



University
of Glasgow

Horsham, Matthew Robert (2011) *Structural and functional analysis of pneumococcal histidine triad D from Streptococcus pneumoniae*. PhD thesis.

<http://theses.gla.ac.uk/2639/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Structural and Functional Analysis of
Pneumococcal Histidine Triad D from
Streptococcus pneumoniae

By

Matthew Robert Horsham



UNIVERSITY
of
GLASGOW

A thesis submitted for the degree of Doctor of Philosophy
in the College of Medical, Veterinary and Life Sciences,
University of Glasgow

Declaration

This thesis has been written in accordance with the University of Glasgow regulations. It has not been presented for a degree at any other university and is the original work of the author except where indicated otherwise by reference in the text. The work contained within is the authors own except where work was done in collaboration as indicated.

© Matthew Robert Horsham, 11th January 2011.

Signed

Date

.....

.....

Matthew Robert Horsham

11th January 2011.

“The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but 'That's funny...’”

Isaac Asimov

This thesis is dedicated to my Grandparents.

Abstract

The pathogenic bacteria *Streptococcus pneumoniae* is one of the major causes of morbidity and mortality in humans in the world today. A Gram-positive facultative anaerobe, under natural conditions it exists as a commensal bacteria residing in the nasopharynx. Upon invasion of the body, *S. pneumoniae* can cause a number of diseases, which range in severity from acute otitis-media to pneumonia, septicaemia and meningitis.

The main virulence factor of *S. pneumoniae* has been identified as the polysaccharide capsule, which coats the outside of the cell. *S. pneumoniae* can be categorised into 92 distinct serotypes based on capsular composition. Current available vaccines utilise a mixture of polysaccharides from the most prevalent capsular serotypes, or capsular polysaccharide conjugated to a carrier protein. Vaccine coverage is therefore serotype specific. Furthermore, vaccine efficiency is lower in those groups most at risk of infection; namely infants, the elderly and the immuno-compromised. As a result investigation is ongoing into a next generation of protein-based vaccines that can provide increased coverage and efficiency.

A novel family of proteins termed Pneumococcal Histidine Triad (Pht) proteins were identified from a whole genome antigen-screen as potential candidates for inclusion in a novel protein-based pneumococcal vaccine.¹ Animal models of infection have shown that immunisation with the Pht protein Pneumococcal Histidine Triad D (PhtD) confers protection against invasion of *S. pneumoniae*.

PhtD was cloned, expressed, purified and subjected to crystallisation trials in an attempt to uncover the function of PhtD through determining the protein structure by macromolecular X-ray crystallography, which yielded some rudimentary crystallographic data. Biophysical analysis of PhtD using a variety of techniques including limited proteolysis and circular dichroism revealed that PhtD exclusively bound the divalent-cation Zn^{2+} and that Zn^{2+} -binding induced a major conformational change in the protein structure, which proved to be a reversible process. A rationalised, targeted analysis of PhtD protein structure by limited proteolysis, Nuclear Magnetic Resonance (NMR),

N-terminal sequencing and mass-spectrometry revealed localised, ordered regions of structure within the protein sequence that were highly stable. These identified protein fragments were subsequently cloned, expressed, purified and subjected to crystallisation trials. Due to their smaller and more ordered nature, it was postulated that these PhtD fragments may prove more readily crystallisable than the full-length molecule. Initial crystals have been obtained for these protein fragments and are being optimised to improve crystal size and quality. Evaluation of the PhtD proteolysis products by Western blotting with anti-PhtD antibody has revealed the dominant epitope for the PhtD protein, localised to a 15 kDa region in the C-terminal half of the PhtD protein sequence. This could be a major advancement in development of a protein based vaccine as only the 15 kDa epitope-containing region need be included in order to elicit an antibody reaction. Furthermore, the small size of the protein fragment is highly conducive to structural determination by a variety of methods. This 15 kDa epitope fragment has been cloned into an expression plasmid allowing recombinant protein expression in order to investigate further.

Additionally, as training for handling of PhtD X-ray diffraction data, macromolecular X-ray crystallography was attempted with a variety of different proteins from Gram-positive bacteria. Two proteins -a novel *cis-trans* isomerase PpmA from *S. pneumoniae*, and the transketolase TktA from *Lactobacillus salivarius*- were successfully crystallised. Diffraction quality crystals of TktA were grown that produced X-ray diffraction to 2.3Å resolution. The structure of TktA was successfully determined using the molecular replacement method. Diffraction quality crystals of PpmA were grown that show X-ray diffraction to 2.5Å resolution, and optimisation of crystallisation conditions should yield better X-ray diffraction, allowing the structure of PpmA to be determined. The data pertaining to these proteins is also included as part of this thesis.

Acknowledgements

Firstly and foremostly, I would like to thank my supervisor Professor Tim Mitchell for the opportunity to work in his lab on this project, and the support and guidance he has provided throughout. I would also like to thank the Biotechnology and Biological Sciences Research Council (BBSRC) for the funding of this project. Thank you to the past and present members of the Mitchell Pneumococcal Group, including Denise Candlish for making the lab such an enjoyable atmosphere to work in, and for helping hands when needed.

I am especially indebted to Alan Riboldi-Tunncliffe for initiating me into the world of proteins and X-ray crystallography, for his patience in answering many inane questions and for (nearly always) managing to get the first round in, no matter how hard I tried. Thanks are also due to Prof. Neil Isaacs and members of the PX-lab for their patience with me in learning a challenging but interesting aspect of science, and to the members of the MDR group at GlaxoSmithKline, in particular Don Somers for making the time during my placement so enjoyable and productive. Thanks are also due to Brian Smith for help with NMR spectroscopy, Richard Burchmore for help with mass-spectrometry, Sharon Kelly for help with CD, and to Margaret Nutley for assistance with ITC.

I have to say a heartfelt thank you to my parents Janette and Graham for their constant support and encouragement throughout this project and before, to my partner Harriet for her love, understanding and support, and to you all for having belief in my abilities, even when I may have doubted them myself.

Abbreviations

Ab	Antibody
ABC	ATP-Binding Cassette
AUC	Analytical Ultra Centrifugation
ATP	Adenosine Triphosphate
CD	Circular Dichroism
Da	Dalton
dH ₂ O	Distilled water
DLS	Dynamic Light Scattering
DNA	Deoxy-Ribonucleic Acid
DNase	Deoxy-Ribonuclease
dNTP	Deoxy-ribonucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
GSK	GlaxoSmithKline
GOI	Gene Of Interest
HRV-C3	Human Rhinoviral C3 protease
IMAC	Immobilised Metal-ion Affinity Chromatography
IEC	Ion-Exchange Chromatography
IPD	Invasive Pneumococcal Disease
IPTG	Isopropyl- β -D-thiogalactopyranoside
ITC	Isothermal Titration micro-Calorimetry
K	Temperature in Kelvin

LB	Luria Broth
LN ₂	Liquid nitrogen
LPS	Lipopolysaccharide
MALDI-ToF	Matrix Assisted Desorption/Ionisation Time-of-Flight
MES	2-(N-morpholino) ethane sulphonic acid
MS-MS	Tandem mass-spectrometry
MWT	Molecular weight
MWCO	Molecular Weight Cut-Off
NMR	Nuclear Magnetic Resonance
OPPF	Oxford Protein Production Facility
ORF	Open Reading Frame
PDB	Protein Data Bank
PEG	Polyethylene glycol
PMF	Peptide Mass Fingerprinting
POI	Protein Of Interest
PPIase	Peptidyl - Proline Isomerase
rpm	Revolutions per minute
SDS-PAGE	Sodium-Dodecyl-Sulphate Polyacrylamide Gel Electrophoresis
Se-Met	Selenomethionine
TAE	Tris Acetate EDTA
TB-Onyx	O/N Express instant TB media
TE	Tris EDTA
T _m	Melting temperature

UV	Ultra-violet
x g	relative centrifugal force in X multiples of gravity
°C	Temperature in degrees Celsius

Contents

Declaration	ii
Abstract.....	iv
Acknowledgements	vi
Abbreviations	vii
Contents	x
List Of Figures	xv
List Of Tables	xviii
<u>Chapter 1</u> : Introduction	1
1.1 <i>Streptococcus pneumoniae</i>	1
1.1.1 Morphology	1
1.1.2 Pneumococcal Carriage	2
1.1.3 Pneumococcal Disease	3
1.1.4 Pneumococcal Virulence Factors	7
1.2 Pneumococcal Vaccines.....	11
1.2.1 At Risk Groups/Vaccine Coverage	11
1.2.2 Evolution of the Pneumococcal vaccine.....	12
1.2.3 Pneumococcal Vaccine Types.....	13
1.2.4 Limitations Of Current Vaccines.....	14
1.2.5 Next-Generation Vaccine Development	15
1.3 Pneumococcal Histidine Triad Proteins.....	17
1.3.1 Identification As A Vaccine Candidate	17
1.3.2 Sequence Variation	18
1.3.3 Putative Role	19
1.4 Pneumococcal Histidine Triad D (PhtD)	23
1.5 Project Aims	24
<u>Chapter 2</u> : Materials And Methods.....	25
2.1 Materials	25
2.1.1 Chemicals	25
2.1.2 Enzymes	25
2.1.3 Competent Cells	25
2.1.4 Oligonucleotides	25
2.1.5 Chromatography Supplies	25
2.1.6 Crystallographic Supplies.....	26
2.2 Media And Buffers	26
2.2.1 Molecular Biology Buffers	26
2.2.2 Growth Media	27
2.2.3 Antibiotics	28
2.2.4 Isopropyl β -D-thiogalactopyranoside (IPTG)	28
2.3 Bacterial Strains And Vectors	28
2.3.1 Storage	28
2.3.2 Bacterial Strains	29
2.3.3 Plasmid DNA	29
2.4 Laboratory Techniques.....	31
2.4.1 General Laboratory Techniques	31
2.5 Polymerase Chain Reaction (PCR)	31
2.5.1 Preparation Of Genomic DNA	31
2.5.2 PCR Primers.....	31
2.5.3 dNTPs	32
2.5.4 PCR.....	32
2.6 DNA Manipulation	34

2.6.1	Agarose Gel Electrophoresis	34
2.6.2	DNA Purification	35
2.7	Protein Expression	36
2.7.1	Bacterial Transformation	36
2.7.2	Large Scale Culture Growth	36
2.7.3	Cell Lysis	37
2.8	Protein Purification.....	37
2.8.1	General Notes	37
2.8.2	Ni-NTA (nitrilotriacetic acid) IMAC Chromatography	37
2.8.3	Ion Exchange Chromatography.....	38
2.8.4	Size Exclusion Chromatography.....	39
2.9	Protein Manipulation	40
2.9.1	Cleavage Of His ₆ Tag With C3-Protease.	40
2.9.2	Buffer Exchange Of Protein Samples	40
2.9.3	Concentration Of Protein.....	40
2.10	Protein Analysis.....	41
2.10.1	SDS-PAGE	41
2.10.2	Measurement Of Protein Concentration	41
2.10.3	Western Blot	42
2.10.4	Dynamic Light Scattering (DLS)	42
2.10.5	Isothermal Titration Microcalorimetry (ITC)	42
2.10.6	Mass Spectrometry	43
2.10.7	N-terminal Protein Sequencing.....	43
2.10.8	Nuclear Magnetic Resonance	43
2.10.9	Circular Dichroism (CD)	43
2.10.10	Protein Melting Temperature (T _m)	44
2.10.11	Analytical Ultra-Centrifugation (AUC)	44
2.10.12	Limited Proteolysis	44
2.11	Protein Crystallisation	44
2.11.1	Crystal Screening.....	44
2.11.2	Crystal Screens	45
2.12	Computational Software.....	46
Chapter 3 : Crystallographic Studies Of Proteins From Gram Positive Bacteria		
.....		48
3.1	Overview	48
Expression, Purification, Crystallisation And Data Collection Of TktA from <i>Lactobacillus salivarius</i>		
.....		51
3.2	Reference	51
3.3	Sequence Search Of The PDB With TktA	51
3.4	Overexpression Of TktA From <i>L. salivarius</i>	51
3.4.1	TktA Clone	51
3.4.2	Transformation And Expression of TktA	52
3.5	Purification Of TktA	53
3.5.1	IMAC Ni-Affinity Chromatography	53
3.5.2	Concentration Of Purified TktA.....	55
3.6	Crystallisation And Data Collection Of TktA	55
3.6.1	In-House Testing Of TktA Crystals.....	58
3.6.2	TktA Crystallographic Data Acquisition And Processing.....	58
Structural Determination Of TktA		61
3.7	Reference	61
3.8	Molecular Replacement	61
3.8.1	Solution To The Phase Problem.....	61
3.8.2	Choice Of Molecular Replacement Model	62
3.8.3	Molecular Replacement Of TktA.....	63

3.8.4	Model Building And Refinement	63
3.8.5	Final Model Refinement And Analysis.....	64
	Expression, Purification, Crystallisation And Data Collection Of PpmA From <i>S. pneumoniae</i>	69
3.9	Reference	69
3.10	Introduction	69
3.11	Overexpression Of PpmA From <i>S. pneumoniae</i>	70
3.11.1	PpmA Clone	70
3.11.2	Transformation And Expression Of PpmA.....	71
3.12	Purification Of Recombinant PpmA.....	71
3.12.1	Cell Lysis.....	71
3.12.2	Initial IMAC Ni-Affinity Purification.....	72
3.12.3	Ion Exchange Chromatography Of PpmA.....	73
3.12.4	Concentration And Storage Of Purified PpmA	75
3.13	Crystallisation And Data Collection Of PpmA	75
3.13.1	In-House Testing Of PpmA Crystals	76
3.13.2	Data Collection And Processing of PpmA	77
	Chapter 4 : Cloning, Expression And Purification Of Pneumococcal Histidine Triad D (PhtD) From <i>Streptococcus pneumoniae</i>	80
4.1	Introduction.....	80
4.2	Human Rhinovirus-C3 (HRV-C3) Protease	80
4.2.1	General Notes	80
4.2.2	Bacterial Transformation	81
4.2.3	Preparation Of Starter Culture	81
4.2.4	Large Scale Expression Of HRV-C3 Protease	81
4.2.5	Purification of HRV-C3 Protease.....	82
4.2.6	Size Exclusion Chromatography.....	82
4.2.7	HRV-C3 Protease Storage	83
4.3	Construction Of PhtD Expression Plasmid	83
4.3.1	Production Of <i>S. pneumoniae</i> gDNA	83
4.3.2	PhtD Nucleotide Sequence	83
4.3.3	Primer Design And Amplification Of PhtD Gene.....	84
4.3.4	Generation Of PhtD Expression Plasmid.....	86
4.4	Expression Of F/L PhtD	89
4.4.1	Transformation.....	89
4.4.2	Protein Expression	89
4.5	Purification Of Recombinant F/L PhtD	89
4.5.1	Initial IMAC Ni-Affinity Chromatography	89
4.5.2	C3-Cleavage Of His ₆ Tag & Reverse Ni-Affinity Chromatography. 91	
4.5.3	Ion Exchange Chromatography (Anion Exchange).....	91
4.5.4	Concentration Of Purified F/L PhtD	93
4.6	Construction Of Truncated Forms of PhtD	93
4.6.1	Overview	93
4.6.2	Fragment Design	93
4.6.3	Primer Design And Amplification Of PhtD Fragments	94
4.6.4	Amplification Of PhtD Fragments	95
4.6.5	Generation Of pMRH Expression Constructs	96
4.7	Expression Of Truncated Forms Of PhtD	103
4.8	Purification Of Recombinant Truncated Forms Of PhtD	103
	Chapter 5 : Biophysical Characterisation Of PhtD	105
5.1	Limited Proteolysis Of F/L PhtD	105
5.1.1	Preliminary Investigation	105
5.1.2	Limited Proteolysis Of Apo- F/L PhtD	106
5.1.3	Limited Proteolysis Of Zn-Bound F/L PhtD.....	108

5.1.4	Limited Proteolysis Of F/L PhtD In The Presence Of Selected Divalent-Cations	111
5.2	Mass-Spectrometry Analysis Of F/L PhtD Limited Proteolysis Profiles 112	
5.2.1	Overview	112
5.2.2	Intact-Mass MALDI-ToF Mass Spectrometry Of Limited Proteolysis Profiles (Peptide Mass Fingerprinting).....	113
5.2.3	Western Blot Analysis Of PhtD	118
5.3	Analysis Of Purified PhtD Proteins By MALDI-ToF Mass-Spectrometry 120	
5.3.1	Overview	120
5.3.2	Preparation Of PhtD Samples For Mass-Spectrometry Analysis ..	121
5.3.3	Analysis Of Purified PhtD Proteins	121
5.4	Analysis Of PhtD Limited Proteolysis Profiles By N-Terminal Sequencing	125
5.4.1	Overview	125
5.4.2	Sample Preparation	126
5.4.3	N-terminal sequencing Analysis Of Apo- F/L PhtD And Zn ²⁺ -bound F/L PhtD.....	127
5.5	ITC Analysis Of Zn ²⁺ Binding With PhtD	130
<u>Chapter 6</u> : Structural Analysis Of PhtD		135
6.1	Overview	135
6.2	Preliminary Investigation	135
6.3	One-Dimensional Nuclear Magnetic Resonance (1D-NMR) Spectroscopy Of PhtD.....	139
6.3.1	Initial 1D-NMR Analysis Of Apo- F/L PhtD	139
6.3.2	1D-NMR Analysis Of Zn ²⁺ Binding	142
6.4	Analysis Of PhtD By Circular Dichroism (CD).....	152
6.4.1	CD Analysis of F/L PhtD.....	153
6.4.2	CD Analysis Of C-term and ΔC-term PhtD Fragments	157
6.5	Crystallisation Of PhtD	160
6.5.1	Crystallisation Of F/L PhtD.....	160
6.5.2	Crystallisation Of C-term And ΔC-term PhtD Fragments	170
6.5.3	PhtD Crystallisation In Collaboration With CSIC Madrid	171
<u>Chapter 7</u> : Final Discussion		177
7.1	Structural Studies Of PhtD.....	177
7.1.1	Crystallisation Of F/L PhtD.....	177
7.1.2	Crystallisation Of PhtD Protein Fragments.....	177
7.1.3	Interaction Of Zn ²⁺ With PhtD	178
7.1.4	Hypothetical Structure of PhtD.....	179
7.1.5	PhtD Hypothesised Mode Of Action	180
7.1.6	PhtD As A Vaccine Candidate	181
7.1.7	Further Work	181
7.2	Towards The Structure Of PpmA	182
7.2.1	Summary Of Results Obtained	182
7.2.2	Future Work	182
Appendices		183
A.1	DNA And Protein Sequences.....	183
A.1.1	PhtD (SP_1003) TIGR4 Natural Sequences	183
A.1.2	pMRH1 PhtD F/L Sequences.....	185
A.1.3	pMRH2 C-term PhtD Truncate Sequences.....	187
A.1.4	pMRH3 ΔC-term PhtD Truncate Sequences	188
A.1.5	pMRH4 5.3 kDa C-terminal PhtD Truncate Sequences	189
A.1.6	pMRH6 34 kDa PhtD Truncate Sequences	190

A.1.7	pMRH7 PhtD Epitope Truncate Sequences	191
A.1.8	PpmA (SP_0981) TIGR4 Natural Sequences	192
A.1.9	PpmA Expression Construct Sequences.....	193
A.1.10	TktA (LSL_1946) L. salivarius Natural Sequences	194
A.1.11	TktA Expression Construct Sequences	195
A.2	DNA And Protein Molecular Weight Standards	197
A.2.1	DNA Molecular Weight Standards	197
A.2.2	Protein Molecular Weight Standards.....	198
A.3	Selected Standard Operating Protocols (SOPs) Devised By The Mitchell Group.....	199
A.3.1	Preparation Of Pneumococcal Genomic DNA.....	199
A.3.2	Western Blot	201
A.4	Molecular Biology Kit Protocols.....	203
A.4.1	QIAquick [®] PCR Purification Kit Protocol	203
A.4.2	QIAquick [®] Gel Extraction Kit Protocol.....	203
A.4.3	QIAprep [®] Miniprep Kit Protocol	204
A.5	Protein Manipulation Protocols	206
A.5.1	PD-10 Desalting Column Protocol	206
A.5.2	Ni-Affinity IMAC Template	207
A.5.3	Anion Exchange Chromatography Template.....	207
	Conference Contributions.....	208
	Publications	209
	TktA Crystallisation Paper	209
	PpmA Crystallisation Paper (Draft Manuscript)	212
	Manuscripts In Preparation.....	216
	References	217

List Of Figures

Figure 1.1: Scanning electron micrograph of <i>S. pneumoniae</i> at x14,500 magnification ²	1
Figure 1.2: Incidence of serious pneumococcal disease worldwide	4
Figure 1.3: Incidence of mortality resulting from serious pneumococcal disease	5
Figure 1.4: Pneumococcal virulence factors	8
Figure 1.5: Worldwide implementation of pneumococcal conjugate vaccine	12
Figure 1.6: Sequence alignment of Pht proteins	19
Figure 1.7: 1.2Å Crystal structure of PhtA fragment	20
Figure 1.8: Zn ²⁺ binding in 1.2Å crystal structure of PhtA, illustrating Zn ²⁺ -ligating residues	21
Figure 1.9: PhtD protein sequence from the <i>S. pneumoniae</i> strain TIGR4 ...	23
Figure 2.1: pOPINF expression vector map	30
Figure 2.2: Outline of a general PCR experiment	34
Figure 3.1: TktA expression construct	52
Figure 3.2: Elution profile for Ni-affinity purification of TktA	54
Figure 3.3: SDS-PAGE gel of selected fractions from Ni-affinity purification of TktA.....	54
Figure 3.4: Typical hexagonal rod TktA crystals.....	57
Figure 3.5: Rhombic plate TktA crystals from PEG/Ion2 36	57
Figure 3.6: Diffraction image for a single crystal of TktA	59
Figure 3.7: Close-up diffraction image for a single crystal of TktA	59
Figure 3.8: Sequence alignment of TktA and 1ITZ	63
Figure 3.9: Modelling of TktA structure to electron density.....	64
Figure 3.10: Ramachandran plot for refined TktA structure.	65
Figure 3.11: Monomeric molecular model of refined TktA structure	66
Figure 3.12: Molecular model of TktA	66
Figure 3.13: Crystal packing in TktA	67
Figure 3.14: Illustration of 2-fold symmetry for TktA	68
Figure 3.15: PpmA expression construct and protein translation.....	70
Figure 3.16: Typical Ni-affinity trace for purification of PpmA.....	72
Figure 3.17: Typical SDS-PAGE of PpmA after Ni-affinity purification.....	73
Figure 3.18: Typical anion exchange trace for PpmA	74
Figure 3.19: Typical SDS-PAGE of PpmA post-anion exchange	74
Figure 3.20: PpmA crystals from PEG/Ion 2.....	76
Figure 3.21: In-House X-ray diffraction image for a single crystal of PpmA .	77
Figure 4.1: SDS-PAGE of purified C3-protease	82
Figure 4.2: PhtD protein sequence (downloaded from CMR database)	84
Figure 4.3: 0.8% Agarose gel showing successful amplification of F/L <i>PhtD</i> gene.....	86
Figure 4.4: pMRH1 F/L PhtD expression construct	88
Figure 4.5: Typical Ni-affinity trace for F/L PhtD.....	90
Figure 4.6: SDS-PAGE of F/L PhtD from Ni-affinity purification	90
Figure 4.7: Typical anion exchange trace of F/L PhtD	92
Figure 4.8: SDS-PAGE for anion exchange of F/L PhtD	92
Figure 4.9: Schematic representation of PhtD proteins.....	94
Figure 4.10: 0.8% agarose gel showing products from PCR amplification of PhtD gene and fragments.....	96
Figure 4.11: pMRH2 expression construct	98
Figure 4.12: pMRH3 expression construct	99

Figure 4.13: pMRH4 expression construct	100
Figure 4.14: pMRH6 expression construct	101
Figure 4.15: pMRH7 expression construct	102
Figure 4.16: Recombinantly expressed and purified PhtD proteins	104
Figure 5.1: Limited proteolysis of Apo-F/L PhtD	107
Figure 5.2: Limited proteolysis of Zn ²⁺ -bound F/L PhtD.....	109
Figure 5.3: Comparison of Apo- and Zn ²⁺ -bound F/L PhtD limited proteolysis	110
Figure 5.4: Limited proteolysis of F/L PhtD in the presence of selected divalent cations	111
Figure 5.5: SDS-PAGE profile of F/L PhtD tryptic digest subjected to PMF .	114
Figure 5.6: PMF analysis of 1min digest sample	115
Figure 5.7: PMF analysis of 3min digest sample	115
Figure 5.8: PMF analysis of 49 kDa fragment excised from SDS-PAGE gel ...	116
Figure 5.9: PMF analysis of 34 kDa band excised from SDS-PAGE gel	117
Figure 5.10: Western blot of Apo- F/L PhtD	119
Figure 5.11: Coomassie stained gel corresponding to western blot (Fig 5.10)	119
Figure 5.12: Sequence of predicted dominant PhtD antibody-binding region	120
Figure 5.13: SDS-PAGE of purified PhtD samples for mass-spectrometry ...	122
Figure 5.14: Complete mass-spectrometry results for purified F/L PhtD ...	123
Figure 5.15: Complete mass-spectrometry results for PhtD C-term fragment	123
Figure 5.16: Complete mass-spectrometry results for ΔC-term PhtD fragment	124
Figure 5.17: Mass-spectrometry results for 34 kDa PhtD protein fragment .	125
Figure 5.18: Western transfer for Apo- F/L PhtD digest	126
Figure 5.19: Western transfer for Zn ²⁺ -bound F/L PhtD digest	126
Figure 5.20: Apo-F/L PhtD N-term sequencing results mapped onto F/L PhtD sequence	127
Figure 5.21: Zn ²⁺ -bound F/L PhtD N-term sequencing results mapped onto F/L PhtD sequence	129
Figure 5.22: ITC analysis of PhtD Zn ²⁺ binding.....	132
Figure 6.1: F/L PhtD disorder prediction	137
Figure 6.2: C-term PhtD disorder prediction.....	137
Figure 6.3: ΔC-term PhtD disorder prediction	138
Figure 6.4: 34 kDa PhtD fragment disorder prediction	138
Figure 6.5: 1D-NMR Spectrum of Apo- F/L PhtD	140
Figure 6.6: Reference 1D NMR spectra.....	141
Figure 6.7: F/L PhtD Zn ²⁺ titration	143
Figure 6.8: Scaled data for F/L PhtD Zn ²⁺ titration	144
Figure 6.9: Close-up of scaled data for F/L PhtD Zn ²⁺ titration.....	145
Figure 6.10: C-term PhtD Zn ²⁺ titration	147
Figure 6.11: Close-up of C-term PhtD Zn ²⁺ titration	148
Figure 6.12: ΔC-term PhtD Zn ²⁺ titration.....	149
Figure 6.13: Close-up of ΔC-term Zn ²⁺ titration	150
Figure 6.14: comparison of amide shoulder-region for F/L, C-term and ΔC- term PhtD.....	151
Figure 6.15: Example far-UV CD spectra	153
Figure 6.16: F/L PhtD Far UV CD spectra NaCl vs. No- NaCl	154
Figure 6.17: Far UV CD spectra of F/L Apo-PhtD vs. F/L PhtD + Zn ²⁺	155
Figure 6.18: Far-UV CD spectra for F/L PhtD	157

Figure 6.19: Far-UV CD spectra of C-term PhtD fragment	158
Figure 6.20: Far-UV CD spectra for Δ C-term PhtD fragment	159
Figure 6.21: Crystal from screening with F/L PhtD	161
Figure 6.22: Crystal from screening with F/L PhtD	161
Figure 6.23: Crystal from screening F/L PhtD with trypsin	162
Figure 6.24: Diffraction image for F/L PhtD crystal	163
Figure 6.25: T_m results for apo- F/L PhtD.	168
Figure 6.26: T_m results for Zn^{2+} -bound F/L PhtD	169
Figure 6.27: C-term PhtD crystal	171
Figure 6.28: Diffraction image for Δ C-term crystal test	171
Figure 6.29: Small Δ C-term PhtD crystals	175
Figure 6.30: Small Δ C-term PhtD crystals	175
Figure 7.1: Theoretical model of PhtD.....	180

List Of Tables

Table 1.1: Pneumococcal virulence factors and their functions	8
Table 1.2: Current licensed available pneumococcal vaccines.....	13
Table 2.1: Commonly used buffers and their composition.....	26
Table 2.2: Growth media and their composition	27
Table 2.3: Antibiotics	28
Table 2.4: List of bacterial strains, including genotypes.....	29
Table 2.5: pOPINF expression vector features	30
Table 2.6: PCR primers used in DNA cloning and sequencing.....	32
Table 2.7: Standard PCR reaction-mix	33
Table 2.8: General PCR experiment parameters	33
Table 2.9: Restriction enzymes used to linearise pOPINF plasmid in preparation for insertion of required gene or DNA fragment.....	35
Table 2.10: Buffers used in Ni-affinity purification	38
Table 2.11: Anion Exchange Buffers.....	39
Table 2.12: Buffers used in SDS-PAGE.....	41
Table 2.13: Commercial crystal screens	46
Table 3.1: Crystal producing conditions from crystallisation screens.....	56
Table 3.2: Data collection statistics for TktA	59
Table 3.3: Statistics for calculated Matthews coefficient of TktA crystal....	60
Table 3.4: Various forms of PpmA protein that were subjected to crystallisation trials	75
Table 3.5: Preliminary data statistics for a single PpmA crystal.....	78
Table 3.6: Solvent content for single crystal of PpmA.....	78
Table 4.1: Primers for amplification of F/L <i>PhtD</i>	85
Table 4.2: PCR conditions for amplification of F/L <i>PhtD</i>	85
Table 4.3: <i>PhtD</i> plasmids and resulting protein products	94
Table 4.4: Primers used for amplification of <i>PhtD</i> fragments.	95
Table 4.5: PCR conditions for <i>PhtD</i> fragments.....	95
Table 4.6: PCR fragment product sizes	96
Table 5.1: Limited proteolysis reaction composition	106
Table 5.2: Proteases used in mass-spectrometry analysis of purified <i>PhtD</i> proteins. ¹¹⁴	121
Table 5.3: N-terminal sequencing results of Apo- F/L <i>PhtD</i>	127
Table 5.4: N-terminal sequencing results for Zn ²⁺ -bound F/L <i>PhtD</i>	128
Table 6.1: Dichroweb ¹⁷⁰ calculation of 2 ^o structure elements	156
Table 6.2: Indexing statistics for single F/L <i>PhtD</i> crystal	164
Table 6.3: Results of MATTHEWS ¹³² for F/L <i>PhtD</i> crystal.....	165
Table 6.4: Selected crystallisation conditions from F/L <i>PhtD</i> trials.....	173
Table 6.5: Selected conditions from C-term <i>PhtD</i> trials.....	174

Chapter 1 : Introduction

1.1 *Streptococcus pneumoniae*

1.1.1 Morphology

The bacterial human pathogen *Streptococcus pneumoniae* (commonly referred to as the pneumococcus) is one of the leading causes of disease in the world today, causing an estimated 1 million deaths per year in children alone. A Gram-positive facultative anaerobe, *S. pneumoniae*, lacks the presence of flagellae, rendering it non-motile. It occurs as a diplococcus, which forms chains during growth, as illustrated in figure 1.1

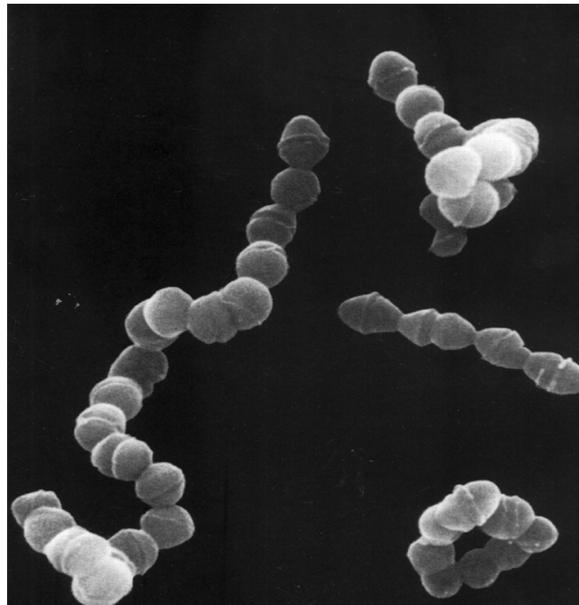


Figure 1.1: Scanning electron micrograph of *S. pneumoniae* at x14,500 magnification²

Although normally present asymptotically as a commensal of the nasopharyngeal flora, *S. pneumoniae* can cause a variety of diseases as a result of relocating to a different ecological niche of the body. These diseases can range from the mild- such as otitis-media, to the life-threatening such as meningitis.

The pneumococcus produces a highly antigenic extracellular capsule comprised of a variety of polysaccharides, which aids the bacteria in evading opsonophagocytosis through masking of Complement C3b deposition sites.^{3, 4}

At present more than 90 different strains of pneumococcus have been identified, based on polysaccharide immunohistochemistry. However, due to cross-reactivity between antibodies used to identify different serotypes, nomenclature has been altered to include sub-grouping, where closely related serotypes are clustered into serogroups. Clinical identification of *S. pneumoniae* has traditionally been performed through the Quellung reaction utilising capsular-polysaccharide specific antibodies,^{5, 6} and by sensitivity to Optochin antibiotic when cultured on blood agar plates- techniques that are still employed in the laboratories of today.

1.1.2 Pneumococcal Carriage

The pneumococcus is a common commensal bacteria which colonises the nasopharynx. In addition to the pneumococcus, the nasopharynx is a competitive environment, being the ecological niche for a diverse range of organisms and including such species as *Haemophilus influenzae*, *Staphylococcus aureus* and *Neisseria meningitidis*.⁷ The pneumococcus is able to produce hydrogen peroxide (H₂O₂), which has been shown to inhibit the growth of other microflora of the nasopharynx, providing the pneumococcus with a selective advantage.⁸ It is thought that every person will be colonised with these organisms at least once during their lifetime. Pneumococcal carriage varies due to factors such as age, geographical location, and genetic background of the population.⁹⁻¹¹ In developed countries, it is estimated that virtually every child will have been colonised with *S.pneumoniae* within their first year of life.

Colonisation is thought to be key to any subsequent presentation of invasive pneumococcal disease (IPD). However due to variation between the current 92 different serotypes, not all serotypes are equally pathogenic,¹² with a number of serotypes known for their high incidence of invasive disease, whilst others appear to favour colonisation over invasion. Although pneumococcal carriage studies are rare, it has been shown that the most common pneumococcal serotypes in Europe are 6A, 6B, 9V, 14, 18C, 19F and 23F.⁷

1.1.3 Pneumococcal Disease

1.1.3.1 Overview

Although usually present asymptotically, the pneumococcus can cause a variety of diseases if it migrates from the nasopharynx to an atypical ecological niche within the body. The diseases that present are dependant on the resultant niche to which the pneumococcus migrates. There are a broad range of diseases which result from pneumococcal invasion which present symptoms of varying severity. It is thought that post-colonisation, the pneumococcus is able to transmigrate across the epithelial barrier in the nasopharynx and disseminate throughout the host, although the mechanism by which this process occurs is yet to be determined.¹³ As a result of this transmigration and dissemination process, the pneumococcus can invade a variety of niches within the host.

1.1.3.2 Pneumococcal Burden Of Disease

World Health Organisation (WHO) statistics report that in the year 2000, there were an estimated 14.5 million cases of serious IPD.¹⁴

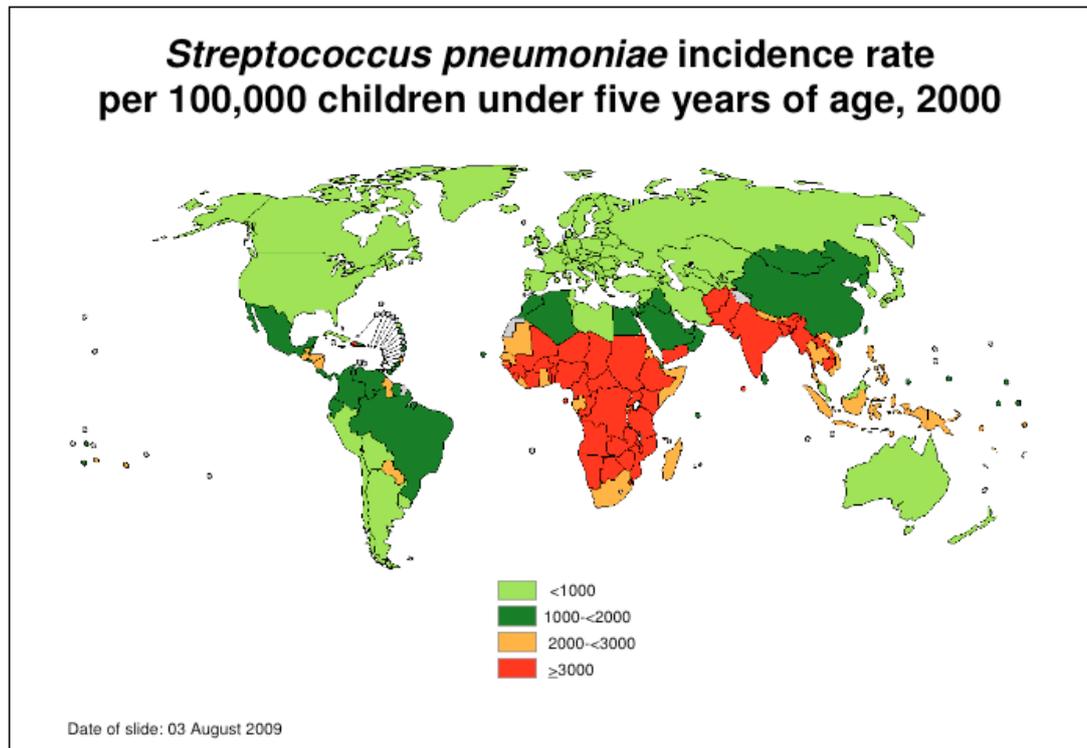


Figure 1.2: Incidence of serious pneumococcal disease worldwide

Global representation illustrating incidence of serious IPD. Figure reproduced from the WHO¹⁴

Looking at the schematic representation illustrating the incidence of serious IPD in figure 1.2, it is evident that the regions worst affected by serious IPD are the developing countries; in particular, a large proportion of African countries show an incidence of disease of ≥ 3000 cases per 100,000 children under 5 years of age.

The corresponding rates of mortality illustrated in figure 1.3 compound these findings, with the developing nations, in particular Africa exhibiting the highest mortality rates.

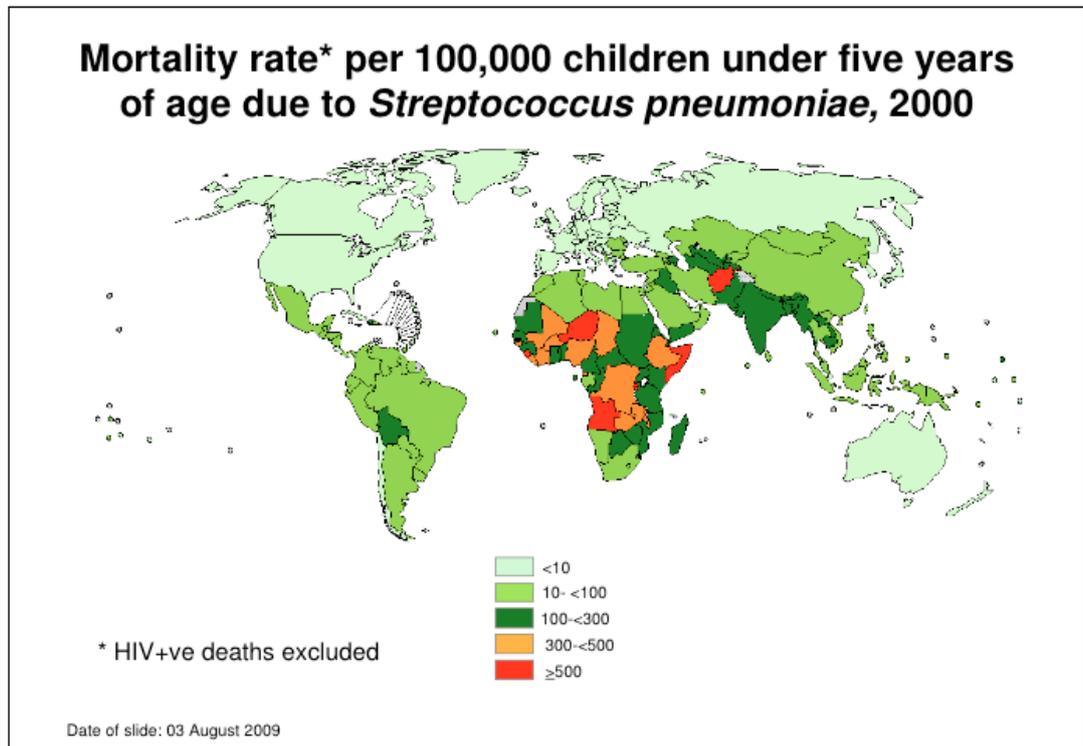


Figure 1.3: Incidence of mortality resulting from serious pneumococcal disease
Global representation illustrating incidence mortality as a result of serious IPD. Figure reproduced from the WHO¹⁴

Recent estimates suggest that there is a mortality rate of 700,000 - 1 million deaths per year for children below 5 years of age every year worldwide.¹⁵ Of these mortalities, an estimated 91,000 were in HIV+ children and an estimated 735,000 were in HIV- children. Of the deaths in HIV- children, over 61% occurred in 10 African and Asian countries. These statistics illustrate the severity of the disease burden caused by the pneumococcus.

1.1.3.3 Acute Otitis-Media (AOM)

Acute otitis-media is the most common complaint for which antibiotics are prescribed for children in the US.¹⁶ The most common pneumococcal disease, AOM is a relatively benign condition characterised by growth of the pneumococcus in the middle-ear cavity resulting in inflammation of the tympanic membrane.¹⁷ It is thought that AOM occurs in co-infection with other diseases, such as viral-respiratory diseases which can alter Eustachian

tube function, leading to negative middle-ear pressure predisposing bacterial middle-ear infection.¹⁸ Of the incidence in cases of AOM, 40-60% are attributed to the pneumococcus,¹⁹⁻²¹ with the other major disease-causing agent being identified as *H. influenzae*. In the UK, around 30% of children each year are diagnosed with AOM, of which 97% are prescribed antibiotic treatment.²² Although not a life-threatening disease, the number of incidences of AOM cause a serious burden on the current healthcare system through outpatient treatment, and would therefore benefit from prevention through vaccination which would also result in reduction in the use of antibiotics, potentially helping to reduce the spread and proliferation of antibiotic-resistant pneumococci.

1.1.3.4 Pneumococcal Pneumonia

The pneumococcus is the leading cause of pneumonia in the world today.¹⁴ Of the estimated 14.5 million incidences of IPD in the year 2000, 95% of cases were attributed to pneumonia.¹⁴ Pneumonia is characterised by a build-up of fluid in the lung cavity, which reduces the area across which oxygen can pass into the bloodstream via the bronchial alveoli. Pneumonia is responsible for the deaths of an estimated 2 million children each year, attributing to around 20% of child mortality, of which the pneumococcus is responsible for a high proportion (~70%).²³ In developed countries where the majority of pneumonia cases are as a result of community-acquired pneumonia. With a mortality rate of 10-20%, the pneumococcus is the major causal agent of community-acquired pneumonia. Pneumonia cases in the developed-world are mainly prevalent in the “at risk” groups comprising the elderly and immunocompromised, where the immune system has been weakened. Although the mortality rate for pneumococcal meningitis is greater, the sheer burden of disease caused by pneumococcal pneumonia makes it by far the most serious of the pneumococcal diseases.

1.1.3.5 Meningitis & Bacteraemia

As a consequence of the migration of the pneumococcus to the lungs and the onset of pneumonia, it is thought that bacteraemia due to IPD may be caused by pneumococcal invasion of the bloodstream across the bronchial alveoli. As a result of the dissemination and proliferation of the pneumococcus in the

bloodstream, it is possible for the pneumococcus to present itself at and cross the blood/brain barrier, resulting in meningitis in the host.

Meningitis is characterised by inflammation of the meninges, which are the protective membranes that surround the brain tissue and spinal-cord. Symptoms of meningitis include headache, stiff neck, fever, confusion or loss of consciousness and intolerance to light.^{24, 25} Left untreated, meningitis can lead to coma and death. In adults, bacterial meningitis has an annual incidence of 4 -6 per 100,000 members of the population; 80% of all bacterial meningitis cases are caused by *S. pneumoniae* or *N. meningitidis*.²⁴ Meningitis has a mortality rate of 20-30%^{26, 27}. Complete recovery is rare however, with a large number of survivors suffering from residual effects; epilepsy, spinal-cord dysfunction and hearing loss have been reported.²⁶ In addition to these observations, the WHO reports that in sub-Saharan Africa, 40-75% of individuals diagnosed with pneumococcal meningitis either die or are permanently disabled.

1.1.4 Pneumococcal Virulence Factors

1.1.4.1 Overview

The pneumococcus has a number of virulence factors that have an effect on the pathogenicity of the organism. As previously mentioned, the main virulence factor of the pneumococcus has been identified as the polysaccharide capsule, which coats the surface of the bacteria. In addition to the polysaccharide capsule, a variety of other pneumococcal virulence factors have been identified and include surface proteins capable of performing a variety of functions, pili thought to play a role in cellular adherence, and a lytic toxin named pneumolysin (Ply). A selection of these virulence factors are illustrated in figure 1.4, and their functions described in table 1.1.

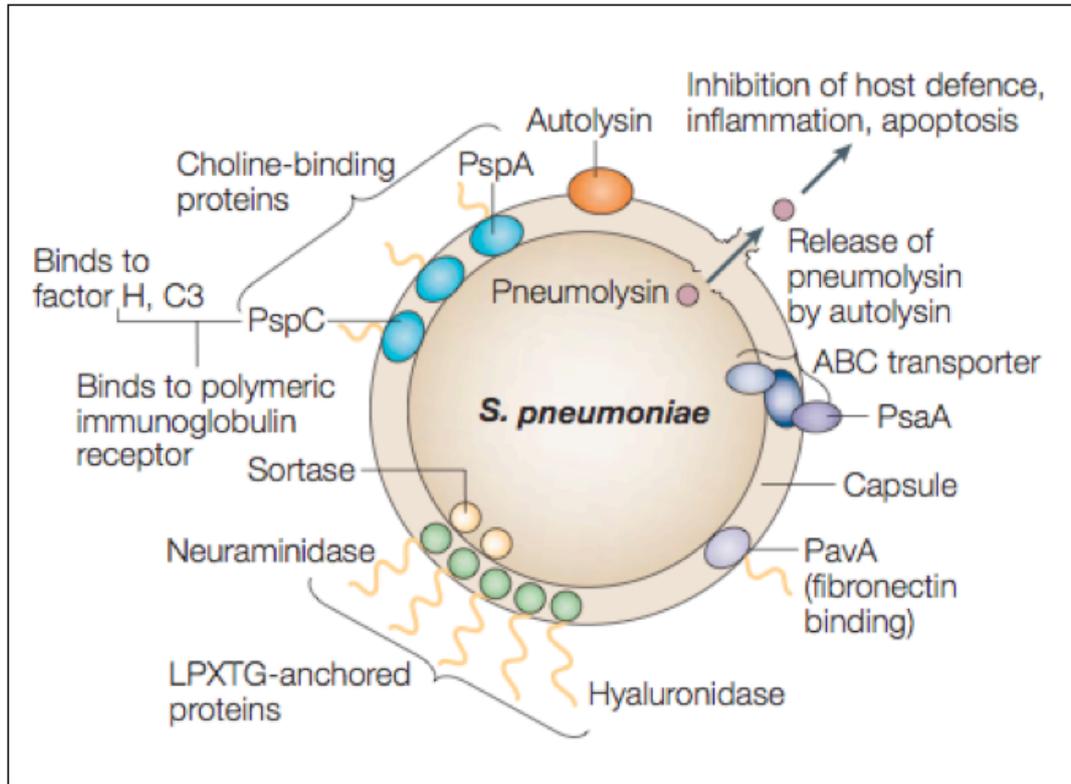


Figure 1.4: Pneumococcal virulence factors

Figure illustrating a selection of virulence factors possessed by *S. pneumoniae*. Figure reproduced from Mitchell, T.J. (2003).²⁸

Virulence Factor	Function/Effect
Capsule	Protection against phagocytosis and opsonisation. ^{4, 29, 30}
Pneumolysin (Ply)	Cytolytic toxin. ^{31, 32} Essential for pneumococcal pathogenesis and survival in the host. ³³⁻³⁶
Autolysin A (LytA)	Degrades pneumococcal peptidoglycan releasing cell-wall products resulting in inflammation, ³⁷ and allowing release of Ply toxin from the inside of the cell. ³⁸ Also involved in migration of pneumococci from nasopharynx to lower respiratory tract. ³⁶
Pneumococcal Surface Protein C (PspC, aka CbpA)	Protects host from opsonisation by recruiting complement Factor H to the cell surface. ³⁹⁻⁴¹ Involved in pneumococcal colonisation. ⁴²
Pneumococcal Surface Protein A (PspA)	Inhibition of Complement binding and activation. ⁴³⁻⁴⁵ Protects pneumococcus against killing effects of lactoferrin. ⁴⁶

Pneumococcal Surface Antigen A (PsaA)	Substrate-binding protein of Mn ²⁺ ABC transporter. ^{47, 48} Required for pneumococcal adherence. ⁴⁹⁻⁵¹
Neuraminidase A (NanA)	Cleaves sialic acid from cell surfaces, revealing receptors for adherence to host. ^{52, 53}
Pilus	Involved in adhesion of pneumococci to host. ⁵⁴⁻⁵⁷ A second pilus recently discovered also appears involved in adhesion. ⁵⁸
Hydrogen Peroxide (H ₂ O ₂)	Kills or inhibits growth of competitive nasopharyngeal microflora. ⁸
Zinc Metalloprotease (ZmpB)	Induces TNF- α production, resulting in lower respiratory tract inflammation. ⁵⁹
Pneumococcal Histidine Triad (Pht) proteins	Unknown function. Proposed metal-scavengers for survival under environmental stress. ⁶⁰ Possible role in Factor H binding. ⁶¹

Table 1.1: Pneumococcal virulence factors and their functions

1.1.4.2 Pneumococcal Polysaccharide Capsule

The capsule protects the pneumococcus from phagocytosis through blocking the sites of Complement C3b binding, which causes opsonisation.^{4, 29} The composition of the pneumococcal capsule is highly variable, which has led to the evolution and proliferation of greater than 90 distinct serotypes.⁶² This capsule is laid down on the outer surface of the cell. Due to the high variability in capsular composition, differences occur in capsule physiology between pneumococcal serotypes (STs); as an example, serotype 3 (ST3) pneumococci exhibit a significantly more highly mucoid capsule than other serotypes. There has been speculation that the type of capsule expression may affect the preference of the bacteria for invasion or colonisation; expression of a capsule is important for survival in the blood, and is strongly associated with the ability of pneumococci to cause invasive disease. Certainly there are differences in invasion and colonisation instances between strains possessing different capsule types.⁶³

Having been identified as the main pneumococcal virulence factor, the polysaccharide capsule is the current target of the developed range of pneumococcal vaccines, with a polysaccharide vaccine that targets 23 different serotypes, an established protein conjugate vaccine targeting 7 serotypes and a recently introduced protein conjugate vaccine which targets 10 serotypes.

1.2 Pneumococcal Vaccines

1.2.1 At Risk Groups/Vaccine Coverage

The main 'at risk' groups susceptible to IPD are the young (children under 5 years of age), the elderly, and the immuno-compromised.⁶⁴ Children in undeveloped countries are particularly at risk from pneumococcal infection. In developed countries, the elderly are particularly at risk from pneumococcal disease resulting in community-acquired pneumonia. Current pneumococcal vaccines are based on the pneumococcal polysaccharide capsule. As there are greater than 90 different serotypes of pneumococcus based on capsule composition,⁶² current vaccines are serotype specific. These vaccines have been highly successful in providing protection against the pneumococcal strains that they target, however due to their serotype-specific design, this efficiency varies depending on geographical location due to varying prevalence of different serotypes around the world.⁶⁵ Current pneumococcal vaccines have shown high efficiency in preventing IPD in developed countries such as the USA,⁶⁵ as the current vaccines were tailored for this market, targeting the most prevalent US serotypes. Worldwide implementation of the pneumococcal conjugate vaccine is illustrated in figure 1.5.

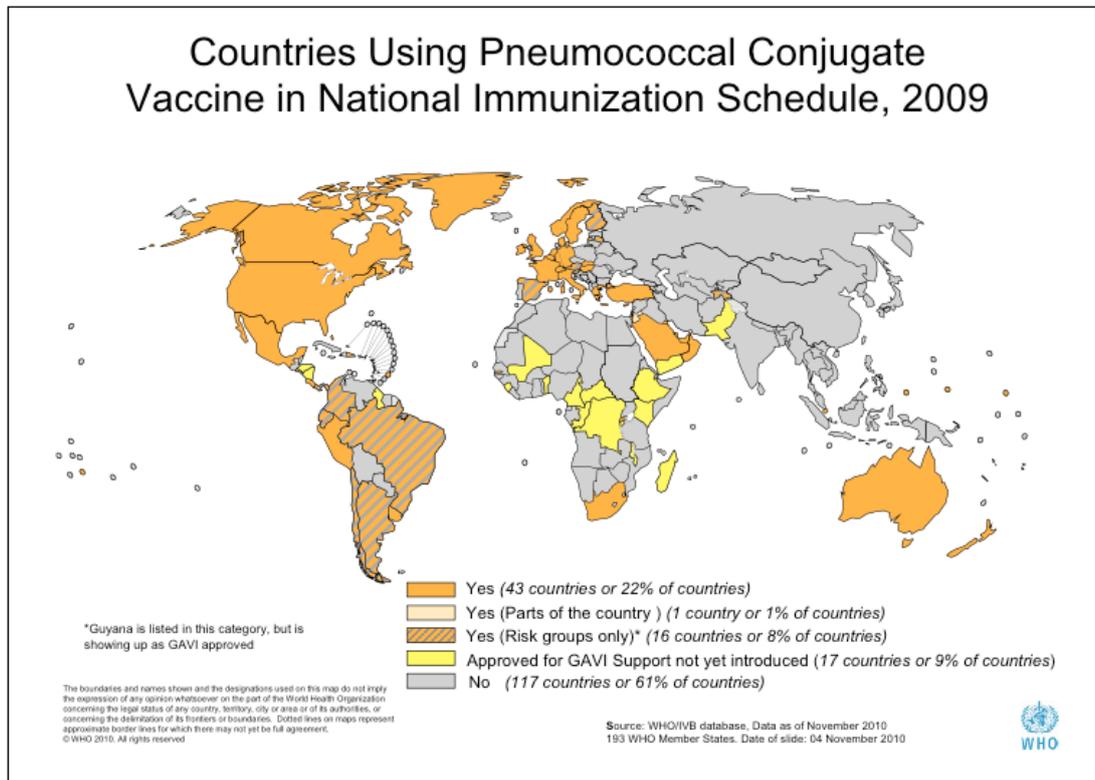


Figure 1.5: Worldwide implementation of pneumococcal conjugate vaccine

From the data above it can be observed that use of the pneumococcal conjugate vaccine is largely confined to the developed world. Figure reproduced from the WHO.¹⁴

Although implementation of the pneumococcal conjugate vaccine has begun to be introduced in developing countries such as regions of South America, use of the pneumococcal conjugate vaccine is predominantly still confined to regions of the developed world such as the USA and Europe. Possible reasons for this could be cost, lack of ease-of-administration (an immunisation-course must be administered by a General Practitioner), or possibly due to the prevalence of serotypes not covered by the vaccine.

1.2.2 Evolution of the Pneumococcal vaccine

Similar to the original smallpox vaccine developed by Edward Jenner, the first pneumococcal vaccine was whole-cell based, utilising whole killed pneumococci.^{66, 67} Further investigation lead to the revelation that the pneumococcal polysaccharide capsule played a key role in causing IPD. A 14-valent polysaccharide vaccine comprising the polysaccharide elements for the 14 most prevalent pneumococcal serotypes was developed and released in 1977. Subsequently, this 14-valent vaccine was updated to a 23-valent

vaccine, which included polysaccharide for an additional 9 serotypes.⁶⁸ However, limitations in these polysaccharide vaccines led to the development of a new generation of vaccines based on the method of fusing antigens to a protein carrier first developed and implemented with success for the influenza HiB vaccine.^{69, 70} Both the pneumococcal polysaccharide vaccine and the protein-conjugate vaccines have advantages and limitations, and are administered accordingly.

1.2.3 Pneumococcal Vaccine Types

As described in chapter 1.2.2, there are currently two distinct types of pneumococcal vaccine on the market at present. An overview of these currently available vaccines is given in table 1.2.

Vaccine Name	Type	Vaccine Coverage (ST)	Details	Manufacturer
PneumoVax [®]	Polysaccharide	1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 19F, 20, 22F, 23F, 33F	Polysaccharide vaccine comprising capsular polysaccharide for 23 pneumococcal STs	Sanofi Pasteur MSD
Prevenar [™] (PCV7)	Protein-conjugate	4, 6B, 9V, 14, 18C, 19F, 23F (Additions for PCV13: 1, 3, 5, 6A, 7F, 19A)	7 most prevalent disease-causing STs conjugated to CRM ₁₇ diphtheria toxoid ⁷¹ (PCV7 has recently been updated to include a further 6 serotypes and named PCV13. ⁷²⁻⁷⁴ These additional serotypes are shown in parenthesis)	Wyeth Vaccines (Pfizer)
Synflorix [™]	Protein-conjugate	1, 4, 5, 6B, 7F, 9V, 14, 18C, 19F, 23F	10 most prevalent disease-causing STs conjugated to NTHiD ^{75, 76}	GlaxoSmithKline

Table 1.2: Current licensed available pneumococcal vaccines

The 23 serotype-polysaccharides included in the PneumoVax polysaccharide vaccine comprise those responsible for 65-86% of pneumococcal disease over

the period from 1978-94 in the US.⁶⁵ PneumoVax is the pneumococcal vaccine usually administered to elderly patients.

Wyeth Vaccines' pneumococcal conjugate vaccine Prevenar™ utilises polysaccharides that recognise the 7 most prevalent pneumococcal serotypes in the US. These polysaccharides are conjugated to the CRM₁₇ diphtheria toxoid. In contrast to the 23-valent polysaccharide vaccine, Prevenar™ is capable of providing protection in young children.⁷⁷

Synflorix™ takes the serotype formulation used in Prevenar™ and incorporates serotype 7F, along with two serotypes -1 and 5- most prevalent in developing countries.¹⁴ In addition to these additional serotypes, the polysaccharides are conjugated to the Non-typable *H. influenzae* D protein (NTHiPD) rather than the CRM₁₇ carrier utilised in Prevenar™, resulting in a degree of protection against *H. influenzae*.⁷⁵

1.2.4 Limitations Of Current Vaccines

Whilst all the above vaccines have proven to be successful in controlling pneumococcal disease, each has their limitations. The 23-valent polysaccharide vaccine is ineffective in children under 2 years of age, one of the most vulnerable groups susceptible to pneumococcal disease, with the poor immunogenic response to the pneumococcal antigens resulting from a lack of T-cell involvement that is a prerequisite for high-level antibody response and induction of immunologic memory.^{65, 78, 79}

Whilst Prevenar™ has been shown to provide protection in children where PneumoVax® does not, it has the disadvantage of only providing protection against the 7 serotypes of pneumococci most prevalent in the US. As a result of this, vaccine efficacy in different geographical areas is likely to vary significantly, depending on the prevalent serotypes of the region. It is likely that this would also occur with the Synflorix™ 10-valent conjugate vaccine, although having been recently introduced, it will be some time before enough data is available to draw any conclusions. Additionally, with regards to the polysaccharide conjugate vaccines covered in this section, there is the added complication of economic value for vaccine production; the conjugation process becomes progressively more difficult with increasing

number of conjugate-antigens, and similarly vaccine production costs also increase. It could therefore be foreseen that the difficulties in continuing this method of vaccine development would eventually outweigh the merits of production of the vaccine.

A further complication that has been foreseen as a potential problem for the current group of pneumococcal vaccines is the possibility of serotype replacement occurring.^{80, 81} It may be that upon successful control of the serotypes targeted by the current vaccines, a gap is made available which is filled by other pneumococcal serotypes which had previously been relatively uncommon in comparison to those targeted by the vaccine, resulting in the need for reformulation of the current vaccines to cover the newly emergent serotypes.^{82, 83} This effect has already been observed in regions which have implemented the PCV7 vaccine;⁸⁴⁻⁸⁶ successful reduction in the serotypes covered by PCV7 has resulted in serotype-replacement of these strains, with the subsequent emergence of serotype 19A pneumococci as a virulent strain in these areas.⁸⁷⁻⁸⁹

1.2.5 Next-Generation Vaccine Development

Although there has been success with administration of the current pneumococcal vaccines, the limitations in the vaccine design at present as outlined above have encouraged research into development of a next-generation protein vaccine. As such, investigation has begun in an attempt to identify suitable candidate proteins that would satisfy the following criteria:

- Highly conserved throughout the various pneumococcal serotypes
- Capability of eliciting an antibody response from the host immune-system upon exposure to the protein vaccine candidate
- Ideally, unique to the pneumococcus
- No homology to any human proteins

Potentially, a protein candidate which satisfied these criteria could have the ability to provide protection against the vast majority of pneumococcal

strains, whilst avoiding any cross-reactivity with host proteins which may have a detrimental effect to the patient.

1.3 Pneumococcal Histidine Triad Proteins

1.3.1 Identification As A Vaccine Candidate

In 2001, a whole genome study of a serotype 4 strain of pneumococcus (named *Norway 4* - isolated from a Norwegian patient with bacteraemia⁹⁰ and subsequently renamed TIGR4⁹¹) performed by Medimmune Inc. (Gathersberg, Maryland US) identified 130 ORFs for proteins encoding secretion motifs or showing similarity to known virulence factors, of which 6 conferred protection against pneumococcal infection and showed broad strain distribution.¹ Utilising flow-cytometry, several of these targets were confirmed as surface localised.¹ Further investigation revealed the presence of a novel family of proteins that localised to the cell-surface,⁹² which were termed Pneumococcal Histidine triad proteins (Pht) after an HxxHxH motif repeated several times throughout the amino acid sequence of the Pht protein family. These proteins were shown to provide protection against pneumococcal disease in a murine model of systemic infection, implying that the Pht proteins were possible candidates for inclusion in a protein-based pneumococcal vaccine.⁹² A subsequent study in 2006, performing a whole genome screen via a lambda-display library was carried out on the pneumococcal strain D39.⁹³ Screening of this library with sera from individuals suffering pneumococcal infection, or with sera from mice immunised with *S. pneumoniae* D39 resulted in the identification of a number of potential vaccine candidates including the Pneumococcal Histidine Triad (Pht) proteins.⁹³ These proteins have also variously been termed Pneumococcal Histidine Proteins (Php),⁹⁴ and BVH proteins.⁹⁵ Throughout this thesis, these proteins will be referred to as Pht proteins, as this is the most widely accepted name for these proteins in the current literature.

In a murine model of pneumococcal infection, immunisation of mice with Pht proteins has been shown to confer protection from pneumococcal challenge.⁹⁶ Additionally, flow-cytometry experiments and immunogold-labelling experiments have shown that the Pht proteins are present on the surface of the pneumococcal cell.^{92, 95} The observations above, and the

potential for antibodies generated against one Pht protein to cross-react with the other Pht proteins (due to high sequence identity between the different Pht proteins) make the pneumococcal histidine triad protein family an attractive option for inclusion in a new protein-based pneumococcal vaccine.

1.3.2 Sequence Variation

There have been 4 identified Pht proteins in the pneumococcus. These have been named PhtA, PhtB, PhtD, and PhtE respectively. To date these Pht proteins have been identified exclusively in Streptococci, and are highly conserved across all pneumococcal serotypes. Sequence homology between the 4 proteins in this family varies from 87%,⁹² with PhtE bearing the lowest sequence identity to the others, being only 37% homologous. Most variation between the different Pht proteins is evident at the C-terminal end of the molecule, whilst the N-terminal sequence is rather more highly conserved with an average of >62% homology between members. In general, the Pht proteins show a high degree of sequence homology, as illustrated in figure 1.6. This sequence homology could be advantageous when considering the Pht proteins as potential vaccine candidates, as the high degree of similarity in sequence could confer cross-reactivity of antibodies generated from one Pht protein with other members of the Pht family.



Figure 1.7: 1.2Å Crystal structure of PhtA fragment

Crystal structure of the PhtA fragment (residues 166-220).⁹⁸ The Zn²⁺ atom is illustrated in grey. Image generated using Jmol.⁹⁹

The Zn²⁺ atom was revealed to be bound into the crystal structure through a novel zinc-binding fold, which is created by the histidines from the HxxHxH motif and an additional aspartate residue; the ligating residues and their interaction with the zinc are shown in figure 1.8.

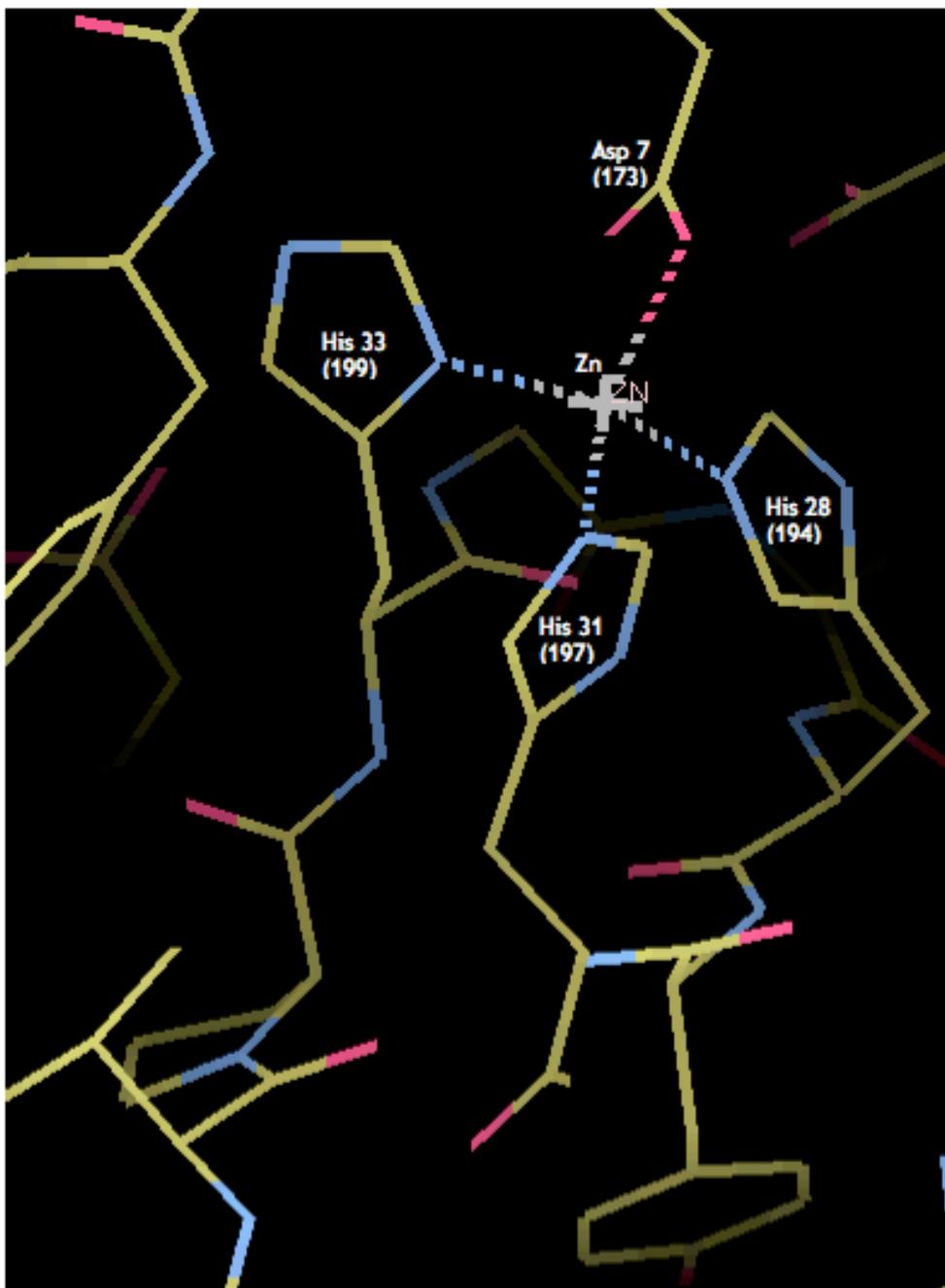


Figure 1.8: Zn^{2+} binding in 1.2Å crystal structure of PhtA, illustrating Zn^{2+} -ligating residues. The coordination of the ligating residues showing their interaction with the Zn^{2+} atom is illustrated. Ligating residues are numbered according to their residue number in the determined crystal structure. Corresponding residue numbers in the full protein sequence of PhtA are given in parenthesis. Image created using COOT.

The Zn^{2+} atom is bound by ND1 of His-194, NE2 of His-197, NE2 of His-199 and OD2 of Asp-173 at distances of 2.02, 2.03, 2.03 and 1.97Å, forming a distorted tetrahedral coordination sphere for the Zn^{2+} atom, with the three histidine residues forming the base of the triangle and the side-chain of aspartate 173 forming the apex. The observed distances were shorter than

expected; typical distances of protein - Zn²⁺ atoms are 2.2Å for histidines and 2.1Å for aspartate residues.⁹⁸

Interestingly, crystallisation trials were begun using full-length PhtA protein, suggesting either that the PhtA protein had subjected itself to auto-proteolysis, or a protease contaminant had inadvertently been introduced during experimental set-up. Data in this thesis supports the mechanism of Zn²⁺ binding by the Pht proteins, and that PhtD appears to elicit Zn²⁺ binding exclusively. Recent publications have variously suggested that PhtD may play a role in protection of the pneumococcus from opsonophagocytosis by the recruiting of Complement Factor-H to the cell surface,^{61, 100} or that the Pht proteins function to scavenge Zn²⁺ or Mn²⁺ from the environment^{60, 101, 102} and act as a reservoir which can be released in nutrient- depleted conditions, effectively playing the role of a siderophore.^{103, 104}

Although 4 different Pht proteins have been described to date in the pneumococcus, the degree of sequence homology outlined above would suggest that only one of the Pht proteins may be required in order to provide protection, due to the possibility of cross-reaction between generated antibodies and the other members of the Pht protein family. To this end, PhtD has been singled out as the candidate protein from this family, as it is the most highly conserved of the Pht proteins, and has been identified in all strains of the pneumococcus that have been analysed to date.⁶⁰

A variety of other pneumococcal proteins have been investigated as potential vaccine candidates, such as PspA, PspC, and PsaA.^{49, 105-107} However in experiments carried out using lung and nasopharyngeal colonisation models, immunisation with the Pht proteins was more successful at preventing pneumococcal colonisation than immunisation with other candidates including PspC, PspA, and PsaA.⁹⁶ Additionally, PhtD has been seen to provide protection against a variety of different pneumococcal strains including ST3, which is becoming increasingly prevalent.¹⁰⁸ Furthermore, in a murine model of intranasal challenge, mice immunised with PhtD showed significant protection against challenge with a lethal dose of pneumococci.⁹⁶ Ply has also been investigated as a possible vaccine candidate.^{109, 110} Since Ply is a member of the cholesterol dependant cytolysins and is toxic, it must be used at

extremely low concentrations; a non-lytic version of Ply has been discovered which allows greater doses to be administered.¹¹¹ However, the potential of Ply as a vaccine is currently focused on making use of its properties as an adjuvant where it is administered in conjunction with another protein as a fusion-protein to boost the immunogenicity of its fusion-partner.^{112, 113}

1.4 Pneumococcal Histidine Triad D (PhtD)

PhtD is the main antigen of interest from the Pht family of proteins for possible inclusion in a protein-based pneumococcal vaccine. Although lacking any classical anchoring region such as an LPxTG motif or choline binding domain, flow-cytometry and immuno-gold labelling have shown PhtD to be located on the cell surface,⁹⁵ giving rise to PhtD being termed “cell-surface associated”.

The protein sequence of PhtD is illustrated in figure 1.9. The predicted signal sequence for targeting PhtD to the surface of the cell,⁹² and the location of the HxxHxH motifs in the amino acid sequence are highlighted.

MKINKKYL	AG	SVAVLALSVC	SYELGRHQAG	QVKKESNRVS	YIDGDQAGQK	AENLTPDEVS
KREGINAEQI	VIKITDQGYV	TS	HGDHYHY	NGKVPYDAII	SEELLMKDPN	YQLKSDIVN
EIKGGYVIKV	DGKYVYVYVKD	AAHADNIRTK	EEIKRQKQEH	SHNHGGGSND	QAVVAARAQK	
RYTTDDGYIF	NASDIIEDTG	DAYIVP	HGDH	YHYIPKNELS	ASELAAAEAY	WNGKQGSRPS
SSSSYNANPA	QPRLSEHNHL	TVTPTYHQNQ	GENISLLRE	LYAKPLSERH	VESDGLIFDP	
AQITSRTARG	VAVP	HGNHYH	FIPYEQMSEL	EKRIARI IPL	RYRSNHWVPD	SRPEQPSQPS
TPEPSPSPQP	APNPQPAPSN	PIDEKLVKEA	VRKVGDDGYVF	EENGVSRYIP	AKDLSAETAA	
GIDSKLAKQE	SLSHKLGAKK	TDLPSSDREF	YNKAYDLLAR	IHQDLLDNKG	RQVDFEALDN	
LLERLKDVP	DKVKLVDDIL	AFLAPIRHPE	RLGKPNQIT	YTDDEIQVAK	LAGKYTTEDG	
YIFDPRDITS	DEGDAYVTPH	MTHSH	WIKKD	SLSEAERAAA	QAYAKEKGLT	PPSTDHQDSG
NTEAKGAEAI	YNRVKAACKV	PLDRMPYNLQ	YTVEVKNGSL	IIPH	HYDHYHN	IKFEWFDEGL
YEAPKGYTLE	DLLATVKYIV	EHPNERPHSD	NGFGNASDHV	RKNKVDQDSK	PDEDKEHDEV	
SEPTHPESDE	KENHAGLNPS	ADNLYKPSTD	TEETEEEAED	TTDEAEIPQV	ENSVINAKIA	
DAEALLEKVT	DPSIRQNAME	TLTGLKSSLL	LGTKDNNTIS	AEVDSLALL	KESQPAPIQ	

Figure 1.9: PhtD protein sequence from the *S. pneumoniae* strain TIGR4

The figure above shows the protein sequence of PhtD from the *S. pneumoniae* strain TIGR4. HxxHxH motifs are highlighted in pink. Predicted signal sequence for targeting PhtD to the pneumococcal cell-surface is highlighted yellow.

Without the signal sequence, mature PhtD protein from the pneumococcal strain TIGR4 has a predicted molecular weight of 91.5 kDa¹¹⁴ and contains 5 of the canonical HxxHxH motifs for which the Pht family is named.

Vaccination studies with PhtD have shown the protein to be a promising candidate for inclusion in a protein-based pneumococcal vaccine, producing an antibody response in mice that provided protection from a lethal challenge of pneumococci. The uniqueness of the Pht proteins is another advantage for inclusion of PhtD in a protein-based vaccine, as to date the Pht proteins are thought to be highly conserved and exclusive to streptococci (genome comparisons have predicted the presence of Pht proteins in *S. mitis* and *S. pyogenes*, which is not surprising, as for example horizontal gene transfer is very common between *S. mitis* and *S. pneumoniae*). Thus a vaccine targeting Pht proteins should be able to provide comprehensive, specific protection against pneumococcal disease.

PhtD is the most prevalent of the Pht proteins, and has been found present in 100% of pneumococcal strains analysed to date.¹⁰² Whilst there has been speculation as to the function of the Pht family of proteins in general,^{61, 100-102} the definitive understanding of the functions of PhtA, PhtB, PhtD, and PhtE are still unknown. As the crystal structure of PhtA showed Zn²⁺ binding to the PhtA fragment through the HxxHxH motifs,⁹⁸ it was postulated that PhtD would also be able to bind metal atoms through the same mechanism.

1.5 Project Aims

The primary aim of this project was to attempt structural determination of the protein Pneumococcal Histidine Triad D by macromolecular X-ray crystallography, in order to use the determined crystal structure for elucidating the biological function of the PhtD protein. Structural analysis of PhtD was also investigated by a variety of complimentary biochemical techniques to uncover regions of the protein that may be conducive to X-ray crystallography, with an aim to providing a crystal structure for a fragment of the PhtD protein in the event that a crystal structure of the complete protein was unattainable. Furthermore, following on from previous work in our laboratory which uncovered the metal-binding ability of the HxxHxH motif in Pht proteins,⁹⁸ biophysical characterisation of PhtD and its role as a metal-binding protein were performed in order to assess the effect of metal-binding on protein structure, and the interaction of PhtD with various different metals.

Chapter 2 : Materials And Methods

2.1 Materials

2.1.1 Chemicals

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich, Gillingham UK or Fisher Scientific, Loughborough UK. Agarose was procured from Invitrogen, Paisley UK. Complete™ EDTA-free Protease Inhibitor tablets and Proteomics Grade Bovine Trypsin were purchased from Roche Diagnostics, West Sussex UK.

2.1.2 Enzymes

Restriction enzymes and Phusion® high-fidelity DNA polymerase were purchased from New England BioLabs, Hitchin UK. GoTaq® DNA polymerase was purchased from Promega UK, Southampton UK.

2.1.3 Competent Cells

Subcloning Efficiency™ DH5α™ cells and MAX Efficiency® DH10B™ cells were purchased from Invitrogen, Paisley UK. Stratagene XL10-Gold™ Ultracompetent cells were obtained from Aligent Technologies UK, Stockport UK. Rosetta™ 2(DE3)pLysS Singles™ and B834 chemically competent cells were obtained from Merck Chemicals, Nottingham UK.

2.1.4 Oligonucleotides

Oligonucleotides were obtained from Sigma-Aldrich, Gillingham UK or Invitrogen, Paisley UK.

2.1.5 Chromatography Supplies

Prepacked HisTrap™ HP affinity columns for Immobilised Metal Ion Affinity Chromatography (IMAC), HiTrap™ Capto™Q columns for anion exchange, and

PD10-Desalting columns were obtained from GE Healthcare, Buckinghamshire UK.

2.1.6 Crystallographic Supplies

A variety of commercially available crystallisation screens were utilised throughout this project. Crystal Screens 1 & 2, PEG/Ion Screen and Grid Screen were obtained from Hampton Research, Aliso Viejo, CA USA. JCSG+ and Wizard 1& 2 screens were purchased from Emerald Biosystems, Bainbridge Island, WA USA. PACT screen was obtained from Qiagen, West Sussex UK. Morpheus™ and MemStart/MemSys screens, Cryocaps and vials mounted Litholoops™ were obtained from Molecular Dimensions, Suffolk UK. Innovaplate™ SD-2 crystallisation plates were obtained from Innovadyne Technologies INC, Santa Rosa, CA USA.

2.2 Media And Buffers

2.2.1 Molecular Biology Buffers

Commonly used buffers for molecular biology are listed in table 2.1.

Buffer	Composition	Comments
50x Tris-Acetate-EDTA (TAE)	242g Tris base, 57.1ml glacial acetic acid, 100ml 0.5M EDTA (pH8.0).	Final volume 1L, pH adjusted to 8.5. Diluted to 1x before use
DNA loading dye	0.25% Orange G (Sigma-Aldrich), 40% (W/V) Sucrose	Made to final vol in dH ₂ O
Tris-EDTA (TE)	10mM Tris HCl pH7.5, 1mM EDTA	N/A

Table 2.1: Commonly used buffers and their composition

2.2.2 Growth Media

The *Escherichia coli* (*E. coli*) bacterial strains DH10B, XL10-Gold, DH5 α , B834, and Rosetta 2(DE3)pLysS were routinely grown in Luria Broth (LB). For large scale protein expression cultures the strains B834 and Rosetta 2(DE3)pLysS were routinely grown in Overnight-Express™ Instant TB medium (TB-Onyx; Merck Chemicals, Nottingham UK). For growth of *E. coli* bacterial strains on agar plates, solid LB medium was obtained by dissolving 1.5% (w/v) micro-agar in liquid LB and autoclaving at 121°C for 15 mins. For growth of *S. pneumoniae* on solid media, Blood Agar Base (Oxoid, Hampshire UK) plates were obtained by dissolving 40g/L BAB in dH₂O and autoclaving at 121°C for 15 mins. The sterile agar was allowed to cool until warm but still molten then supplemented with 5% defibrinated horse blood (E & O Laboratories, Bonnybridge, Scotland UK). Liquid culture of *S. pneumoniae* was performed in Brain Heart Infusion (BHI) broth (Oxoid). Compositions of all growth media used are given in table 2.2.

Media	Composition	Comments
Luria Broth (LB)	Tryptone 10g/L Yeast extract 5g/L NaCl 10g/L	20g/L dissolved in dH ₂ O, then autoclaved at 121°C for 15 mins
Brain Heart Infusion (BHI)	Brain infusion solids 12.5g/L Beef heart infusion solids 5.0g/L Proteose peptone 10g/L Glucose 2g/L NaCl 5g/L Disodium phosphate 2.5g/L pH to 7.4 \pm 0.2 @ 25°C	37g/L dissolved in dH ₂ O then autoclaved at 121°C for 15 mins
LB Agar	LB media + 1.5% (w/v) micro-agar	N/A
Blood Agar Base (BAB)	Proteose peptone 15g/L Liver digest 2.5g/L Yeast extract 5.0g/L NaCl 5.0g/L Agar 12g/L pH to 7.4 \pm 0.2 @ 25°C	40g/L dissolved in dH ₂ O Then autoclaved at 121°C for 15 mins. For blood plates: supplement warm, molten BAB with 5% defibrinated horse blood
Overnight-Express Instant TB Medium (TB-Onyx)	Proprietary formulation	60g/L in dH ₂ O. Microwaved until dissolved and simmering. N.B. Cannot be autoclaved due to presence of sugar supplements in formulation
GS96 Medium	Proprietary formulation	49.1g/L in dH ₂ O autoclaved at 121°C for 15mins. Supplemented with 0.1% (v/v) glycerol to increase plasmid yield.

Table 2.2: Growth media and their composition

2.2.3 Antibiotics

The antibiotics described in table 2.3 were used routinely. Stock solutions of these antibiotics were made at 1000x working strength, filter sterilised and stored at -20°C until required. Sterile broth media was cooled to ambient temperature before addition of antibiotic immediately before starting culture growth. When using antibiotic-supplemented agar, required antibiotics were added to sterile agar that was hand-warm but still molten immediately prior to pouring plates, in order to preserve antibiotic activity.

Antibiotic	Stock concentration	Final concentration	Solvent	Sterilisation
Ampicillin	50mg/ml	50µg/ml	dH ₂ O	0.22µm syringe filter
Chloramphenicol	34mg/ml	34µg/ml	EtOH	0.22µm syringe filter
Carbenicillin	54mg/ml	54µg/ml	dH ₂ O	0.22µm syringe filter

Table 2.3: Antibiotics

2.2.4 Isopropyl β-D-thiogalactopyranoside (IPTG)

A 1M stock of IPTG for a final working concentration of 1mM was prepared in dH₂O and filter sterilised with a 0.22µm syringe filter. The 1M IPTG stock was stored at -20°C until required.

2.3 Bacterial Strains And Vectors

2.3.1 Storage

All bacterial strains were kept as glycerol stocks for long-term storage at -80°C. Glycerol stocks were prepared by addition of 0.5ml 80% (w/v) sterile glycerol to per 1ml of O/N culture. Glycerol stocks were divided into multiple aliquots before storage.

2.3.2 Bacterial Strains

A summary of the bacterial strains utilised is given in table 2.4 For long-term storage and preservation, plasmids were transformed into Subcloning Efficiency DH5 α cells and stored as glycerol stocks at -80°C. Protein expression was carried out in Rosetta 2(DE3)pLysS or B834 chemically competent cells under control of a T7 promoter. Initially, newly constructed plasmids were transformed into XL10-GOLD chemically competent or DH10B electrocompetent cells, and subsequently transformed into DH5 α cells for storage.

Strain	Genotype	Reference
XL10-Gold [®]	Tet ^r Δ (<i>mcrA</i>)183 Δ (<i>mcrCB-hsdSMR-mrr</i>)173 <i>endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte</i> [F' <i>proAB lacZ</i> Δ M15 Tn10 (Tet ^r) Amy Cam ^r]	N/A
MAX Efficiency [®] DH10B [™]	F' <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74 recA1 endA1 araD139 Δ(<i>ara, leu</i>)7697 <i>galU galK</i> λ' <i>rpsL nupG tonA</i></i>	Killmann <i>et al</i> (1996) ¹¹⁵
Subcloning Efficiency [™] DH5 α [™]	F' ϕ 80 <i>lacZ</i> Δ M15 Δ (<i>lacZYA-argF</i>)U169 <i>recA1 endA1 hsdR17</i> (<i>r_k' m_k'</i>) <i>phoA supE44 thi-1 gyrA96 relA1</i> λ '	Rogers <i>et al</i> (2000) ¹¹⁶
B834	F' <i>ompT hsdS_B</i> (<i>r_B' m_B'</i>) <i>gal dcm met</i>	Wood (1966) ¹¹⁷ Studier & Moffat (1986) ¹¹⁸
Rosetta [™] 2(DE3)pLysS Singles [™]	F' <i>ompT hsdS_B</i> (<i>r_B' m_B'</i>) <i>gal dcm</i> (DE3) pLysSRARE2 (Cam ^R)	N/A

Table 2.4: List of bacterial strains, including genotypes

2.3.3 Plasmid DNA

All expression constructs generated were the result of cloning the desired *S. pneumoniae* TIGR4 gDNA into the expression vector pOPINF¹¹⁹ (Fig 2.1). Protein expression is possible through the *lac* operator under control of a T7 promoter and can utilise an IPTG-inducible or auto-induction system. In addition, pOPINF¹¹⁹ has added versatility in that it can be used for expression in multiple hosts without having to shuttle the insert into a different vector. Main vector features are described in table 2.5.

Vector	Features	Reference
pOPINF	T7 promoter, <i>lac</i> operator, N-terminal His ₆ -Tag, HRV-C3 protease site for His ₆ cleavage, Ampicillin resistance	Berrow (2007) ¹¹⁹

Table 2.5: pOPINF expression vector features

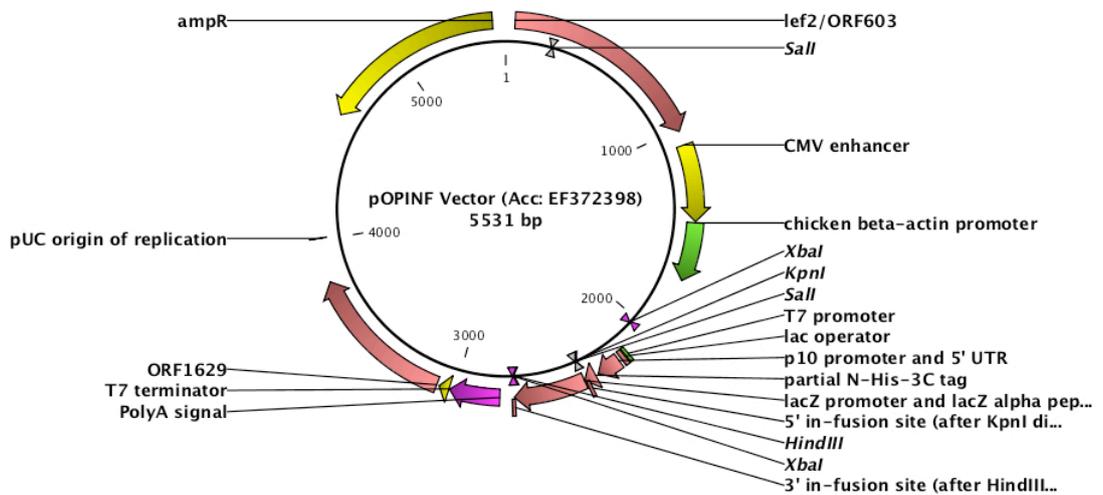


Figure 2.1: pOPINF expression vector map

Expression constructs based on the pOPINF vector¹¹⁹ were generated by insertion of the gene of interest (GOI) or DNA fragment between *KpnI* and *HindIII* restriction sites, fusing a His₆-tag to the N-terminus of the inserted DNA and placing recombinant protein expression under the control of the T7 promoter located upstream from the insertion site. Plasmid map was generated using CLC Genomics Workbench 4.¹²⁰

2.4 Laboratory Techniques

2.4.1 General Laboratory Techniques

Unless otherwise specified, all general laboratory techniques were performed as outlined in Sambrook *et al* (1989)¹²¹ or the set of standard operating procedures devised or optimised by the Mitchell Lab (relevant SOPs are included in appendices).

2.5 Polymerase Chain Reaction (PCR)

2.5.1 Preparation Of Genomic DNA

S. pneumoniae gDNA was prepared from the strain TIGR4 as described in the standard protocols devised by the Mitchell Lab (See Appendix A.3.1). Purified gDNA concentration was measured using a Nanodrop 1000, standardised to a concentration of 100ng/ μ l and stored at 4°C for use as required.

2.5.2 PCR Primers

Primers for PCR were supplied from either Sigma-Aldrich or Invitrogen in solid form. Primers were reconstituted in DNase/RNase-free Ultra-pure water (Invitrogen) to a final concentration of 100 μ M. A working stock of 10 μ M was prepared by dilution of the 100 μ M stock 10-fold with DNase/RNase-free water. Both Concentrations of primers were stored at -20°C until required. Table 2.6 describes the primers used.

Primer Name	Primer Sequence	Comments
PhtD F/L Fwd	<u>AAGTTCTGTTTCAGGGCCCGTCCTA</u> TGAACCTGGTCGTCACCAA	Fwd primer for pMRH1
PhtD F/L Rev	<u>ATGGTCTAGAAAGCTTTACTGTATA</u> GGAGCCGGTTGACTTTC	Rev primer for pMRH1, pMRH2 & pMRH4
PhtD C-Term Fwd	<u>AAGTTCTGTTTCAGGGCCCGCCAG</u> CGCAAATCACAGTCGAAC	Fwd primer for pMRH2, pMRH3 & pMRH6
PhtD ΔC-Term Rev	<u>ATGGTCTAGAAAGCTTTATTTTTCTA</u> GCAAGGCCTCCGCATC	Rev primer for pMRH3 & pMRH7
PhtD 5.3kDa Fwd	<u>AAGTTCTGTTTCAGGGCCCGTAACA</u> GATCCTAGTATTAGAC	Fwd primer for pMRH4
PhtD 34kDa Rev	<u>ATGGTCTAGAAAGCTTTAGTTATGGT</u> AATGGTCATAATGAGG	Rev primer for pMRH6
PhtD Epitope Fwd	<u>AAGTTCTGTTTCAGGGCCCGATCAAA</u> TTTGAGTGGTTTGAC	Fwd primer for pMRH7
PhtD Internal 1	GTTCTAGTTATAATGCAAATCC	Located at 664 - 686bp. For fully sequencing <i>PhtD</i>
PhtD Internal 2	GTCGACAAGTTGATTTTGAGGC	Located at 1348 - 1370bp. For fully sequencing <i>PhtD</i>
PhtD Internal 3	CCAAACGAACGTCCGCATTGAG	Located at 1986 - 2008bp. For fully sequencing <i>PhtD</i>
PhtD/Plasmid join	GCAGAAGTAGATAGTCTCTTGG	Sequences from <i>PhtD</i> into pOPINF plasmid at C- terminal join
Ply Sense (52Q)	ATTTCTGTAACAGCTACCAACGA	For amplification of Ply fragment ¹²²
Ply Anti-sense (52R)	GAATTCCTGTCTTTTCAAAGTC	For amplification of Ply fragment ¹²²

Table 2.6: PCR primers used in DNA cloning and sequencing

2.5.3 dNTPs

A set of dNTPs (Invitrogen), a source of free nucleotides, for use in PCR reaction-mixes was prepared in a ratio of 1:1:1:1 in DNase/RNase-free ultra pure H₂O to a final concentration of 10μM and stored at -20°C until required.

2.5.4 PCR

PCR was performed using a Techgene thermal cycler (Bibby Scientific, Staffordshire UK) using 200μl thin-walled PCR tubes (Starlabs, Milton Keynes UK). A total reaction volume of 50μl was comprised as shown in table 2.7 Where PCR products were to be used for cloning into plasmid DNA, Phusion high fidelity DNA polymerase and HF reaction buffer were utilised. For general PCR, GoTaq DNA polymerase and GoTaq Green reaction buffer were

utilised. Details of a typical PCR experiment are given in table 2.8 The temperature used in the annealing step was varied as required for different PCRs; more details are given in the relevant sections. Reactions were analysed by agarose gel electrophoresis as described in 2.6.1 An overview of the PCR process is given in figure 2.2.

Reaction Component	Volume
Template (100ng/ μ l)	1.5 μ l
FWD primer (10 μ M stock)	1 μ l
REV primer (10 μ M stock)	1 μ l
dNTPs (10 μ M stock)	1 μ l
DNA polymerase	0.5 μ l
Reaction buffer (5x stock)	10 μ l
DNase/RNase free PCR H ₂ O	35 μ l
Total vol.	50 μ l

Table 2.7: Standard PCR reaction-mix

Reaction Step	Temperature ($^{\circ}$ C)	Time
Init. denaturation	98	30 seconds
Melting	98	10 seconds
Annealing	54	30 seconds
Extending	72	90 seconds
Final extension	72	5 minutes

} x35

Table 2.8: General PCR experiment parameters

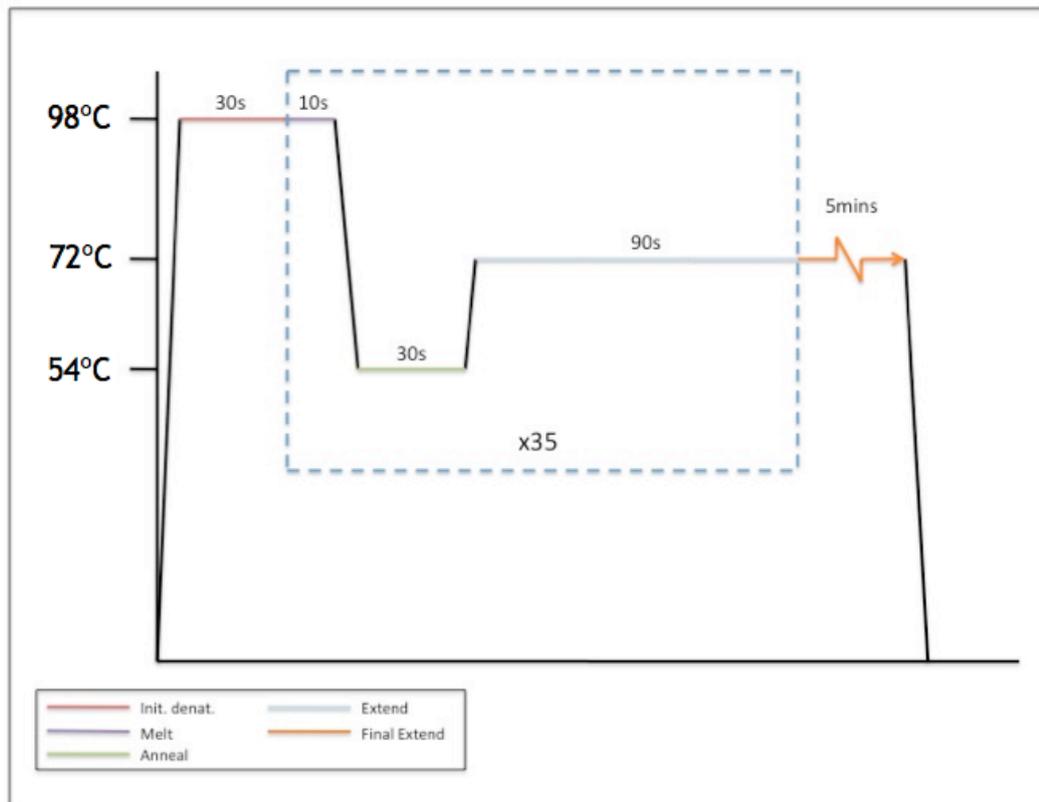


Figure 2.2: Outline of a general PCR experiment

2.6 DNA Manipulation

2.6.1 Agarose Gel Electrophoresis

PCR products, restriction digests, and plasmid DNA were all analysed using agarose gel electrophoresis. In all cases, agarose gels were 0.8% (w/v) agarose in TAE buffer. 5 μ l of SybrSafe[®] per 100ml agarose was added for product staining and visualisation when the agarose had cooled, before pouring. All gels were run in 1xTAE buffer at 100V for 20mins. All gels were loaded with 0.3 μ g 1kbp+ DNA standard (Invitrogen, Paisley UK). Gels were visualised and imaged using a UVPro Gold gel-documentation system (UVtech).

2.6.2 DNA Purification

2.6.2.1 Purification Of PCR Products

Purification of PCR products was carried out using a Qiagen QIAquick PCR Purification Kit¹²³ as described in the accompanying manufacturers protocol (see Appendix A.4.1).

2.6.2.2 Gel Extraction

Where large concentrations of purified PCR or restriction digest product were required, gel extraction was performed using a Qiagen QIAquick Gel Extraction Kit¹²⁴ according to the accompanying manufacturers protocol (see Appendix A.4.2).

2.6.2.3 Restriction Digestion

Restriction digestion was used to prepare the pOPINF plasmid for insertion of desired *S. pneumoniae* gDNA. 5µg of purified pOPINF plasmid was incubated with 50 units each of *HindIII* and *KpnI* in the presence of Bovine Serum Albumin (BSA) in a total reaction volume of 100µl at 37°C for 3 hours in a water-bath. Digests were carried out as double digests using the appropriate supplied buffer. The reaction was split over 5 reaction-tubes of 20µl volume to increase digestion efficiency, and subsequent purification of the resulting products carried out as described in 2.6.2.2 A summary of the restriction enzymes used is outlined in table 2.9.

Restriction Enzyme	Recognition Site	Buffer	Temperature (°C)
<i>HindIII</i>	5'...A↓AGCTT...3' 3'...TTCGA↑A...5'	NE Buffer 2	37
<i>KpnI</i>	5'...GGTAC↓C...3' 3'...C↑CATGG...5'	NE Buffer 2	37

Table 2.9: Restriction enzymes used to linearise pOPINF plasmid in preparation for insertion of required gene or DNA fragment

2.6.2.4 Insertion Of *S. pneumoniae* gDNA Into Expression Vector

Molecular cloning of the required *S. pneumoniae* gDNA into the expression vector pOPINF was carried out using the ligation-independent In-Fusion[®] System (Clontech, Mountain View, CA USA).¹²⁵ The required ratio of insert to linearised pOPINF was calculated using the online tool 'In-Fusion Molar Ratio Calculator' (<http://bioinfo.clontech.com/infusion/molarRatio.do>),¹²⁶ added to 1 vial of In-Fusion drydown mix in a final reaction volume of 10 μ l and reacted at 42°C for 30mins using a Techgene thermal cycler (Bibby Scientific, Staffordshire UK). The reaction was diluted immediately post-incubation to a final volume of 50 μ l with TE buffer and immediately transformed into the required bacterial cell-line.

2.6.2.5 Purification Of Plasmid DNA

When required, plasmid DNA was purified from bacterial culture with the use of a Qiagen QIAspin Miniprep Kit¹²⁷ according to the accompanying manufacturers protocol (see Appendix A.4.3)

2.7 Protein Expression

2.7.1 Bacterial Transformation

An overview of the transformation of bacterial strains with plasmid DNA is given below. Specifics for each transformation are given in the relevant chapters. Briefly, an aliquot of competent cells was incubated on ice in the presence of plasmid DNA. Uptake of plasmid DNA was induced by either electroporation or the heat-shock method. Growth media was added diluting the cells 10-fold and incubated at 37°C for 1hr. Serial dilutions of grown cells were plated out on LB agar containing appropriate antibiotics, allowed to dry and incubated upside-down at 37°C overnight.

2.7.2 Large Scale Culture Growth

A single colony of either Rosetta 2(DE3)pLysS or B834 *E. coli* cells transformed with the appropriate plasmid DNA was used to inoculate 50ml of

sterile LB supplemented with appropriate antibiotics and grown in a shaking incubator at 37°C, 170 revolutions per minute (rpm) overnight (O/N). Large volume cultures of either O/N Express media or LB, supplemented with appropriate antibiotics were inoculated with 0.1% of their volume of O/N culture. Large volume cultures were grown in a Unitron shaking incubator (Infors, Bottmingen Switzerland) at 37°C, 170rpm until O.D₆₀₀ reached 0.6, upon which 1mM IPTG was added (if using LB) and temperature lowered to 25°C and allowed to grow for the remainder of the 24hr period post-inoculation. Bacterial cultures were harvested by centrifugation at 4000rpm, 4°C for 30mins in a Beckman J-6B large volume centrifuge to pellet bacteria. Supernatant was discarded and cells were stored at -20°C until required.

2.7.3 Cell Lysis

Cells expressing the desired recombinant protein were lysed using a French Pressure cell. Cells were resuspended in lysis buffer to a final volume of 100ml per litre of cultured cells. Resuspended culture was supplemented with 2µl DNase1 (Invitrogen, Paisley UK), a Complete EDTA-free protease inhibitor tablet and 25µl 0.5M MgSO₄ to aid in bacterial DNA breakdown and inhibition of undesired protease activity. Lysis was carried out on ice at 950psi. The sample was passed through the French Press x5 to ensure maximum lysis.

2.8 Protein Purification

2.8.1 General Notes

This section is intended to give an overview into the protein purification processes used. Specific details pertaining to the various proteins purified are given in the subsequent relevant chapters.

2.8.2 Ni-NTA (*nitrilotriacetic acid*) IMAC Chromatography

Ni-affinity chromatography was used as the primary protein purification step. Buffers used in the purification process are given in table 2.10 1ml or 5ml HisTrap HP (GE Healthcare) nickel affinity columns were equilibrated by

washing with 5x column volumes Nickel Wash Buffer using a bench-top peristaltic pump (Masterflex C/L) at a rate of 2ml/min. The fractionated sample of bacterial cell lysate containing the soluble protein to be purified was loaded onto the HisTrap HP nickel column with the peristaltic pump at a rate of 2ml/min and allowed to cycle through the column for 30mins to ensure maximum binding to the Ni-resin. During this process the sample was kept on ice. The column was then washed with 5x column volumes of wash buffer to remove any unbound proteins, and transferred to an ÄKTAprime™ Plus (GE Healthcare) which was used to run a Nickel Affinity Purification program (see Appendix A.5.2). The bound protein was eluted from the column using a linear gradient of Nickel Elution Buffer (see table 2.10). Fractions were collected and analysed by SDS-PAGE as described in 2.10.1.

Buffer Name	Composition	Comments
Nickel Wash Buffer	50mM Tris HCl pH7.5 500mM NaCl 20mM Imidazole	pH adjusted to 7.5 with HCl, Filter sterilised with 0.22µm filter
Nickel Elution Buffer	50mM Tris HCl pH7.5 500mM NaCl 500mM Imidazole	pH adjusted to 7.5 with HCl, Filter sterilised with 0.22µm filter
Lysis Buffer	Wash Buffer + 0.2% tween	Filter sterilised with 0.22µm filter

Table 2.10: Buffers used in Ni-affinity purification

2.8.3 Ion Exchange Chromatography

Ion exchange chromatography was used as a subsequent purification step to Ni-affinity chromatography. Isoelectric point (P_i) of the protein being purified was determined by submitting the protein sequence to the into the ProtParam program held on the Expert Protein Analysis Server ‘ExpASY’ (<http://www.expasy.ch/tools/protparam.html>).¹¹⁴ Choice of ion exchange column was determined by P_i of the protein being purified. Protein P_i values allowed 1ml or 5ml CaptopQ anion exchange columns (GE Healthcare) to be used exclusively. Buffers used throughout are described in table 2.11 The

column was washed for 5x column volumes with Anion Exchange Buffer A at a rate of 2ml/min before loading the protein sample. The sample was cycled through the column for 30mins to ensure maximum binding. The sample was kept on ice throughout. The column was then washed with 5x column volumes of Anion Buffer A to remove unbound proteins and transferred to an ÄKTAprime™ Plus (GE Healthcare) which was used to run an Anion Exchange Purification program (see Appendix A.5.3). The bound protein was eluted from the column using a linear gradient of Anion Exchange Buffer B (see table 2.11). Fractions were collected and analysed by SDS-PAGE as described in 2.10.1.

Buffer Name	Composition		Comments
Anion Exchange Buffer A	20mM Tris HCl pH8		pH adjusted to 8.0 with HCl, Filter sterilised with 0.22µm filter
Anion Exchange Buffer B	20mM Tris HCl pH8 1M NaCl		pH adjusted to 8.0 with HCl, Filter sterilised with 0.22µm filter
1 x Dulbecco's Phosphate Buffered Saline (PBS).	NaCl	10g	Make to 1L. pH adjusted to 7.12 - 7.3. Filter sterilised through 0.22µm filter
	KCl	250mg	
	Na ₂ HPO ₄	1.44g	
	KH ₂ PO ₄	410mg	

Table 2.11: Anion Exchange Buffers

2.8.4 Size Exclusion Chromatography

Size exclusion chromatography was performed on an S75 300ml size exclusion column (GE Healthcare) using an ÄktaPrime™ plus. Prior to loading, the protein was buffer exchanged into 20mM Tris HCl pH7.5, 200mM NaCl. The S75 column was washed with 2ml of the above buffer before sample application. The column was then run at a flow rate of 0.2ml per minute for 1.5 column volumes. 5ml fractions were collected and analysed by SDS-PAGE.

2.9 Protein Manipulation

2.9.1 Cleavage Of His₆ Tag With C3-Protease.

The His₆ tag was removed from the recombinantly expressed protein during the purification process. 5µl C3-protease (see chapter 3.2 for generation of HRV C3-protease) was added per 10ml protein sample to be cleaved. This was supplemented with 10mM DTT, 1mM EDTA and incubated at 4°C while stirring with a magnetic stirring bar. Incubation was performed for at least 3hrs and preferably O/N.

2.9.2 Buffer Exchange Of Protein Samples

When required, protein samples were buffer exchanged using PD-10 Desalting Columns (GE Healthcare). Buffer exchange was performed according to the accompanying manufacturers protocol (see appendix A.5.1).¹²⁸

2.9.3 Concentration Of Protein

In all cases, purified protein samples were concentrated by using Amicon Ultra-4 or -15 Centrifugal Concentrators (Millipore, Watford UK). Molecular Weight Cut-Off (MWCO) of concentrator membranes was varied according to molecular weight of purified protein. Samples were concentrated at 3000 x g, 4°C in a Sigma 4K15 tabletop centrifuge (Sigma, Harz Germany) for 10mins. Samples were resuspended after each run to prevent aggregation or precipitation of the protein at the bottom of the concentrator during sample concentration. The process was repeated as necessary until the protein sample reached the desired concentration. Protein samples for immediate use were stored at 4°C, otherwise the purified, concentrated protein was aliquotted to avoid excessive freeze/thaw cycles and stored at -20°C for short-term storage or -80°C for long-term storage.

2.10 Protein Analysis

2.10.1 SDS-PAGE

SDS-PAGE analysis was carried out using the Invitrogen NuPAGE[®] Novex precast gel system which is based upon the Laemmli method.¹²⁹ All analysis was carried out using 4-12% Bis-Tris gels in 1xNuPAGE MES buffer. Protein samples were mixed with NuPAGE loading dye supplemented with 10mM dithiothreitol (DTT) and denatured in a hot-block at 90°C for 5mins. 10µl of each prepared sample was run. 5µl of SeeBlue Plus 2 prestained molecular weight markers were included as standard. Electrophoresis was performed at 200V for 35mins. SDS-PAGE gels were stained with coomassie R250 stain, destained with several changes of destain, then followed by several changes of dH₂O. Coomassie stain and destain are detailed in table 2.12.

Buffer Name	Composition	Comments
1x NuPAGE MES running buffer	Proprietary formulation	Diluted from 25x concentrated stock with dH ₂ O
Coomassie R250 stain	1.25g/l Coomassie R250, 40% EtOH, 10% glacial acetic acid	Gravity-filtered through coarse filter paper before use
Coomassie destain	40% EtOH, 10% glacial acetic acid	N/A

Table 2.12: Buffers used in SDS-PAGE

2.10.2 Measurement Of Protein Concentration

An absorbance reading at 280nm (A_{280}) was taken for the protein sample using a NanoDrop ND-1000 (Fisher Scientific, Loughborough UK). The extinction coefficient of the protein sample was calculated by submitting the protein sequence into the ProtParam program held on the Expert Protein Analysis Server 'ExpASY' (<http://www.expasy.ch/tools/protparam.html>).¹¹⁴ Using this information, a derivation of the Beer-Lambert law was used to calculate protein concentration (illustrated in equation 2.1).

$$[\text{Protein}] = (A_{280} / \text{Extinction Co-eff}) \times \text{Dilution Factor}$$

Equation 2.1: Equation for calculating protein concentration

Where:

[Protein] = protein concentration

A₂₈₀ = Absorbance value of sample taken at 280nm

2.10.3 Western Blot

Protein to be analysed by Western blotting was first subjected to SDS-PAGE as outlined in 2.10.1. Instead of staining the gel, the proteins were transferred to a Hybond™-P PolyVinylidene Fluoride (PVDF) membrane (Amersham Biosciences - GE Healthcare) using an Xcell II blotting module (Invitrogen)¹³⁰ according to the manufacturer's instructions. Transfer was carried out at 30V for 1hr. Before assembly of the module, the Hybond-P membrane was activated by immersing in methanol for 1min. Probing of the membrane with relevant antibodies was carried out as described in the standard operating procedures devised by the Mitchell lab (see appendix A.3.2).

2.10.4 Dynamic Light Scattering (DLS)

DLS was employed to assess the molecular dispersal of the purified protein in solution. 5.5µl samples were analysed using an AvidNano W130i System at GlaxoSmithKline, Stevenage UK.

2.10.5 Isothermal Titration Microcalorimetry (ITC)

ITC was performed at the Biophysical Chemistry Group, University of Glasgow UK with the assistance of Margaret Nutley.

2.10.6 *Mass Spectrometry*

MALDI-ToF mass spectrometry was performed in collaboration with both Dr. Richard Burchmore at the Sir Henry Wellcome Functional Genomics facility, University of Glasgow UK, and Satty Borman at GlaxoSmithKline, Stevenage UK.

2.10.7 *N-terminal Protein Sequencing*

N-terminal sequencing was performed using the Edman Degradation method¹³¹ in collaboration with Neil Freeman at GlaxoSmithKline, Stevenage UK. Samples were prepared for sequencing by SDS-PAGE analysis as described in 2.8.1. and transferred to a PVDF membrane (Invitrogen) as for transfer for Western blot (see 2.8.2). The PVDF membrane was stained with Coomassie R250 stain and destained. The membrane was washed extensively to remove residual acid from the membrane and allowed to air dry. Samples were analysed using a Procise[®] protein sequencer (Applied Biosystems).

2.10.8 *Nuclear Magnetic Resonance*

1-Dimensional NMR was performed in collaboration with Dr. Brian Smith at the Protein Science Group, University of Glasgow UK. Protein was supplied in 1x Dulbecco's PBS buffer (see table 2.11). Further details are given in the relevant chapters.

2.10.9 *Circular Dichroism (CD)*

Far-UV CD was carried out in collaboration with Dr. Sharon Kelly at the Protein Characterisation Facility, University of Glasgow UK. Far-UV CD experiments were also undertaken at GlaxoSmithKline, Stevenage UK. Experiments were carried out at ambient temperature in Tris buffer comprising 50mM Tris HCl pH7.5, 500mM NaCl. Further information is given in the relevant chapters.

2.10.10 *Protein Melting Temperature (T_m)*

The effect of different protein buffers on the T_m of the protein samples was assessed by CD. Samples were measured at a fixed wavelength over increasing temperature at set time-intervals. Increase in the proportion of random coil in the sample was monitored as an indication of protein denaturation. Further details are given in the relevant chapter.

2.10.11 *Analytical Ultra-Centrifugation (AUC)*

AUC was carried out in collaboration with Dr. Olwyn Byron, Glasgow Biomedical Research Centre, University of Glasgow UK. Protein samples were provided at a concentration of 5mg/ml in 1x Dulbecco's PBS buffer.

2.10.12 *Limited Proteolysis*

Limited proteolysis was performed at ambient temperature. Essentially, the experiment was a controlled time-course tryptic digest. Purified F/L PhtD protein was incubated with proteomics grade trypsin (Roche) and samples removed at set time-points. Samples were either denatured by mixing in a 1:1 ratio with NuPAGE SDS loading-dye and analysed by SDS-PAGE as outlined in 2.10.1 or quenched by addition of Trichloroacetic acid (TCA) for intact-mass spectrometry. Specific reaction details are given in the relevant chapters.

2.11 Protein Crystallisation

2.11.1 *Crystal Screening*

Screening of proteins for protein crystallisation was carried out at the University of Glasgow using the sitting drop vapour diffusion method in 96 well plate format. Crystal screens were dispensed using a Hamilton Liquid Handling Robot (Hamilton), crystal trials were performed using a HoneyBee 8+1 robot (Cartesian), stored at 20°C and monitored of crystal growth.

Crystallisation screening was also performed at GlaxoSmithKline, Stevenage UK. Crystal screens were dispensed using a Hydra (Thermo Scientific). Crystal trials were performed using a Mosquito (Molecular Dimensions). Crystal trials were stored at 20°C and monitored for crystal growth.

Latterly, crystallisation screening was performed in collaboration with Prof. Juan Hermoso at the Instituto de Quimica Fisica “Rocasolano”, Spanish National Research Council, Madrid Spain.

2.11.2 *Crystal Screens*

Many commercially available crystal screens were utilised throughout the screening process. A list of the crystal screens used and their manufacturers is given in table 2.13.

Crystal Screen	Manufacturer
Crystal Screen 1	Hampton Research
Crystal Screen 2	Hampton Research
PEG/Ion Screen	Hampton Research
CryoScreen 1	Emerald Biosystems
CryoScreen 2	Emerald Biosystems
Wizard 1	Emerald Biosystems
Wizard 2	Emerald Biosystems
Structure Screen 1	Molecular Dimensions
Structure Screen 2	Molecular Dimensions
Morpheus Screen	Molecular Dimensions
Cations Screen	Qiagen
PEGsII Screen	Qiagen
PACT Screen	Qiagen
Nextal Classics Screen	Qiagen
JCSG+	Qiagen
Ammonium Sulphate Screen	Qiagen

Table 2.13: Commercial crystal screens

2.12 Computational Software

A variety of computer software was used throughout the duration of this project for data analysis and in presentation of data in this thesis.

Crystallographic computer work was performed using programs from the *CCP4*¹³² and *PHENIX*¹³³ suites of crystallographic software. Processing of raw X-ray diffraction data was performed using *MOSFILM*¹³⁴ or *D*TREK*.¹³⁵

Visualisation of electron density maps and molecular models, and the manual manipulation of molecular models were performed using *COOT*.¹³⁶ Diagrams of molecular models determined by macromolecular crystallography in this thesis were generated using *PyMOL*.¹³⁷

Pairwise sequence alignments of DNA and protein sequences for illustrative purposes in this thesis were performed using the program *MULTALIN*.¹³⁸

Physical and chemical parameters (e.g. predicted molecular weight iso-electric point) for the proteins in this thesis were calculated from the respective protein sequence using the program *PROTPARAM* hosted at <http://expasy.org/tools/protparam.html>

Sequence-based disorder prediction of the F/L PhtD protein and its truncated versions was carried out using the Random Order Neural Network (RONN)¹³⁹ hosted at <http://www.strubi.ox.ac.uk/RONN>

Graphical illustrations for analysis of CD and T_m data and illustrative representation of the corresponding results in this thesis were generated using *PRISM 4*.

Analysis of DNA sequencing data was performed using *VectorNTI 11.5* (Invitrogen)¹⁴⁰ or *CLC Genomics Workbench 4* (CLCbio, Aarhus Denmark. www.clcbio.com).¹²⁰ Expression constructs were designed and plasmid maps generated using *CLC Genomics workbench 4*.

Chapter 3 : Crystallographic Studies Of Proteins From Gram Positive Bacteria

3.1 Overview

The data presented in this chapter was generated through the process of learning the technique of X-ray crystallography, including crystal manipulation, handling X-ray diffraction data and structure determination, to provide experience of data processing prior to the event of successful crystallisation of PhtD.

Both of the proteins described in this chapter were chosen for investigation for a number of reasons. Firstly, both could be purified in milligram quantities, a prerequisite for X-ray crystallographic studies due to the high concentrations normally used in crystallisation experiments. The expression plasmid for the transketolase (TktA) protein was provided through collaborative work; a search of the Protein Data Bank (PDB)¹⁴¹ revealed that transketolase protein structures had been successfully determined in the past, indicating that there was a possibility that we would succeed in obtaining crystals, and of determining the structure by the molecular replacement method using a structure present in the PDB. As the protein Putative Protease Maturation protein A (PpmA) from the pneumococcus had previously been investigated in our laboratory, PpmA was included in crystallographic studies in an attempt to further progress with this protein. As a novel protein, molecular replacement would not be possible in this case, requiring incorporation of heavy-metals for successful structural determination.

As such, investigation into the determination of the crystal structure of the transketolase enzyme TktA from *Lactobacillus salivarius* would provide training for how to handle protein crystals, formulate planning of data-collection for the TktA crystals, and to provide experience of processing the raw-data through various crystallographic computer programs, subsequently using the processed data to determine the 3D crystal structure of TktA by the increasingly common method of molecular replacement. In essence, TktA was

used as a positive control or exercise in the process of protein structure determination -from protein purification to final refinement of the 3D-crystal structure- in readiness for handling any PhtD protein crystals that were generated.

In addition, the pneumococcal protein PpmA was investigated due to prior interest in our laboratory, where biochemical data for this protein has already been obtained. As a novel protein, PpmA was included in this study as an exercise in problem solving; the 3D crystal structure of PpmA is desired in order to further understand the protein function. Because it is a novel protein, crystallisation and structural determination of PpmA would provide experience of ways (other than molecular replacement) of solving the 'phase-problem' in X-ray crystallography, such as selenomethionine (Se-Met) labelling of PpmA or incorporation of heavy-metals into native PpmA protein crystals, which were not required in the case of TktA but would be required in the structural determination of PhtD, as although a fragment of one of the other Pht proteins has been successfully determined,⁹⁸ this fragment alone would be insufficient to determine the whole 3D crystal structure of PhtD through molecular replacement, therefore additional methods such as those described above would also need to be implemented.

The transketolase TktA from the lactic acid bacterium *Lactobacillus salivarius* acts as a link between glycolysis and the pentose-phosphate pathway. The transketolase enzyme is able to mediate the reversible transfer of a two-carbon ketol unit from a ketose to an aldose (specifically from D-xylulose 5-phosphate to D-ribose 5-phosphate, generating D-sedulose 7-phosphate and D-glyceraldehyde 3-phosphate in the process).^{142, 143} A key enzyme in sugar synthesis, the transketolase is able to shift excess quantities of fructose-3-phosphate and D-glyceraldehyde-3-phosphate from glycolysis to the pentose-phosphate shunt, thus eliminating an excess of these harmful metabolites from the cytosol.¹⁴⁴

The pneumococcal protein PpmA has been predicted to be a member of the family of proteins classed peptidyl-proline isomerases (PPIases). Specific information on PpmA is outlined later in this chapter (see chapter 3.10). PPIases mediate the *cis-trans* isomerisation of peptide bonds in proline

residues. Unlike other amino acids, proline is unique in its structure; the tetrahedral CH₃ groups of proline are similarly 'crowded' in either *cis* or *trans* state, and therefore a relatively small free energy difference exists between the *cis* and *trans* conformations. This means that unlike other amino acids - which are present predominantly as *trans* isomers- proline residues are able to exist distinctly in either the *cis* or *trans* state.¹⁴⁵ Proline will not spontaneously adopt the intended conformation; a PPlase is needed in order to effect the required isomerisation.¹⁴⁶ PPlases therefore act like chaperones to protein folding, limiting the rate at which folding occurs and preventing the accumulation of reaction intermediates that are prone to form non-native protein conformations.¹⁴⁷

Expression, Purification, Crystallisation And Data Collection Of TktA from *Lactobacillus salivarius*

3.2 Reference

This chapter is a detailed explanation of the work published in Horsham, M., Saxby, H., Blake, J., Isaacs, N.W., Mitchell, T.J. and Riboldi-Tunncliffe, A. “Expression, purification, crystallisation and preliminary X-ray crystallographic data from TktA, A transketolase from the lactic acid bacterium *Lactobacillus salivarius*” *Acta cryst F*, **66**, 899-901 (2010).¹⁴⁸

3.3 Sequence Search Of The PDB With TktA

The protein sequence for TktA (see Appendix A.1.10.2) was downloaded from the CMR database¹⁴⁹ hosted by the JCVI institute at <http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi>

A sequence search of the PDB¹⁴¹ was performed using default settings with the TktA amino acid sequence as the search query, which returned 45 possible matches, including two structures determined by NMR. This showed that the structures of transketolase enzymes from other sources had been successfully determined, and that there was a possibility that crystallisation of TktA from *L. salivarius* would be successful. Furthermore, it revealed that the structure of TktA could be determined by molecular replacement if crystals were successfully grown.

3.4 Overexpression Of TktA From *L. salivarius*

3.4.1 TktA Clone

An expression plasmid containing the full gene encoding the TktA protein (LSL1946) was kindly provided by Dr. Martin Walsh (ESRF BM-14, Grenoble France; Diamond Light Source, Oxford UK).^{150, 151} The *TktA* gene was inserted

in-frame into the vector pOPINF¹¹⁹ (see chapter 2.3.3), which encodes a His₆ tag at the N-terminus, separated by a HRV-C3 protease cleavage site (MAHHHHHSSGLEVL^{FQ↓}GP). The vector map and amino acid sequence is given in figure 3.1 The clone was transformed as described in 2.7.1 into Subcloning Efficiency™ DH5α™ competent cells and stored long term as a glycerol stock at -80°C.

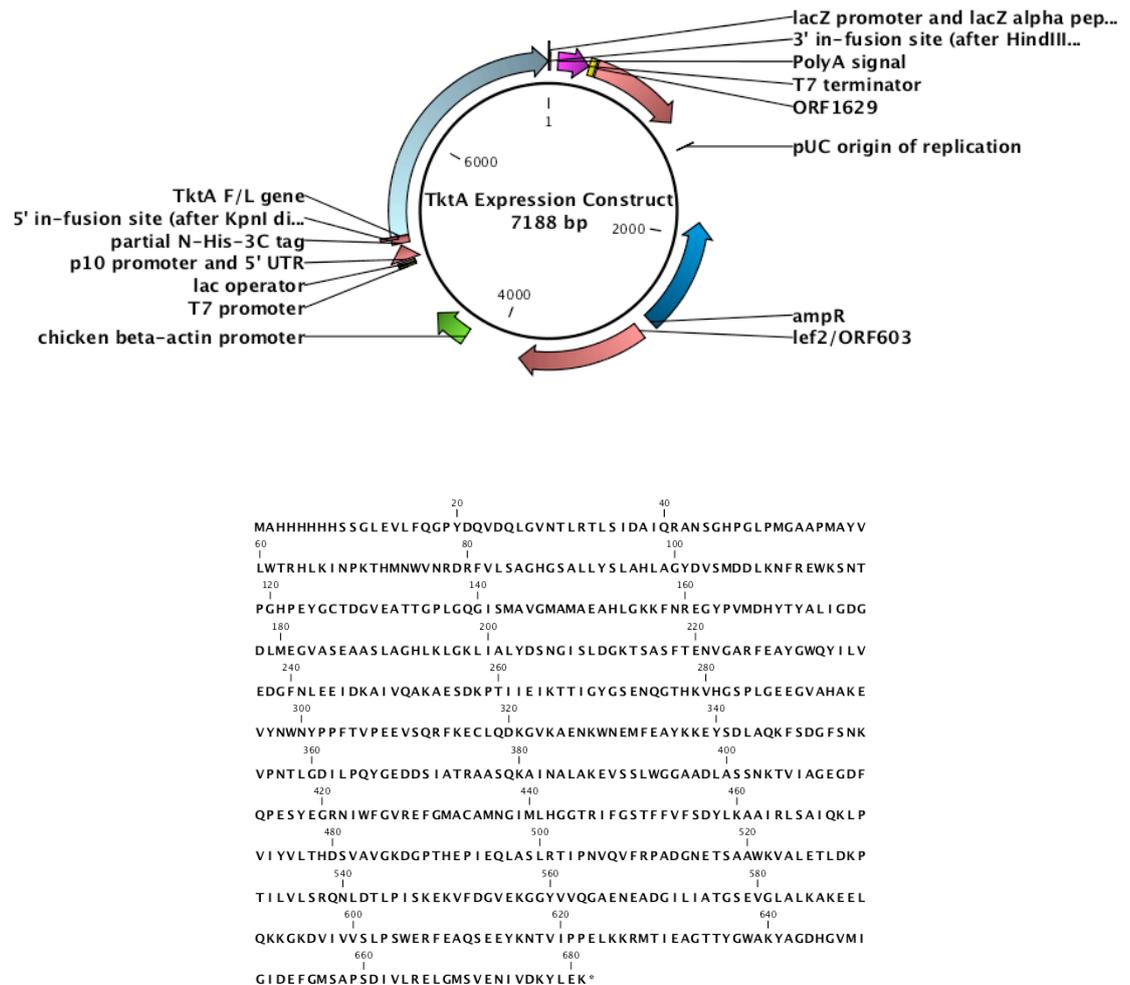


Figure 3.1: TktA expression construct

DNA was fused N-terminally to a His₆-tag that incorporates an HRV-C3 protease site for cleavage of the His₆-tag. Regulation of recombinant protein expression was placed under control of the upstream T7 promoter. The translated protein sequence is also illustrated. The plasmid map was generated using CLC Genomics Workbench 4.¹²⁰

3.4.2 Transformation And Expression of TktA

The TktA expression construct was transformed into the *E. coli* expression strain 2(DE3)pLysS as described in 2.7.1. ~400ng expression plasmid was

incubated with 50 μ l 2(DE3)pLysS cells on ice for 30mins. The cells were heat-shocked at 42°C for 30secs in a water-bath, thereafter the cells were allowed to recover on ice for 2mins. Cells were diluted with 450 μ l GS-96 growth media and incubated in a standing incubator at 37°C for 1hr. Cells were grown on LB agar supplemented with 54 μ g/ml carbenicillin and 34 μ g/ml chloramphenicol O/N at 37°C.

A small volume culture of 50ml sterile LB supplemented with the above antibiotics was inoculated with a single colony of transformed cells and incubated in a shaking incubator at 37°C, 170rpm O/N. The following day, 2x500ml cultures of sterile LB medium supplemented with the above antibiotics were each inoculated with 5ml of the O/N starter culture. These expression cultures were incubated at 37°C, 170rpm until OD₆₀₀ had reached 0.6. Thereafter, 500 μ l 1M IPTG was added to each and the temperature was lowered to 22°C and incubated for the remainder of the 24hr period. The cultures were harvested at 4000rpm, 4°C for 30mins in a Beckman J-6B large volume centrifuge to pellet bacteria.

3.5 Purification Of TktA

3.5.1 IMAC Ni-Affinity Chromatography

Cell pellets were divided in two and re-suspended in 50ml lysis buffer (See table 2.10) and each was supplemented with 2 μ l DNase1, a Complete EDTA-free protease inhibitor tablet and 25 μ l 0.5M MgSO₄ to aid in bacterial DNA breakdown and inhibition of undesired protease activity. Cells were lysed as described in 2.7.3 Lysed cells were centrifuged at 7500 x g, 4°C for 30mins and the pellet discarded. The supernatant was then further centrifuged at 40,000 x g, 4°C for 30mins, the pellet discarded and the supernatant passed through a 0.22 μ m syringe filter.

The cleared lysate was subjected to Ni-affinity IMAC purification as described in 2.8.2 The loaded column was transferred to an ÄKTAprime plus purification system, which was used to implement a linear gradient elution program from Ni-buffer A to B (for composition see table 2.10). A typical elution profile for purification of TktA is shown in figure 3.2 Eluted samples corresponding to

the A₂₈₀ peak were analysed by SDS-PAGE as outlined in 2.10.1. The resulting gel is shown in figure 3.3 Selected fractions were pooled and dialysed back into Ni-buffer A to remove excess imidazole over-night at 4°C in 10 kDa MWCO dialysis tubing which had been pre-treated by boiling for 10mins in dH₂O.

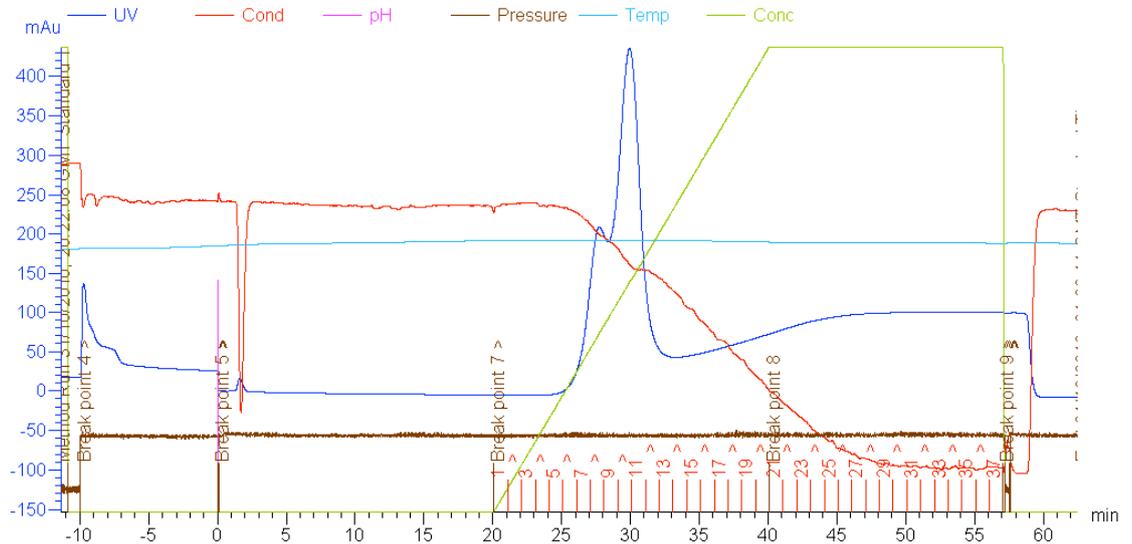


Figure 3.2: Elution profile for Ni-affinity purification of TktA

Purification was carried out on Äkta Prime Plus purification system, using a 1ml HisTrap HP column. Protein was eluted from the column at a flow rate of 2ml/min using a linear gradient elution starting at 0% and ending at 100% of nickel buffer B. (Refer to table 2.10). Fractions corresponding to the peak in the UV trace were collected for analysis by SDS-PAGE. Fractions corresponding to the peak in the UV trace were collected for SDS-PAGE analysis.

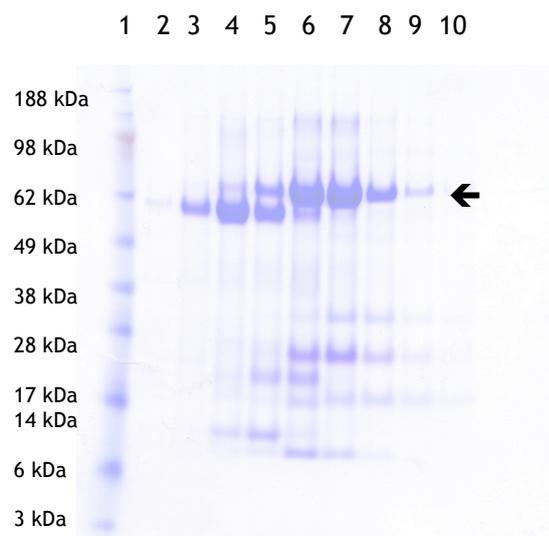


Figure 3.3: SDS-PAGE gel of selected fractions from Ni-affinity purification of TktA

Lane 1: SeeBlue Plus 2 mol wt. marker, Lane 2 - 10: elution fractions. The arrow indicates the recombinantly expressed TktA.

3.5.2 Concentration Of Purified TktA

Purified TktA was subjected to a rigorous routine of concentration by centrifugation as described in 2.9.3 Protein concentration was measured as described in 2.10.2 Concentration of TktA was halted at 13.2mg/ml as it was over the 10mg/ml threshold deemed acceptable for attempting crystallisation trials. Protein was stored at 4°C for immediate use.

3.6 Crystallisation And Data Collection Of TktA

Purified, concentrated TktA was subjected to crystal trials using the sitting-drop vapour diffusion technique at the University Of Glasgow as outlined in 2.11.1. Crystal screening was carried out using the commercial screens JCSG+, PEG/Ion, GridScreen, and Cryo1&2 screens (see table 2.13). Protein was dispensed at 13.2mg/ml and 9.1mg/ml.

After 72hrs incubation at 20°C, crystals were observed in a number of the conditions at both protein concentrations from the JCSG+, PEG/Ion and Cryo1&2 screens, with the vast majority appearing in conditions from the PEG/Ion screen. As a general observation, it was noted that the crystals produced at the lower protein concentration were smaller and “uglier”. Two conditions- JCSG+ 22 (0.2M magnesium chloride, 0.1M sodium cocodylate, 50% (v/v) PEG200) and PEG/Ion screen2 36 (0.07M citric acid, 0.03M Bis-Tris propane pH3.4, 16% (w/v) PEG3350) produced small rhombic plates. All other crystallising conditions produced crystals that had the appearance of hexagonal rods. Typical crystals for the hexagonal form are shown in figure 3.4 Rhombic crystals are shown in figure 3.5 A summary of the conditions that yielded crystals is given in table 3.1.

Screen & Condition	Composition	Protein Concentration (mg/ml)	Crystal Morphology
Cryo1&2 24	0.1M sodium-potassium phosphate pH6.2, 0.2M sodium chloride, 40% (v/v) PEG 400	13.2 & 9.1	Hexagonal rods (edges not well defined)
JCSG+ 8	0.2M ammonium formate, 20% (w/v) PEG 3350	13.2 & 9.1	Hexagonal rods
JCSG+ 10	0.2M potassium formate, 20% (w/v) PEG 3350	13.2	Single Rectangular crystal
JCSG+ 22	0.2M magnesium chloride, 0.1M sodium cocodylate, 50% (v/v) PEG200	13.2	Rhombic plates
JCSG+ 37	24% (w/v) PEG 1500, 20% (w/v) glycerol	13.2 & 9.1	Hexagonal rods
PEG/Ion1 1	0.2M sodium fluoride pH7.3, 20% (w/v) PEG 3350	13.2 & 9.1	Hexagonal rods
PEG/Ion1 2	0.2M potassium fluoride pH7.3, 20% (w/v) PEG 3350	13.2	Hexagonal rods
PEG/Ion1 3	0.2M ammonium fluoride pH6.2, 20% (w/v) PEG 3350	13.2	Hexagonal rods
PEG/Ion1 5	0.2M magnesium chloride pH5.9, 20% (w/v) PEG 3350	13.2 & 9.1	Hexagonal rods
PEG/Ion1 6	0.2M sodium chloride pH6.9, 20% (w/v) PEG 3350	13.2	Hexagonal rods
PEG/Ion1 9	0.2M ammonium chloride pH6.3, 20% (w/v) PEG 3350	13.2 & 9.1	Hexagonal rods
PEG/Ion1 15	0.2M lithium nitrate pH7.1, 20% (w/v) PEG 3350	13.2	Hexagonal rods
PEG/Ion1 21	0.2M sodium formate pH7.2, 20% (w/v) PEG 3350	13.2	Hexagonal rods
PEG/Ion1 22	0.2M potassium formate pH7.3, 20% (w/v) PEG 3350	13.2 & 9.1	Hexagonal rods
PEG/Ion1 23	0.2M ammonium formate pH6.6, 20% (w/v) PEG 3350	13.2 & 9.1	Hexagonal rods
PEG/Ion1 27	0.2M sodium acetate pH8.0, 20% (w/v) PEG 3350	13.2	Hexagonal rods
PEG/Ion1 30	0.2M Ammonium acetate pH7.1, 20% (w/v) PEG 3350	13.2 & 9.1	Hexagonal rods
PEG/Ion2 26	0.2M sodium acetate pH7.0, 20% (w/v) PEG 3350	13.2	Hexagonal rods
PEG/Ion2 27	0.1M sodium formate pH7.0, 12% (w/v) PEG 3350	13.2 & 9.1	Hexagonal rods
PEG/Ion2 28	0.2M sodium formate pH7.0, 20% (w/v) PEG 3350	13.2	Hexagonal rods
PEG/Ion2 34	2% (v/v) tacsimate pH7.0, 0.1M HEPES pH7.5, 20% (w/v) PEG3350	13.2 & 9.1	Hexagonal rods
PEG/Ion2 36	0.07M citric acid, 0.03M Bit-Tris propane/pH3.4, 16% (w/v) PEG 3350	13.2 & 9.1	Rhombic plates
PEG/Ion2 37	0.06M citric acid, 0.04M Bit-Tris propane/pH4.1, 16% (w/v) PEG 3350	13.2 & 9.1	Hexagonal rods

Table 3.1: Crystal producing conditions from crystallisation screens



Figure 3.4: Typical hexagonal rod TktA crystals

Protein = TktA incorporating His₆ tag. Protein concentration = 13.2mg/ml. Drop vol = 1 μ l.

Droplet volume = 1 μ l, comprising 1:1 ratio of protein : reservoir solution.



Figure 3.5: Rhombic plate TktA crystals from PEG/Ion2 36

Protein = TktA incorporating His₆ tag. Protein concentration = 13.2mg/ml.

Droplet volume = 1 μ l, comprising 1:1 ratio of protein : reservoir solution.

3.6.1 In-House Testing Of TktA Crystals

In-house testing of the grown TktA crystals was performed on a Rigaku Micromax-007 HF X-Ray generator (Rigaku Europe, Kent UK). Data was collected on a Mar345 dtb detector. Crystals were flash-cooled in a stream of nitrogen gas maintained at 110K (-163°C). Attempts were made to cryoprotect the protein crystals utilising various cryoprotectants, with dried paraffin oil being most successful, producing fewest ice-rings and maintaining diffraction quality. Crystals were exposed to 2 x 30sec exposures separated by 90° to test for diffraction quality. Crystals grown in PEG/Ion 1 condition 6 provided the best in-house diffraction, therefore replicate trays of this condition were set up. The resulting crystals were cryoprotected with dried paraffin oil and frozen in a stream of nitrogen gas at 110K (-163°C) as before.

3.6.2 TktA Crystallographic Data Acquisition And Processing

Frozen TktA crystals measuring approximately 200 x 40 x 40 μm in size from the replicate trays of PEG/Ion 1 condition 6 were taken to the European Synchrotron Radiation Facility (ESRF - Grenoble, France) and exposed to synchrotron X-rays on the beamline ID14-1.¹⁵⁰ Diffraction data were collected on a single crystal using an ADSC Q210 CCD detector with a crystal rotation of 1° per frame. An example plate of the collected diffraction data is given in figures 3.6 and 3.7.

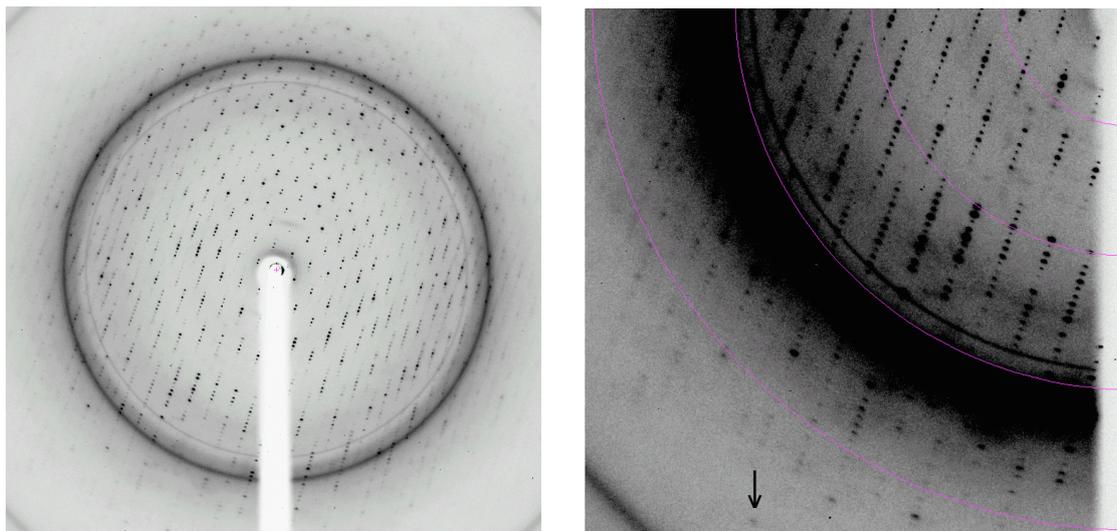


Figure 3.6: Diffraction image for a single crystal of TktA

Figure 3.7: Close-up diffraction image for a single crystal of TktA

The arrow shows the limit of diffraction at 2.3Å resolution. Outer blue arc shows diffraction limit at 2.9Å

The space group to which the crystal belonged was determined with the utilisation of the program *POINTLESS*,^{132, 152} which revealed that the crystal belonged to the space group $P3_221$. The collected data were processed in $P3_221$ with the crystallographic program *MOSFILM*.¹³⁴ A summary of the data collection statistics is given in table 3.2.

Source	ID14-1, ESRF, Grenoble ¹⁵⁰
Wavelength (Å)	0.934
Resolution (Å)	36.94-2.30 (2.42-2.30)
Space group	$P3_221$
Unit-cell parameters (Å, °)	$a = b = 75.43$, $c = 184.11$, $\alpha = \beta = 90.0$, $\gamma = 120.0$
R_{merge}	0.126 (0.352)
Total No. of reflections	55793
No. of unique reflections	25397
Average redundancy	2.2
Completeness (%)	92.8 (93.6)

Table 3.2: Data collection statistics for TktA

(Values in parentheses are for the outer shell)

Additionally, the crystallographic program *MATTHEWS*¹³² was utilised to estimate the solvent content of the crystal. Statistics are given in table 3.3.

Nmol/asym	Matthews Coeff ($\text{\AA}^3\text{Da}^{-1}$)	% Solvent	Probability
1	2.03	39.36	1.00

Table 3.3: Statistics for calculated Matthews coefficient of TktA crystal

Structural Determination Of TktA

3.7 Reference

The work discussed in this section is a version of Horsham, M., Mitchell, T.J & Riboldi-Tunncliffe, A. “2.3 Ångstroms Crystal Structure Of TktA, A transketolase from the lactic acid bacterium *Lactobacillus salivarius*” To be published.

3.8 Molecular Replacement

3.8.1 Solution To The Phase Problem

In order to calculate the electron density $\rho(xyz)$ from the measured intensities of indexed diffraction data for an unknown crystal structure, it is necessary to use the inverse Fourier transform of the structure factor vector $F(hkl)$.¹⁵³ The structure factor vector comprises two separate components; an amplitude, and a phase angle $F(hkl) = |F(hkl)|e^{i\alpha(hkl)}$ (equation 6.1).

$$\rho(xyz) = \frac{1}{V} \sum_{h,k,l} |F(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

Where: $\rho(xyz)$ = electron density at point (xyz) in the unit cell

V = the volume of the unit cell

$|F(hkl)|$ = structure factor amplitude for reflection (hkl)

$i\alpha(hkl)$ = phase angle for reflection (hkl)

Equation 3.1: The electron density equation

The X-ray diffraction experiment provides only the amplitude; the phase angle is lost. This gives rise to what is commonly known as the Phase Problem, as without the phase angle it is not possible to calculate the electron density in the unit cell. There are however, several ways to solve the phase problem, with one of the most popular being molecular replacement.

Molecular replacement relies upon the provision of a suitably homologous model to that of the unknown crystal structure. If a suitable model is available, then it can be used as a template for the unknown crystal structure. Using the known model as a template, the phase angles of the known model are applied to the indexed diffraction data of the unknown crystal structure. Molecular replacement consists of two steps; a rotation process and a translation process, to estimate the location of the target model in the unit cell.¹⁵⁴ With this initial template correctly in place, it is then possible to perform refinement of the model to give a final, chemically acceptable model of the target protein. It should be noted that with the exponential growth in the number of determined structures present in the Protein Data Bank (PDB),¹⁴¹ the method of molecular replacement has become exceedingly powerful, as the number of potential template models has increased drastically over the last few years, currently standing at greater than 65,000 entries with more than 85% of the deposited structures determined by X-ray crystallography.

3.8.2 Choice Of Molecular Replacement Model

A search of the PDB was performed using the amino acid sequence of TktA as before in the preliminary stages of setting up crystallographic trials (see chapter 3.3). With a sequence identity of 51% (illustrated in figure 3.8), the PDB entry 1ITZ (maize transketolase in complex with TPP) was chosen as the template to be used for molecular replacement. The PDB file was edited to remove any additional data such as H₂O molecules or ligands, leaving Chain A of the PDB file (the monomeric transketolase molecule) for use as a search model.

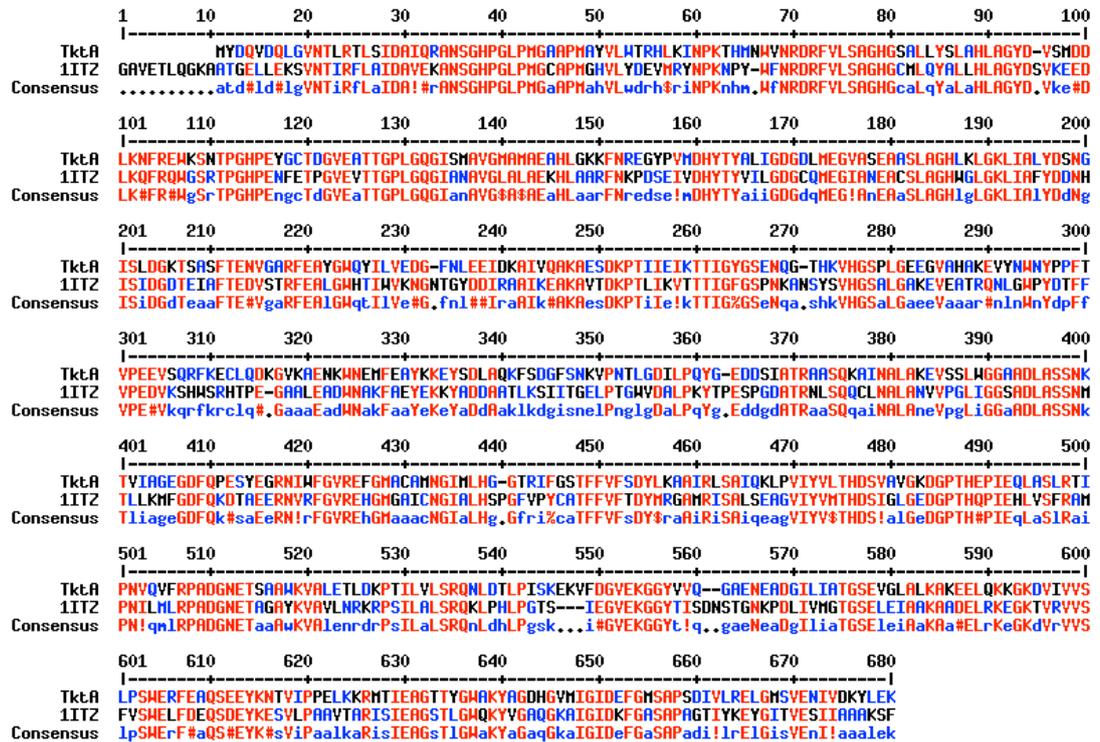


Figure 3.8: Sequence alignment of TktA and 1ITZ

Red = high consensus residues, Blue = low consensus residues, Black = neutral residues.

There is 51% identity between the TktA sequence and the sequence of the 1ITZ model. Alignment was performed using MULTALIN.¹³⁸

3.8.3 Molecular Replacement Of TktA

Molecular replacement of the TktA data was performed using *PHENIX AutoMR* from the *PHENIX*¹³³ suite of crystallographic programs. The edited 1ITZ PDB file comprising of Chain A alone was used as the search model for the molecular replacement of the indexed TktA data over the resolution range 36.94 - 2.30Å. One copy of the molecule was successfully located per asymmetric unit; the biologically relevant dimer is formed through a symmetry operation.

3.8.4 Model Building And Refinement

The molecular replacement solution from *AutoMR* is output as two separate files; the *.mtz* file contains the electron density map, and the *.pdb* file contains the model of the protein molecule. Visualisation of the initial MR

model and the electron density map was performed using *COOT*.¹³⁶ With the sequence alignment of TktA and 1ITZ as a reference, manual manipulation of the model was performed in *COOT*¹³⁶ by mutation, deletion or insertion of residues so that the model sequence was identical that of TktA. Restrained refinement of the TktA model was performed using *REFMAC5* from the *CCP4* suite of crystallographic programs.¹³² Iterative rounds of refinement were performed; in each case the resulting model was checked manually for any required alterations to the model. In order to prevent bias from occurring between the model and data, each round of refinement using *REFMAC5* was performed by refining the latest post-refinement model with the initial *.mtz* file containing the electron density map from the *AutoMR* solution, with R_{work} and R_{free} values successfully decreasing after each round of refinement. Figure 3.9 illustrates the electron density map of TktA and the TktA model fitted to the same region of electron density.

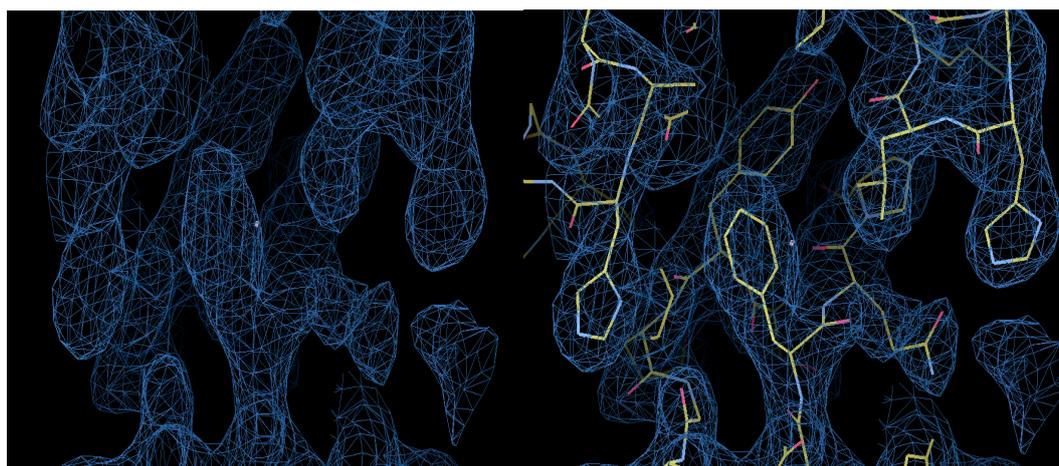


Figure 3.9: Modelling of TktA structure to electron density.

The image on the left shows the electron density for a region of TktA. The image on the right shows how the corresponding refined region of the TktA model fits into the aforementioned region of electron density. Figure generated using COOT.¹³⁶

3.8.5 Final Model Refinement And Analysis

Final model refinement is ongoing at present, with modelling of alternate conformations of residue side-chains and placement of H₂O atoms underway. Current model statistics show that $R_{work} = 20\%$ and $R_{free} = 26.8\%$. Validation of the structure (see figure 3.10) shows 4 outliers in the Ramachandran plot - which calculates acceptable bond geometry of the amino acids in protein

structures- corresponding to 0.68% of the structure, 554 (94.2%) residues lie in preferred regions with the final 30 (5.1%) present in allowed regions.

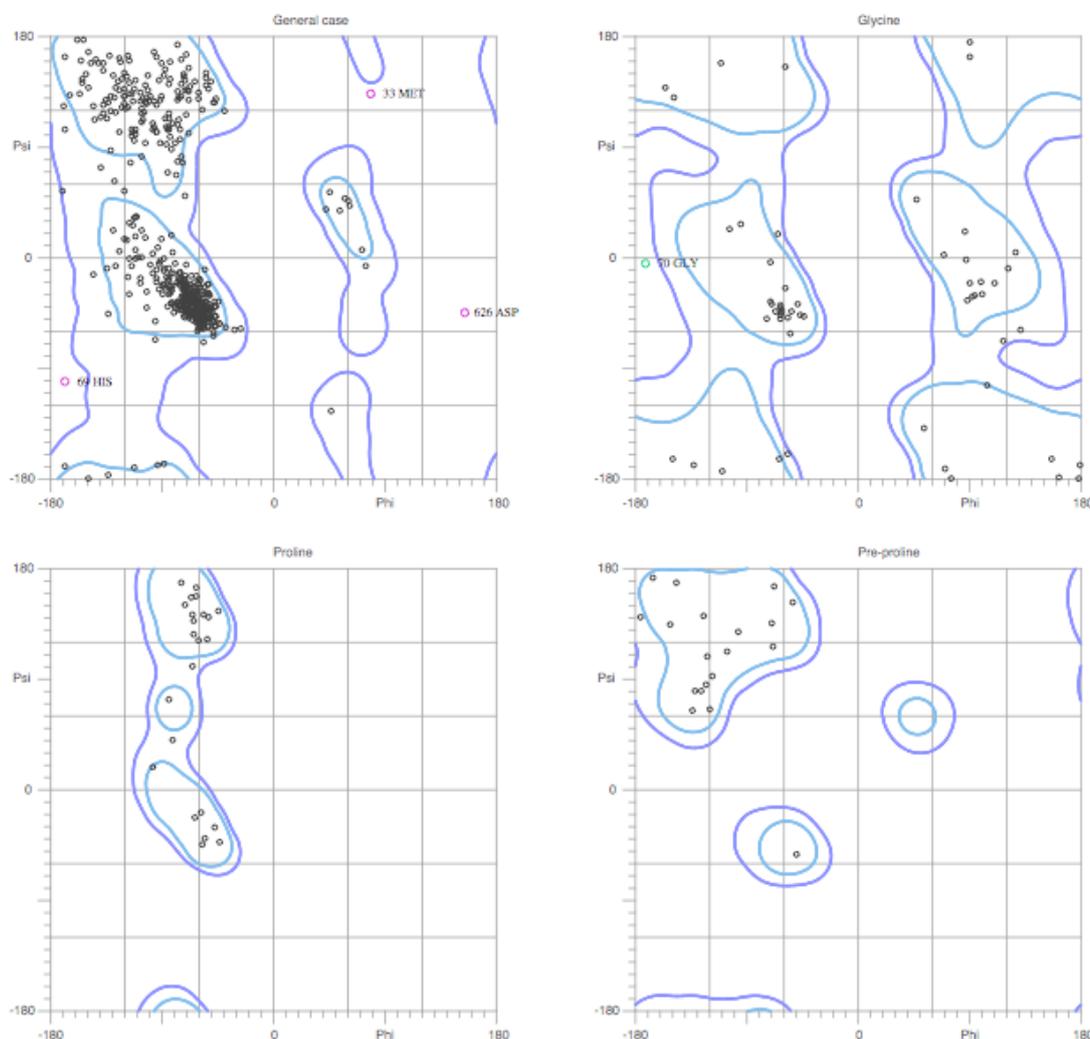


Figure 3.10: Ramachandran plot for refined TktA structure.

Validation of refined TktA structure. Four plots are generated corresponding to residue-class (indicated above each section). Light blue ringed areas encompass favoured regions, dark blue rings encompass allowed regions. Outliers are highlighted as individual named and numbered residues corresponding to protein sequence. The plot indicates that there are 4 outliers, corresponding to 0.68% of all residues, 94.2% of all residues are in favoured regions, 5.1% are in allowed regions. Ramachandran plots were generated using the MolProbity server.¹⁵⁵

172 H₂O molecules have been modelled into the structure to date; further modelling of waters (and additional unattributed density; possibly Chloride atoms, or PEG molecules) is ongoing. A molecular model of the TktA generated using the refined TktA monomer (figure 3.11) is shown in its biologically functional dimeric form in figure 3.12.

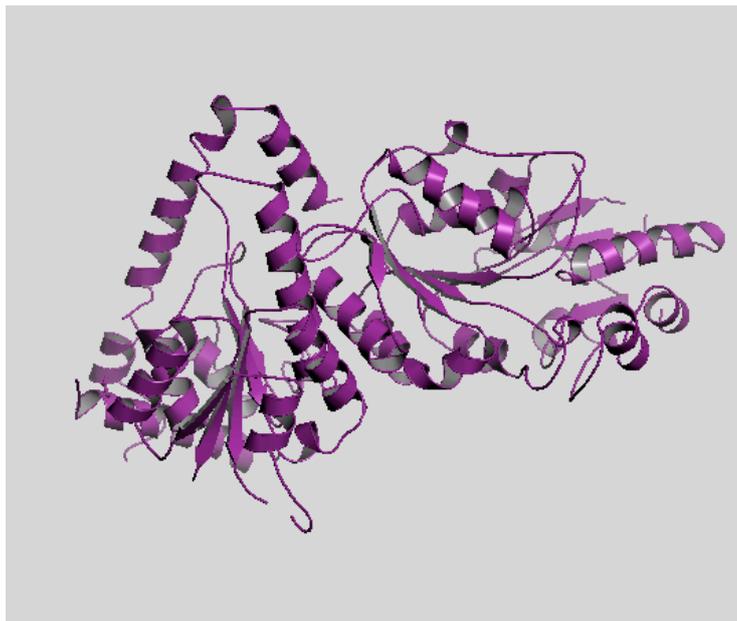


Figure 3.11: Monomeric molecular model of refined TktA structure
The image was generated from the refined TktA structure using PyMol.¹³⁷

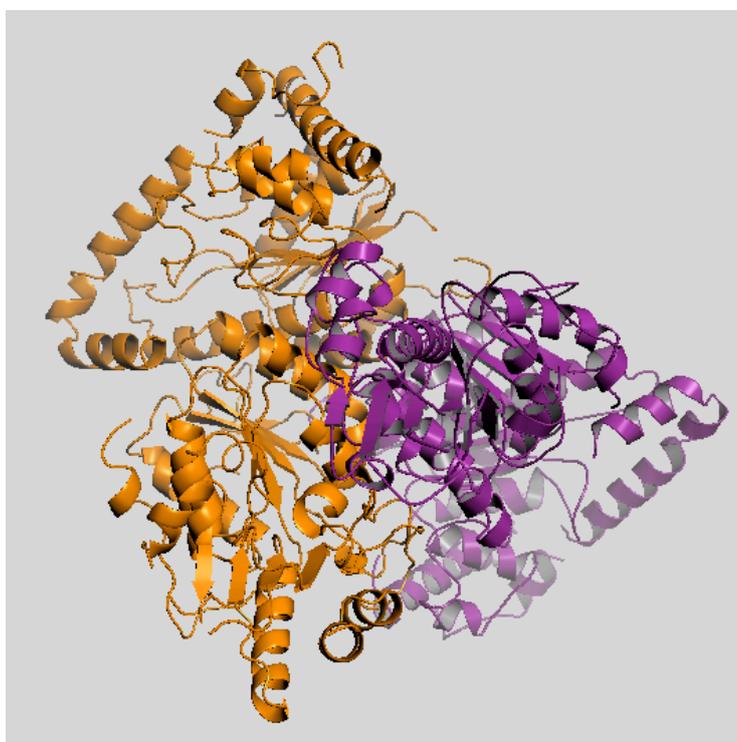


Figure 3.12: Molecular model of TktA
Molecular model of TktA in its biologically functional dimeric form. The two TktA monomers are distinguished from each other using the separate colours purple and orange. The image was generated from the refined TktA structure using PyMol.¹³⁷

The packing order of the TktA molecule in the crystal lattice is shown in figure 3.13

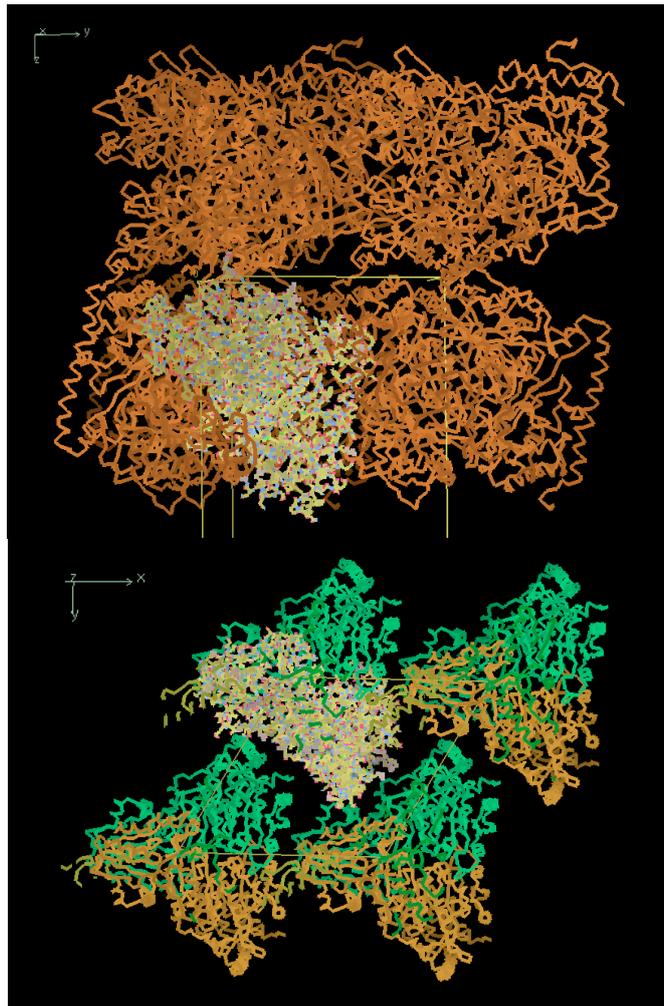


Figure 3.13: Crystal packing in TktA

Top: Crystal packing in TktA viewed along the X-axis. The TktA model is shown bottom-left, coloured by bonds. A two-fold axis of symmetry along the X-axis forms the biologically relevant dimer. The yellow cuboid denotes the unit cell.

Bottom: Crystal packing for TktA viewed along the Z-axis. The TktA model is shown top-left, coloured by bonds. Symmetry related molecules in the crystal lattice are illustrated in matching colours. Images were generated using COOT.¹³⁶

The refined TktA model is illustrated using the “coloured by bonds” representation, showing all side-chains. Symmetry related molecules are shown in orange using the “C-alphas” representation where only the main-chain carbon atoms are used to trace an outline of the model. A two-fold axis of symmetry exists along the X-axis, which allows the formation of the biologically relevant dimer. This symmetry is shown clearly in figure 3.14.

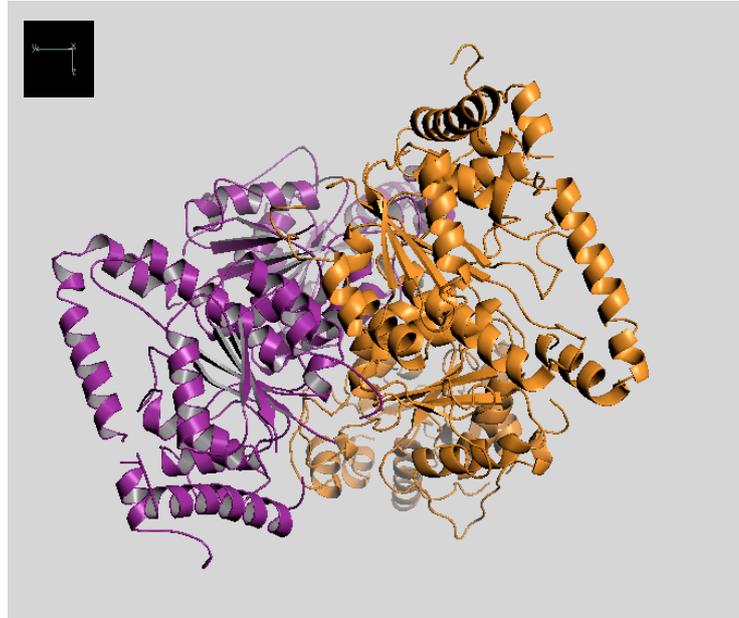


Figure 3.14: Illustration of 2-fold symmetry for TktA

Viewing the model along the X-axis. As a result of the 2-fold axis of symmetry along X, rotating the TktA monomer 180° around the X-axis reveals the dimeric form of TktA molecule responsible for its biological function.

The refined TktA model is represented in purple. The dimer-forming symmetry related molecule is shown in orange. The image was generated from the refined TktA structure using PyMol.¹³⁷

The TktA molecule has been successfully cloned, expressed, purified and crystallised. Crystallisation was performed using the sitting-drop vapour diffusion technique. The structure has been determined using the molecular replacement method to solve the phase problem. Iterative rounds of restrained refinement have successfully improved the quality of the model as seen by acceptable R_{work} and R_{free} values for data of this resolution (2.3Å). Final refinement is ongoing in an attempt to bring outlying residues as calculated in the Ramachandran plot into acceptable geometric conformations; outliers currently comprise 0.68% of the structure with a final target value of 0.2%, as is completion of modelling solvent components (H₂O molecules and other associated elements such as chloride atoms).

Expression, Purification, Crystallisation And Data Collection Of PpmA From *S. pneumoniae*

3.9 Reference

The data presented in this section (to be published) is a detailed account of that which has been drafted for submission to Acta Crystallographica Section F: Crystallisation Communications, with the proposed running title “Expression, Purification, Crystallisation And Preliminary Diffraction Data For PpmA, A Predicted *Cis-Trans* Isomerase From The Pathogenic Bacteria *Streptococcus pneumoniae*” Horsham, M., Isaacs, N.W., Mitchell, T.J. and Riboldi-Tunncliffe, A.

3.10 Introduction

The protein named Putative Protease Maturation Protein A (PpmA) is one of two pneumococcal proteins that have been shown to have homology to a family of proteins known as the peptidyl-protease isomerases (PPlases).¹⁵⁶ These PPlases are responsible for the *cis-trans* isomerisation of proteins at proline residues and act as an aid to protein folding. However, to date PpmA has not been shown to have any PPlase activity, nor is it known which protein PpmA activates.^{156, 157} It has been shown in a murine model that the absence of PpmA results in decreased pneumococcal virulence and an increase in susceptibility to opsonophagocytosis in PpmA-deficient pneumococci,¹⁵⁷ making it a possible candidate for inclusion in a novel protein-based vaccine. It is hoped that structural determination of PpmA could aid the discovery of PpmA function and therefore allow development of inhibitors or mimics of the PpmA protein.

3.11 Overexpression Of PpmA From *S. pneumoniae*

3.11.1 *PpmA* Clone

The PpmA expression construct was kindly provided by Dr. Alan Riboldi-Tunncliffe. The F/L *PpmA* gene from the *S. pneumoniae* strain TIGR4 was inserted into the expression vector pOPINF¹¹⁹ in-frame with an N-terminal His₆ tag separated by an HRV -C3 protease cleavage site using the In-Fusion¹²⁵ technique as detailed in chapter 2.6.2.4 The resulting construct illustrated in figure 3.15 enabled production of recombinantly expressed PpmA protein which could be purified with the aid of the cleavable His₆ tag.

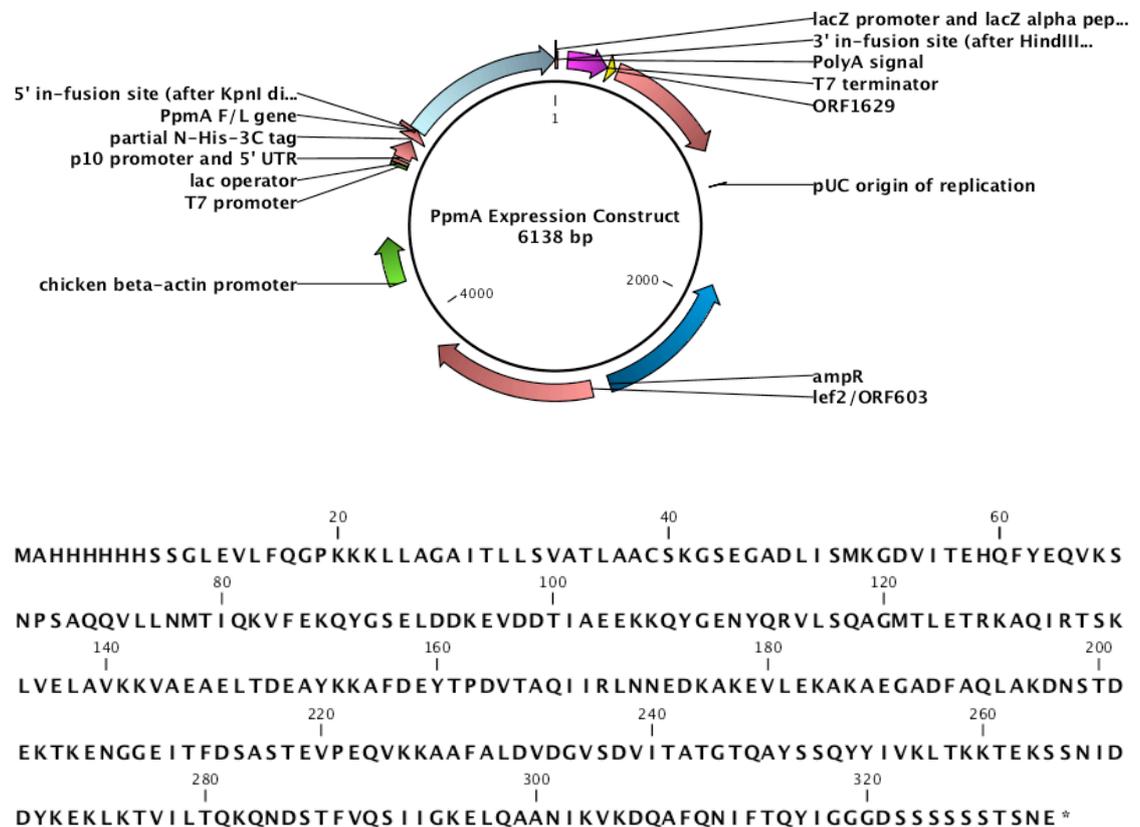


Figure 3.15: PpmA expression construct and protein translation

DNA was fused N-terminally to a His₆-tag that incorporates an HRV-C3 protease site for cleavage of the His₆-tag. Regulation of recombinant protein expression was placed under control of the upstream T7 promoter. The translated protein sequence is also illustrated. The plasmid map was generated using CLC Genomics Workbench 4.¹²⁰

3.11.2 Transformation And Expression Of PpmA

The PpmA expression construct was transformed into the chemically competent *E. coli* expression cell-line Rosetta 2(DE3)pLysS as outlined in chapter 2.7.1 ~400ng of PpmA expression construct were incubated with 50 μ l competent cells for 30mins on ice. Plasmid uptake was induced by heat-shocking the culture at 42°C for 30secs. The culture was allowed to recover on ice for 2mins before being diluted 10-fold with sterile GS96 growth media to a final volume of 500 μ l, and was then incubated at 37°C for 1hr. Post incubation, the culture was plated out onto 1ml LB agar supplemented with 54 μ g/ml carbenicillin and 34 μ g/ml chloramphenicol at dilutions of 5, 10, 15, and 20 μ l, allowed to dry then inverted and incubated at 37°C O/N.

Essentially, protein expression was performed as outlined in chapter 2.7.2. A single colony from the resulting transformed bacteria was used to inoculate a small-volume starter culture of 50ml LB supplemented with the above antibiotics and incubated at O/N at 37°C, 170rpm in a shaking incubator. The following day, 2x500ml large-volume cultures of sterile TB-Onyx auto-induction expression media supplemented with 54 μ g/ml carbenicillin and 3 μ g/ml chloramphenicol were inoculated with 5ml each of the O/N starter culture and incubated at 37°C, 170rpm in a shaking incubator until an O.D₆₀₀ of 0.6 was achieved; thereafter the temperature was lowered to 25°C and incubation was continued for the remainder of the 24hr period. Cells were then harvested as described in chapter 2.7.2.

3.12 Purification Of Recombinant PpmA

3.12.1 Cell Lysis

PpmA was purified using the general purification protocol previously outlined in chapter 2.8 The cell pellets were split into two and each was resuspended in 50ml lysis buffer (see table 2.10). Each was supplemented with 2 μ l DNase1, a complete EDTA-free protease inhibitor tablet and 25 μ l 0.5M MgSO₄ to aid in breakdown of bacterial DNA and inhibit undesired protease activity. The cells were lysed in a French Press as described in chapter 2.7.3, and the cellular debris cleared from the cell lysate containing the soluble PpmA

protein by centrifugation at 7500 x g for 30mins at 4°C and then centrifugation of the resulting lysate at 40,000 x g for a further 30mins at 4°C before being filtered through a 0.22µm syringe filter to remove remaining particulate matter.

3.12.2 Initial IMAC Ni-Affinity Purification

Cleared cell lysate containing expressed PpmA was loaded onto a 1ml HisTrap HP column (GE Healthcare) and Ni-affinity purification was performed using a linear elution gradient as outlined in chapter 2.8.2 1ml fractions were collected and those corresponding to the A₂₈₀ peak and presumed to contain the recombinantly expressed PpmA were analysed by SDS-PAGE as outlined in chapter 2.10.1 A typical trace for Ni-Affinity purification of PpmA is detailed in figure 3.16 and typical SDS-PAGE of analysed fractions is shown in figure 3.17.

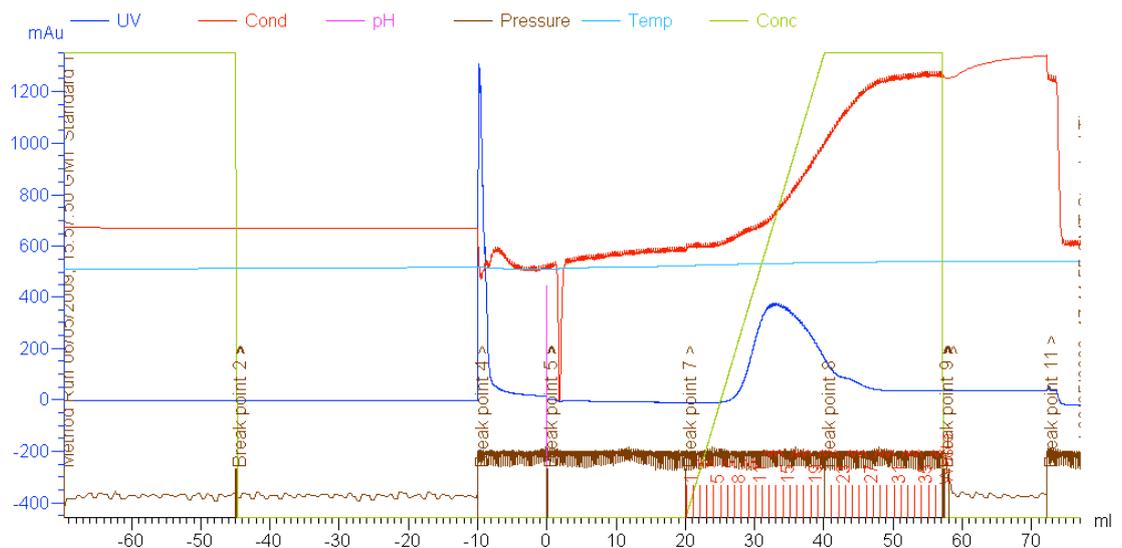


Figure 3.16: Typical Ni-affinity trace for purification of PpmA

Purification was carried out on an Äkta Prime Plus purification system, using a 1ml HisTrap HP column. Protein was eluted from the column at a flow rate of 2ml/min using a linear gradient elution starting at 0% and ending at 100% of nickel buffer B. (Refer to table 2.10). Fractions corresponding to the peak in the UV trace were collected for analysis by SDS-PAGE. Fractions corresponding to the peak in the UV trace were collected and analysed by SDS-PAGE.

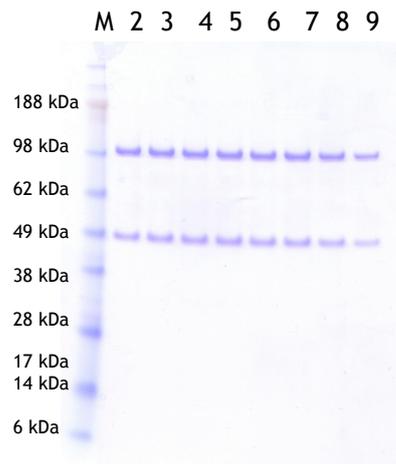


Figure 3.17: Typical SDS-PAGE of PpmA after Ni-affinity purification

Lane 1: SeeBlue Plus 2 mol wt. marker, Lanes 2 - 6: elution fractions

PpmA is known to exist as a dimer when in an oxidised state. The lower band shows the monomer at ~38 kDa (predicted 34 kDa). The Larger band shows the dimer at ~62 kDa (predicted 68 kDa- predictions based on theoretical MW from ProtParam¹¹⁴)

The fractions containing purified PpmA were buffer exchanged back into Ni-buffer A (see table 2.10) as described in chapter 2.9.2 The His₆ tag was then cleaved from the purified PpmA with C3-protease O/N as described in 2.9.1 The Cleaved PpmA was subjected to additional Ni-affinity purification by means of reverse Ni-affinity chromatography. The sample containing the cleaved PpmA, C3-protease and cleaved His₆ tags was cycled through a 1ml HisTrap HP column at a rate of 2ml/min on ice using a peristaltic pump. The cleaved PpmA passed through the column, whilst the His₆ tags and any other remaining bacterial proteins including the non-cleavable His-tagged C3-protease, were bound by the Ni-resin, separating them from the desired PpmA. The his₆ tags and remaining bound proteins were eluted from the column with 10ml Ni-buffer B and discarded. Purified, cleaved PpmA was buffer exchanged into anion exchange buffer-A (see table 2.11) as outlined in 2.9.2 in preparation for anion exchange chromatography.

3.12.3 Ion Exchange Chromatography Of PpmA

The cleaved, Ni-affinity purified PpmA was subjected to further purification by ion-exchange chromatography. In order to determine the type of ion exchanger suitable for PpmA, the PpmA amino acid sequence was

downloaded from the CMR database¹⁴⁹ and input into the ExPASy proteomics server¹¹⁴ to calculate the isoelectric point (P_I) of PpmA. ExPASy¹¹⁴ calculated that the P_I of PpmA was 5.04 and therefore an anion exchanger was chosen for ion exchange purification of PpmA. Anion-exchange chromatography was carried out on a 1ml CaptoQ anion exchange column (GE Healthcare) using a linear gradient elution as outlined in chapter 2.8.3 A typical trace for anion-exchange chromatography of PpmA is shown in figure 3.18 1ml fractions were collected and those corresponding to the A_{280} peak were analysed by SDS-PAGE as outlined in 2.10.1 Typical SDS-PAGE analysis of anion-exchange purified PpmA is illustrated in figure 3.19.

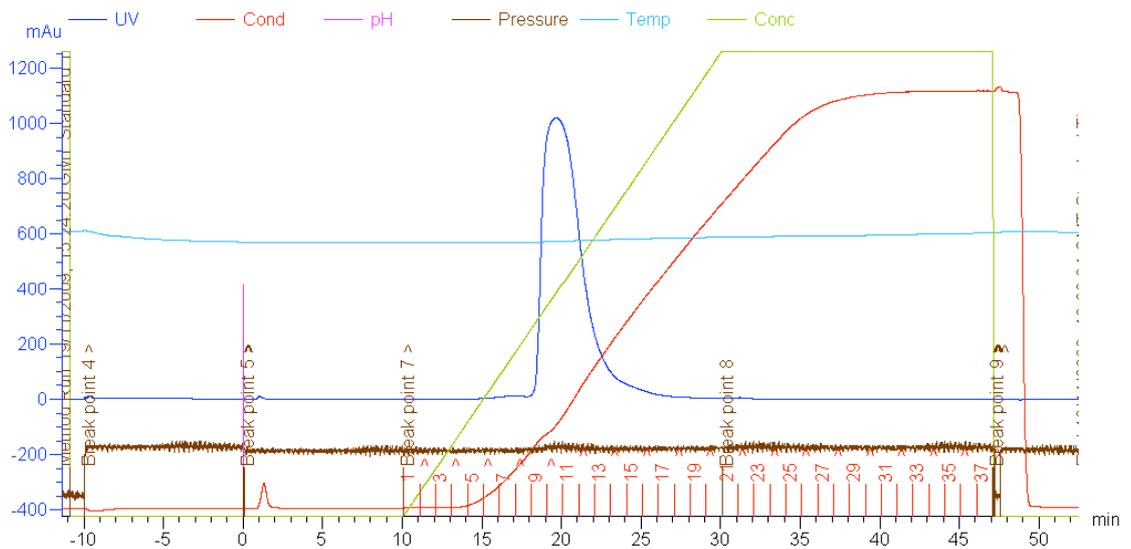


Figure 3.18: Typical anion exchange trace for PpmA

Purification was carried out on an Äkta Prime Plus purification system, using a 1ml CaptoQ column. Protein was eluted from the column at a flow rate of 1ml/min using a linear gradient elution starting at 0% and ending at 100% of anion buffer B. (Refer to table 2.11).

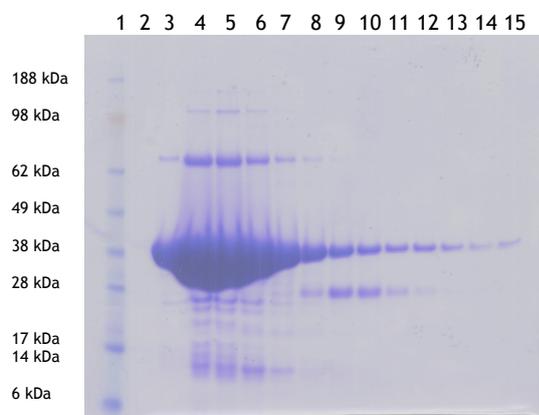


Figure 3.19: Typical SDS-PAGE of PpmA post-anion exchange

Lane 1: SeeBlue Plus 2 mol wt. marker, Lanes 2 -15: elution fractions

3.12.4 Concentration And Storage Of Purified PpmA

Purified PpmA was concentrated using 10 kDa MWCO centrifugal concentrators (Millipore) to a concentration greater than 10mg/ml, aliquotted and stored until required as previously described in 2.9.3

3.13 Crystallisation And Data Collection Of PpmA

Due to the high purity of the PpmA protein even after the initial Ni-affinity purification step, crystal trials were carried out using several purification states of PpmA protein. Four states of PpmA protein were produced and are outlined in table 3.4

Protein Form	Ni-Affinity purified	C3-cleaved (His ₆ tag removed)	AEC Purified
I	✓	X	X
II	✓	✓	X
III	✓	X	✓
IV	✓	✓	✓

Table 3.4: Various forms of PpmA protein that were subjected to crystallisation trials

Forms I and III incorporate the His₆ tag used for purification; forms III and IV have been subjected to additional purification by anion exchange chromatography. These forms explore a wide range of variables that may either help or hinder the crystallisation process. The effects of these differences were fully explored experimentally.

(i) Ni-affinity purified PpmA + His₆ tag; (ii) PpmA - His₆ tag, (iii) PpmA + His₆ tag, LPS removed, (iv) PpmA - His₆ tag, LPS removed. The LPS (lipopolysaccharide) removal is a by-product of the anion exchange chromatography purification process; since LPS has an overall negative charge, the LPS will also bind the anion exchange column, but binds very tightly being left behind when the protein is eluted from the column. An artefact arising from use of bacterial expression systems, most recombinant proteins expressed in this way are coated in LPS, which may be undesirable.

PpmA protein was concentrated to 30mg/ml and subjected to crystal trials at this concentration and also at a diluted concentration of 16mg/ml using the sitting-drop vapour diffusion technique, by mixing protein and reservoir solution 1:1 (500 μ l:500 μ l) to form the drop using a Honeybee 8+1 robot (Digilab, Cambridgeshire UK). Crystal trials included a number of commercially available screens (see table 2.13) including CryoScreens 1&2, PEG/Ion screens 1&2 and Structure Screens 1&2. First crystals grew over a period of 48hrs at 20°C with crystals observed in many of the conditions for PEG/Ion screen 2 (Hampton Research) after 2 weeks. The higher concentration of PpmA produced showers of micro-crystals that formed a large crystalline mass. The lower protein concentration produced thin plate-like crystals of varying size as seen in figure 3.20 Interestingly, PpmA protein crystals were only successfully grown after the bare minimum of purification, as the crystals grew from the PpmA form (i) or (ii). Conversely, no crystals were obtained at either of the two concentrations when using PpmA protein forms (iii) or (iv) where further purification had been performed, including the depletion of LPS levels.

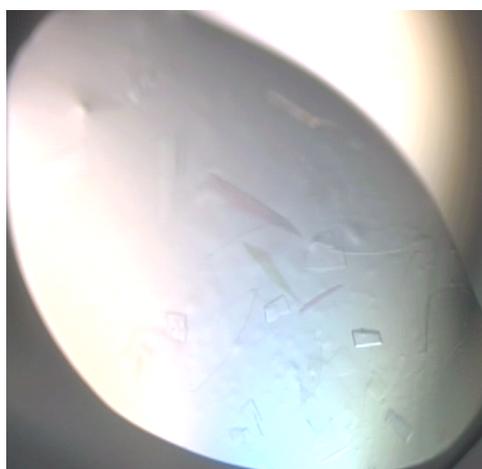


Figure 3.20: PpmA crystals from PEG/Ion 2

Thin plates of PpmA Form I. Protein concentration = 16mg/ml. Total droplet volume = 1 μ l comprising a 1:1 ratio of protein : reservoir solution.

3.13.1 *In-House Testing Of PpmA Crystals*

Testing of the PpmA crystals was initially performed in-house on a Rigaku Micromax 007 HF x-ray generator (Rigaku). Data was collected on a Mar DTB detector. Crystals were tested with a variety of cryoprotectants the best

being 20% w/v PEG 400. The crystals were flash cooled in a stream of nitrogen gas maintained at 110K (-163°C) and subjected to 2x30sec exposures separated by 90° to test diffraction quality. Following optimisation the best crystal was obtained from a reservoir solution comprising 31% (w/v) PEG 4000, 0.18M ammonium acetate and 0.1M sodium acetate and showed diffraction data to beyond 2.5Å. Replicates of this condition were set-up in 96-well sitting-drop plates, and also set-up in larger volume 24 well sitting-drop plates using a total drop size of 2µl in an attempt to obtain larger crystals. Resulting crystals were cryoprotected with 20% (w/v) PEG 400 and flash cooled as before.

3.13.2 Data Collection And Processing of PpmA

PpmA crystals were exposed to synchrotron X-rays at the ESRF (Grenoble, France) on beamline ID14-1. Diffraction data were collected on a single crystal using an ADSC Q210 CCD detector with a crystal rotation of 1° per frame. Diffraction quality was good in one plane, giving strong, clear spots to beyond 2.5Å resolution as shown in figure 3.21, however diffraction data quality decreased upon 90° rotation of the crystal with spots becoming streaky rather than the defined points observed before.

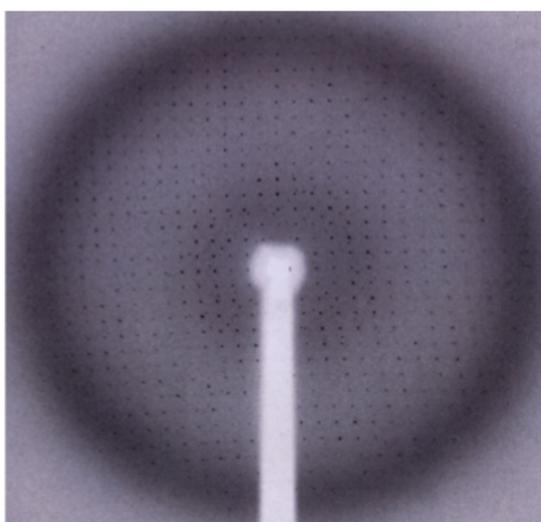


Figure 3.21: In-House X-ray diffraction image for a single crystal of PpmA

Diffraction data acquired at 0° rotation for a single PpmA crystal. Clean X-ray diffraction is evident to beyond 2.5Å resolution. Diffraction quality decreases at 90° rotation of the crystal.

Despite this it was possible to process the diffraction data obtained to gain an indication of space group and the volume of the unit cell for PpmA. Data were processed using the crystallographic program MOSFILM.¹³⁴ As the diffraction pattern was very messy, spots were hand-picked over 10 images. Preliminary analysis of the diffraction data for PpmA is given in table 3.5

Resolution (Å)	2.5Å
Space group	$P2_12_12_1$
Unit cell parameters (Å, °)	$a=65.0, b=105.0,$ $c=120.0,$ $\alpha=\beta=\gamma=90.0$

Table 3.5: Preliminary data statistics for a single PpmA crystal.

To provide additional information, the program *MATTHEWS*¹³² was used to estimate the solvent content of the unit cell for PpmA. This data is given below in table 3.6

Nmol/Asym	Matthews Coeff (Å ³ Da ⁻¹)	% Solvent	Probability
1	5.95	79.33	0.00
2	2.97	58.65	0.48
3	1.98	37.98	0.51
4	1.49	17.31	0.00

Table 3.6: Solvent content for single crystal of PpmA

Predicted solvent content of the PpmA crystal was calculated using the crystallographic program MATTHEWS from the CCP4 suite of crystallographic programs.¹³² Values were calculated using the predicted MWT of the PpmA molecule and unit cell dimensions.

MATTHEWS predicts either 2 or 3 molecules of PpmA per asymmetric unit; previous work in our lab using gel-filtration studies has shown that PpmA exists as a dimeric molecule under non-reduced conditions.

Although PpmA has been successfully crystallised and rudimentary diffraction data obtained, there is insufficient data at present to determine the structure of PpmA. Work is ongoing at present in an attempt to grow crystals that yield better quality X-ray diffraction data. The decrease in X-ray diffraction quality may be due to poor crystal mosaicity caused either by

unsuitable cryo-conditions or possibly crystal bending. It is thought that the latter situation is more likely as the crystals appear as very thin plates and could therefore be susceptible to bending upon manipulation, affecting the mosaicity of the crystal lattice; diffraction quality is good along one plane and therefore cryo-conditions seem acceptable. In addition, since PpmA is predicted to be a novel type of *cis-trans* isomerase, conventional molecular replacement methods will be unsuitable for the determination of the structure of PpmA, with heavy atom soaks or SeMet protein derivatives potentially having to be generated in order to overcome the phase-problem¹⁵³ and successfully determine the structure of PpmA.

Chapter 4 : Cloning, Expression And Purification Of Pneumococcal Histidine Triad D (PhtD) From *Streptococcus pneumoniae*

4.1 Introduction

Initially, purified PhtD protein was supplied by GlaxoSmithKline (GSK-Rixensart, Belgium). To avoid potential problems with supply of PhtD protein, and to clone alternative fragments of the protein the decision was taken to clone the *PhtD* gene encoding the mature protein into an expression vector capable of expressing PhtD recombinantly in a bacterial expression system from which it could then be isolated and purified. This chapter details the production process, from initial isolation of genomic DNA from *S.pneumoniae*, through creation of the expression vector to expression and purification of the recombinantly expressed PhtD.

4.2 Human Rhinovirus-C3 (HRV-C3) Protease

4.2.1 General Notes

All expression constructs used in this thesis were generated using the pOPINF expression vector¹¹⁹ which fuses an HRV-C3 protease cleavable His₆-tag to the N-terminus of the recombinantly expressed protein of interest (POI). The His₆-tag has been engineered specifically to aid during the protein purification process, and must be removed to achieve as close to the desired final protein sequence as possible. It was therefore required that a stock of HRV-C3 protease be generated for this purpose. Plasmid DNA expressing HRV-C3 protease was kindly supplied by Dr. Alan Riboldi-Tunncliffe, and has been engineered to include a non-cleavable His₆-tag for the purpose of removing the C3 protease from the sample post-cleavage.

HRV-C3 protease is a cysteine protease and is able to cleave proteins between the Q and G of the eight amino acid recognition sequence (LEVL[↓]FQGP). A single glycine and proline residue remain at the N-terminus

of the recombinantly expressed protein as an artefact of the cleavage process. HRV-C3 protease is required to be maintained in a reduced environment in order to preserve its activity. This section describes the production of a stock of C3 protease for use in cleavage of His₆ tags from recombinantly expressed proteins as an aid to protein purification.

4.2.2 Bacterial Transformation

A 50 μ l aliquot of Rosetta 2(DE3)pLysS chemically competent *E. coli* cells were incubated on ice with ~400ng of plasmid DNA encoding HRV-C3 protease for 30mins. Cells were then heat-shocked in a water bath at 42°C for 30secs and allowed to recover on ice for 2mins. Cells were diluted 1:50 to a final volume of 500 μ l with GS96 growth media and incubated at 37°C for 1hr. Serial dilutions of this culture were plated out onto LB agar supplemented with 50 μ g/ml ampicillin and 34 μ g/ml chloramphenicol and allowed to dry. Plates were inverted and incubated at 37°C O/N.

4.2.3 Preparation Of Starter Culture

A single colony of transformed Rosetta cells was picked into 50ml of LB supplemented with 50 μ g/ml ampicillin and 34 μ g/ml chloramphenicol. This small volume culture was incubated at 37°C, 170rpm in a shaking incubator and allowed to grow O/N.

4.2.4 Large Scale Expression Of HRV-C3 Protease

4l of LB was prepared as previously described in 4x 2l conical flasks, and supplemented with 54 μ g/ml carbenicillin and 34 μ g/ml chloramphenicol immediately before culture growth. 10ml of O/N starter culture was used to inoculate each 1l of prepared LB. Cultures were grown at 37°C, 170rpm in a shaking incubator until O.D₆₀₀ = 0.6 was achieved. Protein production was then induced by addition of 1mM IPTG to each flask. The temperature was lowered to 22°C and allowed to grow for the remainder of the 24hr period post-inoculation. Cells were harvested as described in 2.7.2.

4.2.5 Purification of HRV-C3 Protease

N.B. Due to the physiological nature of the C3 protease, all buffers used in the purification process were supplemented with 10mM β -mercaptoethanol to preserve the reduced state of the protein. Harvested cells were lysed as described in 2.7.3 Lysed cells were centrifuged at 6.5k x g at 277K for 30mins in a Sigma 4K15 tabletop centrifuge. The supernatant was then centrifuged for a further 30mins at 40 000 x g, 4°C in a Beckman Allegra™ 64R tabletop centrifuge (Beckman-Coulter, High Wycombe UK) to pellet any remaining cell debris. The resulting cleared lysate was syringe filtered through a 0.22 μ m syringe filter before being subjected to Ni-affinity chromatography as described in 2.8.2 10mM Dithiothreitol (DTT) was added to fraction collection tubes keep the protein in a reduced state after elution from the Ni-affinity column (DTT was used as a substitute reducing agent at this stage due to β -mercaptoethanol toxicity).

4.2.6 Size Exclusion Chromatography

Size exclusion chromatography or gel filtration was performed as a final purification step, as described in 2.8.4 The final product, purified C3-protease is illustrated in figure 4.1.

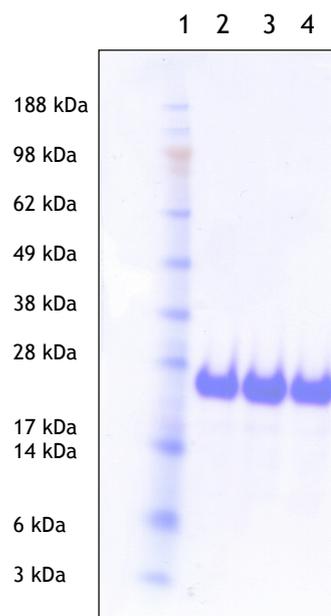


Figure 4.1: SDS-PAGE of purified C3-protease

Lane 1: SeeBlue Plus 2 mol wt. marker, Lanes 2 - 4: final, purified C3-protease

4.2.7 HRV-C3 Protease Storage

The purified C3-protease was concentrated to 1ml volume, and diluted 1:1 with sterile 100% glycerol, giving a final concentration of ~1.5mg/ml. The purified protein was divided into 100µl aliquots and stored at -80°C for long-term storage. A single aliquot was stored at -20°C for use and replaced from the long-term stock when required.

4.3 Construction Of PhtD Expression Plasmid

4.3.1 Production Of *S. pneumoniae* gDNA

Genomic DNA was prepared and extracted from the *S. pneumoniae* TIGR4 strain as described in 2.5.1.

4.3.2 PhtD Nucleotide Sequence

SP_1003, the nucleotide sequence of the gene encoding PhtD was extracted from the Central Microbial Resources (CMR) database (<http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi>)¹⁴⁹ hosted by the J. Craig Venter Institute (JCVI - <http://www.jcvi.org/>). The sequence was trimmed of the first 60 nucleotides -which encode the signal sequence of the protein- in order to replicate the composition of the PhtD previously supplied by GSK. Hereafter this protein shall be referred to as F/L PhtD, as it is essentially the mature protein. An annotated schematic of the PhtD protein sequence is illustrated in figure 4.2.

MKINKKYL	AG	SVAVLALSVC	SYELGRHQAG	QVKKESNRVS	YIDGDQAGQK	AENLTPDEVS
KREGINAEQI	VIKITDQGYV	TS	HGDHYHY	NGKVPYDAII	SEELLMKDPN	YQLKDSDIVN
EIKGGYVIKV	DGKYVYVLKD	AAHADNIRTK	EEIKRQKQEH	SHNHGGGSND	QAVVAARAQG	
RYTTDDGYIF	NASDIIEDTG	DAYIVP	HGDH	YHYIPKNELS	ASELAAAEAY	WNGKQGSRPS
SSSSYNANPA	QPRLSEHNHL	TVTPTYHQNQ	GENISSLLRE	LYAKPLSERH	VESDGLIFDP	
AQITSRTARG	VAVP	HGNHYH	FIPYEQMSEL	EKRIARI IPL	RYRSNHVVPD	SRPEQPSPQS
TPEPSPSPQP	APNPQPAPSN	PIDEKLVKEA	VRKVG DGYVF	EENGVSRYIP	AKDLSAETAA	
GIDSKLAKQE	SLSHKLGAKK	TDL PSSDREF	YNKAYDLLAR	IHQDLLDNKG	RQVDFEALDN	
LLERLKD VPS	DKVKLVDDIL	AFLAPIRHPE	RLGKPN AQIT	YTDDEIQVAK	LAGKYTTEDG	
YIFDPRDITS	DEGDAYVTPH	MTHSHWIKKD	SLSEAERAAA	QAYAKEKGLT	PPSTDHQDSG	
NTEAKGAEAI	YNRVKA AKKV	PLDRMPYNLQ	YTVEVKNGSL	IIPHYDHYHN	IKFEWFDEGL	
YEAPKGYTLE	DLLATVKY YV	EHPNERPHSD	NGFGNASDHV	RKNKVDQDSK	PDEDKEHDEV	
SEPTHPESDE	KENHAGLNPS	ADNLYKPSTD	TEETEEEAED	TTDEAEIPQV	ENSVINAKIA	
DAEALLEKVT	DPSIRQNAME	TLTGLKSSLL	LGTKDNNTIS	AEVDSLLALL	KESQPAPIQ	

Figure 4.2: PhtD protein sequence (downloaded from CMR database)

Annotated PhtD sequence. Residues highlighted in yellow indicate signal sequence to target the protein to the cell surface. Residues highlighted in pink indicate the histidine triad motifs.

4.3.3 Primer Design And Amplification Of PhtD Gene

Primers were ordered from Sigma-Aldrich. Due to the precise nature of the product required, there was no choice as to position of the primer-binding region; in order to gain favourable primer characteristics (~60% G+C content, melting temperature >68°C) the gene specific portion of the primer was lengthened or shortened accordingly. In order to make the primers compatible with the In-Fusion[®] cloning system,¹²⁵ a vector-specific region was included in the primer corresponding to the end regions of the pOPINF expression plasmid. Proposed primers were entered into the Sigma-Aldrich oligonucleotides Design Tool¹⁵⁸ in order to determine melting temperature, 2° structure and primer-dimer prediction

(http://www.sigmaaldrich.com/configurator/servlet/DesignTool?prod_type=S_TANDARD). Resulting primers used in amplification of the *PhtD* gene are given in table 4.1.

Primer Name	Sequence	Melting Temperature (°C)
PhtD F/L FWD	<u>AAGTTCTGTTTCAGGGCCCGTCCTA</u> TGAAC TTGGTCGTCACCAA	85.2
PhtD F/L REV	ATGGTCTAGAAAGCTTTACTGTATA <u>GGAGCCGGTTGACTTTC</u>	75.9

Table 4.1: Primers for amplification of F/L *PhtD*.

Underlined region of sequence denotes vector-specific primer region.

PCR was carried out to amplify *PhtD* as outlined in 2.5.4 Table 4.2 gives the specific conditions under which the reaction was carried out. Primers 52Q and 52R (See table 2.6) that are known to amplify a fragment of the gene encoding the protein pneumolysin (Ply)¹²² were used as positive and negative controls. The positive control substituted PhtD primers with 52Q and 52R to verify that the gDNA was viable in the case that amplification of a PhtD product failed. The negative control omitted template gDNA from the reaction-mix in order to show that contamination of the reaction-mix with undesired DNA had not occurred. PCR results were only deemed viable if a product was obtained in the positive control and absent in the negative control reactions.

Reaction Step	Temperature (°C)	Time
Init. Denaturation	98	30 seconds
Melting	98	10 seconds
Annealing	54	30 seconds
Extending	72	90 seconds
Final extension	72	5 minutes

} x35

Table 4.2: PCR conditions for amplification of F/L *PhtD*

The products from the PCR experiment were analysed by agarose gel electrophoresis as previously outlined in 2.6.1 Results are shown in figure 4.3.

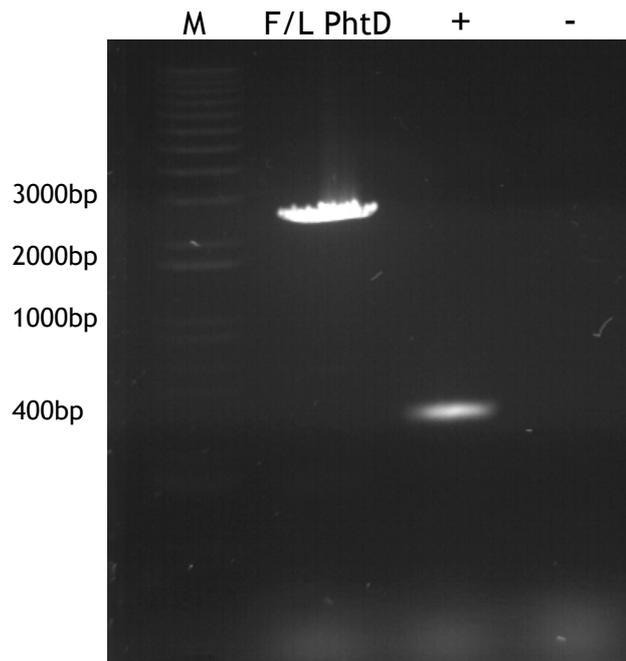


Figure 4.3: 0.8% Agarose gel showing successful amplification of F/L *PhtD* gene

Lane 1: 1kb+ standards, Lane 2: F/L *PhtD* PCR product, Lane 3: Ply frag +ve control, Lane 4: Ply frag -ve control. Amplification of *PhtD* gene was successful.

A clean single band corresponding to the expected product size of 2496bp was obtained. The positive control showed successful amplification of the Ply fragment, while the negative control yielded no product as expected. The successfully amplified *PhtD* was used to generate an expression plasmid for recombinant protein expression.

4.3.4 Generation Of *PhtD* Expression Plasmid

An expression plasmid for *PhtD* was constructed by inserting the amplified *PhtD* into the vector pOPINF¹¹⁹ using the In-Fusion[®] system (Clontech).¹²⁵ Linearised pOPINF¹¹⁹ plasmid at a concentration of 100ng/ μ l was kindly provided by Dr. A. Riboldi-Tunncliffe. The In-Fusion system is a ligation independent cloning system which utilises a viral exonuclease to degrade the ends of the vector and insert and subsequently ‘knits’ them together, hence the need for the long 20bp and 18bp vector/insert overlaps generated during the PCR process. 150ng linear pOPINF¹¹⁹ and 138ng amplified *PhtD* insert were utilised in the In-Fusion reaction, which was carried out as detailed in 2.6.2.4 Transformation of the resulting construct was carried out immediately into

XL-10 GOLD chemically competent cells. 2 μ l of the ligation mix were incubated on ice for 30mins with 20 μ l XL-10 GOLD cells. DNA uptake was induced by heat-shocking at 42°C for 30secs in a water-bath. Cells were allowed to recover on ice for 2mins before being diluted to a final volume of 200 μ l with GS96 growth media and then incubated at 37°C for 1hr. Cells were plated out on LB agar supplemented with 54 μ g/ml carbenicillin, allowed to dry, and incubated O/N at 37°C.

A single colony from the successfully transformed bacteria was used to inoculate 50ml LB supplemented with 54 μ g/ml carbenicillin and incubated in a shaking incubator O/N at 37°C, 170rpm. The following day, plasmid DNA was extracted from this O/N culture using the QIAquick miniprep kit¹²⁷ according to manufacturers instructions, substituting DNase & RNase-free dH₂O for the elution buffer. 2 μ l plasmid DNA was analysed by agarose gel electrophoresis as previously described to confirm successful plasmid generation. Quality control was carried out by having the plasmid sequenced using T7 promoter and terminal primers, the original insertion primers and internal *PhtD* primers in order to obtain full read-through of the gene (refer to table 2.6). Sequencing was carried out at DNA Sequencing & Services (University Of Dundee, Scotland UK).¹⁵⁹ Sequencing results showed no errors in the plasmid sequence. The resulting expression plasmid was named pMRH1 and is shown in figure 4.4 along with the translated protein sequence. For storage, pMRH1 was transformed into DH5 α cells as described above, a glycerol stock prepared as outlined in 2.3.1 and stored at -80°C.

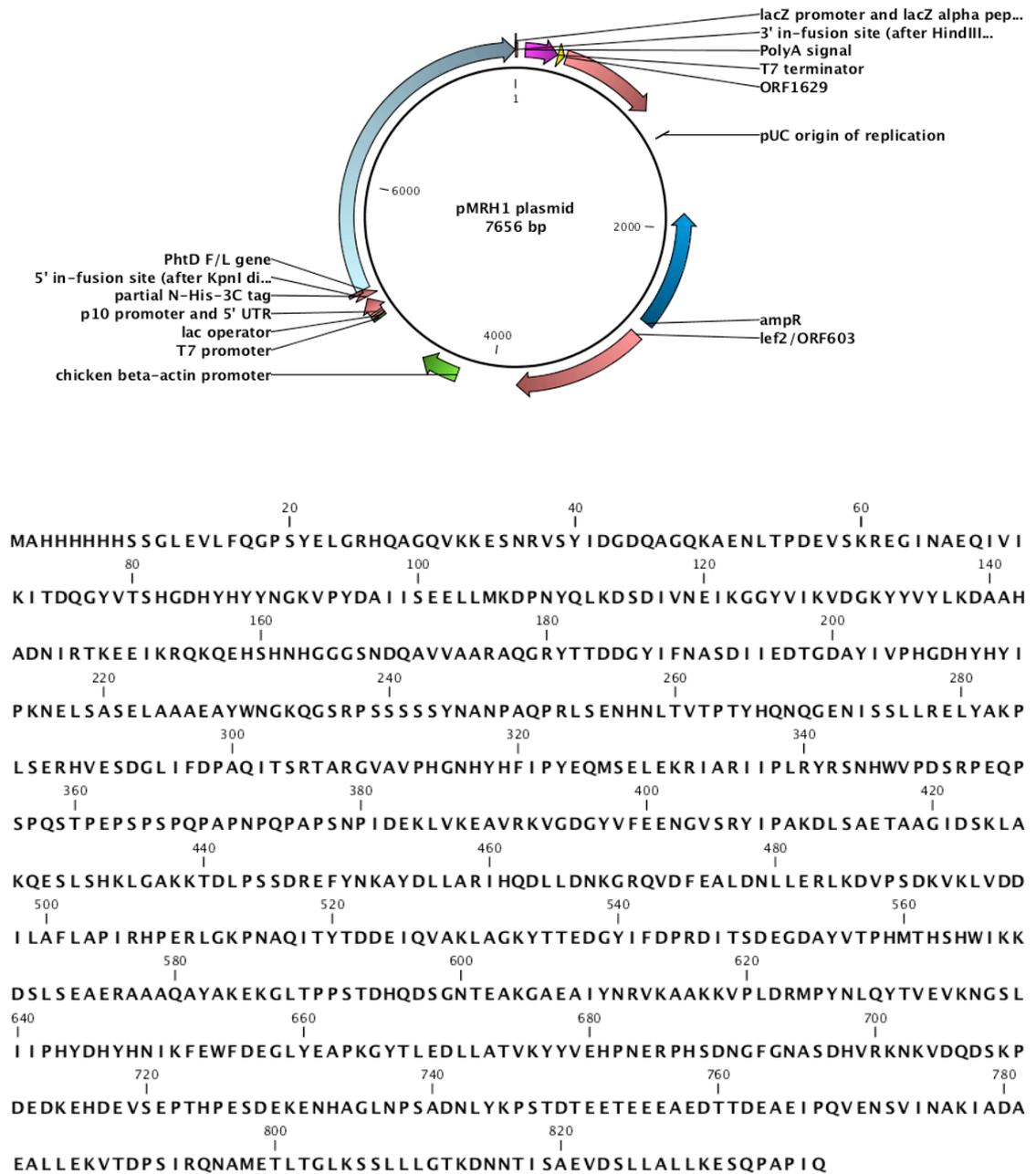


Figure 4.4: pMRH1 F/L PhtD expression construct

DNA was fused N-terminally to a His₆-tag that incorporates an HRV-C3 protease site for cleavage of the His₆-tag. Regulation of recombinant protein expression was placed under control of the upstream T7 promoter. The bottom panel shows the resulting translated protein sequence. Plasmid map was generated using CLC Genomics Workbench 4.¹²⁰

4.4 Expression Of F/L PhtD

4.4.1 Transformation

Essentially, transformation was carried out as previously described. A 50 μ l aliquot of Rosetta™ 2(DE3)pLysS chemically competent cells were incubated on ice with ~400ng of pMRH1 plasmid for 30mins. DNA uptake was induced by heat-shocking at 42°C for 30secs in a water-bath. Cells were allowed to recover on ice for 2mins, diluted to a final volume of 500 μ l with GS96 growth media and incubated at 37°C for 1hr. Cells were plated out on LB agar supplemented with 54 μ g/ml carbenicillin, allowed to dry, and incubated O/N at 37°C.

4.4.2 Protein Expression

A single transformant was used to inoculate 50ml LB supplemented with 54 μ g/ml carbenicillin and 34 μ g/ml chloramphenicol in a 200ml conical flask and incubated at 37°C in a shaking incubator O/N at 170rpm. Large scale protein expression was carried out using 2 x 500ml volumes of TB-Onyx auto-induction media supplemented with 54 μ g/ml carbenicillin and 34 μ g/ml chloramphenicol in 2L conical flasks as described in 2.7.2

4.5 Purification Of Recombinant F/L PhtD

4.5.1 Initial IMAC Ni-Affinity Chromatography

Pelleted cells were lysed in a French pressure cell as outlined in 2.7.3 Lysed cells were centrifuged at 7500 x g, 4°C for 30mins and the pellet discarded. The supernatant was then further centrifuged at 40,000 x g, 4°C for 30mins, the pellet discarded and the supernatant passed through a 0.22 μ m syringe filter to remove as much particulate matter as possible before chromatography. The cleared lysate was subjected to Ni-affinity chromatography as outlined in 2.8.2 Eluted fractions corresponding to the A₂₈₀ peak on the trace were collected and analysed by SDS-PAGE as outlined in 2.10.1 A trace typical of the initial Ni-affinity purification for F/L PhtD is

shown in figure 4.5 Figure 4.6 shows typical SDS-PAGE analysis of fractions eluted from the Ni-affinity column.

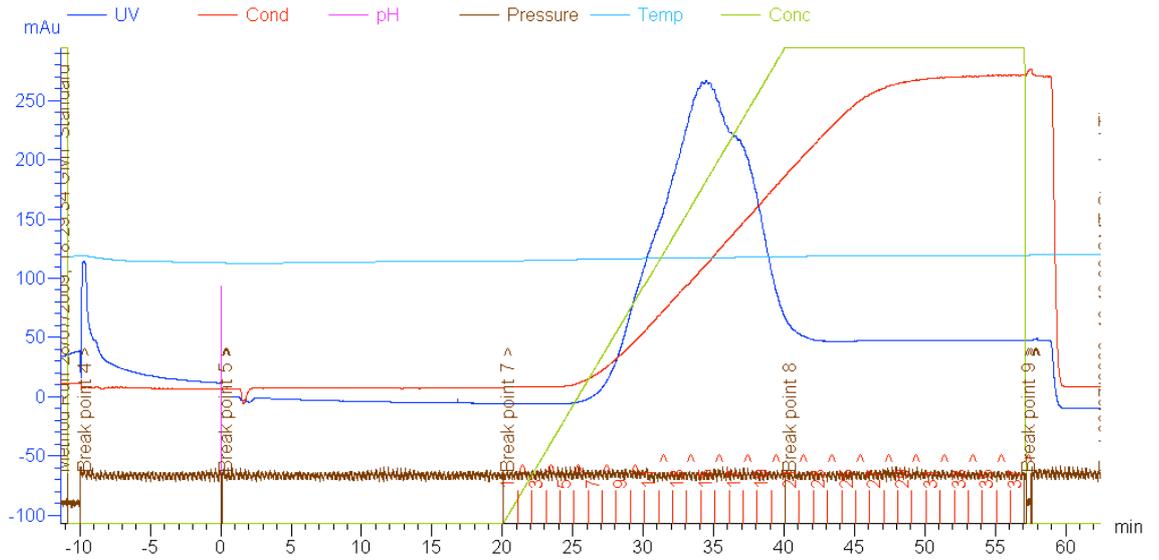


Figure 4.5: Typical Ni-affinity trace for F/L PhtD

Purification was carried out on an Äkta Prime Plus purification system, using a 1ml HisTrap HP column. Protein was eluted from the column at a flow rate of 2ml/min using a linear gradient elution starting at 0% and ending at 100% of nickel buffer B. (Refer to table 2.10). Fractions corresponding to the peak in the UV trace were collected for analysis by SDS-PAGE

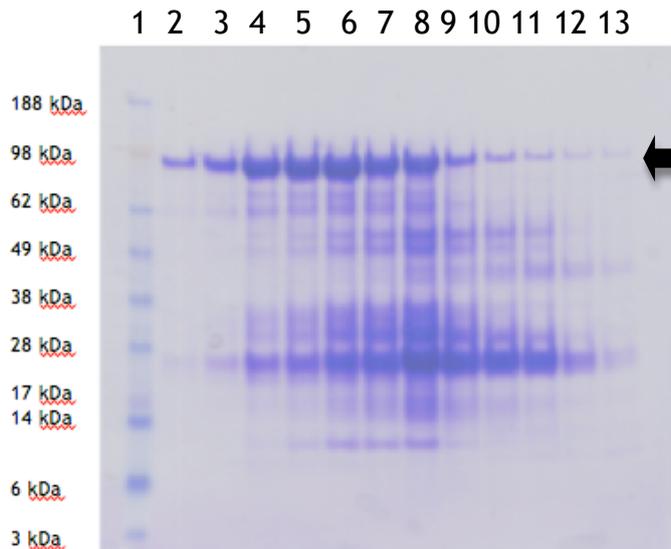


Figure 4.6: SDS-PAGE of F/L PhtD from Ni-affinity purification

Lane1: mol wt. marker; Lanes 2-13: elution fractions. Arrow indicates recombinantly expressed F/L PhtD protein.

4.5.2 C3-Cleavage Of His₆ Tag & Reverse Ni-Affinity Chromatography

Fractions containing the highest abundance of target protein were pooled for continuation of the purification process (e.g. fractions 2-9 in figure 4.6). F/L PhtD was buffer exchanged back into Ni-buffer A using a PD-10 desalting column¹²⁸ as outlined in 2.9.2.

As an added purification step before ion exchange chromatography, the N-terminal His₆ tag was cleaved from F/L PhtD present in the Ni-affinity elution fractions with C3-protease as outlined in 2.9.1. Cleavage was carried out O/N. The cleaved protein was then subjected to reverse Ni-affinity chromatography by cycling the sample through a 1ml HisTrap HP Ni-column using a peristaltic pump at a flow-rate of 2ml/min for 30mins on ice. Because the His₆ tag has been removed from the protein it no longer binds to the resin in the column, instead passing straight through whilst the cleaved His₆ tags and contaminant host-proteins which had previously shown an affinity for the column still bind. The HRV-C3 protease also contains a non-cleavable His-tag, which allows it to be isolated from the cleaved target-protein sample by binding to the Ni-affinity column. The tags, contaminant proteins and C3-protease were eluted from the Ni-column with 10ml of Ni-buffer B and discarded. The cleaved F/L PhtD was carried forward for further purification by ion exchange chromatography.

4.5.3 Ion Exchange Chromatography (Anion Exchange)

Ion exchange chromatography was carried out using a 1ml CaptoQ anion exchange column as previously described in 2.8.3. A typical trace for anion exchange chromatography of F/L PhtD is illustrated in figure 4.7 and a typical SDS-PAGE analysis of the resulting fractions is illustrated in figure 4.8.

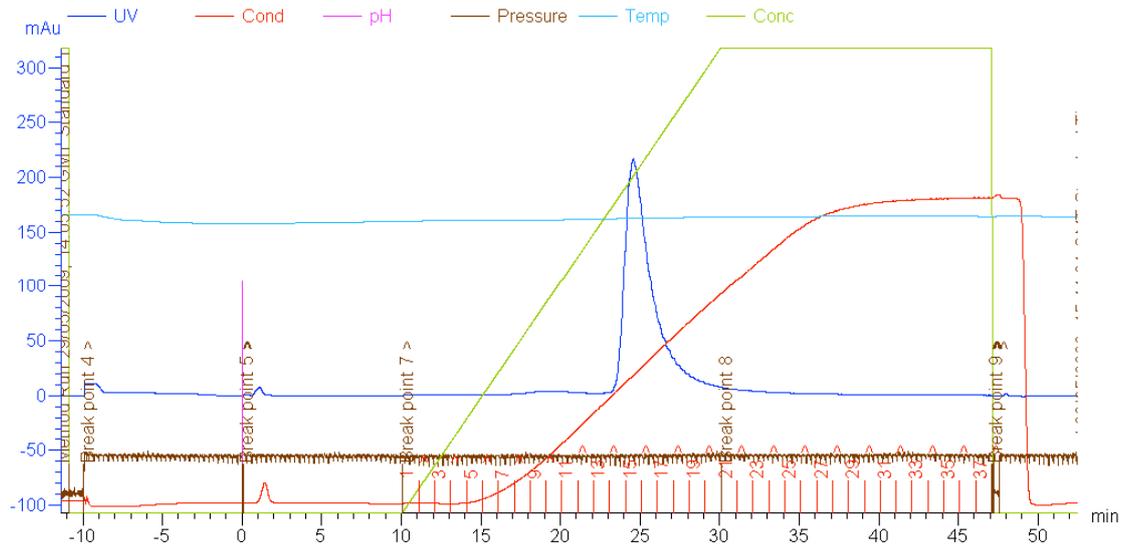


Figure 4.7: Typical anion exchange trace of F/L PhtD

Purification was carried out on an Äkta Prime Plus purification system, using a 1ml CantoQ column. Protein was eluted from the column at a flow rate of 1ml/min using a linear gradient elution starting at 0% and ending at 100% of anion buffer B. (Refer to table 2.11).

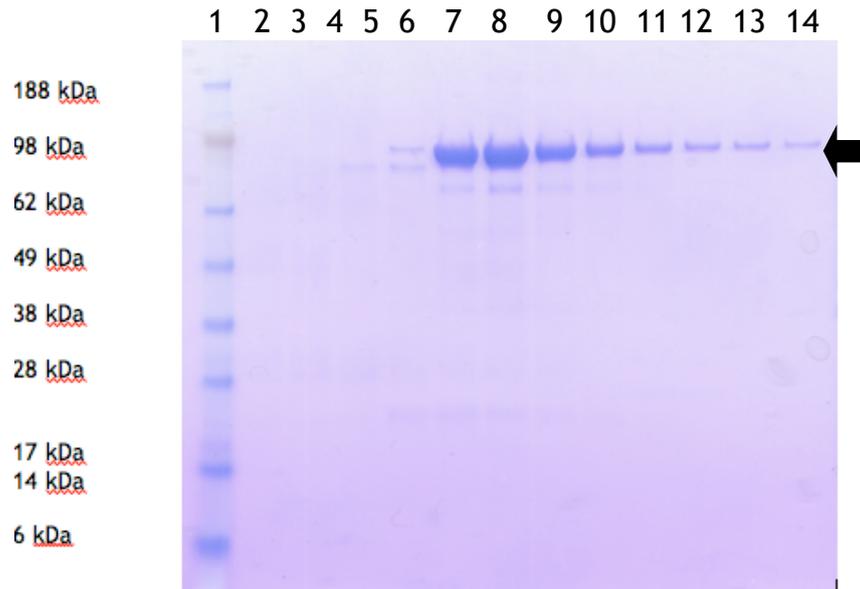


Figure 4.8: SDS-PAGE for anion exchange of F/L PhtD

Lane 1: mol wt. marker; Lanes 2-14: elution fractions. Arrow indicates purified F/L PhtD.

Following purification by anion exchange chromatography and analysis by SDS-PAGE, it was judged that the F/L PhtD was of pure enough quality to be used without any further purification. The purest fractions were pooled, concentrated and stored at -20°C for short-term storage or -80°C for long-term storage until required.

4.5.4 Concentration Of Purified F/L PhtD

The purified F/L PhtD was buffer exchanged from Anion-exchange buffer B into 50mM Tris HCl pH7.5, 500mM NaCl, 20mM imidazole using a PD-10 desalting column¹²⁸ as described in 2.9.2 The purified F/L PhtD was then concentrated as outlined in 2.9.3 until a concentration of ~10mg/ml had been obtained. F/L protein was then aliquotted and stored as previously described in 2.9.3.

4.6 Construction Of Truncated Forms of PhtD

4.6.1 Overview

Structural predictions and protein analysis by 1D-NMR (detailed in chapter 4) indicated that crystallisation and structural determination of the F/L PhtD molecule would prove challenging, and that greater success might be obtained by attempting to crystallise separate fragments of the PhtD molecule. It was therefore decided to use a targeted approach to identify structured regions of the F/L protein and to construct and purify these regions as truncated forms or fragments of the F/L PhtD molecule. To this end, regions of interest were identified by limited proteolysis, mass-spectrometry and N-terminal sequencing (details of these experiments and their findings are given in the relevant chapters) and a set of truncated proteins constructed. The following sections outline the generation of these fragments.

4.6.2 Fragment Design

In total, 7 constructs were planned of which 1 was the F/L protein described previously and another was to attempt insertion of a TEV cleavage site into the middle of the F/L PhtD construct, although this was ultimately abandoned. The other 5 constructs were all fragments of the N-terminal half of the protein as this was seen to be the more highly structured region of the protein. These expression constructs were named pMRH1 - 7 as detailed in table 4.3.

Plasmid Name	N- And C- Term Translated Amino Acids	Protein Name
pMRH1	SYEL - APIQ	F/L
pMRH2	SNHW - APIQ	C-Term
pMRH3	SNHW - LLEK	Δ C-Term
pMRH4	VTDP - APIQ	5.3 kDa
pMRH6	SNHW - HYHN	34 kDa
pMRH7	IKFE - LLEK	Epitope

Table 4.3: PhtD plasmids and resulting protein products

Start and end sequences of translated amino acids for the pMRH constructs are given above. Full protein sequences are given with relevant vector maps.

***pMRH5: F/L with TEV insert excluded*

A schematic diagram for ease of visualising the different fragments is shown below in figure 4.9

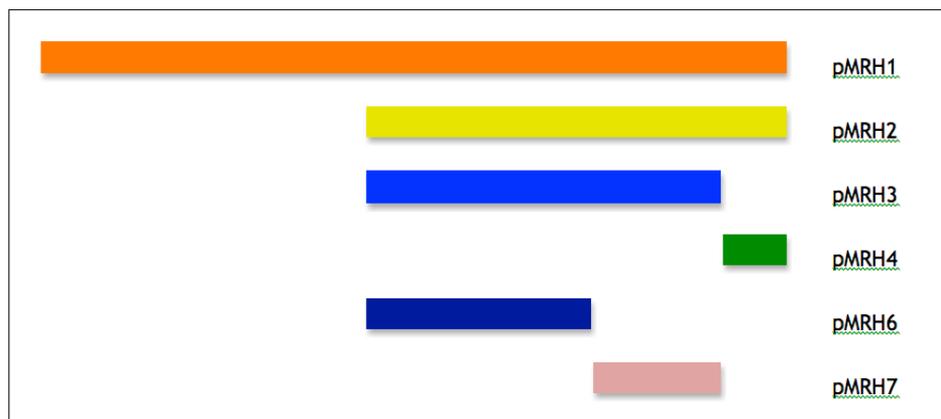


Figure 4.9: Schematic representation of PhtD proteins

4.6.3 Primer Design And Amplification Of PhtD Fragments

As before, primer design was limited because of the exact location that was required to generate each fragment. The gene-specific region of each primer was coupled with the pOPINF¹¹⁹-specific region for In-Fusion cloning.¹²⁵ Primers are given in table 4.4.

Primer Name	Primer Sequence	Comments
PhtD F/L Fwd	<u>AAGTTCTGTTTCAGGGCCCGTCCTA</u> TGAACCTGGTCGTCACCAA	Fwd primer for pMRH1
PhtD F/L Rev	ATGGTCTAGAAAGCTTTACTGTATA GGAGCCGTTGACTTTC	Rev primer for pMRH1, pMRH2 & pMRH4
PhtD C-Term Fwd	<u>AAGTTCTGTTTCAGGGCCCGCCCAG</u> CGCAAATCACAAGTCGAAC	Fwd primer for pMRH2, pMRH3 & pMRH6
PhtD ΔC-Term Rev	ATGGTCTAGAAAGCTTTATTTTTCTA GCAAGGCCTCCGCATC	Rev primer for pMRH3 & pMRH7
PhtD 5.3kDa Fwd	<u>AAGTTCTGTTTCAGGGCCCGTAACA</u> GATCCTAGTATTAGAC	Fwd primer for pMRH4
PhtD 34kDa Rev	ATGGTCTAGAAAGCTTTAGTTATGGT AATGGTCATAATGAGG	Rev primer for pMRH6
PhtD Epitope Fwd	<u>AAGTTCTGTTTCAGGGCCCGATCAAA</u> TTTGAGTGGTTTGAC	Fwd primer for pMRH7

Table 4.4: Primers used for amplification of *PhtD* fragments.

Underlined sequence denotes vector-specific region.

4.6.4 Amplification Of *PhtD* Fragments

PCR was carried out to amplify the *PhtD* fragments as outlined in 2.6.4. Table 4.5 gives the specific conditions under which the reaction was carried out. Primers 52Q and 52R (See table 2.6) for amplifying a fragment of the gene encoding the protein pneumolysin (Ply)¹²² were used as positive and negative controls as previously described in section 3.3.3

Reaction Step	Temperature	Time
Init. Denaturation	98°C	30secs
Melting	98°C	10secs
Annealing	60°C	30secs
Extending	72°C	50secs
Final extension	72°C	5mins

} x35

Table 4.5: PCR conditions for *PhtD* fragments

Products of the PCR experiment were analysed by agarose gel electrophoresis as outlined in 2.6.1 Results for PCR of fragments are shown below in figure 4.10

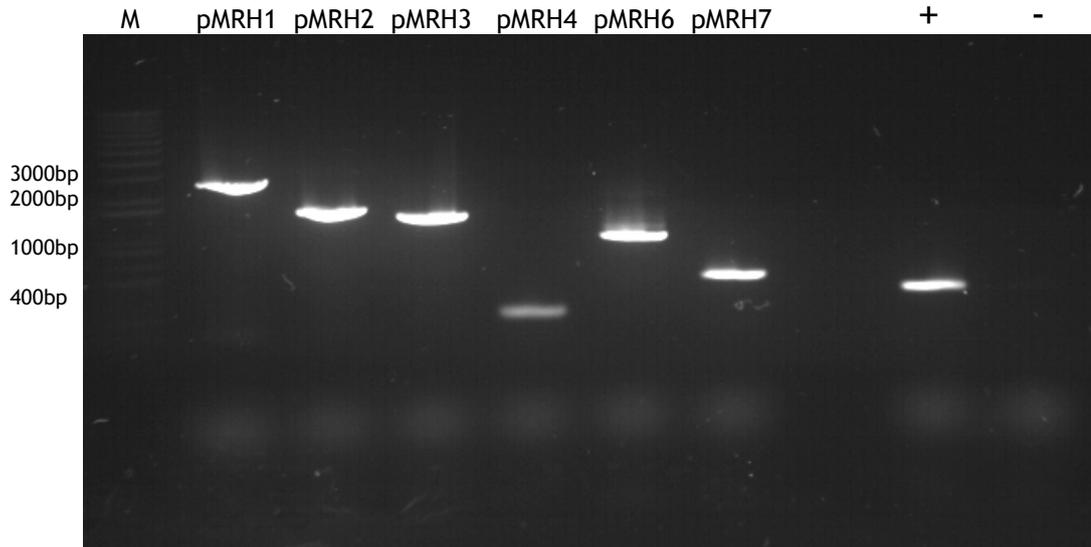


Figure 4.10: 0.8% agarose gel showing products from PCR amplification of PhtD gene and fragments

PCR products corresponding to F/L PhtD and the fragments of the PhtD coding for the selected PhtD protein fragments were all successfully amplified. The positive control yielded the Ply fragment and the negative control contained no product, as expected.

All fragments generated were of the expected size. Fragment sizes are given in table 4.6 PCR products were cleaned using a QIAquick PCR Purification kit¹²⁷ as described in 2.6.2.1 before being inserted into the pOPINF¹¹⁹ expression vector.

PCR fragment	Product Size (bp)
pMRH1	2496
pMRH2	1527
pMRH3	1174
pMRH4	192
pMRH6	960
pMRH7	453

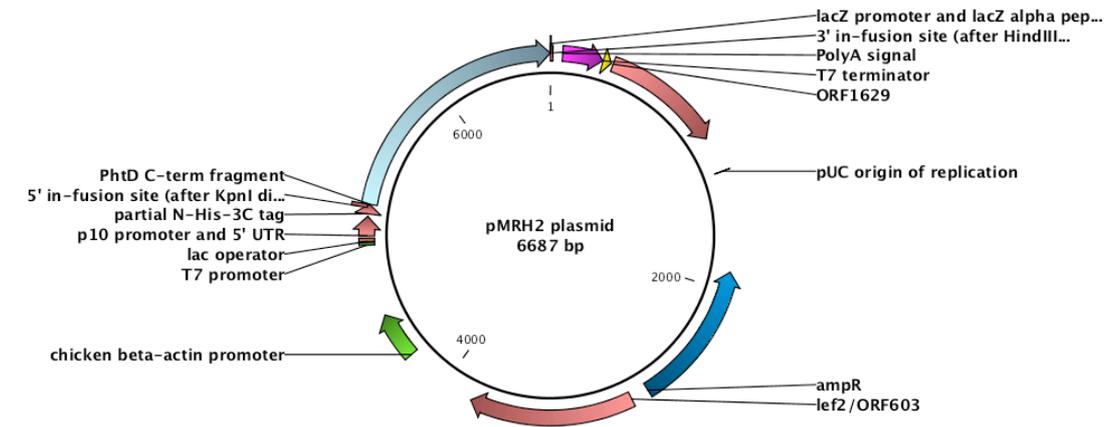
Table 4.6: PCR fragment product sizes

All product sizes include the additional 38bp generated for In-Fusion cloning

4.6.5 Generation Of pMRH Expression Constructs

The process for generating the expression constructs pMRH2, pMRH3, pMRH4, pMRH6 and pMRH7 was exactly the same as that used for generating the

pMRH1 F/L expression construct as detailed in 3.3.4 The cleaned PCR product comprising the relevant fragment was inserted into linearised pOPINF¹¹⁹ vector using the In-Fusion method.¹²⁵ Resulting expression constructs were verified by sequencing and stored as glycerol stocks (see 2.3.1). Vector maps for these constructs and the resulting translated sequences of the protein fragments are shown below in figures 4.11 to 4.15.



MAHHHHHS S G L E V L F Q G P S N H W V P D S R P E Q P S P Q S T P E P S P S P Q P A P N P Q P A P S N P I D E K L V K E A V
 R K V G D G Y V F E E N G V S R Y I P A K D L S A E T A A G I D S K L A K Q E S L S H K L G A K K T D L P S S D R E F Y N K A Y D L L
 A R I H Q D L L D N K G R Q V D F E A L D N L L E R L K D V P S D K V K L V D D I L A F L A P I R H P E R L G K P N A Q I T Y T D D E
 I Q V A K L A G K Y T T E D G Y I F D P R D I T S D E G D A Y V T P H M T H S H W I K K D S L S E A E R A A A Q A Y A K E K G L T P P
 S T D H Q D S G N T E A K G A E A I Y N R V K A A K K V P L D R M P Y N L Q Y T V E V K N G S L I I P H Y D H Y H N I K F E W F D E G
 L Y E A P K G Y T L E D L L A T V K Y Y V E H P N E R P H S D N G F G N A S D H V R K N K V D Q D S K P D E D K E H D E V S E P T H P
 E S D E K E N H A G L N P S A D N L Y K P S T D T E E T E E E A E D T T D E A E I P Q V E N S V I N A K I A D A E A L L E K V T D P S
 I R Q N A M E T L T G L K S S L L L G T K D N N T I S A E V D S L L A L L K E S Q P A P I Q

Figure 4.11: pMRH2 expression construct

DNA was fused N-terminally to a His₆-tag that incorporates an HRV-C3 protease site for cleavage of the His₆-tag. Regulation of recombinant protein expression was placed under control of the upstream T7 promoter. The bottom panel shows the resulting translated protein sequence. Plasmid map was generated using CLC Genomics Workbench 4.¹²⁰

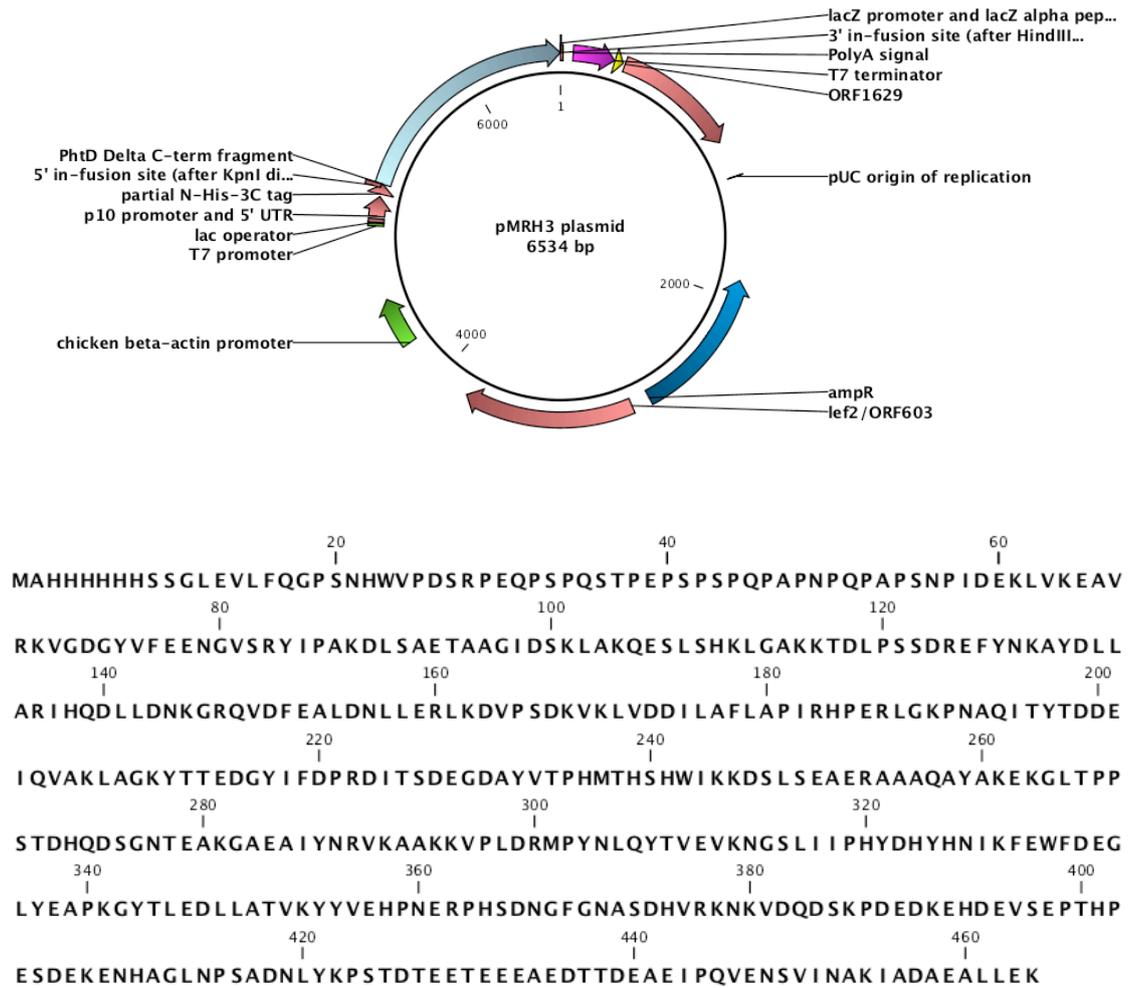


Figure 4.12: pMRH3 expression construct

DNA was fused N-terminally to a His₆-tag that incorporates an HRV-C3 protease site for cleavage of the His₆-tag. Regulation of recombinant protein expression was placed under control of the upstream T7 promoter. The bottom panel shows the resulting translated protein sequence. Plasmid map was generated using CLC Genomics Workbench 4.¹²⁰

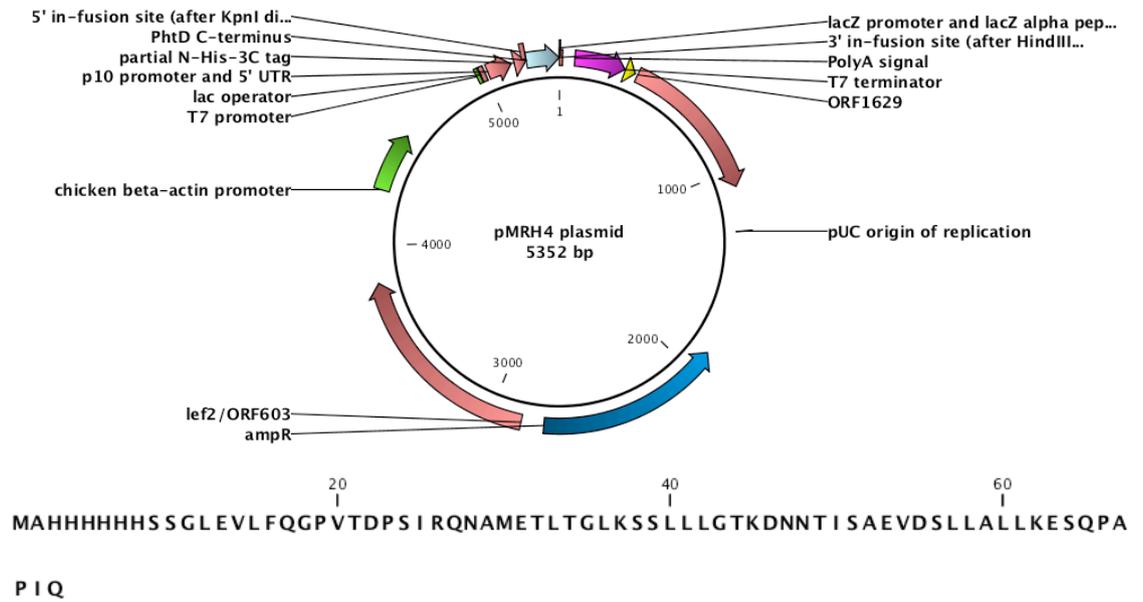


Figure 4.13: pMRH4 expression construct

DNA was fused N-terminally to a His₆-tag that incorporates an HRV-C3 protease site for cleavage of the His₆-tag. Regulation of recombinant protein expression was placed under control of the upstream T7 promoter. The bottom panel shows the resulting translated protein sequence. Plasmid map was generated using CLC Genomics Workbench 4.¹²⁰

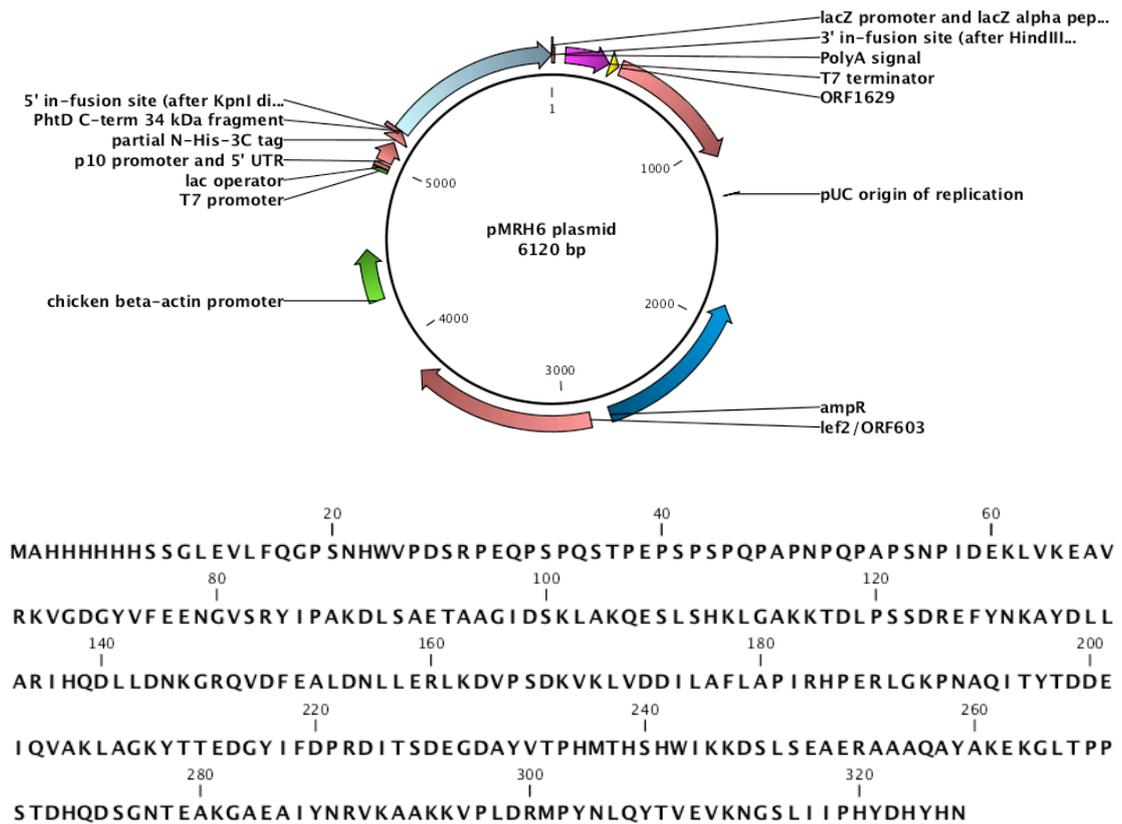


Figure 4.14: pMRH6 expression construct

DNA was fused N-terminally to a His₆-tag that incorporates an HRV-C3 protease site for cleavage of the His₆-tag. Regulation of recombinant protein expression was placed under control of the upstream T7 promoter. The bottom panel shows the resulting translated protein sequence. Plasmid map was generated using CLC Genomics Workbench 4.¹²⁰

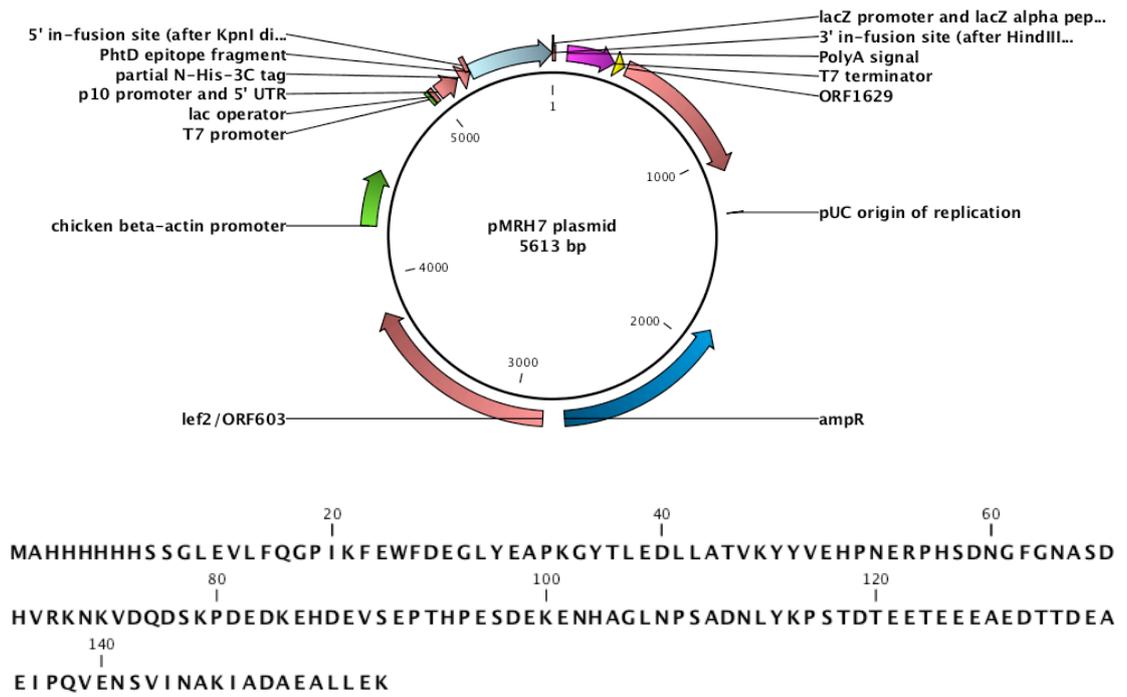


Figure 4.15: pMRH7 expression construct

DNA was fused N-terminally to a His₆-tag that incorporates an HRV-C3 protease site for cleavage of the His₆-tag. Regulation of recombinant protein expression was placed under control of the upstream T7 promoter. The bottom panel shows the resulting translated protein sequence. Plasmid map was generated using CLC Genomics Workbench 4.¹²⁰

4.7 Expression Of Truncated Forms Of PhtD

Expression of all truncated forms of PhtD was carried out in exactly the same way as expression of the F/L PhtD protein as described in 4.4 Plasmids were transformed into Rosetta2(DE3)pLysS competent cells, a starter culture was grown in LB, which was then used to inoculate the expression cultures of TB-Onyx.

4.8 Purification Of Recombinant Truncated Forms Of PhtD

The process for purification of the various PhtD protein fragments was carried out in exactly the same way as the purification of the F/L PhtD protein as described in 4.5 Pelleted cells were lysed using a French Pressure cell, centrifuged and filtered to remove particulate matter then subjected to a process of IMAC Ni-affinity chromatography, His₆ tag cleavage accompanied by reverse Ni-affinity chromatography, and finally anion exchange chromatography. Samples were analysed between each stage of the purification process by SDS-PAGE as outlined in 2.10.1 Purified PhtD protein fragments were concentrated to a concentration >10mg/ml using Amicon Ultra centrifugal concentrators with a MWCO of 10 kDa as outlined in 2.9.3 Samples of resulting purified, concentrated PhtD protein fragments are shown in figure 4.16.

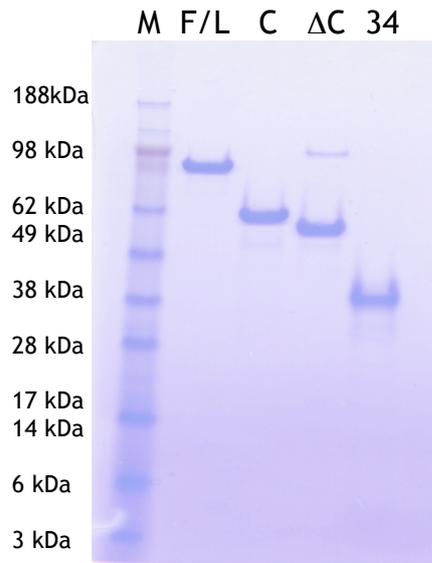


Figure 4.16: Recombinantly expressed and purified PhtD proteins
Lane 1: SeeBlue +2 Marker, Lane 2: F/L, Lane 3: C-Term, Lane 4: ΔC-term,
Lane 5: 34 kDa fragment. Each lane was loaded with 1μg of purified protein.

Although the expression constructs pMRH4 and pMRH7 which were designed to encode the 5.3 kDa extreme C-terminus of PhtD and a 15 kDa region of interest were successfully created (and verified through sequencing), and could be cultured in bacterial cell lines allowing recombinant protein expression, it was not possible to obtain any recombinant protein for these two fragments. Initial attempts at protein purification were unsuccessful, and subsequent Western blotting of cell lysates from the expression cultures with anti-PhtD showed that the protein fragments encoded by these constructs were not expressed using the devised expression protocol. It is possible that variation of expression culturing conditions for these constructs may have resulted in successful protein expression; however due to time constraints this was not attempted and these two constructs were not taken forward any further.

Chapter 5 : Biophysical Characterisation Of PhtD

5.1 Limited Proteolysis Of F/L PhtD

Data presented in this thesis provides an insight into the structure of the PhtD molecule. Initially, bioinformatic analysis of the PhtD protein sequence was performed using structural prediction tools (RONN disorder prediction server). Structural investigation was carried out utilising a variety of techniques including Nuclear Magnetic Resonance (NMR) spectroscopy and Circular Dichroism (CD) to experimentally test the bioinformatics structural prediction results. The resulting data from the NMR analysis of F/L PhtD compounded the RONN¹³⁹ prediction that PhtD has structured regions, but that PhtD as a whole is largely mobile, with regions moving independently of each other resulting in an overall very flexible or mobile protein structure (outlined in chapter 6.3). In order to assess whether it was possible to isolate the more highly structured regions of the protein, F/L PhtD was subjected to limited proteolysis by performing a controlled time-course tryptic digest.¹⁶⁰ In conjunction with structural predictions (see chapter 6.2) and the NMR data, the results in this chapter were used to identify structured regions of the PhtD molecule to be cloned and expressed that may be more conducive crystallisation than the F/L PhtD molecule.

5.1.1 Preliminary Investigation

Before any wet labwork was performed, the PhtD AA sequence was input into the program PeptideCutter (<http://www.expasy.ch/tools/peptidecutter/>) hosted at the ExPASy proteomics server¹¹⁴ to identify potential protease cleavage sites. Trypsin is a serine protease, which is able to cleave proteins C-terminally to lysine or arginine residues. PeptideCutter identified 85 possible cleavage sites, which would reduce the PhtD protein to very small fragments if 100% cleavage occurred. It was hypothesised that if there were highly structured regions within the molecule, these would intrinsically be rigid, folded regions of the protein and so any trypsin cleavage sites present in these regions could be inaccessible to the protease, resulting in stable fragments of the protein which were resistant to proteolytic degradation.

5.1.2 Limited Proteolysis Of Apo- F/L PhtD

Limited proteolysis experiments were designed based on previous research.¹⁶⁰ For the limited proteolysis experiment, 1mg of purified F/L PhtD protein was exposed to 25 μ g of proteomics grade trypsin in a total reaction volume of 500 μ l as outlined in table 5.1 The reaction was carried out at ambient temperature. 40 μ l samples were removed at 1, 3, 5, 10, 15, 20, 30, 45 and 60min time intervals. Before the reaction was begun, a sample of the undigested protein was taken in order to confirm that degradation of the protein had not occurred prior to addition of the protease.

Reaction Component	Quantity
Recombinant purified F/L PhtD	1mg
Proteomics grade trypsin	25 μ g
	Final Vol
	500 μ l

Table 5.1: Limited proteolysis reaction composition

Components were made up to final volume of 500 μ l with protein buffer (50mM Tris HCl pH7.5, 500mM NaCl, 20mM imidazole.)

As a negative control to show that the F/L PhtD was not undergoing self-degradation, the exact same reaction was performed using F/L PhtD protein from the same stock, but the trypsin protease was excluded from the reaction mixture. The two experiments were performed at the same time to ensure consistency between reactions. All samples were mixed 1:1 with NuPAGE SDS loading buffer and denatured in a boiling water-bath as soon as the samples were removed from the reaction in order to gain as accurate a profile of protein digestion as possible for each time-point. Samples were subsequently analysed by SDS-PAGE as outlined in 2.10.1. The resulting profile of limited proteolysis for Apo- F/L PhtD is shown in figure 5.1

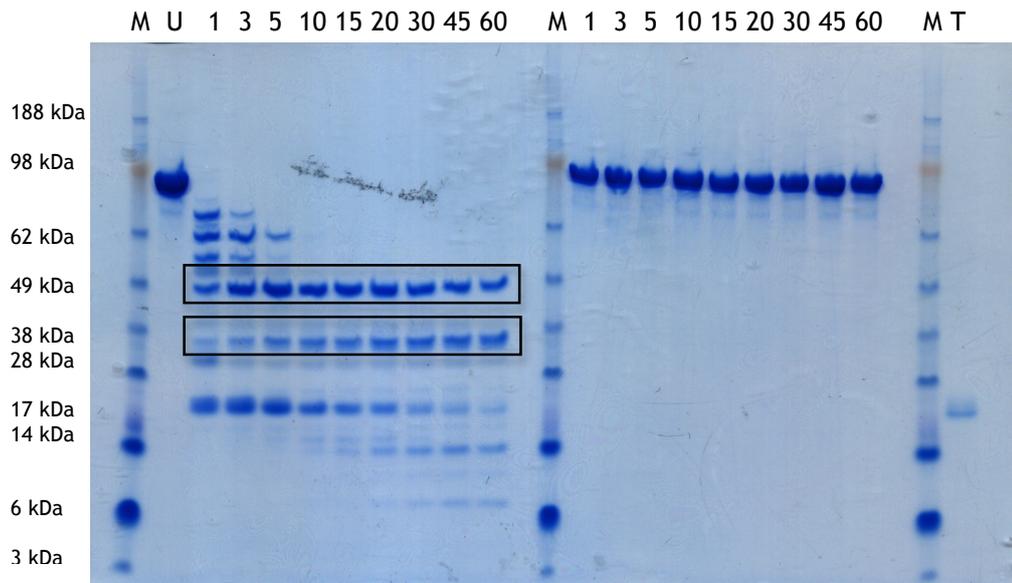


Figure 5.1: Limited proteolysis of Apo-F/L PhtD

The limited proteolysis experiment was carried out at ambient temperature in Tris buffer (50mM Tris HCl pH7.5, 500mM NaCl, 20mM imidazole). 1mg of protein at a concentration of 2mg/ml was utilised for the experiment. A 40 μ l sample was removed at each time-point with each lane of the SDS-PAGE gel therefore theoretically comprising 80 μ g total protein.

Lanes 1, 13 & 14: SeeBlue Plus 2 Mol. Wt. markers; Lane 2: Undigested F/L PhtD; Lanes 3 - 11: Time-points for F/L PhtD digest with trypsin; Lanes 14 -22: Time-points for F/L PhtD without trypsin; Lane 25: Trypsin. Boxes on the left show the two stable fragments throughout the proteolysis experiment.

The limited proteolysis profile showed that the F/L PhtD was degraded very rapidly; after 1 minute there is no evidence of any F/L protein intact, and that after 5 minutes, the most abundant fragment of the protein is around half the size of the F/L molecule, with the largest visible product only two-thirds the size of the F/L molecule. However, it can be seen that after 5 minutes two stable bands (indicated in figure 5.1) have been obtained and appear resistant to any further degradation. These bands are approximately 49 and 34 kDa in size respectively by SDS-PAGE. The negative control has confirmed that the F/L molecule is relatively stable under the same conditions when trypsin is absent, and that the presence of trypsin is required for proteolysis to occur. The identification of these two stable, structured regions of PhtD was successfully performed by utilising a combination of mass-spectrometry and N-terminal protein sequencing (discussed in chapter 5.2, 5.3 and 5.4). This data was subsequently used to generate expression constructs which allowed expression and purification of

these fragments in order to attempt to determine their structure by macromolecular x-ray crystallography (detailed in chapter 3).

5.1.3 Limited Proteolysis Of Zn-Bound F/L PhtD

From work carried out previously at the University of Glasgow and data in this thesis, there is evidence that the Pneumococcal Histidine Triad family of proteins bind Zn^{2+} through the HxxHxH motifs for which the family is named. This was seen to be true in the crystal structure of a fragment of Pneumococcal Histidine Triad A (PhtA) which revealed a Zn^{2+} atom bound in the HxxHxH motif.⁹⁸ It was hypothesised that the binding of Zn^{2+} atoms may induce some form of conformational change in the F/L PhtD molecule resulting in a more overall ordered protein and therefore more resistant to proteolytic activity. To investigate this, the limited proteolysis experiment outlined in 5.1.2 was repeated but this time a source of Zn^{2+} atoms was introduced to the F/L PhtD protein before exposure to trypsin. The reaction was set up and performed as described above, but before addition of the trypsin, $ZnSO_4$ was added to the F/L PhtD to a final concentration of 10mM. Samples were again taken at set time-points and analysed by SDS-PAGE. The resulting SDS-PAGE is shown in figure 5.2.

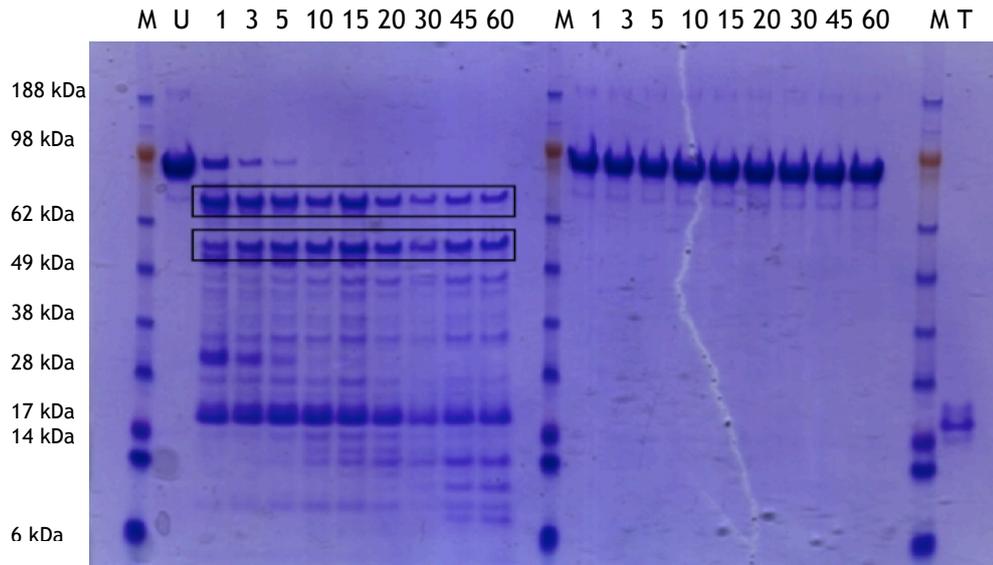


Figure 5.2: Limited proteolysis of Zn^{2+} -bound F/L PhtD

The limited proteolysis experiment was carried out at ambient temperature in Tris buffer (50mM Tris HCl pH7.5, 500mM NaCl, 20mM imidazole). 1mg of protein at a concentration of 2mg/ml was utilised for the experiment. A 40 μ l sample was removed at each time-point with each lane of the SDS-PAGE gel therefore theoretically comprising 80 μ g total protein.

Lanes 1, 13 & 14: SeeBlue Plus 2 Mol. Wt. markers; Lane 2: Undigested Zn-bound F/L PhtD; Lanes 3 - 11: Time-points for Zn-bound F/L PhtD digest with trypsin; Lanes 14 - 22: Time-points for Zn-bound F/L PhtD without trypsin; Lane 25: Trypsin.

Boxed areas show the two stable protein fragments in the presence of Zn^{2+} showing the increase in MWT of the two fragments when compared to the Apo-form.

A side-by-side comparison of the two digests (Figure 5.3) shows that there is a marked difference in the limited proteolysis profile of F/L PhtD when comparing the Apo- and Zn^{2+} -bound forms of the protein.

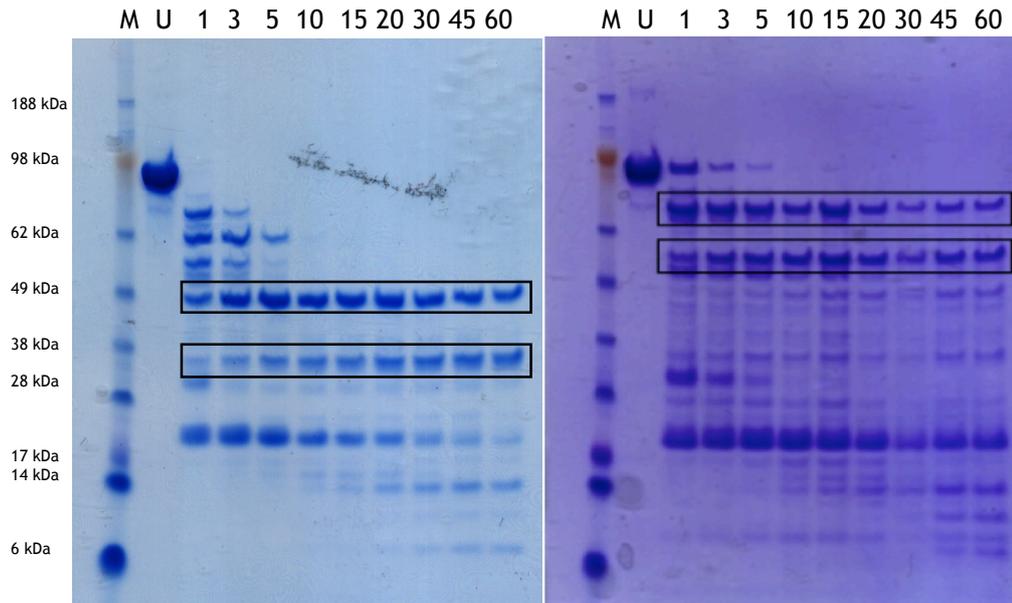


Figure 5.3: Comparison of Apo- and Zn^{2+} -bound F/L PhtD limited proteolysis

The limited proteolysis experiment was carried out at ambient temperature in Tris buffer (50mM Tris HCl pH7.5, 500mM NaCl, 20mM imidazole). 1mg of protein at a concentration of 2mg/ml was utilised for the experiment. A 40 μ l sample was removed at each time-point with each lane of the SDS-PAGE gel therefore theoretically comprising 80 μ g total protein.

Boxed out regions denote the stable fragments in the Apo-form on the left and in the Zn^{2+} -bound form on the right.

The first major observation is that there are still two proteolytically resistant bands on the SDS-PAGE gel for the Zn^{2+} -bound F/L PhtD, but that they have increased in size by approximately 25 kDa for the larger protein fragment from ~49 kDa to ~75 kDa, and by approximately 21 kDa for the smaller protein fragment from ~34 kDa to ~55 kDa. In relation to this stabilising effect on the fragments, it has been observed that whilst there is a complete absence of intact F/L PhtD after 1min under Apo- conditions, in contrast it is evident that under Zn^{2+} -bound conditions intact F/L molecule is still present after 5mins in the presence of the protease. This reinforces the hypothesis that structural changes induced by the binding of Zn^{2+} atoms are conferring a more highly ordered proteolytically resistant structure on the F/L PhtD molecule.

5.1.4 Limited Proteolysis Of F/L PhtD In The Presence Of Selected Divalent-Cations

Due to the ability of PhtD to bind Zn^{2+} atoms via its canonical HxxHxH motifs, the possibility that the protein would be able to bind other divalent cations was investigated.

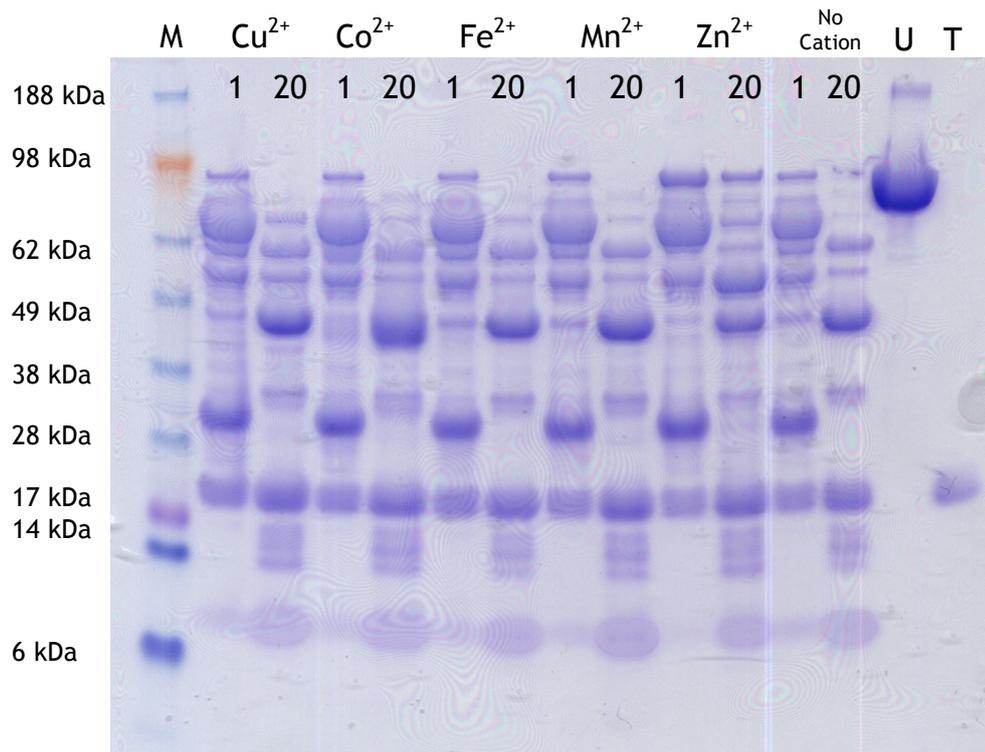


Figure 5.4: Limited proteolysis of F/L PhtD in the presence of selected divalent cations

Two samples were analysed for each cation; an initial time-point and then 20mins post incubation with trypsin. These profiles were compared to the apo-form where no cation was present. Zn^{2+} is the only cation which appears to have a visible effect on protein stability, with F/L PhtD still present after 20mins. All others show the same profile as the apo-digest suggesting they have no effect on protein structure.

As before, purified F/L PhtD was subjected to limited proteolysis in the same way as outlined in 5.1.2 and 5.1.3 with the relevant metal being added to the F/L PhtD before incubation with the protease (trypsin). A selection of divalent-cations were used to test the hypothesis that PhtD was able to selectively bind divalent cations and thus had an exclusive affinity for Zn^{2+} atoms. Selected metals were as follows: $CuSO_4$, $CoCl_2$, $FeSO_4$, and $MnSO_4$. All solutions were made up as a 0.5M stock. To conserve protein stocks, all values were reduced by a factor of 10 from the original limited proteolysis

experiments and only 2 time-points were taken; one at 1 minute post-incubation and one at 20 minutes post-incubation. These time-points were chosen as they were expected to show a clear difference between profiles. To provide comparable digest profiles, limited proteolysis digests were also performed with an equivalent concentration of ZnSO_4 to act as a positive control, and without any divalent-cation present to act as a negative control. 40 μl samples were removed for each per time point, mixed 1:1 with NuPAGE SDS-loading dye and denatured immediately in a boiling water-bath for 5mins. Samples were run on SDS-PAGE as before to visualise digest profiles. These results are illustrated in figure 5.4.

Analysis of the digest profiles for the various selected divalent-cations shows that of the divalent-cations tested, all profiles were visually identical to the profile of the Apo- F/L PhtD limited proteolysis digest, with no F/L molecule evident in the second time-point sample for any of the tested Cu^{2+} , Co^{2+} , Fe^{2+} or Mn^{2+} divalent cations, and a large band present of ~55 kDa. In contrast there is an absence of this strong ~55 kDa band at the second time-point, and clearly a significant quantity of F/L PhtD molecule present in the second time-point sample for the PhtD digest in the presence of ZnSO_4 supporting the hypothesis that PhtD may have a ability to bind Zn^{2+} atoms to the exclusion of other divalent-cations. It is possible however, that any of these metals tested may have an effect at a higher concentration, although it was expected that the concentrations used would be sufficient to saturate any binding sites and therefore that any effect would have been noticed at the concentration used.

5.2 Mass-Spectrometry Analysis Of F/L PhtD Limited Proteolysis Profiles

5.2.1 Overview

The technique of mass spectrometry was used in two ways. Primarily, coupled with N-terminal protein sequencing, it was used as an aid to identification of the stable, structured regions of PhtD observed in the limited proteolysis experiments. Furthermore, once these regions had been identified and a cloning strategy designed and adopted, mass spectrometry

was again employed; this time, coupled with DNA sequencing of the expression constructs, the resulting expressed PhtD fragments were also analysed by mass spectrometry in a 'belt and braces' approach to make doubly sure that the correct product had been obtained from the protein expression process. Mass-spectrometry was carried out at the University of Glasgow with the kind assistance of Dr. Richard Burchmore, and at GlaxoSmithKline Biologicals with the assistance of Satty Borman.

5.2.2 Intact-Mass MALDI-ToF Mass Spectrometry Of Limited Proteolysis Profiles (Peptide Mass Fingerprinting)

An Apo- F/L PhtD limited proteolysis experiment was performed exactly as outlined in 5.1.2 Duplicate samples were removed at each time-point; one sample was mixed 1:1 with NuPAGE SDS loading dye and denatured as before for analysis by SDS-PAGE. With the second sample the reaction was quenched by addition of 2 μ l TCA, as the serine-protease trypsin is rendered inactive when under acidic conditions.¹⁶¹ The addition of 2 μ l neat TCA proved enough to lower the pH in the reaction below pH7 and into an acidic environment. This was required as intact-mass measurement requires that the protein/s to be analysed are in their native-state to obtain accurate mass measurements of the sample, hence the reaction could not be stopped in the usual way by denaturing with SDS.

In contrast to traditional MALDI-ToF mass spectrometry for an in-gel sample, the samples were not subjected to further tryptic digestion before running through the mass spectrometer in an attempt to identify each fragment based on its accurate mass-fingerprint. The samples from three distinct time-points were analysed in this way. Once the accurate masses had been obtained for the various fragments in each of the time-point samples, results were analysed using the MASCOT search engine (www.matrixscience.com/).¹⁶² The partner SDS-PAGE gel is shown in figure 5.5 Samples taken at time-points 1, 3, and 60mins were subjected to PMF.

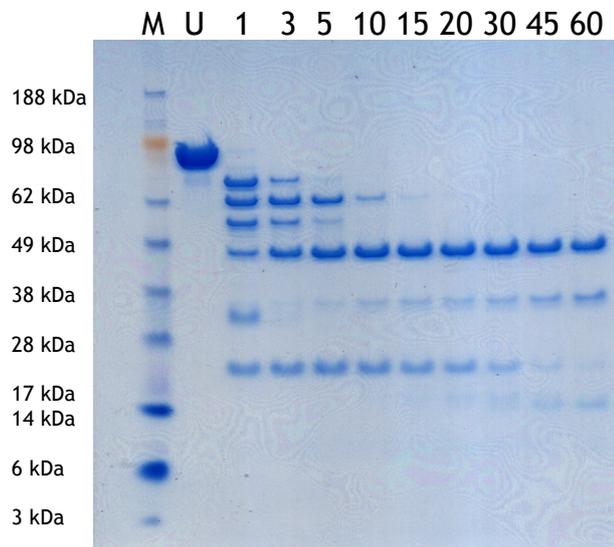


Figure 5.5: SDS-PAGE profile of F/L PhtD tryptic digest subjected to PMF

Samples from the TCA quenched reaction corresponding to the 1, 3, and 60min time-points on the SDS-PAGE gel were analysed by PMF.

Analysing the resultant mass spectrometry data with MASCOT¹⁶² successfully identified the protein from which the fragments were derived, giving the expected gene number SP_1003, which corresponds to PhtD in the pneumococcal strain TIGR4. Disappointingly, no accurate masses were obtained which corresponded to the 49 kDa or 34 kDa stable fragments. However, accurate masses were obtained for many much smaller protein fragments, with coverage of these smaller fragments increasing over time that could be mapped back to the N-terminal half of the protein and the extreme C-terminus of the protein, leaving a large area in the C-terminal half of estimated molecular weight 35.7 kDa, corresponding roughly to one of the stable fragments which appeared resistant to digestion. It was elucidated that either the samples had failed to ionise sufficiently to fly in the mass-spectrometer, or that they were simply too large to fly without being further digested. The results of the MASCOT¹⁶² search comparing the 1 and 3min samples are illustrated in figures 5.6 and 5.7.

SYELGRHQAG	QVKKESNRVS	YIDGDQAGQK	AENLTPDEVS	KREGINAEQI
VIKITDQGYV	TSHGDHYHY	NGKVPYDAII	SEELLMKDPN	YQLKSDIVN
EIKGGYVIKV	DGKYVYVYVKD	AAHADNIRTK	EEIKRQKQEH	SHNHGGGSND
QAVVAARAQG	RYTTDDGYIF	NASDIIEDTG	DAYIVPHGDH	YHYIPKNELS
ASELAAAEAY	WNGKQGSRPS	SSSSYNANPA	QPRLSENHNL	TVTPTYHQNQ
GENISSLLRE	LYAKPLSERH	VESDGLIFDP	AQITSRTARG	VAVPHGNHYH
FIPYEQMSEL	EKRIARI IPL	RYRSNHVVPD	SRPEQPSPQS	TPEPSPSPQP
APNPQPAPSN	PIDEKLVKEA	VRKVG DGYVF	EENGVSRYIP	AKDLSAETAA
GIDSKLAKQE	SLSHKLGAKK	TDLPSSDREF	YNKAYDLLAR	IHQDLLDNKG
RQVDFEALDN	LLERLKD VPS	DKVKLVDDIL	AFLAPIRHPE	RLGKPN AQIT
YTDDEIQVAK	LAGKYTTEDG	YIFDPRDITS	DEGDAYVTPH	MTHSHWIKKD
SLSEAERAAA	QAYAKEKGLT	PPSTDHQDSG	NTEAKGAEAI	YNRVKA AKKV
PLDRMPYNLQ	YTVEVKNGSL	IIPHYDHYHN	IKFEWFDEGL	YEAPKGYTLE
DLLATVKYYV	EHPNERPHSD	NGFGNASDHV	RKNKVDQDSK	PDEDKEHDEV
SEPTHPESDE	KENHAGLNPS	ADNLYKPSTD	TEETEEEAED	TTDEAEIPQV
ENSVINAKIA	DAEALLEKVT	DPSIRQNAME	TLTGLKSSLL	LGTKDNNTIS
AEVDSLLALL	KESQPAPIQ			

Figure 5.6: PMF analysis of 1min digest sample
Identified peptides are shown in blue.

SYELGRHQAG	QVKKESNRVS	YIDGDQAGQK	AENLTPDEVS	KREGINAEQI
VIKITDQGYV	TSHGDHYHY	NGKVPYDAII	SEELLMKDPN	YQLKSDIVN
EIKGGYVIKV	DGKYVYVYVKD	AAHADNIRTK	EEIKRQKQEH	SHNHGGGSND
QAVVAARAQG	RYTTDDGYIF	NASDIIEDTG	DAYIVPHGDH	YHYIPKNELS
ASELAAAEAY	WNGKQGSRPS	SSSSYNANPA	QPRLSENHNL	TVTPTYHQNQ
GENISSLLRE	LYAKPLSERH	VESDGLIFDP	AQITSRTARG	VAVPHGNHYH
FIPYEQMSEL	EKRIARI IPL	RYRSNHVVPD	SRPEQPSPQS	TPEPSPSPQP
APNPQPAPSN	PIDEKLVKEA	VRKVG DGYVF	EENGVSRYIP	AKDLSAETAA
GIDSKLAKQE	SLSHKLGAKK	TDLPSSDREF	YNKAYDLLAR	IHQDLLDNKG
RQVDFEALDN	LLERLKD VPS	DKVKLVDDIL	AFLAPIRHPE	RLGKPN AQIT
YTDDEIQVAK	LAGKYTTEDG	YIFDPRDITS	DEGDAYVTPH	MTHSHWIKKD
SLSEAERAAA	QAYAKEKGLT	PPSTDHQDSG	NTEAKGAEAI	YNRVKA AKKV
PLDRMPYNLQ	YTVEVKNGSL	IIPHYDHYHN	IKFEWFDEGL	YEAPKGYTLE
DLLATVKYYV	EHPNERPHSD	NGFGNASDHV	RKNKVDQDSK	PDEDKEHDEV
SEPTHPESDE	KENHAGLNPS	ADNLYKPSTD	TEETEEEAED	TTDEAEIPQV
ENSVINAKIA	DAEALLEKVT	DPSIRQNAME	TLTGLKSSLL	LGTKDNNTIS
AEVDSLLALL	KESQPAPIQ			

Figure 5.7: PMF analysis of 3min digest sample
Identified peptides are shown in blue. There is an increase in peptides from the N-terminal half of the protein, whilst the C-term half remains resistant to proteolytic digestion.

The largely intact area starting at residue 472 Q (glutamine) where no peptides have been recovered is clearly visible for both datasets.

To reinforce these findings, a selection of bands including the 49 kDa and 34 kDa stable bands were excised from the SDS-PAGE gel and subjected to PMF. As these samples were isolated from each other as a result of the SDS-PAGE process, each sample was further digested with trypsin prior to PMF analysis. In this instance, the problem initially observed with the larger fragments was overcome due to the fact that the protein had been denatured and so any previously protected trypsin cleavage sites should be easily accessible resulting in smaller, processible peptide fragments. As before, resulting data was processed using the MASCOT¹⁶² search engine, and protein identity was successfully confirmed as being derived from SP_1003.

As postulated, when the PMF profiles for the 49 kDa and 34 kDa bands were mapped onto the PhtD protein sequence, they fell exactly into the region which was shown to be protected from proteolysis in the non-denatured samples. The MASCOT¹⁶² results for these two protein fragments are illustrated in figures 5.8 and 5.9.

SYELGRHQAG	QVKKESNRVS	YIDGDQAGQK	AENLTPDEVS	KREGINAEQI
VIKITDQGYV	TSHGDHYHYY	NGKVPYDAII	SEELLMKDPN	YQLKDSDIVN
EIKGGYVIKV	DGKYVYVLKD	AAHADNIRTK	EEIKRQKQEH	SHNHGGGSND
QAVVAARAQG	RYTTDDGYIF	NASDIIEDTG	DAYIVPHGDH	YHYIPKNELS
ASELAAAEAY	WNGKQGSRPS	SSSSYNANPA	QPRLSENHNL	TVTPTYHQNQ
GENISSLLRE	LYAKPLSERH	VESDGLIFDP	AQITSRTARG	VAVPHGNHYH
FIPYEQMSEL	EKRIARI IPL	RYRSNHWVPD	SRPEQPSPQS	TPEPSPSPQP
APNPQPAPSN	PIDEKLVKEA	VRKVG DG YVF	EENGVSRYIP	AKDLSAETAA
GIDSKLAKQE	SLSHKLGAKK	TDLPS SDREF	YNKAYDLLAR	IHQDLLDNKG
RQVDFEALDN	LLERLKD VPS	DKVKLVDDIL	AFLAPIRHPE	RLGKPN AQIT
YTDDEIQVAK	LAGKYTTEDG	YIFDPRDITS	DEGDAYVTPH	MTHSHWIKKD
SLSEAERAAA	QAYAKEKGLT	PPSTDHQDSG	NTEAKGAEAI	YNRVKA AKKV
PLDRMPYNLQ	YTVEVKNGSL	IIPHYDHYHN	IKFEWFDEGL	YEAPKGYTLE
DLLATVKY YV	EHPNERPHSD	NGFGNASDHV	RKNKVDQDSK	PDEDKEHDEV
SEPTHPESDE	KENHAGLNPS	ADNLYKPSTD	TEETEEEAED	TTDEAEIPQV
ENSVINAKIA	DAEALLEKVT	DPSIRQNAME	TLTGLKSSLL	LGTKDNNTIS
AEVDSL LALL	KESQPAPIQ			

Figure 5.8: PMF analysis of 49 kDa fragment excised from SDS-PAGE gel

Identified peptides are shown in blue. The identified peptides correspond to those absent from the matching region of sequence from analysis under non-denatured conditions.

SYELGRHQAG	QVKKESNRVS	YIDGDQAGQK	AENLTPDEVS	KREGINAEQI
VIKITDQGYV	TSHGDHYHY	NGKVPYDAII	SEELLMKDPN	YQLKSDIVN
EIKGGYVIKV	DGKYYVYLKD	AAHADNIRTK	EEIKRQKQEH	SHNHGGGSND
QAVVAARAQG	RYTTDDGYIF	NASDIIEDTG	DAYIVPHGDH	YHYIPKNELS
ASELAAAEAY	WNGKQGSRPS	SSSSYNANPA	QPRLSEHNHL	TVTPTYHQNQ
GENISSLLRE	LYAKPLSERH	VESDGLIFDP	AQITSRTARG	VAVPHGNHYH
FIPYEQMSEL	EKRIARIIPL	RYRSNHWVPD	SRPEQPSPQS	TPEPSPSPQP
APNPQPAPSN	PIDEKLVKEA	VRKVG DG YVF	EENGVSRYIP	AKDLSAETAA
GIDSKLAKQE	SLSHKLGAKK	TDLPSSDREF	YNKAYDLLAR	IHQDLLDNKG
RQVDFEALDN	LLERLKD VPS	DKVKLVDDIL	AFLAPIRHPE	RLGKPN AQIT
YTDDEIQVAK	LAGKYTTEDG	YIFDPRDITS	DEGDAYVTPH	MTHSHWIKKD
SLSEAERAAA	QAYAKEKGLT	PPSTDHQDSG	NTEAKGAEAI	YNRVKAACKV
PLDRMPYNLQ	YTVEVKNGSL	IIPHVDHYHN	IKFEWFDEGL	YEAPKGYTLE
DLLATVKYYV	EHPNERPHSD	NGFGNASDHV	RKNKVDQDSK	PDEDKEHDEV
SEPTHPESDE	KENHAGLNPS	ADNLYKPSTD	TEETEEEAED	TTDEAEIPQV
ENSVINAKIA	DAEALLEKVT	DPSIRQNAME	TLTGLKSSLL	LGTKDNNTIS
AEVDSLLALL	KESQPAPIQ			

Figure 5.9: PMF analysis of 34 kDa band excised from SDS-PAGE gel

Identified peptides are shown in blue. The identified peptides correspond to those absent from the matching region of sequence from analysis under non-denatured conditions.

It can clearly be seen that the peptides recovered from these two samples under denatured conditions corresponds to the region in the non-denatured samples that previously showed no mapped peptides.

As a result of these experiments, it was possible to successfully identify the region of the protein that seemed the most proteolytically resistant and therefore hypothetically was highly structured; a physiological characteristic which would contribute to proteolytic resistance due to the highly folded nature of this region making cleavage sites inaccessible.

In order to determine a definitive starting region for the 49 kDa and 34 kDa stable fragments, samples were subjected to Western transfer, prepared as outlined in 2.10.3 and sent for N-terminal sequencing at the University of Dundee. Analysis of the two fragments revealed that both fragments started at serine 344, with the leading sequence SNHWVPD. This corroborated the PMF data, as comparable peptides were recovered for both bands at the N-terminus, however the 49 kDa fragment contained extra peptides not present in the 34 kDa fragment at the C-terminal end. Using the sum of all this data, these two fragments have been successfully identified, allowing for a targeted cloning approach to be used in generation of protein fragments,

which can be employed in attempts to determine protein structure by macromolecular X-ray crystallography.

5.2.3 Western Blot Analysis Of PhtD

In an attempt to uncover specific regions of interest in the PhtD protein, the limited proteolysis profile of Apo- F/L PhtD was subjected to Western blotting as described in 2.10.3 using a polyclonal antibody raised in rabbits against F/L PhtD. It was hypothesised that, as the protein was digested over time, it may have been possible to elucidate antibody-binding regions of the protein.

Western blotting was carried out on Hybond-P PVDF membrane (Amersham Biosciences - GE Healthcare) according to the SOP optimised in the Mitchell Pneumococcal Lab (outlined in appendix A.3.2), using primary (1^o) and secondary (2^o) antibody dilutions of 1 in 2000. 1^o polyclonal antibody (provided by GSK, Rixensart Belgium) was raised against F/L PhtD protein in rabbits. 2^o antibody consisted of HRP-conjugated donkey anti-rabbit antibody (Amersham Biosciences - GE Healthcare). Reaction of the protein with the antibody was detected directly on the PVDF membrane by exposure to 4-chloro-1-naphthol, which visually reacts with HRP to produce a blue/purple colour. The developed membrane is shown in figure 5.10 and the corresponding coomassie stained SDS-PAGE gel is shown in figure 5.11.

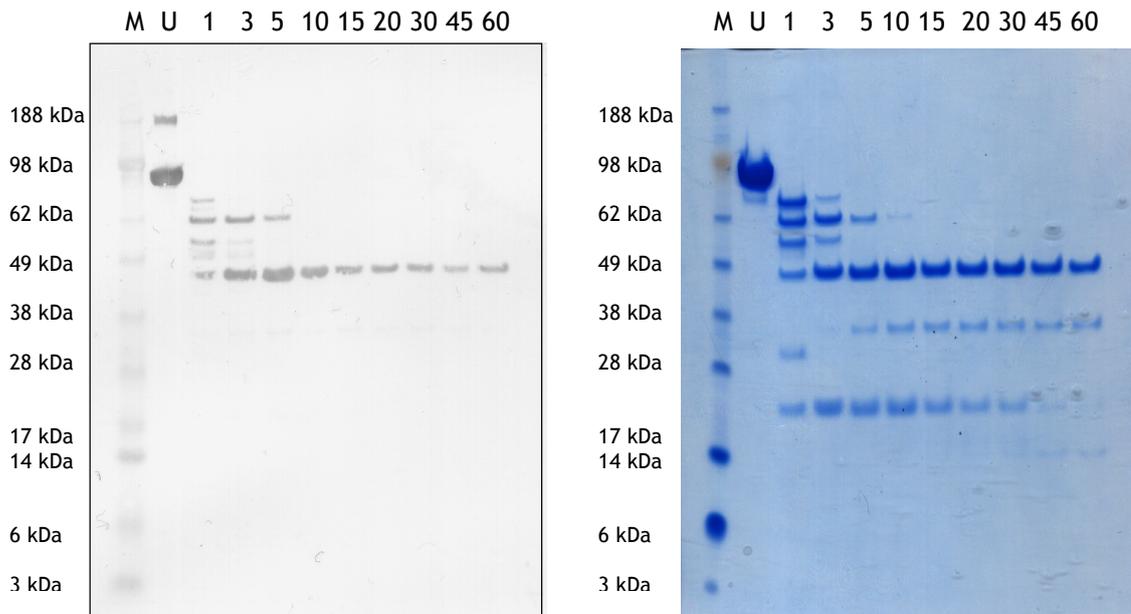


Figure 5.10: Western blot of Apo- F/L PhtD

Blotting with anti-PhtD polyclonal antibody reveals that the 34 kDa fragment does not react with the anti-body whereas the 49 kDa fragment is reactive. Using this information and prior mass-spectrometry data regarding the identity of the two stable fragments, it is elucidated that the additional 15 kDa of sequence present in the 49 kDa fragment contains the major antibody-binding site.

Figure 5.11: Coomassie stained gel corresponding to western blot (Fig 5.10)

This SDS-PAGE gel is a duplicate of the samples analysed by Western blot. The gel was stained with coomassie R250 visualising the range of fragments in the samples to allow direct comparison with the Western blot result.

It is clear from the comparison of the developed membrane and the accompanying coomassie stained gel that the 34 kDa band is not recognised by the anti-PhtD antibody.

Because the identity of these two protein fragments has been successfully identified using the mass-spectrometry PMF and N-terminal sequencing techniques, it is possible therefore to elucidate the dominant antibody binding epitope region. As the two fragments have the same starting sequence, it has been hypothesised that the 15 kDa region of the protein which differentiates the two fragments is contains this antibody binding domain. This region begins at iso-leucine 651 with the leading sequence IKFEW and ending at lysine 788 with the ending sequence DAEALLEK. The resulting protein sequence of the predicted antibody-binding domain is illustrated in figure 5.12.

```

IKFEWFDEGLYEAPKGYTLEDLLATVKYYVEHPNERPHSDNGFGNASDHVRKNKVDQDS
KPDEDKEHDEVSEPTHPESDEKENHAGLNPSADNLYKPSTDTEETEEEAEDTTDEAEIPQ
VENSVINAKIADAEALLEK

```

Figure 5.12: Sequence of predicted dominant PhtD antibody-binding region

As a result of these findings, it has been possible to locate pinpoint the dominant antibody-binding region for PhtD to within a 15 kDa fragment of the full 92 kDa protein. Utilising these findings, an expression construct for this fragment has been created (see chapter 3.6) so that structural determination of this small important fragment comprising the antibody-binding region may be possible if the full protein structure is not achieved.

5.3 Analysis Of Purified PhtD Proteins By MALDI-ToF Mass-Spectrometry

5.3.1 Overview

The results obtained from the data described above allowed a targeted approach to generating a series of expression constructs encoding selected fragments of the F/L PhtD protein (previously described in chapter 3.6). Following their construction, all newly created expression constructs (pMRH1-7) were sequenced in order to confirm that inserts had been successfully introduced into the pOPINF¹¹⁹ vector in-frame, without the presence of any mutations. Upon confirmation that the expression construct sequences were correct, expression and purification of the various fragments was carried out, detailed in chapter 3.8. As an additional precaution, samples of the various purified PhtD fragments were analysed by mass-spectrometry. This section details the results from the mass-spectrometry analysis of the various purified PhtD protein fragments.

5.3.2 Preparation Of PhtD Samples For Mass-Spectrometry Analysis

The 4 successfully recombinantly expressed and purified proteins (F/L, C-term, Δ C-term, and 34 kDa fragment) were each mixed 1:1 with NuPAGE SDS-loading dye and denatured at 90°C for 5mins in a hot-block. Samples of each were run in triplicate by SDS-PAGE against SeeBlue Plus 2 mol wt. markers as outlined in 2.10.1, loading 5 μ g of purified protein per lane in order to provide the maximum possible quantity of protein for analysis. The SDS-PAGE gel was stained with coomassie-R250 and destained with extensive changes of destain, followed by copious washing with dH₂O in order to remove residual destain.

5.3.3 Analysis Of Purified PhtD Proteins

Mass spectrometry of the purified F/L, C-term and Δ C-term proteins was performed with the kind assistance of Satty Borman at GlaxoSmithKline Biologicals, Stevenage, UK. Each of the bands corresponding to the purified proteins were excised from the gel with a scalpel. Three proteases were utilised in the analysis of the samples. Proteases used and their properties are given in table 5.2.

Protease	Properties
Trypsin	Cleaves C-terminally to K or R residues (unless downstream residue is P)
Chymotrypsin	Cleaves C-terminally to Y, W, or F residues
Asp-N	Cleaves N-terminally to D or E residues

Table 5.2: Proteases used in mass-spectrometry analysis of purified PhtD proteins.¹¹⁴

In order to gain the maximum possible coverage for each purified protein, the above three enzymes which all recognise different cleavage sites were

employed in order to digest the samples as much as possible. The SDS-PAGE gel corresponding to the samples analysed is shown in figure 5.13.

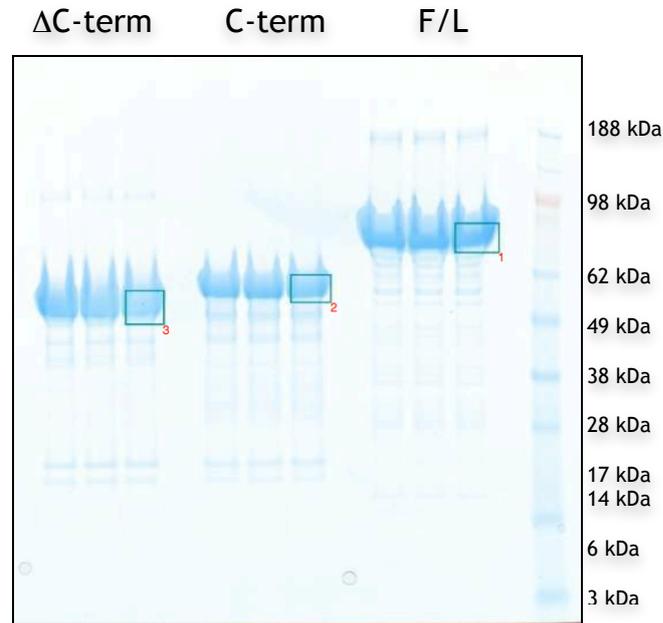


Figure 5.13: SDS-PAGE of purified PhtD samples for mass-spectrometry

Boxed samples indicate bands to be analysed. Each sample analysed in triplicate. Sample 1: F/L, Sample 2: C-term, Sample 3: Δ C-term. N.B. Samples are overloaded at $5\mu\text{g}$ per lane. Standard loading of $1\mu\text{g}$ per lane shows a clean, single band.

In order to gain as much data as possible for each protein, two forms of mass-spectrometry were employed; peptide mass fingerprinting (PMF) and tandem mass-spectrometry (MS-MS). The resulting mass-spec data was analysed using MASCOT.¹⁶² Figures 5.14, 5.15 and 5.16 show the combined coverage obtained from mass-spectrometry for F/L, C-term and Δ C-term PhtD respectively.

SYELGRHQAG	<u>QVKKESNRVS</u>	<u>YIDGDQAGQK</u>	AENLTPDEVS	KREGINAEQI
VIKITDQGYV	TSHGDHYHYH	NGKVPYDAII	SEELLMKDPN	YQLKDSDIVN
EIKGGYVIKV	DGKYVYVYVKD	AAHADNIRTK	EEIKRQKQEH	SHNHGGGSND
<u>QAVVAARAQG</u>	<u>RYTTDDGYIF</u>	<u>NASDIIEDTG</u>	<u>DAYIVPHGDH</u>	<u>YHYIPKNELS</u>
<u>ASELAAAEAY</u>	<u>WNGKQGSRPS</u>	<u>SSSSYNANPA</u>	<u>QRLSENHNL</u>	<u>TVTPTYHQNQ</u>
<u>GENISSLLRE</u>	<u>LYAKPLSERH</u>	<u>VESDGLIFDP</u>	<u>AQITSRTARG</u>	<u>VAVPHGNHYH</u>
<u>FIPYEQMSEL</u>	<u>EKRIARI IPL</u>	<u>RYRSNHVVPD</u>	<u>SRPEQSPQS</u>	<u>TPEPSPSPQP</u>
<u>APNPQPAPSN</u>	<u>PIDEKLVKEA</u>	<u>VRKVG DGYVF</u>	<u>EENGVSRYIP</u>	<u>AKDLSAETAA</u>
<u>GIDSKLAKQE</u>	<u>SLSHKLGAKK</u>	<u>TDLPS DREF</u>	<u>YNKAYDLLAR</u>	<u>IHQDLLDNKG</u>
<u>RQVDFEALDN</u>	<u>LLERLKD VPS</u>	<u>DKVKLVDDIL</u>	<u>AFLAPIRHP E</u>	<u>RLGKPN AQIT</u>
<u>YTDDEIQVAK</u>	<u>LAGKYTTEDG</u>	<u>YIFDPRDITS</u>	<u>DEGDAYVTPH</u>	<u>MTHSHWIKKD</u>
<u>SLSEAERAAA</u>	<u>QAYAKEKGLT</u>	<u>PPSTDHQDSG</u>	<u>NTEAKGAEAI</u>	<u>YNRVKA AKKV</u>
<u>PLDRMPYNLQ</u>	<u>YTVEVKN GSL</u>	<u>IIPHYDHYHN</u>	<u>IKFEWFDEGL</u>	<u>YEAPKGYTLE</u>
<u>DLLATVKY YV</u>	<u>EHPNERPHSD</u>	<u>NGFGNASDHV</u>	<u>RKNKVDQDSK</u>	<u>PDEDKEHDEV</u>
<u>SEPTHPESDE</u>	<u>KENHAGLNPS</u>	<u>ADNLYKPSTD</u>	<u>TEETEEEAED</u>	<u>TTDEAEIPQV</u>
<u>ENSVINAKIA</u>	<u>DAEALLEKVT</u>	<u>DPSIRQNAME</u>	<u>TLTGLKSSLL</u>	<u>LGTKDNNTIS</u>
<u>AEVDSLLALL</u>	<u>KESQPAPIQ</u>			

Figure 5.14: Complete mass-spectrometry results for purified F/L PhtD

Identified peptides are highlighted in blue. Peptides identified by MS-MS are underlined.

N.B. MS-MS identified peptides may duplicate PMF recovered peptides for some regions of sequence.

SYELGRHQAG	<u>QVKKESNRVS</u>	<u>YIDGDQAGQK</u>	AENLTPDEVS	KREGINAEQI
VIKITDQGYV	TSHGDHYHYH	NGKVPYDAII	SEELLMKDPN	YQLKDSDIVN
EIKGGYVIKV	DGKYVYVYVKD	AAHADNIRTK	EEIKRQKQEH	SHNHGGGSND
<u>QAVVAARAQG</u>	<u>RYTTDDGYIF</u>	<u>NASDIIEDTG</u>	<u>DAYIVPHGDH</u>	<u>YHYIPKNELS</u>
<u>ASELAAAEAY</u>	<u>WNGKQGSRPS</u>	<u>SSSSYNANPA</u>	<u>QRLSENHNL</u>	<u>TVTPTYHQNQ</u>
<u>GENISSLLRE</u>	<u>LYAKPLSERH</u>	<u>VESDGLIFDP</u>	<u>AQITSRTARG</u>	<u>VAVPHGNHYH</u>
<u>FIPYEQMSEL</u>	<u>EKRIARI IPL</u>	<u>RYRSNHVVPD</u>	<u>SRPEQSPQS</u>	<u>TPEPSPSPQP</u>
<u>APNPQPAPSN</u>	<u>PIDEKLVKEA</u>	<u>VRKVG DGYVF</u>	<u>EENGVSRYIP</u>	<u>AKDLSAETAA</u>
<u>GIDSKLAKQE</u>	<u>SLSHKLGAKK</u>	<u>TDLPS DREF</u>	<u>YNKAYDLLAR</u>	<u>IHQDLLDNKG</u>
<u>RQVDFEALDN</u>	<u>LLERLKD VPS</u>	<u>DKVKLVDDIL</u>	<u>AFLAPIRHP E</u>	<u>RLGKPN AQIT</u>
<u>YTDDEIQVAK</u>	<u>LAGKYTTEDG</u>	<u>YIFDPRDITS</u>	<u>DEGDAYVTPH</u>	<u>MTHSHWIKKD</u>
<u>SLSEAERAAA</u>	<u>QAYAKEKGLT</u>	<u>PPSTDHQDSG</u>	<u>NTEAKGAEAI</u>	<u>YNRVKA AKKV</u>
<u>PLDRMPYNLQ</u>	<u>YTVEVKN GSL</u>	<u>IIPHYDHYHN</u>	<u>IKFEWFDEGL</u>	<u>YEAPKGYTLE</u>
<u>DLLATVKY YV</u>	<u>EHPNERPHSD</u>	<u>NGFGNASDHV</u>	<u>RKNKVDQDSK</u>	<u>PDEDKEHDEV</u>
<u>SEPTHPESDE</u>	<u>KENHAGLNPS</u>	<u>ADNLYKPSTD</u>	<u>TEETEEEAED</u>	<u>TTDEAEIPQV</u>
<u>ENSVINAKIA</u>	<u>DAEALLEKVT</u>	<u>DPSIRQNAME</u>	<u>TLTGLKSSLL</u>	<u>LGTKDNNTIS</u>
<u>AEVDSLLALL</u>	<u>KESQPAPIQ</u>			

Figure 5.15: Complete mass-spectrometry results for PhtD C-term fragment

Identified peptides are highlighted in blue. Peptides identified by MS-MS are underlined.

N.B. MS-MS identified peptides may duplicate PMF recovered peptides. (Italicised residues show start and end of fragment).

SYELGRHQAG	QVKKESNRVS	YIDGDQAGQK	AENLTPDEVS	KREGINAEQI
VIKITDQGYV	TSHGDHYHY	NGKVPYDAII	SEELLMKDPN	YQLKSDIVN
EIKGGYVIKV	DGKYVYVLKD	AAHADNIRTK	EEIKRQKQEH	SHNHGGGSND
QAVVAARAQG	RYTTDDGYIF	NASDIIEDTG	DAYIVPHGDH	YHYIPKNELS
ASELAAAEAY	WNGKQGSRPS	SSSSYNANPA	QPRLSENHNL	TVTPTYHQNQ
GENISSLLRE	LYAKPLSERH	VESDGLIFDP	AQITSRTARG	VAVPHGNHYH
FIPYEQMSEL	EKRIARI IPL	RYRSNHWVPD	<u>SRPEQPSPQS</u>	<u>TPEPSPSPQP</u>
<u>APNPQPAPSN</u>	<u>PIDEKLVKEA</u>	<u>VRKVG DGYVF</u>	<u>EENGVSRYIP</u>	<u>AKDLSAETAA</u>
<u>GIDSKLAKQE</u>	<u>SLSHKLGAKK</u>	<u>TDLPS DREF</u>	<u>YNKAYDLLAR</u>	<u>IHQDLLDNKG</u>
<u>RQVDFEALDN</u>	<u>LLERLKD VPS</u>	<u>DKVKLVDDIL</u>	<u>AFLAPIR HPE</u>	<u>RLGKPN AQIT</u>
<u>YTDDEIQVAK</u>	<u>LAGKYTTEDG</u>	<u>YIFDPRDITS</u>	<u>DEGDAYVTPH</u>	<u>MTHSHWIKKD</u>
<u>SLSEAERAAA</u>	<u>QAYAKEKGLT</u>	<u>PPSTDHQDSG</u>	<u>NTEAKGAEAI</u>	<u>YNRVKA AKKV</u>
<u>PLDRMPYNLQ</u>	<u>YTVEVKNGLS</u>	<u>IIPHYDHYHN</u>	<u>IKFEWFDEGL</u>	<u>YEAPKGYTLE</u>
<u>DLLATVKYYV</u>	<u>EHPNERPHSD</u>	<u>NGFGNASDHV</u>	<u>RKNKVDQDSK</u>	<u>PDEDKEHDEV</u>
<u>SEPTHPESDE</u>	<u>KENHAGLNPS</u>	<u>ADNLYKPSTD</u>	<u>TEETEEEAED</u>	<u>TTDEAEIPQV</u>
<u>ENSVINAKIA</u>	<u>DAEALLEKVT</u>	<u>DPSIRQNAME</u>	<u>TLTGLKSSLL</u>	<u>LGTKDNNTIS</u>
<u>AEVDSLLALL</u>	<u>KESQPAPIQ</u>			

Figure 5.16: Complete mass-spectrometry results for Δ C-term PhtD fragment

Identified peptides are highlighted in blue. Peptides identified by MS-MS are underlined. N.B. MS-MS identified peptides may duplicate PMF recovered peptides. (Italicised residues show start and end of fragment).

As a result of the extensive analysis through the use of 3 different proteases, coupled with the two mass-spectrometry techniques, it was possible to recover ~90% of the peptides present in each of the purified PhtD proteins, and showed definitively that the proteins had been correctly expressed without any mutations or unexpected alterations.

Mass-spectrometry analysis of the 34 kDa PhtD fragment was kindly performed with the assistance of Dr. Richard Burchmore at the University of Glasgow, as there was no stock of this protein fragment available during the time at GlaxoSmithKline when the previous proteins were analysed. 5 μ g of the 34 kDa PhtD protein fragment was run on an SDS-PAGE gel, stained and the band excised as before. The excised sample was digested with trypsin and analysed by PMF. The resulting mass-spectrometry data was analysed using MASCOT.¹⁶² The interpreted data of recovered peptides mapped onto the PhtD sequence is shown in figure 5.17.

```

SYELGRHQAG QVKKESNRVS YIDGDQAGQK AENLTPDEVS KREGINAEQI
VIKITDQGYV TSHGDHYHY NGKVPYDAII SEELLMKDPN YQLKSDIVN
EIKGGYVIKV DGKYYVYLKD AAHADNIRTK EEIKRQKQEH SHNHGGSND
QAVVAARAQG RYTDDGYIF NASDIIDTG DAYIVPHGDH YHYIPKNELS
ASELAAAEAY WNGKQGSRPS SSSSYNANPA QPRLSENHNL TVTPTYHQNQ
GENISSLLRE LYAKPLSERH VESDGLIFDP AQITSRTARG VAVPHGNHYH
FIPYEQMSEL EKRIARI IPL RYRSNHWVVD SRPEQSPQS TPEPSPSPQP
APNPQPAPSN PIDEKLVKEA VRKVG DGYVF EENGVSRYIP AKDLSAETAA
GIDSKLAKQE SLSHKLGAKK TDL P SSDREF YNKAYDLLAR IHQDLLDNKG
RQVDFEALDN LLERLKDVPS DKVKLVDDIL AFLAPIRHPE RLGKPNQAQIT
YTDEIQVAK LAGKYTTEDG YIFDPRDITS DEGDAYVTPH MTHSHWIKKD
SLSEAEAAA QAYAKEKGLT PPSTDHQDSG NTEAKGAEAI YNRVKAACKV
PLDRMPYNLQ YTVEVKNGSL IIPHYDHYHN IKFEWFDEGL YEAPKGYTLE
DLLATVKYYV EHPNERPHSD NGFGNASDHV RKNKVDQDSK PDEDKEHDEV
SEPTHPESDE KENHAGLNPS ADNLYKPSTD TEETEEEAED TTDEAEIPQV
ENSVINAKIA DAEALLEKVT DPSIRQNAME TLTGLKSSLL LGTKDNNTIS
AEVDSLALL KESQPAPIQ

```

Figure 5.17: Mass-spectrometry results for 34 kDa PhtD protein fragment

Identified peptides are highlighted in blue. (Italicised residues show start and end of fragment).

N.B. Sequence coverage is lower due to the use of 1 enzyme (trypsin) compared to 3 for the other samples (trypsin, chymotrypsin, Asp-N).

Due to the use of only one enzyme rather than three, coverage of recovered peptides was less than before, but still gave enough evidence that this protein fragment had also been correctly expressed without any unexpected alterations.

5.4 Analysis Of PhtD Limited Proteolysis Profiles By N-Terminal Sequencing

5.4.1 Overview

N-terminal protein sequencing was kindly performed with the assistance of Neil Freeman at GlaxoSmithKline Biologicals, UK. As a complimentary approach to the mass-spectrometry analysis for the limited proteolysis profiles of the F/L PhtD protein, N-terminal sequencing was employed to identify the starting sequence of each of the bands produced by the Apo-proteolysis experiment. Additionally, the N-terminal sequencing technique

was also used to identify the starting sequence of bands from the Zn^{2+} -bound samples to allow comparison between the Apo- and Zn^{2+} -bound datasets.

5.4.2 Sample Preparation

Using the SDS-PAGE samples generated from the F/L PhtD Apo- and Zn^{2+} -bound limited proteolysis experiments, samples were prepared for N-terminal sequencing as outlined in 2.10.7 N-terminal sequencing was performed on a Procise[®] protein sequencing system (Applied Biosystems) utilising the Edman degradation method.¹³¹ The protein on the PVDF membrane is washed with Edman reagent (phenylisothiocyanate); using anhydrous acid the terminal amino residue is cleaved as an unstable t-amino acid. The t-amino acid form is treated with aqueous acid to form a stable p-amino acid form. This single stable peptide is then analysed by High Performance Liquid Chromatography (HPLC) and the process repeated cyclically. N-terminal sequencing was set up to read 7 amino acids for each sample. A total of 11 bands were sequenced; 7 from the blot of the apo-digest and 4 from the blot of the Zn^{2+} -bound digest. The stained membranes of the two blotted digests are illustrated in figures 5.18 and 5.19. Samples used for analysis are indicated on the relevant membranes.

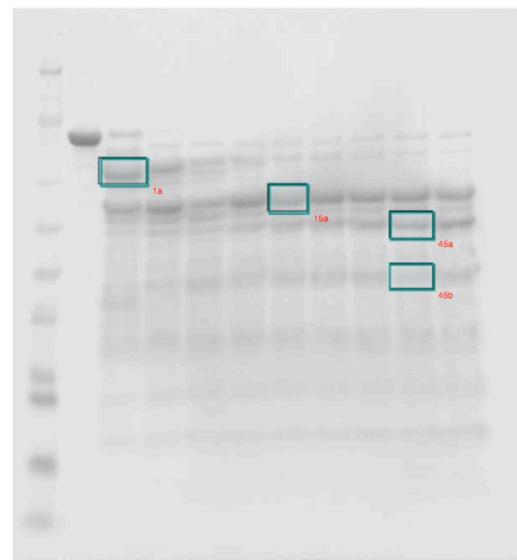
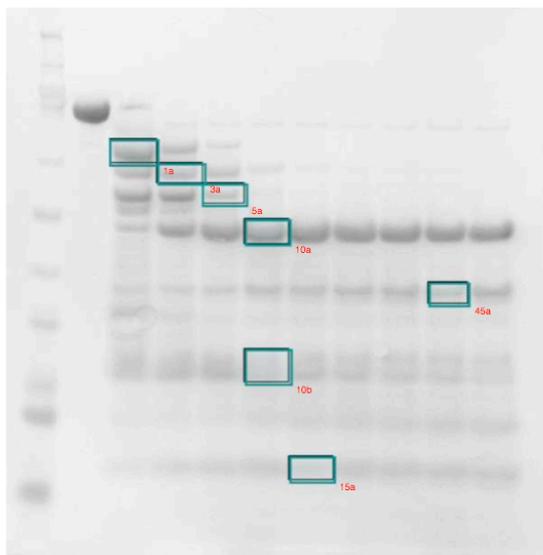


Figure 5.18: Western transfer for Apo- F/L PhtD digest

Bands analysed by N-terminal sequencing are highlighted in boxes.

Figure 5.19: Western transfer for Zn^{2+} -bound F/L PhtD digest

Bands analysed by N-terminal sequencing are highlighted in boxes.

5.4.3 N-terminal sequencing Analysis Of Apo- F/L PhtD And Zn²⁺-bound F/L PhtD

Table 5.3 presents the sequence identification of the bands selected for N-terminal sequencing of Apo- F/L PhtD. The identified sequences are mapped onto the amino acid sequence for F/L PhtD in figure 5.20

Band Identity	Sequence Observed	pmol
1a	LSEHNHL	3.90
3a	SNHWVPD	1.40
5a	LSEHNHL	2.60
10a	SNHWVPD	9.90
45a	LAKQESL	2.10
	SNHWVPD	1.70
10b	VSYIDGD	7.60
	KNKVDQ	6.60
	NKVDQD	6.00
15a	LSEHNHL	13.80
	AQGRYTT	5.00
	EGINAEQ	0.70

Table 5.3: N-terminal sequencing results of Apo- F/L PhtD

Sequence identities of bands N-terminally sequenced from Apo- F/L PhtD limited proteolysis experiment. Band identities correspond to those given in figure 5.18. Larger molecular weight fragments show either LSEHNHL or SNHWVPD as starting sequences for stable regions. Smaller molecular weight bands show a variety of sequences indicating that the band consists of a variety of breakdown products from the digestion process.

SYELGRHQAG	QVKKESNRVS	<u>YIDGDQAGQK</u>	AENLTPDEVS	KREGINAEQI
VIKITDQGYV	TSHGDHYHY	NGKVPYDAII	SEELLMKDPN	YQLKSDIVN
EIKGGYVIKV	DGKYVYVYLD	AAHADNIRTK	EEIKRQKQEH	SHNHGGGSND
QAVVAARAQG	<u>RYTTDDGYIF</u>	NASDIIEDTG	DAYIVPHGDH	YHYIPKNELS
ASELAAAEAY	WNGKQGSRPS	SSSSYNANPA	QPR <u>LSEHNHL</u>	TVTPTYHQNQ
GENISLLRE	LYAKPLSERH	VESDGLIFDP	AQITSRTARG	VAVPHGNHYH
FIPYEQMSEL	EKRIARIIPL	<u>RYRSNHWVPD</u>	SRPEQPSPQS	TPEPSPSPQP
APNPQPAPSN	PIDEKLVKEA	VRKVG DGYVF	EENGVSRYIP	AKDLSAETA
GIDSKLAKQE	<u>SLSHKLGAKK</u>	TDLPSSDREF	YNKAYDLLAR	IHQDLLDNKG
RQVDFEALDN	LLERLKDVP	DKVKLVDDIL	AFLAPIRHP	RLGKPNQIT
YTDDEIQVAK	LAGKYTTEDG	YIFDPRDITS	DEGDAYVTPH	MTHSHWIKKD
SLSEAERAAA	QAYAKEKGLT	PPSTDHQDSG	NTEAKGAEAI	YNRVKAARKV
PLDRMPYNLQ	YTVEVKNGSL	IIPHYDHYHN	IKFEWFDEGL	YEAPKGYTLE
DLLATVKYYV	EHPNERPHSD	NGFGNASDHV	<u>RKNKVDQDSK</u>	PDEDKEHDEV
SEPTHPESDE	KENHAGLNPS	ADNLYKPSTD	TEETEEEAED	TTDEAEIPQV
ENSVINAKIA	DAEALLEKVT	DPSIRQNAME	TLTGLKSSLL	LGTKDNNTIS
AEVDSLLALL	KESQPAPIQ			

Figure 5.20: Apo-F/L PhtD N-term sequencing results mapped onto F/L PhtD sequence

Identified N-terminal protein sequences are highlighted in blue. Underlined residues illustrate sequences corresponding to larger molecular weight bands. Dominant stable band 10a begins with the sequence SNHWVPD. Smaller molecular weight bands consist of a mixture of the remaining sequences, indicating their presence as smaller breakdown products as a result of proteolysis.

It is apparent that the stable bands from the Apo-limited proteolysis experiment begin with the sequence SNHWVPD. The smaller molecular weight bands comprise a mixture of fragments with various N-terminal sequences, which identify them as proteolytic breakdown products of the unstructured N-terminal half of the F/L PhtD molecule.

The identities of the bands N-terminally sequenced from the limited proteolysis of F/L PhtD in the presence of Zn^{2+} are given in table 5.4. The identified N-terminal sequences are mapped onto the F/L PhtD amino acid sequence in figure 5.21.

Band Identity	Sequence Observed	pmol
1a	LSEHNHL	5.80
15a	LSEHNHL	8.30
45a	SNHWVPD	6.10
45b	LAKQESL	4.40

Table 5.4: N-terminal sequencing results for Zn^{2+} -bound F/L PhtD

Sequence identities of bands N-terminally sequenced from Zn^{2+} -bound F/L PhtD limited proteolysis experiment. Band identities correspond to those given in figure 5.19.

Products of the proteolysis experiment are much more resistant to degradation in Zn^{2+} -bound form, with the majority of the proteins stabilising at a higher molecular weight than in the Apo- proteolysis experiment. The dominant stable band 15a begins with sequence LSEHNHL and is of a higher molecular weight than the dominant stable band in the Apo-experiment.

SYELGRHQAG	QVKKESNRVS	YIDGDQAGQK	AENLTPDEVS	KREGINAEQI
VIKITDQGYV	TSHGDHYHY	NGKVPYDAII	SEELMKDPN	YQLKDSDIVN
EIKGGYVIKV	DGKYVYVLKD	AAHADNIRTK	EEIKRQKQEH	SHNHGGGSND
QAVVAARAQG	RYTTDDGYIF	NASDIIEDTG	DAYIVPHGDH	YHYIPKNELS
ASELAAAEAY	WNGKQGSRPS	SSSSYNANPA	QPR <u>LSEHNHL</u>	TVTPTYHQNQ
GENISSLLRE	LYAKPLSERH	VESDGLIFDP	AQITSRTARG	VAVPHGNHYH
FIPYEQMSEL	EKRIARIIPL	RYR <u>SNHWVPD</u>	SRPEQPSPQS	TPEPSPSPQP
APNPQPAPSN	PIDEKLVKEA	VRKVG DGYVF	EENGVSRYIP	AKDLSAETAA
GIDSKLAKQE	<u>SLSHKLGAKK</u>	TDLPSSDREF	YNKAYDLLAR	IHQDLLDNKG
RQVDFEALDN	LLERLKDVP	DKVKLVDDIL	AFLAPIRHPE	RLGKPNQAQIT
YTDDEIQVAK	LAGKYTTEDG	YIFDPRDITS	DEGDAYVTPH	MTHSHWIKKD
SLSEAERAAA	QAYAKEKGLT	PPSTDHQDSG	NTEAKGAEAI	YNRVKAACKV
PLDRMPYNLQ	YTVEVKNGSL	IIPHYDHYHN	IKFEWFDEGL	YEAPKGYTLE
DLLATVKYVV	EHPNERPHSD	NGFGNASDHV	RKNKVDQDSK	PDEDKEHDEV
SEPTHPESDE	KENHAGLNPS	ADNLYKPSTD	TEETEEEAED	TTDEAEIQV
ENSVINAKIA	DAEALLEKVT	DPSIRQNAME	TLTGLKSSLL	LGTKDNNTIS
AEVDSLLALL	KESQPAPIQ			

Figure 5.21: Zn²⁺-bound F/L PhtD N-term sequencing results mapped onto F/L PhtD sequence

Identified N-terminal protein sequences are highlighted in blue. Underlined residues illustrate sequences corresponding to larger molecular weight bands. Dominant stable band 15a begins with the sequence LSEHNHL.

Direct comparison between the Apo- and Zn²⁺-bound N-terminal sequencing experiments shows that upon addition of Zn²⁺ proteolysis profile changes, with a much lower abundance of smaller molecular weight products present in the Zn²⁺-bound experiment. Additionally, there is a change in the proportion of the stable bands seen in the two experiments; the two stable bands from the Apo- experiment had a molecular weight of 49 kDa and 34 kDa (10a and 45a). The dominant band 10a had the N-terminal sequence SNHWVPD. In contrast in the Zn²⁺-bound experiment, the two stable bands are 55 kDa and 49 kDa (15a and 45a). The dominant band 15a is of a larger molecular weight than in the Apo- experiment and has the N-terminal sequence LSEHNHL, which is closer to the N-terminus of the F/L molecule than SNHWVPD. This concurs with the limited proteolysis and mass-spec results that the binding of Zn²⁺ in F/L PhtD is having a stabilising effect on the protein, protecting the protein from the action of proteolysis. It is thought that this protection is a result of the change in structure induced by the addition of Zn²⁺, as observed in the CD experiments (detailed in chapter 4.4).

5.5 ITC Analysis Of Zn²⁺ Binding With PhtD

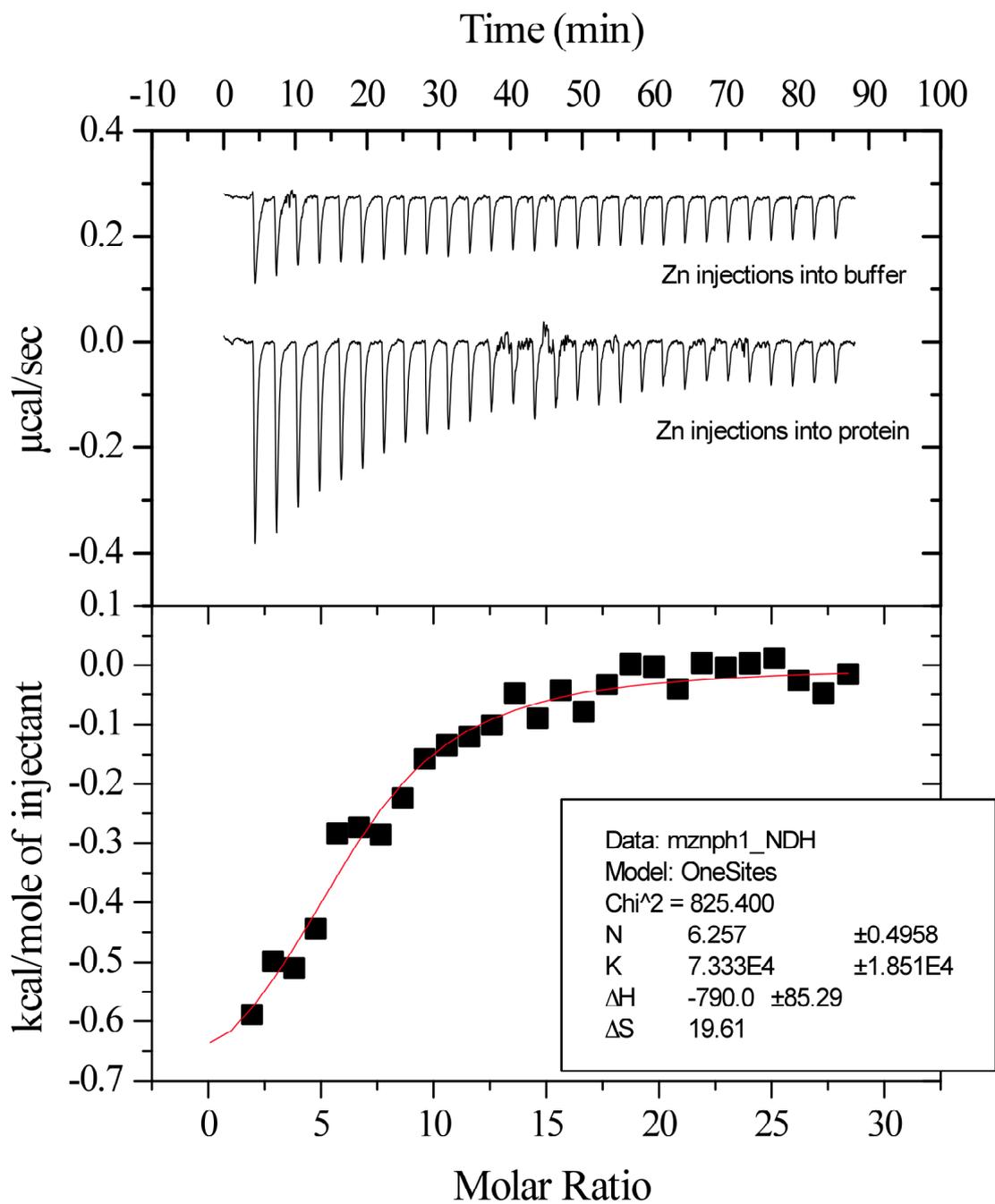
ITC was performed at the department of physical chemistry, University of Glasgow with the kind assistance of Mrs. Margaret Nutley.

As previous investigation at the University of Glasgow had revealed through macromolecular X-ray crystallography that PhtA is able to bind Zn²⁺ atoms via the HxxHxH motifs which give the Pht protein family their name, and experiments discussed elsewhere in this thesis show that PhtD is also able to bind Zn²⁺ atoms resulting in extensive structural reorganisation,⁹⁸ Isothermal Titration micro-Calorimetry (ITC) was employed in an attempt to reveal the stoichiometric relationship between PhtD and the Zn²⁺ atoms. It was hypothesised based on the evidence from the partial PhtA crystal structure that each HxxHxH motif would be able to bind 1 Zn²⁺ atom, and that the number of binding sites would correlate to the number of HxxHxH motifs present in the protein sequence. As PhtD contains 5 HxxHxH motifs, it was therefore postulated that ITC would return a stoichiometry value of 5 Zn²⁺ per PhtD molecule.

2mg of purified F/L PhtD protein was used for analysis in 1ml of 20mM Tris HCl pH7.5, 200mM NaCl, which equated to a concentration of ~20 μ M. This was diluted 1:1 with the same buffer to give a final volume of 2ml (required to fill the sample-cell of the calorimeter) with a concentration of ~10 μ M (when measured, protein concentration was found to be 9 μ M). 0.5M ZnSO₄ was prepared from the same buffer stock as the protein in order to minimise differences between solutions when titration was performed.

ITC was performed using a VP-ITC Isothermal Titration Calorimeter (MicroCal, GE Healthcare, Buckinghamshire UK). The instrument was cleaned by purging multiple times with buffer from the same stock used to prepare the protein (F/L PhtD) and ligand (ZnSO₄). A blank titration was carried out by injecting the ligand into buffer without any protein in order to obtain a background reading, which could then be subtracted from the experimental titration during analysis to factor in contributions from the background signal. Titration was carried out at 25°C. 10 μ l of ZnSO₄ solution at a concentration of 1.16mM (dilution was carried out with the same stock of protein buffer) was

injected into the 2ml solution of F/L PhtD at a concentration of $9\mu\text{M}$ every 3mins for an experimental duration of 90mins and each exothermic heat-pulse recorded. Data were analysed and integrated with a one-sites model and are illustrated in figure 5.22.



Date: 05/11/2010
 File: M1Znph1.opj
 Origin v. 5.046

[Syringe] = 1.16 mM
 [Cell] = 0.00902 mM
 T = 24.99631 °C

Figure 5.22: ITC analysis of PhtD Zn²⁺ binding

Upper-panel: Exothermic heat-pulse raw data.

Lower panel: Integrated heat data. Background has been accounted for.

From the raw data, it is evident that binding is occurring, as an increased exothermic heat pulse was observed compared to the blank reading. As expected, a gradual decrease in exothermic heat pulse was observed, due to decreasing free binding sites in the protein over time due to Zn^{2+} uptake. The values recorded for the exothermic heat pulse readings are relatively small at a maximum of $\sim 0.4 \mu\text{cal}/\text{sec}$ (compared to Hen Egg White Lysozyme binding tri-N-acetyl-glycosamine with a reading of $\sim 1.5 \mu\text{cal}/\text{sec}$).¹⁶³ Due to the fact that a small value such as this was observed with a 100-fold increase of ligand to protein ($1.16 \text{mM ZnSO}_4 : 9 \mu\text{M F/L PhtD}$) coupled with the observation in the raw data that there was no definitive sharp decrease in heat-pulse recorded, it was postulated that the interaction between PhtD and Zn^{2+} was through weak binding. This was confirmed upon integration of the raw heat-data (figure 5.22, lower panel). Analysis of the integrated data with a one-sites standard model resulted in a binding affinity $K_{\text{ass}} = 7.3 \times 10^4 \text{ M}^{-1} \pm 1.9 \times 10^4$ with the dissociation value $K_{\text{diss}} (=) 1/K_{\text{ass}} = 14 \mu\text{M}$ confirming as hypothesised that the binding between PhtD and Zn^{2+} was weak. The stoichiometry value $N = 6.3 \pm 0.50$ confirms the presence of multiple binding sites, and suggests that there may be 6 Zn^{2+} binding sites per PhtD molecule. However, as the association between protein and ligand is through weak binding, no definitive point of saturation can be seen leading to decreased accuracy when N is calculated, and it is therefore quite possible that the original hypothesis of 5 Zn^{2+} binding sites per PhtD molecule is correct.

It is interesting to compare this occurrence of zinc-binding in PhtD with other known zinc-binding proteins. The binding of zinc atoms by PhtD is somewhat unusual, as a weak interaction is observed between the protein and ligand; in contrast, other known zinc-binding proteins illicit a strong interaction between protein and ligand (ie tight binding) such as the ABC transporters ZnuA and TroA.¹⁶⁴ In instances such as these, the binding of zinc atoms may be required to 'lock' a protein into a specific conformation in order to allow transport of substances across the cell membrane. Similarly, zinc-dependant metalloproteases may require a high affinity to Zn-atoms in order to remain active when performing proteolytic activity. The binding of zinc atoms may also act as a stimulus or regulator of a biological process and therefore high-affinity binding may be advantageous in order for the organism to keep the relevant pathway active or inactive (depending on whether the zinc atom

acts as a promoter or repressor). In contrast, the incidence of weak binding observed by PhtD fits well with the hypothesis of the protein acting as a siderophore (metal scavenger - for use by the bacterium under stress environments), as the bacterium would be required expend less energy in order to free the scavenged zinc atoms when required, a clear advantage for the hypothetical function of PhtD as a siderophore, compared to the large amount of energy that would need to be expended in order to free a tightly bound zinc atom in an instance where the bacterium is already disadvantaged by its environment.

Chapter 6 : Structural Analysis Of PhtD

6.1 Overview

This chapter details an attempt to uncover the biological function of PhtD using a structural approach. The primary method employed for this investigation was macromolecular x-ray crystallography, due to previous success in determining the structure for a section of Pneumococcal Histidine Triad A (PhtA) from the same protein family.⁹⁸ Additionally the techniques of Nuclear Magnetic resonance (NMR), Circular Dichroism (CD) and Analytical Ultra-Centrifugation (AUC) were also utilised to gain as much of an insight into the structure of PhtD as possible, and are also described in this chapter.

6.2 Preliminary Investigation

Before any experiments were performed, the F/L PhtD protein sequence was input into the disorder prediction tool Regional Order Neural Network (RONN - <http://www.strubi.ox.ac.uk/RONN>)¹³⁹ hosted at the Division of Structural Biology (STRUBI) at the University of Oxford.

Where two proteins have a similar biological function, then the primary sequences of the two molecules usually elicit a significant similarity, which can be detected by sequence alignment programs such as BLAST.¹⁶⁵ Based on this observation, RONN utilises a 'bio-basis neural network pattern recognition' algorithm¹⁶⁶ in order to predict the probability of order or disorder for regions of an input protein sequence. This 'neural network' is trained using an extensive dataset (in the case of RONN, generated from a full annotation of entries in the Molecular Structure Database hosted at the European Bioinformatics Institute).¹⁶⁷ It is expected that the biological function of the query sequence will be the same as that of a protein sequence of known function if the alignment score is high enough. RONN implements this to predict order or disorder in a query protein sequence by performing alignments of the query sequence with those of known folding states; the resulting alignment homology scores are used for statistical pattern recognition to classify the query sequence as ordered or disordered

using the trained 'neural network', which are then averaged to give disorder probability scores for each individual residue in the query sequence.¹³⁹ The RONN program outputs the results of protein disorder prediction in graphical format.

The resulting disorder prediction graph for F/L PhtD is shown in figure 6.1. The fragments C-term PhtD, Δ C-term PhtD and the 34 kDa fragment were also subjected to disorder prediction analysis using RONN; the resulting disorder prediction graphs are illustrated in figures 6.2, 6.3 and 6.4.

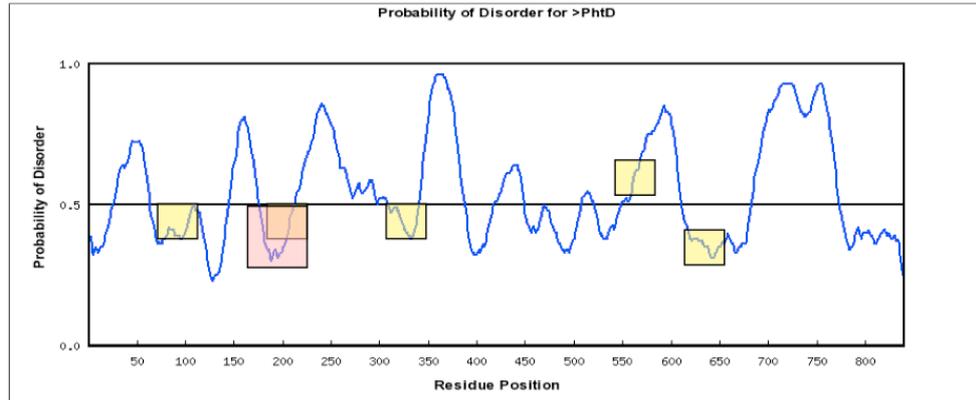


Figure 6.1: F/L PhtD disorder prediction

Disorder prediction generated using the RONN server.¹³⁹ Probability of disorder increases as value approaches 1.0. The location of each HxxHxH motif is boxed out in yellow. The location of the corresponding known crystal structure from PhtA is boxed out in pink.

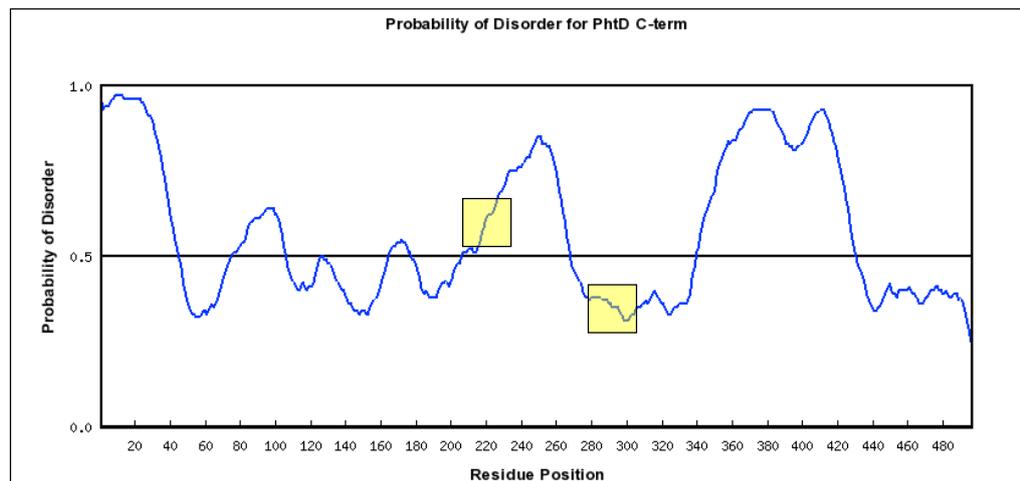


Figure 6.2: C-term PhtD disorder prediction

Disorder prediction generated using the RONN server.¹³⁹ Probability of disorder increases as value approaches 1.0. HxxHxH motifs are boxed out in yellow.

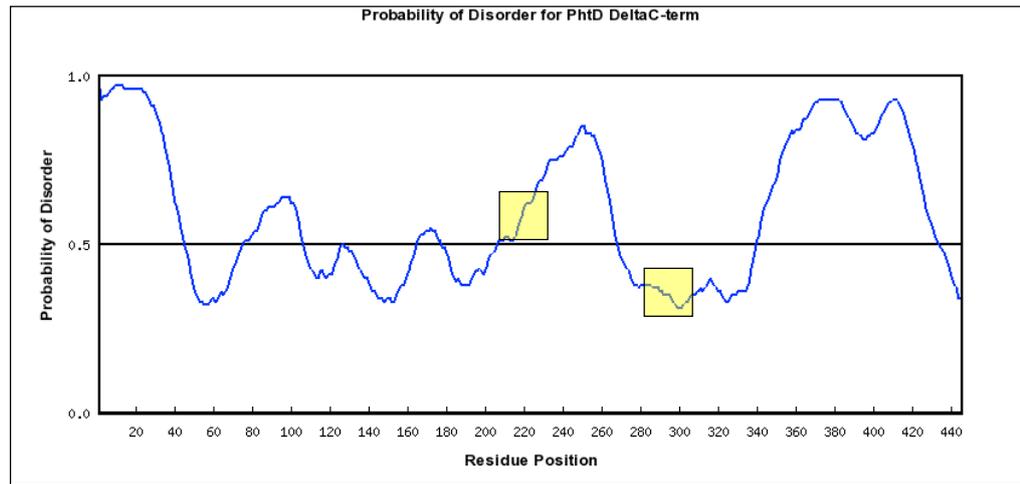


Figure 6.3: Δ C-term PhtD disorder prediction

Disorder prediction generated using the RONN server.¹³⁹ Probability of disorder increases as value approaches 1.0. HxxHxH motifs are boxed out in yellow.

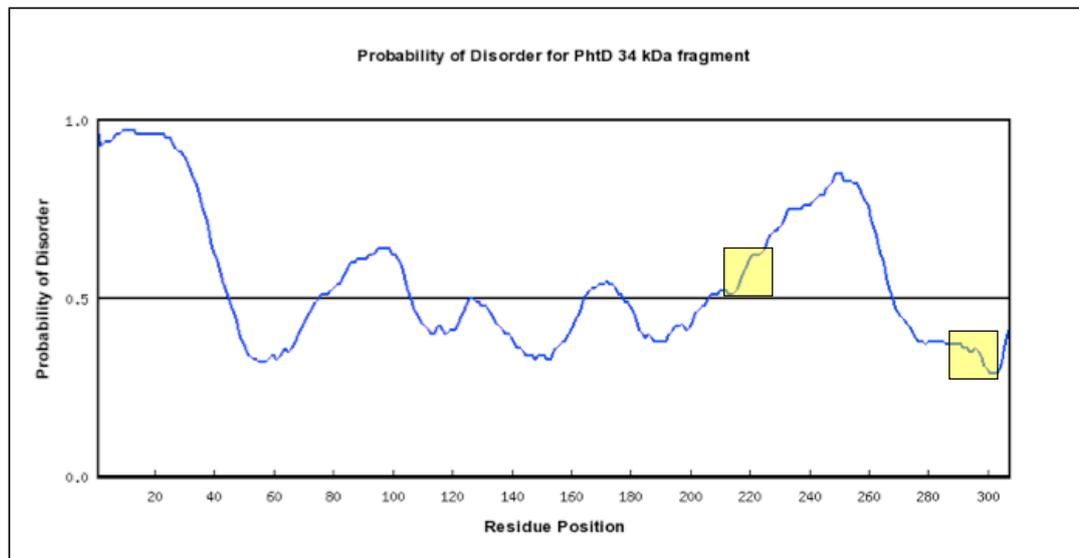


Figure 6.4: 34 kDa PhtD fragment disorder prediction

Disorder prediction generated using the RONN server.¹³⁹ Probability of disorder increases as value approaches 1.0. HxxHxH motifs are boxed out in yellow.

Residue number is shown along the X-axis, with probability of disorder shown along the Y-axis. Probability is measured on a scale of 0.0 to 1.0 where 0.0 indicates a probability that the analysed residue is highly ordered and 1.0 indicates a probability of the analysed residue being natively unordered. A threshold of 0.5 is imposed by the program, with anything <0.5 classed natively structured, and anything >0.5 classed natively unordered. Looking at the results of the RONN prediction for PhtD, it can be seen that a large

proportion of the protein (~60%) falls above the set 0.5 threshold and is predicted unstructured by the RONN program. It is apparent from the data that the protein appears to contain extensive regions of natively unordered protein interspersed with short regions of structured protein. However, as the program only uses sequence data applied to an algorithm, it should be noted that the results of the RONN analysis can only give an indication of how the protein may behave. It is unable to take into account other factors such as the incorporation of ligand-binding -in the case of PhtD, Zn²⁺ atoms- and the change in structure which the binding process may induce. As a starting point however, it indicated that determining the structure of the F/L PhtD protein may be difficult due to the large flexible regions apparently present, thus preventing the generation of an ordered protein lattice, which is required for crystal formation.

6.3 One-Dimensional Nuclear Magnetic Resonance (1D-NMR) Spectroscopy Of PhtD

6.3.1 Initial 1D-NMR Analysis Of Apo- F/L PhtD

NMR spectroscopy was done in collaboration with Dr. Brian Smith at the Protein Science Group, University of Glasgow.

In an attempt to obtain some physical evidence to relate to the preliminary indication by sequence analysis with RONN that PhtD may be relatively unordered in its apo- form, 1D-NMR was utilised in order to ascertain the nature of Apo- F/L PhtD in solution. 1mg of F/L PhtD was used at a concentration of 1mg/ml. 1D-NMR was carried out on this sample in 1xDulbecco's PBS buffer. The resulting spectrum is shown in figure 6.5.

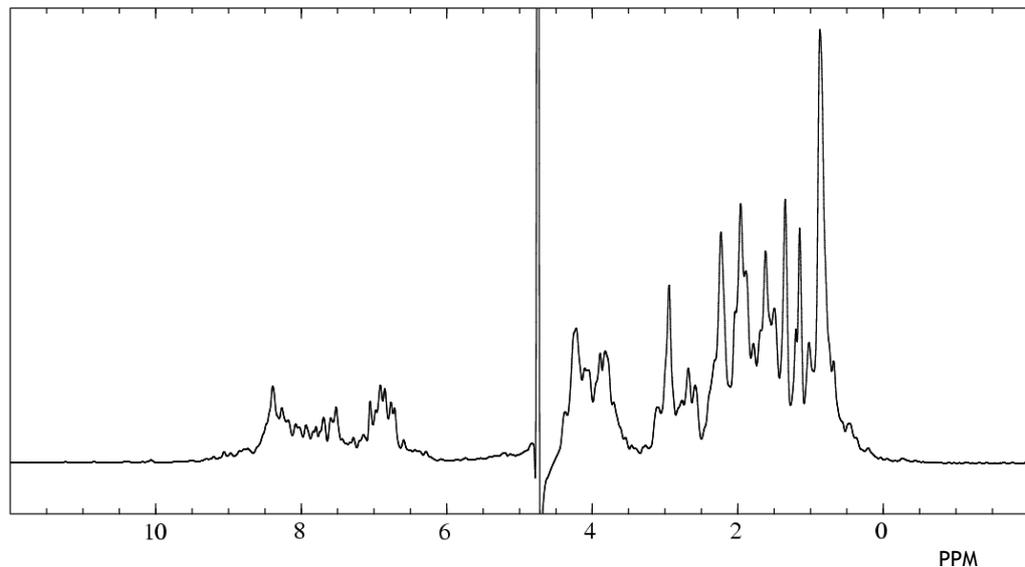


Figure 6.5: 1D-NMR Spectrum of Apo- F/L PhtD

There are two easily discernible discrete regions in the spectrum; the first region from 0 - 5ppm is defined the aliphatic region and corresponds to signals generated by CH_3 groups. The second region, which is present in the spectrum between 6 - 9ppm is classed as the amide region, and as such corresponds to signals generated by NH_2 groups. As 1D-NMR does not require the protein to be labelled in any way (unlike 2D-NMR which requires the protein of interest to be ^{15}N and ^{13}C labelled) it is a very quick way of determining rudimentary information about a protein. However, the NMR spectrometer does require a 'lock on' signal during experimental setup and therefore a small quantity of D_2O (deuterium) is added to the sample to be analysed. The D_2O appears as a very strong peak at $\sim 4.75\text{ppm}$ and is evident in the spectrum for the Apo- F/L PhtD sample analysed above.

In a highly structured protein, the extreme ends of the aliphatic and amide regions show discrete, sharp peaks due to the restricted movement of the residues resulting from intra-molecular bonding. Conversely, in a natively unstructured protein, these regions broaden out showing no peaks, as the residues in the protein are able to move about and so give a less well-defined spectrum. The differences in spectra generated from varying grades of protein are illustrated in the reference spectra shown in figure 6.6.

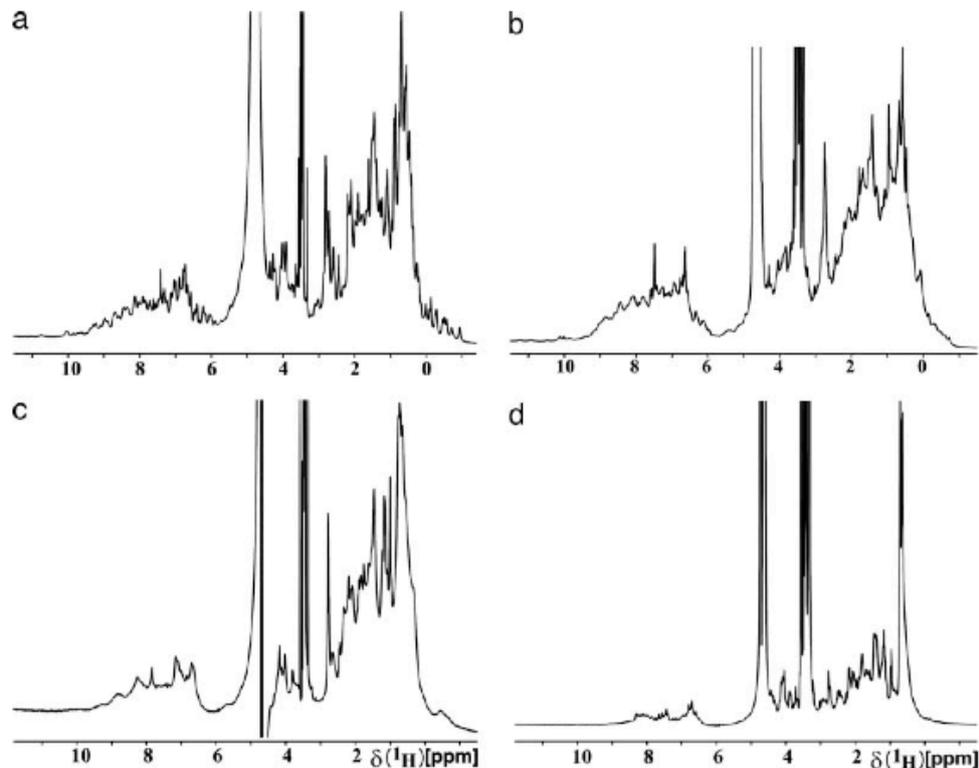


Figure 6.6: Reference 1D NMR spectra

Reference spectra for various grades of protein classed “A” to “D”. (a) High quality “A”-grade proteins were candidates for structure determination by NMR or X-ray crystallography, (b) “B”-grade proteins were earmarked for X-ray crystallography, (c) “C”-grade indicated folded globular proteins and (d) “D”-grade are nonglobular, unfolded polypeptides. Figure reproduced from Page et al (2005).¹⁶⁸

In the spectrum generated for the Apo- F/L PhtD protein, the edges of the aliphatic and amide regions of the spectrum show small peaks to be present, reminiscent of the spectrum for “B” or “C”-grade protein suggesting that although the Apo- PhtD protein appears largely unstructured, it should be a globular protein, and have small regions which are highly structured. These results concur with the RONN¹³⁹ sequence analysis data, giving rise to the analogy that PhtD may be thought of as having a ‘beads on a string’ appearance, where localised, highly structured regions of the protein are connected by regions of random coil, allowing a structured region to act independently of the others or flex around and interact with one or more of the other structured regions. The presence of structured regions within the protein has been shown through the use of limited proteolysis experiments, and that the addition of Zn²⁺ causes a stabilising effect thought to be brought about by a change in protein structure upon ligand-binding (Limited proteolysis results are discussed in chapter 5).

6.3.2 1D-NMR Analysis Of Zn²⁺ Binding

In order to investigate the effect that the binding of Zn²⁺ conferred on PhtD, 1D-NMR was performed on purified F/l PhtD protein, and also the purified C-terminal and ΔC-terminal fragments. In each case NMR was carried out as before in 4.3.1 to give an apo- spectrum. A source of Zn²⁺ in the form of ZnSO₄ was titrated in, the sample was re-examined in the NMR spectrometer and spectra were generated after each titration. Comparison of the sequential spectra allowed any changes in protein structure to be observed. The two PhtD fragments above were analysed to provide a comparison with the F/L to see if changes were as evident in the C-terminal half of PhtD compared to the F/L molecule.

In all experiments, a spectrum of the Apo-protein was generated, the sample was recovered from the NMR tube and the required quantity of Zn²⁺ introduced, and a subsequent Zn²⁺-bound spectrum was generated. This procedure was repeated for all titrations of each protein. In all cases, the same ZnSO₄ solution was used for titration of Zn²⁺ atoms into each sample.

PhtD has 5 HxxHxH Zn²⁺ binding motifs, and it is thought that each motif binds 1 Zn²⁺ atom as seen in the PhtA crystal structure⁹⁸, giving a ratio of 5 Zn²⁺ atoms per 1 copy of the PhtD protein. To this end, an initial addition of ZnSO₄ at 5x the concentration of PhtD sample was used in order to provide the required 5:1 molar ratio, giving the required 1 Zn²⁺ atom per HxxHxH binding motif. A subsequent addition of an equivalent concentration of ZnSO₄ was performed in order to double the molar ratio, resulting in 2 Zn²⁺ atoms per HxxHxH motif and any further effects noted.

6.3.2.1 Zn²⁺ Binding in Purified F/L PhtD

Initially, a 1D-NMR spectrum was generated for the Apo protein using 0.3mM F/L PhtD. A spectrum was then generated for 0.3mM F/L PhtD with 1.5mM ZnSO₄. Finally, a further 1.5mM ZnSO₄ was added, giving a 3mM ZnSO₄ concentration and doubling the molar ration of Zn²⁺ : F/L PhtD. The 3 resulting spectra were overlaid for analysis as seen in figure 6.7.

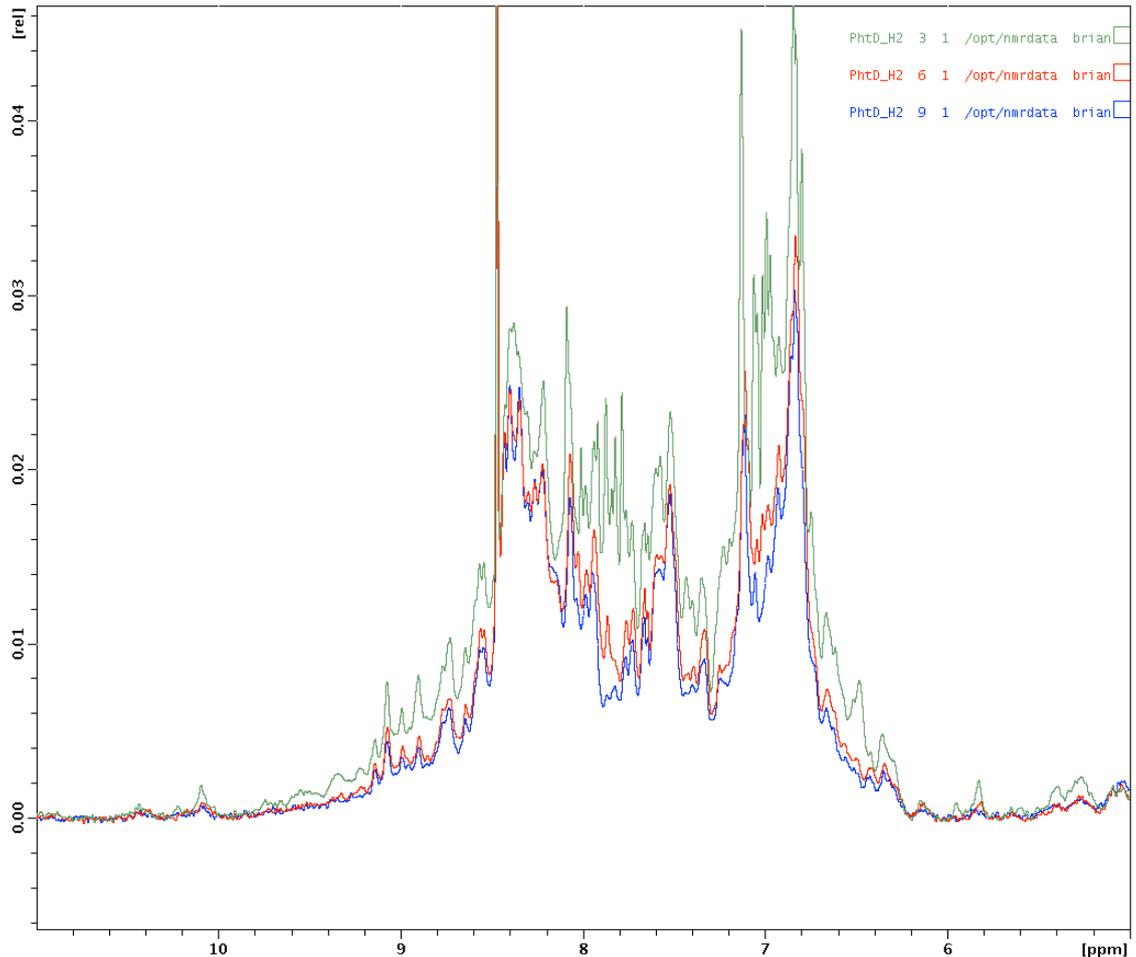


Figure 6.7: F/L PhtD Zn^{2+} titration

The image above shows the amide region for the 1D-NMR spectra.

Green: Apo- F/L PhtD **Red:** F/L PhtD + 1.5mM ZnSO_4 **Blue:** F/L PhtD + 3mM ZnSO_4

The green spectrum corresponds to a close-up view of the amide region from figure 6.5. Upon initial addition of Zn^{2+} (red spectrum) a significant decrease in signal is observed, suggesting the presence of a much larger structure.

Looking at the spectra above in figure 6.7, the signal generated by the F/L PhtD sample drops drastically upon addition of the 1st titration of Zn^{2+} atoms (red spectrum) when compared to the Apo-form (green spectrum). This would suggest that a much larger molecule -at the upper measurable limit by NMR- is tumbling in the NMR spectrometer, resulting in the production of the reduced signal seen in the Zn^{2+} -bound form. Since PhtD is a large protein of 92 kDa, it is possible that the addition of the Zn^{2+} atoms and subsequent binding, have caused a large change in structure of the protein, altering it from a flexible molecule with structured regions to a molecule comprising of

a more singular, ordered structure. The presence of this larger single molecule would be enough to cause the difference seen between the spectra.

When the data for the three spectra (apo-, 1.5mM Zn^{2+} , 3mM Zn^{2+}) are normalised as shown in figure 6.8 and the close-up in figure 6.9, additional changes can be seen.

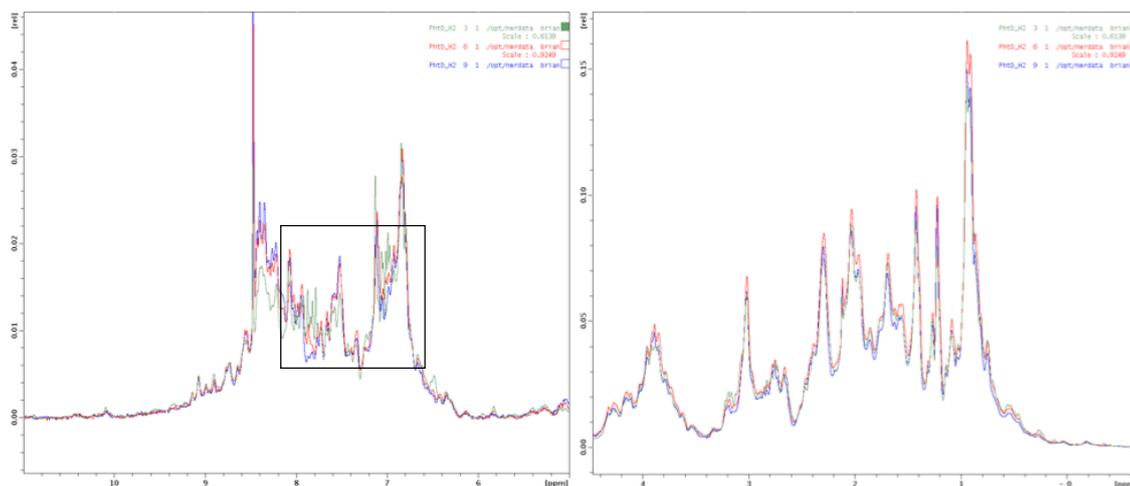


Figure 6.8: Scaled data for F/L PhtD Zn^{2+} titration

The above images show normalised spectra for the amide and aliphatic regions region for the 1D-NMR spectra of F/L PhtD.

Green: Apo- F/L PhtD

Red: F/L PhtD + 5:1 Zn^{2+}

Blue: F/L PhtD + 10:1 Zn^{2+}

Specific chemical shifts can be observed between 7 and 8 ppm (indicated by box). These changes are illustrated in greater detail in figure 6.9.

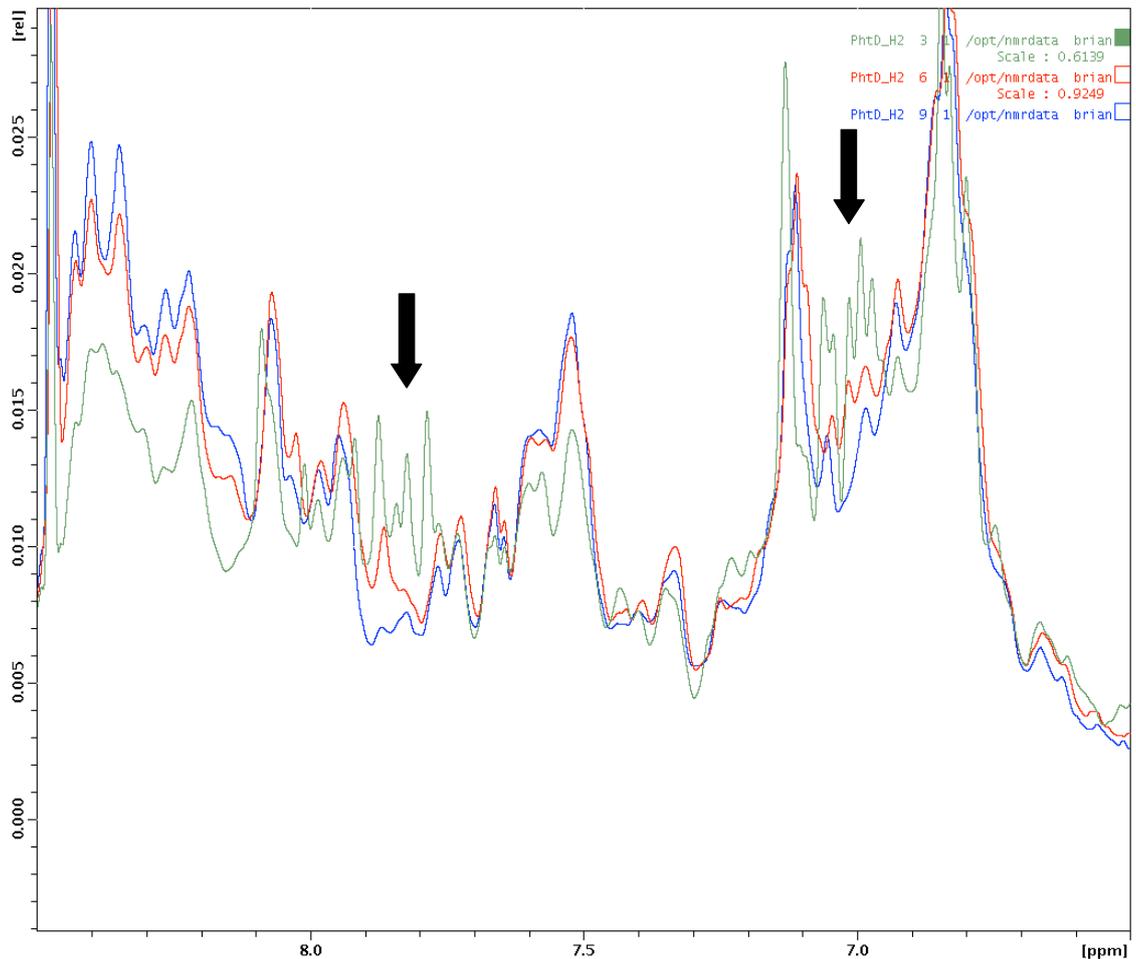


Figure 6.9: Close-up of scaled data for F/L PhtD Zn^{2+} titration

The above image shows a close-up of the normalised amide region for the 1D-NMR spectra from 6.5 - 8.5ppm.

Green: Apo- F/L PhtD

Red: F/L PhtD + 5:1 Zn^{2+}

Blue: F/L PhtD + 10:1 Zn^{2+}

The arrows indicate the regions that undergo a chemical shift when Zn^{2+} is added to the protein. Peaks present in the Apo-form (green spectrum) corresponding to His-sidechains, are observed to disappear in the Zn^{2+} -bound forms (red and blue spectra).

Comparing the Apo- F/L PhtD spectrum with the two spectra for Zn^{2+} -bound F/L PhtD in figure 6.9, there are two main observations. The first is that there are a series of peaks present in the apo- spectrum (green) that have completely disappeared in the Zn^{2+} -bound spectra (red and blue), indicated by the arrows in figure 6.9. These peaks correspond to those typically seen to be generated by histidine residues, and their complete absence in the other Zn^{2+} -bound spectra is indicative of a complete chemical shift resulting from major structural re-arrangement of the histidine residues. Additionally, it is apparent that the remainder of the spectra in this region appears sharper and more defined in the Zn-bound spectra than in that generated by the Apo- F/L PhtD, suggesting a change to a more highly ordered structure. This is

evidence that in F/L PhtD, interaction with Zn^{2+} atoms is occurring, and that a substantial structural change has occurred as a result of this interaction with Zn^{2+} . It is interesting to note however that though a major difference can be seen when comparing the two Zn^{2+} -bound spectra with that for the apo- F/L PhtD, the differences between the 1.5mM and 3mM titration are minimal, implying that a molar ratio of 5:1 Zn^{2+} atoms per molecule may be enough to saturate the Zn^{2+} -binding sites of PhtD, and that subsequent Zn^{2+} addition will have little further effect on the protein.

6.3.2.2 Zn^{2+} Binding In Purified C-term PhtD Protein Fragment

Spectra for the purified C-term fragment of PhtD were generated exactly the same way as previously described in 6.3.2.1 and overlaid for comparison as seen in figures 6.10 and 6.11. As there was little difference between addition of a 5:1 and 10:1 molar ratio of Zn^{2+} atoms : PhtD molecules in the experiments using F/L PhtD, it was deemed that the binding had reached saturation after Zn^{2+} addition at a 5:1 molar ratio, therefore only addition of a 5:1 molar ratio was performed for both the purified C-term and Δ C-term PhtD protein fragments.

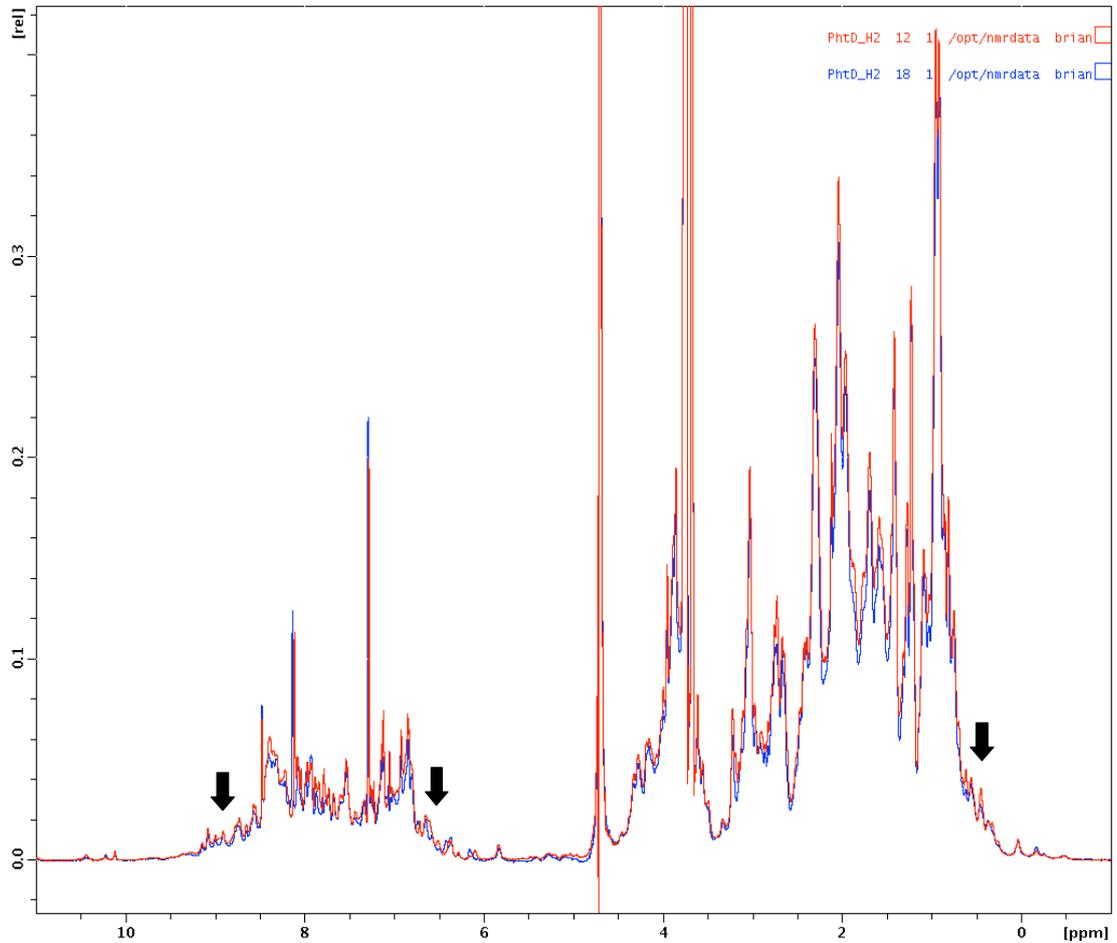


Figure 6.10: C-term PhtD Zn^{2+} titration

Red: Apo- C-term PhtD **Blue:** C-term PhtD + 5:1 Zn^{2+}

General observation of the spectra for C-term PhtD (either in apo- or Zn^{2+} -bound form) shows it to be much closer in appearance to the reference spectrum for “A”-grade protein in figure 6.6, with much more defined peaks at the ends of the amide and aliphatic regions of the spectra (indicated by arrows). This is in contrast with the spectrum for the Apo-form F/L PhtD molecule, suggesting that the C-terminal half of PhtD has an intrinsically more defined structure than the N-terminal half of the F/L molecule. Chemical shifts caused by the addition of Zn^{2+} are illustrated in greater detail in the corresponding close-up image in figure 6.11.

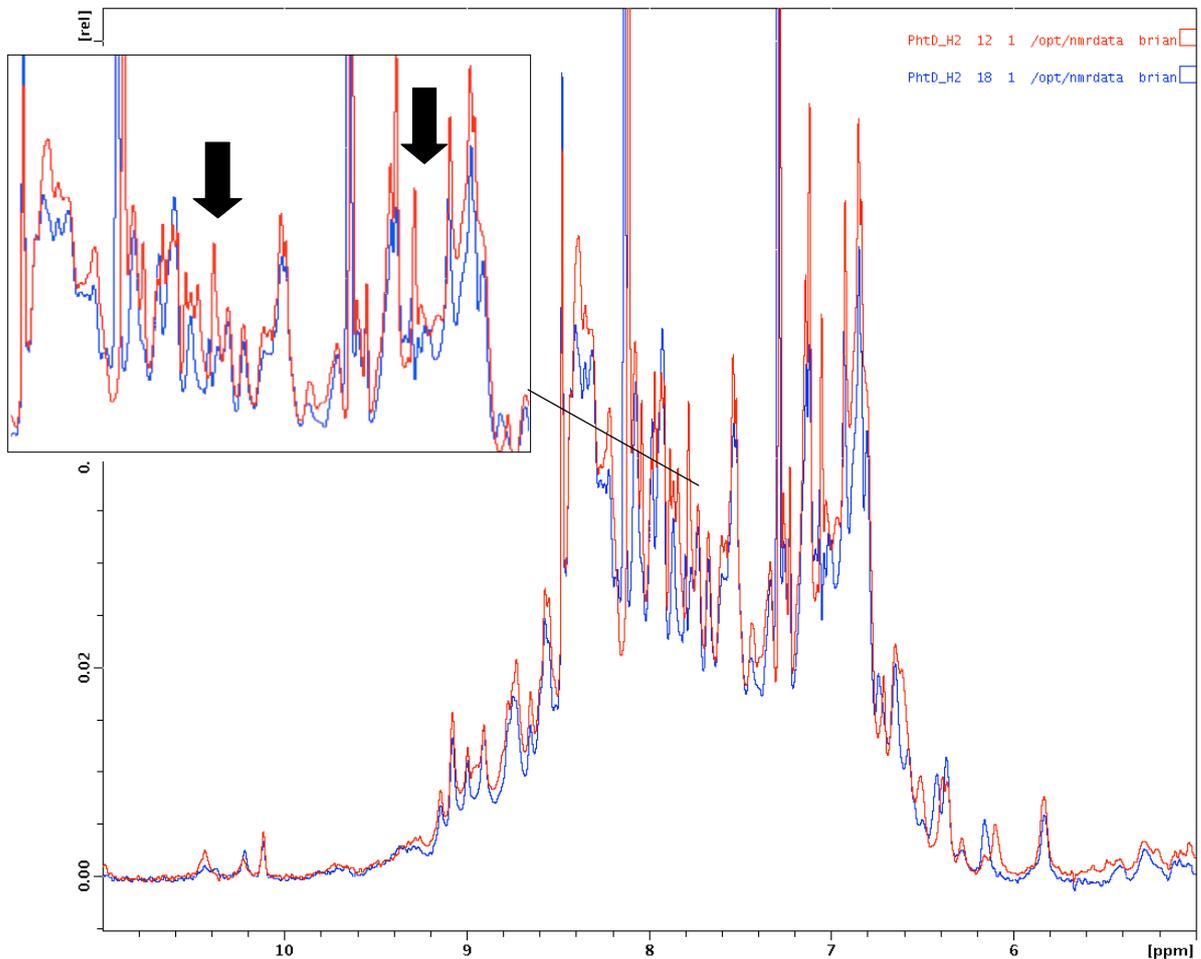


Figure 6.11: Close-up of C-term PhtD Zn^{2+} titration

The above image shows the amide region of the 1D-NMR spectra for C-term PhtD protein

Red: Apo- C-term PhtD **Blue:** C-term PhtD + 5:1 Zn^{2+}

The two spectra show a much larger number of discrete peaks throughout the entire amide region when compared to the F/L molecule suggesting that it has a more intrinsically ordered structure than the F/L molecule.

The boxed out region in the top left illustrates a close-up of the region showing the chemical shifts induced by the introduction of Zn^{2+} . As this fragment of the protein contains only 2 of the 5 HxxHxH motifs, the effect of the addition of Zn^{2+} is not as significant as when Zn^{2+} is introduced to F/L PhtD (see figure 6.9 in comparison).

In contrast with the differences seen in the NMR experiments using F/L PhtD, there is no large change in signal upon addition of Zn^{2+} . However, the abundance of discrete, sharp peaks in the sample even before the introduction of Zn^{2+} to the sample suggest that the C-terminal half of the protein is much more highly structured and therefore that much of the N-terminal half of the protein is present as flexible loops or random coiled in structure. This concurs with the observations noted from the limited proteolysis and mass-spectrometry results discussed in chapter 5. It is still possible to see that the peaks in the amide region corresponding to histidine

residues (indicated in figure 6.11 by arrows in the detailed boxed out region) undergo a chemical shift, but as only two HxxHxH motifs are present (3 HxxHxH motifs are present in the N-terminal half of the protein and therefore not present in this protein fragment) this shift is much less notable than that observed with the F/L protein.

6.3.2.3 Zn²⁺ Binding In Purified ΔC-term PhtD Protein Fragment

As for the previous two proteins, spectra for the purified ΔC-term fragment of PhtD were generated exactly the same way as previously described in 6.3.2.1. The two spectra (apo- and Zn²⁺-bound) were overlaid for comparison as shown in figures 6.12 and 6.13.

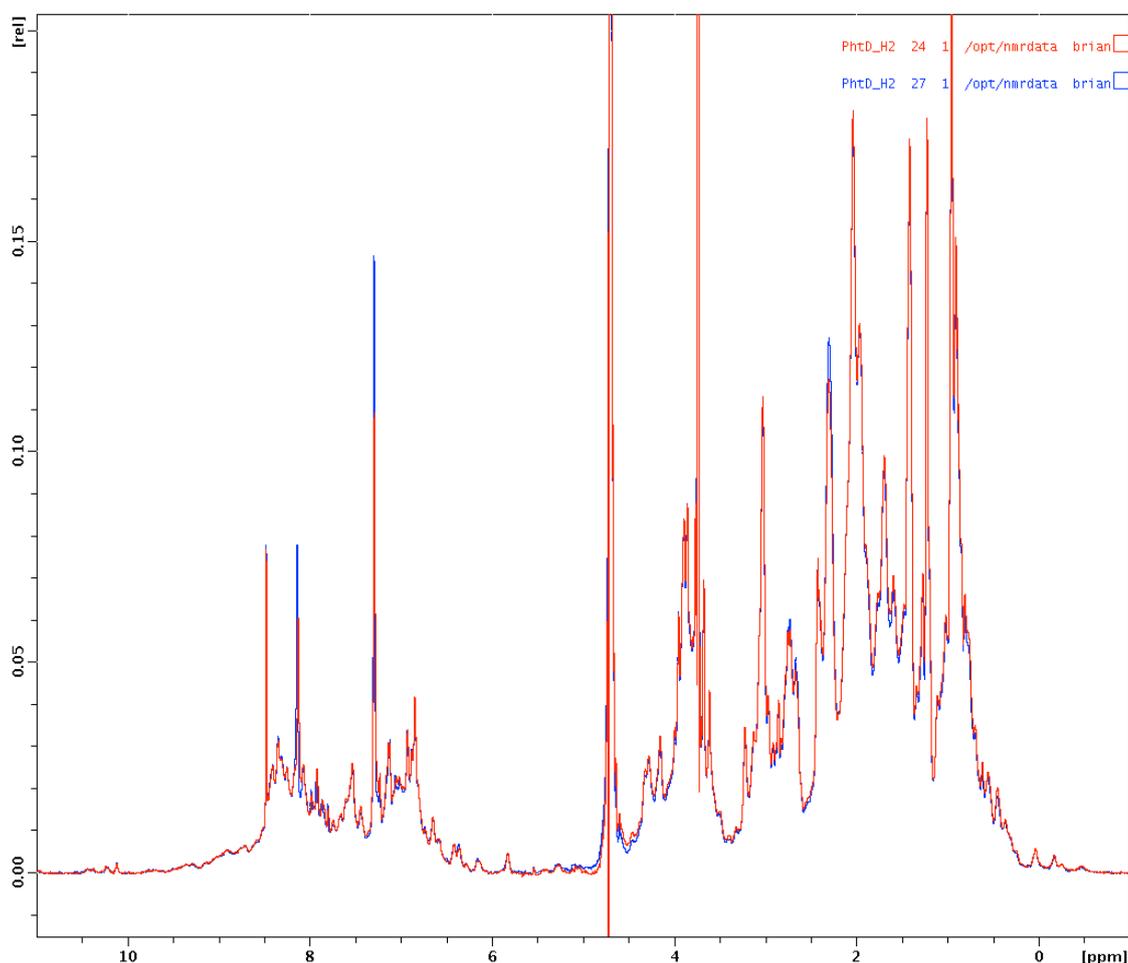


Figure 6.12: ΔC-term PhtD Zn²⁺ titration

Red: Apo- ΔC-term PhtD **Blue:** ΔC-term PhtD + 5:1 Zn²⁺

There are minimal differences between the Apo- and Zn²⁺-bound spectra for the ΔC-term PhtD fragment. Additionally, it can be observed that much of the definition observed for the C-term PhtD spectra (figures 6.10 & 6.11) has been lost, suggesting that the 5 kDa of protein sequence that is included in the C-term PhtD fragment either plays a role in maintaining the overall protein structure of the molecule, or is highly structured in itself.

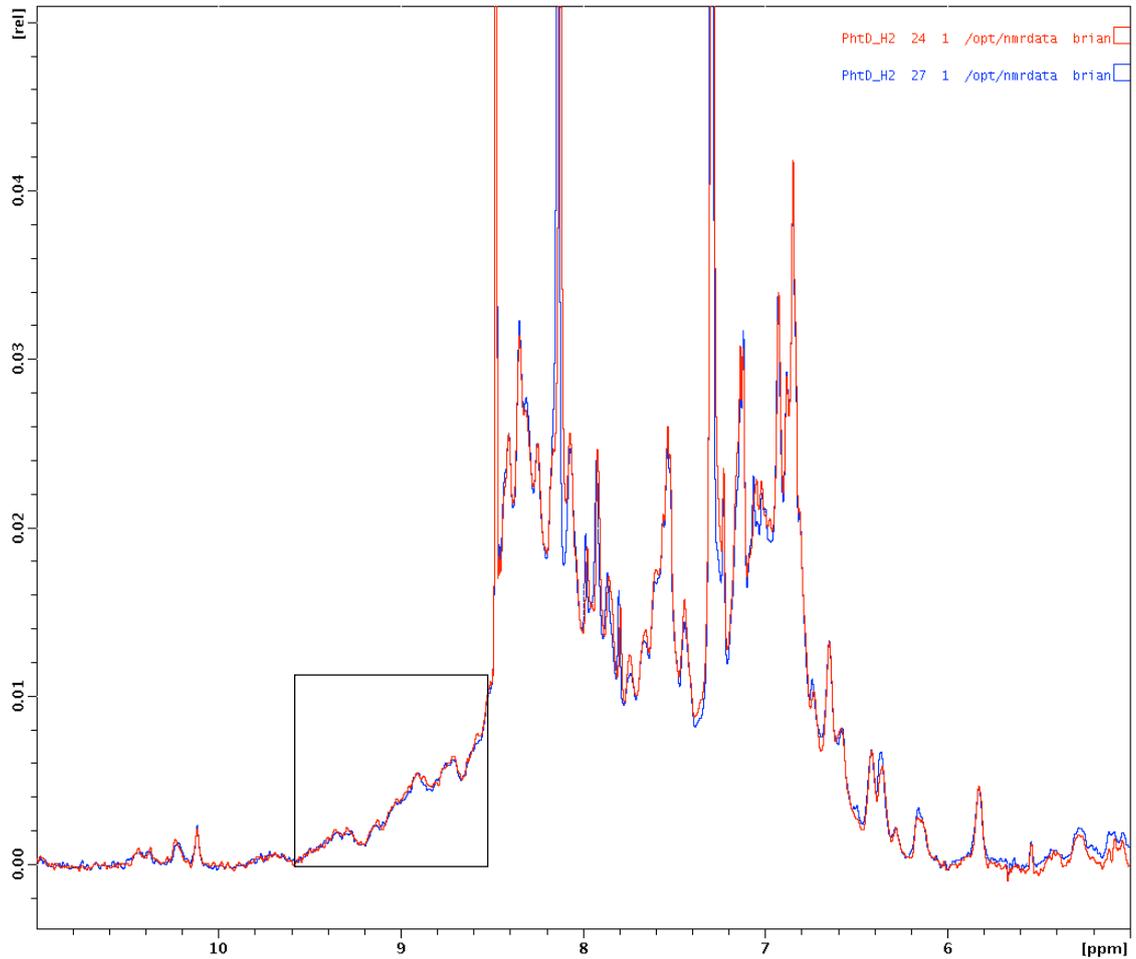


Figure 6.13: Close-up of Δ C-term Zn^{2+} titration

The above image shows the amide region of the 1D-NMR spectra

Red: Apo- Δ C-term PhtD **Blue:** Δ C-term PhtD + 5:1 Zn^{2+}

There are no obvious chemical shifts to be observed between the Apo- and Zn^{2+} -bound forms of Δ C-term PhtD. The boxed out region illustrates the loss of peak definition in the Δ C-term PhtD fragment.

As expected, the differences between the two spectra before and after addition of Zn^{2+} are minimal, since the only difference between the two PhtD fragments is that the Δ C-term excludes the last 5.3 kDa of the protein. Interestingly, the Δ C-term fragment appears to be less highly ordered than the C-term fragment as the edges of the amide and aliphatic regions show fewer, much broader and less defined peaks (example boxed out in figure 6.13; figure 6.14 illustrates a comparison of the same region with the corresponding regions from F/L PhtD and C-term PhtD).

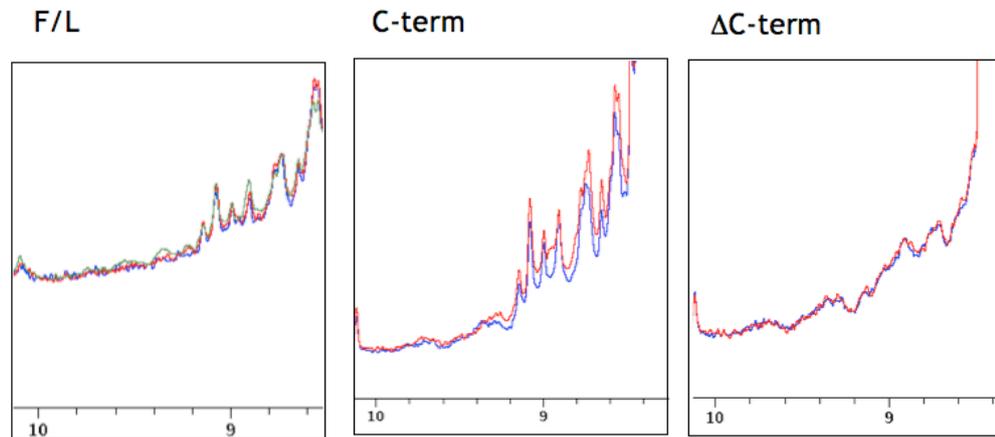


Figure 6.14: comparison of amide shoulder-region for F/L, C-term and Δ C-term PhtD

The above figure shows the boxed out region in figure 6.13 for Δ C-term PhtD and the corresponding regions in the F/L and C-term PhtD spectra.

It can clearly be seen that the C-term fragment has the most discrete peaks, suggesting that it is the most highly structured of the three proteins investigated.

It is possible therefore that the terminal 5.3 kDa of the protein is responsible for some important intramolecular interactions, which aid in keeping the C-terminal half of the protein in a rigid conformation, and that if this region is removed, structural integrity is compromised, or that the 5 kDa of protein sequence by which C-term and Δ C-term fragments differ is in itself intrinsically structured.

These preliminary NMR experiments confirm that the PhtD molecule is relatively flexible and it is likely that it has structured regions interspersed with flexible regions. Comparison between Apo- and Zn^{2+} bound forms of the protein reveals that the C-terminal half of the protein is more structured, and therefore Zn-binding brings the N-terminal region of the molecule into a more ordered conformation, either by itself or with some interaction with the C-terminal half of the molecule. Additionally, it would appear that the very end of the molecule has some role to play in providing the structural integrity of the C-terminal half of the molecule.

With 1D-NMR analysis of the C-term fragment of PhtD revealing that it is relatively highly structured, easily expresses in milligram quantities and the fact that at 55 kDa it is within manageable size, it is possible that this fragment of the molecule could lend itself to structural determination by

NMR, if ^{15}N and ^{13}C labelled forms of the protein fragment could be successfully generated.

6.4 Analysis Of PhtD By Circular Dichroism (CD)

Circular Dichroism (CD) was used for further investigation into the structural changes to the PhtD molecule as seen in the limited proteolysis (see chapter 5) and NMR experiments. Initial CD experiments were performed in collaboration with Dr. Sharon Kelly at the University of Glasgow. Subsequent CD experiments were performed at GSK, Stevenage UK.

It is possible to perform CD at two different ranges in the spectrum of light, both of which are in the Ultra-Violet (UV) region. Near-UV ranges from 260 - 320nm and can be used to determine tertiary structure of a protein molecule based on signals emitted by aromatic residues such as Phe, Tyr and Trp, whilst secondary structure composition can be estimated by CD analysis in the far-UV region (170 - 260nm).¹⁶⁹ By examining the resulting CD traces it is possible to determine the overall dominant secondary structure of the protein of interest. Further estimation of the ratios of the various secondary structural elements (α -helix, β -sheet and random coil) can be calculated using analysis software. Secondary structure ratios presented in this thesis were calculated using the Dichroweb structural prediction server.¹⁷⁰ Figure 6.15 illustrates the typical CD traces observed for the various protein forms.

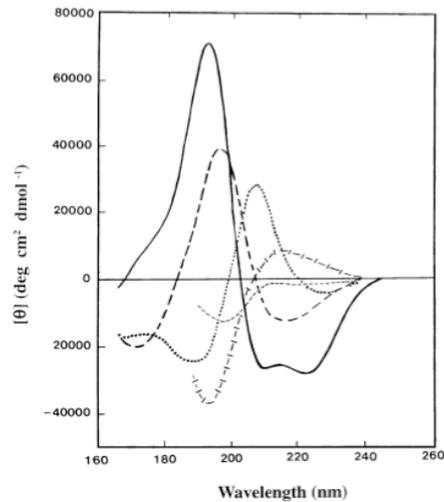


Figure 6.15: Example far-UV CD spectra

Far-UV CD spectra associated with various types of secondary structure. Solid line = α -helix; long dashed line = anti-parallel β -sheet; dotted line = type-I β -turn; cross-dashed line = extended 3_1 -helix or poly (Pro) II helix; short dashed line = irregular structure. Figure reproduced from Kelly et al (2005).¹⁶⁹

As analysis was being performed to observe changes in the secondary structure of PhtD, CD experiments were carried out in the far-UV range, using a wavelength range (λ) from 185-260nm. Initial experiments were carried out on purified F/L PhtD to investigate the changes induced by the binding of Zn^{2+} . Subsequent experiments were performed at GSK Stevenage UK, and included repetition of the original F/L PhtD experiments. In addition to repetition of the F/L CD, experiments investigating the effect of Zn^{2+} binding on the two PhtD fragments named 'C-term' and ' Δ C-term' were performed in order to compare the differences between the F/L molecule and these truncated forms.

6.4.1 CD Analysis of F/L PhtD

Initial CD analysis was performed at the University of Glasgow. 2 aliquots each comprising 1mg of purified F/L PhtD were supplied in 50mM Tris HCl pH7.5. The presence of chloride ions can cause signal-noise during CD experiments, however it was noted that protein stability could be compromised in the absence of NaCl. To this end, the above buffer was supplemented with 500mM NaCl for one of the aliquots.

In order to test the requirement of NaCl in the buffer, far-UV CD was carried out on the protein sample minus NaCl. The same sample was re-tested after incubation at 4°C for 2 weeks. These were then compared with data from protein sample in the presence of NaCl, as seen in figure 6.16.

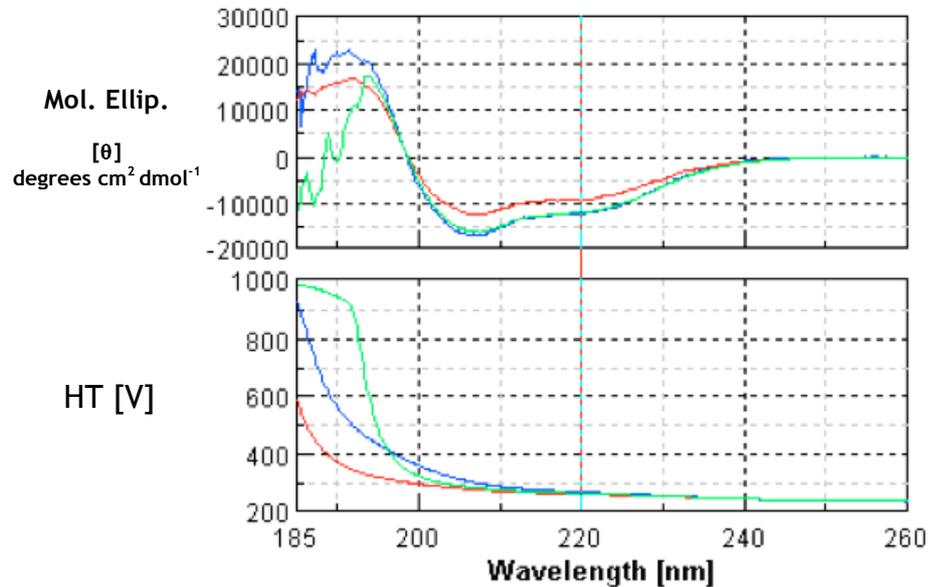


Figure 6.16: F/L PhtD Far UV CD spectra NaCl vs. No- NaCl

(i) Green: F/L PhtD + buffer + 500mM NaCl (ii) Blue: F/L PhtD + buffer, no NaCl

(iii) Red: F/L PhtD + buffer, no NaCl. Re-analyses of sample (ii) after incubation for 2 weeks at 4°C.

Experiments were carried out at 20°C in 50mM Tris HCl pH7.5 (supplemented with 500mM NaCl where indicated), using a protein concentration of 0.5mg/ml.

The top spectra show the CD signal of the three samples. Due to the presence of the chloride ions, data are unreliable below 200nm, due to excessive noise generated by the chloride ions. Comparing the three spectra from the region 240 - 200nm it can be seen that the initial data for F/L PhtD + NaCl, and F/L PhtD - NaCl are comparable. However, Re-analysis of the F/L PhtD - NaCl sample after incubation at 4°C for 2 weeks showed a decrease in CD signal, suggesting a degree of degradation had occurred.

Between the range 240 - 200nm, the CD spectrum for F/L PhtD + NaCl and the initial spectrum for F/L PhtD - NaCl are comparable, as evidenced by their superimposable CD spectra between these wavelengths as seen in figure 6.16. Below 200nm the CD data for the F/L PhtD + NaCl sample is compromised by signal-noise generated by the presence of the chloride ions in the buffer and thus the three spectra cannot be reliably compared below this wavelength. Subsequent CD experiments stopped data acquisition at 195 or 200nm. Re-analysis of the F/L PhtD - NaCl sample after incubation at 4°C for a period of 2 weeks showed that the stability of the F/L PhtD - NaCl

sample has been compromised, with the new spectrum no longer matching that generated by the PhtD + NaCl sample over the same wavelength range of 240 - 200nm. Furthermore, prior to sample re-testing it was noted that after the 2-week incubation period at 4°C the F/L PhtD - NaCl sample was lightly turbid, suggesting a degree of localised precipitation. In contrast to this, the F/L PhtD + NaCl sample had remained clear. It was therefore decided that NaCl should be included in the buffer for subsequent CD experiments.

In order to assess the effect of Zn^{2+} binding on the F/L PhtD molecule, far-UV CD was performed on the sample in order to gain an initial spectrum. The sample was then removed and $ZnSO_4$ added. The sample was then re-analysed in the CD spectrometer under the same conditions as before to generate a spectrum for the Zn^{2+} bound form of PhtD and the spectra overlaid for analysis. The results of this experiment are shown in figure 6.17.

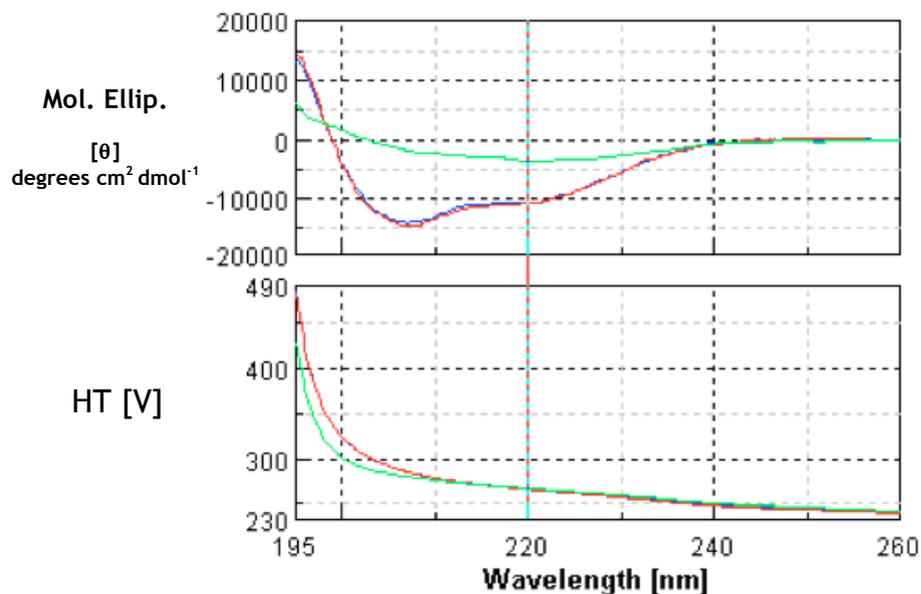


Figure 6.17: Far UV CD spectra of F/L Apo-PhtD vs. F/L PhtD + Zn^{2+}

Blue: F/L Apo-PhtD **Red:** F/L PhtD + 0.8mM $ZnSO_4$ **Green:** F/L PhtD + 10mM $ZnSO_4$

Experiments were carried out at 20°C in 50mM Tris HCl pH7.5, 500mM NaCl (supplemented with 0.8mM or 10mM $ZnSO_4$ where indicated), using a protein concentration of 0.5mg/ml.

Initial CD analysis of Apo- F/L PhtD (blue) showed that the molecule was predominantly α -helical in structure (spectrum compared with example spectrum in figure 6.15). Addition of 0.8mM $ZnSO_4$ (red) showed little evidence of change in the secondary structure of the protein, with the resulting spectrum comparable to that of the Apo- sample. However, addition of an excess of $ZnSO_4$ (green) induced a major change in protein structure, to a predominantly β -sheet conformation.

The spectrum generated by the apo-PhtD sample is typical for a protein that is predominantly α -helical in structure. It is apparent from the above spectra that the addition of 0.8mM Zn^{2+} has had no noticeable effect on the protein structure, with the spectrum almost identical to that of the Apo- F/L PhtD sample. In contrast, a significant change is observed in the spectrum generated with the addition of 10mM Zn^{2+} where the spectrum generated is much more typical of a β -sheet like structure. Comparing these two spectra, it would appear that upon binding Zn^{2+} PhtD changes from an α -helical to a β -sheet like structure. In order to try and quantify this change the CD data were submitted to the Dichroweb structural prediction server (<http://dichroweb.cryst.bbk.ac.uk/>).¹⁷⁰ Data returned from Dichroweb showing the proportion of structural changes is given in table 6.1.

Protein Sample	α -helix(%)	β -sheet(%)	Turns (%)	Random Coil (%)
Apo- F/L PhtD	49	13	15	23
F/L PhtD + 10mM Zn^{2+}	9	40	21	30

Table 6.1: Dichroweb¹⁷⁰ calculation of 2^o structure elements

Secondary structure estimates predict a change from 13% β -sheet in the Apo- form to 40% in the Zn^{2+} -bound form of the protein.

From the analysis, it would appear that although the proportion of random coil increases slightly in the Zn^{2+} -bound sample, the most noticeable changes are how the proportions of α -helix and β -sheet almost completely switch between the apo- and Zn^{2+} -bound samples. Furthermore, after analysis it was observed that the Zn^{2+} -bound sample remained clear, indicating that the protein had not precipitated and that therefore these results were viable. In order to assess whether the binding process was reversible, EDTA was added to the Zn^{2+} -bound sample in an attempt to chelate the metal ion from the protein, and the sample re-analysed which resulted in the original trace being reproduced, suggesting that Zn^{2+} -binding was a reversible process. A significant change in the CD spectrum indicating conformational change was observed during repetition of these experiments during a placement of work carried out at GSK, Stevenage UK (figure 6.18). The CD experiments in the following sections were also carried out at GSK, however secondary structure proportions were unable to be calculated, due to analysis software requiring

data down to a wavelength of 190nm (previously excluded due to signal-noise from chloride ions). Repetition of these experiments to include readings at 190nm should allow these ratios to be calculated.

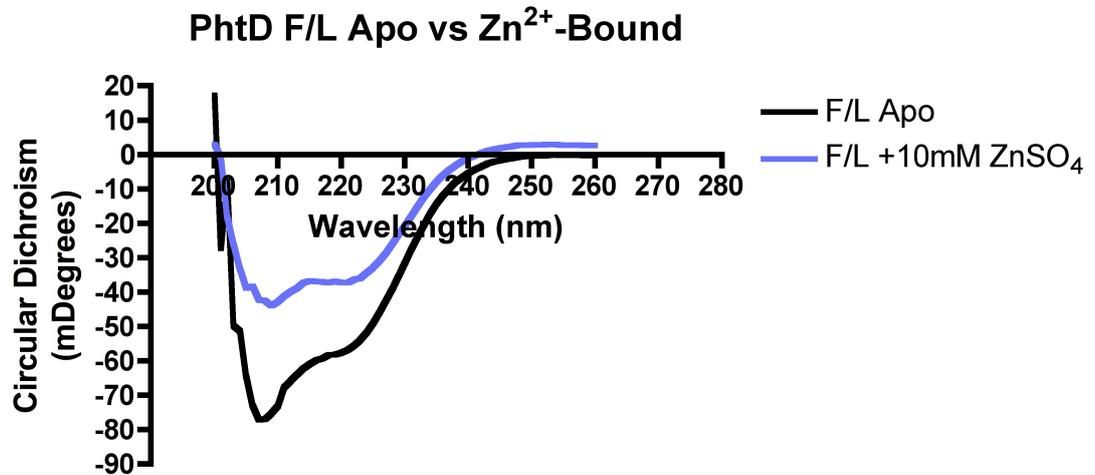


Figure 6.18: Far-UV CD spectra for F/L PhtD

Experiments were carried out in Tris buffer (50mM Tris HCl pH7.5, 500mM NaCl) at 20°C.

Protein concentration = 0.5mg/ml

Black: Apo- F/L PhtD Blue: Zn²⁺-bound F/L PhtD (Zn²⁺-bound sample supplemented with ZnSO₄ to a final concentration of 10mM before measuring).

6.4.2 CD Analysis Of C-term and ΔC-term PhtD Fragments

To try to determine the extent of the structural change induced by Zn²⁺-binding, the same far-UV CD analysis was carried out on the two PhtD fragments 'C-term' and 'ΔC-term' respectively. In each experiment far-UV CD was performed using a Chirascan™ Circular Dichroism spectrometer (Applied Photophysics, Surrey UK) between 200 - 260nm at 20°C, with a measurement increment of 1nm/s. Multiple datasets were collected for each sample and the average for each of these was calculated for use in analysis. Far-UV CD was performed on the apo-samples. The sample was then removed from the cell, ZnSO₄ was added at to a final concentration of 10mM and the sample was re-analysed using the same parameters as for the apo-sample. Multiple datasets were again collected for each sample, and the average used in analysis. Spectra were produced for visualisation of data using Prism 4. Figure 6.19 shows a comparison of apo- C-term PhtD fragment and Zn²⁺-bound C-term PhtD fragment.

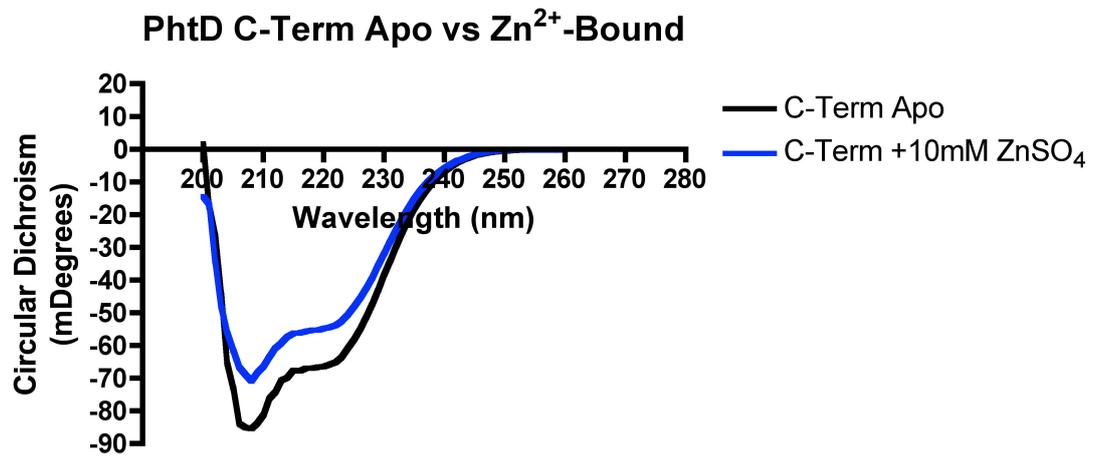


Figure 6.19: Far-UV CD spectra of C-term PhtD fragment

Experiments were carried out in Tris buffer (50mM Tris HCl pH7.5, 500mM NaCl) at 20°C.

Protein concentration = 0.5mg/ml

Black: Apo- C-term PhtD fragment Blue: Zn²⁺-bound C-term PhtD fragment (Zn²⁺-bound sample supplemented with ZnSO₄ to a final concentration of 10mM before measuring).

Although a change is still noticeable between the spectra generated by the apo- and Zn²⁺-bound C-term PhtD fragment, the difference is much less pronounced than when compared to the same conditions for the F/L PhtD molecule.

When the same experiment is performed using the smaller ΔC-term fragment the results (shown in figure 6.20) are similar to those arising from the C-term fragment.

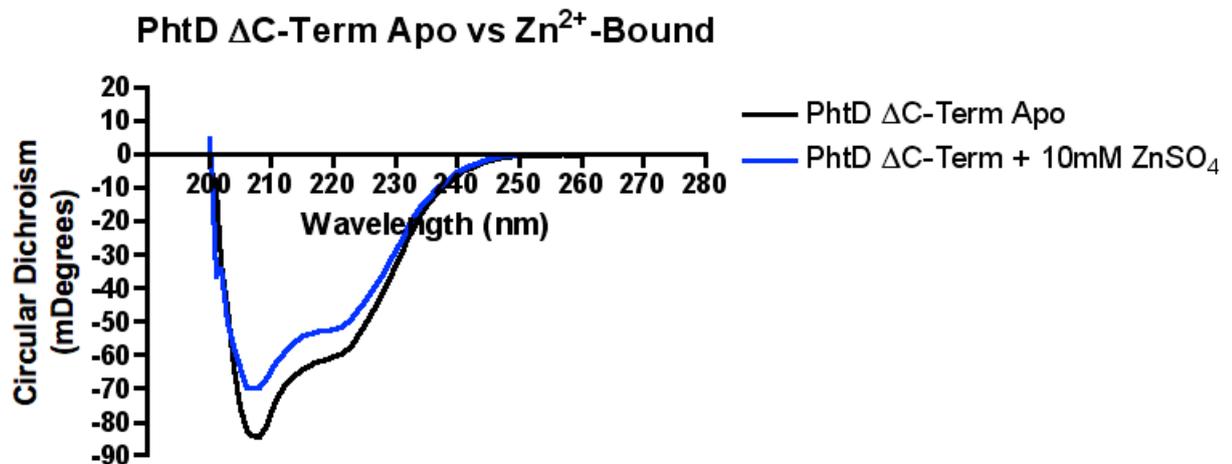


Figure 6.20: Far-UV CD spectra for Δ C-term PhtD fragment

Experiments were carried out in Tris buffer (50mM Tris HCl pH7.5, 500mM NaCl) at 20°C.

Protein concentration = 0.5mg/ml

Black: Apo- Δ C-term PhtD fragment Blue: Zn^{2+} -bound Δ C-term PhtD fragment (Zn^{2+} -bound sample supplemented with $ZnSO_4$ to a final concentration of 10mM before measuring).

The close similarity between the spectra generated by the two different fragments is unsurprising, as they both comprise largely the same part of the F/L molecule, and as a result both have the same two remaining HxxHxH motifs and so each should be able to perform the same binding as the other.

The F/L PhtD molecule may undergo such a dramatic conformational change as a result of the three remaining HxxHxH motifs, which are present in the N-terminal half of the molecule. As the N-terminal half of the PhtD molecule appears to be much less natively structured than the C-terminal half, it could be hypothesised that the binding of Zn^{2+} at the N-terminal sites may cause much more extensive structural re-organisation. It would also possibly be of merit to systematically remove each of the HxxHxH motifs independently and assess the effect on Zn^{2+} binding, and whether there is a dose-dependant relationship between these motifs and the Zn^{2+} atoms, or whether there is a key interaction between one or more of the motifs in the N-terminal half with those in the C-terminal half of the protein.

6.5 Crystallisation Of PhtD

In an attempt to derive function from structure, the technique of macromolecular crystallography was utilised. Initial attempts were made to crystallise the F/L PhtD protein, followed by attempted crystallisation of truncated forms of PhtD, comprising various portions of the C-terminal half of the molecule. Crystal trials were carried out at three locations; the majority of work was performed at the University of Glasgow, with additional crystallisation attempts carried out during industrial placement at GSK, Stevenage. Finally, crystal trials were also carried out in collaboration with Prof. Juan Hermoso at the Spanish National Research Council, Madrid Spain.

6.5.1 Crystallisation Of F/L PhtD

All crystal trials were carried out using the sitting-drop vapour diffusion method¹⁷¹ in 96-well plate format with each well containing a different crystallisation condition effectively resulting in 96 separate experiments in one plate. This method, combined with the use of several commercially available crystallisation-screening kits allows a large number of crystallisation conditions to be tested. The protein of interest is mixed 1:1 with reservoir solution to form the droplet; the plate is sealed with a plastic film and stored at constant temperature whilst being monitored for crystal growth. If successful, crystals should begin to grow as the protein concentration in the droplet changes over time due to the equilibrium relationship between droplet and reservoir. The majority of crystal trials in this work were carried out at 20°C. Initially the possibility of crystallisation at 4°C was also explored, but yielded no positive results and was subsequently excluded.

Initial screening of F/L PhtD was carried out using the original stock of purified F/L PhtD provided in PBS by GSK. Crystal trials were performed using a variety of commercially available screens (see table 2.13) and observed daily for the first week then periodically thereafter. Initial screening at 10mg/ml in PBS yielded a few large crystals in various different crystallisation conditions after a short period of a few days. Testing on an In-house X-ray source (Rigaku 007 X-ray generator) resulted in few large dark spots on the diffraction image indicating that crystals were salt, probably as

an artefact arising from the interaction of the phosphate buffer with components of the relevant crystallisation screen. Crystallisation screening was therefore carried out using a Tris buffer comprising 50mM Tris-HCl pH7.5, 20mM Imidazole, and concentrations of NaCl varying from 500 - 200mM. Additional variations included increasing protein concentrations along a range from 10 - 25mg/ml, presence or absence of His₆ tag, presence or absence of Zn²⁺.

After several months, a few very small crystals were observed in a couple of conditions using F/L PhtD in Tris buffer with 500mM NaCl, - His₆ tag, at a concentration of 10mg/ml (shown in figure 6.21 and 6.22).



Figure 6.21: Crystal from screening with F/L PhtD

Crystallisation conditions: 0.1M ammonium citrate. Protein concentration: 10mg/ml

Figure 6.22: Crystal from screening with F/L PhtD

Crystallisation conditions: 0.1M potassium citrate. Protein concentration: 10mg/ml

Crystals were tested with X-rays in-house, however no diffraction was obtained at all. There could be several possible reasons for this; it may be that the protein forming the crystals is too randomly ordered, or that the crystals are too small to produce a diffraction pattern. Although no protein diffraction was obtained for these crystals, no salt diffraction was obtained either. These findings, coupled with the observation that the crystals were very fragile, with several disintegrating upon attempted manipulation suggested that they were protein crystals. Attempts were made to reproduce or improve upon these crystals by screening around the initial hit conditions, but were unsuccessful.

As it had been observed that exposing F/L PhtD to trypsin resulted in the generation of stable portions of the molecule, this technique was applied to the crystallisation trial process, with the intention to observe whether gradual induced-degradation of the protein during the crystallisation process resulted in crystal formation of a fragment of the protein where a suitable quantity of random coil had been digested away allowing the more structured regions to pack into a crystal lattice more readily. To this end, proteomics grade trypsin (Roche) was added to an 8.5mg/ml F/L PhtD protein solution at a final concentration of 0.21mg/ml immediately prior to setting up crystal trials. This technique had previously been seen to yield success for challenging crystallography targets.^{172, 173}

After a period of 14 months, a single small crystal measuring approximately 20 x 20 x 20µm was observed in one of the trypsin-supplemented crystallisation conditions resulting from screening of F/L PhtD in Tris buffer with 500mM NaCl as seen in figure 6.23.

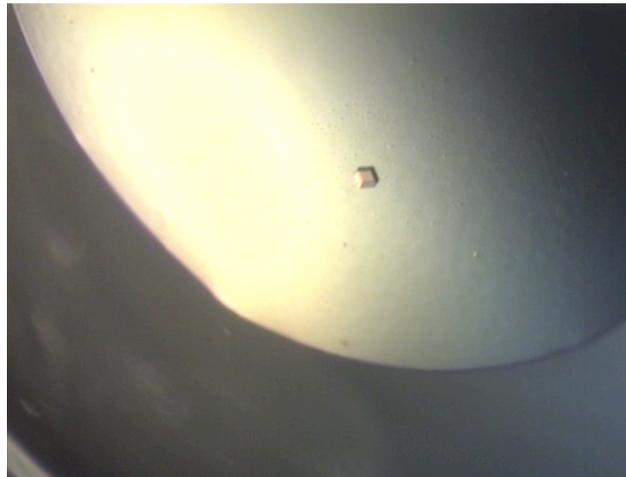


Figure 6.23: Crystal from screening F/L PhtD with trypsin

Crystallisation conditions: 2.4M ammonium sulphate, 0.1M HEPES pH7.0

Protein concentration: 8.5mg/ml (supplemented with 0.34mg/ml trypsin)

In parallel with the crystal trials that yielded the above crystal, a duplicate set of trays were produced without the presence of the PhtD protein in order to provide negative controls. No crystals were observed in any of the conditions from the set of negative control trays.

The crystal was exposed to X-rays in-house. The crystal was flash cooled in a stream of nitrogen gas maintained at 110K (-163°C). Due to the small size of the crystal, no cryoprotectant was used. A single exposure of 1hr was acquired, as the crystal was of such small dimensions. The resulting diffraction image illustrated in figure 6.24 showed very weak spots diffracting to beyond 3Å resolution, with the spots clearly forming lattice lines, confirming that the crystal was protein, and not salt.

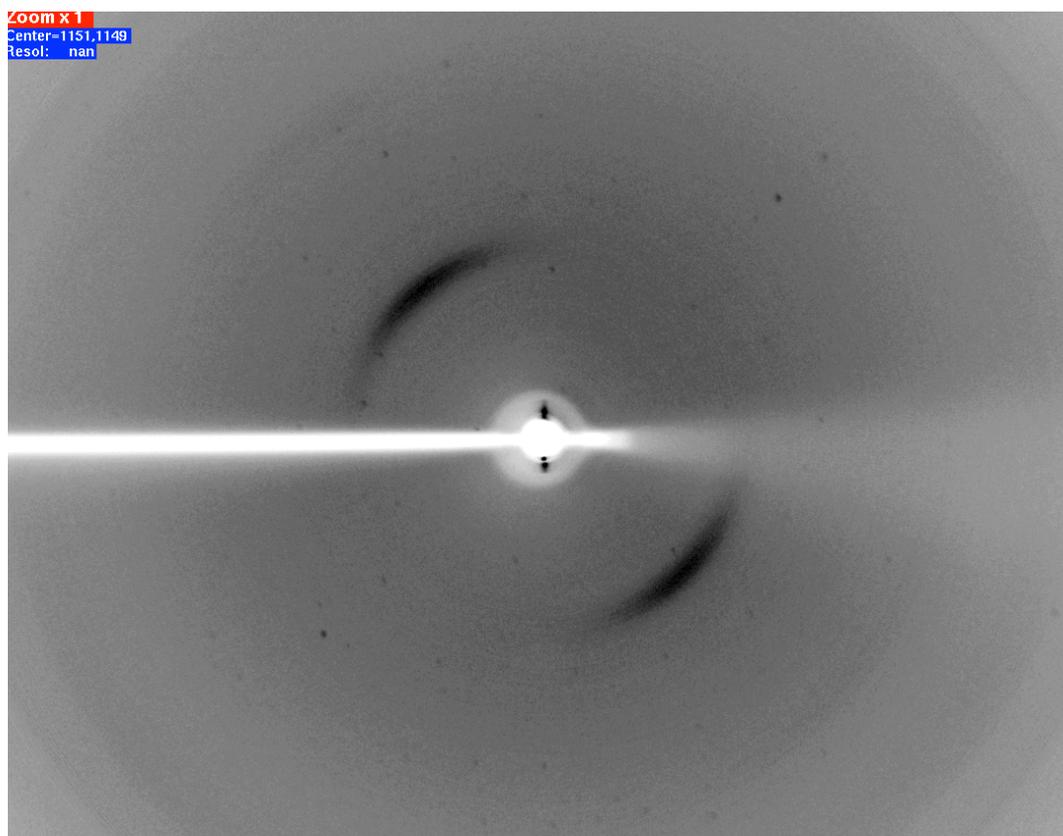


Figure 6.24: Diffraction image for F/L PhtD crystal

As the diffraction image was very weak, the crystal was stored under LN_2 until the next available beam-time at a synchrotron radiation facility, which is capable of producing a much stronger, highly focused X-ray beam than an in-house source and which should enable more extensive acquisition of data.

As it required over one year to obtain the original crystal, 192 replicates of this condition were made, along with 192 replicates of the condition supplemented with double the trypsin concentration in an attempt to accelerate the crystallisation process. Unfortunately, it was not possible to reproduce the original crystal in the replicate conditions. It is possible that

due to the uncontrolled nature of the in-drop proteolysis process where it is unlikely that the same sequence happens in every drop, that the specific conditions were not achieved and therefore crystallisation did not occur in these cases. Other attributing factors may result in using subsequent protein batches or crystallisation screening kits, although precaution was taken to ensure minimal variation between preparation of protein batches, and commercial kits are quality controlled during factory preparation.

Using the initial diffraction data acquired from the single in-house exposure, it was possible to gain some rudimentary information; namely indexing of the crystallographic data to obtain the unit cell parameters and a prediction of the space-group to which the protein belonged. Indexing was performed with the crystallographic program MOSFILM.¹³⁴ The resulting data are illustrated in table 6.2.

Source	Rigaku micromax 007
Space group	C222
Unit-cell parameters (Å, °)	$a=142.19$, $b=207.05$, $c=209.31$, $\alpha=\beta=\gamma=90.0$
Mosaic spread (°)	0.77

Table 6.2: Indexing statistics for single F/L PhtD crystal

This rudimentary data, specifically the unit-cell parameters can give an indication about how the protein is appearing when packed into the crystal. Using the unit-cell parameters above and the molecular weight of the protein (purified F/L PhtD -His₆ tag molecular weight = 91.735 kDa), the program MATTHEWS¹³² was run in order to obtain a Matthews co-efficient value, estimate number of protein molecules per asymmetric unit and corresponding solvent content¹⁷⁴ for the PhtD crystal. These results are illustrated in table 6.3.

Nmol/asym	Matthews Coeff ($\text{\AA}^3 \text{ Da}^{-1}$)	%solvent	Prob
1	8.40	85.36	0.00
2	4.20	70.72	0.02
3	2.80	56.08	0.41
4	2.10	41.44	0.56
5	1.68	26.80	0.00
6	1.40	12.16	0.00

Table 6.3: Results of MATTHEWS¹³² for F/L PhtD crystal

The results of running MATTHEWS¹³² suggest that there are 3 or 4 protein molecules per asymmetric unit based on the probability (41% and 56%), suggesting that PhtD in this case has crystallised as a trimer or an oligomer consisting of 4 copies of the individual molecule. This seems plausible as it is generally observed that in most cases, protein crystals are ~50% solvent,¹⁷⁴ which concurs with the results obtained from MATTHEWS.¹³² The crystal was taken to the Diamond Light Source synchrotron (Didcot, Oxfordshire UK)¹⁵¹ for further data acquisition. Unfortunately the low-temperature transportation unit containing all the samples for the trip was damaged in transport to the facility resulting in the samples being destroyed; no more data was therefore able to be gathered for the PhtD crystal.

F/L crystal trials at GSK continued the work performed at the University of Glasgow, in parallel with crystallisation trials of C-term and Δ C-term PhtD fragments (see sections 6.5.2. and 6.5.3). All crystallisation trials of F/L PhtD at GSK were performed with purified PhtD protein minus His₆-tag, investigating a range of different protein concentrations, from 7.7 - 24.96 mg/ml and lowering NaCl concentration to 200mM as standard.

Before setting up crystal trials, samples were analysed by Dynamic Light Scattering (DLS) in order to determine whether the proteins were present as a monodisperse solution. Of the three proteins, F/L and Δ C-term were present as monodisperse solutions, whilst the C-term PhtD sample appeared to contain a large aggregate. The protein sample was filtered in order to try

and remove the aggregation and achieve a monodisperse solution for the C-term PhtD protein before using for subsequent crystallisation trials.

Crystal trials were performed using a number of commercially available screens (see table 2.13) including Morpheus and PACT and the Nextal Classics sparse matrix screen. Crystal trials were performed by the sitting-drop vapour diffusion method, and set up using a Mosquito[®] nanolitre liquid handling robot (TTP Labtech, Herts UK) to create 200nl droplets comprising of a 1:1 mixture of protein : reservoir solution. Initial one-off testing with the 'PCT' pre-crystallisation test (Hampton Research) indicated that concentrations 10-12mg/ml should be the optimal range. In addition each crystallisation trial had two available drop sites; apo-protein was used for the droplet in site 1, Zn²⁺-bound protein (ZnSO₄ added to apo-protein at a final concentration of 1mM) was used for the droplet in site 2. Although lowering the NaCl concentration appeared to have an effect, with a larger proportion of crystallisation trial droplets showing variation from their initially clear appearance, no crystals were evident during the duration of the placement and there has been no evidence of crystal formation in the time since.

Following protein-buffer screening experiments (see 6.5.1.1), crystallisation trials were performed on F/L PhtD using 50mM MES pH7.0, 200mM NaCl as the protein buffer. Crystal trials were set-up with a protein concentration of 12.06mg/ml supplemented with ZnSO₄ to a final concentration of 1mM in site 1, and 10mM in site 2. These trials were initiated shortly before the end of the placement, at the end of which no crystals had been observed. Crystallisation trials continued to be monitored with updates provided; however to date no crystals have been obtained from any of the above conditions.

6.5.1.1 Crystallisation Buffer Screening Of F/L PhtD

As little success was being obtained using the Tris buffer for crystallisation, investigation was made into the possibility of finding a more suitable buffer that might better promote crystallisation of PhtD. Crystallisation buffer screening was performed by measuring protein melting-temperature (T_m) by circular dichroism. By repeatedly measuring the CD signal at a constant wavelength over time, whilst performing a linear gradient temperature

increase, it is possible to measure the increase in random coil present in the protein as it denatures due to the influence of increasing temperature. The stability of the protein of interest may vary under different buffer conditions, and by screening a number of different conditions, it may be possible to find the most thermostabilising buffer that could then be used as the protein solution buffer in crystallisation attempts. It has been previously observed increasing protein thermostability has resulted in greater crystallisation success.¹⁷⁵

For performing the T_m experiment, 325 μ l of F/L apo-PhtD protein or Zn^{2+} -bound F/L PhtD (2mM $ZnSO_4$ final concentration) at a concentration of 0.1mg/ml was measured at a constant $\lambda=222$ nm over a temperature range from 20 - 90°C. Temperature was increased at a linear rate of 1°C min^{-1} with data acquisition for 20secs after every incremental increase. Due to time constraints only two separate buffers were tested both with the apo-form of F/L PhtD and also with the Zn^{2+} -bound form. Because of previous experience with the formation of phosphate crystals, PBS was excluded. Additionally, due to the phenomenon of the temperature-dependant pH of Tris, which makes measuring by this method difficult, and previous limited success in crystallisation using this buffer it was also excluded. The two buffers tested by T_m were 50mM HEPES pH6.5, and 50mM MES pH7.0.

Screening of apo-protein by this method failed to produce the expected sigmoidal curve normally seen with this method, instead showing a gradual increase throughout the experiment as illustrated in figure 6.25.

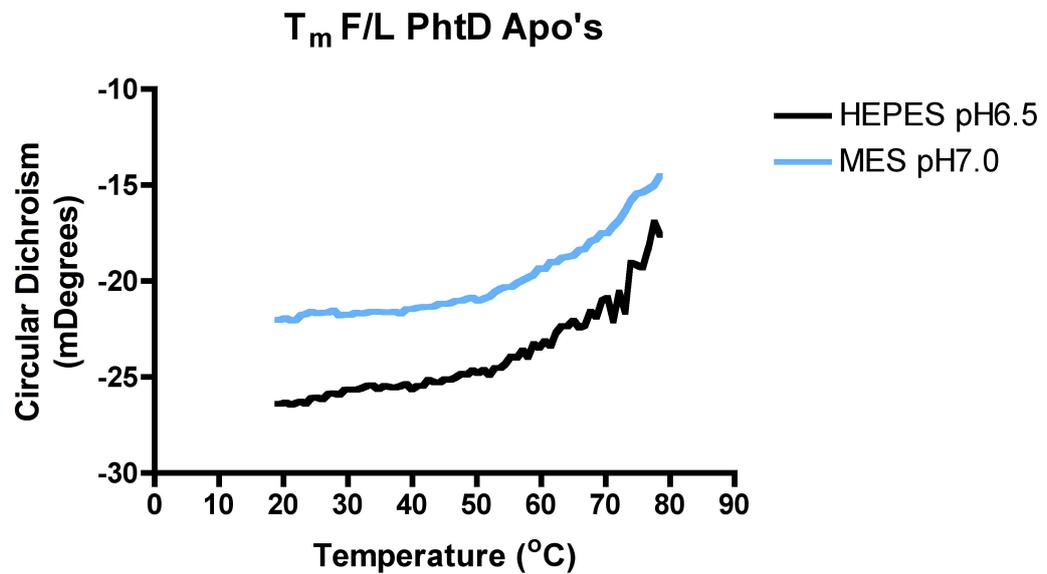


Figure 6.25: T_m results for apo- F/L PhtD.

Protein concentration = 0.1mg/ml

T_m data were not interpretable for Apo- F/L PhtD as no melting curve was obtained. It is possible that due to the largely unstructured, flexible nature of the Apo- form there is no definitive point at which structure is lost, as a large proportion is natively present as random coil.

Previous experiments using NMR and Far-UV CD methods showed the apo-form of the F/L PhtD molecule to be largely unordered in structure. Since a large portion of the F/L PhtD molecule exists naturally as random coil, it is understandable that there is no definitive point at which the protein becomes disordered; the localised sections of the molecule which possess a defined structure gradually denature leading to an increase in the proportion of random coil as seen in the results without a definitive structured-protein to unstructured-protein change observed.

In contrast, the results from experiments performed with the same parameters on the Zn²⁺-bound form of F/L PhtD exhibit the normally observed trend, resulting in the typical melting-curve as seen in figure 6.26.

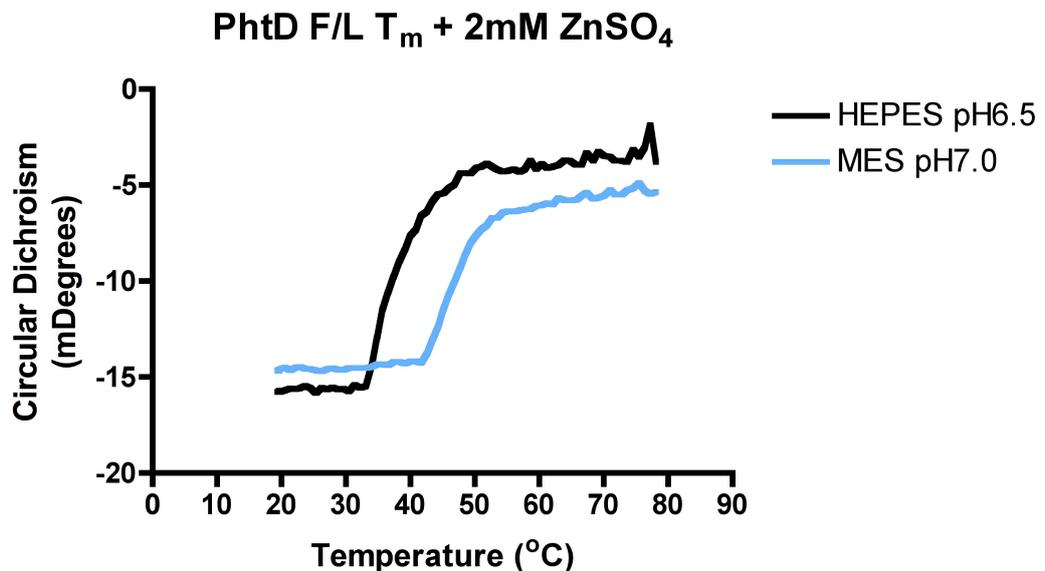


Figure 6.26: T_m results for Zn²⁺-bound F/L PhtD

Protein concentration = 0.1mg/ml, sample was supplemented with ZnSO₄ to a final concentration of 2mM before beginning experiment.

T_m of the Zn²⁺-bound form of F/L PhtD results in the typical melting-curve. Results show that F/L PhtD in HEPES buffer (black) has a T_m around 38°C, contrasting with F/L PhtD in MES buffer which has a T_m around 48°C and is therefore more thermostable in MES.

The expected sigmoidal melting-curve was observed in the Zn²⁺-bound F/L PhtD samples. This could be because as other experiments have indicated, a large change in the structural organisation of the F/L PhtD molecule is induced upon Zn²⁺-binding. If the Zn²⁺ has resulted in the F/L PhtD protein becoming a more structured molecule than the apo-form, then this structure could be denatured with increasing temperature, leading to the more conventional results seen above, rather than those previously illustrated for the apo-form where the protein existed naturally in a largely unstructured conformation.

In the above graph, it can be seen that with both buffers, the protein remains stable up to ~30°C. Above this temperature, there begins a rapid increase in the proportion of random coil for the samples, which decreases in rate at higher temperatures, presumably because the proportion remaining, structured protein is greatly decreased. The melting temperature T_m was attributed to the mid-point of the curve, which corresponded to ~38°C in the HEPES buffer and ~48°C in the MES buffer. It was observed therefore, that the same protein in MES buffer was 10°C more thermally stable than when in

HEPES buffer. As a result of these experiments, protein crystallisation was attempted using 50mM HEPES pH7.0, 200mM NaCl as the protein buffer solution.

6.5.2 Crystallisation Of C-term And Δ C-term PhtD Fragments

Crystallisation of the two fragments of PhtD which encode the structured N-terminal half of the PhtD molecule were subjected to crystallisation trials at GSK during industrial placement. Attempted crystallisation of these fragments was performed in parallel with crystal trials of the F/L PhtD molecule. As such, the same selection of crystallisation screens were utilised, with the two protein fragments being screened over a range of different concentrations varying from 10.01 - 20.02mg/ml for the C-term fragment and 12.15 - 24.3mg/ml for the Δ C-term fragment. Similarly to the approach used with the F/L PhtD protein initial crystallisation trials were performed using Tris buffer with either 500mM or 200mM NaCl. Each crystallisation experiment utilised both available droplet sites, with site one containing the apo- and site 2 the Zn^{2+} -bound forms of the proteins respectively. As before, crystallisation trials were set up using the Mosquito[®] robot to give a 200nl droplet comprising 1:1 ratio of protein : reservoir solution, stored at 20°C and monitored for crystal growth.

After a period of 4 days, a small crystal from the Δ C-term PhtD fragment trial at 12.34mg/ml with 1mM $ZnSO_4$ (i.e. Zn^{2+} bound form) was observed beginning to grow in a single condition from the PACT screen which comprised 20%(w/v) PEG 3350, 0.2M sodium potassium phosphate, 0.1M Bis-Tris propane pH8.5. Over the following 48hrs this crystal was observed to have grown in size, and is illustrated in figure 6.27 Preliminary testing was carried out using a PX-Scanner (Oxford Diffraction, Abingdon UK). An initial, short 20sec exposure yielded no diffraction, however a longer exposure of 6 frames at 0.5° increments with a 100sec exposure per frame showed the presence of few, large dark spots in the diffraction image as seen in figure 6.28 indicating that the crystal was a salt and not protein.

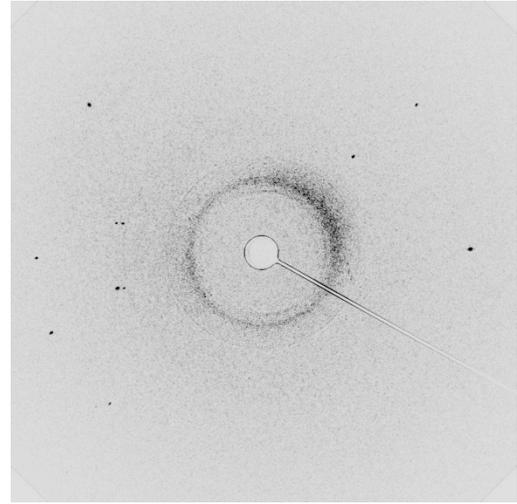
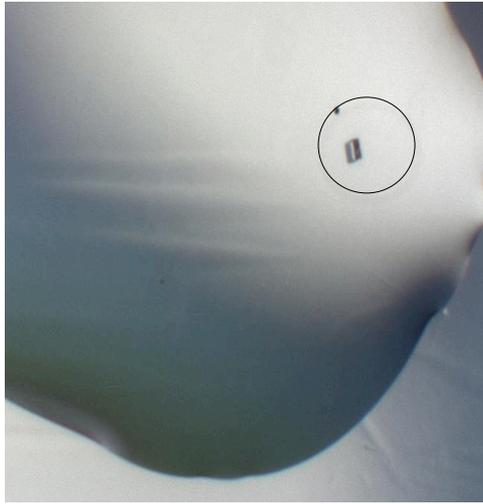


Figure 6.27: C-term PhtD crystal

Crystal circled.

Figure 6.28: Diffraction image for Δ C-term crystal test

Large black spots identify crystal as salt.

As a result of performing buffer screening experiments on F/L PhtD which identified MES buffer as a potentially good thermostabilising buffer for PhtD (see section 4.5.1.1.) Replication of the F/L PhtD crystal trials in the new MES buffer was performed for the C-term and Δ C-term fragments. Proteins were buffer exchanged into the 50mM MES pH 7.0 buffer with 200mM NaCl, concentrated to 17.6 and 20.1mg/ml respectively. Crystallisation trials were set up as previously outlined. Proteins were supplemented with 1mM and 10mM final concentration $ZnSO_4$ for sites 1 and 2 respectively, stored at 20°C and monitored for crystal growth. As these crystallisation trials were set up at the same time as those for the F/L PhtD protein, observation time was limited due to conclusion of the placement, however no crystals had been observed at the end of the placement, and continued observation reported that no crystals have been observed to this date.

6.5.3 PhtD Crystallisation In Collaboration With CSIC Madrid

Ongoing partnerships between the Mitchell Pneumococcal Group and other European laboratories through the Seventh Framework Program¹⁷⁶ instigated by the European Commission (http://cordis.europa.eu/fp7/home_en.html) provided the opportunity for collaboration with Prof. J. Hermoso of the Spanish National Research Council on the crystallisation of PhtD. The

Hermoso Group¹⁷⁷ has extensive experience of determining the structure of pneumococcal proteins by macromolecular X-ray crystallography.¹⁷⁷⁻¹⁸¹ By providing purified protein and additional experimental data outlined in this thesis, it was hoped that this collaboration would yield diffraction quality protein crystals which could either be used to determine the structure, or perhaps obtain initial crystals which could provide a starting point for PhtD crystallisation, allowing crystal optimisation in order to provide crystals which could subsequently be used to determine the protein structure.

Crystallisation trials were performed using the sitting-drop vapour diffusion method. Trials were set up using a Nanodrop™ liquid handling robot (Santa Rosa, California US) mixing a 1:1 ratio of protein : reservoir solution, to give a final drop volume of 500nl. Crystallisation trials were stored at 25°C and monitored for crystal growth.

6.5.3.1 F/L PhtD

F/L PhtD protein was buffer exchanged into 10mM HEPES pH7.5 using Amicon 10K MWCO centrifugal concentrators (Millipore) and concentrated to 16mg/ml. Crystallisation trials were performed as outlined above using a variety of commercially available crystallisation screens (See table 2.13) including Crystal Screen 1& 2. 7 months after initial setup, evidence of phase separation was observed in several conditions, illustrated in table 6.4 Phase separation may give an indication as to preferred crystallisation conditions where the protein and precipitant are reacting favourably and therefore provide a platform for further condition screening utilising components of these conditions.

Buffer (0.1M)	Precipitant	Additive
SPG pH7.0	25% PEG 1500	N/A
SPG pH8.0	25% PEG 1500	N/A
SPG pH9.0	25% PEG 1500	N/A
HEPES pH7.0	20% PEG 6000	0.2M (NH ₄) ₂ Cl ₂ , 0.2M MgCl ₂ , 0.2M LiCl ₂
Tris pH8.0	20% PEG 6000	0.2M (NH ₄) ₂ Cl ₂ , 0.2M MgCl ₂ , 0.2M LiCl ₂
Bis-Tris pH8.5	20% PEG 3350	0.2M Na-malonate
HEPES pH7.5	10% PEG 8000	8% Ethylene glycol
Bicine pH9.0	50% PEG 200	2% Dioxane
HEPES pH7.0	10% PEG 6000	N/A
Tris pH8.0	40% MPD	N/A
Na CaCo pH6.5	30% PEG 8000	0.2M Ammonium sulphate
N/A	30% MPD	0.2M Mg-acetate
Tris pH8.0	30% PEG 4000	0.2M Na-acetate
MES pH6.5	1.6M MgSO ₄ , 30% Jeffamine M-600	0.05M Caesium chloride

Table 6.4: Selected crystallisation conditions from F/L PhtD trials

Crystallisation conditions that gave phase separation after 7 months for F/L PhtD

Although initial screening provided some interesting observations as noted above, no further change in these conditions was observed and no crystals have been obtained to date, although trials are continuing.

6.5.3.2 C-term PhtD

For the C-term PhtD protein fragment, buffer exchange was carried out as described in 6.5.3.1 buffer exchanging into 10mM Tris pH7.5, 20mM imidazole, and concentrated to 12.2mg/ml. Crystallisation trials were performed as before using a variety of commercial crystallisation screens (see table 2.13) and additionally, SaltRx screen (Hampton Research). Trials were stored at 25°C and monitored for crystal growth. 3 months after initial setup, 18 trials showed evidence of changes in the droplet, outlined in table 6.5 with condition 19 (highlighted) producing a visually poor quality crystal.

Buffer (0.1M)	Precipitant	Additive
Na-cacodylate pH6.5	30% MPD	Mg-acetate
MES pH6.5	12% PEG 20000	1.6M Ammonium sulphate
MES pH6.5	25% PEG 550	0.01M ZnSO ₄
HEPES pH7.5	10% PEG 6000	0.5M Ammonium sulphate
HEPES pH7.5	10% PEG 8000	N/A
Tris pH8.5	1M LiSO ₄	0.01M NiCl ₂
Na/K phosphate pH6.2	25% 1-2-propanediol	10% Glycerol
Phosphate citrate pH4.5	20% PEG 8000	0.2M NaCl
Citric acid pH4.0	20% PEG 6000	1M LiSO ₄
Bis-Tris pH7.0	2.2M DL-malic acid	N/A
Tris pH8.5	2.5M Ammonium sulphate	N/A
SPG pH4.0	25% PEG 1500	N/A
Na-acetate pH5.0	20% PEG 6000	0.2M LiCl ₂
PCB pH4.0	25% PEG 1500	N/A
PCB pH5.0	25% PEG 1500	N/A
MNT pH4.0	25% PEG 1500	N/A
Tris pH8.0	20% PEG 6000	0.01M ZnCl ₂
Bis-Tris pH7.5	20% PEG 3350	0.2M potassium thiocyanate
Tris pH8.5	20% PEG 2000	0.01M NiCl₂

Table 6.5: Selected conditions from C-term PhtD trials

Crystallisation conditions which showed evidence of changes in droplet morphology after 3 months. The condition that produced the crystal is highlighted in yellow.

In preparation for testing the crystal in-house by exposure to X-rays, a cryo-protecting solution (to protect the crystal during the flash-cooling process) was prepared by mixing the reservoir solution with glycerol to a final glycerol concentration of 20%. The crystal immediately cracked upon contact with the cryo-solution and partial melting of the crystal was also observed. It was therefore not possible to measure and acquire X-ray data, however the fragility of the crystal is a good indication of the crystal being a protein crystal. Unfortunately, to date reproducibility of the crystal obtained from this condition has not been achieved, although trials are continuing.

6.5.3.3 Δ C-term PhtD

The purified Δ C-term PhtD protein fragment was buffer exchanged into 10mM Tris pH7.0 using a PD10-desalting column (Amersham, GE Healthcare) and concentrated to 16mg/ml with amicon 10K centrifugal concentrators. Despite the absence of NaCl, the protein was judged to be stable in this buffer as no precipitation was observed. Crystal trials were performed as for the F/L PhtD

and C-term PhtD proteins by the sitting-drop vapour diffusion method, using a variety of commercial crystallisation screens (see table 2.13) with a total drop volume of 500nl. Among the conditions tested, two different conditions - 0.05M citrate pH5.0, 15% PEG 600 and 0.05M MES pH6.0, 2.5% PEG 600 - show evidence of crystal formation. Small crystals appear after 1 week of incubation at 20°C as illustrated in figures 6.29 and 6.30.



Figure 6.29: Small Δ C-term PhtD crystals

Crystals appear as very small, thin rods.

Crystallisation conditions: 0.05M citrate pH5.0, 15% PEG 600. Protein concentration: 16mg/ml.

Figure 6.30: Small Δ C-term PhtD crystals

Crystals appear as very small, thin rods.

Crystallisation conditions: 0.05M MES pH6.0, 2.5% PEG 600. Protein concentration: 16mg/ml.

Due to their extremely small size and morphology these crystals are too small and fragile to be useful for X-ray diffraction experiments as it would not be possible to manipulate the crystals without destroying them. Several trials to scale-up or optimise these conditions are ongoing in an attempt to grow bigger more robust crystals from which X-ray protein diffraction data could be acquired. Additionally, crystallisation trials have recently begun to be performed with the purified 34 kDa PhtD fragment and are being monitored for crystal growth.

It is evident from the wide range of different crystallisation attempts carried out at the University of Glasgow, during placement at GlaxoSmithKline, and in collaboration with the CSIC in Madrid, Spain that crystallisation of the F/L PhtD protein molecule has been challenging. However, a rational approach to

evaluating the protein structure, uncovering and generating isolated more highly structured fragments of the F/L PhtD protein has lead to greater success with the crystallisation process, with in-situ proteolysis during the crystallisation trials and crystallisation of recombinantly expressed PhtD fragments both showing promising results which are continuing to be investigated and built upon.

Chapter 7 : Final Discussion

7.1 Structural Studies Of PhtD

7.1.1 Crystallisation Of F/L PhtD

The gene encoding the F/L PhtD protein was successfully isolated from the TIGR4 strain of *S. pneumoniae*, cloned into the expression vector pOPINF¹¹⁹ and subsequently successfully purified to a quality such that analysis of 1µg of purified F/L PhtD by SDS-PAGE showed a single band on the SDS-PAGE gel when stained with coomassie R-250, and used in crystallisation trials. Structural prediction software, coupled with 1D-NMR data suggested that the F/L PhtD molecule comprised many flexible regions interspersed with structured sections of protein. Although a large number of crystallisation formulations were investigated under a varied range of protein conditions such as inclusion of the PhtD binding-partner Zn²⁺ and substitution of different protein buffers, crystallisation of the F/L molecule remained challenging, resulting in the limited success of growing a few poor quality protein crystals which were unsuitable for X-ray diffraction experiments due to their fragility, as they could not be manipulated without being destroyed. The addition of the protease trypsin to the crystallisation experiments in an attempt to crystallise the PhtD protein as unstructured loops were digested, providing more regimented crystal-contacts and therefore a higher likelihood of crystal formation proved more successful. This process resulted in obtaining a protein crystal that yielded rudimentary crystallographic data for PhtD, including a predicted space-group. This approach could be used more extensively, but would first need a degree of optimisation due to the long time-frame between the start of the experiment and crystal formation.

7.1.2 Crystallisation Of PhtD Protein Fragments

A rational approach to selecting fragments of the PhtD molecule, which included limited proteolysis and mass-spectrometry to identify structured regions of interest within the F/L molecule was used to generate a set of PhtD protein fragments. By removing much of the unstructured protein in the

F/L molecule and expressing the regions which were observed to have a more ordered structure it was postulated that these fragments would prove more conducive to the crystallisation process, by providing more regular crystal-contacts and promoting crystal lattice formation. These fragments were successfully purified and subjected to crystallisation trials which resulted in successful crystallisation of two of these fragments (named C-term and Δ C-term) where very small crystals were obtained over a relatively short time-period -weeks compared to months for any F/L PhtD crystals. Although too small for X-ray diffraction experiments, this initial success is encouraging and has given an indication of the crystallisation conditions that should be explored in order to produce larger diffraction-quality crystals, which could then be used to determine the protein structure of a fragment of the PhtD molecule. The successful crystallisation of these fragments also validates the method devised here of targeting specific regions of the protein molecule through computational and biological techniques to generate suitable candidates for structural analysis.

Generation of the protein fragments has also opened up new possibilities into investigating the structure of PhtD, as even the largest of these fragments at 55 kDa in size falls inside the limits of determining the protein structure using NMR.

7.1.3 Interaction Of Zn^{2+} With PhtD

Data in this thesis has shown that PhtD is able to bind Zn^{2+} as previously seen in the crystal structure of a fragment of the related protein PhtA from work previously performed in our laboratory.⁹⁸ Limited proteolysis experiments investigating the interaction of a variety of divalent-cations with PhtD lead us to believe that PhtD is capable of binding Zn^{2+} atoms exclusively. CD analysis has shown that upon binding Zn^{2+} , the protein undergoes a large conformational change from a primarily α -helical structure to a predominantly β -sheet structure. Further CD experiments have shown that this binding process is reversible, with the protein returning to its original structure. Repetition of these CD experiments using the purified PhtD fragments, which exclude a number of the HxxHxH motifs shows a reduction in the magnitude of structural change which occurs upon Zn^{2+} -binding, and is

backed-up by a similar observation in the NMR experiments. Further investigation would be needed in order to ascertain whether Zn^{2+} -binding induces a “dose-dependant” conformational change in the protein structure, or if it promotes the formation of intra-molecular interactions by bringing key residues into proximity with each other. Crystal structures of the F/L PhtD protein in both apo- and Zn^{2+} -bound forms would be the definitive way to reveal this mechanism as the two protein models could be visually compared.

Analysis of the interaction of PhtD with Zn^{2+} by ITC reveals that weak binding occurs at multiple sites in the protein, reinforcing the theory that each HxxHxH motif is responsible for binding a Zn^{2+} atom. Recently published work has suggested that PhtD may perform a role as a zinc-scavenger, acting as a Zn^{2+} -reservoir for the pneumococcus to survive in a zinc-depleted environment.¹⁰² The work presented in this thesis where we observe evidence of weak Zn^{2+} -binding, which induces a reversible change in protein structure appears to corroborate this theory.

7.1.4 Hypothetical Structure of PhtD

A schematic representation of the hypothesised structure of PhtD, resulting from the findings presented in this thesis is outlined in figure 7.1. The full PhtD molecule has been represented in linear format. The locations of key structural components have been colour-coded on the molecule as indicated; the signal sequence responsible for targeting PhtD to the cell-surface is illustrated in yellow, HxxHxH motifs are illustrated in pink, the region thought to contain the dominant epitope is illustrated in green, the stable region of PhtD identified from limited proteolysis experiments and subsequently cloned as pMRH2 - 7 is illustrated in purple. Predicted disordered regions (e.g. flexible loops) are illustrated in red; one disordered region (res 550 - 610) contains one of the HxxHxH motifs, whilst another (res 683 - 773) is present in the hypothesised epitope region.

as metal scavengers that store metal-ions for use by the organism in metal depleted environments.

7.1.6 PhtD As A Vaccine Candidate

Previously published work has shown that immunisation with PhtD provides protection against pneumococcal invasion in a murine model, and that antibodies generated are cross-reactive with the other Pht family members.⁹² The work presented in this thesis has uncovered the major PhtD epitope, which is localised to a 15 kDa region of the F/L molecule. It could therefore be advantageous to modify the vaccine strategy such that immunisation need only be carried out using this small 15 kDa fragment PhtD, cutting out the remaining 77kDa to produce a much more streamlined vaccine component.

7.1.7 Further Work

The data generated during this project and presented in this thesis has laid a solid foundation into the structural analysis of the protein PhtD from the Pht family of proteins, and has opened up new avenues of investigation which should eventually lead to the determination of part of the PhtD molecule; specifically an important section of the protein which encompasses the major PhtD epitope. Further experiments leading towards this goal should include:

- Continued crystallisation trials of PhtD protein fragments, using the initial crystallisation conditions which produced the small crystals as a starting point for condition screening in order to obtain better quality protein crystals.
- NMR spectroscopy of the PhtD protein fragments, especially the 55 kDa C-term and 15 kDa Epitope PhtD fragments in an attempt to uncover the structure of this important region of the protein, in parallel with the continued crystallisation experiments.
- Further extensive screening of PhtD interaction with divalent-cations (supplementary to those investigated in this thesis) by ITC to uncover whether PhtD has affinity for any divalent-cations other than zinc.

- Investigation into whether PhtD : Zn²⁺ binding is “dose-dependant” through manipulation at the genetic level by mutagenesis techniques to systematically remove the HxxHxH motifs and test their ability to bind Zn²⁺ compared to the native form of F/L PhtD.

7.2 Towards The Structure Of PpmA

7.2.1 Summary Of Results Obtained

PpmA protein was successfully expressed and purified for crystallisation trials. Crystals were successfully grown which showed X-ray diffraction to ~2.5Å resolution. However, although diffraction quality was good along one plane of the crystal, rotation through 90° reduced diffraction quality. Indexing of the X-ray diffraction acquired to date has made it possible to reveal rudimentary crystallographic data, however a large body of work remains in order to determine the 3D crystal structure of PpmA.

7.2.2 Future Work

Future work in an attempt to determine the structure of PpmA should definitely focus on attempting to obtain better crystals, which would yield higher quality X-ray diffraction data. Screening around the current crystallising conditions in order to try improve crystal quality, or slowing down protein nucleation by varying incubation temperature in order to try and control crystal growth-rate may yield the desired results. It is thought that as current crystals appear as thin plates, improvement of crystal morphology to give a more 3-dimensional profile would result in the presence of a more substantial lattice in all dimensions, reducing crystal mosaicity, thereby hopefully improving diffraction quality. In order to determine the protein structure, a method of obtaining the phases for the crystal, either by soaking in heavy-metals or producing selenomethionine derivatives of PpmA will then have to be investigated, as molecular replacement is not possible in this instance due to the fact that PpmA has been predicted to be a novel protein.

Appendices

A.1 DNA And Protein Sequences

A.1.1 *PhtD (SP_1003) TIGR4 Natural Sequences*

PhtD Nucleotide Sequence

ATGAAAATTAATAAAAAATATCTAGCAGGTTTCAGTGGCAGTCCTTGCCCTAAGTGTTTGTTCCTATGAA
 CTTGGTCGTCACCAAGCTGGTCAGGTTAAGAAAAGAGTCTAATCGAGTTTCTTATATAGATGGTGATCAG
 GCTGGTCAAAGGCAGAAAACCTTGACACCAGATGAAGTCAGTAAGAGGGAGGGGATCAACGCCGAACA
 AATCGTCATCAAGATTACGGATCAAGGTTATGTGACCTCTCATGGAGACCATTATCATTACTATAATGG
 CAAGGTCCCTTATGATGCCATCATCAGTGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGA
 TTCAGACATTGTCAATGAAATCAAGGGTGGTTATGTTATCAAGGTAGATGGAAAATACTATGTTTACCT
 TAAGGATGCAGCTCATGCGGATAATATTCGGACAAAAGAAGAGATTAACGTCAGAAGCAGGAACACA
 GTCATAATCACGGGGGTGGTTCTAACGATCAAGCAGTAGTTGCAGCCAGAGCCCAAGGACGCTATACA
 ACGGATGATGGTTATATCTTCAATGCATCTGATATCATTGAGGACACGGGTGATGCTTATATCGTTCCT
 CACGGCGACCATTACCATTACATTCCTAAGAATGAGTTATCAGCTAGCGAGTTAGCTGCTGCAGAAGCC
 TATTGGAATGGGAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGCAAATCCAGCTCAACCA
 AGATTGTCAGAGAACCACAATCTGACTGTCACTCCAACCTTATCATCAAATCAAGGGGAAAAACATTTCAA
 GCCTTTTACGTGAATTGTATGCTAAACCCTTATCAGAACGCCATGTGGAATCTGATGGCCTTATTTTCG
 ACCCAGCGCAAATCACAAGTCGAACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCACTTTA
 TCCCTTATGAACAAATGTCTGAATTGGAAAAACGAATTGCTCGTATTATCCCTTCGTTATCGTTCAAA
 CCATTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCCAAGTCC
 GCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATTGGTCAAAGAAGCTGT
 TCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTCTCGTTATATCCAGCCAAGGATC
 TTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACCTGGCCAAGCAGGAAAGTTTATCTCATAAGCTAG
 GAGCTAAGAAAACCTGACCTCCCATCTAGTGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCAA
 GAATTCACCAAGATTTACTTGATAATAAAGGTCGACAAGTTGATTTTGAAGCTTTGGATAACCTGTTGG
 AACGACTCAAGGATGTCCCAAGTGATAAAGTCAAGTTAGTGGATGATATCTTGCCTTCTTAGCTCCGA
 TTCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGATTCAAGTAG
 CCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCGTGATATAACCAGTGATG
 AGGGGGATGCCTATGTAACCTCCACATATGACCCATAGCCACTGGATTAAGAAAGATAGTTTGTCTGAAG
 CTGAGAGAGCGGCAGCCAGGCTTATGCTAAAGAGAAAGGTTTGACCCCTCCTTCGACAGACCATCAG
 GATTCAGGAAATACTGAGGCAAAGGAGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGGT
 GCCACTTGATCGTATGCCTTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACCT
 CATTATGACCATTACCATAACATCAAATTTGAGTGGTTTACGAAGGCCTTTATGAGGCACCTAAGGGG
 TATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACGTCCGCATTCA
 GATAATGGTTTTGGTAACGCTAGCGACCATGTTTCGTAATAAAGGTAGACCAAGACAGTAAACCTGAT
 GAAGATAAGGAACATGATGAAGTAAGTGAGCCAACCTCACCCTGAATCTGATGAAAAAGAGAATCACGCT
 GGTTTAAATCCTTCAGCAGATAATCTTTATAAACCAGCACTGATACGGAAGAGACAGAGGAAGAAGCT
 GAAGATAACACAGATGAGGCTGAAATTCCTCAAGTAGAGAATTCTGTTATTAACGCTAAGATAGCAGAT
 GCGGAGGCCTTGCTAGAAAAAGTAACAGATCCTAGTATTAGACAAAATGCTATGGAGACATTGACTGGT
 CTAATAAAGTAGTCTTCTTCTCGGAACGAAAGATAATAACACTATTTTCAGCAGAAGTAGATAGTCTCTTG
 CTTTGTATAAAGAAAGTCAACCGGCTCCTATACAGTAG

PhtD Protein Sequence

MKINKKYLKAGSVAVLALSVCSEYELGRHQAGQVKKESNRVSYIDGDQAGQKAENLTPDEVSKREGINAEQVIKIT
DQGYVTSHGHDYHYNGKVPYDAIIEELLMKDPNYQLKSDIVNEIKGGYVIKVDGKYYVYLKDAAHADNIRT
KEEIKRQKQEHSHNHGGGSNDQAVVAARAQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNELSASE
LAAAEAYWNGKQGSRPSSSSSYNANPAQPRLESENHNLTVPTYHQNQGENISSLLRELYAKPLSERHVESDGLI
FDPAQITSRTARGVAVPHGNHYHFIPYEQMSELEKRIARIIPLYRSNHWVPDSRPEQPSPQSTPEPSPSPQPAP
NPQPAPSNPIDEKLVKEAVRKVGDGYVFEENGVSRYIPAKDLSAETAAGIDSKLAKQESLSHKLGAKKTDLPSSD
REFYNKAYDLLARIHQDLLDNKGRQVDFEALDNLLERLKDVP SDKVKLVDDILAF LAPIRHPERLGKPN AQITYT
DDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDAYVTPHMTSHWIKKDSLSEAERAAAQAYAKEKGLTPPSTD
HQDSGNT EAKGAEAIYNRVKAAKKVPLDRMPYNLQYTVEVKNGSLIIPHYDHYHNIKFEWFDEGLYEAPKGYT
LEDLLATVKYYYVEHPNERPHSDNGFGNASDHVRKNKVDQDSKPDEDKEHDEVSEPTHPESDEKENHAGLNPS
ADNLYKPSTDTEETEEEAEDTTDEAEIPQVENSVINAKIADAEALLEKVTDP SIRQNAMETLTGLKSSLLLGTKD
NNTISAEVDSLLALLKESQPAPIQ

A.1.2 *pMRH1 PhtD F/L Sequences*

pMRH1 Nucleotide Sequence

ATGGCACACCATCACCACCATCACAGCAGCGGTCTGGAAGTTCTGTTTTCAGGGCCCCGTCT
 ATGAACTTGGTCGTCACCAAGCTGGTCAGTTAAGAAAGAGTCTAATCGAGTTTTCTATATAGATGGTG
 ATCAGGCTGGTCAAAAGGCAGAAAACCTTGACACCAGATGAAGTCAGTAAGAGGGAGGGGATCAACGCC
 GAACAAATCGTCATCAAGATTACGGATCAAGTTATGTGACCTCTCATGGAGACCATTATCATTACTAT
 AATGGCAAGGTCCCTTATGATGCCATCATCAGTGAAGAGCTCCCTCATGAAAGATCCGAATTATCAGTTG
 AAGGATTCAGACATTGTCAATGAAATCAAGGGTGGTTATGTTATCAAGGTAGATGGAAAATACTATGTT
 TACCTTAAGGATGCAGCTCATGCGGATAATATTCGGACAAAAAGAGATTAACGTCAGAAGCAGGAA
 CACAGTCATAATCACGGGGGTGGTTCTAACGATCAAGCAGTAGTTGCAGCCAGAGCCCAAGGACGCTA
 TACAACGGATGATGGTTATATCTTCAATGCATCTGATATCATTGAGGACACGGGTGATGCTTATATCGT
 TCCTCACGGCGACCATTACCATTACCTTCAAGAATGAGTTATCAGCTAGCGAGTTAGCTGCTGCAGA
 AGCCTATTGGAATGGGAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGCAAATCCAGCTCA
 ACCAAGATTGTCAGAGAACCACAATCTGACTGTCACTCCAATTATCATCAAAATCAAGGGGAAAAACATT
 TCAAGCCTTTTACGTGAATTGTATGCTAAACCCTTATCAGAACGCCATGTGGAATCTGATGGCCTTATT
 TTCGACCCAGCGCAAATCACAAGTCGAACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTACCAC
 TTTATCCCTTATGAACAAATGTCTGAATTGGAACCAAGCAATTGCTCGTATTATCCCTTTCGTTATCGTT
 CAAACCATTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCCAA
 GTCCGCAACCTGCACCAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATTTGGTCAAGAAG
 CTGTTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTTCTCGTTATATCCCAGCCAAG
 GATCTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACCTGGCCAAGCAGGAAAGTTTATCTCATAAG
 CTAGGAGCTAAGAAAACCTGACCTCCCATCTAGTGATCGAGAATTTTACAATAAGGCTTATGACTTACTA
 GCAAGAATTCACCAAGATTTACTTGATAATAAAGGTCGACAAGTTGATTTTGGAGCTTTGGATAACCTG
 TTGGAACGACTCAAGGATGTCCCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTTAGCT
 CCGATTCGTCATCCAGAAGTTTAGGAAAACCAATGCGCAAATTACCTACACTGATGATGAGATTCAA
 GTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCGTGATATAACCAAGT
 GATGAGGGGGATGCCTATGTAACCTCCACATATGACCCATAGCCACTGGATTAAGAAAGATAGTTTGTCT
 GAAGCTGAGAGAGCGGCAGCCAGGCTTATGCTAAAGAGAAAGGTTTGACCCCTCCTTCGACAGACCA
 TCAGGATTCAGGAAATACTGAGGCAAAAGGAGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGA
 AGGTGCCACTTGATCGTATGCCTTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCA
 TACCTCATTATGACCATTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTA
 AGGGGTATACTTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAACGTCCCG
 ATTCAGATAATGGTTTTGGTAACGCTAGCGACCATGTTTCGTAAAAATAAGGTAGACCAAGACAGTAAAC
 CTGATGAAGATAAGGAACATGATGAAGTAAGTGAGCCAACCTACCCTGAATCTGATGAAAAAGAGAATC
 ACGCTGGTTTTAAATCCTTCAGCAGATAATCTTTATAAACCAAGCACTGATACGGAAGAGACAGAGGAAG
 AAGCTGAAGATACCACAGATGAGGCTGAAATTCCTCAAGTAGAGAATTCGTTATTAACGCTAAGATAG
 CAGATGCGGAGGCCTTGCTAGAAAAAGTAACAGATCCTAGTATTAGACAAAATGCTATGGAGACATTGA
 CTGGTCTAAAAAGTAGTCTTCTTCTCGGAACGAAAGATAATAACACTATTTTCAGCAGAAGTAGATAGTC
 TCTTGGCTTTGTTAAAAGAAAGTCAACCGCTCCTATACAGTAG

pMRH1 F/L PhtD Protein Sequence With His₆ Tag

MAHHHHHSSGLEVLFGPSYELGRHQAGQVKKESNRVSYIDGDQAGQKAENLTPDEVSKREGINAEQIVIKI
 TDQGYVTSHGHDHYHYNGKVPYDAIISEELLMKDPNYQLKDSDIVNEIKGGYVIKVDGKYVYVYLKDAHADNIR
 TKEEIKRQKQEHSHNHGGGSDNDQAVVAARAQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNELSA
 SELAAAEAYWNGKQGSRPSSSSSYNANPAQPRLSEHNLTVTPTYHQNGENISSLLRELYAKPLSERHVESD
 GLIFDPAQITSRTARGVAVPHGNHYHFIPYEQMSELEKRIARIPLRYRSNHWWPDSRPEQSPQSTPEPSPSPQ
 PAPNPQPAPSNPIDEKLVKEAVRKVGDGYVFEENGVSRYIPAKDLSAETAAGIDSKLAKQESLSHKLGAKKTDL
 PSSDREFYNKAYDLLARIHQDLLDNKGRQVDFEALDNLLERLKDVPDKVYVLDLAFAPIRHPERLKGPN
 QITYTDDDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDAYVTPHMTSHSHWIKKDSLSEAAAAQAYAKEKGLTP
 PSTDHDQDSGNTEAKGAEAIYNRVKAARKVPLDRMPYNLQYTVKNGSLIIPHYDHYHNIKFEWFDEGLYEAP
 KGYTLEDLLATVKYVVEHPNERPHSDNGFGNASDHVRKNKVDQDQSKPDEKHEHDEVSEPTHPESDEKENHAG
 LNPSADNLYKPSDTEETEEEAEDTTDEAEIPQVENSVINAKIADAEEALLEKVTDPISIRQNAMETLTGLKSSLLL
 TKDNNTISAEVDSLLALLKESQPAPIQ

pMRH1 F/L PhtD Protein Sequence Minus His₆ Tag

GPSYELGRHQAGQVKKESNRVSYIDGDQAGQKAENLTPDEVSKREGINAEQJVIKITDQGYVTSHGDHYHYYN
GKVPYDAIISEELMKDPNYQLKDSDIVNEIKGGYVIKVDGKYVYVKDAAHADNIRTKEEIKRQKQEHSHNHG
GGSNDQAVVAARAQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNELSAELAAAEAYWNGKQGSR
PSSSSYNANPAQPRLSENHNLTVTPTYHQNQGENISSLLRELYAKPLSERHVESDGLIFDPAQITSRTARGVAV
PHGNHYHFIPYEQMSELEKRIARIIPLYRSNHWVPDSRPEQSPQSTPEPSPSPQPAPNPQPAPSNPIDEKLVK
EAVRKVGDGYVFEENGVSRYIPAKDLSAETAAGIDSKLAKQESLSHKLGAKKTDLPSSDREFYNKAYDLLARIHQ
DLLDNKGRQVDFEALDNLLERLKDVP SDKV KLVDDILAF LAPIRH PERLGKPN AQITYTDDEIQVAKLAGKYTTE
DGYIFDPRDITSDEGDAYVTPHMTSHWIKKDSLSEAERAAAQAYAKEKGLTPSTDHQDSGNTEAKGAEAIY
NRVKAACKVPLDRMPYNLQYTVKNGSLIIPHYDHYHNIKFEWFDEGLYEAPKGYTLEDLLATVKYYYVEHPN
ERPHSDNGFGNASDHVRKNKVDQDSKPEDKEHDEVSEPTHPESDEKENHAGLNPSADNLYKPSTDTEETEEE
AEDTTDEAEIPQVENSVINAKIADAEALLEKVTDP SIRQNAMETLTGLKSSLLLGTKDNNTISAEVDSLLALLKES
QPAPIQ

A.1.3 *pMRH2 C-term PhtD Truncate Sequences*

pMRH2 C-term PhtD Truncate Nucleotide Sequence

ATGGCACACCATCACCACCATCACAGCAGCGGTCTGGAAGTTCTGTTTTAGGGCCCCGTCAA
 ACCATTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCCAAGTC
 CGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATTGGTCAAAGAGCTG
 TTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTCTCGTTATATCCCAGCCAAGGAT
 CTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACCTGGCCAAGCAGGAAAGTTTATCTCATAAGCTA
 GGAGCTAAGAAAACCTGACCTCCCATCTAGTGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCA
 AGAATTCACCAAGATTTACTTGATAATAAAGGTCGACAAGTTGATTTTGAGGCTTTGGATAACCTGTTG
 GAACGACTCAAGGATGTCCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTTAGCTCCG
 ATTCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGATTCAAGTA
 GCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCGTGATATAACCAGTGAT
 GAGGGGGATGCCTATGTAACCTCCACATATGACCCATAGCCACTGGATTAATAAAGATAGTTTGTCTGAA
 GCTGAGAGAGCGGCAGCCAGGCTTATGCTAAAGAGAAAGGTTTGACCCCTCCTTCGACAGACCATCA
 GGATTCAGGAAATACTGAGGCAAAAGGAGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGG
 TGCCACTTGATCGTATGCCTTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAAATCATAAC
 TCATTATGACCATTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGG
 GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAACGTCCGCATTC
 AGATAATGGTTTTGGTAACGCTAGCGACCATGTTTCGTAATAAAGGTAGACCAAGACAGTAAACCTGA
 TGAAGATAAGGAACATGATGAAGTAAGTGAGCCAACCTACCCTGAATCTGATGAAAAAGAGAATCACGC
 TGGTTTAAATCCTTCAGCAGATAATCTTTATAAACCAAGCACTGATACGGAAGAGACAGAGGAAGAAGC
 TGAAGATACCACAGATGAGGCTGAAATTCCTCAAGTAGAGAATTCTGTTATTAACGCTAAGATAGCAGA
 TGCGGAGGCCTTGCTAGAAAAAGTAACAGATCCTAGTATTAGACAAAATGCTATGGAGACATTGACTGG
 TCTAAAAAGTAGTCTTCTTCGGAACGAAAGATAATAACACTATTTTCAGCAGAAGTAGATAGTCTCTT
 GGCTTTGTTAAAAGAAAGTCAACCGGCTCCTATACAG

pMRH2 C-term PhtD Truncate Protein Sequence With His₆ Tag

MAHHHHHSSGLEVLVFGPSNHWVPSRPEQSPQSTPEPSPSPQPAPNPQPAPSNPIDEKLVKEAVRKVGD
 GYVFEENGVSRYIPAKDLSAETAAGIDSKLAKQESLSHKLGAKKTDLPSSDREFYNKAYDLLARIHQDLLDNKGR
 QVDFEALDNLLERLKDVP SDKVKLVDDILAF LAPIRHPERLGKPN AQITYTDDEIQVAKLAGKYTTEDGYIFDPR
 DITSDEGDAYVTPHMTSHWIKKDSLSEAERAAAQAYAKEKGLTPPSTDHQDSGNTEAKGAEIYNRVKAAKK
 VPLDRMPYNLQYTVEVKNGLIIPHYDHYHNIKFEWFDEGLYEAPKGYTLEDLLATVKYVEHPNERPHSDNG
 FGNASDHVRKNKVDQDSKPDEDKEHDEVSEPTHPESDEKENHAGLNPSADNLYKPSTDTEETEEEAEDTTDEA
 EIPQVENSVINAKIADAEALLEKVTDPSIRQ_NAMETLTGLKSSLLLGTKDNNTISAEVDSLLALLKESQPAPIQ

pMRH2 C-term PhtD Truncate Protein Sequence Minus His₆ Tag

GPSNHWVPSRPEQSPQSTPEPSPSPQPAPNPQPAPSNPIDEKLVKEAVRKVGDGYVFEENGVSRYIPAKDLS
 AETAAGIDSKLAKQESLSHKLGAKKTDLPSSDREFYNKAYDLLARIHQDLLDNKGRQVDFEALDNLLERLKDVP
 SDKVKLVDDILAF LAPIRHPERLGKPN AQITYTDDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDAYVTPHMTSH
 HWIKKDSLSEAERAAAQAYAKEKGLTPPSTDHQDSGNTEAKGAEIYNRVKAAKKVPLDRMPYNLQYTVEVK
 NGLIIPHYDHYHNIKFEWFDEGLYEAPKGYTLEDLLATVKYVEHPNERPHSDNGFGNASDHVRKNKVDQDS
 KPDEDKEHDEVSEPTHPESDEKENHAGLNPSADNLYKPSTDTEETEEEAEDTTDEAIPQVENSVINAKIADAE
 LLEKVTDPSIRQ_NAMETLTGLKSSLLLGTKDNNTISAEVDSLLALLKESQPAPIQ

A.1.4 *pMRH3* ΔC -term *PhtD* Truncate Sequences

pMRH3 ΔC -term *PhtD* Truncate Nucleotide Sequence

ATGGCACACCATCACCACCATCACAGCAGCGGTCTGGAAGTTCTGTTTCAGGGCCCCGTCAA
 ACCATTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCCAAGTC
 CGAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATTGGTCAAAGAAGCTG
 TTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTCTCGTTATATCCCAGCCAAGGAT
 CTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACCTGGCCAAGCAGGAAAGTTTATCTCATAAGCTA
 GGAGCTAAGAAAAGTACCTCCCATCTAGTGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCA
 AGAATTCACCAAGATTTACTTGATAATAAAGGTGACAAGTTGATTTTGGAGCTTTGGATAACCTGTTG
 GAACGACTCAAGGATGTCCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTAGCTCCG
 ATTCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGATTCAAGTA
 GCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCGTGATATAACCAGTGAT
 GAGGGGGATGCCTATGTAACCTCCACATATGACCCATAGCCACTGGATTAATAAAGATAGTTTGTCTGAA
 GCTGAGAGAGCGGCAGCCAGGCTTATGCTAAAGAGAAAGGTTTACCCTCCTTCGACAGACCATCA
 GGATTCAGGAAATACTGAGGCAAAAGGAGCAGAAGCTATCTACAACCGCTGAAAGCAGCTAAGAAGG
 TGCCACTTGATCGTATGCCTTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAAATCATAAC
 TCATTATGACCATTACCATAACATCAAATTTGAGTGGTTTACGAAGGCCTTTATGAGGCACCTAAGGG
 GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAACGTCCGCATTC
 AGATAATGGTTTTGGTAACGCTAGCGACCATGTTTCGTAATAAAGGTAGACCAAGACAGTAAACCTGA
 TGAAGATAAGGAACATGATGAAGTAAGTGAGCCAACCTCACCTGAATCTGATGAAAAAGAGAATCACGC
 TGGTTTAAATCCTCAGCAGATAATCTTTATAAACCAAGCACTGATACGGAAGAGACAGAGGAAGAAGC
 TGAAGATACCACAGATGAGGCTGAAATTCCTCAAGTAGAGAATTCGTTATTAACGCTAAGATAGCAGA
 TCGGAGGCCTTGCTAGAAAAA

pMRH3 ΔC -term *PhtD* Truncate Protein Sequence With His₆ Tag

MAHHHHHSSGLEVLFGPSNHWVPDSRPEQSPQSTPEPSPSPQPAPNPQPAPSNPIDEKLVKEAVRKVGD
 GYVFEENGVSRYIPAKDLSAETAAGIDSKLAKQESLSHKLGAKKTDLPSSDREFYNKAYDLLARIHQDLLDNKGR
 QVDFEALDNLLERLKDVP SDKV KLVDDILAFLAPIRHPERLGKPN AQITYTDDEIQVAKLAGKYTTEDGYIFDPR
 DITSDEGDAYVTPHMTSHWIKKDSLSEAERAAAQAYAKEKGLTPPSTDHQDSGNTEAKGAEIYNRVKAAKK
 VPLDRMPYNLQYTVEVKNGSLIIPHYDHYHNIKFEWFDEGLYEAPKGYTLEDLLATVKYYYVEHPNERPHSDNG
 FGNASDHVRKNKVDQDSKPDEDKEHDEVSEPTHPESDEKENHAGLNPSADNLYKPSTDTEETEEEAEDTTDEA
 EIPQVENSVINAKIADAEALLEK

pMRH3 ΔC -term *PhtD* Truncate Protein Sequence Minus His₆ Tag

GPSNHWVPDSRPEQSPQSTPEPSPSPQPAPNPQPAPSNPIDEKLVKEAVRKVGDGYVFEENGVSRYIPAKDLS
 AETAAGIDSKLAKQESLSHKLGAKKTDLPSSDREFYNKAYDLLARIHQDLLDNKGRQVDFEALDNLLERLKDVP
 SDKV KLVDDILAFLAPIRHPERLGKPN AQITYTDDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDAYVTPHMTSH
 HWIKKDSLSEAERAAAQAYAKEKGLTPPSTDHQDSGNTEAKGAEIYNRVKAAKKVPLDRMPYNLQYTVEVK
 NGSLIIPHYDHYHNIKFEWFDEGLYEAPKGYTLEDLLATVKYYYVEHPNERPHSDNGFGNASDHVRKNKVDQDS
 KPDEDKEHDEVSEPTHPESDEKENHAGLNPSADNLYKPSTDTEETEEEAEDTTDEAEIPQVENSVINAKIADAE
 LLEK

A.1.5 pMRH4 5.3 kDa C-terminal PhtD Truncate Sequences

pMRH4 5.3 kDa C-terminal PhtD Truncate Nucleotide Sequence

ATGGCACACCATCACCACCATCACAGCAGCGGTCTGGAAGTTCTGTTTCAGGGCCCGGTAACAGATCCT
AGTATTAGACAAAATGCTATGGAGACATTGACTGGTCTAAAAAGTAGTCTTCTTCTCGGAACGAAAGAT
AATAACACTATTTTCAGCAGAAGTAGATAGTCTCTTGGCTTTGTTAAAAGAAAAGTCAACCGGCTCCTATA
CAG

pMRH4 5.3 kDa C-terminal PhtD Truncate Protein Sequence With His₆ Tag

MAHHHHHHSSGLEVLFGQGPVTDPSIRQNAMELTGLKSSLLLGTKDNNTISAEVDSLLALLKESQPAPIQ

pMRH4 5.3 kDa C-terminal PhtD Truncate Protein Sequence Minus His₆ Tag

GPVTDPSIRQNAMELTGLKSSLLLGTKDNNTISAEVDSLLALLKESQPAPIQ

A.1.6 pMRH6 34 kDa PhtD Truncate Sequences

pMRH6 34 kDa PhtD Truncate Nucleotide Sequence

ATGGCACACCATCACCACCATCACAGCAGCGGTCTGGAAGTTCTGTTTCAGGGCCCGTCAA
 ACCATTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAGTCCAAGTC
 CGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATTGGTCAAAGAAGCTG
 TTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTCTCGTTATATCCCAGCCAAGGAT
 CTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACCTGGCCAAGCAGGAAAGTTTATCTCATAAGCTA
 GGAGCTAAGAAAACCTGACCTCCCATCTAGTGATCGAGAATTTTACAATAAGGCTTATGACTTACTAGCA
 AGAATTCACCAAGATTTACTTGATAATAAAGGTCGACAAGTTGATTTTGAGGCTTTGGATAACCTGTTG
 GAACGACTCAAGGATGTCCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCTTCTTAGCTCCG
 ATTCGTCATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGATTCAAGTA
 GCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCGTGATATAACCAGTGAT
 GAGGGGGATGCCTATGTAACCTCACATATGACCCATAGCCACTGGATTAATAAAGATAGTTTGTCTGAA
 GCTGAGAGAGCGGCAGCCAGGCTTATGCTAAAGAGAAAGGTTTGACCCCTCCTTCGACAGACCATCA
 GGATTCAGGAAATACTGAGGCAAAAAGGAGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAAGAAGG
 TGCCACTTGATCGTATGCCTTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAATCATACC
 TCATTATGACCATTACCATAAC

pMRH6 34 kDa PhtD Truncate Protein Sequence With His₆ Tag

MAHHHHHHSSGLEVLFGPSNHWVPDSRPEQSPQSTPEPSPSPQPAPNPQPAPSNPIDEKLVKEAVRKVGD
 GYVFEENGVSRYIPAKDLSAETAAGIDSKLAKQESLSHKLGAKKTDLPSSDREFYNKAYDLLARIHQDLLDNKGR
 QVDFEALDNLLERLKDVP SDKV KLVDDILAF LAPIRHPERLGKPN AQITYTDDEIQVAKLAGKYTTEDGYIFDPR
 DITSDEGDAYVTPHMT HSHWIKKDSLSEAERAAAQAYAKEKGLTPPSTDHQDSGNTEAKGAEIYNRVKAAKK
 VPLDRMPYNLQYTVEVKNGSLIIPHYDHYHN

pMRH6 34 kDa PhtD Truncate Protein Sequence Minus His₆ Tag

GPSNHWVPDSRPEQSPQSTPEPSPSPQPAPNPQPAPSNPIDEKLVKEAVRKVGDGYVFEENGVSRYIPAKDLS
 AETAAGIDSKLAKQESLSHKLGAKKTDLPSSDREFYNKAYDLLARIHQDLLDNKGRQVDFEALDNLLERLKDVP
 SDKV KLVDDILAF LAPIRHPERLGKPN AQITYTDDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDAYVTPHMT H
 HWIKKDSLSEAERAAAQAYAKEKGLTPPSTDHQDSGNTEAKGAEIYNRVKAAKKVPLDRMPYNLQYTVEVK
 NGSLIIPHYDHYHN

A.1.7 *pMRH7 PhtD Epitope Truncate Sequences*

pMRH7 PhtD Epitope Truncate Nucleotide Sequence

ATGGCACACCATCACCACCATCACAGCAGCGGTCTGGAAGTTCTGTTTCAGGGCCCGATCAAATTTGAG
 TGGTTTGACGAAGGCCTTTATGAGGCACCTAAGGGGTATACTCTTGAGGATCTTTTGGCGACTGTCAA
 GTAATAATGTCGAACATCCAAACGAACGTCCGCATTGAGATAATGGTTTTGGTAACGCTAGCGACCATGT
 TCGTAAAAATAAGGTAGACCAAGACAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAGTGAGCC
 AACTCACCTGAATCTGATGAAAAAGAGAATCACGCTGGTTTAAATCCTTCAGCAGATAATCTTTATAAA
 CCAAGCACTGATACGGAAGAGACAGAGGAAGAAGCTGAAGATACCACAGATGAGGCTGAAATTCCTCA
 AGTAGAGAATTCTGTTATTAACGCTAAGATAGCAGATGCGGAGGCCTTGCTAGAAAAA

pMRH7 PhtD Epitope Truncate Protein Sequence With His₆ Tag

MAHHHHHSSGLEVLFGPIKFEWFDEGLYEAPKGYTLEDLLATVKYYVEHPNERPHSDNGFGNASDHVRKN
 KVDQDSKPDDEDKEHDEVSEPTHPESDEKENHAGLNPSADNLYKPSTDTEETEEEAEDTTDEAEIPQVENS
 VINA
 KIADAEALLEK

pMRH7 PhtD Epitope Truncate Protein Sequence Minus His₆ tag

GPIKFEWFDEGLYEAPKGYTLEDLLATVKYYVEHPNERPHSDNGFGNASDHVRKNKVDQDSKPDDEDKEHDEV
 SEPTHPESDEKENHAGLNPSADNLYKPSTDTEETEEEAEDTTDEAEIPQVENSVINAKIADAEALLEK

A.1.8 PpmA (SP_0981) TIGR4 Natural Sequences

PpmA Nucleotide Sequence

ATGAAGAAAAAATTATTGGCAGGTGCCATCACACTATTATCAGTAGCAACTTTAGCAGCTTGTTGAAA
 GGGTCAGAAGGTGCAGACCTTATCAGCATGAAAGGGGATGTCATTACAGAACATCAATTTTATGAGCAA
 GTGAAAAGCAACCCTTCAGCCCAACAAGTCTTGTTAAATATGACCATCCAAAAAGTTTTTAAAAACAAT
 ATGGCTCAGAGCTTGATGATAAAGAGGTTGATGATACTATTGCCGAAGAAAAAAAACAATATGGCGAAA
 ACTACCAACGTGTCTTGTGACAAGCAGGTATGACTCTTGAAACACGTAAAGCTCAAATTCGTACAAGTA
 AATTAGTTGAGTTGGCAGTTAAGAAGGTAGCAGAAGCTGAATTGACAGATGAAGCCTATAAGAAAAGCC
 TTTGATGAGTACACTCCAGATGTAACGGCTCAAATCATCCGTCTTAATAATGAAGATAAGGCCAAAGAA
 GTTCTCGAAAAAGCCAAGGCAGAAGGTGCTGATTTTGTCTCAATTAGCCAAAGATAATTCAACTGATGAA
 AAAACAAAAGAAAATGGTGGAGAAATTACCTTTGATTCTGCTTCAACAGAAGTACCTGAGCAAGTCAAA
 AAAGCCGCTTTTCGCTTTAGATGTGGATGGTGTCTGATGTGATTACAGCAACTGGCACACAAGCCTAC
 AGTAGCCAATATTACATTGTA AAACTCACTAAGAAAACAGAAAAATCATCTAATATTGATGACTACAAAG
 AAAAATTA AAAACTGTTATCTTGACTCAAAAACAAAATGATTCAACATTTGTTCAAAGCATTATCGGAAA
 AGAATTGCAAGCAGCCAATATCAAGGTTAAGGACCAAGCCTTCCAAAATATCTTTACCCAATATATCGG
 TGGTGGAGATTCAAGCTCAAGCAGTAGTACATCAAACGAATAG

PpmA Protein Sequence

MKKKLLAGAITLLSVATLAACSKGSEGADLISMKGDVITEHQFYEQVKSNPAAQVLLNMTIQKVFQYKQYSEL
 DDKEVDDTIAEKKQYGENYQVLSQAGMTLETRKAQIRTSKLVAVKKVAEAEALTDEAYKKAFDEYTPDVT
 AQIIRLNNEKAKEVLEKAKAEGADFAQLAKDNSTDEKTKENGGEITFDSASTEVPQVKKAAFALDVGVSVDV
 ITATGTQAYSSQYYIVKLTKKTEKSNIDYKEKLTIVILTQKQNDSTFVQSIIGKELQAANIKVKDQAFQNIPTQ
 YIGGGDSSSSSSTSNE

A.1.9 PpmA Expression Construct Sequences

PpmA Expression Construct Nucleotide Sequence

ATGGCACACCATCACCACCATCACAGCAGCGGTCTGGAAGTTCTGTTTCAGGGCCCCGAAGAAAAAATTA
 TTGGCAGGTGCCATCACACTATTATCAGTAGCAACTTTAGCAGCTTGTTTCGAAAGGGTCAGAAGGTGC
 AGACCTTATCAGCATGAAAGGGGATGTCATTACAGAACATCAATTTTATGAGCAAGTAAAAGCAACCC
 TTCAGCCCAACAAGTCTTGTTAAATATGACCATCCAAAAAGTTTTTAAAAACAATATGGCTCAGAGCTT
 GATGATAAAGAGGTTGATGATACTATTGCCGAAGAAAAAAAACAATATGGCGAAAACCTACCAACGTGTC
 TTGTCACAAGCAGGTATGACTCTTGAAACACGTAAAGCTCAAATTCGTACAAGTAAATTAGTTGAGTTG
 GCAGTTAAGAAGGTAGCAGAAGCTGAATTGACAGATGAAGCCTATAAGAAAGCCTTTGATGAGTACACT
 CCAGATGTAACGGCTCAAATCATCCGTCTTAATAATGAAGATAAGGCCAAAGAAGTTCTCGAAAAAGCC
 AAGGCAGAAGGTGCTGATTTTGCTCAATTAGCCAAAGATAATTCAACTGATGAAAAACAAAAAGAAAT
 GGTGGAGAAATTACCTTTGATTCTGCTTCAACAGAAGTACCTGAGCAAGTCAAAAAAGCCGCTTTTCGCT
 TTAGATGTGGATGGTGTCTGATGTGATTACAGCAACTGGCACACAAGCCTACAGTAGCCAATATTAC
 ATTGTA AAACTCACTAAGAAAAACAGAAAAATCATCTAATATTGATGACTACAAAGAAAAATTA AAACTG
 TTATCTTGACTCAAAAAACAAAATGATTCAACATTTGTTCAAAGCATTATCGGAAAAGAATTGCAAGCAGC
 CAATATCAAGGTTAAGGACCAAGCCTTCCAAAATATCTTTACCCAATATATCGGTGGTGGAGATTCAAG
 CTC AAGCAGTAGTACATCAAACGAATAG

PpmA Expression Construct Protein Sequence With His₆ Tag

MAHHHHHSSGLEVLFGPKKKLLAGAITLLSVATLAACSKGSEGADLISMKGDVITEHQFYEQVKSNPAAQQ
 VLLNMTIQKVFQYGSSELDDEKVDITAEKKQYGENYQRVLSQAGMTLETRKAQIRTSKLVAVKKVAEAE
 LTDEAYKKAFFDEYTPDVTAQIIRLNNEKAKEVLEKAKAEGADFAQLAKDNSTDEKTKENGGEITFDSASTEVP
 EQVKKAALFDVVDGVSVDVITATGTQAYSSQYIVKLTKKTEKSNIDYKEKLTQKQNDSTFVQSIIGKEL
 QAANIKVKDQAFQNIPTQYIGGGDSSSSSSSNE

PpmA Expression Construct Protein Sequence Minus His₆ Tag

PKKKLLAGAITLLSVATLAACSKGSEGADLISMKGDVITEHQFYEQVKSNPAAQQVLLNMTIQKVFQYGSSE
 LDDKEVDDITAEKKQYGENYQRVLSQAGMTLETRKAQIRTSKLVAVKKVAEAEELTDEAYKKAFFDEYTPDV
 TAQIIRLNNEKAKEVLEKAKAEGADFAQLAKDNSTDEKTKENGGEITFDSASTEVPQVKKAALFDVVDGVS
 VITATGTQAYSSQYIVKLTKKTEKSNIDYKEKLTQKQNDSTFVQSIIGKELQAANIKVKDQAFQNIPT
 QYIGGGDSSSSSSSNE

A.1.10 *TktA* (LSL_1946) *L. salivarius* Natural Sequences

TktA Nucleotide Sequence

ATGTATGATCAAGTAGACCAATTGGGAGTAAATACTCTTAGAACATTATCAATCGACGCAATCCAAAGA
 GCTAACTCAGGTCATCCAGGTTTACCAATGGGAGCGGCACCAATGGCATATGTCCTATGGACACGTCAT
 TAAAAATTAATCCTAAAACACATATGAACTGGGTAATAGAGATAGATTTGTTTTATCGGCAGGACAT
 GGTTTCAGCTTTACTATATAGTTTGGCACATTTAGCTGGATATGATGTTTCAATGGATGATTTGAAGAAT
 TTTAGAGAATGGAAATCTAATACACCAGGACATCCTGAATATGGTTGTACCGATGGAGTGGAAAGCTACA
 ACAGGGCCATTAGGTCAAGGAATTTCAATGGCTGTAGGAATGGCGATGGCTGAAGCTCATCTAGGTAA
 GAAATTTAACCGTGAAGGTTACCCAGTAATGGATCATTATACATACGCTTTAATTGGCGATGGTGATTT
 AATGGAAGGTGTCGCTAGTGAAGCAGCGTCATTGGCAGGTCATTTGAAATTAGGTAAGTTAATTGCTT
 TGTATGATTCAAATGGAATTTCTTTAGATGGTAAACTTTCAGCATCATTTACAGAAAATGTAGGAGCTC
 GTTTTGAAGCTTATGGCTGGCAATATATTCTAGTAGAAGATGGTTTCAATCTCGAAGAGATTGATAAAG
 CAATTGTTCAAGCAAAGCAGAAAGTGATAAGCCAACAATTATTGAAATAAAAACTACAATTGGTTATG
 GTTCAGAAAATCAAGGTACACATAAAGTACATGGAAGTCTCTTGGTGAAGAAGGAGTAGCACACGCTA
 AAGAAGTATATAACTGGAATTATCCACCATTTACAGTACCTGAAGAAGTAAGTCAAAGATTTAAAGAAT
 GCTTACAAGATAAGGGTGTTAAAGCAGAAAATAAGTGAATGAAATGTTTGAAGCATATAAAAAAGAAT
 ATTCAGATCTTGCACAAAAGTTTTCCGACGGCTTTTCTAACAAGGTTCAAATACTCTTGGAGACATTTT
 ACCACAATATGGAGAAGATGACAGTATCGCAACACGTGCAGCTAGTCAAAAAGCAATTAATGCTTTGGC
 TAAAGAAGTATCATCACTTTGGGGTGGGGCAGCTGATTTAGCTAGTTCGAATAAAAACAGTAATTGCTG
 GAGAAGGTGATTTCAAACCTGAATCATATGAAGGAAGAAATATTTGGTTTGGGGTTCGAGAATTTGGA
 ATGGCATGTGCGATGAATGGAATCATGTTACATGGAGGGACACGAATATTCGGAAGTACGTTCTTTGT
 ATTTAGTGATTATTTAAAAGCGGCAATTCGTTTATCTGCCATTCAAAAATTACCAGTAATTTATGTTTTA
 ACTCACGATTCAGTGGCAGTAGGAAAAGACGGTCCAACCTCATGAACCAATTGAACAACCTAGCAAGTTTA
 AGAACGATTCCAAATGTCCAGGTATTTGCGCCAGCAGATGGAAATGAAACATCAGCAGCCTGGAAAAGT
 GCATTAGAAACCTTAGATAAAACCAACAATCTTAGTTTTGAGTCGCCAAAACCTTAGATACTACTACCAATTT
 CAAAAGAAAAAGTTTTTATGATGGGGTAGAAAAGGTGGGTATGTAGTTCAAGGAGCTGAAAATGAAGCC
 GATGGAATTTTTGATTGCCACAGGTTGAGAAGTAGGTTTAGCTTTGAAAGCTAAGAAGAACTACAAAAG
 AAAGGTAAAGATGTAATTGTAGTTTTCATTACCTAGTTGGGAAAAGATTTGAAGCTCAAAGTGAAGAATAT
 AAGAATACGGTCATCCCTCCAGAACTTAAGAAACGCATGACAATTGAAGCAGGTACAACATATGGTTGG
 GCTAAATATGCAGGAGATCATGGAGTGATGATAGGGATAGATGAATTCGGAATGTCAGCTCCAAGTGA
 CATAGTCCTAAGAGAACTAGGAATGAGTGTAGAAAATATCGTTGATAAATATTTGGAAAAATAA

TktA Protein Sequence

MYDQVDQLGVNTRLRLTSLDAIQRANSQHPLPMGAAPMAYVLWTRHLKINPKTHMNWVNRDRFVLSAGHGS
 ALLYSLAHLAGYDVSMDDLKNFREWKSNTPGHPEYGCTDGVEATTGPLGQGISMVGMAMAEHLGKFFNR
 EGYPMVDHYTYALIGDGLMEGVASEAASLAGHLKLGKLIALLYDSNGISLDGKTSASFTENVGARFEAYGWQY
 ILVEDGFNLEEIDKAIQAKAESDKPTIIEIKTTIGYSENQGTHKVGHSPLGEEGVHAHAKEVYNWNYPPFTVPE
 EVSQRKFKECLQDKGVKAENKWNEMFEAYKKEYSDLAQKFSDFSNKVPNTLGDILPQYGEDDSIATRAASQKA
 INALAKEVSSLWGAADLASSNKTVIAGEGDFQPESYEGRNIWFGVREFGMACAMNGIMLHGGTRIFGSTFFV
 FSDYLKAAIRLSAIQKLPVIYVLTHDSVAVGKDGPTHEPIQLASLRTPNVQVFRPADGNETSAAWKVALETLD
 KPTILVLSRQNLDTLPISKEKVFVGGYVQGAENEADGILIAATGSEVGLALKAKEELQKKGKDVIVVSLPS
 WERFEAQSEEYKNTVIPPELKKRMTIEAGTTYGWAKYAGDHGVMIGIDFGMSAPSDIVLRELGMVENIVDKY
 LEK

A.1.11 *TktA* Expression Construct Sequences

TktA Expression Construct Nucleotide Sequence

ATGGCACACCATCACCACCATCACAGCAGCGGTCTGGAAGTTCTGTTTCAGGGCCCCGTATGATCAAGTA
 GACCAATTGGGAGTAAATACTCTTAGAACATTATCAATCGACGCAATCCAAAGAGCTAACTCAGGTCAT
 CCAGGTTTACCAATGGGAGCGGCACCAATGGCATATGTCCTATGGACACGTCATTTAAAAATTAATCCT
 AAAACACATATGAACTGGGTAAATAGAGATAGATTTGTTTTATCGGCAGGACATGGTTCAGCTTTACTA
 TATAGTTTGGCACATTTAGCTGGATATGATGTTTCAATGGATGATTTGAAGAATTTAGAGAATGGAAA
 TCTAATACACCAGGACATCCTGAATATGGTTGTACCGATGGAGTGGAAGCTACAACAGGGCCATTAGG
 TCAAGGAATTTCAATGGCTGTAGGAATGGCGATGGCTGAAGCTCATCTAGGTAAGAAATTTAACCGTG
 AAGGTTACCCAGTAATGGATCATTATACATACGCTTTAATTGGCGATGGTGATTTAATGGAAGGTGTCG
 CTAGTGAAGCAGCGTCATTGGCAGGTCATTTGAAATTAGGTAAGTTAATTGCTTTGTATGATTCAAATG
 GAATTTCTTTAGATGGTAAAACCTCAGCATCATTTACAGAAAATGTAGGAGCTCGTTTTGAAGCTTATG
 GCTGGCAATATATTCTAGTAGAAGATGGTTTTCAATCTCGAAGAGATTGATAAAGCAATTTGTTCAAGCAA
 AAGCAGAAAAGTATAAGCCAACAATTATTGAAATAAAAACACAATTGGTTATGGTTTCAGAAAAATCAAG
 GTACACATAAAGTACATGGAAGTCCTCTTGGTGAAGAAGGAGTAGCACACGCTAAAGAAGTATATAACT
 GGAATTATCCACCATTTACAGTACCTGAAGAAGTAAAGTCAAAGATTTAAAGAATGCTTACAAGATAAGG
 GTGTTAAAGCAGAAAATAAGTGAATGAAATGTTTGAAGCATATAAAAAAGAATATTCAGATCTTGAC
 AAAAGTTTTCCGACGGCTTTTCTAACAAGTTTCAAATACTCTTGAGACATTTTACCACAATATGGAG
 AAGATGACAGTATCGCAACACGTGCAGCTAGTCAAAAAGCAATTAATGCTTTGGCTAAAGAAGTATCAT
 CACTTTGGGGTGGGGCAGCTGATTTAGCTAGTTTCAATAAAAACAGTAATTGCTGGAGAAGGTGATTT
 CAACCTGAATCATATGAAGGAAGAAATATTTGGTTGGGGTTCGAGAATTTGGAATGGCATGTGCGAT
 GAATGGAATCATGTTACATGGAGGGACACGAATATTCGGAAGTACGTTCTTTGTATTTAGTGATTATTT
 AAAAGCGGCAATTCGTTTATCTGCCATTCAAAAATTACCAGTAATTTATGTTTTAACTCACGATTCAGTG
 GCAGTAGGAAAAGACGGTCCAACCTCATGAACCAATTGAACCACTAGCAAGTTTAAAGAACGATTTCAAAT
 GTCCAGGATTTTCCGCCAGCAGATGGAAATGAAACATCAGCAGCCTGGAAAGTGGCATTAGAAACCTTA
 GATAAACCAACAATCTTAGTTTTGAGTCGCCAAAACCTTAGATACACTACCAATTTCAAAGAAAAAGTTT
 TTGATGGGGTAGAAAAAGGTGGGTATGTAGTTCAAGGAGCTGAAATGAAGCCGATGGAATTTTGATT
 GCCACAGTTTCAAGAAGTAGTTTTAGCTTTGAAAGCTAAAGAAGAACTACAAAAGAAAGGTAAAGATGTA
 ATTGATGTTTTATTACCTAGTTGGGAAAGATTTGAAGCTCAAAGTGAAGAATATAAGAATACGGTCATC
 CCTCCAGAATTAAGAAACGCATGACAATTGAAGCAGGTACAACATATGGTTGGGCTAAATATGCAGGA
 GATCATGGAGTGATGATAGGGATAGATGAATTCGGAATGTCAGCTCCAAGTGACATAGTCCTAAGAGA
 ACTAGGAATGAGTGTAGAAAATATCGTTGATAAATATTTGGAAAAATAA

TktA Expression Construct Protein Sequence With His₆ Tag

MAHHHHHSSGLEVLFGPPYDQVDQLGVNTRLRLTSLIDAIQRANSRHPGLPMGAAPMAYVLWTRHLKINPKT
 HMNVVNRDRFVLSAGHGSALLYSLAHLAGYDVSMDDLKNFREWKSNTPGHPEYGCTDGV EATTGPLGQGIS
 MAVGMAMAEHLGKKNREGYPVMDHYTYALIGDGLMEGVASEAASLAGHLKLGKLIALYDSNGISLDGKT
 SASFTENVGARFEAYGWQYILVEDGFNLEEIDKAIVQAKAESDKPTIIEIKTTIGYSENQGTHKVVHGSPLGEEG
 VAHAKEVYNWNYPPTVPEEVSQRKECLQDKGVKAENKWNEMFEAYKKEYSDLAQKFSDFSNKVPNTLG
 DILPQYGEDDSIATRAASQKAINALAKEVSSLWGGAADLASSNKTVIAGEGDFQPESEYGRNIWFGVREFGMAC
 AMNGIMLHGGTRIFGSTFFVFSYDLKAAIRLSAIQKLPVIYVLTHTDSVAVGKDGPTHEPIEQLASLRTIPNVQVFR
 PADGNETSAAWKVALETLDKPTILVLSRQNLDTLPISKEKVFDFGVEKGGYVVQGAENEADGILIATGSEVGLAL
 KAKEELQKKGKDVIVVSLPSWERFEAQSEYKNTVIPPELKKRMTIEAGTTYGWAKYAGDHGVMIGIDFEGMS
 APSDIVLRELGMSVENIVDKYLEK

TktA Expression Construct Protein Sequence Minus His₆ Tag

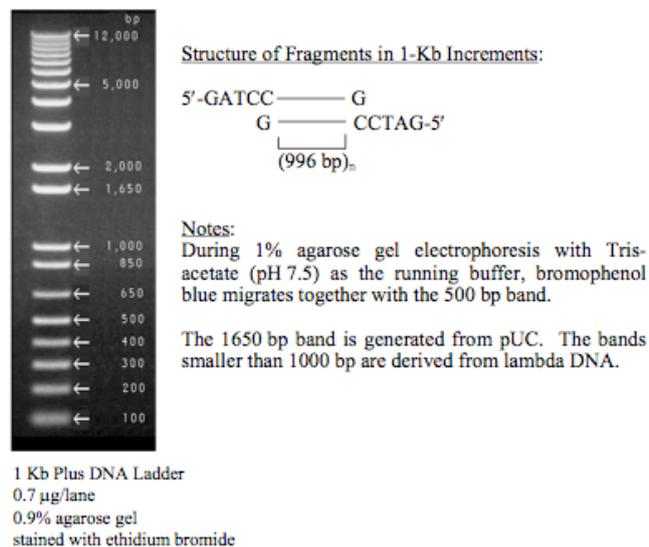
GPYDQVDQLGVNTRLRLSIDAIQRANSGHPGLPMGAAPMAYVLWTRHLKINPKTHMNWVNRDRFVLSAGHG
SALLYSLAHLAGYDVSMDLKNFREWKSNTPGHPEYGCTDGVEATTGPLGQGISMVGMAMAEHLGKKFN
REGYPVMDHYTYALIGDGLMEGVASEAASLAGHLKLGKLIYDSNGISLDGKTSASFTENVGARFEAYGWQ
YILVEDGFNLEEIDKAIVQAKAESDKPTIIEIKTTIGYGENQGTHKVHGSPLGEEGVAHAKEVYNWNYPPFTVP
EEVSQRKECLQDKGVKAENKWNEMFEAYKKEYSDLAQKFSDGFSNKVPNTLGDILPQYGEDDSIATRAASQK
AINALAKEVSSLWGGAADLASSNKTVIAGEGDFQPESYEGRNIWFGVREFGMACAMNGIMLHGGTRIFGSTFF
VFSDYLKAAIRLSAIQKLPVIYVLTHDSVAVGKDGPTHEPIEQLASLRTIPNVQVFRPADGNETSAAWKVALETL
DKPTILVLSRQNLDTLPISKEKVFVDFGVEKGGYVVQGAENEADGILIATGSEVGLALKAKEELQKKGKDVIVVSLP
SWERFEAQSEYKNTVIPPELKKRMTIEAGTTYGWAKYAGDHGVMIGIDFGMSAPSDIVLRELGMSVENIVDK
Y
LEK

A.2 DNA And Protein Molecular Weight Standards

A.2.1 DNA Molecular Weight Standards

1Kb Plus DNA Ladder (Invitrogen)

1Kb Plus DNA ladder (figure A.1) is supplied at a concentration of $1\mu\text{g}/\mu\text{l}$ and can be used to size DNA products ranging from 100bp - 12Kb. A working stock was made by diluting the marker 1 in 10 to a final concentration of $0.1\mu\text{g}/\mu\text{l}$ with 1xDNA loading dye. $0.3\mu\text{g}$ marker was used when running agarose gels for DNA analysis.



Cat. No. 10787-018

Figure A. 1: 1Kb Plus DNA ladder

(Image reproduced from Invitrogen product documentation)

A.2.2 Protein Molecular Weight Standards

SeeBlue® Plus2 Pre-Stained Standards (Invitrogen)

SeeBlue® Plus2 molecular weight standards (figure A.2) are supplied as a pre-stained solution of 12 proteins of known molecular weight for visualisation of sample migration and estimation of molecular weight by SDS-PAGE analysis. 5µl of SeeBlue® Plus2 standards were run for 15-lane 4-12% Bis-Tris SDS gels. All SDS gels were run in 1xMES buffer.

Protein	Approximate Molecular Weights (kDa)			
	Tris-Glycine	Tricine	NuPAGE® MES	NuPAGE® MOPS
Myosin	250	210	188	191
Phosphorylase	148	105	98	97
BSA	98	78	62	64
Glutamic Dehydrogenase	64	55	49	51
Alcohol Dehydrogenase	50	45	38	39
Carbonic Anhydrase	36	34	28	28
Myoglobin Red	22	17	17	19
Lysozyme	16	16	14	14
Aprotinin	6	7	6	n/a
Insulin, B Chain	4	4	3	n/a

NuPAGE® Novex®
Bis-Tris 4-12% Gel

Figure A. 2: SeeBlue Plus2 protein molecular weight standards
(Image reproduced from Invitrogen product documentation)

A.3 Selected Standard Operating Protocols (SOPs)

Devised By The Mitchell Group

A.3.1 Preparation Of Pneumococcal Genomic DNA

Notes

To avoid contamination it is best to use filter tips throughout this work.

It may be useful to prepare DNA from sterile medium in parallel with proper samples to confirm the absence of contamination during the process e.g. when investigating the presence/absence of gene(s) in several strains at one time.

Autoclave buffers beforehand to limit contamination of samples.

Protocol

1. Grow culture. An overnight culture of 20ml BHI typically gives sufficient yield for most purposes. Aseptically streak a loopful onto a BAB plate and add optochin disc (incubate O/N, this will confirm the purity and identity of the culture).
2. Spin down culture, 4000rpm, 15mins, 4°C. Sometimes culture doesn't pellet well after first spin. If so, a second spin as above usually works.
3. Remove supernatant. Pellet is fragile so be careful not to disturb cells.
4. Resuspend pellet in 1ml lysis buffer (see below).
5. Transfer to eppendorf and incubate at 37°C for 1hr.
6. Add proteinase K to final concentration of 100µg/ml (5µl/ml from a 20mg/ml stock).
7. Incubate in H₂O bath at 50°C for 3hrs.
8. Add RNase A to a final concentration of 20µg/ml (2µl/ml of a 10mg/ml stock).***
9. Incubate at 37°C for 30mins.
10. Add equal volume of phenol:chloroform:IAA (25:24:1). Due to volume involved may have to split samples into two tubes. Several types of eppendorf are not suitable for use with phenol (they degrade/melt leading to loss of sample and phenol contamination of lab equipment).

If in doubt, cryotubes are usually fine. Take from the lower phase of the phenol:chloroform:IAA solution or mix well first.

11. Mix by inverting the tubes sharply several times. Vortex mixing is not required and may cause shearing of the DNA and loss of quality.
12. Spin at 13000rpm, 3mins.
13. Remove as much of the upper phase as possible without disturbing the lower phase, and add to fresh eppendorf. Pool samples if previously split.
14. Add 0.2 volumes 10M ammonium acetate (e.g. 20 μ l to 100 μ l).
15. Add ~600 μ l absolute ethanol (analytical reagent grade)
16. Gently invert the tubes and the 'clouds' of DNA often but not always become visible.
17. Spin for 30mins at 13000rpm. This pellets the DNA (It may be useful to put all the tubes in at the same angle so you know where the pellet will be).
18. Remove supernatant. Be careful to leave DNA pellet.
19. Air-dry for 15-20mins upside-down on a paper towel. Over-drying may make it hard to re-dissolve the DNA.
20. Remove any remaining ethanol.
21. Re-suspend in TE buffer (100-300 μ l). If gently pipetting up and down to re-suspend take care not to remove DNA.
22. Incubate in 65°C H₂O bath for 10mins. Gently flick bottom of tube with fingers every few minutes to help dissolve DNA. As well as helping to dissolve the DNA it may also inactivate DNase present in the sample.
23. Store at 4°C. Store at -20°C long-term, but avoid frequent freeze thawing, which leads to shearing of the DNA (store in small aliquots).

Lysis Buffer (1ml. Make up fresh before use)

10 μ l 1M Tris pH 8.0, 200 μ l 500mM EDTA, 50 μ l 10% SDS (w/v), 740 μ l ddH₂O

***RNase must be boiled to ensure it is DNase free.

A.3.2 Western Blot

Buffers

Transfer Buffer: 25ml 1xNuPAGE transfer buffer (Invitrogen), 50ml methanol, 425ml dH₂O.

Tris NaCl pH7.4: Tris base 1.2g, NaCl 8.7g. Make to 1L with dH₂O and pH with 800µl conc. HCl.

Developer: Dissolve 30mg 4chloro-1-naphthol in 10ml methanol, add 40ml Tris NaCl pH7.4, add 30µl H₂O₂ (30% w/v)

Day 1

1. Run samples on SDS-PAGE
2. Equilibrate gel for 30mins in transfer buffer
3. Cut 2 x (gel size) filter paper and 1 x membrane per blot
4. Soak membrane, filter-paper and pads in transfer buffer.
5. Assemble cassette as follows: pads/paper/gel/membrane/paper/pads. Place in gel-tank and fill with transfer buffer. Surround with cold water to keep cool.
6. Blot at 30V for 1hr using Xcell II blotting module (Invitrogen).
7. Transfer membrane to 3% skimmed milk (1.5g in 50ml Tris NaCl pH 7.4) and block O/N at 4°C.

Day 2

1. Transfer membrane to 3% skimmed milk containing 1/x dilution of 1^o antibody*
2. Shake at 37°C for 2-3hr

3. Wash x 4 with 40ml Tris NaCl pH7.4 for 5mins
4. Transfer membrane to 3% skimmed milk containing 1/x dilution of 2° antibody* (HRP labelled)
5. Shake at 37°C for 1hr
6. Wash x 4 with 40ml Tris NaCl pH7.4 for 5mins.
7. Develop in dark and stop with dH₂O

X = dilution factor appropriate for Ab being used - 1° Ab should be titrated, instructions for use of 2° Ab are provided by supplier.

* Ab solutions can be stored at -20°C and reused.

A.4 Molecular Biology Kit Protocols

N.B. All buffers used were supplied with the relevant kits. In all cases, the supplied elution buffer was substituted with DNase/RNase free PCR grade H₂O.

13,000rpm = 17,900 x g

A.4.1 QIAquick[®] PCR Purification Kit Protocol

- Add 5 volumes of buffer PB to 1 volume of the PCR sample and mix.
- Place a QIAquick spin column in a provided 2ml collection tube.
- To bind DNA, apply the sample to the QIAquick column and centrifuge at 13,000rpm for 30-60secs.
- Discard flow-through. Place QIAquick column back into the same tube.
- To wash, add 0.75ml buffer PE to the QIAquick column and centrifuge at 13,000rpm for 30-60secs.
- Discard flow-through and place the column back in the same tube. Centrifuge the column for an additional 1min.
- Place the QIAquick column in a clean 1.5ml microcentrifuge tube.
- To elute DNA, add 50µl warmed sterile DNase/RNase free H₂O to the centre of the QIAquick membrane, let stand for 1min, then centrifuge at 13,000rpm for 1min.

A.4.2 QIAquick[®] Gel Extraction Kit Protocol

- Excise the DNA fragment from the agarose gel with a clean, sharp scalpel.
- Weigh the gel slice in a colourless tube. Add 3 volumes of buffer QG to 1 volume of gel (100mg ~ 100µl).
- Incubate at 50°C for 10min (or until the gel slice has completely dissolved). To help dissolve gel, mix by vortexing every 2-3min during the incubation.

- After the gel slice has dissolved completely, check that the colour of the mixture is yellow (similar to buffer QG without dissolved agarose).
- Add 1 gel volume of isopropanol to the sample and mix. Do not centrifuge the sample at this stage.
- Place a QIAquick spin column in a provided 2ml collection tube.
- To bind DNA, apply the sample to the QIAquick column and centrifuge at 13,000rpm for 1min.
- Discard flow-through and place QIAquick column back in the same collection tube.
- Recommended: Add 0.5ml of buffer QG to QIAquick column and centrifuge at 13,000rpm for 1min. This will remove all traces of agarose.
- To wash, add 0.75ml of buffer PE to the QIAquick column and centrifuge at 13,000rpm for 1min.
- Discard the flow-through and centrifuge the QIAquick column for an additional 1min at 13,000rpm.
- Place the QIAquick column into a clean 1.5ml microcentrifuge tube.
- To elute DNA add 50µl warmed sterile DNase/RNase free H₂O to the centre of the QIAquick membrane, let stand for 1min, then centrifuge at 13,000rpm for 1min.

A.4.3 QIAprep[®] Miniprep Kit Protocol

Note

This protocol is designed for purification of up to 20µg of high-copy plasmid DNA from 1-5ml overnight cultures of *E. coli* in LB medium.

Procedure

- Resuspend pelleted bacterial cells in 250µl buffer P1 and transfer to a microcentrifuge tube.
- Add 250µl buffer P2 and mix thoroughly by inverting the tube 4-6 times. Mix gently by inverting the tube. Do not vortex, as this will result in shearing of genomic DNA. If necessary, continue inverting the

tube until the solution becomes viscous and slightly clear. Do not allow the lysis reaction to proceed for more than 5min.

- Add 350µl buffer N3 and mix immediately and thoroughly by inverting the tube 4-6 times. To avoid localised precipitation, mix the solution thoroughly immediately after addition of buffer N3. The solution should become cloudy.
- Centrifuge for 10min at 13,000rpm in a table-top microcentrifuge. A compact white pellet will form.
- Apply the supernatants from step 4 to the QIAprep spin column by decanting or pipetting.
- Centrifuge at 13,000rpm for 30-60secs. Discard the flow-through.
- Recommended: Wash the QIAprep spin column by adding 0.5ml buffer PB and centrifuging at 13,000rpm for 30-60secs. Discard the flow-through. This step is necessary to remove trace nuclease activity.
- Wash the QIAprep spin column by adding 0.75µl buffer PE and centrifuging at 13,000rpm for 30-60secs.
- Discard the flow-through, and centrifuge for an additional 1min to remove residual wash buffer.
- Place the QIAprep spin column into a clean 1.5ml microcentrifuge tube. To elute DNA add 50µl warmed sterile DNase/RNase free H₂O to the centre of the QIAquick membrane, let stand for 1min, then centrifuge at 13,000rpm for 1min.

A.5 Protein Manipulation Protocols

A.5.1 PD-10 Desalting Column Protocol

N.B. The following is a copy of the Gravity Protocol for buffer exchange of proteins from the instruction documentation supplied by the manufacturer (GE Healthcare) with the product.

1. PD-10 desalting column preparation

- Remove the top cap and pour off the column storage solution
- Cut the sealed end of the column at notch

2. Column equilibration

- Fill up the column with equilibration buffer and allow the equilibration buffer to enter the packed bed completely
- Repeat 4 times
- Discard the flow-through

Note: about 25ml equilibration buffer should be used in total for all three steps

3. Sample application

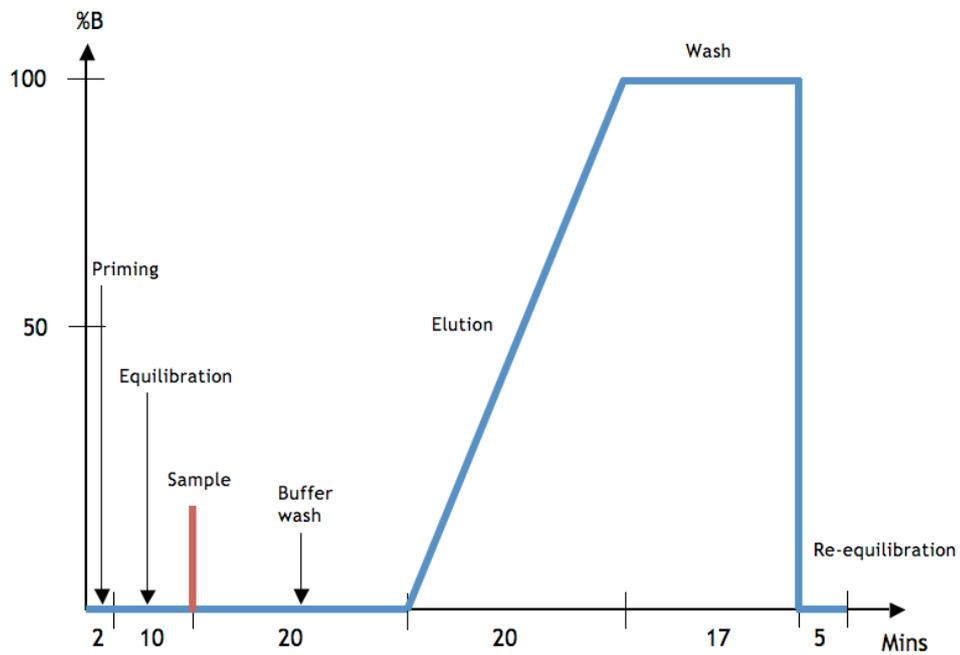
- Add maximum 2.5ml of sample to the column
- For sample volumes less than 2.5ml, add equilibration buffer to adjust the volume up to 2.5ml *after* the sample has entered the packed bed completely
- Let the sample or equilibration buffer enter the packed bed completely
- Discard the flow-through

4. Elution

- Place a test tube for sample collection under the column
- Elute with 3.5ml buffer and collect the eluate.

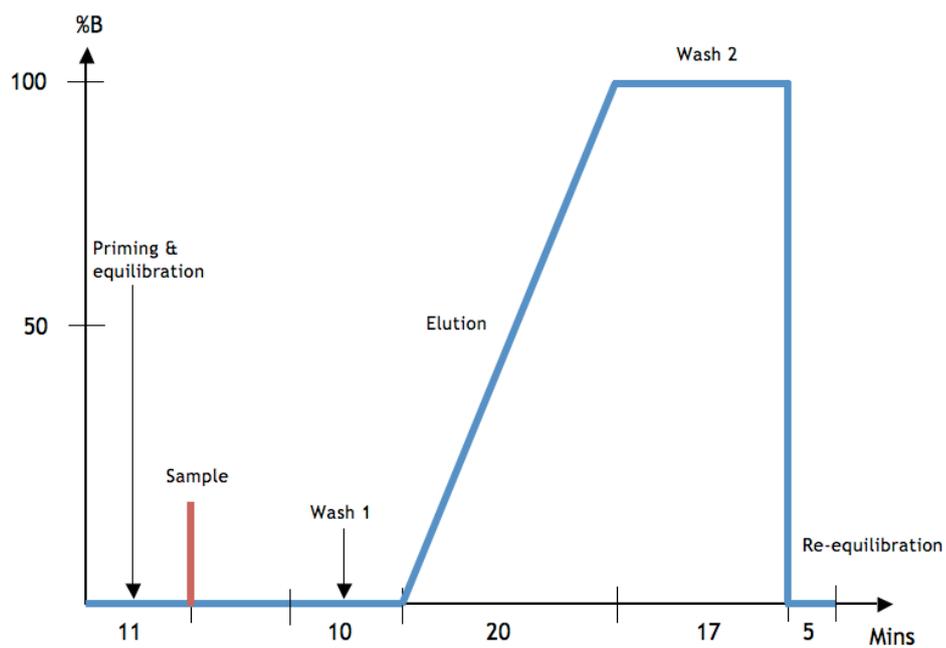
A.5.2 Ni-Affinity IMAC Template

N.B. These templates are replicated from the ÄKTAprime plus cue cards.



Total separation time = 74 min + sample application time.

A.5.3 Anion Exchange Chromatography Template



Total separation time = 63 mins + sample application time.

Conference Contributions

Europneumo 9. 2009. Bern, Switzerland. **Oral and poster presentations given:** *Structural Studies Of PhtD, A Member Of The Pneumococcal Histidine Triad Family of Proteins.*

12th, 13th, 14th Northern Protein Structure Workshops. 2007, 2008, 2009. Carlisle, UK. **Oral Presentations.**

9th International School on the Crystallography of Macromolecules. 2008. Como, Italy. **Poster Presentation:** *Using Macromolecular Protein Crystallography As An Approach To Studying Virulence In Streptococcus pneumoniae.*

University of Glasgow FBLs Divisional Poster Sessions 2007, 2008, 2009. **Poster Presentations.**

Australia Synchrotron 2009. **Poster Presentation:** *Targeting Proteins From Pathogenic Bacteria* (contributing author).

TktA Crystallisation Paper

crystallization communications

Acta Crystallographica Section F
Structural Biology
and Crystallization
Communications
ISSN 1744-3091

Matt Horsham,^a Harriet Saxby,^b
James Blake,^b Neil W. Isaacs,^b
Tim J. Mitchell^a and Alan
Riboldi-Tunncliffe^{b*}

^aIBLS I&I, University of Glasgow,
120 University Place, Glasgow G12 8TA,
Scotland, and ^bDepartment of Chemistry,
University of Glasgow, 120 University Place,
Glasgow G12 8TA, Scotland

Correspondence e-mail:
alan.r.t@synchrotron.org.au

Received 23 December 2009
Accepted 31 March 2010



© 2010 International Union of Crystallography
All rights reserved

Expression, purification, crystallization and preliminary X-ray crystallographic data from TktA, a transketolase from the lactic acid bacterium *Lactobacillus salivarius*

The enzyme transketolase from the lactic acid bacterium *Lactobacillus salivarius* (subsp. *salivarius* UCC118) has been recombinantly expressed and purified using an *Escherichia coli* expression system. Purified transketolase from *L. salivarius* has been crystallized using the vapour-diffusion technique. The crystals belonged to the trigonal space group $P3_121$, with unit-cell parameters $a = b = 75.43$, $c = 184.11$ Å, and showed diffraction to 2.3 Å resolution.

1. Introduction

The transketolase enzyme was first identified in *Saccharomyces cerevisiae* (Turner, 2000). The enzyme acts as a link between glycolysis and the pentose-phosphate pathway and relies on divalent magnesium ions and thiamine pyrophosphate for activity. Transketolase catalyses the reversible synthesis of higher carbon sugars in the pentose-phosphate pathway, mediating the transfer of a two-carbon ketol unit from a ketose to an aldose (specifically from *D*-xylulose 5-phosphate to *D*-ribose 5-phosphate, generating *D*-sedulose 7-phosphate and *D*-glyceraldehyde 3-phosphate in the process; Turner, 2000). Here, we present work on the protein LSL_1946 from the lactic acid bacterium *L. salivarius*, which has been identified as the transketolase enzyme TktA. Here, we present data on the over-expression, purification, crystallization and preliminary X-ray diffraction data analysis of LSL_1946.

2. Materials and methods

2.1. Protein expression and purification

The PCR product coding for the transketolase gene LSL_1946 was cloned into the expression vector pOPIN⁺ using In-Fusion cloning technology (Clontech) in-frame with an N-terminal His₆ tag separated by a HRV-C3 protease cleavage site. The vector was transformed into competent *Escherichia coli* DH5- α cells and grown at 310 K on LB agar plates supplemented with 54 $\mu\text{g ml}^{-1}$ carbenicillin. A single colony was used to inoculate 50 ml LB supplemented with 54 $\mu\text{g ml}^{-1}$ carbenicillin and the culture was grown overnight in a shaking incubator operated at 310 K and 170 rev min^{-1} . Plasmids harvested from the overnight culture using a Qiagen Plasmid Mini-prep kit (Qiagen) were transformed into the protein-expression strain *E. coli* 2(DE3)pLysS and grown at 310 K on LB agar plates supplemented with 54 $\mu\text{g ml}^{-1}$ carbenicillin. A single colony was used to inoculate 50 ml LB supplemented with 54 $\mu\text{g ml}^{-1}$ carbenicillin and 34 $\mu\text{g ml}^{-1}$ chloramphenicol and the culture was grown overnight at 310 K and 170 rev min^{-1} . Two 500 ml cultures of LB medium were each inoculated with 5 ml of this starter culture and incubated at 310 K and 170 rev min^{-1} until an OD₆₀₀ of 0.6 was reached. The cultures were then induced with 0.5 ml 1 M isopropyl β -*D*-1-thiogalactopyranoside (IPTG) and incubated at 295 K for the remainder of the 24 h period. Cells were harvested by centrifugation at 5000g and resuspended in 50 ml buffer A (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole) supplemented with 0.1% Tween. Cells were

crystallization communications

Table 1
Data-collection statistics for TktA.

Values in parentheses are for the outer shell.	
Source	ID14-1, ESRF, Grenoble
Wavelength (Å)	0.904
Resolution (Å)	36.54-2.30 (2.42-2.30)
Space group	$P3_121$
Unit-cell parameters (Å, °)	$a = b = 75.43, c = 184.11,$ $\alpha = \beta = 90.0, \gamma = 120.0$
R_{merge}	0.126 (0.302)
Total No. of reflections	5179
No. of unique reflections	2197
Average redundancy	2.2
Completeness (%)	92.8 (93.6)

lysed by five passes through a French pressure cell set at 6.5 MPa. The cell lysate was centrifuged at 6500g to remove cell debris and the supernatant was then centrifuged at 40 000g for a further 30 min. The cleared lysate was applied onto a 1 ml HisTrap HP nickel-affinity column (GE Healthcare) and the protein was eluted from the column with a linear gradient to buffer B (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 500 mM imidazole) on an ÄKTAprius plus (GE Healthcare). The purified protein was dialysed overnight into buffer A to remove excess imidazole.

2.2. Crystallization and data collection

The protein was concentrated to 13.2 mg ml^{-1} and crystallized using the sitting-drop vapour-diffusion technique, mixing equal volumes (500 μl) of protein solution and reservoir solution to form the drop (using a HoneyBee 8-1 robot). Crystallization trials included Cryo Screens 1 and 2, Crystal Screens 1 and 2 and PEG/Ion Screens 1 and 2. Crystals grew over a period of 3 d at 293 K in many conditions. The best-quality crystals were observed in condition 6 of PEG/Ion Screen 1 (Hampton Research), which consisted of 0.2 M sodium chloride, 20% (w/v) PEG 3350. The crystals appeared as hexagonal rods of approximately $200 \times 40 \times 40 \mu\text{m}$ in size (Fig. 1). The crystals were flash-cooled in a stream of nitrogen gas maintained at 110 K, using dried paraffin oil as the cryoprotectant (Riboldi-Tunciciliffe & Hilgenfeld, 1999).

Diffraction data were collected from a single crystal on beamline ID14-1 at the ESRF using an ADSC Q210 CCD detector with a crystal rotation of 1° per frame. A summary of the data-collection statistics is given in Table 1. The data were processed using the



Figure 1
TktA crystals. The drop volume was 1 μl .

program *MOSFLM* (Leslie, 1992) and scaled using the program *SCALA* (Evans, 1993).

3. Results and discussion

The crystal belonged to the trigonal space group $P3_121$, with unit-cell parameters $a = b = 75.43, c = 184.11 \text{ \AA}$, and diffracted to 2.3 Å resolution (Figs 2a and 2b). The calculated Matthews coefficient (V_M) for one molecule of the protein in the asymmetric unit is $2.03 \text{ \AA}^3 \text{ Da}^{-1}$, with 39.36% of the unit cell occupied by solvent

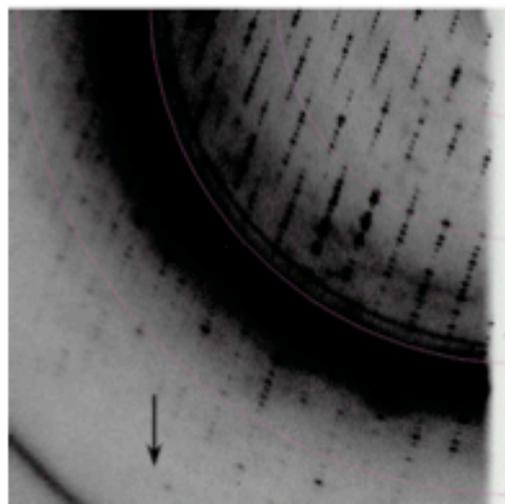
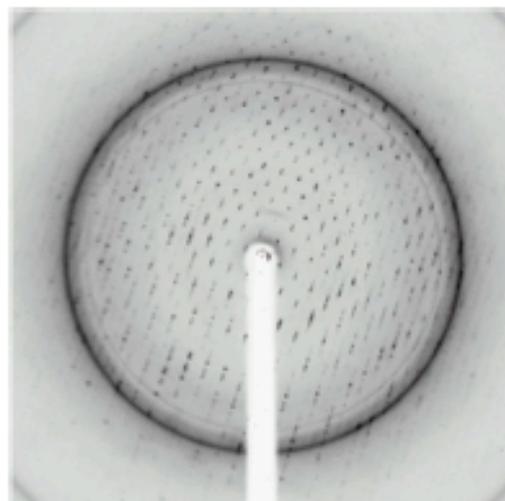


Figure 2
Diffraction image for a single crystal of TktA. The arrow indicates the limit of diffraction at 2.3 Å resolution. The outer arc shows the diffraction limit at 2.5 Å.

(Matthews, 1968); the biologically relevant dimer is created through a symmetry operation. The structure has been solved by the molecular-replacement method using the *PHENIX* suite of crystallographic programs (Adams *et al.*, 2002) and is currently undergoing refinement.

We would like to thank the beamline staff of ID14-1 at the ESRF, Grenoble for help and support during data collection and Dr Martin Walsh (ESRF BM-14) for supplying the expression construct.

References

- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst. D* **58**, 1948–1954.
- Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 114–122. Warrington: Daresbury Laboratory.
- Leslie, A. G. W. (1992). *Int. CCP4/ESF-EACBM News. Protein Crystallogr.* **26**.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Robold-Tunncliffe, A. & Hilgenfeld, R. (1999). *J. Appl. Cryst.* **32**, 1003–1005.
- Turner, N. J. (2000). *Curr. Opin. Biotechnol.* **11**, 527–531.

PpmA Crystallisation Paper (Draft Manuscript)

Expression, Purification, Crystallisation and Preliminary Diffraction Data For PpmA, A predicted Cis-Trans Isomerase From The Pathogenic Bacteria *Streptococcus pneumoniae*

Matt Horsham¹, Neil W. Isaacs², Tim Mitchell¹ and Alan Riboldi-Tunncliffe^{3*}

¹FBLIS I&I, Glasgow Biomedical Research Centre, University of Glasgow, Glasgow G12 8TA

²Dept of Chemistry, Glasgow Biomedical Research Centre, University of Glasgow, Glasgow G12 8TA

³ Australian Synchrotron, 800 Blackburn Rd Clayton, VIC 3168, Melbourne, Australia

1. Abstract

The surface associated protein PpmA (Putative Protease Maturation Protein A) from the human pathogen *Streptococcus pneumoniae* has been recombinantly expressed using an *Escherichia coli* expression system. Purified PpmA from *S. pneumoniae* has been crystallised utilising the sitting-drop vapour diffusion technique. The crystals, which appear as large plates belong to the space-group $P2_12_12_1$ with unit cell parameters $a=65.0\text{\AA}$, $b=105.0\text{\AA}$, $c=120.0\text{\AA}$, $\alpha=\beta=\gamma=90.0^\circ$. The crystals show diffraction to beyond 2.5 Angstroms resolution.

2. Introduction

S. pneumoniae is a common commensal bacteria found in the nasopharynx of the vast majority of the human population. A gram positive facultative anaerobe, it can however cause a variety of diseases ranging from mild (otitis-media) to life-threatening (pneumonia, meningitis) when relocated to an atypical niche within the body.

Current available pneumococcal vaccines in use are either the 23-valent capsular polysaccharide vaccine, or a 7-valent capsular polysaccharide - protein conjugation vaccine, where the capsular polysaccharide elements are linked to a carrier protein such as diphtheria toxoid. However, the latter technique is limited by the number of capsular polysaccharides that can be successfully conjugated to the carrier-protein, raising the possibility of a population shift to strains not encompassed by the current vaccine, whilst the former method is less effective in the major at risk groups where the immune system is underdeveloped or weakened¹⁸². As a result there has been increased interest in the possibility of developing a protein-based vaccine which could overcome current vaccine limitations⁶⁵.

S. pneumoniae produces two proteins which exhibit homology to peptidyl-prolyl isomerases (PPIases) of which the putative proteinase maturation protein A (PpmA) is one; the other being streptococcal rotamase A (SlrA)¹⁵⁶. These PPIases are responsible for the *cis-trans* isomerisation of proteins at proline residues, aiding protein folding¹⁸³. PpmA shows homology to PrtM, a PPIase from lactic acid bacteria. However PpmA has not been shown to have any PPIase activity, nor is it currently known which protein PpmA activates^{156, 157}. It has been shown that in a murine model, absence of PpmA results in

decreased virulence of the pneumococcus mediated by a decrease in the ability of adherence and an increase in susceptibility to opsonophagocytosis in PpmA deficient pneumococci¹⁵⁷. PpmA is widely conserved throughout various strains of *S. pneumoniae*¹⁵⁷ making it a possible candidate for inclusion in a novel protein-based pneumococcal vaccine.

3. Materials and methods

3.1 Protein expression and purification

The gene coding for PpmA was amplified from gDNA of the TIGR4 strain of *S. pneumoniae* by PCR. The obtained PCR product was cleaned using a PCR Purification Kit (Qiagen) and cloned into the expression vector pOPINF¹¹⁹ using the InFusion cloning system (Clontech)¹²⁵ in-frame with an N-terminal His₆ tag separated by a HRV-C3 protease cleavage site. The plasmid was transformed into competent *E. coli* DH5- α cells (Invitrogen) and grown at 37°C on LB agar plates supplemented with 54 μ g/ml carbenicillin. A single colony was used to inoculate 50ml of LB supplemented with 54 μ g/ml carbenicillin, and the culture was grown overnight in a shaking incubator operating at 37°C and 170rpm. Plasmids, harvested from the overnight culture using the Qiagen Plasmid Miniprep kit (Qiagen), were transformed into the protein expression strain *E. coli* Rosetta 2(DE3)pLysS and grown at 37°C on LB agar plates supplemented with 54 μ g/ml carbenicillin. A single colony was used to inoculate 50ml of LB supplemented with 54 μ g/ml carbenicillin and 34 μ g/ml chloramphenicol, and the culture was grown overnight in a shaking incubator at 37°C and 170rpm. Two 500ml cultures of LB media were each inoculated with 5ml of the starter culture and incubated in a shaking incubator at 37°C and 170rpm until an O.D₆₀₀ of 0.6 was reached. Cultures were then induced by addition of 0.5ml 1M isopropyl- β -D-1-thiogalactopyranoside (IPTG), and incubated at 22°C for the remainder of the 24-hour period. Cells were harvested at 4000g for 30mins at 4°C and re-suspended in 50ml buffer A (50mM Tris-HCl pH7.5, 500mM NaCl, 20mM Imidazole) supplemented with 0.1% Tween. Cells were lysed by passing 5 times through a French Pressure Cell set at 950psi. The cell lysate was centrifuged at 6500g for 30mins at 4°C and the supernatant then centrifuged at 40000g for a further 30mins at 4°C. The resulting supernatant was filtered through a 0.2 μ m syringe filter. The cleared cell lysate was then applied to a 1ml HisTrap HP Nickel affinity column (GE Healthcare) and the protein was eluted from the column with buffer B (50mM Tris-HCl pH7.5, 500mM NaCl, 500mM Imidazole) on an Äkta PrimePlus (GE Healthcare) using a linear gradient. The purified protein was transferred back into buffer A using a PD-10 Desalting Column (GE Healthcare) to remove excess imidazole. The Ni-affinity purified batch of PpmA was divided into two equal aliquots; one retained the His₆ tag during crystal trials, whilst the other had the tag removed. Cleavage of the His₆ tag was performed overnight at 4°C by addition of 2 μ l HRV-C3 protease, Dithiothreitol (DTT) and EDTA at a final concentration of 10mM and 1mM respectively. Reverse IMAC was used to separate the pure PpmA from the cleaved His-tags. Purified PpmA was concentrated for crystal trials using centrifugal concentrators with a MWCO of 10kDa (Millipore). Concentrated PpmA was then used to set up crystal trials.

3.2 Crystallisation and data-collection

Recombinantly expressed, purified PpmA protein was subjected to crystal trials at concentrations of 30mg/ml and 16mg/ml using the sitting-drop vapour diffusion technique mixing a 1:1 ratio of protein to reservoir solution, stored at 20°C and monitored for crystal growth. Screening for crystals showed crystals to be present in around 50% of conditions from PEG/Ion screen (Hampton Research). The higher concentration of PpmA produced showers of small micro-crystals all in a large crystalline mass. The lower protein concentration produced single plates varying in size (*Fig1*). Crystals grew within a varying period of 2 days - 2 weeks. Crystals were flash cooled in a liquid nitrogen stream at 110K, cryo-protected with dried paraffin oil¹⁸⁴ and tested for protein diffraction in-house on a Rigaku 007 x-ray generator (Rigaku). Following further optimisation of crystallisation conditions the best crystal was obtained from the condition comprising 31% w/v PEG 4000, 0.18M ammonium acetate and 0.1M sodium acetate. The crystal showed diffraction to beyond 2.5Å resolution (*Fig2*). Good, clean diffraction was obtained when shooting through the large, flat face of the crystal, but diffraction quality decreased greatly when shooting along the thin edge. Despite this, a full dataset was collected and it was possible to successfully index the crystal.

Indexing of the PpmA crystal was performed using the crystallographic program MOSFILM.¹³⁴ Spots were picked by hand over 10 images due to poor diffraction quality in a number of the images. The preliminary data-collection statistics for this crystal are given in Table 1.

Resolution (Å)	2.5Å
Space group	$P2_12_12_1$
Unit cell parameters (Å, °)	$a=65.0, b=105.0,$ $c=120.0,$ $\alpha=\beta=\gamma=90.0$

Table 1 Preliminary data collection statistics for single crystal of PpmA

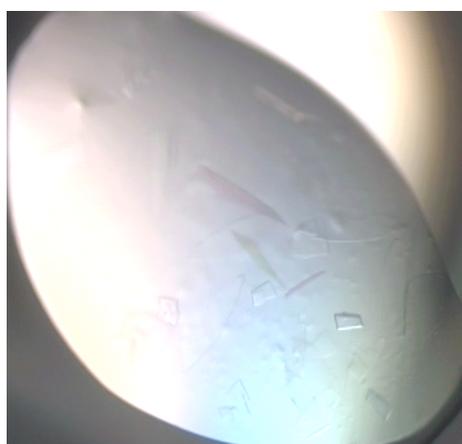


Fig1 Crystals of PpmA from *S. pneumoniae*

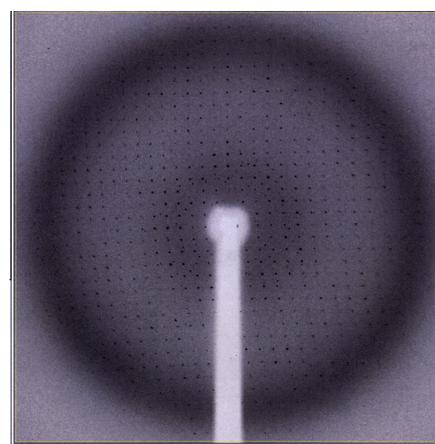


Fig2 In-house diffraction image for PpmA shot at 0° rotation (crystal shot through large flat surface)

4. Results and discussion

We have successfully crystallised the pneumococcal protein PpmA using an *E. coli* recombinant expression system. PpmA crystals grew over a period of days using the sitting-drop vapour diffusion method. Crystals were present as thin plates, which show diffraction to beyond 2.5Å. Indexing of the crystal has been possible, showing PpmA to belong to the space-group $P2_12_12_1$, with unit-cell dimensions $a=65.0\text{Å}$, $b=105.0\text{Å}$ $c=120.0\text{Å}$, $\alpha=\beta=\gamma=90.0^\circ$. Estimation of solvent content of the PpmA crystals by derivation of the Matthews coefficient suggests 2 molecules of PpmA per asymmetric unit. This suggests that PpmA is present as a dimer; previous experiments in our lab using gel-filtration studies confirm this to be true. Attempts are ongoing at present to obtain phases and improve crystal quality for PpmA, which will be required to determine the crystallographic structure of PpmA from *S. pneumoniae*.

5. References

1. Ogunniyi, A.D., et al., *Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of Streptococcus pneumoniae*. Infect Immun, 2007. **75**(1): p. 350-7.
2. Obaro, S.K., *The new pneumococcal vaccine*. Clin Microbiol Infect, 2002. **8**(10): p. 623-33.
3. Hermans, P.W., et al., *The streptococcal lipoprotein rotamase A (SlrA) is a functional peptidyl-prolyl isomerase involved in pneumococcal colonization*. J Biol Chem, 2006. **281**(2): p. 968-76.
4. Rahfeld, J.U., et al., *Confirmation of the existence of a third family among peptidyl-prolyl cis/trans isomerases. Amino acid sequence and recombinant production of parvulin*. FEBS Lett, 1994. **352**(2): p. 180-4.
5. Overweg, K., et al., *The putative proteinase maturation protein A of Streptococcus pneumoniae is a conserved surface protein with potential to elicit protective immune responses*. Infect Immun, 2000. **68**(7): p. 4180-8.
6. Berrow, N.S., et al., *A versatile ligation-independent cloning method suitable for high-throughput expression screening applications*. Nucleic Acids Res, 2007. **35**(6): p. e45.
7. Zhu, B., et al., *In-fusion assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations*. Biotechniques, 2007. **43**(3): p. 354-9.
8. Riboldi-Tunicliffe, A. and R. Hilgenfeld, *Cryocrystallography with oil - an old idea revived*. J. Appl. Cryst, 1999(**32**): p. 1003-1005.
9. Pflugrath, J.W., *The finer things in X-ray diffraction data collection*. Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 10): p. 1718-25.

Manuscripts In Preparation

N.B. All author lists and titles are proposed running titles at present and may be subject to change.

1. Horsham, M; Riboldi-Tunncliffe, A; Smith, B; Burchmore, R; Kelly, S; Nutley, M; Mitchell, T.J. “Structural Interactions Of Pneumococcal Histidine Triad D (PhtD) from *S. pneumoniae* With Zinc Atoms.”
2. Horsham, M; Mitchell, T.J & Riboldi-Tunncliffe, A. “2.3 Ångstroms Crystal Structure Of TktA, A Transketolase From The Lactic Acid Bacterium *Lactobacillus salivarius*.”
3. Riboldi-Tunncliffe, A; Horsham, M; Mitchell, T.J. “1.3 Ångstroms Crystal Structure Of EF0337, An Ankyrin Repeat Protein From *Enterococcus faecalis*.”

References

1. Wizemann, T.M., et al., Use of a whole genome approach to identify vaccine molecules affording protection against *Streptococcus pneumoniae* infection. *Infect Immun*, 2001. **69**(3): p. 1593-8.
2. Phillips, D.M., Images In Clinical Medicine. *Streptococcus pneumoniae*. *N Engl J Med*, 1993. **329**(447).
3. Jonsson, S., et al., Phagocytosis and killing of common bacterial pathogens of the lung by human alveolar macrophages. *J Infect Dis*, 1985. **152**(1): p. 4-13.
4. Paterson, G.K. and T.J. Mitchell, Innate immunity and the pneumococcus. *Microbiology*, 2006. **152**(Pt 2): p. 285-93.
5. Lund, E., and J. Henrichsen. , Laboratory diagnosis, serology and epidemiology of *Streptococcus pneumoniae*. In T. Bergan and J. R.Norris (ed.), *Methods in microbiology*. . 1978.
6. Henrichsen, J., Six newly recognized types of *Streptococcus pneumoniae*. *J Clin Microbiol*, 1995. **33**(10): p. 2759-62.
7. Bogaert, D., R. De Groot, and P.W. Hermans, *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect Dis*, 2004. **4**(3): p. 144-54.
8. Pericone, C.D., et al., Inhibitory and bactericidal effects of hydrogen peroxide production by *Streptococcus pneumoniae* on other inhabitants of the upper respiratory tract. *Infect Immun*, 2000. **68**(7): p. 3990-7.
9. Kluytmans, J., A. van Belkum, and H. Verbrugh, Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. *Clin Microbiol Rev*, 1997. **10**(3): p. 505-20.
10. Principi, N., et al., Risk factors for carriage of respiratory pathogens in the nasopharynx of healthy children. Ascanius Project Collaborative Group. *Pediatr Infect Dis J*, 1999. **18**(6): p. 517-23.
11. Bogaert, D., et al., Pneumococcal carriage in children in The Netherlands: a molecular epidemiological study. *J Clin Microbiol*, 2001. **39**(9): p. 3316-20.
12. Park, I.H., et al., Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. *J Clin Microbiol*, 2007. **45**(4): p. 1225-33.
13. Attali, C., et al., The interaction of *Streptococcus pneumoniae* with plasmin mediates transmigration across endothelial and epithelial monolayers by intercellular junction cleavage. *Infect Immun*, 2008. **76**(11): p. 5350-6.
14. World Health Organisation (WHO). Available from: <http://www.who.int/en/>.
15. O'Brien, K.L., et al., Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet*, 2009. **374**(9693): p. 893-902.
16. Coker, T.R., et al., Diagnosis, microbial epidemiology, and antibiotic treatment of acute otitis media in children: a systematic review. *JAMA*. **304**(19): p. 2161-9.
17. Okitsu, N., et al., Binax NOW(R) *Streptococcus pneumoniae* test of middle ear fluid for detecting causative pathogens in children with acute otitis media. *J Microbiol Methods*. **84**(2): p. 341-2.
18. Pelton, S.I., Acute otitis media in the era of effective pneumococcal conjugate vaccine: will new pathogens emerge? *Vaccine*, 2000. **19 Suppl 1**: p. S96-9.
19. Jacobs, M.R., et al., Prevalence of antimicrobial-resistant pathogens in middle ear fluid: multinational study of 917 children with acute otitis media. *Antimicrob Agents Chemother*, 1998. **42**(3): p. 589-95.

20. Luotonen, J., et al., *The bacteriology of acute otitis media in children with special reference to Streptococcus pneumoniae as studied by bacteriological and antigen detection methods. Scand J Infect Dis*, 1981. **13**(3): p. 177-83.
21. Karma, P., et al., *Pneumococcal bacteriology after pneumococcal otitis media with special reference to pneumococcal antigens. Int J Pediatr Otorhinolaryngol*, 1985. **10**(2): p. 181-90.
22. O'Neill, P., *Acute otitis media. BMJ*, 1999. **319**(7213): p. 833-5.
23. Mulholland, K., *Magnitude of the problem of childhood pneumonia. Lancet*, 1999. **354**(9178): p. 590-2.
24. van de Beek, D., et al., *Community-acquired bacterial meningitis in adults. N Engl J Med*, 2006. **354**(1): p. 44-53.
25. Ostergaard, C., H.B. Konradsen, and S. Samuelsson, *Clinical presentation and prognostic factors of Streptococcus pneumoniae meningitis according to the focus of infection. BMC Infect Dis*, 2005. **5**: p. 93.
26. Kastenbauer, S. and H.W. Pfister, *Pneumococcal meningitis in adults: spectrum of complications and prognostic factors in a series of 87 cases. Brain*, 2003. **126**(Pt 5): p. 1015-25.
27. Saez-Llorens, X. and G.H. McCracken, Jr., *Bacterial meningitis in children. Lancet*, 2003. **361**(9375): p. 2139-48.
28. Mitchell, T.J., *The pathogenesis of streptococcal infections: from tooth decay to meningitis. Nat Rev Microbiol*, 2003. **1**(3): p. 219-30.
29. Brown, E.J., et al., *Localization of complement component 3 on Streptococcus pneumoniae: anti-capsular antibody causes complement deposition on the pneumococcal capsule. Infect Immun*, 1983. **39**(1): p. 403-9.
30. Melin, M., et al., *The capsular serotype of Streptococcus pneumoniae is more important than the genetic background for resistance to complement. Infect Immun*. **78**(12): p. 5262-70.
31. Gilbert, R.J., *Cholesterol-dependent cytolysins. Adv Exp Med Biol*. **677**: p. 56-66.
32. Feldman, C., et al., *The effect of Streptococcus pneumoniae pneumolysin on human respiratory epithelium in vitro. Microb Pathog*, 1990. **9**(4): p. 275-84.
33. Winter, A.J., et al., *A role for pneumolysin but not neuraminidase in the hearing loss and cochlear damage induced by experimental pneumococcal meningitis in guinea pigs. Infect Immun*, 1997. **65**(11): p. 4411-8.
34. Rubins, J.B., et al., *Distinct roles for pneumolysin's cytotoxic and complement activities in the pathogenesis of pneumococcal pneumonia. Am J Respir Crit Care Med*, 1996. **153**(4 Pt 1): p. 1339-46.
35. Kadioglu, A., et al., *Upper and lower respiratory tract infection by Streptococcus pneumoniae is affected by pneumolysin deficiency and differences in capsule type. Infect Immun*, 2002. **70**(6): p. 2886-90.
36. Orihuela, C.J., et al., *Tissue-specific contributions of pneumococcal virulence factors to pathogenesis. J Infect Dis*, 2004. **190**(9): p. 1661-9.
37. Tuomanen, E., *Molecular and cellular biology of pneumococcal infection. Curr Opin Microbiol*, 1999. **2**(1): p. 35-9.
38. Howard, L.V. and H. Gooder, *Specificity of the autolysin of Streptococcus (Diplococcus) pneumoniae. J Bacteriol*, 1974. **117**(2): p. 796-804.
39. Cheng, Q., D. Finkel, and M.K. Hostetter, *Novel purification scheme and functions for a C3-binding protein from Streptococcus pneumoniae. Biochemistry*, 2000. **39**(18): p. 5450-7.
40. Dave, S., et al., *PspC, a pneumococcal surface protein, binds human factor H. Infect Immun*, 2001. **69**(5): p. 3435-7.
41. Dave, S., et al., *Interaction of human factor H with PspC of Streptococcus pneumoniae. Indian J Med Res*, 2004. **119** Suppl: p. 66-73.

42. Rosenow, C., et al., Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. *Mol Microbiol*, 1997. **25**(5): p. 819-29.
43. Briles, D.E., et al., The potential to use PspA and other pneumococcal proteins to elicit protection against pneumococcal infection. *Vaccine*, 2000. **18**(16): p. 1707-11.
44. Jedrzejewski, M.J., Pneumococcal virulence factors: structure and function. *Microbiol Mol Biol Rev*, 2001. **65**(2): p. 187-207.
45. Ren, B., et al., Effects of PspA and antibodies to PspA on activation and deposition of complement on the pneumococcal surface. *Infect Immun*, 2004. **72**(1): p. 114-22.
46. Shaper, M., et al., PspA protects *Streptococcus pneumoniae* from killing by apolactoferrin, and antibody to PspA enhances killing of pneumococci by apolactoferrin [corrected]. *Infect Immun*, 2004. **72**(9): p. 5031-40.
47. Novak, R., et al., Penicillin tolerance genes of *Streptococcus pneumoniae*: the ABC-type manganese permease complex Psa. *Mol Microbiol*, 1998. **29**(5): p. 1285-96.
48. Dintilhac, A., et al., Competence and virulence of *Streptococcus pneumoniae*: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. *Mol Microbiol*, 1997. **25**(4): p. 727-39.
49. Berry, A.M. and J.C. Paton, Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of *Streptococcus pneumoniae*. *Infect Immun*, 1996. **64**(12): p. 5255-62.
50. Romero-Steiner, S., et al., Inhibition of pneumococcal adherence to human nasopharyngeal epithelial cells by anti-PsaA antibodies. *Clin Diagn Lab Immunol*, 2003. **10**(2): p. 246-51.
51. Briles, D.E., et al., Intranasal immunization of mice with a mixture of the pneumococcal proteins PsaA and PspA is highly protective against nasopharyngeal carriage of *Streptococcus pneumoniae*. *Infect Immun*, 2000. **68**(2): p. 796-800.
52. Tong, H.H., et al., Effect of lacto-N-neotetraose, asialoganglioside-GM1 and neuraminidase on adherence of otitis media-associated serotypes of *Streptococcus pneumoniae* to chinchilla tracheal epithelium. *Microb Pathog*, 1999. **26**(2): p. 111-9.
53. Tong, H.H., et al., Evaluation of the virulence of a *Streptococcus pneumoniae* neuraminidase-deficient mutant in nasopharyngeal colonization and development of otitis media in the chinchilla model. *Infect Immun*, 2000. **68**(2): p. 921-4.
54. Hillerigmann, M., et al., Pneumococcal pili are composed of protofilaments exposing adhesive clusters of RrgA. *PLoS Pathog*, 2008. **4**(3): p. e1000026.
55. Hillerigmann, M., et al., Molecular architecture of *Streptococcus pneumoniae* TIGR4 pili. *EMBO J*, 2009. **28**(24): p. 3921-30.
56. Kang, H.J., et al., Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. *Science*, 2007. **318**(5856): p. 1625-8.
57. Nelson, A.L., et al., RrgA is a pilus-associated adhesin in *Streptococcus pneumoniae*. *Mol Microbiol*, 2007. **66**(2): p. 329-40.
58. Bagnoli, F., et al., A second pilus type in *Streptococcus pneumoniae* is prevalent in emerging serotypes and mediates adhesion to host cells. *J Bacteriol*, 2008. **190**(15): p. 5480-92.
59. Blue, C.E., et al., ZmpB, a novel virulence factor of *Streptococcus pneumoniae* that induces tumor necrosis factor alpha production in the respiratory tract. *Infect Immun*, 2003. **71**(9): p. 4925-35.
60. Rioux, S., et al., Transcriptional regulation, occurrence and putative role of the Pht family of *Streptococcus pneumoniae*. *Microbiology*, 2011. **157**(Pt 2): p. 336-48.
61. Ogunniyi, A.D., et al., Pneumococcal histidine triad proteins are regulated by the Zn²⁺-dependent repressor AdcR and inhibit complement deposition through the recruitment of complement factor H. *Faseb J*, 2009. **23**(3): p. 731-8.
62. Bentley, S.D., et al., Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet*, 2006. **2**(3): p. e31.

63. Weiser, J.N., *The pneumococcus: why a commensal misbehaves*. *J Mol Med*. **88**(2): p. 97-102.
64. Austrian, R., *The pneumococcus at the millennium: not down, not out*. *J Infect Dis*, 1999. **179 Suppl 2**: p. S338-41.
65. Obaro, S.K., *The new pneumococcal vaccine*. *Clin Microbiol Infect*, 2002. **8**(10): p. 623-33.
66. Wright, A.E., Parry Morgan, W., Colebrook, L., and Dodgson, R.W. , *Prophylactic inoculation against pneumococcal infection*. *Lancet*, 1914. **87**(95).
67. Chien, Y.W., K.P. Klugman, and D.M. Morens, *Efficacy of whole-cell killed bacterial vaccines in preventing pneumonia and death during the 1918 influenza pandemic*. *J Infect Dis*. **202**(11): p. 1639-48.
68. Watson, D.A., et al., *A brief history of the pneumococcus in biomedical research: a panoply of scientific discovery*. *Clin Infect Dis*, 1993. **17**(5): p. 913-24.
69. Fritzell, B. and S. Plotkin, *Efficacy and safety of a Haemophilus influenzae type b capsular polysaccharide-tetanus protein conjugate vaccine*. *J Pediatr*, 1992. **121**(3): p. 355-62.
70. Barra, A., et al., *Immunogenicity of Haemophilus influenzae type b conjugate vaccine in allogeneic bone marrow recipients*. *J Infect Dis*, 1992. **166**(5): p. 1021-8.
71. Butler, J.C., et al., *Serotype distribution of Streptococcus pneumoniae infections among preschool children in the United States, 1978-1994: implications for development of a conjugate vaccine*. *J Infect Dis*, 1995. **171**(4): p. 885-9.
72. Duggan, S.T., *Pneumococcal polysaccharide conjugate vaccine (13-valent, adsorbed) [Prevenar 13(R)]*. *Drugs*. **70**(15): p. 1973-86.
73. Anon, *Prevenar 13 authorized in Europe*. *Hum Vaccin*, 2010. **6**(3): p. 227-31.
74. *Licensure of a 13-valent pneumococcal conjugate vaccine (PCV13) and recommendations for use among children - Advisory Committee on Immunization Practices (ACIP)*, 2010. *MMWR Morb Mortal Wkly Rep*. **59**(9): p. 258-61.
75. Prymula, R. and L. Schuerman, *10-valent pneumococcal nontypeable Haemophilus influenzae PD conjugate vaccine: Synflorix*. *Expert Rev Vaccines*, 2009. **8**(11): p. 1479-500.
76. Croxtall, J.D. and G.M. Keating, *Pneumococcal polysaccharide protein D-conjugate vaccine (Synflorix; PHiD-CV)*. *Paediatr Drugs*, 2009. **11**(5): p. 349-57.
77. Ruckinger, S., et al., *Efficacy of 7-valent pneumococcal conjugate vaccination in Germany: An analysis using the indirect cohort method*. *Vaccine*. **28**(31): p. 5012-6.
78. Siber, G.R., *Pneumococcal disease: prospects for a new generation of vaccines*. *Science*, 1994. **265**(5177): p. 1385-7.
79. Stein, K.E., *Thymus-independent and thymus-dependent responses to polysaccharide antigens*. *J Infect Dis*, 1992. **165 Suppl 1**: p. S49-52.
80. Pilishvili, T., et al., *Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine*. *J Infect Dis*. **201**(1): p. 32-41.
81. McIntosh, E.D. and R.R. Reinert, *Global prevailing and emerging pediatric pneumococcal serotypes*. *Expert Rev Vaccines*. **10**(1): p. 109-29.
82. Nunes, S., et al., *Emergence of a serotype 1 Streptococcus pneumoniae lineage colonising healthy children in Portugal in the seven-valent conjugate vaccination era*. *Clin Microbiol Infect*, 2008. **14**(1): p. 82-4.
83. Lipsitch, M., *Vaccination against colonizing bacteria with multiple serotypes*. *Proc Natl Acad Sci U S A*, 1997. **94**(12): p. 6571-6.
84. Pelton, S.I., et al., *Emergence of 19A as virulent and multidrug resistant Pneumococcus in Massachusetts following universal immunization of infants with pneumococcal conjugate vaccine*. *Pediatr Infect Dis J*, 2007. **26**(6): p. 468-72.
85. Ansaldi, F., et al., *Increasing incidence of Streptococcus pneumoniae serotype 19A and emergence of two vaccine escape recombinant ST695 strains in Liguria, Italy*, 7

- years after implementation of the 7-valent conjugated vaccine. *Clin Vaccine Immunol.* **18**(2): p. 343-5.
86. van Gils, E.J., et al., Pneumococcal conjugate vaccination and nasopharyngeal acquisition of pneumococcal serotype 19A strains. *JAMA.* **304**(10): p. 1099-106.
 87. Moore, M.R., et al., Population snapshot of emergent *Streptococcus pneumoniae* serotype 19A in the United States, 2005. *J Infect Dis,* 2008. **197**(7): p. 1016-27.
 88. Brueggemann, A.B., et al., Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog,* 2007. **3**(11): p. e168.
 89. Hoberman, A., et al., Pneumococcal resistance and serotype 19A in Pittsburgh-area children with acute otitis media before and after introduction of 7-valent pneumococcal polysaccharide vaccine. *Clin Pediatr (Phila).* **50**(2): p. 114-20.
 90. Aaberge, I.S., et al., Virulence of *Streptococcus pneumoniae* in mice: a standardized method for preparation and frozen storage of the experimental bacterial inoculum. *Microb Pathog,* 1995. **18**(2): p. 141-52.
 91. Tettelin, H., et al., Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science,* 2001. **293**(5529): p. 498-506.
 92. Adamou, J.E., et al., Identification and characterization of a novel family of pneumococcal proteins that are protective against sepsis. *Infect Immun,* 2001. **69**(2): p. 949-58.
 93. Beghetto, E., et al., Discovery of novel *Streptococcus pneumoniae* antigens by screening a whole-genome lambda-display library. *FEMS Microbiol Lett,* 2006. **262**(1): p. 14-21.
 94. Tai, S.S., *Streptococcus pneumoniae* protein vaccine candidates: properties, activities and animal studies. *Crit Rev Microbiol,* 2006. **32**(3): p. 139-53.
 95. Hamel, J., et al., Prevention of pneumococcal disease in mice immunized with conserved surface-accessible proteins. *Infect Immun,* 2004. **72**(5): p. 2659-70.
 96. Godfroid, F., et al., Preclinical evaluation of the Pht proteins as potential cross-protective pneumococcal vaccine antigens. *Infect Immun,* 2011. **79**(1): p. 238-45.
 97. Riboldi-Tunncliffe, A., et al., Expression, purification and X-ray characterization of residues 18-230 from the pneumococcal histidine triad protein A (PhtA) from *Streptococcus pneumoniae*. *Acta Crystallogr D Biol Crystallogr,* 2004. **60**(Pt 5): p. 926-8.
 98. Riboldi-Tunncliffe, A., N.W. Isaacs, and T.J. Mitchell, 1.2 Angstroms crystal structure of the *S. pneumoniae* PhtA histidine triad domain a novel zinc binding fold. *FEBS Lett,* 2005. **579**(24): p. 5353-60.
 99. Jmol: An open-source Java viewer for chemical structures in 3D.; Available from: <http://www.jmol.org>.
 100. Melin, M., et al., Interaction of pneumococcal histidine triad proteins with human complement. *Infect Immun.* **78**(5): p. 2089-98.
 101. Jedrzejewski, M.J., Unveiling molecular mechanisms of bacterial surface proteins: *Streptococcus pneumoniae* as a model organism for structural studies. *Cell Mol Life Sci,* 2007. **64**(21): p. 2799-822.
 102. Rioux, S., et al., Transcriptional regulation, occurrence and putative role of the Pht Family of *Streptococcus pneumoniae*. *Microbiology,* 2011. **157**(2): p. 336-48.
 103. Braun, V. and H. Killmann, Bacterial solutions to the iron-supply problem. *Trends Biochem Sci,* 1999. **24**(3): p. 104-9.
 104. Chu, B.C., et al., Siderophore uptake in bacteria and the battle for iron with the host; a bird's eye view. *Biometals,* 2010. **23**(4): p. 601-11.
 105. Crain, M.J., et al., Pneumococcal surface protein A (PspA) is serologically highly variable and is expressed by all clinically important capsular serotypes of *Streptococcus pneumoniae*. *Infect Immun,* 1990. **58**(10): p. 3293-9.
 106. Darrieux, M., et al., Fusion proteins containing family 1 and family 2 PspA fragments elicit protection against *Streptococcus pneumoniae* that correlates with

- antibody-mediated enhancement of complement deposition. *Infect Immun*, 2007. **75**(12): p. 5930-8.
107. Brooks-Walter, A., D.E. Briles, and S.K. Hollingshead, The *pspC* gene of *Streptococcus pneumoniae* encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. *Infect Immun*, 1999. **67**(12): p. 6533-42.
 108. Denoel, P., et al., Combined protective effects of anti-PhtD and anti-pneumococcal polysaccharides. *Vaccine*, 2011. [epub ahead of print].
 109. Alexander, J.E., et al., Immunization of mice with pneumolysin toxoid confers a significant degree of protection against at least nine serotypes of *Streptococcus pneumoniae*. *Infect Immun*, 1994. **62**(12): p. 5683-8.
 110. Ferreira, D.M., et al., DNA vaccines based on genetically detoxified derivatives of pneumolysin fail to protect mice against challenge with *Streptococcus pneumoniae*. *FEMS Immunol Med Microbiol*, 2006. **46**(2): p. 291-7.
 111. Kirkham, L.A., et al., Construction and immunological characterization of a novel nontoxic protective pneumolysin mutant for use in future pneumococcal vaccines. *Infect Immun*, 2006. **74**(1): p. 586-93.
 112. Paton, J.C., et al., Purification and immunogenicity of genetically obtained pneumolysin toxoids and their conjugation to *Streptococcus pneumoniae* type 19F polysaccharide. *Infect Immun*, 1991. **59**(7): p. 2297-304.
 113. Douce, G., et al., Novel mucosal vaccines generated by genetic conjugation of heterologous proteins to pneumolysin (PLY) from *Streptococcus pneumoniae*. *Vaccine*, 2010. **28**(18): p. 3231-7.
 114. ExPASy Proteomics Server. <http://expasy.org/>.
 115. Killmann, H., R. Benz, and V. Braun, Properties of the FhuA channel in the *Escherichia coli* outer membrane after deletion of FhuA portions within and outside the predicted gating loop. *J Bacteriol*, 1996. **178**(23): p. 6913-20.
 116. Rogers, L.M., et al., Requirement for either a host- or pectin-induced pectate lyase for infection of *Pisum sativum* by *Nectria hematococca*. *Proc Natl Acad Sci U S A*, 2000. **97**(17): p. 9813-8.
 117. Wood, W.B., Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the restriction and modification of DNA. *J Mol Biol*, 1966. **16**(1): p. 118-33.
 118. Studier, F.W. and B.A. Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*, 1986. **189**(1): p. 113-30.
 119. Berrow, N.S., et al., A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res*, 2007. **35**(6): p. e45.
 120. CLC Genomics Workbench 4. <http://www.clcbio.com/index.php?id=1240>.
 121. Sambrook, J., Fritsch, E.F & Maniatis, T., *Molecular Cloning: A Laboratory Manual*. Cold Spring harbour Press, New York.
 122. Suzuki, N., et al., Discrimination of *Streptococcus pneumoniae* from viridans group streptococci by genomic subtractive hybridization. *J Clin Microbiol*, 2005. **43**(9): p. 4528-34.
 123. QIAquick PCR Purification Kit Handbook. <http://www.qiagen.com/products/dnacleanup/gelpcrsicleanupsystems/qiaquickpcrurificationkit.aspx#Tabs=t2>.
 124. QIAquick Gel Extraction Handbook. <http://www.qiagen.com/products/dnacleanup/gelpcrsicleanupsystems/qiaquickgelxtractionkit.aspx#Tabs=t2>.
 125. Zhu, B., et al., In-fusion assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *Biotechniques*, 2007. **43**(3): p. 354-9.
 126. In-Fusion Molar Ratio Calculator. <http://bioinfo.clontech.com/infusion/molarRatio.do>.

127. QIAprep Miniprep Handbook.
<http://www.qiagen.com/products/plasmid/qiaprepminiprepsystem/qiaprepminiprepkit.aspx#Tabs=t2>.
128. PD-10 Desalting Column Handbook.
http://www.gelifesciences.com/apatrix/upp00919.nsf/content/689F675DA7323088C1257628001CCFF3?OpenDocument&Path=Catalog&Hometitle=Catalog&entry=5&newrel&LinkParent=C1256FC4003AED40-0AFD63437B764AADC125701900490C98_RelatedLinksNew-C821BEC677D8448BC1256EAE002E3030&newrel&hidesearchbox=yes&moduleid=42278.
129. Laemmli, U.K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 1970. **227**(5259): p. 680-5.
130. Invitrogen X Cell II Blotting Module Handbook.
<http://products.invitrogen.com:80/ivgn/product/EI9051>.
131. Edman, P., A method for the determination of amino acid sequence in peptides. *Arch Biochem*, 1949. **22**(3): p. 475.
132. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, 1994. **50**(Pt 5): p. 760-3.
133. Adams, P.D., et al., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. **66**(Pt 2): p. 213-21.
134. Leslie, A.G.W., Recent changes to the MOSFLM package for processing film and image plate data. *Jnt CCP4/ESF-EAMCB Newsl. Protein Crystallogr.*, 1992. **26**: p. 27-33.
135. Pflugrath, J.W., The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr*, 1999. **55**(Pt 10): p. 1718-25.
136. Emsley, P. and K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
137. The PyMOL Molecular Graphics System, Version 1.3. <http://www.pymol.org/>.
138. Corpet, F., Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*, 1988. **16**(22): p. 10881-90.
139. Yang, Z.R., et al., RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, 2005. **21**(16): p. 3369-76.
140. Vector NTi 11.5 Website. <http://www.invitrogen.com/site/us/en/home/LINNEA-Online-Guides/LINNEA-Communities/Vector-NTI-Community/vector-nti-software.html>.
141. RCSB Protein Data Bank (PDB). Available from:
<http://www.pdb.org/pdb/home/home.do>.
142. Turner, N.J., Applications of transketolases in organic synthesis. *Curr Opin Biotechnol*, 2000. **11**(6): p. 527-31.
143. Stryer, L., Berg, J., Tymoczko, J., *Biochemistry*. (5th ed). 2002: W.H Freeman and Co.
144. Beltramo, E., et al., Effects of thiamine and benfotiamine on intracellular glucose metabolism and relevance in the prevention of diabetic complications. *Acta Diabetol*, 2008. **45**(3): p. 131-41.
145. Gothel, S.F. and M.A. Marahiel, Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts. *Cell Mol Life Sci*, 1999. **55**(3): p. 423-36.
146. Lu, K.P., et al., Prolyl cis-trans isomerization as a molecular timer. *Nat Chem Biol*, 2007. **3**(10): p. 619-29.
147. Schiene, C. and G. Fischer, Enzymes that catalyse the restructuring of proteins. *Curr Opin Struct Biol*, 2000. **10**(1): p. 40-5.
148. Horsham, M., et al., Expression, purification, crystallization and preliminary X-ray crystallographic data from TktA, a transketolase from the lactic acid bacterium

- Lactobacillus salivarius*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 2010. **66**(Pt 8): p. 899-901.
149. Central Microbial Resources (CMR) Database. <http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi>.
 150. European Synchrotron Radiation Facility (ESRF). Available from: <http://www.esrf.eu/>.
 151. Diamond Light Source. Available from: <http://www.diamond.ac.uk/>.
 152. Evans, P., Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*, 2006. **62**(Pt 1): p. 72-82.
 153. Taylor, G., The phase problem. *Acta Crystallogr D Biol Crystallogr*, 2003. **59**(Pt 11): p. 1881-90.
 154. Blow, D., *Outline Of Crystallography For Biologists*. 2002: Oxford University Press.
 155. Chen, V.B., et al., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, 2010. **66**(Pt 1): p. 12-21.
 156. Hermans, P.W., et al., The streptococcal lipoprotein rotamase A (SlrA) is a functional peptidyl-prolyl isomerase involved in pneumococcal colonization. *J Biol Chem*, 2006. **281**(2): p. 968-76.
 157. Overweg, K., et al., The putative proteinase maturation protein A of *Streptococcus pneumoniae* is a conserved surface protein with potential to elicit protective immune responses. *Infect Immun*, 2000. **68**(7): p. 4180-8.
 158. Sigma Oligonucleotide Design Tool. http://www.sigmaaldrich.com/configurator/servlet/DesignTool?prod_type=STANDARD.
 159. DNA Sequencing And Services. <http://www.dnaseq.co.uk/>.
 160. Smith, D.D.S., *A Study Of The Aron Enzyme Complex By Limited Proteolysis*. 1980, University Of Glasgow.
 161. Trypsin, Recombinant, Proteomics Grade Materials Sheet. <http://www.roche-applied-science.com/pack-insert/3708985a.pdf>.
 162. MASCOT Search Engine.; Available from: http://www.matrixscience.com/search_form_select.html.
 163. Cooper, A. ITC User Notes. Available from: <http://www.chem.gla.ac.uk/staff/alanc/itcnotes.pdf>.
 164. Desrosiers, D.C., et al., The general transition metal (Tro) and Zn²⁺ (Znu) transporters in *Treponema pallidum*: analysis of metal specificities and expression profiles. *Mol Microbiol*, 2007. **65**(1): p. 137-52.
 165. BLAST - Basic Local Alignment Search Tool. Available from: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.
 166. Thomson, R., et al., Characterizing proteolytic cleavage site activity using bio-basis function neural networks. *Bioinformatics*, 2003. **19**(14): p. 1741-7.
 167. EMBL-EBI European Bioinformatics Institute. Available from: <http://www.ebi.ac.uk/>.
 168. Page, R., et al., NMR screening and crystal quality of bacterially expressed prokaryotic and eukaryotic proteins in a structural genomics pipeline. *Proc Natl Acad Sci U S A*, 2005. **102**(6): p. 1901-5.
 169. Kelly, S.M., T.J. Jess, and N.C. Price, How to study proteins by circular dichroism. *Biochim Biophys Acta*, 2005. **1751**(2): p. 119-39.
 170. Whitmore, L. and B.A. Wallace, DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res*, 2004. **32**(Web Server issue): p. W668-73.
 171. Rhodes, G., *Crystallography Made Crystal Clear*. Third ed. 2006: Elsevier Academic Press.

172. Dong, A., et al., *In situ* proteolysis for protein crystallization and structure determination. *Nat Methods*, 2007. **4**(12): p. 1019-21.
173. Wernimont, A. and A. Edwards, *In situ* proteolysis to generate crystals for structure determination: an update. *PLoS One*, 2009. **4**(4): p. e5094.
174. Matthews, B.W., Solvent content of protein crystals. *J Mol Biol*, 1968. **33**(2): p. 491-7.
175. Warne, T., et al., Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature*, 2008. **454**(7203): p. 486-91.
176. Seventh Framework Program. Available from: http://cordis.europa.eu/fp7/home_en.html.
177. Hermoso Crystallography Group. Available from: <http://www.xtal.iqfr.csic.es/grupo/xjuan/>.
178. Molina, R., et al., Crystallization and preliminary X-ray diffraction studies of choline-binding protein F from *Streptococcus pneumoniae*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 2007. **63**(Pt 9): p. 742-5.
179. Molina, R., et al., Crystal structure of CbpF, a bifunctional choline-binding protein and autolysis regulator from *Streptococcus pneumoniae*. *EMBO Rep*, 2009. **10**(3): p. 246-51.
180. Perez-Dorado, I., et al., Crystallization of the pneumococcal autolysin LytC: in-house phasing using novel lanthanide complexes. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 2010. **66**(Pt 4): p. 448-51.
181. Perez-Dorado, I., et al., Insights into pneumococcal fratricide from the crystal structures of the modular killing factor LytC. *Nat Struct Mol Biol*, 2010. **17**(5): p. 576-81.
182. Ogunniyi, A.D., et al., Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of *Streptococcus pneumoniae*. *Infect Immun*, 2007. **75**(1): p. 350-7.
183. Rahfeld, J.U., et al., Confirmation of the existence of a third family among peptidyl-prolyl *cis/trans* isomerases. Amino acid sequence and recombinant production of parvulin. *FEBS Lett*, 1994. **352**(2): p. 180-4.
184. Riboldi-Tunicliffe, A. and R. Hilgenfeld, Cryocrystallography with oil - an old idea revived. *J. Appl. Cryst*, 1999(**32**): p. 1003-1005.