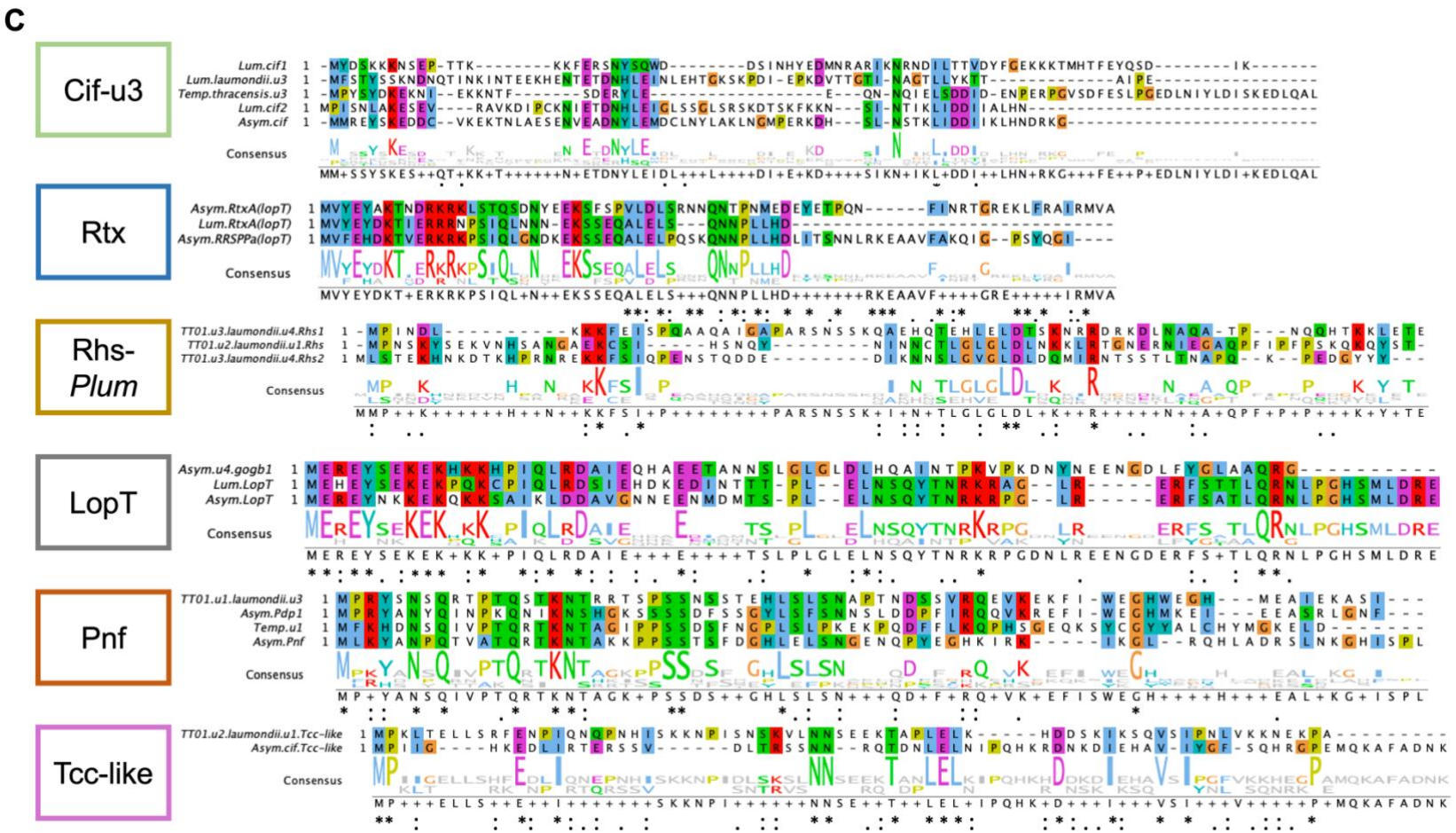

c
Cif-u3
Rtx
Rhs-
Plum
LopT
Pnf
Tcc-like
Consensus

Figure 3.3 | Putative PVC effector LSs show conserved motifs within clustered groups.

Putative PVC effector ORFs were identified from downstream of each operon entry in the dbeCIS database (therefore only including strains *P. asymbiotica* ATCC43949, *P. luminescens* TT01, *P. luminescens laumondii* DPSV002N, and *P. temperata thracensis* DSM 15199). AlphaFold2 was used to identify ORFs encoding highly disordered leader sequences. **(A)** Clustal Omega was used to align all sequences across these strains which show some rare residue conservation (e.g Pnf's Lys16). **(B)** From the Clustal Omega alignment, the percent identity matrix, which is ordered according to pairwise comparisons of each entry, was coloured according to its values to make a heatmap of sequence identity. Groups were selected based on effectors with identities of typically greater than 20% to its neighbours in the matrix. **(C)** For each of the 6 groups identified from the heatmap, Clustal Omega was used again to align sequences which appeared to share certain amino acid motifs.

3.4.2 Effector Predictions

3.4.2.1 Pnf and Cif Effector Homology

To assess the accuracy of AlphaFold2's 3-dimensional predictions for the effectors of focus in this work, the predicted structure of the catalytic domain of *E. coli* CNF1 was compared to the experimentally resolved structure by Buetow and colleagues (2001) (PDB: 1HQ0). Additionally, the Cif homologues from enteropathogenic and enterohemorrhagic *E. coli* (Cif_{Ec}; PDB: 3EFY), *B. pseudomallei* (Cif_{Bp}; PDB: 3GQM), and *P. luminescens* (Cif_{Pl}; PDB: 3GQJ) were compared against their predicted structures (Hsu *et al.*, 2008; Crow *et al.*, 2009) (Figure 3.4). Whilst a structure is resolved for *Y. pseudotuberculosis* (PDB: 4F8C), this structure is in complex with its NEDD8 target protein so it was not used for comparison to the output AlphaFold2 .pdb files (Crow *et al.*, 2012).

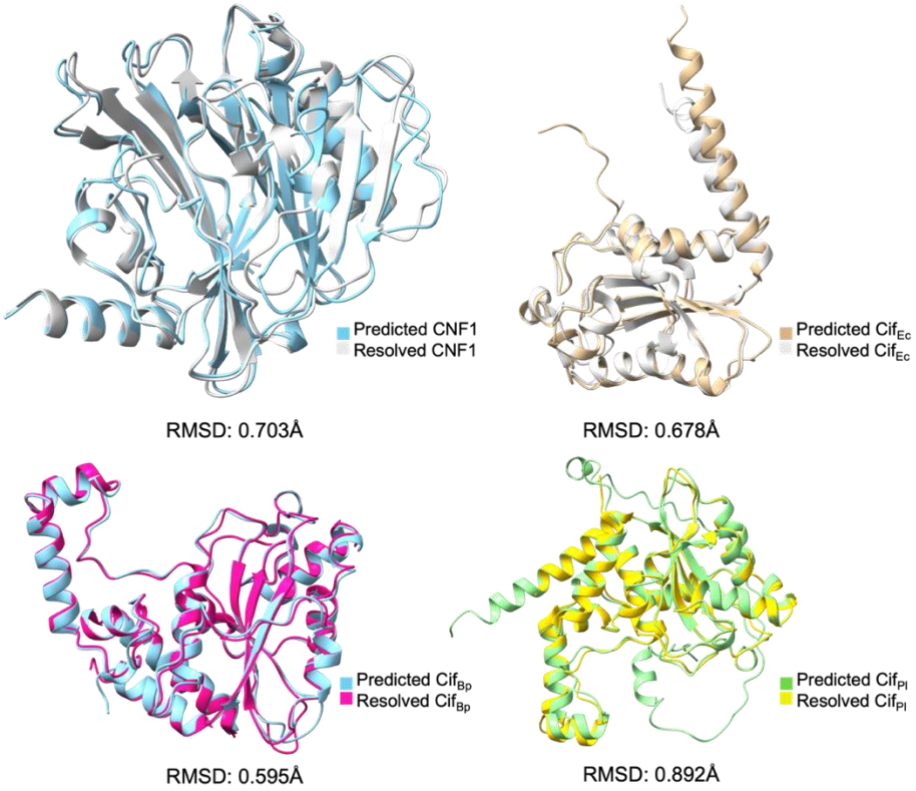


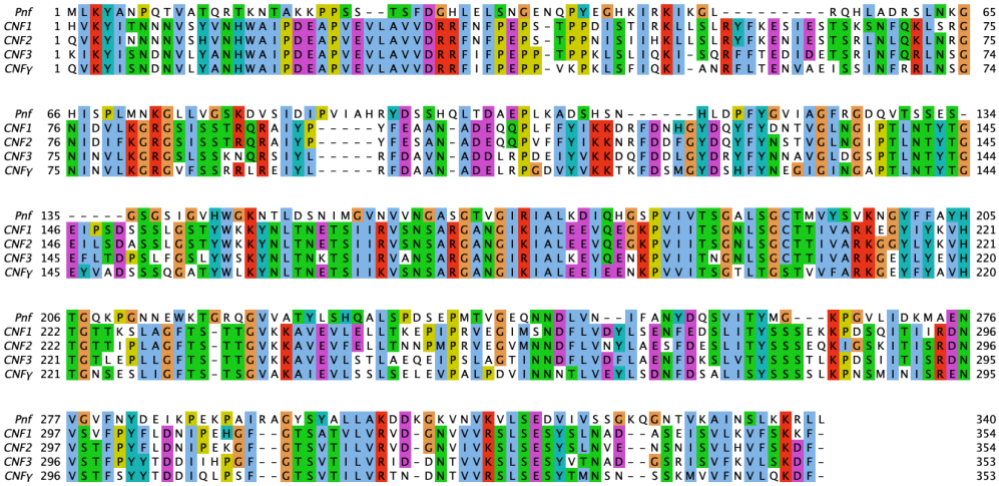
Figure 3.4 | AlphaFold2 is an accurate predictor of effector structure.

The predicted *E. coli* CNF1 and Cif$_{Ec}$ structures were compared to their resolved structures (1HQ0 and 3EFY, respectively) as well as the *B. pseudomallei* Cif$_{Bp}$ and the *P. luminescens* Cif$_{Pl}$ (3GQM and 3GQJ, respectively) and overlayed using Matchmaker in the ChimeraX suite. Root mean square deviations (RMSD) is indicated below each overlay and was consistently shown to be below 0.9Å. Overall pLDDT scores for each predicted structure was consistently above 80%, indicating a confident prediction from AlphaFold. The resolved structures lack leader sequences (residues 1- ~49), presumably because they have low confidence and are omitted from the resolved PDB models; the leaders of the predicted structures, therefore, were also omitted for the purpose of this alignment.

After confirming good alignment between resolved and predicted structures, Pnf from *P. asymbiotica* (Pnf$_{Pa}$) was aligned with its CNF homologues from *E. coli* (CNF1, 2, and 3) and *Yersinia pseudotuberculosis* (CNFγ). The overall structures of the CNFs were far larger than that of Pnf: roughly 1000 amino acids versus only 340 amino acids for the full-length Pnf (full sequence alignments and the percent identity matrix can be seen in Appendices 10.2). Pnf

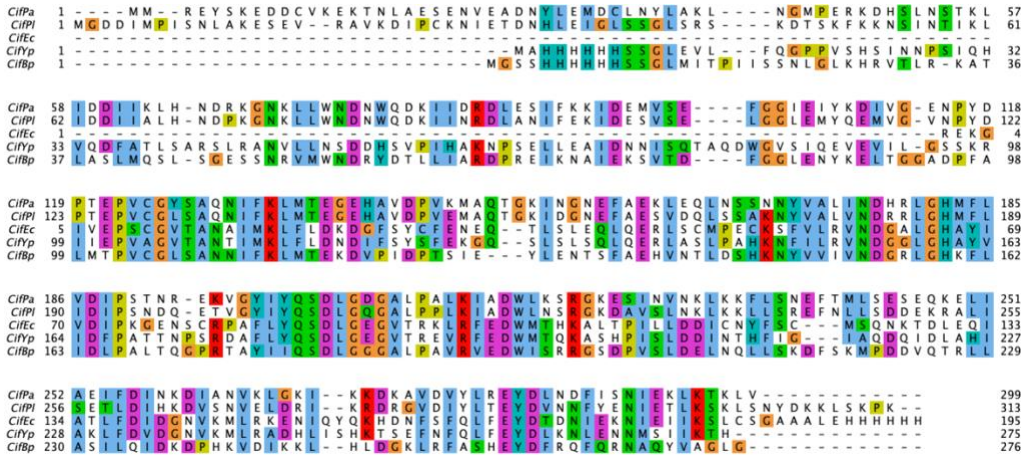
aligns most closely to the D5 catalytic deamidase domain of CNF so only this region of each CNF was used for alignment; each is roughly 350 amino acids and is designated here as CNF1-cat, CNF2-cat, CNF3-cat, and CNFγ-cat. Overall, Pnf shares very low sequence identity with the CNF-cat homologues; the same is true for the Cif homologues (Figure 3.5). Despite the apparent distinguishment between the catalytic domains of the CNFs and that of full-length Pnf, however, the predicted tertiary structures were found to be strikingly similar. Pnf does, indeed, encode a leader sequence which displayed low pLDDT (confidence) score in the AlphaFold prediction, and which does not appear to fold as part of the more globular catalytic structure (Figure 3.6).

A



% ID	Pnf	CNF1	CNF2	CNF3	CNFγ
Pnf	100.00	26.07	26.99	27.69	25.85
CNF1	26.07	100.00	84.75	68.56	59.77
CNF2	26.99	84.75	100.00	68.56	59.77
CNF3	27.69	68.56	68.56	100.00	68.84
CNFγ	25.85	59.77	59.77	68.84	100.00

B



% ID	CifPa	CifPl	CifEc	CifYp	CifBp
CifPa	100.00	63.64	27.78	22.56	28.57
CifPl	63.64	100.00	26.32	24.81	37.78
CifEc	27.78	26.32	100.00	61.33	27.37
CifYp	22.56	24.81	61.33	100.00	28.57
CifBp	28.57	37.78	27.37	28.57	100.00

Figure 3.5 | Local multiple sequence alignment of Pnf and CNF catalytic domain homologues and Cif homologues.

Local Clustal Omega multiple sequence alignment (MSA) was done for **(A)** Pnf and the catalytic domain of its CNF homologues and **(B)** the Cif homologues. Alignments are coloured according to the nature of each amino acid given in Clustalx form in Jalview. Also displayed are the percent identities (% ID) for each of the homologues. Far greater identity is observed between the CNF homologues than to that of Pnf from *Photorhabdus asymbiotica*, whilst the *Photorhabdus* Cif effectors had greater identity to one another than that of the other Cif homologues; only Cif_{Ec} and Cif_{Yp} had a similar % ID. Cells are coloured according to their value between 0 and 100. Percent identities were calculated using Clustal Omega.

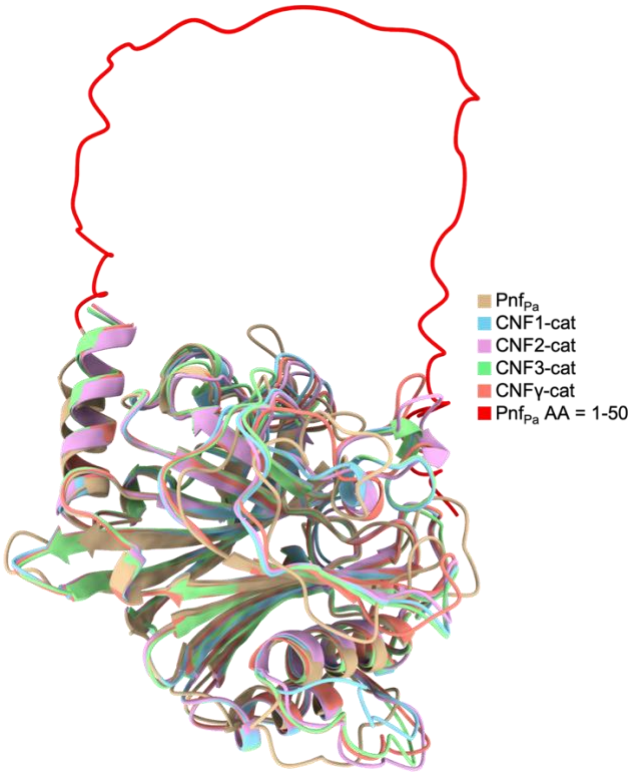


Figure 3.6 | Pnf and CNF catalytic domains form similar protein structures.
AlphaFold v2.0 was used to predict the 3-dimensional structures of Pnf from *P. asymbiotica*, CNF1, 2, and 3 from *E. coli*, and CNFγ from *Y. pseudotuberculosis*. The first 50 amino acids of Pnf are shown in red; a region of low confidence from the prediction and which exists outside of the more folded, globular regions of the catalytic domain.

The same low score pLDDT sequence scores were found for the first 50 amino acids of the Cif effector homologues. Unlike CNF, however, these effector homologues were much smaller and are loaded into secretion systems in their native species – as well as the PVC*cif* chassis in *P. luminescens* and *P. asymbiotica* (Figure 3.7). This implicates the LSs as inherently disordered structures which likely owes to a mechanism of loading in the respective needle-like structures. Intriguingly, only Cif_{Ec} lacked a low-pLDDT N-terminal extended LS.

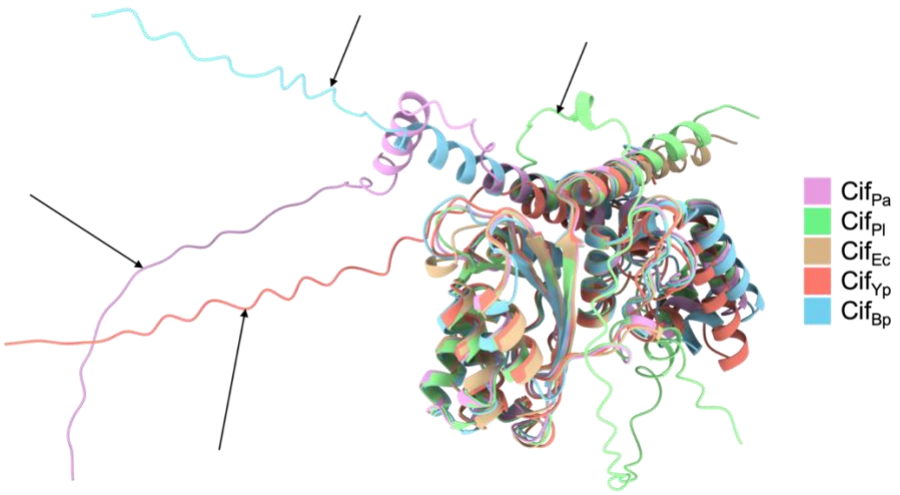


Figure 3.7 | Cif homologues form similar protein structures and encode LSs.
AlphaFold v2.0 gave low-confidence scores to the first ~50 N-terminal amino acids for each of the Cif effector homologues, despite large similarity for the rest of the structures. Structures were superimposed using Matchmaker within the ChimeraX suite. Arrows indicate the low-pLDDT (confidence) region found at the first ~50 amino acids of each structure, referred to as the leader sequence (LS); only Cif_{Ec} lacked this low-confidence region.

It was necessary to affirm that LSs are disordered structures as these would later be the subject of alterations in later experiments, such as truncations and switching the LS to a different type whilst retaining the epitope-tagged payload protein. For example, Pnf(C190A) is non-functional with its catalytic cysteine mutated to alanine for the purposes of expressing and purifying the protein, then verifying its abundance via its epitope tag in western blot. Modulating the 50-amino acid LS from the usual Pnf50 to the first 50 amino acids of Cif, Cif50, would also be useful for experiments. Hence, AlphaFold was used to predict the structure of the heterologous Cif50-Pnf(C190A)-HA and the native Pnf50-Pnf from *P. asymbiotica* (Pnf_{Pa}) (Figure 3.8). Both native and heterologous LSs formed low confidence predictions, indicating that the LS may not fold within the catalytic globular domain of Pnf.

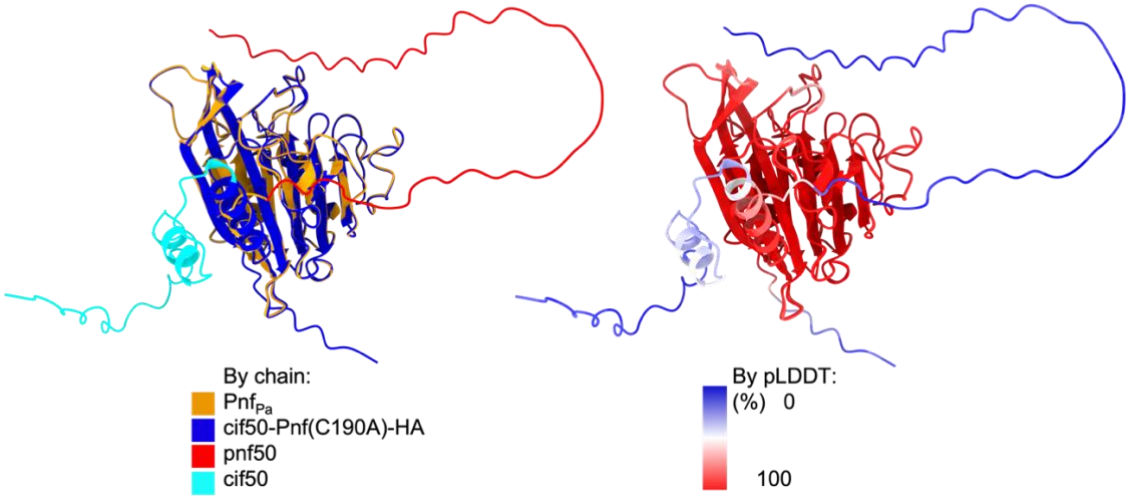
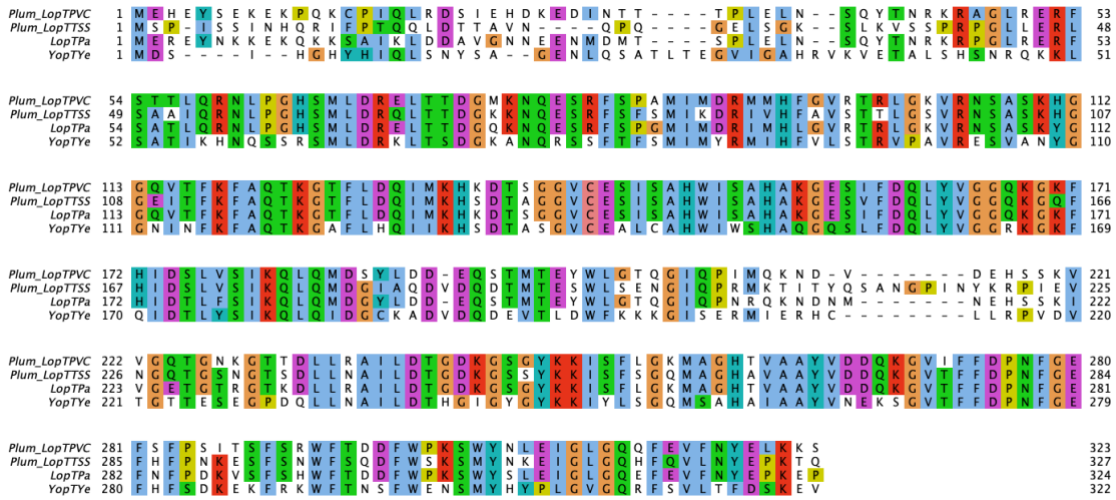


Figure 3.8 | N-terminal disorder predicts LSs through low pLDDT.

The native Pnf50 LS represents the first 50 amino acids of the native *P. asymbiotica* Pnf$_{Pa}$, whereas Cif50 represents the first 50 amino acids of the *P. asymbiotica* Cif effector. Predictions of the heterologous Cif50-Pnf(C190A)-HA structure and indicate that the leader sequence does not fold within the rest of the Pnf structure. Predicted LDDT (pLDDT) is shown in the right-hand figure where high confidence residue conformations are redder and low confidence residue conformations are bluer.

3.4.2.2 Comparison of LopT-PVC with LopT-TSS

In *P. luminescens* TT01, the LopT effector is encoded at two loci: one is loaded into the PVC*lopT* complex (encoded as gene *plu2400*) whilst the other is used as an effector for a T3SS (*plu3788*). To assess whether they differed in overall structure, AlphaFold predictions were overlayed and showed good similarity by local Clustal Omega alignments (Figure 3.9).



% ID	Plum_LopT-PVC	Plum_LopT-TSS	LopTPa	YopTYe
Plum_LopT-PVC	100.00	67.30	84.52	45.54
Plum_LopT-TSS	67.30	100.00	66.46	52.56
LopTPa	84.52	66.46	100.00	46.67
YopTYe	45.54	52.56	46.67	100.00

Figure 3.9 | LopT and YopT local sequence alignments.

Local Clustal Omega multiple sequence alignment (MSA) was done for LopT encoded in *Photorhabdus luminescens* and *P. asymbiotica* as well as the YopT homologue from *Yersinia enterocolitica*. The YopT homologue differs most significantly from the other LopT homologues. YopT from *Bacillus subtilis* has a sequence far shorter than the other homologues so could not be fairly compared using these alignments.

LopT and the T3SS-associated YopT homologues also displayed low-confidence 50-60 aa LSs at their N-termini (Figure 3.10). There was also high structural similarity between the PVC- and T3SS-associated LopT proteins from *P. luminescens*; both displayed ~50 aa LSs and retained a highly similar globular structure. Altogether, these predicted structures indicate a pattern of payload effectors of both extracellular contractile- and T3SS-associated structures to contain low-confidence, highly disordered LSs at the N-terminus.

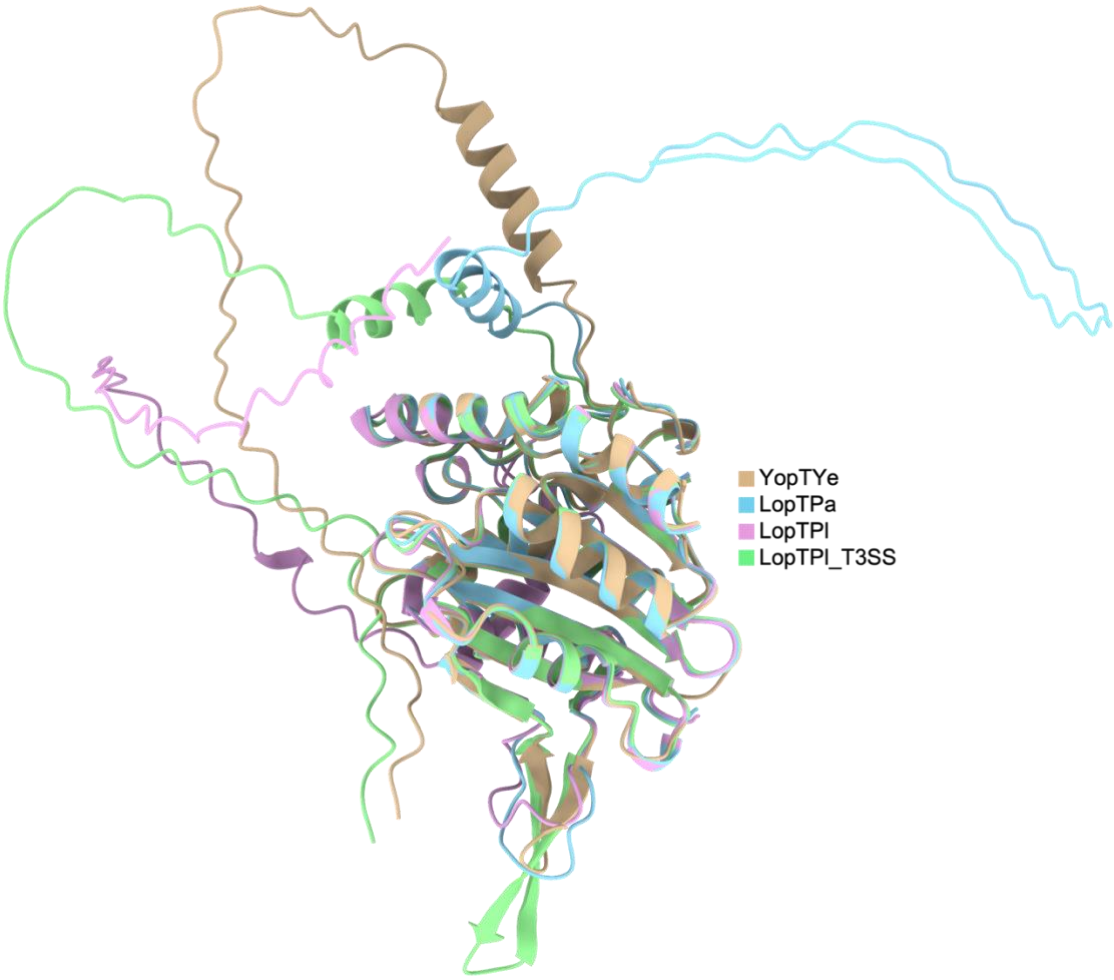


Figure 3.10 | LopT and YopT form similar protein structures and encode LSs.

YopT from *Yersinia enterocolitica* was overlayed with LopT homologues from *Photorhabdus luminescens* and *P. asymbiotica*. Both PVC and T3SS-associated loci in *P. luminescens* (*plu2400* and *plu3877*, respectively) showed the presence of low-confidence (disordered) leader sequences within the first 50-60 amino acids.

3.4.3 Predicting the Structure and Function of Pvc15

It was desirable to understand more about the structure and function of the elusive Pvc15 ATPase associated with diverse cellular activities (AAA+). Therefore, it was first investigated how similar the different Pvc15 genes are between *Photorhabdus* loci, as well as the homologous Afp15 from *S. entomophila* and ClpV, also termed TssH, from *Pseudomonas spp*.

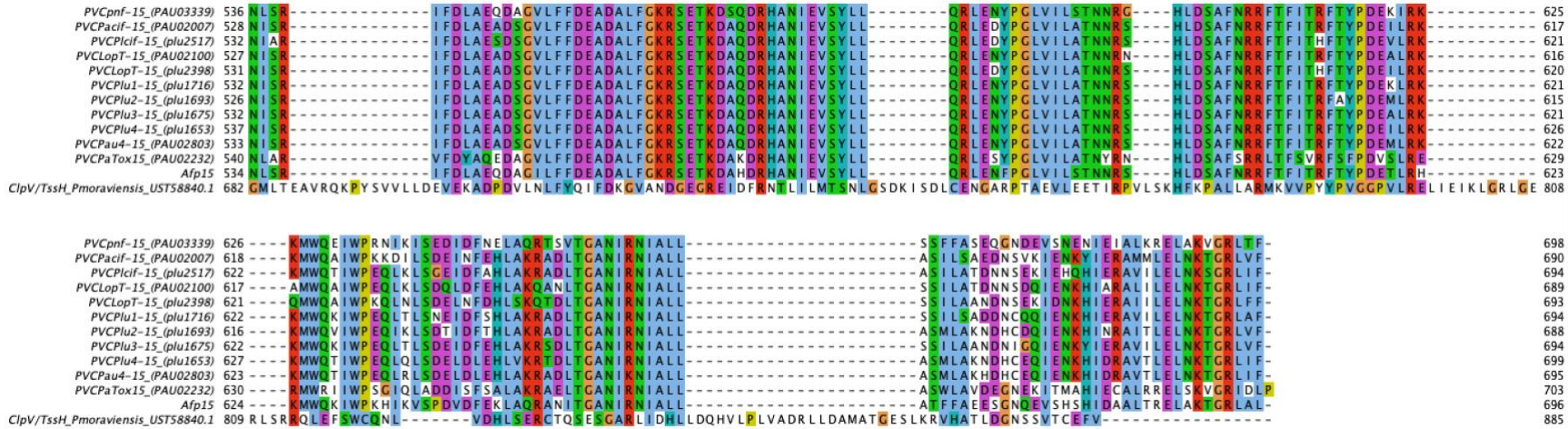
Pvc15 is well-retained between each locus in both *P. asymbiotica* and *P. luminescens*; in particular, the Pvc15 located in the PaTox locus (hereafter designated Pvc15$_{PaTox}$) was most dissimilar to all other ATPases compared, according to its percent identity of which the N-terminal residues appeared to differ the most (Figure 3.11). Pnf Pvc15 (Pvc15$_{pnf}$) was also more dissimilar to the other homologues; in fact, the *S. entomophila* Afp15 homologue displayed greater percent identity to Pvc15$_{pnf}$ than any of the other homologues (65.66%; the second largest identity is Pvc15$_{unit1}$ at 57.62%). Meanwhile, the T6SS-associated ATPase, ClpV from *Pseudomonas moraviensis*, is far larger than the other ATPases by almost 200 amino acids; it is also one of the most distantly-related of bacterial ATPases (Jiang *et al.*, 2022).

PVCpnf-15_(PAU03339)
PVCPacif-15_(PAU02007)
PVCPlcif-15_(plu2517)
PVCLopT-15_(plu2100)
PVCLopT-15_(plu2398)
PVCPlu1-15_(plu1716)
PVCPlu2-15_(plu1693)
PVCPlu3-15_(plu1675)
PVCPlu4-15_(plu1653)
PVCPau4-15_(PAU02803)
PVCPaTox15_(PAU02232)
Afp15
ClpV/TssH_Pmoraviensis_UST58840.1

PVCpnf−15_(PAU03339) 229 KQVLGNVLLSDNDNIRPIVLLRGMAGSARAYTIT-----NMMASEGKQTLLVDISKLADSDEKN-------IIL----------QIKHILRE----TRMHGACLLLRNFCLLVEQNKQLLDSL 325
PVCPacif−15_(PAU02007) 220 CHSLEKILLNETDKVRPLVILRGKQASARELTVS-----NIMAFHSINTLVLDLARLPNEDNST-------TIS----------DLLKNAVRE----ARLHDACLLIRNFSLLAEEKKILNSEL 317
PVCPlcif−15_(plu2517) 224 YHSFEKILLNETDEIRPLVILRGKQGSARELAVS-----NIMAFHSISTLILDLARLPEEENLT-------VIA----------SLLADAVRE----ARLHNACLLIRHFSLLADEKKTLNNLL 321
PVCLopT−15_(PAU02100) 219 GHAFEKILLNETDEIRPLVVLKGKQDSARELAVS-----NIMGIHGINTLTFDLFHLPDEECTT-------SIL----------NLLIDAIRE----TRLHNACLLIRNFSLLAEEKRISHREL 316
PVCLopT−15_(plu2398) 223 YHSLNKILLNEIDEICPLVILKGKQDSARELAVS-----NIMALHNISTLIVDLAHLPDEENTN-------TKP----------CLLTDAIRE----ARLHDACLLIRNFSLLAEEKQIWLSKL 320
PVCPlu1−15_(plu1716) 224 YHSFKKILLNDTNEIQPLIILRGKQDSARELAVS-----NIMASHGINTLTLDLALLPDDENPT-------EIT----------HLLADTIRE----TRLHNACLLIRNFSLLTEAKKILHRNL 321
PVCPlu2−15_(plu1693) 222 YPSLEKILLNKTDEIRPLVILRGKQDSARELAVS-----NMMALHDISTLTLDLAQLPEENNS----------NLLIDAVRE----ARLHDACLLIRNFSLLKDEKKILHSEF 315
PVCPlu3−15_(plu1675) 224 YHSLEKILLNETDKIRPLVILRGKEGSARELAVS-----NIMAYHGIHTLILDLACLPDEENSV-------KIS----------SLLADVIRE----TRLHDACLLIRKFSLLAAEKKILHHEL 321
PVCPlu4−15_(plu1653) 229 YHSFEKILLNQTDEIRPLVILRGKQDSARELAVS-----HIMASHNIRTLTLDLARLPDEDNIT-------VIT----------NLLTDAVRE----TRLYDGCLLIRNFSLFTEKKKILNNEL 326
PVCPau4−15_(PAU02803) 225 YHSFEKILLNETDEVRPLVILRGKPGSARELAVS-----HIMASHDISTLTLDLARLPDEDNIR-------AIA----------NLLTDAVRE----ARLHNACLLIRNFSLLAEKKKILNNEL 322
PVCPaTox15_(PAU02232) 235 TNQL--IELCQMEQEGMLTIIAGGAGSGSKTSVA-----RAASQVGRSVLLLSLASVTLSEHET-------I----------TLITLALRE----AQLRNACLMFEALDEFCEARPALQLWL 328
Afp15 227 KHALAGVLLTDDIEPRPVVLLRGMPGSARAHAVS-----SIFASENRQTLQVDIDKCTESDDDA-------LLL----------HLTCLMRE----ARMRGAGMVIRNLQARMEKSTSLPERL 323
ClpV/TssH_Pmoraviensis_UST58840.1 336 ARRFQPVQLHEPTVSEAVTILRGLAQVYEKSHGIYLRDDAVSAAELSARYLAGRQLPQKAVDVLDTACARVRISLAAAPESLERLRGELAEGGRQRQALRRDAEAGLLIDHEALEA---------L 453

PVCpnf−15_(PAU03339) 326 SELLNQPELRIVCLIEPYSPLVWLKKIPVLLIEMPLLTPAEKAR----LLIASLPDNCSEDIDTITLSQRYTFNPETLPLILQEAQLYQQQYDPLDILQQCDIRQ--------ALNLR--------- 431
PVCPacif−15_(PAU02007) 318 SILLSQPKLRVVCLVESGDALVWIKHLSMVQINMPVATLAEKEI----MLKRSLPSHSSDQINIAELCRRFSFTAETLPLIIKEACQYQILRQLEGQLEEIDLYR--------ALSFR--------- 423
PVCPlcif−15_(plu2517) 322 SILLHQPKLPVVCLVEPGDALVWIKHLSMVQIEMPVATRVDKES----MLKECLPDNAADKVNLAQLCQRFSFTAETLPLIIKEAYQYQILRQPDDQLEEIDLRK--------ALSFR--------- 427
PVCLopT−15_(PAU02100) 317 SALLNQPKLRVVCLAESEESLAWVKHLPIVQINMPPATLADKKT----MLEASLPDNVTKGINITQLCQRFSFTAETLPLIIKEAHQYQILRQPEDQLKESDLRK--------ALNCR--------- 422
PVCLopT−15_(plu2398) 321 STLLNQSKLRVVCLVEPGDSFVWIQHLPMVQIEMPVATLAEKKA----ILTENLPNNADREINLTQLCQRFSFTPETLPLILKEACQYQILRQPEDQLKEIDLRK--------ALNFR--------- 426
PVCPlu1−15_(plu1716) 322 SILLNQPKLRVVCLAEPGDILVWIKHQPMVQINMPIATLADKQT----LLKASLPENIAQKINITRLCQRFSFTAETLPLIVEEAHQYQMIRQPEGQLEETDLRK--------ALSFR--------- 427
PVCPlu2−15_(plu1693) 316 STLLNQAKLRVVCLAEPQDSLVWIKHLPIVQINMPALTLADKKA----MLEANLSDNIAQKINITQLCQRFSFTAETLPQILQEANQYQILRQPEGQLEETDLRK--------ALSFR--------- 421
PVCPlu3−15_(plu1675) 322 SILLNQPKLRVVCLVESGNGLAWLKHLPMVQIDMPVMTLAERKT----ILKANLPENVAQEINITQLCQRFSFTAETLPLILAEAHQYQIIRQPEDQIKEIDLRK--------ALSFR--------- 427
PVCPlu4−15_(plu1653) 327 STLLNQPKLRVVCLAEPQDSLIWIKHLPMVQIDMPAATLADKKT----MLEASLPDNIVQKINIPQLCQRFSFTAETLPLILKEAHQYQILRQPAGQLEETDLRK--------ALSFR--------- 432
PVCPau4−15_(PAU02803) 323 STLLNHPKLRVVCLIEPQDSLIWIEHLPMVQIDMPALTLADKKI----MLETNLPANIIQKINIAQLCQRFSFTAETLPQILQEANQYQILRQPEGQLEETDLRK--------ALNFR--------- 428
PVCPaTox15_(PAU02232) 329 GNRLARCSIPLFCQLPKQASLLPLDAISQVVLSMPMPSLMVKAAALASMMTNYFPDNS---LDVESLVTCFHPSPLILKKALSEAEIYRRLRQETASLRLDDVQM--------SLRFR--------- 435
Afp15 324 SEHLNQPGLRVVCLVEPYAPPLWLKKTPTLQTEMPVLTQEKTH----LLMGCLPERRAADIDLISLSQRYTFTPESLPLIVQEAELYRQQKDPADVLQQCDIRQ--------ALNLR--------- 429
ClpV/TssH_Pmoraviensis_UST58840.1 454 EARLDEAEEEMVAL-----ETLWTEQKELAERLLELRQQLAKAREAA----AVEPTVS-----VEE-----DAEGTVIETVVAEVEEGQSVEALEAQLHETHAALTAAQVKERLVSFEVCPRLVAEV 561

PVCpnf−15_(PAU03339) 432 ----AQQNFGQLAQRIIPKR-SLKDLLVSDEIAQQLREILIAIKYRE---QVLAGGFKDKIAYGTGISALFYGDSGTGKTMAAEVIADHIGV-D-LI-KVDL-----STVVNK-------YIGETEK 535
PVCPacif−15_(PAU02007) 424 ----AQQNFGKLAQRITPKR-HFNDLVISDELVQQLKEIIAAINYRE---QILSSGFKEKIGYGTGISVLFYGESGTGKTMAAEVIAGHLGV-D-LI-KVDL-----STVVNK-------YIGETEK 527
PVCPlcif−15_(plu2517) 428 ----AQQNLGKLARRITPKR-NFNDLVVSDELAQQLKEIIATINYRD---QILGAGFQEKVGYGTGISALFHGESGTGKTMAAEVIAGHLGV-D-LI-KVDL-----STVVNK-------YIGETEK 531
PVCLopT−15_(PAU02100) 423 ----AQQNFGKLAQRMTPKR-SFNDLVISADLTQQLKEIIAAINYRD---QILGAGFREKISYGTGISALFYGESGTGKTMAAEVIASYLGV-D-LI-KVDL-----STVVNK-------YIGETEK 526
PVCLopT−15_(plu2398) 427 ----AQQNFGKLAQRITPKR-SFNDLVVSEGLTQQLREIIATINYRN---QVLNTGFQEKIGYGTGISALFYGESGTGKTMAAEVIAGYLGV-D-LI-KVDL-----STVVNK-------YIGETEK 530
PVCPlu1−15_(plu1716) 428 ----AQQNFGKLAQRITPKR-SFSDLVVSDMLAQQLKEIVAAINYRD---QILSAGFQEKISYGTGISALFYGESGTGKTMAAEVIAGYLGV-D-LI-KVDL-----STVVNK-------YIGETEK 531
PVCPlu2−15_(plu1693) 422 ----AQQNFGKLAQRITPKR-SLNDLVISDVLTQQLKEIIAAIHYRD---QILATGFQEKIGYGTGISALFYGESGTGKTMAAEVIAGHLGV-D-LI-KVDL-----STVVNK-------YIGETEK 525
PVCPlu3−15_(plu1675) 428 ----AQQNFGKLAQRITPKL-SFNDLVVSETLAQQLKEIIAAIHYRD---QILSAGFQEKISYGTGISALFYGESGTGKTMAAEVIAGYLGV-D-LI-KVDL-----STVVNK-------YIGETEK 531
PVCPlu4−15_(plu1653) 433 ----AQQNFDKLAQRITPKR-RFNDLVISDALTQQLKEIIAAIHYRD---QILSAGFQEKIGYGTGISALFYGESGTGKTMAAEVIAGHLGV-D-LV-KVDL-----SSVVNK-------YIGETEK 536
PVCPau4−15_(PAU02803) 429 ----AQQNFGKLAQRITPKR-CFNDLVISDVLAQQLKEIIATIHYRD---QILGAGFQQKIGYGTGISALFYGESGTGKTMAAEVIAGHLGV-D-LI-KVDL-----STVVNK-------YIGETEK 532
PVCPaTox15_(PAU02232) 436 ----LQQNFGRLAQRITPQR-TFDDLIISESQQQQLQEILAAIRQRD---RMLEQGFARKVSYGTGISTLFFGESGTGKTMVAEVLAGVLGV-D-LI-KVDL-----STVVNK-------YIGETEK 539
Afp15 430 ----TQQNFGALAQRITPKR-TLNDLLVSESLLQQLQEMLTAIRYRE---KVLAGGFKDKVGYGVGISALFYGDSGTGKTMAAEVLAHTLGV-D-LI-KVDL-----STVINK-------YIGETEK 533
ClpV/TssH_Pmoraviensis_UST58840.1 562 ISAWTGVPLAQLARERHNAKVASFA-----TDLRTRIRGQEQAVHALDRSMRATAAGLNK-PDAPVG-VFLLVGPSGVGKTETALALADLLYGGDRFITTINMSEFQEKHTVSRLIGAPPGYVGYGEG 681



ATPase %ID	Pnf	PaCif	PlCif	PaLopT	PlLopT	Pl u1	Pl u2	Pl u3	Pl u4	Pa u4	PaTox	Afp15	ClpV/TssH
Pnf	100.00	56.14	54.80	56.04	56.48	57.12	57.62	56.10	54.55	55.73	42.75	65.66	21.23
PaCif	56.14	100.00	75.77	75.33	78.28	78.02	78.59	78.17	78.55	77.25	44.53	52.35	22.70
PlCif	54.80	75.77	100.00	75.77	78.70	78.58	78.13	79.16	78.67	79.68	44.56	52.19	23.55
PaLopT	56.04	75.33	75.77	100.00	78.92	78.80	77.58	78.07	77.50	75.91	43.35	51.97	21.93
PlLopT	56.48	78.28	78.70	78.92	100.00	77.78	79.62	79.51	77.78	77.42	43.93	53.14	22.76
Pl u1	57.12	78.02	78.58	78.80	77.78	100.00	81.10	83.86	82.42	80.06	44.61	53.35	23.05
Pl u2	57.62	78.59	78.13	77.58	79.62	81.10	100.00	79.51	84.59	83.84	44.66	53.68	22.78
Pl u3	56.10	78.17	79.16	78.07	79.51	83.86	79.51	100.00	80.55	79.19	44.16	52.62	24.38
Pl u4	54.55	78.55	78.67	77.50	77.78	82.42	84.59	80.55	100.00	88.78	43.68	54.55	23.19
Pa u4	55.73	77.25	79.68	75.91	77.42	80.06	83.84	79.19	88.78	100.00	44.11	55.73	23.84
PaTox	42.75	44.53	44.56	43.35	43.93	44.61	44.66	44.16	43.68	44.11	100.00	43.26	20.43
Afp15	65.66	52.35	52.19	51.97	53.14	53.35	53.68	52.62	54.55	55.73	43.26	100.00	20.63
ClpV/TssH	21.23	22.70	23.55	21.93	22.76	23.05	22.78	24.38	23.19	23.84	20.43	20.63	100.00

Figure 3.11 | Local sequence alignment of *Photorhabdus* Pvc15 ATPase homologues.

The *pvc15* gene is present in all of the documented PVC operons; Clustal Omega multiple sequence alignment (MSA) was used to compare sequence similarities and percent identities (%ID).

According to HHPred, the C-terminal region of Pvc15 is a putative AAA+ ATPase with greatest alignment to a Cdc48 homologue from *Mycobacterium smegmatis* (Msm0858; PDB: 5E7P) (Pvc15: amino acids 243-688; Msm0858: 262-725). Msm0858 is a magnesium-dependent ATPase with diverse functionality which includes the control of cell division, ubiquitin chain synthesis, and DNA repair (Unciuleac, Smith and Shuman, 2016). AlphaFold2 was used to elucidate the structure of Pvc15$_{pnf}$ which revealed that it, like Msm0858, Pvc15 contains tandem AAA domains; these are designated D1 (243-433) and D2 (434-688), respectively as is the case for Msm0858 (Figure 3.12; Ramachandran plot found in Appendices Figure 10.3). It is unclear why Cdc48 and its homologues possess two ATPase domains though D1 may stabilise the hexamer and be used for non-proteolytic functions whilst D2 is likely involved in translocation after association with the 20S proteasome to degrade polyubiquitylated proteins (Wang *et al.*, 2003).

In the work by Unciuleac, Smith, and Shuman (2016), Msm0858 was found to be monomeric in solution by zonal velocity sedimentation but lacked ATPase activity when the Arg finger was mutated, implying that the protein oligomerises when bound to ATP and/or substrate. Similarly, a homohexameric structure for Msm0858 could not be obtained although dimers were observed and facilitated when the N-domain was deleted. Intriguingly, both AAA domains appeared to possess ATPase activity, seemingly independently, such that only mutation of both P-loop lysines abolished ATPase activity completely.

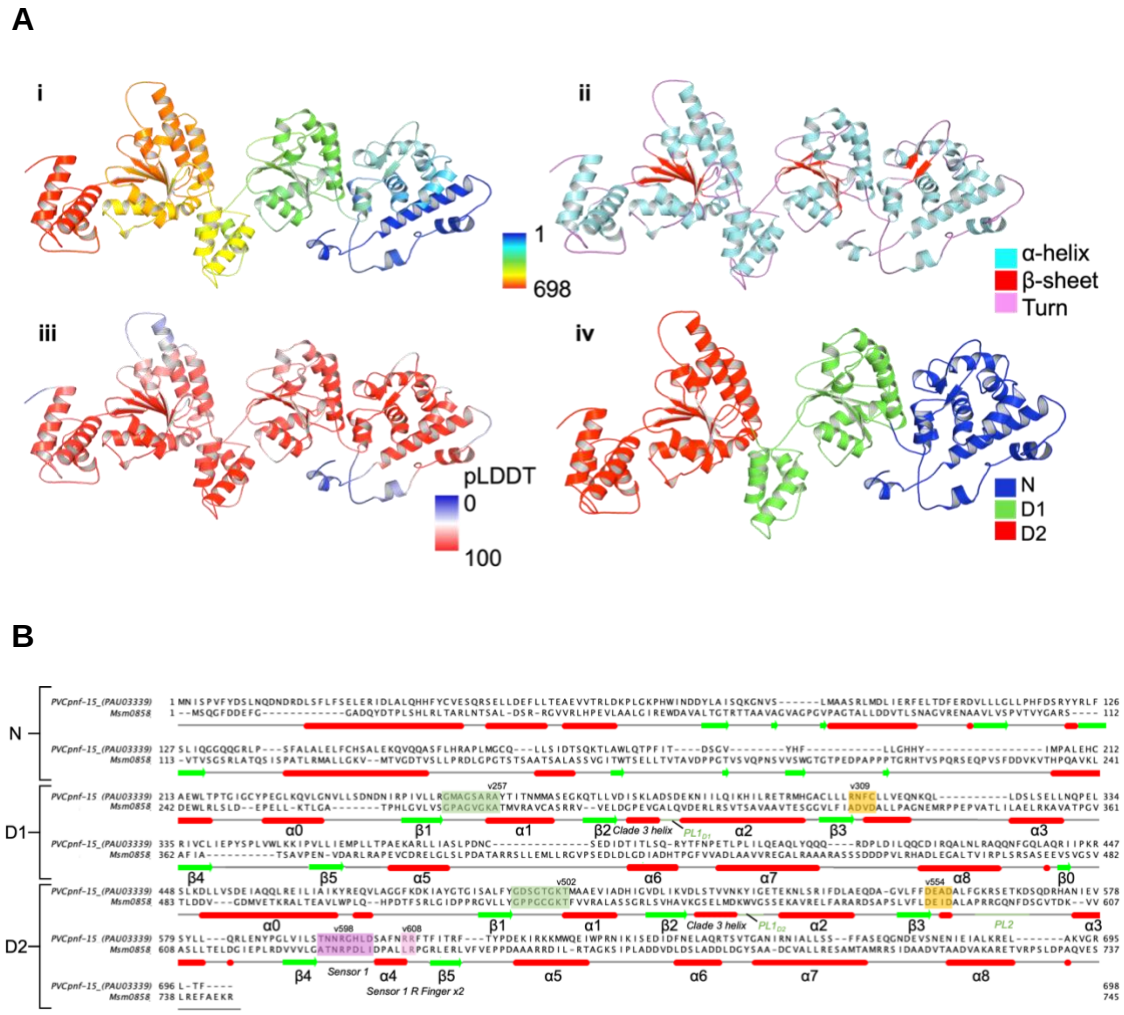


Figure 3.12 | Pvc15 is a Cdc48-like type II ATPase precluded by a single N-domain.

The AlphaFold prediction of Pvc15 as found in the PaPVC*pnf* operon (PAU03339) is shown by **(i)** residue, **(ii)** secondary structure, and **(iii)** inferred domains. AlphaFold's per-residue confidence score, pLDDT, is shown as a percentage in **(iv)** with higher scores being a lower error. Using MMSeqs2 multiple sequence alignment (MSA), two regions of Pvc15 were identified as an N-terminal (N) spanning residues 1-205 and two AAA+ domains: D1 (residues 206-434), and D2 (residues 435-698). **(B)** Local alignment with *M. smegmatis* Cdc48, Msm0858, indicates a generally conserved secondary structure despite differences in primary sequence. The Walker A and B motifs are coloured in green and orange, respectively. The Pvc15 D1 domain lacks 'β0' as well as α4 found in its D2 domain.

Whilst D2 appears to represent a classic AAA domain, D1 lacks any obvious characteristics in its Walker B motif including any aspartate or glutamate residues, encoding 'RNFC' instead of 'DEAD' (Figure 3.12E). A deviant motif is

also true of Msm0858 D1 which, like Pvc15, lacks a sensor 1 asparagine and arginine finger (Baek *et al.*, 2013; Unciuleac, Smith and Shuman, 2016). Whilst Msm0858 encodes two β-barrel modules in the N domain, the Pvc15 N domain appears to consist mostly of α-helices, and the Walker A motif in D1 appears even more deviant from the consensus 'GXGK(T/S)' than Msm0858 ('GVGKA') as Pvc15 D1 encodes an arginine in place of lysine: 'GSARA'. In addition, pore loop 1 (PL1) of D2 contains an aromatic residue, tyrosine, which have consistently been shown to be essential for substrate binding in the hexamer (Hinnerwisch *et al.*, 2005; Puchades, Sandate and Lander, 2020); D1, on the other hand, and similarly to its close homologue from *M. smegmatis*, does not contain this aromatic residue. Additionally, Pvc15 D1 superimposes well with D2 (3.83Å RMSD across 225 residue pairs) but lacks α4 and has a shorter α0 (Figure 3.13).

Altogether, it appears that D1 lacks obvious features sufficient for ATPase activity and, thus, it is likely that D1 is only used for stabilising oligomeric assembly of the Pvc15 complex.

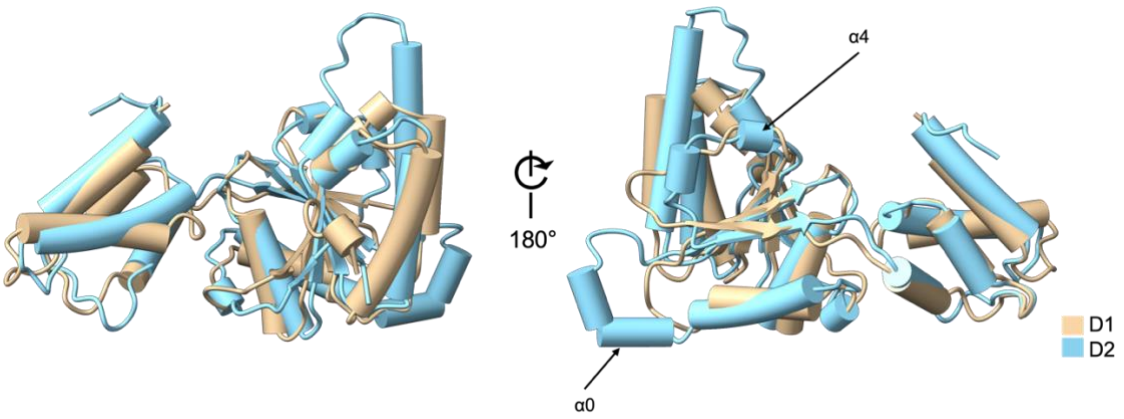
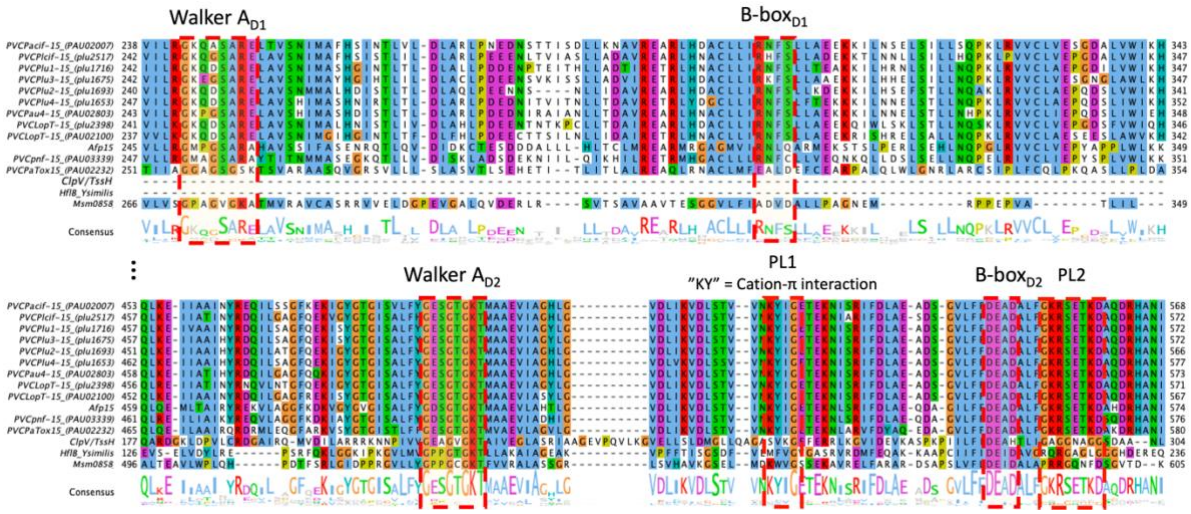


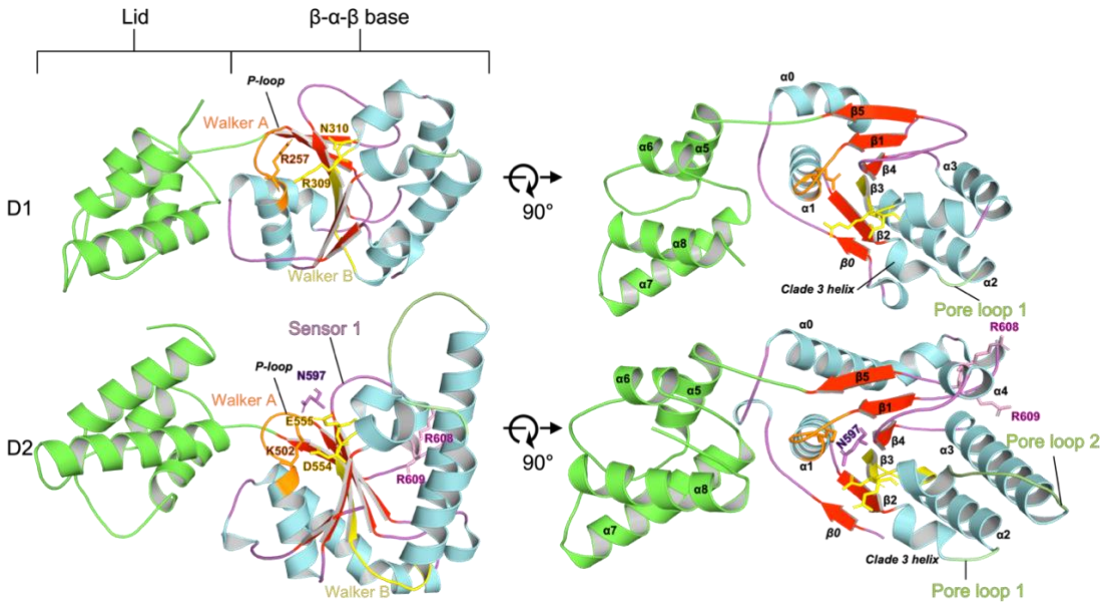
Figure 3.13 | Superimposition of Pvc15 D1 and D2 domains reveals tertiary similarities. AAA+ domains D1 and D2 from Pvc15$_{pnf}$ differ most significantly in their tertiary structures by the insertion of α4 into D2 as well as an extended α0. Both domains appears to encode the short β0 strand which extends the β-sandwich supersecondary structure. Superimposition was done using Matchmaker in ChimeraX-1.5; RMSD from 88 Cα pairs was 1.30Å and 3.83Å across all 225 pairs.

A closer look at the structure of the Pvc15 homologues indicates that the conserved Walker A and B motifs are, indeed, present in D2 of each of the homologues (Figure 3.14A). D1, on the other hand, has a different consensus in both motifs: the Walker A 'GKT' is replaced with 'ARA' and the Walker B 'DEAD' motif is replaced with 'RNFC'. The AAA+ domain is made up of a 'base' and a 'lid'; at the centre of the ATPase domain is a classical β-α-β 'sandwich' or Rossmannoid fold 'base': at least 5 parallel β-sheets are found sandwiched between at least two α-helices on either side (Hanukoglu, 2015) and is of the classical order: β5-β1-β4-β3-β2 whilst the C-terminal 'lid' domain consists of 4 α-helices (α5-α8). Together, these two domains make up the minimal features of a AAA+ domain; only one crucial feature separates Pvc15 from archetypal clade 1 ATPases (i.e. DNA clamp loaders), such as replication factor C (RFC1; PDB: 2CHG) (Seybert *et al.*, 2006): a very short α-helix found inserted between β2 and α2 is characteristic of clade 3 AAA ATPases: the 'classic' clade to which FtsH and Cdc48 belong (Miller and Enemark, 2016).

A


Walker A$_{D1}$
B-box$_{D1}$
Walker A$_{D2}$
PL1
"KY" = Cation-π interaction
B-box$_{D2}$
PL2
Consensus

B


Lid
β-α-β base
D1
D2
P-loop
Walker A
Walker B
N310
R257
R309
Sensor 1
N597
E555
K502
D554
R608
R609
90°
α0
α1
α2
α3
α4
α5
α6
α7
α8
β0
β1
β2
β3
β4
β5
Clade 3 helix
Pore loop 1
Pore loop 2

C

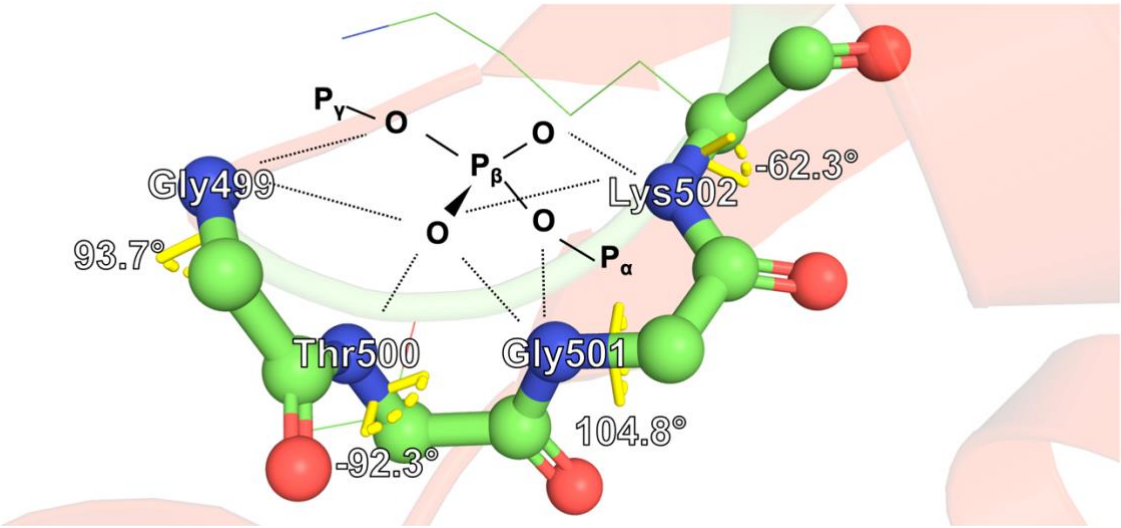


Figure 3.14 | Pvc15 ATPase motifs are highly conserved in D2.

(A) Walker A and B motifs are highly conserved ATP and GTP-binding sites and are found in the C-terminal region of Pvc15 homologues (with the exception of $Pvc15_{PaTox}$) and including Afp and ClpV/TssH. Also annotated are the secondary structure predictions from JPred and their confidence scores (JNETCONF) as well as burial scores where lower scores indicate more exposed residues. **(B)** The α/β sandwich fold found in the C-terminal region of Pvc15 with labelled Walker A and B motifs; this fold is conserved in ATPases including Ras. Notable motifs include the Walker A and B motifs which are highly conserved sequences of amino acids in AAA+ ATPases; these include the 'nest' residues Gly499-Lys502 in the Walker A motif which likely coordinate ATP, and Asp554-Glu555 in the Walker B motif which likely coordinate a divalent cation and catalyse ATP hydrolysis. Other notable residues exist in the second region of homology (SRH) which have putative functions in facilitating ATPase activity and stabilising conformational changes in the structure during ATP binding and hydrolysis: the sensor 1 residue Asn597, the sensor 2 residue Arg659, and the more distant Arg-finger motif which coordinates neighbouring protomers in the quaternary structure which may be either Arg608 or Arg609. **(C)** The Walker A motif "LRLR" nest with proposed coordination of phosphate as documented for the P-loop of Ras by Watson and Milner-White (2002). Phi bond angles are also shown; carbon atoms are shown in green, nitrogen in blue, and oxygen in red; Amber-relaxed side chains are depicted as thin sticks.

The ATPase domain supersecondary structure itself is also typical of an AAA+ ATPase (Figure 3.14B). The Walker A motif exists as a glycine-rich loop

situated between β1 and α1 of the super-secondary structure. In D2, the "GTGK" motif itself forms a compound nest of the form "LRLR", given by the alternating signs of the phi angles, which likely coordinates the β-phosphate of ATP or ADP within the concave of the NH bonds, as in the case with Ras (Figure 3.14C) (Bugreev and Mazin, 2004). Lys502 is likely responsible for nucleotide binding; a role which can be abrogated by mutating to alanine in homologous structures (Hanson and Whiteheart, 2005).

Walker B motifs tend to be less well conserved but often feature a set of up to 4 hydrophobic amino acids followed by a 'DE'; in Pvc15 these residues are part of a 'DEAD' motif which is often found in proteins with helicase activity; the relevance of this to Pvc15, however, remains to be ascertained (Watson and Milner-White, 2002; Cordin *et al.*, 2004). The aspartate residue (D554) may coordinate a divalent cation such as magnesium (II), whilst the glutamate residue (E555) is likely responsible for ATP hydrolysis; mutation of the latter is known to reduce ATPase activity in various ATPases such as the Lon protease and bacterial SecA (Kim *et al.*, 2013; Shin *et al.*, 2020). On the other hand, Jiang and colleagues (2022) also identified the Walker A and B motifs of Pvc15$_{pnf}$ but apparently failed to abolish its unfolding activity upon leader-associated fluorescent proteins by point mutations; notably, however, they did not specify which point mutations were tested. Walker B motifs are often situated in the β3-α3 loop (Hanson and Whiteheart, 2005); with Pvc15, however, the Walker B motif appears to begin within the α2-β3 loop.

Additionally, the second region of homology (SRH) is a defining characteristic of AAA+ ATPases (Figure 3.14B; right). The sensor 1 motif is located between residues 595-599 in the loop between β4 and α4; Asn597, in particular, is likely responsible for orientation of the nucleophilic water used in hydrolysis by the Walker B motif and is ideally positioned directly between the Walker A P-loop residues and the Walker B motif (best illustrated in Figure 3.14B, right). Finally, two R-fingers (R608 and R609) are found, as expected, in

the α4 helix which mediate ATP interactions in the adjacent protomer in the oligomeric complex.

Overall, Pvc15 is a clade 3 AAA ATPase due to the presence of the short α-helix before α2, lack of sensor 2, and presence of an additional sensor 1 R-finger. The Pvc15 AAA domain has one particular characteristic that makes it stand out amongst its clade: though it contains a 'Box II' sequence (α0), Pvc15 lacks a Box VII which would normally appear after α4, as is the case for DnaA, for example (Erzberger and Berger, 2006; Snider, Thibault and Houry, 2008); this sequence is often involved in inter-subunit communication in the multimer. Additionally, the AAA domain appears to be encoded C-terminal of a short additional β-strand which inserts antiparallel to β2; the significance of this uniquely appended strand – designated here as 'β0' – to the main β-sheet of the sandwich 'base' is currently unknown.

Hitherto, the Pvc15 used as a model is the one encoded at the PVC*pnf* operon (Pvc15$_{pnf}$) though Pvc15 homologues are one of the most well conserved ORFs across eCIS operons. To get an idea of how similar the overall three-dimensional structures of Pvc15 homologues are, each were separately predicted using AlphaFold2 and superimposed using the MatchMaker suite in ChimeraX with Pvc15$_{pnf}$ as a reference (Figure 3.15). As expected, all homologues overlaid very well with one another with an average number of atom pairs to the reference of 396 with a mean RMSD of 0.989Å across all homologues. Therefore, the supersecondary nature of Pvc15$_{pnf}$ discussed thus far also applies well to the other homologues.

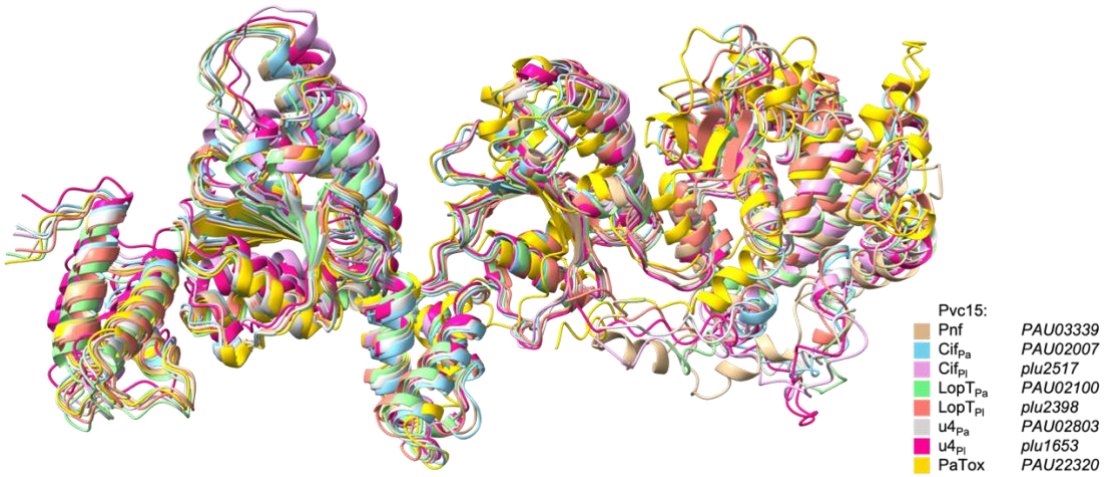


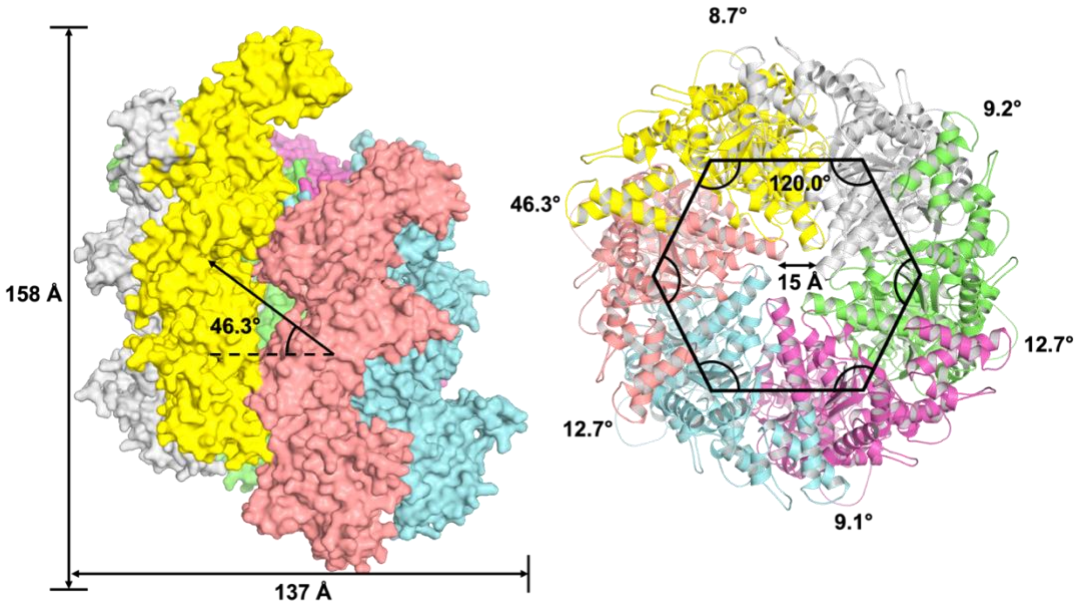
Figure 3.15 | Pvc15 homologues have vast structural similarities.
Amino acid sequences for each of the Pvc15 units in *Photorhabdus asymbiotica* (Pa) or *P. luminescens* (Pl) were predicted using AlphaFold. The highest-scoring model was then overlaid with homologues using Matchmaker within the ChimeraX suite; the PVC*pnf* Pvc15 (PAU03339) was used as a reference.

3.4.3.1 Inferring Hexameric Assemblies of Pvc15

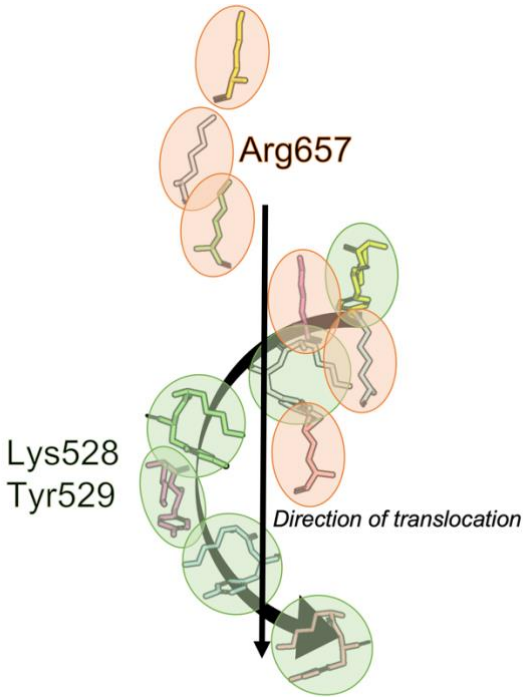
Since most ATPases appear to form oligomeric rings – hexamers being the most common – it was tested how likely such an assembly would be for Pvc15 in the absence of experimental data. MoLPC (Modelling of Large Protein Complexes) was used via a Jupyter Notebook script run on a high RAM GPU (>40GiB), according to the instructions from the authors and designating a stoichiometry of Pvc15 subcomponents of 6 (Bryant *et al.*, 2022). The hexameric structure resembles a classical right-handed open spiral staircase and the multiple-interface predicted DockQ (mpDockQ) value was 0.23 which, though seemingly low, is classified as 'acceptable' (≥0.23) (Basu and Wallner, 2016; Johansson-Åkhe and Wallner, 2022). The hexamer has a pLDDT score of interfacing residues (pLDDT$_{IF}$) of 61 and an overall pLDDT of roughly 74. The authors note, however, that the program has difficulty in defining asymmetrical structures; the asymmetry of the Pvc15 hexamer may reflect the altered

conformations that subcomponents can assume depending on whether they are bound to ATP, ADP + Pi, or ADP alone as well as substrate: inputs which are not possible for MoLPC to account for. For some AAA ATPases, the right-handed spiral staircase conformation is assumed until binding ATP and substrate which likely causes a reassembly to a more symmetrical closed spiral staircase conformation, such as the Lon protease which switches from left-handed open to right-handed closed (Shin *et al.*, 2020).

A


158 Å
46.3°
137 Å
8.7°
9.2°
46.3°
120.0°
15 Å
12.7°
12.7°
9.1°

B


Arg657
Lys528
Tyr529
Direction of translocation

C

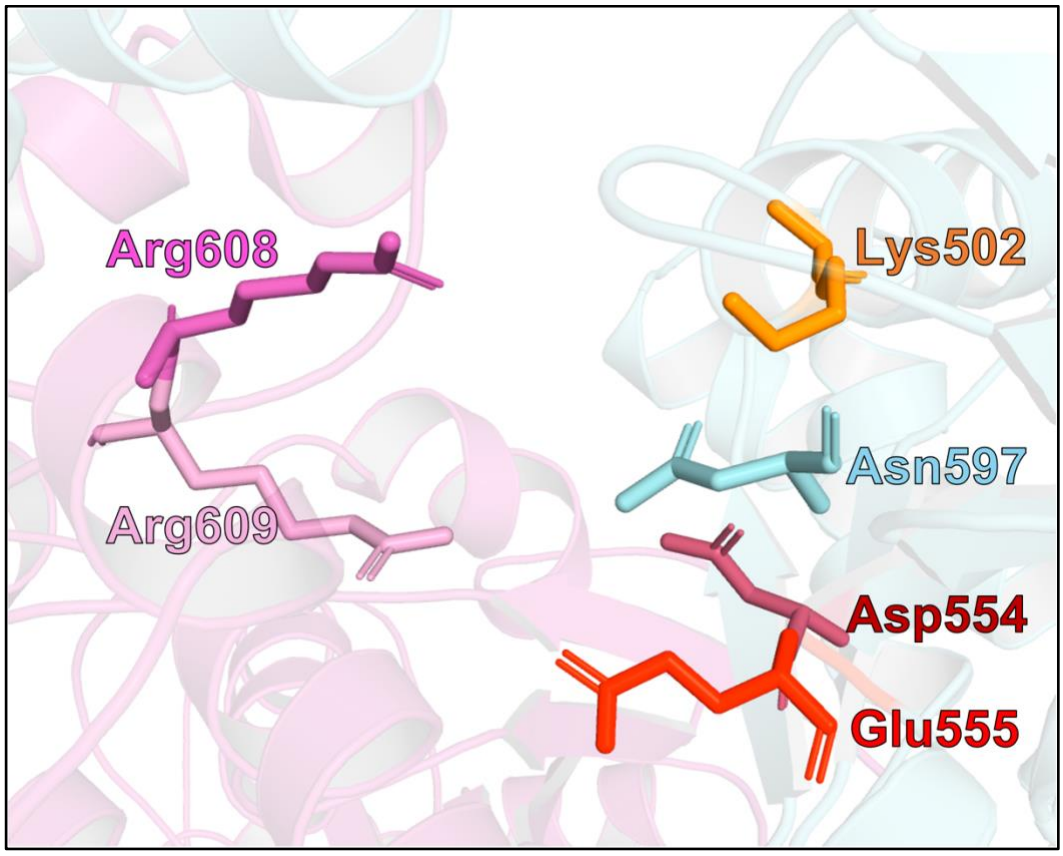


Figure 3.16 | A complete prediction of the Pvc15 hexamer.

The hexameric Pvc15 structure was predicted with high confidence by MoLPC with a predicted local distance difference test (pLDDT) at interfacing residues of 61, indicated by the mpDockQ score of 0.233. **(A)** Pvc15x6 assumes an open, right-handed spiral staircase conformation. The C-terminal D2 domain (top of the left-hand structure) forms a roughly 15Å-diameter pore, similarly to other AAA ATPases. **(B)** Pore loops 1 (green) and 2 (orange) can interface with substrate from different sides; each subunit sequentially provides an interaction that may translocate substrate polypeptide upon ATP hydrolysis. In other classical ATPases, interfaces 1 and 2 are more intimately adjacent one another only upon binding of both a single substrate and ATP in up to 4 of the 6 subunits simultaneously. **(C)** Arginine fingers (R-fingers) R608 and R609 are situated *in trans* 5-10Å of the predicted ATP binding site of its adjacent subunit. These residues are expected to coordinate the γ-phosphate of ATP to communicate its hydrolysis into conformational changes around the hexameric ring.

PL1 and PL2 spiral through the hexameric structure, matching the spiralling domains, and bind the unfolded β-strand-like substrate via non-residue-specific

hydrogen bonding of the amino acid backbone in the central pore. Meanwhile, the conserved aromatic amino acid in PL1, in Pvc15: Tyr529, coordinates a tight grip on the substrate every 2 amino acids facilitated by the cation-π interaction with its neighbouring cationic Lys528 side chain. This arrangement is similar to that of Vps4 (PDB: 6BMF) though the environment surrounding the aromatic residue can be specific for the ATPase's intended function; the grip would be looser when neighbouring glycine (e.g. ClpB D2 (PDB: 6OAY)) or made more hydrophobic when neighbouring aliphatic residues such as valine (e.g. YME1 (PDB: 6AZ0)) (Puchades, Sandate and Lander, 2020). Finally, the arginine fingers (R-fingers) Arg608 and Arg609 are found within 5-10Å of the predicted ATP binding site of its adjacent subunit (Figure 3.16C); this is an abundant feature in many ATPases in all 7 AAA+ clades (Jessop, Felix and Gutsche, 2021).

Following from this analysis, the Pvc15$_{pnf}$ N-domain was assessed by homologous atomic coordinates using the DALI server (Figure 3.17) (Holm, 2022). It was found that the N-domain has highest homology hits for DNA-directed transcription factor subunits such as RNA polymerase II from *S. cerevisiae* with a Z-score of 7.5 at 2.9Å R.M.S.D (PDB: 5OQJ) and the repressor of the MepA efflux pump, MepR, with a Z-score of 6.8 and 3.5Å R.M.S.D over 90 aligned residues (PDB: 4LD5) (Holm and Rosenström, 2010; Birukou *et al.*, 2013; Schilbach *et al.*, 2017; Holm, 2022). Together with the potential relevance of the Walker B 'DEAD' box helicase motif, it is not unreasonable to hypothesise some nucleic acid-directed activity of Pvc15 mediated by the N-domain; effects on transcription could be investigated from PVC-producing cell lysates using reverse transcriptase quantitative PCR (RT-qPCR) with and without Pvc15, for example.

Job: Pvc15_N_domain
Query: s001A

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	5oqj-2	7.5	2.9	82	395	12		MOLECULE: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB1;
2:	7o4j-2	7.2	2.7	81	445	9		MOLECULE: GENERAL TRANSCRIPTION AND DNA REPAIR FACTOR IIH H
3:	7o4i-2	7.2	2.8	82	445	9		MOLECULE: GENERAL TRANSCRIPTION AND DNA REPAIR FACTOR IIH H
4:	7ml3-2	7.2	3.4	84	460	10		MOLECULE: BJ4_G0050160.MRNA.1.CDS.1;
5:	7ml2-2	6.9	3.4	83	460	10		MOLECULE: DNA-DIRECTED RNA POLYMERASE SUBUNIT;
6:	7o4l-2	6.9	2.7	81	445	9		MOLECULE: GENERAL TRANSCRIPTION AND DNA REPAIR FACTOR IIH H
7:	4ld5-G	6.8	3.5	93	139	8		MOLECULE: MEPR;
8:	7o4k-2	6.7	2.7	81	445	9		MOLECULE: GENERAL TRANSCRIPTION AND DNA REPAIR FACTOR IIH H
9:	7m2u-2	6.7	2.8	80	460	9		MOLECULE: DNA REPAIR HELICASE RAD25;
10:	7k04-2	6.7	2.8	80	460	9		MOLECULE: DNA REPAIR PROTEIN RAD33;
11:	3bdd-A	6.6	6.4	84	140	6		MOLECULE: REGULATORY PROTEIN MARR;
12:	7o72-2	6.6	2.8	82	445	9		MOLECULE: GENERAL TRANSCRIPTION AND DNA REPAIR FACTOR IIH H
13:	5jlu-A	6.6	6.1	81	137	16		MOLECULE: ADHESIN COMPETENCE REPRESSOR;
14:	6nmi-D	6.5	3.0	82	452	12		MOLECULE: GENERAL TRANSCRIPTION AND DNA REPAIR FACTOR IIH H
15:	6cmv-E	6.5	3.0	82	111	12		MOLECULE: TRANSCRIPTIONAL REGULATOR LRS14-LIKE PROTEIN;
16:	6cmv-D	6.5	3.0	83	111	13		MOLECULE: TRANSCRIPTIONAL REGULATOR LRS14-LIKE PROTEIN;
17:	6cmv-F	6.5	3.0	82	111	13		MOLECULE: TRANSCRIPTIONAL REGULATOR LRS14-LIKE PROTEIN;
18:	6k4y-M	6.5	2.3	75	209	9		MOLECULE: DNA-DIRECTED RNA POLYMERASE SUBUNIT ALPHA;
19:	4ld5-C	6.5	3.3	89	139	6		MOLECULE: MEPR;
20:	7o75-2	6.4	2.7	81	445	9		MOLECULE: GENERAL TRANSCRIPTION AND DNA REPAIR FACTOR IIH H
21:	3bdd-C	6.4	6.1	85	140	6		MOLECULE: REGULATORY PROTEIN MARR;
22:	7k01-2	6.4	2.8	81	460	10		MOLECULE: GENERAL TRANSCRIPTION AND DNA REPAIR FACTOR IIH S
23:	4eju-B	6.4	4.1	94	147	11		MOLECULE: TRANSCRIPTIONAL REGULATOR TCAR;
24:	2xig-B	6.4	2.6	77	148	5		MOLECULE: FERRIC UPTAKE REGULATION PROTEIN;
25:	5yi0-D	6.4	5.9	83	146	13		MOLECULE: ZINC TRANSPORT TRANSCRIPTIONAL REGULATOR;
26:	5eri-A	6.3	5.4	84	153	8		MOLECULE: MARR FAMILY TRANSCRIPTIONAL REGULATOR;
27:	6ro4-C	6.3	2.8	75	386	13		MOLECULE: DNA1;
28:	3bdd-B	6.3	6.6	81	133	6		MOLECULE: REGULATORY PROTEIN MARR;
29:	7nvz-2	6.3	3.0	76	390	13		MOLECULE: TFIIH BASAL TRANSCRIPTION FACTOR COMPLEX HELICASE
30:	7ml1-2	6.3	2.8	82	460	10		MOLECULE: TFB1;
31:	2d1h-B	6.3	2.8	81	99	15		MOLECULE: 109AA LONG HYPOTHETICAL TRANSCRIPTIONAL REGULATOR
32:	7jpq-D	6.3	6.9	102	403	12		MOLECULE: ORIGIN RECOGNITION COMPLEX SUBUNIT 2;
33:	2xig-D	6.3	2.8	78	148	5		MOLECULE: FERRIC UPTAKE REGULATION PROTEIN;
34:	6j05-B	6.3	3.1	78	90	9		MOLECULE: TRANSCRIPTIONAL REGULATOR ARSR;
35:	6cmv-A	6.3	3.3	84	112	12		MOLECULE: TRANSCRIPTIONAL REGULATOR LRS14-LIKE PROTEIN;
36:	4ld5-A	6.3	3.7	91	138	5		MOLECULE: MEPR;
37:	7nvr-2	6.2	3.0	76	390	13		MOLECULE: TFIIH BASAL TRANSCRIPTION FACTOR COMPLEX HELICASE
38:	3bdd-D	6.2	6.4	79	132	6		MOLECULE: REGULATORY PROTEIN MARR;
39:	7nvy-2	6.2	3.0	76	390	13		MOLECULE: TFIIH BASAL TRANSCRIPTION FACTOR COMPLEX HELICASE
40:	3lmm-C	6.2	3.8	88	477	9		MOLECULE: UNCHARACTERIZED PROTEIN;
41:	3bpv-A	6.2	5.0	85	137	9		MOLECULE: TRANSCRIPTIONAL REGULATOR;
42:	2xig-A	6.2	2.4	75	146	5		MOLECULE: FERRIC UPTAKE REGULATION PROTEIN;
43:	6cmv-C	6.2	3.1	83	112	12		MOLECULE: TRANSCRIPTIONAL REGULATOR LRS14-LIKE PROTEIN;
44:	2fxa-C	6.2	4.9	87	160	13		MOLECULE: PROTEASE PRODUCTION REGULATORY PROTEIN HPR;
45:	5yi0-A	6.2	5.9	83	146	13		MOLECULE: ZINC TRANSPORT TRANSCRIPTIONAL REGULATOR;
46:	2qww-E	6.2	3.9	78	143	10		MOLECULE: TRANSCRIPTIONAL REGULATOR, MARR FAMILY;
47:	2hzt-C	6.2	2.2	72	95	6		MOLECULE: PUTATIVE HTH-TYPE TRANSCRIPTIONAL REGULATOR YTCD;
48:	4ld5-F	6.2	3.8	93	139	6		MOLECULE: MEPR;
49:	4ld5-H	6.2	3.4	88	139	8		MOLECULE: MEPR;
50:	7jgr-D	6.1	5.6	99	441	13		MOLECULE: ORIGIN RECOGNITION COMPLEX SUBUNIT 2;

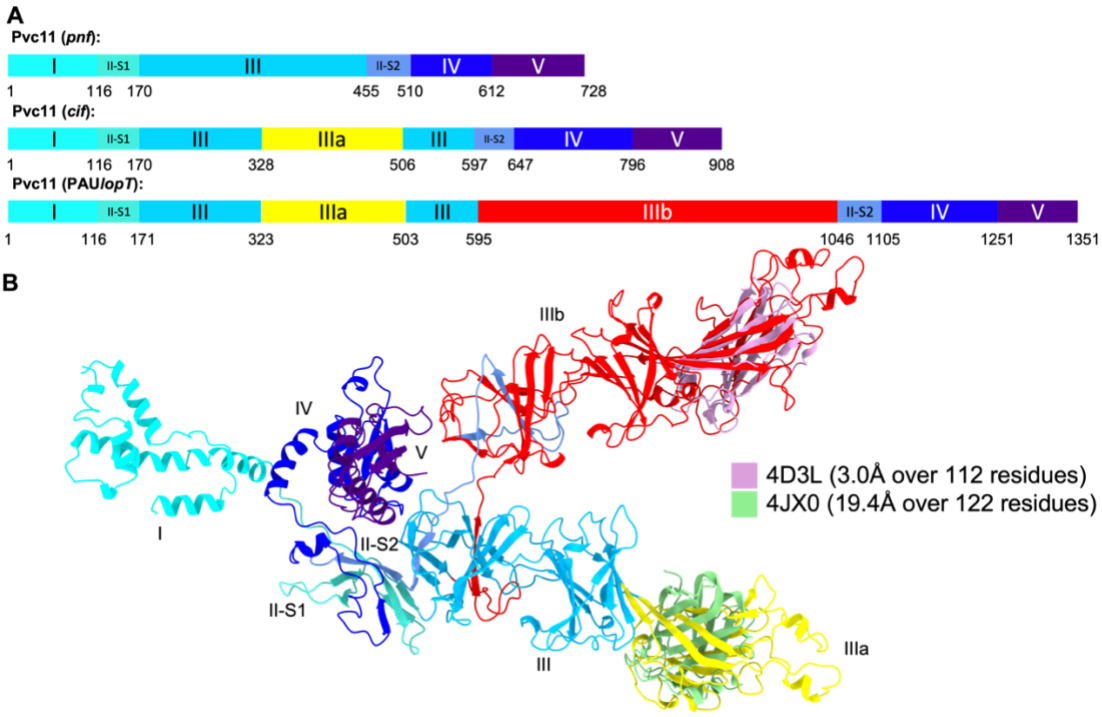
Figure 3.17 | DALI server results suggests the Pvc15 N-domain may bind nucleic acids.

The DALI server, used for the homology search of atomic coordinates in a protein structure file, has many significant hits for DNA-binding proteins, primarily transcriptional regulators.

3.4.4 Inferring the Binding Capacity of the PVC*lopT* Chassis

Though this subchapter is somewhat detached from the overall narrative of the role of Pvc15 and payload LSs in loading into the mature PVC, it is worthwhile reporting my brief findings here on building a hypothesis to describe the potential for the Pvc11$_{lopT}$ baseplate to substitute the role of Pvc13 tail fibres for glycan binding.

Both *P. asymbiotica* and *P. luminescens* Pvc11$_{lopT}$ were found to have a long stretch of amino acids which did not align to the rest of the Pvc11 homologues, spanning residues ~595-1046 and situated after residue 412 of the native Pvc11$_{pnf}$; neither HHPred nor pBLAST searches could allude to the function of these amino acids (Figure 3.18A). Also of note, residues 323 to 503 are found in many other Pvc11 homologues including Pvc11$_{cif}$ and unit 4, but not Pvc11$_{pnf}$ which is comparably shorter. These as-yet-unreported domains are denoted, in this work, as IIIa and IIIb which are separated by the remaining aligning region of domain III.



C

# No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	6rao-I	17.7	4.4	368	443	39		MOLECULE: AFP1;
2:	4unm-B	8.7	9.2	130	609	8		MOLECULE: SECRETED PROTEIN;
3:	5hx2-D	8.0	10.1	187	647	13		MOLECULE: BASEPLATE WEDGE PROTEIN GP7;
4:	1us2-A	6.9	3.7	129	507	12		MOLECULE: ENDO-BETA-1,4-XYLANASE;
5:	4jx0-A	6.9	19.4	122	302	10		MOLECULE: HYPOTHETICAL PROTEIN;
6:	6f2p-A	6.8	26.4	183	1040	10		MOLECULE: PAENIBACILLUS XANTHAN LYASE;
7:	7cik-A	6.6	21.4	96	156	8		MOLECULE: FLAGELLAR M-RING PROTEIN;
8:	7p82-C	6.6	6.8	111	408	10		MOLECULE: S-ADENOSYLMETHIONINE SYNTHASE;
9:	6s04-B	6.6	13.5	98	497	8		MOLECULE: EXOSIALIDASE FROM UNCULTURED BACTERIUM PG7;
10:	4d3l-A	6.6	3.0	112	132	6		MOLECULE: CARBOHYDRATE BINDING MODULE;
11:	5euv-A	6.5	20.6	142	731	2		MOLECULE: BETA-D-GALACTOSIDASE;
12:	5ces-A	6.5	3.4	81	91	11		MOLECULE: PA0618;
13:	7bxo-E	6.5	3.8	88	125	10		MOLECULE: TOXIN-ANTITOXIN SYSTEM ANTIDOTE MNT FAMILY;
14:	2d1r-A	6.4	5.9	102	539	8		MOLECULE: LUCIFERIN 4-MONOOXYGENASE;
15:	3gku-A	6.3	23.5	95	206	4		MOLECULE: PROBABLE RNA-BINDING PROTEIN;
16:	3lax-A	6.3	3.2	76	106	18		MOLECULE: PHENYLACETATE-COENZYME A LIGASE;
17:	4qaw-A	6.2	10.6	159	537	10		MOLECULE: XYN30D;
18:	6ds5-A	6.0	3.1	116	160	9		MOLECULE: SEIPIN;
19:	7qe7-L	6.0	4.7	133	183	11		MOLECULE: ANAPHASE-PROMOTING COMPLEX SUBUNIT 10;
20:	3wnk-A	6.0	12.7	167	712	7		MOLECULE: CYCLOISOMALTOOLIGOSACCHARIDE GLUCANOTRANSFERASE;
21:	4uoj-A	6.0	20.9	164	917	4		MOLECULE: BETA-MANNOSIDASE GH2;
22:	7bk0-a	6.0	11.6	84	250	8		MOLECULE: FLAGELLAR M-RING PROTEIN;
23:	4a42-A	6.0	3.4	105	127	9		MOLECULE: ALPHA-N-ACETYLGLUCOSAMINIDASE FAMILY PROTEIN;
24:	6ijb-B	5.9	7.9	120	538	9		MOLECULE: AMP-BINDING DOMAIN PROTEIN;
25:	6qfj-A	5.8	2.7	69	84	12		MOLECULE: MAGNETOSOME MEMBRANE PROTEIN MAMB, PUTATIVE CO/ZN
26:	4xup-A	5.7	6.9	151	330	10		MOLECULE: ENDO-1,4-BETA-XYLANASE C;
27:	5m7h-A	5.7	23.5	116	420	6		MOLECULE: GTPASE DER;
28:	6q2f-A	5.7	32.0	185	1055	9		MOLECULE: GLYCOSIDE HYDROLASE FAMILY PROTEIN;
29:	5gl6-A	5.6	3.6	91	169	10		MOLECULE: RIBOSOME MATURATION FACTOR RIMP;
30:	5zru-A	5.6	16.5	157	576	11		MOLECULE: ALPHA-1,3-GLUCANASE;
31:	5tcp-0	5.6	17.4	96	184	9		MOLECULE: LIPOPROTEIN PRGK;
32:	2wz8-A	5.6	3.6	108	135	6		MOLECULE: CELLULOSOME PROTEIN DOCKERIN TYPE I;
33:	6a48-A	5.5	7.4	141	651	1		MOLECULE: REELIN;
34:	4qhw-A	5.5	20.1	130	400	8		MOLECULE: UNCHARACTERIZED PROTEIN;
35:	6ulw-B	5.4	6.0	103	1209	8		MOLECULE: AMINO ACID ADENYLATION DOMAIN-CONTAINING PROTEIN;
36:	4ni3-B	5.4	14.7	143	583	10		MOLECULE: ALPHA-FUCOSIDASE GH29;
37:	6vda-A	5.4	2.7	69	86	10		MOLECULE: ZT_DIMER DOMAIN-CONTAINING PROTEIN;
38:	7t7y-A	5.4	3.5	116	154	10		MOLECULE: ENDOGLUCANASE;
39:	5ho1-A	5.4	2.5	64	86	6		MOLECULE: MAGNETOSOME PROTEIN MAMB;
40:	4ycg-B	5.3	4.9	137	198	9		MOLECULE: BONE MORPHOGENETIC PROTEIN 9 GROWTH FACTOR DOMAIN
41:	4lo4-B	5.3	19.9	98	421	9		MOLECULE: HA-70;
42:	3cqo-A	5.3	7.5	158	293	2		MOLECULE: FBP32;
43:	5mmj-c	5.3	14.5	92	216	7		MOLECULE: 50S RIBOSOMAL PROTEIN L31;
44:	7toh-A	5.2	8.8	120	379	6		MOLECULE: SGNH HYDROLASE;
45:	2vzs-A	5.2	22.0	164	857	9		MOLECULE: EXO-BETA-D-GLUCOSAMINIDASE;
46:	2mcq-A	5.2	3.0	68	77	19		MOLECULE: UNCHARACTERIZED PROTEIN RP812;
47:	1o59-A	5.2	10.9	125	318	4		MOLECULE: ALLANTOICASE;
48:	1pnf-A	5.2	3.3	115	314	10		MOLECULE: PEPTIDE-N(4)-(N-ACETYL-BETA-D-GLUCOSAMINYL)ASPARA
49:	6ed1-A	5.2	16.6	156	860	5		MOLECULE: GLYCOSYL HYDROLASE FAMILY 2, SUGAR BINDING DOMAIN
50:	1no5-B	5.2	3.1	79	102	13		MOLECULE: HYPOTHETICAL PROTEIN HI0073;

Figure 3.18 | DALI server results suggests Pvc11$_{PAUlopT}$ contains a carbohydrate binding module within the extended domain III.

Pvc11 from *P. asymbiotica* ATCC43949 was predicted using AlphaFold2. **(A)** Pvc11$_{PAU\mathit{lopT}}$ is much larger in size than its other analogues: two additional regions, denoted as domain IIIa and, the larger, IIIb. These domains were assigned according to their alignment with Pvc11$_{\mathit{pnf}}$. **(B)** Superimposition of Pvc11-LopT with other resolved PDB structures; RMSD values are shown in the legend. Domain V is shown in the perpendicular plane facing towards the reader. In three-dimensional space, domain IIIa superimposes well with chain A of 4JX0: a protein domain of unknown function but which was used to predict the carbohydrate binding module within the spike cage of Alg11 (Xu *et al.*, 2022). Additionally, xylanase 10C (PDB: 1US2), an endo-β-1,4-xylanase aligns very well with this domain with a Z-score of 6.9 (3.7Å over 129 residues). Another carbohydrate binding module, a single chain of CBM79-1$_{RfGH9}$ (PDB: 4D3L), was shown to align well with domain IIIb (Z-score = 6.6): a larger protrusion from the base structure. **(C)** The DALI search alignments of Pvc11$_{\mathit{lopT}}$ represented the top 10 tertiary protein alignments using the PDB90 database with Z-scores consistently above 6.5 ($p < 0.0001$). Many other proteins also involved in carbohydrate binding were identified as being significant. In contrast, Pvc11$_{\mathit{pnf}}$ did not have any 'carbohydrate binding family' alignments within the first 50 alignments (not shown).

AlphaFold was used to generate an Amber relaxed .pdb file of PaPVC*lopT* Pvc11 (pLDDT = 82.95%) which indicated two additional extensions from the structure of Pvc11$_{\mathit{pnf}}$ (Figure 3.18B; blue colours). One of these is a smaller extension of the already-present 'arm' which is found in most Pvc11 homologues other than PVC*pnf* and is denoted as domain IIIa (yellow). Additionally, a much larger extension of roughly the same length as the other full 'arm' is denoted as domain IIIb (red) and is present only in the *P. asymbiotica* and *P. luminescens* Pvc11$_{\mathit{lopT}}$ structures. The DALI server was used to find homologues based on the atom coordinates (Holm and Rosenström, 2010), as was done by Xu and colleagues (2022) (Figure 3.18C). The DALI search indicated strong alignment at domain IIIb to a single chain of the non-catalytic CBM79 from the GH9 family of glycoside hydrolases encoded in *Ruminococcus flavefaciens* FD-1 (CBM79-1$_{RfGH9}$; PDB: 4D3L; Z-score = 6.6, RMSD = 3.0Å) (Figure 3.18B and C). Moreover, the same hypothetical CBM protein which aligned to

Alg11 (PDB: 4JX0) was found to superimpose well with Pvc11$_{lopT}$ but had its highest alignment with domain IIIa, along with another carbohydrate binding protein: xylanase 10C from *Cellvibrio japonicus* (PDB: 1US2 and 1US3; Xyn10C). Xyn10C contains a CBM15 which has been shown to bind internal glycan chains (i.e *endo*-type or type B) of xylo-oligosaccharides and β-1,4 glucans (Pell *et al.*, 2004); the open conformation of the concave region of Pvc11$_{lopT}$ IIIb concurs with likely being a type B CBM class. In contrast, Pvc11$_{pnf}$ showed a lack of CBMs for its best molecular comparisons via the DALI server.

One of the highest scoring CBM classes of the top 50 structures were those of CBM classes 79 (PDB: 4D3L) and 80 (PDB: 4V17) which have been shown to bind linked β-1,3-1,4-glucans as well as β-1,4-glucans and mannans (Venditto *et al.*, 2016). Indeed, both domain IIIa and IIIb appear to be rich in β-sheets; domain IIIb, in particular, appears to contain a total of 2 sheets of 4 antiparallel β-strands separated by 4 short α-helices (of the form 2x [ββαβαβ] with β-sheets ordered 1-8-3-6 and 2-7-4-5), as well as a further 7 β-strands with 2 short α-helices in its neighbouring proximal motif) (Figure 3.19). β-sandwich folds are a key feature of many classes of CBM in the "fold 1" superfamily and which often coordinate a metal ion (reviewed by Boraston and colleagues (2004)); if the α-helices in this fold are significant, however, then it would be one of very few CBM motifs to encode a Rossman-like fold aside CBM87 (Bamford *et al.*, 2020).

On the other hand, there is a great diversity of ligand binding targets between CBMs such that inferring this specificity from its structure alone is particularly difficult; though the orientation of aromatic amino acids at the solvent-accessible surface plays a particularly important role (Simpson *et al.*, 2000). In CBM79-1$_{RfGH9}$, Trp564 and Trp606 play an essential role in binding since alanine substitution at either of these sites abrogates binding with all key ligands (Venditto *et al.*, 2016); these amino acids are located in the concave surface of the β-sandwich fold: a structure also found in the Pvc11$_{lopT}$ domain IIIb of which the closest aromatic residues to this fold are Tyr863 and Tyr868

(Figure 3.19). These residues may be key in any binding activity that this CBM candidate may have: a hypothesis which should be tested further.

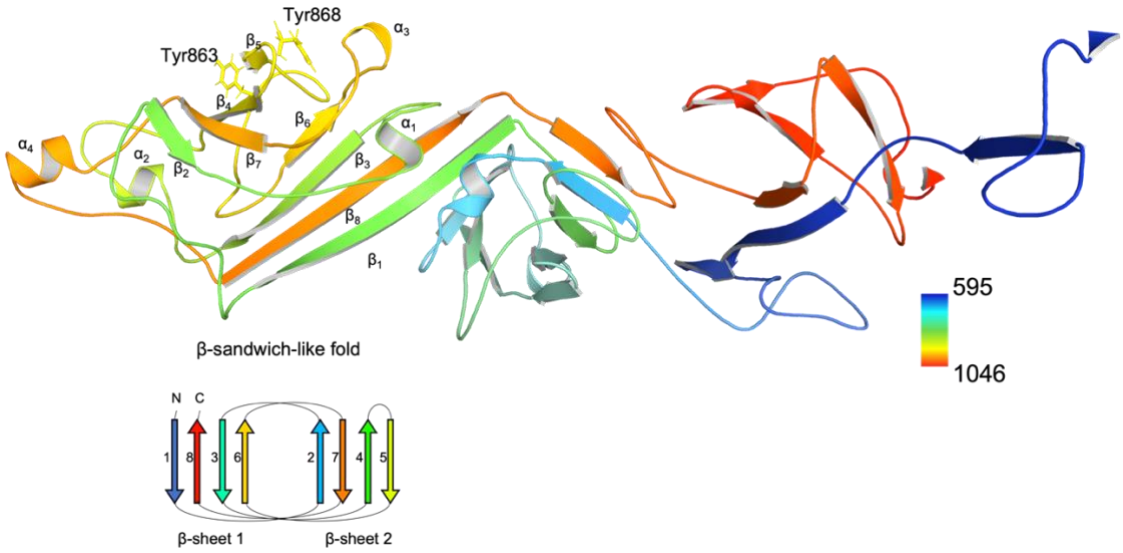


Figure 3.19 | Pvc11$_{lopT}$ domain IIIb is rich in β-sheet supersecondary structures.
AlphaFold2.0 prediction indicates an abundance of supersecondary structures present in domain IIIb which may contribute to carbohydrate binding activity. Aromatic amino acids, Y863 and Y868, are found at the concave surface of the β-sandwich at the most distal end of the domain which aligned most significantly with CBM79-1$_{RfGH9}$ (PDB: 4D3L), in which the closest aromatic residues are essential for ligand binding. Underneath the figure is a schematic of the supersecondary structure of the β-sandwich-like fold; coloured by residue number.

Due to the complexity of structure to functional binding with regards to protein-glycan interactions, a non-structure based approach was sought to assess the ability of Pvc11$_{lopT}$ to bind a range of glycan targets. To this end, a machine learning extension of the "glycowork" framework, LectinOracle, was used to predict Z-scores of AI-predicted fluorescence values of an *in silico* microarray (Bojar *et al.*, 2021; Burkholz, Quackenbush and Bojar, 2021; Thomès, Burkholz and Bojar, 2021; Lundstrøm *et al.*, 2022). The dataset used contains over 55,000 glycan targets analysing more than 560,000 protein-glycan interactions and, by using raw array data from the Consortium for Functional Glycomics database, includes abundant true negative interactions which

protein-glycan co-crystallisation data does not provide. A customised Jupyter Notebook script was used for testing a range of protein targets against the glycan target dataset provided by Thomès, Burkholz, and Bojar (2021); the link can be found in Materials and Methods chapter 3.3.1 on page 112.

After background correction, however, none of the Pvc13 tail fibres nor Pvc11 baseplate sequences were significantly predicted to bind any glycans in the dataset. The authors note that lectins with heterogeneous binding properties are more difficult for this tool to interpret. The sequence for concanavilin A (ConA) is generally well predicted when compared to experimental glycan microarray data, but LectinOracle fails to predict binding for lectins such as DC-SIGN which possess only a few binding sites in their CBMs (Lundstrøm *et al.*, 2022). Therefore, since predicting lectin-glycan binding remains to be a non-trivial task in glycobiology, one cannot rule out the possibility of false negative predictions for the class of structural PVC components using this AI-driven tool. Future improvements in this field may, indeed, support the hypothesis that Pvc11$_{lopT}$ contains one or more CBMs.

Taken together, the Pvc11$_{lopT}$ may, indeed, bind target sugars in the cell glycocalyx in a similar fashion to that inferred of MACs, T6SS^{iv}, and AlgoCIS which all lack tail fibres but form cages around the spike protein with a protrusion for injection of the tip (Xu *et al.*, 2022). Unpublished TEM data from the Waterfield lab indicate that PVC*lopT*, expressed from an *E. coli* cosmid, encodes a 'double baseplate' which fits with predictions here of the added extension in its Pvc11 (i.e., domain IIIb) which acts as another 'arm' in the molecular model. Contrary to the method of PVC*pnf*, the double baseplate may facilitate binding to the target cell glycocalyx using either or both of domains IIIa and IIIb. If true and correct, this may prove to be an important feature for the potential future biotechnological use of PVCs for delivery of therapeutic proteins since modulating these CBMs could alter cell-specific binding activity in a manner which does not require modified homotrimeric tail fibres.

3.5 Discussion

3.5.1 Implications of Functional Diversity of eCIS-associated ORFs

During the search through the genomes of the eCIS database, many ORFs were predicted to be transposases which hints at the previous stipulations that eCIS operons frequently undergo horizontal gene transfer (HGT): a hypothesis further suggested by the incongruence between genomic phylogeny and that of the homology of eCIS core elements *afp8* and *afp11*, as well as the sporadic appearances of eCIS operons across diverse clades (Chen *et al.*, 2019; Geller *et al.*, 2021). Despite the widespread integration of eCIS loci into genomes across the microbial kingdoms and the common tendency for HGT, eCIS elements are depleted in other common human pathogens such as *Escherichia*, *Salmonella*, *Staphylococcus*, and *Streptococcus*. Indeed, eCISs are enriched in microbe lifestyles that are symbiotic in aquatic or terrestrial aerobic settings and depleted in anaerobic settings or tissues of mammals or birds (Geller *et al.*, 2021).

In addition to the findings in the diversity of effector functions across lineage Ia, *Photorhabdus* appears enriched in rearrangement hot spot (Rhs) repeat-containing toxins: a class of polymorphic toxins with a specialised PAAR domain (Jurėnas, Rosa, *et al.*, 2021).

The N-termini of Rhs toxins associate the toxin with its secretion apparatus. An example of an Rhs protein in *P. luminescens* is the Rhs1 protein, encoded at the gene *plu0353*. The N-terminus of this protein is stabilised by its effector-associated gene (Eag) family chaperone, EagR (*plu0354*), and uses its pre-PAAR and PAAR domains to bind directly with the T6SS VgrG (*plu0355*) stabilised by the tetrahedral coordination of a Zn^{2+} ion via the histidine in the PAAARxxDxxxH motif (Jurėnas, Payelleville, *et al.*, 2021).

Similarly to Tc toxins, Rhs1 encapsulates its toxic C-terminal domain, Rhs1^{CT} or Tre23, inside of a β-barrel shell which translocates the toxin across the membrane and releases it by aspartyl-autoproteolysis on both sides of the shell (Jurėnas, Payelleville, *et al.*, 2021). Tre23 blocks protein synthesis through ADP-ribosylation of the 23S rRNA in the target bacterium unless the target cell is expressing an immunity protein, Tri23, encoded by the gene *plu0352*. Since Rhs1 is associated with a T6SS, *Photorhabdus* likely uses the toxin for inter-bacterial competition much like how Rhs proteins are employed by *Dickeya* (Koskiniemi *et al.*, 2013) and *Serratia* (Diniz and Coulthurst, 2015). Many other novel Rhs proteins are discussed by Ma and colleagues (2017).

Geller and colleagues also found that not all putative effector genes are encoded downstream of the final structural gene in the operon, such as the putative restriction endonuclease within the eCIS operon of *Moorea producens* PAL 15AUG08-1. This represents a limitation to the method used here for assessing the effectors of eCIS operons. In particular, more complex operons such as the MAC from the HI strain of *Pseudomonas luteoviolacea* has many ORFs encoded prior to the majority of structural genes, one of which is the metamorphosis inducing factor (Mif1) (Figure 3.20; adapted from Alker and colleagues (2020)). This ORF is a good example of an effector which would not be acquired by looking solely downstream of the operon; it is important to stress, therefore, that the analysis done in this work on the dbeCIS database is for eCIS-downstream ORFs only which may or may not include the primary effectors used by the eCIS. Whilst this could have implications for how eCIS operons were analysed in this work, the presence of LSs in many of the *Photorhabdus*, *Yersinia*, and *Serratia* species for eCIS-downstream ORFs suggests that this separation of core and modular payload components tends to be the more common occurrence.

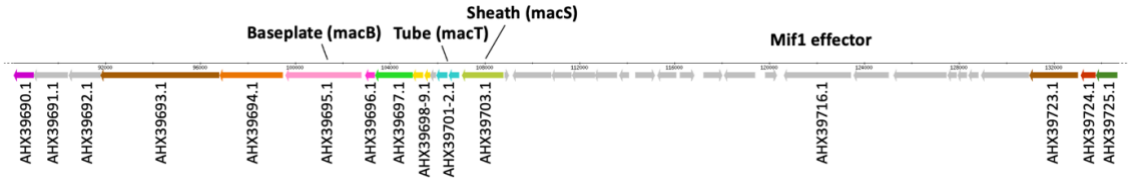


Figure 3.20 | The *P. luteoviolacea* HI MAC has an effector encoded within the operon.
Like other lineage Ib eCIS, the MAC cluster genes are spaced farther apart than most lineage Ia eCIS operons. The afp15 and afp16 orthologues are encoded upstream of the other structural components including the effector Mif1, the sheath, macS, tube proteins macT1 and 2, and the afp11 orthologue macB. Imaged from the dbeCIS webpage and adapted from Alker *et al*. (2020).

3.5.2 Leader Sequence Homologies

The dbeCIS database contains many genomes of bacteria possessing lineage Ia eCIS, but it does not include all of them. European strains such as *P. asymbiotica* HIT and JUN which, exceptionally to the species, do not cause human photorhabdosis, as well as the exceptional Texas strain of *P. luminescens* and the significant *P. australis* (a.k.a Kingscliff strain) are not included in the database. The analysis done on the included strains is a useful interpretation of the lineage as a whole and the likely functions of eCIS-downstream ORFs but could be enhanced by the addition of these exceptional strains.

In particular for leader sequence homology, the LS of PaTox within the Kingscliff strain is markedly more similar to the ATCC43949 Pnf LS (Figure 3.21) which would expand upon the groupings of homologous LSs identified from dbeCIS if included in the dataset.

```
         v10       v20       v30
MLKYANPQAVPTQRTKNTAKKPSSSSSFDGQLELSNGE  PAU_03332 ATCC43949: Pnf
MLKYANPQ-V-TQRTKNTAKKP-SS-SFDG-LELSNGE
MLKYANPQTVATQRTKNTAKKPPSSTSFDGHLELSNGE  PAK_03185 Kingscliff: PaTox
M--------------------------G--------
MKGIEGVIMLSHDILPEKLLVSEKKHENVGSYFSDDIG  PAU_02230 ATCC43949: PaTox
```

Figure 3.21 | Alignment of LSs of PaTox and ATCC43949 Pnf.

Evidence of horizontal gene transfer (HGT) is exemplified by the homology of the leader sequences of *P. australis* PaTox (Kingscliff) and that of the North American *P. asymbiotica* ATCC43949 Pnf. Aligned residues are indicated between homologous sequences.

3.5.3 Implications of Machine Learning Methodologies

Overall, machine learning and neural networks serve a great purpose in testing well established hypotheses in molecular biology. However, one should be cautious about using these softwares to make inferences. For example, AlphaFold2 predicted an interaction of the Pnf LS with the N-terminal domain of monomeric Pvc15 with ambiguous PAE values representing uncertain multimer interactions. Similarly, LectinOracle failed to predict glycan targets for even positive control lectins such as DC-SIGN and concanavalin A (Lundstrøm *et al.*, 2022), as well as Pvc13 tail fibres which could, therefore, represent false negative results.

3.6 Conclusions

In summary, the eCIS effectome appears to be enriched in RHS repeat-containing toxins in *Photorhabdus*. PVC homologues, specifically, are enriched in LS-like highly disordered N-termini. This suggests other mechanisms may be in play when loading other eCISs in other genera, or that some eCIS putative payloads are not always encoded immediately downstream of the operon, as is typical of the PVC and Afp. Lineage Ib has a greater diversity of functions for eCIS-downstream ORFs including metallopeptidases, regulators, and transporters. Experimental evidence should be used to confirm the relevance of these genes in the eCIS mechanisms of their respective species.

Finally, whilst it was known that Pvc15 was involved in the process of PVC-payload association to some extent, its protein structure and function had yet to be characterised. AlphaFold predicted three distinct domains involved in the ATPase activity which may be exclusively carried out by the D2 domain.

4. Modulating PVC*pnf* Payloads by Encoding Native Leader Sequences

4.1 Introduction

In addition to the work by Jiang and colleagues (2022), the Waterfield lab had identified leader sequences (LSs) as a loading signal for PVC payloads. The capacity for the PVC to load small and large protein payloads had yet to be explored, however, and it was desirable to identify that LS-tagged proteins were still functional after injection. Such a finding would have implications for the use of the PVC as a biotechnological tool for the direct delivery of proteinaceous 'drugs' into human cells.

4.2 Aims

i) Verify that LS-tagged heterologous payload proteins associate with the PVC by co-immunoprecipitation (Co-IP).
ii) Verify the functional injection of heterologously-loaded PVCs by assaying cell lines for fluorescence output or cytotoxic effects.

4.3 Methods

4.3.1 PVC Purification

The typical workflow for purifying PVCs loaded with any given payload is detailed in this chapter. In this workflow, one of the PVC*pnf* proteins was labelled with an epitope tag for purification. The PVC*pnf* operon was expressed by the addition of arabinose on a pBAD plasmid whilst the payload was expressed simultaneously with IPTG from a pVTRa plasmid vector. Some methods used were done in collaboration with Nanosyrinx Ltd; details of these methods have been omitted due to potential conflicts of interest regarding their patented works.

4.3.1.1 Expression and Harvesting

A starter culture was prepared as 10 mL BL21(DE3) 'NiCo21' cells (New England Biolabs) containing the pBAD-PVC*pnf* and pVTRa-payload vectors and grown in LB + Ampicillin + Chloramphenicol and 37 °C shaking at 180 rpm overnight. The next day, some of these starting cultures were frozen in glycerol stocks (see chapter 2.1.2) whilst the remaining volume was used to inoculate 1:100 in a larger volume to be grown again in the same conditions. 10 mL into 1L was typically used for purification whilst 2.2 mL into 250 mL was used for generating an IPTG gradient, as below. Cultures were induced by the addition of filter-sterilised arabinose at a final concentration of 0.2% and grown overnight at 25 °C and 200 rpm.

4.3.1.1.1 IPTG Gradient

Before purifying the loaded PVCs, it was necessary to identify the optimal concentration for induction by IPTG for each strain of the PVC-payload transformants since too much IPTG can negatively affect cell growth (Dvorak *et*

al., 2015). The 250 mL culture was first induced with arabinose, as above, then split into five 50 mL tubes. A gradient of IPTG concentrations was performed at final concentration values of 0.1, 0.25, 0.5, and 1 mM, as well as an arabinose-only sample. These samples were grown overnight at 25 °C and 200 rpm, as above.

The next day, OD_{600} was measured and samples were diluted to normalise to OD_{600} = 1 for each in a total of 100 μL in 1x LDS sample buffer with 50 mM dithiothreitol (DTT) and boiled for 20 minutes. The amount of PVC syringe as a fraction of the total cell content was assessed by SDS-PAGE and staining with Coomassie blue (see chapter 2.5.1). The optimal concentration of IPTG could then be found for each PVC-payload combination.

4.3.1.2 Lysis

To prepare for purifying the PVCs using optimal induction conditions, lysis buffer was prepared as follows:

- Lysis buffer
    - 200 μg/mL lysozyme (Sigma Aldrich, L6876)
    - 0.5% Triton X-100
    - 1 tablet complete EDTA-free protease inhibitor per 50 mL (Sigma Aldrich; cOmplete EDTA-free protease inhibitor cocktail, 11873580001)
    - 70 μL (140 units) per 100 mL DNase I (NEB, M0303S)

Starting cultures were grown once again overnight at 37 °C and 180 rpm, inoculated 1:100 volume the next day, and, upon reaching OD_{600} = 1, induced with both arabinose and the optimal IPTG concentration overnight at 25 °C and 200 rpm. Cultures were then spun at 3,000 g for 20 minutes at 4 °C. The cell pellet was weighed and 10 mL of lysis buffer was added per gram of pellet. After resuspension, the sample was sonicated using a needle sonicator on ice in intervals of 10 seconds spaced 30 seconds apart for a total of 60 seconds 'on'

time or using a cell homogenizer at 30 kpsi (One Shot Model; Constant Systems LTD).

4.3.1.3 Lysate Preparation

To measure the protein content at each stage of the process, 25 μL samples were taken for each of the following steps and topped up to 50 μL with 2x LDS sample buffer with 100 mM DTT, unless otherwise specified. These samples were later used for SDS-PAGE and WB assays to assess the purity of samples throughout the purification process. For the purpose of this protocol, each sample to be saved for later assaying will be designated a bolded and italicised letter. For instance, 25 μL of the 'total' sonicated sample was aliquoted and is designated as ***T***.

The sample was then spun at 12,000 g for 35 minutes at 4 °C. The supernatant was separated and kept, whilst 25 μL was aliquoted as above (***S***). Some of the pellet was scraped out and was resuspended in 50 μL of 1x LDS sample buffer with 50 mM DTT (***P***). The remaining supernatant was then filtered through a syringe filter with a 0.45 μm pore size.

4.3.1.3.1 Resin Immunoprecipitation

To purify the PVC*pnf* syringes, immunoprecipitation was used via a gravimetric flow column (Bio-Rad; Econo-column®; cat. no.: 7371512). The flow-through was collected in fractions (***FT***), washed with 10-15 mL PBS (***W***), and eluted using competitive elution buffer (***E1-3***). Protein-protein interactions were determined qualitatively by co-immunoprecipitation using either an α-His resin (Roche cOmplete™ His-tag Purification Resin; Cat. No.: 5893682001) or an α-Myc resin (Pierce c-Myc Tag IP/Co-IP Kit; Cat. No: 20168). FT samples were collected from both methods and SDS-PAGE was used to assess protein content and the presence of the protein of interest in the purified sample. The α-c-Myc

kit was also used as per the manufacturer's instructions with only minor optimisations.

For each aliquoted sample, SDS-PAGE gels and WBs were prepared, loading 10 μL of each sample (see chapter 2.5.1). Lanes were loaded as shown in the table below along with the desired results (Table 4.1).

Table 4.1 | Gel loading samples for assaying PVC purification.
Each aliquot containing LDS sample buffer and dithiothreitol (DTT) was added to protein gels for assaying by SDS-PAGE and WB. Sample letters indicate solutions taken at stages of the purification process.

		Desired Result	
Sample	**Volume (μL)**	**Coomassie (bands)**	**Western blot (band(s))**
T	5	Many	Good
S	5	Many	Good
P	5	Many	None
FT	5	Many	None
W	10	Some	None
E1	10	1+, bright	Good
E2	10	1+, fair	Fair
E3	10	1+, faint	Faint

4.3.1.4 Formaldehyde Cross-linking

For co-immunoprecipitating transient protein-protein interactions, the non-specific cross linker formaldehyde (FA) was made after boiling paraformaldehyde (pFA) for 1 h at 80 °C and subsequently filtered using a 0.22 μm pore filter; FA was stored in the dark at room temperature and discarded after 4 weeks. After harvesting 350 mL of induced bacterial cultures and

pelleting at 3,500 g for 20 minutes, cells were resuspended in 14 mL PBS + 0.75% FA for 30 minutes at room temperature: a modified protocol from that of Chowdhury and colleagues (2009). The cross-linking reaction was quenched by the addition of glycine to a final concentration of 0.125M. Cells were centrifuged again and the remaining protocol was followed after the addition of lysis buffer.

4.3.2 Cloning Heterologous Payloads into a PVC-compatible Vector

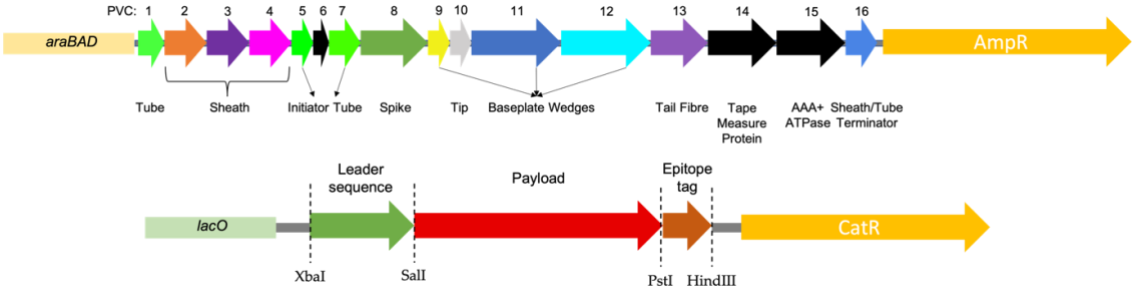


Figure 4.1 | Schematic of the pBAD-PVC*pnf* and pVTRa vectors.
The pBAD (top) and pVTRa (bottom) vectors are used to express payload proteins in a PVC nanosyringe, heterologously expressed in *E. coli*. A double-transformed strain is resistant to ampicillin and chloramphenicol at working concentrations and PVCs can be purified after lysing cells in a cell disruptor. Cloning of payloads, antibody tags, and leader sequences is conducted using restriction enzymes whose cut sites are depicted as vertical dashed lines. Expression of the pVTRa components is activated by the presence of isopropyl β-D-1 thiogalactopyranoside (IPTG) whilst pBAD-PVC*pnf* components are activated by the addition of arabinose.

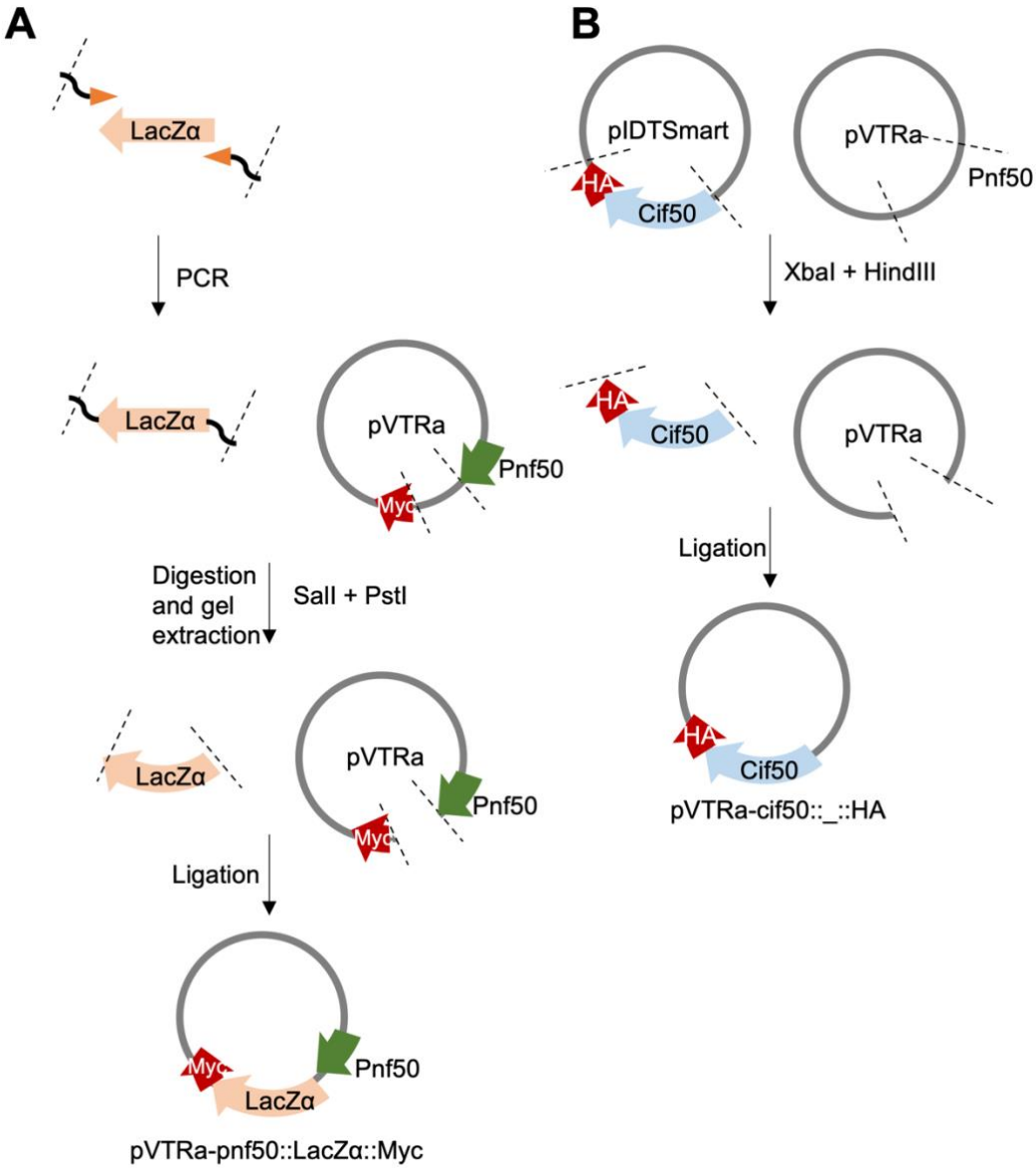
To express the payload protein as well as the PVC nanosyringe complex heterologously in *E. coli*, two vectors were used: a derivative of pBAD30 but with a ColE1 origin of replication (termed pBAD-PVC*pnf*) and pVTRa, which has a pSC101 low copy-number origin of replication, where the payload was encoded (Figure 4.1). The modularity of the pVTRa vector by restriction ligation procedures makes it an ideal candidate for expression of payloads for loading

into the PVCs. Expression of chloramphenicol acetyltransferase (CatR) encoded on the pVTRa vector and β-lactamase (AmpR) on the pBAD-PVC*pnf* vector (both provided by the Waterfield lab) enables double transformants to express PVC loaded with a payload of choice. The pVTRa backbone is ~5.5 kb in size and encodes a leader sequence, between the BamHI/XbaI and SalI sites, derived from one of the *P. asymbiotica* PVC genes required for loading as well as an epitope-tagged payload protein immediately downstream; the payload is derepressed by the presence of 0.1 mM isopropyl β-D-1 thiogalactopyranoside (IPTG). The pBAD-PVC*pnf* vector is significantly larger than pVTRa as it encodes the structural components of the PVC which are induced with 0.2% arabinose.

Firstly, it was desirable to assess the loading diversity of the mature, purified nanostructures; this experiment, however, has been done more extensively in the work by Jiang and colleagues (2022). Nonetheless, it was decided that fluorescent payloads could be used for verifying injection by the nanosyringe so GFP and LacZ (β-galactosidase) were chosen as candidates for LS-tagged payloads. Since it was unsure whether these payloads may be too large or structured to load, a 'split' system was also devised for each: a larger protein component (GFP(1-10) and LacZω) could be transfected into the target cell and complemented by the injection of the smaller components (GFP11 and LacZα) (Cabantous, Terwilliger and Waldo, 2005; Hamed, Khedr and Abdelraof, 2020).

Genes intended as payloads for the PVC such as GFP11, GFP, LacZα, *P. temperata*_PVCu1, and *S. entomophila*_Afp17 were synthesised as gBlocks or inserted into the pIDTSmart vector, courtesy of Integrated DNA Technologies (IDT). Inserts were cloned into the IPTG-inducible pVTRa vector using a standard PCR-digest-ligate workflow (Figure 4.2). The pVTRa template vector already encoded the pnf50 LS and a Myc tag, so splicing using SalI and PstI was sufficient to produce functional plasmid vectors. To alter the LS and epitope

tag, an XbaI-HindIII digest could be used whilst retaining the SalI site in between them for splicing of other payloads into this vector (Figure 4.2A and B); restriction digestion and ligation was used as detailed in Methods and Materials chapter 2.4.4 on page 100.



C
pHR'-mScarlet-GFP11
LacZα
mScarlet
GFP11
PCR
LacZα
mScarlet
Overlap Extension PCR
LacZα
mScarlet
pHR'-mScarlet-GFP11
mScarlet
GFP11
Digestion and gel extraction
BamHI + NotI
mScarlet
LacZα
pHR'-mScarlet-GFP11
Ligation
pHR'-mScarlet-LacZα
mScarlet
LacZα
D
9 A V V L … E A 43…
LacZ GCC GTG GTt tta… gag GCC…
LacZω GCC GTG GT- ---… --G GCC…
9 A V V A 12…
pD864-LacZ
LacZ
PCR
-LacZω
PCR
LacZω
pD864-LacZ
LacZ
PCR
LacZ
pHR'-GFP(1-10)
GFP(1-10)
Digestion and gel extraction
BamHI + NotI
Digestion and gel extraction
BamHI + NotI
LacZω
pHR'-GFP(1-10)
LacZ
Ligation
Ligation
pHR'-LacZω
LacZω
pHR'-LacZ
LacZ

Figure 4.2 | Workflow schematic of plasmid construction.

Standard and overlap extension polymerase chain reaction (PCR), restriction enzymes, and T4 ligase were used for splicing. Construction of pVTRa vectors are shown in **(A)** and **(B)**, whilst pHR' vector construction is shown in **(C)** and **(D)**. Genes of interest are depicted as block arrows with primers depicted as triangles, coloured as complementary to their relative template; overhangs are indicated as thin segments. Restriction sites are depicted as dashed lines.

(A) Most inserts were either amplified from their pIDTSmart vectors (ordered from IDT), or as individually synthesised gBlocks to be spliced into the pVTRa vector. The α-complement subunit of β-galactosidase (LacZα) is depicted, though other payloads including GFP, GFP11, and the heterologous payloads *P. temperata* PVCu1 and *S. entomophila* Afp17 were all spliced into their pVTRa vectors using this method.

(B) The modulation of leader sequence and epitope tag was achieved using the XbaI and HindIII restriction enzymes. Splicing of the cif50 leader sequence and HA tag was done using this method and modulated payload proteins are cloned between the leader and tag using the SalI and PstI restriction sites.

(C) To integrate LacZα into the pHR' vector for mammalian expression, it was desirable to retain the N-terminal mScarlet to verify successful transfection or transduction by microscopy. Since there was no restriction enzyme site encoded between mScarlet and GFP11 in the pHR'-mScarlet-GFP11 vector, overlap extension PCR was used to fuse LacZα downstream of mScarlet. Splicing was then conducted using BamHI and NotI.

(D) Full-length LacZ (LacZ) and ΔM15 LacZ (LacZω) were amplified from the pD864-LacZ vector (Addgene #109380). Amino acids 12-42 were removed in LacZω using the sequence change shown in the top text box, in accordance with the sequence found in ΔM15 *E. coli* (Prentki, 1992). In this sequence, the final thymine of codon 11 (i.e., GTT; valine) is replaced with the final guanine of nucleotide 42 (i.e., GAG; glutamate), thus removing codons 12 to 42. After the first PCR, the amplified LacZω is missing its first 11 amino acids (-LacZω). The overhang region of -LacZω is homologous to the primer used for amplification of full-length LacZ; the second subsequent PCR results in the correct LacZω and the same primers are used for amplification of full-length LacZ.

Plasmids for mammalian expression were constructed using the pHR vector (Naldini, Blömer, Gage, *et al.*, 1996). In particular, GFP(1-10) and LacZω were of interest to splice into this vector so that fluorescence would result from injection of their complementary GFP11 and LacZα PVC payloads, respectively (the latter also requiring the fluorescent probe BioTracker 543). As a positive control for the 'split-LacZ' system, mScarlet and LacZα were amplified with overhangs homologous to the linker region found at the 3' end of mScarlet in this vector (a 'GSGS' amino acid motif). Another PCR using both amplified products as a template enables fusion of these two regions when the homologous region anneals during the reaction; only the fused product can be amplified when using primers found at the 5' of mScarlet and 3' of LacZα. The fused products, as well as the pHR'-mScarlet-GFP11 vector (gifted from the John James lab), were then cut using BamHI and NotI (Figure 4.2C).

It was desirable to express full-length LacZ as another positive control for the 'split-LacZ' system, so it was amplified with a BamHI restriction site and Kozak sequence from the pD864-LacZ vector (Addgene #109380) (Figure 4.2D). In order to construct LacZω, the gene was split into two parts by amplifying 5' from the final 'G' in codon 42, with an overhang homologous to the upstream sequence of the second 'T' in codon 11. Amplifying using the same primers as that of LacZ results in the sequence of LacZω as found in ΔM15 strains of *E. coli* including DH5α and DH10β (Prentki, 1992). Digestions were done using BamHI and NotI to splice these constructs into pHR'-GFP(1-10), also a gift from the John James lab.

4.4 Results

4.4.1 Leader Sequences as Packaging Signals for Payloads

To test the modularity of PVC payloads, the PVC*pnf* chassis was expressed via the arabinose-inducible pBAD-PVC*pnf* plasmid with the corresponding payload co-expressed on the IPTG-inducible pVTRa plasmid (see Methods 2.4.1). Experiments were conducted to analyse loading and injection of payloads into eukaryotic cell lines including *Drosophila* S2 macrophage-like cells and human cell lines: Jurkat T cells, Raji B cells, HEK293T cells, and THP-1 monocytes. Table 4.2 details the relative sizes of natural and heterologous payloads in the PVC*pnf* chassis from previous works as well as this one.

Table 4.2 | Validations of natural and heterologous PVC*pnf* payloads.
Unless otherwise specified, the table specifies whether each payload has been experimentally validated for loading and delivery via the mature PVC*pnf* multi-protein complex.

PVC payload	Size (kDa)	Loads?	Reference
Natural payloads			
Pnf	31.1	Yes	(Yang *et al.*, 2006; Wang *et al.*, 2021)
Cif	29.5	Yes	(Chavez *et al.*, 2010)
LopT	36.8 (theoretical)	Yes	Unpublished work
$RRSP_{Pa}$	61.0 (theoretical)	*Unknown*	(Wang *et al.*, 2020)
Pdp1 / SepC-like	37.2	Yes	(Wang *et al.*, 2021)
PaTox	334.9 (theoretical)	*Unknown*	(Bogdanovic *et al.*, 2019)
$PaTox^{P,G,D}$	47.0, 37.5, 25.1 (theoretical)	*Unknown*	
Heterologous payloads			
pnf50::Pnf(C190A)::Myc	38.5	Yes	(Vlisidou *et al.*, 2019)

pnf50::GFP::Myc	34.2	Yes	This work
pnf50::GFP11x1::Myc	9.4	Yes	This work
pnf50::LacZα::Myc	19.4	Yes	This work
pnf70::LacZ	123.8	Yes	(Jiang *et al*., 2021)
pnf50::*P.temperata*-VY86_19225::Myc	18.1	Yes	This work
pnf50::*S.entomophila*-Afp17::Myc	48.1	No	This work

Though the resazurin reduction assay was a tractable means of measuring injection and cytotoxicity for the toxic payload Pnf, it was desirable to identify a new means of verifying injection of heterologous payloads such as by fluorescence of a GFP payload after injection, or through enzymatic reaction with a probe which would catalyse fluorescence via external means, such as by using a β-galactosidase payload with X-Gal or BioTracker 543 fluorescent probes.

WB and SDS-PAGE protocols were used to verify that the PVCs and their associated payloads were present at each step of the purification process. Heterologous payloads were encoded with an N-terminal pnf50 LS and co-expressed with the PVC*pnf* operon. LacZα and *P. temperata* VY86_19225 appeared to co-purify well with low amounts of payload in the flow-through and high levels in the elution as analysed by α-Myc primary antibodies used in WB (Figure 4.3). GFP and GFP11 payloads, however, were more difficult to discern whether they had correctly associated with the PVCs given the low intensity of elution bands compared to flow-through bands which could suggest non-specific association with the resin. HEK293T cells incubated with loaded PVCs were lysed and WB was attempted to observe payloads in cell lysate, but this was never successfully shown even with the cognate PVC*pnf*-Pnf (data not shown). It is possible, therefore, that the amount of injected protein

was too small to be recovered from lysates or that payloads were degraded within 1 h of incubation with these PVCs.

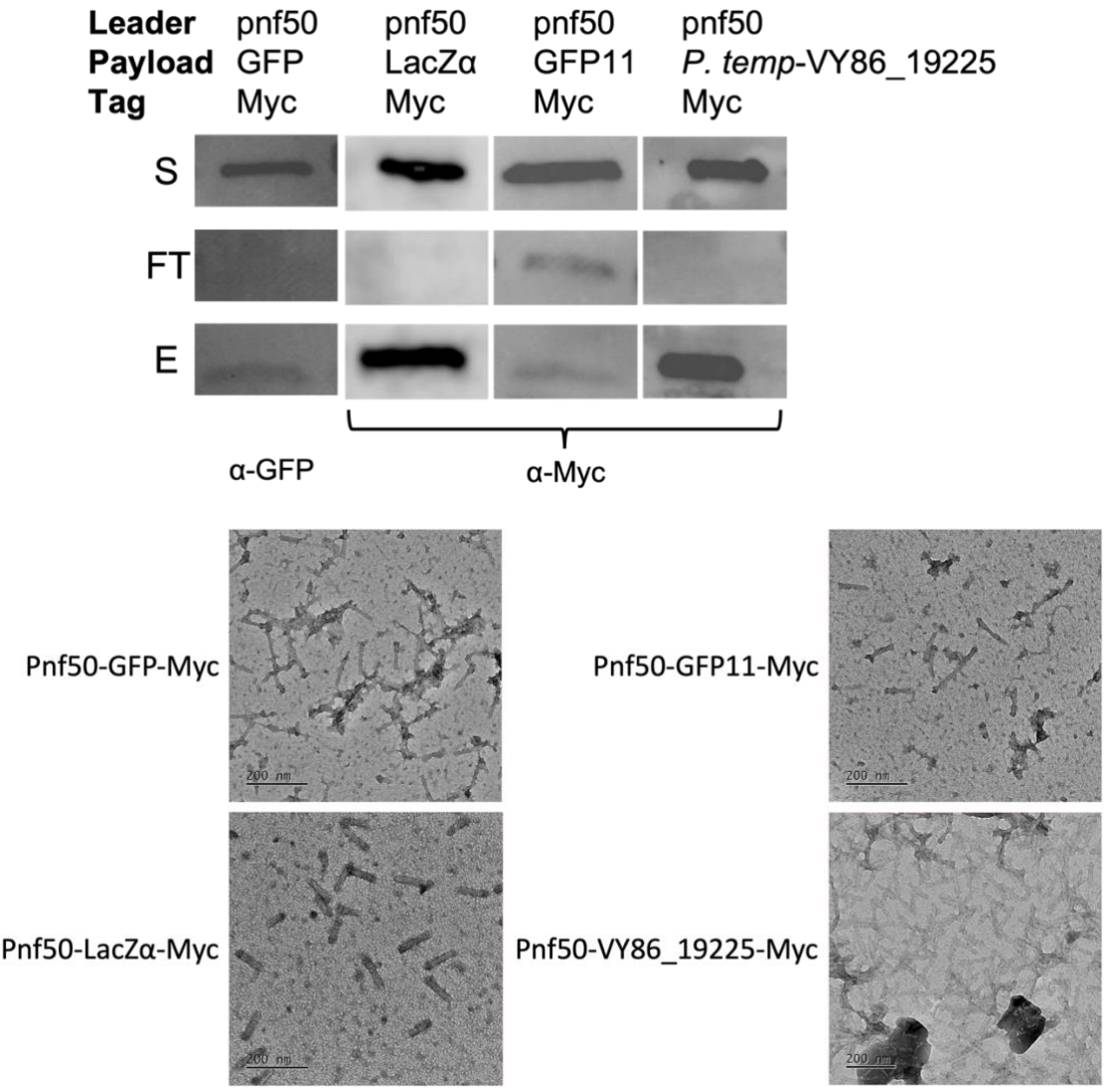


Figure 4.3 | Western blot of aliquots of the heterologous payload-PVC*pnf* co-purification process.

Immunoprecipitation of PVC*pnf* chassis (pBAD-PVC*pnf*) co-expressing pVTRa-Pnf50-GFP-Myc, pVTRa-Pnf50-LacZα-Myc, pVTRa-Pnf50-GFP11-Myc, or pVTRa-Pnf50-VY86_19225-Myc were analysed at each step of the purification process: supernatant (S), flow-through after equilibration with the resin (FT), and elution from the addition of competitive elution buffer (E). Primary antibodies used are indicated at the bottom. TEM images of each elution are shown in the lower panel.

It was next tested whether GFP fluorescence would serve as a reliable readout of cellular injection; after multiple repeated experiments in various cell lines, no significant levels of fluorescence could be observed by fluorescence microscopy or plate reader analysis after incubation with PVC*pnf*-Pnf50-GFP-Myc. Lack of readout was also observed for the 'split GFP' system with GFP(1-10)-mScarlet transfected into the cell lines and GFP11 loaded into the PVC*pnf* chassis; this possibly echoed the previous finding of too low of a level of injection to be identified as a reliable readout. Subsequently, it was hypothesised that the 'split LacZ' system (i.e., LacZα and LacZω) would serve as a better means of verifying injection due to the amplification of signal via an enzymatic reaction.

The Pnf50-LacZ-Myc payload construct could not be cloned in the pVTRa vector, likely due to the large size of the insert, so the smaller LacZα component was encoded instead. Purified LacZα-loaded PVC*pnf* and a cell-permeable BioTracker 543 probe (Sigma Aldrich; Cat. No.: SCT025) were incubated with Jurkat and Raji cell lines transduced with the complementary LacZω. Fluorescence was measured after fixing with 4% paraformaldehyde using flow cytometry. Once again, no significant levels of fluorescence could be observed in any of these cells but were observed in positive control cells transduced with both LacZα and LacZω (data not shown). Interestingly, the LacZα component appeared to catalyse the probe independently of LacZω when fluorescence was measured in a plate reader but this could not be reproduced in flow cytometry.

4.4.2 *Serratia entomophila* Afp17 is Toxic to *E. coli*

After adding IPTG to induce *Serratia entomophila* Afp17 (gene and protein ID: AAT48354.1) in *E. coli* NiCo21 (DE3) cells, neither the FLAG (DYKDDDDK)-tagged Pvc16 nor the Myc-tagged payload were detected by WB with any of the final concentrations of IPTG (Figure 4.4). Induced cell lysates were also carried

through to the protein purification process and distinctly lacked PVC particles from elutions when viewed using transmission electron microscopy (TEM).

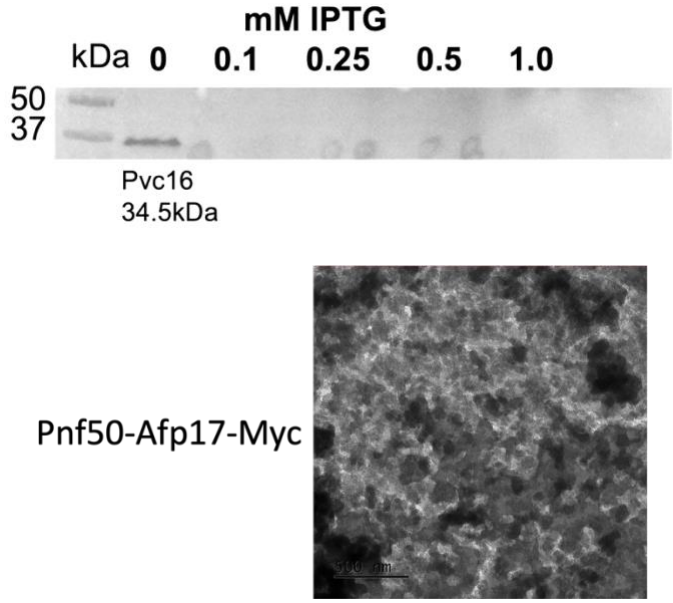


Figure 4.4 | *S. entomophila* Afp17 is toxic when induced in *E. coli*.
A primary α-FLAG (DYKDDDDK) WB of *E. coli* lysates over a gradient of IPTG induction. Lower panel shows a TEM of immunoprecipitation elution samples; no PVCs could be seen.

4.4.3 Verifying Injection of Payloads by their Effect on Respiration

Verification of injection using heterologous payloads, such as GFP and LacZ, did not appear to give meaningful readouts. It was hypothesised that this could be because of low rates of 'injection' or signal sensitivity, or a lack of binding of wild-type PVC*pnf* to the Jurkat, Raji, and HEK293T cells that were tested.

To gain insight into PVC activity upon cell lines, a plate reader experiment was conducted to quantify cytotoxic effects. Cell health and viability were measured through the effects on respiratory activity using the reduction of resazurin as an indicator of the action of the native PVC effector upon the mitochondrion (see Materials and Methods chapter 2.3.4). The fluorescences of resorufin were compared to that of untreated cells over 24 h (Czekanska, 2011).

Each cell line was first calibrated for the optimal number of starting cells such that resorufin fluorescence plateaus near the end of the experiment. A calibration curve was constructed for *Drosophila* S2 macrophage-like cells by serial dilution of cells from a 200,000 starting cell number (Figure 4.5A). High cell densities resulted in loss of resorufin fluorescence which is likely due to its over-reduction to the non-fluorescent dihydroresorufin. An optimal time point was chosen to illustrate a difference in relative respiration levels, as done by Oja and colleagues (2014). Cells were incubated with PVCs for 4 h before resazurin addition, then a further 6 h to measure relative resorufin levels. PVC*pnf* loaded with its cognate Pnf effector (PVC*pnf*-Pnf) inhibited respiration in S2 cells from a starting cell number of 15,000 (Figure 4.5B) (Student's Two-sample T test: t(4) = -9.00, p-value = 0.000850 ***). PVC*pnf* loaded with pro-apoptotic BAX and truncated BID (tBID) did not appear to affect cellular resazurin reduction.

Calibrations for Jurkat and Raji mammalian cell lines are shown in Figure 4.6, indicating optimal starting cell numbers of around 80,000 cells for both Jurkat and Raji cell lines. The subsequent experiment is shown in Figure 4.7. Student's T tests were conducted using the adjusted α value of 0.025 (i.e., 0.05 divided by two cell types). Only PVC*pnf*-Pnf showed a significant difference in resorufin fluorescence (Jurkat: t(4) = -16.36, p-value = 8.17e-05 ***; Raji: t(4) = -4.63, p-value = 0.010 *). Heterologous pro-apoptotic payloads tBID and BAX (gifted by Joe Healey) showed no significant change in resazurin reduction.

To affirm that PVC*pnf*-Pnf reduced resorufin production by dysregulating actin filaments, a follow-up experiment aimed to observe differences in actin filamentation across a gradient of inoculated PVC*pnf* concentrations. S2 and HEK293T cells were incubated with PVCs then fixed, permeabilised, and stained with Alexa Fluor™ 568-conjugated phalloidin. Unfortunately, neither S2 cells nor HEK293T cells indicated phalloidin-conjugate localisation using fluorescence microscopy; this effect has been observed when transfecting cells with Pnf using BioPORTER previously (Vlisidou *et al.*, 2019).

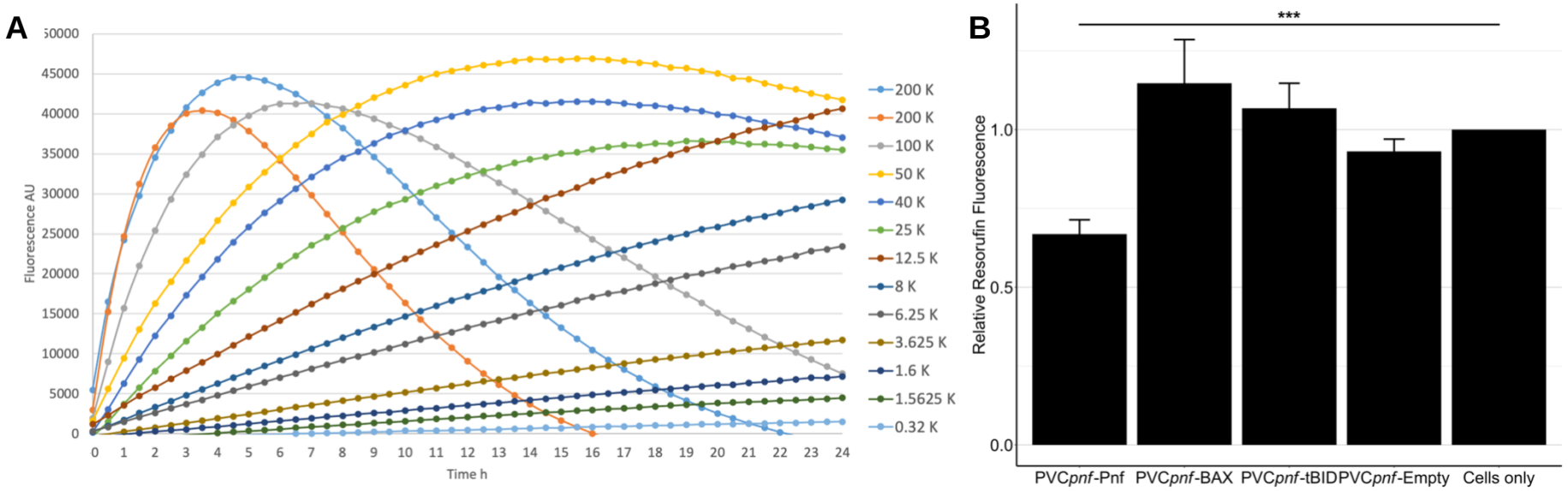


Figure 4.5 | *Drosophila* S2 cells are sensitive to native PVC*pnf*.

(**A**) *Drosophila* macrophages (S2 cells) were grown over a 24 h time course in Schneider's medium, initially to optimise for the correct number of starting cells per well. (**B)** Relative resorufin fluorescence was measured at the 6 h time point after incubation with different PVC preparations. A Student's T test revealed a significant reduction in respiration at 6 h when compared to the 'Cells only' control (t(4) = -9.00, unadjusted p-value = 0.000850 ***; n = 3 biological replicates of separately cultured cells. Error bars show standard deviation. No significant effect was observed for any of the other payloads including the pro-apoptotic BAX, truncated BID (tBID), or empty PVC*pnf* chassis.

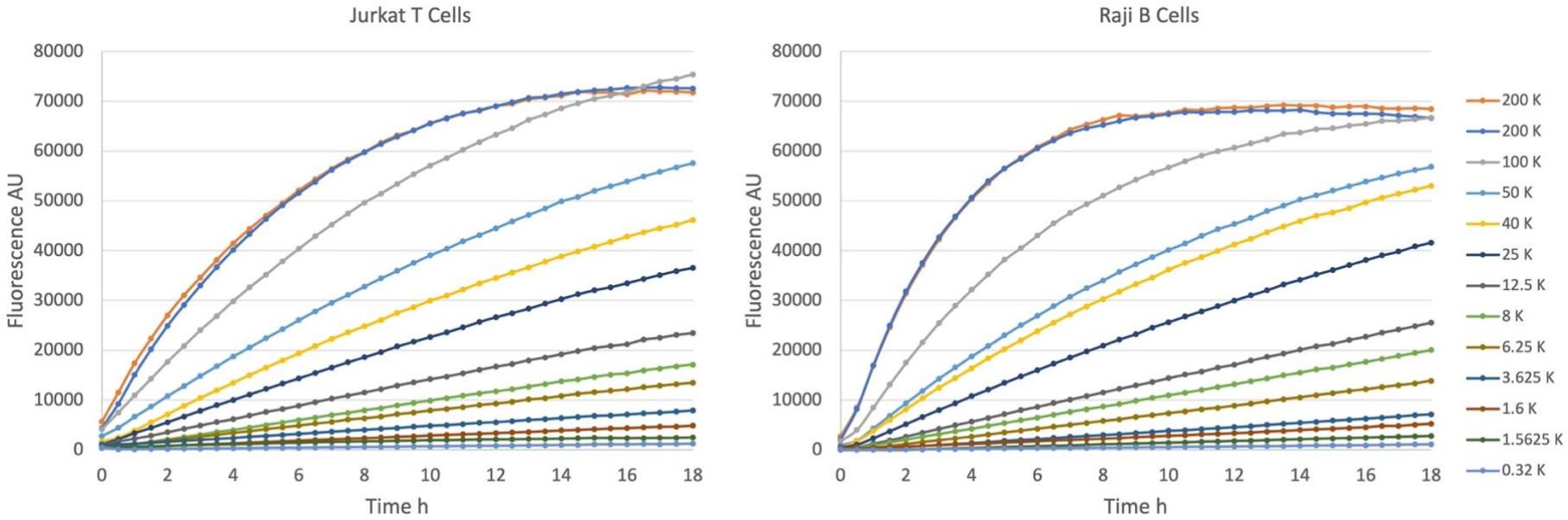


Figure 4.6 | Optimal starting cell numbers using resazurin calibration growth curves.

A gradient of starting cell numbers in a 96-well plate (100 µL per well) was measured over an 18-hour overnight time course. Black, flat-bottomed 96-well plates were used and fluorescence was measured at λ_{ex} = 530-570 nm and λ_{em} = 580-620 nm as a measurement of resorufin fluorescence. The optimal starting cell number is estimated as the number to which fluorescence just plateaus at 18 h; for both the Jurkat and Raji cell lines, optimal starting cell number was estimated at 80,000 starting cells per well, or 100 µL of 800,000 cells mL^{-1}.

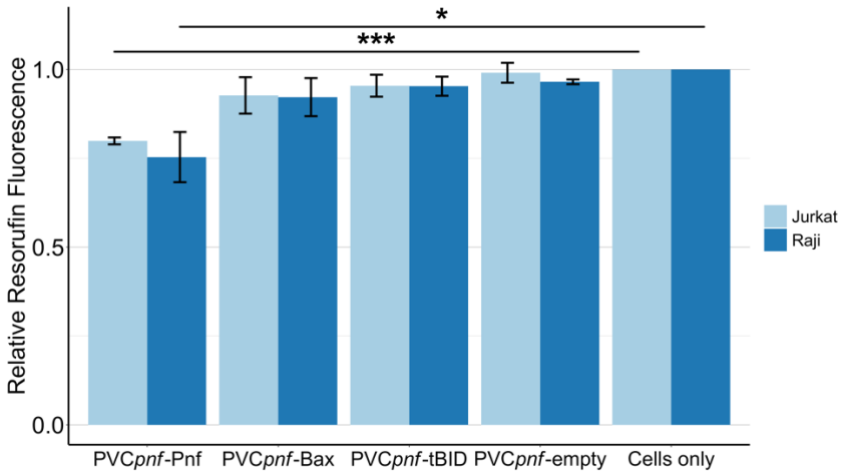


Figure 4.7 | PVC*pnf* perturbs respiration of mammalian immune cell lines.

Jurkat T cells and Raji B cells were incubated in a plate reader inoculated with purified PVC*pnf*-payload constructs; resorufin fluorescence was measured at 6 h after inoculation. A Student's T test was conducted using R (n = 3 biological replicates) with a group-wise p-value adjustment performed by dividing the α value (0.05) by the number of tested cell types (α_{adj} = 0.025). * = p < 0.025, ** = p < 0.005, *** = p < 0.0005. PVC*pnf* loaded with native Pnf (PVC*pnf*-Pnf) significantly reduced the level of resazurin reduction from decreased mitochondrial activity (Jurkat: t(4) = -16.36, p-value = 8.17e-05 ***; Raji: t(4) = -4.63, p-value = 0.0010 *. Empty PVC*pnf* had no significant effect on cell respiration for either cell type, nor did the pro-apoptotic payloads tBID and BAX. Error bars indicate standard deviations relativised to the mean values of the respective 'Cells only' fluorescences. One-way analysis of variance (ANOVA) indicated that cell type did not significantly affect relative respiration levels (F(1) = 3.07, p-value = 0.150).

4.5 Discussion

This chapter has demonstrated that the PVC payload that is associated / loaded into the tube lumen is modular when encoded C-terminal to the leader sequence (LS) which is native to the chosen operon: in this case, PVC*pnf*. However, there is nuance when attempting to observe the functionality of the payload after injection into a cell line. In particular, PVC*pnf* encoding wild-type Pvc13 tail fibres injects payload at low levels which means fluorescent readouts and cytotoxic effects, especially apoptotic ones, can be difficult to obtain.

4.5.1 Heterologous Payloads

The continued lack of readouts via the Pnf50-LacZα-Myc system could indicate a number of things: i) the payload abrogates injection via the PVC, ii) the payload fails to refold after injection, or iii) the LSs act as a potent signal of degradation upon entering eukaryotic cells. The answer to why the Pnf50-GFP-Myc system was not fluorescent was found by Jiang and colleagues (2022) where Pvc15 was found to unfold GFP in the *E. coli* cytosol. It is possible, therefore, that low levels of GFP injection did take place in this thesis. Future work should aim to characterise whether GFP becomes refolded after injection into the target cell and using the modified tail fibres from the work by Kreitz and colleagues (2023) should be used to ensure adequate levels of targeted injection.

4.5.2 Verifying Injection by Observing Cellular Effects

As shown by recent data, PVCs can be artificially targeted to certain cell receptors to ensure optimal payload injection and cell type specificity; in this case, Pnf is cytotoxic to the target cells (Kreitz *et al.*, 2023). However, some background binding to non-targeted cells is also possible as demonstrated in this thesis. Since the PVCs often failed to kill the incubated cells, cell health and

viability was measured using the reduction of resazurin as an indicator of effector potency after injection (see Materials and Methods chapter 2.3.4). Resazurin is irreversibly reduced by mitochondrial diaphorase enzymes to the fluorescent resorufin as a measure of cell viability (Talbot *et al.*, 2008), but can also be more heavily reduced to the non-fluorescent dihydroresorufin if cell density is too high, as demonstrated by the high starting cell numbers of S2 cells (Figure 4.5).

The reduction of resazurin was used to assess relative levels of mitochondrial respiratory activity. Jurkat cells have been shown to undergo mitochondrial disintegration, ROS production, and cell death with high levels of resazurin (>22 μM after 24 h) (Erikstein *et al.*, 2010) which could skew results. Thus, a low concentration of 5 μM was used to ensure resazurin did not interfere with the health of the cells over the 24 h time course (Erikstein *et al.*, 2010).

Truncated BID (tBID) and BAX did not have a similar effect on cell respiration as that of Pnf. Actin microfilaments act as reservoirs for Drp1: a protein which drives mitochondrial fission events (Hatch *et al.*, 2016); microfilaments also act to regulate mitochondrial trafficking and quality control (Kim and Cheong, 2020). Crucially, Guzun and colleagues (2009) demonstrated that the cytoskeleton directly affects mitochondrial energy and anionic fluxes by its interaction with mitochondrial outer membrane proteins such as voltage-dependent anion channels (VDAC). Mitochondrial creatine kinase (MtCK) also localises to actomyosin fibres and limits VDAC permeability; cells overexpressing MtCK exhibit a reduction in ATP and ADP levels (Guzun *et al.*, 2009). Altogether, Pnf-induced stress fibre formation and maintenance through RhoA deamidation (Vlisidou *et al.*, 2019) may have a more potent effect on the mitochondrion when injected at low levels than individual molecules of BAX and tBID which act at the mitochondrion directly (Guzun *et al.*, 2009). As such, higher levels of BAX and tBID may be required for cytotoxicity which could be

achieved by modifying the tail fibres to target the cells with BAX and tBID-loaded PVCs.

Actin microfilaments could not be observed after S2 and HEK293T cell incubation with PVC*pnf*-Pnf. Similarly to the lack of readouts for other heterologous payloads, it may be that 'background' injection capacity was too low to observe the effect on actomyosin maintenance from the payload. Kreitz and colleagues (2023) showed that J774A.1 macrophages are susceptible to injection by PVCs either possessing or lacking wild-type tail fibres; future experiments should aim to test actomyosin filamentation using this cell line and phalloidin fluorescent dye conjugates to verify Pnf injection and function.

4.5.3 PVC Production and 'Regulatory' Genes

In contrast to the findings of Wang, Cheng, and colleagues (2021), the LysR transcriptional regulator (PAU_RS16560) was not required for the production of intact PVCs nor for the expression of the payload Pnf since PAU_RS16560 was not included in either of the pVTRa or pBAD-PVC*pnf* vectors. In addition, Kreitz and colleagues (2023) showed that inclusion of the 'regulatory genes' found downstream of the PVC*pnf* operon (PAU_RS16560-70 + PAU_RS24015) resulted in a 100-fold increase in PVC yield. In their work, reverse transcriptase quantitative PCR (RT-qPCR) revealed that mRNA levels of structural PVC components were not affected by deleting the regulatory genes, implying a role for these proteins in PVC assembly and not transcriptional regulation. Given that intact PVCs were produced in the data presented in this thesis, the so-called 'regulatory' genes appear dispensable for production of wild-type PVC*pnf*.

5. Uncovering Pvc15-Leader Sequence Interactions

5.1 Introduction

Jiang and colleagues (2022) first identified the Pvc15-leader sequence (LS) interaction through its capacity to unfold LS-associated mRFP, and as a requirement to load Pnf into PVC*pnf*. Given the first structural characterisation of Pvc15 in this thesis, it was desirable to test the hypotheses laid out in previous chapters. Notably, Pvc15 has both sequence and structural homology to AAA+ ATPases, though its activity has never been documented. Additionally, a more in-depth exploration of its interaction with the Pnf LS would be ideal. Jiang and colleagues also found that mutations in Pvc15 did not abolish its unfolding activity, which suggests that the protein may possess multiple roles.

5.2 Aims

i) Explore the role of Pvc15 in the Pnf LS interaction in *E. coli* lysates. Explore whether such interactions are also conserved for other LSs such as Cif50.
ii) Define the parameters for Pvc15-mediated PVC*pnf*-Pnf association by comparing the wild-type system with truncated LSs and Pvc15 mutants.
iii) Assay Pvc15 for its ATPase activity and infer the relevance of wild-type Pvc15 in PVC assembly and payload loading.

5.3 Methods

5.3.1 Bacterial Lysate Protein Abundance and Degradation

An *E. coli* expression strain, such as NiCo21, was grown in a starter culture overnight at 37 °C. Cells were diluted 1:100 in 50 mL of LB with the relevant antibiotics and grown to OD_{600} = 1.0 at which point cells were induced with IPTG (0.5 mM) and/or arabinose (0.2%). After the optimal time of protein expression, determined earlier using western blot (WB) analysis, culture density was assessed once again and added to 2x LDS + 50 mM DTT to total 50 μL volume such that OD_{600} in the solution would be 0.5. Samples were boiled for 20 minutes before running on SDS-PAGE gel.

For assessing protein degradation, cultures were, instead, centrifuged at 2,000 g for 10 minutes to remove IPTG and arabinose and pellets were resuspended in the same volume of LB + 20 μg/mL Kanamycin to halt translation. Aliquots could then be taken at regular intervals to assay protein abundance over time, ensuring the same OD_{600} at each time point.

5.3.2 ATPase assay

The colorimetric ATPase assay kit (Abcam, cat. no.: ab234055) was used to assess Pvc15 ATPase activity in cell lysates as per the manufacturer's instructions. 15 mL of *E. coli* NiCo21 cultures grown to OD_{600} = 1.0 were induced with 0.2% arabinose and 0.5 mM IPTG to co-express PVC*pnf*Δ*pvc15* and a pVTRa vector encoding the LS-associated payload with or without Pvc15 for 16 h then processed according to the manufacturer's instructions.

For assaying, 30 μL ATPase assay developer was added to every well on the plate, including standards, in quick succession. OD_{650} was measured at the start and after incubation at room temperature for 30 mins. For analysis, a standard curve was plotted: phosphate amount (nmol) against OD_{650} and fitted with a

linear regression. After adjusting for spontaneous changes in the malachite green probe over the time course, each sample background was subtracted from their corresponding reaction after adjusting each by subtracting the maximum corresponding change in OD_{650} of either the background or reaction mix reagent controls:

$$Adjusted\ \Delta OD_{rxn} = \left(\Delta OD_{rxn} - max\ (Control\ \Delta OD_{rxn})\right) - \left(\Delta OD_{background} - max\left(Control\ \Delta OD_{background}\right)\right)$$

Phosphate amount generated, B, could be found using the standard curve for each mean change in adjusted OD_{650} with the intercept of the curve adjusted to 0 by subtracting the '0' standard reading from all standard and background readings. After converting OD_{650} to phosphate amount by dividing the adjusted ΔOD_{rxn} by the slope, ATPase activity was calculated as per the manufacturer's instructions:

$$ATPase\ activity\ (U/mL) = \frac{B}{(30 * V)}$$

where B is the phosphate amount generated and V is the sample volume initially added to the wells (5 or 20 μL), and where one ATPase-specific unit, U, represents the amount of enzyme that generates 1 μmol of phosphate per min at pH 7.5 and 25 °C.

5.4 Results

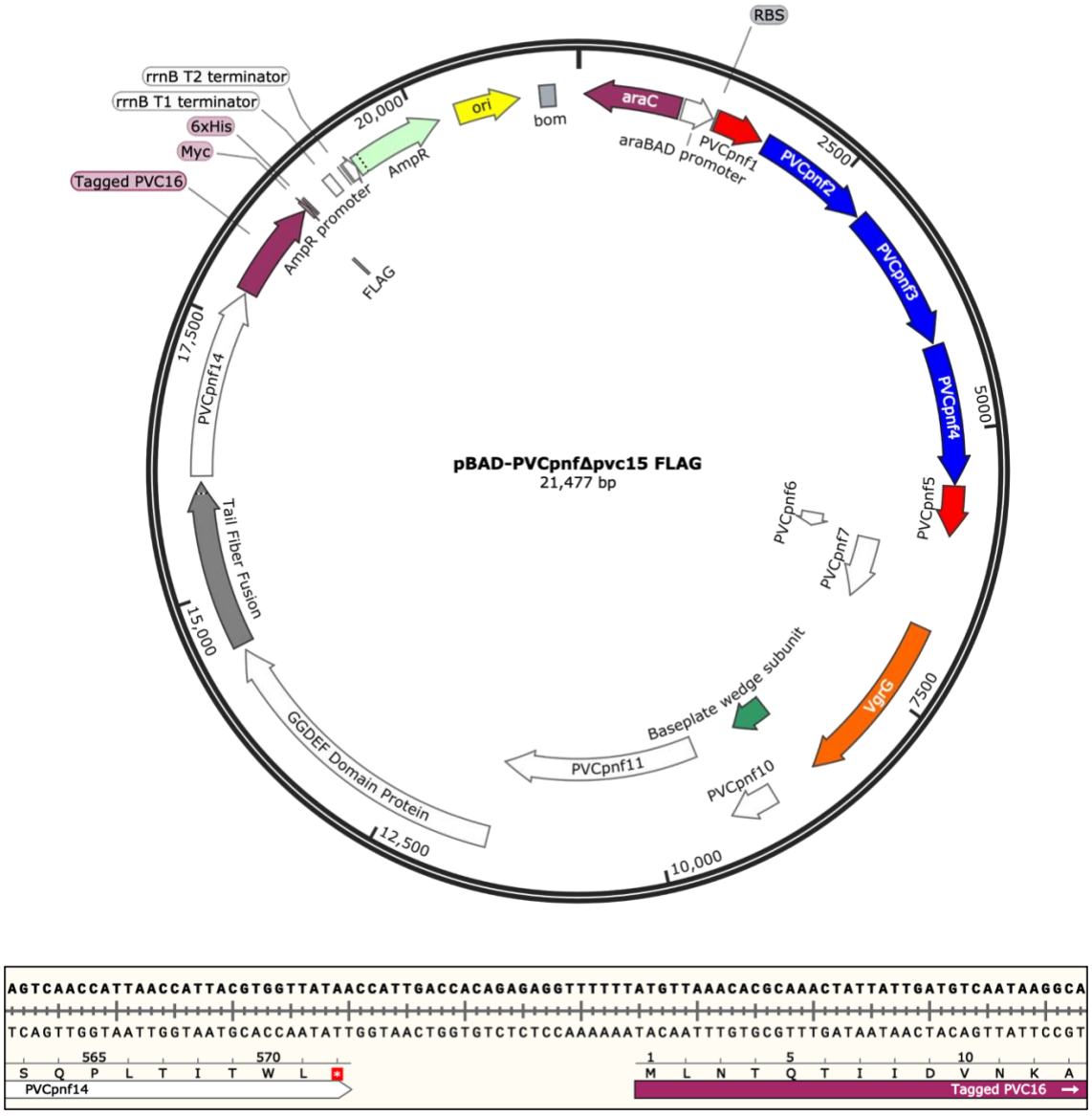


Figure 5.1 | Whole-plasmid sequencing confirms the PVC*pnf*Δ*pvc15* vector.

Oxford Nanopore sequencing obtained a consensus sequence and confirmed each gene in the plasmid was correctly encoded. As expected, the *pvc15* gene was absent in the vector and the 5' UTR of *pvc16* was retained.

To uncover the role of Pvc15 in PVC*pnf* production, a Pvc15 knock-out operon (PVC*pnf*Δ*pvc15*) was constructed using overlap extension PCR and was confirmed by sequencing of the full plasmid (Figure 5.1). Assessing the

expression of double-6x-His-tagged Pvc15 from the pVTRaDuet vector was difficult as only faint bands were observed 4 h after induction using western blot (WB) (Figure 5.2A). Various concentrations of IPTG for induction indicated uniform levels of expression of both the FLAG-tagged Pvc16 and the full-length Myc-tagged Pnf toxin (Figure 5.2B; left panel).

Western blot analysis of eluates indicated good levels of Pvc16-FLAG as a marker for PVC*pnf* purification, given that it is the final protein associated with the mature complex (Figure 5.2B; middle gel, bottom). In the absence of PVC expression, slight background levels of Pnf were detected in eluates (Figure 5.2B; middle gel, lane 6). However, Pnf was absent after more thoroughly purifying PVC*pnf*Δ*pvc15* samples (Figure 5.2B; right panel); Jiang and colleagues (2022) also noted that Pvc15 is "essential" for loading. Interestingly, it was found that Pnf was depleted in the supernatant of cells lacking Pvc15 (middle gel; lane 4). PVC*pnf* and PVC*pnf*Δ*pvc15* syringes did not appear to differ in appearance when compared in negative stain transmission electron microscopy (TEM) (Figure 5.2B; right panel).

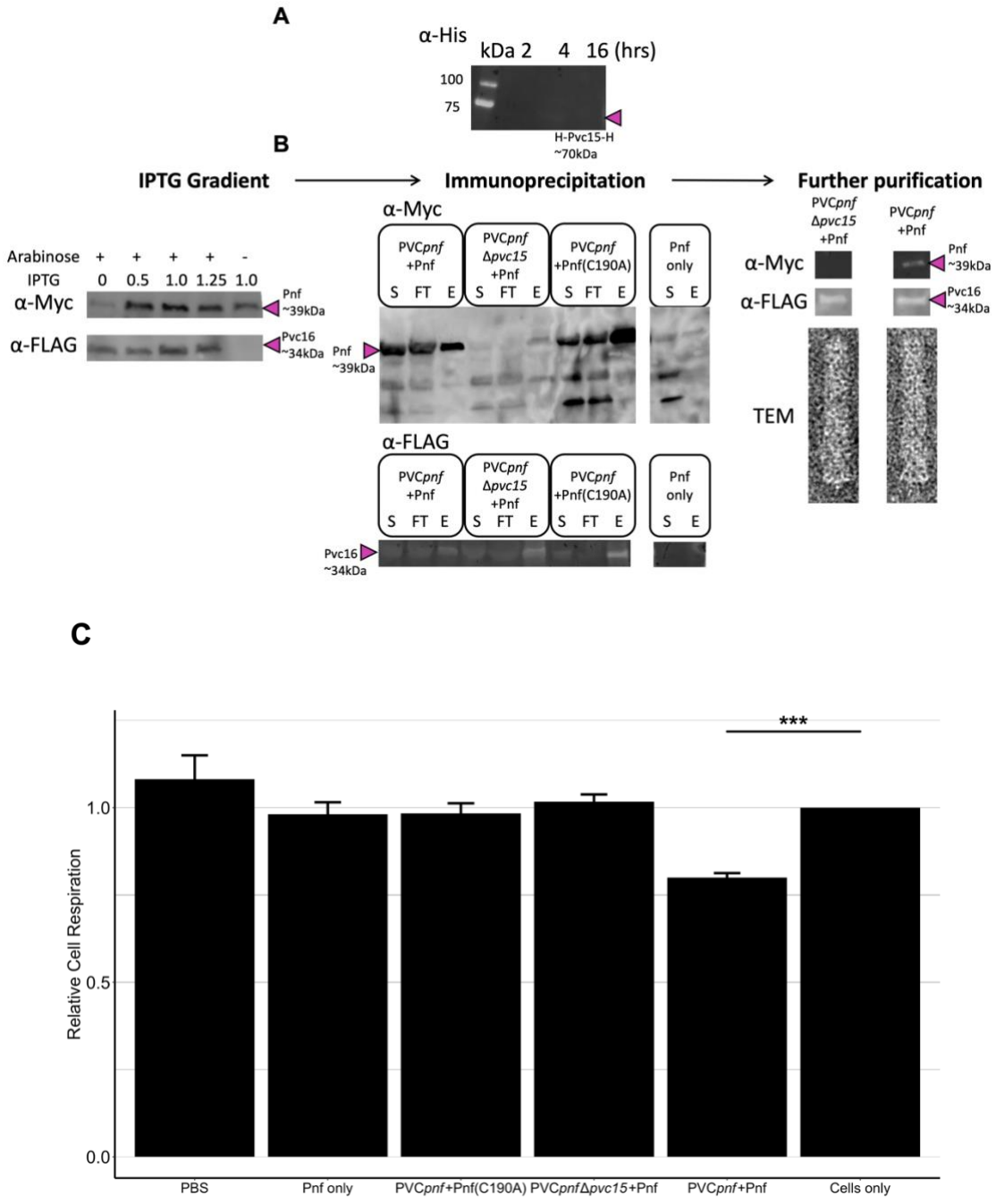


Figure 5.2 | Pvc15 is required for loading and functional injection of the cognate payload.

(A) α-His primary antibody WB for detection of double hexahistidine tagged Pvc15 from *E. coli* lysates of pVTRaDuet-6xHis-Pvc15-6xHis-Pnf50-Pnf(C190A)-Myc expressed over a time course. **(B)** Left panel: a gradient of IPTG induction was used to find the optimal concentration of IPTG for PVC*pnf*-Pnf induction from pVTRa-Pnf50-Pnf-Myc, including a control without arabinose to observe that the payload can be induced independently of the PVC. Some leakage of Pnf expression is seen when no IPTG is added.

Middle panel: Immunoprecipitation of PVC*pnf* (S = supernatant; FT = flow-through after resin binding; E = elution after addition of competitive elution buffer). Pnf co-purifies only when Pvc15 is expressed. No significant differences were observed when purifying PVC*pnf* loaded with Pnf(C190A) to that of the native Pnf.

Right panel: further purification of PVC*pnf* removed any detected 'background' payload.

(C) Raji cells were incubated overnight with 10 µg of each preparation. Resazurin was added to the 96-well plate and incubated in a plate reader at 37 °C and 5% CO_2. Samples were compared 4 h after the addition of PVC*pnf* and relativised to the "Cells only" control. Statistical significance was measured using a Student's T-test; error bars show standard deviation; n = 4 biological replicates of separately cultured cells. The native PVC*pnf*-Pnf sample was found to impede the reduction of resazurin (t(6) = -9.95, p-value = 5.569e-05 ***) whilst other samples did not affect cell viability, including PVC*pnf* purified from lysates expressing pBAD-PVC*pnf* and pVTRa-Pnf50-Pnf(C190A)-Myc which contained Pnf with a C190A active site mutation. PVC*pnf* purified from lysates expressing pVTRa-Pnf50-Pnf-Myc and pBAD-PVC*pnf*Δ*pvc15* did not affect cell viability. In addition, PVC*pnf* was shown to be important for reduction of cell viability since Pnf-Myc alone (from lysates expressing pVTRa-Pnf50-Pnf-Myc only), purified using a Ni^{+}-NTA resin, did not affect cells without first being loaded into the PVC*pnf* chassis.

To further affirm whether Pnf was still associated with the PVC in such a manner for successful injection, a resazurin reduction assay was conducted in Raji B cells (Figure 5.2C). Indeed, cells incubated with PVC*pnf*Δ*pvc15* purified after co-expression with Pnf were found to have similar levels of resorufin fluorescence as that of the PBS and "Cells only" controls (Figure 5.2C). This suggests that Pvc15 was required for the effective loading and delivery of functional payloads into target cells. Relative respiration for Raji cells alone was distinctly higher than that of the native PVC*pnf*-Pnf sample (t(6) = -9.95, p-value = 5.569e-05 ***). Unfortunately, Pnf could not be purified using an α-Myc resin even after repeated attempts and various protocols. However, the Ni^{+}-NTA resin did appear to purify Pnf despite that no obvious positively charged regions were identified in the protein. Nonetheless, native Pnf did not affect

cells on its own and required PVC*pnf* co-expression; alanine-mutation of Pnf's catalytic cysteine also abolished its toxic activity.

5.4.1 Effects of Leader Truncation on Pnf Payload Stability

Pvc15$_{pnf}$ (referred to as simply Pvc15 for this chapter, unless specified) seemed to stabilise Pnf in the supernatant during PVC*pnf* purification (Figure 5.2B). Jiang and colleagues (2022) experimented with Pnf LSs, termed "signal peptides" (SPs) in their work. The authors found that deletion of the first 70 amino acids of the Pnf LS abrogates its production to undetectable levels, whilst deletion of the first 50 only decreases its abundance. Therefore, it was hypothesised that the N-terminal tip and LS length may play a role in stabilising the associated 'payload' protein, such as via an interaction with Pvc15.

To first ascertain the optimal time to measure protein expression after induction in NiCo21 cells, a time course was done using aliquots at T=0, 2, 4, 6, and 16 h after induction. Samples were normalised to the same OD_{600} and through a Coomassie blue stain; Pnf-Myc was measured on an α-Myc WB and quantified as in chapter 2.5.1.3.1 on page 105. The calibration indicated 2-4 h post-induction had the highest levels of protein expression (Figure 5.3). Thus, subsequent experiments on payload stabilisation were conducted at 3 h post-induction with IPTG and arabinose. Myc bands at ~17 and 25 kDa were observed as with other experiments where Pnf was expressed; these may have been degraded/cleaved Pnf, similar to those found by Jiang and colleagues (2022).

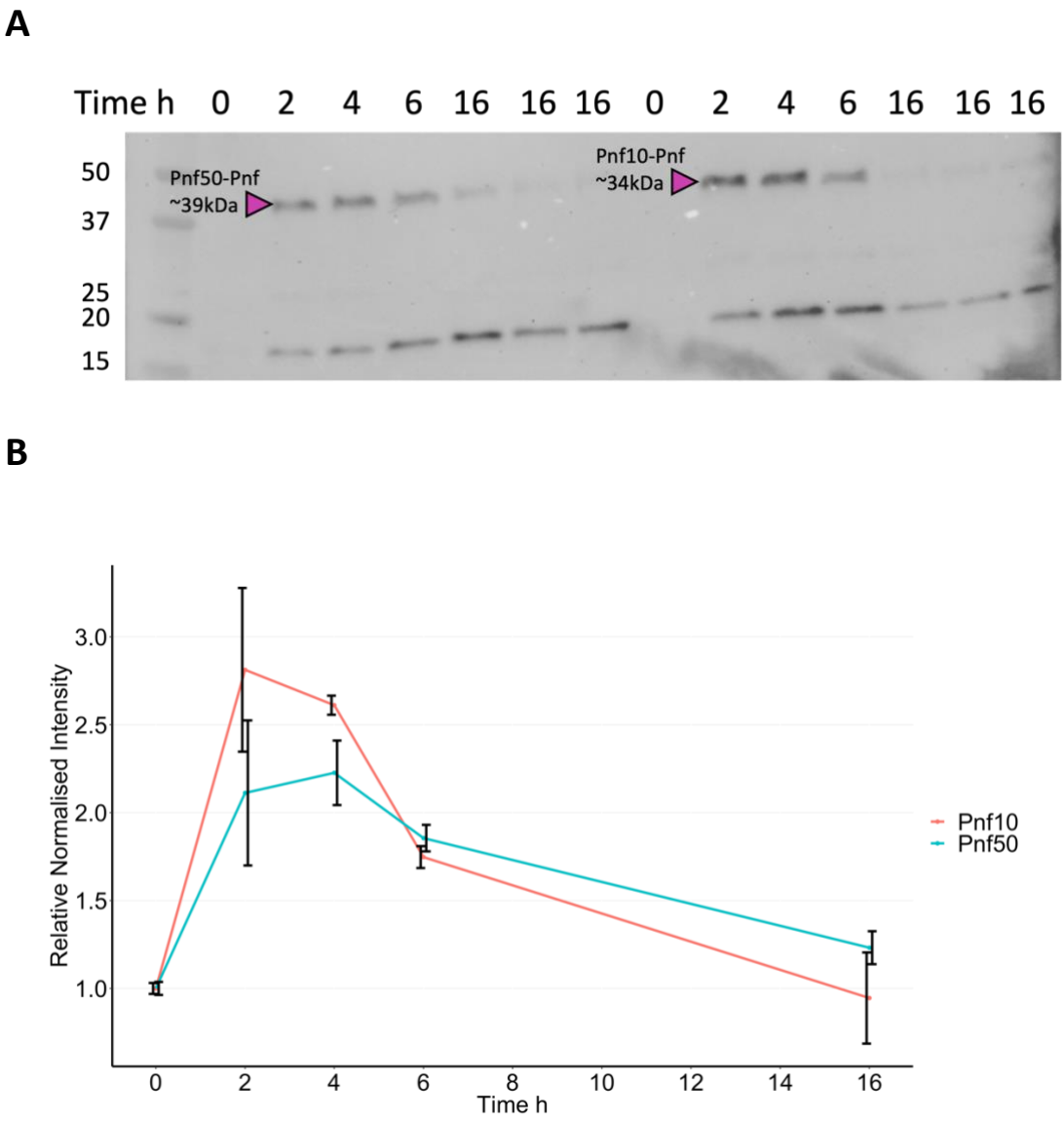


Figure 5.3 | Time course of Pnf abundance post-induction.

(A) α-Myc primary antibodies were used to obtain the relative band intensities (normalised by Coomassie blue stain) of Pnf-Myc after induction with IPTG via western blot (WB). *E. coli* containing both pBAD-PVC*pnf* and pVTRa-Pnf-Myc with either the full-length LS (Pnf50) or the C-terminal truncated LS (Pnf10). Bands at 15-20 kDa are uncharacterised putative Pnf-Myc degradants. **(B)** 3-4 h was chosen as the optimal time point to measure Pnf abundance. Error bars represent standard deviations for three technical replicates.

Pnf(C190A) and PVC*pnf* were induced in the presence or absence of Pvc15 and normalised using the FLAG-tagged Pvc16 in WB (Figure 5.4A). Normalised band intensities of lysates 3 h post-induction with IPTG and arabinose on WB

showed a significant decrease in the abundance of Pnf when Pvc15 was deleted (t(4) = -5.55, p = 0.0052 **). Truncation of the LS from its C-terminus (i.e., retaining its original 10 amino acids at the LS 'tip') (Pnf10) resulted in a greater abundance of Pnf compared to the native Pnf50 LS, regardless of the presence of Pvc15 (with Pvc15: t(4) = 3.30, p-value = 0.030 *; without Pvc15: t(4) = 4.26, p-value = 0.013 *).

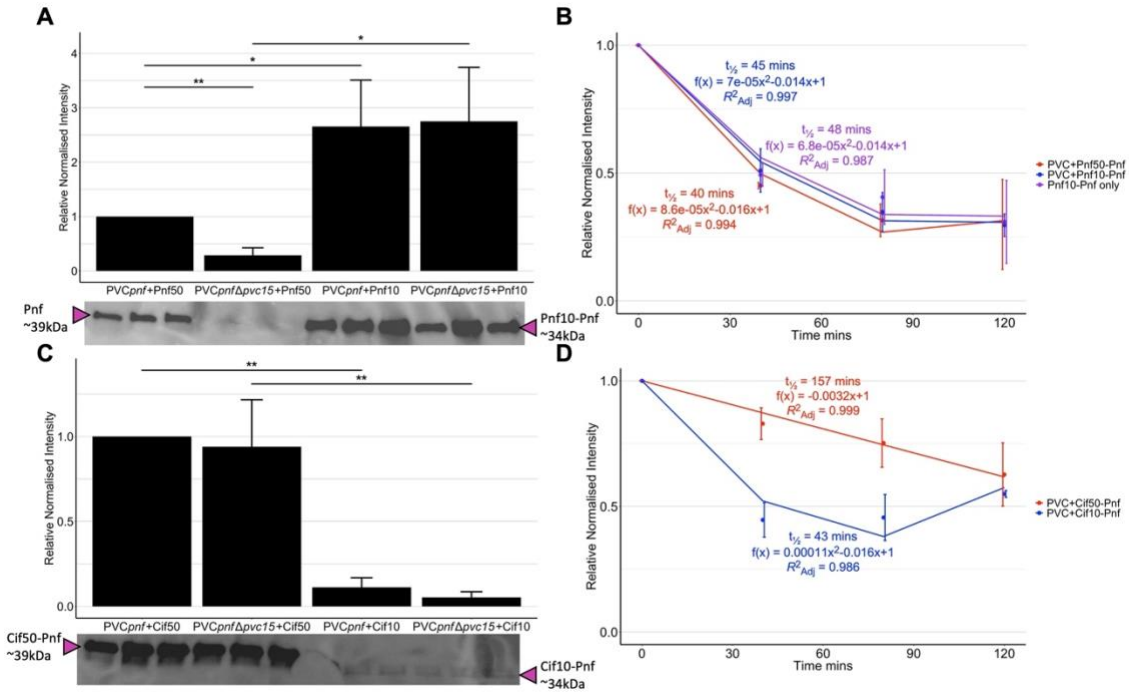


Figure 5.4 | LS C-terminal truncations differentially affect the stability of the payload.

E. coli lysates from cells induced with IPTG and arabinose for 3 h were aliquoted. Protein abundance was quantified using α-Myc primary antibodies in western blot (WB). Bands were normalised either by the mean of lanes on a Coomassie blue stain or by fixed point via an α-β-RNA polymerase primary antibody.

(A) Cells co-expressed pBAD-PVC*pnf* or pBAD-PVC*pnf*Δ*pvc15* with pVTRa-Pnf50-Pnf-Myc (LS = MLKYANPQTVAKAQRTKNTAKKPPSSTSFDGHLELSNGENQPYEGHKIRKI) or pVTRa-Pnf10-Pnf-Myc (LS = MLKYANPQTV). Truncation of the LS resulted in greater prevalence of Pnf-Myc in lysates when co-expressed with pBAD-PVC*pnf* (t(4) = 3.30, unadjusted p-value = 0.030 *) as well as with pBAD-PVC*pnf*Δ*pvc15* (t(4) = 4.26, unadjusted p-value = 0.013 *). The Pnf50 LS co-expressed with pBAD-PVC*pnf*Δ*pvc15* results in a dramatic decrease to the prevalence of the payload compared to with pBAD-PVC*pnf* (t(4) = -5.55, p = 0.005 **); n = 3 biological replicates

of separately cultured and induced cells. Each sample was done in biological triplicates and normalised by sum of lanes on a Coomassie stain. * = $p < 0.05$; ** = $p < 0.01$; Student's T-test.

(B) A time course experiment after induction of Pnf with 0.5 mM IPTG and PVC*pnf* with either 0.2% arabinose or further repressed using 0.5% glucose. After 3 h of induction, these components were removed by centrifugation and cells were resuspended in LB with 20 μg/mL kanamycin to halt protein synthesis. Aliquots were analysed by WB and normalised by the intensity of *E. coli* β-RNAP. Linear quadratic regressions indicate similar coefficients and half-lives for each lysate of 40-48 mins. Adjusted R^2 was consistently greater than 0.98.

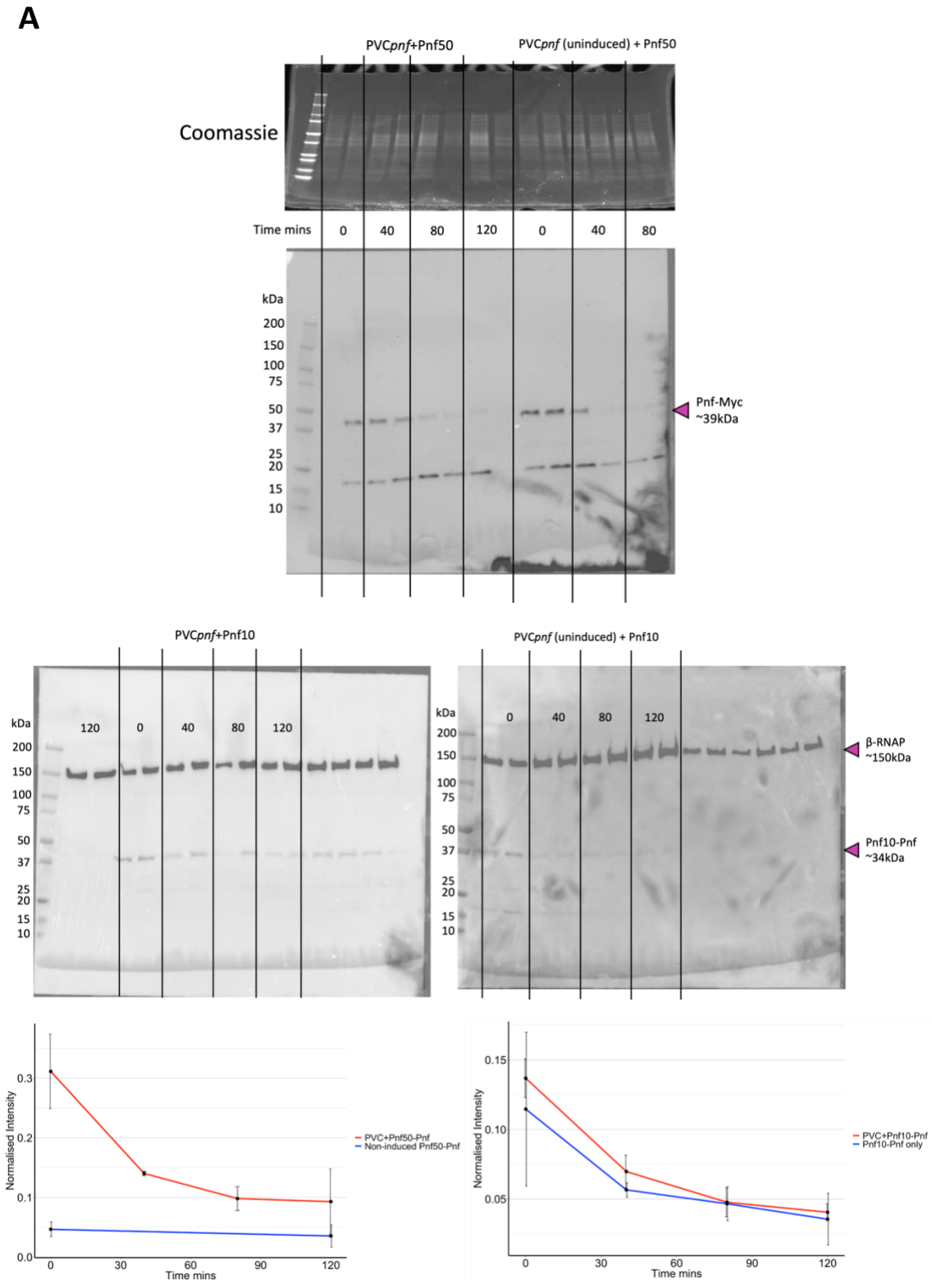
(C) pBAD-PVC*pnf* or pBAD-PVC*pnf*Δ*pvc15* were co-expressed with pVTRa-Cif50-Pnf-Myc (LS = MREYSKEDDCVKEKTNLAESENVEADNYLEMDCLNYLAKLNGMPERKDHS) or pVTRa-Cif10-Pnf-Myc (LS = MREYSKEDD). Truncation of the Cif LS led to a decrease in abundance of the Pnf cargo when co-expressed with both pBAD-PVC*pnf* ($t(3) = -11.86$, $p = 0.0013$ **) and pBAD-*PVCpnfΔpvc15* ($t(4) = -5.51$, $p = 0.005$ **). n = 3 biological replicates of separately cultured and induced cells; one outlier was excluded.

(D) A time course experiment after induction of Pnf with 0.5 mM IPTG and PVC*pnf* with either 0.2% arabinose or further repressed using 0.5% glucose. After 3 h of induction, these components were removed by centrifugation and cells were resuspended in LB with 20 μg/mL kanamycin to halt protein synthesis. Aliquots were analysed by WB and normalised by the intensity of *E. coli* β-RNAP. Coefficients were significantly different: the Cif10-Pnf-Myc quadratic coefficient was 4.18 σ from Cif50-Pnf-Myc ***, and the Cif10-Pnf-Myc linear coefficient was -4.1 σ from Cif50-Pnf-Myc ***. Half-lives differed by almost 4-fold.

To test whether the difference in band intensity between the different LS lengths was due to increased stability or expression, IPTG was removed by centrifugation and 20 μg/mL of kanamycin was added to stop protein synthesis and measure Pnf band intensities over a time course (Figure 5.4B). Non-linear regression models indicated similar quadratic coefficients and half-lives across each of the lysates tested (40-48 minutes). This indicated that the stabilities of Pvc15-stabilised Pnf50-Pnf and standalone Pnf10-Pnf were not dissimilar over the 2 h after kanamycin addition. Further interpretations are discussed in Discussion chapter 5.5.3 on page 224.

Conversely, Pvc15 had seemingly no effect on Pnf encoded with a Cif LS (Cif50 and Cif10), though the C-terminal truncated Cif LS displayed a decrease to the stability of the payload both with ($t(3) = -11.86$, $p = 0.0013$ **) and without Pvc15 ($t(4) = -5.51$, $p = 0.005$ **) (Figure 5.4C). Degradation profiles between Cif50 and the C-terminal truncated Cif10 were vastly different: the Cif50 LS followed a linear regression with a half-life of 157 minutes, whilst Cif10 followed a quadratic regression with a half-life of 43 minutes.

Thus, Pvc15 encoded by the PVC*pnf* operon interacts differently between the different PVC payload LSs and each LS differentially mediates the payload's stability and takes precedence over Pvc15-mediated stabilisation. In addition, the difference in band intensity observed between Pnf10 and Cif10 indicates that the 10 amino acid tip is an important variable for determining payload stability. In contrast, the tip appears to be dispensable for Pvc15-dependent stabilisation since only Pnf50 displayed a differential effect with or without Pvc15 as a chaperone. This implies that the Pnf10 tip may confer an innate level of stability whilst some of the amino acids 11-50 are required for the Pvc15-dependent stability. Data from WB for the degradation experiments are shown in Figure 5.5.


A
PVCpnf+Pnf50
PVCpnf (uninduced) + Pnf50
Coomassie
Time mins
0
40
80
120
0
40
80
kDa
200
150
100
75
50
37
25
20
15
10
Pnf-Myc
~39kDa
PVCpnf+Pnf10
PVCpnf (uninduced) + Pnf10
120
0
40
80
120
0
40
80
120
β-RNAP
~150kDa
Pnf10-Pnf
~34kDa
Normalised Intensity
0.3
0.2
0.1
0.0
0
30
60
90
120
Time mins
PVC+Pnf50-Pnf
Non-induced Pnf50-Pnf
0.15
0.10
0.05
PVC+Pnf10-Pnf
Pnf10-Pnf only

B

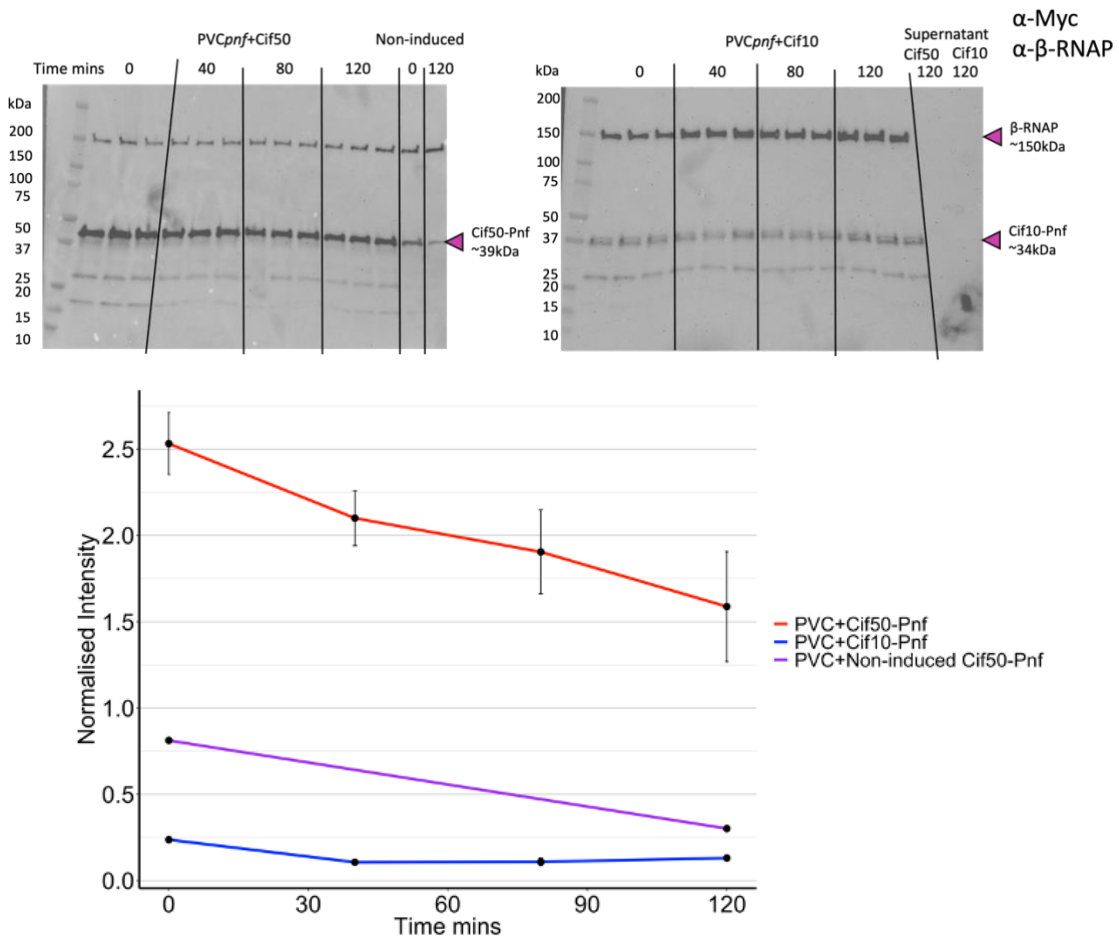


Figure 5.5 | Collected data for Pnf stability and degradation experiments.

WBs were stained with the appropriate primary antibodies and recorded for the data presented above for the **(A)** Pnf LS and **(B)** Cif LS. Note that 'Pnf50 only' used a strain containing but not inducing PVC*pnf*; the presence of Pnf at 0 hrs here but not in the previous figure with PVC*pnf*Δ*pvc15* suggests that the payload was stabilised by the low levels of Pvc15 expressed by leakage from the pBAD-PVC*pnf* plasmid.

Having tested C-terminal truncations of the Pnf LS, it was desirable to also test whether N-terminal truncations of Pnf50 had similar effects. Using a straightforward restriction ligation workflow splicing into the standard pVTRa vector, BL21(DE3) strains expressed N-terminal truncations of 24 and 39 amino acids of the Pnf LS (N25 and N40) with or without the PVC*pnf* operon. Co-

transformation with the PVC*pnf*Δ*pvc15* operon plasmid was unsuccessful so these truncations could not be tested in the context of Pvc15, specifically.

Strains were induced for 3 h and cell lysates were assessed by α-β-RNAP + α-Myc WB, as done with the C-terminal truncated LSs. In contrast to the C-terminal Pnf50 truncations, N-terminal truncated Pnf LSs PnfN25 and PnfN40 were more abundant than that of Pnf50 only when co-expressed with the PVC*pnf* operon (Figure 5.6). Thus, N-terminal truncations appeared to have a PVC*pnf*-dependent stabilising effect. The lack of innate stability lends further evidence to the potential of the N-terminal 'tip' of the Pnf LS to serve an innately stabilising role to Pnf, but which is far less stable if trailed by a sufficiently long LS.

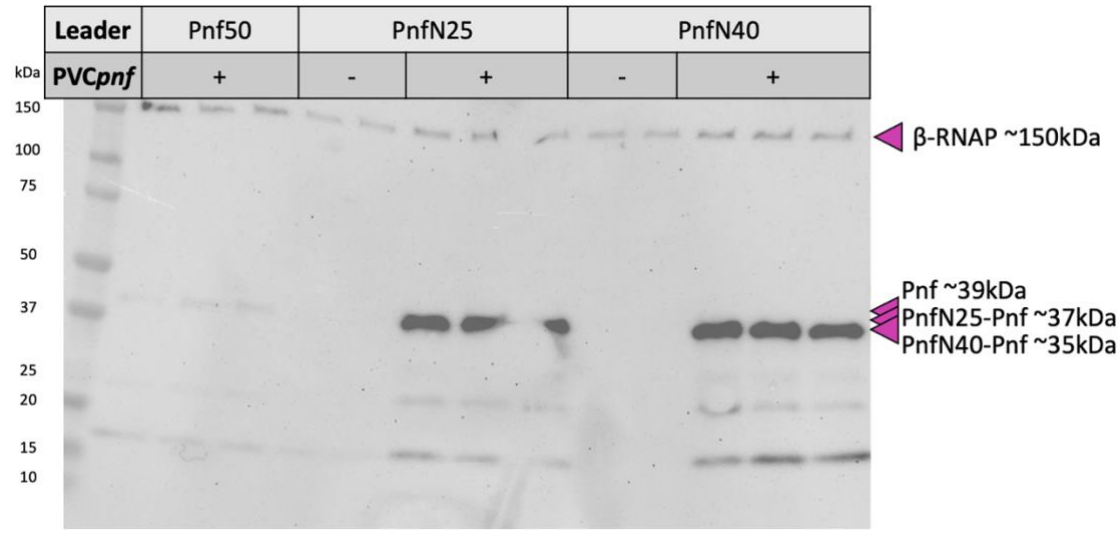


Figure 5.6 | N-terminal Pnf50 truncations enhance Pnf abundance when co-expressed with PVC*pnf*.

Western blot using α-Myc and α-β-RNAP primary antibodies. Lysates of *E. coli* BL21(DE3) strains expressing pVTRa-Pnf50-Pnf-Myc, pVTRa-PnfN25-Pnf-Myc, or pVTRa-PnfN40-Pnf-Myc were aliquoted 3 h after induction. PnfN25 is an LS lacking the first 24 amino acids and PnfN40 is an LS lacking the first 39 amino acids.

To test whether Pvc15 was the only component of PVC*pnf* which could act as a chaperone, a time course experiment was conducted using each of the

strains containing the native PVC*pnf* + Pnf system, PVC*pnf*Δ*pvc15* + Pnf, and Pnf alone. Aliquots were taken from replicate 50 mL cultures at 4, 8, and 20 h post-induction, normalised to OD_{600} = 1 as before, and normalised further by staining with α-β-RNAP primary antibody along with α-Myc and α-FLAG. Whilst Pvc16-FLAG was, indeed, expressed at higher levels at 8 and 20 h post-induction, Pnf-Myc was only found at high levels when Pvc15 was co-expressed at 3-4 h or 16-20 h stages (Figure 5.7); this lends to the hypothesis that Pvc15 is the sole chaperone protein for native Pnf in the PVC operon. Oddly, when co-expressed with Pvc15, neither of Pvc16 nor Pnf appeared to be present in any of the 8 h samples.

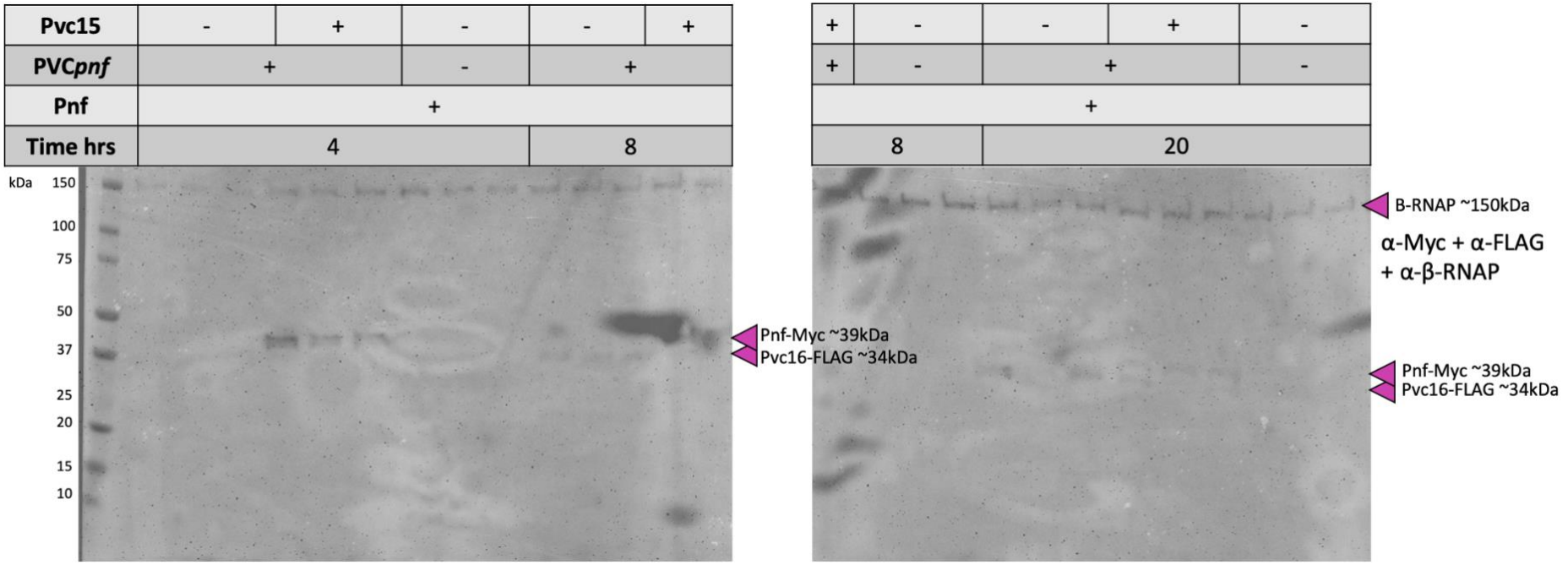


Figure 5.7 | Pvc15 is the sole PVC*pnf* chaperone protein for its native payload.

Aliquots for biological replicates (n=3) were taken from 50 mL cultures of *E. coli* BL21(DE3) at time points 4, 8, and 20 h post-induction with IPTG and arabinose. Strains encoding either i) pBAD-PVC*pnf* + pVTRa-Pnf50-Pnf-Myc, ii) pBAD-PVC*pnf*Δ*pvc15* + pVTRa-Pnf50-Pnf-Myc, or iii) pVTRa-Pnf50-Pnf-Myc alone were normalised to OD_{600} = 1 and WB was conducted using primary antibodies plus an α-β-RNAP antibody. α-Myc and α-FLAG primary antibodies were used to stain Pnf-Myc and Pvc16-FLAG, respectively.

5.4.2 PVC*pnf* Loading Depends on Pvc15 and the Leader Sequence

If native Pnf can be stabilised by encoding a shorter LS, presumably making it a less desirable target for protein degradation machinery, then the longer LS may be preferable for its role in loading. Thus, it was hypothesised that payloads encoding C-terminal truncated LSs would be far less capable of being loaded into the PVC*pnf* chassis; this, then, would justify the bacterium having evolved longer, unstable LSs in its native payloads to optimise loading whilst also encoding a nearby chaperone, Pvc15, to mitigate the accompanying instability.

To compare loading efficiencies of the different LS-truncated payloads, PVC*pnf* was purified using immunoprecipitation of a PVC structural protein, as before (Figure 5.8). α-Myc WB indicated a decrease in the relative co-purified levels of Pnf (Figure 5.8A) which was confirmed when each band was normalised to the mean band intensity of the corresponding α-FLAG (Pvc16) WB band (Figure 5.8B). Overall, either truncation of the LS or the deletion of Pvc15 resulted in a ~50-75% decrease in the loading efficiency of Pnf into PVC*pnf*. Since Figure 5.2 indicated that a further round of PVC purification removes background levels of payload in samples of purified PVC*pnf*Δ*pvc15*, the true drop in abundance of loaded nanosyringes may be greater still. Altogether, these data confirm the hypothesis that shorter LSs, whilst conferring stability to Pnf in the bacterium, fail to recruit payload to the PVC at comparable levels to that of the full-length LS.

Additional biological replicates for Pnf loading are included in Figure 5.16.

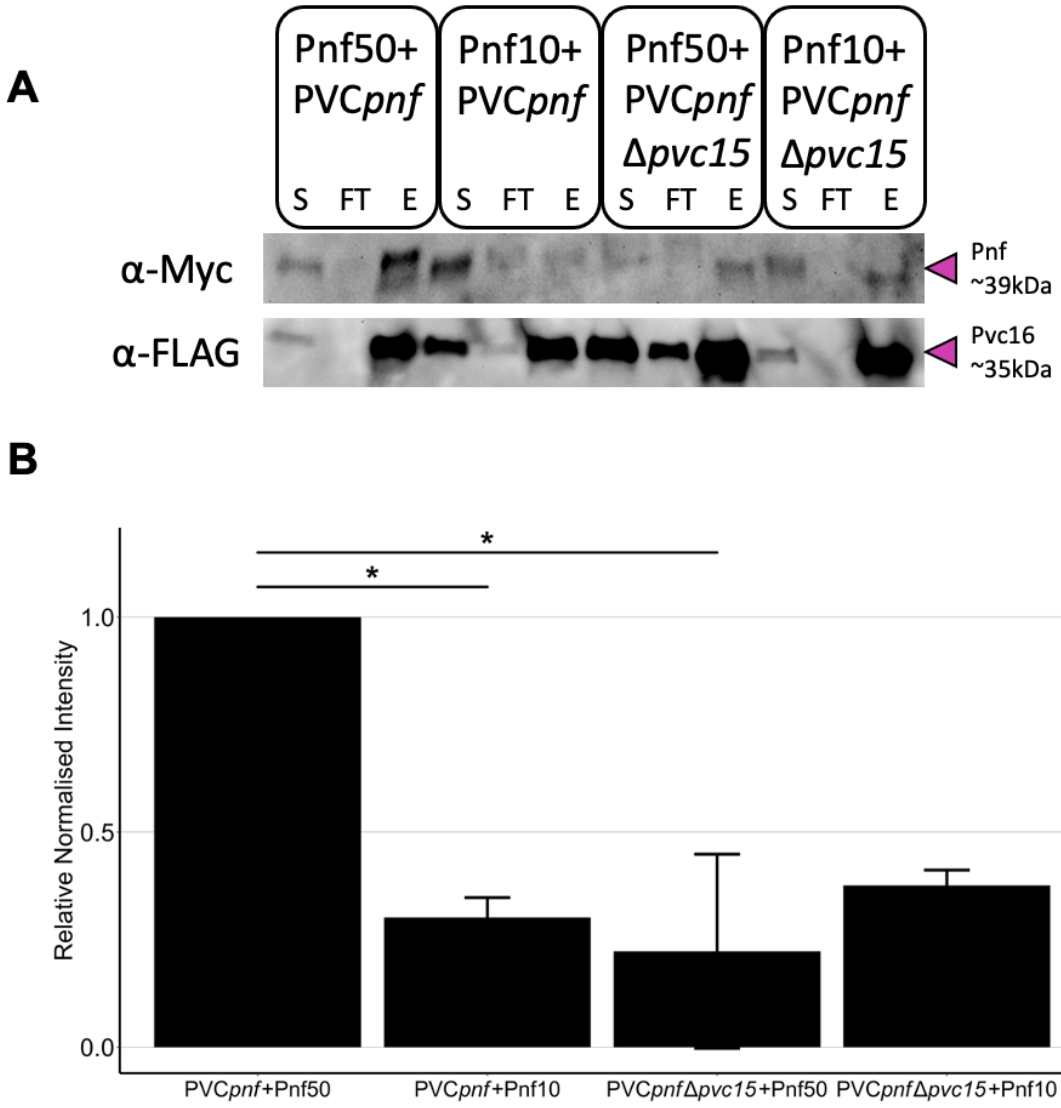
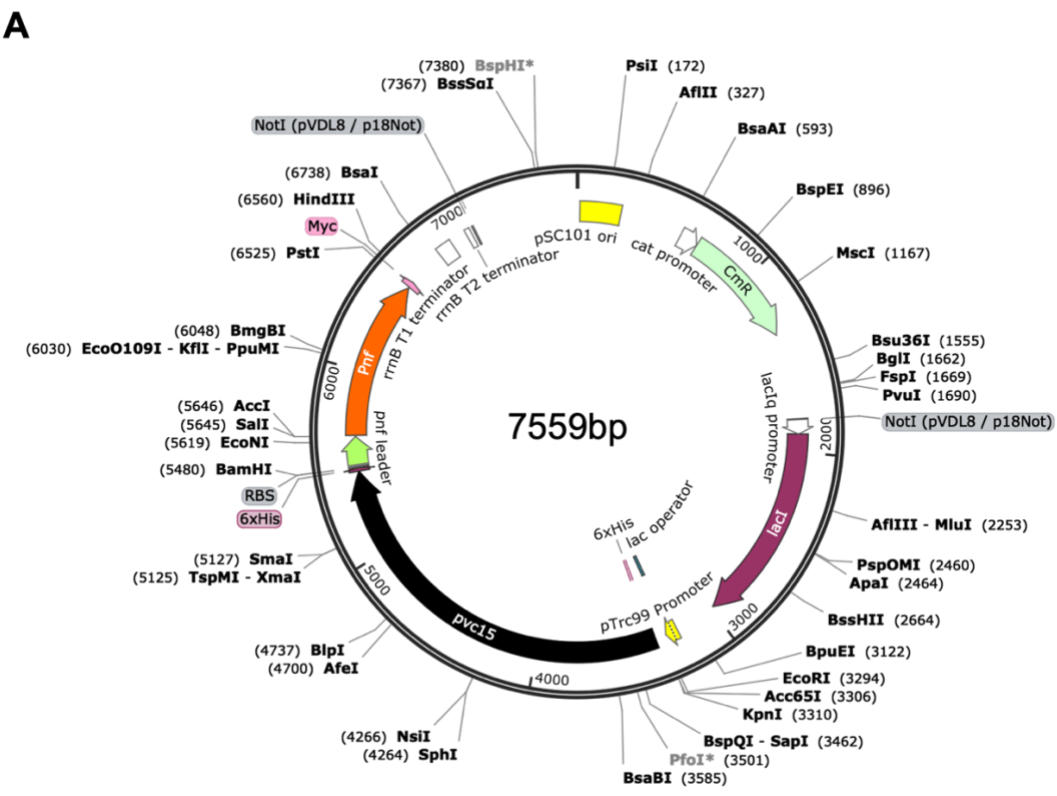


Figure 5.8 | An intact LS and Pvc15 are required for PVC*pnf*-payload association.

pVTRa vectors expressing Myc-tagged Pnf with either Pnf50 or the C-terminal truncated Pnf10 LS were co-expressed with pBAD-PVC*pnf* or PVC*pnf*Δ*pvc15*. PVC*pnf* nanosyringes were purified by immunoprecipitation and aliquoted for analysis in WB using α-Myc and α-FLAG primary antibodies. **(A)** Band intensities of the supernatant (S), flow-through (FT), and elution (E) were aliquoted for each strain. **(B)** Each α-Myc band intensity was normalised according to its corresponding α-FLAG mean band intensity, each **elution**-step band was compared quantitatively and relativised. Additional biological replicates are also included, though not depicted here; n = 3. PVC*pnf*+Pnf50 vs. PVC*pnf*Δ*pvc15*+Pnf50: t(4) = -4.23, unadjusted p-value = 0.013 *; PVC*pnf*+Pnf50 vs. PVC*pnf*+Pnf10: t(3) = -3.22, unadjusted p-value = 0.049 *. Statistical inferences were made using a two-tailed Student's T-test.

5.4.3 The Effect of Mutations on the Roles of Pvc15

It was necessary to explore the effects of mutations in Pvc15 without the necessary co-expression of the rest of the PVC*pnf* operon. Thus, a pVTRa vector encoding the non-catalytic Pnf(C190A) downstream of double hexahistidine-tagged Pvc15 was constructed and termed pVTRaDuet (Figure 5.9A). With the aim of expressing both Pvc15 and Pnf simultaneously, each is under the action of the same promoter. α-His WB still did not indicate successful expression of Pvc15, although Pnf was expressed at high levels at 4 h post-induction (Figure 5.9B). These results show, once again, that the stabilising effect of Pvc15 is present though its detectable protein content in WB analysis is not.



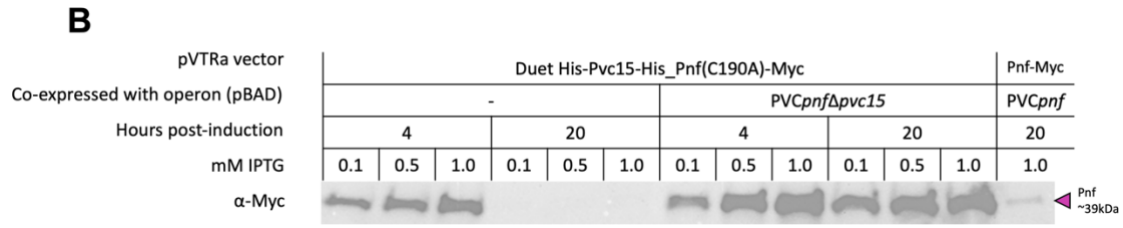


Figure 5.9 | pVTRaDuet vector: simultaneous Pnf-Myc and Pvc15 expression.

The pVTRaDuet vector was constructed using the KpnI + BamHI sites within the original pVTRa Pnf(C190A)-Myc vector. Sequencing revealed an accurate sequence of the native N- and C-terminal 6xHis tagged Pvc15$_{pnf}$ as well as the C-terminal Myc-tagged Pnf(C190A). **(A)** Labelled sequence map of the pVTRaDuet-6xHis-Pvc15-6xHis-Pnf50-Pnf-Myc expression vector (double hexahistidine tagged Pvc15 and Pnf50-Pnf(C190A)-Myc under the control of the same promoter) with chloramphenicol resistance sites (*catR*) and *lacI* encoded upstream of the *lacO cis* repressor element within the pTrc99-type promoter. **(B)** *E. coli* lysates of cells encoding pVTRaDuet-6xHis-Pvc15-6xHis-Pnf50-Pnf(C190A)-Myc were analysed by WB using an α-Myc primary antibody and compared to lysates of cells which also expressed pBAD-PVC*pnf*Δ*pvc15* at various time points after induction with IPTG and arabinose. These were also compared to the *cis* system where Pvc15 is encoded in the pBAD-PVC*pnf* vector along with pVTRa-Pnf50-Pnf-Myc.

Interestingly, Pnf co-expressed with Pvc15 alone resulted in loss of detectable payload in supernatant at 20 h post-induction. On the other hand, Pnf co-expressed with Pvc15 as well as the rest of the PVC*pnf* operon retains payload stability at this time point (lanes 4-6 compared to lanes 10-12). This could imply that a further set of stabilising interactions exists between the payload, Pvc15, and the remaining PVC*pnf* structural components in the later stages of protein expression (i.e., 20 h). There was no difference in Pnf-Myc band intensity at 4 h post-induction of the pVTRaDuet vector with or without the co-expression of the PVC*pnf*Δ*pvc15* operon ($t(4) = 1.43$, $p = 0.28$). Notably, some differences in Pnf expression are observed when comparing Pvc15 expression from the pVTRaDuet vector to that of pBAD-PVC*pnf* (lane 12 and 13), which may be indicative of differences in Pnf expression affected by being encoded under the control of the same promoter as Pvc15 in pVTRaDuet.

5.4.4 Pvc15 Domain D2 Deletion Reduces Pnf's Optimal Abundance

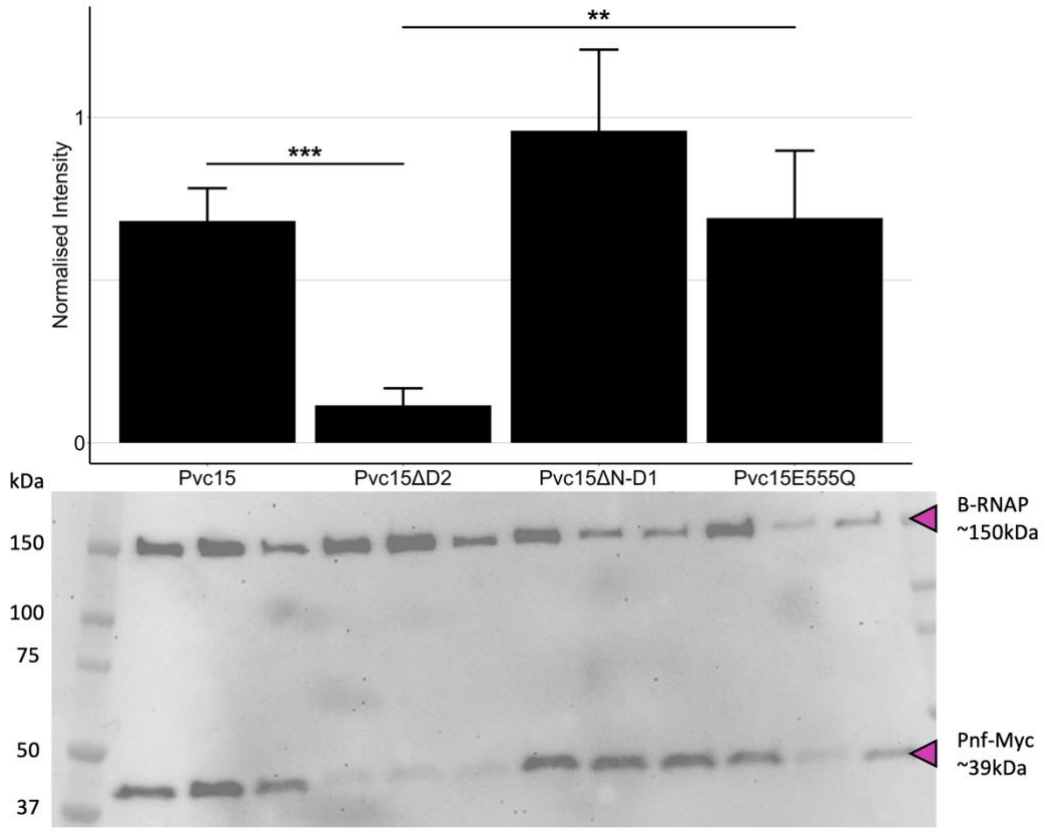


Figure 5.10 | Pnf abundance is depleted when the Pvc15 D2 domain is deleted.

E. coli encoding pVTRaDuet-6xHis-Pvc15-6xHis-Pnf50-Pnf(C190A)-Myc, pVTRaDuet-6xHis-Pvc15NTD-Pnf50-Pnf(C190A)-Myc, pVTRaDuet-Pvc15CTD-6xHis-Pnf50-Pnf(C190A)-Myc, or pVTRaDuet-6xHis-Pvc15E555Q-6xHis-Pnf50-Pnf(C190A)-Myc were induced with IPTG and lysates were aliquoted after 3 h of growth and analysed by WB using α-Myc and α-β-RNAP primary antibodies. Pnf abundance is depleted when co-expressed with Pvc15NTD (Pvc15ΔD2) ($t(4) = -8.66$, p-value = 0.0010 ***). The ATPase point mutation, E555Q, also retains a high level of Pnf in lysates compared to the Pvc15ΔD2 mutant ($t(4) = -4.68$, p-value = 0.0094 **). *** = $p < 0.005$, ** = $p < 0.01$, * = $p < 0.05$.

Given the bioinformatic analysis of Pvc15, it was determined that each of the N, D1, and D2 domains may play different roles in Pvc15's activity. To investigate the potential roles, mutants of Pvc15 were cloned into the pVTRaDuet vector to be co-expressed with the payload Pnf. These mutants, retaining the respective

N- or C-terminal hexahistidine tags, had domains N-D1 (residues 1-340) deleted or D2 (residues 434-698) deleted. Additionally, a point mutation in the catalytic site of D2, E555Q, traps ATP in the active site by preventing its hydrolysis.

When induced for 3 h, only deletion of the entire D2 domain appeared to decrease the prevalence of Pnf in *E. coli* lysates since point mutation of the active site had seemingly no effect (Figure 5.10) (wild-type vs. Pvc15ΔD2: t(4) = -8.66, p-value = 0.0010 ***; E555Q vs. Pvc15ΔD2: t(4) = -4.68, p-value = 0.0094 **). Some Pnf abundance was retained when co-expressed with Pvc15ΔD2, implicating D2 as the primary chaperone domain for the Pnf LS, though domains N-D1 may have some stabilising activity given that deletion of Pvc15 from previous figures would normally abolish Pnf abundance completely (Figure 5.4). The co-expression of the remaining PVC*pnf* operon (PVC*pnf*Δ*pvc15*) did not appear to recover Pnf abundance, as shown in Figure 5.11.

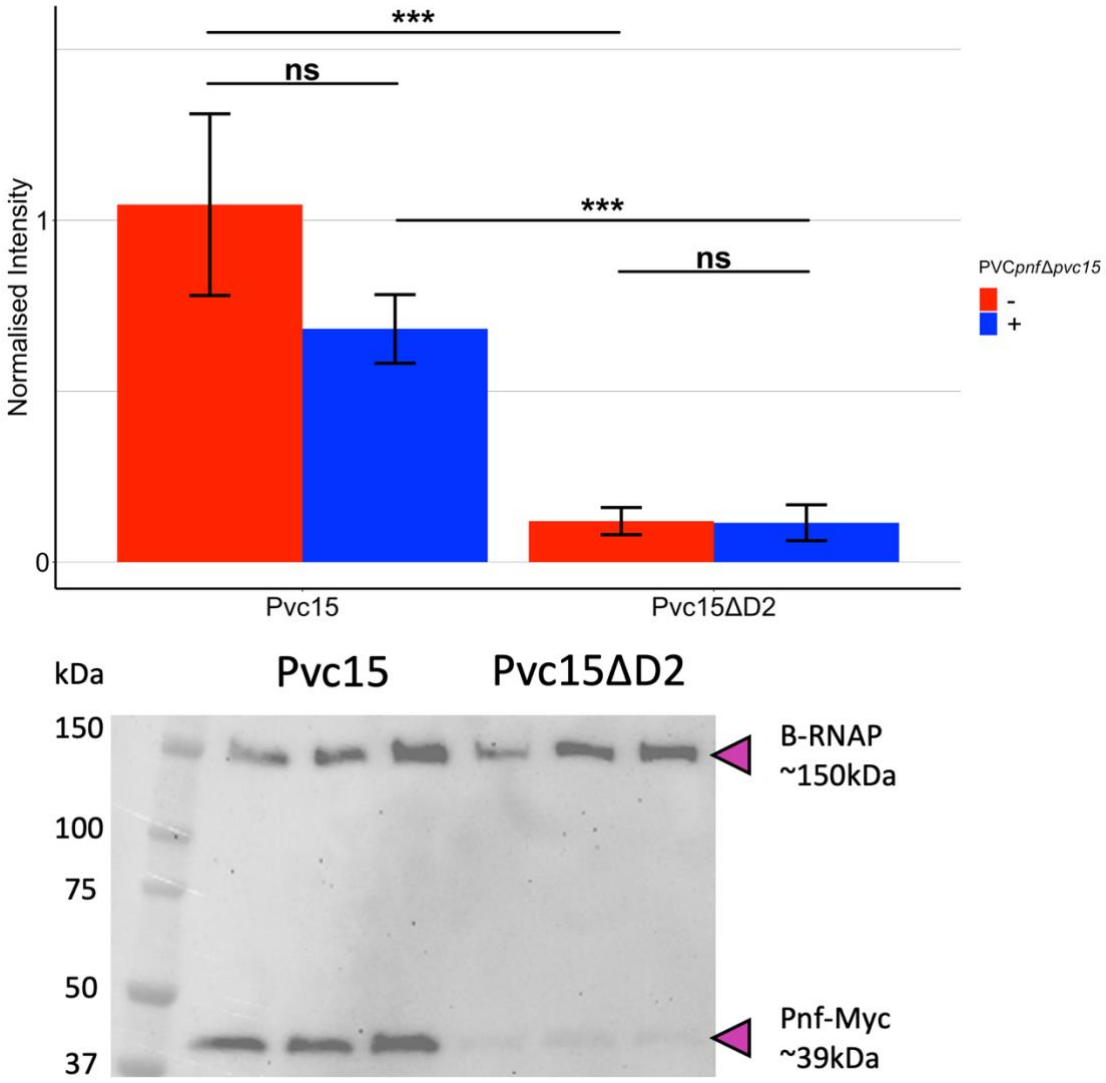


Figure 5.11 | Co-expression of PVC*pnfΔpvc15* does not affect Pvc15's role in Pnf stabilisation. Pnf abundance was measured at optimal expression of 3 h using WB with α-Myc and α-β-RNAP primary antibodies. Each Pnf50-Pnf-Myc band was quantified by normalising by the intensity of β-RNAP in the same lane, as in the previous data. No difference was found in cells encoding pVTRaDuet-6xHis-Pvc15-6xHis-Pnf50-Pnf(C190A)-Myc or pVTRaDuet-Pvc15NTD-6xHis-Pnf50-Pnf(C190A)-Myc with lysates co-expressing pBAD-PVC*pnfΔpvc15* (i.e., red bars vs blue bars). The Pvc15ΔD2 strain (encoding pVTRaDuet-Pvc15NTD-6xHis-Pnf50-Pnf(C190A)-Myc) displayed less Pnf abundance regardless of the presence of the induced operon (without PVC*pnfΔpvc15*: t(4) = -5.97, p-value = 0.0040 ***; with PVC*pnfΔpvc15*: t(4) = -8.66, p-value = 0.0010 ***). WB (bottom) is from cell lysates after induction of the respective pVTRaDuet vector as well as the PVC*pnfΔpvc15* operon. A two-sample T-test was used to assess significance; *** = $p < 0.005$; ns = not significant.

5.4.4.1 Synergistic Interference of Arabinose with IPTG Induction

To ensure correct controls between *E. coli* inductions, it was tested whether arabinose affected Pnf expression from the pVTRaDuet vector since this would be used to distinguish the relevance of the PVC*pnf* operon to the various Pvc15 mutants. Pnf abundance was significantly greater in the presence of arabinose despite the intention of the vector to be induced with IPTG (Figure 5.12ii) ($t(4) = 7.54$, p-value = 0.0064 **). This increase in abundance could be seen for the degraded fragments of Pnf which are often observed in lysates (Figure 5.12i) ($t(2) = 12.49$, p-value = 0.0017 ***). Thus, when comparing payload expression with or without the PVC*pnf* operon, it was important to compare strains which do or do not contain the operon, instead of using the positive strain with or without arabinose. As indicated by Figure 5.5A, uninduced Pvc15 has a stabilising effect similar to the induced Pvc15, rather than the deleted Pvc15.

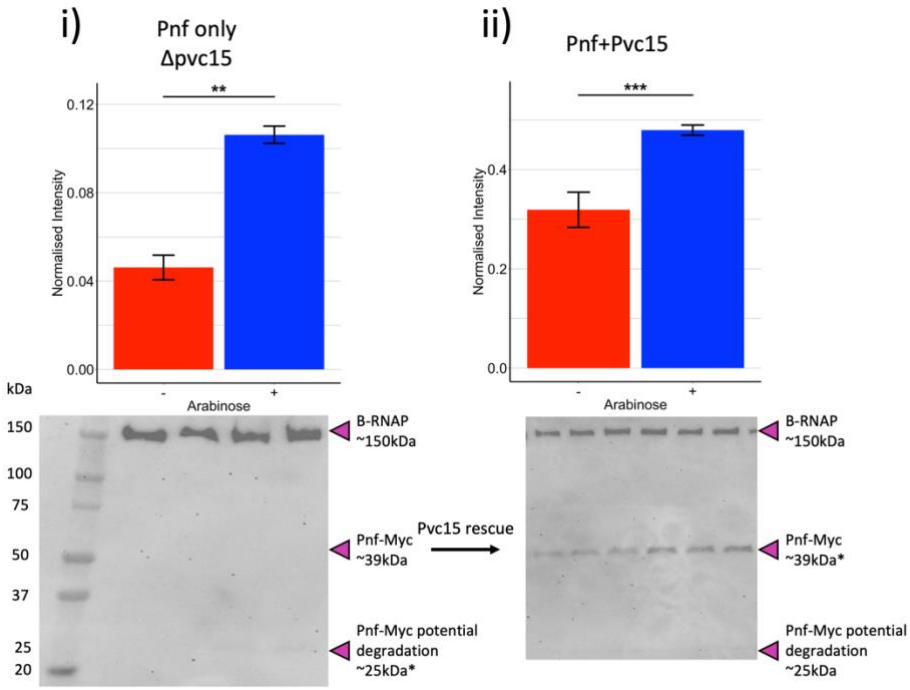


Figure 5.12 | Arabinose enhances IPTG-induced Pnf expression.

E. coli strains encoding i) pVTRa-Pnf50-Pnf-Myc, or ii) pVTRaDuet-6xHis-Pvc15-6xHis-Pnf50-Pnf(C190A)-Myc were grown and lysates were taken at 3 h post-induction with IPTG and with

or without arabinose. Aliquots were analysed by WB using α-Myc and α-β-RNAP primary antibodies for normalisation. The addition of arabinose increased the expression of Pnf-Myc in both its i) degraded form (~25 kDa; t(2) = 12.49, p-value = 0.0064 **) and ii) native form (~39 kDa; t(4) = 7.54, p-value = 0.0017 ***), despite the absence of PVC*pnf*. Asterisks indicate which band intensities are measured in the bar chart above. For significance, a two-sample T-test was used; ** = $p < 0.01$, *** = $p < 0.005$.

5.4.5 ATPase Activity is Depleted in Pvc15 Mutants

Jiang and colleagues (2022) documented the unfolding of LS-associated mRFP via Pvc15 which occurred most predominantly around 14-20 h post-induction. To investigate how the ATPase activity of Pvc15 was related to its capacity for payload loading, cell lysates at 16 h post-induction were assayed using a colorimetric malachite green ATPase activity kit (Abcam). Samples were corrected for the carryover of phosphates from lysates and background activity of the probe as in chapter 5.3.2. The standard curve constructed to convert OD_{650} to amount of phosphate is shown in Figure 5.13.

It was found that only wild-type Pvc15 in the presence of native Pnf and the PVC*pnf* operon displayed ATPase activity over the 30-minute experiment (Figure 5.14). Both Pvc15 domain deletions (ΔN-D1 and ΔD2) as well as the E555Q point mutation resulted in almost complete abolishment of Pvc15 ATPase activity. Since E555 is part of the Walker B box in D2, it can be interpreted that D2 is the sole functional ATPase domain and though the D1 domain lacks ATPase activity, it is still required for catalysis from D2. In addition, this activity depended on the presence of other PVC*pnf* operon proteins.

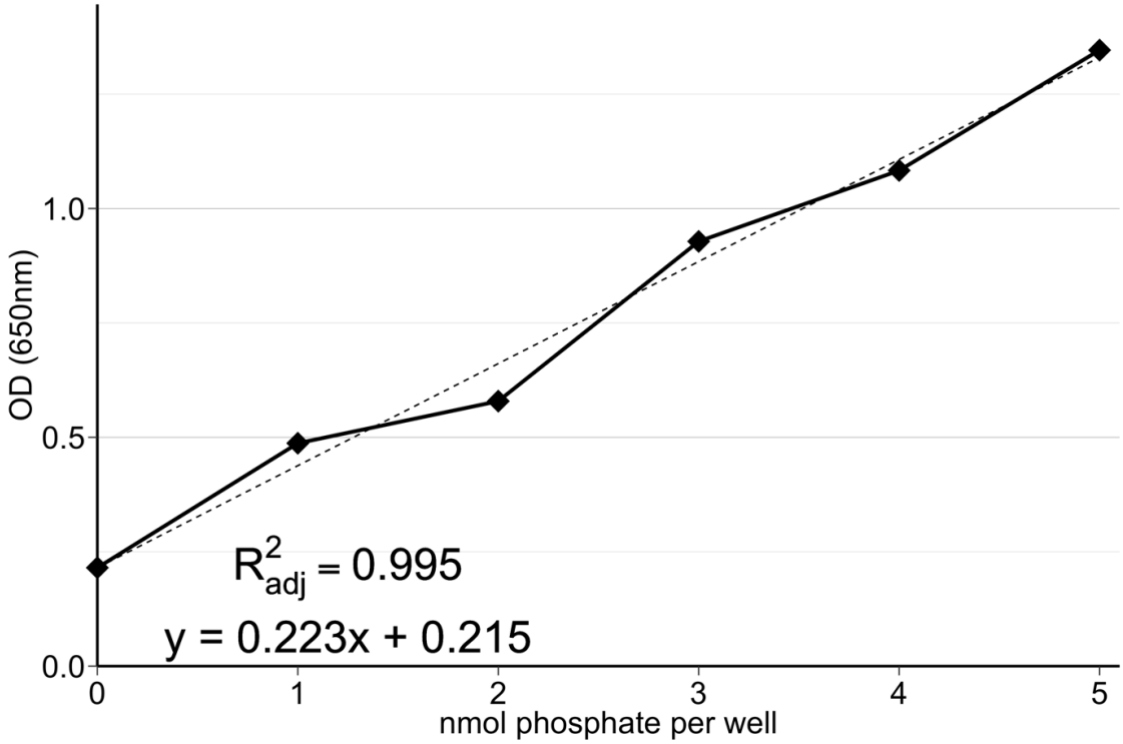


Figure 5.13 | Standard curve for conversion of OD (650 nm) to nmol of phosphate.

OD_{650} was measured for phosphate standards of known quantity to assay how the amount of phosphate affected the conversion of the malachite green probe. The equation of the linear regression at time t=0 (adjusted R^2 = 0.995) was used to find the phosphate from a change in OD_{650}: y = 0.223x + 0.215.

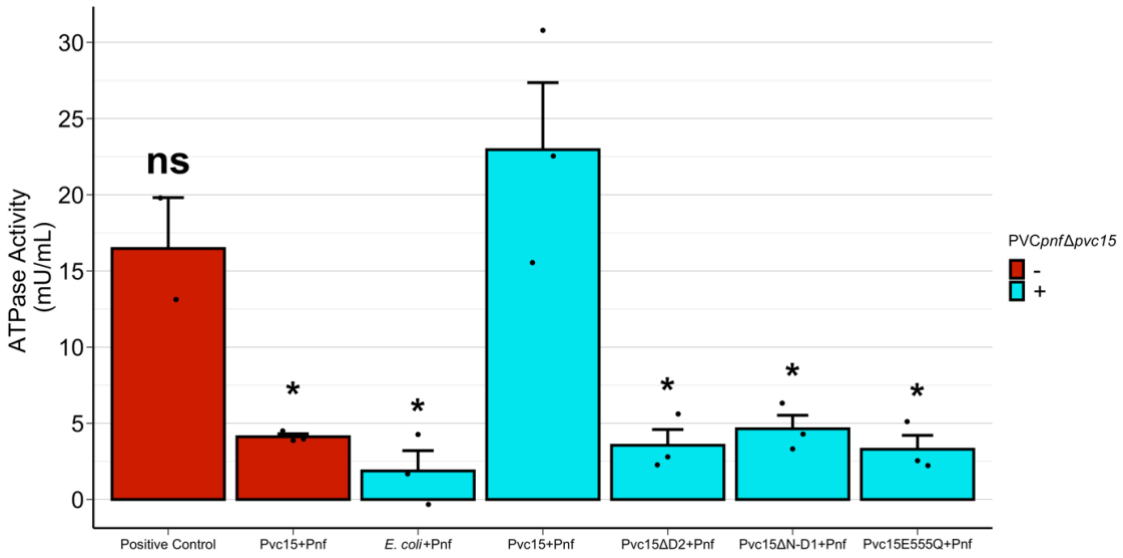


Figure 5.14 | D2 is the sole functional ATPase domain with PVC*pnf*-dependent activity.

A colorimetric ATPase assay using malachite green (Abcam; Cat. No.: ab234055) was used to assess the amount of inorganic phosphate released from cell lysates at 16 h post-induction. *E.*

coli encoded either pVTRa-Pnf50-Pnf-Myc (*E. coli*+Pnf), pVTRaDuet-6xHis-Pvc15-6xHis-Pnf50-Pnf(C190A)-Myc (Pvc15+Pnf), pVTRaDuet-6xHis-Pvc15NTD-Pnf50-Pnf(C190A)-Myc (Pvc15ΔD2+Pnf), pVTRaDuet-Pvc15CTD-6xHis-Pnf50-Pnf(C190A)-Myc (Pvc15ΔN-D1+Pnf), or pVTRaDuet-6xHis-Pvc15E555Q-6xHis-Pnf(C190A)-Myc (Pvc15E555Q+Pnf) with or without the co-expression of pBAD-PVC*pnf*Δ*pvc15* (without = red, with = cyan). OD (650 nm) was measured over after 30 minutes and corrected for the carryover of phosphate from lysates, as well as for the spontaneous change in OD_{650} from the malachite green. A cell-free positive control was also provided.

All of the Pvc15 mutants, as well as the *E. coli*+Pnf negative control, displayed decreased ATPase activity compared to the wild-type PVC*pnf*+Pnf lysate (*E. coli*+Pnf: t(4) = -4.59, unadjusted p-value = 0.010 *; Pvc15ΔD2: t(4) = -4.29, unadjusted p-value = 0.013 *; Pvc15ΔN-D1: t(4) = -4.08, unadjusted p-value = 0.015 *; Pvc15E555Q: t(4) = -4.37, unadjusted p-value = 0.012 *). The mutants were not significantly different from that of the *E. coli*+Pnf control whilst the wild-type lysate was not significantly different from the positive control (t(3) = -1.05, p-value = 0.37). Wild-type Pvc15 retained its ATPase activity only when co-expressed with the PVC*pnf*Δ*pvc15* operon (t(4) = -4.28, unadjusted p-value = 0.013 *).

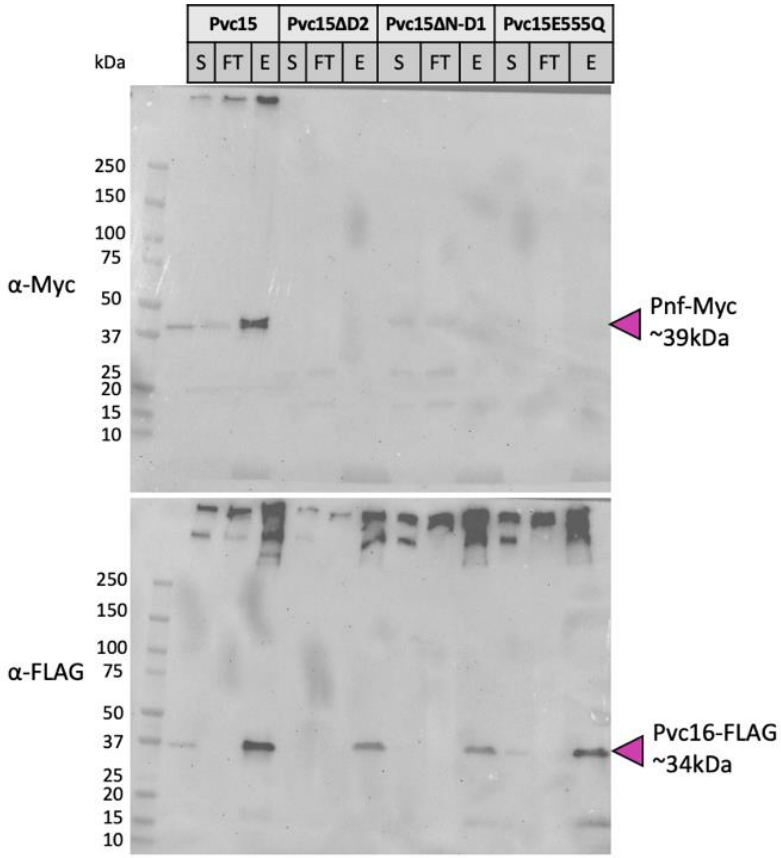
Error bars show **standard error of the mean**. Asterisks represent p-values of less than 0.05 in a two-sided Student's T-test when compared to the Pvc15+Pnf PVC*pnf*Δ*pvc15*$^{+}$ strain lysate; ns = not significant, * = $p < 0.05$. Comparison of positive control to negative control: t(3) = 4.83, unadjusted p-value = 0.017 *.

5.4.6 ATPase Competence is Required for Payload Loading

ATPase activity by Pvc15 only occurs in the presence of the induced PVC*pnf* operon. This led to the hypothesis that the ATPase activity itself was essential for payload loading, specifically, and that the unfolding activity characterised by Jiang and colleagues (2022) may be a separate process. The capacity for these Pvc15 mutants to load payload into mature PVC*pnf* was analysed by IP of lysates of cells co-expressing the pVTRaDuet encoding either the wild-type, ΔN-D1, ΔD2, or E555Q mutants with pBAD-PVC*pnf*Δ*pvc15*.

As predicted, PVCs expressed with mutant Pvc15 did not co-purify with Pnf, suggesting that the ATPase activity is required for translocating Pnf50 LS-associated proteins into the PVC tube lumen (Figure 5.15). Altogether, Pvc15 ATPase activity i) is mediated solely by D2: the primary ATP-independent chaperone for the LS-tagged payload, ii) is likely facilitated by hexamerisation via the N and D1 domains, iii) is enabled by interactions with other PVC*pnf* operon proteins, and lastly, iv) is required for translocation of the payload into the PVC*pnf* tube lumen.

A



B

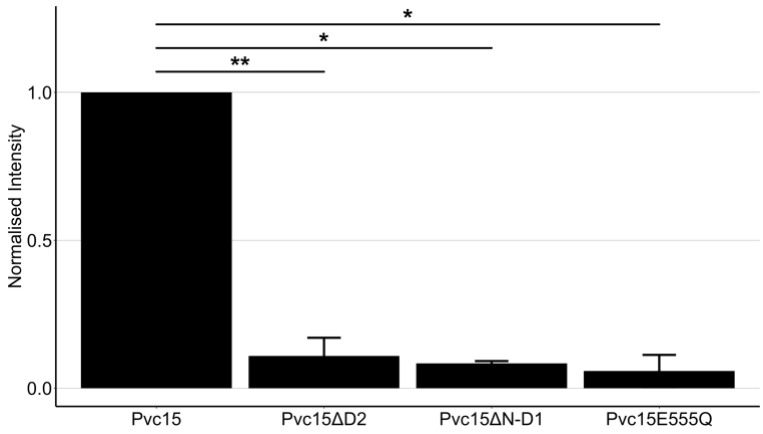


Figure 5.15 | PVC*pnf* cargo loading requires ATPase competent Pvc15.

PVC*pnf* was purified by IP from *E. coli* lysates of cells encoding pBAD-PVC*pnf*Δ*pvc15* and pVTRaDuet-6xHis-Pvc15-6xHis-Pnf50-Pnf(C190A)-Myc (Pvc15), pVTRaDuet-6xHis-Pvc15NTD-Pnf50-Pnf(C190A)-Myc (Pvc15ΔD2), pVTRaDuet-Pvc15CTD-6xHis-Pnf50-Pnf(C190A)-Myc (Pvc15ΔN-D1), or pVTRaDuet-6xHis-Pvc15E555Q-6xHis-Pnf(C190A)-Myc (Pvc15E555Q) and aliquoted and analysed by WB using α-Myc and α-FLAG primary antibodies. **(A)** Supernatants (S), flow-throughs (FT), and elutions (E) from the IP are shown. Gels are representative of n = 3, 3, 2, and 2 independent reiterations of PVC*pnf* purifications, respectively. **(B)** Quantitative representation of **(A)**: a ratio of Pnf-Myc : Pvc16-FLAG was obtained for each elution and then normalised to wild-type PVC*pnf*. Wild-type Pvc15 vs. Pvc15ΔD2: t(3) = -6.01, unadjusted p-value = 0.00092 **; Pvc15ΔN-D1: t(2) = -4.84, unadjusted p-value = 0.040 *; Pvc15E555Q: t(2) = -4.88, unadjusted p-value = 0.039 *.

5.4.7 Additional PVC*pnf* Loading Biological Replicates

Biological replicates for some PVC loading experiments are depicted in Figure 5.16.

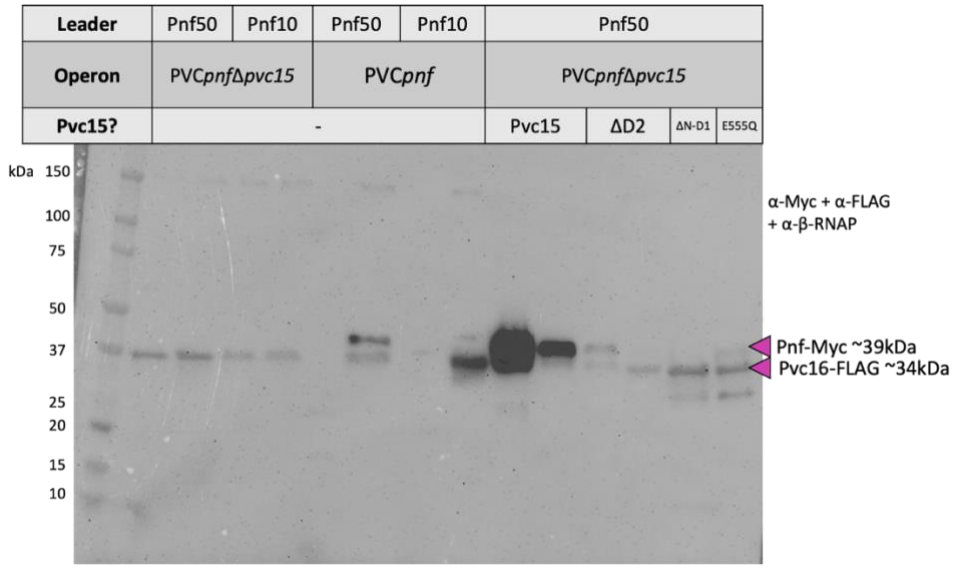


Figure 5.16 | Biological replicates for statistical analysis in previous loading figures.

Figure 5.8 and Figure 5.15 used quantitative analysis to assess the ratio of Pnf-Myc payload per Pvc16-FLAG. Biological replicates are depicted here. Lanes 6, 8, and 10 were not used in statistical analysis due to either the absence of appropriate bands or smearing. WB using α-Myc, α-FLAG, and α-β-RNAP primary antibodies was performed on PVCs purified by IP from *E. coli* lysates of cells co-expressing vectors pVTRa-Pnf50-Pnf-Myc (lanes 2, 3, 6, and 7) or pVTRa-Pnf10-Pnf-Myc (lanes 4, 5, 8, and 9) with pBAD-PVC*pnf*Δ*pvc15* (lanes 2-5) or pBAD-PVC*pnf* (lanes 6-9). Lanes 10-15 co-expressed the pBAD-PVC*pnf*Δ*pvc15* vector and pVTRaDuet-6xHis-Pvc15-6xHis-Pnf50-Pnf(C190A)-Myc (Pvc15), pVTRaDuet-6xHis-Pvc15NTD-Pnf50-Pnf(C190A)-Myc (Pvc15ΔD2), pVTRaDuet-Pvc15CTD-6xHis-Pnf50-Pnf(C190A)-Myc (Pvc15ΔN-D1), or pVTRaDuet-6xHis-Pvc15E555Q-6xHis-Pnf(C190A)-Myc (Pvc15E555Q).

5.5 Discussion

Jiang and colleagues (2022) investigated the effects of point mutations within the Pnf LS to identify any single residues which may have altered the association with the PVC using alanine scanning, yet they failed to identify any key residues altering the loading of Pnf50-*Renilla* luciferase (Rluc). However, their data does not reveal the nature of Pvc15-specific interactions nor how stability of the native payloads, such as Pnf, are affected by point mutations since Rluc is stable regardless of the presence or absence of LSs (Leclerc *et al.*, 2000).

This chapter demonstrated the importance of intact leader sequences (LSs) and the Pvc15 ATPase for payload stability and PVC association and loading. Whilst both of the N- and C-terminal truncations of the Pnf LS in this work are found to confer some level of stability, a consequence of doing so is a large depletion in loaded PVCs. Hence, long, disordered LSs may be suboptimal for payload stability and cytosolic abundance, but this trade-off is more than compensated for when co-expressed with the Pvc15 chaperone, regardless of its ATPase competence. Furthermore, it was demonstrated that the ATPase activity is enabled only when co-expressed with some elements of the native PVC*pnf* operon; Pvc15 mutants also failed to sufficiently load Pnf into PVC*pnf*.

Indeed, Pvc15 homologues may be so well conserved in eCISs because they complement the inherent instability that some LSs may bestow to their payloads. This opens many questions as to the nature of this conservation, such as why some Pvc15 homologues are encoded where LSs are not found, and why some LSs, such as Cif50, appeared to confer more stability than their truncated counterpart regardless of Pvc15 presence. Future work should aim to resolve the atomic structure of Pvc15 and observe whether hexamerisation takes place in wild-type Pvc15 and each of the mutants tested here. Given the

difficulty in purifying Pvc15 using a Ni^{+} NTA resin, a more dynamic approach to characterise its structure may be required.

5.5.1 Cif Leader Interactions

Whilst the abundance of Cif50-associated Pnf was not affected by the presence of $Pvc15_{pnf}$ (Figure 5.4), it is possible that the $Pvc15_{cif}$ would interact differently with this LS. Though $Pvc15_{pnf}$ and $Pvc15_{cif}$ share 56% amino acid sequence identity, this is still less than that of $Pvc15_{pnf}$ and Afp15 at 66% which, as discussed in chapter 1.3.1.6, display differences regarding expression and likely cognate protein targets. The interplay between different LSs and different Pvc15 homologues should next be investigated.

5.5.2 Interpretation of Protein Abundance Experiments

Degradation time course experiments conducted from *E. coli* lysates identified that after halting translation, chaperoned Pnf50-Pnf degrades at roughly the same rate as unchaperoned Pnf10-Pnf (Figure 5.4). This experiment represents the dynamics of protein expression and stability balanced by the intrinsic bacterial degradation machinery acting upon the payload proteins. Though the conclusions drawn are persuasive for Pvc15's role in stabilising LS-associated Pnf, Pvc6 and Pvc14 are yet to be characterised in any capacity and may serve as potential candidates for follow-up knock-out experiments to investigate Pnf payload lysate abundance and loading capacity. The work could also be used to infer whether Pvc14, Pvc6, Pvc15, and the payload associate to influence protein interactions in the same vein as that of the jettisoning of Afp14 from the Afp16-Afp3 complex to form the Afp15-Afp16-Afp3-Afp2 complex (Bhardwaj, Mitra and Hurst, 2021).

In the work by Jiang and colleagues (2022), a band on WBs of epitope-tagged Pnf can be found roughly 8 kDa lower than that of the native Pnf (~39

kDa for Myc-tagged Pnf). The authors reason that this lower band is the result of PVC-specific cleavage during the loading process and that the first 70 amino acids are cleaved in the resulting product. In this thesis, multiple lower bands are observed in native and truncated Pnf LSs which are consistently found at roughly 25 kDa and 17 kDa. This implies much more C-terminal, and more frequent, cleavage events. In addition, these lower bands are observed in Δ*pvc15* strains, including those where stable Pnf cannot be observed and where the Pnf LS is replaced with the Cif LS. In the authors' work, only the larger 39 kDa molecular weight Pnf bands are observed to co-purify with PVCs, implying that cleavage of the LS is unimportant for loading Pnf; the same is observed in this thesis since any lower bands disappear when doing WB on purified PVC elutions.

Instead, these bands are likely the result of other degradative activities acting upon Pnf, supported by the fact that the intensity of these lower bands increases when adding arabinose to artificially increase Pnf induction even in the absence of any other PVC components (Figure 5.12). Therefore, cleavage of the LS may, indeed, take place as Jiang and colleagues (2022) suggested, as verified by their mass spectrometry analysis. However, the data in this thesis suggests that LS cleavage events do not play an important role in Pnf loading. Lower molecular weight bands imply a specific action of bacterial proteases upon the payload protein itself regardless of the sequence of the leader or presence of other PVC proteins in the lysate.

5.5.3 Alternative Interpretations of Pnf LS-Pvc15 Interactions

The abundance of Pnf10-Pnf-Myc with PVC*pnf* was demonstrated to be higher from cell lysates after 3 h of induction with IPTG than that of Pnf50-Pnf-Myc with PVC*pnf* (Figure 5.4A), but these initial levels differed at the 0 h time point in non-relativised data (Figure 5.4B). The Pnf50 degradation experiment could

not be conducted for the Δ*pvc15* mutant since no Pnf50-Pnf bands were present to measure an initial time point due to its instability.

Despite the lower initial abundance of chaperoned Pnf50 compared to unchaperoned Pnf10, the degradation rates appeared similar. With this information, it is possible to put together a simplified mass action kinetics model of the Pnf expression-degradation cycle (Figure 5.17). In this model, Pnf50-Pnf (referred to as simply Pnf) and Pnf10-Pnf are transcribed and translated at a rate k_{exp} and k_{exp_Pnf10}, respectively (equations 1a and 1b; it is possible that these two rates are synonymous). Pvc15 has a potent ability to increase Pnf abundance even when actively repressed with 0.5% glucose (Figure 5.12); it is thus represented in the model as present at saturation, and associates with Pnf at a rate of k_a and dissociating at a rate of k_d (equations 2a and 2b). Pnf decays very quickly in the absence of Pvc15 (equation 3a), as shown by artificially increasing its expression using arabinose and observing a significant increase in Pnf's degradative band intensity (Figure 5.12). Thus, it can be interpreted that bands measured for the degradation experiment were, in fact, Pnf associated with Pvc15 until the addition of LDS for running in SDS-PAGE. Given Pvc15's putative chaperoning ability, its complexation with Pnf leads to a far lesser rate of decay (k_{decay_2}) which, as shown by data in Figure 5.4B, are roughly the same as that of Pnf10 regardless of the presence of Pvc15 (equations 3b and 3c and interpretation [1]). This gives a simple kinetics equation to describe the concentration of Pnf over time (interpretation [2]).

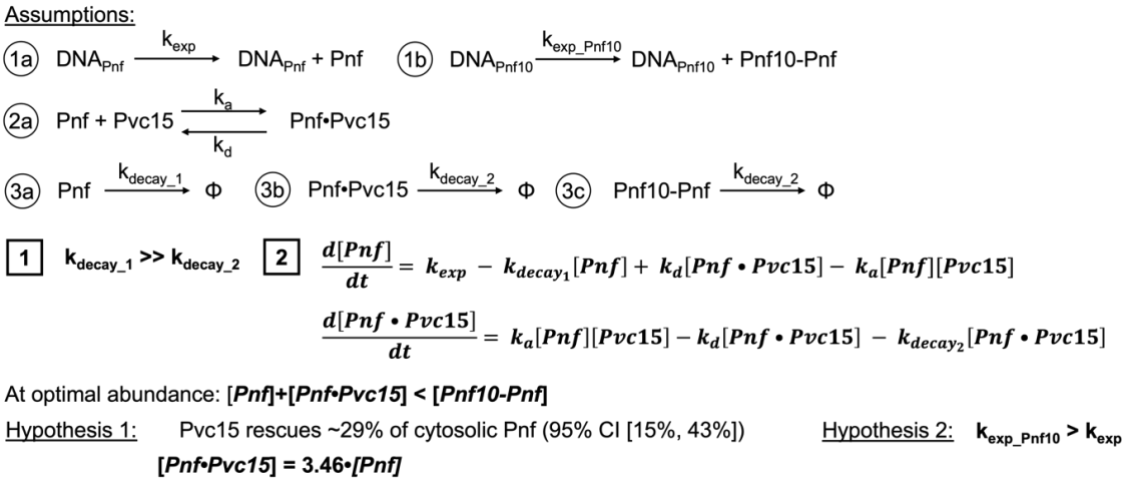


Figure 5.17 | Inferred mass action kinetics model of Pnf abundance.

Native Pnf50-Pnf (shortened here to simply 'Pnf') and Pnf10-Pnf, expressed via transcription and translation machinery, are represented here as condensed rates: k_{exp} and k_{exp_Pnf10}, respectively (equations 1a and 1b). Pnf, but not Pnf10-Pnf, can associate with Pvc15 according to the rates of association, k_a, and dissociation, k_d (equation 2a). Even actively repressed Pvc15 via glucose addition causes a potent increase in Pnf abundance such that Pvc15 can be considered present in saturating amounts (equation 2b).

Native Pnf decays at rate k_{decay_1} which is markedly larger than k_{decay_2} since Pnf in the absence of Pvc15 is undetectable (equations 3a-c; interpretation 1). Therefore, the similar degradation rates observed in the degradation experiment implies that bands observed were derived from Pnf•Pvc15 (dissociated only after the addition of LDS) and free Pnf10-Pnf. The change in concentration of Pnf can, therefore, be defined according to these base assumptions (interpretation 2).

Building on these base assumptions, it is conclusive that Pnf without Pvc15 decays rapidly but that Pvc15 rescues roughly 29% of cytosolic Pnf at the optimal abundance time point (i.e. 1-(1/2.6)).

The added complexity of Pvc15 having an unfolding activity upon the payload, as shown by Jiang and colleagues (2022), may complicate the interpretation from the degradation experiment described in Figure 5.4B since unfolded Pnf may be less stable than its folded isoform. Pnf10-Pnf's abundance is seemingly independent of the presence of Pvc15, so if Pvc15 can unfold the native Pnf50-Pnf but not Pnf10-Pnf, then this 'multiple isoforms' hypothesis

may explain why Pnf is found at relatively lower levels of abundance compared to Pnf10-Pnf, particularly if this unfolded state has a higher propensity to be targeted to the proteasome. Such a hypothesis could be tested using a thermal shift assay, such as using the thiol-specific fluorochrome N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM) or static light scattering techniques.

Nonetheless, if Pnf10-Pnf levels represent the maximal Pnf abundance and native Pnf in the absence of Pvc15 represents the lowest level, then the results from this experiment estimate that Pvc15 rescues 29% of cytosolic Pnf (95% CI [15%, 43%]) from rapid degradation.

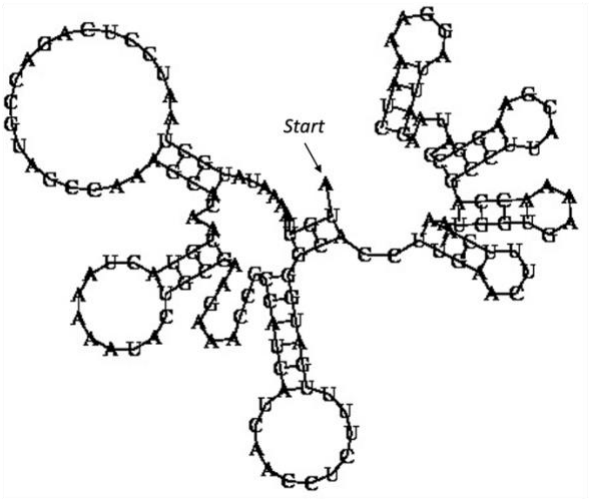
Alternatively, Pvc15 may bind a higher proportion of Pnf but the difference in abundance results from differences in the rate of expression which may be greater for the shorter LS. Given that Pnf10-Pnf abundance is seemingly independent of Pvc15, this suggests that Pvc15 would not regulate either transcription or translation of the payload despite the potential for the N-domain to have DNA-binding properties according to the results from the DALI server. Indeed, the shorter LS may have a lower propensity to form DNA or RNA secondary structures and subsequently interfere with the initiation complex (Studer and Joseph, 2006; Campo *et al.*, 2015). Predictions using vsfold5 (Dawson, Fujiwara and Kawai, 2007) may support this hypothesis (Figure 5.18): total ΔG for pnf50 mRNA = -33.08 kcal/mol; Pnf10 = +0.29 kcal/mol. These structures are even more likely to form when considering the 5′ untranslated region (5′ UTR) of the mRNA (~40 nucleotides) from the pVTRa vectors from which they were expressed where total ΔG values are -38.40 and -8.29 kcal/mol, respectively. The pnf50 LS mRNA forms far more hairpin loops than that of pnf10 mRNA which could implicate differences in ribosomal binding and initiation in Pnf50 expression compared to the truncated LSs.

Interestingly, a large hairpin loop is found immediately 5′ of the AUG start codon when considering the natural 5′ UTR found in the *pnf* ORF in wild-type

P. asymbiotica ATCC43949. Additional folds occur between this region and that of pnf50 (total ΔG = -34.38 kcal/mol), which do not occur in pnf10 (total ΔG = -2.28).

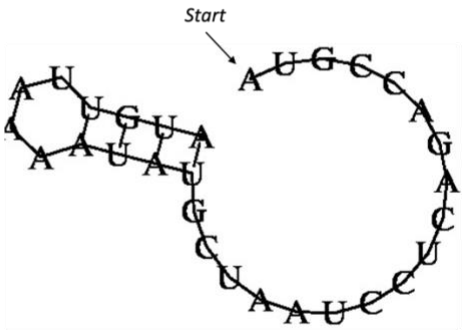
Pnf50:

ΔG = -33.08 kcal/mol



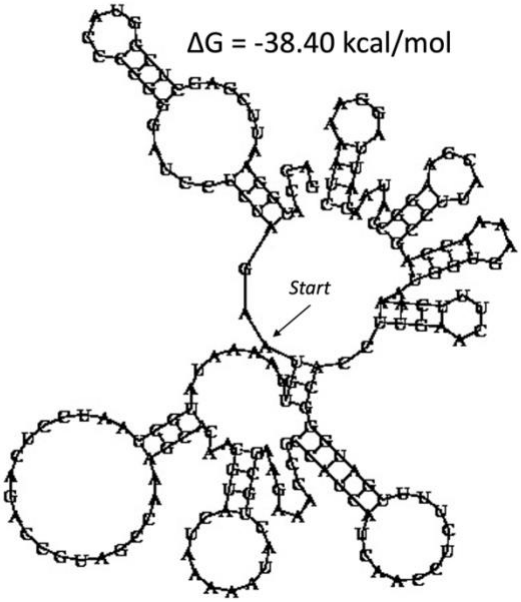
Pnf10:

ΔG = +0.29 kcal/mol



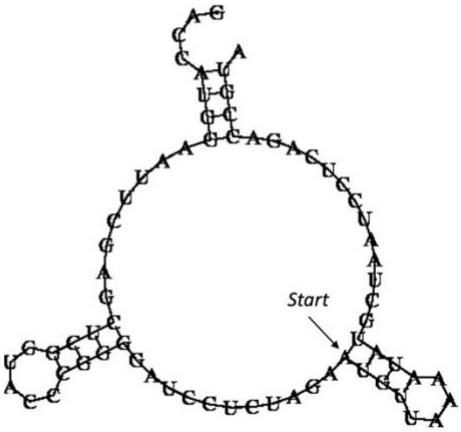
pVTRaRBS_Pnf50:

ΔG = -38.40 kcal/mol



pVTRaRBS_Pnf10:

ΔG = -8.29 kcal/mol



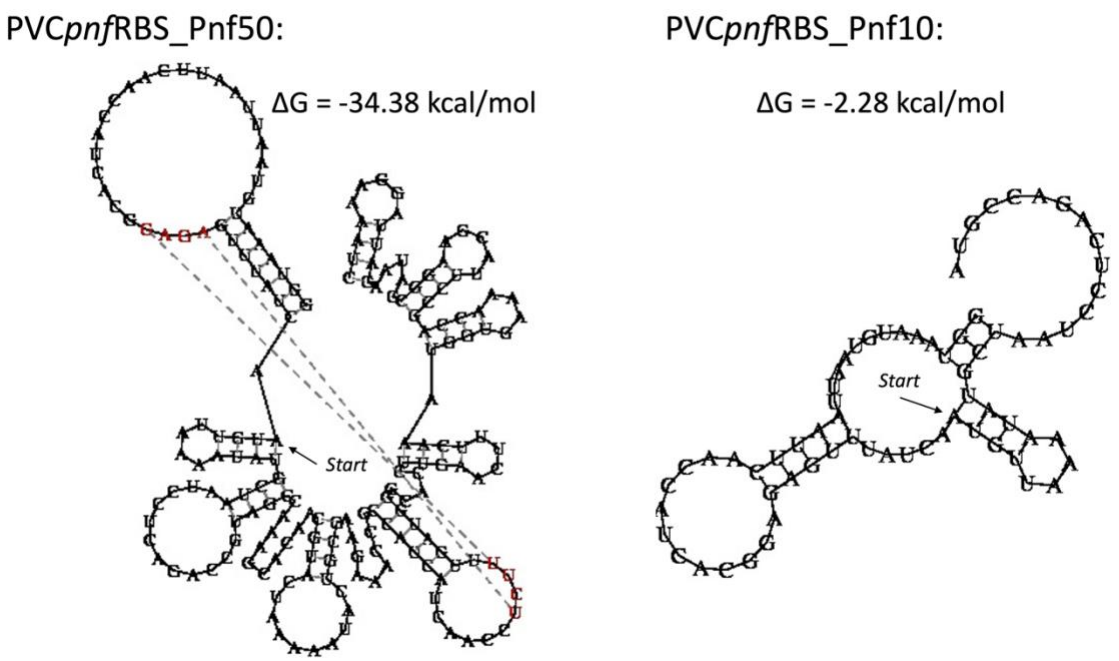


Figure 5.18 | RNA secondary structures of wild-type and C-terminal truncated LSs.
vsFold5 using default parameters (T = 37 °C, minimum stem length = 4 bps, persistence length = 4 nts, Jacobsen-Stockmeyer method, gamma = 1.75). Top: mRNA secondary structure of wild-type pnf50 LS with a total ΔG of -33.08 kcal/mol; the C-terminal truncated pnf10 LS is less prone to secondary structure formation with a total ΔG of +0.29 kcal/mol. Middle: Same as top panel but with an additional 40 nucleotides from the 5' UTR in the pVTRa vector from which it was expressed. Bottom: Same as top panel but with an additional 40 nucleotides from the 5' UTR of *pnf* in the genome of *P. asymbiotica* ATCC43949. Dashed lines represent additional three-dimensional folds constituting further nucleotide binding between the genome 5' UTR and the pnf50 mRNA.

Another hypothesis is that Pnf remains unfolded to be degraded unless chaperoned by Pvc15. The data presented in Figure 5.4 do not support this hypothesis, however, since unfolded Pnf10-Pnf would also not be well tolerated by the bacterial protein degradation machinery so a low abundance would be expected. Instead, this data supports the idea that Pnf folds correctly initially to then be unfolded by Pvc15 at a later time appropriate for loading into the PVC tube lumen, in agreement with the results of the unfolding of mRFP in the work by Jiang and colleagues (2022). The lack of unfolding ability upon Pnf10-Pnf,

and its lack of association with the Pvc15 ATPase, would then explain why it cannot be optimally loaded into the PVC as in Figure 5.8.

5.5.4 Additional Binding Roles of PVC*pnf* Components

Intact Pvc15 ATPase activity was found to depend upon the co-expression of the remaining PVC*pnf* operon (Figure 5.14), implicating a further set of interactions of Pvc15, the LS-tagged protein, and some other native PVC*pnf* operon proteins. Together, these interactions may regulate Pvc15 activity such as to translocate protein only to intact PVC*pnf* during assembly. From the work by Jiang and colleagues (2022) documenting that unspecified Pvc15 mutations do not abolish its unfolding activity, it can be hypothesised that the ATPase activity may not be required for the unfolding activity, specifically. Given its reliance on other PVC*pnf* proteins the ATPase activity may, instead, be responsible for translocating payloads to the PVC tube lumen for loading or, at the very least, associating Pnf strongly enough with the mature PVC that they co-purify (Figure 5.2). This hypothesis that Pnf may be stabilised by translocation into the PVC*pnf* tube lumen may be supported by the data in Figure 5.9B as Pnf stability is increased when co-expressed with the PVC*pnf*Δ*pvc15* operon.

Notably, abundance investigated for the N-terminal truncations could not be investigated solely in the context of Pvc15 due to repeated difficulties in getting positive transformants. However, it is interesting to note that, unlike C-terminal truncations of the Pnf LS, N-terminal truncations resulted in a PVC*pnf*-dependent increase in abundance, which may or may not relate to its binding with Pvc15 or other PVC*pnf* components. If so, the Pvc15-independent stability exerted by the C-terminal truncation of Pnf10-Pnf could further implicate the 'tip' of the leader as conferring innate stability to the payload and that Pvc15 binds the region between 40-50 amino acids of the LS. Jiang and colleagues (2022) identified that N-terminal deletion of Pnf50 resulted in a decrease in Pnf

WB band intensity from cell lysate; deleting a further 20 amino acids caused complete loss of the protein (tested only in the presence of the operon). This suggests that the Pnf leader switches from a *stabilising* to *destabilising* role in the range of the 40th-50th amino acids; this sensitivity would not be too surprising considering the difference between Pnf10 and Cif10 abundance shown in Figure 5.4.

6. Conclusions and Future Directions

This thesis has covered an in-depth background to the PVC, its host bacterium, and the capabilities and predicted structures of its conserved ATPase.

In Chapter 3, it was found that *Photorhabdus* encodes many RHS repeat-containing toxins and that many encoded highly disordered leader sequences (LSs) at their N-termini. Given that certain genera such as *Burkholderia*, *Enterobacter*, *Pseudomonas*, and *Xenorhabdus* do not appear to encode LSs in their putative effectors, it would be fruitful to investigate if there exist alternative modes of eCIS loading in these structures or if, indeed, other ORFs encode effectors at more distant sites or within the structural operons themselves. Future experiments should also aim to test the hypothesis that Pvc11$^{PAU lopT}$ has affinity for glycan motifs or, more broadly, that class II PVCs bind cell membranes through an alternative mechanism to the Pvc13 tail fibres.

Additionally, the N-domain of Pvc15 was predicted to be involved in nucleic acid binding and could act to regulate other PVC*pnf* components; this should be investigated using RT-qPCR to assess mRNA levels of each of the PVC*pnf* genes affected by the presence or absence of the domain. One could also conduct co-evolutionary analysis to demonstrate how Pvc15 homologues may co-evolve with their associated payload LSs.

Chapter 4 revealed many structural conservations between effector homologues as well as the Pvc15 ATPase. More distant genes, such as ClpV in

Pseudomonas, share less sequence identity due to large insertions between sites. These sequence and functional differences should be explored further, and interactions of Pvc15 proteins with their cognate effectors and their LSs would be very fruitful to investigate. Understanding how best to load the PVC, and which PVC operon to use, would be crucial information for optimising its potential in biotherapeutics.

Whilst the work of Chapter 5 had aimed to characterise new ways to assess injection of functional payloads via fluorescent readouts, altering cell binding tropisms by encoding retargeted tail fibres may bring about more meaningful results, as demonstrated by recent literature. Attempts should now be made to demonstrate proof-of-concept experiments which verify injection of therapeutic proteins or a working synthetic circuit using only PVC delivery.

Chapter 6 represents the first in-depth investigation of Pvc15 regarding its interaction with the Pnf LS, its reliance on other PVC*pnf* operon components, and its capacity to chaperone and stabilise the payload. This opens up many opportunities to follow up experiments. Firstly, it should be assessed whether different LSs, such as Cif, play a role in PVC*pnf* loading and whether this perturbs Pvc15 ATPase activity during nanosyringe assembly. Secondly, more fundamental processes of PVC action should be investigated: how payloads might refold after injection into the target cell, and whether the PVC is prone to being internalised by the cell once bound.

Finally, the capacity for certain eCIS, such as that of *Moorea producens*, to have associated antibacterial effectors may implicate these structures in binding to bacteria; such a prospect opens up a new ideal for protein delivery: expanding the definition of 'antibiotic' from small, diffusible molecules to large anti-bacterial proteins which can be packaged into eCIS for use in treating antimicrobial resistant bacterial infections. Opening the repertoire of antibiotics to any toxic protein with an associated resistance gene could transform how scientists and doctors deal with the global threat of antibiotic resistance.

Whilst detection of Pvc15 via WB seems unlikely using traditional means, one could look to develop an antibody specific to the protein to yield better binding to the protein itself. In this work, even individual domains N-D1 and D2 alone could not yield bands via WB, so other means of expression may be worthwhile such as using a cell-free system to remove degradative processes. Purification of Pvc15 would open up many opportunities to investigate specific protein-protein interactions by co-immunoprecipitation, a yeast two-hybrid assay, biophysical techniques such as isothermal titration calorimetry or microscale thermophoresis, or by proximity labelling such as by using BioID to characterise any potential interactions with core components or the elusive Pvc6 or Pvc14 tape measure protein. In addition, one could then characterise the mode of action of its ATPase activity, its hexamerised structure, its point of interaction with those components including the LS using cross-linking mass spectrometry and kinetic experiments, and how LSs may be shuttled through the pore via the pore loops.

In summary, there is much that remains elusive about the PVC 'nanosyringes' and how the payloads associate with them. This work has demonstrated the importance of Pvc15's chaperoning and ATPase activities to the packaging process, though exploring this protein further will require optimised methods to purify and assay it. With current knowledge, development of a novel synthetic biology toolkit for intracellular delivery of therapeutic proteins is plausible, though fully understanding the intricacies to these devices will prove the deciding factor as to whether they can compete with current *in situ* protein delivery methods. Both PVC research and development are at a key point in their history to seeing their use in wider applications; one should keep an eye on how and when the potential of the PVC to biotechnology may come to fruition.

7. Bibliography

Abd-Elgawad, M.M.M. (2021) '*Photorhabdus* spp.: An Overview of the Beneficial Aspects of Mutualistic Bacteria of Insecticidal Nematodes', *Plants*, 10(8), p. 1660. Available at: https://doi.org/10.3390/plants10081660.

Aepfelbacher, M. *et al.* (2003) 'Characterization of YopT effects on Rho GTPases in *Yersinia enterocolitica*-infected cells', *The Journal of Biological Chemistry*, 278(35), pp. 33217–33223. Available at: https://doi.org/10.1074/jbc.M303349200.

Alcoforado Diniz, J., Liu, Y. and Coulthurst, S.J. (2015) 'Molecular weaponry: diverse effectors delivered by the Type VI secretion system', *Cellular Microbiology*, 17(12), pp. 1742–1751. Available at: https://doi.org/10.1111/cmi.12532.

Alker, A.T. *et al.* (2020) 'Genetic examination of the marine bacterium *Pseudoalteromonas luteoviolacea* and effects of its metamorphosis-inducing factors', *Environmental microbiology*, 22(11), pp. 4689–4701. Available at: https://doi.org/10.1111/1462-2920.15211.

Antic, I. *et al.* (2015) 'Site-specific processing of Ras and Rap1 Switch I by a MARTX toxin effector domain', *Nature Communications*, 6, p. 7396. Available at: https://doi.org/10.1038/ncomms8396.

Argemi, X. *et al.* (2018) 'Comparative genomic analysis of Staphylococcus lugdunensis shows a closed pan-genome and multiple barriers to horizontal gene transfer', *BMC Genomics*, 19(1), p. 621. Available at: https://doi.org/10.1186/s12864-018-4978-1.

Baek, G.H. *et al.* (2013) 'Cdc48: A Swiss Army Knife of Cell Biology', *Journal of Amino Acids*, 2013, p. 183421. Available at: https://doi.org/10.1155/2013/183421.

Bamford, N.C. *et al.* (2020) 'Structural and biochemical characterization of the exopolysaccharide deacetylase Agd3 required for Aspergillus fumigatus biofilm formation', *Nature Communications*, 11(1), p. 2450. Available at: https://doi.org/10.1038/s41467-020-16144-5.

Barret, M. *et al.* (2011) 'Genomic analysis of the type VI secretion systems in *Pseudomonas* spp.: novel clusters and putative effectors uncovered', *Microbiology*, 157(6), pp. 1726–1739. Available at: https://doi.org/10.1099/mic.0.048645-0.

Basu, S. and Wallner, B. (2016) 'DockQ: A Quality Measure for Protein-Protein Docking Models', *PLOS ONE*, 11(8), p. e0161879. Available at: https://doi.org/10.1371/journal.pone.0161879.

Beauvois, S.G. *et al.* (2023) 'The tip protein PAAR is required for the function of the type VI secretion system', *Microbiology Spectrum*, 0(0), pp. e01478-23. Available at: https://doi.org/10.1128/spectrum.01478-23.

Bebenek, K. and Kunkel, T.A. (1995) 'Analyzing fidelity of DNA polymerases', in *Methods in Enzymology*. Academic Press (DNA Replication), pp. 217–232. Available at: https://doi.org/10.1016/0076-6879(95)62020-6.

Becker, Y. *et al.* (2022) 'Bacteria producing contractile phage tail-like particles (CPTPs) are promising alternatives to conventional pesticides', *Journal für Kulturpflanzen*, pp. 85-93 Seiten. Available at: https://doi.org/10.5073/JFK.2022.03-04.06.

Bhardwaj, P., Mitra, A.K. and Hurst, M.R.H. (2021) 'Investigating the Process of Sheath Maturation in Antifeeding Prophage: a Phage Tail-Like Protein Translocation Structure', *Journal of Bacteriology*, 203(20), pp. e00104-21. Available at: https://doi.org/10.1128/JB.00104-21.

Birukou, I. *et al.* (2013) 'The Molecular Mechanisms of Allosteric Mutations Impairing MepR Repressor Function in Multidrug-Resistant Strains of *Staphylococcus aureus*', *mBio*, 4(5), pp. e00528-13. Available at: https://doi.org/10.1128/mBio.00528-13.

Böck, D. *et al.* (2017) 'In situ architecture, function, and evolution of a contractile injection system', *Science (New York, N.Y.)*, 357(6352), pp. 713–717. Available at: https://doi.org/10.1126/science.aan7904.

Bogdanovic, X. *et al.* (2019) 'A cysteine protease-like domain enhances the cytotoxic effects of the *Photorhabdus asymbiotica* toxin PaTox', *The Journal of Biological Chemistry*, 294(3), pp. 1035–1044. Available at: https://doi.org/10.1074/jbc.RA118.005043.

Bojar, D. *et al.* (2021) 'Deep-Learning Resources for Studying Glycan-Mediated Host-Microbe Interactions', *Cell Host & Microbe*, 29(1), pp. 132-144.e3. Available at: https://doi.org/10.1016/j.chom.2020.10.004.

Bomberger, J.M. *et al.* (2011) 'A *Pseudomonas aeruginosa* Toxin that Hijacks the Host Ubiquitin Proteolytic System', *PLoS Pathogens*, 7(3), p. e1001325. Available at: https://doi.org/10.1371/journal.ppat.1001325.

Bomberger, J.M. *et al.* (2014) '*Pseudomonas aeruginosa* Cif Protein Enhances the Ubiquitination and Proteasomal Degradation of the Transporter Associated with Antigen Processing (TAP) and Reduces Major Histocompatibility Complex (MHC) Class I Antigen Presentation', *The Journal of Biological Chemistry*, 289(1), pp. 152–162. Available at: https://doi.org/10.1074/jbc.M113.459271.

Booysen, E., Malan, A.P. and Dicks, L.M.T. (2022) 'Colour of *Heterorhabditis zealandica*-infected-*Galleria mellonella* dependent on the Photorhabdus symbiont, with two new nematode-symbiotic associations reported', *Journal of Invertebrate Pathology*, 189, p. 107729. Available at: https://doi.org/10.1016/j.jip.2022.107729.

Boraston, A.B. *et al.* (2004) 'Carbohydrate-binding modules: fine-tuning polysaccharide recognition', *Biochemical Journal*, 382(3), pp. 769–781. Available at: https://doi.org/10.1042/BJ20040892.

Bravo, A. *et al.* (2011) '*Bacillus thuringiensis*: A story of a successful bioinsecticide', *Insect Biochemistry and Molecular Biology*, 41(7), pp. 423–431. Available at: https://doi.org/10.1016/j.ibmb.2011.02.006.

Brugirard-Ricaud, K. *et al.* (2005) 'Site-specific antiphagocytic function of the *Photorhabdus luminescens* type III secretion system during insect colonization', *Cellular Microbiology*, 7(3), pp. 363–371. Available at: https://doi.org/10.1111/j.1462-5822.2004.00466.x.

Bryant, F.R. (1988) 'Construction of a recombinase-deficient mutant recA protein that retains single-stranded DNA-dependent ATPase activity', *The Journal of Biological Chemistry*, 263(18), pp. 8716–8723.

Bryant, P. *et al.* (2022) 'Predicting the structure of large protein complexes using AlphaFold and Monte Carlo tree search', *Nature Communications*, 13(1), p. 6028. Available at: https://doi.org/10.1038/s41467-022-33729-4.

Buetow, L. *et al.* (2001) 'Structure of the Rho-activating domain of *Escherichia coli* cytotoxic necrotizing factor 1', *Nature Structural Biology*, 8(7), pp. 584–588. Available at: https://doi.org/10.1038/89610.

Bugreev, D.V. and Mazin, A.V. (2004) 'Ca2+ activates human homologous recombination protein Rad51 by modulating its ATPase activity', *Proceedings of the National Academy of Sciences*, 101(27), pp. 9988–9993. Available at: https://doi.org/10.1073/pnas.0402105101.

Burkholz, R., Quackenbush, J. and Bojar, D. (2021) 'Using graph convolutional neural networks to learn a representation for glycans', *Cell Reports*, 35(11), p. 109251. Available at: https://doi.org/10.1016/j.celrep.2021.109251.

Cabantous, S., Terwilliger, T.C. and Waldo, G.S. (2005) 'Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein', *Nature Biotechnology*, 23(1), pp. 102–107. Available at: https://doi.org/10.1038/nbt1044.

Campo, C.D. *et al.* (2015) 'Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function', *PLoS Genetics*, 11(10). Available at: https://doi.org/10.1371/journal.pgen.1005613.

Carim, S. *et al.* (2021) 'Systematic discovery of pseudomonad genetic factors involved in sensitivity to tailocins', *The ISME Journal*, 15(8), pp. 2289–2305. Available at: https://doi.org/10.1038/s41396-021-00921-1.

Ceroni, F. *et al.* (2018) 'Burden-driven feedback control of gene expression', *Nature Methods*, 15(5), pp. 387–393. Available at: https://doi.org/10.1038/nmeth.4635.

Chaoprasid, P. *et al.* (2021) 'Crystal structure of bacterial cytotoxic necrotizing factor CNF$_Y$ reveals molecular building blocks for intoxication', *The EMBO Journal*, 40(4). Available at: https://doi.org/10.15252/embj.2020105202.

Chavez, C.V. *et al.* (2010) 'The cyclomodulin Cif of *Photorhabdus luminescens* inhibits insect cell proliferation and triggers host cell death by apoptosis', *Microbes and Infection*, 12(14), pp. 1208–1218. Available at: https://doi.org/10.1016/j.micinf.2010.09.006.

Chen, L. *et al.* (2019) 'Genome-wide Identification and Characterization of a Superfamily of Bacterial Extracellular Contractile Injection Systems', *Cell Reports*, 29(2), pp. 511-521.e2. Available at: https://doi.org/10.1016/j.celrep.2019.08.096.

Chowdhury, S.M. *et al.* (2009) 'A method for investigating protein-protein interactions related to *Salmonella Typhimurium* pathogenesis', *Journal of proteome research*, 8(3), pp. 1504–1514. Available at: https://doi.org/10.1021/pr800865d.

Ciche, T.A. *et al.* (2003) 'Photobactin: a Catechol Siderophore Produced by *Photorhabdus luminescens*, an Entomopathogen Mutually Associated with *Heterorhabditis bacteriophora* NC1 Nematodes', *Applied and Environmental Microbiology*, 69(8), pp. 4706–4713. Available at: https://doi.org/10.1128/AEM.69.8.4706-4713.2003.

Ciche, T.A. *et al.* (2008) 'Cell Invasion and Matricide during *Photorhabdus luminescens* Transmission by *Heterorhabditis bacteriophora* Nematodes', *Applied and Environmental Microbiology*, 74(8), pp. 2275–2287. Available at: https://doi.org/10.1128/AEM.02646-07.

Clarke, D.J. (2017) 'The Regulation of Secondary Metabolism in *Photorhabdus*', in R.H. ffrench-Constant (ed.) *The Molecular Biology of Photorhabdus Bacteria*. Cham: Springer International Publishing, pp. 81–102. Available at: https://doi.org/10.1007/82_2016_21.

Collier-Hyams, L.S. *et al.* (2005) 'Cutting Edge: Bacterial Modulation of Epithelial Signaling via Changes in Neddylation of Cullin-1', *The Journal of Immunology*, 175(7), pp. 4194–4198. Available at: https://doi.org/10.4049/jimmunol.175.7.4194.

Cordin, O. *et al.* (2004) 'The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity', *The EMBO Journal*, 23(13), pp. 2478–2487. Available at: https://doi.org/10.1038/sj.emboj.7600272.

Cormack, B.P., Valdivia, R.H. and Falkow, S. (1996) 'FACS-optimized mutants of the green fluorescent protein (GFP)', *Gene*, 173(1), pp. 33–38. Available at: https://doi.org/10.1016/0378-1119(95)00685-0.

Crennell, S.J. *et al.* (2000) 'The predicted structure of photopexin from Photorhabdus shows the first haemopexin-like motif in prokaryotes', *FEMS microbiology letters*, 191(1), pp. 139–144. Available at: https://doi.org/10.1111/j.1574-6968.2000.tb09330.x.

Crow, A. *et al.* (2009) 'Crystal Structures of Cif from Bacterial Pathogens *Photorhabdus luminescens* and *Burkholderia pseudomallei*', *PLoS ONE*. Edited by B. Kobe, 4(5), p. e5582. Available at: https://doi.org/10.1371/journal.pone.0005582.

Crow, A. *et al.* (2012) 'The molecular basis of ubiquitin-like protein NEDD8 deamidation by the bacterial effector protein Cif', *Proceedings of the National Academy of Sciences*, 109(27), pp. E1830–E1838. Available at: https://doi.org/10.1073/pnas.1112107109.

Cui, J. *et al.* (2010) 'Glutamine deamidation and dysfunction of ubiquitin/NEDD8 induced by a bacterial effector family', *Science (New York, N.Y.)*, 329(5996), pp. 1215–1218. Available at: https://doi.org/10.1126/science.1193844.

Czekanska, E.M. (2011) 'Assessment of Cell Proliferation with Resazurin-Based Fluorescent Dye', in M.J. Stoddart (ed.) *Mammalian Cell Viability: Methods and Protocols*. Totowa, NJ: Humana Press, pp. 27–32. Available at: https://doi.org/10.1007/978-1-61779-108-6_5.

Daborn, P.J. *et al.* (2002) 'A single *Photorhabdus* gene, makes caterpillars floppy (mcf), allows *Escherichia coli* to persist within and kill insects', *Proceedings of the National Academy of Sciences of the United States of America*, 99(16), pp. 10742–10747. Available at: https://doi.org/10.1073/pnas.102068099.

Dawson, W.K., Fujiwara, K. and Kawai, G. (2007) 'Prediction of RNA Pseudoknots Using Heuristic Modeling with Mapping and Sequential Folding', *PLoS ONE*, 2(9), p. e905. Available at: https://doi.org/10.1371/journal.pone.0000905.

Degasperi, A. *et al.* (2014) 'Evaluating Strategies to Normalise Biological Replicates of Western Blot Data', *PLoS ONE*. Edited by K.E. Vrana, 9(1), p. e87293. Available at: https://doi.org/10.1371/journal.pone.0087293.

Derzelle, S. *et al.* (2002) 'Identification, Characterization, and Regulation of a Cluster of Genes Involved in Carbapenem Biosynthesis in Photorhabdus luminescens', *Applied and Environmental Microbiology*, 68(8), pp. 3780–3789. Available at: https://doi.org/10.1128/AEM.68.8.3780-3789.2002.

Desfosses, A. *et al.* (2019) 'Atomic structures of an entire contractile injection system in both the extended and contracted states', *Nature microbiology*, 4(11), pp. 1885–1894. Available at: https://doi.org/10.1038/s41564-019-0530-6.

Diabate, M. *et al.* (2015) 'Escherichia coli α-hemolysin counteracts the anti-virulence innate immune response triggered by the Rho GTPase activating toxin CNF1 during bacteremia', *PLoS pathogens*, 11(3), p. e1004732. Available at: https://doi.org/10.1371/journal.ppat.1004732.

Diniz, J.A. and Coulthurst, S.J. (2015) 'Intraspecies Competition in Serratia marcescens Is Mediated by Type VI-Secreted Rhs Effectors and a Conserved Effector-Associated Accessory Protein', *Journal of Bacteriology*, 197(14), p. 2350. Available at: https://doi.org/10.1128/JB.00199-15.

Dowling, A.J. *et al.* (2004) 'The insecticidal toxin Makes caterpillars floppy (Mcf) promotes apoptosis in mammalian cells', *Cellular Microbiology*, 6(4), pp. 345–353. Available at: https://doi.org/10.1046/j.1462-5822.2003.00357.x.

Dubendorf, J.W. and Studier, F.W. (1991) 'Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor', *Journal of Molecular Biology*, 219(1), pp. 45–59. Available at: https://doi.org/10.1016/0022-2836(91)90856-2.

Duchaud, E. *et al.* (2003) 'The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*', *Nature Biotechnology*, 21(11), pp. 1307–1313. Available at: https://doi.org/10.1038/nbt886.

Dunn, O.J. (1961) 'Multiple Comparisons among Means', *Journal of the American Statistical Association*, 56(293), pp. 52–64. Available at: https://doi.org/10.1080/01621459.1961.10482090.

Dutta, A. *et al.* (2018) 'Neonatal Bacteremia and Cutaneous Lesions Caused by Photorhabdus luminescens: A Rare Gram-Negative Bioluminescent Bacterium', *Journal of the Pediatric Infectious Diseases Society*, 7(3), pp. e182–e184. Available at: https://doi.org/10.1093/jpids/piy064.

Dutta, P. *et al.* (2019) 'Presence of actin binding motif in VgrG-1 toxin of Vibrio cholerae reveals the molecular mechanism of actin cross-linking', *International Journal of Biological Macromolecules*, 133, pp. 775–785. Available at: https://doi.org/10.1016/j.ijbiomac.2019.04.026.

Dvorak, P. *et al.* (2015) 'Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway', *Microbial Cell Factories*, 14(1), p. 201. Available at: https://doi.org/10.1186/s12934-015-0393-3.

Eddy, S.R. (2004) 'What is a hidden Markov model?', *Nature Biotechnology*, 22(10), pp. 1315–1316. Available at: https://doi.org/10.1038/nbt1004-1315.

Enemark, E.J. and Joshua-Tor, L. (2008) 'On helicases and other motor proteins', *Current Opinion in Structural Biology*, 18(2), pp. 243–257. Available at: https://doi.org/10.1016/j.sbi.2008.01.007.

Ericson, C.F. *et al.* (2019) 'A contractile injection system stimulates tubeworm metamorphosis by translocating a proteinaceous effector', *eLife*, 8, p. e46845. Available at: https://doi.org/10.7554/eLife.46845.

Erikstein, B.S. *et al.* (2010) 'Cellular Stress Induced by Resazurin Leads to Autophagy and Cell Death Via Production of Reactive Oxygen Species and Mitochondrial Impairment', *Journal of cellular biochemistry*, 111(3), pp. 574–584. Available at: https://doi.org/10.1002/jcb.22741.

Erzberger, J.P. and Berger, J.M. (2006) 'EVOLUTIONARY RELATIONSHIPS AND STRUCTURAL MECHANISMS OF AAA+ PROTEINS', *Annual Review of Biophysics and Biomolecular Structure*, 35(1), pp. 93–114. Available at: https://doi.org/10.1146/annurev.biophys.35.040405.101933.

Esaki, M. and Ogura, T. (2010) 'ATP-bound form of the D1 AAA domain inhibits an essential function of Cdc48p/p97This paper is one of a selection of papers published in this special issue entitled 8th International Conference on AAA Proteins and has undergone the Journal's usual peer review process.', *Biochemistry and Cell Biology*, 88(1), pp. 109–117. Available at: https://doi.org/10.1139/O09-116.

Evans, R. *et al.* (2021) *Protein complex prediction with AlphaFold-Multimer*. preprint. Bioinformatics. Available at: https://doi.org/10.1101/2021.10.04.463034.

Farmer, J.J. *et al.* (1989) '*Xenorhabdus luminescens* (DNA hybridization group 5) from human clinical specimens.', *Journal of Clinical Microbiology*, 27(7), pp. 1594–1600.

Fei, X. *et al.* (2020) 'Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate', *eLife*. Edited by J.M. Berger et al., 9, p. e52774. Available at: https://doi.org/10.7554/eLife.52774.

Ffrench-Constant, R.H. *et al.* (2000) 'A Genomic Sample Sequence of the Entomopathogenic Bacterium *Photorhabdus luminescens* W14: Potential Implications for Virulence', *Applied and Environmental Microbiology*, 66(8), pp. 3310–3329.

Fiorentini, C. *et al.* (1998) 'Toxin-Induced Activation of Rho GTP-Binding Protein Increases Bcl-2 Expression and Influences Mitochondrial Homeostasis', *Experimental Cell Research*, 242(1), pp. 341–350. Available at: https://doi.org/10.1006/excr.1998.4057.

Forst, S. *et al.* (1997) '*XENORHABDUS* AND *PHOTORHABDUS* SPP.: Bugs That Kill Bugs', *Annual Review of Microbiology*, 51(1), pp. 47–72. Available at: https://doi.org/10.1146/annurev.micro.51.1.47.

Foucault, M.-L. *et al.* (2010) 'In Vivo Bioluminescence Imaging for the Study of Intestinal Colonization by *Escherichia coli* in Mice', *Applied and Environmental Microbiology*, 76(1), pp. 264–274. Available at: https://doi.org/10.1128/AEM.01686-09.

Fraser, A. *et al.* (2021) 'Quantitative description of a contractile macromolecular machine', *Science Advances*, 7(24), p. eabf9601. Available at: https://doi.org/10.1126/sciadv.abf9601.

Garcia-Mata, R., Boulter, E. and Burridge, K. (2011) 'The invisible hand: regulation of RHO GTPases by RHOGDIs', *Nature Reviews. Molecular Cell Biology*, 12(8), pp. 493–504. Available at: https://doi.org/10.1038/nrm3153.

Ge, P. *et al.* (2015) 'Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states', *Nature Structural & Molecular Biology*, 22(5), pp. 377–382. Available at: https://doi.org/10.1038/nsmb.2995.

Geller, A.M. *et al.* (2021) 'The extracellular contractile injection system is enriched in environmental microbes and associates with numerous toxins', *Nature Communications*, 12(1), p. 3743. Available at: https://doi.org/10.1038/s41467-021-23777-7.

Gerrard, J. *et al.* (2004) 'Human infection with Photorhabdus asymbiotica: an emerging bacterial pathogen', *Microbes and Infection*, 6(2), pp. 229–237. Available at: https://doi.org/10.1016/j.micinf.2003.10.018.

Gerrard, J.G. and Stevens, R.P. (2017) 'A Review of Clinical Cases of Infection with Photorhabdus Asymbiotica', in R.H. ffrench-Constant (ed.) *The Molecular Biology of Photorhabdus Bacteria*. Cham: Springer International Publishing (Current Topics in Microbiology and Immunology), pp. 179–191. Available at: https://doi.org/10.1007/82_2016_56.

Ghosh, I., Hamilton, A.D. and Regan, L. (2000) 'Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein', *Journal of the American Chemical Society*, 122(23), pp. 5658–5659. Available at: https://doi.org/10.1021/ja994421w.

Glover, D.M. (ed.) (1985) *DNA cloning: a practical approach*. Oxford ; Washington, DC: IRL Press (Practical approach series).

Goldman, A., Harper, S. and Speicher, D.W. (2016) 'Detection of Proteins on Blot Membranes', *Current protocols in protein science*, 86, p. 10.8.1-10.8.11. Available at: https://doi.org/10.1002/cpps.15.

Green, E.R. and Mecsas, J. (2016) 'Bacterial Secretion Systems – An overview', *Microbiology spectrum*, 4(1), p. 10.1128/microbiolspec.VMBF-0012–2015. Available at: https://doi.org/10.1128/microbiolspec.VMBF-0012-2015.

Guzman, L.M. *et al.* (1995) 'Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.', *Journal of Bacteriology*, 177(14), pp. 4121–4130.

Guzun, R. *et al.* (2009) 'Regulation of respiration controlled by mitochondrial creatine kinase in permeabilized cardiac cells in situ: Importance of system level properties', *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1787(9), pp. 1089–1105. Available at: https://doi.org/10.1016/j.bbabio.2009.03.024.

Hamed, A.A., Khedr, M. and Abdelraof, M. (2020) 'Activation of LacZ gene in Escherichia coli DH5α via α-complementation mechanism for β-galactosidase production and its biochemical characterizations', *Journal of Genetic Engineering & Biotechnology*, 18, p. 80. Available at: https://doi.org/10.1186/s43141-020-00096-w.

Han, R. and Ehlers, R.-U. (2000) 'Pathogenicity, Development, and Reproduction of *Heterorhabditis bacteriophora* and *Steinernema carpocapsae* under Axenic in Vivo Conditions', *Journal of Invertebrate Pathology*, 75(1), pp. 55–58. Available at: https://doi.org/10.1006/jipa.1999.4900.

Hanson, P.I. and Whiteheart, S.W. (2005) 'AAA+ proteins: have engine, will work', *Nature Reviews Molecular Cell Biology*, 6(7), pp. 519–529. Available at: https://doi.org/10.1038/nrm1684.

Hanukoglu, I. (2015) 'Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites', *Biochemistry and Molecular Biology Education*, 43(3), pp. 206–209. Available at: https://doi.org/10.1002/bmb.20849.

Hapeshi, A. and Waterfield, N.R. (2017) '*Photorhabdus asymbiotica* as an Insect and Human Pathogen', in R.H. ffrench-Constant (ed.) *The Molecular Biology of Photorhabdus Bacteria*. Cham: Springer International Publishing, pp. 159–177. Available at: https://doi.org/10.1007/82_2016_29.

Hatch, A.L. *et al.* (2016) 'Actin filaments as dynamic reservoirs for Drp1 recruitment', *Molecular Biology of the Cell*, 27(20), pp. 3109–3121. Available at: https://doi.org/10.1091/mbc.e16-03-0193.

Heo, Y.-J. *et al.* (2007) 'R-type pyocin is required for competitive growth advantage between *Pseudomonas aeruginosa* strains', *Journal of Microbiology and Biotechnology*, 17(1), pp. 180–185.

Hilbert, B.J. *et al.* (2015) 'Structure and mechanism of the ATPase that powers viral genome packaging', *Proceedings of the National Academy of Sciences of the United States of America*, 112(29), pp. E3792–E3799. Available at: https://doi.org/10.1073/pnas.1506951112.

Hinnerwisch, J. *et al.* (2005) 'Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation', *Cell*, 121(7), pp. 1029–1041. Available at: https://doi.org/10.1016/j.cell.2005.04.012.

Hodge, R.G. and Ridley, A.J. (2016) 'Regulating Rho GTPases and their regulators', *Nature Reviews Molecular Cell Biology*, 17(8), pp. 496–510. Available at: https://doi.org/10.1038/nrm.2016.67.

Hoffmann, C. *et al.* (2004) 'The *Yersinia pseudotuberculosis* Cytotoxic Necrotizing Factor (CNFY) Selectively Activates RhoA', *Journal of Biological Chemistry*, 279(16), pp. 16026–16032. Available at: https://doi.org/10.1074/jbc.M313556200.

Holm, L. (2022) 'Dali server: structural unification of protein families', *Nucleic Acids Research*, 50(W1), pp. W210–W215. Available at: https://doi.org/10.1093/nar/gkac387.

Holm, L. and Rosenström, P. (2010) 'Dali server: conservation mapping in 3D', *Nucleic Acids Research*, 38(Web Server issue), pp. W545–W549. Available at: https://doi.org/10.1093/nar/gkq366.

Howard, S.A. *et al.* (2021) 'The Breadth and Molecular Basis of Hcp-Driven Type VI Secretion System Effector Delivery', *mBio* [Preprint]. Available at: https://doi.org/10.1128/mBio.00262-21.

Hsu, Y. *et al.* (2008) 'Structure of the Cyclomodulin Cif from Pathogenic *Escherichia coli*', *Journal of molecular biology*, 384(2), pp. 465–477. Available at: https://doi.org/10.1016/j.jmb.2008.09.051.

Hurst, M.R.H. *et al.* (2007) 'Isolation and characterization of the Serratia entomophila antifeeding prophage', *FEMS Microbiology Letters*, 270(1), pp. 42–48. Available at: https://doi.org/10.1111/j.1574-6968.2007.00645.x.

Hurst, M.R.H. *et al.* (2018) '*Serratia proteamaculans* Strain AGR96X Encodes an Antifeeding Prophage (Tailocin) with Activity against Grass Grub (Costelytra giveni) and Manuka Beetle (Pyronota Species) Larvae', *Applied and Environmental Microbiology*, 84(10). Available at: https://doi.org/10.1128/AEM.02739-17.

Hurst, M.R.H., Glare, T.R. and Jackson, T.A. (2004) 'Cloning *Serratia entomophila* Antifeeding Genes—a Putative Defective Prophage Active against the Grass Grub *Costelytra zealandica*', *Journal of Bacteriology*, 186(15), pp. 5116–5128. Available at: https://doi.org/10.1128/JB.186.15.5116-5128.2004.

James, J.R. (2018) 'Tuning ITAM multiplicity on T-cell receptors can control potency and selectivity to ligand density', *Science signaling*, 11(531). Available at: https://doi.org/10.1126/scisignal.aan1088.

James, J.R. and Vale, R.D. (2012) 'Biophysical mechanism of T-cell receptor triggering in a reconstituted system', *Nature*, 487(7405), pp. 64–69. Available at: https://doi.org/10.1038/nature11220.

Jank, T. *et al.* (2013) 'A bacterial toxin catalyzing tyrosine glycosylation of Rho and deamidation of Gq and Gi proteins', *Nature Structural & Molecular Biology*, 20(11), pp. 1273–1280. Available at: https://doi.org/10.1038/nsmb.2688.

Jank, T. *et al.* (2015) 'Tyrosine glycosylation of Rho by *Yersinia* toxin impairs blastomere cell behaviour in zebrafish embryos', *Nature Communications*, 6(1), p. 7807. Available at: https://doi.org/10.1038/ncomms8807.

Janska, H., Kwasniak, M. and Szczepanowska, J. (2013) 'Protein quality control in organelles — AAA/FtsH story', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1833(2), pp. 381–387. Available at: https://doi.org/10.1016/j.bbamcr.2012.03.016.

Jensen, M.A., Fukushima, M. and Davis, R.W. (2010) 'DMSO and betaine greatly improve amplification of GC-rich constructs in de novo synthesis', *PloS One*, 5(6), p. e11024. Available at: https://doi.org/10.1371/journal.pone.0011024.

Jeruzalmi, D., O'Donnell, M. and Kuriyan, J. (2001) 'Crystal Structure of the Processivity Clamp Loader Gamma (γ) Complex of *E. coli* DNA Polymerase III', *Cell*, 106(4), pp. 429–441. Available at: https://doi.org/10.1016/S0092-8674(01)00463-9.

Jessop, M., Felix, J. and Gutsche, I. (2021) 'AAA+ ATPases: structural insertions under the magnifying glass', *Current Opinion in Structural Biology*, 66, pp. 119–128. Available at: https://doi.org/10.1016/j.sbi.2020.10.027.

Jiang, F. *et al.* (2019) 'Cryo-EM Structure and Assembly of an Extracellular Contractile Injection System', *Cell*, 177(2), pp. 370-383.e15. Available at: https://doi.org/10.1016/j.cell.2019.02.020.

Jiang, F. *et al.* (2022) 'N-terminal signal peptides facilitate the engineering of PVC complex as a potent protein delivery system', *Science Advances*, 8(17), p. eabm2343. Available at: https://doi.org/10.1126/sciadv.abm2343.

Johansson-Åkhe, I. and Wallner, B. (2022) 'Improving peptide-protein docking with AlphaFold-Multimer using forced sampling', *Frontiers in Bioinformatics*, 2, p. 959160. Available at: https://doi.org/10.3389/fbinf.2022.959160.

Jubelin, G. *et al.* (2009) 'Cycle Inhibiting Factors (CIFs) Are a Growing Family of Functional Cyclomodulins Present in Invertebrate and Mammal Bacterial Pathogens', *PLoS ONE*, 4(3), p. e4855. Available at: https://doi.org/10.1371/journal.pone.0004855.

Jubelin, G. *et al.* (2010) 'Pathogenic Bacteria Target NEDD8-Conjugated Cullins to Hijack Host-Cell Signaling Pathways', *PLoS Pathogens*, 6(9), p. e1001128. Available at: https://doi.org/10.1371/journal.ppat.1001128.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. Available at: https://doi.org/10.1038/s41586-021-03819-2.

Jurėnas, D., Rosa, L.T., *et al.* (2021) 'Mounting, structure and autocleavage of a type VI secretion-associated Rhs polymorphic toxin', *Nature Communications*, 12(1), p. 6998. Available at: https://doi.org/10.1038/s41467-021-27388-0.

Jurėnas, D., Payelleville, A., *et al.* (2021) 'Photorhabdus antibacterial Rhs polymorphic toxin inhibits translation through ADP-ribosylation of 23S ribosomal RNA', *Nucleic Acids Research*, 49(14), pp. 8384–8395. Available at: https://doi.org/10.1093/nar/gkab608.

Jurėnas, D. and Journet, L. (2021) 'Activity, delivery, and diversity of Type VI secretion effectors', *Molecular Microbiology*, 115(3), pp. 383–394. Available at: https://doi.org/10.1111/mmi.14648.

Kalnins, A. *et al.* (1983) 'Sequence of the lacZ gene of Escherichia coli.', *The EMBO Journal*, 2(4), pp. 593–597.

Khan, Y.A., White, K.I. and Brunger, A.T. (2022) 'The AAA+ superfamily: a review of the structural and mechanistic principles of these molecular machines', *Critical Reviews in Biochemistry and Molecular Biology*, 57(2), pp. 156–187. Available at: https://doi.org/10.1080/10409238.2021.1979460.

Kim, D.M. *et al.* (2013) 'ATPase active-site electrostatic interactions control the global conformation of the 100 kDa SecA translocase', *Journal of the American Chemical Society*, 135(8), pp. 2999–3010. Available at: https://doi.org/10.1021/ja306361q.

Kim, J. and Cheong, J.-H. (2020) 'Role of Mitochondria-Cytoskeleton Interactions in the Regulation of Mitochondrial Structure and Function in Cancer Stem Cells', *Cells*, 9(7), p. 1691. Available at: https://doi.org/10.3390/cells9071691.

Knust, Z. and Schmidt, G. (2010) 'Cytotoxic Necrotizing Factors (CNFs)−A Growing Toxin Family', *Toxins*, 2(1), pp. 116–127. Available at: https://doi.org/10.3390/toxins2010116.

Koskiniemi, S. *et al.* (2013) 'Rhs proteins from diverse bacteria mediate intercellular competition', *Proceedings of the National Academy of Sciences of the United States of America*, 110(17), pp. 7032–7037. Available at: https://doi.org/10.1073/pnas.1300627110.

Kowarz, E., Löscher, D. and Marschalek, R. (2015) 'Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines', *Biotechnology Journal*, 10(4), pp. 647–653. Available at: https://doi.org/10.1002/biot.201400821.

Kreitz, J. *et al.* (2023) 'Programmable protein delivery with a bacterial contractile injection system', *Nature*, pp. 1–8. Available at: https://doi.org/10.1038/s41586-023-05870-7.

Kube, S. and Wendler, P. (2015) 'Structural comparison of contractile nanomachines', *AIMS Biophysics*, 2(2), pp. 88–115. Available at: https://doi.org/10.3934/biophy.2015.2.88.

Kudryashev, M. *et al.* (2015) 'Structure of the type VI secretion system contractile sheath', *Cell*, 160(5), pp. 952–962. Available at: https://doi.org/10.1016/j.cell.2015.01.037.

Lamoureux, A. (2018) *Why Some Wounded Civil War Soldiers Actually Glowed In The Dark – And Lived Because Of It*, *All That's Interesting*. Available at: https://allthatsinteresting.com/angels-glow (Accessed: 28 October 2020).

Lazic, S.E., Clarke-Williams, C.J. and Munafò, M.R. (2018) 'What exactly is "N" in cell culture and animal experiments?', *PLoS Biology*, 16(4), p. e2005282. Available at: https://doi.org/10.1371/journal.pbio.2005282.

Leclerc, G.M. *et al.* (2000) 'Development of a Destabilized Firefly Luciferase Enzyme for Measurement of Gene Expression', *BioTechniques*, 29(3), pp. 590–601. Available at: https://doi.org/10.2144/00293rr02.

Lee, F.K.N. *et al.* (1999) 'The R-Type Pyocin of Pseudomonas aeruginosa C Is a Bacteriophage Tail-Like Particle That Contains Single-Stranded DNA', *Infection and Immunity*, 67(2), pp. 717–725.

Lee, G. *et al.* (2016) 'F-Type Bacteriocins of Listeria monocytogenes: a New Class of Phage Tail-Like Structures Reveals Broad Parallel Coevolution between Tailed Bacteriophages and High-Molecular-Weight Bacteriocins', *Journal of Bacteriology*, 198(20), pp. 2784–2793. Available at: https://doi.org/10.1128/JB.00489-16.

Lee, Y. *et al.* (2019) 'Makes caterpillars floppy-like effector-containing MARTX toxins require host ADP-ribosylation factor (ARF) proteins for systemic pathogenicity', *Proceedings of the National Academy of Sciences*, 116(36), pp. 18031–18040. Available at: https://doi.org/10.1073/pnas.1905095116.

Lee, Y.-W. *et al.* (2019) 'Protein Delivery into the Cell Cytosol using Non-Viral Nanocarriers', *Theranostics*, 9(11), pp. 3280–3292. Available at: https://doi.org/10.7150/thno.34412.

Leiman, P.G. *et al.* (2009) 'Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin', *Proceedings of the National Academy of Sciences*, 106(11), pp. 4154–4159. Available at: https://doi.org/10.1073/pnas.0813360106.

Lin, C.-C. *et al.* (2016) 'Structural Insights into the Allosteric Operation of the Lon AAA+ Protease', *Structure*, 24(5), pp. 667–675. Available at: https://doi.org/10.1016/j.str.2016.03.001.

Lin, J.J. (1992) 'Endonuclease A degrades chromosomal and plasmid DNA of *Escherichia coli* present in most preparations of single stranded DNA from phagemids', *Proceedings of the National Science Council, Republic of China. Part B, Life Sciences*, 16(1), pp. 1–5.

Linder, P. *et al.* (1989) 'Birth of the D-E-A-D box', *Nature*, 337(6203), pp. 121–122. Available at: https://doi.org/10.1038/337121a0.

Linder, P. (2006) 'Dead-box proteins: a family affair—active and passive players in RNP-remodeling', *Nucleic Acids Research*, 34(15), pp. 4168–4180. Available at: https://doi.org/10.1093/nar/gkl468.

Liu, X., Rao, L. and Gennerich, A. (2020) 'The regulatory function of the AAA4 ATPase domain of cytoplasmic dynein', *Nature Communications*, 11, p. 5952. Available at: https://doi.org/10.1038/s41467-020-19477-3.

Lo, Y.-H. *et al.* (2019) 'Cryo-EM structure of the essential ribosome assembly AAA-ATPase Rix7', *Nature Communications*, 10(1), p. 513. Available at: https://doi.org/10.1038/s41467-019-08373-0.

Lundstrøm, J. *et al.* (2022) 'LectinOracle: A Generalizable Deep Learning Model for Lectin–Glycan Binding Prediction', *Advanced Science*, 9(1), p. 2103807. Available at: https://doi.org/10.1002/advs.202103807.

Ma, J. *et al.* (2017) 'PAAR-Rhs proteins harbor various C-terminal toxins to diversify the antibacterial pathways of type VI secretion systems', *Environmental Microbiology*, 19(1), pp. 345–360. Available at: https://doi.org/10.1111/1462-2920.13621.

Machado, R.A.R. *et al.* (2018) 'Whole-genome-based revisit of *Photorhabdus* phylogeny: proposal for the elevation of most *Photorhabdus* subspecies to the species level and description of one novel species Photorhabdus bodei sp. nov., and one novel subspecies *Photorhabdus laumondii* subsp. *clarkei* subsp. nov.', *International Journal of Systematic and Evolutionary Microbiology*, 68(8), pp. 2664–2681. Available at: https://doi.org/10.1099/ijsem.0.002820.

Machado, R.A.R. *et al.* (2020) '*Photorhabdus heterorhabditis* subsp. *aluminescens* subsp. nov., *Photorhabdus heterorhabditis* subsp. *heterorhabditis* subsp. nov., *Photorhabdus australis* subsp. *thailandensis* subsp. nov., *Photorhabdus australis* subsp. *australis* subsp. nov., and *Photorhabdus aegyptia* sp. nov. isolated from *Heterorhabditis* entomopathogenic nematodes', *International Journal of Systematic and Evolutionary Microbiology*, 71(1), p. 004610. Available at: https://doi.org/10.1099/ijsem.0.004610.

Malik, E.M. and Müller, C.E. (2016) 'Anthraquinones As Pharmacological Tools and Drugs', *Medicinal Research Reviews*, 36(4), pp. 705–748. Available at: https://doi.org/10.1002/med.21391.

Manicassamy, S. *et al.* (2010) 'Activation of β-Catenin in Dendritic Cells Regulates Immunity Versus Tolerance in the Intestine', *Science (New York, N.Y.)*, 329(5993), pp. 849–853. Available at: https://doi.org/10.1126/science.1188510.

Marschall, A.L.J. (2021) 'Targeting the Inside of Cells with Biologicals: Chemicals as a Delivery Strategy', *BioDrugs*, 35(6), pp. 643–671. Available at: https://doi.org/10.1007/s40259-021-00500-y.

Mazal, H. *et al.* (2021) 'Ultrafast pore-loop dynamics in a AAA+ machine point to a Brownian-ratchet mechanism for protein translocation', *Science Advances*, 7(36), p. eabg4674. Available at: https://doi.org/10.1126/sciadv.abg4674.

McInerney, J.O., McNally, A. and O'Connell, M.J. (2017) 'Why prokaryotes have pangenomes', *Nature Microbiology*, 2(4), p. 17040. Available at: https://doi.org/10.1038/nmicrobiol.2017.40.

Menéndez-Conejero, R. *et al.* (2017) 'Structure of a Reptilian Adenovirus Reveals a Phage Tailspike Fold Stabilizing a Vertebrate Virus Capsid', *Structure*, 25(10), pp. 1562-1573.e5. Available at: https://doi.org/10.1016/j.str.2017.08.007.

Meusch, D. *et al.* (2014) 'Mechanism of Tc toxin action revealed in molecular detail', *Nature*, 508(7494), pp. 61–65. Available at: https://doi.org/10.1038/nature13015.

Michel-Briand, Y. and Baysse, C. (2002) 'The pyocins of *Pseudomonas aeruginosa*', *Biochimie*, 84(5–6), pp. 499–510. Available at: https://doi.org/10.1016/s0300-9084(02)01422-0.

Miller, J.M. and Enemark, E.J. (2016) 'Fundamental Characteristics of AAA+ Protein Family Structure and Function', *Archaea*, 2016, p. 9294307. Available at: https://doi.org/10.1155/2016/9294307.

Mirdita, M. *et al.* (2021) *ColabFold - Making protein folding accessible to all*, p. 2021.08.15.456425. Available at: https://doi.org/10.1101/2021.08.15.456425.

Miyashiro, T. and Ruby, E.G. (2012) 'Shedding light on bioluminescence regulation in *Vibrio fischeri*', *Molecular Microbiology*, 84(5), pp. 795–806. Available at: https://doi.org/10.1111/j.1365-2958.2012.08065.x.

Monjarás Feria, J. and Valvano, M.A. (2020) 'An Overview of Anti-Eukaryotic T6SS Effectors', *Frontiers in Cellular and Infection Microbiology*, 10. Available at: https://doi.org/10.3389/fcimb.2020.584751.

Morikawa, H. *et al.* (2010) 'The bacterial effector Cif interferes with SCF ubiquitin ligase function by inhibiting deneddylation of Cullin1', *Biochemical and Biophysical Research Communications*, 401(2), pp. 268–274. Available at: https://doi.org/10.1016/j.bbrc.2010.09.048.

Mout, R. *et al.* (2017) 'In Vivo Delivery of CRISPR/Cas9 for Therapeutic Gene Editing: Progress and Challenges', *Bioconjugate chemistry*, 28(4), pp. 880–884. Available at: https://doi.org/10.1021/acs.bioconjchem.7b00057.

Mulley, G. *et al.* (2015) 'From Insect to Man: *Photorhabdus* Sheds Light on the Emergence of Human Pathogenicity', *PLOS ONE*. Edited by M. Skurnik, 10(12), p. e0144937. Available at: https://doi.org/10.1371/journal.pone.0144937.

Nakamura, F. (2013) 'FilGAP and its close relatives: a mediator of Rho–Rac antagonism that regulates cell morphology and migration', *Biochemical Journal*, 453(1), pp. 17–25. Available at: https://doi.org/10.1042/BJ20130290.

Naldini, L., Blömer, U., Gage, F.H., *et al.* (1996) 'Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector.', *Proceedings of the National Academy of Sciences of the United States of America*, 93(21), pp. 11382–11388.

Naldini, L., Blömer, U., Gallay, P., *et al.* (1996) 'In Vivo Gene Delivery and Stable Transduction of Nondividing Cells by a Lentiviral Vector', *Science*, 272(5259), pp. 263–267. Available at: https://doi.org/10.1126/science.272.5259.263.

Ndolo Obonyo, D. *et al.* (2019) 'Research and Development of Biopesticides: Challenges and Prospects', *Outlooks on Pest Management*, 30, pp. 267–276. Available at: https://doi.org/10.1564/v30_dec_08.

Neuwald, A.F. *et al.* (1999) 'AAA+: A Class of Chaperone-Like ATPases Associated with the Assembly, Operation, and Disassembly of Protein Complexes', *Genome Research*, 9(1), pp. 27–43. Available at: https://doi.org/10.1101/gr.9.1.27.

Nguyen, V.S. *et al.* (2018) 'Towards a complete structural deciphering of Type VI secretion system', *Current Opinion in Structural Biology*, 49, pp. 77–84. Available at: https://doi.org/10.1016/j.sbi.2018.01.007.

Nirwan, N. *et al.* (2019) 'Structure-based mechanism for activation of the AAA+ GTPase McrB by the endonuclease McrC', *Nature Communications*, 10(1), p. 3058. Available at: https://doi.org/10.1038/s41467-019-11084-1.

Nougayrède, J.-P. *et al.* (2005) 'Cyclomodulins: bacterial effectors that modulate the eukaryotic cell cycle', *Trends in Microbiology*, 13(3), pp. 103–110. Available at: https://doi.org/10.1016/j.tim.2005.01.002.

O'Connor, K. and Chen, M. (2013) 'Dynamic functions of RhoA in tumor cell migration and invasion', *Small GTPases*, 4(3), pp. 141–147. Available at: https://doi.org/10.4161/sgtp.25131.

Oja, S.M. *et al.* (2014) 'Fluorescence-Enabled Electrochemical Microscopy with Dihydroresorufin as a Fluorogenic Indicator', *Analytical Chemistry*, 86(12), pp. 6040–6048. Available at: https://doi.org/10.1021/ac501194j.

Olszewski, M.M. *et al.* (2019) 'The Cdc48 unfoldase prepares well-folded protein substrates for degradation by the 26S proteasome', *Communications Biology*, 2(1), pp. 1–8. Available at: https://doi.org/10.1038/s42003-019-0283-z.

Orozco, R.A. *et al.* (2016) 'Bioprospecting for secondary metabolites in the entomopathogenic bacterium *Photorhabdus luminescens* subsp. *sonorensis*', *Journal of Invertebrate Pathology*, 141, pp. 45–52. Available at: https://doi.org/10.1016/j.jip.2016.09.008.

Oswald, E. *et al.* (2005) 'Bacterial toxins that modulate host cell-cycle progression', *Current Opinion in Microbiology*, 8(1), pp. 83–91. Available at: https://doi.org/10.1016/j.mib.2004.12.011.

Page, A.N. *et al.* (2011) 'Structure and Biochemical Activities of Escherichia coli MgsA', *The Journal of Biological Chemistry*, 286(14), pp. 12075–12085. Available at: https://doi.org/10.1074/jbc.M110.210187.

Parihar, R.D. *et al.* (2022) 'Heterorhabditis and Photorhabdus Symbiosis: A Natural Mine of Bioactive Compounds', *Frontiers in Microbiology*, 13, p. 790339. Available at: https://doi.org/10.3389/fmicb.2022.790339.

Park, Y.H. *et al.* (2016) 'Pyrin Inflammasome Activation and RhoA Signaling in the Autoinflammatory Diseases FMF and HIDS', *Nature immunology*, 17(8), pp. 914–921. Available at: https://doi.org/10.1038/ni.3457.

Park, Y.-J. *et al.* (2018) 'Structure of the type VI secretion system TssK–TssF–TssG baseplate subcomplex revealed by cryo-electron microscopy', *Nature Communications*, 9(1), p. 5385. Available at: https://doi.org/10.1038/s41467-018-07796-5.

Pedulla, M.L. *et al.* (2003) 'Origins of Highly Mosaic Mycobacteriophage Genomes', *Cell*, 113(2), pp. 171–182. Available at: https://doi.org/10.1016/S0092-8674(03)00233-2.

Pell, G. *et al.* (2004) 'Structural and Biochemical Analysis of Cellvibrio japonicus Xylanase 10C: HOW VARIATION IN SUBSTRATE-BINDING CLEFT INFLUENCES THE CATALYTIC PROFILE OF FAMILY GH-10 XYLANASES * [boxs]', *Journal of Biological Chemistry*, 279(12), pp. 11777–11788. Available at: https://doi.org/10.1074/jbc.M311947200.

Pérez-Martin, J. and de Lorenzo, V. (1996) 'VTR expression cassettes for engineering conditional phenotypes in Pseudomonas: activity of the Pu promoter of the TOL plasmid under limiting concentrations of the XylR activator protein', *Gene*, 172(1), pp. 81–86. Available at: https://doi.org/10.1016/0378-1119(96)00193-X.

Pettersen, E.F. *et al.* (2004) 'UCSF Chimera—A visualization system for exploratory research and analysis', *Journal of Computational Chemistry*, 25(13), pp. 1605–1612. Available at: https://doi.org/10.1002/jcc.20084.

Pettersen, E.F. *et al.* (2021) 'UCSF ChimeraX: Structure visualization for researchers, educators, and developers', *Protein Science*, 30(1), pp. 70–82. Available at: https://doi.org/10.1002/pro.3943.

Prentki, P. (1992) 'Nucleotide sequence of the classical lacZ deletion ΔM15', *Gene*, 122(1), pp. 231–232. Available at: https://doi.org/10.1016/0378-1119(92)90056-U.

Pryor, J.L., Xu, W. and Hamilton, D.W. (1992) 'Immunodetection after complete destaining of Coomassie blue-stained proteins on immobilon-PVDF', *Analytical Biochemistry*, 202(1), pp. 100–104. Available at: https://doi.org/10.1016/0003-2697(92)90213-Q.

Puchades, C., Sandate, C.R. and Lander, G.C. (2020) 'The molecular principles governing the activity and functional diversity of AAA+ proteins', *Nature reviews. Molecular cell biology*, 21(1), pp. 43–58. Available at: https://doi.org/10.1038/s41580-019-0183-6.

R Core Team (2020) 'R: A language and environment for statistical computing.' Available at: http://www.r-project.org/index.html.

Reich, K.A. and Schoolnik, G.K. (1996) 'Halovibrin, secreted from the light organ symbiont Vibrio fischeri, is a member of a new class of ADP-ribosyltransferases.', *Journal of Bacteriology*, 178(1), pp. 209–215.

ResearchAndMarkets.com (2020) *Biopesticides Market by Type, Source, Mode of Application, Formulation, Crop Application And Region - Global Forecast to 2025*. Available at: https://www.reportbuyer.com/product/4364825/ (Accessed: 30 October 2020).

Rivera-Ramírez, A. *et al.* (2022) 'Comparative Genomics and Pathogenicity Analysis of Two Bacterial Symbionts of Entomopathogenic Nematodes: The Role of the GroEL Protein in Virulence', *Microorganisms*, 10(3), p. 486. Available at: https://doi.org/10.3390/microorganisms10030486.

Robichon, C. *et al.* (2011) 'Engineering *Escherichia coli* BL21(DE3) Derivative Strains To Minimize E. coli Protein Contamination after Purification by Immobilized Metal Affinity Chromatography', *Applied and Environmental Microbiology*, 77(13), pp. 4634–4646. Available at: https://doi.org/10.1128/AEM.00119-11.

Rocchi, I. *et al.* (2019) 'A Bacterial Phage Tail-like Structure Kills Eukaryotic Cells by Injecting a Nuclease Effector', *Cell Reports*, 28(2), pp. 295-301.e4. Available at: https://doi.org/10.1016/j.celrep.2019.06.019.

Roderer, D. *et al.* (2019) 'Towards the application of Tc toxins as a universal protein translocation system', *Nature Communications*, 10(1), p. 5263. Available at: https://doi.org/10.1038/s41467-019-13253-8.

Rodou, A., Ankrah, D.O. and Stathopoulos, C. (2010) 'Toxins and secretion systems of *Photorhabdus luminescens*', *Toxins*, 2(6), pp. 1250–1264. Available at: https://doi.org/10.3390/toxins2061250.

Rojas, M.I. *et al.* (2020) 'A Distinct Contractile Injection System Gene Cluster Found in a Majority of Healthy Adult Human Microbiomes', *mSystems*, 5(4), pp. e00648-20. Available at: https://doi.org/10.1128/mSystems.00648-20.

Rothman, K.J. (1990) 'No Adjustments Are Needed for Multiple Comparisons', *Epidemiology*, 1(1), pp. 43–46.

RStudio Team (2020) 'RStudio: Integrated Development for R'.

Russo, T.A. *et al.* (2005) 'E. coli virulence factor hemolysin induces neutrophil apoptosis and necrosis/lysis in vitro and necrosis/lysis and lung injury in a rat pneumonia model', *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 289(2), pp. L207–L216. Available at: https://doi.org/10.1152/ajplung.00482.2004.

Rybakova, D. *et al.* (2013) 'Role of antifeeding prophage (Afp) protein Afp16 in terminating the length of the Afp tailocin and stabilizing its sheath', *Molecular Microbiology*, 89(4), pp. 702–714. Available at: https://doi.org/10.1111/mmi.12305.

Rybakova, D. *et al.* (2015) 'Afp14 is involved in regulating the length of Anti-feeding prophage (Afp)', *Molecular Microbiology*, 96(4), pp. 815–826. Available at: https://doi.org/10.1111/mmi.12974.

Rybakova, D., Mitra, A. and Hurst, M. (2014) 'Purification and TEM of Afp and Its Variants', *BIO-PROTOCOL*, 4(10). Available at: https://doi.org/10.21769/BioProtoc.1132.

Santin, Y.G. *et al.* (2019) 'Cell Width Dictates Type VI Secretion Tail Length', *Current Biology*, 29(21), pp. 3707-3713.e3. Available at: https://doi.org/10.1016/j.cub.2019.08.058.

Santos Amaral, L. *et al.* (2022) 'Heterorhabditis bacteriophora (Rhabditida: Heterorhabditidae), isolate HP88, induces reproductive and physiological alterations in Biomphalaria glabrata (Gastropoda: Planorbidae): an alternative for biological control of schistosomiasis', *Acta Tropica*, 230, p. 106396. Available at: https://doi.org/10.1016/j.actatropica.2022.106396.

Sarris, P.F. *et al.* (2014) 'A Phage Tail-Derived Element with Wide Distribution among Both Prokaryotic Domains: A Comparative Genomic and Phylogenetic Study', *Genome Biology and Evolution*, 6(7), pp. 1739–1747. Available at: https://doi.org/10.1093/gbe/evu136.

Schaefer, A., Reinhard, N.R. and Hordijk, P.L. (2014) 'Toward understanding RhoGTPase specificity: structure, function and local activation', *Small GTPases*, 5(2), p. e968004. Available at: https://doi.org/10.4161/21541248.2014.968004.

Schilbach, S. *et al.* (2017) 'Structures of transcription pre-initiation complex with TFIIH and Mediator', *Nature*, 551(7679), pp. 204–209. Available at: https://doi.org/10.1038/nature24282.

Schindelin, J. *et al.* (2012) 'Fiji: an open-source platform for biological-image analysis', *Nature Methods*, 9(7), pp. 676–682. Available at: https://doi.org/10.1038/nmeth.2019.

Scholl, D. *et al.* (2009) 'An Engineered R-Type Pyocin Is a Highly Specific and Sensitive Bactericidal Agent for the Food-Borne Pathogen Escherichia coli O157:H7', *Antimicrobial Agents and Chemotherapy* [Preprint]. Available at: https://doi.org/10.1128/AAC.01660-08.

Seybert, A. *et al.* (2006) 'Communication between subunits within an archaeal clamp-loader complex', *The EMBO Journal*, 25(10), pp. 2209–2218. Available at: https://doi.org/10.1038/sj.emboj.7601093.

Shao, F. *et al.* (2002) 'A Yersinia Effector and a Pseudomonas Avirulence Protein Define a Family of Cysteine Proteases Functioning in Bacterial Pathogenesis', *Cell*, 109(5), pp. 575–588. Available at: https://doi.org/10.1016/S0092-8674(02)00766-3.

Sheets, J. and Aktories, K. (2017) 'Insecticidal Toxin Complexes from *Photorhabdus luminescens*', in R.H. ffrench-Constant (ed.) *The Molecular Biology of Photorhabdus Bacteria*. Cham: Springer International Publishing, pp. 3–23. Available at: https://doi.org/10.1007/82_2016_55.

Shi, Y.-M. and Bode, H. (2018) 'Chemical language and warfare of bacterial natural products in bacteria–nematode–insect interactions', *Natural Product Reports*, p. 28.

Shikuma, N.J. *et al.* (2014) 'Marine Tubeworm Metamorphosis Induced by Arrays of Bacterial Phage Tail–Like Structures', *Science (New York, N.Y.)*, 343(6170), p. 529. Available at: https://doi.org/10.1126/science.1246794.

Shin, M. *et al.* (2020) 'Structural basis for distinct operational modes and protease activation in AAA+ protease Lon', *Science Advances*, 6(21), p. eaba8404. Available at: https://doi.org/10.1126/sciadv.aba8404.

Shneider, M.M. *et al.* (2013) 'PAAR-repeat proteins sharpen and diversify the type VI secretion system spike', *Nature*, 500(7462), pp. 350–353. Available at: https://doi.org/10.1038/nature12453.

Sievers, F. *et al.* (2011) 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular Systems Biology*, 7, p. 539. Available at: https://doi.org/10.1038/msb.2011.75.

Silva, C.P. *et al.* (2002) 'Bacterial infection of a model insect: *Photorhabdus luminescens* and *Manduca sexta*', *Cellular Microbiology*, 4(6), pp. 329–339. Available at: https://doi.org/10.1046/j.1462-5822.2002.00194.x.

da Silva, W.J. *et al.* (2020) 'The great potential of entomopathogenic bacteria *Xenorhabdus* and *Photorhabdus* for mosquito control: a review', *Parasites & Vectors*, 13(1), p. 376. Available at: https://doi.org/10.1186/s13071-020-04236-6.

Silverman, J.M. *et al.* (2013) 'Haemolysin Co-regulated Protein is an Exported Receptor and Chaperone of Type VI Secretion Substrates', *Molecular cell*, 51(5), p. 10.1016/j.molcel.2013.07.025. Available at: https://doi.org/10.1016/j.molcel.2013.07.025.

Simpson, P.J. *et al.* (2000) 'The Structural Basis for the Ligand Specificity of Family 2 Carbohydrate-binding Modules *', *Journal of Biological Chemistry*, 275(52), pp. 41137–41142. Available at: https://doi.org/10.1074/jbc.M006948200.

Snider, J., Thibault, G. and Houry, W.A. (2008) 'The AAA+ superfamily of functionally diverse proteins', *Genome Biology*, 9(4), p. 216. Available at: https://doi.org/10.1186/gb-2008-9-4-216.

Somvanshi, V.S. *et al.* (2010) '*Photorhabdus* phase variants express a novel fimbrial locus, mad, essential for symbiosis', *Molecular Microbiology*, 77(4), pp. 1021–1038. Available at: https://doi.org/10.1111/j.1365-2958.2010.07270.x.

Somvanshi, V.S. *et al.* (2012) 'A Single Promoter Inversion Switches *Photorhabdus* Between Pathogenic and Mutualistic States', *Science*, 337(6090), pp. 88–93. Available at: https://doi.org/10.1126/science.1216641.

Sosnowski, P. *et al.* (2018) 'The CryoEM structure of the Saccharomyces cerevisiae ribosome maturation factor Rea1', *eLife*, 7, p. e39163. Available at: https://doi.org/10.7554/eLife.39163.

Sprague, L.D. and Neubauer, H. (2014) 'Genome Sequence of Yersinia similis Y228T, a Member of the Yersinia pseudotuberculosis Complex', *Genome Announcements*, 2(2), pp. e00216-14. Available at: https://doi.org/10.1128/genomeA.00216-14.

Stabb, E.V., Reich, K.A. and Ruby, E.G. (2001) 'Vibrio fischeri Genes hvnA andhvnB Encode Secreted NAD+-Glycohydrolases', *Journal of Bacteriology*, 183(1), pp. 309–317. Available at: https://doi.org/10.1128/JB.183.1.309-317.2001.

Starr, T. *et al.* (2018) 'The phorbol 12-myristate-13-acetate differentiation protocol is critical to the interaction of THP-1 macrophages with Salmonella Typhimurium', *PLOS ONE*, 13(3), p. e0193601. Available at: https://doi.org/10.1371/journal.pone.0193601.

Studer, S.M. and Joseph, S. (2006) 'Unfolding of mRNA Secondary Structure by the Bacterial Translation Initiation Complex', *Molecular Cell*, 22(1), pp. 105–115. Available at: https://doi.org/10.1016/j.molcel.2006.02.014.

Suryawanshi, A. *et al.* (2020) 'Wnt Signaling Cascade in Dendritic Cells and Regulation of Anti-tumor Immunity', *Frontiers in Immunology*, 11, p. 122. Available at: https://doi.org/10.3389/fimmu.2020.00122.

Tada, S. (2007) 'Cdt1 and geminin: role during cell cycle progression and DNA damage in higher eukaryotes', *Frontiers in Bioscience*, 12(1), p. 1629. Available at: https://doi.org/10.2741/2175.

Tai, W., Zhao, P. and Gao, X. (2020) 'Cytosolic delivery of proteins by cholesterol tagging', *Science Advances*, 6(25), p. eabb0310. Available at: https://doi.org/10.1126/sciadv.abb0310.

Taieb, F. *et al.* (2006) '*Escherichia coli* cyclomodulin Cif induces G2 arrest of the host cell cycle without activation of the DNA-damage checkpoint-signalling pathway', *Cellular Microbiology*, 8(12), pp. 1910–1921. Available at: https://doi.org/10.1111/j.1462-5822.2006.00757.x.

Taieb, F., Nougayrède, J.-P. and Oswald, E. (2011) 'Cycle Inhibiting Factors (Cifs): Cyclomodulins That Usurp the Ubiquitin-Dependent Degradation Pathway of Host Cells', *Toxins*, 3(4), pp. 356–368. Available at: https://doi.org/10.3390/toxins3040356.

Talbot, J.D. *et al.* (2008) 'Rapid, stimulation-induced reduction of C12-resorufin in motor nerve terminals: linkage to mitochondrial metabolism', *Journal of Neurochemistry*, 105(3), pp. 807–819. Available at: https://doi.org/10.1111/j.1471-4159.2007.05176.x.

Taylor, N.M.I., van Raaij, M.J. and Leiman, P.G. (2018) 'Contractile injection systems of bacteriophages and related systems', *Molecular Microbiology*, 108(1), pp. 6–15. Available at: https://doi.org/10.1111/mmi.13921.

Thomas, W. *et al.* (2001) 'Cytotoxic Necrotizing Factor from Escherichia coli Induces RhoA-Dependent Expression of the Cyclooxygenase-2 Gene', *Infection and Immunity*, 69(11), pp. 6839–6845. Available at: https://doi.org/10.1128/IAI.69.11.6839-6845.2001.

Thomès, L., Burkholz, R. and Bojar, D. (2021) 'Glycowork: A Python package for glycan data science and machine learning', *Glycobiology*, 31(10), pp. 1240–1244. Available at: https://doi.org/10.1093/glycob/cwab067.

Unciuleac, M.-C., Smith, P.C. and Shuman, S. (2016) 'Crystal Structure and Biochemical Characterization of a Mycobacterium smegmatis AAA-Type Nucleoside Triphosphatase Phosphohydrolase (Msm0858)', *Journal of Bacteriology*, 198(10), pp. 1521–1533. Available at: https://doi.org/10.1128/JB.00905-15.

Unni, R. *et al.* (2022) 'Presence and absence of type VI secretion systems in bacteria: This article is part of the Bacterial Cell Envelopes collection.', *Microbiology*, 168(4). Available at: https://doi.org/10.1099/mic.0.001151.

Uratani, Y. and Hoshino, T. (1984) 'Pyocin R1 inhibits active transport in Pseudomonas aeruginosa and depolarizes membrane potential', *Journal of Bacteriology*, 157(2), pp. 632–636. Available at: https://doi.org/10.1128/jb.157.2.632-636.1984.

Venditto, I. *et al.* (2016) 'Complexity of the Ruminococcus flavefaciens cellulosome reflects an expansion in glycan recognition', *Proceedings of the National Academy of Sciences*, 113(26), pp. 7136–7141. Available at: https://doi.org/10.1073/pnas.1601558113.

Vladimirov, M. *et al.* (2023) 'A contractile injection system is required for developmentally regulated cell death in Streptomyces coelicolor', *Nature Communications*, 14(1), p. 1469. Available at: https://doi.org/10.1038/s41467-023-37087-7.

Vlisidou, I. *et al.* (2019) 'The *Photorhabdus asymbiotica* virulence cassettes deliver protein effectors directly into target eukaryotic cells', *eLife*, 8, p. e46259. Available at: https://doi.org/10.7554/eLife.46259.

Wald, J. *et al.* (2022) 'Mechanism of AAA+ ATPase-mediated RuvAB–Holliday junction branch migration', *Nature*, 609(7927), pp. 630–639. Available at: https://doi.org/10.1038/s41586-022-05121-1.

Wang, Q. *et al.* (2003) 'D1 Ring Is Stable and Nucleotide-independent, whereas D2 Ring Undergoes Major Conformational Changes during the ATPase Cycle of p97-VCP', *Journal of Biological Chemistry*, 278(35), pp. 32784–32793. Available at: https://doi.org/10.1074/jbc.M303869200.

Wang, R. and Brattain, M.G. (2007) 'The maximal size of protein to diffuse through the nuclear pore is larger than 60kDa', *FEBS letters*, 581(17), pp. 3164–3170. Available at: https://doi.org/10.1016/j.febslet.2007.05.082.

Wang, T. *et al.* (2015) 'Type VI Secretion System Transports Zn2+ to Combat Multiple Stresses and Host Immunity', *PLoS Pathogens*, 11(7), p. e1005020. Available at: https://doi.org/10.1371/journal.ppat.1005020.

Wang, X. *et al.* (2020) 'The *Photorhabdus* Virulence Cassettes RRSP-Like Effector Interacts With Cyclin-Dependent Kinase 1 and Causes Mitotic Defects in Mammalian Cells', *Frontiers in Microbiology*, 11, p. 366. Available at: https://doi.org/10.3389/fmicb.2020.00366.

Wang, X. *et al.* (2021) 'Characterization of *Photorhabdus* Virulence Cassette as a causative agent in the emerging pathogen Photorhabdus asymbiotica', *Science China Life Sciences* [Preprint]. Available at: https://doi.org/10.1007/s11427-021-1955-4.

Waterfield, N.R. *et al.* (2008) 'Rapid Virulence Annotation (RVA): Identification of virulence factors using a bacterial genome library and multiple invertebrate hosts', *Proceedings of the National Academy of Sciences of the United States of America*, 105(41), pp. 15967–15972. Available at: https://doi.org/10.1073/pnas.0711114105.

Waterhouse, A.M. *et al.* (2009) 'Jalview Version 2—a multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25(9), pp. 1189–1191. Available at: https://doi.org/10.1093/bioinformatics/btp033.

Watson, J.D. and Milner-White, E.J. (2002) 'A novel main-chain anion-binding site in proteins: the nest. A particular combination of φ,ψ values in successive residues gives rise to anion-binding sites that occur commonly and are found often at functionally important regions11Edited by J. Thornton', *Journal of Molecular Biology*, 315(2), pp. 171–182. Available at: https://doi.org/10.1006/jmbi.2001.5227.

Weiss, G.L. *et al.* (2022) 'Structure of a thylakoid-anchored contractile injection system in multicellular cyanobacteria', *Nature Microbiology*, pp. 1–11. Available at: https://doi.org/10.1038/s41564-021-01055-y.

Weissfeld, A.S. *et al.* (2005) '*Photorhabdus asymbiotica*, a Pathogen Emerging on Two Continents That Proves that There Is No Substitute for a Well-Trained Clinical Microbiologist', *Journal of Clinical Microbiology*, 43(8), pp. 4152–4155. Available at: https://doi.org/10.1128/JCM.43.8.4152-4155.2005.

Welinder, C. and Ekblad, L. (2011) 'Coomassie Staining as Loading Control in Western Blot Analysis', *Journal of Proteome Research*, 10(3), pp. 1416–1419. Available at: https://doi.org/10.1021/pr1011476.

Wickham, H. (2007) 'Reshaping data with the reshape package', 21(12). Available at: http://www.jstatsoft.org/v21/i12/paper.

Wickham, H. (2016) 'ggplot2: Elegant Graphics for Data Analysis'. Available at: https://ggplot2.tidyverse.org.

Wolf, D.H. and Stolz, A. (2012) 'The Cdc48 machine in endoplasmic reticulum associated protein degradation', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1823(1), pp. 117–124. Available at: https://doi.org/10.1016/j.bbamcr.2011.09.002.

Xiao, Y. and Wu, K. (2019) 'Recent progress on the interaction between insects and Bacillus thuringiensis crops', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1767), p. 20180316. Available at: https://doi.org/10.1098/rstb.2018.0316.

Xu, J. *et al.* (2022) 'Identification and structure of an extracellular contractile injection system from the marine bacterium Algoriphagus machipongonensis', *Nature Microbiology*, 7(3), pp. 397–410. Available at: https://doi.org/10.1038/s41564-022-01059-2.

Yang, G. *et al.* (2006) '*Photorhabdus* virulence cassettes confer injectable insecticidal activity against the wax moth', *Journal of Bacteriology*, 188(6), pp. 2254–2261. Available at: https://doi.org/10.1128/JB.188.6.2254-2261.2006.

Ye, Q. *et al.* (2015) 'TRIP13 is a protein-remodeling AAA+ ATPase that catalyzes MAD2 conformation switching', *eLife*. Edited by A.T. Brunger, 4, p. e07367. Available at: https://doi.org/10.7554/eLife.07367.

You, J. *et al.* (2006) 'Nutritive significance of crystalline inclusion proteins of *Photorhabdus luminescens* in *Steinernema* nematodes: Crystalline inclusion proteins of Photorabdus luminescens', *FEMS Microbiology Ecology*, 55(2), pp. 178–185. Available at: https://doi.org/10.1111/j.1574-6941.2005.00015.x.

Zadeh, J.N. *et al.* (2011) 'NUPACK: Analysis and design of nucleic acid systems', *Journal of Computational Chemistry*, 32(1), pp. 170–173. Available at: https://doi.org/10.1002/jcc.21596.

Zhang, Z. *et al.* (2021) 'PAAR Proteins Are Versatile Clips That Enrich the Antimicrobial Weapon Arsenals of Prokaryotes', *mSystems* [Preprint]. Available at: https://doi.org/10.1128/mSystems.00953-21.

Zhou, Q. *et al.* (2019) 'Molecular mechanism of polyketide shortening in anthraquinone biosynthesis of Photorhabdus luminescens †Electronic supplementary information (ESI) available. See DOI: 10.1039/c9sc00749k', *Chemical Science*, 10(25), pp. 6341–6349. Available at: https://doi.org/10.1039/c9sc00749k.

Zinke, M., Schröder, G.F. and Lange, A. (2022) 'Major tail proteins of bacteriophages of the order Caudovirales', *Journal of Biological Chemistry*, 298(1), p. 101472. Available at: https://doi.org/10.1016/j.jbc.2021.101472.

8. Appendices

Please see attached documents: 'Appendices' and the Supplementary file 'Effector_function_table.xlsx'