

Structure and Function of Bacterial Proteins Secreted by the Type Three Secretion and Twin Arginine Translocation Pathways

James E. D. Lillington

**Inorganic Chemistry Laboratory
Sir William Dunn School of Pathology
New College**



**A thesis submitted in partial fulfilment of the requirements for the degree of Doctor
of Philosophy at the University of Oxford**

Trinity Term 2011

Structure and Function of Bacterial Proteins Secreted by the Type Three Secretion and Twin Arginine Translocation Pathways

Abstract

The Type Three Secretion Systems (T3SSs) of Gram-negative bacteria, including *Shigella*, *Salmonella*, and Enteropathogenic/Enterohaemorrhagic *Escherichia coli* (EPEC/EHEC), pass virulence factors directly into the host to mediate invasion. Prior to secretion down the narrow T3SS channel, effector proteins associate with chaperone proteins. The binding enables the T3SS to keep effectors soluble and partially unfolded for secretion. In the first part of this thesis, the association of one promiscuous chaperone, Spa15 of *Shigella flexneri*, with three of its cognate effectors has been studied. In addition to the role this plays in secretion, the binding of one particular substrate leads to Spa15 being involved in the regulation of the T3SS. The oligomerisation and impact of substrate binding upon Spa15 has been determined by crystallography and EPR. Once secreted, T3SS effectors subvert the host cytoskeleton for the benefit of the bacteria. Soluble homologues of Spa15 effectors from EHEC and *Salmonella* have been purified, and their interactions with host GTPases which lead to stress fibre phenotypes observed.

The Twin Arginine Translocation (Tat) pathway provides a contrasting view of bacterial secretion. Instead of preventing folding in the cytoplasm, it is a criterion of transport that the protein be folded. One of the reasons for internal folding is the necessity to insert cofactors which could not be incorporated externally. In the second part of this thesis, a protein which exemplifies this necessity is studied. This is PhoD, the model protein for Tat export from *Bacillus subtilis*. PhoD is an alkaline phosphodiesterase expressed to scavenge phosphate in times of phosphate deficiency. The structure of PhoD has been solved, and the protein is shown to be able to cleave a component of its own cell wall. It uses an unusual catalytic site more reminiscent of the eukaryotic purple acid phosphatases than of other currently known alkaline phosphatases. Furthermore this site appears to require metal binding before export from the bacterial cytoplasm.

Acknowledgements

I thank Susan Lea and Christiane Timmel for their guidance and supervision throughout my DPhil. Particular thanks go to Susan who allowed me to work in her lab, fostering my interests in protein crystallography and biochemistry.

Thanks to the Postdocs who helped me with various techniques: I thank Steven Johnson for both initial 'PCR is a reaction where...' type talks and continued biochemical/crystallographic insight throughout my DPhil. Janet Lovett sat at the EPR spectrometer for many hours with me and I am very grateful for her help and guidance. Pietro Roversi has answered many crystallographic questions, and also spent a large amount of time collecting and helping with the interpretation of data. Patrizia Abrusci, Janet Deane and Devon Sheppard have been fonts of biochemical and laboratory knowledge (and Patrizia an excellent thesis proof-reader!).

I thank my fellow DPhil students (past and present) for work and non-work conversation, and for making the lab always a fun place to be. Thanks to Bev Prosser, Kirstin Leath, Joe Caesar, Mel McDowell, Sarah Rollauer, Rachel Everett and Matthew Cottee.

My thanks go to our collaborators for their roles in the research. The EspM2 study was carried out with Ana Arbeloa in the lab of Gad Frankel, and James Garnett in the lab of Steve Matthews, both at Imperial College. Ana provided constructs and performed biochemical assays, and James took NMR data. The PhoD and PhoX work was carried out with Fernanda Rodriguez and Shee Chien Yong in the lab of Ben Berks, Department of Biochemistry, Oxford. Fernanda and Chien supplied me with purified protein and carried out biochemical assays to aid my data.

My thanks to the EPSRC for financial support, and to Susan Lea for 'topping me up' at the end.

Finally, I thank Nienke, my family and my friends for their love and support.

Abbreviations

β-MC	β-Mercaptoethanol
CBD	Chaperone Binding Domain
CCP4	Collaborative Computational Project Number 4
CW	Continuous Wave
DEER	Double Electron-Electron Resonance
EDTA	Ethylenediaminetetraacetic acid
EHEC	Enterohaemorrhagic <i>Escherichia coli</i>
EPEC	Enteropathogenic <i>Escherichia coli</i>
EPR	Electron Paramagnetic Resonance
FFAS	Fold and Function Assignment Server
FFT	Fast Fourier Transform
FOM	Figure of Merit
GDP	Guanosine-5'-diphosphate
GEF	Guanine nucleotide Exchange Factor
GTP	Guanosine-5'-triphosphate
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HICUP	Hetero-Compound Information Centre – Uppsala
His	Hexa-histidine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JCSG	Joint Center for Structural Genomics
LB	Luria-Bertani
LTA	Lipo Teichoic Acid
MAD	Multi-Wavelength Anomalous Dispersion
MALLS	Multangle Laser Light Scattering
MMM	Multiscale Modelling of Macromolecular Systems
MOPS	3-(N-morpholino)propanesulphonic acid
mT	Millitesla

MTSL	(S-(2,2,5,5-Tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonylthioate)
NCBI	National Center for Biotechnology Information
NCS	Non-Crystallographic Symmetry
PAGE	Polyacrylamide Gel Electrophoresis
PAP	Purple Acid Phosphatase
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PhoD	Alkaline Phosphatase D
PhoX	Alkaline Phosphatase X
PIXE	Proton Induced X-ray Emission
PSI-BLAST	Position-Specific Iterated - Basic Local Alignment Search Tool
R.m.s.d.	Root mean square deviation
SCV	<i>Salmonella</i> Containing Vacuole
SDS	Sodium Dodecyl Sulphate
SEC	Size Exclusion Chromatography
Sec	Secretory
SUMO	Small Ubiquitin-like Modifier
T3SS	Type Three Secretion System
Tat	Twin arginine translocation
TEMED	Tetramethylethylenediamine
TMAO	Trimethylamine <i>N</i> -oxide
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol

Single letter and triplet codes for amino acid residues are used. Other abbreviations are defined in the text where first encountered.

Publications resulting from the work of this thesis

Shigella flexneri Spa15 crystal structure verified in solution by double electron-electron resonance.

J.E.D. Lillington, J.E. Lovett, S. Johnson, P. Roversi, C.R. Timmel, S.M. Lea. 2011. Journal of Molecular Biology 405(2), 407-425.

EspM2 is a RhoA guanine nucleotide exchange factor.

A. Arbeloa, J. Garnett and J. Lillington, R.R. Bulgin, C. N. Berger, S.M. Lea, S. Matthews, G. Frankel. 2010. Cellular Microbiology 12, 654-664.

In preparation:

Structure of PhoD: A bacterial alkaline phosphatase with the structure of a purple acid phosphatase.

F.M. Rodriguez and J.E.D. Lillington, S. Johnson, P. Roversi, J.E Lovett, S.M. Lea, B.C. Berks.

Structural and biochemical characterization of the alkaline phosphatase PhoX.

S.C. Yong and P. Roversi, J.E.D. Lillington, O.B. Zeldin, E.F. Garman, S.M. Lea, B.C. Berks.

Table of Contents

1	Introduction: Bacterial Protein Secretion.....	1
1.1	A large secretion repertoire requires many secretion systems.	1
1.2	Type III secretion.	3
1.2.1	<i>The role of the T3SS in Shigella, Salmonella and EPEC/EHEC virulence.</i>	3
1.2.1.1	<i>Virulence mechanisms.</i>	3
1.2.2	T3SS architecture.	7
1.2.3	Regulation and timing.	9
1.2.3.1	<i>Building the apparatus.</i>	9
1.2.3.2	<i>First wave secretion.</i>	10
1.2.3.3	<i>Second wave secretion.....</i>	11
1.2.4	Chaperone usage.....	11
1.2.4.1	<i>Classes of chaperones.</i>	12
1.2.4.2	<i>Structure and binding of Class I chaperones.</i>	12
1.2.4.3	<i>Structure and binding of Class II chaperones.</i>	14
1.3	Tertiary structures of secreted proteins.	15
1.3.1	T3SS: The requirement for unfolded effectors.	15
1.3.2	Tat secretion: The requirement for folded substrates.	16
1.4	Secretion by the Tat translocon.....	17
1.4.1	The apparatus for Tat transport.....	17
1.4.2	The mechanism of Tat transport.....	19
1.4.3	The Gram-positive Tat apparatus.....	20
1.4.4	Proofreading and quality control.	22
1.5	The elucidation of secretion biochemistry with physical methods.	23
1.6	Project aims.	25
2	Materials and Methods.	27
2.1	Molecular biology.	27
2.1.1	Constructs and primers.	27
2.1.2	Polymerase chain reaction.	28
2.1.3	Cloning protocol.....	28
2.2	Protein expression and purification.	31
2.3	Protein solution characterisation.	33
2.4	Protein/DNA construct optimisation.....	35
2.5	Crystallisation.....	36
2.6	X-ray crystallography.	37
2.6.1	Diffraction in protein crystallography.	37
2.6.1.1	<i>The phase problem.</i>	39
2.6.2	Data collection.	40
2.6.2.1	<i>Indexing and strategising a data collection.</i>	40
2.6.3	Structure solution.....	41
2.6.3.1	<i>Data processing.....</i>	41
2.6.3.1.1	<i>Integration of data: iMosflm and XDS.</i>	41

2.6.3.1.2	<i>Symmetry checking: Pointless.</i>	42
2.6.3.1.3	<i>Scaling and merging files: Scala.</i>	42
2.6.3.1.4	<i>Calculation of F from intensities: Truncate.</i>	43
2.6.3.1.5	<i>Automatic data processing: Process and Xia2.</i>	44
2.6.3.2	<i>Determination of the asymmetric unit: Matthews coefficient and Patterson function.</i>	44
2.6.3.3	<i>Experimental phasing and model building.</i>	45
2.6.3.3.1	<i>X-ray absorption.</i>	46
2.6.3.3.2	<i>Heavy atom finding: autoSHARP.</i>	46
2.6.3.3.3	<i>Heavy atom refinement and phasing.</i>	47
2.6.3.3.4	<i>Density modification: Solomon.</i>	48
2.6.3.3.5	<i>Automatic model building: Buccaneer.</i>	48
2.6.3.4	<i>Molecular replacement.</i>	49
2.6.3.4.1	<i>MolRep.</i>	49
2.6.3.4.2	<i>Phaser.</i>	50
2.6.3.5	<i>Refinement and validation.</i>	50
2.7	Electron paramagnetic resonance.	52
2.7.1	Continuous wave-EPR.	52
2.7.1.1	<i>Theory.</i>	52
2.7.1.2	<i>Sample preparation.</i>	53
2.7.1.3	<i>Room temperature CW-EPR.</i>	54
2.7.1.4	<i>Base temperature CW-EPR.</i>	55
2.7.2	Pulsed EPR and double electron-electron resonance.	55
2.7.2.1	<i>DEER measurements.</i>	59
2.7.2.2	<i>Multiscale modelling of macromolecular systems.</i>	60
2.8	Surface plasmon resonance.	60
3	Structure of Spin labelled Spa15; The Use of EPR to Complement X-ray Crystallography.	62
3.1	Spa15 – a promiscuous chaperone.	64
3.1.1	Expression and purification of Spa15 and Spa15:lpdB1.	64
3.1.2	Spin labelling of Spa15 and Spa15:lpdB1.	66
3.2	Continuous Wave EPR studies.	67
3.2.1	Quantitative calibration.	69
3.3	Double Electron-Electron Resonance studies.	70
3.3.1	Observer and pump spins.	71
3.3.2	DEER measurements.	72
3.3.3	Satellite peaks.	73
3.3.4	Rationalisation of DEER distances with spin label conformation.	74
3.3.5	Multiscale modelling of macromolecular systems.	75
3.4	Spin-labelled Spa15 X-ray crystal structure.	76
3.4.1	Agreement of the structure with DEER.	82
3.4.2	Disagreement of the structure with MMM.	82
3.5	The effect of lpdB1 binding.	84
3.6	Conclusions.	86
4	T3SS - Spa15 WxxxE Effectors and their Homologues.	87
4.1	WxxxE proteins and their function.	87
4.2	The Spa15 chaperoned effector lpdB1.	91

4.2.1	Expression of IpgB1.	91
4.2.2	Characterisation of the Spa15:IpgB1 complex.	92
4.2.3	Removal of the IpgB1 N-terminus.	93
4.2.4	Spa15:IpgB1 crystallographic attempts.	94
4.3	WxxxE proteins which target RhoA.	97
4.3.1	The role of EspM2 and SifA.	97
4.3.2	Expression of EspM2 and crystallographic attempts.	98
4.3.3	RhoA binds EspM2.	100
4.3.3.1	<i>Biochemical support for an EspM2-RhoA interaction.</i>	<i>102</i>
4.3.4	SifA binds RhoA.	105
4.3.5	Insight from the structure of IpgB2.	106
4.4	Conclusions.	110

5 The Spa15 Chaperoned Effector OspD1 and its Role in Transcriptional Regulation. 112

5.1	The regulatory side to Spa15 effectors.	112
5.2	The Spa15:OspD1 complex.	115
5.2.1	Spa15:OspD1 purification.	115
5.2.2	The nature of the complex.	116
5.3	The Spa15:OspD1:MxiE complex.	118
5.3.1	The frameshifting origins of MxiE.	119
5.3.2	Spa15:OspD1:MxiE purification.	119
5.3.3	Increasing the yield of the complex.	121
5.3.4	Crystal trials.	123
5.3.5	Separation of MxiE.	125
5.4	The co-anti-activators IpaB and IpaC.	127
5.5	Binding MxiE and IpgC.	130
5.6	Conclusions.	133

6 *Bacillus subtilis* PhoD: An Alkaline Phosphatase/Phosphodiesterase Produced under Phosphate Starvation by the Tat System. 134

6.1	Stress mechanisms of <i>B. subtilis</i>.	134
6.1.1	The Pho regulon: phosphate deficient response.	135
6.2	PhoD construct.	138
6.2.1	Functional behaviour.	140
6.3	EPR spectrum.	141
6.3.1	Origin of the PhoD EPR spectrum.	143
6.4	Crystallisation of PhoD.	145
6.4.1	Data collection and processing.	146
6.4.2	Phasing.	152
6.4.3	Refinement.	154
6.5	Structure of PhoD.	157
6.6	The catalytic site of PhoD.	160
6.6.1	An overall view of the catalytic site.	160
6.6.2	Metal I: Fe(III).	164
6.6.2.1	<i>Comparison with the PAP Fe(III) site.</i>	<i>164</i>
6.6.3	Metals II and III.	165

6.6.3.1	<i>Calcium occupancy.</i>	165
6.6.3.2	<i>Geometry.</i>	167
6.7	Fe(III)-cysteine coordination.	169
6.7.1	The coordination rarity.	169
6.7.2	PhoX.	170
6.7.3	Fe(III)-Cys function.	173
6.8	A C-terminal α-helix occludes the active site.	174
6.9	PhoD mechanism.	181
6.9.1	Alkaline phosphatase mechanism.	181
6.9.2	Purple acid phosphatase mechanism.	183
6.10	PhoD target.	185
6.11	Conclusions.	187
7	Overall Conclusions.	190
	Bibliography.	193
	Appendices.	224
A.	CW-EPR experimental setup.	224
i.	Spectrometer tuning.	224
ii.	Spectrometer cooling.	224
iii.	Easyspin.	225
B.	DEER experimental setup.	225
C.	SPR experimental setup.	226

Index of Figures and Tables

Figure 1-1 Overview of protein secretion systems.....	2
Figure 1-2 Virulence of <i>Shigella</i> , <i>Salmonella</i> and EPEC.	6
Figure 1-3 Structure and composition of the <i>S. flexneri</i> T3SS.....	8
Figure 1-4 The sequential stages of T3SS apparatus building.....	10
Figure 1-5 Conformation of Class I and Class II chaperones in complex with their substrates.	14
Figure 1-6 Model of Tat export.	19
Figure 1-7 The Tat translocon of <i>B. subtilis</i>	21
Figure 1-8 Comparison of the chromosomal organisation between <i>E. coli</i> and <i>B. subtilis</i>	22
Table 2.1 Constructs used.....	27
Table 2.2 PCR protocols.	28
Table 2.3 Agarose gel and TAE running buffer composition.....	29
Table 2.4 Cloning strategy.	30
Table 2.5 Protein purification columns used.	31
Table 2.6 Induction, expression and purification conditions of the proteins produced.	32
Table 2.7 SDS-PAGE solutions.	33
Table 2.8 Construct optimisation methods.	35
Table 2.9 Molecular dimensions crystal trial broad screens.....	36
Figure 2-1 The Ewald Sphere.	38
Table 2.10 Data quality statistics.	43
Figure 2-2 Mechanism of cysteine labelling by MTSL.....	53
Figure 2-3 The Hahn Echo sequence displayed in the rotating frame.	56
Figure 2-4 DEER pulse sequence.	59
Figure 3-1 Spin labelled Spa15.	63
Figure 3-2 Spa15 and Spa15:IpgB1 purification profiles.....	65
Figure 3-3 Mass spectrometry showing successful labelling of Spa15 cysteines	67
Table 3.1 EPR spectral parameters used for Spa15 and Tempol nitroxide radical detection.....	68
Figure 3-4 Characteristic EPR spectrum for a freely rotating nitroxide radical.	69
Figure 3-5 EPR verification of 100% Spa15 labelling.	70
Figure 3-6 Separation of chemically identical nitroxide radicals into different species for DEER detection.	71
Figure 3-7 Spa15 experimental DEER data.	72
Figure 3-8 Probing the origin of the DEER satellite peaks.	73
Figure 3-9 Spa15 crystal optimisation and crystal morphology.....	76
Table 3.2 Data collection strategy and statistics.....	77
Figure 3-10 Spa15 self rotation function.	78
Table 3.3 Spa15 refinement statistics.	79
Figure 3-11 Primary sequence and secondary structure of PDB 2XGA.	79
Figure 3-12 MTSL spin label observed in the Spa15 crystal structure.....	81
Figure 3-13 SEC-MALLS of Spa15:IpgB1 complex.....	84
Figure 3-14 The impact of IpgB1 binding on Spa15 DEER.....	85
Figure 4-1 Multiple sequence alignment of the WxxxE domain of the effectors involved in this study....	88
Figure 4-2 Solubility of IpgB1 brought about through coexpression with Spa15.	92
Figure 4-3 Characterisation of Spa15:IpgB1.	94
Figure 4-4 Protein manipulations made to encourage crystal formation.	96

Figure 4-5 Size exclusion chromatography and SDS-PAGE from each of the proteins EspM2 ²⁹⁻¹⁹⁶ , SifA, RhoA and Rac1.	99
Table 4.1 Attempted EspM2 crystallisation conditions.	100
Figure 4-6 EspM2-RhoA binding.	101
Figure 4-7 EspM2 enhances nucleotide exchange in RhoA.	103
Figure 4-8 Identification of EspM2 residues key for structure and function.	104
Figure 4-9 SPR binding of SifA to RhoA.	106
Figure 4-10 A comparison of RhoA binding between IpgB2 and the eukaryotic PDZ-RhoGEF.	108
Figure 4-11 The insoluble expression of IpgB2, both with and without Spa15 coexpression.	110
Figure 5-1 Sequence of events leading to transcriptional activation by MxiE.	114
Figure 5-2 OspD1 purification profiles.	115
Figure 5-3 Identification of Spa15:OspD1 aggregation.	116
Figure 5-4 Partial proteolysis of Spa15:OspD1 complex.	117
Figure 5-5 Spa15 is required to bring about the soluble binding of recombinant OspD1 by MxiE.	120
Figure 5-6 Purification of the SOM complex.	121
Figure 5-7 Oligomeric MxiE.	123
Figure 5-8 Growth and morphology of Spa15:OspD1:MxiE crystals.	124
Figure 5-9 Separation of free MxiE from remaining MxiE:OspD1.	126
Figure 5-10 Purification of IpaB:IpgC and IpaC:IpgC complexes.	128
Figure 5-11 IpaC:IpgC crystals.	129
Figure 5-12 IpgC:SOM will not coexpress, and IpgC cannot sequester MxiE from SOM.	131
Figure 5-13 The interaction of MxiE with IpgC.	132
Figure 6-1 Regulation of the Pho regulon.	137
Figure 6-2 Sequence of PhoD.	139
Figure 6-3 Purity and nature of PhoD in solution.	140
Figure 6-4 Alkaline phosphatase behaviour of PhoD.	141
Table 6.1 EPR spectral parameters used for PhoD and PhoX data collections.	142
Figure 6-5 EPR spectra of PhoD highlighting the g = 4.3 peak.	143
Figure 6-6 Origin of PhoD EPR peaks.	145
Figure 6-7 PhoD optimisation screen.	146
Figure 6-8 Morphology of PhoD crystals.	146
Figure 6-9 Iron K-edge scan.	147
Figure 6-10 Observation of the systematic absences which led to space group determination.	148
Table 6.2 Pointless result log.	148
Figure 6-11 Scala graphs.	149
Table 6.3 Data collection and phasing statistics.	150
Table 6.4 Top five translational Patterson peaks as calculated by FFT for Patterson.	151
Figure 6-12 Harker sections showing the translational NCS at (0.50, 0.50, 0.08).	152
Table 6.5 Highest FFAS scores for the PhoD sequence.	153
Figure 6-13 Schematic of phase solution.	154
Figure 6-14 The improvement of R/R _{free} with iterative refinement and crystal packing diagram.	155
Table 6.6 Refinement statistics.	156
Figure 6-15 PEG ordering.	156
Figure 6-16 Global structure of PhoD.	158
Figure 6-17 DALI structural homology with PhoD.	160
Figure 6-18 The PhoD active site.	161
Figure 6-19 Active site conservation.	163
Figure 6-20 The Fe(III) site.	165

Figure 6-21 Variation in anomalous signal at the active site with increasing wavelength.....	167
Figure 6-22 The geometry of Ca_2^{+1} and Ca_2^{+2} sites.	169
Figure 6-23 Non-haem proteins with solitary Fe(III)-cysteine coordination in the PDB.	170
Figure 6-24 EPR spectra of PhoX.	171
Figure 6-25 UV-visible wavelength absorption spectra of PhoX upon addition of sodium dithionite and EDTA	172
Figure 6-26 Non-conservation of the PhoD C-terminal α -helix.	175
Figure 6-27 Electrostatic surface model of PhoD highlighting the region of the C-terminal α -helix.	176
Figure 6-28 Salt bridges to the main chain and interactions to the phosphate formed by α -helical residues.....	177
Figure 6-29 Flexibility of the C-terminal α -helix.....	178
Figure 6-30 Clashing of pNPP with C-terminal α -helix residues in possible orientations of pNPP at the catalytic site.	179
Figure 6-31 EPR spectra of 100 μM Tempol and spin labelled PhoD solutions at room temperature. ...	182
Figure 6-32 Stepwise mechanisms for <i>E. coli</i> alkaline phosphatase, kidney bean phosphatase and PhoD. There is no evidence for a one step vs. two step mechanism for the PhoD phosphodiester hydrolysis.	185
Figure 6-33 <i>B. subtilis</i> teichoic acid structure.	186
Figure 6-34 Non-specificity of PhoD.	187
Figure 6-35 Cartoon of the putative function of PhoD.....	189
Table A.i An m.file for the plotting of data through Easyspin.....	225
Table C.i SPR protocol.....	226

1 Introduction: Bacterial Protein Secretion.

1.1 A large secretion repertoire requires many secretion systems.

Following synthesis, all proteins are targeted to the location of their function. In bacteria, a significant percentage - typically around 30% - of the expressed proteins have roles external to the cytoplasm, in the membrane, periplasm, outer envelope or external milieu (Wooldridge, 2009). The roles of these translocated proteins are wide-ranging, covering virulent and non-virulent niche adaptation purposes, nutrient acquisition, and cell-cell communication (Snyder and Champness, 2007). To meet the needs of the bacteria, translocated proteins must all pass (at least) a hydrophobic phospholipid bilayer without compromising the integrity of the membrane, in the process of secretion (Gennity and Inouye, 1991; Papanikou *et al.*, 2007)

A protein's structure is crucial to its function. In order for its function to be fulfilled, the protein must be delivered correctly folded (or with the ability to be correctly folded), and must be delivered to the correct location. The task of secretion in meeting these conditions for the multitude of secreted proteins is achieved by the array of secretion apparatuses which bacteria provide. A number of these are introduced in Figure 1-1, with each conferring unique translocation abilities to the bacteria (Tseng *et al.*, 2009). For example, correct protein folding may require a cytoplasmically inserted cofactor, requiring a secretion system such as the Twin arginine translocation (Tat) pathway which is specialised in secreting such substrates (Palmer *et al.*, 2005). Alternatively, the final destination of a virulence protein may be the host cytoplasm, requiring a secretion system which passes the bacterial protein directly into the host. The Type III, Type IV and Type VI secretion systems are optimised for this role.

Bacterial architectures require differing logistics for transport; the two membranes of Gram-negative bacteria provide a double obstacle for secreted proteins compared to the cytoplasmic membrane of

Gram-positive bacteria. This has resulted in a number of secretion systems which are found in Gram-negative bacteria only (Types I – Type VI in Figure 1-1), as well as the ubiquitous Secretory (Sec) and Tat systems which are found in both Gram-positive and Gram-negative bacteria (Wooldridge, 2009).

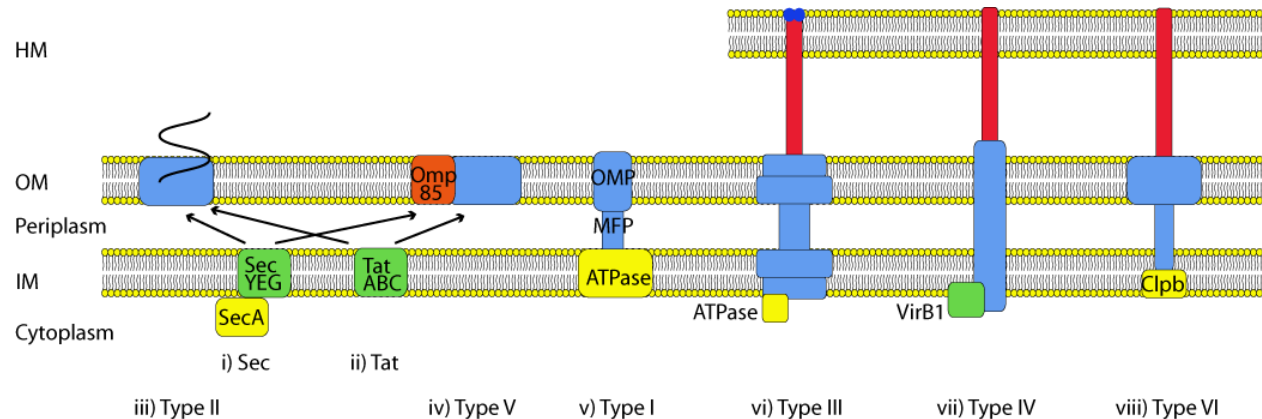


Figure 1-1 Overview of protein secretion systems. The two membranes of Gram-negative bacteria are shown, although the Sec and Tat pathways are found in Gram-positive bacteria as well. In Gram-negative bacteria, translocation mechanisms may be double step (Sec, Tat followed by Type II, V) or single step (Type I, III, IV, VI). IM (inner membrane), OM (outer membrane), HM (host membrane). **i)** The Sec pathway transports unfolded proteins through the inner membrane, traversing the protein conducting channel SecYEG using the motor component SecA. This method is homologous to translocases in the chloroplast thylakoid membrane (Driessen and Nouwen, 2008). **ii)** The Tat pathway provides the same function for folded proteins and is described in greater detail in the text. **iii)** Type II secretion allows the passing through the outer membrane of large multimeric proteins (including virulence factors, toxins and proteases) which are folded in the periplasm (Wooldridge, 2009). It is an assembly of twelve to sixteen general secretion pathway proteins which form a pseudopilin piston-like structure (Filloux, 2004). **iv)** The Type V secretion pathway are the so-called autotransporters, secreting substrates capable of mediating their own transport from the periplasm with the aid of a helper protein, Omp85 (Bernstein, 2007). Substrates contain a β -barrel forming translocation domain (cleaved upon transport), and effector domain (Jose *et al.*, 1995). Occasionally these domains are located on separate polypeptides (Jacob-Dubuisson *et al.*, 2001). **v)** Type I secretion uses an ABC transporter, a membrane fusion protein, and an outer membrane protein in a continuous channel to transport unfolded substrates from cytoplasm to exterior of bacteria (Wooldridge, 2009). The translocon secretes mostly polypeptides involved in virulence, of which the best characterised are RTX (repeats in toxins) toxins and lipases (Delepelaire, 2004; Linhartova *et al.*, 2010). **vi)** The Type III secretion pathway passes virulence factors directly into the host cytoplasm. This pathway is described in the text below. **vii)** Type IV secretion systems provide the same function as Type III, but secrete DNA in addition to proteins. This fulfils virulence functions or contributes to horizontal gene transfer in bacteria. Components are divided into the cytoplasmic energising body, periplasmic channel, and surface exposed proteins (Wooldridge, 2009). **viii)** The Type VI secretion system is thought to transport substrates directly from bacterial cytoplasm via a tubular structure (Filloux, 2009), with two groups of substrates (Hcp and VgrG) currently known (Filloux *et al.*, 2008). (Figure reproduced (Tseng *et al.*, 2009)).

The proteins explored in this study are secreted from the Type III Secretion System (T3SS) and Tat pathway. These are systems with substantial differences: the T3SS must pass proteins reduced to their secondary structure through the narrow channel of the T3SS. In contrast however, a stringent criterion for successful passing via the Tat system is that the substrates are folded prior to entering the

translocon. An introduction to each of these secretory mechanisms and the environment in which they exist is presented below.

1.2 Type III secretion.

The T3SS is a key virulence factor in numerous Gram-negative bacterial pathogens of both animals and plants (Block et al., 2008; Deane et al., 2010; Wooldridge, 2009); it allows the direct passage of virulence effectors from bacteria to host cytosol. These effectors may target the cytoskeleton, innate defences, autophagy, phosphoinositide mediated signalling or other novel targets (Mattoo *et al.*, 2007), allowing the successful colonisation of the host cell by the bacteria.

1.2.1 The role of the T3SS in *Shigella*, *Salmonella* and EPEC/EHEC virulence.

Shigella, *Salmonella* and Enteropathogenic/Enterohaemorrhagic *Escherichia coli* EPEC/EHEC are closely related pathogenic bacteria, which are transmitted via the faecal-oral route. They cause gastroenteropathy, mediated by the use of their T3SSs upon the intestinal epithelium of the host cell (Elliott, 2007; Fukushima et al., 2002). Despite use of the common T3SS system, substantial differences exist between their virulence mechanisms.

1.2.1.1 Virulence mechanisms.

Shigella is unable to target the luminal side of the epithelial cells of the gastrointestinal tract (Ashida *et al.*, 2009). It therefore traverses the Microfold cells (M-cells), inducing apoptosis of macrophages which it encounters, allowing access to the basolateral side of the epithelium (Schroeder and Hilbi, 2008).

The first contact between *Shigella* and the epithelial cells is from the T3SS tip ‘translocator’ proteins to lipid raft membrane domains (Hayward *et al.*, 2005), mediated by receptors such as the hyaluronan receptor CD44 and $\alpha_5\beta_1$ integrin (Skoudy *et al.*, 2000; Watarai *et al.*, 1996). Effectors secreted by the T3SS subvert the host cytoskeleton; they activate small Rho GTPases and tyrosine kinases resulting in membrane ruffling and engulfment of the bacterium (Boquet and Lemichez, 2003). *Shigella* then lyses

the phagosomal membrane which confines it and escapes into the cytoplasm. *Shigella* travels through the cells by propulsion due to directed actin polymerisation by the protein IscA (Bernardini *et al.*, 1989). When encountering the cell boundary, the T3SS mediates breakdown of the double membrane allowing the passing of *Shigella* into the adjacent cell (Figure 1-2a).

Host detection of *Shigella* leads to an upregulation of the chemokine IL-8, recruiting polymorphonuclear neutrophil leukocytes (PMN) during the immune response (Philpott *et al.*, 2000; Sansonetti *et al.*, 1999). *Shigella* infiltration is initially aided by the inflammatory response and attraction of PMN, which breaks down the epithelial barrier. This allows bacteria to enter via a non-M-cell route, however they should eventually be entrapped and killed (Perdomo *et al.*, 1994).

The invasion pathway of *Salmonella* is similar to that of *Shigella* (Figure 1-2b). However, in addition to basolateral attack, *Salmonella* may also enter the apical side of epithelial cells through adherence via fimbrial adhesions to cells (Haraga *et al.*, 2008). Once within the phagosome upon cellular entry, rather than rupturing the vacuolar compartment like *Shigella*, *Salmonella* is retained within it, although it prevents the usual phagolysosomal recycling. This niche is the *Salmonella* containing vacuole (SCV) (Alpuche-Aranda *et al.*, 1994). Another distinctive difference between the *Salmonella* and *Shigella* virulence pathways is the existence of two distinct T3SSs in *Salmonella*. These are encoded by *Salmonella* pathogenicity islands 1 and 2 (SPI1 and SPI2). SPI1 and SPI2 are used for differing stages of infection (Hansen-Wester and Hensel, 2001); the SPI1 T3SS enables initial uptake of the bacteria into the cell, whereas the SPI2 T3SS is expressed when *Salmonella* is contained within the SCV (Haraga *et al.*, 2008).

In contrast to *Shigella* and *Salmonella*, EPEC and EHEC are not internalised, and instead attach to the intestinal epithelial cell surface. They belong to a group of pathogenic *E. coli* termed the attaching and effacing pathogens, due to the localised effacement of microvilli and the development of raised

polymerised actin pedestals which they form at the cell surface (Croxen and Finlay, 2009; Schmidt, 2010). From this surface, they may subvert the host cell cytoskeletal networks, and disrupt the intestinal barrier tight functions, amongst many effects (McNamara *et al.*, 2001). The host immune response, as for *Shigella* and *Salmonella*, is the production of IL-8 and the recruitment of PMN cells to the site of infection (Savkovic *et al.*, 1997).

Initial bacterial-host binding is mediated by a translocated intimin receptor, Tir, which is passed through the T3SS (Sal-Man *et al.*, 2009) to allow attachment of focal adhesion proteins including α -actinin, talin, and vinculin (Cantarelli *et al.*, 2001; Freeman *et al.*, 2000). Tir is anchored back to the cytoplasmic-membrane-located intimin. The T3SS is encoded on a pathogenicity island called the 'locus of enterocyte effacement' (LEE), although some effectors which are located externally to the LEE are secreted (McDaniel *et al.*, 1995).

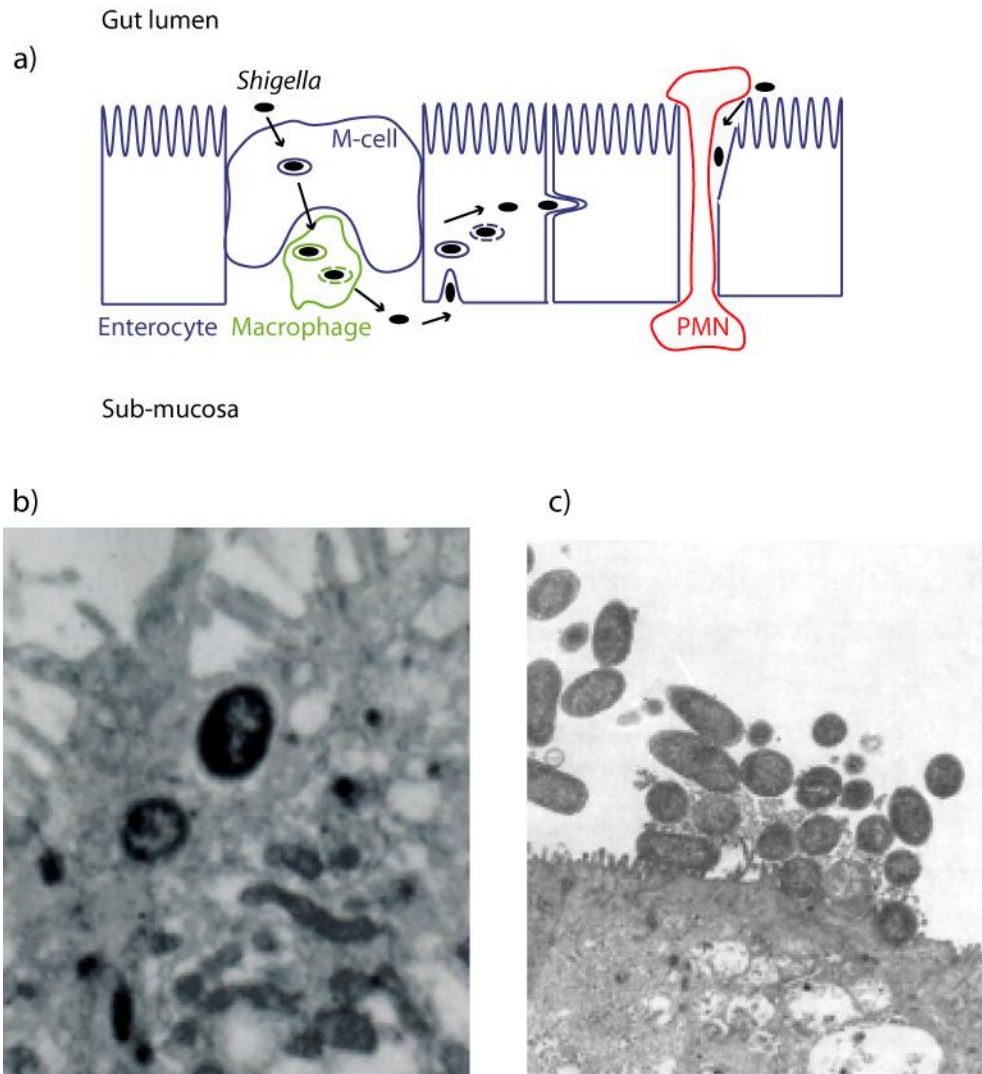


Figure 1-2 Virulence of *Shigella*, *Salmonella* and EPEC. **a)** The invasion of *Shigella* begins with its passage through the M-cells. Whilst the M-cells sample material in the gut lumen for immune surveillance, the unwitting passing of *Shigella* to macrophages in an intra-epithelial pocket allows it to induce apoptosis and be liberated at the basolateral side (Schroeder and Hilbi, 2008). This action does not go unnoticed however; macrophage death leads to the release of the proinflammatory cytokines interleukin-1 β (IL-1 β) and IL-18 (Sansonetti *et al.*, 2000). At the basolateral side, *Shigella* T3SS effectors cause host membrane ruffling. The bacterium is internalised, and breaks out from the phagosome into the cytoplasm. Here it uses IscA actin propulsion for cell to cell spread. Whilst *Shigella* attempts to conceal its presence in the cell, for example by masking of an IscA autophagy-inducing recognition site by IscB (Ogawa *et al.*, 2005), peptidoglycan microbial identifiers (Nod1) lead to the activation of the host immune defence NF- κ B pathway. This upregulates the chemokine IL-8, recruiting PMN. Recruitment of PMN makes it possible for other bacteria to access the basolateral side via a non M-cell route. (Figure reproduced (Parsot, 2005)). **b)** Electron micrograph of *S. typhimurium* attack. Unlike *Shigella*, *Salmonella* may attack either the apical or basolateral side of the epithelium. For invasion via the M-cell, entry into the epithelium proceeds via the scheme as described for *Shigella*. (Figure reproduced (Goosney *et al.*, 1999)). **c)** Attaching and effacing lesion formed by EPEC. The destruction of microvilli and pedestal formation are observable. Pedestals may be 10 μ m high and are not static, moving across the cell surface at 0.1 μ m/s (Sanger *et al.*, 1996). (Figure reproduced (Hecht, 2001)).

1.2.2 T3SS architecture.

The different pathogenic invasion strategies between *Shigella*, *Salmonella* and EPEC/EHEC are reflected in subtle structural differences across their respective T3SSs. However, these structures all share homology with flagella (Kubori *et al.*, 1998), a comparison which has led to insight into the working of the T3SS. They also share the common features of a membrane spanning basal body, hollow extra cellular needle and tip (Blocker *et al.*, 1999), which shall in this section be examined through the representative *Shigella* architecture.

The *Shigella* T3SS apparatus is encoded by the *mxi/spa* operon, found within a 31 kb section of the *Shigella* virulence plasmid alongside genes for most of the regulators, translocators and effectors of the system (Parsot, 1994). A cytoplasmic bulb (C-ring) shown in Figure 1-3 consists of the proteins Spa47, MxiN, Spa32, Spa33, MxiK and Spa13 (Morita-Ishihara *et al.*, 2006). This is a region implicated in the assembly of the T3SS and preparation of effector secretion. Next to this, in the basal body, the inner membrane is spanned by rings of MxiG and MxiJ, although MxiJ is mostly periplasmic (Allaoui *et al.*, 1992; McDowell *et al.*, 2011). The outer membrane is traversed by the secretin ring MxiD (Burghout *et al.*, 2004) accompanied by its pilotin lipoprotein MxiM (Schuch and Maurelli, 2001). Other membrane proteins include MxiA and Spa40 (and Spa9, Spa24, Spa29) which have C-terminal globular domains exposed in the cytosol (Abrusci *et al.*, (to be published); Botteaux *et al.*, 2010). The needle structure starts in the periplasm, where a MxiI rod extends towards the MxiD ring (Marlovits *et al.*, 2006). At this ring it is replaced by a helical needle assembly of MxiH subunits (Cordes *et al.*, 2005; Cordes *et al.*, 2003; Deane *et al.*, 2006). Approximately 120 MxiH copies are built in the needle extending 600 Å out from the bacteria, containing a channel of ca. 25 Å (Blocker *et al.*, 2001). Given that an α -helix has a diameter of approximately 12 Å, it is thought that effectors secreted through this channel must be unfolded to the extent of the secondary structure before assembling in the host. For control of secretion activity and insertion of the translocon into the host, a tip complex made of hydrophobic IpaB and hydrophilic IpaD

moieties associates with the needle (Espina *et al.*, 2006; Johnson *et al.*, 2007; Veenendaal *et al.*, 2007). Sensing of the host leads to recruitment of IpaC to the tip. IpaB and IpaC are inserted into the host cell membrane to form the secretion pore (Epler *et al.*, 2009; Veenendaal *et al.*, 2007).

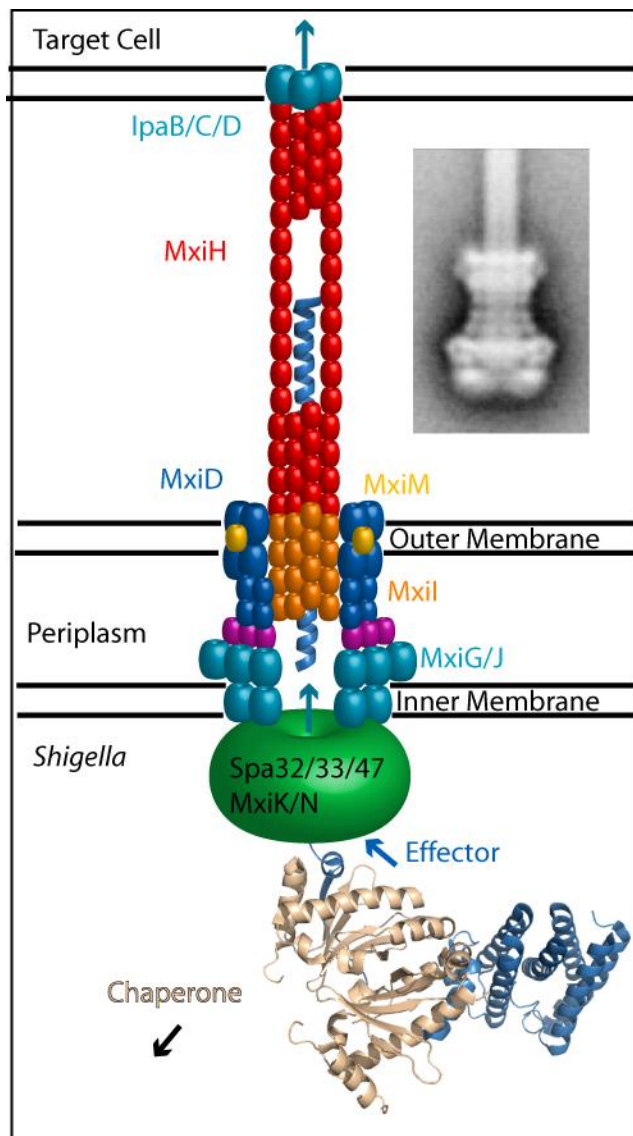


Figure 1-3 Structure and composition of the *S. flexneri* T3SS. A multitude of proteins make up the apparatus, comprising a basal body, needle and tip. The apparatus spans both bacterial membranes and the host membrane, creating a pathway from bacterial to host cytoplasm. A narrow channel allows passage of unfolded effectors, which are bound in the cytoplasm to chaperones. The chaperones are not exported. The C-ring is shown in green, the remainder of the basal body in blues/purple, and needle in red/orange. The electron microscopy insert shows the global morphology of the apparatus (Blocker *et al.*, 2001). (Figure modified (Schroeder and Hilbi, 2008)).

1.2.3 Regulation and timing.

1.2.3.1 Building the apparatus.

Due to the energetically costly nature of a fully functioning T3SS, transcription, translation and secretion of the components (apparatus and effectors) are controlled through a number of regulatory mechanisms (Brutinel and Yahr, 2008; Deane *et al.*, 2010). The stages towards apparatus completion are shown in Figure 1-4. Building of the T3SS is activated upon ingestion into the host. The major virulence regulator VirF signals the transcriptional activator VirB at 37°C, enabling protein production (Schroeder and Hilbi, 2008).

The initial inner membrane components are Sec transported (MxiG, MxiJ, MxiD), before the apparatus is completed using components secreted through the T3SS itself. The needle components MxiI and MxiH are transported until the needle reaches a length of ca. 60 nm. This value changes for differing T3SS utilising species, but is well defined within a species, according to the length required to traverse the host cell surface (Mota *et al.*, 2005). The mechanism for length regulation is disputed, but knockout of *spa32* leads to loss of length control (Tamano *et al.*, 2002). At full length, an oxygen sensing switch dependent on Spa32 and Spa40 (Botteaux *et al.*, 2010; Deane *et al.*, 2008; Marteyn *et al.*, 2010) allows secretion of the translocator proteins, and the tip complex (IpaB, IpaD) is added. However, the T3SS is not active until contact with the host cell is made.

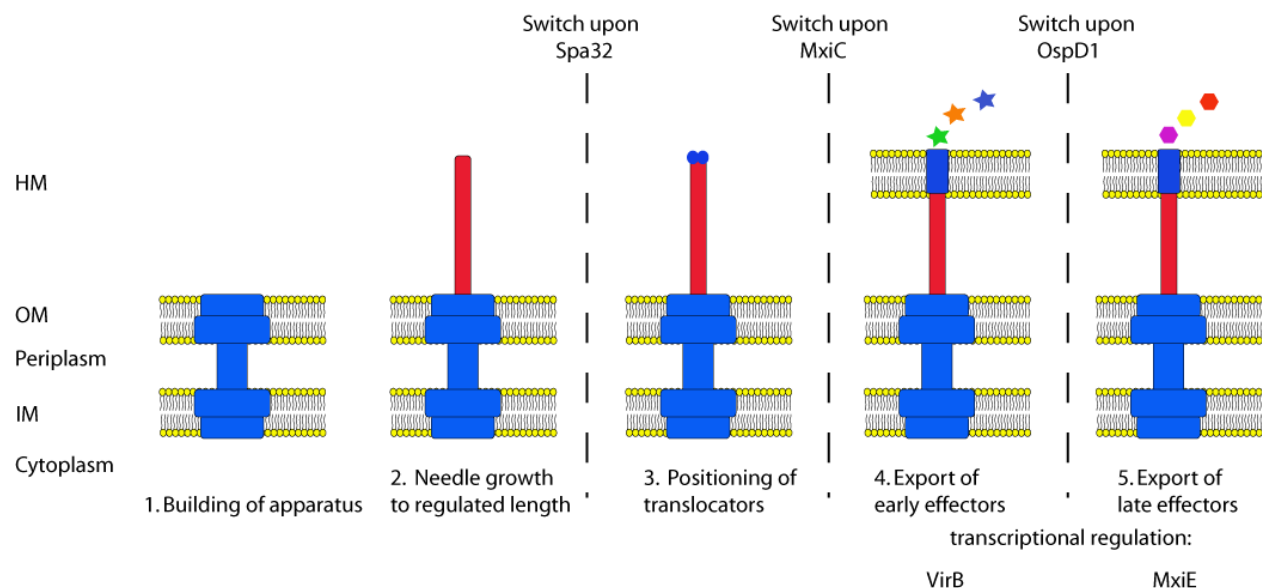


Figure 1-4 The sequential stages of T3SS apparatus building. Building of the apparatus is a multi-stage process, where the basal body, needle and tip are added sequentially as described in the text. Several switches occur regarding components secreted through the apparatus. The first is the Spa32/40 switch preventing further elongation of the needle, and inducing secretion of the translocators. The second is the MxiC switch stopping further translocator secretion, and inducing the first wave of effectors. The OspD1 switch allows transcription and subsequent secretion of the second wave effectors. (Figure reproduced (Deane *et al.*, 2010)).

1.2.3.2 First wave secretion.

The early effectors secreted are those which enable entry of *Shigella*. They are expressed under the control of VirB, independently of the T3SS apparatus (Le Gall *et al.*, 2005). This allows rapid secretion upon host cell contact. Upon host sensing, IpaC is recruited and the translocon is inserted. An undefined signal causes export of MxiC (Martinez-Argudo and Blocker, 2010). This export frees the ATPase Spa47 which MxiC previously binds (Botteaux *et al.*, 2009). Spa47 provides the impetus for release for many effectors which are cytoplasmically locked to chaperones (see Figure 1-3 and below (Akeda and Galan, 2005)). Effector transport, and *Shigella* attack may now begin.

The early effectors contain those which manipulate the host cytoskeleton, including IpaA, IpgB1 and IpgD (Izard *et al.*, 2006; Niebuhr *et al.*, 2002; Ohya *et al.*, 2005). These effectors localise to the plasma membrane and facilitate *Shigella* entry by causing membrane ruffling, either directly or via G-protein signaling cascades (Mattoo *et al.*, 2007). Early stage effectors are also used in the post-invasion stage

however. An example is IscB, which camouflages against autophagic recognition of *Shigella*, whilst the bacterium engages in cell to cell spread (Ogawa *et al.*, 2005). A secretion signal targets proteins to be translocated. However this signal does not appear to take a common form, with both mRNA and non-conserved peptide signals reported (Anderson and Schneewind, 1997; Ramamurthi and Schneewind, 2002; Samudrala *et al.*, 2009).

1.2.3.3 Second wave secretion.

One of the early effectors is OspD1, the subject of Chapter 5. Its secretion is required for the production of a second set of effectors. These are termed the 'late' effectors, expressed only after first wave effector secretion has commenced, under the control of the activator MxiE. This late set is involved at a more advanced stage of invasion, possibly in different cells. The late effectors include those which dampen the host immune response; OspG prevents NF κ B activation (Kim *et al.*, 2005) and the E3 ubiquitin ligase IpaH9.8 modulates the host inflammatory response (Ashida *et al.*, 2010).

There are some effectors whose regulation is not so clearly defined – those which are produced when the T3SS is inactive, but whose expression is increased upon activation of the T3SS. They bring about both structural and immune system subversion. Examples include VirA, required for cell entry via microtubule stabilisation and cell-cell spread (Germane *et al.*, 2008), and OspF, a phosphothreonine lyase of the mitogen-activated protein kinases, MAPKs (Kim *et al.*, 2008).

1.2.4 Chaperone usage.

Prior to the signals controlling export of translocators and first wave effector secretion, proteins which are already expressed are stored until they are required. They are often associated in a semi-unfolded state with chaperones which may be specific to one or two substrates (see IpgC in Chapter 5), or bind to multiple substrates (see Spa15 Chapter 3-5) (Parsot *et al.*, 2003; Wilharm *et al.*, 2007). The different properties and structures led to their division into classes.

1.2.4.1 Classes of chaperones.

Three types of chaperones exist in T3SSs (Parsot *et al.*, 2003). Class I are the chaperones of effector proteins, Class II the chaperones of translocators (IpaB/IpaC), and Class III the chaperones of needle component proteins (MxiH homologues). Class I effector chaperones are acidic, small molecular weight (< 20 kDa) homodimeric proteins and the class is subdivided into two: IA chaperones transport one effector and are encoded in the vicinity of this substrate (e.g. SycE of *Yersinia pseudotuberculosis* (Birtalan and Ghosh, 2001) and SicP of *S. typhimurium* (Stebbins and Galan, 2001a)). Class IB carry multiple effectors and are encoded with the operons for the structural T3SS apparatus (e.g. Spa15 of *S. flexneri* (Van Eerde *et al.*, 2004) and InvB of *S. typhimurium* (Lilic *et al.*, 2006)). Class II translocator chaperones are encoded, as are IA chaperones, in the vicinity of their substrate (e.g. SycD of *Y. enterocolitica* (Buttner *et al.*, 2008) and IpgC of *S. flexneri* (Lunelli *et al.*, 2009)). They are homodimeric when not bound to a substrate, although unlike the Class I chaperones they have a structure that does not require dimerisation for effector binding (Lunelli *et al.*, 2009). The third class of chaperones is defined from homology of the T3SS with the flagella system, where the class protects extracellular filament proteins. The EPEC T3SS protein CesA is of this class, featuring all-helical topology with an overall hairpin shape (Yip *et al.*, 2005).

1.2.4.2 Structure and binding of Class I chaperones.

The homodimeric Class I chaperones consist of monomers with five β -strands surrounded by three α -helices (Page and Parsot, 2002). Exceptions to this homodimeric nature do occur, for example in *Yersinia*, where the heterodimeric pair YscB:SycN chaperones the regulatory protein YopN - the homologue of *Shigella* MxiC (Schubot *et al.*, 2005). The chaperone C-terminus may be found immediately after the third α -helix, as is the case for SicP of Figure 1-5, or have a β -strand containing region beyond this (Luo *et al.*, 2001; Trame and McKay, 2003). Where an extended C-terminus exists, for

example in CesT of EPEC, its removal appears to prevent T3SS targeting (Thomas *et al.*, 2005). Both Class IA and IB chaperones share this fold (SicP and InvB of *S. typhimurium* share 16% sequence identity), although the subunit orientation differs between the classes by approx 30° (Van Eerde *et al.*, 2004).

Class IA and IB chaperones bind effectors by an N-terminal chaperone binding domain (CBD). This has been characterised in a number of examples including SycE:YopE, residues 15-75 (Birtalan *et al.*, 2002), SicP:SptP, residues 15-100 (Stebbins and Galan, 2001a) and InvB:SipA, residues 48-158 (Lilic *et al.*, 2006). A number of effectors bind using a common motif (Lilic *et al.*, 2006), however this is not present in all effectors. The crystal structures of the above examples show the CBD is stretched in an extended morphology, wrapping around the chaperone dimer. Two conserved interaction patches are of particular interest: The first of these is where the N-terminus of the CBD donates a β -strand to a β -sheet groove of the chaperone. The second patch sees the filling of a hydrophobic groove in the chaperone. These interactions are shown in Figure 1-5. They contribute to the extensive burial of hydrophobic surface, and to a tight interaction, calculated as $K_d = 0.3$ nM for SycE:YopE (Cheng and Schneewind, 1999).

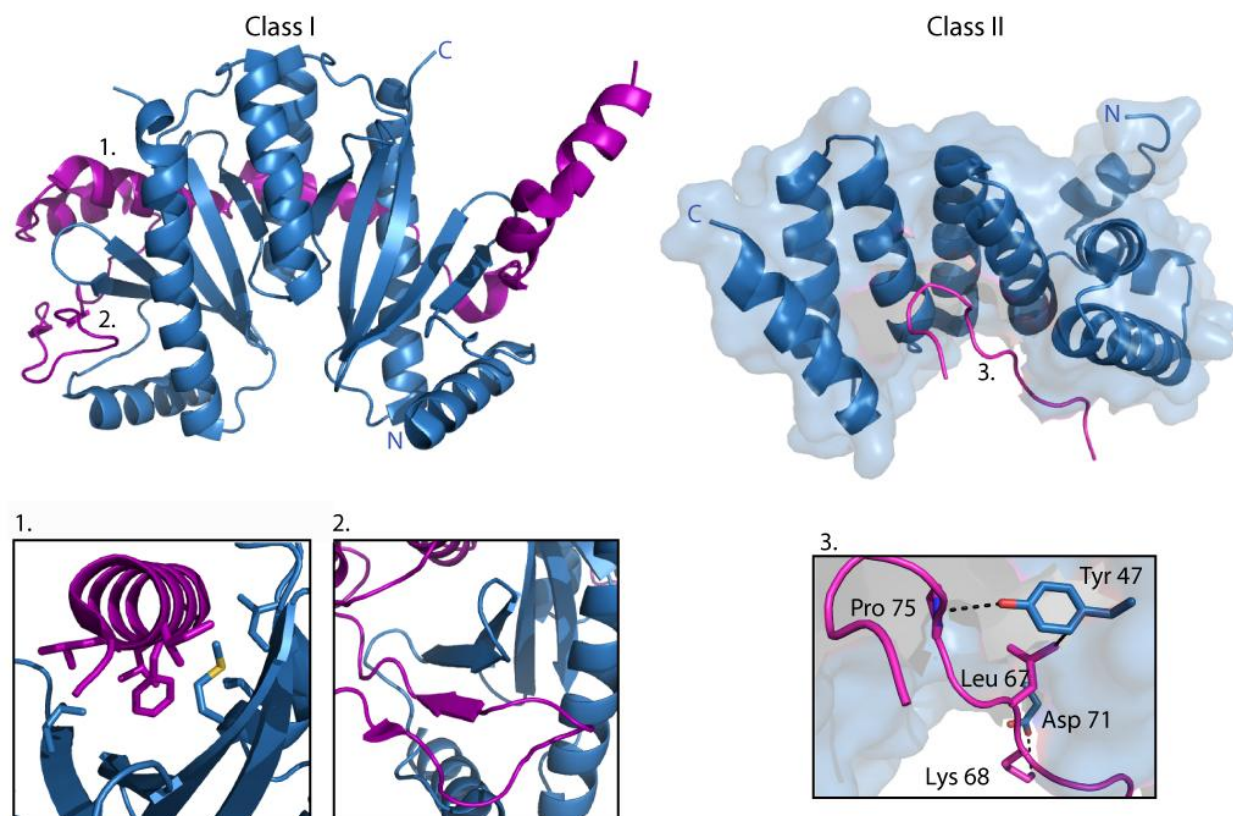


Figure 1-5 Conformation of Class I and Class II chaperones in complex with their substrates. **Left:** SicP:SptP structure showing the homodimeric chaperone SicP in blue and the CBD of the effector SptP in purple. **1.** Interactions with a conserved chaperone hydrophobic groove, including chaperone residues Leu28, Met51, Leu92, Ile85 and effector residues Leu80, Phe83, Leu87, Ile79. **2.** The donation of a β -strand from the effector to a β -sheet groove of the chaperone. **Right:** IpaB:IpgC structure showing the all α -helical monomeric IpgC structure, with concave morphology. IpgC is shown in blue and IpaB in purple. **3.** The concave cleft allows binding with the substrate via a variety of interactions, including those between Pro75-Tyr47, Leu67-Tyr47 and Lys68-Asp71 as shown.

1.2.4.3 Structure and binding of Class II chaperones.

The Class II chaperones are all α -helical moieties. IpgC of *Shigella* displays eight α -helices, containing three 34 amino acid anti-parallel α -helical motifs (Barta *et al.*, 2010). These are tetratricopeptide repeats, known to be involved in protein-protein interactions, and which form the scaffold of a concave surface. This surface is the peptide binding cleft, which is 10.5 Å wide at its narrowest point.

Upon binding the translocator substrate, the narrow cleft is used to protect the extended conformation of the N-terminal CBD, in a fashion similar to the action of the Class I chaperone (residues 51-72 for IpaB). This is achieved via a multitude of interactions, including salt-bridge, hydrogen bond, and

hydrophobic interactions. The K_d appears to be somewhat weaker than for the chaperone:effector complexes, calculated as 625 μM for the chaperone binding sequence of IpaB binding the chaperone IpgC (Lokareddy et al., 2010; Lunelli et al., 2009).

Whilst translocator proteins precede the passage of effectors in the T3SS and have roles in the initial host interaction, they may also have effector like roles. IpaC has a role in eliciting actin polymerisation, which aids the internalisation of *Shigella* by inducing membrane ruffling (Tran Van Nhieu *et al.*, 1999). The pressures and requirements on translocators and effectors to be kept chaperoned in these extended conformations may therefore be the same.

1.3 Tertiary structures of secreted proteins.

1.3.1 T3SS: The requirement for unfolded effectors.

The binding of T3SS effectors by chaperones in the bacterial cytosol had led to the emergence of alternative theories relating to the necessity of maintaining these effectors in a partially unfolded state (Ghosh, 2004). The first theory is that of passive protection, where the intended host cell function of the CBD is prevented when within the bacteria, by keeping the CBD in an unfolded state prior to secretion. Evidence for this comes from the T3SS of *Yersinia*, whose effector YopH binds phosphotyrosines of host cell proteins with its CBD (Montagna *et al.*, 2001). Furthermore, *Salmonella* SopE has a CBD which functions as a ubiquitination target (Kubori and Galan, 2003). Passive protection highlights the prevention of effector aggregation when at a high concentration level within the bacteria. For example, the CBD of YopE is highly prone to aggregation when uncomplexed (Birtalan *et al.*, 2002). However the idea of passive protection suffers from the lack of identification of a role in the CBD of many effectors.

The second theory is that of secretion competency: effectors are kept in a partially unfolded state in preparation for passage through the 25 Å T3SS channel. This view also has its opponents – they point to the effectors which have no requirement for a chaperone, and the maintenance of effector C-terminal

folding during chaperone binding. The CBD of some effectors is also naturally disordered in the absence of chaperones (Rodgers *et al.*, 2008). The proponents however suggest that a partial unfolding of the effector would lead to a priming of unfolding for the rest of the protein. This would be carried out upon interaction with the ATPase allowing release and immediate transport into the host.

1.3.2 Tat secretion: The requirement for folded substrates.

In contrast to a system which keeps its substrates partially unfolded prior to secretion and fully unfolded during secretion, the Tat system is used in bacteria due to the necessity to pass some substrates fully folded through the membrane (Wooldridge, 2009). The translocon is conserved through many bacteria and archaea (Yuan *et al.*, 2010) and is homologous to a system found in the thylakoid membrane of chloroplasts (Wooldridge, 2009). Different species utilise their systems to varying degrees; *E. coli* exports only 26 proteins via the Tat pathway, whereas halophilic archaea secrete the majority of their proteins this way, since high salt conditions in the cytosol lead to the rapid folding of proteins (Rose *et al.*, 2002). The more common requirement for the export of fully folded proteins however is where a cytoplasmic cofactor is incorporated. Examples are often redox active, including molybdenum, FAD and iron-sulfur clusters, involved in such processes as anaerobic respiration, periplasmic ligand binding, and virulence (Berks *et al.*, 2003; Lee *et al.*, 2006b).

The ability to pass folded proteins has been explored for commercial purposes (Kolkman *et al.*, 2008). Currently a number of proteases, lipases and amylases are industrially obtained by the *Bacillus subtilis* Sec pathway, due to the non-pathogenic bacteria's ability to produce large quantities (20 – 25 g/l of cells) of heterologous protein (Harwood, 1992; Schallmey *et al.*, 2004). This is an imperfect system however, due to the export by the Sec-pathway of unfolded proteins. These may be sensitive to post-translational proteolysis, and folding defects can limit the protein yield (Schallmey *et al.*, 2004). The ability of *B. subtilis* to Tat-export commercially desirable targets such as the protease subtilisin (with the

addition of a Tat signal sequence) has been explored. Subtilisin was secreted in a large quantity and shown to be active. However, subtilisin may not be representative of the *B. subtilis* Tat pathway's ability to pass non-Tat substrates, since a number of other targets (amylase, tobacco etch virus protease) were limited by cytoplasmic folding. These were either not passed by the Tat system, or inactive upon secretion (Kolkman *et al.*, 2008).

1.4 Secretion by the Tat translocon.

Tat substrates, in common with their Sec partners, have an N-terminal tripartite signal sequence. This contains a short basic n-region which incorporates the twin arginine namesake motif, a longer hydrophobic h-region, and a carboxyl terminal c-region which includes the recognition sequence for the signal peptidase (Berks *et al.*, 2003). Compared to Sec signal sequences, Tat signal sequences are longer and contain basic residues in the c-region – this ensures the avoidance of Sec translocation (Blaudeck *et al.*, 2003). Not all Tat exported proteins contain the signal sequence; some are transported in multimeric complexes where other components contain the sequence, in a 'hitchhiker' fashion. An example is the export of the dimeric DMSO reductase, containing components DmsA and DmsB. Only DmsA contains the signal sequence (Sambasivarao *et al.*, 2000).

1.4.1 The apparatus for Tat transport.

The general translocon, as exemplified by the model system *E. coli*, is composed of three essential inner-membrane proteins – TatA, TatB, and TatC. Another two Tat proteins (TatD and TatE) are non-essential for transport. Indeed, the cytoplasmic TatD appears to have no effect on Tat translocation (Wexler *et al.*, 2000). TatE, however, duplicates the functional properties of TatA, a protein with which it shares 53% sequence identity. Deletion of both TatA and TatE is required to abolish Tat transport (Sargent *et al.*, 1998; Sargent *et al.*, 1999). The *tatABCD* operon is shown in Figure 1-8. The *tatE* gene is located elsewhere on the chromosome from the *tatABCD* operon.

TatA oligomerises to form the Tat translocation pore. It has a single transmembrane α -helix, which is connected via a hinge to an amphiphatic helix and disordered region at the C-terminus (Porcelli *et al.*, 2002). The cytoplasmic-periplasmic orientation of TatA is disputed, with publications suggesting both possibilities (Chan *et al.*, 2007; Porcelli *et al.*, 2002), although a C-terminal cytoplasmic amphiphatic helix may explain a lid structure observed by electron microscopy (Gohlke *et al.*, 2005). The oligomers seen by photophysical analysis of fluorescent labelled TatA have variable stoichiometry, averaging at 25 units in a pore (Leake *et al.*, 2008). This variability is supported by single-particle electron microscopy, with observation of channel diameters from 30 – 70 Å (Gohlke *et al.*, 2005).

TatB shares 20% sequence identity with TatA, displaying a similar single transmembrane and amphiphatic helix (De Leeuw *et al.*, 2001), however it has a longer C-terminal domain. It is implicated in binding to TatC, the substrate, and to TatA. TatB and TatC bind to form the recognition complex for the substrate signal sequence. Following binding of the substrate, TatB is the component of the TatBC complex which binds the pore protein TatA (Cline and Mori, 2001). TatB is essential in Gram-negative bacteria and in the thylakoid pathway, with both transmembrane and amphiphatic α -helix domains required for function (Lee *et al.*, 2002). However, as detailed below, most Gram-positive bacteria have a Tat pathway consisting of only TatA and TatC.

TatC is expected to be composed of six transmembrane helices, with both termini located at the cytoplasmic side of the membrane (Mori *et al.*, 2001). TatC is the part of the TatBC complex which initially identifies the substrate (Cline and Mori, 2001; Holzapfel *et al.*, 2007). The TatBC complex is a 1:1 stoichiometry high order multimer (Bolhuis *et al.*, 2001; Maldonado *et al.*, 2011), probably containing 5-7 TatBC subunits (Oates *et al.*, 2003). TatB transmembrane helices reside at the core of the complex (Lee *et al.*, 2006a).

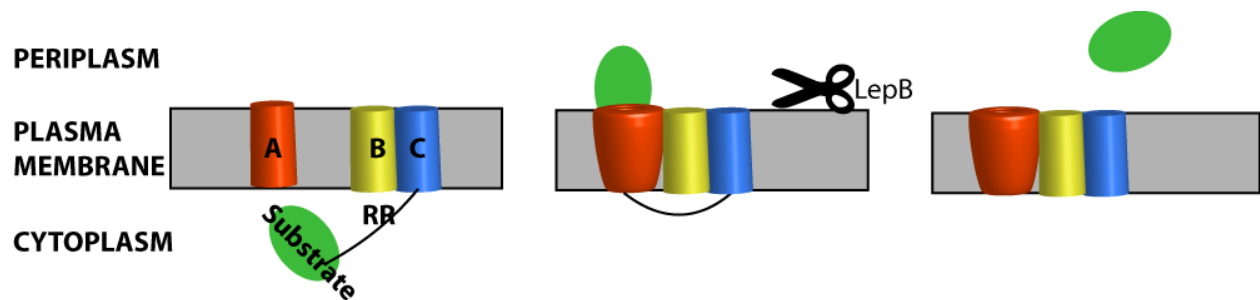


Figure 1-6 Model of Tat export. Recognition of the Tat signal peptide by a TatBC complex leads to recruitment of TatA. TatA provides the pore for translocation. Upon export to the periplasm, the signal peptidase LepB removes the signal peptide.

1.4.2 The mechanism of Tat transport.

The Tat translocon achieves the transport of folded proteins ranging from 2-7 nm in diameter through a membrane of 5-10 nm thickness without compromising ionic sealing. The process first involves binding of the substrate to the TatBC complex (Tarry *et al.*, 2009). The signal peptide is recognised by the N-terminal half of TatC (De Leeuw *et al.*, 2002; Holzapfel *et al.*, 2007). Particularly important is the first cytosolic loop, mutation of which can switch the specificity of the signal peptide from twin arginines to twin lysines (Strauch and Georgiou, 2007). Additional contacts are made to the substrate by TatB, which interacts with the folded part of the precursor protein (Maurer *et al.*, 2010). The TatBC complex can bind multiple substrates, with two copies of the Tat exported SufI shown to bind TatBC concurrently in adjacent positions by negative stain electron microscopy (Tarry *et al.*, 2009).

Active TatA next associates with the substrate bound TatBC complex (Mori and Cline, 2002), which requires an intact proton gradient across the membrane. The substrate is transported via a TatA channel (Gohlke *et al.*, 2005). Consistent with the concurrent binding of multiple SufI substrates is the observation of multiple substrates being passed collectively (Ma and Cline, 2010). A model for translocation exists where the amphiphatic helices of TatA undergo an infolding, forming a basket in which the substrate is accepted and exported (Robinson *et al.*, 2011). This is yet to be directly detected. However, it is supported by an inability to observe the arrested secretion of substrates in any stable channel, something often possible in protein-conducting channels (Cline and McCaffery, 2007).

Following substrate translocation the signal peptide is cleaved by a signal peptidase, LepB (Yahr and Wickner, 2001), and the apparatus complexes break apart (Berks *et al.*, 2000).

Unlike most bacterial translocons, ATP hydrolysis is not required, with the proton motive force sufficient for translocation energy provision. This is necessary for binding of TatA to the substrate bound TatBC complex. A study determined that 7.9×10^4 protons are released per translocated protein (Alder and Theg, 2003).

1.4.3 The Gram-positive Tat apparatus.

As with many Gram-positive bacteria (with the notable exception of *Streptomyces* (De Keersmaecker *et al.*, 2005)), *B. subtilis* is deficient of a TatB protein. TatA binds directly to TatC, fulfilling the roles of both TatA and TatB from the above described Gram-negative system (Barnett *et al.*, 2008). *B. subtilis* expresses three TatA and two TatC isoforms. These make up two distinctive Tat translocons, TatAdCd and TatAyCy, which transport specific effectors (Eijlander *et al.*, 2009b; Jongbloed *et al.*, 2004). When a TatAdCd system was overburdened, a TatAdCy translocon was observed, however the pairings are generally specific. The role of the third TatA (TatAc) is unknown, but it has been shown to have no role in Tat-dependent secretion (Eijlander *et al.*, 2009a).

The NMR structure of TatAd in *B. subtilis* was recently determined. The N-terminal transmembrane helix, amphiphatic helix and disordered C-terminus domains were observed (Hu *et al.*, 2010). Insertion of TatAd into planar bicelles shows a slanted alignment of the amphiphatic helix, forcing a number of hydrophilic residues into the membrane. The slant is shown in Figure 1-7. It is speculated that the hydrophilic residues are stabilised by intra- or intermolecular salt-bridges in the TatAd pore (Walther *et al.*, 2010). Regions of TatAd which pertain to TatB (extreme N-terminus) and TatA function (C-terminal end of the periplasmic helix and hinge) from the Gram-negative Tat model have been identified. This has allowed a homologous model of Tat transport to be proposed (Barnett *et al.*, 2011).

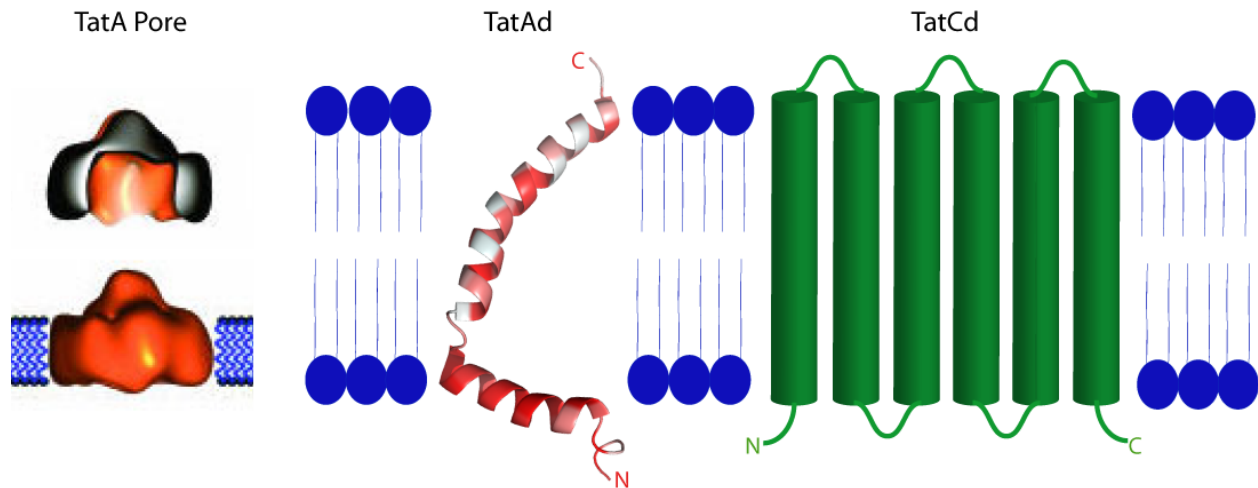


Figure 1-7 The Tat translocon of *B. subtilis*. Left: TatA pore (Gohlke *et al.*, 2005) The top figure shows a cross-sectional view of the bottom pore ©. Middle: NMR structure of TatAd (Hu *et al.*, 2010) and its alignment in the membrane (Walther *et al.*, 2010). Right: TatCd is predicted to contain six transmembrane helices (Nolandt *et al.*, 2009).

The Tat translocon is not extensively used in *B. subtilis*. There are two proteins known to be exported (PhoD and YwbN). Two further targets have been putatively identified via a *Streptomyces coelicolor* agarase assay (Widdick *et al.*, 2008) (YkuE and QcrA – homologue of the Tat exported Reiske protein in *Legionella pneumophila*). Finally an additional 69 proteins have been identified which are predicted to be secreted via the *B. subtilis* Tat pathway, based on their signal sequence properties (Jongbloed *et al.*, 2002). One of the known substrates, PhoD, is the subject of Chapter 6. This is the only known protein exported by the system TatAdCd. It has been suggested that the translocon is specifically constructed for PhoD. PhoD is a protein required in times of phosphate starvation, and TatCd has been shown to only be expressed under these conditions (Jongbloed *et al.*, 2000). Proteins containing the unusual PhoD signal sequence have been shown to be exported by the TatAdCd pathway, including β -galactosidase. (Pop *et al.*, 2002; Xia *et al.*, 2010).

The other pathway of *B. subtilis*, TatAyCy, is used by the iron-dependent peroxidase YwbN. The gene for YwbN is located apart from its transport apparatus, unlike the *phoD*, *tatAd* and *tatAy* genes which are located together on the chromosome (Figure 1-8) (Jongbloed *et al.*, 2000).

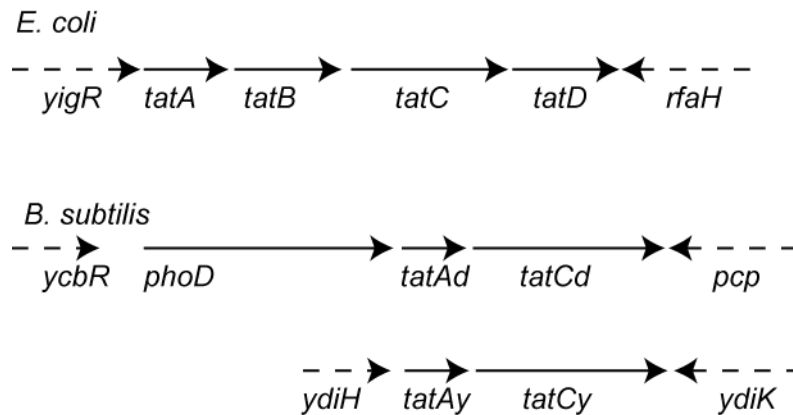


Figure 1-8 Comparison of the chromosomal organisation between *E. coli* and *B. subtilis*. In *E. coli*, the *tatABCD* genes are organised in an operon. *tatE* is located elsewhere on the chromosome. *B. subtilis* has two functional Tat translocons. The first, encoded by *tatAd* and *tatCd* lies directly downstream of the *phoD* gene, the only known substrate of this translocon. The second translocon, encoded by *tatAy* and *tatCy*, do not lie near that of their substrate *ywbN*, which is situated elsewhere on the chromosome. (Figure reproduced (Jongbloed *et al.*, 2000)).

1.4.4 Proofreading and quality control.

There is evidence that the Tat system prevents export of incorrectly folded substrates, although the field remains divided as to how strict this control is. A number of non-Tat secreted proteins have been shown to pass the translocon when given the correct signal sequence and cytoplasmic folding, including cytochrome c (Sanders *et al.*, 2001) and the disulphide bond containing PhoA (DeLisa *et al.*, 2003). In support of strict control, the exposure of hydrophobic patches in incorrectly bound substrates has been shown to exclude Tat export for some tested substrates (Richter *et al.*, 2007). Conversely, the transport of unstructured peptides up to 120 residues has been reported (Cline and McCaffery, 2007).

The exact method of proofreading control is also the subject of debate (DeLisa *et al.*, 2003; Sanders *et al.*, 2001). Discrimination may occur prior to or post TatBC binding (Richter and Bruser, 2005). The former looks more likely, with reports of a Tat-dependent proofreading mechanism for the *E. coli* Fe/S protein NrfC (Matos *et al.*, 2009; Matos *et al.*, 2008) now revoked, in favour of an unknown Tat-independent pathway (Lindenstrauss *et al.*, 2010).

It appears that a significant number of Tat-dependent substrates do reach the translocon correctly folded, partly due to the roles of specific and general chaperones which they bind. These assist in folding, cofactor insertion and association with partner proteins, as well as preventing interactions with the translocation machinery prior to folding (Robinson *et al.*, 2011; Sargent, 2007).

The first known chaperone example was DmsD of *E. coli*. This interacts with the Tat signal sequence of the DMSO reductase DmsA (Oresnik *et al.*, 2001), but without targeting DmsA to the Tat system. This lack of a targeting role suggested an alternative one in enzyme maturation for Tat chaperones, such as aiding the insertion of the DmsA molybdopterin cofactor (Ray *et al.*, 2003).

Indeed, such a role has been determined for the *E. coli* TorD chaperone. This too binds the signal sequence of its substrate TorA (trimethylamine N-oxide reductase subunit A), as well as to the cofactor biosynthesis component, and to the cofactor. The coincident binding led to conclusions that TorD acts to facilitate the cofactor insertion to its substrate (Genest *et al.*, 2008). This interaction is TorA signal peptide specific, however other chaperones (for example *E. coli* SlyD and DnaK) interact with numerous Tat-exported proteins. This is independent of the cofactor acquisition process (Graubner *et al.*, 2007; Perez-Rodriguez *et al.*, 2007).

Similarly to the T3SS therefore, substrates have different requirements of chaperones, although with the Tat chaperones a potential role is that of folding assistance. For both T3SS and Tat chaperones, the protection of substrates from proteolysis or unwanted interactions is a common function.

1.5 The elucidation of secretion biochemistry with physical methods.

As with many biological questions, physical methods originally (and still) used in the study of inorganic and simple organic systems are now powerful in providing structural and functional answers for such complicated machines as secretion systems and their substrates.

Electron microscopy and nuclear magnetic resonance (NMR) are techniques which have enabled visualisation of the T3SS needle apparatus and provided structures where crystals proved unobtainable (Hodgkinson *et al.*, 2009; McDowell *et al.*, 2011). However the spectroscopic tools of X-ray crystallography with electron paramagnetic resonance (EPR) and surface plasmon resonance (SPR) are the techniques which have been used in this study. These techniques are introduced in Chapter 2.

From its initial application in the field of protein structure with Hodgkin, Perutz and Kendrew, X-ray crystallography accounts for approximately 90% of protein structures in the PDB (Rupp, 2010). These include *Shigella* T3SS structures for the apparatus - MxiH (Deane *et al.*, 2006), translocators - IpaD (Johnson *et al.*, 2007) and effectors - IpgB2 (Klink *et al.*, 2010). However it is often necessary to supplement this impressive information with solution state research, to avoid false conclusions from crystal effects.

EPR allows the oligomerisation and complexation of proteins to be examined, along with the determination of the cofactor and its environment. It was applied to biological systems in the 1960s, however initially had limited applications. These were increased upon the introduction of site-directed spin labelling (Berliner, 1984), the technical development of EPR spectrometers (Bennati and Prisner, 2005), and the introduction of pulse sequences to measure interesting parameters such as distances (Schiemann and Prisner, 2007; Prisner *et al.*, 2001). Relevant to this body of work, EPR studies have shown interactions between TatA helices forming the ring-like pore of the Tat system (White *et al.*, 2010), and observed the membrane insertion dependency of the Fe containing Riske protein ISP upon the Tat translocon (Bachmann *et al.*, 2006).

Another light resonance technique widely used to determine protein-protein interactions is SPR. This stemmed from initial observations relating to the odd reflection of polarised light from a metal backed diffraction grating by Wood and Lord Raleigh at the beginning of the 20th century (Rayleigh, 1907; Wood,

1902). Since the 1990s, the technique applied to biochemical interactions has observed binding of *Pseudomonas aeruginosa* PcrV (the T3SS homologue to *Shigella* IpaD) to host antigen binding fragments, and of the *Salmonella* T3SS effector SopE to GTPases which engender cytoskeletal alteration (Papalia *et al.*, 2006; Rudolph *et al.*, 1999).

1.6 Project aims.

The first aim of this project was to examine effectors which are secreted by the T3SS and Tat translocon, and to highlight the role of the chaperone in the T3SS passing unfolded peptides, while the Tat translocon transports ready to function, folded, substrates.

The T3SS is a virulence organelle, responsible for transit of proteins directly to the host. As shown above, this is achieved by constructing a narrow needle structure and puncturing the host cell membrane. The first part of this thesis will examine the bacterial environment prior to secretion where the effectors are kept in a bound state by chaperones. Once the unfolded proteins are passed through the T3SS, they fold and interact with host targets, often with novel functional, rather than structural mimicry.

The latter part of the thesis focuses on a protein secreted by the Tat pathway. Since the role of this pathway is not to target host cytosolic effectors, this protein need not be transported unfolded through a thin channel into a host. The requirement for folded transport is actually greater than this; the unusual nature of the protein's cofactor environment indicates that folding must be carried out in the bacterial cytoplasm. The thermodynamic preference for the protein to fold is thus allowed in a way that is explicitly prevented by the type three effector chaperones.

The second aim of this thesis is to highlight how the different physical techniques of X-ray crystallography and EPR may be used to complement each other. Specifically, a study is presented in Chapter 3 showing how EPR can determine distances between spin labels in a protein oligomer which

may be verified using the more established method of X-ray crystallography. This lends weight to its future use as an independent technique in cases where crystals are unobtainable. As an example of the independent use of EPR, the oxidation state of the cofactor in the crystal structure determined in Chapter 6 has been established. Elucidation of metal identity and oxidation states is not always possible using X-ray crystallography, however in the case of paramagnetic examples, EPR may conclusively identify them.

2 Materials and Methods.

2.1 Molecular biology.

2.1.1 Constructs and primers.

Primers (Sigma-Aldrich Co.) were designed to contain restriction enzyme cleavage sites for inclusion into Novagen's T7 promoted pET, pACYC and pCOLA plasmids via common sticky ends (Studier and Moffatt, 1986). Table 2.1 gives details of the constructs used in this study.

No.	Vector	Reference/primer	Insertion
3.1	pET28::spa15	A. Veenendaal	
4.1	pET28::espM2	Arbeloa <i>et al.</i> , 2010	
4.2	pET28::espM229-196	Arbeloa <i>et al.</i> , 2010	
4.3	pMW172-His-RhoA F25N	Valderrama <i>et al.</i> , 2006	
4.4	pMW172-His-Rac1	Valderrama <i>et al.</i> , 2006	
4.5	pET28::espM229-196 W70A	Arbeloa <i>et al.</i> , 2010	
4.6	pET28::espM229-196 L118A	Arbeloa <i>et al.</i> , 2010	
4.7	pET28::espM229-196 Q124A	Arbeloa <i>et al.</i> , 2010	
4.8	pET28::espM229-196 I127A	Arbeloa <i>et al.</i> , 2010	
4.9	pET28::sifA	Arbeloa <i>et al.</i> , 2010	
4.10	pET14b::ipgB1	F:CGCGGCAGCCATATGCAAATTCTAAACAAAATACTTCCA R:TTAGCAGCCGGATCCTTAATTTGTATTGCTTTGACG	Nde1/BamH1
4.11	pACYC::ipgB1	F:CGCGGCAGCCATATGCAAATTCTAAACAAAATACTTCCA R:TTAGCAGCCCTCGAGTTAATTTGTATTGCTTTGACG	Nde1/Xho1
4.12	pET15b:: ipgB158-208	F.CGCGGCAGCCATATGGACAGCGTCATAGACAAAATAAAA R.TTAGCAGCCCTCGAGTTAATTTGTATTGCTTTGACG	Nde1/Xho1
4.13	pACYC::ipgB158-208	As 4.12	Nde1/Xho1
4.14	pACYC::ipgB1 K115A K116A	F:GTTTTGACCTCAGATAACGCAGCAAACATATTTGCAGCTATT R:AATAGCTGCAAATATGTTTGCTGCGTTATCTGAGGTCAAAC	Mutagenesis
4.15	pACYC::ipgB1 K123A, K124A	F: ATATTTGCAGCTATTGAAGCAGCATTCCCGATATAAACTT R:AAGTTTTATATCCGGGAATGCTGCTCAATAGCTGCAAATAT	Mutagenesis
4.16	pACYC::ipgB1 K153A, R153A	F: CTCAGAGCAAAGATTTTAGCAGCTTATAGCAGTGACATGGAT R:ATCCATGTCACTGCTATAAGCTGCTAAAATCTTTGCTCTCTGAG	Mutagenesis
4.17	pACYC::ipgB2	F:CGCGGCAGCCATATGCTTGAACATCTTTAATAATTTTGA R:TTAGCAGCCCTCGAGTTAGAAAGCGATTCTAAATTTGTAAT	Nde1/Xho1
4.18	pET15::ipgB2	As 4.17	Nco1/Xho1
5.1	pET15::spa15	F:CGCGGCAGCACCATGGGCAGTAACATTAATTTAGTTCAA R:TTAGCAGCCGAATTCTTATAAGACCCCATTTAAGATTTC	Nco1/EcoR1
5.2	pACYC::His-spa15	Digested from 3.1	Nco1/EcoR1
5.3	pACYC::ospD1	F:CGCGGCAGCCATATGTCAATAAATAACTATGGATTACATCCA R: TTAGCAGCCCTCGAGTTAAACAACCTCTTGATTTTGTGAGA	Nde1/Xho1
5.4	pACYC::spa15::ospD1	spa15 as 5.1 ospD1 as 5.3	Nco1/EcoR1 Nde1/Xho1
5.5	pET15b::ospD1	as 5.3	Nde1/Xho1
5.6	pCOLA::ospD1::mxiE	ospD1 as 5.3 mxiE digested from 5.7	Nde1/Xho1 Nco1/Sal1
5.7	pET28b::mxiE	S. Johnson	Nco1/Sal1
5.8	pCOLA::mxiE	Digested from 5.7	Nco1/Sal1
5.9	pET15(lpgC)	S. Johnson	
5.10	pACYC(lpaC)	S. Johnson	
5.11	pACYC(lpaB)	S. Johnson	

Table 2.1 Constructs used. Labels denote the relevant chapter. Donated constructs are labelled according to the benefactor.

2.1.2 Polymerase chain reaction.

To obtain high concentrations of DNA insert (50 – 100 ng/μl), polymerase chain reactions (PCR) were carried out using PFU turbo (Agilent) or KOD polymerase (Novagen). All genes were amplified from *S. flexneri* (M90T) virulence plasmid (purified via a miniprep kit (QIAprep)), using the forward and reverse primers shown in Table 2.1. Deoxynucleotide triphosphates, magnesium chloride and a suitable buffering agent completed the reaction mixture. This was cycled through the denaturation, annealing and extension steps as detailed in Table 2.2. In some cases, non-specific amplification required the use of the Touchdown protocol. Here a decreasing annealing temperature ensured that the initial reaction cycles were achieved at a temperature close to the primer melting temperature (Don *et al.*, 1991). This temperature was calculated using a calculator for the particular oligonucleotide (Promega_Biomath).

Step	Construct					
	4.10 PFU ultra	4.11 PFU ultra	4.12 PFU ultra - Touchdown	4.17 PFU ultra	5.1 PFU ultra - Touchdown	5.3 KOD
1. Initilisation (°C, secs)	98, 180	98, 180	94, 300	98, 180	94, 300	94, 300
2. Denaturation (°C, secs)	98, 30	98, 30	95, 30	98, 30	95, 30	98, 15
3. Annealing (°C, secs)	60, 30	60, 30	55 – 0.3°C / cycle, 30	60, 30	55 – 0.3°C / cycle, 30	68, 2
4. Extension (°C, secs)	72, 90	72, 90	72, 4	72, 90	72, 4	72, 20
5. Repeat Steps 2-4 (no. of times)	30	30	10, before 20 x regular PCR	30	10, before 20 x regular PCR	30
6. Final Elongation (°C, secs)	72, 300	72, 300	72, 600	72, 300	72, 600	72, 300

Table 2.2 PCR protocols. Steps for the individual constructs are detailed.

2.1.3 Cloning protocol.

The stepwise insertion of the PCR product into a plasmid is shown in Table 2.4. The inserted plasmid was transformed into a cloning strain of *E. coli* to allow its purification. Upon DNA sequencing to verify the insert, the plasmid was transformed into an *E. coli* expression strain.

An alternative pathway was used when creating mutants from a construct already inserted into a plasmid. Since this is considered a construct optimisation method, protocol details are given in section 2.4.

At each stage of construct and vector manipulation, products were checked for purity and size using gel electrophoresis. An applied electric field causes the negatively charged DNA to pass through the gel, retarding the larger molecules relative to the smaller. Addition of 6 x Blue/Orange Loading Dye (Promega) made the size separation progression of the gel easy to follow. Final length determination of the DNA size was by comparison to standard DNA size markers (1 kb DNA Ladder (New England Biolabs)) relative to the protein which was dyed with Ethidium Bromide or SYBR safe dye, and viewed under UV or visible light (respectively). The gel composition and running buffer is described in Table 2.3, and the gels were run in a Bio–Rad Mini Sub Cell GT at 60 V.

Gel: 50 ml deionised H ₂ O, 0.5 g agarose (0.1% w/v), 40 mM Tris-acetate pH 8.3, 2 mM disodium EDTA, 0.5 µg/ml Ethidium Bromide/ SYBR safe	Running buffer: 40 mM Tris-acetate pH 8.3, 2 mM disodium EDTA, 7 µl Ethidium Bromide (10 mg/ml).
---	--

Table 2.3 Agarose gel and TAE running buffer composition.

Recombinant cloning techniques enable overexpression of large quantities of protein relative to natural abundance, allow the addition of affinity tags (Rupp, 2010), and have the advantage of limiting contact to potentially harmful pathogens such as *S. flexneri* explored in this study. All bacterial strains were grown in Luria Bertani (LB) broth, containing 1% (w/v) tryptone, 0.5% (w/v) yeast extract and 1% (w/v) NaCl. Use of antibiotic resistance (either 100 µg/ml ampicillin, 25 µg/ml chloramphenicol or 50 µg/ml kanamycin according to the plasmid) ensured growth of only the resistant plasmid containing bacteria. After insertion of the construct, all plasmids were sequenced to confirm their identity (Source Bioscience, University of Oxford).

Step	Description	Kit/enzyme/stain used.
1.PCR purification	Post PCR, any contaminants/reagents remaining are removed so as to not inhibit future reactions. Obtain concentrated pure DNA.	PCR purification kit (QIAquick).
2.Restriction digests	Use of two restriction endonucleases reduces the risk of self ligation. DNA & plasmid are incubated at 37°C for two hours with the first enzyme before heat inactivation (65°C 2 mins, 4°C 2 mins) and addition of a second restriction enzyme. This is incubated for a further two hours at 37°C.	Both Fermentas Fast-digest and New England Biolabs restriction enzymes used.
3a.Gel purification	Undigested, singly digested and doubly digested DNA are separated by gel electrophoresis at 60 V. The required doubly cut DNA is visualised with Ethidium Bromide/UV light or Invitrogen SYBR safe/visible (470 nm) light, excised with a blade and purified with the Qiagen DNA extraction kit.	Invitrogen SYBR safe dye. DNA extraction kit (QIAquick).
3b.Alkaline phosphatase	Alternatively to gel purification, self ligation of singly cut vector can be prevented by removal of 5' phosphates of the vector. Reaction incubates at 37°C for 10 minutes before heat inactivation.	Fermentas FastAP™ Thermosensitive Alkaline Phosphatase.
4.Ligation	DNA insert becomes bonded within the plasmid by reaction between 5'-phosphate and the 3'-hydroxyl groups of the adjacent nucleotides in a ligase catalysed reaction requiring incubation at 21°C for 30 minutes.	Promega (LigaFast)T4 DNA ligase.
5.Plasmid uptake by high transformation cloning cells	The plasmid is taken up by cells in the process of transformation. Selectivity is conferred by exploiting antibiotic resistance of cells which contain the insert plasmid. 1-10 ng of plasmid are chilled with 20 µl NovaBlue cells, before heat shock at 42°C for 1 minute. Chill before plating onto LB + 1.5% Agar and incubating at 37°C overnight. Resulting colonies picked and grown at 37°C overnight in LB broth.	Novagen NovaBlue cells.
6.Minipreparation	DNA is purified from cloning cells in steps of cell lysis, neutralisation and high salt precipitation of proteins/chromosomal DNA/cell debris. Soluble plasmid DNA is then caught and purified on a silica membrane before elution in 10 mM Tris-HCl, pH 8.5.	Minipreparation kit (QIAprep).
7.Expression bacteria transformation	The plasmid is taken up into strain B834(DE3), and selected for antibiotic resistance. Transformation is as for the cloning cells, except for a 37°C, 1 hour incubation prior to plating allows sufficient antibiotic resistance to be attained.	Novagen B834(DE3).

Table 2.4 Cloning strategy. Stepwise protocol used, starting with < 10 µg DNA construct and circular plasmid.

2.2 Protein expression and purification.

Each protein was expressed in *E. coli* B834(DE3) with selective antibiotic, the exceptions being the IpaB:IpgC and IpaC:IpgC complexes for which Tuner(DE3) (Novagen) cells were used. Hexahistidine affinity tags were used for purification. The cells were grown at 310 K until $A_{600\text{ nm}} = 0.6$, where protein expression was induced upon addition of 1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG), and cells harvested by centrifugation after 15 hours at 293 K (4000 x g, 20 min, Beckman JLA-8.1000 rotor). The cells were resuspended in HisA buffer containing complete EDTA-free protease inhibitor cocktail (Roche) with 250 μ l of 10 mg/ml DNase, and lysed using an Emulsiflex-C5 Homogeniser (GC Technology) at 15000 psi. After separation of the soluble fraction by centrifugation (13000 x g, 30 mins, Beckman JA-25.50 rotor), the eluant was flowed through a Ni-NTA Superflow cartridge (Qiagen – Table 2.5) where the protein was eluted by competition with an imidazole gradient (between the imidazole concentrations for HisA and HisB buffers given in Table 2.6. The final purification step was either size exclusion chromatography using a HiLoad Superdex 75 or 200 column (GE Biosciences), or ionic exchange chromatography using a salt gradient on a Mono Q 5/50 GL column (GE Healthcare) – Table 2.5. All columns were attached to an ÄKTA FPLC system (GE). Purification buffers are given in Table 2.6.

Ni-NTA affinity

The affinity of histidine to nickel ions is exploited by attaching a polyhistidine-tag to the protein (Hochuli *et al.*, 1988). Removal is via displacement by imidazole, which competitively binds the nickel. This is achieved by a gradient between 10 – 1000 mM imidazole, after washing of approximately twelve column volumes containing 10 mM imidazole to remove other nonspecific bound proteins (Ni-NTA).

Size exclusion chromatography

Separation of proteins according to size/hydrodynamic volume. The Superdex columns used in this study contain pores between the polymer beads of differing sizes (composed of cross-linked agarose and dextran) which allows fastest transfer of large proteins, with later elution of the smaller proteins. Proteins too large to enter any of the pores are eluted in the void volume. S75 and S200 columns separate globular proteins of approx < 75 kDa, and 600 kDa respectively (Superdex_75/200).

Ion exchange chromatography

Separation of proteins according to charge. In this study anion exchange was performed using a MonoQ column, binding the charged protein in 20 mM NaCl buffer pHs at 7.5. This was eluted with an increasing salt gradient which shields the protein charge from the resin (Mono_Q).

Table 2.5 Protein purification columns used.

Protein (Chapter)	Expression conditions	Plasmid	Cell yield mg/L	Purification steps	HisA/B buffer	Size exclusion/ion exchange buffer
Spa15 (3)	1mM IPTG, 20°C O/N	pET28b	40	Ni-NTA then S75	A = 20 mM Tris pH 7.5, 150 mM NaCl, 20 mM imidazole B = A + 300 mM imidazole	SE: 20 mM Tris pH 7.5, 150 mM or 50 mM NaCl
EspM2 ²⁹⁻¹⁹⁶ (4)	1mM IPTG, 20°C O/N	pET28b	20	Ni-NTA then S75	A = 20 mM Tris pH 8.0, 500 mM NaCl, 1mM DTT, 20 mM imidazole B = A + 300 mM imidazole	SE: 50 mM Tris pH 8.0, 150 mM NaCl, 5 mM DTT
RhoA (4)	37°C constitutive expression	pMW172	20	Ni-NTA then S75	A = 50 mM Tris pH 7.5, 500 mM NaCl, 1 mM MgCl ₂ 5 mM imidazole B = A + 200 mM imidazole	SE: 25 mM Tris pH 7.5, 0.5 M NaCl, 5 mM MgCl ₂
Rac1 (4)	37°C constitutive expression	pMW172	20	Ni-NTA then S75	As above	As above
SifA (4)	1mM IPTG, 20°C O/N	pET28b	5	Ni-NTA then S200	A = 50 mM Tris pH 7.5, 500 mM NaCl, 5 mM imidazole, 5 mM DTT B = A + 750 mM imidazole	SE: 50 mM Tris pH 7.5, 500 mM NaCl, 5 mM DTT
Spa15:lpgB1 (4)	1mM IPTG, 20°C O/N	pET28b(<i>spa15</i>) pACYC(<i>lpgB1</i>)	2	Ni-NTA then QSepharose/ MonoQ (salt gradient)	A = 50 mM Tris pH 7.5, 500 mM NaCl, 5 mM imidazole, 5 mM DTT B = A + 300 mM imidazole	MQ: A: 50 mM Tris pH 7.5, 20 mM NaCl, 5 mM DTT B: 50 mM Tris pH 7.5, 1M NaCl, 5 mM DTT
Spa15:OspD1 (5)	1mM IPTG, 20°C O/N	pET28b(<i>spa15</i>) pACYC(<i>ospD1</i>)	2	Ni-NTA then S200	A = 25 mM Tris pH 7.5, 500 mM NaCl, 10 mM imidazole, 10 mM β-MC B = A + 750 mM imidazole	SE: 25 mM Tris pH 7.5, 150 mM NaCl, 10 mM β-MC
Spa15:OspD1: MxiE (5)	1mM IPTG, 20°C O/N	pACYC(<i>spa15</i> , <i>ospD1</i>), pET28b(<i>mxiE</i>)	2	Ni-NTA then S200 or MQ	As above	SE: As above, alternatively MQ: A: 50 mM Tris pH 7.5, 20 M NaCl, 5 mM DTT B: 50 mM Tris, 500 mM NaCl, 10 mM β-MC
lpgC:lpaB, lpgC:lpaC (5)	1mM IPTG, 20°C O/N	pET15b(<i>lpgC</i>) pACYC(<i>lpaB</i> or <i>lpaC</i>)	5	Ni-NTA then S200	A = 25 mM Tris pH 7.5, 150 mM NaCl, 10 mM imidazole, 10 mM β-MC B = A + 750 mM imidazole	25 mM Tris pH 7.5 150 mM NaCl

Table 2.6 Induction, expression and purification conditions of the proteins produced.

2.3 Protein solution characterisation.

Prior to crystallisation the integrity of each protein was tested using a variety of techniques, including SDS-PAGE, Mass Spectrometry and Multiangle Laser Light Scattering (MALLS).

SDS-PAGE is a widely used technique which separates proteins on the basis of electrophoretic mobility, which translates approximately as weight. The gel specifications used for reducing SDS-PAGE are given in Table 2.7. Gels were run in a Bio-Rad Mini Protean Tetra Cell, at 25 mA/gel through the stacking gel and 50 mA through the separating gel. Protein bands were visualised by Coomassie Brilliant Blue staining, or for small amounts of protein silver staining was used. Size comparisons were made using Bio-Rad Standard Broad Range Markers.

Solution	Component
SDS-PAGE separating gel 15% (Acrylamide percentage varied 10 – 20%)	Acrylamide (30% w/v National Diagnostics) Tris-HCl pH 8.8 (0.33 M) SDS (0.1% w/v) NNN'N'tetramethylenediamine (TEMED 0.01% (w/v) – Sigma) Ammonium persulphate (APS – 0.1% w/v - Sigma)
SDS-PAGE stacking gel	Acrylamide (3%) Tris-HCl pH 6.8 (0.125 M) SDS (0.1 M w/v) TEMED (0.01%) APS (0.1% w/v)
5 x SDS-PAGE sample buffer	Tris-HCl pH 6.8 (0.25 M) Glycerol (50% v/v) DTT (10 mM) SDS (10%) Bromophenol blue (0.5%)
SDS –PAGE running buffer	Tris – HCl (25 mM) Glycine (250 mM) SDS (0.1% (w/v)
Coomassie Brilliant Blue stain	Methanol (50% v/v) Acetic acid (10% v/v) Coomassie Brilliant Blue R250 (0.25 % v/v)
Destain	Methanol (20% v/v) Acetic acid (7% v/v)
Silver stain	Invitrogen Silver Quest Staining kit

Table 2.7 SDS-PAGE solutions.

Mass spectrometry confirmed the mass of proteins accurately, using the method of electrospray ionisation. This sprays the sample through a fine charged nozzle, which then is directed towards the detector by the repeller electrode (Ashton *et al.*, 1994). Distinctive fragments may also be observed. Mass Spectrometry was carried out at the OPPF and Department of Biochemistry, Oxford, by Joanne Nettleship and David Staunton.

MALLS exploits the refraction of light (by Raleigh scattering) from a molecule in solution to give a shape independent mass calculation for the protein or oligomeric species in solution. Two adjacent molecules will refract, and the resultant waves from these may constructively or destructively interfere. In time, molecular motion will cause a fluctuation from constructive to destructive interference. The rate of this fluctuation depends on size – i.e. the frequency of intensity change is plotted as molar mass. A number of detectors are placed around the sample because the intensity, approximately given by equation 2.1, will not be isotropic for a non-spherical protein,

$$I(\theta) \propto MC\left(\frac{dn}{dc}\right)^2 \quad 2.1.$$

where M is the mass, C the concentration, and dn/dc the change in refractive index with concentration (Xie *et al.*, 2002).

The SEC-MALLS equipment used in this study was a Multiangle Light Photometer (Dawn Heleos II, Wyatt) connected to a high pressure liquid chromatography (HPLC) system (Agilent Technologies 1200 series) onto which a Superdex 200 10/300 column (GE) was attached to allow separation of the sample material for detection. The flowrate used was 0.4 ml/min, and the buffer was specific to the protein sample, as given in the final column of Table 2.6.

2.4 Protein/DNA construct optimisation.

A number of modifications to the constructs were made throughout this study to aid crystallisation. Specific examples and results will be mentioned in the individual chapters, however the more common methods are described below.

Methods	Protocol
Partial proteolysis	• 1:80 – 1:10000 protease:protein (w/w) of trypsin and subtilisin carlsberg (Sigma-Aldrich Co.) are added to the protein and incubated at 21°C for two hours. • These are serine proteases (the former more specific than the latter). Trypsin cuts at the carboxyl side of lysine and arginine, except when either is bound to a C-terminal proline. • The purpose is to identify the general stability of the protein, and any more stable core regions which may be suitable crystal targets in the event of the full length protein not crystallising.
Lysine methylation	• Dimethylaminoborane and formaldehyde will methylate exposed lysine residues whose flexibility can inhibit crystallisation through entropic factors. • 1M dimethylaminoborane complex (20 µl/ml protein) and 1 M formaldehyde (40 µl formaldehyde/ml protein) are added to the protein, in three steps overnight. The protein is kept in a non-amine containing buffer at < 1 mg/ml to avoid significant precipitation, until reaction quenching in 20 mM Tris- HCl. • Protocol described in (Walter <i>et al.</i>, 2006).
Surface entropy mutants	• High surface entropy of residue patches may be reduced by replacement of lysine/arginine rich patches with small, hydrophobic residues, typically alanines. This makes them more conducive to forming crystal contacts (Cieslik and Derewenda, 2009; Derewenda and Vekilov, 2006). Protocol: 1) Primers designed to be complimentary to the two strands of the insert-containing-plasmid of interest. Primer length 25-45 bases with the mutation in the middle are used. 2) A PCR reaction (20 cycles of 95°C 30 secs denaturation, 55°C 60 secs annealing, 68°C 1min/kb of plasmid length extension) separates the plasmid strands, and extends the primers to the complete length of the primer to create a plasmid with staggered nicks. 3) Addition of DpnI restriction endonuclease digests the original dam methylated plasmid (1 µl of the DpnI restriction enzyme (10 U/µl) added to a 50 µl reaction). 4) Transformation of the vector - 1 µl - into ultra competent cells (NovaBlue Gigasingles - 50 µl), required for uptake of non-supercoiled DNA, with transformation as Table 2.4 section 7. Nicks are repaired here. 5) Minipreparation and DNA sequencing of resultant plasmid, before transformation into expression lines. • Detailed protocols are described in the Agilent mutagenesis kit handbook (QuikChange).

Table 2.8 Construct optimisation methods.

2.5 Crystallisation.

Crystallisation may occur when a supersaturated protein solution phase separates, resulting in an ordered crystal array existing in equilibrium with the solution. Achieving the specific metastable supersaturated phase is a delicate balance between many variables. Manipulation of these variables, such as water content, pH, or concentration of a precipitant are carried out in the crystallisation experiment. This employs a buffered solution, the 'mother liquor', designed to encourage supersaturation and nucleation, via salt, organic precipitant, or other additives. In this thesis, the sitting drop vapour diffusion method was used for crystallisation. For this method, a well of the mother liquor is placed next to a well containing a mixed protein and mother liquor drop. These are equilibrated over time to allow dehydration of the smaller protein drop (Rupp, 2010). Temperature may affect the protein solubility and crystallisation. In this work, crystallisation trials were performed at 277, 285 and 294 K.

Initial screening was carried out with Molecular Dimensions broad screens, which provide ready-made solutions shown to previously yield protein crystals, with both systematic and 'sparse' screening of conditions (Table 2.9). A gauge for a suitable initial protein concentration used was the 'Stura test': 1 μ l Stura Footprint condition 2C (0.2 M imidazole malate pH 7.0, 20 % w/v PEG 4000) was added to 1 μ l protein sample. Evidence of precipitation between 30 and 300 seconds indicated a potentially suitable concentration level.

MD Broad screen	Type
Structure Screen I and II	Sparse matrix screen. Samples salts, polymers, organics and pH.
JCSG+	Sparse matrix screening, PEG and salt conditions at the pH range 4.0 – 10.0.
Pact Premier	Systematic screen designed for pH screening with a PEG/Ion environment.
Stura Footprint and MacroSol	Systematic screen observing the protein precipitant solubility curve.
Morpheus, ProPlex, Wizard, Index	Other screens used infrequently.

Table 2.9 Molecular dimensions crystal trial broad screens.

The concentration used would be refined upon observation of extensive precipitation/lack of precipitation in the initial crystal trays over the subsequent days/weeks. The protein was mixed in differing ratio drops with mother liquor using an OryxNano robot (Douglas Instruments) – typically 50% and 25% protein initially in the two drop, 96 well MRC plates. Initial drop size was 0.4 μ l, with subsequent enlargement up to 0.8 μ l used in attempts to obtain larger crystals. Initial crystals would be optimised with homemade screens. These commonly took the broad screen condition which yielded crystals as a basis, and varied the pH and precipitant concentration systematically above and below the condition's values. This would tweak the environment to allow the largest growth possible. Details of particular optimisation screens are specified in the relevant chapters.

2.6 X-ray crystallography.

2.6.1 Diffraction in protein crystallography.

In a crystal, the proteinaceous motif is arranged in a periodic lattice, which is divided into repeating unit cells. Rotational symmetry operations which act upon a motif describe its point group. Eleven point groups result from the chiral nature of proteins. These may be fitted into fourteen possible Bravais lattices (seven unique crystal systems). Addition to the point group of the translational requirements of the repeating unit describes the crystal space group, of which there are 65 chiral groups (Hahn, 1978; Rupp, 2010).

In an X-ray diffraction experiment, constructive interference of interatomic wavelength X-rays which pass through the lattice bring about observable diffraction. The conditions for constructive interference come from the Laue equations. These are often interpreted through Bragg's law - scattering occurs 'in phase' from a set of planes when the path difference between waves is an integer number of wavelengths. The X-ray diffraction experiment yields diffraction spots in the reciprocal lattice representation. Reciprocal lattice points are perpendicular to the real lattice planes. A geometric figure,

the Ewald Sphere, uses the centrosymmetric reciprocal lattice to show which real-space Miller planes satisfy the Bragg diffraction condition (Figure 2-1).

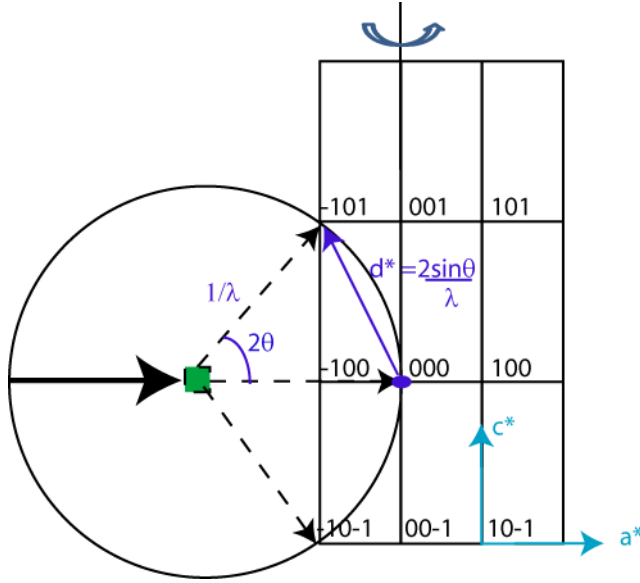


Figure 2-1 The Ewald Sphere. The Bragg condition $n\lambda = 2d \sin \theta$ is fulfilled where an Ewald circle of radius $1/\lambda$ intersects with a reciprocal lattice point (h, k, l) . More points may be brought into coincidence with the circle by rotation of the crystal around the axis shown. (Figure reproduced (Rupp, 2010)).

The crystallographic task is to extract the real space electron density from the reciprocal data collected.

A scattered wave (atomic scattering factor) from an atom or a volume element, f_s , is a wave dependent on the Bragg phase difference $2\pi i \mathbf{S} \mathbf{r}$ (where $\mathbf{S}$ is the path difference vector between two incoming waves at a distance $\mathbf{r}$), with a magnitude corresponding to the electron density in this volume element, $\rho(\mathbf{r})$.

$$f_s = \int_{\mathbf{r}}^{V(atom)} \rho(\mathbf{r}) \exp(2\pi i \mathbf{S} \mathbf{r}) d\mathbf{r} \quad 2.2.$$

Every atomic scattering factor $f_{s,j}$ (from every atom j) is summed to give the unit cell structure factor, with each contribution having phase $2\pi(hx_j + ky_j + lz_j)$, given in Miller indices.

$$\mathbf{F}_{hkl} = \sum_{j=1}^{\infty} f_{s,j} \exp(2\pi i(hx_j + ky_j + lz_j)) \quad 2.3.$$

Structure factors (generally) obey Friedel's Law: the centrosymmetric nature of reciprocal space causes the related reflection pairs (h, k, l) and $(-h, -k, -l)$ to have equal structure factor magnitude. This results in equal intensity observed for these reflection pairs. Breakdown of this rule via anomalous scattering is widely used to overcome the phase problem (see below).

The electron density is the Fourier transform of equation 2.3 (Blow, 2002; Rhodes, 2006; Rupp, 2010). The electron density gives the structure of the crystal, and so equation 2.4 is the fundamental equation of protein crystallography.

$$\rho(x, y, z) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \mathbf{F}(hkl) \exp(-2\pi i(hx + ky + lz)) \quad 2.4.$$

2.6.1.1 The phase problem.

The above equations assume that all the reciprocal data is known in order that the real space electron density be extracted. This is not the case, since the observed intensity is the hermitian product of the structure factor amplitude with its complex conjugate, $\mathbf{F}(hkl)\mathbf{F}(hkl)^*$. The necessarily real result is the square of the structure factor amplitudes only, $|F(hkl)|^2$, with the phase information required for solution of equation 2.4 lost (Rupp, 2010; Taylor, 2003).

Much of the difficulty with obtaining a structural solution relates to finding this phase information. Experimental determination and molecular replacement are two methods which have been used in this thesis. Upon completion, the electron density directly shows the structure and fold of the protein.

2.6.2 Data collection.

Crystals were mounted for X-ray data collection using Hampton Research CrystalCap nylon CryoLoops. They were cryoprotected with a suitable cryoprotectant (McFerrin and Snell, 2002), and immediately cooled with liquid nitrogen. X-ray diffraction data were collected at 100 K in an attempt to minimise radiation damage at both Diamond Light Source, Oxfordshire, UK and the European Synchrotron Radiation Facility, Grenoble, France.

Diffraction images for (h, k, l) reflection observation were collected by rotation of the mounted and centred crystal around a single axis while it was exposed to the X-ray beam. Initial images 90° apart allowed the determination of unit cell parameters and crystal orientation (indexing) using iMosflm.

2.6.2.1 Indexing and strategising a data collection.

iMosflm indexes the crystal by finding spots within the resolution radius, using the Ewald sphere reconstruction to derive the scattering vector of a reflection from the spots it gives (Leslie, 1992). Only when a cell axis lies along the beam direction will a Fourier transform of the scattering vectors lead to regularly spaced maxima (Steller *et al.*, 1997). iMosflm therefore associates maxima with the directions of the principle axes. From the list of lattices possible with this indexing, iMosflm selects the highest symmetry with reasonable penalty option. The mosaic spread of the spots predicted by this model is estimated (Leslie, 2006; Rupp, 2010).

Prior to data collection, from the symmetry chosen, iMosflm advises on a strategy for data collection. The rotation range for each image ($\delta\phi$) is often set to 1°, although this can be altered based on the cell dimensions and mosaicity – in Chapter 6 the large cell edge led to a $\delta\phi$ of 0.5° being used. The strategy is frequently used as a guide taking into account the requirements of the data – large multiplicity was required for anomalous signal in the PhoD MAD dataset so extra degrees were taken.

2.6.3 Structure solution.

A battery of programs were used in the processing, phasing and refinement of data. These are described in the next sections and include a number from the CCP4 assembly of software (Collaborative Computational Project Number 4, 1994).

2.6.3.1 Data processing.

2.6.3.1.1 Integration of data: iMosflm and XDS.

After data collection, the cell dimensions and experimental parameters may be refined using a number of small batches of images, spread apart in rotation angle. This is done by seeing how the intensity distribution of partially recorded spots varies over the images, to determine the exact angle where the reciprocal lattice point lies on the Ewald sphere.

Once the cell has been optimised, integration can be carried out. Here, the positions of the spots are predicted (using the derived symmetry) and the spot position integral taken with an associated standard deviation (Leslie, 2006). Integration may be two- or three-dimensional: iMosflm uses 2D integration, where the partially recorded reflections on each image are calculated separately and only summed to give the total integral at the point of scaling and merging (Leslie, 1992). XDS however uses 3D integration, where partial reflections are assembled by the integration software and a 3D profile evaluates the total intensity (Kabsch, 2010b).

The actual integration takes a background signal around the spot and sums the pixel values at the spot relative to the background. This is effective for strong spots, whereas for weaker spots a least-squares summation of peak pixels called profile-fitting achieves better results (Diamond, 1969).

Once the reflections have been integrated, an error associated with these measurements is calculated and the reflection intensities output in an ‘.mtz’ file ready for scaling and merging.

2.6.3.1.2 Symmetry checking: Pointless.

Once the data have been integrated, the true Laue symmetry (the point group plus centre of inversion) and an attempt at space group symmetry can be determined using the program Pointless (Evans, 2006). This is necessary since iMosflm uses only the lattice geometry for indexing, while Pointless has the already integrated diffraction pattern at its disposal. Pointless scores the symmetry operators in the diffraction pattern (via a correlation coefficient and an R-factor), and ranks the point groups accordingly. Further inspection of axial systematic absences allows determination of the space group, by multiplication of the Laue group probability with the systematic absence probability.

2.6.3.1.3 Scaling and merging files: Scala.

The integrated data is presented as a list of indices, reflection intensities and errors. The CCP4 program Scala (Evans, 2006; Evans, 1997) is used in the process of data reduction. Here the redundancy of multiple measurements of symmetry related observations is used to generate a scaling model with which to scale the frames (Kabsch, 2010a). This compensates for physical experimental differences in scales of spot intensities. The scaling model accounts for variation in rotation angle, B-factor, absorption of secondary beams, and errors due to diffuse scattering.

The declaration of the presence of anomalous scatterers ensures that Bijvoet related observations are not considered equivalent (Rupp, 2010). A file is output which merges multiple observations of the same reflection, giving the merged observation and averaged intensity. A number of statistics are also output (Table 2.10) which help in analysis of the data quality.

Data quality statistic	Description	
$\langle I / \sigma(I) \rangle = \frac{1}{N} \sum_h \frac{ I_{(h)} }{\sigma(I_{(h)})}$	The resolution of the data is given for merging N observations of a reflection h within a resolution shell. A measure of signal: noise ratio across the resolution bins.	2.5.
$R_{merge} = \frac{\sum_h \sum_{i=1}^N I_{(h)i} - \bar{I}_{(h)} }{\sum_h \sum_{i=1}^N I_{(h)i}}$	$\bar{I}_{(h)}$ = averaged intensity of each reflection. This term may be denoted R_{sym} when merging symmetry related reflections. This statistic increases with increased multiplicity of the data.	2.6.
$R_{rim} = R_{meas} = \frac{\sum_h \frac{N}{N-1} \sum_{i=1}^N I_{(h)i} - \bar{I}_{(h)} }{\sum_h \sum_{i=1}^N I_{(h)i}}$	The redundancy independent merging R-value does not suffer from the multiplicity problem of R_{merge} .	2.7.
$R_{pim} = \frac{\sum_h \frac{1}{N-1} \sum_{i=1}^N I_{(h)i} - \bar{I}_{(h)} }{\sum_h \sum_{i=1}^N I_{(h)i}}$	The precision indicating merging R-value. The merging of more observations leads to an increase in precision which is accounted for with R_{pim} .	2.8.
$R_{anom} = 2 \frac{\sum_h I_{(-h)} - I_{(+h)} }{\sum_h (I_{(-h)} + I_{(+h)})}$	Anomalous merging R-value. $I_{(-h)}$ and $I_{(+h)}$ are the mean intensities of the Bijvoet pairs for the observation h .	2.9.
$B_{iso} = 8\pi^2 \langle u_{iso} \rangle^2$	The B-factor, B_{iso} , or isotropic displacement factor, accounts for disorder in atoms encompassing a variety of reasons, including temperature related vibrations, and crystal defects. It may be expressed as the mean square displacement of an atom from its mean path $\langle u_{iso} \rangle$, and accounts for the greater reduction in intensity seen than expected.	2.10.
Data Completeness	Some loss from the ideal 100% data completeness can be tolerated in the highest resolution shell. Systematically missing data will cause problems with structure solution.	

Table 2.10 Data quality statistics.

2.6.3.1.4 Calculation of $|F|$ from intensities: Truncate.

The program Truncate takes the merged intensities and converts them to structure amplitudes, which is useful with respect to negative intensities which are otherwise poorly dealt with (French and Wilson, 1978). It also separates a Free R set of data from the remaining data which is not to be used in structural solution, to ensure that data do not become over-biased during refinement (Brunger, 1992).

Truncate outputs intensity statistic plots to help estimate data quality. These include the Wilson plot, which shows how the B-factor varies with resolution. Truncate is furthermore used to identify anisotropic and twinned data.

2.6.3.1.5 Automatic data processing: Process and Xia2.

All the above processes may be carried out using Xia2 (CCP4 (Winter, 2010)) and Process (part of the autoPROC pipeline - Global Phasing Ltd (Vonrhein *et al.*, 2011)). These may be run in a fully automatic or semi-automatic way. Xia2 will use Mosflm and Scala if “-2d” integration is requested, or XDS and XSCALE (Kabsch, 2010b) for “-3d”. Process uses XDS and Scala.

2.6.3.2 Determination of the asymmetric unit: Matthews coefficient and Patterson function.

Multiple protein copies may exist in the asymmetric unit, related via rotational or translational non crystallographic symmetry (NCS). Their presence may be determined with knowledge of only the Laue group and Bravais centring, through inspection of the matthews coefficient or patterson function. Knowledge of NCS may aid in phasing.

The Matthews volume, V_M , of the sample is the ratio of the volume of the asymmetric unit to the molecular weight of the protein (Matthews, 1968). The average solvent content of a protein crystal, 43%, corresponds to an average V_M of $2.5 \text{ \AA}^3\text{Da}^{-1}$, giving an indication of the number of protein molecules in the asymmetric unit.

$$\text{Solventcontent} = 1 - \frac{1.23}{V_M} \quad 2.11.$$

However, the coefficient is only a probabilistic guide, and for high symmetry space groups several solutions for numbers of molecules in the asymmetric unit may appear possible.

The Patterson function $P(u)$ may be used to show the nature of the NCS, since it shows all interatomic vectors in the unit cell (Rupp, 2010). Mathematically, $P(u)$ is a convolution integral of the electron density at position r and the electron density at position $r + u$:

$$P(u) = \int_R \rho(r)\rho(r+u)dr \quad 2.12.$$

The Patterson function is calculated using intensities F_h^2 (no phases needed) by Fourier transformation of the reflections hkl at positions uvw , with knowledge of the cell volume V :

$$P(uvw) = \frac{2}{V} \sum_{h=0}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F_h^2 \cos 2\pi(hu + kv + lw) \quad 2.13.$$

A native Patterson map, where the map coefficients are native intensities, may be used to detect translational NCS or rotational NCS parallel to a crystallographic axis. Whilst there are many interatomic vectors in the Patterson map, the translation of all atoms in a molecule by the same vector causes a sharp peak to be observed. This allows observation of the NCS, and determination of the translation vector.

Proper NCS may be found via the rotation Patterson function. Here the native Patterson is rotated upon itself. The convolution function is maximised upon overlap at the observed rotation angle. Self rotation functions may be drawn as stereographic projections with the CCP4 program MolRep (Vagin and Teplyakov, 2010).

2.6.3.3 Experimental phasing and model building.

In Chapter 6, the phase problem is overcome by a Multiple Anomalous Dispersion (MAD) experiment. The following sections describe the method and program used for this particular solution to the phase problem.

2.6.3.3.1 X-ray absorption.

Materials absorb X-rays such that the greater the X-ray energy, the less the absorption. Heavier atoms absorb more energy at a given wavelength. There are certain resonances which correspond to inner shell electron ejection. These lead to an increase in absorption, at the K edge (1s electrons), L edge (2s/p electrons) and M edge (3s, p, d electrons). The size of the edge depends on the number of excitable electrons. The total scattering factor $f_{(S,\lambda)}$ for an atom may be described by

$$f_{(S,\lambda)} = f_{(S)}^0 + f'_{(\lambda)} + if''_{(\lambda)} \quad 2.14.$$

where normal angle dependent (S) scattering is supplemented by real and complex wavelength dependent anomalous scattering. The dispersive or anomalous differences which heavy atoms provide can be used. Dispersive differences are the intensity difference recorded by a particular reflection when the wavelength changes. Anomalous differences are the intensity differences between Friedel pairs of centrosymmetric related reflections.

Practically, a MAD data collection takes a common form. The first wavelength to be selected gives the maximal anomalous f'' signal, since the largest anomalous difference will be observable here. The second wavelength is the high energy remote set, which has some anomalous signal, and a large dispersive difference against the “inflexion set”. This inflexion set is the third set, so called because it is the point of inflexion of the f'' part. The minima of the f' curve is also located here, giving the greatest dispersive signal when compared to the remote set. It is usually recorded last due to ease of missing it (Hendrickson, 1991; Hendrickson and Ogata, 1997).

2.6.3.3.2 Heavy atom finding: autoSHARP.

Isomorphous and anomalous methods of phase reconstruction are similar in that the presence of a heavy atom is used as a marker. autoSHARP is an automated solution structure pipeline, finding and

refining heavy atom sites from reflection data, which allows phasing and potential building of an initial map (Vonrhein, 2007). Within autoSHARP, SHELXD determines the heavy atom substructure for the reflection data, using dual space direct methods and experimental information such as the chemical nature of the sites. This method is used since it is more achievable to solve the structure of the heavy atom model alone for a particular resolution of data. Patterson based seeding provides starting phases—interatom vectors are taken and a translational search is performed. The best initial scores are expanded to give the number of expected heavy atoms. Alternation between reciprocal and real space allows chemically sensible restraints (distances, number of atoms) to influence the phases which are expanded by the tangent formula, which relates phase angles φ_k between normalised structure factors E for reflections h, k and $h-k$ (Rupp, 2010).

$$\tan(\varphi_h) = \frac{\sum_k |E_k \cdot E_{h-k}| \sin(\varphi_k + \varphi_{h-k})}{\sum_k |E_k \cdot E_{h-k}| \cos(\varphi_k + \varphi_{h-k})} \quad 2.15.$$

2.6.3.3.3 Heavy atom refinement and phasing.

In Sharp, initial phases are used to refine the positions of the heavy atoms via maximum likelihood statistics (Bricogne *et al.*, 2003). A likelihood function is generated based upon the probability of the trial native structure factors given the heavy model parameters, which is maximised by varying the heavy atom model. The independence of structure factor probabilities is accounted for by integrating out the phase, to leave a likelihood function which is dependent on the model parameters only (De La Fortelle and Bricogne, 1997). Sharp outputs heavy atom refinement statistics: R_{Cullis} and phasing power P_{iso} of the heavy atoms.

$$R_{Cullis} = \frac{\sum_h \left| F_{PH(obs)} \pm F_{P(obs)} - F_{H(calc)} \right|}{\sum_h \left| F_{PH(obs)} \pm F_{P(obs)} \right|} \quad 2.16.$$

$$P_{iso} = \frac{\sum_h F_{H(calc)}}{\sum_h \left| F_{PH(obs)} - F_{PH(calc)} \right|} \quad 2.17.$$

Upon convergence of the heavy atom model, the initial protein map is made taking the best structure factors F_{best} as Fourier coefficients:

$$F_{best} = mF_p \exp(i\phi_{best}) \quad 2.18.$$

Which are Fourier summed to give the electron density:

$$\rho(x) = \frac{1}{V} \sum_h F_{best} \exp(-2\pi i h x) \quad 2.19.$$

2.6.3.3.4 Density modification: Solomon.

Initial phases may be greatly improved by incorporation of known properties of a protein and the solvent around it. The electron density maps are density modified within autoSHARP by Solomon (Abrahams and Leslie, 1996). This is a phase improvement program which uses solvent flipping to invert the density values in the solvent region and hence improve phase probabilities. Solomon is a classical (as opposed to statistical) modification program - it implements real space modifications before back-transforming the flattened map into modified structure factor amplitudes and phases (Cowtan, 2010).

2.6.3.3.5 Automatic model building: Buccaneer.

Solvent flattened phases are used by the program Buccaneer to trace a polypeptide chain into the modified density using the protein sequence (Cowtan, 2006). Buccaneer will output a PDB file of the model coordinates obtained which is suitable for refinement. Initial C α positions are determined and chains extended. The amino acid sequence is docked into position, and optimised. A key feature is the

ability to use a partially built model as input, for example a previous Buccaneer job model or a partial molecular replacement solution.

2.6.3.4 Molecular replacement.

Phases may be obtained by using a known structure model in the process of molecular replacement, provided there is sequence identity > 30% and a r.m.s.d. of < 2.0 Å between model and structure Cα atoms.

2.6.3.4.1 MolRep.

The program MolRep finds the real space electron density using the known reciprocal structure factor amplitudes $|F_{hkl}|$ (experimental), which are correlated to those of a suitable model $|F_{hkl}|$ (model) at a series of trial orientations and translations (Vagin and Teplyakov, 2010). When they are similar, the correct positioning of the model is known, so the known phases of the model may be used as a starting point for the experimental structure.

The model orientations and translations are compared via Patterson methods calculable from the intensities, maximising the convolution functions 2.20 and 2.21 in the rotational and translational search respectively (Evans and Mccoy, 2008; MolRep documentation):

$$R(\kappa, \phi, \psi) = \int_{r \min}^{r \max} P_{(Exp)}(u) P_{(Model)}(\kappa, \phi, \psi, u) du \quad 2.20.$$

$$T(t) = \int_{cell} P_{2 \rightarrow 1}(u^r - t^r) P_{native}(u^r) du^r \quad 2.21.$$

$P_{2 \rightarrow 1}(u^r)$ is the intermolecular vector set. A correlation coefficient may also score the positions, calculated between the observed and calculated square amplitudes.

$$c(t) = \frac{\sum(|F_{obs}|^2 - \langle |F_{obs}|^2 \rangle)(|F_{calc}|^2 - \langle |F_{calc}|^2 \rangle)}{\sqrt{\sum(|F_{obs}|^2 - \langle |F_{obs}|^2 \rangle)(|F_{calc}|^2 - \langle |F_{calc}|^2 \rangle)}} \quad 2.22.$$

Solutions are assigned a score, which accounts for the packing of molecules (with penalty for collisions). These scores are ranked by contrast, the ratio of the top score with the mean score.

2.6.3.4.2 Phaser.

An alternative MR program is Phaser (McCoy *et al.*, 2007). This uses maximum likelihood functions rather than correlation. However it carries out a similar order of rotational then translational search before accounting for packing than MolRep does. Allowance for model incompleteness and model errors increases the chance of a successful solution. Two statistics are output, the log likelihood gain (LLG) and the Z-score (Read, 2001). The LLG shows how much better than random the orientated model is, using a Wilson distribution, and the Z-score is the number of standard deviations above the LLG mean value. These statistics are given for both the rotational and translational searches. The best solution is rigid body refined before being output in PDB format.

2.6.3.5 Refinement and validation.

Crystallographic refinement is an iterative process of real space modelling (fitting the model to the electron density) followed by reciprocal space refinement (optimising the fit between observed and calculated structure factor amplitudes (Rupp, 2010)). The ease of the process depends on the quality of the phases.

Refinement was carried out with both Refmac (Murshudov *et al.*, 2011) and autoBUSTER programs (Blanc *et al.*, 2004). These both use maximum likelihood functions to optimise the coordinates, B-factors and TLS of the input model against reflection data. Reciprocal space refinement may help to correct the many stereochemical errors introduced in the real space refinement by giving rise to the R-factor:

$$R = \frac{\sum_h F_{obs} - F_{calc}}{\sum_h F_{obs}} \quad 2.23.$$

Model bias is prevented by comparing the R-factor for the refined data with the R_{free} -factor, the same equation 2.23 calculated for the 5% of unrefined data which were separated from the working set.

During refinement, R and R_{free} factors should decrease with cycle and not diverge. However, refinement will also account for the geometric properties of proteins – including bond lengths and angles, by outputting two geometry statistics, the Z-scores of the deviations of bond lengths and angles from ideal values (Tronrud, 2004).

In this study, refined models were viewed and manipulated in real space with Coot (Emsley *et al.*, 2010) before re-refinement using the manipulated model. Coot plots the differences between NCS copies, Ramachandran plots and unmodelled or overmodelled density to detect addressable errors in real space. Once the chain and residues approach completion, solvent molecules which form hydrogen bonds to the protein/other solvent molecules can be added.

Stereochemical validity of the model was checked with Molprobit (Chen *et al.*, 2010). Global parameters such as the R-factor fail to observe the local problems in the structure, such as individual Asn/Gln sidechain amide conformations. Molprobit performs analysis of the protein stereochemistry – including clashscore and all-atom contact analysis.

2.7 Electron paramagnetic resonance.

2.7.1 Continuous wave-EPR.

2.7.1.1 Theory.

The observable measured in EPR, in common with all magnetic resonance, is the quantum mechanical property of spin. Whereas in NMR magnetic nuclei are detected, in EPR the unpaired electron is the species of interest. The spin of a particle is characterised by its spin quantum number, which takes integer or half-integer values. Here the single electron is considered, whose spin quantum number is $s = 1/2$. This shall be regarded in the absence of other electrons or magnetic nuclei. Spin is quantised in both magnitude ($\hbar\sqrt{s(s+1)}$, where $\hbar$ is Planck's constant/ 2π) and space ($m_s = \pm 1/2$, where m_s gives the values of s projected onto the z-axis) for the electron. In the absence of a strong magnetic field B_0 however, these are degenerate. Application of the magnetic field down the z-axis in the EPR experiment leads to a lifting of degeneracy; spins will align parallel ($m_s = -1/2$) or antiparallel ($m_s = +1/2$) with B_0 separated in energy according to the Zeeman Effect by $g_e\mu_B B_0$. A resonant microwave $h\nu$ reconnects these energy levels (Atkins and De Paula, 2006; Atkins and Friedman, 2005):

$$\Delta E = h\nu = g_e\mu_B B_0 \quad 2.24.$$

The frequency independent g -factor g_e may be reported from an experiment. Elucidation of this characteristic value can allow the simplest continuous wave EPR experiment to give information about a paramagnetic species, such as its identity and oxidation state, as achieved in Chapter 6 (Bruker, 2011). The CW-EPR spectra can furthermore be used to measure concentration of paramagnetic species (Chapter 3).

2.7.1.2 Sample preparation.

An unpaired electron moiety, associated with a protein for EPR measurement, may take a number of forms (Prisner *et al.*, 2001). In this study, they take the form of paramagnetic cofactors (high-spin Fe^{3+} in PhoD and PhoX of Chapter 6), or an artificially added spin label (labelling the cysteine residues in Spa15 and the Spa15:IpgB1 complex of Chapter 3).

To spin label the cysteine residues of Spa15 and Spa15:IpgB1 samples, a hundred fold excess of the spin label *S*-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonylthioate (MTSL – Toronto Research Chemicals) dissolved into dimethyl sulfoxide was added to the sample and left overnight. This ensured maximal labelling although only a ten fold excess is required. Excess label was washed out via repeated centrifugal concentrator washing into 25 mM Tris-HCl pH 7.5, 50 mM NaCl using Amicon Ultracel Centrifugal filters (10 kDa cut-off) (Berliner *et al.*, 1982).

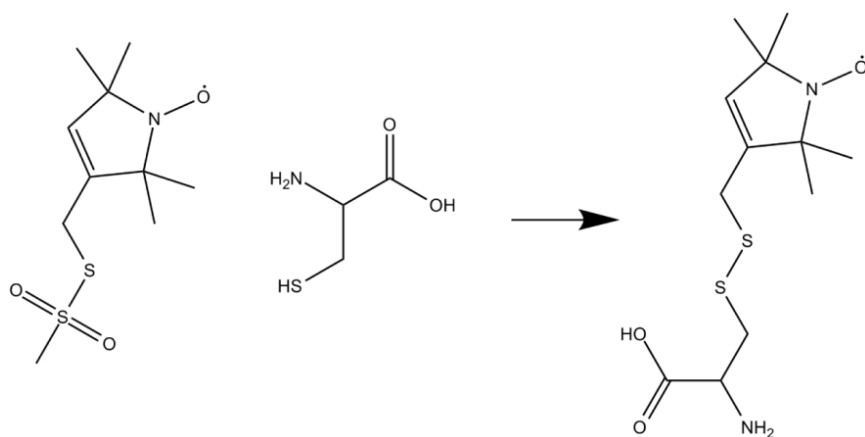


Figure 2-2 Mechanism of cysteine labelling by MTSL. Attachment to the cyteine free thiol via a disulphide bond is accompanied by leaving of the MTSL sulphinic acid leaving group. Successfully labelled proteins show a mass increase of 186 Da.

For all EPR studies the final protein concentrations used were 50 – 600 μM . The sample volumes used were 50 μl . All frozen samples had 30 – 50% glycerol added to reduce relaxation (with the same final concentration). Glycerol has this effect since glass formation is enhanced upon freezing, preventing the formation of short-range molecular clusters (Schiemann and Prisner, 2007).

The effects of such a preparative procedure may be observed by passing the glycerol containing spin labelled sample through a gel filtration column. This will ensure that aggregation and hence spin crowding do not influence the result of the EPR experiment. This step was not performed in this thesis.

2.7.1.3 Room temperature CW-EPR.

CW-EPR was performed using an X-band CW-EPR spectrometer (EMX Bruker Biospin GmbH, Germany) using an X-band Super High Sensitivity Probehead (Bruker).

Sample cuvettes sit in a cavity which enables amplification of a signal. A standing (resonating) microwave is set up in the cavity. The sample is positioned at the minimum of the electric field component - to reduce absorbance, and at the maximum of the magnetic field component of the standing wave – to maximise the sensitivity. The detector measures the radiation reflected back from the microwave cavity (Zhang *et al.*, 2009). The protocol by which the cavity and waveguide may be tuned is given in the Appendix.

The magnetic field is swept using a magnetic field controller which sets the magnetic field values and timings of the sweep. An EPR signal arises when the magnetic field splits the energy levels in the sample such that the sample absorbs the standing microwave set up around it in the cavity. This change leads to loss of coupling of the cavity, resulting in microwaves being reflected back to the microwave bridge and detector.

A common method to increase the EPR signal size is use of phase sensitive detection via a lock-in amplifier (LIA) (Berliner, 2003). The magnetic field strength experienced by the sample is sinusoidally varied at a particular modulation frequency. This has the effect of modulating any resultant EPR signal. The LIA detects only signals oscillating at the same modulating frequency and discards all other signals as artefacts. The amplitude of the modulation must be selected so as to maximise the signal size without washing out small effects or inducing broadening of the spectra. The response time of the spectrometer

is determined by a time constant. This is chosen to balance a satisfactory signal/noise ratio with high spectral resolution.

Data were analysed using the program EasySpin (Stoll and Schweiger, 2006). This is a Matlab toolbox written by Dr Stefan Stoll (Easyspin) which may be used for simple experimental plotting of data, as well as simulation and fitting. An example m.file for plotting the data with respect to both magnetic field and g-value is given in the Appendix.

2.7.1.4 Base temperature CW-EPR.

Whereas nitroxide radicals may be observed by EPR at room temperature, certain paramagnetic metals (including Fe^{3+} - Chapter 6) must be observed under cooler conditions (Ubbink *et al.*, 2002), since their relaxation is too fast otherwise. Prior to spectrometer tuning, the apparatus is cooled to cryogenic temperatures for operation.

Several parts are added to the EPR spectrometer to cater for the low temperature capability. A low temperature helium flow cryostat (Oxford Instruments CF935) encloses the sample. A liquid helium dewar and helium transfer line (Oxford instruments GFS) deliver the cryogen to the cryostat. A vacuum pump (Adixen Packtel 1025) evacuates the outer vacuum chamber. An exhaust gas tube and second pump (GAST 72R555-V37_C222TX) draw the helium through the system. A gas flow controller (Oxford instruments VC 41) and temperature controller (Oxford instruments ITC 503S) maintain regulation. The protocol for cooling the system is given in the Appendix. Once cooled, the cavity and waveguide are critically coupled as at room temperature.

2.7.2 Pulsed EPR and double electron-electron resonance.

Pulsed EPR is the application of pulsed microwave radiation rather than searching for resonance by slowly sweeping the field at fixed microwave frequency (CW-EPR). It has led to the development of a whole armoury of techniques offering insight into the structure and dynamics of many paramagnetic

systems. These include the ability to accurately determine distances between labels attached to interacting proteins over the range 2 – 8 nm, with significant range ability over the analogous CW-EPR experiments ($< 20 \text{ \AA}$) (Bruker, 2011).

The microwave pulses span a finite bandwidth of frequencies which excite the sample at its resonant frequencies at the same time. Figure 2-3 shows the spin states upon application of a Hahn Echo sequence, the common basis for a number of experiments (Berliner, 2003; Lovett (nee Banham) J, 2006; Poole, 1996).

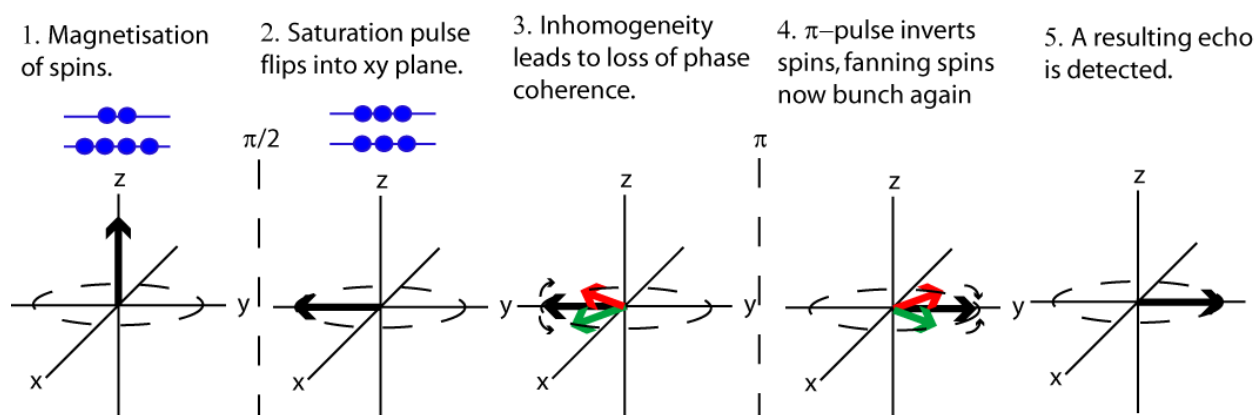


Figure 2-3 The Hahn Echo sequence displayed in the rotating frame. 1. Initially a magnetic field B_0 applied in the laboratory z direction leads all spins to precess around that axis. Magnetisation results for greater population in the parallel rather than anti-parallel direction. 2. A microwave pulse, B_1 , is applied in order to achieve resonance. This is typically achieved with a $\pi/2$ pulse resulting in the equalisation of the populations within the spin state. 3. The spins continue to precess (different spins shown in red and green) gradually losing phase coherence (as precession frequencies differ caused by spin interactions, both coherent and incoherent in nature). 4. As long as these spin interactions have been constant throughout the spin echo experiment the phases of the spins can be refocused by application of a π pulse which serves to reflect the spins on the x-axis. 5. This leads finally to a refocusing on the spin packet creating an observable echo. (Figure reproduced (Ernst *et al.*, 1987)).

Importantly, not all processes affecting the precession frequencies of the spins (Figure 2-3) lead to coherent spin evolution. Incoherent processes will also affect the spins. Spin-lattice relaxation, characterised by a time constant T_1 , will return the system to its initial spin population whilst spin-spin relaxation, characterised by T_2 , lead to loss of spin coherence in the xy plane. Determination of long range distances can be achieved by Double Electron-Electron Resonance (DEER), which exploits the

inhomogeneous refocusing of spins due to dipolar interactions. Milov first carried out DEER as a pulse technique on spin-labelled peptides (Milov *et al.*, 1999).

DEER is a double resonance (or double frequency) technique, where the inhomogeneous refocusing of a group of spins is perturbed by excitation of a second group of spins, via application of a second microwave 'pump' frequency.

The interaction which connects these two sets of spins is the dipolar coupling, which decays with the inverse of the interspin distance. The Hamiltonian which describes the interaction energy between the two spins via the dipolar coupling mechanism may be approximated to first order by equation 2.25; g is the g-value and S_z the spin angular momentum operator for the radicals A and B along the field direction. The spins are separated by the distance R . θ defines the angle between the interspin vector and the applied magnetic field (Schiemann and Prisner, 2007).

$$H_{dip} \approx \frac{g^A g^B \mu_B S_z^A S_z^B (1 - 3 \cos^2 \theta)}{R^3} \quad 2.25.$$

Observation of the dipolar coupling thus allows the distance between two spins to be determined. This may be detected in the EPR lines which are broadened in solids and powders by the dipolar interaction. The anisotropic nature of the dipolar coupling prevents observation for solutions, and hence frozen glasses are required for their experimentation. The necessity for DEER in dipolar coupling determination is in the lack of resolution of line broadening with CW-EPR.

The DEER experiment produces a time-domain spectrum of the dipolar interactions. This may be Fourier transformed into the frequency domain, giving intensity as a function of dipolar interaction (the Pake Pattern). The spectrum will be a doublet in the isolated free radicals case (a Pake Doublet), where the frequency difference between the maxima gives twice the dipolar frequency, a quantity both distance and orientation dependent (Bowman *et al.*, 2004). However, experimental cases are usually not so

simple, and determination of many distance distributions is an ill-posed problem. As a result, dipolar coupling, and hence distance measurement, is usually carried out by Tikhonov regularisation. This works by minimising the deviation between theoretical and experimental form factors, given by the below function:

$$G_{\alpha}(P) = \rho + \alpha\eta \quad 2.26.$$

ρ is the mean square deviation between theoretical and experimental form factors, α is the regularisation parameter calculated via the L-curve method, and η is the roughness of the distribution, for cutting out noise artefacts (Jeschke and Polyhach, 2007).

The actual DEER experiment involves modulation of the echo intensity of one spin population as the second population is inverted by a pump pulse. The timing of three ‘observer’ pulses stays the same (top line in Figure 2-4). However, a pump pulse is varied in time relative to the maximum of an undetected Hahn echo throughout the experiment (bottom line in Figure 2-4). Oscillations in the observed echo will occur at the dipolar frequency. These oscillations are separated out in a ‘form factor’ by subtracting out the decay from intermolecular spin interactions (Jeschke and Polyhach, 2007).

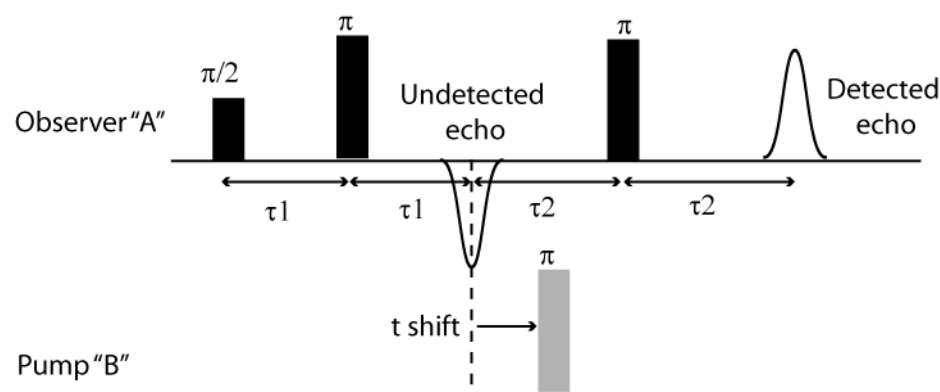


Figure 2-4 DEER pulse sequence. The A spins are saturated with a $\pi/2$ pulse and inhomogeneities in the field lead to dephasing. One of these inhomogeneities is the dipolar coupling. A spins are flipped by a π pulse which leads to an undetected echo. At a time t , the dipolar coupling is inverted by flipping the B spins. This alters the rate of dephasing and so when the A spins are flipped again, the resulting echo varies as a function of both the time of the pump pulse and the dipolar coupling. By altering the time of the pump pulse, the dipolar contribution is obtained.

2.7.2.1 DEER measurements.

Whereas in CW-EPR critical coupling of the waveguide and cavity is required, in pulsed EPR overcoupling is necessary (Bruker_Biospin). This results in some of the microwaves being reflected back, however overcoupling reduces the dead time of the experiment because the long lived standing wave is disrupted. Coupling is also influenced by the shape of the resonator – a split ring example was used in this study to minimise dead time (Jeschke, 2007). Setup for the DEER experiment was carried out as described in the Appendix.

All samples were flash frozen and operation temperatures were cryogenic. This kept the spin labelled molecules immobilised, preventing the averaging out of the anisotropic dipolar coupling between them. Deuteration of the solution matrix has been shown to increase T2 relaxation times (Jeschke *et al.*, 2004), allowing for shorter echo decay and longer distance accessibility, so in this work D₂O buffers and deuterated glycerol were used throughout.

The 50% glycerol flash frozen samples were subjected to the standard four pulse DEER sequence (Figure 2-4) on an X-band spectrometer (Bruker ElexSys E680) using a 3 mm split ring resonator (Bruker EN

4118X-MS3) at 50 K. The resonator was overcoupled and the pump pulse coincided with the centre of the microwave mode of the resonator and the maximum of the nitroxide radical spectrum.

Analysis was carried out using DEERAnalysis2009 (Jeschke *et al.*, 2006). This is a Matlab derived program by Gunnar Jeschke which can obtain distance distributions from DEER data.

2.7.2.2 Multiscale modelling of macromolecular systems.

MMM is a program for visualisation and inspection of proteins (Jeschke and Polyhach, 2009). It allows simulation of spin label conformations which may be used to verify DEER distance measurements, and is used in Chapter 3 to explain the experimental DEER trace of Spa15. The spin labelled PDB obtained from crystallography was uploaded to the multiscale modelling of macromolecular systems 2009 program (Jeschke and Polyhach, 2009). The spin label was removed from the PDB and re-added using the program's labelling toolbutton. This was to allow the conformation calculation to be independent of the experimental result. 50 different possible conformations for the spin label were calculated, with their associated probabilities at the glass transition temperature for water/glycerol of 175 K. DEER distances were calculated based upon the most likely conformations of spin label. Boltzmann scaling of the full density functional theory calculation of the label at 175 K was applied to predict conformations at 50 K and 10 K.

2.8 Surface plasmon resonance.

To a first order approximation, incident light onto a surface of different refractive index reflects when the incident angle is greater than the critical angle. However, this reflection is usually not total, and in particular the mobile electrons of a gold surface can lead to energy absorption. This absorption is maximised at a particular angle; the plasmon resonance angle, θ_{spr} . Here, the wavevector of the incident light is resonant with the wavelengths of the oscillating mobile electrons. These are alternatively called surface plasmons and propagate near to the surface ($\lambda \approx 300$ nm) (Van der Merwe, 2001). The angle at

which this reduction in reflected light occurs is recorded in the sensorgram. Binding of protein near the surface upon which the light impinges changes the refractive index. This alters the momentum of the plasmons. In turn, the angle of photon absorption is changed, since momentum must be conserved upon the transfer of energy from photon to plasmon (Markey, 1999). The binding can be detected of a ligand to a protein which has already been coupled to the surface. The common measurement is that of response units, where 1 RU is equivalent to a shift of 10^{-4} degrees (Van der Merwe, 2001).

Experiments were carried out using a BIAcore™ 2000 System (Biacore AB, Sweden) at 20°C with a general scheme as described in the Appendix. The ligand protein was coupled to a CM5 sensor chip leading to a rise of 500 - 2000 RU using standard amine coupling protocols (BIAcore Amine Coupling Kit BR-1000-50). Hepes-buffered saline containing 5 mM MgCl_2 was flowed throughout the experiments at 50 ml min^{-1} . Samples of the analyte protein were injected in the same buffer for 180 secs duration over the ligand coupled channel and an analyte control channel.

Dissociation was followed for 1000 secs from the end of injection after which regeneration of the surface was carried out. Full retention of activity was checked after regeneration. Experiments (repeated three times) were carried out with (i) varying concentrations of analyte (concentration determined by NanoDrop Spectrophotometer) and with (ii) other conditions as detailed in the results sections. Blank and analyte bound negative control channels were used to identify experimental artefacts and self-association. Data was analysed with BIAevaluation software (GE Healthcare).

3 Structure of Spin labelled Spa15; The Use of EPR to Complement X-ray Crystallography.

The Type Three Secretion System (T3SS) of many Gram-negative bacteria, including *S. flexneri*, is an essential component of bacterial virulence which was introduced in Chapter 1. In contrast to the Tat pathway, where fully folded proteins with intact cofactor are passed, T3SS effector proteins are transported in an unfolded state through a 25 Å channel connecting bacterial to host cytoplasm (Blocker *et al.*, 2001; Johnson *et al.*, 2005b). Before transport across the T3SS, in the bacterial cytoplasm, many effectors require a chaperone. A chaperone is needed to prevent any premature or incorrect interactions, to hold the effector in a state competent for secretion, and also to aid secretion itself (Page and Parsot, 2002; Parsot *et al.*, 2003).

This chapter examines a T3SS chaperone - Spa15 of *S. flexneri* - and also presents a case of how EPR may be of use in proteins where no paramagnetic metals naturally occur, particularly when a crystal structure proves unobtainable. Attachment of the MTSL spin label introduces a paramagnetic moiety to Spa15 enabling structural information to be determined. In this study, Spa15 has been probed by the distance obtaining experiment DEER. Interpretation of the DEER data was sought through Multiscale Modelling of Macromolecular Systems (MMM), the simulation program developed to provide such structural interpretation. MMM could not generate the correct DEER distance from the MTSL orientations which it predicted.

Observation of the MTSL orientation was thus achieved by determining the spin labelled Spa15 crystal structure (Figure 3-1). This demonstrated that the distances between labels measured from X-ray crystallography and from EPR were in agreement. It also confirmed that the MMM method produced an MTSL orientation which is not experimentally observed. An agreement between the crystallography and

DEER methods confirmed the crystallographic dimer to be a relevant solution state, and provides a demonstration of DEER's accuracy in distance determination. A complex of Spa15 with one of its effectors, IpgB1, has also been co-purified and spin labelled. The DEER traces obtained have enabled the impact of effector binding on Spa15 to be observed. The conclusions are that IpgB1 binding has no global structural effect on Spa15, and that a pocket which is utilised by the Spa15 *Salmonella* spp. homologue InvB to bind its effectors is not used by Spa15 in this complex.

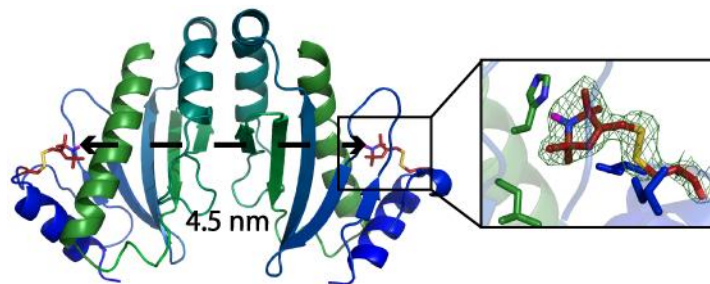


Figure 3-1 Spin labelled Spa15. The DEER distance between spin labels in homodimeric Spa15 is verified by the crystal structure. The crystal structure was solved in response to the lack of prediction of this 4.5 nm distance by the modelling program MMM.

This chapter will be structured as follows: An introduction to Spa15 is given, and its role as chaperone for IpgB1 explained. Next, details of the Spa15 spin labelling and the results of the DEER experiment yielding the distance between the attached spin labels are described. The MMM results that show discrepancy with the DEER results are shown, before moving to the crystallographic solution of spin labelled Spa15. This corroborated the experimental DEER results. Finally, the DEER data of the Spa15:IpgB1 complex are presented. EPR work and MMM modelling were performed with Dr Janet Lovett, and crystallography with Dr Steven Johnson, Dr Pietro Roversi and Professor Susan Lea.

3.1 Spa15 – a promiscuous chaperone.

Spa15 belongs to the promiscuous subset of the effector chaperone class IB, which was introduced in Chapter 1. It chaperones several effectors of unrelated sequence including IpaA, IpgB1, IpgB2, OspB, OspC2, OspC3 and OspD1 (Page *et al.*, 2002; Schmitz *et al.*, 2009). The structure of Spa15 has been previously reported (PDB 1RY9) (Van Eerde *et al.*, 2004). The fold is typical of a Class I chaperone, existing as a homodimer of the 15 kDa moieties. Spa15 shares striking similarity with the *Salmonella* spp. homologue InvB (Lilic *et al.*, 2006) with whom it shares a sequence similarity of 58% and with which comparisons are drawn at the end of the chapter.

One of the effectors chaperoned by Spa15, IpgB1, was selected to observe the consequence of effector binding on Spa15 conformation. IpgB1 is a member of a bacterial guanine exchange factor protein family defined by a conserved WxxxE motif which will be fully discussed in the next chapter (Alto *et al.*, 2006).

3.1.1 Expression and purification of Spa15 and Spa15:IpgB1.

Spa15 from construct 3.1 as defined in Table 2.1 was purified with an N-terminal hexahistidine tag via an Ni-NTA column followed by size exclusion chromatography (Table 2.6). Protein purity was checked by SDS-PAGE (Figure 3-2). Mass spectrometry confirmed the molecular weight of the tagged protein to be 17150 Da. This agrees with the full length construct minus the N-terminal methionine, giving a predicted mass of 17149 Da.

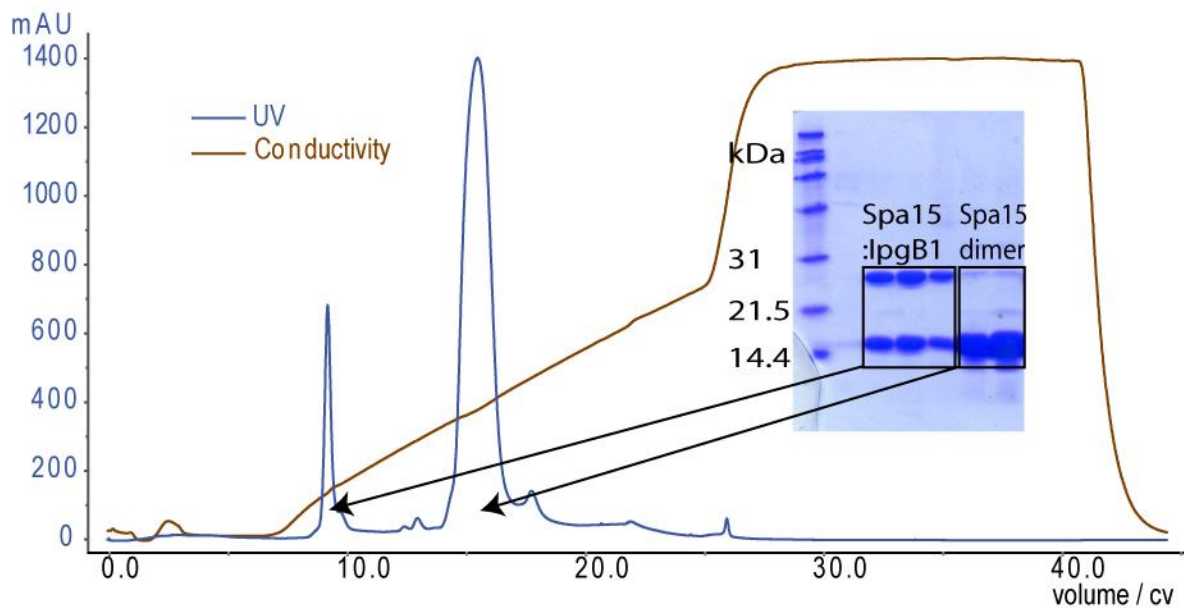
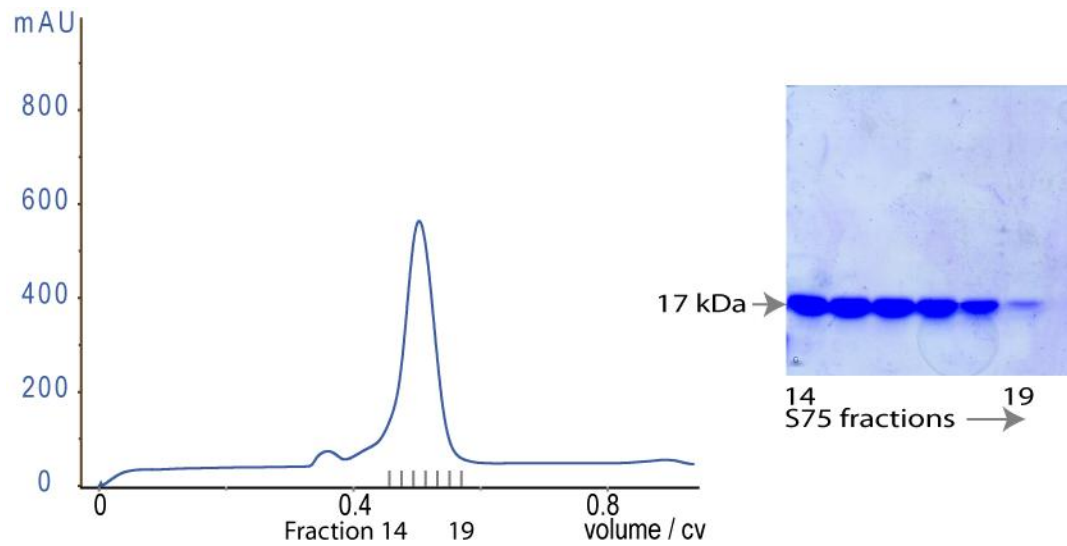


Figure 3-2 Spa15 and Spa15:lpgB1 purification profiles. Top: Spa15 purification via S75 16/60 and resulting SDS-PAGE. **Bottom:** Spa15:lpgB1 purification via MonoQ and resulting SDS-PAGE.

Coexpression of a soluble Spa15:lpgB1 complex was achieved by transformation of the N-terminal His-tagged Spa15 construct 3.1 and untagged lpgB1 construct 4.11 into B834(DE3) cells. As described in Table 2.6 purification was completed via ion exchange chromatography (MonoQ) which separated a single Spa15:lpgB1 complex peak away from a Spa15 dimer peak. As for Spa15, SDS-PAGE and mass

spectrometry were used to check the purity and integrity of the sample (Figure 3-2 and Figure 3-3). The mass spectrometry result of 23666 Da was 3 Da from the 23669 Da expected from IpgB1 containing an N-formylmethionine, which remained uncleaved by a formylase or methionine aminopeptidase (Sherman *et al.*, 1985).

3.1.2 Spin labelling of Spa15 and Spa15:IpgB1.

Spa15 contains a single cysteine residue at position 19 (highlighted in Figure 3-11), and IpgB1 two, Cys49 and Cys75. MTSL spin label was added to both purified Spa15 dimer and to Spa15:IpgB1 by the protocol described in Section 2.7.1.2. Mass spectrometry confirmed the 100% labelling of the single Cys19 of Spa15, both with IpgB1 bound or unbound, and the 0% labelling of IpgB1 showing that its cysteines are not surface exposed when complexed with Spa15 (Figure 3-3).

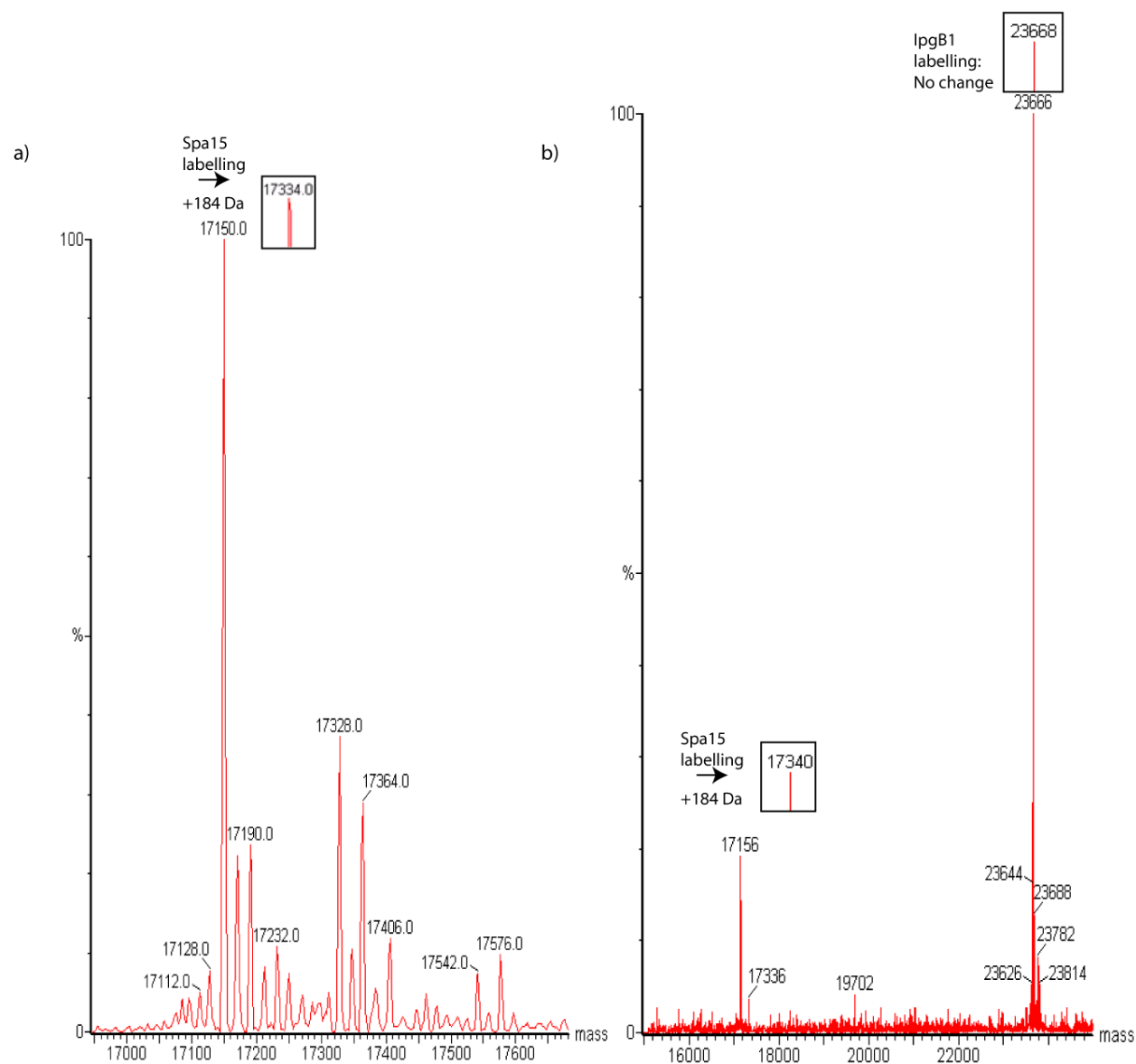


Figure 3-3 Mass spectrometry showing successful labelling of Spa15 cysteines in a) Spa15 dimer and b) when Spa15 complexed with IpgB1. IpgB1 cysteines remain unlabelled however. Spin labelling indicated by addition in mass of ca. 186 Da.

3.2 Continuous Wave EPR studies.

CW-EPR was carried out on the MTSL labelled Spa15 sample. This allowed an initial assessment of the freedom of the Spa15 label environment whilst a quantitative analysis against the standard Tempol (4-Hydroxy-2,2,6,6-tetramethylpiperidine 1-oxyl, - Sigma) was also used to verify the 100% labelling. The EPR spectral parameters used are given in Table 3.1.

Parameter	Value
Centre of field / G	3340
Width of Field / G	150
Time of scan / s	60
g value	2
Gain / dB	60
Modulation amplitude / G	0.5
Modulation frequency / kHz	100
Number of scans	3
Offset / %	0
Attenuation / dB	25
Microwave power / mW	0.63
Microwave frequency / GHz	9.374
Temperature / K	293

Table 3.1 EPR spectral parameters used for Spa15 and Tempol nitroxide radical detection.

Tempol (chemical structure shown in Figure 3-4) is a common quantitative standard chemical used in nitroxide EPR. Its single unpaired electron ($S = \frac{1}{2}$) experiences a hyperfine interaction with the nearby ^{14}N nucleus (with spin quantum number $I = 1$). The ^{14}N spin is equally likely to be found in the $M_I = -1, 0$ or $+1$ states. There is no magnetic moment associated with the $M_I = 0$ state. However, when the ^{14}N nucleus is in either $M_I = +1$ or -1 , magnetic fields of equal magnitude but opposite sign are induced at the site of the electron spin, leading to the splitting pattern depicted in Figure 3-4. As the selection rule for EPR is $\Delta M_S = \pm 1$, a characteristic triplet nitroxide spectrum is seen. This can be an oversimplified view however; both hyperfine and Zeeman (by virtue of the g-tensor) contributions have anisotropic components which will not average out if the rotational correlation time is slow. Thus, if the combined contributions from molecular tumbling, local backbone dynamics and internal label motion lead to rotational correlation times exceeding a few nanoseconds, there is incomplete averaging of the anisotropic components at X-band frequencies (9 GHz). The result is an EPR spectrum which is not dominated solely by the isotropic interactions but which includes anisotropic interactions. This is experimentally viewed as a broadening of the EPR spectrum.

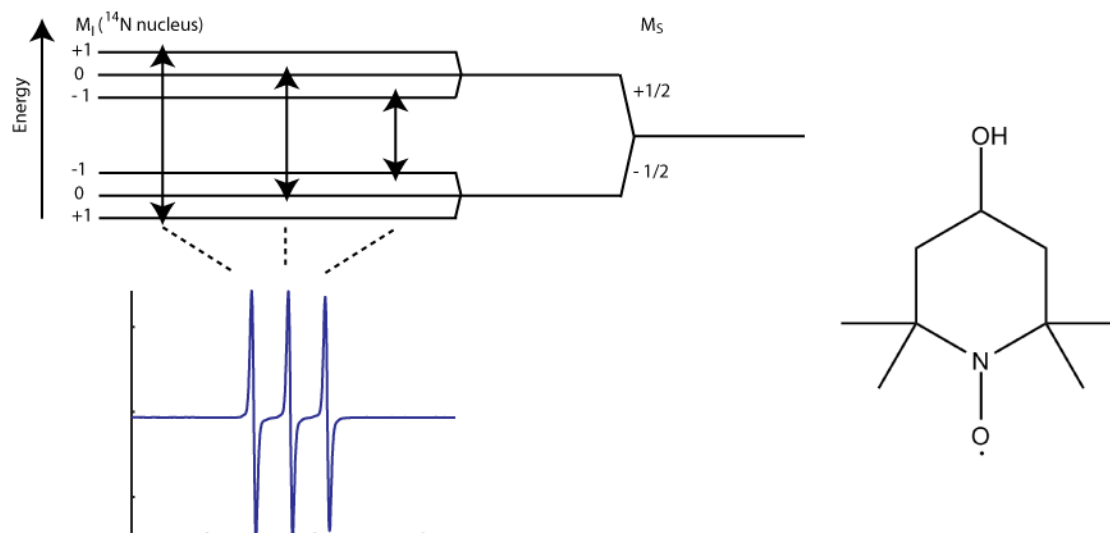


Figure 3-4 Characteristic EPR spectrum for a freely rotating nitroxide radical. This spectrum would be expected for Tempol (pictured), and results from the three permitted transitions described in the text. (Figure modified (Sowa and Qin, 2008)).

3.2.1 Quantitative calibration.

The EPR spectrum of a 230 μM sample of Tempol was shown to have the same number of spins as a 230 μM sample of Spa15 under the same conditions, indicating that each Spa15 in the population was spin labelled. The top panel of Figure 3-5 gives the EPR spectrum of Tempol. This is the derivative spectrum due to the applied modulation field, but may be integrated to give the 'Integral' column. The number of spins is found by taking the area under the Integral spectra, giving the 3rd column - 'Double Integral'. This was repeated for spin-labelled Spa15 in the middle row of Figure 3-5. The bottom row shows the overlay of the Tempol and Spa15 results. In this row the raw data look quite different in magnitude, although integrating this twice to give the number of spins gives the result that Spa15 is 100% labelled. This indicates a substantial broadening of the spin labels of Spa15 compared to the free spin labels of Tempol, which is indicative of motional constraint as described in Section 3.2. It became apparent in the crystallography of spin labelled Spa15, shown below, that this is certainly the case.

The broadening of the Spa15 spectrum is clearly observed within the plot by comparison of the broad triplet peaks with a set of sharp peaks superposed in the same figure. The sharp peaks are coincident

with those of the the (freely tumbling) Tempol EPR spectrum, as shown by dotted lines in Figure 3-5. They are due to a small amount of residual free label which was not successfully washed out from the sample. The broad peaks are due to the spin labels attached to the Spa15.

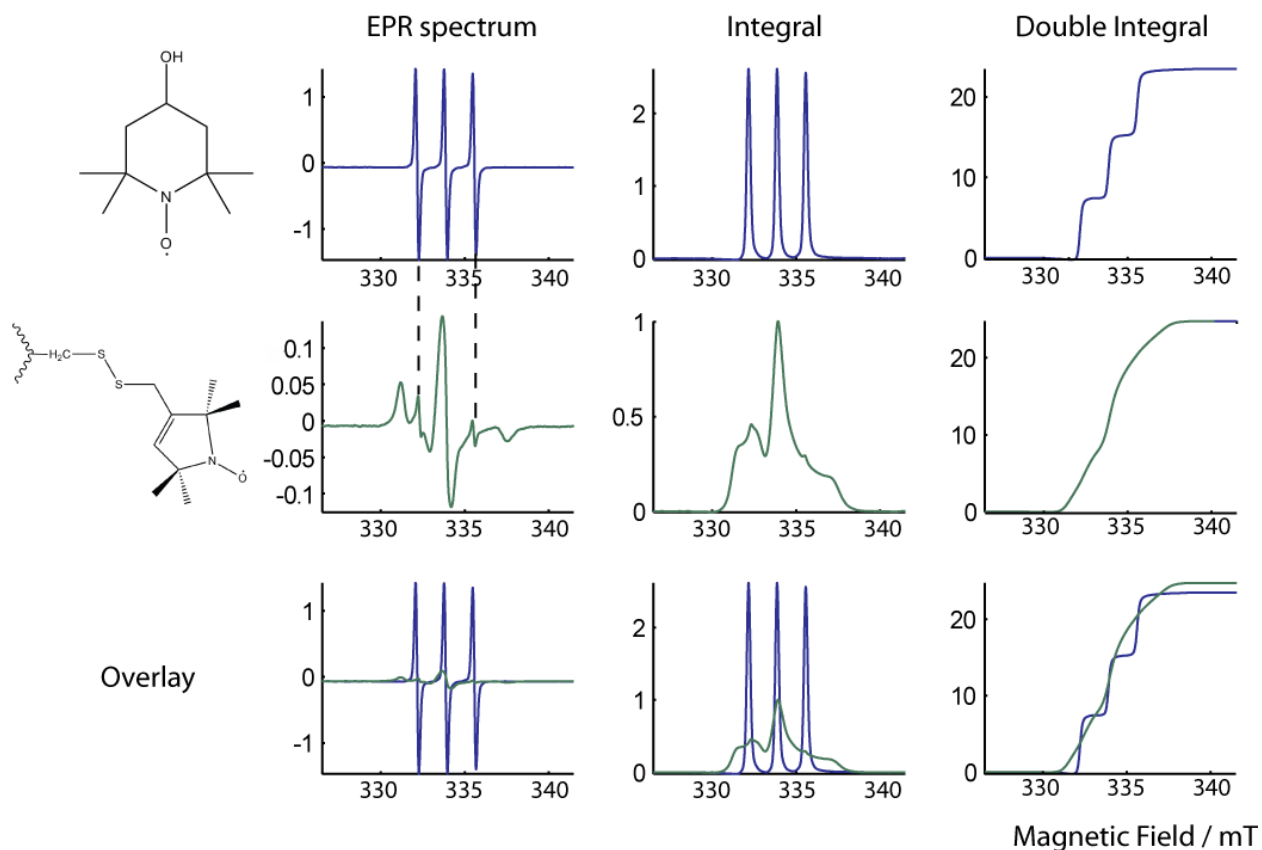


Figure 3-5 EPR verification of 100% Spa15 labelling. The number of spins in a 230 μ M Tempol solution (1st row) and a 230 μ M spin labelled Spa15 solution (2nd row) are compared and found to be equal (3rd row). Each successive column shows a further integral, starting with the 1st derivative raw data, and finishing with the spin count in the third column.

3.3 Double Electron-Electron Resonance studies.

With the result that the Spa15 Cys19 residues were spin labelled and most likely motionally constrained, DEER was used to determine the distance distributions of the spin labels in dimeric Spa15 in solution.

3.3.1 Observer and pump spins.

As discussed in Chapter 2, in the DEER experiment the two-pulse Hahn echo sequence leads to detection of an echo intensity of the observation spin species from an observation-pump pair. The refocusing of the observation spins is perturbed by flipping the pump spins at varying times.

In this study, both observer and pump spins are (chemically) identical MTSL nitroxide species, but are separated into “observation” and “pump” spins by their differing position in the EPR spectrum. It was shown above that a nitroxide EPR spectrum is a triplet due to hyperfine coupling of ^{14}N ($I=1$). This hyperfine coupling is anisotropic, 100 MHz if the molecular axis is parallel to the external magnetic field and 10 MHz if perpendicular (Schiemann and Prisner, 2007). For the parallel spins, this is broad enough to allow the necessary observation and pumping of different spin (M_I) states – DEER would not work if the same spins were observed and pumped. The observer and pump spins are usually defined as low field wing and centre field as shown in Figure 3-6. The excitation bandwidth must be sufficiently thin to avoid excitation overlap, but the pump should be as wide as possible to excite sufficient spins to achieve a deep modulation.

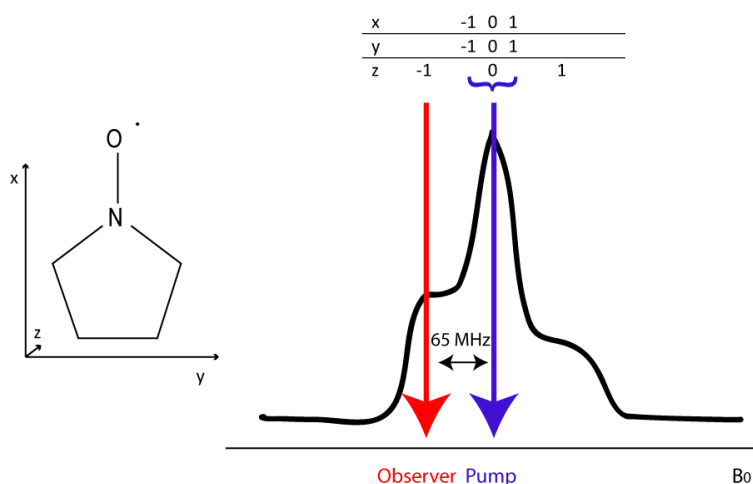


Figure 3-6 Separation of chemically identical nitroxide radicals into different species for DEER detection. The observer spins are those with $M_I = -1$ which lie parallel (z) to the applied field B_0 . Pump spins are $M_I = -1(xy), 0(xyz), 1(xy)$. That is, pump spins are both those with $M_I = 0$ which lie parallel (z) to the field and all states ($M_I = -1, 0, 1$) perpendicular (x, y) to B_0 . (Figure modified (Schiemann and Prisner, 2007)).

3.3.2 DEER measurements.

The 50% glycerol flash frozen sample was subjected to the standard four pulse DEER sequence $\pi/2(\text{obs})-t_1-\pi(\text{obs})-t-\pi(\text{pump})-(t_1+t_2-t)-\pi(\text{obs})-t_2$ -echo (see Figure 2-4) with 32-ns pulse lengths for observer pulses and 12-ns duration for pump pulses, where t was incremented by 8-ns each time. $t_1 = 400$ -ns, $t_2 = 4$ - μ s. t_1 was varied eight times by 56-ns to average out any deuterium esem modulations and the initial $\pi/2$ observation pulse was phase cycled (Lillington *et al.*, 2011).

The experiment was repeated three times at 200 μ M with independent samples, taken from different protein purifications on different days. DEER experiments at 50 K reproducibly identified the major distance between Spa15 labels as 4.5 nm (± 0.1 nm) (Figure 3-7a, b). A fast dephasing T_2 relaxation required overnight measurement despite the use of deuterated solvent. A reduction of temperature below 50 K in an attempt to lengthen T_2 led to saturation of signal due to long T_1 relaxation. Since there was just one labelled site per monomer, the dominance of the 4.5 nm peak indicated it as the intra-dimer label distance.

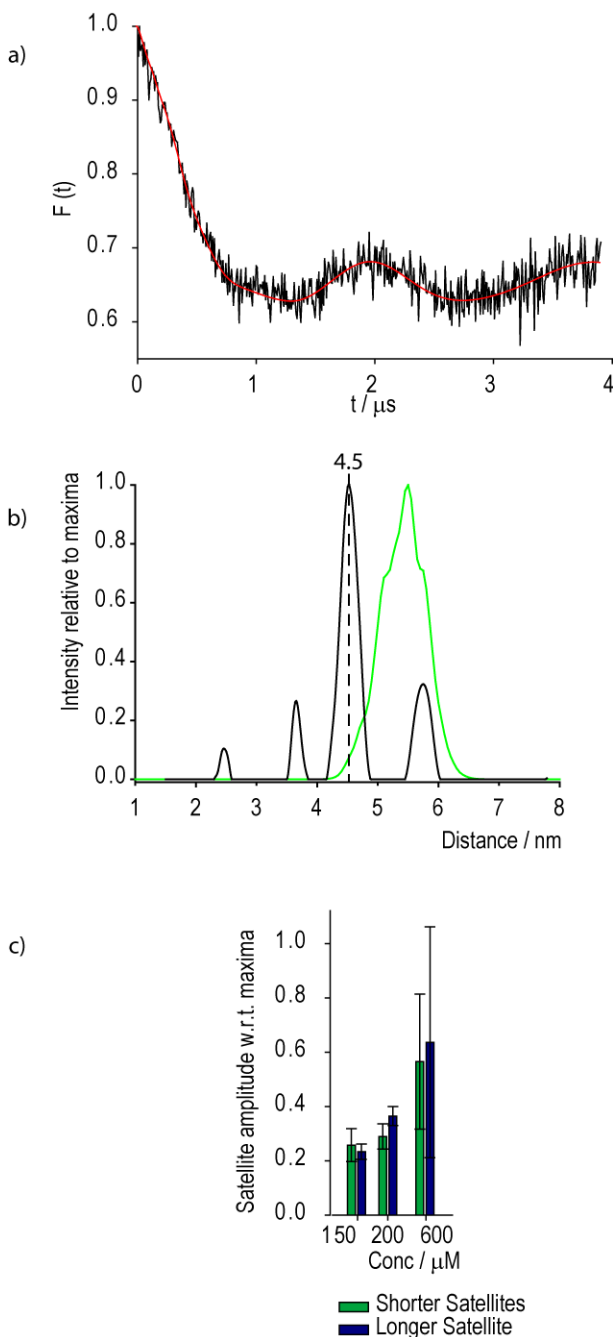


Figure 3-7 Spa15 experimental DEER data. **a)** Form factor (as fitted by DeerAnalysis (Red)), the dipolar evolution function after background correction for the 200 μ M sample. **b)** Tikhonov regularisation of the form factor (regularisation parameter = 1), primary distance of 4.5 nm seen with additional minor distance components. Shown in green is the 5.3 nm DEER distance for the MMM most probable conformation (see Section 3.3.5). **c)** The minor distance components changed in amplitude with the concentration of sample relative to the major 4.5 nm peak. This was determined by integration of the peaks at 3.8 nm and 5.6 nm.

3.3.3 Satellite peaks.

A feature of the DEER spectrum was the inclusion of distances at either side of the 4.5 nm peak, which changed with sample and with concentration (Figure 3-7c). Other researchers have observed satellite signals to be a result of the orientational constraining of a label. This situation may be identified by a variant DEER trace in an experiment changing the frequency difference between observer and pump frequency pulses (Margraf *et al.*, 2007). This possibility was ruled out since such DEER experiments yielded identical signals between 40-80 MHz (Figure 3-8).

The increase in side peaks relative to the major intra-dimer peak at higher concentration led us to conclude that these side peaks were due to aggregation, a product of the preparative conditions of the EPR experiment. This was supported by MALLS of a 200 μ M Spa15 sample, which indicated that at the concentration level used in EPR, protein aggregates began to be evident (Figure 3-8). Preparation of the DEER sample required exposure to organic solvent, repeated concentration and dilution, addition of glycerol and flash freezing. On occasions, precipitate was observable during sample preparation.

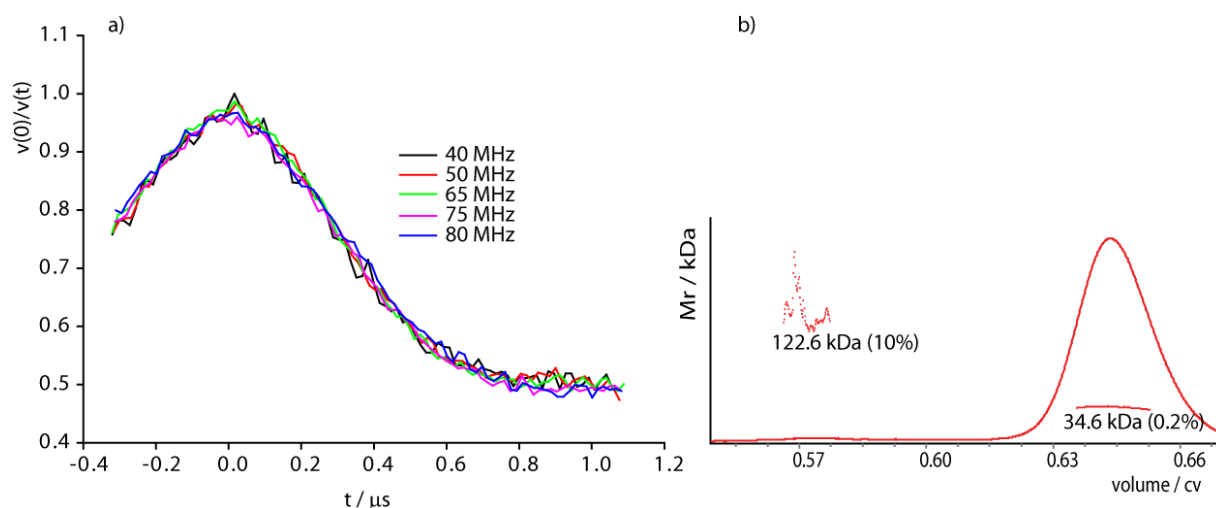


Figure 3-8 Probing the origin of the DEER satellite peaks. **a)** Time trace of Spa15 using DEER with variable frequency difference between observer and pump frequencies. No change in spectra were observed. **b)** SEC-MALLS of Spa15. Flow rate 0.4 ml/min. A limited quantity of high weight aggregate was observed.

3.3.4 Rationalisation of DEER distances with spin label conformation.

Insertion of MTSL into the protein will affect the conformation of the label via interactions with the surrounding protein, altering it from the unbound state (Zielke *et al.*, 2008). In order to correctly interpret EPR data, the key interactions of MTSL which lead to its preferred conformations need to be understood, since a differing conformation can lead to a substantially different DEER distance. A number of researchers have approached this from both an experimental angle and a theoretical point of view:

Experimental efforts include examinations of how EPR spectra vary according to the flexibility of the MTSL binding site, as well as the impact of the exact chemical composition of the spin label (Mchaourab *et al.*, 1999; Mchaourab *et al.*, 1996). In other studies, spin labelled lysozyme was crystallised and the observed MTSL constraint compared to that seen in the breadth of the EPR spectra (Guo *et al.*, 2007; Langen *et al.*, 2000). A relaxed conformation of MTSL was identified, denoted $\chi_4\chi_5$ (angles defined in Figure 3-12b), where the nitroxide is unperturbed by interactions from other residues (Langen *et al.*, 2000). Where there are few/weak interactions, the dominant contributor to conformation is a weak pseudo-hydrogen bond interaction between the S_D disulphide with C_α-H on the bound cysteine. χ_4 and χ_5 angles remain flexible. This 'relaxed state' is found in lysozyme α -helical residues 72 and 80, making it ideal for 'simplest case' simulations upon which rotary diffusion, backbone fluctuations and/or side chain interactions can be added as a perturbation (Columbus *et al.*, 2001).

More strained conformations are seen in α -helices when the *i*+ 4th residue is polar (Guo *et al.*, 2007), however the greatest interaction of residues with the label are those where MTSL is found in a crystal contact. These conformations have angles furthest from the $\chi_4\chi_5$ label, stabilised by a number of generally hydrophobic, inter- and intramolecular interactions.

Theoretical studies aimed at explaining spin label structure have continued alongside experimental investigations. Molecular dynamics has successfully modelled EPR spectra on α -helices of lysozyme

(Sezer *et al.*, 2009) using forcefield parameters (Sezer *et al.*, 2008a), stochastic trajectories (Sezer *et al.*, 2008b) and differing solvent distributions (Stoica, 2004). Other α -helical based studies have been *ab initio*, looking at the torsional profiles of the linkers (Tombolato *et al.*, 2006), or semi empirical (Columbus and Hubbell, 2002) / stochastic dynamics simulations (Steinhoff and Hubbell, 1996). These studies make it clear that calculating the conformation of the simplest scenario – the rigid α -helix with no interactions to neighbouring residues – is a complex process, albeit one to which a high degree of success has been attained.

There are therefore extensive studies upon labels in α -helical environments – where the label may be freely flexible under the $\chi^4\chi^5$ model, or interact with the $i + 4$ th residue. Other than for crystal contact examples, where the solution and crystal state have quite different interactions, the non-helical spin label has been understudied.

An exception to this is a study of the *E. coli* periplasmic translocation channel protein Wza (Hagelueken *et al.*, 2009). Two spin label conformations were seen in both PELDOR and in a spin labelled crystal structure in a loop of the protein. This shows the complementarity of the techniques for a non-constrained protein label.

In this study DEER showed one intramolecular distance between spin labels of Spa15. Interpretation of this was sought from the only freely available DEER simulation program, MMM, which predicts spin label orientations and the resultant DEER spectra (Jeschke and Polyhach, 2009).

3.3.5 Multiscale modelling of macromolecular systems.

Initially, The unlabelled Spa15 PDB 1RY9 was uploaded to the MMM 2009 program, and the spin label added using the program's labelling toolbutton. 50 different possible conformations for the spin label with their associated probabilities were calculated at the glass transition temperature for water/glycerol of 175 K. DEER distances were calculated based upon the most likely conformation of spin label.

Interestingly, the most likely conformation in this example would lead to a distance of 5.4 nm (with standard deviation 0.54 nm and relative width 9.9%), not the 4.5 nm as observed experimentally (Figure 3-7 b). The disagreement of MMM with the DEER distances therefore required a crystallographic solution to explain the spin label conformation which provides the experimental DEER distance.

3.4 Spin-labelled Spa15 X-ray crystal structure.

Crystallisation trials were set up with the JSCG and MD 1&2 broad screens at 21°C. Initial hits at the condition 100 mM MES pH 5.5, 100 mM guanidium chloride, 50 mM ammonium sulphate, 5% ethylene glycol were optimised according to the grid in Figure 3-9. Specifically, the pH and ethylene glycol content were systematically varied, at two salt concentrations, to obtain the largest crystals possible. The 3 mg/ml protein was plated in 50:50 ratio drops with mother liquor and yielded 75 µm disc like crystals. These were cryoprotected in 15% glycerol and X-ray diffracted at beamline I02, Diamond Light Source, Oxfordshire, UK. Other crystal conditions, 30% PEG 8000, 0.2 M ammonium sulphate and 20% PEG 8000, 0.05 M potassium phosphate led to crystals of lower diffraction quality.

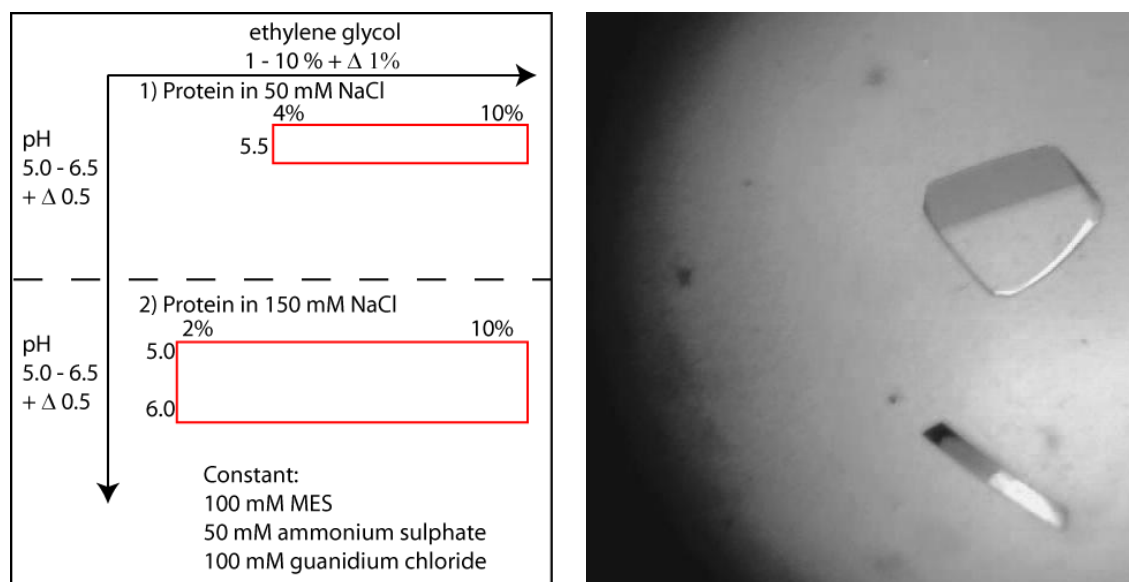


Figure 3-9 Spa15 crystal optimisation and crystal morphology. Left: Optimisation grid varying pH (5.0 - 6.5) against ethylene glycol concentration (1 - 10 %) at two different NaCl concentrations (50 mM and 150 mM). Right: Optimised Spa15 crystals. Approximate dimensions 75 µm x 75 µm x 10 µm.

Data were collected and processed using CCP4 programs; indexing, strategising and integrating via iMosflm (Leslie, 1992) and scaling with Scala to resolution 2.3 Å (Evans, 1997). Data collection strategy and statistics are given in Table 3.2.

Data Collection Strategy:	
Initial images	3 x images: 25°, 70°, 115°
Initial estimated mosaicity	0.38
Advised strategy	25° – 170° (110°)
Strategy completeness	99.7% (unique) 99.9% (anomalous)
Actual collection	25° - 205° (180°)
$\delta\phi$	1°
Exposure level and time	100%, 1 sec
Data Collection Summary:	
Space group	P2 ₁
Molecules in asymmetric unit	2
Unit-cell parameters (Å, °)	a = 56.1, b = 50.4, c = 57.8 $\alpha = \gamma = 90.0$, $\beta = 117.2$
Resolution range (Å)	51.4 – 2.3 (2.4 – 2.3)
Number of Unique reflections	12955 (1848)
Number of Reflections (R-Free)	650 (51)
Solvent content (%)	41
Rmerge (%)	11 (31)
Rpim (%)	10 (28)
Mean I/ σ (I)	5.2 (3.0)
Completeness (%)	99 (99)
Redundancy	3.5 (3.6)
B(Isotropic) From Wilson Plot	17.9

Table 3.2 Data collection strategy and statistics. Brackets denote outer shell statistics. R factors are defined in Table 2.10.

The asymmetric unit was in agreement with the solution state dimer, with a Matthews's coefficient 2.11 for the case of two monomers in the asymmetric unit, with probability 0.96. The self rotation function search returned two-fold rotational non crystallographic symmetry (NCS), with score 19.66 σ (relative to the mean), which was similar in magnitude to the identity operation score (24.4 σ) and significantly higher than the next best peak score (3.73 σ). Figure 3-10 shows the self rotation function for the $\kappa = 180^\circ$ section, as output by MolRep. This was the only κ section to display observable peaks, showing the two fold relation. The peaks at $\phi = 90^\circ$ and -90° are a result of the space group symmetry (Sawaya, 2007), and the two fold NCS peak is also duplicated by this crystallographic symmetry operation.

The two NCS copies are related by rotation (ϕ, ψ, κ) 59.92° , 179.79° , 179.59° and translation (x, y, z) $29.62 \text{ \AA}$, $48.70 \text{ \AA}$, $50.55 \text{ \AA}$. This rotation function was used in the structural solution by molecular replacement, using MolRep (Vagin and Teplyakov, 1997) with search model PDB 1RY9, the Spa15 structure (Van Eerde *et al.*, 2004). The output solution had score 0.656 and contrast 21.36, unsurprisingly a very large number given the use of the same protein as a search model.

RF(theta,phi,chi)_max : 1234. rms : 50.59
 Rad : 25.19 Resmax : 2.30

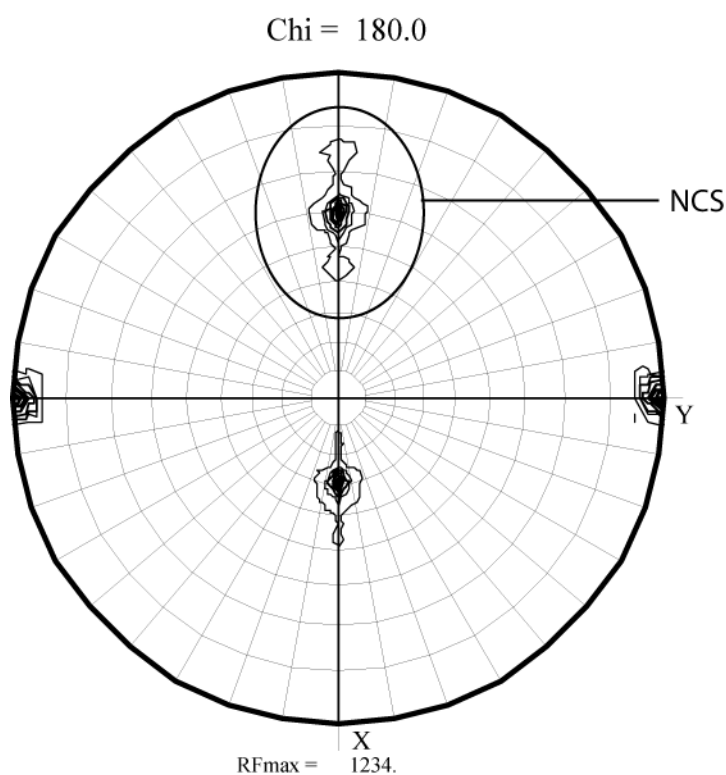


Figure 3-10 Spa15 self rotation function. Calculated by MolRep. A single two-fold NCS peak is shown.

Iterative refinement (Refmac (Murshudov *et al.*, 1997)) and model building (Coot (Emsley and Cowtan, 2004)) to resolution $2.3 \text{ \AA}$ produced the final structure with R and R_{free} 22.4% and 26.4% respectively. The spin label residue R1A (PDB and dictionary file) was obtained from the Hetero-compound Information Centre – Uppsala (HICUP) (Kleywegd).

The quality of the model was checked using Molprobit (Davis *et al.*, 2007), with final Molprobit score 1.08 (100th percentile). Table 3.3 gives the refinement statistics.

Refinement summary:	
Resolution (Å)	51.4 – 2.3 (2.33)
R factor (%)	22.4 (39)
Free R factor (%)	26.4 (37)
R.m.s.d. bond lengths (Å)	0.008
R.m.s.d. bond angles (°)	1.285
Protein atoms	2062
Solvent atoms	80
Ramachandran plot:	
Preferred (%)	99.2
Allowed (%)	0.8
Outliers (%)	0.0
Clashscore	2.95
PDB code	2XGA

Table 3.3 Spa15 refinement statistics. Brackets denote outer shell statistics. R-factors are as defined in eqn 2.23. Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.

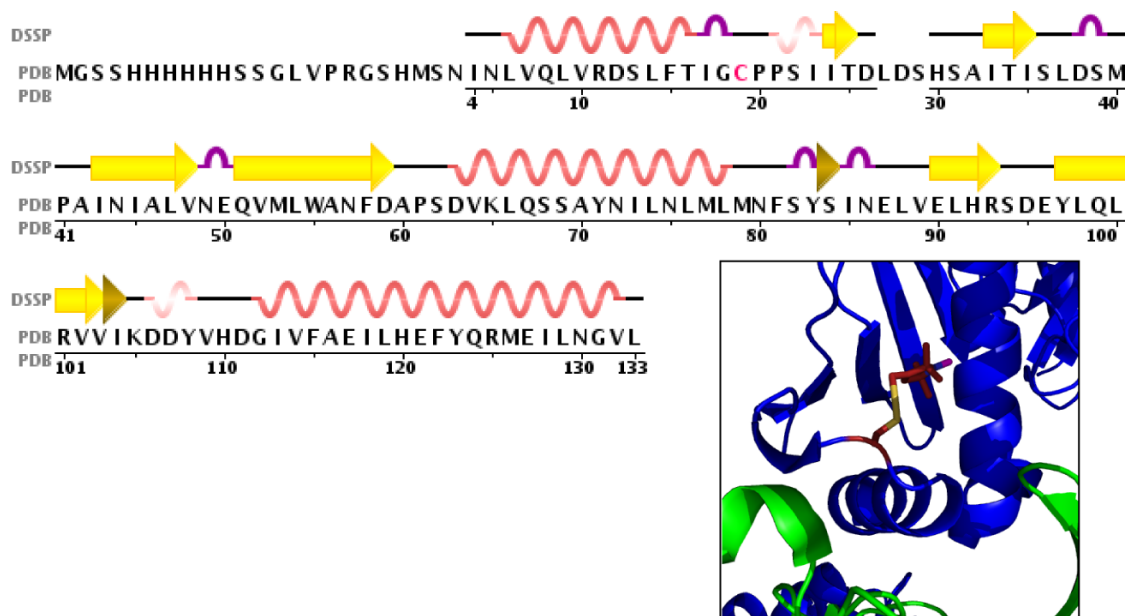


Figure 3-11 Primary sequence and secondary structure of PDB 2XGA. Calculated with DSSP (Kabsch and Sander, 1983) from PDBsum. The MTSL label at position 19 (highlighted) was shown not to interact with neighbours in the crystal packing (inset).

Spa15 has three α -helices and a twisted six stranded β -sheet, interacting with a partner via a right angled α to α -helix hydrophobic interface, involving Leu74 and Leu78 with Ile73, Leu91 and Leu98. An

extended solvent network forms at the dimer interface. The hexahistidine tag and first four N-terminal residues, as well as the loop encompassing residues 27-29 were seen to be disordered, as shown in Figure 3-11. The single spin labelled cysteine residue at position 19 did not interact with other Spa15 neighbours in the crystal (Figure 3-11). It could be modelled as a single conformation with B-factors close to that of the main chain (average B factor for residue R1A = $40.6 \text{ \AA}^2$, whole chain = $28.4 \text{ \AA}^2$, loop on which R1A resides (but excluding R1A) = $50.2 \text{ \AA}^2$) as shown in Figure 3-12b. The two equivalent residues are positioned at the very widest part of the heart shaped dimer relative to one another. Due to favourable energetics the multi-methyl substituted pyrrol ring lies solely within a largely hydrophobic pocket, containing the residues Ile17, Ile23, Ile35, Ile43, Leu14, Leu37 and Leu119. The closest interactions lie from the MTSL C₆ to C_{D1} of Ile17, and from the MTSL O to N_{D1} of His120. The addition of MTSL has no significant effect upon the backbone structure on the loop around residue 19 relative to the unlabelled Spa15 PDB 1RY9 (Van Eerde et al., 2004), nor upon the residues with which contacts are made, with the exception of Ile17 (Figure 3-12c). In contrast with a number of the reported lysozyme structures, no polar interactions from MTSL S_G or S_D are apparent with the main chain (Langen *et al.*, 2000).

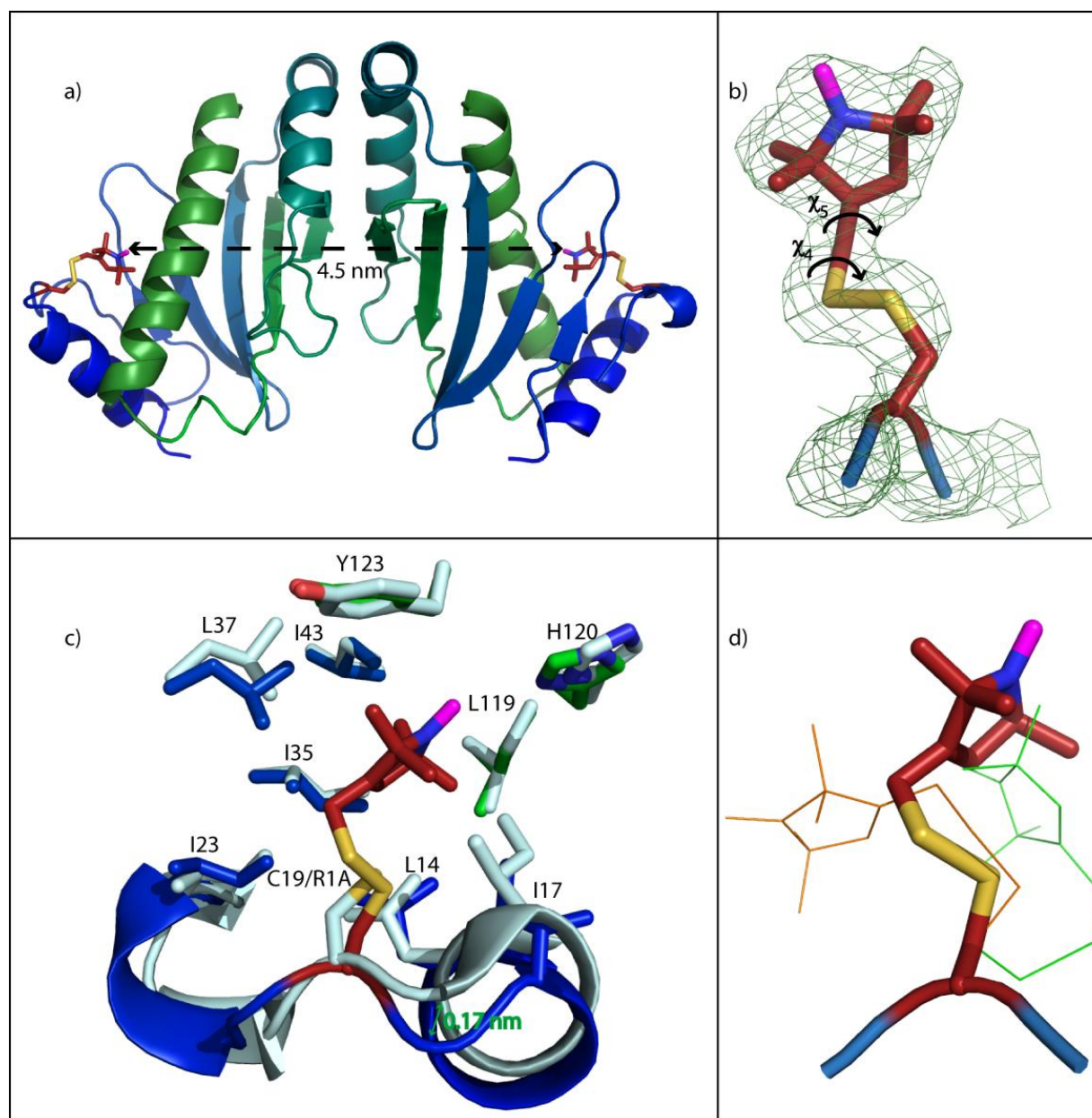


Figure 3-12 MTSL spin label observed in the Spa15 crystal structure. **a)** Ribbon diagram of spin-labelled Spa15 dimer crystal structure, showing the two Spa15 subunits coloured blue to green from N to C terminus respectively. The MTSL spin label is atom coloured, highlighting the S-S bond, with the blue nitrogen bound to purple oxygen. The distance between the two spin labels is shown. **b)** The spin label of Cys19 was highly ordered in the electron density at 1σ background. The label is shown to have only one conformation, due to particular stabilisation by the MTSL C₆ to C_{D1} of Ile17, from the MTSL O to N_{D1} and due to the steric restraints of the pocket. **c)** The hydrophobic pocket within which the spin label (red) resides. Neighbouring Leu14, Ile17, Ile23, Ile35, Leu37, Ile43, Leu119, His120 and Tyr123 are shown. There is little difference between backbone structure for unlabelled (PDB 1RY9) (light blue) and labelled (colour scheme as for a), dark blue – green) structures, with perhaps a subtle closing around the label in the latter. Side chains are similarly unaffected by the spin label, an exception being Ile17 which is placed further into the hydrophobic pocket in 1RY9 than in the labelled 2XGA structure. **d)** Experimentally observed MTSL conformation (stick) compared to MMM calculated conformations. The most probable at 175 K (orange line, $P = 0.17$) is unlike the experimentally observed conformation, and points away from the hydrophobic pocket. The most probable conformation at 50 K produced a DEER distance similar to that experimentally observed. However, comparison of this conformation ($P = 0.65$) with the experimental conformation shows that this is not due to the correct prediction of the experimental conformation.

3.4.1 Agreement of the structure with DEER.

The distance between spin labelled cysteines observable in the crystal structure, taking the measurement from the centremost point of the N-O bond, was 4.5 nm and corresponds exactly to that determined by DEER. This shows that the same conformation of dimer exists in the frozen solution state for the DEER measurements as in the crystalline environment. It provides a further demonstration of the ability of EPR to correctly measure distances between residues of an interacting species.

3.4.2 Disagreement of the structure with MMM.

The crystal structure allowed a comparison of the actual spin label conformation with those calculated by MMM. The simulations were repeated using the spin labelled crystal structure (PDB 2XGA) to aid this comparison. In agreement with the original simulations, a DEER distance of 5.3 nm was predicted (standard deviation 0.37 nm, relative width 7.0%) and unless stated otherwise, results presented are for these 2XGA simulations rather than those using PDB 1RY9. All of the higher probability conformations placed the label in a less sterically hindered position than the hydrophobic pocket at 175 K. The program did produce a conformation similar to that which was experimentally observed, however only giving it a probability of 0.01.

The top probability conformation at 175 K has a somewhat different χ_4 of 91.0 ($\pm 0.1^\circ$) to the experimentally determined value of -51° ($\pm 0.5^\circ$) as a result of the stabilisation from the MTSL O to the N_{D1} of His120 (angle defined in Figure 3-12b). Furthermore, the absence of an MTSL S_D and H- C_α interaction allows a greater extension of the label into the hydrophobic pocket than predicted, for the small price of the loss of this interaction.

MMM simulations are usually performed at the glass transition temperature as experimental DEER data demonstrates the freezing of conformational distributions below this temperature (Jeschke *et al.*, 2010). However, repeating the MMM simulations at 50 K and 10 K gave a different result with the top

probability conformation giving a DEER distance of around 4.6 nm, close to the experimentally observed value. Specifically, at 50 K the DEER prediction was 4.73 nm with standard deviation 0.22 nm and relative width 4.7%. At 10 K the DEER prediction was 4.63 nm with standard deviation 0.10 nm and relative width 2.2%. Despite the similarity in predicted distance, comparison of the MTSL conformation with the crystallographic one showed that they were still dissimilar, meaning that the agreement in DEER distance was due to chance rather than by correct prediction of the MTSL conformation (Figure 3-12).

A comparison between the top probability rotamers calculated at 175 K by MMM using the current crystal structure (spin labelled Spa15 PDB 2XGA) and the original unlabelled Spa15 (PDB 1RY9 (Van Eerde et al., 2004)) allowed insight into the effect of small side chain reorientations upon the calculations. When using unlabelled Spa15 (1RY9) as the basis model, the top probability spin label rotamer ($P = 0.16$) was similar to that of the labelled Spa15 (2XGA) derived MMM rotamers, pointing away from the hydrophobic pocket and predicting a similar incorrect DEER distance of 5.4 nm. There is a small perturbation in conformation of the label however towards Ile17 when the unlabelled Spa15 (1RY9) model is used. This is a residue which encroaches on the pocket more in unlabelled Spa15 (1RY9) than in labelled Spa15 (2XGA), as is visible in Figure 3-12c. This is a hydrophobic factor possibly weighting the MMM change, but emphasis must be drawn to the fact that this label conformation is also not observed in the labelled Spa15 (2XGA) crystal structure.

This experiment has successfully demonstrated the use of EPR data in studies of protein interactions. Previous conformational work on MTSL has largely focussed upon labels within α -helices, and the resulting interactions. These are obviously rigid and thus observable in a crystal structure. Trends have therefore been identified pertaining to helix environments, noting particularly such interactions as that with the $i + 4$ residue (Langen *et al.*, 2000). The spin labelled Spa15 structure demonstrates a very

different type of MTSL conformation, dictated not by residues close in sequence, but rather those close in space. The DEER distance measurements agree closely with distances calculated from the crystal structure of the spin labelled protein, demonstrating it as a useful method for the elucidation of structural phenomena.

3.5 The effect of IpgB1 binding.

An observable of interest was whether DEER could detect a structural change upon binding of the effector IpgB1 to Spa15. The co-purification of Spa15:IpgB1 complex in a 2:1 ratio was confirmed using MALLS (Figure 3-13), showing that IpgB1 has no effect upon the stoichiometry of the dimeric Spa15. This is to be expected for a Class I chaperone from homology with other chaperone:effector complexes (such as InvB/SipA (Lilic *et al.*, 2006), YopE/SycE (Birtalan *et al.*, 2002)).

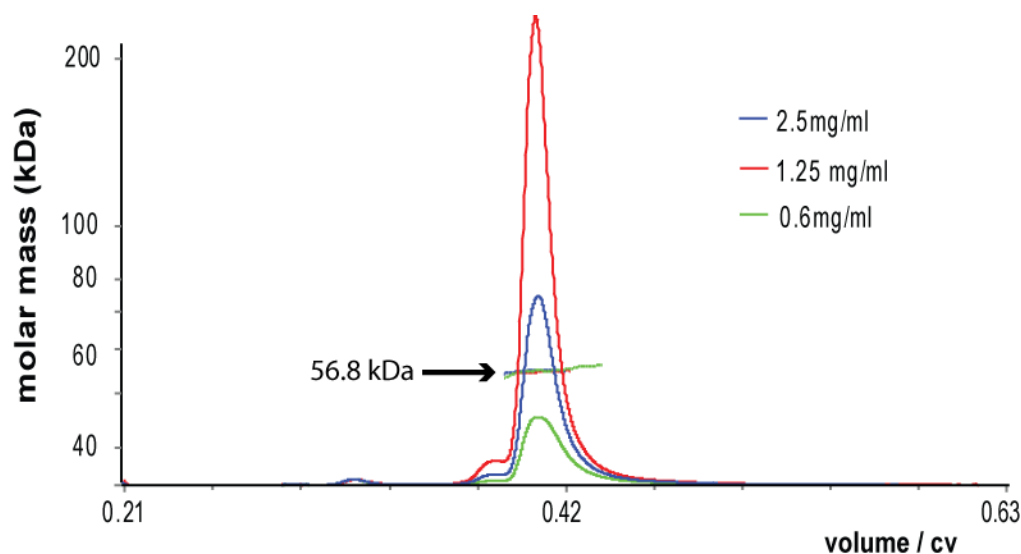


Figure 3-13 SEC-MALLS of Spa15:IpgB1 complex. The recorded mass is in agreement with a 2:1 ratio of Spa15 (17 kDa) to IpgB1 (23.6 kDa). SEC flowrate 1.0 ml/min.

As stated above, upon spin labelling of the complex, Spa15 was labelled but IpgB1 remained unlabelled. Spin labelling of the Spa15:IpgB1 complex had no significant effect (± 0.1 nm) upon the major DEER distance relative to that of the Spa15 dimer (Figure 3-14). Satellite peaks were again observable due to the high concentration conditions.

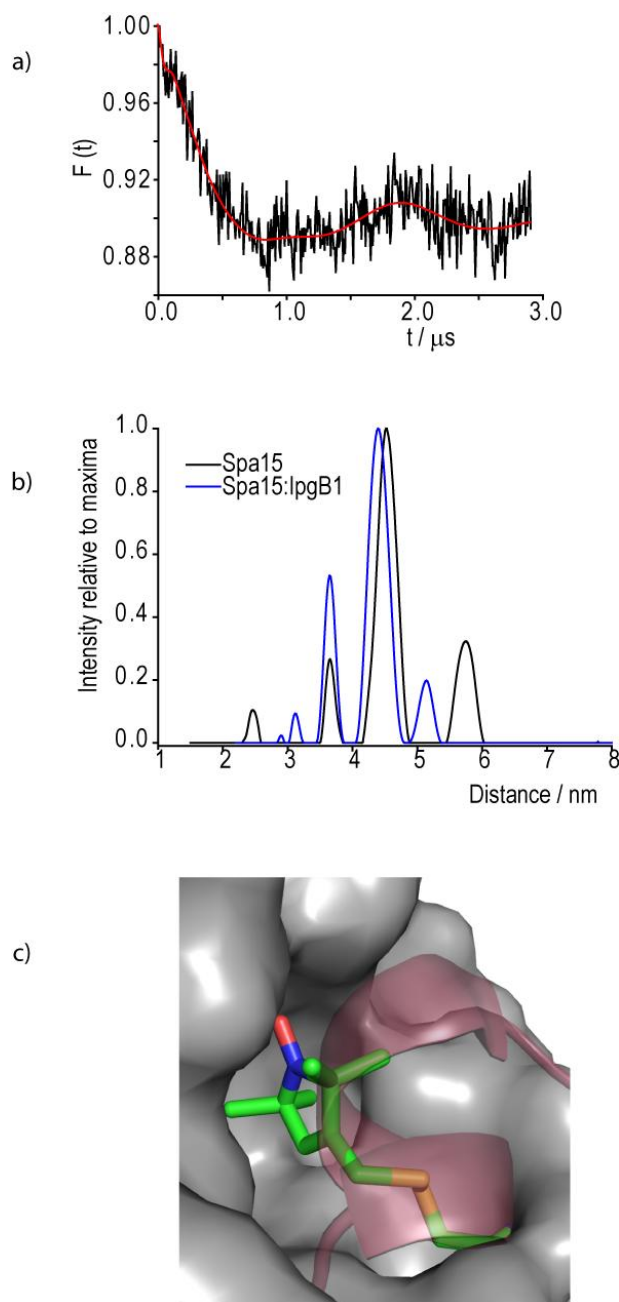


Figure 3-14 The impact of IpgB1 binding on Spa15 DEER. **a)** DEER trace for Spa15:IpgB1. Note: Modulation depths between Figure 3-7 and Figure 3-14 are not directly comparable due to different technical conditions when the samples were measured. **b)** DEER derived distances for Spa15 dimer (black) compared to Spa15:IpgB1 (blue) show no significant difference in distance between the Cys19 residues of Spa15 when the effector is bound and unbound. **c)** Part of the InvB:SipA structure (PDB 2FM8) with spin labelled Spa15 (grey surface) replacing InvB. SipA (red ribbon) occludes the binding pocket from the spin label. MTSL is shown in green (atom coloured stick), clashing with SipA.

As is detailed in Chapter 4, the Spa15:IpgB1 complex underwent extensive crystallography trials in order to yield a crystal allowing structural examination around the Spa15 Cys19 residues, but no crystals were forthcoming. The absence of any crystals or other Spa15:effector structures required us to look at a homologous protein complex for DEER distance explanation. The *Salmonella* spp. Spa15 homologue InvB (33% sequence identity) bound to the chaperone binding domain of its effector SipA allowed us to examine how an effector is expected to bind around the chaperone. In this example, SipA wraps around InvB, donating a β -strand to a β -sheet groove of the chaperone, and filling a hydrophobic groove of the chaperone in the common style of Class 1 chaperone:effector complexes (introduced in Chapter 1). In addition however, SipA uses leucine residues to bind the currently discussed hydrophobic pocket of InvB, thus presumably occluding the site from a spin labelled cysteine residue which is conserved at position 19 (Figure 3-14c). This is clearly not the case with Spa15:IpgB1, as seen by 100% labelling and a DEER

distance unchanged by complexation. Furthermore, from this DEER data, it is shown that the Spa15 dimer does not seem to be 'compressed' or changed in global conformation upon IpgB1 binding, as a result of the invariant inter-dimer distance.

3.6 Conclusions.

The invariance of the distance between Spa15 spin labelled Cys19 residues when uncomplexed and when bound to IpgB1 has been analysed. The effector binding has no constraining effect when it wraps around the chaperone, and does not alter the spin label conformation demonstrating that Spa15 and IpgB1 cannot interact in the same way as the complex of the InvB/SipA homologue. The spin label conformation itself has been verified by DEER and crystallography, and shown to remain in one favoured position. This position is not that predicted on the basis of the MMM criteria. However, despite the apparent steric limitations of its position relative to a position pointing away from the protein, DEER and crystallography both show that due to hydrophobic interactions it is the one that is favoured.

The purification of the Spa15:IpgB1 complex is a result which allowed the DEER experiment showing the effect on Spa15 to be shown. It was also the first purification of this complex, which had only previously been shown by a two-hybrid screen in yeast (Page *et al.*, 2002) and the first soluble purification of IpgB1, which untagged is highly insoluble (Klink *et al.*, 2010).

4 T3SS - Spa15 WxxxE Effectors and their Homologues.

In Chapter 3, the identity of Spa15 as a promiscuous chaperone was introduced. Two of the effectors which are chaperoned by Spa15 are IpgB1 and IpgB2. These belong to a group of proteins found in *Shigella*, *Salmonella* and EPEC/EHEC with a common fold, stabilised by a WxxxE motif, with corresponding roles in bacterial invasion.

In this chapter, studies upon some of the members of this group are presented. Spa15 is shown to be an effective chaperone for IpgB1, bringing it into solution when it is insoluble as a single species in the recombinant expression system used. This is in contrast to IpgB2, which is insoluble with or without Spa15 coexpression.

Functional studies were carried out on the soluble EHEC and *Salmonella* homologues EspM2 and SifA, which are members of the WxxxE group. SPR allowed their binding to host proteins to be shown, adding them to the growing list of WxxxE proteins shown to be functional mimics of human proteins, despite having no structural conservation with their eukaryotic counterparts.

4.1 WxxxE proteins and their function.

Manipulation of cytoskeletal actin is used by cells for a variety of functions including motility and cell to cell communication. The Ras-like GTPases signal this manipulation, they switch from their inactive GDP bound state to the active GTP bound state in order that they set off a chain of processes enabling specific actin architectures to be built (Hall, 1998; Valderrama *et al.*, 2006).

Pathogens have been found to manipulate these GTPases, to aid their tight binding or internalisation into the cell. In 2006 the WxxxE family was identified (Alto *et al.*, 2006). Conservation of the characteristic motif in some of the group members is shown in Figure 4-1. Secondary structure prediction anticipated that each member consisted of 6-8 α -helices, connected by a common protein V-

shape fold. The study reported that different members of the group could mimic the Rho GTPases Cdc42, Rac1 and RhoA, which led to the formation of filopodia, lamellipodia and stress fibre phenotypes (respectively) upon subjection of the effector on HEK293A cells. Subsequent work has extended this group in number and clarified the role as being that of a Guanine Exchange Factor (GEF), activating the Rho-GTPases rather than mimicking them (Arbeloa et al., 2008; Bulgin et al., 2009; Ohya et al., 2005). This activation is achieved through removal of the GDP from the inactive GTPase in order that GTP might bind.

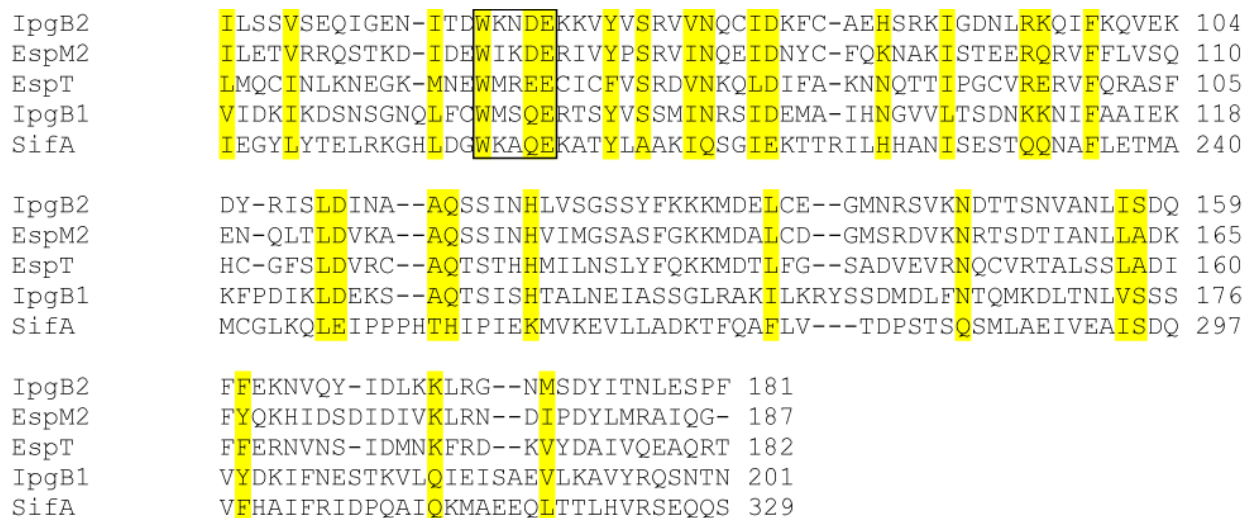


Figure 4-1 Multiple sequence alignment of the WxxxE domain of the effectors involved in this study. The WxxxE motif is boxed, with similar residues coloured in yellow. Figure prepared with ClustalW2.

Currently, the structures of three members of the WxxxE group are known, and indeed conform to the predicted V-shape architecture: Map (*E. coli*), SifA (*S. typhimurium*) and IpgB2 (*S. flexneri*) (Huang et al., 2009a; Klink et al., 2010; Ohlson et al., 2008). Elucidation of their function and mechanism initially relied upon a homologous bacterial GEF which was found previously (Buchwald *et al.*, 2002). SopE of *S. typhimurium* enables bacterial internalisation by inducing pronounced rearrangement of the actin cytoskeleton and membrane ruffling through its action on the Rho GTPases Cdc42 and Rac1. After internalisation of the bacteria, its complimentary partner SptP downregulates the signaling from these

GTPases, contributing to the recovery of the host cell (Buchwald *et al.*, 2002; Stebbins and Galan, 2001b).

The first identified mammalian Rho GEF was Dbl (isolated from Diffuse B lymphoma cells), and many subsequent eukaryotic GEFs were found to contain the same domain, termed the Dbl homology (DH) domain (Rossman *et al.*, 2005). The domain forms the catalytically active part of these proteins, responsible for binding the GTPase regions which undergo major conformational change upon GTP and GDP binding - switch 1 and switch 2. The other mechanistically important part of the GTPase in GEF action is the β 123 strand patch, which has non-conserved residues between RhoA, Rac1, and Cdc42. Non-conservation allows the β 123 region to be the basis for selection between the three GTPases (Huang and Chai, 2010).

Upon solution of the crystal structure of the complex SopE/Cdc42, comparisons with the eukaryotic GEFs were drawn (Buchwald *et al.*, 2002):

Firstly, there is no similarity between the overall sequence or structure of SopE and the Dbl like GEFs. Both are largely α -helical, but consist of different tertiary folds and use different amino acids to interact with Cdc42. Furthermore, the catalytic cores are very different, SopE has a GAGA small residue catalytic motif, in contrast to the bulky side chains employed by its eukaryotic counterparts.

Secondly, despite the differences in structure, upon binding either bacterial or eukaryotic GEF, the switch regions of Cdc42 are conformationally altered with very similar local rearrangements. SopE binds Cdc42 by pushing switch 1 aside and pulling switch 2 towards its GAGA motif. Insertion of the GAGA motif causes two consecutive peptide flips in switch 1, in a conformation that is stabilised by hydrophobic burying and a salt bridge between SopE and Cdc42.

SopE is not a member of the WxxxE family. However, it is a bacterial GEF, which is composed of six α -helices held in two bundles in a V-shape, with a catalytic loop. This is exactly the architecture predicted of the WxxxE family, and thus it was unexpected to find that when crystallised, Map (EHEC) was shown to have seven α -helices, in one 4 and one 3 bundle, with the WxxxE motif located at the interface of the bundles. This motif is not catalytic; it provides structural support by hydrogen bonding to amino acids (often two serines) on the catalytic loop. Map-Cdc42 was crystallised and shown to interact with the GTPase switches 1 and 2 similarly to SopE (Huang et al., 2009b).

GTPase specificity for the bacterial GEFs was highlighted by Huang *et al.* As mentioned above, Dbl GEFs distinguish between GTPase isoforms by pairing with their β 123 specificity patch. Huang saw similar conformational changes induced by Map in the β 123 region as for the eukaryotic ITSN-Cdc42 complex. Although Map will not bind wild type Rac1, upon mutation of the Rac1 specificity patch to the Cdc42 residues, Map could be induced to bind.

Recent mechanistic insight to the bacterial GEF activation has been provided by way of a number of IpgB2:RhoA structures (Klink *et al.*, 2010). These show the initial binding of IpgB2 to RhoA-GDP-Mg²⁺, and follow the displacement of the Mg²⁺, a step which reduces the affinity of GDP to RhoA before the GDP is completely removed (as seen in the Map:Cdc42 structure).

With the IpgB2 structures adding to the already known SifA and Map examples, it appears that SopE and the WxxxE proteins are acting as GEFs, but in a structurally different way to eukaryotic proteins. Convergent evolution, creating functional mimicry via structural novelty, is not always the method used in bacterial pathogenesis - *Yersinia* and *Pseudomonas* GAPs mimic eukaryotes with the same arginine finger used as a catalytic motif (Stebbins and Galan, 2001b). Despite the structural homology seen between SopE and WxxxE type prokaryotic GEFs, there is a key difference: SopE utilises a GAGA catalytic motif which the WxxxE proteins do not possess.

4.2 The Spa15 chaperoned effector IpgB1.

Initially, the WxxxE member IpgB1 was selected as a structural target. This is a bacterial mimic of the eukaryotic RhoG, binding to the guanine exchange factor ELMO-Dock180 complex to activate Rac1 GTPase. This brings about lamellipodia and membrane ruffling of the cell, aiding the internalisation of *Shigella* (Handa *et al.*, 2007; Ohya *et al.*, 2005). IpgB1 is chaperoned by Spa15, and in Chapter 3, the purification of this complex was detailed.

4.2.1 Expression of IpgB1.

Full length solitary recombinant IpgB1 proved to be insoluble. Expression was carried out with both N-terminal His-tagged and untagged constructs (4.10 and 4.11 in Table 2.1 respectively - Figure 4-2a) trialing expression conditions 20°C, 30°C and 37°C. These variations were made since the presence of affinity tags and temperature changes are known to affect the expression levels of soluble protein (Terpe, 2006). The EHEC homologue of IpgB1, EspT, which too leads to observation of lamellipodia architecture in cells (Bulgin *et al.*, 2009) was also found to be insoluble in a His-tagged construct donated by Ana Arbeloa (Figure 4-2).

A number of published studies have commented upon IpgB1 insolubility (Handa *et al.*, 2007; Klink *et al.*, 2010). Soluble IpgB1 has therefore been obtained by researchers through fusion protein expression. Sizable tags such as GST (26 kDa) and MBP (43 kDa) have been attached, where the tag protein is greater than the size of IpgB1 (23.6 kDa). It is possible for protein functional behaviour to be modified when taking this approach. An alternative approach used here has been the utility of the cognate chaperone. As shown in the previous chapter, coexpression of IpgB1 with Spa15 brought IpgB1 into the soluble fraction. The chromatographic trace at the end of the purification is shown in Figure 3-2. After expression trials at 20°C, 30°C and 37°C, overnight expression at 20°C was selected as giving the best yield of protein (Figure 4-2b).

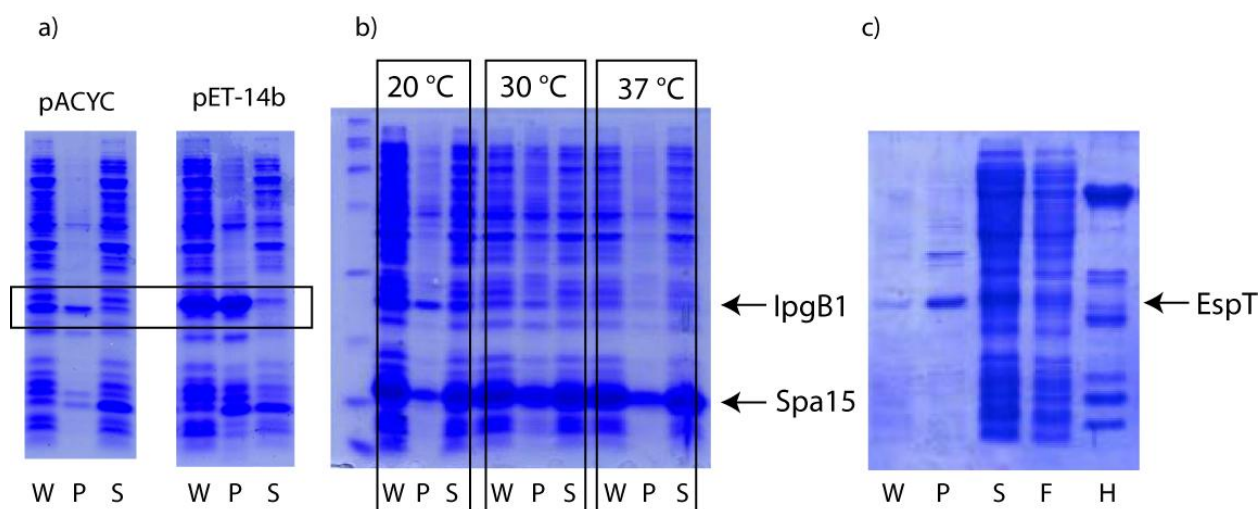


Figure 4-2 Solubility of IpgB1 brought about through coexpression with Spa15. **a)** IpgB1 expression trial - insolubility of IpgB1 with (pET-14b) and without (pACYC) His-tag. **b)** Spa15:IpgB1 expression trials - optimal expression seen at 20°C. **c)** Insolubility of His tagged EspT. Whereas overexpression of a correct weight protein was observed, this remained in the insoluble pellet after centrifugation. W (whole cell), P (cell pellet), S (cell supernatant), F (Ni-NTA flow through), H (Ni-NTA 'His trap' elution).

4.2.2 Characterisation of the Spa15:IpgB1 complex.

In Chapter 3 the 2:1 stoichiometry of Spa15:IpgB1 was determined using MALLS. The purified complex was stable in 50 mM Tris pH 7.5, 150 mM NaCl, 5 mM DTT for a week at 4°C post purification.

Since a putative role of chaperones is that of passive protection, a partial proteolysis experiment with trypsin was carried out to examine the protection of IpgB1 in the complex. Only the first thirteen residues of IpgB1 were not resistant to trypsin proteolysis by Spa15 (Figure 4-3a), which is common for chaperone:effector complexes. Furthermore this region is expected to be somewhat disordered by the disorder predictor Ronn (Yang *et al.*, 2005). This shows an increased probability of disorder at the very beginning of the sequence relative to the rest of the protein – based on the primary sequence (Figure 4-3b). N-terminal sequencing also revealed the stability of Spa15 under the proteolytic conditions, with cleavage of the N-terminal His-tag leaving the stable protein. The more vigorous protease subtilisin carlsberg however cut both Spa15 and IpgB1 completely (Figure 4-3a).

The complex forms a very tight interaction and the components could not be separated via ionic strength of the buffer. This was demonstrated in the ion exchange chromatographic purification

protocol which saw a buffer exchange of 20 – 500 mM NaCl to exploit the electrostatic difference between IpgB1 (pI = 9.5) and Spa15 (pI = 4.4) at pH 7.5. Throughout this process the complex stayed intact. Assessment of the complex stability with pH was further seen by placing the complex in different pH solutions; precipitation was observed above pH 9.5 in the buffer 50 mM glycine-NaOH, 200 mM NaCl, or below pH 6.0 in the buffer 50 mM MES-HCl, 200 mM NaCl. The remaining solution was found to contain neither soluble Spa15 nor IpgB1.

4.2.3 Removal of the IpgB1 N-terminus.

In an attempt to obtain a soluble IpgB1 construct, an N-terminally truncated mutant, lacking the first 57 residues of the full length protein, was designed. This truncated mutant corresponds to the conserved C-terminal domain of the WxxxE group, from homology with the crystal structures of SifA, Map and IpgB2. Upon recombinant expression this domain was found to be insoluble, as shown in Figure 4-3c.

As detailed in Chapter 1, the non-conserved N-termini of type III effectors has widely been found to contain the 25-100 residue chaperone binding domain, following or containing a putative signal sequence (Phan et al., 2004; Rodgers et al., 2010). This has been shown in such examples as the *Yersinia* chaperones SycE and SycH which bind their unwound effectors YopE and YopH at residues 15-75 and 20-70 respectively (Birtalan *et al.*, 2002; Woestyn *et al.*, 1996). In agreement with this, the truncated mutant IpgB1 Δ 1-57 was unable to be coexpressed with Spa15. Repetition of the purification protocol as for the wild type, with N-terminal His-tagged Spa15 and untagged IpgB1 Δ 1-57 led to purification of Spa15 only.

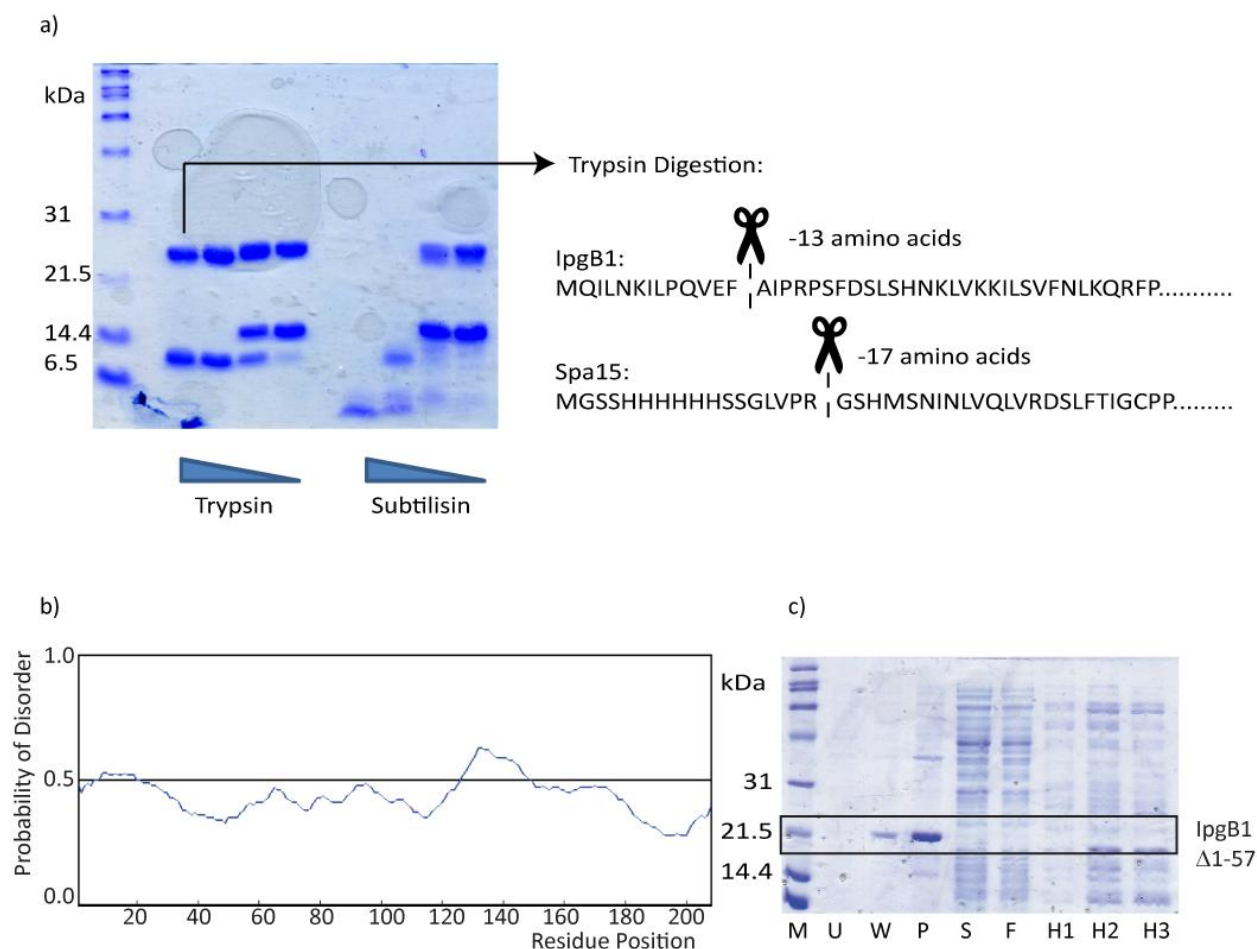


Figure 4-3 Characterisation of Spa15:IpgB1. **a)** Partial proteolysis in the range 1:80 – 1:10000 protease:protein (w/w). A stable fragment resulted from incubation with trypsin after 2 hours at room temperature; however subtilisin carlsberg degraded the proteins fully at the higher protease concentrations. The tryptic fragments were determined by N-terminal sequencing as shown. N-terminal sequencing was carried out by Tony Willis, Department of Biochemistry, University of Oxford. **b)** RONN disorder predictor indicates that the expected disorder profile of IpgB1, with the first 20 amino acids predicted to have higher probability of disorder than the immediately following sequence. **c)** Purification attempt of IpgB1Δ1-57 demonstrated the insolubility of this construct – protein corresponding to the weight of IpgB1Δ1-57 is observable in the insoluble cell pellet only. M (markers); U (uninduced cell); W (whole cell); P (pellet); S (supernatant); F (His-trap flow-through); H1-3 (His-trap fractions).

4.2.4 Spa15:IpgB1 crystallographic attempts.

Crystallisation was attempted with the Spa15:IpgB1 complex as purified. A number of approaches were taken:

- i) Full length and proteolysed complex.

The high level of complex purity obtainable made the unmodified species a reasonable first target. High purity is often necessary (but not sufficient) for crystallisation, and is often cited as the most important

factor to a successful crystallisation (Rhodes, 2006). Trays were set down at 5.7 and 2.8 mg/ml, at 21°C and 4°C, in the broad screens JCSG+, MD1&2, Stura Macrosol and Pact Premier. The two concentrations ensured variation in the environment for the many conditions tested, and the two temperatures altered the solubility of the protein. These trays were repeated with addition of 10 mM β -mercaptoethanol, to prevent short term non-specific aggregation.

The proteolysed complex was also put into crystal trials at 3.0 mg/ml. The four broad screens were set up with material previously proteolysed by trypsin (1:80 w/w for 2 hours at 21°C) and purified by size exclusion chromatography. Trays of full length protein with in-drop trypsin (1:5000 w/w) were also set up. In drop proteolysis has been shown in certain cases to bring about crystallisation – the flexible parts of the protein are cleaved and at such point as a rigid fragment remains, the protein crystallises (Dong *et al.*, 2007).

ii) Surface Entropy Mutants.

Three mutants with double point mutations were designed and purified. These were 108NKKN→NAAN, 117EKKF→EAAF and 152LKRY→LAAY. These sites were chosen as suitable by identification of groups of adjacent conformationally free lysines or lysine/arginines, which were predicted as suitable sites by the server SER (Goldschmidt *et al.*, 2007). Surface entropy mutants were cloned using the methods outlined in Table 2.8.

Each mutant was purified using the same method as the wild type. In every case, the yields were somewhat lower than the wild type protein (WT|2mg/l, M108|0.75mg/l, M117|0.05mg/l, and M152|1mg/l). Prior to crystallisation, each of the mutants was truncated by trypsin (which was removed by size exclusion chromatography). Purity of the protein is shown in Figure 4-4a. Crystal trials were setup at 2.8-3.0 mg/ml.

iii) Reductive Methylation.

Wild type IpgB1 contains 21 lysine residues, which have a significant influence on the resulting alkaline pI of the protein. Methylation of these within the Spa15:IpgB1 complex was attempted to reduce the entropy of crystallisation as detailed in Table 2.8. During methylation, the concentration of protein was kept at 0.3mg/ml, since attempts at higher concentrations led to complete precipitation of the sample. The mass spectrometry result in Figure 4-4b shows an increase in mass of 610 Da for IpgB1 and 51 Da for Spa15. This indicates methylation of both Spa15 and a significant number of IpgB1 lysine residues, from an increase of 14 Da per methyl group. Crystal trials were set up at 2.2mg/ml, the reduced solubility compared with WT protein being in accordance with the increase in hydrophobicity.

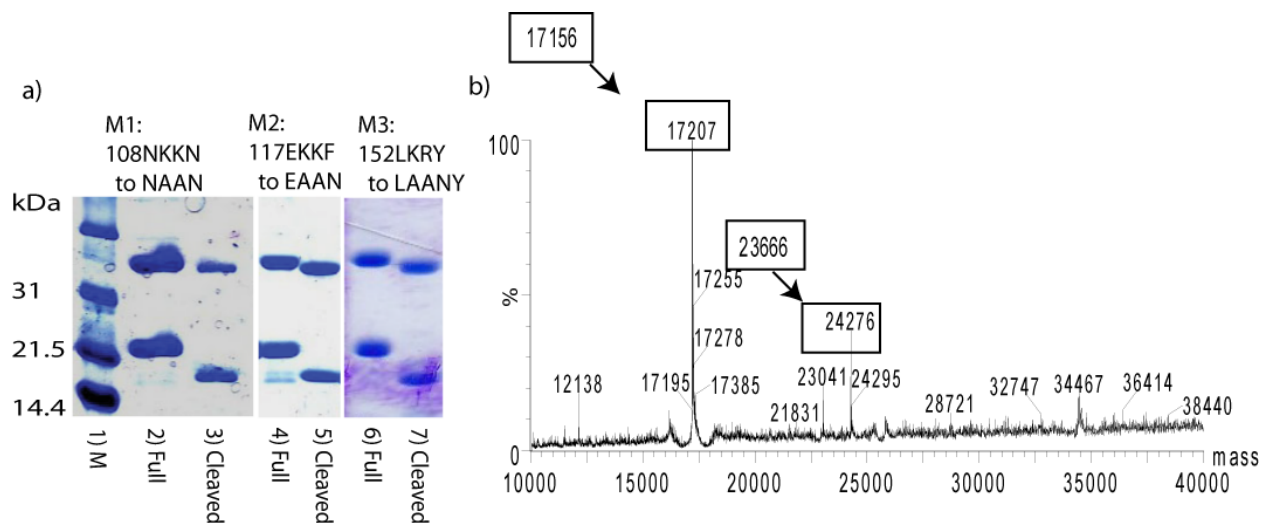


Figure 4-4 Protein manipulations made to encourage crystal formation. a) Three surface entropy mutants were purified and proteolysed with trypsin before setting up crystal trials. **b)** Lysine methylation added 51 Da to Spa15 and 610 Da to IpgB1.

Despite the above attempts, Spa15:IpgB1 failed to produce any crystals. IpgB1 is predicted to be a difficult target to crystallise by Xtalpred (Slabinski *et al.*, 2007), placing it in class 4/5 of predicted difficulty. One of the reasons for this label is the high isoelectric point – structures with pI at 9.5 statistically are shown by Xtalpred to have a low history of success. Whilst prediction programs certainly should not be taken at face value, in this case, the prediction proved correct.

IpgB1 thus proved difficult to work with. Attention was turned to other members of the WxxxE family, which were soluble and with which some collaborators were interested in working. These were EspM2 and SifA, and are described in the next section.

4.3 WxxxE proteins which target RhoA.

4.3.1 The role of EspM2 and SifA.

The EHEC O157:H7 effector EspM2 is only found in half of EHEC (and EPEC) strains (Arbeloa *et al.*, 2009). This lack of conservation causes it to be considered an accessory effector to the core T3SS effectors in these bacteria, allowing different infection strategies across strains (Iguchi *et al.*, 2009). Globally, the attack by these bacteria are characterised by a close association with the enterocyte plasma membrane and destruction of brush border microvilli, as described in Chapter 1 (Garmendia *et al.*, 2005).

The precise role of EspM2 is thought to relate to location changes of the tight junctions in the epithelial monolayer (Simovitch *et al.*, 2010). Actin stress fibre formation and occurrence of a circular cell morphology was observed upon exposure of HeLa and Madin darby canine kidney epithelial cells to EspM2. This changed the tight junction barrier localisation – but did not break the barrier – possibly to avoid severe disruption of intestinal barrier functions. As noted above, the phenotype of stress fibre formation is brought about via the activation of RhoA, which then activates the coiled coil containing protein kinase ROCK and subsequent downstream effects (Hall, 1998).

Stress fibre formation is also seen in *Salmonella* transfection due to the combined *Salmonella* effectors SifA and SseJ (Ohlson *et al.*, 2008). The N-terminus of SifA binds the host SifA kinesin interacting protein (SKIP) and with its partner SseJ activates RhoA to bring about phagosomal tubulations (Brumell *et al.*, 2002; Ohlson *et al.*, 2008). These maintain the integrity of the phagosomal membrane and prevent *Salmonella* being released from its SCV niche into the cytoplasm (Ruiz-Albert *et al.*, 2002). Previous work

has demonstrated an interaction between SseJ and RhoA, and noted the absence of a direct SseJ-SifA interaction.

This study was completed in collaboration with Dr Ana Arbeloa and Dr James Garnett (Imperial College, London). It allowed characterisation of EspM2 as a RhoA guanine nucleotide exchange factor. SPR showed the binding of RhoA to EspM2. This result is in agreement with assays performed by Ana Arbeloa showing GDP release upon incubation of EspM2 with exclusively RhoA, and stress fibre formation upon EspM2 infection of Swiss 3T3 cells done previously (Arbeloa *et al.*, 2008). Furthermore, SifA was shown to bind RhoA and to confirm that, contrary to EspM2, it cannot act alone as a RhoA GEF.

4.3.2 Expression of EspM2 and crystallographic attempts.

The bacterial effectors EspM2 & SifA, and the human GTPases RhoA & Rac1 were purified using the conditions described in Table 2.6. Each was expressed with an N-terminal His-tag. Identity and purity of the proteins were judged from the size exclusion chromatography traces, SDS-PAGE (Figure 4-5) and with mass spectrometry.

Each of the proteins expressed were full length except for EspM2. Full length EspM2 was found to degrade after a day into a smaller fragment and after three days to have widely precipitated. The same initial fragment was also observed upon tryptic digest of full length EspM2. EspM2 was recloned as EspM2²⁹⁻¹⁹⁶, corresponding to the fragment, in a bid to express a more enduring protein. The protein remained somewhat unstable however, requiring the presence of 500 mM NaCl to stabilise it.

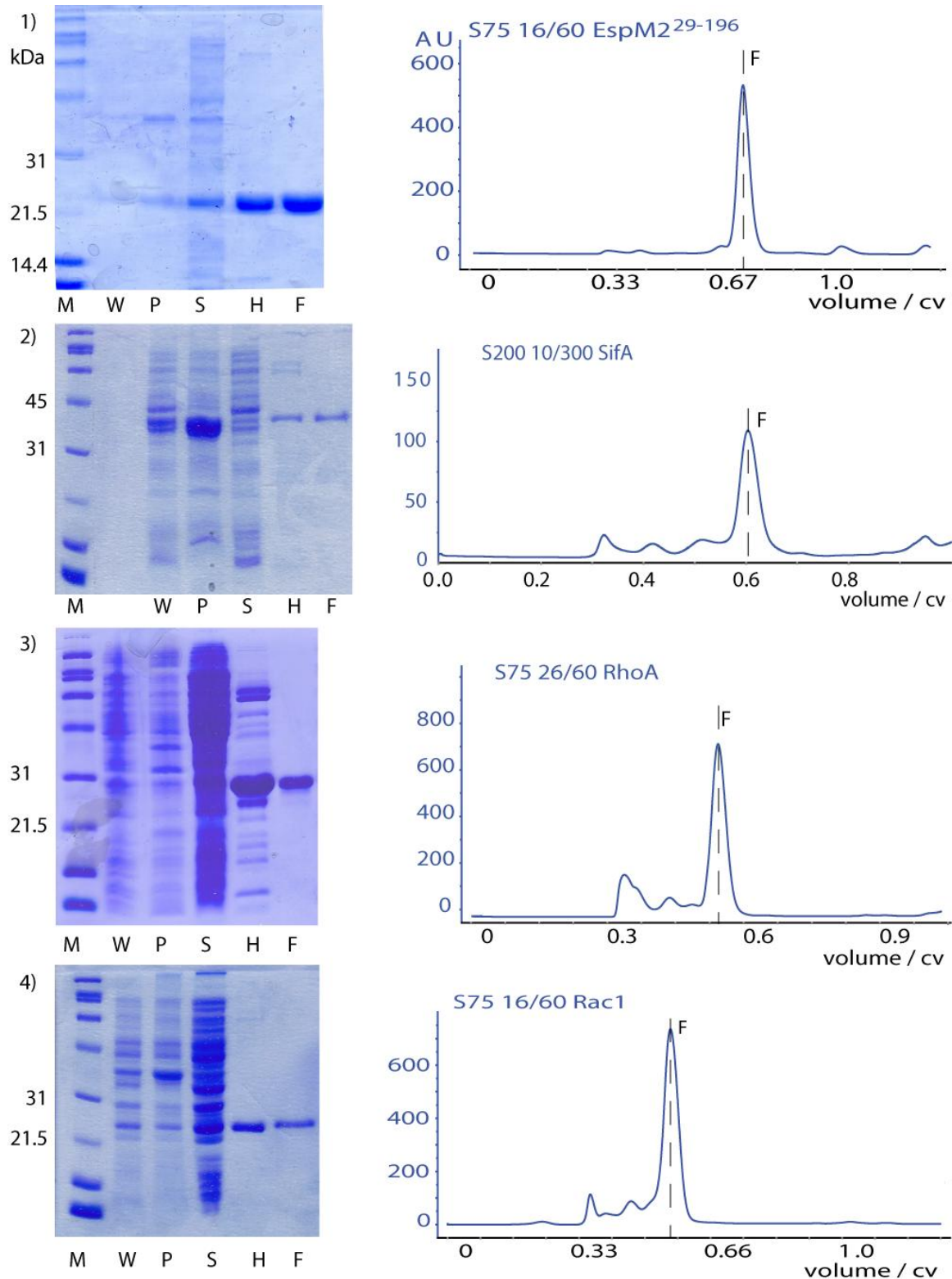


Figure 4-5 Size exclusion chromatography and SDS-PAGE from each of the proteins EspM2²⁹⁻¹⁹⁶, SifA, RhoA and Rac1. (Top to bottom respectively). SDS-PAGE column key: M (markers), W (whole cell), P (cell pellet), S (cell supernatant), H (Ni-NTA 'His trap' pooled fractions), F (purified protein fraction).

Crystallisation trials for EspM2 were no more forthcoming than those of IpgB1. The protein was subject to a number of modifications in an attempt to crystallise it (Table 4.1), but both the full length EspM2 WT and truncated EspM2²⁹⁻¹⁹⁶ were unreceptive to crystallisation.

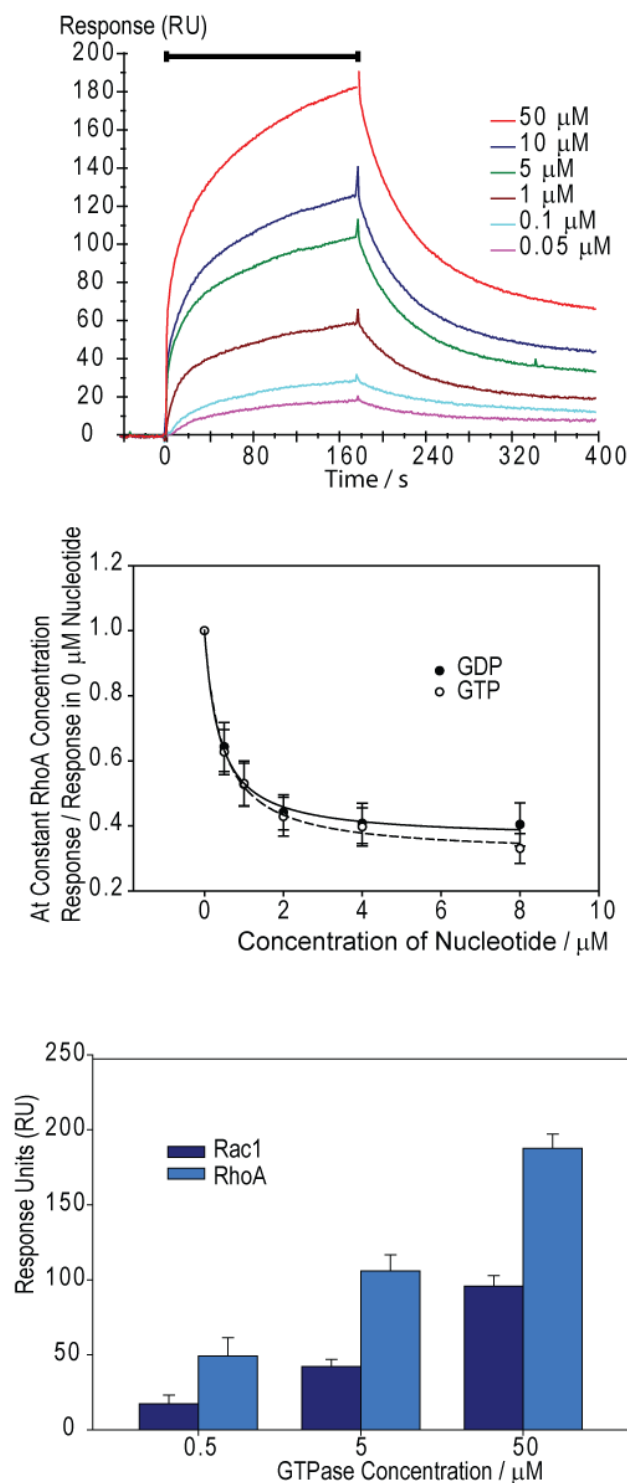
EspM2 WT EspM2 ²⁹⁻¹⁹⁶	1) Concentrations: 4.6 and 2.5 mg/ml. 2) In drop proteolysis: 1:10000 and 1:5000 trypsin:protein (w/w). 3) Methylation: addition of 406 Da to EspM2 and 350 Da to EspM2 ²⁹⁻¹⁹⁶ indicates methylation of 2/3 of lysines (averaging both di and tri methylated possibilities). 4) Temperature: 4 °C, 12°C, 21°C. 5) 0.5 M TMAO –Additive to stabilise protein (Jiang <i>et al.</i> , 2006). 6) 3 mM Iodoacetamide: Irreversibly block Cys to reduce non-specific aggregate formation.
-------------------------------------	---

Table 4.1 Attempted EspM2 crystallisation conditions.

4.3.3 RhoA binds EspM2.

Surface Plasmon Resonance was used to probe the interaction between EspM2²⁹⁻¹⁹⁶ and RhoA. Nucleotide free RhoA was flowed (sequential injections ranging from 0.05 µM to 50 µM) over an EspM2²⁹⁻¹⁹⁶ bound surface and displayed an increased rate of binding with concentration confirming a specific interaction (Figure 4-6 - top). Rac1 was flowed over the same EspM2²⁹⁻¹⁹⁶ surface and for a given GTPase concentration showed substantially less binding (Figure 4-6 bottom), confirming the specificity of EspM2²⁹⁻¹⁹⁶ for RhoA over other GTPases.

Although high RhoA concentrations (µM range) were required to see an interaction, EspM2²⁹⁻¹⁹⁶/RhoA binding was shown to be long lived. Interaction response was followed for 1000 s after injections, by which time the rate of dissociation appeared to have fallen to zero in an incompletely dissociated state. Dissociation of EspM2²⁹⁻¹⁹⁶/RhoA could not be brought about by addition of 500 µM free GDP in running buffer (in the presence or absence of 5 mM MgCl₂), in contrast to the rapid dissociation of SopE/Cdc42 reported in the literature (Rudolph *et al.*, 1999). Dissociation of EspM2²⁹⁻¹⁹⁶/RhoA was achieved in high pH conditions (25 mM NaOH, pH 12.4) to bring the response back to the pre-experiment level.



Whilst the presence of GDP could not dissociate an already formed complex, free GDP or GTP added to the running buffer did inhibit formation of an EspM2²⁹⁻¹⁹⁶/RhoA complex. This was shown to be a (guanine nucleotide) concentration dependent process (Figure 4-6 - middle). GDP and GTP both have the same effect as one another (within experimental error), reducing the level of binding of EspM2²⁹⁻¹⁹⁶/RhoA by up to 67% in the nucleotide concentration range 0.5 – 8 μ M. This was for a given concentration of RhoA analyte (4 mM) and is exactly as has been seen for other small GTPase/GEF complexes (Hart and Powers, 1995; Miki, 1995).

Figure 4-6 EspM2-RhoA binding. **Top** - SPR demonstrates RhoA concentration dependence for binding EspM2²⁹⁻¹⁹⁶. Concentrations of RhoA varying from 0.05 μ M to 50 μ M were flowed at 50 μ l min⁻¹ (duration indicated by black bar) over a CM5 sensor chip with EspM2²⁹⁻¹⁹⁶ covalently bound to the surface. Control substituted signals for 400 secs of the experiment are shown. **Middle** - Inhibition of EspM2²⁹⁻¹⁹⁶/RhoA binding by GDP or GTP. For RhoA (4 μ M) flowing over an EspM2²⁹⁻¹⁹⁶ bound surface in buffer containing GDP/GTP, the response maxima relative to response maxima in the absence of nucleotide are plotted with respect to GDP/GTP concentration. **Bottom** - GTPase binding of EspM2²⁹⁻¹⁹⁶ to RhoA and Rac1. Averaged response maxima for three representative concentrations compare the strength of EspM2 binding with the two GTPases, standard deviations are shown.

4.3.3.1 Biochemical support for an EspM2-RhoA interaction.

Despite many efforts using a variety of experimental conditions it was not possible to obtain SPR data which could be reasonably fit using standard affinity or kinetic interaction models. Specifically the data were fitted using models assuming either simple 1:1 binding (and allowing for refractive index changes, mass transport or baseline problems), 1:1 binding followed by a conformational change, assumption of a bivalent or heterogenous ligand or two state reactions. This implies that the binding event is more complex than that which can be reliably quantified using SPR data (BiaEvaluation 3.0). In particular the extremely slow dissociation rate meant that the off rate was impossible to define within a reasonable experimental time frame. As a result, the difference in binding of EspM2 to RhoA and Rac1 has been compared directly rather than via comparisons of affinity or kinetic constants.

The absolute ability of the GTPases to bind EspM2 under the physiologically relevant conditions correlates well with the functional data of Ana Arbeloa. EspM2 was shown to cause stress fibres in HeLa cells (Arbeloa *et al.*, 2008). The ability of EspM2²⁹⁻¹⁹⁶ to induce guanine nucleotide exchange in RhoA was demonstrated by a spectroscopic assay: the fluorescence emission of *N*-methylantraniloyl (mant)-GTP upon insertion into the GTPase binding pocket was monitored. As shown in Figure 4-7, emission increased in an EspM2²⁹⁻¹⁹⁶ concentration dependent manner upon nucleotide exchange in RhoA, but emission did not increase significantly in the Rac1 experiment. 250 mM EDTA was used as a positive control for exchange. This experiment confirms that EspM2 is able to function as a GEF for RhoA, activating it for GTP binding.

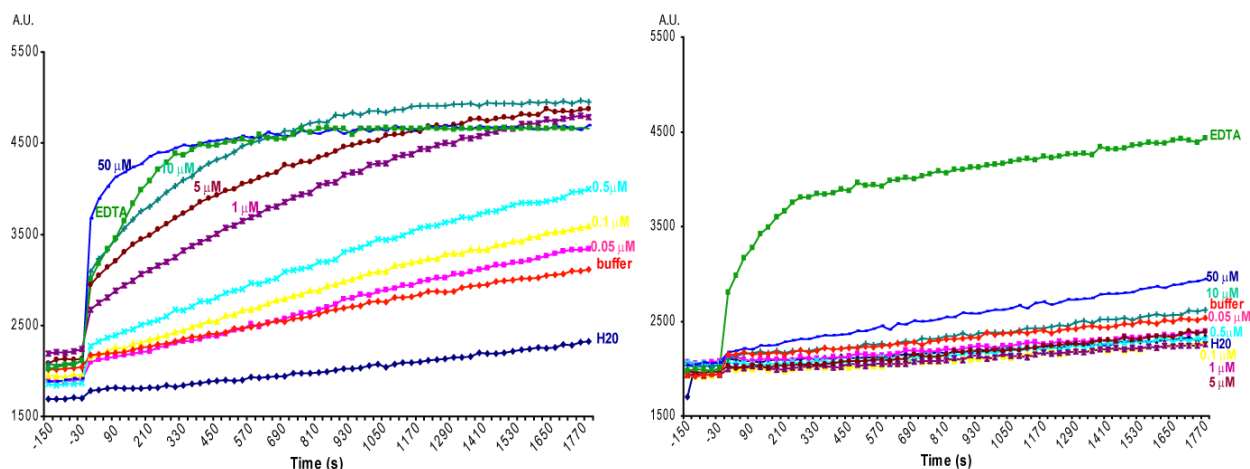


Figure 4-7 EspM2 enhances nucleotide exchange in RhoA. Left: EspM2²⁹⁻¹⁹⁶ mediates loading of (mant)-GTP into RhoA. (Mant)-GTP (0.5 μM) was incubated with 2 μM RhoA in the presence of 250 mM EDTA (squares), or 0.05–50 μM EspM2²⁹⁻¹⁹⁶ (circles) or in the presence of buffer only (triangles). The insertion of the (mant)-GTP into the nucleotide binding pocket of RhoA in presence of EspM2²⁹⁻¹⁹⁶ detected by an increase in the fluorescent emission was found to be concentration dependent. Right: EspM2²⁹⁻¹⁹⁶ does not induce nucleotide exchange for Rac1. Experimental conditions are the same as for RhoA with the substitution 2 μM Rac1. The traces show the average of three independent experiments. Work of Ana Arbeloa.

Three mutants were designed based upon i) alignment with the SopE and SifA catalytic loops and ii) an NMR study by Dr James Garnett which highlighted chemical shift changes in certain EspM2²⁹⁻¹⁹⁶ residues upon titration with RhoA (Arbeloa *et al.*, 2010). The aim was to identify those residues crucial for stability and activity.

Of the mutants, EspM2²⁹⁻¹⁹⁶W70A, I127A and Q124A (Figure 4-8b), the only one which was stably expressed was Q124A, as shown in Figure 4-8d. This correlates with a knocked out ability of ectopically produced W70A and I127A to produce stress fibres in Swiss 3T3 cells (Figure 4-8a). However for Q124A (which was stably expressed), stress fibre formation was attenuated rather than abolished (35% of Swiss 3T3 cells exhibited stress fibres compared to WT).

Purification yield levels of EspM2²⁹⁻¹⁹⁶ Q124A were essentially unaffected compared to WT (approximately 20 mg/l of bacterial culture), which allowed comparison in SPR binding to RhoA. This showed no significant change of EspM2²⁹⁻¹⁹⁶ Q124A to WT (Figure 4-8c). The architecture and binding of this mutant were thus unaffected, however the GEF abilities of EspM2²⁹⁻¹⁹⁶ Q124A, detected by the in

vitro (mant)-GTP spectroscopic assay were also found to be attenuated - 50 μ M EspM2²⁹⁻¹⁹⁶ Q124A was required to observe same GTP exchange as 1 μ M of wild type.

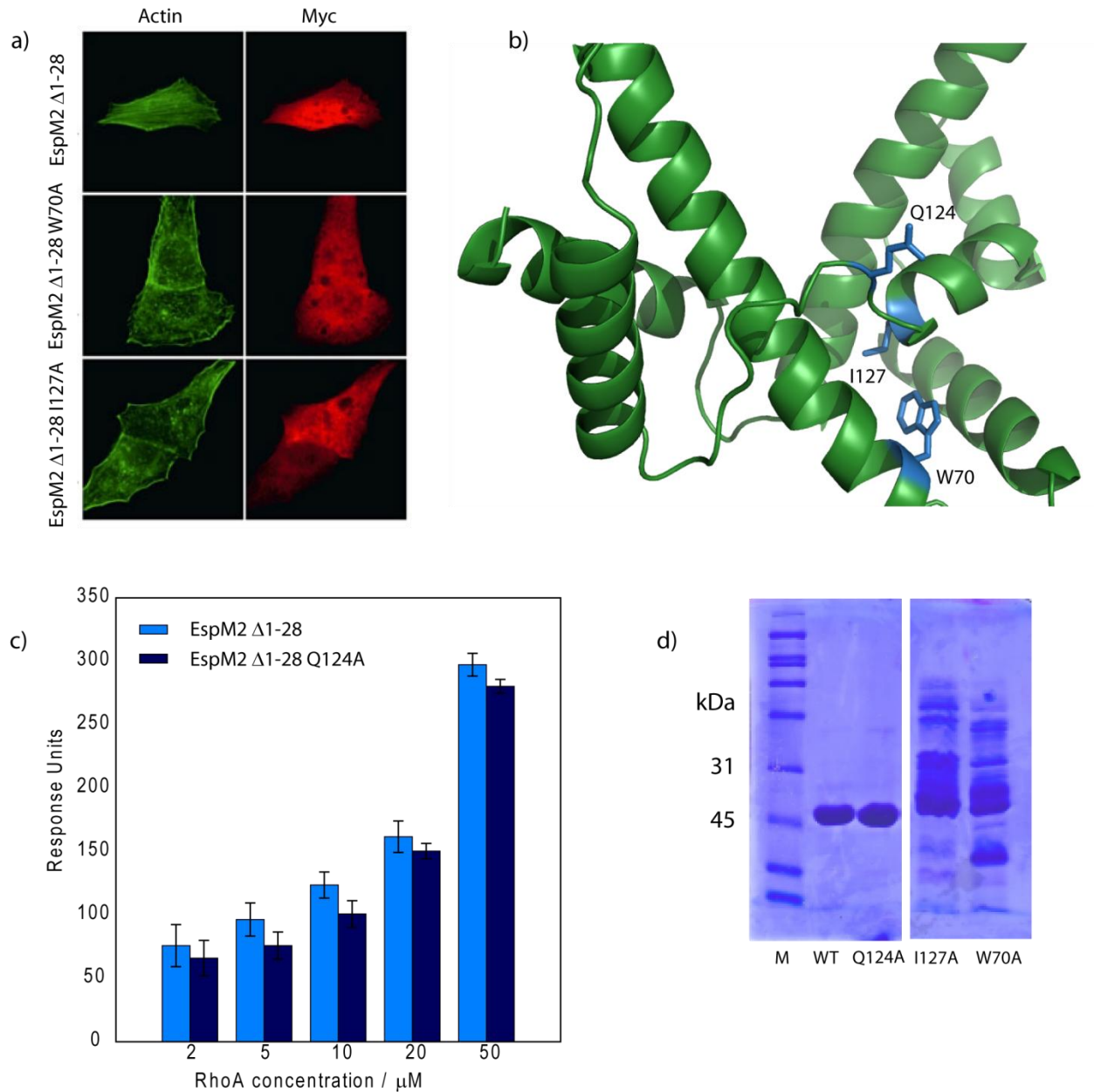


Figure 4-8 Identification of EspM2 residues key for structure and function. **a)** Swiss 3T3 cells were transfected with the mammalian expression vector pRK5 encoding myc-tagged EspM2²⁹⁻¹⁹⁶, EspM2²⁹⁻¹⁹⁶ W70A and EspM2²⁹⁻¹⁹⁶ I127A. Actin was stained with Oregon green phalloidin and the myc tag was detected with monoclonal antibody. (Work of Ana Arbeloa). **b)** Homology model (SQRL) of EspM2 using IpgB2 as a basis structure highlighting the positions of the mutated residues W70A, I127A and Q124A. **c)** SPR comparison of RhoA binding to EspM2 WT and to EspM2 Q124A. No significant difference in binding was detected over a range of RhoA concentrations. Figure shows the averaged response of three repeats. **d)** Best attained purity of EspM2²⁹⁻¹⁹⁶ WT, W70A, I127A and Q124A.

4.3.4 SifA binds RhoA.

As an accompaniment to the EspM2-RhoA work, a SifA-RhoA interaction was explored. Binding between RhoA and SifA was suggested by a pull-down experiment in the literature, but not characterised (Ohlson *et al.*, 2008).

A concentration dependent interaction was also observed between SifA and RhoA by SPR (Figure 4-9). SifA was flowed over the RhoA bound surface (1200 RU) of a CM5 chip, with concentrations ranging from 0.01 to 6 μ M. A SifA bound control lane was used to safeguard against nonspecific binding. Whilst incomplete dissociation was seen 1000 secs after the injection, regeneration of the chip could be easily achieved by passing a solution of increased ionic strength over the chip (50 μ l sodium tetraborate pH 8.5, 1 M NaCl). Similarly to EspM2-RhoA, the complex could not be disrupted by passing guanine nucleotides (500 μ M) over the chip (in the presence of absence of Mg^{2+}).

Measurement of the steady state affinity gave an indication of interaction strength for SifA-RhoA. A K_d of 250 nM with χ^2 value 4.12 resulted from a plot of equilibrium response with concentration (Figure 4-9). The 'kinject' mode was selected to minimise sensorgram distortion during the data acquisition, and after an initial experiment, the concentration of SifA flowed over the chip was over the range 0.1 to 10 times K_d , for independently repeated experiments.

This K_d reports a weaker interaction than that reported by many GEF-GTPases in the literature. The most directly comparable numbers come from studies of the *Salmonella* GEF SopE to the GTPases Cdc42 and Rac1, which report affinities of 0.3 nM (Friebel *et al.*, 2001; Rudolph *et al.*, 1999). Other studies of eukaryotic GEFs report values ranging from similar magnitude binding (150 nM for DBS-Cdc42) (Smith *et al.*, 2005), to a tighter interaction (3.3 nM for Cdc25-Ras) (Lenzen *et al.*, 1998).

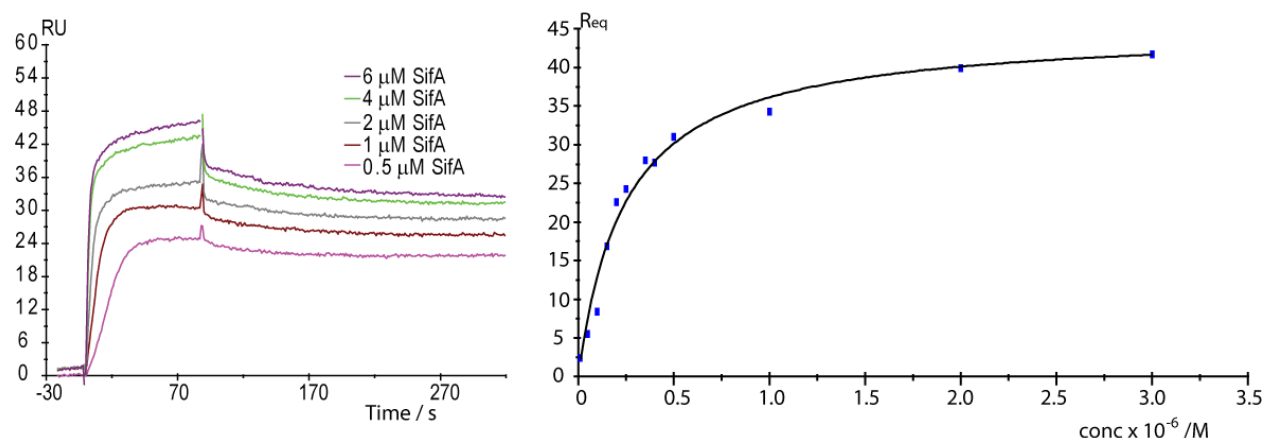


Figure 4-9 SPR binding of SifA to RhoA. **Left:** Concentrations of SifA varying from 0.01 μM to 6 μM were flowed at $35 \mu\text{Lmin}^{-1}$ over a CM5 sensor chip with RhoA covalently bound to the surface. Control substituted signals are shown. **Right:** Equilibrium binding analysis varying with concentration as calculated by BiaEvaluation. Independent experiments were repeated.

Interestingly, the (mant)-GTP spectroscopic assay used to characterise the GEF activity of EspM2 showed that addition of SifA did not induce nucleotide exchange in RhoA (work of Ana Arbeloa). Addition of SifA alone to cells has been shown to be insufficient to promote stress fibre formation (Ohlson *et al.*, 2008). The role of the cooperation between SseJ and SifA in inducing stress fibre formation characteristic of RhoA is thus unclear, however it is apparent that while binding RhoA, SifA is unlike EspM2 in being able to act alone as a RhoA GEF.

4.3.5 Insight from the structure of IpgB2.

The binding and promotion of guanine nucleotide exchange of EspM2 to RhoA led to the categorisation of EspM2 as a RhoA GEF. This study was presented alongside NMR characterisation by James Garnett, showing via $^{13}\text{C}^{\alpha}$, $^{13}\text{C}^{\beta}$ and $^{13}\text{C}^{\gamma}$ resonances that EspM2 is primarily helical, with helix content and structure homologous to the WxxxE domain of SifA, for whom a crystal structure had already been determined.

With this data, the residues I127 and W70 were expected to have a structural role, from their equivalents in the SifA structure. I127 is expected to hydrogen bond with the conserved E of the WxxxE motif, and W70 to interact with the carbonyl oxygen of a catalytic loop residue. In both cases the WxxxE

residues are used to stabilise the catalytic loop. It was therefore unsurprising that purification of the destabilised mutants I127A and W70A was impossible.

EspM2 proving refractory to crystallisation, however, did inhibit the interpretation of the role of Gln124. Mutation of this amino acid to alanine led to partial loss of function, but did not alter the binding of EspM2 for RhoA. Subsequent solution of the *Shigella* homologue IpgB2 in 2010 however elucidated the role of this residue, since EspM2 and IpgB2 are 43% sequence identical, and this residue is conserved (Klink *et al.*, 2010).

In the structure of IpgB2:RhoA, the conserved glutamine residue is shown to be a catalytically important residue, which interacts with the Mg^{2+} hydration shell. This brings about the dissociation of Mg^{2+} which enables nucleotide release. Similarly to our work, Klink *et al.* mutated this Gln residue to Ala, and saw a reduction in stress fibre formation. Further mutation to glutamate led to a complete abolition of stress fibre formation. The crystal structure shows that it is not necessary for complex formation, but supports nucleotide release, in complete agreement with the EspM2 data.

The IpgB2-RhoA crystal structures further elucidate the residues which are important for binding of WxxxE GEFs to Rho GTPases. As shown in Figure 4-10, IpgB2 interacts with precisely the same RhoA residues as the eukaryotic PDZ GEFs. This is despite having a completely different structure, and is an excellent example of how pathogenic bacteria have convergently evolved to subvert their hosts. Sequence alignment of IpgB2 with EspM2 shows that every one of these contact residues is conserved making a similar mechanism of GEF action probable. That is, binding by the EspM2 Q85 and D88 residues reorients switch I of RhoA, leading to a coordination reduction for the associated Mg^{2+} . In a second step, interaction with EspM2 D119 and D73 causes a switch II reorientation, whereupon Mg^{2+} is lost altogether. This paves the way for GDP removal, since Mg^{2+} binds GDP.

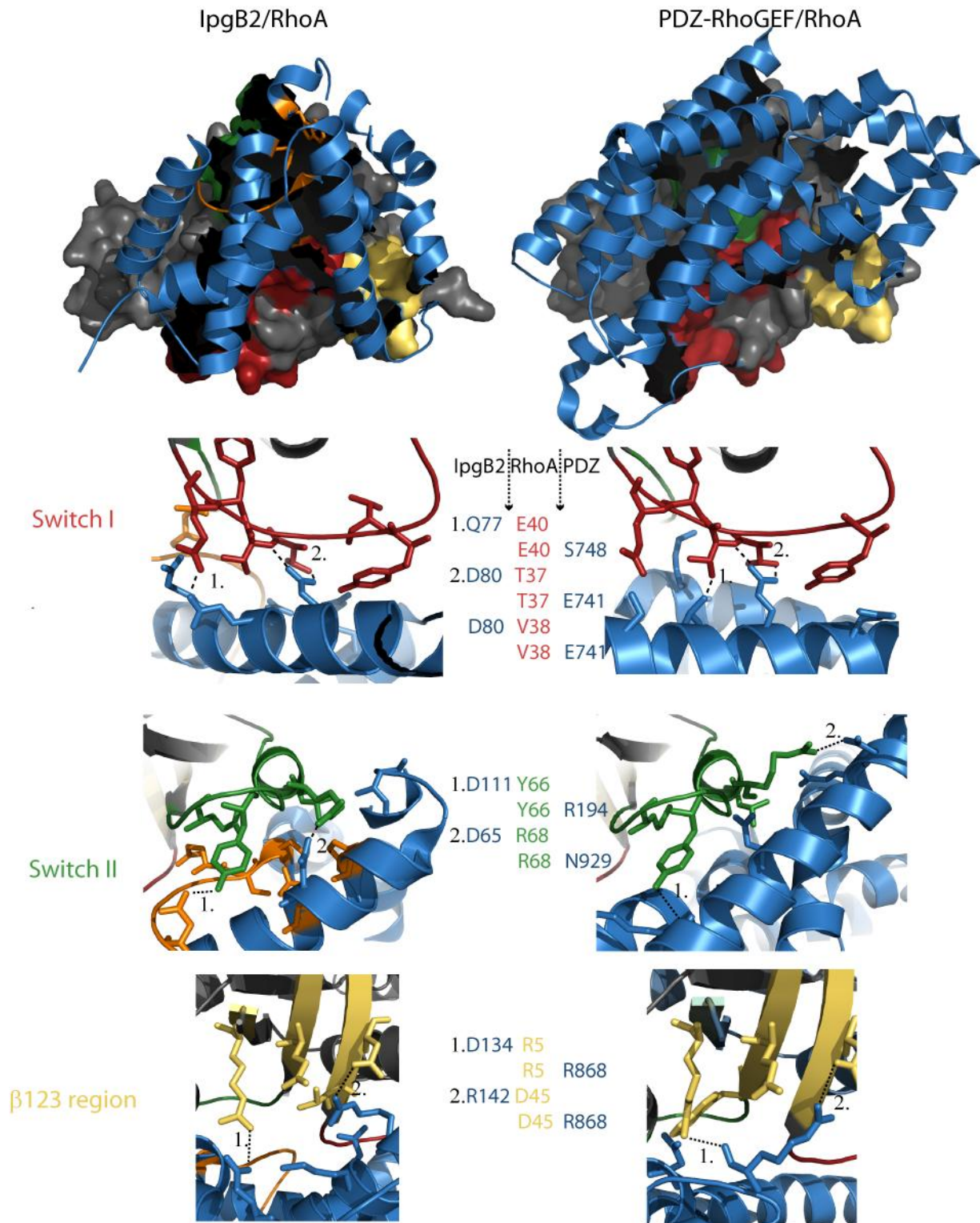


Figure 4-10 A comparison of RhoA binding between IpgB2 and the eukaryotic PDZ-RhoGEF. RhoA is displayed as a grey surface with switch I highlighted in red, switch II in green and the β 123 region in yellow. The WxxxE motif and catalytic loop of IpgB2 are shown in orange. Despite very different overall architecture, the prokaryotic and eukaryotic GEFs (shown in blue, PDB 3LWN and 1XCG respectively) act upon the same RhoA residues, suggesting convergent evolution.

For the preparation of IpgB2 for crystallisation, IpgB2 displayed the precarious solubility seen in other members of the WxxxE group. Klink *et al.* obtained purified IpgB2 upon expression of a His-tagged MBP-IpgB2 fusion, since IpgB2 expressed alone is insoluble. A buffer optimisation was carried out with cleavage of the MBP occurring in each of the conditions of a broad crystallisation screen. The conditions were tested for tag cleavage, IpgB2 solubility and IpgB2 stability after three days. Klink discovered that 100 mM sodium citrate conditions stabilised the otherwise insoluble IpgB2, an additive which was henceforth used in soluble IpgB2 production. This insightful approach led to the crystal structures of IpgB2, but the authors concede that the same approach failed for IpgB1, which remains resolutely insoluble after MBP cleavage.

In this thesis, IpgB1 was made soluble upon coexpression with Spa15. A similar tack was attempted to bring IpgB2 (25% sequence identity with IpgB1) into solution, since Spa15 is also reported to chaperone this effector. However, despite high overexpression, IpgB2 remained insoluble when coexpressed with Spa15, just as it was when solitarily expressed (Figure 4-11).

This result highlights the multifaceted role of Spa15 chaperoning. IpgB1 has been shown to require Spa15 for both its stability and secretion in *S. flexneri*, however IpgB2 requires Spa15 for secretion only: IpgB2 was shown by Hachani *et al.* to be present in the crude extract and not the culture supernatant of a *S. flexneri* Δ *spa15* mutant, whereas IpgB1 was detectable in neither (Hachani *et al.*, 2008). Similarly, the well characterised Spa15 effector IpaA (Hamiaux *et al.*, 2006; Park *et al.*, 2011) is required for secretion only (Page *et al.*, 2002), with viable cytoplasmic storage of stable IpaA shown in a *S. flexneri* Δ *spa15* mutant. Recombinant expression of Spa15 with IpgB2 might therefore be expected to not show an interaction, since an acceptable natural cytoplasmic state of IpgB2 is an uncomplexed one.

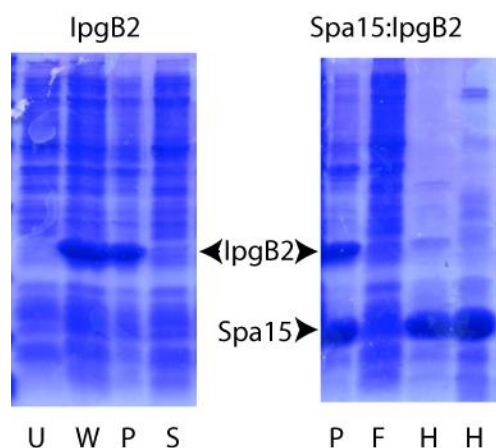


Figure 4-11 The insoluble expression of IpgB2, both with and without Spa15 coexpression. U (uninduced cell), W (whole cell), P (cell pellet), S (cell supernatant), F (Ni-NTA flow through), H (Ni-NTA 'His trap' fractions).

4.4 Conclusions.

A number of functional clarifications have been made over the lifetime of the WxxxE group. When initially collected together, their role was expected to be one of functional mimicry of the GTPases (Alto *et al.*, 2006). Other research carried out on IpgB1 (Ohya *et al.*, 2005) and Map (Berger *et al.*, 2009), however, appeared to suggest a role in activation of the GTPases rather than mimicry. Different members of the group target different GTPases. It is clear that the different invasion mechanisms of *Shigella*, *Salmonella* and EPEC/EHEC require different effector functions – actin polymerisation structures are required to promote *Shigella* internalisation, maintain the integrity of the *Salmonella* containing vacuole and allow tight *E. coli* surface binding. It is currently unknown how activation of particular GTPases aids a specific invasion mechanism. However, the elucidation of WxxxE group structures as well as functional studies make it apparent that use of the conserved domain as a prokaryotic GEF plays an important role in bacterial virulence.

The work of this chapter has enabled the addition of another member to the bacterial GEF list. EspM2 has been shown to bind RhoA and to bring about GDP dissociation and stress fibre formation in HeLa cells. The binding of SifA to RhoA has also been shown.

Another GEF component has been examined, IpgB1, which mimics RhoG in order to bind ELMO. This interacts with Dock180 to act as a GEF for Rac1 (Handa *et al.*, 2007). The stable complex of IpgB1 with its chaperone, Spa15 has been shown in solution, with stoichiometry as expected for a Class 1 chaperone. The N-terminus of IpgB1 is required for successful complexation with Spa15, except for initial disordered residues which are cleavable by trypsin. The roles of Spa15 as chaperone are clearly not identical for each of its effectors, with it being necessary for the solubilisation and stability of IpgB1, but not IpgB2.

5 The Spa15 Chaperoned Effector OspD1 and its Role in Transcriptional Regulation.

Another effector which is chaperoned by Spa15 is OspD1. This effector has no known role in subverting the host, and its secretion appears to serve the purpose of simply relocating the protein away from the bacterial cytoplasm. Here, whilst chaperoned by Spa15, OspD1 binds a transcriptional activator protein, MxiE. The binding of Spa15 to OspD1 therefore has implications in the regulation of the T3SS. In this chapter Spa15 has been expressed with OspD1 and a purification of the complex detailed. This complex has in turn been purified with MxiE, showing the advantages of complexation as a potential purification technique. In order to function as transcriptional activator, MxiE interacts with another protein, IpgC. In this chapter the interaction is shown.

5.1 The regulatory side to Spa15 effectors.

The *Shigella* T3SS effectors have been identified as being in three classes (Le Gall *et al.*, 2005): those under the control of the activator VirB, those under MxiE control, and those under the control of both. The thirteen exclusively MxiE controlled effectors include OspD3, OspG and members of the IpaH family (Kane *et al.*, 2002). These are all secreted after activation of the T3SS, in a second wave of secretion (as detailed in Chapter 1), possibly at the stage of *Shigella* infection where the bacteria is moving to different cells (Le Gall *et al.*, 2005).

MxiE is a member of the AraC family, encoded on the T3SS apparatus operon (Allaoui *et al.*, 1993), with identified counterparts performing similar roles in *Salmonella* (InvF) (Kaniga *et al.*, 1994) and *Yersinia* (VirF) (Cornelis *et al.*, 1998). These transcriptional activators contain a C-terminal DNA binding region and a variable N-terminal domain.

MxiE requires the cytoplasmic chaperone IpgC to activate transcription (Mavris *et al.*, 2002). The two proteins form a complex to enable MxiE to activate its target promoters. To prevent premature DNA binding, MxiE is held in an inactive state until such time as it is required. Parsot postulated a sequence of events carried out by contributing regulatory proteins as denoted in Figure 5-1 (Parsot *et al.*, 2005).

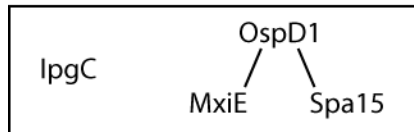
Initially, IpgC binds to the translocators IpaB or IpaC, and MxiE is bound by OspD1. IpaB and IpaC are secreted prior to any T3SS effectors since they form structural components of the translocation pore. After secretion, IpgC is free to bind MxiE. MxiE however is still competitively bound by the Spa15 chaperoned effector OspD1. Only upon secretion of OspD1 is MxiE free to bind IpgC and commence transcription.

By immunoblotting for MxiE in *Shigella* lysate, the binding of OspD1 to MxiE has been observed (Parsot *et al.*, 2005). However, the IpgC-MxiE interaction remained speculative, and attempts to observe a MxiE-IpgC interaction by a yeast two-hybrid experiment failed (Mavris *et al.*, 2002). Parsot also pulled OspD1 out of *Shigella* lysate with Spa15, however it was unknown whether these interactions would be observed in up-scaled *E. coli* recombinant coexpression. The following study shows that this is the case.

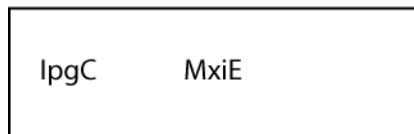
1) Pre-secretion



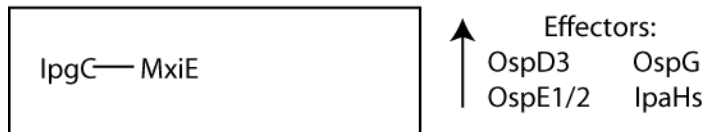
2) Secretion of IpaB and IpaC



3) Secretion of OspD1



4) Binding of IpgC and MxiE activates transcription



5) Following closure of T3SS, neosynthesized OspD1 re-sequesters MxiE

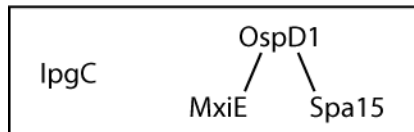


Figure 5-1 Sequence of events leading to transcriptional activation by MxiE. (Figure modified (Parsot *et al.*, 2005)).

5.2 The Spa15:OspD1 complex.

5.2.1 Spa15:OspD1 purification.

Tagless recombinant OspD1 is soluble, in contrast with the Spa15 effectors IpgB1 and IpgB2. For observation of a complex with Spa15, however, N-terminal His-Spa15 was coexpressed with OspD1 (constructs 3.1 and 5.3). The pair were shown to form a soluble complex (Figure 5-2).

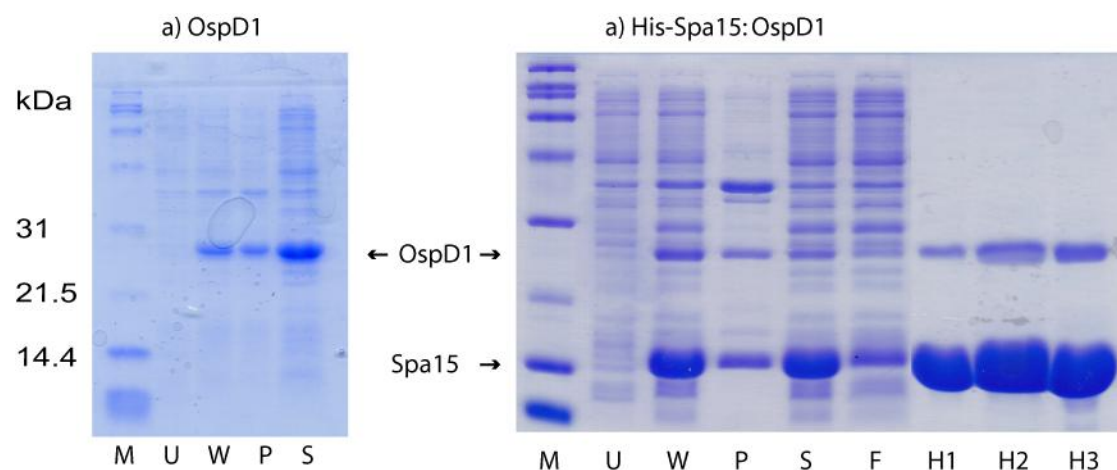


Figure 5-2 OspD1 purification profiles. Left: Solubility of OspD1 (with no affinity tag) upon overexpression. Right: Soluble Spa15:OspD1 complex, using His-Spa15 and untagged OspD1 as the basis for purification. M (markers), U (uninduced cell), W (whole cell), P (cell pellet), S (cell supernatant), F (Ni-NTA flow through), H (Ni-NTA 'His trap' fractions 1, 2, and 3).

A number of species were observed by size exclusion chromatography. Spa15 dimer was generated in large excess. An initial SEC run enabled much of this to be removed. A second SEC column of the useful pooled fractions, shown in Figure 5-3, allowed separation of Spa15:OspD1 aggregates as well as the 2:1 stoichiometric complex and residual Spa15 dimer by-product.

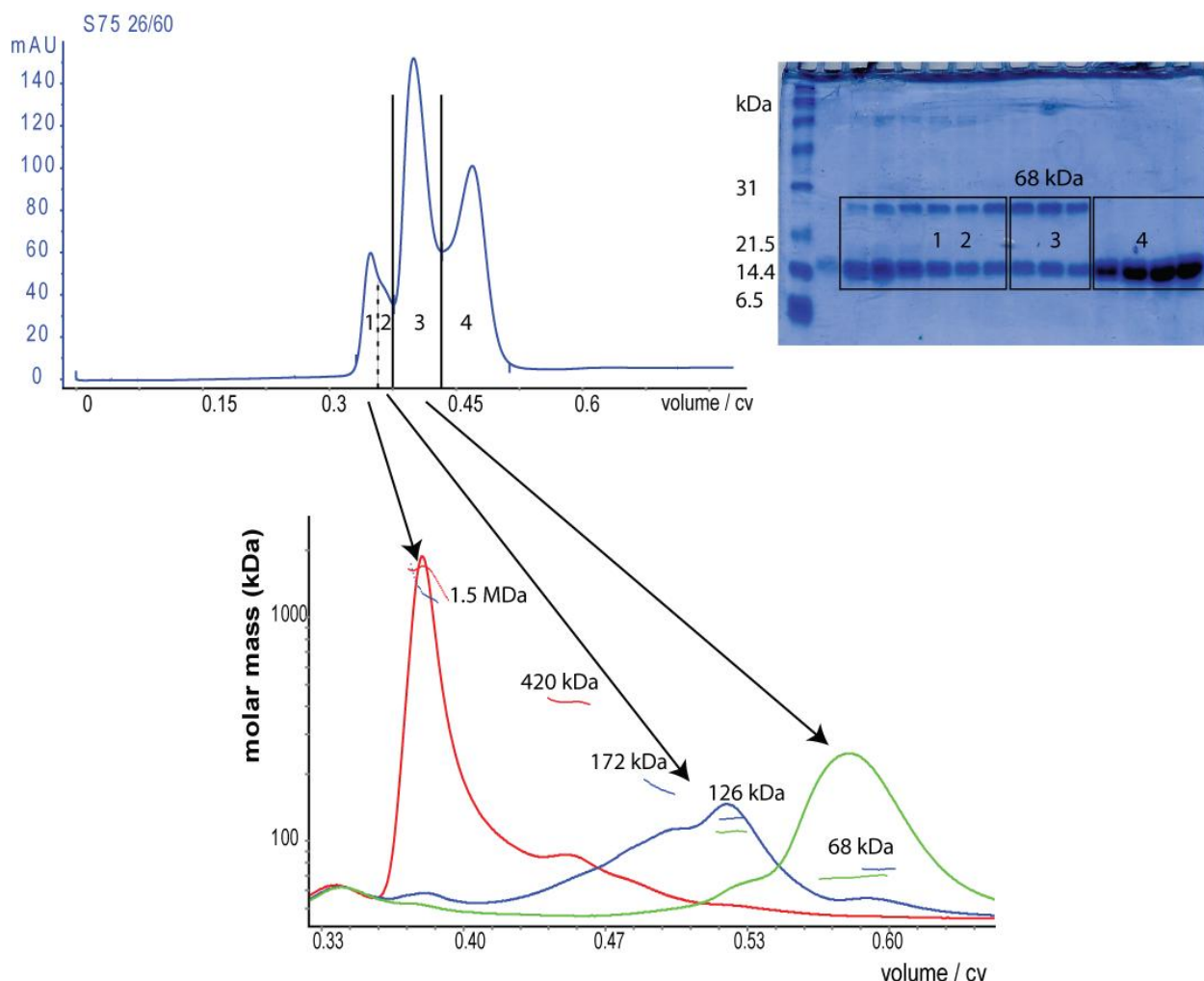


Figure 5-3 Identification of Spa15:OspD1 aggregation. Size exclusion chromatography (S75 26/60), SDS-PAGE and MALLS showing the species present during Spa15:OspD1 purification. Peak 1 (red) and 2 (blue) contains aggregates of the Spa15:OspD1 complex. Peak 3 is the stoichiometric Spa15:OspD1 2:1 complex (green). Peak 4 is a Spa15 dimer byproduct. Size exclusion chromatography for purification was carried out at flowrate 3 ml/min. Samples were pooled and concentrated, before loading onto SEC-MALLS at flowrate 1.6 ml/min. The buffer used was 25 mM Tris pH 7.5, 150 mM NaCl, 10 mM β -MC.

5.2.2 The nature of the complex.

Whilst inclusion of 10 mM β -MC in the Spa15:OspD1 purification reduced the quantities of non-specific aggregation, it proved impossible to prevent. MALLS was used to quantify this tendency to aggregate: peak 3 in Figure 5-3 mostly contained the minimal Spa15:OspD1 complex with stoichiometry 2:1 (68 kDa). Samples taken from peaks 1 and 2 showed a dimer of this minimal complex (4:2 - 126 kDa), complex-trimer (6:4 - 170 kDa) and larger aggregates.

Fractions were pooled from peak 3. A similar partial proteolysis experiment to that of Spa15:IpgB1 (Section 4.2.2) showed that Spa15:OspD1 was much more susceptible to proteolytic cleavage. No stable fragments of OspD1 were apparent from trypsin and subtilisin carlsberg cleavage (Figure 5-4).

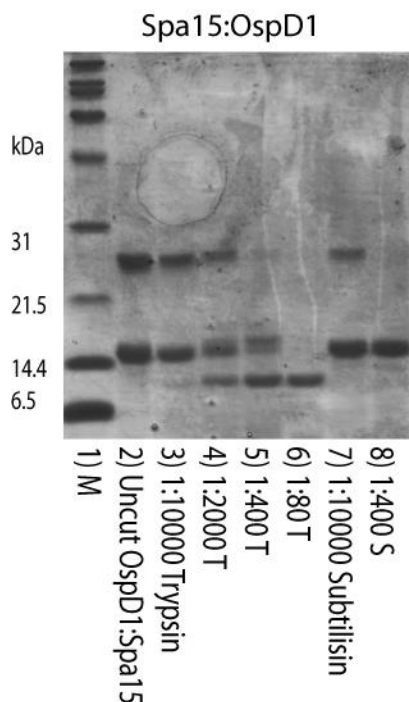


Figure 5-4 Partial proteolysis of Spa15:OspD1 complex. The range shown is 1:80 – 1:10000 protease:protein (w/w). No stable fragments of OspD1 are observable upon incubation with trypsin or subtilisin after 2 hours at room temperature. Spa15 had the first 17 amino acids removed by trypsin similarly to that shown in Figure 4-3.

Crystal trials were set up for Spa15:OspD1 but no crystals were forthcoming. Trials were conducted with protein concentration at 2.4 mg/ml, 21°C in the broad screens MD1&2, JSCG+, Pact Premier, Stura Matrix and Index.

Predicted structural motifs for OspD1 were obtained using the Fold and Function Assignment Server (FFAS). This compares a primary protein sequence for structural homology with structures deposited in the protein databank (PDB) or other databases (Jaroszewski *et al.*, 2005). The NCBI PSI-BLAST alignment is used as a starting basis, where alignment profiles with *E*-value < 0.005 are weighted according to their

similarity with other PDB sequences included in the alignment. The statistical significance is reflected in a score, which is deemed significant if less than -9.5. FFAS sequence alignment of OspD1 with known structures in the PDB shows structural similarity with ankyrin-repeat containing proteins such as Notch proteins (Coleman *et al.*, 2007) (score -52.0) and euchromatin-associated methyltransferases (Collins *et al.*, 2008) (score -51.5). This anti-parallel helix-loop-helix pair motif is well known as a mediator of protein-protein interactions, where the repeats cooperatively fold, stabilised by short-range hydrophobic and other interactions (Mosavi *et al.*, 2004). The motif is certainly consistent with a species whose role is to bind MxiE, with no other apparent function upon secretion. Whilst a β -motif identified for some chaperone binding effectors is not present in OspD1 (Lilic *et al.*, 2006), it is expected that Spa15 binds OspD1 via an N-terminal CBD in the first 100 residues, a region which is thought likely to be disordered by Xtalpred and RONN. Interaction with MxiE could then simultaneously occur (as is demonstrated in the next section), via two of the three repeats of the 44 residue motif (xELLxAxx....xxGxPGLFxALQNGHxDAlxAYGxIL.KxxxLTxExI) located in the C-terminus (Buchrieser *et al.*, 2000).

5.3 The Spa15:OspD1:MxiE complex.

To observe the protein binding abilities of OspD1, it was coexpressed with MxiE. The *mxiE* gene was initially identified as 210 codons, however, this gene was redefined as *mxiEb* (Allaoui *et al.*, 1993). This is because the full length 29.9 kDa MxiE protein observed had greater mass than expected on the basis of this gene (Penno *et al.*, 2005). The N-terminus of MxiE actually derives from an additional 59 codons in a different reading frame to *mxiEb*, termed *mxiEa*. *mxiEa* contains the translation site and *mxiEb* the DNA binding domain. Both are required for MxiE to be a functional activator (Penno *et al.*, 2005), with production dependent on transcriptional slippage.

5.3.1 The frameshifting origins of MxiE.

Although maintenance of the reading frame is a vital part of the information transfer from DNA to protein, there are instances, such as codon redefinition, translational bypassing and frameshifting where the normal rules are not adhered to (Namy *et al.*, 2004). Frameshifting may reposition the ribosome forwards or backwards, before continuing translation in the new reading frame.

One mechanism of frameshifting is transcriptional slippage, where a non-template nucleotide is incorporated into the mRNA. During transcription of *mxiE* the frameshifting (by slippage of RNA polymerase) incorporates an additional nucleotide into the mRNA. The exact mechanism for this slippage is unclear, but comparison with the similar case of *Thermus thermophilus dnaX* indicates that slippage requires both a 5' purine rich sequence and a series of nine Ts (Penno *et al.*, 2005).

The slippage achieves ca. 30% efficiency, raising the question of its purpose. It has been postulated that frameshifting is a regulation tool in this case (Penno and Parsot, 2006): MxiE is not required for entry into epithelial cells, but is a translational regulator for effectors required later in the invasion. Thus, keeping the MxiE concentration initially low provides an indirect control of the premature production of the effectors which are not necessary for invasion. Conditions conducive to transcriptional slippage may increase the amount of MxiE to the system later, as the effectors are required - although this is speculative.

5.3.2 Spa15:OspD1:MxiE purification.

In common with many AraC proteins, MxiE is an insoluble species in an *E. coli* recombinant expression system, and overexpression of the N-terminal His-tag construct 5.7. produced a large quantity of insoluble protein. Similarly, coexpression of this construct with untagged soluble OspD1 (construct 5.3) did not generate soluble MxiE, as shown in Figure 5-5.

However, inclusion of Spa15 provided a resolution. A stable species of His-Spa15:OspD1:MxiE (SOM) was found to be soluble (constructs 5.1 and 5.6). The purification used the Spa15 N-terminal His-tag in immobilised metal affinity chromatography, before removal of excess Spa15 dimer by size exclusion chromatography. The SOM complex was separated from any remaining Spa15, and from by-product Spa15:OspD1 via anion exchange chromatography with a salt gradient from 25 mM – 500 mM NaCl.

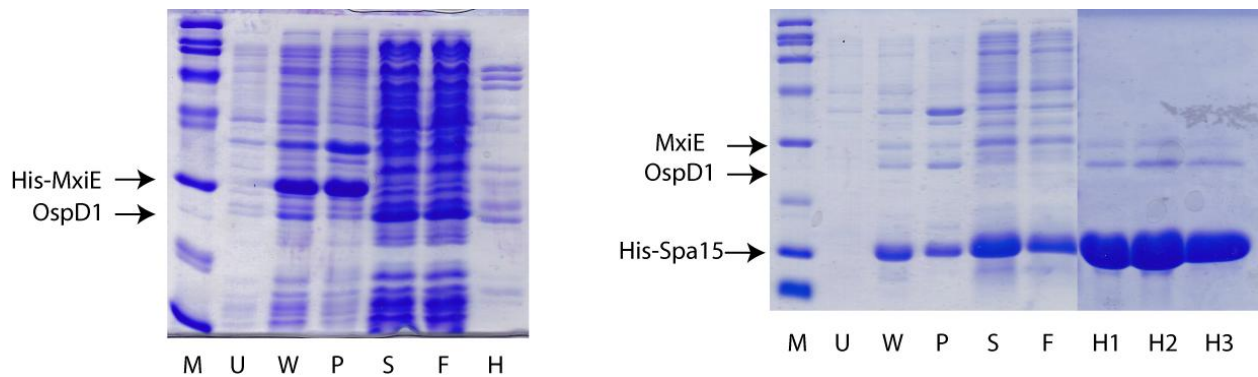


Figure 5-5 Spa15 is required to bring about the soluble binding of recombinant OspD1 by MxiE. **Left:** Unsuccessful purification of His-MxiE:OspD1 complex. **Right:** Successful solubilisation of a His-Spa15:OspD1:MxiE complex. M (markers), U (uninduced cell), W (whole cell), P (cell pellet), S (cell supernatant), F (Ni-NTA flow through), H (Ni-NTA 'His trap' fractions). SDS-PAGE stained with Coomassie Brilliant Blue.

The stoichiometry of the MonoQ separated species was verified by MALLS (Figure 5-6). The SOM complex exists in solution with stoichiometry 2:1:1, and the Spa15:OspD1 and Spa15 species with the same stoichiometry as previously seen. The molecular masses of OspD1 and MxiE were verified by mass spectrometry (Figure 5-6).

Binding of MxiE to OspD1 rather than to Spa15 was verified in a coexpression trial of His-Spa15 and MxiE, in which MxiE remained insoluble, and only Spa15 alone could be purified (data not shown).

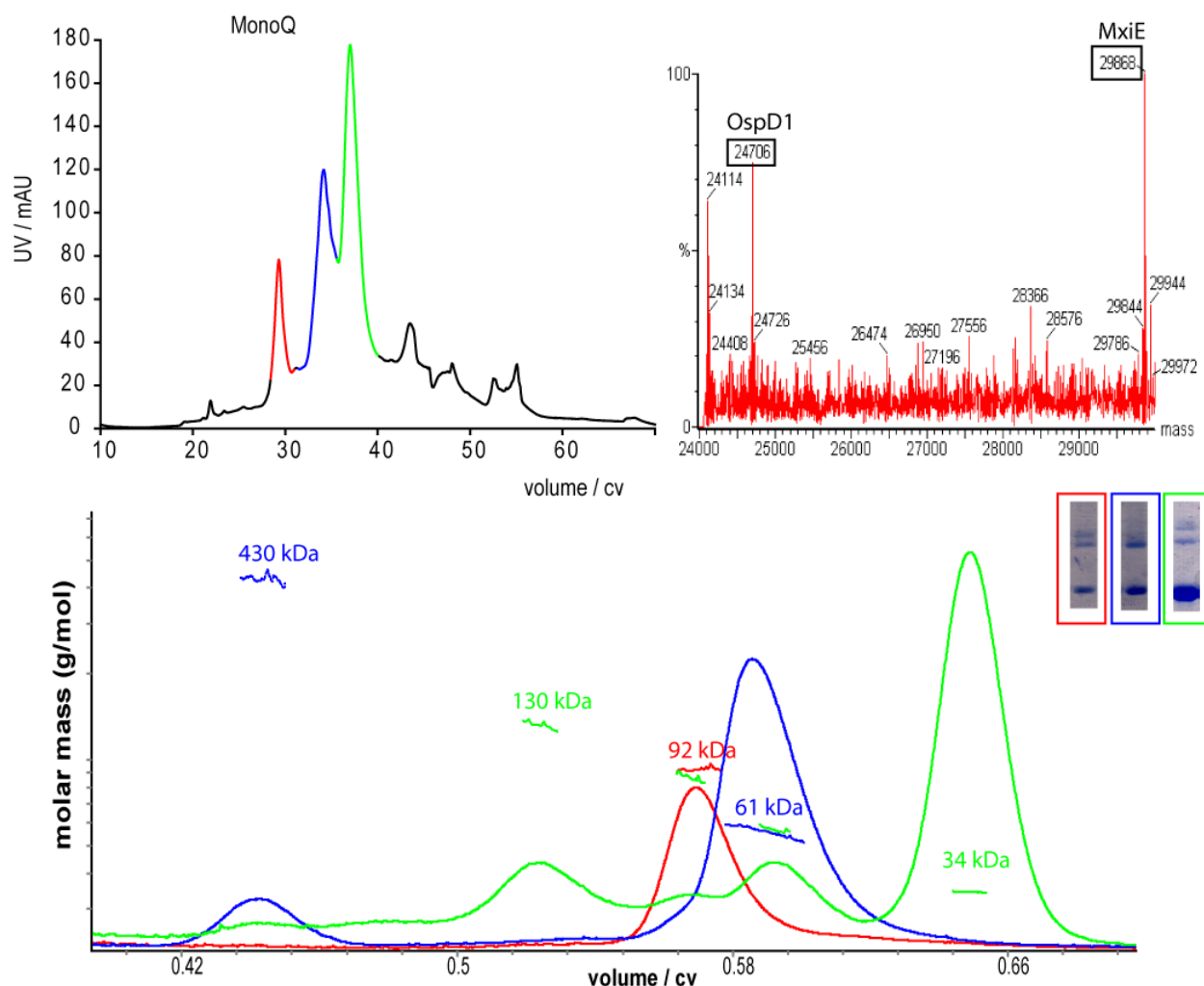


Figure 5-6 Purification of the SOM complex. **Top left:** Anion exchange chromatography purification of SOM, between an NaCl gradient of 25 mM to 500 mM at pH 7.5. **Bottom:** The separated peaks were identified by MALLS and SDS-PAGE. The red peak species had a single mass of 92 kDa, corresponding to SOM stoichiometry 2:1:1. All three species were seen in the SDS-PAGE. The blue peak had a predominant mass of 61 kDa, corresponding to the Spa15:OspD1 complex with stoichiometry 2:1. Again this was verified by SDS-PAGE. The green peak has a predominant mass of 34 kDa, the dimeric weight of Spa15, and the SDS-PAGE showed the presence of Spa15 in this peak. **Top right:** The SOM fractions were also characterised by mass spectrometry. OspD1 and MxiE are seen minus N-terminal Met. The Spa15 weight was assumed based upon previous samples. Masses (minus N-terminal Met for OspD1 and MxiE): His-Spa15: 17280 Da, OspD1: 24706 Da, MxiE: 29870 Da.

5.3.3 Increasing the yield of the complex.

Optimisation of the protocol was carried out to increase the relative quantity of SOM to Spa15 by-product. After recloning into a number of plasmid combinations the best results were achieved by placing Spa15 into a lower copy number plasmid (pET-15b (40 copies/cell) → pACYC (10 copies/cell)) and

targeting purification of the SOM complex by N-terminally His-tagging MxiE into a high copy number plasmid (pET-28b (40 copies/cell)).

Separation of the SOM from Spa15 and SO species was unnecessary with this tag combination. However, the more highly expressed MxiE showed an ability to oligomerise. Following affinity purification, size exclusion chromatography revealed two peaks. The masses of these peaks were determined by MALLS; 90 kDa and 150 kDa are consistent with S:O:M species of stoichiometry 2:1:1 and 2:1:3 respectively.

Support for the 2:1:3 stoichiometry species came from SDS-PAGE: in addition to the unique Spa15, OspD1 and MxiE protein bands, a high molecular weight species was seen on reducing SDS-PAGE gels accompanying the S200 left shoulder (Figure 5-7). This species was identified as MxiE by proteomic mass spectrometry analysis using a Mascot search (Dr Ben Thomas, Central Proteomics Facility, University of Oxford). The SDS-PAGE band is of size 66 kDa, indicating a MxiE dimer which is clearly resistant to degradation under the harsh conditions of a reducing SDS-PAGE. The complex thus appears to be an expected 2:1:1 SOM complex, with an additional two MxiE copies bound.

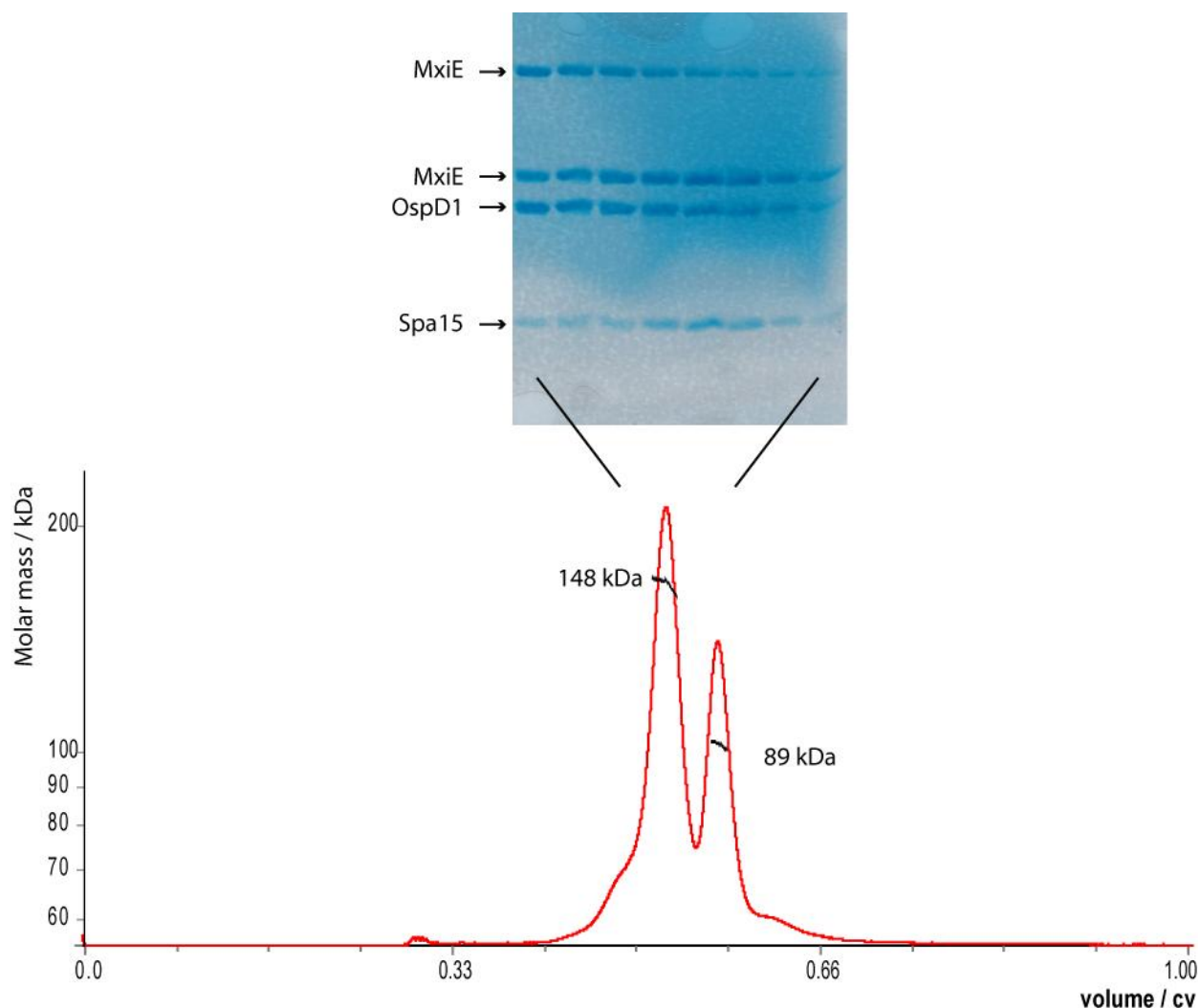


Figure 5-7 Oligomeric MxiE. SDS-PAGE and SEC-MALLS showing the presence of a higher MxiE oligomeric state in addition to that seen in the stoichiometric complex. The left-hand peak shows a greater amount of this oligomer. Running conditions: Flowrate 0.4 ml/min, 25 mM Tris pH 7.5, 150 mM NaCl, 10 mM β -MC.

5.3.4 Crystal trials.

Fractions of the SOM 2:1:1 complex were concentrated to 1.62 mg/ml. Crystal trials were set up in JCSG+, Stura Macrosol and ProPlex screens at 12°C and 21°C. Identical trials containing in-drop subtilisin carlsberg (1:10000 w/w) were also set up.

After one month, small crystals appeared (Figure 5-8) in subtilisin containing ProPlex (Molecular Dimensions) conditions B5 (0.1 M sodium hepes pH 7.5, 10% PEG 4000) and E8 (0.1 M MOPS pH 7.5, 0.1

M magnesium acetate, 12% PEG 8000). These conditions were screened more closely via the optimisation screen shown in Figure 5-8 and the crystals were reproduced.

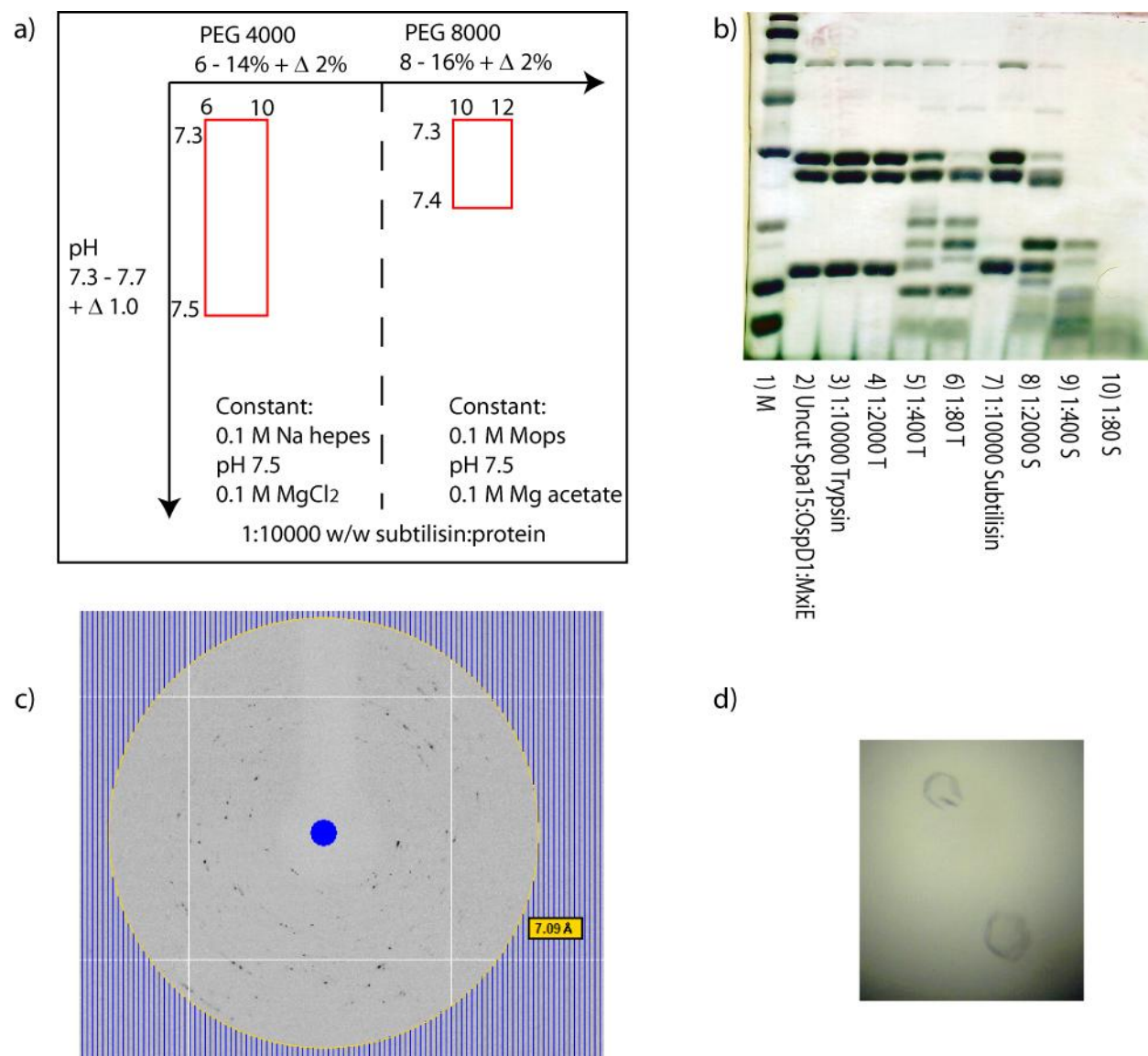


Figure 5-8 Growth and morphology of Spa15:OspD1:MxiE crystals. **a)** Optimisation screen constructed on the basis of crystals obtained in ProPlex B5 and E8 conditions. pH is systematically varied with PEG concentration. Red boxes indicate regions of crystal growth. **b)** Partial proteolysis of Spa15:OspD1:MxiE complex in range 1:80 – 1:10000 protease:protein (w/w). No stable fragments of OspD1 or MxiE are observable upon incubation with trypsin or subtilisin after 2 hours at room temperature. **c)** Diffraction image showing low level diffraction. **d)** Crystal shape - the multiple diffraction pattern recorded was in accordance with multiple looking crystals.

An attempt to assess the protease stability of the complex via partial proteolysis is shown in Figure 5-8. Similarly to Spa15:OspD1, no stable domains of MxiE or OspD1 became apparent upon proteolysis with trypsin or subtilisin.

The crystals were exposed to X-rays at ID23-1. The spots show a very multiple diffraction image with maximum resolution 7Å. A tentative indexing was given by iMosflm by setting $I/\sigma(I)$ to 5, with cell dimensions 59.0Å, 78.7Å, 100.9Å, 90°, 101°, 90° in C2. Phaser, however, failed to find a solution using Spa15 as a search model.

A repeated data collection at a later date with newly grown crystals failed to index. No other crystals were obtained from this complex.

5.3.5 Separation of MxiE.

To obtain MxiE alone for interaction studies with its putative transcriptional binding partner IpgC, a number of conditions were tested to separate MxiE from the Spa15:OspD1:MxiE complex:

- Ionic Strength: as for the Spa15:IpgB1 complex, the SOM complex was shown in anion exchange chromatography to be stable over the NaCl range 25 mM to 500 mM.
- Detergents: addition of 0.1% - 1% Triton X-100 and Tween-20 to a 1 mg/ml solution of SOM did not disrupt the complex. Some precipitation occurred in the Triton samples, this precipitate was shown by SDS-PAGE to be equally distributed amongst the species.
- pH: a reduction in pH below pH 6 led to complete precipitation of the complex. At conditions more alkaline than pH 10, fragmentation of the proteins was observed.

A composite approach of addition of 0.1% Tween-20 and alkaline pH were used to successfully remove MxiE from the Spa15 and OspD1 components. The bacterial cell supernatant was loaded onto a Ni-NTA column. The column was washed with 40 column volumes of buffer A (25 mM Tris pH 7.5, 500 mM NaCl,

10 mM Imidazole) containing 0.1% Tween-20 at pH 11.5. The flow was stopped and the loaded column incubated for 2 hours at 4°C. A further 12 column volumes of buffer A were flowed through before imidazole elution. The protein immediately underwent size exclusion chromatography at pH 7.5. Two peaks were observed corresponding to MxiE, and MxiE:OspD1, as shown by SDS-PAGE (Figure 5-9).

The purification of soluble MxiE is therefore achieved via the indirect route of chaperone complex purification, followed by removal. The SEC fractions were collected, and used in MxiE-IpgC interaction studies, described in section 5.5.

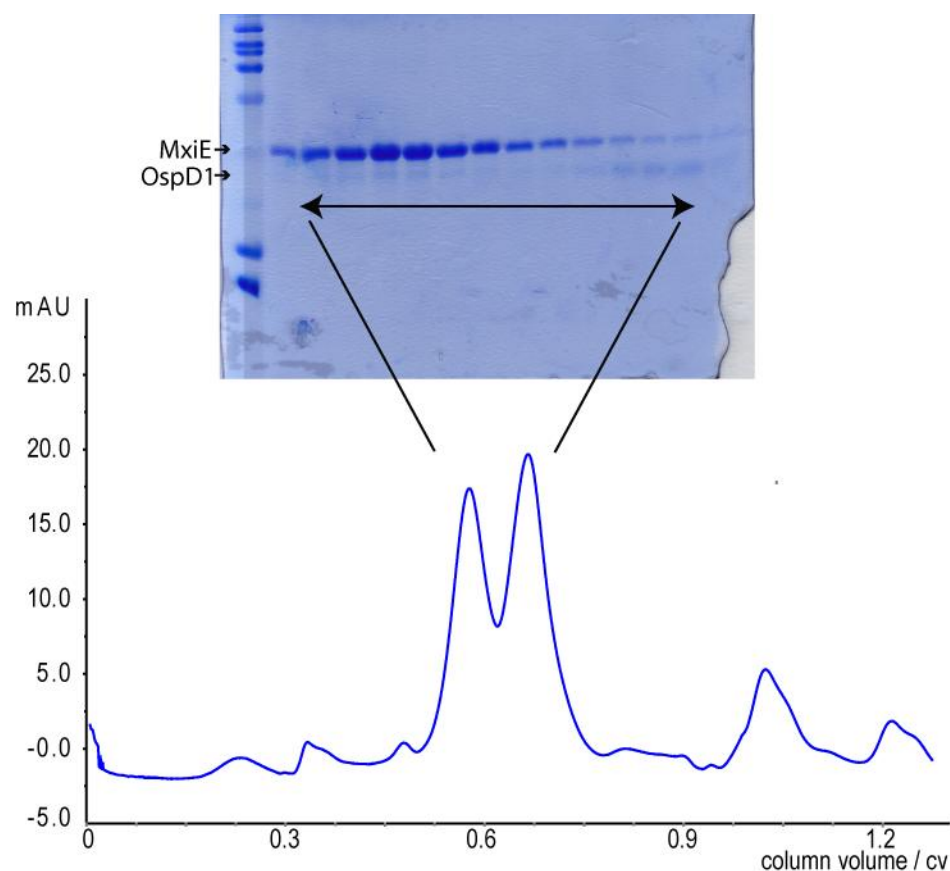


Figure 5-9 Separation of free MxiE from remaining MxiE:OspD1. Peaks from the size exclusion chromatographic separation were identified by SDS-PAGE. S200 26/60 column running conditions: 5ml sample loaded at flowrate 2ml/min in buffer 25 mM Tris pH 7.5, 150 mM NaCl, 10 mM β -MC, 0.1% Tween-20.

5.4 The co-anti-activators IpaB and IpaC.

IpgC was expressed for interaction studies with MxiE. As discussed at the beginning of the chapter and in Chapter 1, IpgC (18 kDa) is a Class II chaperone which chaperones two translocator proteins; IpaB (62 kDa) and IpaC (39 kDa). These translocators associate with the host membrane, forming a pore complex that allows translocation of the other effectors (Epler *et al.*, 2009; Olive *et al.*, 2007).

As an aside for this project, both IpaB:IpgC and IpaC:IpgC complexes were purified. Coexpression of His-tagged IpgC (pET-15b construct 5.9) and IpaB (pACYC-duet construct 5.11) was carried out in Tuner cells. The analogue for IpaC (construct 5.10) was also coexpressed. An excess of IpgC was generated when compared to the respective IpaB:IpgC and IpaC:IpgC complexes, which was used in MxiE interaction studies (see below).

Separation of the complex from IpgC was achieved by size exclusion chromatography. This resolved IpaB:IpgC from IpgC, however in the case of the smaller IpaC:IpgC complex, separation required two S200 runs: the IpaC:IpgC shoulder peak of a size exclusion run identified by a box in Figure 5-10 was pooled and rerun. The stoichiometry was verified by MALLS as dimeric for IpgC, and 1:1 for the complexes; consistently with the literature (Birket *et al.*, 2007; Lunelli *et al.*, 2009).

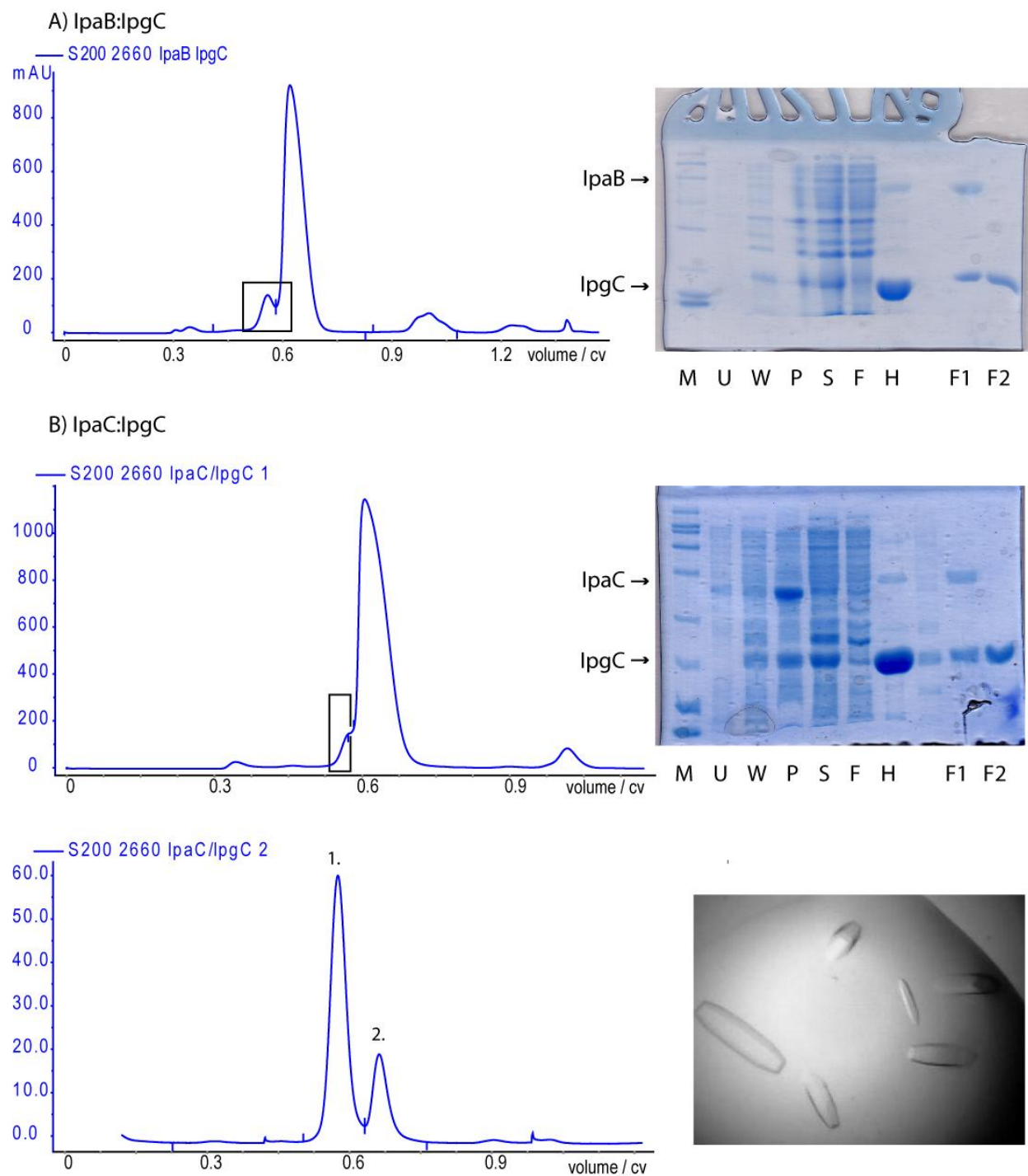


Figure 5-10 Purification of IpaB:IpgC and IpaC:IpgC complexes. **A)** S200 26/60 separation of IpaB:IpgC (boxed) from IpgC dimer. SDS-Page shows the purity of the sample obtained. **B)** Separation of IpaC:IpgC required two S200 26/60 runs, due to the similarity in size between the species and the excess quantity of IpgC obtained. Fractions in the boxed region from the first S200 run were pooled and re-run. Sample purity is shown by SDS-PAGE. The morphology of the crystals is shown. SDS running conditions: Flowrate 0.4 ml/min. The buffer used was 25 mM Tris pH 7.5 150 mM NaCl. SDS-PAGE: M (markers), U (uninduced cell), W (whole cell), P (cell pellet), S (cell supernatant), F (Ni-NTA flow through) H (Ni-NTA 'His trap' fractions), F1 (purified IpaB/C:IpgC complex), F2 (IpgC dimer).

Crystal trials were set up. For IpaB:IpgC, 11.7 mg/ml protein (assuming an absorbance 0.1% (optical density) of 0.350) was added to 1M Trimethylamine N-oxide. IpaC:IpgC was plated at 7.6 mg/ml (assuming an absorbance 0.1% of 0.398). Both complexes underwent identical trials at 21°C with drops containing 1:10000 protease:protein (w/w) in-drop subtilisin in the MD1&2 and JCSG+ broad screens.

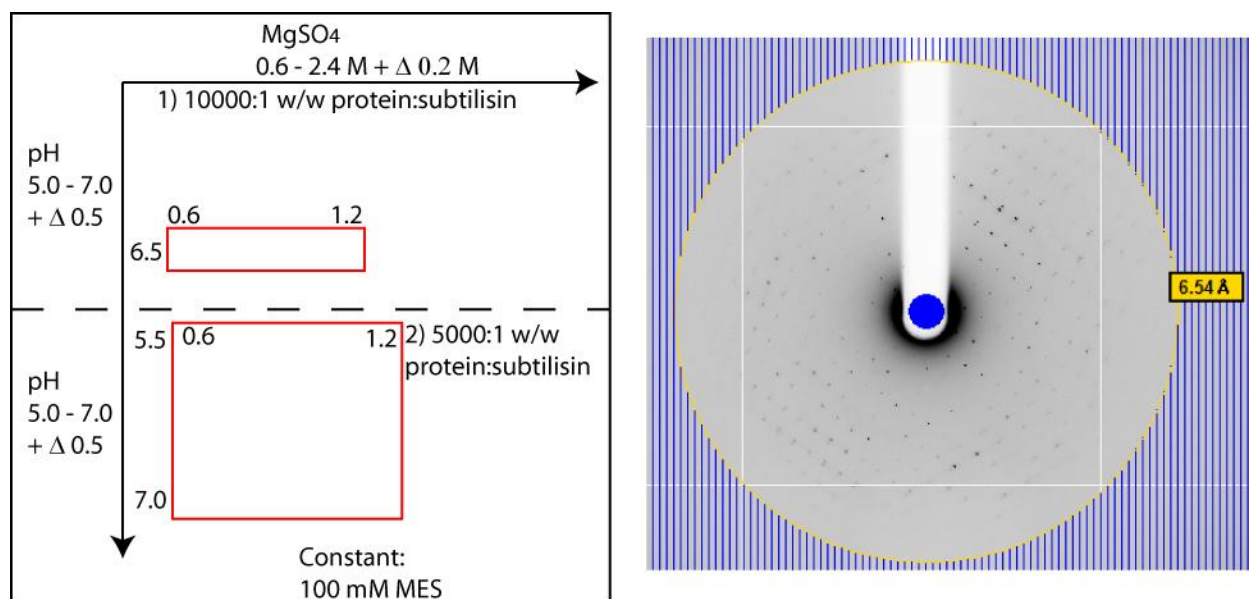


Figure 5-11 IpaC:IpgC crystals. **Left:** Optimisation screen. This was constructed based around hits achieved in JCSG+ Condition E9. pH and magnesium sulphate concentration were systematically screened, with differing amounts of in-drop subtilisin. Red boxes indicate regions of crystal growth. **Right:** Diffraction image showing low resolution diffraction.

For IpaC:IpgC, after three weeks, large crystals were generated in the JCSG+ E9 condition – 0.1 M MES pH 6.5, 1.6 M magnesium sulphate. These were optimised to 250 μ m in size using the conditions shown in Figure 5-11. Given the proteolytic nature of the sample conditions, the crystals had a limited lifetime, visibly degrading within three weeks of growing.

However, the crystal beauty did not correlate with diffraction ability (Figure 5-11). Crystals cryoprotected in 15% glycerol diffracted to a maximum resolution of 6.5 Å. Other potential cryos (including formate) were not tested. Mosflm indexed the crystals as primitive hexagonal, with cell edges (100 Å, 100 Å, 374 Å, 90°, 90°, 119°).

A different morphology crystal which appeared in the same JCSG+ E9 condition in the IpaB:IpgC tray after 10 months diffracted to 7Å, with cell dimensions (224 Å, 206 Å, 51 Å, 90°, 98°, 90°) in C2.

After confirmation by SDS-PAGE that these crystals contained only IpgC, the decision was made to suspend this as a crystallisation target since the structure of IpgC, and IpgC with short IpaB fragments had already been determined (Barta et al., 2010; Lokareddy et al., 2010; Lunelli et al., 2009). However, the interaction of IpgC with MxiE remained an observation of interest.

5.5 Binding MxiE and IpgC.

In an attempt to create a complex of the activated transcriptional activator, MxiE and His-IpgC (constructs 5.8 and 5.9) were coexpressed in B834(DE3), however no interaction could be seen, with purification of His-IpgC only (Figure 5-12).

Furthermore, to see whether IpgC could sequester MxiE from OspD1, a competition experiment between SOM and IpgC was carried out: SOM (50 µl at 5 µM) and (excess) IpgC (50 µl at 25 µM) were incubated overnight and passed through an analytical size exclusion column together. The presence of any IpgC-MxiE binding was searched for in the chromatographic traces and SDS-PAGE. Figure 5-12 shows that the individual components were unchanged by this overnight incubation, and excess IpgC was unable to bind MxiE in the presence of OspD1. It would be unexpected to observe IpgC:MxiE binding in these conditions, even if the OspD1:MxiE binding were dynamic, due to the physiological role of OspD1 being the obstructor of an MxiE-IpgC interaction.

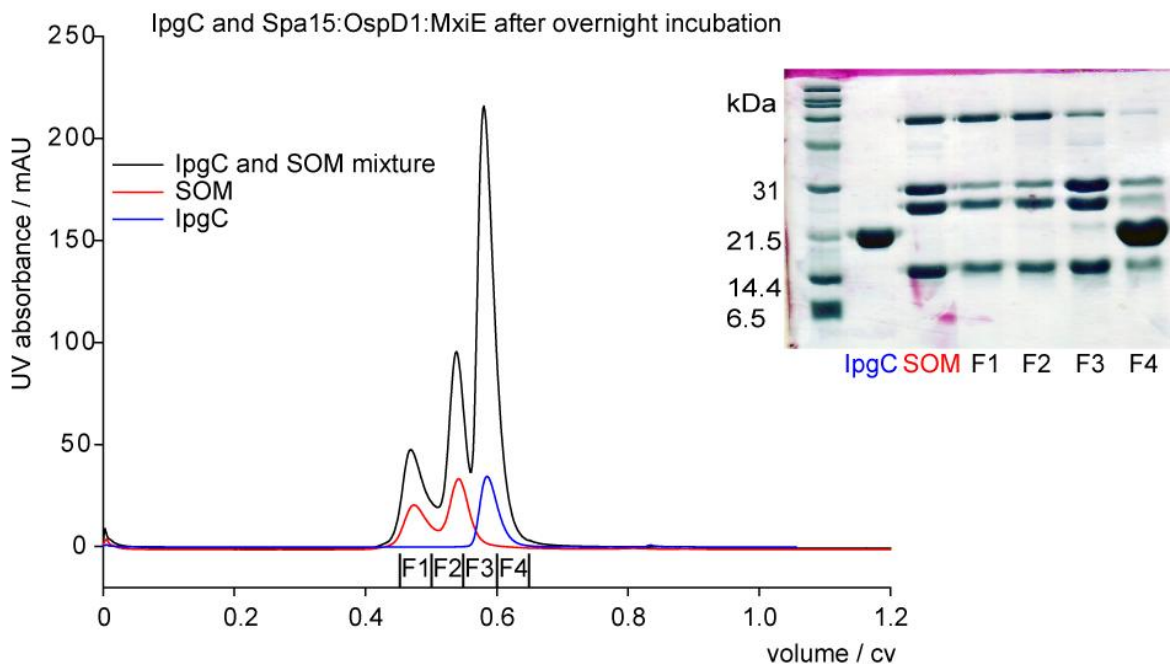
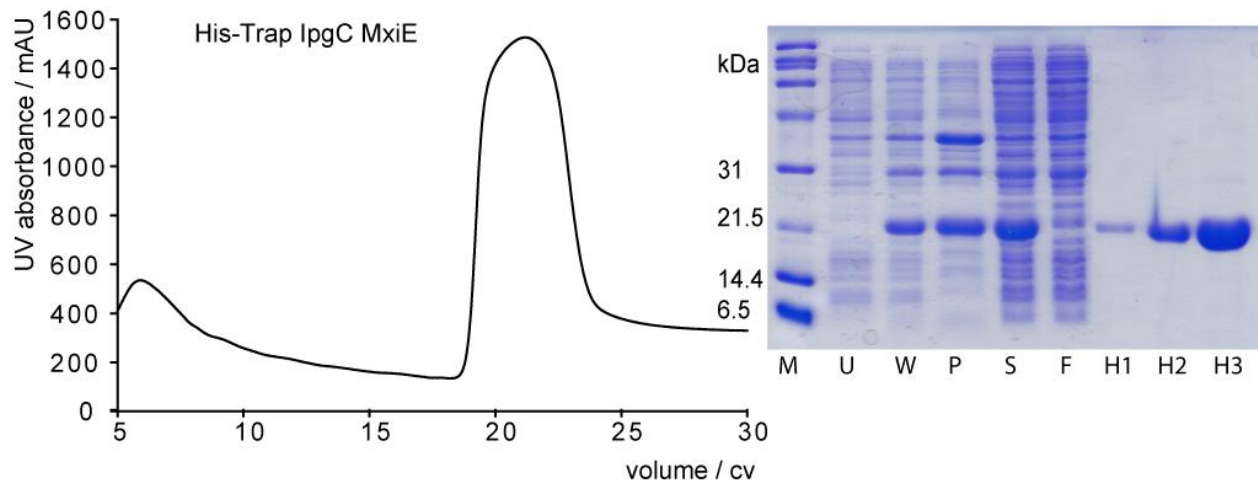


Figure 5-12 IpgC:SOM will not coexpress, and IpgC cannot sequester MxiE from SOM. **Top:** Coexpression of His-IpgC with MxiE, IMAC trace and purification SDS-PAGE shown - no interaction is observed between the two proteins. SDS-PAGE: M (markers), U (uninduced cell), W (whole cell), P (cell pellet), S (cell supernatant), F (Ni-NTA flow through), H1-3 (Ni-NTA 'His trap' fractions). Column running conditions: His-trap 5ml: Flowrate 3 ml/min, Buffer 25 mM Tris pH 7.5, 150 mM NaCl, 10 – 750 mM Imidazole, 10 mM β -MC. **Bottom:** After overnight incubation of IpgC and the Spa15:OspD1:MxiE complex, no interaction could be seen between IpgC and MxiE. The S200 trace after this incubation (black) shows no change in species compared to the species present (SOM - red, IpgC - blue) prior to incubation. Column running conditions: S200 10/300 Flowrate 0.5 ml/min, Buffer 25 mM Tris pH 7.5, 150 mM NaCl, 10 mM β -MC. The SDS-PAGE shows the identity of the peaks from the SEC trace.

However, an interaction was detected between MxiE and IpgC when the separate species were incubated together. MxiE separated from Spa15:OspD1 (as detailed in section 5.3.5) in 0.1% Tween-20 (50 μ l at 10 μ M) was incubated overnight with an excess of His-IpgC (50 μ l at 50 μ M) before analytical size exclusion chromatography. Figure 5-13 shows the resulting trace. The greater magnitude peak is shown by SDS-PAGE to be IpgC alone, and the smaller to contain MxiE and IpgC species. This result is in agreement with the demonstration of a MxiE:IpgC interaction with MBP-MxiE (Pilonieta and Munson, 2008). However, whilst large fusion proteins can overcome protein insolubility, this result provides reassurance that the interaction is specific to the proteins alone.

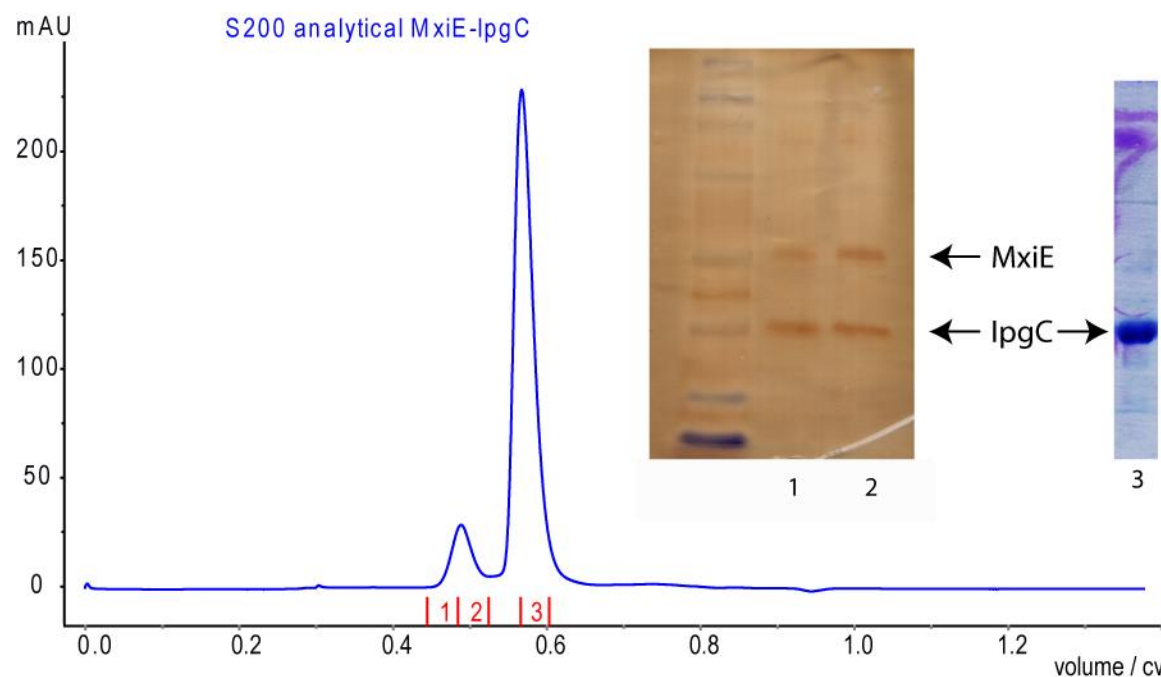


Figure 5-13 The interaction of MxiE with IpgC. S200 10/300 peaks were identified via SDS-PAGE (silver stained and Coomassie Brilliant Blue) shows presence of the bound MxiE:IpgC species in fractions 1 and 2, and excess IpgC in fraction 2. Flowrate 0.5 ml/min, Buffer 25 mM Tris pH 7.5, 150 mM NaCl, 10mM β -MC, 0.1% Tween-20.

Whilst a pH 12, 0.1% Tween-20 environment is clearly not a physiological environment in which to isolate MxiE from Spa15:OspD1, a pH trigger could free it *in vivo* to enable the binding to IpgC. MxiE is inactivated rather than chaperoned by OspD1. Whereas it is thought that the energy for chaperone release comes from the ATPase Spa47 (Akeda and Galan, 2005; Wilharm *et al.*, 2007), release of MxiE

from OspD1 does not necessarily come from the same response. An example of pH control has been shown in the tightly controlled *Salmonella* SPI2 T3SS machinery (Yu *et al.*, 2010). Here, effector secretion is brought about by the breakdown of an SsaL:SsaM:SpaC complex in an increased pH environment (growth conditions of pH 5.0, expression at pH 7.2). SsaL is the MxiC homologue in *Shigella*, which also regulates effector secretion (Martinez-Argudo and Blocker, 2010). Expression of effectors could also be a pH dependent process.

5.6 Conclusions.

In this chapter the complex of Spa15 with its effector OspD1 has been purified. This is another example where the purification of an effector is achieved by binding with a chaperone, when overexpression with the directly tagged protein is difficult. In the case of OspD1, no overexpression was seen for a His-tagged construct (construct 5.5). OspD1 binds Spa15 with a different region to that used in its function, meaning that when chaperoned OspD1 can concurrently bind its substrate. This uses its ankyrin like C-terminus, binding the AraC member transcriptional activator MxiE. The interaction prevents premature transcription of the late T3SS effectors, and is strong enough to be maintained in a purification procedure. Furthermore, the binding of MxiE to OspD1 prevents the binding of the activating partner IpgC. This is itself a chaperone of two translocator proteins (IpaB and IpaC), which must itself be freed before it may bind MxiE. Upon secretion of IpaB, IpaC, and OspD1, the interaction between MxiE and OspD1 may occur, an interaction which has been shown by size exclusion chromatography. This work provides results which confirms earlier speculation that IpgC and MxiE do indeed bind (Parsot *et al.*, 2005), and that the interactions of OspD1 and Spa15 with MxiE do serve to prevent this interaction until it is required.

6 *Bacillus subtilis* PhoD: An Alkaline Phosphatase/Phosphodiesterase Produced under Phosphate Starvation by the Tat System.

The Tat pathway, as detailed in the introduction, is significantly different from the T3SS. This mode of protein secretion transports fully folded proteins, complete with their cofactors (Berks *et al.*, 2005), and prevents the export of incorrectly folded proteins or apoproteins (DeLisa *et al.*, 2003; Matos *et al.*, 2008; Robinson *et al.*, 2011). For this reason, a number of iron-sulphur coordinating proteins are exported via this pathway, including Reiske proteins (Aldridge *et al.*, 2008), and examples involved in anaerobic bacterial metabolism, such as *E. coli* NrfC (Tullman-Ercek *et al.*, 2007). This chapter describes structural insights into a further two Fe-S coordinating proteins which both utilise the system for translocation. These are PhoD, an alkaline phosphatase/phosphodiesterase of *Bacillus subtilis* whose structure has been determined and which shall be described in detail, and PhoX, an alkaline phosphatase of *Pseudomonas fluorescens*, whose catalytic site was characterised alongside the study of PhoD.

PhoD is expressed as a response to phosphate deficiency in *B. subtilis*. The chapter begins with an overview of the environment in which it is expressed, before proceeding to its structural determination. Subsequently, the protein fold and active site shall be viewed. A detailed examination of the active site can allow mechanistic detail to be inferred, and can possibly explain why the relatively uncommon Tat method of translocation is employed for PhoD. This is done through comparison with other examples, notably the alkaline phosphatase PhoX. Finally an examination into the putative function of PhoD is conducted.

6.1 Stress mechanisms of *B. subtilis*.

Survival of the soil dwelling bacterium *B. subtilis* depends on its ability to weather changes in conditions and nutrient levels. For this reason, the bacterium is highly developed to adapt to these changes and

many different regulons may be induced by a particular stress or starvation stimulant (Marles-Wright and Lewis, 2007). The most comprehensive of responses is through induction of the general stress proteins (gsps), under the general stress regulon, σ^B (Hecker *et al.*, 2007). This provides a response to a host of maladies, including heat, osmotic stress and alkaline or acidic treatment. In total, there are 150 regulon members, expression of which creates a considerable burden on the bacterium, and so σ^B is tightly silenced during growth conditions, only being induced when necessary. An example of a more specific response is the induction of the Pho regulon, which fulfils the requirement of a bacterial response to conditions of phosphate starvation (Antelmann *et al.*, 2000; Hulett, 1996; Lamarche *et al.*, 2008; Vershinina and Znamenskaya, 2002).

6.1.1 The Pho regulon: phosphate deficient response.

Unlike the general stress regulon, the Pho regulon is constitutively expressed in *B. subtilis*. However, in conditions of < 0.08 M phosphate, its activity increases by three-fold (Pragai *et al.*, 2004). The regulon is under the control of three systems – PhoP/R, ResD/E and Spo0A. The first two are positive regulators, whereas Spo0A downregulates induction. Spo0A wields the ultimate control, regulating genes expressed in the post-exponential growth phase, including those responsible for spore formation. The transcriptional regulators which bind the Pho regulon promoters are PhoP in Gram-positive bacteria, and PhoB in Gram-negative. In both cases, these are phosphorylated by the histidine kinase PhoR, a transmembrane protein which appears to sense the phosphate concentration. This leads to expression of different groups of proteins which aim to provide sufficient phosphate to the bacteria to prevent the necessity for sporulation (Antelmann *et al.*, 2000). Three protein groups are illustrated in Figure 6-1.

The first of these groups contains three alkaline phosphatases (PhoA, PhoB and PhoD); these are degradative enzymes which scavenge phosphates from the environment (Vershinina and Znamenskaya, 2002).

Secondly, the Pho regulon downregulates TagA/TuaD and upregulates TuaA. These are synthases of the cell wall teichoic acids (former) and teichuronic acid (latter). Teichoic acids (Armstrong *et al.*, 1959) are anionic polymers of glycerol or ribitol phosphate, widely found in Gram-positive bacteria, which are anchored to the peptidoglycan (wall teichoic acid) or cell membrane (lipo teichoic acid) and can make up around 50% of the wall (Hancock, 1997; Lovering *et al.*, 2010). Their composition varies between species (*B. subtilis*' are glycerol phosphate polymers with D-alanyl ester substituents (Neuhaus and Baddiley, 2003)) and whether they are lipo or wall teichoic acids. In times of phosphate starvation, teichoic acids are replaced by the non phosphate containing substitute teichuronic acid, a process under control of the Pho regulon (Liu *et al.*, 1998; Liu and Hulett, 1998).

Thirdly, The Pho regulon leads to production of a phosphate specific transport system for efficient movement of the phosphate. The Pst system allows enhanced uptake of inorganic phosphate from the extracellular milieu (Qi *et al.*, 1997).

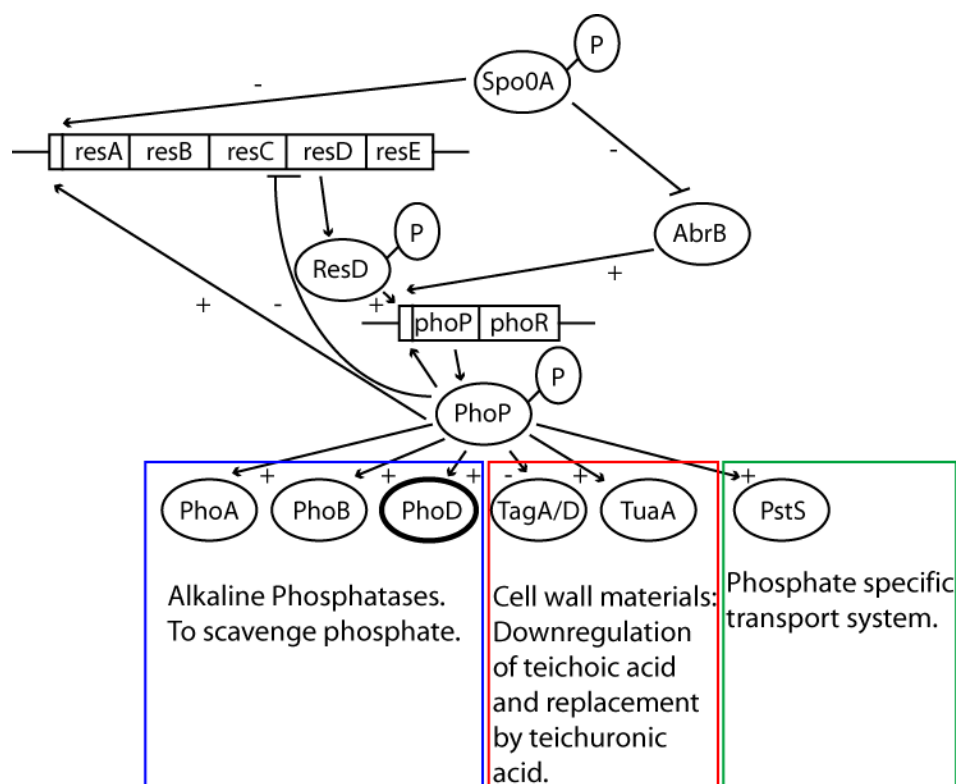


Figure 6-1 Regulation of the Pho regulon. (Figure modified (Vershina and Znamenskaya, 2002)).

The alkaline phosphatase PhoD was selected to be the target for this structural study. The literature reports that PhoD exhibits alkaline phosphatase: alkaline phosphodiesterase activity with ratio 1.0:1.4 at its optimum pH 9.5 (Yamane and Maruo, 1978a, b). Of the PhoP/R regulated proteins, it is the only one which is exported by the *B. subtilis* Tat pathway (Jongbloed *et al.*, 2000). This initially distinguished it from the other alkaline phosphatases expressed, PhoA and PhoB, which share 59.7% sequence identity and exhibit mono-esterase activity only (Hulett *et al.*, 1990; Hulett *et al.*, 1991). Indeed, PhoD is the only known protein exported via the TatAdCd system (Jongbloed *et al.*, 2000). For this reason, PhoD has been used as a model to examine Tat export in *B. subtilis*. It has been used to distinguish between the *B. subtilis* TatAdCd and TatAyCy pathways (Jongbloed *et al.*, 2000), to show that Tat specificity is species dependent (Pop *et al.*, 2002), and to identify binding site residues in TatC (Eijlander *et al.*, 2009b).

Upon export, the unusually long Tat signal sequence (with an n-domain extended to 30 amino acids) of PhoD is cleaved to leave a mature protein of 56 kDa whose concentration in the surrounding medium

depends on phosphate limiting conditions (Eder *et al.*, 1996). There has been speculation that the target of PhoD is the teichoic acid of the *B. subtilis* cell wall (Eder *et al.*, 1996). The Pho regulon would thus play an additional role related to teichoic acids: as well as downregulating the teichoic acid synthases to prevent further phosphate usage, existing teichoic acids may also be used as a phosphate source, being replaced by phosphate-less teichuronic acid for maintenance of structural integrity (Grant, 1979).

In this study the substrate specificity of PhoD has been explored, as well as the metal supplementation conditions which lead to its optimal activity. The structure has been solved by X-ray crystallography, in order to shed light on why it is Tat exported when so many other *B. subtilis* proteins are efficiently exported by other means. A number of unusual details have emerged from this study: PhoD has a catalytic site quite different to other alkaline phosphatases known, but with similarities to the eukaryotic purple acid phosphatases. A C-terminal α -helix completely occludes the active site, whose truncation led to inactivation of the enzyme. An unusual Fe(III)-cysteine coordination is seen at the active site, and this may hint at the necessity for Tat export. PhoD has been shown to cleave teichoic acid, lending weight to the theory that this is its target. The work of this chapter was conducted in partnership with Dr Fernanda Rodrigez, who carried out purifications and biological assays.

6.2 PhoD construct.

Purified PhoD⁵⁷⁻⁵⁸³ was kindly donated by Fernanda Rodriguez. This had been expressed in *E. coli* BL21 cells as a solubility enhancing SUMO fusion protein (Peroutka Iii *et al.*, 2011) before removal of the SUMO sequence. The construct was designed without the disordered Tat signal sequence to aid crystallography (Figure 6-2). The protein obtained was confirmed as pure by size exclusion chromatography and SDS-PAGE. SEC-MALLS showed that PhoD exists as a monomer in solution, with mean mass 59.7 kDa (Figure 6-3).

10	20	30	40	50	60
MAYDSRFDEW	VQKLKEESFQ	NNTFDRRKFI	QGAGKIAGLS	LGLTIAQSVG	AFEVNAAPNF
70	80	90	100	110	120
SSYPFTLGVA	SGDPLSDSVV	LWTRLAPDPL	NGGGMPKQAV	PVKWEVAKDE	HFRKIVRKGT
130	140	150	160	170	180
EMAKPSLAHS	VHVEADGLEP	NKVYYYRFKT	GHELSPVGKT	KTLPAPGANV	PQMTFAFASC
190	200	210	220	230	240
QQYEHGYTA	YKHMAKEKLD	LVFHLGDYIY	EYGPNEYVSK	TGNVRTHNSA	EIITLQDYRN
250	260	270	280	290	300
RHAQYRSDAN	LKAAHAAPFW	VVTWDDHEVE	NNYANKIPEK	GQSVEAFVLR	RAAAYQAYYE
310	320	330	340	350	360
HMPLRISSLP	NGPDMQLYRH	FTYGNLASFN	VLDTRQYRDD	QANNDGNKPP	SDESRNPVRT
370	380	390	400	410	420
LLGKEQEQL	FNNLGSSTAH	WNVLAQQIFF	AKWNFGTSAS	PIYSMDSWDG	YPAQRERVIN
430	440	450	460	470	480
FIKSKNLNNV	VVLTGDVHAS	WASNLHVDFF	KTSSKIFGAE	FVGTSITSGG	NGADKRADTD
490	500	510	520	530	540
QILKENPHIQ	FFNDYRGYVR	CTVTPHQWKA	DYRVMPFVTE	PGAAISTRAS	FVYQKDQTGL
550	560	570	580		
RKVSSTTIQG	GVKQSDEVEE	DRFFSHNKAH	EKQMIKKRAK	ITN	

Figure 6-2 Sequence of PhoD. The Tat signal peptide is shown in red, the residues observable in the electron density shown in black, and those unobservable shown in green.

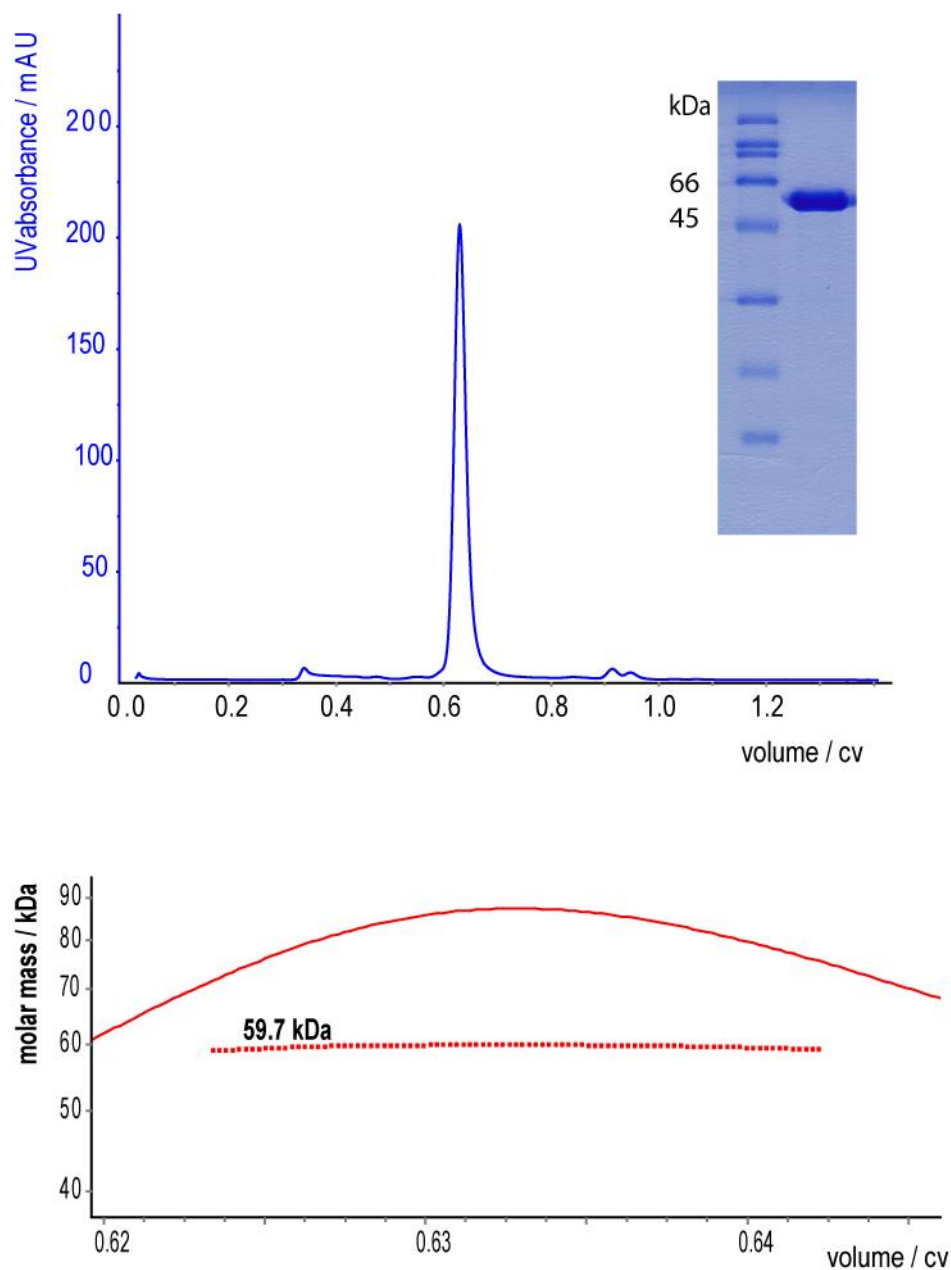


Figure 6-3 Purity and nature of PhoD in solution. Top: Size exclusion chromatography (Superdex 200 30/100) and resulting reducing SDS-PAGE of the single peak. **Bottom:** SEC-MALLS of PhoD. Running conditions: Buffer 20 mM Hepes pH 7.5, 400 mM NaCl, 2mM calcium acetate, SEC-MALLS flow-rate 0.4 ml/min.

6.2.1 Functional behaviour.

The purified PhoD behaved as an alkaline phosphatase. It was shown to cleave para-nitrophenyl phosphate (pNPP) with pH specificity. In this experiment, the yellow product para-nitrophenol is detected at 405 nm as a measure of phosphatase activity (Montalibet *et al.*, 2005) (Figure 6-4).

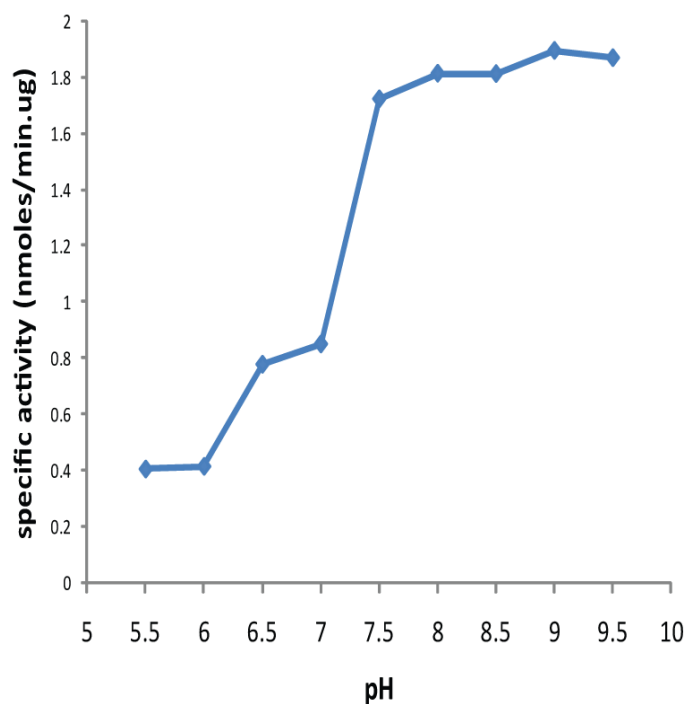


Figure 6-4 Alkaline phosphatase behaviour of PhoD. (Work of Fernanda Rodriguez).

As reported in the literature however, PhoD also functions as an alkaline phosphodiesterase, in contrast to the alkaline mono-phosphatases PhoA and PhoB (Hulett *et al.*, 1990). Furthermore, a clear physical difference between PhoA/B from PhoD is that of colour; where the former are colourless (Hulett *et al.*, 1990), the latter exhibits a purple colour, with value $\lambda_{\text{max}} \text{PhoD} = 527 \text{ nm}$. While PhoA/B are homologous with the well studied *E. coli* PhoA, the apparent divergence in behaviour of PhoD led to our pursuit of structural information.

6.3 EPR spectrum.

Since both the purple colour and use of the Tat export pathway are indicative of the binding of metallic cofactors, the identity of such a species was sought with EPR. The experimental parameters are given in Table 6.1. An EPR spectrum was only observable at cryogenic temperatures. The figure in Table 6.1 shows the signal decreased with increasing temperature. Collection at 10 K allowed a significant signal

to be observed, albeit at half the size of that observed at base temperature, without overuse of liquid helium.

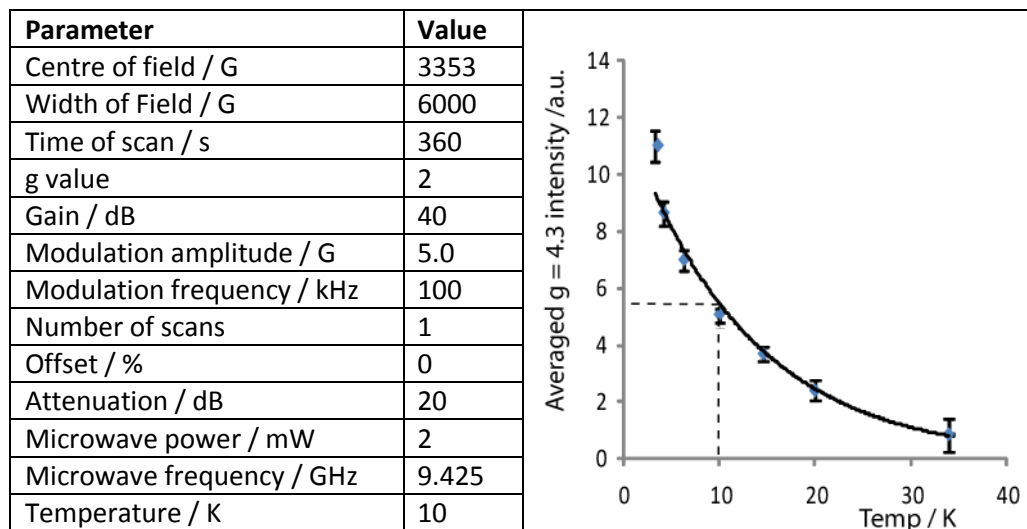


Table 6.1 EPR spectral parameters used for PhoD and PhoX data collections. The figure shows the fall in the observed ($g = 4.3$) peak height with temperature. The operating temperature of 10K was chosen.

The observed EPR spectrum is shown in Figure 6-5; a single peak was seen at $g = 4.3$, with a peak at 1/100th of this size at $g = 9.6$. In order to determine the metal origin through the redox and removal properties of this paramagnetic peak, as well as attempting to promote observable oxidation states of other metals, chemical additives were tested (conditions as used in other studies (Merks and Averill, 1998a, b). The peaks were readily reduced by addition of 5 mM sodium dithionite, but unchanged by addition of 2 mM hydrogen peroxide. Addition of 50 mM EDTA had no effect on the EPR peak, nor did it alter the purple colour of the protein solution, indicating that the paramagnetic metal could not be stripped. In parallel with these experiments, addition of sodium dithionite and EDTA was found to remove phosphatase activity (work of Fernanda Rodriguez).

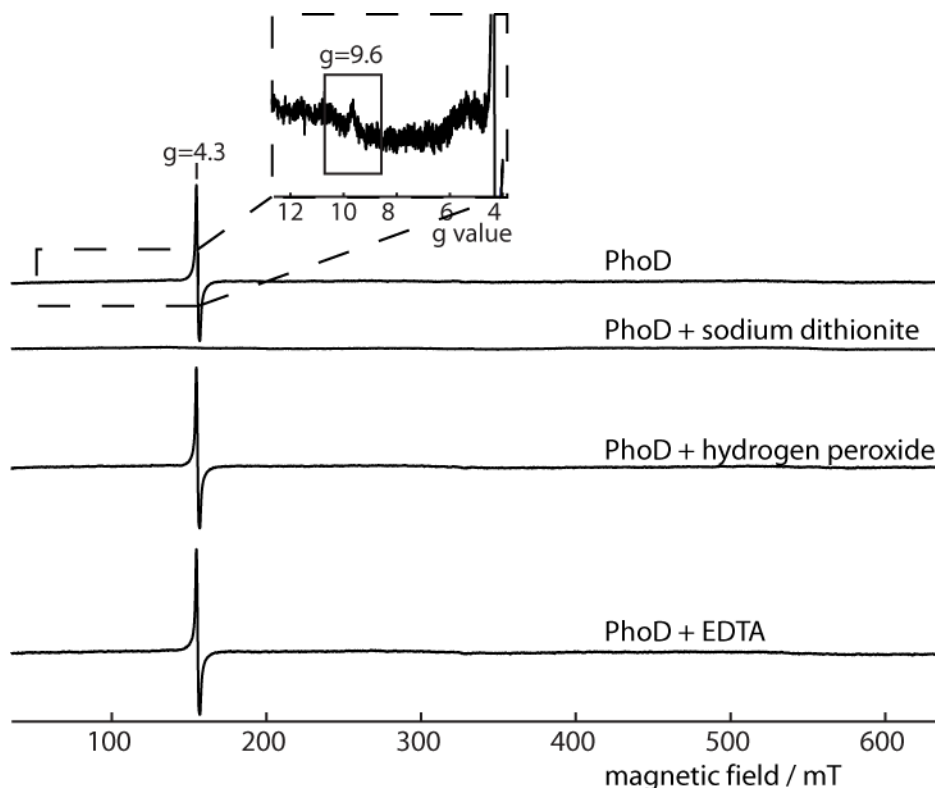


Figure 6-5 EPR spectra of PhoD highlighting the $g = 4.3$ peak. Inset - small $g = 9.6$ peak. Spectra are shown with native PhoD, 5 mM sodium dithionite, 2 mM hydrogen peroxide, and 50 mM EDTA. EPR operating conditions: microwave power, 2 mW; microwave frequency, 9.425 GHz; modulation, 5.0 G at 100 kHz; temperature 10 K. Spectra are buffer subtracted, with negative controls taken for sodium dithionite, hydrogen peroxide and EDTA solutions. A repeat with no glycerol added showed no change in lineshape.

6.3.1 Origin of the PhoD EPR spectrum.

The $g = 4.3$ peak is extensively defined in the literature as originating from a solitary high spin rhombic Fe(III) species (Blumberg, 1967; Oosterhuis, 1974). For this reason it is easily reduced and not oxidisable.

The peak arises from the symmetry of the coordination environment:

The EPR spectrum for a high spin Fe(III) may be explained via a Hamiltonian acting upon the electronic states $|m_s\rangle$. Nuclear spin states are neglected since the interactions are many times smaller than that with the electrons.

$$\hat{H} = g\beta(B_x S_x + B_y S_y + B_z S_z) + D(S_z^2 - \frac{S^2}{3}) + E(S_x^2 - S_y^2)$$

6.1

The first term is the Zeeman interaction: g is the electron g -factor (assumed to be isotropic (Bou-Abdallah and Chasteen, 2008)), B is the component of the applied magnetic field in the coordinate system of the spin, S is the spin angular momentum operator, D and E are the axial and rhombic zero field splitting parameters, respectively. This expression is an approximation: a quartic contribution to the zero field splitting is usually small at lower applied magnetic fields. The ground state for the high spin ferric of effective spin 5/2 is 6S . The resulting spin states $\pm 1/2, \pm 3/2, \pm 5/2$ (the three Kramer's doublets) are separated according to the zero field splitting. This zero field splitting arises from the symmetry of the coordination environment, specifically where the ground state 6S mixes with coordination sensitive higher orbital states. Whereas at high magnetic field, the selection rule is $\Delta m_s = \pm 1$, at lower field this mixing leads to resonance absorptions within the Kramer doublets. For rhombic symmetry the contribution comes particularly from the rhombic zero field splitting parameter E , and less from the axial $D = 3E$.

Each of these Kramer's doublets is assigned a fictitious $S' = 1/2$ and an effective g' value which may be observable in the EPR spectrum (Figure 6-6). These g' values are highly orientation dependent and thus, in a frozen solution sample such as that used in this study, only the isotropic $3 \rightarrow 4$ transition $g' = 4.3$ will be observable. They may however be a feature at $g' = 9.7$ with contributions from $1 \rightarrow 2$ and $5 \rightarrow 6$ transitions (due to their similar energy), in the molecular frame B_y and B_z directions respectively (Bou-Abdallah and Chasteen, 2008).

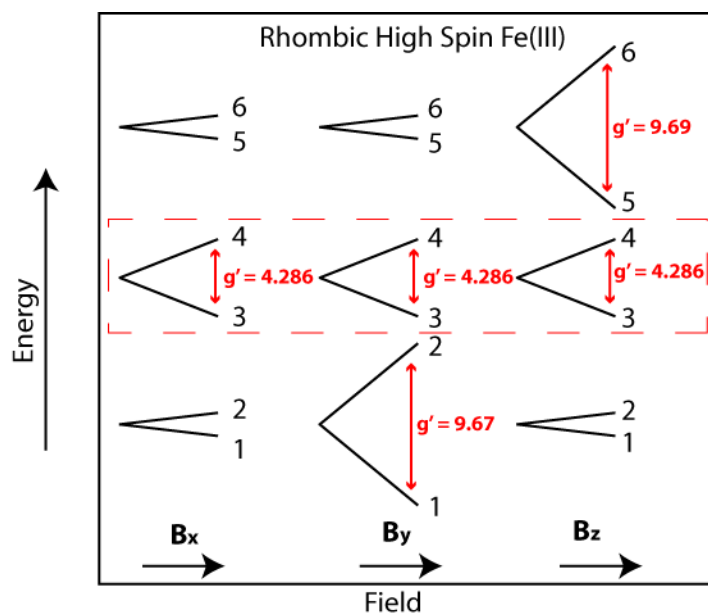


Figure 6-6 Origin of PhoD EPR peaks. Energy level diagram of the three Kramer's doublets which give rise to the Fe(III) EPR signal along the principal x, y and z axes for a rhombic $S = 5/2$ spin system. Energy splittings are not drawn to scale. (Figure modified (Bou-Abdallah and Chasteen, 2008)).

Quantitative measurement of $\text{FeCl}_3 + 50 \text{ mM EDTA}$ against a single PhoD batch (which yielded crystals) indicated ion occupancy of one Fe^{3+} per molecule. This calibration compared the peak height of the two solutions at $200 \mu\text{M FeCl}_3$ and PhoD, but the non-closed nature of the broad PhoD bound integral makes the value approximate.

6.4 Crystallisation of PhoD.

The active purified protein was put into crystal trials. Initial crystals grown in the JCSG+ broad screen condition 1.39 were characterised as being protein at the Diamond beamline I03 and optimised with homemade solutions around this condition (0.1 M sodium/potassium phosphate pH 6.2, 50% PEG 200, 0.2 M sodium chloride) as detailed in Figure 6-7. Specifically, the pH and PEG 200 concentrations were systematically changed to obtain the largest possible crystals.

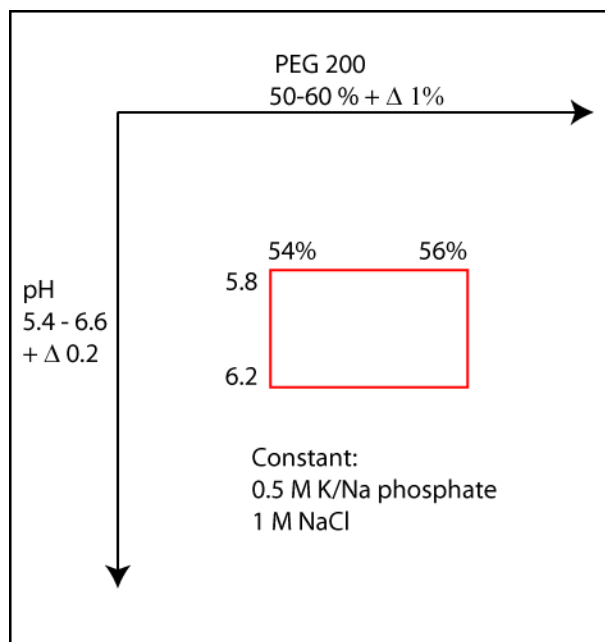


Figure 6-7 PhoD optimisation screen. This was constructed on the basis of crystals obtained in broad screens. The red box indicates the region of crystal growth.

Some difficulty was encountered with crystal reproducibility. Crystal growth was highly dependent on the particular purification batch, and a successful batch would become unreceptive to crystallisation after flash-frozen storage at -80°C . An exception to this was for a previously frozen sample which had a nucleant (Molecular Dimensions) added (Saridakis and Chayen, 2009).

The purpose of this is to provide a porous bio-glass surface on which nucleation may occur. The largest crystals grew at 12°C of the temperatures tested (4°C , 12°C , 21°C).

Crystals were cryoprotected in 10% ethylene glycol, with key datasets collected at beamlines ID23_1 (phasing) and I03 (maximising resolution). Details of the data collection are given in Table 6.3.

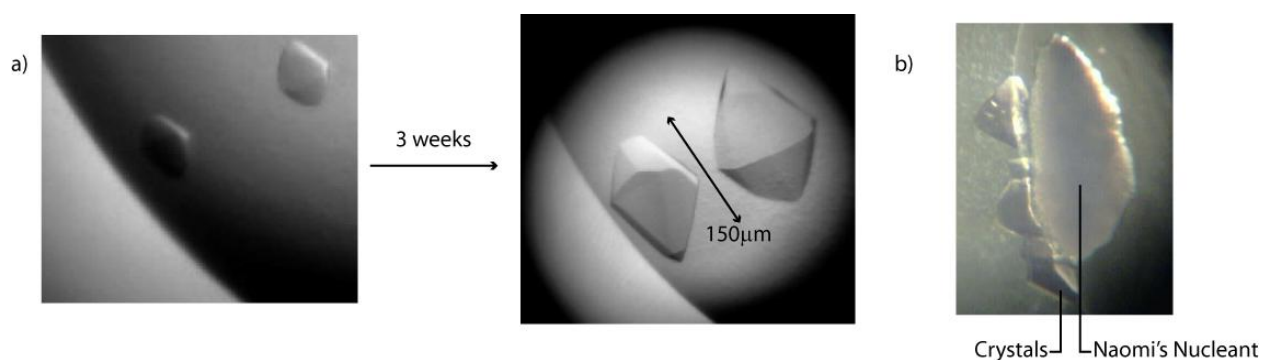


Figure 6-8 Morphology of PhoD crystals. a) PhoD crystal growth in three weeks. b) Use of a nucleant to facilitate crystallisation.

6.4.1 Data collection and processing.

From EPR it was known that PhoD in solution contains iron. This highlighted the possibility of anomalous diffraction methods for X-ray structural solution of the native crystal. A fluorescence scan across the iron

K-edge confirmed the presence of iron signal in the crystal (Figure 6-9), and the program CHOOCH was used to determine the appropriate peak, inflexion and high energy remote values (7110.6 eV, 7103.6 eV and 7140.6 eV respectively) of the X-ray wavelength for a MAD experiment (Evans and Pettifer, 2001). MAD data were collected at the ID23-1 beamline.

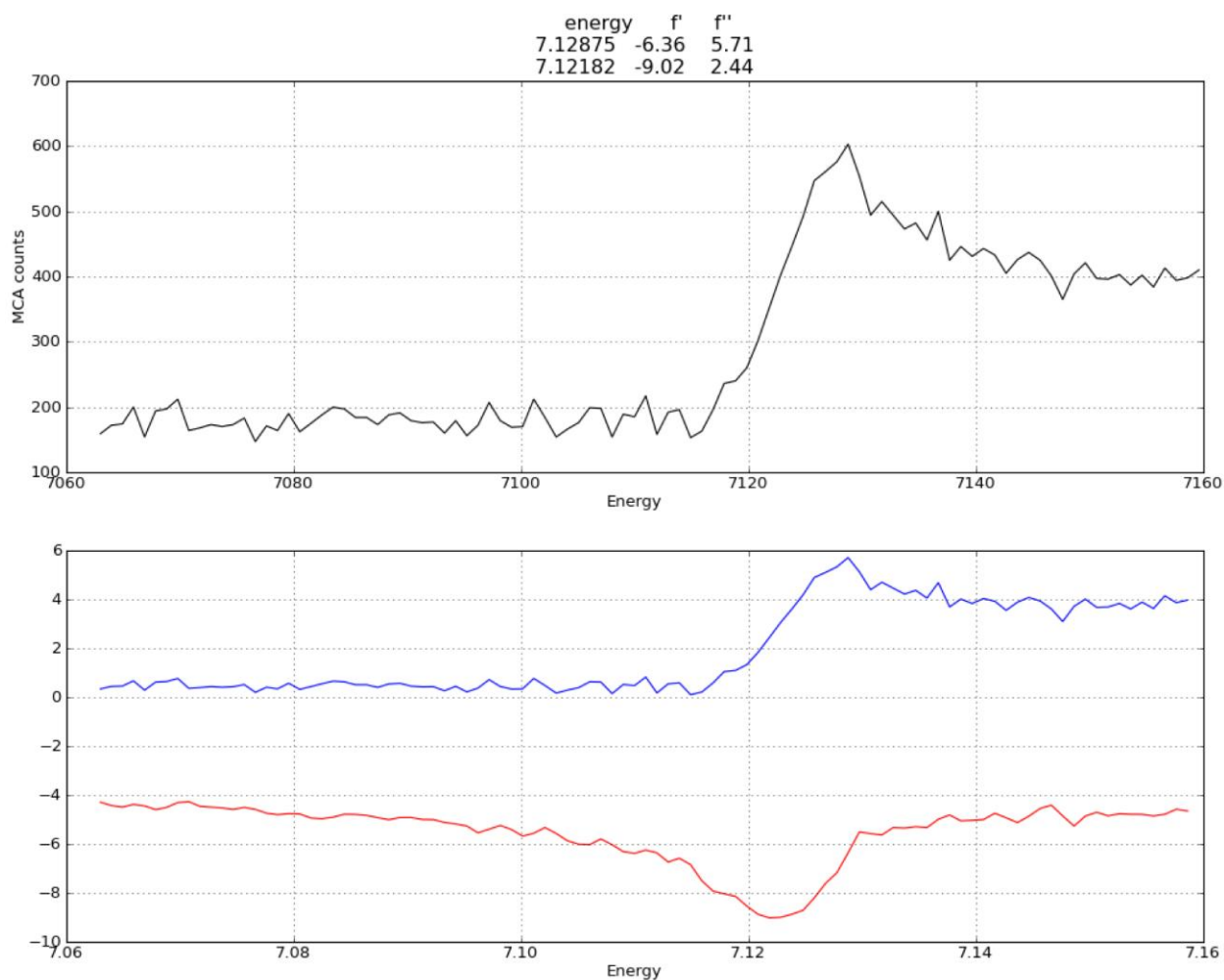


Figure 6-9 Iron K-edge scan. Top: Fluorescence scan. **Bottom:** f' and f'' spectra calculated from the experimentally measured X-ray fluorescence data.

The data were processed using XDS and Scala/Truncate via the Process pipeline. A body centred tetragonal lattice with cell dimensions (148.05 Å, 148.05 Å, 347.2 Å, 90.0°, 90.0°, 90.0°) was initially calculated. The symmetry of the diffraction pattern viewed after data integration allowed assignment by Pointless of the Laue group as 4/mmm. This was achieved by observing how well the intensities for the

purported symmetry agreed, and gave a net Z-score for the intensity correlation coefficients of 9.85. The possible space groups given this lattice and Laue group were identified as $I422$ (97) or $I4_122$ (98). Systematic absences were used to differentiate between these possibilities. Figure 6-10 shows that the reflection conditions were only fulfilled when $l = 4n$, which is characteristic of the space group $I4_122$ (and not $I422$). The space group $I4_122$ was assigned with confidence, given the Pointless probability statistics shown in Table 6.2.

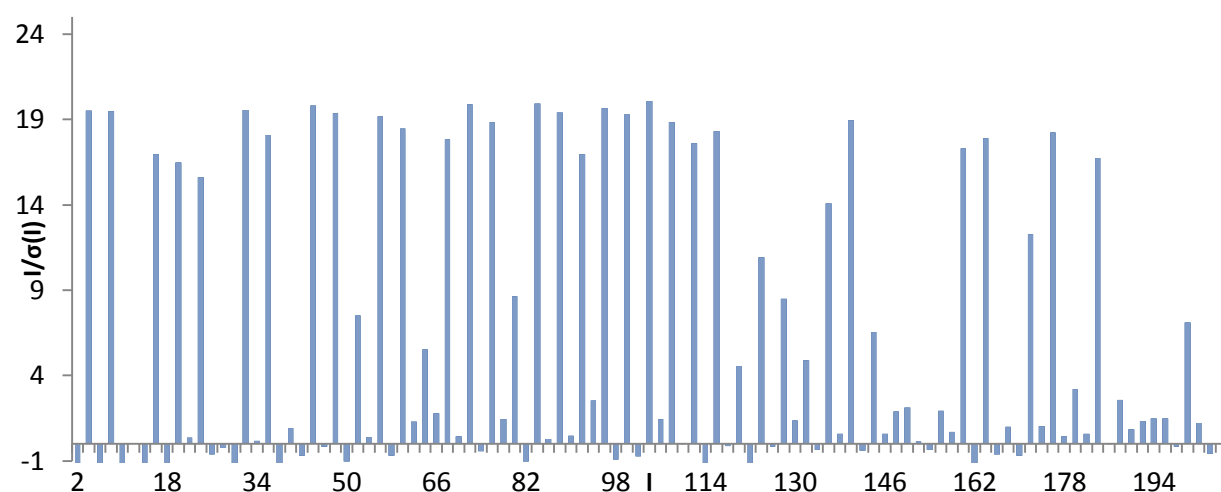


Figure 6-10 Observation of the systematic absences which led to space group determination.

Best Solution space group:	$I4_122$
Laue group probability:	1.000
Systematic absence probability:	0.934
Total probability:	0.934
Space group confidence:	0.901
Laue group confidence:	1.000

Table 6.2 Pointless result log. (Evans, 2006).

Higher resolution datasets were collected at beamline BM14 (ESRF) and I03 (Diamond). The best of these (I03) scaled to 1.9 Å, as shown in the Scala graphs (Figure 6-11). The data are shown to be consistently scaled, with only slightly increasing R_{merge} with batch – interpreted as low level radiation damage. The data are complete to 1.9 Å, with scaling statistics detailed in Table 6.3.

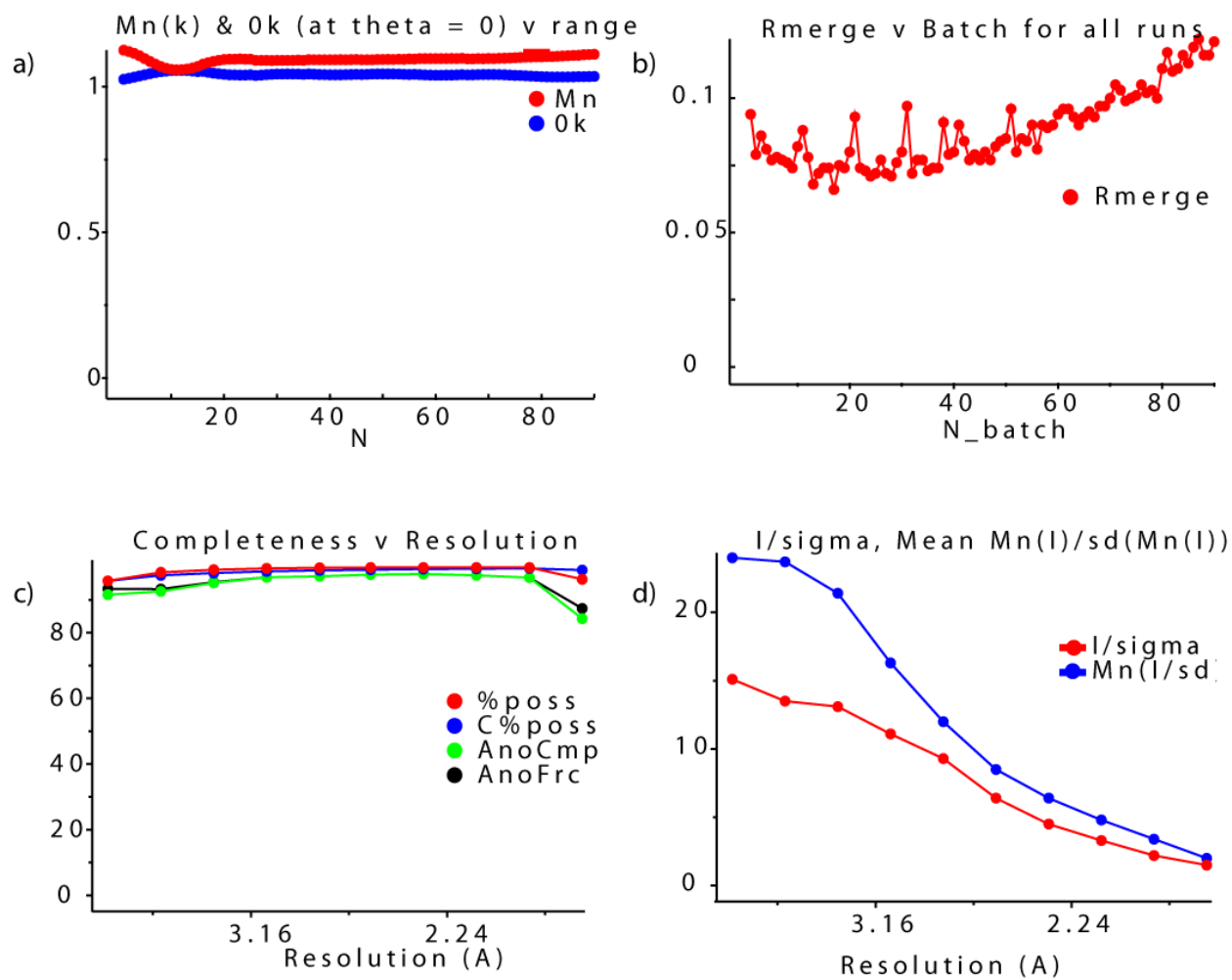


Figure 6-11 Scala graphs. **a)** Consistency of scaling across all N frames. **b)** R_{merge} varying with batch. **c)** Data completeness. **d)** Resolution of the data.

Data Collection:	ID23_1	ID14_4	IO3
Initial images	2 x images 1°, 91°	2 x images 1°, 91°	2 x images 1°, 91°
Initial estimated mosaicity	0.35	0.26	0.24
Initial index group	I4	I4	I4
Advised strategy	185° - 275°	218° - 308°	224° - 314°
Strategy completeness	100% (unique and anomalous)	100% (unique & anomalous)	100% (unique & anomalous)
	Multiplicity 3.9	Multiplicity 4.1	Multiplicity 4.1
Actual collection	185° - 285°	218° - 308°	224° - 314°
δφ	1°	1°	0.5°
Exposure level and time	100%, 1.0 sec	100%, 1.0 sec	100%, 0.5 sec
Space group	I4122		
Molecules in asymmetric unit	2		
Unit-cell parameters (Å, °)	148.05, 148.05, 347.2, 90.0, 90.0, 90.0		
Solvent content (%)	38		
Wavelength (Å)	Peak (λ=1.73920 Å)	Inflexion (λ=1.74090 Å)	Remote (λ = 1.73190 Å)
Resolution range (Å)	173.6 – 2.7	174.7 – 3.7	174.2 – 3.2
Number of Unique reflections	54642	20953	32169
Rmerge (%)	11 (48)	16 (49)	13 (48)
Rpim (%)	12 (29)	18 (25)	14 (25)
Mean I/σ (I)	15.6 (4.0)	11.0 (4.2)	14.1 (4.4)
Completeness (%)	100 (100)	100 (100)	100 (100)
Redundancy	8.4 (6.8)	8.5 (8.5)	8.5 (8.4)
			0.9393
			136.1 – 2.10
			102604 (14800)
			24 (45)
			29 (44)
			6.2 (1.9)
			86 (86)
			3.4 (2.1)
			1.28180
			135.6 - 1.9 (2.0 – 1.9)
			141401 (19895)
			7.5 (49)
			5.3 (38)
			9.6 (2.0)
			99 (96)
			3.5 (3.1)
MAD Phasing Statistics			
Sharp FOM (acentrics)	0.585 (136.19-11.94); 0.22 (136.19 -2.75); 0.11 (2.82-2.68)		
(Repeat with correct metals)	0.592 (136.19-11.94); 0.27 (136.19 -2.75); 0.22 (2.82-2.68)		
Sharp FOM (centrics)	0.199 (136.19-11.94); 0.07 (136.19 – 2.75); 0.07 (2.82-2.68)		
(Repeat with correct metals)	0.206 (136.19-11.94); 0.13 (136.19 -2.75); 0.20 (2.82-2.68)		
Sharp phasing power (iso/ano):			
Acentric	Peak (-)/0.487; Inflexion 0.086/0.328; Remote 0.023/0.250		
Centric	Peak (-)/ (-); Inflexion 0.082/ (-); Remote 0.021/ (-)		
Rcullis (acentrics)	Peak (-)/0.947; Inflexion 0.535/0.981; Remote 0.680/0.99		
Rcullis (centric)	Peak (-)/ (-); Inflexion 0.490/ (-); Remote 0.640/ (-)		
Solvent flattened FOM (overall)	0.864		
Solomon contrast start/end	0.5723/0.2681		
Sharp SigmaA	0.3807/0.7956		

Table 6.3 Data collection and phasing statistics.

The possibility of multiple copies of protein in the asymmetric unit was explored. This was not unlikely when considering the large volume of the cell – the unique edge having a dimension of 347 Å.

The Matthews coefficients gave probabilities for 2, 3 and 4 copies as 0.02, 0.5 and 0.4 respectively, suggesting three copies in the asymmetric unit. This low confidence probability, with significant probability of four copies making up the asymmetric unit, is characteristic of the problems that this coefficient can have at high symmetry, with the solvent content differing relatively little when incrementing the copies.

Patterson Maps proved informative however. Whereas the rotational Patterson showed no non-crystallographic peaks, the native Patterson (Figure 6-12) showed a significant peak at fractional coordinates (0.5, 0.5, 0.08) of the unit cell vector. Table 6.4 and Figure 6-12, calculated with the FFT (Ten Eyck, 1973) for Patterson program, gave the peak height at 185, an order of magnitude greater than subsequent suggested peaks. This indicated the presence of two-fold translational NCS in the PhoD unit cell.

a	b	c	Peak height
0.00	0.00	0.00	429
0.50	0.50	0.08	185
0.03	0.01	0.00	18
0.00	0.00	0.01	16
0.50	0.50	0.07	8

Table 6.4 Top five translational Patterson peaks as calculated by FFT for Patterson. Peaks are located with respect to their fractional distance down a, b and c axes.

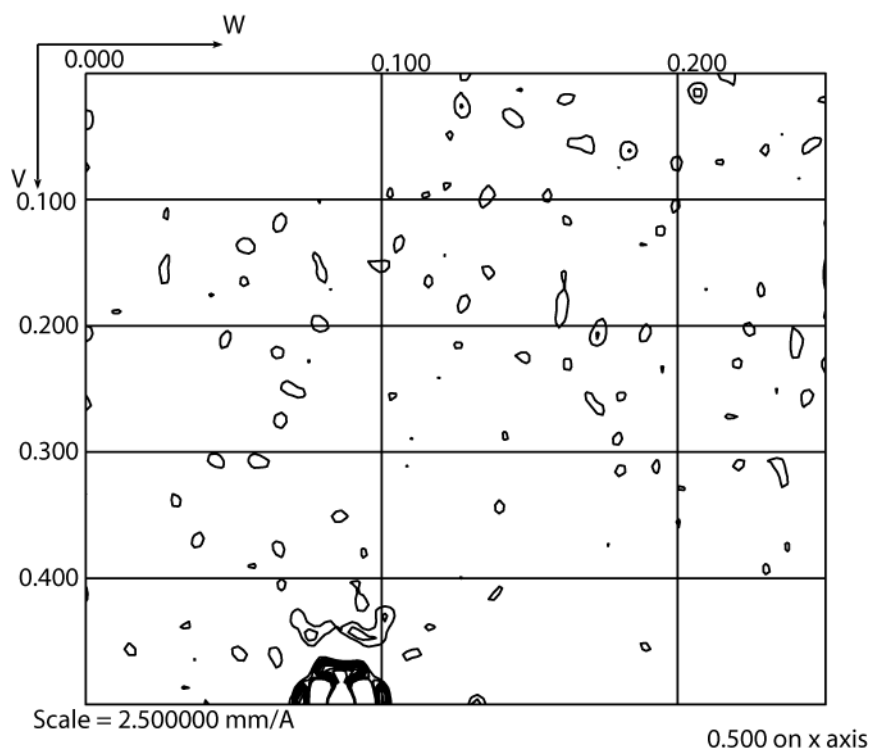


Figure 6-12 Harker sections showing the translational NCS at (0.50, 0.50, 0.08).

6.4.2 Phasing.

The same translational NCS was detected via the direct method approach of SHELXD within autoSHARP (Bricogne *et al.*, 2003). Two groups of three metal sites were found, initially all defined as Fe^{3+} . MAD phasing statistics are given in Table 6.3. Later the metals were redefined as Fe^{3+} and 2 x Ca^{2+} in accordance with chemical determination, although there is little difference to the phasing statistics. Substructure refinement was carried out for both possible hands, although in $I4_122$ the origin is not located on an enantiomorph axis, and the inversion operator for the metal coordinates is $(x, \frac{1}{2}-y, -z)$. Determination of the metal sites enabled phases to be calculated which were solvent flattened to 62% using Solomon (Abrahams and Leslie, 1996). This process allowed determination of the correct substructure hand. For one hand, the ratio of standard deviation of solvent to standard deviation of protein – contrast - decreased from 57% to 27%, whereas in the inverted hand it decreased from 63% to 39%, suggesting a better solvent mask for the former. Agreement with the data was also better for the

first hand, with the correlation coefficient on $|E^2|$ (squared normalised structure factor amplitudes) – SigmaA - increasing from 38% to 80%, whereas in the inverted hand the rise during twenty cycles of solvent flattening was 31% to 64%. Model building by Buccaneer (Cowtan, 2006) initially built 338 residues in six clustered polypeptide chains into this map, surrounded by many short chains.

Alongside this experiment, predicted structural motifs for PhoD were found by insertion of its amino acid sequence into FFAS, with the results shown in Table 6.5. Rather than having significant homology with other alkaline phosphatases of known structure, the PDBs which were predicted to have most similar structure were those of the eukaryotic purple acid phosphatases (PAPs).

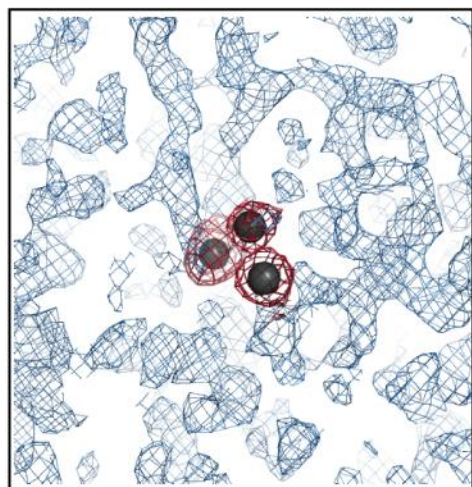
Protein	PDB	Score	% identity
Kidney bean purple acid phosphatase	1KBP	-32.6	13
Sweet Potato purple acid phosphatase	1XZW	-31.2	15
Rat Purple acid phosphatase	1QHW	-25.5	11
Human Purple acid phosphatase	1WAR	-23.8	12
Pig Purple acid phosphatase	1UTE	-22.9	13

Table 6.5 Highest FFAS scores for the PhoD sequence.

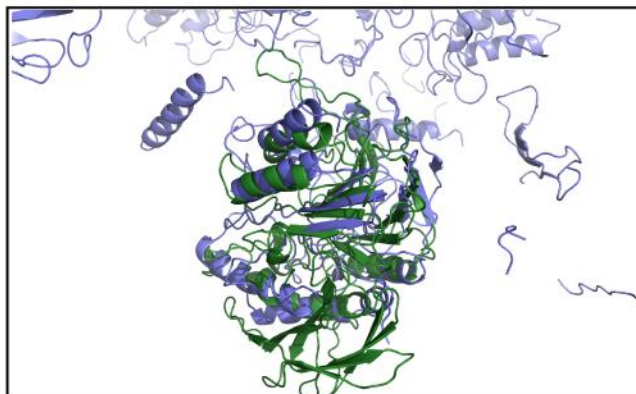
It was noticed that the six clustered polypeptides built by Buccaneer bore similarity to the kidney bean PAP structure (Figure 6-13). The model was thus manually manipulated based around this structure, removing all loose chains. This manual model was used as a starting chain for a second Buccaneer run. The output was a 524 residue chain. This was used as the model for molecular replacement via Molrep (Vagin and Teplyakov, 1997), where a second copy of the model was placed at the translational NCS

vector.

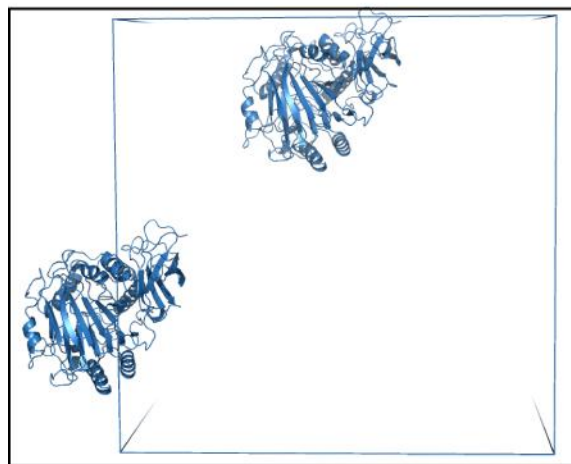
1) Sharp heavy atom model Solomon
solvent flattened phases 62%.



2) Buccaneer model alike kidney bean PAP



4) Two NCS copies separated by translation vector



3) Manual removal of single Buccaneer chain.

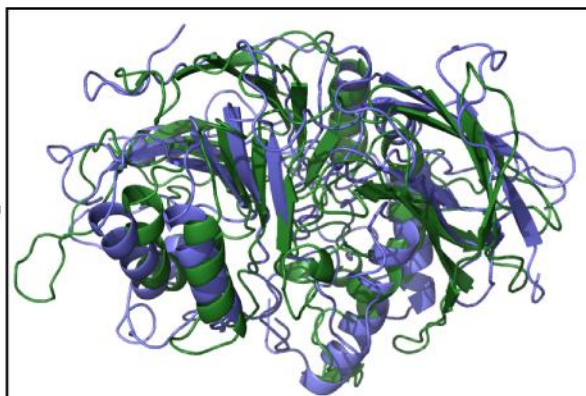


Figure 6-13 Schematic of phase solution. 1) Difference map (red, 5σ) shows metal sites against solvent flattened density (Solomon, 2.25σ). 2) Initial Buccaneer output (blue) was compared against kidney bean PAP (green). 3) A second Buccaneer job produced a more complete chain. 4) Molrep placed the translational NCS copy.

6.4.3 Refinement.

Iterative refinement and model building was carried out with autoBuster (Blanc *et al.*, 2004)/Refmac (Murshudov *et al.*, 1997) and Coot (Emsley and Cowtan, 2004), leading to a final structure at resolution $1.9\text{ \AA}$ with overall R and R_{free} 16.5/18.9% (Table 6.6). Figure 6-14 shows the progression of refinement

with number of refinement cycles. The quality of the model was checked using Molprobit (Davis *et al.*, 2007), with a final Molprobit score 1.19 (99th percentile). Figure 6-14 also shows a crystal packing diagram.

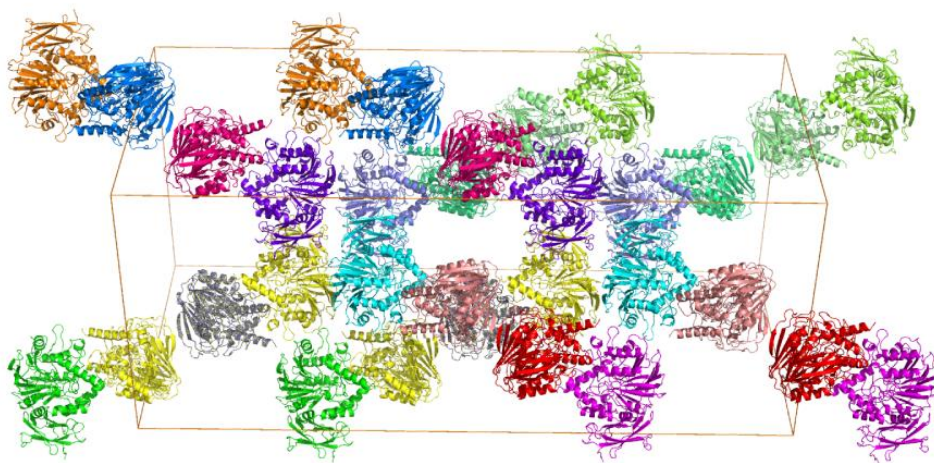
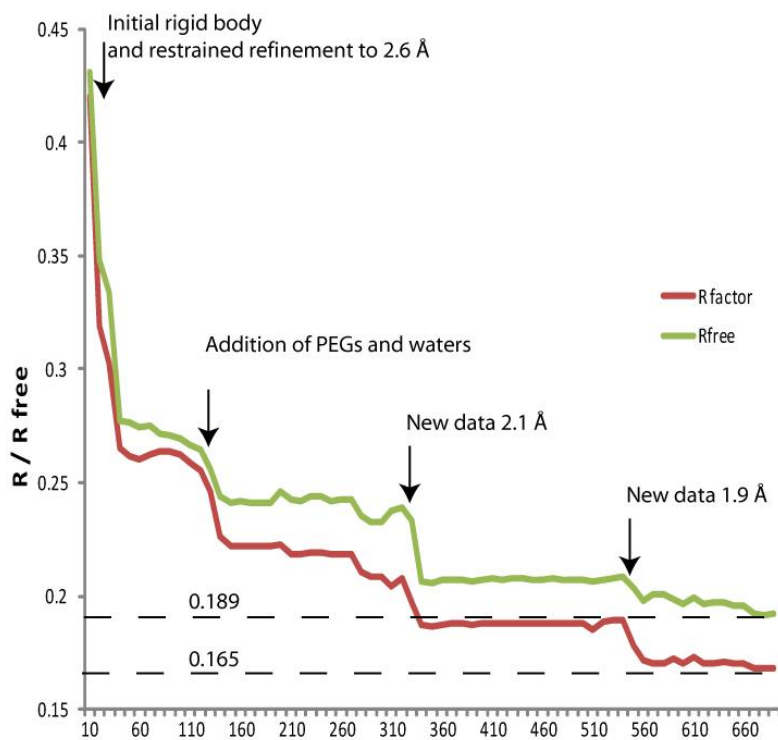


Figure 6-14 The improvement of R/R_{free} with iterative refinement and crystal packing diagram.

Refinement summary from data collection:	ID14_4	IO3
Resolution (Å)	136-2.1 (2.2-2.1)	136-1.9 (2.0 – 1.9)
R factor (%)	18.8 (29.0)	16.5 (23.3)
Free R factor (%)	21.0 (32.0)	18.9 (24.8)
R.m.s.d. bond lengths (Å)	0.008	0.010
R.m.s.d. bond angles (°)	0.96	1.010
No. atoms in asymmetric unit		20841
Of which: Protein		19772
Fe ³⁺		2
Ca ²⁺		4
Na ⁺		6
PEG200/EG		420 (73 units)
Phosphate		10 (2 units)
H ₂ O		627
Ramachandram plot:		
Preferred (%)	97	97
Allowed (%)	3	3
Outliers (%)	0	0
PDB code		2YEQ

Table 6.6 Refinement statistics.

During refinement, a reduction in R-factor by 1.6% was seen as a result of adding 73 PEG molecules to the PhoD structure. The PEG 200 PDB was obtained from HICUP. These mainly occur wrapped around lysine residues (for example 124, 159, 425) although proximity to arginines (147, 528) was also found. Hydrogen bonded PEGs are reported in other structures in the PDB, as well as being described as ‘participating in crystal contacts’ in the literature (Rupp, 2010). Whilst some are undoubtedly taking up that role here, there are a number of PEG molecules which bind to the protein not at a crystal contact.

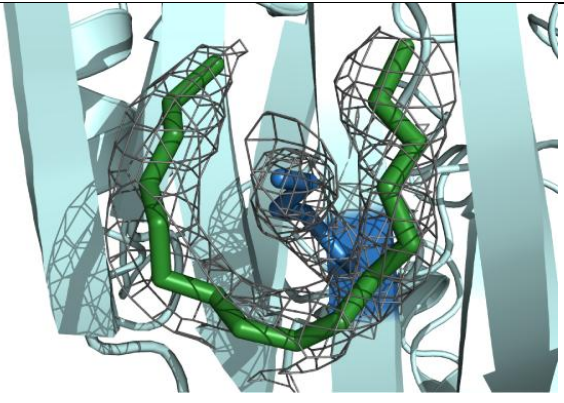
PDB	pH and PEG content of condition	
2WOQ	pH 7.5, 30% PEG 400	
3C1Q	pH 7.0, 12.5% PEG 400	
3BX8	pH 5.5, no details of PEG	
2R09	pH 8.0, 15% PEG 6000	
2OA5	pH 7.0, 80% PEG 400	
1Z5P	pH 7.0, PEG2000 MME	
1U3A	pH 6.5, PEG 550	
1WMA	pH 6.0, no details of PEG	
1Y89	pH 7.5, 40% PEG 300	

Figure 6-15 PEG ordering. Left: Structures in the PDB which contain PEG. Right: Cartoon of PEG wrapped around PhoD Lys509.

At pH 5.8, PhoD crystallised in a lower pH than many of the PDBs in Figure 6-15, however it is not exceptional. Where it is unusual is the sheer number of molecules attached, along with the entropic penalty that this presumably imposes - the above PDBs contain four molecules at most. The concentration (50%) of PEG used here will encourage this binding.

Whilst it is difficult to compare different environment lysines, it appears that introduction of a PEG molecule around a lysine residue stabilises it relative to a free lysine. This is shown in a comparison between sidechain nitrogen B-factors of Lys425 ($B = 61 \text{ \AA}^2$) and Lys124 ($B = 68 \text{ \AA}^2$) with non PEG surrounded analogues Lys118 ($B = 73 \text{ \AA}^2$) and Lys103 ($B = 77 \text{ \AA}^2$). The B-factors are comparable to those of the PEGs themselves, at around $66 \text{ \AA}^2$.

6.5 Structure of PhoD.

Three orientations of PhoD ribbon diagrams, related by 90° , are shown in Figure 6-16. A topology diagram generated by PDBsum is also shown in the same colouring and orientation as the main ribbon diagrams. The global structure is a double β -sheet sandwich backbone whose connecting turns and loops provide the coordinating residues for the metal atoms, which is flanked on either side by a number of variable length α -helices. Quantitatively, there are 14 α -helices comprising 20% of the amino acid residues. A similar percentage (26%) of residues are β -strand material, making up four β -sheets. This leaves a high proportion of residues for other secondary structure, and it is observed in Figure 6-16 that the structure is rich in loops. Two β - α - β motifs (residues 173-178 and 201-204) as well as a more unusual psi-loop (a β -strand placed between the antiparallel β -strands 429-434 and 440-446) are observed. Every residue from 57 to 578 was observed in the electron density.

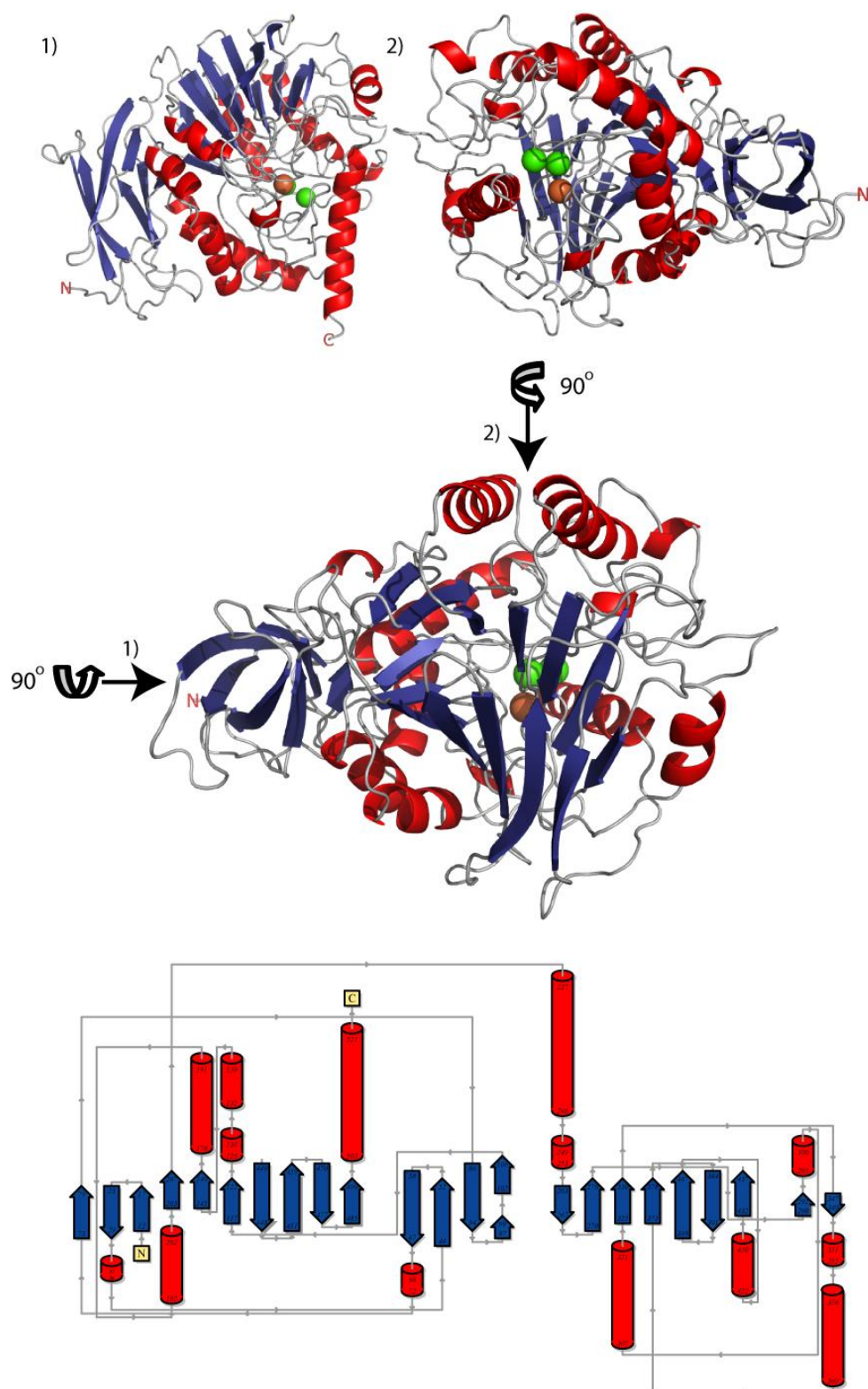


Figure 6-16 Global structure of PhoD. A ribbon diagram is shown in three orientations related by the 90° axes 1) and 2) which are defined in relation to the middle larger primary diagram by the arrows shown. α -helices are coloured in red and β -sheets in blue. The three metals of the catalytic site are presented as spheres. A topology diagram produced by PDBsum is shown with the same colouring scheme, in the same orientation as the primary ribbon diagram.

The structure is representative of a motif seen in a number of metallophosphoesterases, across bacteria, eukaryotes and archaea (Bhadra et al., 2005; Galperin et al., 1998; Koonin, 1994). These are functionally diverse enzymes, acting as protein phosphatases, nucleotidases and nucleases. Examples include *E. coli* alkaline phosphatase, calcineurin, *Listeria monocytogenes* cyclic nucleotide phosphodiesterase (Lmo2642) and pyruvate dehydrogenase phosphatases (Barford et al., 1998; Holtz et al., 2000; Kim et al., 2011; Vassilyev and Symersky, 2007). A number of sulfatases also share this fold, including arylsulfatase A and N-acetylgalactosamine-4-sulfatase (Bond et al., 1997; Lukatela et al., 1998). The metals are always coordinated by loop residues at the C-terminal side of the β -sheets (Miller et al., 2007), although the nature and quantity of metal vary from one (aryl sulfatases, Mg^{2+}) through to three (alkaline phosphatase, two Zn^{2+} , one Mg^{2+}). A two metal catalytic site is common in the metallophosphoesterases, often including iron or zinc ions which are octahedrally coordinated by histidine, aspartate, and asparagine residues.

The active site of PhoD was found to contain one Fe^{3+} , and two Ca^{2+} ions, as well as a bound phosphate, density for which is shown in Figure 6-18. Additional Na^+ ions were identified in the density, bound with octahedral coordination to waters and negatively charged ligands (Glu231, Asp340).

Despite the global fold of PhoD being a member of the metallophosphoesterase family, containing a large number of known structure alkaline phosphatases and phosphodiesterases, structural alignment of PhoD via the DALI server (Holm and Rosenstrom, 2010) confirmed that the PAPs are the most homologous known PDB structures. This is shown in Figure 6-17, which shows the highest DALI Z-score results, as well as an alignment of PhoD to kidney bean PAP.

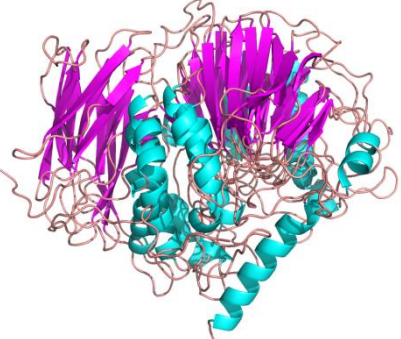
Protein	PDB	Z-Score	R.m.s.d.	Alignment length	
Kidney Bean PAP	2qfp	26.1	3.0	367	
Sweet Potato PAP	1xzw	25.3	3.1	363	
Human PAP	2bq8	22.0	2.9	250	
Rat PAP	1qhw	21.9	2.9	247	
Pig PAP	1ute	21.7	2.9	248	
Cyclic AMP Phosphodiesterase	3ib7	18.5	3.2	295	

Figure 6-17 DALI structural homology with PhoD. PAPs have the most similar structures to PhoD in the PDB. The alignment of PhoD and kidney bean PAP is shown in a ribbon diagram, with β -sheets coloured purple, α -helices coloured cyan, and loops coloured maroon.

6.6 The catalytic site of PhoD.

In the same way that PhoD shows global structural homology with the eukaryotic purple acid phosphatases, similarities were expected at the catalytic site. This initially appeared to be the case, however there are substantial differences in both the number and nature of metal ions.

6.6.1 An overall view of the catalytic site.

The catalytic site of PhoD is shown in two dimensions in Figure 6-18 to allow coordination of the three metal ions to be viewed easily, as well as in three dimensions with stereo-related views. The three metal ions are each coordinated to a single bound phosphate, and were characterised as an Fe^{3+} and two Ca^{2+} ions by the methods outlined in the next sections. Phosphate is a known inhibitor of PhoD (Yamane and Maruo, 1978b), as well as being the product of its phosphatase action, and clearly has a high propensity to bind in the active site. As discussed in section 6.8 it also binds directly to Arg562 of a C-terminal α -helix which caps the active site.

The phosphate is shown to be doubly coordinated with two of the metals, in a tripodal position analogous to that observed in the sweet potato PAP (Schenk *et al.*, 2005). This is thought to be a reaction intermediate position used by the PAPs prior to dissociation of the inorganic phosphate. The smaller ionic radius of Fe^{3+} (78.5 pm) compared to Ca^{2+} (114 pm) allow for the slightly shorter iron to phosphate distances shown in Figure 6-18.

Attempts to generate crystals replacing phosphate with the transition state analogue vanadate (Davies and Hol, 2004) by substitution of the 0.1 M sodium/potassium phosphate buffer with 0.1 M 2-(*N*-morpholino) ethanesulfonic acid and addition of 1-50 mM sodium vanadate yielded no crystals. Attempts to generate apoprotein crystals by addition of 10-50 mM EDTA similarly proved unsuccessful.

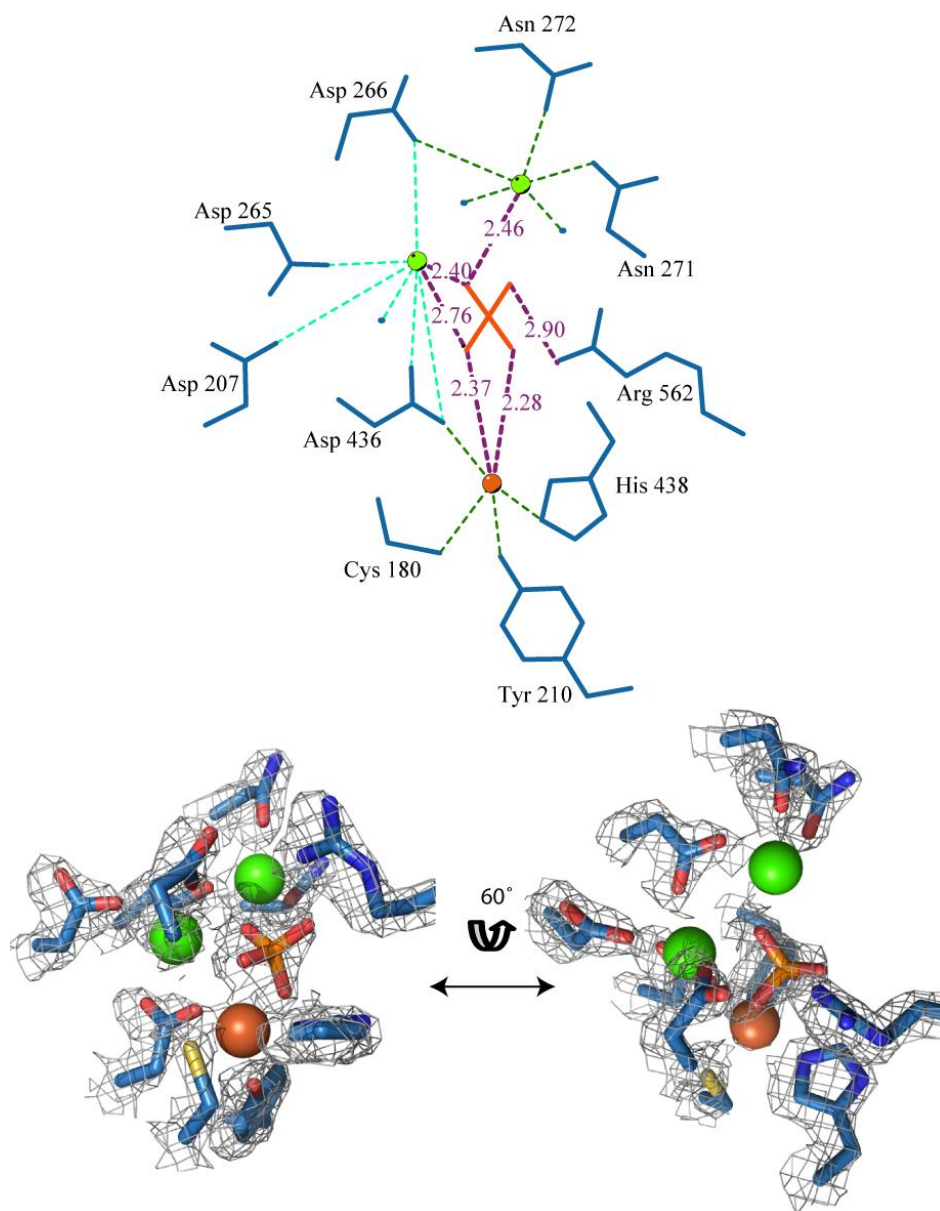


Figure 6-18 The PhoD active site. **Top:** A 2D view of the active site, showing the bonds made by the metals with the bound phosphate (distances shown in Å) and with coordinating residues. Figure drawn by Ligplot. (Wallace *et al.*, 1995) **Bottom:** 3D stereo view of the catalytic site, with the two images rotated by 60° along the axis shown. Figure prepared in Pymol, with $\sigma=2.0$ and $\text{carve} = 1.4$. Density of the metal sites is omitted for clarity. Coordinating residues are shown in blue, phosphate in orange, Fe³⁺ in brown and Ca²⁺ in green.

The coordinating residues of the three metal ions seen in the crystal structure were found to be widely conserved amongst known and putative alkaline phosphatases identified as homologous to PhoD by NCBI Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) (Altschul and Koonin, 1998). These homologous sequences were found in alkaline phosphatases from a variety of different bacteria. After the unsurprisingly high homology with other members of the *Bacillus* genus, significant results with alkaline phosphatases with circa 50% sequence identity from other diverse bacteria are observed. Figure 6-19 shows the catalytic site conservation in tabular and graphical form via Consurf for the 100 most similar sequences. One protein which did not appear in the PSI-BLAST search due to its lesser (31%) sequence identity, but which has the same active site as PhoD by sequence alignment, is the phospholipase PLD of *Streptomyces chromofuscus*. Whilst the structure of this protein is unknown, some published mutational work (Zambonelli *et al.*, 2003) can be combined with the PhoD structure to speculate on the function of certain residues (section 6.9.2).

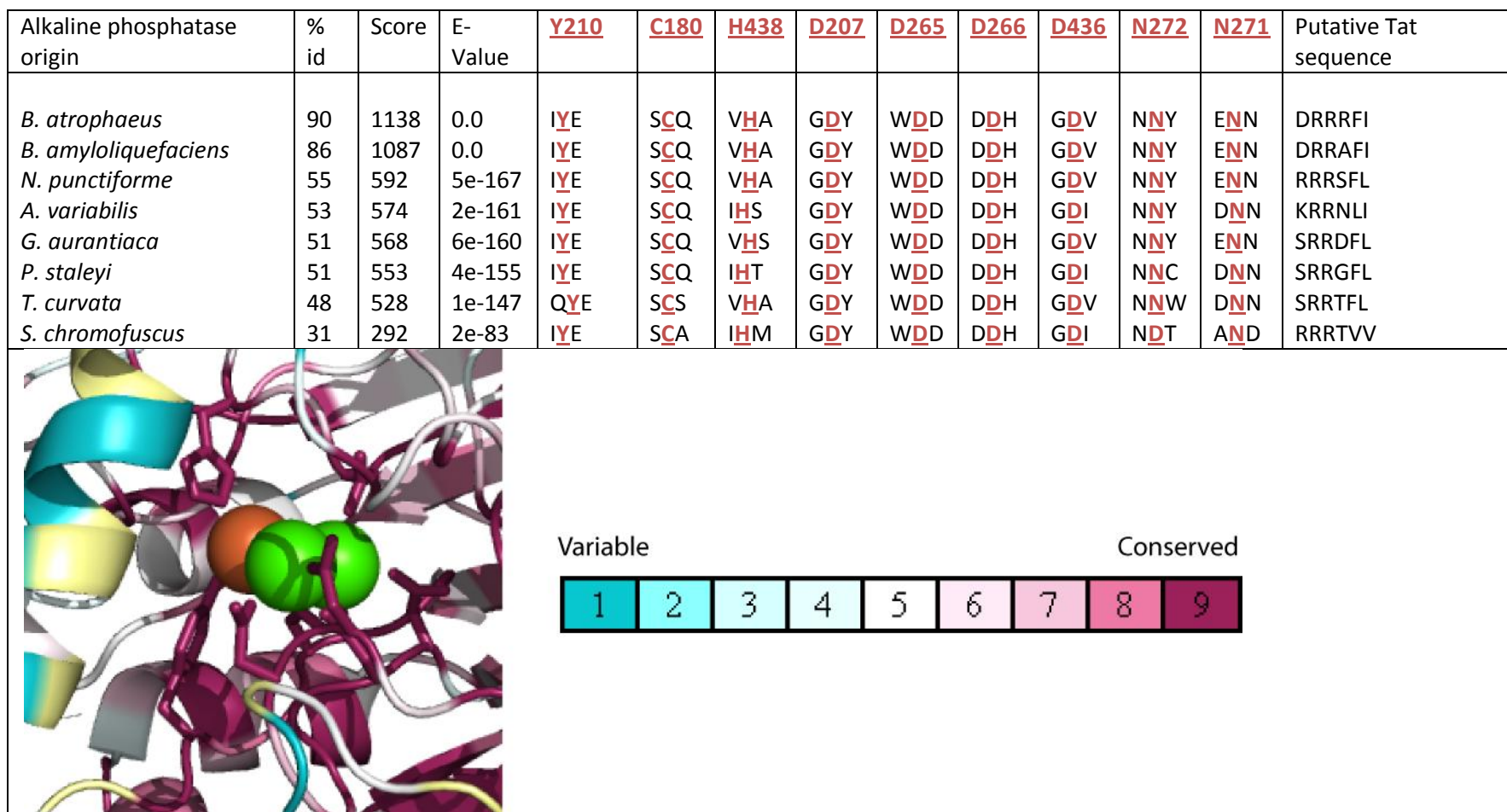


Figure 6-19 Active site conservation. Top: Conservation of all metal coordinating residues within PhoD sequences identified by PSI-BLAST. This is a particular BLAST derivative which identifies significant features in the sequence in the search for evolutionarily related proteins. Related sequences are scored via a number of criteria including the comparison of residues, gap penalties and sequence composition, and the score is assigned an E-value (Altschul and Gish, 1996). The E-value is the number of times a false hit would be expected to give a similar score, so details the level of homology significance. A putative Tat signal sequence is found for each of the sequences. **Bottom:** Conservation of coordinating residues in graphical form (Consurf (Ashkenazy *et al.*, 2010)) from the 100 most similar sequences. The key shows the colouring for the scale from variable to conserved residues.

Despite the high level of active site conservation in alkaline phosphatases, the greatest catalytic site homology with currently deposited structures is seen with the purple acid phosphatases (see below for details), however this group have two metals only at the active site. Other alkaline phosphatases of deposited structure are known to contain three metal ions, although they only use two ions in binding the phosphate substrate. Examples include *E. coli* PhoA (Holtz and Kantrowitz, 1999; Stec et al., 2000) and tPphA from *Thermosynechococcus elongates*. In these examples, it is possible to remove the third metal by mutating coordinating residues and retain some activity. In PhoD however, this was not the case. It proved impossible to express PhoD with mutations at any of the active site residues. This, and the positioning of the phosphate coordinated to all three metals in the crystal structure, implies a binding role for each metal.

6.6.2 Metal I: Fe(III).

The presence of the first metal, a solitary Fe³⁺ ion, was identified by the EPR experiment given in section 6.3. This ion is observable as the strongest single feature in the electron density. It refined to full occupancy with B-factor 17.6 Å², and shares remarkable similarity with the conserved Fe(III) site of the purple acid phosphatases.

6.6.2.1 Comparison with the PAP Fe(III) site.

As shown in Figure 6-20, comparison of PhoD and PAP Fe(III) site shows that the octahedral coordination site is almost entirely conserved, including tyrosine 210, whose visible charge transfer band with Fe(III) leads to the purple colour of both PhoD and the PAPs (λ_{max} PhoD = 527 nm and λ_{max} PAP = 499 – 560 nm) (Klabunde and Krebs, 1997; Schenk et al., 2005). The one difference in Fe³⁺ coordination between this alkaline phosphatase and the purple acid phosphatases is the replacement of an aspartic acid residue with a cysteine (Cys180). The replacement of a hard negatively charged oxygen with softer

sulphur is an unusual replacement leading to a rare Fe(III)-S coordination. This is discussed in Section 6.7.

The octahedral coordination does deviate slightly from regularity; an axial (Cys180 and PO₄) set with longer coordination distances may be separated from the remaining equatorial set. In addition, the angles show variation from ideal 90° coordination, as shown in Figure 6-20.

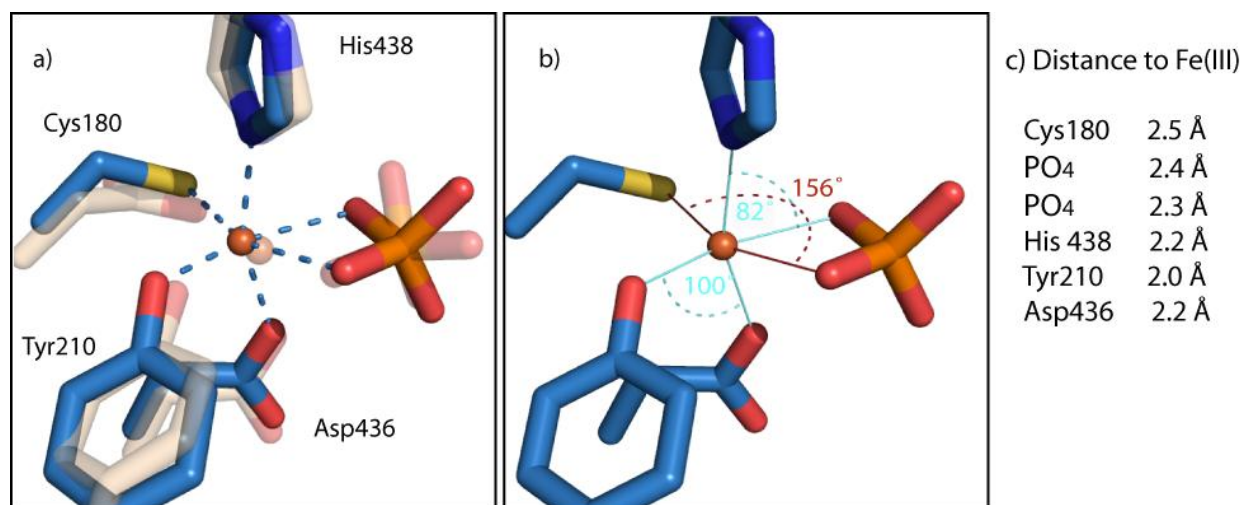


Figure 6-20 The Fe(III) site. **a)** Comparative coordination of Fe(III) in PhoD (foreground, blue) with Fe(III) in the PAPs. Alignment carried out with LSQKAB (Kabsch, 1976). **b)** Geometric coordination of Fe(III). The ion is approximately octahedrally coordinated, with slightly elongated axial distances to Cys180 and PO₄. **c)** Fe(III)-ligand distances.

6.6.3 Metals II and III.

Initially, there were a number of candidates for the second and third metal site identities. The second metal site of the purple acid phosphatases is either a Fe(II) or Zn(II), therefore these were initially pursued as the most likely candidates for the second metal site in PhoD. However, the following evidence allowed these possibilities to be disregarded, with both metals being identified as Ca²⁺ ions, refined with B-factors 18.4 Å² and 20.4 Å² to full occupancy.

6.6.3.1 Calcium occupancy.

1) EPR: The EPR spectra (Figure 6-5) show the existence of a *solitary* Fe(III). EPR spectra from any other paramagnetic species were absent. Further Fe(III) would lead to anti-ferromagnetic spin coupling and

loss of signal. The presence of Fe(II) would lead to observed peaks around $g = 1.8$ rather than $g = 4.3$ because of coupling to the original Fe(III) (Dietrich *et al.*, 1991).

2) Anomalous signal: During the course of data collection, a number of datasets at different wavelengths were collected. Given that the Fe(III) site had been characterised, the variation in anomalous signal across these different wavelength data sets could be followed to identify the other metals, using the anomalous signal of Fe(III) as a benchmark (Figure 6-21).

The first column of Figure 6-21 (0.91731 Å) is a typical native wavelength dataset. At 5σ (above the noise) the iron shows a sizeable anomalous signal. None of the other metals shown display anomalous signal above the noise level. At this wavelength, the f'' (Fe^{3+}) = 1.3386 (as calculated by Crossec), so the other metals would be expected to have f'' lower than this. On the contrary, that of Zn^{2+} is greater - f'' (Zn^{2+}) = 2.2208.

In the second column of Figure 6-21, at 1.28180 Å, the iron anomalous signal has become more pronounced at 5σ (f'' (Fe^{3+}) = 2.3695). Anomalous signal is observable for two other metals, but is smaller than for the iron – again in contrast to that expected from f'' for Zn^{2+} . The ratios of the observed signal sizes iron:metal II and iron:metal III correspond well to the theoretical f'' (Fe^{3+})/ f'' (Ca^{2+}) ratio calculated by Crossec. This data supports both metal sites as being calcium ions.

Increasing the wavelength further to $\lambda = 1.7392$ Å gives the maximum for the Fe^{3+} signal. This wavelength is at the peak of Fe^{3+} absorbance before decreasing to a minimum in the pre-edge around $\lambda = 1.74820$ Å. The fourth column of Figure 6-21 shows the disappearance of all Fe^{3+} anomalous signal. At this low energy, the lighter elements absorb more significantly, as shown by the cysteine anomalous signal. In the same way as the Fe(III)/M(II) ratio indicated M(II) and M(III) as Ca^{2+} , at $\lambda = 1.74820$ Å the f'' of M(II) and M(III) may be correlated as Ca^{2+} using the f'' of the Fe^{3+} coordinating cysteine.

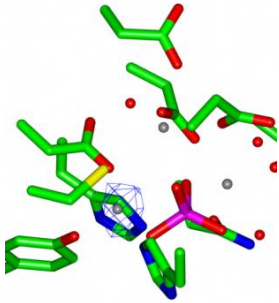
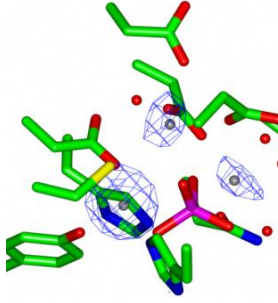
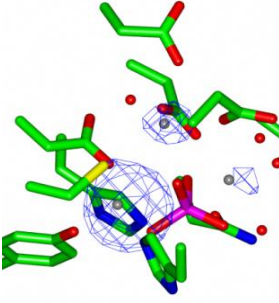
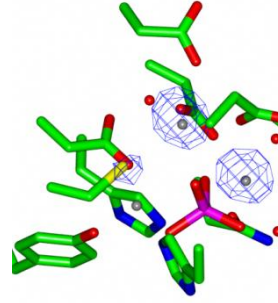
0.91731 Å	1.28180 Å	1.7392 Å	1.74820 Å
			
f'/f'' Fe 0.2734/1.3386 Ca 0.2736/0.4997 Zn -0.1203/2.2208 S 0.1645/0.2058 P 0.1368/0.1580	f'/f'' Fe -0.1698/2.3695 Ca 0.3534/0.9261 Zn -6.2999/3.8861 S 0.2618/0.3940 P 0.2282/0.3051	f'/f'' Fe -5.8626/ 3.9332 Ca 0.2758/1.5863 Zn -1.2203/ 0.8423 S 0.3488/ 0.6961 P 0.3175/0.5439	f'/f'' Fe -5.6696/0.4702 Ca 0.2719/1.6004 Zn -1.2072/0.8500 S 0.3500/0.7027 P 0.3189/0.5491

Figure 6-21 Variation in anomalous signal at the active site with increasing wavelength.

3) Metal supplementation experiments: Addition of Fe^{3+} and Ca^{2+} to PhoD was shown to significantly (greater than two-fold) increase the phosphatase activity of PhoD whereas other metals such as Zn^{2+} , Mg^{2+} , Co^{2+} and Cu^{2+} did not (work of Fernanda Rodriguez).

4) A microPIXE experiment identified the presence of Fe^{3+} and Ca^{2+} , although quantitation of the metals was not possible (work of Oliver Zeldin and Elspeth Garman).

Whilst these data all support the conclusion that the two bound metal ions are calcium, other identification methods may be used for clarification. These might include flame atomic absorption spectrometry and mass spectrometry methods such as inductively coupled plasma mass spectrometry.

6.6.3.2 Geometry.

Examination of the geometry of the Ca^{2+} sites shows the characteristic hard Ca^{2+} binding ligands; aspartic acid and asparagine residues (Figure 6-22).

The M(II) Ca^{2+} site in PhoD (subsequently denoted as Ca^{2+}_1) is 8-fold coordinate, with an irregular geometry, closest to a bicapped trigonal prism. Unlike the Fe^{3+} site, comparison with the PAP M(II) site

shows little similarity with coordination or residue conservation. That of the PAPs is a six coordinate $\text{Fe}^{3+}/\text{Zn}^{2+}$ site with conserved coordinating residues (Olczak *et al.*, 2003). Probably due to the increased coordination number the distances to Ca^{2+}_1 are larger than for the M(II) site in the PAPs (2.0-2.2 Å in kidney bean PAP (Klabunde *et al.*, 1996)). The distances in PhoD are shown in Figure 6-22.

The M(III) site of PhoD (subsequently denoted as Ca^{2+}_2) has no analogue in the PAPs. Ca^{2+}_2 has regular octahedral coordination with all angles approximately 90°, and all bond lengths between 2.3-2.4 Å (Figure 6-22). It is associated with one oxygen of the bound phosphate in contrast to other metals, which are coordinated by two phosphate oxygens each.

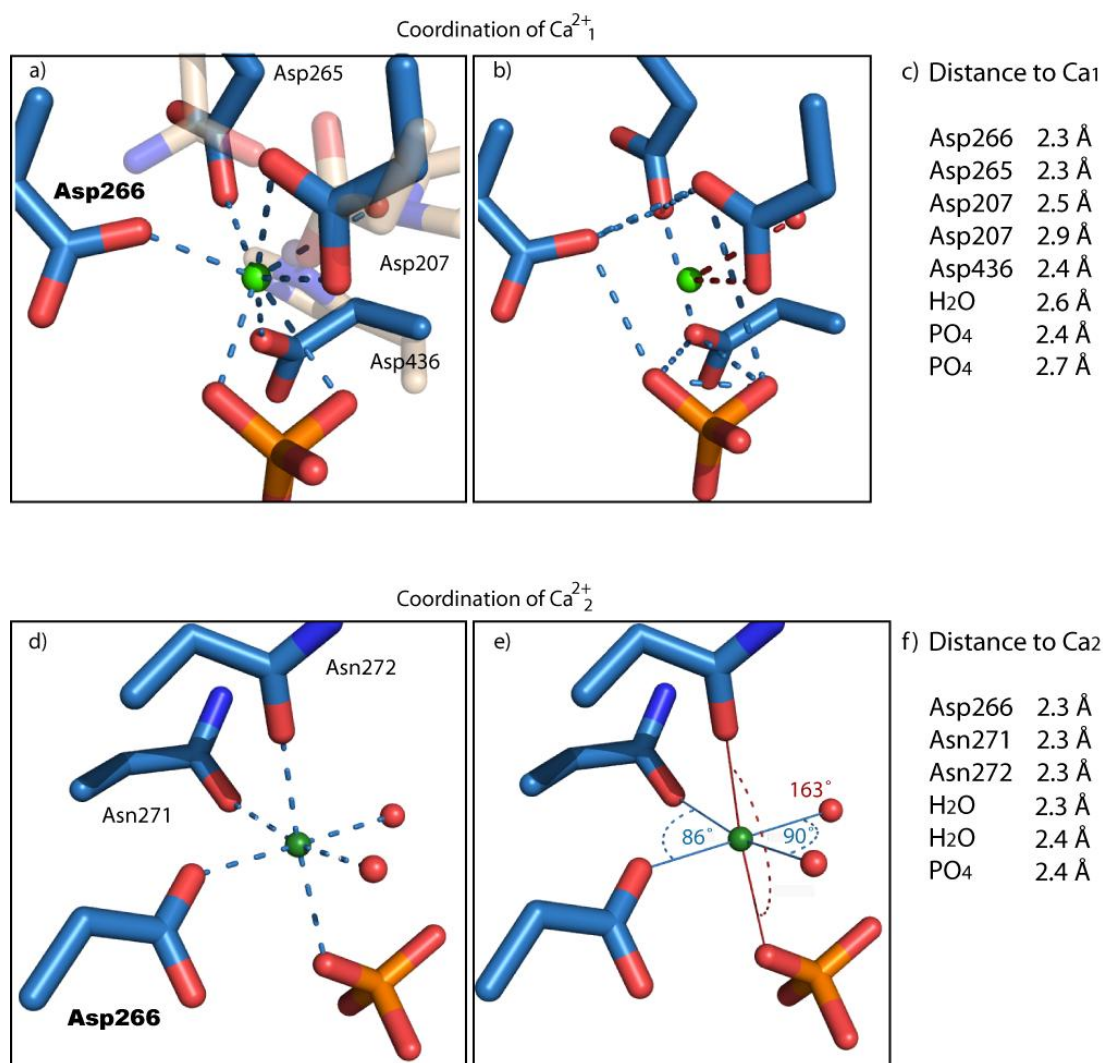


Figure 6-22 The geometry of Ca_1^{2+} and Ca_2^{2+} sites. **a)** Comparison of the Ca_1^{2+} site (coloured in blue) with the PAP M(II) (coloured in beige, background) site shows significant dissimilarity. Alignment carried out with LSQKAB (Kabsch, 1976). **b)** The Ca_1^{2+} site is irregular in shape, conforming most closely to a bicapped trigonal prism. **c)** Distances from ligand to Ca_1^{2+} . **d)** Residue coordination of Ca_2^{2+} site. **e)** The Ca_2^{2+} coordination geometry is regular hexagonal. **f)** Distances from ligand to Ca_2^{2+} .

6.7 Fe(III)-cysteine coordination.

6.7.1 The coordination rarity.

Fe(III)-cysteine coordination is common for Fe-S clusters, where delocalisation of electrons allows huge versatility of function including roles in electron transfer, catalysis and iron storage (Johnson *et al.*, 2005a). Cytochromes use a cysteine coordinated haem group for similar electron transfer reasons (Shriver and Atkins, 2002). However, solitary non-haem Fe(III)-cysteine coordination is rarer. Examples in

the PDB include the *Desulfoferrodoxins*, which are redox active superoxide reductases (Rusnak *et al.*, 2002), and a putative structural protein of *Rhodopseudomonas palustris*, whose structure is part of the ferretin superfamily (Osipiuk, (to be published)). There appears to be no relationship between the function or origins of these proteins and PhoD. The active sites and PDB codes of these examples are shown in Figure 6-23. During this study, the structure of a Fe(III)-cysteine coordinating protein was solved by our group (work of Pietro Roversi and Shee Chien Yong) (Yong, (to be published)) - PhoX of *pseudomonas fluorescens*, a Tat exported alkaline phosphatase. The active site was characterised in this study by EPR for comparison with PhoD.

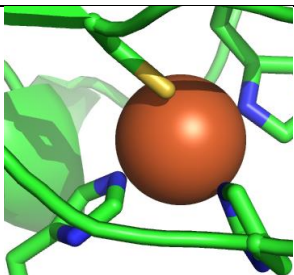
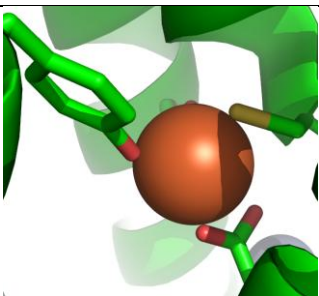
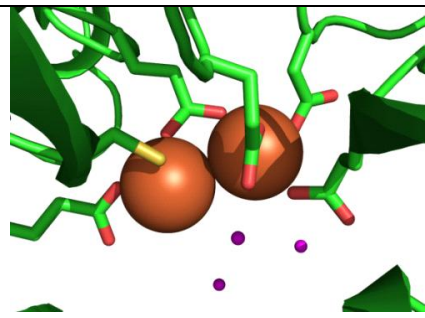
Protein	<i>Desulfovibrio desulfuricans</i> Desulfoferredoxin, PDB 1DFX (Coelho <i>et al.</i> , 1997)	<i>Rhodopseudomonas palustris</i> Ycfl, PDB 2GYQ (Osipiuk, (to be published))	<i>Pseudomonas fluorescens</i> PhoX, PDB unreleased (Yong, (to be published))
Active site			
Action	Superoxide Reductase – redox active.	Putative structural protein. Member of iron storage ferretin superfamily (redox inactive).	Alkaline phosphatase produced upon phosphate starvation – Tat exported. Purported mechanism redox inactive.

Figure 6-23 Non-haem proteins with solitary Fe(III)-cysteine coordination in the PDB.

6.7.2 PhoX.

The active site of PhoX contains two irons as shown in Figure 6-23. The cysteine coordinated Fe shall be referred to as Fe₁, and the non-cysteine Fe as Fe₂. Their presence was determined by microPIXE. EPR was required to determine the oxidation states of these ions.

Unlike PhoD, the EPR spectrum of PhoX showed no peaks. Partial reduction with sodium dithionite led to the occurrence of a peak at $g = 4.3$ with 5 mM dithionite. Further reduction with the addition of 20 mM dithionite led to near loss of this peak.

In a separate experiment, the addition of 50 mM EDTA in the absence of air (to prevent accidental oxidation) led to the occurrence of the same $g = 4.3$ peak. The EPR spectra for both sodium dithionite and EDTA EPR experiments are shown in Figure 6-24. EPR operating conditions were as for PhoD, described in Table 6.1.

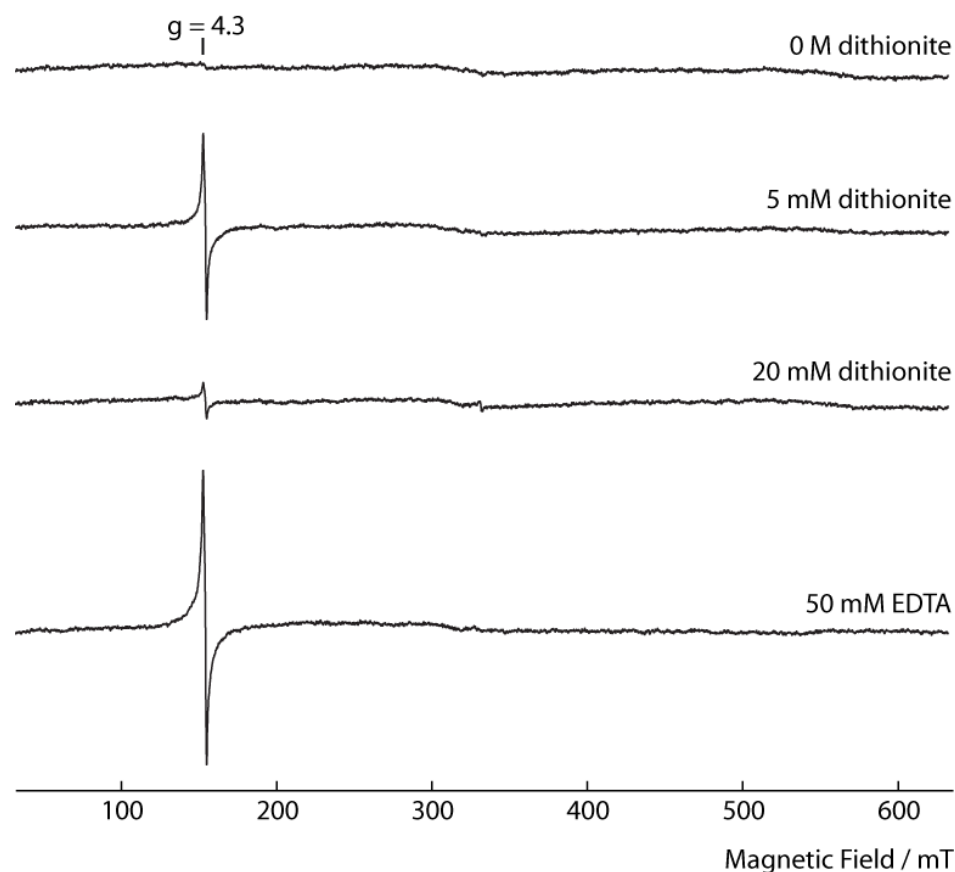


Figure 6-24 EPR spectra of PhoX. Spectra recorded with a concentration of 10 mg ml^{-1} in 20 mM MOPS pH 7.5, 200 mM NaCl, 30% glycerol (v/v). Spectra are shown with no additions, 5 mM sodium dithionite, 20 mM sodium dithionite, and 50 mM EDTA. EPR operating conditions: microwave power, 2 mW; microwave frequency, 9.425 GHz; modulation, 5.0 G at 100 kHz; temperature 10 K. Spectra are buffer subtracted with negative controls taken for sodium dithionite and EDTA solutions.

The results indicate that in the native state, both PhoX Fe ions are ferric: The difference between the EPR spectra of PhoD and PhoX highlights the role of the second Fe in PhoX. Whereas a solitary Fe(III) in PhoD gives a peak at $g = 4.3$, the Fe(III)Fe(III) couple bridged by the carboxylate E194 antiferromagnetically interact to remove the EPR signal, as shown in other examples such as the ferroxidase centre in bacterial ferritin (Le Brun *et*

al., 1995). Whilst the same absence of signal would be expected in an Fe(II)Fe(II) couple, the appearance of a high spin Fe(III) signal upon partial reduction with sodium dithionite precludes this possibility.

The appearance of the $g = 4.3$ peak upon addition of sodium dithionite implies that with 5 mM dithionite, the two Fe ions become decoupled. This is supported by UV–Visible spectroscopy data (Figure 6-25), showing that when treated with 5 mM dithionite, the absorbance of purple PhoX shifted from 550 to 575 nm, visually observed by a colour change to light blue. Further addition of sodium dithionite led to bleaching of the PhoX sample.

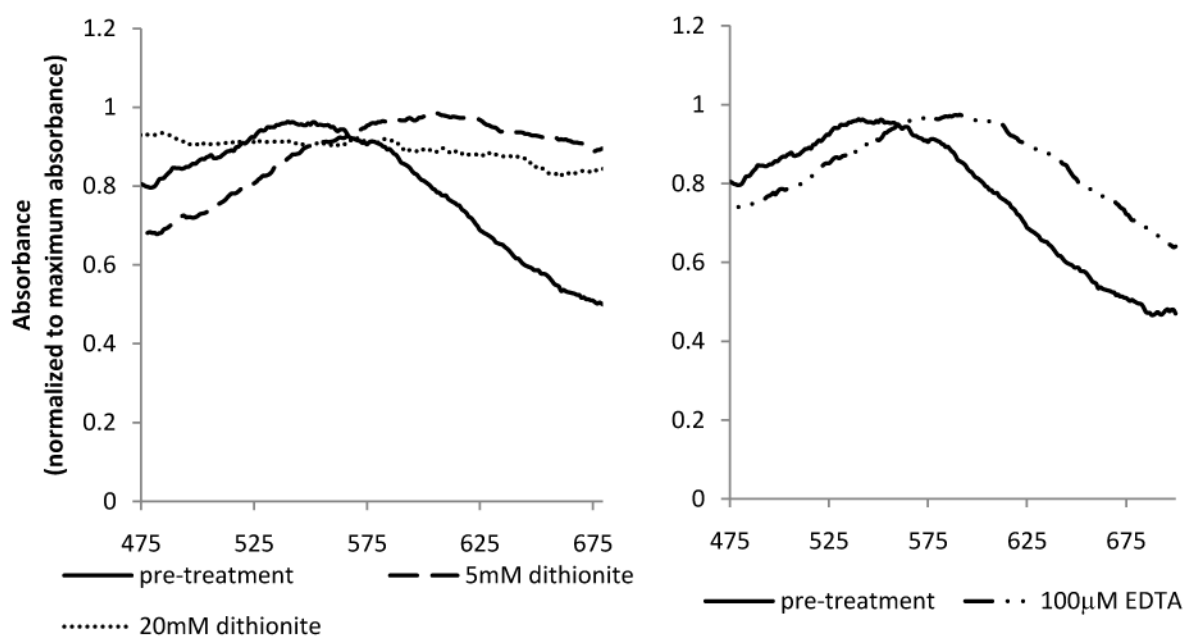


Figure 6-25 UV-visible wavelength absorption spectra of PhoX upon addition of sodium dithionite and EDTA (Shee Yong).

The purple colour of native PhoX is expected to arise from the thiolate-Fe₁(III)Fe₂(III) charge transfer band (Thompson, 2010), since Fe(III)-carboxylate charge transfer bands are in the near UV range (Le Brun *et al.*, 1995). Reduction to the blue colour indicates that the thiolate-Fe₁(III) charge transfer band stays intact, as shown by other blue solitary thiolate-Fe(III) structures such as the Neelaredoxins (Jovanovic *et al.*, 2000; Silva *et al.*, 2001) This implies a reduction of Fe₂ from oxidation state (III) to (II), an event which appears to labilise this ion, leading to a $g = 4.3$ EPR signal from the remaining Fe₁(III).

Such labilisation after reduction of other non haem Fe are found in the literature (Keough et al., 1980; Yasmin et al., 2011).

The postulate that the oxidation states of Fe₁ and Fe₂ are (III) in the absence of any additives, and that Fe₂ is more labile than the Fe₁ is supported by the PhoX-EDTA EPR experiment. Here the same EPR signal corresponding to the ferric state (Figure 6-24) is seen, and the same blue colour shift indicating loss of only Fe₂ occurs (Figure 6-25). The B-factors for the metals Fe₁ and Fe₂ in the crystal structure are 19.6 Å² and 21.1 Å² respectively. Soaking of crystals in sodium dithionite (50 mM concentration, 10 minutes prior to cooling) before X-ray diffraction led to loss of atomic resolution at the active site, implying metal mobility. Addition of sodium dithionite or EDTA to PhoX removed alkaline phosphatase activity (Yong, (to be published)).

6.7.3 Fe(III)-Cys function.

Given the rarity of solitary Fe(III)-cysteine coordination, what might be the purpose of its occurrence in PhoD (and PhoX)? There are a number of potential reasons:

1: The juxtaposition of Tyr210 with Cys180 in PhoD next to a strongly polarising ion such as Fe(III) is an arrangement which can lead to covalent bonding of cysteine to the C₃ position of tyrosine. This is observed in galactose oxidase, where it lowers the tyrosine redox potential to facilitate electron transfer during the oxidation of primary alcohols (Ito *et al.*, 1991; Rogers *et al.*, 2008). There is no evidence of this occurring in the PhoD electron density however, and little expectation that it might occur since the reaction is redox inactive.

2: The Lewis acidity of Fe(III) and thus the effectiveness for hydrolysis is altered by the chemical nature of the ligands and their geometries. The importance of cysteine to the Fe(III) Lewis acidity of this catalytic site is reflected in the phospholipase PLD of *S. chromofuscus*, which has the same active site as PhoD by sequence alignment (overall sequence identity 31%, catalytic site alignment details given in

Figure 6-19). Yang and Roberts showed that mutation of the Fe(III) coordinated cysteine to a serine led to loss of phosphatase activity (Yang and Roberts, 2002). In both PhoD and PhoX, the cysteine-metal-OPO₃ arrangement is linear. This will allow a particular arrangement of the molecular orbitals such that charge donation from the readily polarised sulphur may allow more effective transferral of the hydroxide (Shriver and Atkins, 2002), in a manner that a harder anion such as a serine oxygen would not allow.

3: Both PhoX and PhoD are Tat exported Fe(III)-Cys coordinated species. The PSI-BLAST PhoD sequence alignment shows that both the cysteine and potential Tat signal sequence are widely conserved among the alkaline phosphatases (Figure 6-19), and that they are both features of the above mentioned *S. chromofuscus* PLD. A number of mechanisms are debated for Tat discrimination of correctly folded proteins for transport, including substrate specific chaperones and detection of hydrophobicity (Robinson *et al.*, 2011). Despite the lack of agreement into the mechanism of proofreading, it has been found that the correct folding of a number of Fe-S cluster proteins is a requirement for their transport by the Tat pathway (Matos *et al.*, 2009; Matos *et al.*, 2008). Such a criterion could explain the impetus for conservation of Fe-S in PhoX and PhoD, in order that they be transported by an apparatus that ensures their correct folding.

6.8 A C-terminal α -helix occludes the active site.

A number of alkaline phosphatases exhibit conserved features with PhoD, such as the active site, Tat signal sequence and expected global structure. This was shown above in Figure 6-19. One novel feature of PhoD however is the clear presence in the electron density of a C-terminal α -helix (residues 560-577) which occludes the active site. This feature is not present in the purple acid phosphatases, nor in the *non-Bacilli* phosphodiesterase/alkaline phosphatases which have high sequence similarity to PhoD. The only sequences which are not truncated before this helix are five other *Bacilli* PhoD sequences. Figure

6-26 compares the final section of the PSI-BLAST sequence alignment of these *Bacilli* and non-*Bacilli* PhoD sequences.

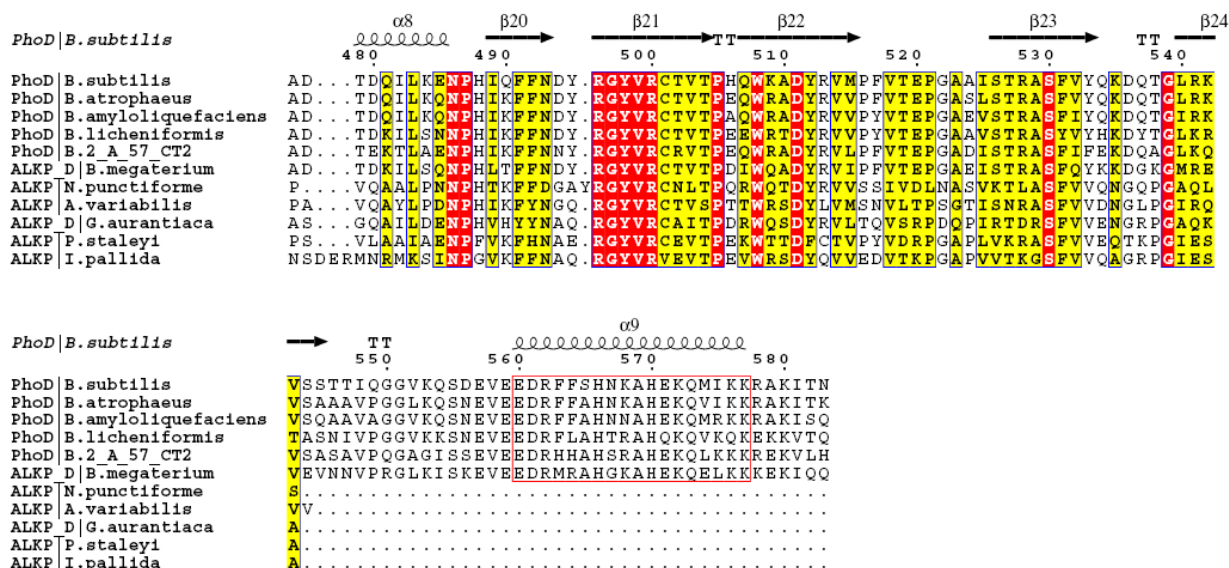


Figure 6-26 Non-conservation of the PhoD C-terminal α-helix. Final part of the multiple sequence alignment from Esript (Gouet *et al.*, 2003), highlighting the presence of a C-terminal α-helix in *Bacilli* sequences via a red box which is absent for sequences of any other bacteria.

Figure 6-27 shows the occlusion of the active site by the α-helix, as well as the charge environment at this position. Unsurprisingly, the environment coordinating the metal ions is negatively charged, but this negative charge continues from the active site down the length of the site in which the α-helix sits, allowing interactions between positively charged groups on the α-helix with the residues which lie beneath.

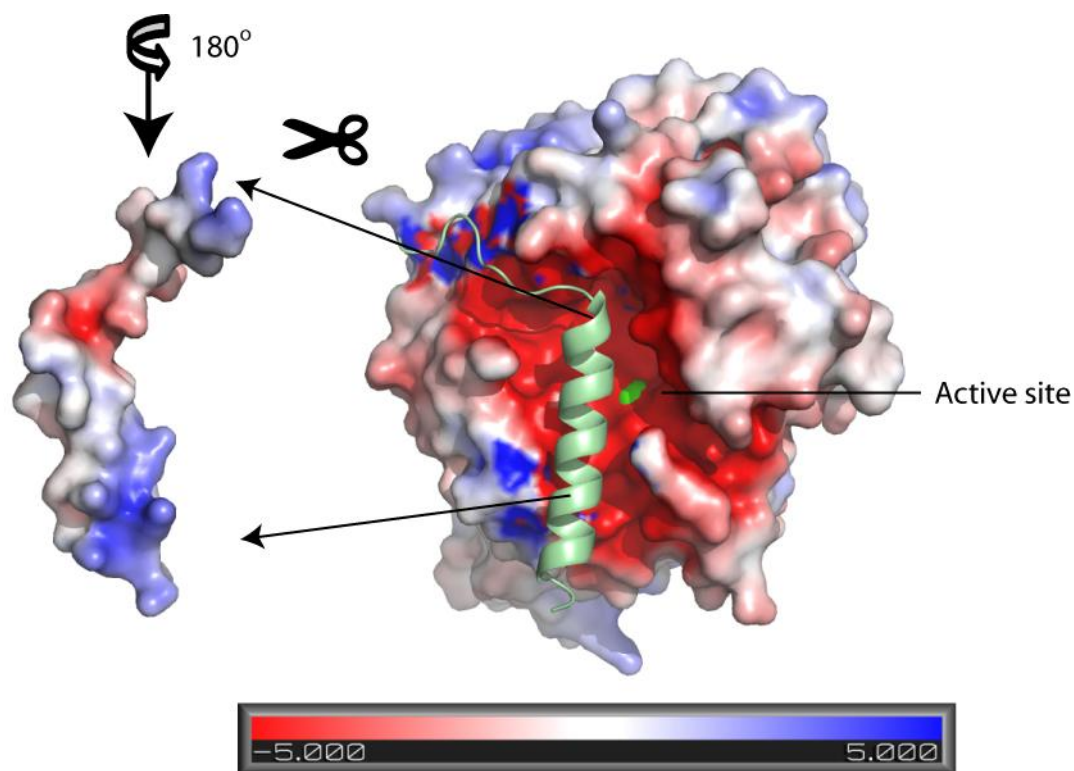


Figure 6-27 Electrostatic surface model of PhoD highlighting the region of the C-terminal α -helix. Regions of negative charge are shown in red and regions of positive charge indicated with blue. The charges of the metals in the active site are not included. The density of the helix has been translated to allow the active site to be viewed and rotated by 180° to allow observation of the positively charged residues which the underside contains. The C-terminal α -helix lies upon a region of negative charge which contains the active site. This region is completely occluded by the helix in the crystal structure.

Salt bridges are found between the helical Arg562, Lys576 and non helical Asp436, Glu216 respectively, as shown in Figure 6-28. These interactions may have led to the stabilisation of the α -helix which enabled crystals to form.

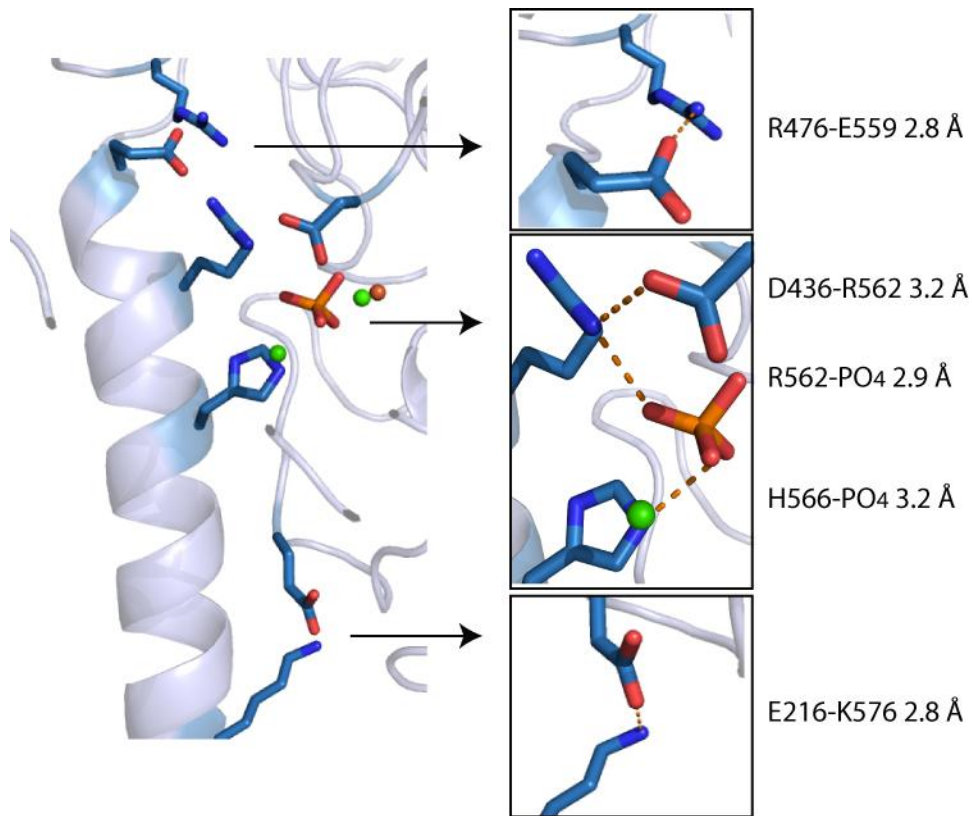


Figure 6-28 Salt bridges to the main chain and interactions to the phosphate formed by α -helical residues.

The presence of the α -helix led to questions as to its purpose and the physiological relevance of the crystal structure. Whereas the crystal structure shows the helix occluding the site, there is the possibility that in solution it is more mobile, and opens to allow a substrate to enter to the active site. This would correlate with the ribbon diagram of Figure 6-29, where the girth of the protein ribbon represents the B-factor.

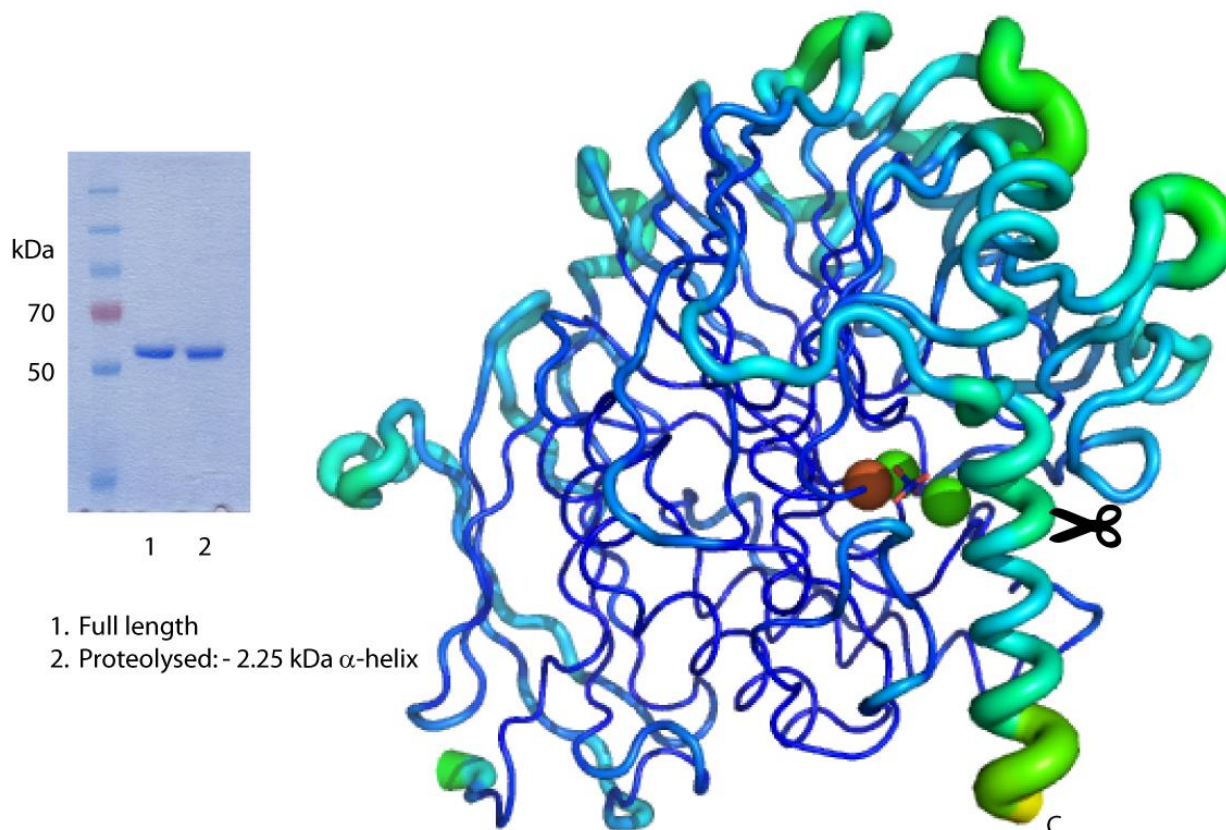


Figure 6-29 Flexibility of the C-terminal α -helix. B-factor putty diagram generated by Pymol, which indicates that the C-terminal α -helix is among the least rigid parts of the crystal structure. Flexibility is indicated by ribbon girth, and increasing lightening of colour. Part of the helix may be removed in a partial proteolysis experiment, as shown in the reducing SDS-PAGE of the experiment. The chymotrypsin cleaved section is indicated by scissors on the B-factor putty diagram.

It is clear that towards the end of the C-terminal α -helix, the temperature factors of the residues increase, and the chain as a whole is more mobile than most other parts of the structure. Specifically, whereas the average mainchain B-factor for the whole protein is $28.2 \text{ \AA}^2$, that of the α -helix is $49.1 \text{ \AA}^2$. This is despite the aforementioned salt bridges between the α -helix and the main structure (Figure 6-28). The full length construct is able to function as an alkaline phosphatase. This was shown by the pNPP assay discussed in 6.2.1, as well as the malachite green molybdate assay – a chemical which reacts with freed phosphate and absorbs at 620-640 nm (work of Fernanda Rodriguez). Manual insertion of pNPP into the crystal structure shows that in its current position, the α -helix prevents access to the catalytic site, and even when in position (superimposed onto the observed phosphate) pNPP would

clash with residues Arg562 and His566 as shown in Figure 6-30. It therefore appears intuitive that movement must be possible.

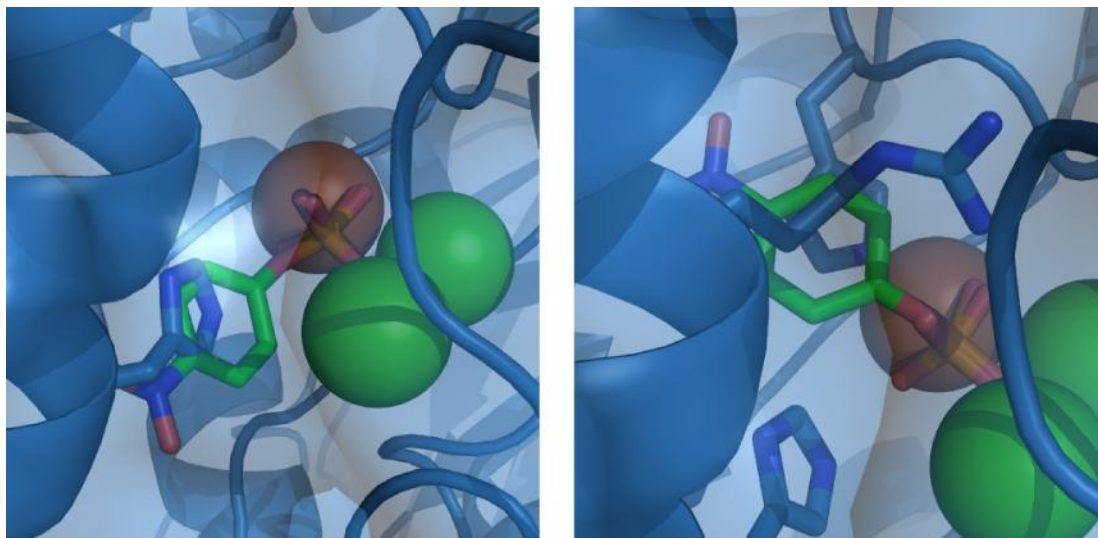


Figure 6-30 Clashing of pNPP with C-terminal α -helix residues in possible orientations of pNPP at the catalytic site. pNPP (green, PDB obtained from HICUP) was manually superposed onto the phosphate shown in the crystal structure. The α -helix is on the left hand side.

Crystal trials attempting to observe this movement in a structure were set up. The α -helix residue Arg562 directly interacts with the bound phosphate, so it was hoped that replacement of phosphate in the buffer with more bulky non-cleavable phosphate derivatives would force movement of the helix and form a co-crystal containing these substrates. Phosphonoacetic acid and phosphonohexanoic acid were selected, since these were not cleaved by PhoD but were found to bind, as detected by inhibition of phosphatase activity on pNPP. Crystals were grown in solutions mimicking the original, except for the replacement of 0.1 M sodium/potassium phosphate with 0.1 M phosphocarboxylic substrate. In case of the acidic crystallisation pH altering substrate binding 20 mM carboxylic acid substrate was also added to the protein (in non-phosphate containing buffer) at pH 7.5. Upon diffraction however the crystals showed a lack of any of the attempted alternative derivatives. Instead, it proved impossible to remove the phosphate, which must have been obtained from the growth medium.

Whilst movement of the α -helix might enable AP activity, there is the possibility that the native protein has the helix removed post transport, allowing a greater level of activity. A 'repression loop' mechanism is used by the mammalian purple acid phosphatases. A loop containing an aspartic acid residue interacts with the Fe(II) cofactor, partially occluding the active site. This does not silence activity completely, but its removal by a cathepsin cysteine protease increases phosphatase activity by four-fold (Strater *et al.*, 2005). Attempts to make a series of C-terminally truncated deletion mutants, removing four residues for each successive mutant, led to inactive colourless protein with very poor yields for any mutants past residue 577 (the end of the helix). This indicates a loss of cofactor binding. A pivotal role for the α -helix may therefore be keeping the metals bound, however the loss of activity may be indicative of a more global folding issue when expressing the truncated construct.

Comparison with the literature revealed however that upon cleavage of the Tat leader sequence, mature PhoD was previously determined by N-terminal sequencing to have a molecular weight of 56 kDa (Eder *et al.*, 1996). This is 3.7 kDa shorter than the construct used in this study, which was recombinantly expressed. Whereas the N-terminus of our construct matches that of the N-terminal sequencing, no details of the C-terminal sequence are given. It appears possible then, that removal of the C-terminus may be a post translational modification. Partial proteolysis supports this, with mass spectrometry showing a mass decrease of 2.25 kDa upon incubation with chymotrypsin. This weight reduction corresponds to removal of the amino acids from Ser565 onward, cutting at the carboxyl side of Phe564. A reducing SDS-PAGE in Figure 6-29 shows that the rest of the protein was resistant to proteolytic degradation, a characteristic of many *B. subtilis* proteins (Jongbloed *et al.*, 2000). Future work will determine whether the proteolysed protein retains activity.

6.9 PhoD mechanism.

There are a number of general features which metallophosphatase mechanisms share: Firstly, the decreased pKa of a water bound to the metal ion allows hydroxide to be used as a nucleophile at physiological pH. Secondly, metal binding of the phosphate bound oxygens enables greater electrophilic susceptibility of the phosphate centre. Thirdly, metals are used to assist the departure of a leaving group by stabilising the partial negative charge in the transition state (Kimura, 2000).

In this section, insight into the potential mechanism of PhoD phosphatase activity shall draw upon knowledge of the active sites and mechanisms of other known phosphatases. The mechanisms described are shown in Figure 6-32.

6.9.1 Alkaline phosphatase mechanism.

The catalytic site of the *E. coli* alkaline phosphatase is widely conserved, and is found in PhoA and PhoB of the *B. subtilis* Pho regulon (Hulett *et al.*, 1991), which are expected to catalyse phosphomonoester hydrolysis under the same well studied mechanism (Holtz and Kantrowitz, 1999; Kim and Wyckoff, 1991; Stec *et al.*, 2000). Homodimeric *E. coli* PhoA has a three metal catalytic site, containing Zn^{2+}_1 , Zn^{2+}_2 and Mg^{2+} . Prior to attack, the substrate coordinates Zn^{2+}_1 and Zn^{2+}_2 . A serine nucleophile, which is coordinated by Zn^{2+}_2 and deprotonated by Mg^{2+} , attacks the substrate, whose leaving group departs. A second nucleophile, a water coordinated to Zn^{2+}_1 , attacks the serine-phosphate intermediate, whereupon the serine-phosphorus covalent bond breaks. The hydrolysed product leaves upon either its protonation, or the protonation of the serine nucleophile, or it leaves due to the increased mobility of Arg166. The rate determining step for the reaction under alkaline conditions is the leaving of the product phosphate (Hull *et al.*, 1976).

Whereas the *B. subtilis* alkaline phosphatases PhoA and PhoB share the catalytic site of *E. coli* PhoA, containing the serine nucleophile, PhoD does not. Other phosphatases, including R/PMH of *Rhizobium*

leguminosarum (Jonas *et al.*, 2008), use cysteines as the basis of their nucleophilic attack. Three points indicate that this is not the case in PhoD: Firstly, the aforementioned cysteine mutation to serine in homologous *S. chromofuscus* PLD would not be expected to fully remove phosphatase activity if this residue were behaving as a nucleophile. Secondly, the crystal structure shows that the phosphate moiety is inaccessible to the cysteine, due to Fe^{3+} lying in the line of sight. This was confirmed in the third point, where the inaccessibility of cysteine to a chemical approximately the size of pNPP, the spin label MTSL, was found. The EPR spectra for equimolar amounts of PhoD (following the spin labelling protocol as described in section 2.7.1.2) and Tempol demonstrate the lack of spin labelling, a result confirmed by mass spectrometry.

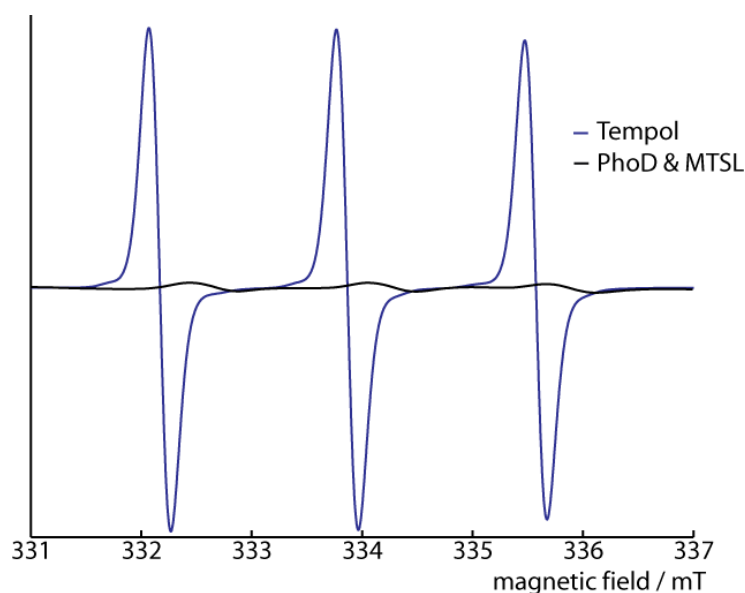


Figure 6-31 EPR spectra of 100 μM Tempol and spin labelled PhoD solutions at room temperature. EPR conditions: microwave power, 0.63 mW; microwave frequency, 9.374 GHz; modulation, 5.0 G at 100 kHz. Other parameters as described in Table 3.1. No EPR signal is observed from the PhoD & MTSL sample.

The mechanism of PhoD phosphoester hydrolysis is therefore unlikely to be that of the alkaline phosphomonoesterases PhoA and PhoB.

6.9.2 Purple acid phosphatase mechanism.

The bimetallic (Fe(III) M(II)) purple acid phosphatase mechanism does not involve direct covalent linkages with amino acids of the protein. The precise stepwise details are contested, with two possible mechanisms. The first involves binding of the substrate to M(II) and nucleophilic attack by an Fe(III) coordinated hydroxide, before release of the leaving group. This mechanism has been reported by EPR, where phosphate was observed to bind the M(II) only, and it is thought that only a terminally coordinated hydroxide would be a strong enough nucleophile to attack the phosphorous centre (Kimura, 2000; Merkx *et al.*, 1999). The second mechanism invokes binding of the substrate to both metals prior to attack by a μ -OH bridge, and completion as before. This mechanism is supported by electron nuclear double resonance and a crystal structure with a fluoride analogue of the μ -OH bridge (Schenk *et al.*, 2008; Smoukov *et al.*, 2002). At pH 7, the rate limiting step of the PAP hydrolysis mechanism is the hydrolysis of the substrate, and not the release of phosphate (Merkx *et al.*, 1999).

The similarities between PAP and PhoD active sites make a similar mechanism probable. For a phosphodiesterase reaction such as in PhoD, it is thought that the single anionic charge of the substrate makes the second putative PAP mechanism less likely, with the diester unable to form a bridge between two metal ions (Richter, 2002). The prospect of the first putative mechanism, with coordination to the Ca^{2+}_1 prior to attack by a Fe(III) coordinated hydroxide is supported by mutational work on the PhoD homologue PLD (Zambonelli *et al.*, 2003). Here, phosphodiesterase activity was decreased more significantly than monoesterase activity upon mutation of the PLD equivalent of the Ca^{2+}_1 coordinating Asp436 ligand.

The role of Ca^{2+}_2 is unknown. In a number of three metal phosphatases the third metal takes a supporting role in the mechanism (Stec *et al.*, 2000; Su *et al.*, 2011), ensuring the correct binding of a substrate to position it for attack (El Kirat *et al.*, 2002), or to help stabilise the partial negative charge in

the transition state (Zambonelli and Roberts, 2003). This may explain the shift from acid to alkaline preference of PhoD despite having a PAP like catalytic site: In the PAPs, where optimal activity is at pH 3-6 (Schenk *et al.*, 2005), the rate-limiting step is reported to be the hydrolysis of the substrate. This may accelerate upon a shift to alkaline conditions to improve the quality of the nucleophile – above pH 4.9 the predominant species will be Fe(III)-OH^- rather than $\text{Fe(III)-H}_2\text{O}$ (Aquino *et al.*, 1994). However hydrolysis also depends on the ability of the leaving group, which is impaired in more basic conditions. This detriment to the substrate leaving group on becoming more negatively charged may be overcome by addition of Ca^{2+}_2 to help stabilise the transition state. The balance between hydrolysis dependence and product leaving is thus achieved at more alkaline conditions than for the PAPs.

In the PhoD structure however, Ca^{2+}_2 is observed directly bound to the phosphate. The tripodal binding of the phosphate to Fe^{3+} and Ca^{2+}_1 as seen in PAP structures combined with the extra coordination to the Ca^{2+}_2 indicates a more important role of the third metal than for previously described mechanisms. This is supported by the lack of expression of Ca^{2+}_2 coordinating residue mutants.

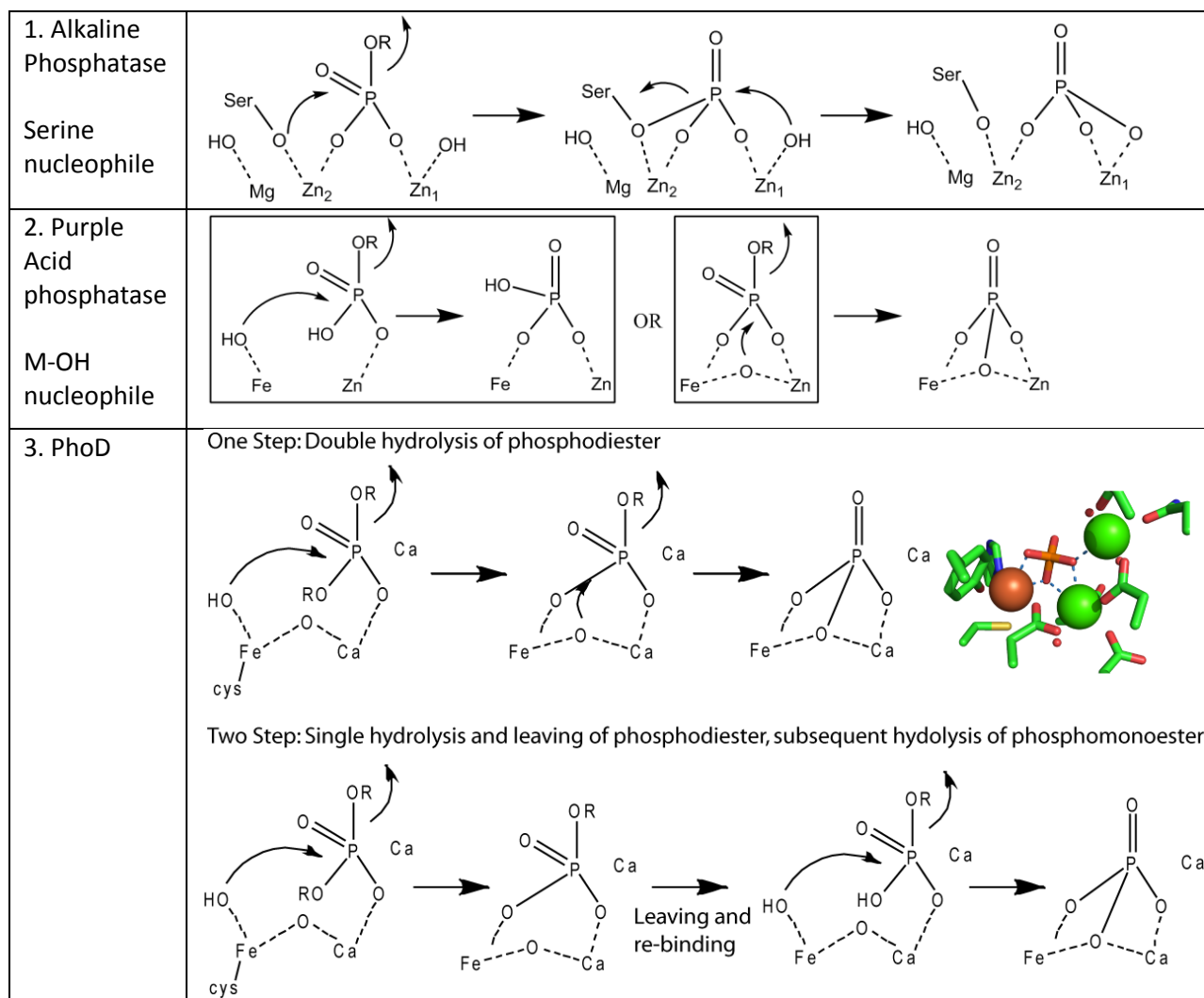


Figure 6-32 Stepwise mechanisms for *E. coli* alkaline phosphatase, kidney bean phosphatase and PhoD. There is no evidence for a one step vs. two step mechanism for the PhoD phosphodiester hydrolysis.

6.10 PhoD target.

As part of the PhoD study, the ability of PhoD to cleave a variety of substrates was examined. Fernanda Rodriguez showed that PhoD can cleave a wide range of substrates, including phosphomonoesters (*para*-nitrophenylphosphate (pNPP), Phosphoenolpyruvic acid), phosphodiester (Bis (p-nitrophenyl) phosphate) and nucleotides (Adenosine-5'-triphosphate), but not pseudophosphate substrates such as the phosphonocarboxylic acids. As mentioned above, a number of the phosphonocarboxylic acids were used as a substrate mimic in crystallography attempts, but without success.

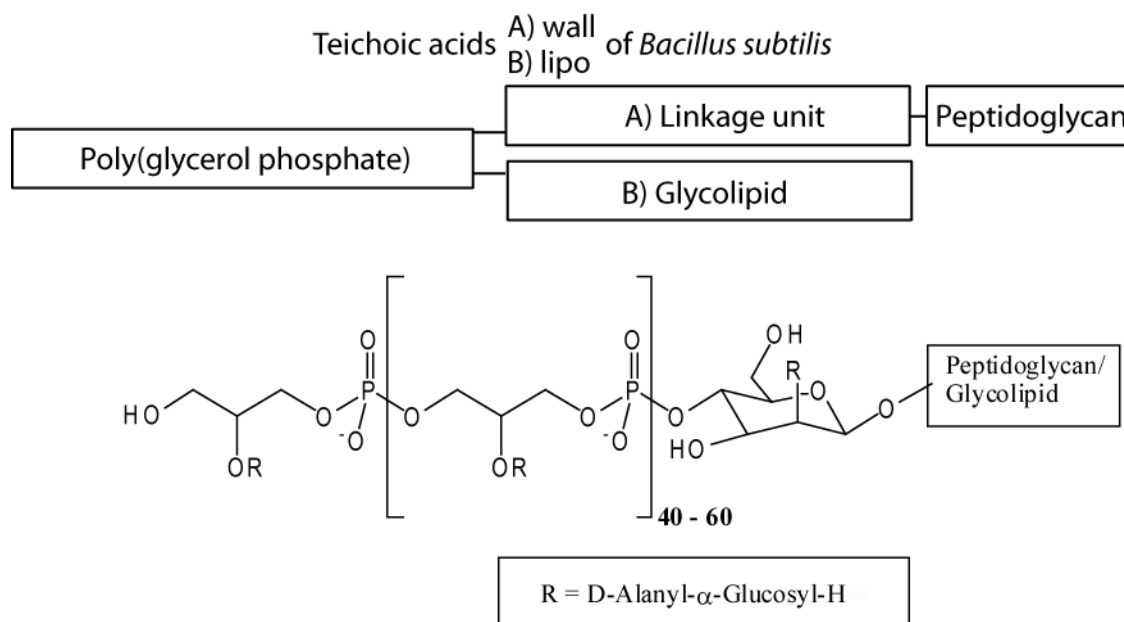


Figure 6-33 *B. subtilis* teichoic acid structure. Teichoic acid is the putative target for PhoD (Eder *et al.*, 1996). While PhoD is *able* to free phosphate from this, the physiological role could equally be to break one of the phosphoester bonds, allowing the more ubiquitous PhoA/PhoB monoesterases to complete the task.

A broad substrate range is a common feature of both alkaline phosphatases and the PAPs, making identification of the intended physiological target difficult (Kim and Wyckoff, 1991; Zambonelli and Roberts, 2003). It is interesting to note however that this broad substrate range does not extend in the PAPs to phosphodiesteres, except for in a few cases where steric inhibition of this hydrolysis is relieved due to the presence of a small second ester group (Cox *et al.*, 2007). Despite the promiscuous nature of the phosphatases, the suggestion of teichoic acid (Figure 6-33) as the substrate for PhoD was investigated. Figure 6-34 shows that PhoD does indeed cleave teichoic acid. The assay is imperfect since phosphate release is measured, which does not account for a composite role of PhoD (phosphodiesterase activity) with PhoA/PhoB (phosphomonoesterase activity) in teichoic acid breakdown, but does provide qualitative evidence that PhoD is able to cleave this substrate.

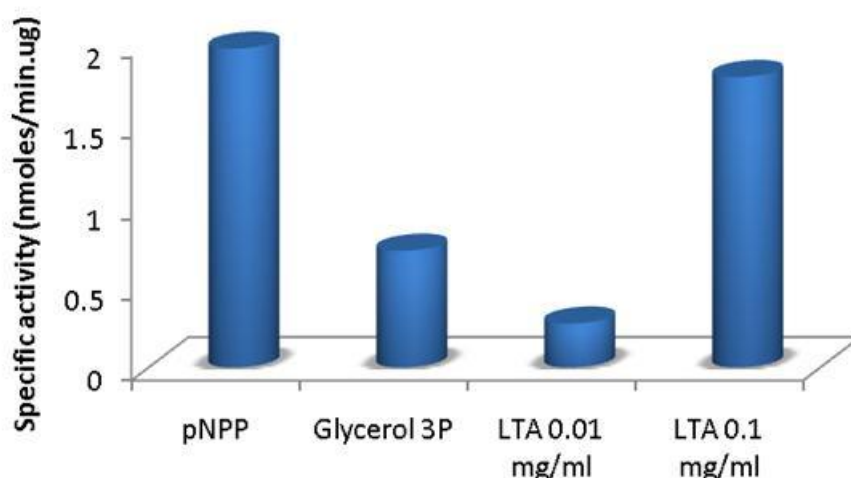


Figure 6-34 Non-specificity of PhoD. The ability of PhoD to cleave para-Nitrophenylphosphate, glyceroltriphosphate, and lipoteichoic acid. Malachite green-molybdate assay performed by Fernanda Rodriguez.

Support for teichoic acid as the physiological target of PhoD is provided by the work of Muller *et al.*, who demonstrate the association and clustering of PhoD with the cell wall by immunogold labelling. This is postulated as occurring due to inefficient cleavage of the unusually long signal sequence upon Tat secretion of PhoD. Furthermore, experiments upon bis-NPP showed that PhoD was able to act as a phosphodiesterase in this uncleaved state, enabling it to potentially act whilst anchored in position at the site of the teichoic acids (Muller and Wagner, 1999).

6.11 Conclusions.

PhoD of *Bacillus subtilis* is an alkaline phosphatase/phosphodiesterase which is expressed by the Pho regulon under phosphate limiting conditions. In this chapter, the solution state protein was first characterised, showing that PhoD exists as a monomer. Secondly, the structure of PhoD has been solved by X-ray crystallography using the anomalous signal from the iron which it contains in its active site. PhoD is a member of the metallophosphoesterase family, with a common fold adopted by a number of other phosphatases. Of these phosphatases, PhoD shows particular structural similarity to the eukaryotic purple acid phosphatases, despite biochemical activity differences between them.

Part of this structural similarity extends to the Fe(III) site, which shows remarkable similarity to that of the purple acid phosphatases. The six fold coordination is almost identical, including the tyrosine residue which gives PhoD the same characteristic purple colour as the PAPs. Where this similarity deviates however is the presence of an unusual Fe(III)-Cys coordination. Previous sequence alignment work which anticipated this coordination in *S. chromofuscus* PLD showed that this residue is required for enzymatic function, since mutation to serine led to loss of activity. This unusual combination was also shown in *P. fluorescens* PhoX, and may furthermore provide an explanation into the use of the Tat system in PhoD/PhoX export.

The similarity with the PAPs however does not extend to two calcium ions which complete the active site in PhoD, providing the first structure for a conserved catalytic site found in many alkaline phosphatases from varying species. These were determined by a combination of EPR, X-ray crystallography, supplementation assay and microPIXE methods. Use of Ca^{2+} in the catalytic site is unusual in phosphatases, with only a handful of other Ca^{2+} structures deposited in the PDB (Kristensen *et al.*, 2004; Kumaran *et al.*, 2006).

This catalytic site affords PhoD phosphodiesterase activity, whereas the other alkaline phosphatases (PhoA/PhoB) regulated by the Pho regulon of *B. subtilis* show monoesterase activity only. For PhoA/PhoB this activity most likely stems from a covalent phosphoserine intermediate, from homology with *E. coli* PhoA, however the mechanism of PhoD action is likely to be more similar to that of the PAPs, not involving protein residues directly.

A number of questions remain over the structure and function of PhoD. The first concerns the role of the α -helix at the C-terminus which occludes the active site. Activity is not precluded by this cap, leading us to believe that it must move, although attempts to truncate the helix via mutants or via a TEV cleavage site led to poor expression of colourless protein. Future work using PhoD expressed directly in

B. subtilis will observe whether any post-translational cleavage of the helix occurs, as is suggested possible by a partial proteolysis experiment with chymotrypsin. The definitive physiological role of PhoD is closer to being determined. The phosphodiesterase activity of PhoD extends to the ability to cleave the teichoic acids of their cell wall, leading to the possibility that PhoD is expressed under phosphate starvation to free these structural phosphates. This has however not been observed directly in *Bacillus*. A cartoon of this action is shown in Figure 6-35.

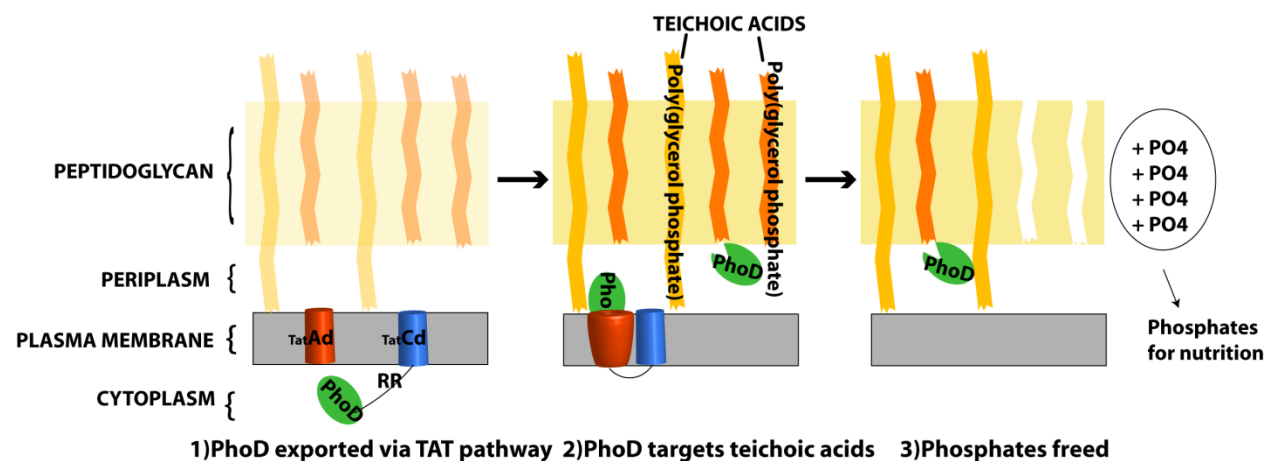


Figure 6-35 Cartoon of the putative function of PhoD. After export by the Tat pathway, it remains located at the cell wall where it may cleave teichoic acid.

A recent study of methicillin resistant *Staphylococcus aureus* showed that bacteria deficient in wall teichoic acids were made susceptible to β -lactam antibiotics (Campbell *et al.*, 2011). Whilst native PhoD is certainly not specific enough to be considered to have therapeutic utility, the ability to manipulate teichoic acids could prove useful for such a reason.

7 Overall Conclusions.

The first section of this thesis concentrated on the *S. flexneri* chaperone Spa15 and its effectors. Spa15's role as a chaperone is multifaceted, changing according to the requirements of its substrates. The promiscuous chaperone binds numerous effectors. Even when the roles of the effector are similar (e.g. host cytoskeleton subversion by the effectors IpaA, IpgB1 and IpgB2), the chaperone is required for different functions. For example: IpgB1 requires Spa15 for cytoplasmic stability (Hachani *et al.*, 2008). This requirement may be the basis for a recombinantly expressed Spa15:IpgB1 complex being soluble when the solitary IpgB1 is insoluble. Conversely, the effector IpgB2 only requires Spa15 for secretion and not cytoplasmic stability, despite being structurally and functionally related to IpgB1 (Hachani *et al.*, 2008). As a result, IpgB2 is still insoluble when coexpressed with Spa15, in contrast to IpgB1.

Spa15 therefore appears less important in cytoplasmic stability with some substrates, being necessary for secretion only. Despite these differences, in binding each of its substrates Spa15 performs a regulatory role. The maintenance of effector secretion competency is vital for a functioning T3SS, and chaperones are widely found to ensure this maintenance, be it through prevention of aggregation, proteolysis or targeting to the translocon.

Above the common regulation by a T3SS chaperone, Spa15 has an additional regulatory role due to its binding of OspD1. This effector is different from IpgB1 and IpgB2 in that it fulfils its function, the binding and controlling of the transcriptional activator MxiE, while in the bacterial cytoplasm. As a result it functions whilst still bound to Spa15. In the work presented, Spa15 was necessary to achieve the binding of OspD1 to MxiE in a soluble complex. OspD1:MxiE binding prevents the constitutive expression of protein effectors in the T3SS, saving large amounts of energy and resources for the dormant bacteria. The timing of protein expression and secretion are vital; MxiE regulates proteins which are important later in infection, but are somewhat useless in the early stages before *Shigella* internalisation. An

example of a MxiE regulated late effector is IpaH9.8, which is targeted to the host nucleus (Toyotome *et al.*, 2001).

Effector proteins are removed from their chaperones at the base of the T3SS via a mechanism which has not yet been fully elucidated. From homology with *Salmonella* however, effector removal probably involves the ATPase Spa47. Transport of the partially unwound effectors into the host cytoplasm may ensue, before effector assembly and host subversion. The effect of IpgB1, IpgB2, and other members of a group with a common WxxxE structural motif on the host is that of actin polymerisation and membrane ruffling. This is a phenotype commonly associated with Rho GTPase activation.

In this study, a member of the WxxxE group which causes stress fibres, EspM2 of EHEC, has been shown to activate the RhoA GTPase (Arbeloa *et al.*, 2010). The binding of EspM2 to RhoA was presented, a result in agreement with the finding that EspM2 induces nucleotide release from RhoA. This has enabled the addition of EspM2 to the members of the WxxxE group already defined as being guanine exchange factors for Rho GTPases. These prokaryotic GEFs functionally mimic the role played by eukaryotic GEFs, but with an unrelated structure. The most recent addition to the prokaryotic GEF group is IpgB2 of *S. flexneri* (Klink *et al.*, 2010). Crystal structures of IpgB2 with its target GTPase RhoA provide further insight into the mechanism of EspM2, since the homologous proteins share 43% sequence identity.

Effectors therefore function within the host, and so are passed via a single step mechanism through the T3SS which serves this requirement. Chaperones are used prior to effector secretion to keep the effectors competent for transit. In a different way, the Tat pathway also caters exactly to the requirements of its substrates.

In the second section of this thesis the structure of an alkaline phosphatase, PhoD of *B. subtilis*, has been elucidated. In common with a number of Tat exported substrates, PhoD contains metal cofactors. One of these is an Fe(III) ion, which is coordinated by a cysteine residue. This is an unusual coordination,

however it is observed in a second Tat exported alkaline phosphatase – PhoX of *P. fluorescens*. Coordination of this cofactor in the cytoplasm may be required, making the Tat pathway necessary for the export of PhoD and PhoX.

The structure of PhoD is typical of many alkaline phosphatases, with a double β -sheet backbone flanked by α -helices. However, it has closer structural similarity to the purple acid phosphatases than to any alkaline phosphatase yet structurally elucidated. The aforementioned catalytic site shares some similarities with the PAPs, including the Fe(III)-Tyr from which the common purple colour arises, but also shows substantive differences, including the binding of three metals.

A feature of the PhoD structure is the appearance of a C-terminal α -helix, which clamps down upon the active site and traps a phosphate molecule. Future work should elucidate the role of this helix. High B-factors in the crystal structure and the fact that the full length PhoD construct is functional indicate that the helix is mobile in solution. However, the question remains as to what role it has, and whether it is cleaved upon transport. The α -helix does bind the cofactors indirectly via the bound phosphate, but whether it has a further role in phosphatase function is unknown.

In this thesis, a number of different biophysical methods have been used to study the proteins. EPR has accompanied the more widely used biochemical methods, proving to be a useful accompaniment when crystallography does not work. Distance determination of interacting moieties can allow observation of protein conformation and oligomeric state. Spin label addition may be employed if no naturally occurring EPR observable species are available. However, EPR can also be a useful tool to complement a crystallographic structural determination. An example of this is the elucidation of a cofactor's oxidation state, information about which can aid mechanistic and biological insight.

Bibliography

- Abrahams, J.P., and Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Crystallographica Section D-Biological Crystallography* 52, 30-42.
- Abrusci, P., Johnson, S., Roversi, P., and Lea, S.M. ((to be published)). Structural studies of membrane components of the *Shigella flexneri* type three secretion system.
- Akeda, Y., and Galan, J.E. (2005). Chaperone release and unfolding of substrates in type III secretion. *Nature* 437, 911-915.
- Alder, N.N., and Theg, S.M. (2003). Energetics of protein transport across biological membranes. a study of the thylakoid DeltapH-dependent/cpTat pathway. *Cell* 112, 231-242.
- Aldridge, C., Spence, E., Kirkilionis, M.A., Frigerio, L., and Robinson, C. (2008). Tat-dependent targeting of Rieske iron-sulphur proteins to both the plasma and thylakoid membranes in the cyanobacterium *Synechocystis* PCC6803. *Mol Microbiol* 70, 140-150.
- Allaoui, A., Sansonetti, P.J., and Parsot, C. (1992). MxiJ, a Lipoprotein Involved in Secretion of *Shigella* Ipa Invasins, Is Homologous to YscJ, a Secretion Factor of the *Yersinia* Yop Proteins. *Journal of Bacteriology* 174, 7661-7669.
- Allaoui, A., Sansonetti, P.J., and Parsot, C. (1993). MxiD, an Outer-Membrane Protein Necessary for the Secretion of the *Shigella-Flexneri* Ipa Invasins. *Mol Microbiol* 7, 59-68.
- Alpuche-Aranda, C.M., Racoosin, E.L., Swanson, J.A., and Miller, S.I. (1994). *Salmonella* stimulate macrophage macropinocytosis and persist within spacious phagosomes. *J Exp Med* 179, 601-608.
- Alto, N.M., Shao, F., Lazar, C.S., Brost, R.L., Chua, G., Mattoo, S., McMahon, S.A., Ghosh, P., Hughes, T.R., Boone, C., and Dixon, J.E. (2006). Identification of a bacterial type III effector family with G protein mimicry functions. *Cell* 124, 133-145.
- Altschul, S.F., and Gish, W. (1996). Local alignment statistics. *Methods Enzymol* 266, 460-480.
- Altschul, S.F., and Koonin, E.V. (1998). Iterated profile searches with PSI-BLAST--a tool for discovery in protein databases. *Trends Biochem Sci* 23, 444-447.
- Anderson, D.M., and Schneewind, O. (1997). A mRNA signal for the type III secretion of Yop proteins by *Yersinia enterocolitica*. *Science* 278, 1140-1143.
- Antelmann, H., Scharf, C., and Hecker, M. (2000). Phosphate starvation-inducible proteins of *Bacillus subtilis*: Proteomics and transcriptional analysis. *Journal of Bacteriology* 182, 4478-4490.
- Aquino, A., Lim, J.-S., and Sykes, A., G. (1994). Mechanism of the Reaction of Different Phosphates with the Iron(ii)iron(iii) form of Purple Acid Phosphatase from Porcine Uteri (Uteroferrin). *Journal of the Chemical Society Dalton Trans.*, 429.

Arbeloa, A., Blanco, M., Moreira, F.C., Bulgin, R., Lopez, C., Dahbi, G., Blanco, J.E., Mora, A., Alonso, M.P., Mamani, R.C., *et al.* (2009). Distribution of *espM* and *espT* among enteropathogenic and enterohaemorrhagic *Escherichia coli*. *Journal of Medical Microbiology* 58, 988-995.

Arbeloa, A., Bulgin, R.R., MacKenzie, G., Shaw, R.K., Pallen, M.J., Crepin, V.F., Berger, C.N., and Frankel, G. (2008). Subversion of actin dynamics by *EspM* effectors of attaching and effacing bacterial pathogens. *Cellular Microbiology* 10, 1429-1441.

Arbeloa, A., Garnett, J., Lillington, J., Bulgin, R.R., Berger, C.N., Lea, S.M., Matthews, S., and Frankel, G. (2010). *EspM2* is a RhoA guanine nucleotide exchange factor. *Cell Microbiol* 12, 654-664.

Armstrong, J.J., Baddiley, J., Buchanan, J.G., Davison, A.L., Kelemen, M.V., and Neuhaus, F.C. (1959). Composition of Teichoic Acids from a Number of Bacterial Walls. *Nature* 184, 247-248.

Ashida, H., Kim, M., Schmidt-Supprian, M., Ma, A., Ogawa, M., and Sasakawa, C. (2010). A bacterial E3 ubiquitin ligase *IpaH9.8* targets NEMO/IKKgamma to dampen the host NF-kappaB-mediated inflammatory response. *Nat Cell Biol* 12, 66-73; sup pp 61-69.

Ashida, H., Ogawa, M., Mimuro, H., and Sasakawa, C. (2009). *Shigella* infection of intestinal epithelium and circumvention of the host innate defense system. *Curr Top Microbiol Immunol* 337, 231-255.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38, W529-533.

Ashton, D.S., Beddell, C.R., Green, B.N., and Oliver, R.W.A. (1994). Rapid Validation of Molecular-Structures of Biological Samples by Electrospray Mass-Spectrometry. *Febs Letters* 342, 1-6.

Atkins, P.W., and De Paula, J. (2006). *Atkins' Physical chemistry*, 8th edn (Oxford ; New York: Oxford University Press).

Atkins, P.W., and Friedman, R. (2005). *Molecular quantum mechanics*, 4th edn (New York: Oxford University Press).

Bachmann, J., Bauer, B., Zwicker, K., Ludwig, B., and Anderka, O. (2006). The Rieske protein from *Paracoccus denitrificans* is inserted into the cytoplasmic membrane by the twin-arginine translocase. *Febs Journal* 273, 4817-4830.

Barford, D., Das, A.K., and Egloff, M.P. (1998). The structure and mechanism of protein phosphatases: Insights into catalysis and regulation. *Annual Review of Biophysics and Biomolecular Structure* 27, 133-164.

Barnett, J.P., Eijlander, R.T., Kuipers, O.P., and Robinson, C. (2008). A minimal *tat* system from a gram-positive organism - A bifunctional *TatA* subunit participates in discrete *TatAC* and *TatA* complexes. *Journal of Biological Chemistry* 283, 2534-2542.

Barnett, J.P., Lawrence, J., Mendel, S., and Robinson, C. (2011). Expression of the bifunctional *Bacillus subtilis* *TatAd* protein in *Escherichia coli* reveals distinct *TatA/B*-family and *TatB*-specific domains. *Arch Microbiol* 193, 583-594.

Barta, M.L., Zhang, L., Picking, W.L., and Geisbrecht, B.V. (2010). Evidence for alternative quaternary structure in a bacterial Type III secretion system chaperone. *BMC Struct Biol* 10, 21.

Bennati, M., and Prisner, T.F. (2005). New developments in high field electron paramagnetic resonance with applications in structural biology. *Reports on Progress in Physics* 68, 411-448.

Berger, C.N., Crepin, V.F., Jepson, M.A., Arbeloa, A., and Frankel, G. (2009). The mechanisms used by enteropathogenic *Escherichia coli* to control filopodia dynamics. *Cell Microbiol* 11, 309-322.

Berks, B.C., Palmer, T., and Sargent, F. (2003). The Tat protein translocation pathway and its role in microbial physiology. *Advances in Microbial Physiology* 47, 187-254.

Berks, B.C., Palmer, T., and Sargent, F. (2005). Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Current Opinion in Microbiology* 8, 174-181.

Berks, B.C., Sargent, F., and Palmer, T. (2000). The Tat protein export pathway. *Mol Microbiol* 35, 260-274.

Berliner, L.J. (1984). *Biological Magnetic Resonance - Spin Labeling: The Next Millennium*.

Berliner, L.J. (2003). *In vivo EPR (ESR) : theory and application* (New York: Kluwer Academic/Plenum Publishers).

Berliner, L.J., Grunwald, J., Hankovszky, H.O., and Hideg, K. (1982). A Novel Reversible Thiol-Specific Spin Label - Papain Active-Site Labeling and Inhibition. *Analytical Biochemistry* 119, 450-455.

Bernardini, M.L., Mounier, J., d'Hauteville, H., Coquis-Rondon, M., and Sansonetti, P.J. (1989). Identification of icsA, a plasmid locus of *Shigella flexneri* that governs bacterial intra- and intercellular spread through interaction with F-actin. *Proc Natl Acad Sci U S A* 86, 3867-3871.

Bernstein, H.D. (2007). Are bacterial 'autotransporters' really transporters? *Trends Microbiol* 15, 441-447.

Bhadra, R., Srinivasan, N., and Pandit, S.B. (2005). A new domain family in the superfamily of alkaline phosphatases. *In Silico Biol* 5, 379-387.

Birket, S.E., Harrington, A.T., Espina, M., Smith, N.D., Terry, C.M., Darboe, N., Markham, A.P., Middaugh, C.R., Picking, W.L., and Picking, W.D. (2007). Preparation and characterization of translocator/chaperone complexes and their component proteins from *Shigella flexneri*. *Biochemistry* 46, 8128-8137.

Birtalan, S., and Ghosh, P. (2001). Structure of the *Yersinia* type III secretory system chaperone SycE. *Nature Structural Biology* 8, 974-978.

Birtalan, S.C., Phillips, R.M., and Ghosh, P. (2002). Three-dimensional secretion signals in chaperone-effector complexes of bacterial pathogens. *Mol Cell* 9, 971-980.

Blanc, E., Roversi, P., Vornrhein, C., Flensburg, C., Lea, S.M., and Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallographica Section D-Biological Crystallography* 60, 2210-2221.

Blaudeck, N., Kreutzenbeck, P., Freudl, R., and Sprenger, G.A. (2003). Genetic analysis of pathway specificity during posttranslational protein translocation across the *Escherichia coli* plasma membrane. *Journal of Bacteriology* 185, 2811-2819.

Block, A., Li, G., Fu, Z.Q., and Alfano, J.R. (2008). Phytopathogen type III effector weaponry and their plant targets. *Curr Opin Plant Biol* 11, 396-403.

Blocker, A., Gounon, P., Larquet, E., Niebuhr, K., Cabiaux, V., Parsot, C., and Sansonetti, P. (1999). The tripartite type III secretin of *Shigella flexneri* inserts IpaB and IpaC into host membranes. *Journal of Cell Biology* 147, 683-693.

Blocker, A., Jouihri, N., Larquet, E., Gounon, P., Ebel, F., Parsot, C., Sansonetti, P., and Allaoui, A. (2001). Structure and composition of the *Shigella flexneri* 'needle complex', a part of its type III secretin. *Mol Microbiol* 39, 652-663.

Blow, D.M. (2002). *Outline of crystallography for biologists* (Oxford ; New York: Oxford University Press).

Blumberg, W.E. (1967). The EPR of high spin Fe³⁺ in rhombic fields. *Magnetic Resonance in Biological Systems* 9, 119-134.

Bolhuis, A., Mathers, J.E., Thomas, J.D., Barrett, C.M.L., and Robinson, C. (2001). TatB and TatC form a functional and structural unit of the twin-arginine translocase from *Escherichia coli*. *Journal of Biological Chemistry* 276, 20213-20219.

Bond, C.S., Clements, P.R., Ashby, S.J., Collyer, C.A., Harrop, S.J., Hopwood, J.J., and Guss, J.M. (1997). Structure of a human lysosomal sulfatase. *Structure* 5, 277-289.

Boquet, P., and Lemichez, E. (2003). Bacterial virulence factors targeting Rho GTPases: parasitism or symbiosis? *Trends Cell Biol* 13, 238-246.

Botteaux, A., Kayath, C.A., Page, A.L., Jouihri, N., Sani, M., Boekema, E., Biskri, L., Parsot, C., and Allaoui, A. (2010). The 33 carboxyl-terminal residues of Spa40 orchestrate the multi-step assembly process of the type III secretion needle complex in *Shigella flexneri*. *Microbiol-Sgm* 156, 2807-2817.

Botteaux, A., Sory, M.P., Biskri, L., Parsot, C., and Allaoui, A. (2009). MxiC is secreted by and controls the substrate specificity of the *Shigella flexneri* type III secretion apparatus. *Mol Microbiol* 71, 449-460.

Bou-Abdallah, F., and Chasteen, N.D. (2008). Spin concentration measurements of high-spin ($g' = 4.3$) rhombic iron(III) ions in biological samples: theory and application. *J Biol Inorg Chem* 13, 15-24.

Bowman, M.K., Maryasov, A.G., Kim, N., and DeRose, V.J. (2004). Visualization of distance distribution from pulsed double electron-electron resonance data. *Applied Magnetic Resonance* 26, 23-39.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallographica Section D-Biological Crystallography* 59, 2023-2030.

Bruker (2011). <http://www.bruker-biospin.com/> (last visited July 2011).

Bruker_Biospin <http://www.bruker-biospin.com/pulsepractice.html>.

Brumell, J.H., Goosney, D.L., and Finlay, B.B. (2002). SifA, a type III secreted effector of *Salmonella typhimurium*, directs *Salmonella*-induced filament (Sif) formation along microtubules. *Traffic* 3, 407-415.

Brunger, A.T. (1992). Free R-Value - a Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures. *Nature* 355, 472-475.

Brutinel, E.D., and Yahr, T.L. (2008). Control of gene expression by type III secretory activity. *Current Opinion in Microbiology* 11, 128-133.

Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., d'Hauteville, H., Kunst, F., Sansonetti, P., and Parsot, C. (2000). The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol Microbiol* 38, 760-771.

Buchwald, G., Friebe, A., Galan, J.E., Hardt, W.D., Wittinghofer, A., and Scheffzek, K. (2002). Structural basis for the reversible activation of a Rho protein by the bacterial toxin SopE. *Embo Journal* 21, 3286-3295.

Bulgin, R.R., Arbeloa, A., Chung, J.C.S., and Frankel, G. (2009). EspT triggers formation of lamellipodia and membrane ruffles through activation of Rac-1 and Cdc42. *Cellular Microbiology* 11, 217-229.

Burghout, P., van Boxtel, R., Van Gelder, P., Ringler, P., Muller, S.A., Tommassen, J., and Koster, M. (2004). Structure and electrophysiological properties of the YscC secretin from the type III secretion system of *Yersinia enterocolitica*. *Journal of Bacteriology* 186, 4645-4654.

Buttner, C.R., Sorg, I., Cornelis, G.R., Heinz, D.W., and Niemann, H.H. (2008). Structure of the *Yersinia enterocolitica* type III secretion translocator chaperone SycD. *Journal of Molecular Biology* 375, 997-1012.

Campbell, J., Singh, A.K., Santa Maria, J.P., Jr., Kim, Y., Brown, S., Swoboda, J.G., Mylonakis, E., Wilkinson, B.J., and Walker, S. (2011). Synthetic lethal compound combinations reveal a fundamental connection between wall teichoic acid and peptidoglycan biosyntheses in *Staphylococcus aureus*. *ACS Chem Biol* 6, 106-116.

Cantarelli, V.V., Takahashi, A., Yanagihara, I., Akeda, Y., Imura, K., Kodama, T., Kono, G., Sato, Y., and Honda, T. (2001). Talin, a host cell protein, interacts directly with the translocated intimin receptor, Tir, of enteropathogenic *Escherichia coli*, and is essential for pedestal formation. *Cell Microbiol* 3, 745-751.

Chan, C.S., Zlomislic, M.R., Tieleman, D.P., and Turner, R.J. (2007). The TatA subunit of *Escherichia coli* twin-arginine translocase has an N-in topology. *Biochemistry* 46, 7396-7404.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D-Biological Crystallography* 66, 12-21.

Cheng, L.W., and Schneewind, O. (1999). *Yersinia enterocolitica* type III secretion. On the role of SycE in targeting YopE into HeLa cells. *J Biol Chem* 274, 22102-22108.

Cieslik, M., and Derewenda, Z.S. (2009). The role of entropy and polarity in intermolecular contacts in protein crystals. *Acta Crystallographica Section D-Biological Crystallography* 65, 500-509.

Cline, K., and McCaffery, M. (2007). Evidence for a dynamic and transient pathway through the TAT protein transport machinery. *Embo Journal* 26, 3039-3049.

Cline, K., and Mori, H. (2001). Thylakoid Delta pH-dependent precursor proteins bind to a cpTatC-Hcf106 complex before Tha4-dependent transport. *Journal of Cell Biology* 154, 719-729.

Coelho, A.V., Matias, P., Fulop, V., Thompson, A., Gonzalez, A., and Carrondo, M.A. (1997). Desulfoferrodoxin structure determined by MAD phasing and refinement to 1.9-angstrom resolution reveals a unique combination of a tetrahedral FeS₄ centre with a square pyramidal FeSN₄ centre. *Journal of Biological Inorganic Chemistry* 2, 680-689.

Coleman, M.L., McDonough, M.A., Hewitson, K.S., Coles, C., Mecinovic, J., Edelmann, M., Cook, K.M., Cockman, M.E., Lancaster, D.E., Kessler, B.M., *et al.* (2007). Asparaginyl hydroxylation of the notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *Journal of Biological Chemistry* 282, 24027-24038.

Collaborative Computational Project Number 4 (1994). The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography* 50, 760-763.

Collins, R.E., Northrop, J.P., Horton, J.R., Lee, D.Y., Zhang, X., Cheng, X.D., and Stallcup, M.R. (2008). The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nature Structural & Molecular Biology* 15, 245-250.

Columbus, L., and Hubbell, W.L. (2002). A new spin on protein dynamics. *Trends in Biochemical Sciences* 27, 288-295.

Columbus, L., Kalai, T., Jeko, J., Hideg, K., and Hubbell, W.L. (2001). Molecular motion of spin labeled side chains in alpha-helices: Analysis by variation of side chain structure. *Biochemistry* 40, 3828-3846.

Cordes, F.S., Daniell, S., Kenjale, R., Saurya, S., Picking, W.L., Picking, W.D., Booy, F., Lea, S.M., and Blocker, A. (2005). Helical packing of needles from functionally altered shigella type III secretion systems. *Journal of Molecular Biology* 354, 206-211.

Cordes, F.S., Komoriya, K., Larquet, E., Yang, S.X., Egelman, E.H., Blocker, A., and Lea, S.M. (2003). Helical structure of the needle of the type III secretion system of *Shigella flexneri*. *Journal of Biological Chemistry* 278, 17103-17107.

Cornelis, G.R., Boland, A., Boyd, A.P., Geuijen, C., Iriarte, M., Neyt, C., Sory, M.P., and Stainier, I. (1998). The virulence plasmid of *Yersinia*, an antihost genome. *Microbiology and Molecular Biology Reviews* 62, 1315-1352.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D-Biological Crystallography* 62, 1002-1011.

Cowtan, K. (2010). Recent developments in classical density modification. *Acta Crystallographica Section D-Biological Crystallography* 66, 470-478.

Cox, R.S., Schenk, G., Mitic, N., Gahan, L.R., and Hengge, A.C. (2007). Diesterase activity and substrate binding in purple acid phosphatases. *J Am Chem Soc* 129, 9550-9551.

Croxen, M.A., and Finlay, B.B. (2009). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol* 8, 26-38.

Davies, D.R., and Hol, W.G. (2004). The power of vanadate in crystallographic investigations of phosphoryl transfer enzymes. *FEBS Lett* 577, 315-321.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall III, W.B., Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35, W375-W383.

De Keersmaecker, S., Van Mellaert, L., Schaerlaekens, K., Van Dessel, W., Vrancken, K., Lammertyn, E., Anne, J., and Geukens, N. (2005). Structural organization of the twin-arginine translocation system in *Streptomyces lividans*. *FEBS Lett* 579, 797-802.

De La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Macromolecular Crystallography, Pt A* 276, 472-494.

De Leeuw, E., Granjon, T., Porcelli, I., Alami, M., Carr, S.B., Muller, M., Sargent, F., Palmer, T., and Berks, B.C. (2002). Oligomeric properties and signal peptide binding by *Escherichia coli* Tat protein transport complexes. *J Mol Biol* 322, 1135-1146.

De Leeuw, E., Porcelli, I., Sargent, F., Palmer, T., and Berks, B.C. (2001). Membrane interactions and self-association of the TatA and TatB components of the twin-arginine translocation pathway. *Febs Letters* 506, 143-148.

Deane, J.E., Abrusci, P., Johnson, S., and Lea, S.M. (2010). Timing is everything: the regulation of type III secretion. *Cell Mol Life Sci* 67, 1065-1075.

Deane, J.E., Graham, S.C., Mitchell, E.P., Flot, D., Johnson, S., and Lea, S.M. (2008). Crystal structure of Spa40, the specificity switch for the *Shigella flexneri* type III secretion system. *Mol Microbiol* 69, 267-276.

Deane, J.E., Roversi, P., Cordes, F.S., Johnson, S., Kenjale, R., Daniell, S., Booy, F., Picking, W.D., Picking, W.L., Blocker, A.J., and Lea, S.M. (2006). Molecular model of a type III secretion system needle: Implications for host-cell sensing. *P Natl Acad Sci USA* 103, 12529-12533.

Delepelaire, P. (2004). Type I secretion in gram-negative bacteria. *Biochimica Et Biophysica Acta-Molecular Cell Research* 1694, 149-161.

DeLisa, M.P., Tullman, D., and Georgiou, G. (2003). Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *P Natl Acad Sci USA* 100, 6115-6120.

Derewenda, Z.S., and Vekilov, P.G. (2006). Entropy and surface engineering in protein crystallization. *Acta Crystallographica Section D-Biological Crystallography* 62, 116-124.

Diamond, R. (1969). Profile Analysis in Single Crystal Diffractometry. *Acta Crystallographica Section a-Crystal Physics Diffraction Theoretical and General Crystallography A* 25, 43-45.

Dietrich, M., Munstermann, D., Suerbaum, H., and Witzel, H. (1991). Purple acid phosphatase from bovine spleen. Interactions at the active site in relation to the reaction mechanism. *Eur J Biochem* 199, 105-113.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., and Mattick, J.S. (1991). Touchdown Pcr to Circumvent Spurious Priming during Gene Amplification. *Nucleic Acids Res* 19, 4008.

Dong, A., Xu, X., Edwards, A.M., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E., *et al.* (2007). In situ proteolysis for protein crystallization and structure determination. *Nat Methods* 4, 1019-1021.

Driessen, A.J.M., and Nouwen, N. (2008). Protein translocation across the bacterial cytoplasmic membrane. *Annual Review of Biochemistry* 77, 643-667.

Easyspin <http://www.easyspin.org/> (last visited July 2011).

Eder, S., Shi, L., Jensen, K., Yamane, K., and Hulett, F.M. (1996). A *Bacillus subtilis* secreted phosphodiesterase alkaline phosphatase is the product of a Pho regulon gene, *phoD*. *Microbiology-Uk* 142, 2041-2047.

Eijlander, R.T., Jongbloed, J.D.H., and Kuipers, O.P. (2009a). Relaxed Specificity of the *Bacillus subtilis* TatAdCd Translocase in Tat-Dependent Protein Secretion. *Journal of Bacteriology* 191, 196-202.

Eijlander, R.T., Kolbusz, M.A., Berendsen, E.M., and Kuipers, O.P. (2009b). Effects of altered TatC proteins on protein secretion efficiency via the twin-arginine translocation pathway of *Bacillus subtilis*. *Microbiol-Sgm* 155, 1776-1785.

El Kirat, K., Besson, F., Prigent, A.F., Chauvet, J.P., and Roux, B. (2002). Role of calcium and membrane organization on phospholipase D localization and activity. Competition between a soluble and insoluble substrate. *J Biol Chem* 277, 21231-21236.

Elliott, E.J. (2007). Acute gastroenteritis in children. *British Medical Journal* 334, 35-40.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography* 60, 2126-2132.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* 66, 486-501.

Epler, C.R., Dickenson, N.E., Olive, A.J., Picking, W.L., and Picking, W.D. (2009). Liposomes Recruit IpaC to the *Shigella flexneri* Type III Secretion Apparatus Needle as a Final Step in Secretion Induction. *Infect Immun* 77, 2754-2761.

Ernst, R.R., Bodenhausen, G., and Wokaun, A. (1987). Principles of nuclear magnetic resonance in one and two dimensions (Oxford [Oxfordshire]; New York)).

Espina, M., Olive, A.J., Kenjale, R., Moore, D.S., Ausar, S.F., Kaminski, R.W., Oaks, E.V., Middaugh, C.R., Picking, W.D., and Picking, W.L. (2006). IpaD localizes to the tip of the type III secretion system needle of *Shigella flexneri*. *Infect Immun* 74, 4391-4400.

Evans, G., and Pettifer, R.F. (2001). CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *Journal of Applied Crystallography* 34, 82-86.

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallographica Section D-Biological Crystallography* 62, 72-82.

Evans, P., and McCoy, A. (2008). An introduction to molecular replacement. *Acta Crystallographica Section D-Biological Crystallography* 64, 1-10.

Evans, P.R. (1997). SCALA! *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* 33, 22-24.

Filloux, A. (2004). The underlying mechanisms of type II protein secretion. *Biochim Biophys Acta* 1694, 163-179.

Filloux, A. (2009). The type VI secretion system: a tubular story. *EMBO J* 28, 309-310.

Filloux, A., Hachani, A., and Bleves, S. (2008). The bacterial type VI secretion machine: yet another player for protein transport across membranes. *Microbiology* 154, 1570-1583.

Freeman, N.L., Zurawski, D.V., Chowrashi, P., Ayoob, J.C., Huang, L., Mittal, B., Sanger, J.M., and Sanger, J.W. (2000). Interaction of the enteropathogenic *Escherichia coli* protein, translocated intimin receptor (Tir), with focal adhesion proteins. *Cell Motil Cytoskeleton* 47, 307-318.

French, S., and Wilson, K. (1978). Treatment of Negative Intensity Observations. *Acta Crystallographica Section A* 34, 517-525.

Friebe, A., Ilchmann, H., Aepfelbacher, M., Ehrbar, K., Machleidt, W., and Hardt, W.D. (2001). SopE and SopE2 from *Salmonella typhimurium* activate different sets of RhoGTPases of the host cell. *J Biol Chem* 276, 34035-34040.

Fukushima, M., Kakinuma, K., and Kawaguchi, R. (2002). Phylogenetic analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. *Journal of Clinical Microbiology* 40, 2779-2785.

Galperin, M.Y., Bairoch, A., and Koonin, E.V. (1998). A superfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with alkaline phosphatases and sulfatases. *Protein Science* 7, 1829-1835.

Garmendia, J., Frankel, G., and Crepin, V.F. (2005). Enteropathogenic and enterohemorrhagic *Escherichia coli* infections: Translocation, translocation, translocation. *Infect Immun* 73, 2573-2585.

Genest, O., Neumann, M., Seduk, F., Stocklein, W., Mejean, V., Leimkuhler, S., and Iobbi-Nivol, C. (2008). Dedicated metallochaperone connects apoenzyme and molybdenum cofactor biosynthesis components. *J Biol Chem* 283, 21433-21440.

- Gennity, J.M., and Inouye, M. (1991). Protein secretion in bacteria. *Curr Opin Biotechnol* 2, 661-667.
- Germane, K.L., Ohi, R., Goldberg, M.B., and Spiller, B.W. (2008). Structural and functional studies indicate that Shigella VirA is not a protease and does not directly destabilize microtubules. *Biochemistry* 47, 10241-10243.
- Ghosh, P. (2004). Process of protein transport by the type III secretion system. *Microbiology and Molecular Biology Reviews* 68, 771-795.
- Gohlke, U., Pullan, L., McDevitt, C.A., Porcelli, I., de Leeuw, E., Palmer, T., Saibil, H.R., and Berks, B.C. (2005). The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *P Natl Acad Sci USA* 102, 10482-10486.
- Goldschmidt, L., Cooper, D.R., Derewenda, Z.S., and Eisenberg, D. (2007). Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Science* 16, 1569-1576.
- Goosney, D.L., Knoechel, D.G., and Finlay, B.B. (1999). Enteropathogenic E. coli, Salmonella, and Shigella: masters of host cell cytoskeletal exploitation. *Emerg Infect Dis* 5, 216-223.
- Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31, 3320-3323.
- Grant, W.D. (1979). Cell-Wall Teichoic-Acid as a Reserve Phosphate Source in Bacillus-Subtilis. *Journal of Bacteriology* 137, 35-43.
- Graubner, W., Schierhorn, A., and Bruser, T. (2007). DnaK plays a pivotal role in Tat targeting of CueO and functions beside SlyD as a general Tat signal binding chaperone. *J Biol Chem* 282, 7116-7124.
- Guo, Z.F., Cascio, D., Hideg, K., Kalai, T., and Hubbell, W.L. (2007). Structural determinants of nitroxide motion in spin-labeled proteins: Tertiary contact and solvent-inaccessible sites in helix G of T4 lysozyme. *Protein Science* 16, 1069-1086.
- Hachani, A., Biskri, L., Rossi, G., Marty, A., Menard, R., Sansonetti, P., Parsot, C., Van Nhieu, G.T., Bernardini, M.L., and Allaoui, A. (2008). IpgB1 and IpgB2, two homologous effectors secreted via the Mxi-Spa type III secretion apparatus, cooperate to mediate polarized cell invasion and inflammatory potential of Shigella flexneri. *Microbes and Infection* 10, 260-268.
- Hagelueken, G., Ingledew, W.J., Huang, H., Petrovic-Stojanovska, B., Whitfield, C., ElMkami, H., Schiemann, O., and Naismith, J.H. (2009). PELDOR Spectroscopy Distance Fingerprinting of the Octameric Outer-Membrane Protein Wza from Escherichia coli. *Angewandte Chemie-International Edition* 48, 2904-2906.
- Hahn, T. (1978). New Edition of International Tables for Crystallography. *Acta Crystallographica Section A* 34, S3-S3.
- Hall, A. (1998). Rho GTPases and the actin cytoskeleton. *Science* 279, 509-514.

- Hamiaux, C., van Eerde, A., Parsot, C., Broos, J., and Dijkstra, B.W. (2006). Structural mimicry for vinculin activation by IpaA, a virulence factor of *Shigella flexneri*. *Embo Rep* 7, 794-799.
- Hancock, I.C. (1997). Bacterial cell surface carbohydrates: structure and assembly. *Biochem Soc Trans* 25, 183-187.
- Handa, Y., Suzuki, M., Ohya, K., Iwai, H., Ishijima, N., Koleske, A.J., Fukui, Y., and Sasakawa, C. (2007). *Shigella* IpgB1 promotes bacterial entry through the ELMO-Dock180 machinery. *Nat Cell Biol* 9, 121-U165.
- Hansen-Wester, I., and Hensel, M. (2001). *Salmonella* pathogenicity islands encoding type III secretion systems. *Microbes Infect* 3, 549-559.
- Haraga, A., Ohlson, M.B., and Miller, S.I. (2008). *Salmonellae* interplay with host cells. *Nat Rev Microbiol* 6, 53-66.
- Hart, M., and Powers, S. (1995). Ras-Cdc25 and Rho-Dbl Binding Assays - Complex-Formation in-Vitro. *Small Gtpases and Their Regulators*, Pt A 255, 129-135.
- Harwood, C.R. (1992). *Bacillus subtilis* and its relatives: molecular biological and industrial workhorses. *Trends Biotechnol* 10, 247-256.
- Hayward, R.D., Cain, R.J., McGhie, E.J., Phillips, N., Garner, M.J., and Koronakis, V. (2005). Cholesterol binding by the bacterial type III translocon is essential for virulence effector delivery into mammalian cells. *Mol Microbiol* 56, 590-603.
- Hecht, G. (2001). Microbes and microbial toxins: paradigms for microbial-mucosal interactions. VII. Enteropathogenic *Escherichia coli*: physiological alterations from an extracellular position. *Am J Physiol Gastrointest Liver Physiol* 281, G1-7.
- Hecker, M., Pane-Farre, J., and Volker, U. (2007). SigB-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria. *Annual Review of Microbiology* 61, 215-236.
- Hendrickson, W.A. (1991). Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation. *Science* 254, 51-58.
- Hendrickson, W.A., and Ogata, C.M. (1997). Phase determination from multiwavelength anomalous diffraction measurements. *Macromolecular Crystallography*, Pt A 276, 494-523.
- Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R., and Stuber, D. (1988). Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Bio-Technology* 6, 1321-1325.
- Hodgkinson, J.L., Horsley, A., Stabat, D., Simon, M., Johnson, S., da Fonseca, P.C.A., Morris, E.P., Wall, J.S., Lea, S.M., and Blocker, A.J. (2009). Three-dimensional reconstruction of the *Shigella* T3SS transmembrane regions reveals 12-fold symmetry and novel features throughout. *Nature Structural & Molecular Biology* 16, 477-485.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38, W545-W549.

Holtz, K.M., and Kantrowitz, E.R. (1999). The mechanism of the alkaline phosphatase reaction: insights from NMR, crystallography and site-specific mutagenesis. *FEBS Lett* 462, 7-11.

Holtz, K.M., Stec, B., Myers, J.K., Antonelli, S.M., Widlanski, T.S., and Kantrowitz, E.R. (2000). Alternate modes of binding in two crystal structures of alkaline phosphatase-inhibitor complexes. *Protein Science* 9, 907-915.

Holzapfel, E., Eisner, G., Alami, M., Barrett, C.M., Buchanan, G., Luke, I., Betton, J.M., Robinson, C., Palmer, T., Moser, M., and Muller, M. (2007). The entire N-terminal half of TatC is involved in twin-arginine precursor binding. *Biochemistry* 46, 2892-2898.

Hu, Y.F., Zhao, E.W., Li, H.W., Xia, B., and Jin, C.W. (2010). Solution NMR Structure of the TatA Component of the Twin-Arginine Protein Transport System from Gram-Positive Bacterium *Bacillus subtilis*. *J Am Chem Soc* 132, 15942-15944.

Huang, Z., and Chai, J. (2010). Mapping the selection mechanisms by bacterial GEFs. *Virulence* 1, 93-96.

Huang, Z.W., Sutton, S.E., Wallenfang, A.J., Orchard, R.C., Wu, X.J., Feng, Y.C., Chai, J.J., and Alto, N.M. (2009a). Structural insights into host GTPase isoform selection by a family of bacterial GEF mimics. *Nature Structural & Molecular Biology* 16, 853-860.

Huang, Z.W., Sutton, S.E., Wallenfang, A.J., Orchard, R.C., Wu, X.J., Feng, Y.C., Chai, J.J., and Alto, N.M. (2009b). Structural insights into host GTPase isoform selection by a family of bacterial GEF mimics. *Nature Structural & Molecular Biology* 16, 853-U879.

Hulett, F.M. (1996). The signal-transduction network for Pho regulation in *Bacillus subtilis*. *Mol Microbiol* 19, 933-939.

Hulett, F.M., Bookstein, C., and Jensen, K. (1990). Evidence for 2 Structural Genes for Alkaline-Phosphatase in *Bacillus-Subtilis*. *Journal of Bacteriology* 172, 735-740.

Hulett, F.M., Kim, E.E., Bookstein, C., Kapp, N.V., Edwards, C.W., and Wyckoff, H.W. (1991). *Bacillus-Subtilis* Alkaline Phosphatase-iii and Phosphatase-iv - Cloning, Sequencing, and Comparisons of Deduced Amino-Acid-Sequence with *Escherichia-Coli* Alkaline-Phosphatase 3-Dimensional Structure. *Journal of Biological Chemistry* 266, 1077-1084.

Hull, W.E., Halford, S.E., Gutfreund, H., and Sykes, B.D. (1976). ³¹P nuclear magnetic resonance study of alkaline phosphatase: the role of inorganic phosphate in limiting the enzyme turnover rate at alkaline pH. *Biochemistry* 15, 1547-1561.

Iguchi, A., Thomson, N.R., Ogura, Y., Saunders, D., Ooka, T., Henderson, I.R., Harris, D., Asadulghani, M., Kurokawa, K., Dean, P., *et al.* (2009). Complete Genome Sequence and Comparative Genome Analysis of Enteropathogenic *Escherichia coli* O127:H6 Strain E2348/69. *Journal of Bacteriology* 191, 347-354.

Ito, N., Phillips, S.E., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D., and Knowles, P.F. (1991). Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature* **350**, 87-90.

Izard, T., Tran Van Nhieu, G., and Bois, P.R. (2006). Shigella applies molecular mimicry to subvert vinculin and invade host cells. *J Cell Biol* **175**, 465-475.

Jacob-Dubuisson, F., Loch, C., and Antoine, R. (2001). Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. *Mol Microbiol* **40**, 306-313.

Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., and Godzik, A. (2005). FFAS03: a server for profile-profile sequence alignments. *Nucleic Acids Res* **33**, W284-288.

Jeschke, G. (2007). Instrumentation and experimental setup. *ESR Spectroscopy in Membrane Biophysics* **27**, 17-47.

Jeschke, G., Bender, A., Paulsen, H., Zimmermann, H., and Godt, A. (2004). Sensitivity enhancement in pulse EPR distance measurements. *J Magn Reson* **169**, 1-12.

Jeschke, G., Chechik, V., Ionita, P., Godt, A., Zimmermann, H., Herbert, Banham, J., Timmel, C.R., Hilger, D., and Jung, H. (2006). Deer Analysis 2006 – a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson* **30**, 473-498.

Jeschke, G., and Polyhach, Y. (2007). Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Physical Chemistry Chemical Physics* **9**, 1895-1910.

Jeschke, G., and Polyhach, Y. (2009). MMM <http://www.epr.ethz.ch/software/> (last visited July 2011).

Jeschke, G., Sajid, M., Schulte, M., Ramezani, N., Volkov, A., Zimmermann, H., and Godt, A. (2010). Flexibility of Shape-Persistent Molecular Building Blocks Composed of p-Phenylene and Ethynylene Units. *J Am Chem Soc* **132**, 10107-10117.

Jiang, J., Lafer, E.M., and Sousa, R. (2006). Crystallization of a functionally intact Hsc70 chaperone. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62**, 39-43.

Johnson, D.C., Dean, D.R., Smith, A.D., and Johnson, M.K. (2005a). Structure, function, and formation of biological iron-sulfur clusters. *Annu Rev Biochem* **74**, 247-281.

Johnson, S., Deane, J.E., and Lea, S.M. (2005b). The type III needle and the damage done. *Curr Opin Struc Biol* **15**, 700-707.

Johnson, S., Roversi, P., Espina, M., Olive, A., Deane, J.E., Birket, S., Field, T., Picking, W.D., Blocker, A.J., Galyov, E.E., *et al.* (2007). Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD. *Journal of Biological Chemistry* **282**, 4035-4044.

Jonas, S., van Loo, B., Hyvonen, M., and Hollfelder, F. (2008). A new member of the alkaline phosphatase superfamily with a formylglycine nucleophile: structural and kinetic characterisation of a phosphonate monoester hydrolase/phosphodiesterase from *Rhizobium leguminosarum*. *J Mol Biol* **384**, 120-136.

- Jongbloed, J.D.H., Antelmann, H., Hecker, M., Nijland, R., Bron, S., Airaksinen, U., Pries, F., Quax, W.J., van Dijl, J.M., and Braun, P.G. (2002). Selective contribution of the twin-arginine translocation pathway to protein secretion in *Bacillus subtilis*. *Journal of Biological Chemistry* 277, 44068-44078.
- Jongbloed, J.D.H., Grieger, U., Antelmann, H., Hecker, M., Nijland, R., Bron, S., and van Dijl, J.M. (2004). Two minimal Tat translocases in *Bacillus*. *Mol Microbiol* 54, 1319-1325.
- Jongbloed, J.D.H., Martin, U., Antelmann, H., Hecker, M., Tjalsma, H., Venema, G., Bron, S., van Dijl, J.M., and Muller, J. (2000). TatC is a specificity determinant for protein secretion via the twin-arginine translocation pathway. *Journal of Biological Chemistry* 275, 41350-41357.
- Jose, J., Jahnig, F., and Meyer, T.F. (1995). Common structural features of IgA1 protease-like outer membrane protein autotransporters. *Mol Microbiol* 18, 378-380.
- Jovanovic, T., Ascenso, C., Hazlett, K.R., Sikkink, R., Krebs, C., Litwiller, R., Benson, L.M., Moura, I., Moura, J.J., Radolf, J.D., *et al.* (2000). Neelaredoxin, an iron-binding protein from the syphilis spirochete, *Treponema pallidum*, is a superoxide reductase. *J Biol Chem* 275, 28439-28448.
- Kabsch, W. (1976). Solution for Best Rotation to Relate 2 Sets of Vectors. *Acta Crystallographica Section A* 32, 922-923.
- Kabsch, W. (2010a). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D-Biological Crystallography* 66, 133-144.
- Kabsch, W. (2010b). Xds. *Acta Crystallographica Section D-Biological Crystallography* 66, 125-132.
- Kabsch, W., and Sander, C. (1983). Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* 22, 2577-2637.
- Kane, C.D., Schuch, R., Day, W.A., and Maurelli, A.T. (2002). MxiE regulates intracellular expression of factors secreted by the *Shigella flexneri* 2a type III secretion system. *Journal of Bacteriology* 184, 4409-4419.
- Kaniga, K., Bossio, J.C., and Galan, J.E. (1994). The *Salmonella*-Typhimurium Invasion Genes *Invf* and *Invg* Encode Homologs of the *Arac* and *Puld* Family of Proteins. *Mol Microbiol* 13, 555-568.
- Keough, D.T., Dionysius, D.A., de Jersey, J., and Zerner, B. (1980). Iron-containing acid phosphatases: characterization of the metal-ion binding site of the enzyme from pig allantoin fluid. *Biochem Biophys Res Commun* 94, 600-605.
- Kim, D.W., Chu, H., Joo, D.H., Jang, M.S., Choi, J.H., Park, S.M., Choi, Y.J., Han, S.H., and Yun, C.H. (2008). *OspF* directly attenuates the activity of extracellular signal-regulated kinase during invasion by *Shigella flexneri* in human dendritic cells. *Mol Immunol* 45, 3295-3301.
- Kim, D.W., Lenzen, G., Page, A.L., Legrain, P., Sansonetti, P.J., and Parsot, C. (2005). The *Shigella flexneri* effector *OspG* interferes with innate immune responses by targeting ubiquitin-conjugating enzymes. *Proc Natl Acad Sci U S A* 102, 14046-14051.

Kim, E.E., and Wyckoff, H.W. (1991). Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. *J Mol Biol* 218, 449-464.

Kim, Y.G., Jeong, J.H., Ha, N.C., and Kim, K.J. (2011). Structural and functional analysis of the Lmo2642 cyclic nucleotide phosphodiesterase from *Listeria monocytogenes*. *Proteins-Structure Function and Bioinformatics* 79, 1205-1214.

Kimura, E. (2000). Dimetallic hydrolases and their models. *Current Opinion in Chemical Biology* 4, 207-213.

Klabunde, T., and Krebs, B. (1997). The dimetal center in purple acid phosphatases. *Metal Sites in Proteins and Models* 89, 177-198.

Klabunde, T., Strater, N., Frohlich, R., Witzel, H., and Krebs, B. (1996). Mechanism of Fe(III)-Zn(II) purple acid phosphatase based on crystal structures. *J Mol Biol* 259, 737-748.

Kleywegd, G., J. <http://xray.bmc.uu.se/hicup/>.

Klink, B.U., Barden, S., Heidler, T., Borchers, C., Ladwein, M., Stradal, T.E.B., Torner, K., and Heinz, D.W. (2010). Structure of *Shigella* IpgB2 in complex with human RhoA: Implications for the mechanism of bacterial GEF-mimicry. *J. Biol. Chem.* 28, 17197-17208.

Kolkman, M.A., van der Ploeg, R., Bertels, M., van Dijk, M., van der Laan, J., van Dijk, J.M., and Ferrari, E. (2008). The twin-arginine signal peptide of *Bacillus subtilis* YwbN can direct either Tat- or Sec-dependent secretion of different cargo proteins: secretion of active subtilisin via the *B. subtilis* Tat pathway. *Appl Environ Microbiol* 74, 7507-7513.

Koonin, E.V. (1994). Conserved sequence pattern in a wide variety of phosphoesterases. *Protein Sci* 3, 356-358.

Kristensen, O., Laurberg, M., Liljas, A., Kastrup, J.S., and Gajhede, M. (2004). Structural characterization of the stringent response related exopolyphosphatase/guanosine pentaphosphate phosphohydrolase protein family. *Biochemistry* 43, 8894-8900.

Kubori, T., and Galan, J.E. (2003). Temporal regulation of *Salmonella* virulence effector function by proteasome-dependent protein degradation. *Cell* 115, 333-342.

Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galan, J.E., and Aizawa, S.I. (1998). Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* 280, 602-605.

Kumaran, D., Bonanno, J.B., Burley, S.K., and Swaminathan, S. (2006). Crystal structure of phosphatidylglycerophosphatase (PGPase), a putative membrane-bound lipid phosphatase, reveals a novel binuclear metal binding site and two "proton wires". *Proteins* 64, 851-862.

Lamarche, M.G., Wanner, B.L., Crepin, S., and Harel, J. (2008). The phosphate regulon and bacterial virulence: a regulatory network connecting phosphate homeostasis and pathogenesis. *Fems Microbiology Reviews* 32, 461-473.

Langen, R., Oh, K.J., Cascio, D., and Hubbell, W.L. (2000). Crystal structures of spin labeled T4 lysozyme mutants: Implications for the interpretation of EPR spectra in terms of structure. *Biochemistry* 39, 8396-8405.

Le Brun, N.E., Andrews, S.C., Guest, J.R., Harrison, P.M., Moore, G.R., and Thomson, A.J. (1995). Identification of the ferroxidase centre of *Escherichia coli* bacterioferritin. *Biochem J* 312 (Pt 2), 385-392.

Le Gall, T., Mavris, M., Martino, M.C., Bernardini, M.L., Denamur, E., and Parsot, C. (2005). Analysis of virulence plasmid gene expression defines three classes of effectors in the type III secretion system of *Shigella flexneri*. *Microbiol-Sgm* 151, 951-962.

Leake, M.C., Greene, N.P., Godun, R.M., Granjon, T., Buchanan, G., Chen, S., Berry, R.M., Palmer, T., and Berks, B.C. (2008). Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by in vivo single-molecule imaging. *P Natl Acad Sci USA* 105, 15376-15381.

Lee, P.A., Buchanan, G., Stanley, N.R., Berks, B.C., and Palmer, T. (2002). Truncation analysis of TatA and TatB defines the minimal functional units required for protein translocation. *J Bacteriol* 184, 5871-5879.

Lee, P.A., Orriss, G.L., Buchanan, G., Greene, N.P., Bond, P.J., Punginelli, C., Jack, R.L., Sansom, M.S.P., Palmer, T., and Berks, B.C. (2006a). Cysteine-scanning mutagenesis and disulfide mapping studies of the conserved domain of the twin-arginine translocase TatB component. *Journal of Biological Chemistry* 281, 34072-34085.

Lee, P.A., Tullman-Ercek, D., and Georgiou, G. (2006b). The bacterial twin-arginine translocation pathway. *Annual Review of Microbiology* 60, 373-395.

Lenzen, C., Cool, R.H., Prinz, H., Kuhlmann, J., and Wittinghofer, A. (1998). Kinetic analysis by fluorescence of the interaction between Ras and the catalytic domain of the guanine nucleotide exchange factor Cdc25Mm. *Biochemistry* 37, 7420-7430.

Leslie, A.G.W. (1992). MOSFLM! *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* 26.

Leslie, A.G.W. (2006). The integration of macromolecular diffraction data. *Acta Crystallographica Section D-Biological Crystallography* 62, 48-57.

Lilic, M., Vujanac, M., and Stebbins, C.E. (2006). A common structural motif in the binding of virulence factors to bacterial secretion chaperones. *Molecular Cell* 21, 653-664.

Lillington, J.E., Lovett, J.E., Johnson, S., Roversi, P., Timmel, C.R., and Lea, S.M. (2011). *Shigella flexneri* Spa15 crystal structure verified in solution by double electron electron resonance. *J Mol Biol* 405, 427-435.

Lindenstrauss, U., Matos, C.F., Graubner, W., Robinson, C., and Bruser, T. (2010). Malfolded recombinant Tat substrates are Tat-independently degraded in *Escherichia coli*. *FEBS Lett* 584, 3644-3648.

Linhartova, I., Bumba, L., Masin, J., Basler, M., Osicka, R., Kamanova, J., Prochazkova, K., Adkins, I., Hejnova-Holubova, J., Sadilkova, L., *et al.* (2010). RTX proteins: a highly diverse family secreted by a common mechanism. *Fems Microbiology Reviews* 34, 1076-1112.

Liu, W., Eder, S., and Hulett, F.M. (1998). Analysis of *Bacillus subtilis* tagAB and tagDEF expression during phosphate starvation identifies a repressor role for PhoP-P. *J Bacteriol* 180, 753-758.

Liu, W., and Hulett, F.M. (1998). Comparison of PhoP binding to the *tuaA* promoter with PhoP binding to other Pho-regulon promoters establishes a *Bacillus subtilis* Pho core binding site. *Microbiology* 144 (Pt 5), 1443-1450.

Lokareddy, R.K., Lunelli, M., Eilers, B., Wolter, V., and Kolbe, M. (2010). Combination of Two Separate Binding Domains Defines Stoichiometry between Type III Secretion System Chaperone IpgC and Translocator Protein IpaB. *Journal of Biological Chemistry* 285, 39965-39975.

Lovering, A.L., Lin, L.Y., Sewell, E.W., Spreter, T., Brown, E.D., and Strynadka, N.C. (2010). Structure of the bacterial teichoic acid polymerase TagF provides insights into membrane association and catalysis. *Nat Struct Mol Biol* 17, 582-589.

Lovett (nee Banham) J, E. (2006). DPhil thesis - Exploring Long Range Structure in Chemical and Biological Systems Using Electron Paramagnetic Resonance.

Lukatela, G., Krauss, N., Theis, K., Selmer, T., Giesemann, V., von Figura, K., and Saenger, W. (1998). Crystal structure of human arylsulfatase A: The aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. *Biochemistry* 37, 3654-3664.

Lunelli, M., Lokareddy, R.K., Zychlinsky, A., and Kolbe, M. (2009). IpaB-IpgC interaction defines binding motif for type III secretion translocator. *P Natl Acad Sci USA* 106, 9661-9666.

Luo, Y., Bertero, M.G., Frey, E.A., Pfuetzner, R.A., Wenk, M.R., Creagh, L., Marcus, S.L., Lim, D., Sicheri, F., Kay, C., *et al.* (2001). Structural and biochemical characterization of the type III secretion chaperones CesT and SigE. *Nat Struct Biol* 8, 1031-1036.

Ma, X.Y., and Cline, K. (2010). Multiple precursor proteins bind individual Tat receptor complexes and are collectively transported. *Embo Journal* 29, 1477-1488.

Maldonado, B., Buchanan, G., Muller, M., Berks, B.C., and Palmer, T. (2011). Genetic Evidence for a TatC Dimer at the Core of the *Escherichia coli* Twin Arginine (Tat) Protein Translocase. *J Mol Microbiol Biotechnol* 20, 168-175.

Margraf, D., Bode, B.E., Marko, A., Schiemann, O., and Prisner, T.F. (2007). Conformational flexibility of nitroxide biradicals determined by X-band PELDOR experiments. *Molecular Physics* 105, 2153-2160.

Markey, F. (1999). What is SPR anyway? *BIAjournal*.

Marles-Wright, J., and Lewis, R.J. (2007). Stress responses of bacteria. *Curr Opin Struct Biol* 17, 755-760.

Marlovits, T.C., Kubori, T., Lara-Tejero, M., Thomas, D., Unger, V.M., and Galan, J.E. (2006). Assembly of the inner rod determines needle length in the type III secretion injectisome. *Nature* 441, 637-640.

- Marteyn, B., West, N.P., Browning, D.F., Cole, J.A., Shaw, J.G., Palm, F., Mounier, J., Prevost, M.C., Sansonetti, P., and Tang, C.M. (2010). Modulation of *Shigella* virulence in response to available oxygen in vivo. *Nature* **465**, 355-358.
- Martinez-Argudo, I., and Blocker, A.J. (2010). The *Shigella* T3SS needle transmits a signal for MxiC release, which controls secretion of effectors. *Mol Microbiol* **78**, 1365-1378.
- Matos, C.F., Di Cola, A., and Robinson, C. (2009). TatD is a central component of a Tat translocon-initiated quality control system for exported FeS proteins in *Escherichia coli*. *Embo Rep* **10**, 474-479.
- Matos, C.F.R.O., Robinson, C., and Di Cola, A. (2008). The Tat system proofreads FeS protein substrates and directly initiates the disposal of rejected molecules. *Embo Journal* **27**, 2055-2063.
- Matthews, B.W. (1968). Solvent Content of Protein Crystals. *Journal of Molecular Biology* **33**, 491-497.
- Mattoo, S., Lee, Y.M., and Dixon, J.E. (2007). Interactions of bacterial effector proteins with host proteins. *Current Opinion in Immunology* **19**, 392-401.
- Maurer, C., Panahandeh, S., Jungkamp, A.C., Moser, M., and Muller, M. (2010). TatB functions as an oligomeric binding site for folded Tat precursor proteins. *Mol Biol Cell* **21**, 4151-4161.
- Mavris, M., Page, A.L., Tournebize, R., Demers, B., Sansonetti, P., and Parsot, C. (2002). Regulation of transcription by the activity of the *Shigella flexneri* type III secretion apparatus. *Mol Microbiol* **43**, 1543-1553.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674.
- McDaniel, T.K., Jarvis, K.G., Sonnenberg, M.S., and Kaper, J.B. (1995). A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci U S A* **92**, 1664-1668.
- McDowell, M.A., Johnson, S., Deane, J.E., Cheung, M., Roehrich, A.D., Blocker, A.J., McDonnell, J.M., and Lea, S.M. (2011). Structural and functional studies on the N-terminal domain of the *Shigella* type III secretion protein MxiG. *J Biol Chem*.
- McFerrin, M.B., and Snell, E.H. (2002). The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. *Journal of Applied Crystallography* **35**, 538-545.
- Mchaourab, H.S., Kalai, T., Hideg, K., and Hubbell, W.L. (1999). Motion of spin-labeled side chains in T4 lysozyme: Effect of side chain structure. *Biochemistry* **38**, 2947-2955.
- Mchaourab, H.S., Lietzow, M.A., Hideg, K., and Hubbell, W.L. (1996). Motion of spin-labeled side chains in T4 lysozyme, correlation with protein structure and dynamics. *Biochemistry* **35**, 7692-7704.
- McNamara, B.P., Koutsouris, A., O'Connell, C.B., Nougayrede, J.P., Sonnenberg, M.S., and Hecht, G. (2001). Translocated EspF protein from enteropathogenic *Escherichia coli* disrupts host intestinal barrier function. *J Clin Invest* **107**, 621-629.

Merkx, M., and Averill, B.A. (1998a). The activity of oxidized bovine spleen purple acid phosphatase is due to an Fe(III)Zn(II) 'impurity'. *Biochemistry* 37, 11223-11231.

Merkx, M., and Averill, B.A. (1998b). Ga³⁺ as a functional substitute for Fe³⁺: preparation and characterization of the Ga³⁺Fe²⁺ and Ga³⁺Zn²⁺ forms of bovine spleen purple acid phosphatase. *Biochemistry* 37, 8490-8497.

Merkx, M., Pinkse, M.W.H., and Averill, B.A. (1999). Evidence for nonbridged coordination of p-nitrophenyl phosphate to the dinuclear Fe(III)-M(II) center in bovine spleen purple acid phosphatase during enzymatic turnover. *Biochemistry* 38, 9914-9925.

Miki, T. (1995). Interaction of Ect2 and Dbl with Rho-Related Gtpases. *Small Gtpases and Their Regulators*, Pt B 256, 90-98.

Miller, D.J., Shuvalova, L., Evdokimova, E., Savchenko, A., Yakunin, A.F., and Anderson, W.F. (2007). Structural and biochemical characterization of a novel Mn²⁺-dependent phosphodiesterase encoded by the yfcE gene. *Protein Science* 16, 1338-1348.

Milov, A.D., Maryasov, A.G., Tsvetkov, Y.D., and Raap, J. (1999). Pulsed ELDOR in spin-labeled polypeptides. *Chemical Physics Letters* 303, 135-143.

MolRep documentation <http://www.ysbl.york.ac.uk/~alexei/molrep.html#PRF> (last visited July 2011).

Mono_Q

<http://www.gelifesciences.com/aptrix/upp01077.nsf/Content/Products?OpenDocument&moduleid=165400>. Website last visited October 2011.

Montagna, L.G., Ivanov, M.I., and Bliska, J.B. (2001). Identification of residues in the N-terminal domain of the Yersinia tyrosine phosphatase that are critical for substrate recognition. *Journal of Biological Chemistry* 276, 5005-5011.

Montalibet, J., Skorey, K.I., and Kennedy, B.P. (2005). Protein tyrosine phosphatase: enzymatic assays. *Methods* 35, 2-8.

Mori, H., and Cline, K. (2002). A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid Delta pH/Tat translocase. *Journal of Cell Biology* 157, 205-210.

Mori, H., Summer, E.J., and Cline, K. (2001). Chloroplast TatC plays a direct role in thylakoid Delta pH-dependent protein transport. *Febs Letters* 501, 65-68.

Morita-Ishihara, T., Ogawa, M., Sagara, H., Yoshida, M., Katayama, E., and Sasakawa, C. (2006). Shigella Spa33 is an essential C-ring component of type III secretion machinery. *Journal of Biological Chemistry* 281, 599-607.

Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., and Peng, Z.Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. *Protein Science* 13, 1435-1448.

Mota, L.J., Journet, L., Sorg, I., Agrain, C., and Cornelis, G.R. (2005). Bacterial injectisomes: Needle length does matter. *Science* 307, 1278.

Muller, J.P., and Wagner, M. (1999). Localisation of the cell wall-associated phosphodiesterase PhoD of *Bacillus subtilis*. *Fems Microbiology Letters* 180, 287-296.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D-Biological Crystallography* 67, 355-367.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Biological Crystallography* 53, 240-255.

Namy, O., Rousset, J.P., Naphine, S., and Brierley, I. (2004). Reprogrammed genetic decoding in cellular gene expression. *Molecular Cell* 13, 157-168.

Neuhaus, F.C., and Baddiley, J. (2003). A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in gram-positive bacteria. *Microbiol Mol Biol Rev* 67, 686-723.

Ni-NTA

<http://www.qiagen.com/products/protein/purification/qiaexpressproteinpurificationsystem/nintasuperflowcartridges.aspx>. Website last visited October 2011.

Niebuhr, K., Giuriato, S., Pedron, T., Philpott, D.J., Gaits, F., Sable, J., Sheetz, M.P., Parsot, C., Sansonetti, P.J., and Payrastre, B. (2002). Conversion of PtdIns(4,5)P(2) into PtdIns(5)P by the *S.flexneri* effector IpgD reorganizes host cell morphology. *EMBO J* 21, 5069-5078.

Nolandt, O.V., Walther, T.H., Roth, S., Burck, J., and Ulrich, A.S. (2009). Structure analysis of the membrane protein TatC(d) from the Tat system of *B. subtilis* by circular dichroism. *Biochimica Et Biophysica Acta-Biomembranes* 1788, 2238-2244.

Oates, J., Mathers, J., Mangels, D., Kuhlbrandt, W., Robinson, C., and Model, K. (2003). Consensus structural features of purified bacterial TatABC complexes. *J Mol Biol* 330, 277-286.

Ogawa, M., Yoshimori, T., Suzuki, T., Sagara, H., Mizushima, N., and Sasakawa, C. (2005). Escape of intracellular *Shigella* from autophagy. *Science* 307, 727-731.

Ohlson, M.B., Huang, Z.W., Alto, N.M., Blanc, M.P., Dixon, J.E., Chai, J.J., and Miller, S.I. (2008). Structure and Function of *Salmonella* SifA Indicate that Its Interactions with SKIP, SseJ, and RhoA Family GTPases Induce Endosomal Tubulation. *Cell Host & Microbe* 4, 434-446.

Ohya, K., Handa, Y., Ogawa, M., Suzuki, M., and Sasakawa, C. (2005). IpgB1 is a novel *Shigella* effector protein involved in bacterial invasion of host cells. *Journal of Biological Chemistry* 280, 24022-24034.

Olczak, M., Morawiecka, B., and Watorek, W. (2003). Plant purple acid phosphatases - genes, structures and biological function. *Acta Biochim Pol* 50, 1245-1256.

Olive, A.J., Kenjale, R., Espina, M., Moore, D.S., Picking, W.L., and Picking, W.D. (2007). Bile salts stimulate recruitment of IpaB to the *Shigella flexneri* surface, where it colocalizes with IpaD at the tip of the type III secretion needle. *Infect Immun* 75, 2626-2629.

Oosterhuis, W.T. (1974). The electronic states of iron in some natural iron compounds: Determination by Mossbauer and ESR spectroscopy. *Structure and Bonding* 20, 59 - 99.

Oresnik, I.J., Ladner, C.L., and Turner, R.J. (2001). Identification of a twin-arginine leader-binding protein. *Mol Microbiol* 40, 323-331.

Osipiuk, J., Evdokimova, E., Kudritska, M., Savchenko, A., Edwards, A., Joachimiak, A. ((to be published)). Ycfl, a putative structural protein from *Rhodopseudomonas palustris* (PDB 2GYQ).

Page, A.L., and Parsot, C. (2002). Chaperones of the type III secretion pathway: jacks of all trades. *Mol Microbiol* 46, 1-11.

Page, A.L., Sansonetti, P., and Parsot, C. (2002). Spa15 of *Shigella flexneri*, a third type of chaperone in the type III secretion pathway. *Mol Microbiol* 43, 1533-1542.

Palmer, T., Sargent, F., and Berks, B.C. (2005). Export of complex cofactor-containing proteins by the bacterial Tat pathway. *Trends Microbiol* 13, 175-180.

Papalia, G.A., Baer, M., Luehrsen, K., Nordin, H., Flynn, P., and Myszka, D.G. (2006). High-resolution characterization of antibody fragment/antigen interactions using Biacore T100. *Anal Biochem* 359, 112-119.

Papanikou, E., Karamanou, S., and Economou, A. (2007). Bacterial protein secretion through the translocase nanomachine. *Nat Rev Microbiol* 5, 839-851.

Park, H., Valencia-Gallardo, C., Sharff, A., Tran Van Nhieu, G., and Izard, T. (2011). A novel Vinculin binding site of the IpaA invasin of *Shigella*. *J Biol Chem* 286, 23214-23221.

Parsot, C. (1994). *Shigella flexneri*: genetics of entry and intercellular dissemination in epithelial cells. *Curr Top Microbiol Immunol* 192, 217-241.

Parsot, C. (2005). *Shigella* spp. and enteroinvasive *Escherichia coli* pathogenicity factors. *Fems Microbiology Letters* 252, 11-18.

Parsot, C., Ageron, E., Penno, C., Mavris, M., Jamoussi, K., d'Hauteville, H., Sansonetti, P., and Demers, B. (2005). A secreted anti-activator, OspD1, and its chaperone, Spa15, are involved in the control of transcription by the type III secretion apparatus activity in *Shigella flexneri*. *Mol Microbiol* 56, 1627-1635.

Parsot, C., Hamiaux, C., and Page, A.L. (2003). The various and varying roles of specific chaperones in type III secretion systems. *Current Opinion in Microbiology* 6, 7-14.

Penno, C., and Parsot, C. (2006). Transcriptional slippage in *mxiE* controls transcription and translation of the downstream *mxiD* gene, which encodes a component of the *Shigella flexneri* type III secretion apparatus. *J Bacteriol* 188, 1196-1198.

Penno, C., Sansonetti, P., and Parsot, C. (2005). Frameshifting by transcriptional slippage is involved in production of MxiE, the transcription activator regulated by the activity of the type III secretion apparatus in *Shigella flexneri*. *Mol Microbiol* 56, 204-214.

Perdomo, J.J., Gounon, P., and Sansonetti, P.J. (1994). Polymorphonuclear leukocyte transmigration promotes invasion of colonic epithelial monolayer by *Shigella flexneri*. *J Clin Invest* 93, 633-643.

Perez-Rodriguez, R., Fisher, A.C., Perlmutter, J.D., Hicks, M.G., Chanal, A., Santini, C.L., Wu, L.F., Palmer, T., and DeLisa, M.P. (2007). An essential role for the DnaK molecular chaperone in stabilizing over-expressed substrate proteins of the bacterial twin-arginine translocation pathway. *J Mol Biol* 367, 715-730.

Peroutka Iii, R.J., Orcutt, S.J., Strickler, J.E., and Butt, T.R. (2011). SUMO fusion technology for enhanced protein expression and purification in prokaryotes and eukaryotes. *Methods Mol Biol* 705, 15-30.

Phan, J., Tropea, J.E., and Waugh, D.S. (2004). Structure of the *Yersinia pestis* type III secretion chaperone SycH in complex with a stable fragment of YscM2. *Acta Crystallogr D Biol Crystallogr* 60, 1591-1599.

Philpott, D.J., Yamaoka, S., Israel, A., and Sansonetti, P.J. (2000). Invasive *Shigella flexneri* activates NF-kappa B through a lipopolysaccharide-dependent innate intracellular response and leads to IL-8 expression in epithelial cells. *J Immunol* 165, 903-914.

Pilonieta, M.C., and Munson, G.P. (2008). The chaperone IpgC copurifies with the virulence regulator MxiE. *J Bacteriol* 190, 2249-2251.

Poole, C.P. (1996). Electron spin resonance : a comprehensive treatise on experimental techniques (Mineola, N.Y.: Dover Publications).

Pop, O., Martin, U., Abel, C., and Muller, J.P. (2002). The twin-arginine signal peptide of PhoD and the TatAd/Cd proteins of *Bacillus subtilis* form an autonomous Tat translocation system. *J Biol Chem* 277, 3268-3273.

Porcelli, I., de Leeuw, E., Wallis, R., van den Brink-van der Laan, E., de Kruijff, B., Wallace, B.A., Palmer, T., and Berks, B.C. (2002). Characterization and membrane assembly of the TatA component of the *Escherichia coli* twin-arginine protein transport system. *Biochemistry* 41, 13690-13697.

Pragai, Z., Allenby, N.E., O'Connor, N., Dubrac, S., Rapoport, G., Msadek, T., and Harwood, C.R. (2004). Transcriptional regulation of the phoPR operon in *Bacillus subtilis*. *J Bacteriol* 186, 1182-1190.

Prisner, T., Rohrer, M., and MacMillan, F. (2001). Pulsed EPR spectroscopy: Biological applications. *Annual Review of Physical Chemistry* 52, 279-313.

Promega_Biomath <http://www.promega.com/resources/tools/biomath-calculators/calculators-and-additional-conversions/tm-melting-temperature-calculations-for-oligos/>.

Qi, Y., Kobayashi, Y., and Hulett, F.M. (1997). The pst operon of *Bacillus subtilis* has a phosphate-regulated promoter and is involved in phosphate transport but not in regulation of the Pho regulon. *Journal of Bacteriology* 179, 2534-2539.

QIAprep <http://www.qiagen.com/literature/handbooks/literature.aspx?id=1000248> (last visited July 2011).

QIAquick www.qiagen.com/literature/render.aspx?id=103715 (last visited July 2011)

QuikChange <http://www.genomics.agilent.com/files/Manual/200523.pdf> (last visited July 2011).

Ramamurthi, K.S., and Schneewind, O. (2002). *Yersinia enterocolitica* type III secretion: Mutational analysis of the yopQ secretion signal. *Journal of Bacteriology* 184, 3321-3328.

Ray, N., Oates, J., Turner, R.J., and Robinson, C. (2003). DmsD is required for the biogenesis of DMSO reductase in *Escherichia coli* but not for the interaction of the DmsA signal peptide with the Tat apparatus. *FEBS Lett* 534, 156-160.

Rayleigh, L. (1907). On the Dynamical Theory of Gratings. *Proceedings of the Royal Society of London. Series A* 79, 399-416.

Read, R.J. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallographica Section D-Biological Crystallography* 57, 1373-1382.

Rhodes, G. (2006). *Crystallography made crystal clear : a guide for users of macromolecular models*, 3rd edn (Amsterdam; Boston: Elsevier/Academic Press).

Richter, S., and Bruser, T. (2005). Targeting of unfolded PhoA to the TAT translocon of *Escherichia coli*. *Journal of Biological Chemistry* 280, 42723-42730.

Richter, S., Lindenstrauss, U., Lucke, C., Bayliss, R., and Bruser, T. (2007). Functional tat transport of unstructured, small, hydrophilic proteins. *Journal of Biological Chemistry* 282, 33257-33264.

Richter, W. (2002). 3',5' Cyclic nucleotide phosphodiesterases class III: members, structure, and catalytic mechanism. *Proteins* 46, 278-286.

Robinson, C., Matos, C.F.R.O., Beck, D., Ren, C., Lawrence, J., Vasisht, N., and Mendel, S. (2011). Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. *Biochimica Et Biophysica Acta-Biomembranes* 1808, 876-884.

Rodgers, L., Gamez, A., Riek, R., and Ghosh, P. (2008). The type III secretion chaperone SycE promotes a localized disorder-to-order transition in the natively unfolded effector YopE. *Journal of Biological Chemistry* 283, 20857-20863.

Rodgers, L., Mukerjea, R., Birtalan, S., Friedberg, D., and Ghosh, P. (2010). A solvent-exposed patch in chaperone-bound YopE is required for translocation by the type III secretion system. *J Bacteriol* 192, 3114-3122.

Rogers, M.S., Hurtado-Guerrero, R., Firbank, S.J., Halcrow, M.A., Dooley, D.M., Phillips, S.E., Knowles, P.F., and McPherson, M.J. (2008). Cross-link formation of the cysteine 228-tyrosine 272 catalytic cofactor of galactose oxidase does not require dioxygen. *Biochemistry* 47, 10428-10439.

Rose, R.W., Bruser, T., Kissinger, J.C., and Pohlschroder, M. (2002). Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Mol Microbiol* 45, 943-950.

Rossman, K.L., Der, C.J., and Sondek, J. (2005). GEF means go: Turning on Rho GTPases with guanine nucleotide-exchange factors. *Nature Reviews Molecular Cell Biology* 6, 167-180.

Rudolph, M.G., Weise, C., Miold, S., Hillenbrand, B., Bader, B., Wittinghofer, A., and Hardt, W.D. (1999). Biochemical analysis of SopE from *Salmonella typhimurium*, a highly efficient guanosine nucleotide exchange factor for RhoGTPases. *J Biol Chem* 274, 30501-30509.

Ruiz-Albert, J., Yu, X.J., Beuzon, C.R., Blakey, A.N., Galyov, E.E., and Holden, D.W. (2002). Complementary activities of SseJ and SifA regulate dynamics of the *Salmonella typhimurium* vacuolar membrane. *Mol Microbiol* 44, 645-661.

Rupp, B. (2010). *Biomolecular crystallography* (New York, NY: Garland Science).

Rusnak, F., Ascenso, C., Moura, I., and Moura, J.J.G. (2002). Superoxide reductase activities of neelaredoxin and desulfoferrodoxin Metalloproteins. *Superoxide Dismutase* 349, 243-258.

Sal-Man, N., Biemans-Oldehinkel, E., and Finlay, B.B. (2009). Structural microengineers: pathogenic *Escherichia coli* redesigns the actin cytoskeleton in host cells. *Structure* 17, 15-19.

Sambasivarao, D., Turner, R.J., Simala-Grantz, J.L., Shaw, G., Hu, J., and Weiner, J.H. (2000). Multiple roles for the twin arginine leader sequence of dimethyl sulfoxide reductase of *Escherichia coli*. *Journal of Biological Chemistry* 275, 22526-22531.

Samudrala, R., Heffron, F., and McDermott, J.E. (2009). Accurate Prediction of Secreted Substrates and Identification of a Conserved Putative Secretion Signal for Type III Secretion Systems. *Plos Pathogens* 5.

Sanders, C., Wethkamp, N., and Lill, H. (2001). Transport of cytochrome c derivatives by the bacterial Tat protein translocation system. *Mol Microbiol* 41, 241-246.

Sanger, J.M., Chang, R., Ashton, F., Kaper, J.B., and Sanger, J.W. (1996). Novel form of actin-based motility transports bacteria on the surfaces of infected cells. *Cell Motil Cytoskeleton* 34, 279-287.

Sansonetti, P.J., Arondel, J., Huerre, M., Harada, A., and Matsushima, K. (1999). Interleukin-8 controls bacterial transepithelial translocation at the cost of epithelial destruction in experimental shigellosis. *Infect Immun* 67, 1471-1480.

Sansonetti, P.J., Phalipon, A., Arondel, J., Thirumalai, K., Banerjee, S., Akira, S., Takeda, K., and Zychlinsky, A. (2000). Caspase-1 activation of IL-1beta and IL-18 are essential for *Shigella flexneri*-induced inflammation. *Immunity* 12, 581-590.

Sargent, F. (2007). Constructing the wonders of the bacterial world: biosynthesis of complex enzymes. *Microbiol-Sgm* 153, 633-651.

Sargent, F., Bogsch, E.G., Stanley, N.R., Wexler, M., Robinson, C., Berks, B.C., and Palmer, T. (1998). Overlapping functions of components of a bacterial Sec-independent protein export pathway. *Embo Journal* 17, 3640-3650.

Sargent, F., Stanley, N.R., Berks, B.C., and Palmer, T. (1999). Sec-independent protein translocation in *Escherichia coli* - A distinct and pivotal role for the TatB protein. *Journal of Biological Chemistry* 274, 36073-36082.

Saridakis, E., and Chayen, N.E. (2009). Towards a 'universal' nucleant for protein crystallization. *Trends Biotechnol* 27, 99-106.

Savkovic, S.D., Koutsouris, A., and Hecht, G. (1997). Activation of NF-kappaB in intestinal epithelial cells by enteropathogenic *Escherichia coli*. *Am J Physiol* 273, C1160-1167.

Sawaya, M.R. (2007). Characterizing a crystal from an initial native dataset. *Methods Mol Biol* 364, 95-120.

Schallmeyer, M., Singh, A., and Ward, O.P. (2004). Developments in the use of *Bacillus* species for industrial production. *Can J Microbiol* 50, 1-17.

Schenk, G., Elliott, T.W., Leung, E., Carrington, L.E., Mitic, N., Gahan, L.R., and Guddat, L.W. (2008). Crystal structures of a purple acid phosphatase, representing different steps of this enzyme's catalytic cycle. *Bmc Structural Biology* 8, 6.

Schenk, G., Gahan, L.R., Carrington, L.E., Mitic, N., Valizadeh, M., Hamilton, S.E., de Jersey, J., and Guddat, L.W. (2005). Phosphate forms an unusual tripodal complex with the Fe-Mn center of sweet potato purple acid phosphatase. *Proc Natl Acad Sci U S A* 102, 273-278.

Schiemann, O., and Prisner, T.F. (2007). Long-range distance determinations in biomacromolecules by EPR spectroscopy. *Quarterly Reviews of Biophysics* 40, 1-53.

Schmidt, M.A. (2010). LEEways: tales of EPEC, ATEC and EHEC. *Cell Microbiol* 12, 1544-1552.

Schmitz, A.M., Morrison, M.F., Agunwamba, A.O., Nibert, M.L., and Lesser, C.F. (2009). Protein interaction platforms: visualization of interacting proteins in yeast. *Nature Methods* 6, 500-502.

Schroeder, G.N., and Hilbi, H. (2008). Molecular pathogenesis of *Shigella* spp.: controlling host cell signaling, invasion, and death by type III secretion. *Clin Microbiol Rev* 21, 134-156.

Schubot, F.D., Jackson, M.W., Penrose, K.J., Cherry, S., Tropea, J.E., Plano, G.V., and Waugh, D.S. (2005). Three-dimensional structure of a macromolecular assembly that regulates type III secretion in *Yersinia pestis*. *J Mol Biol* 346, 1147-1161.

Schuch, R., and Maurelli, A.T. (2001). MxiM and MxiJ, base elements of the Mxi-Spa type III secretion system of *Shigella*, interact with and stabilize the MxiD secretin in the cell envelope. *Journal of Bacteriology* 183, 6991-6998.

Sezer, D., Freed, J.H., and Roux, B. (2008a). Parametrization, molecular dynamics simulation, and calculation of electron spin resonance spectra of a nitroxide spin label on a polyalanine alpha-helix. *Journal of Physical Chemistry B* 112, 5755-5767.

Sezer, D., Freed, J.H., and Roux, B. (2008b). Simulating electron spin resonance spectra of nitroxide spin labels from molecular dynamics and stochastic trajectories. *Journal of Chemical Physics* 128, 165.

Sezer, D., Freed, J.H., and Roux, B. (2009). Multifrequency Electron Spin Resonance Spectra of a Spin-Labeled Protein Calculated from Molecular Dynamics Simulations. *J Am Chem Soc* 131, 2597-2605.

Sherman, F., Stewart, J.W., and Tsunasawa, S. (1985). Methionine or Not Methionine at the Beginning of a Protein. *Bioessays* 3, 27-31.

Shriver, and Atkins (2002). *Inorganic Chemistry* ((Oxford University Press).

Silva, G., LeGall, J., Xavier, A.V., Teixeira, M., and Rodrigues-Pousada, C. (2001). Molecular characterization of *Desulfovibrio gigas* neelaredoxin, a protein involved in oxygen detoxification in anaerobes. *J Bacteriol* 183, 4413-4420.

Simovitch, M., Sason, H., Cohen, S., Zahavi, E.E., Melamed-Book, N., Weiss, A., Aroeti, B., and Rosenshine, I. (2010). EspM inhibits pedestal formation by enterohaemorrhagic *Escherichia coli* and enteropathogenic *E. coli* and disrupts the architecture of a polarized epithelial monolayer. *Cell Microbiol* 12, 489-505.

Skoudy, A., Mounier, J., Aruffo, A., Ohayon, H., Gounon, P., Sansonetti, P., and Tran Van Nhieu, G. (2000). CD44 binds to the *Shigella* IpaB protein and participates in bacterial invasion of epithelial cells. *Cell Microbiol* 2, 19-33.

Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I.A., Lesley, S.A., and Godzik, A. (2007). XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* 23, 3403-3405.

Smith, W.J., Hamel, B., Yohe, M.E., Sondek, J., Cerione, R.A., and Snyder, J.T. (2005). A Cdc42 mutant specifically activated by intersectin. *Biochemistry* 44, 13282-13290.

Smoukov, S.K., Quaroni, L., Wang, X., Doan, P.E., Hoffman, B.M., and Que, L., Jr. (2002). Electro-nuclear double resonance spectroscopic evidence for a hydroxo-bridge nucleophile involved in catalysis by a dinuclear hydrolase. *J Am Chem Soc* 124, 2595-2603.

Snyder, L., and Champness, W. (2007). *Molecular genetics of bacteria*, 3rd edn (Washington, D.C.: ASM Press).

Sowa, G.Z., and Qin, P.Z. (2008). Site-directed Spin Labeling Studies on Nucleic Acid Structure and Dynamics. *Progress in Nucleic Acid Research and Molecular Biology* 82, 147-197.

Stebbins, C.E., and Galan, J.E. (2001a). Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* 414, 77-81.

Stebbins, C.E., and Galan, J.E. (2001b). Structural mimicry in bacterial virulence. *Nature* 412, 701-705.

Stec, B., Holtz, K.M., and Kantrowitz, E.R. (2000). A revised mechanism for the alkaline phosphatase reaction involving three metal ions. *J Mol Biol* 299, 1303-1311.

Steinhoff, H.J., and Hubbell, W.L. (1996). Calculation of electron paramagnetic resonance spectra from Brownian dynamics trajectories: Application to nitroxide side chains in proteins. *Biophysical Journal* 71, 2201-2212.

Steller, I., Bolotovskiy, R., and Rossmann, M.G. (1997). An algorithm for automatic indexing of oscillation images using Fourier analysis. *Journal of Applied Crystallography* 30, 1036-1040.

Stoica, I. (2004). Solvent interactions and protein dynamics in spin-labeled T4 lysozyme. *Journal of Biomolecular Structure & Dynamics* 21, 745-760.

Stoll, S., and Schweiger, A. (2006). EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J Magn Reson* 178, 42-55.

Strater, N., Jasper, B., Scholte, M., Krebs, B., Duff, A.P., Langley, D.B., Han, R.L., Averill, B.A., Freeman, H.C., and Guss, J.M. (2005). Crystal structures of recombinant human purple acid phosphatase with and without an inhibitory conformation of the repression loop. *Journal of Molecular Biology* 351, 233-246.

Strauch, E.M., and Georgiou, G. (2007). *Escherichia coli* tatC mutations that suppress defective twin-arginine transporter signal peptides. *J Mol Biol* 374, 283-291.

Studier, F.W., and Moffatt, B.A. (1986). Use of Bacteriophage-T7 Rna-Polymerase to Direct Selective High-Level Expression of Cloned Genes. *Journal of Molecular Biology* 189, 113-130.

Su, J., Schlicker, C., and Forchhammer, K. (2011). A Third Metal Is Required for Catalytic Activity of the Signal-transducing Protein Phosphatase M tPphA. *J Biol Chem* 286, 13481-13488.

Superdex_75/200

<http://www.gelifesciences.com/aptrix/upp01077.nsf/Content/Products?OpenDocument&moduleid=165422>. Website last visited October 2011.

Tamano, K., Katayama, E., Toyotome, T., and Sasakawa, C. (2002). Shigella Spa32 is an essential secretory protein for functional type III secretion machinery and uniformity of its needle length. *Journal of Bacteriology* 184, 1244-1252.

Tarry, M.J., Schafer, E., Chen, S., Buchanan, G., Greene, N.P., Lea, S.M., Palmer, T., Saibil, H.R., and Berks, B.C. (2009). Structural analysis of substrate binding by the TatBC component of the twin-arginine protein transport system. *Proc Natl Acad Sci U S A* 106, 13284-13289.

Taylor, G. (2003). The phase problem. *Acta Crystallographica Section D-Biological Crystallography* 59, 1881-1890.

Ten Eyck, L.F. (1973). Crystallographic Fast Fourier-Transforms. *Acta Crystallographica A* 29, 183-191.

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 72, 211-222.

Thomas, N.A., Deng, W., Puente, J.L., Frey, E.A., Yip, C.K., Strynadka, N.C., and Finlay, B.B. (2005). CesT is a multi-effector chaperone and recruitment factor required for the efficient type III secretion of both LEE- and non-LEE-encoded effectors of enteropathogenic *Escherichia coli*. *Mol Microbiol* 57, 1762-1779.

Thompson, A. (2010). personal communication.

Tombolato, F., Ferrarini, A., and Freed, J.H. (2006). Dynamics of the nitroxide side chain in spin-labeled proteins. *Journal of Physical Chemistry B* **110**, 26248-26259.

Toyotome, T., Suzuki, T., Kuwae, A., Nonaka, T., Fukuda, H., Imajoh-Ohmi, S., Toyofuku, T., Hori, M., and Sasakawa, C. (2001). Shigella protein IpaH(9.8) is secreted from bacteria within mammalian cells and transported to the nucleus. *Journal of Biological Chemistry* **276**, 32071-32079.

Trame, C.B., and McKay, D.B. (2003). Structure of the Yersinia enterocolitica molecular-chaperone protein SycE. *Acta Crystallogr D Biol Crystallogr* **59**, 389-392.

Tran Van Nhieu, G., Caron, E., Hall, A., and Sansonetti, P.J. (1999). IpaC induces actin polymerization and filopodia formation during Shigella entry into epithelial cells. *EMBO J* **18**, 3249-3262.

Tronrud, D.E. (2004). Introduction to macromolecular refinement. *Acta Crystallographica Section D-Biological Crystallography* **60**, 2156-2168.

Tseng, T.T., Tyler, B.M., and Setubal, J.C. (2009). Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol* **9 Suppl 1**, S2.

Tullman-Ercek, D., DeLisa, M.P., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T., and Georgiou, G. (2007). Export pathway selectivity of Escherichia coli twin arginine translocation signal peptides. *Journal of Biological Chemistry* **282**, 8309-8316.

Ubbink, M., Worrall, J.A.R., Canters, G.W., Groenen, E.J.J., and Huber, M. (2002). Paramagnetic resonance of biological metal centers. *Annual Review of Biophysics and Biomolecular Structure* **31**, 393-422.

Vagin, A., and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025.

Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallographica Section D-Biological Crystallography* **66**, 22-25.

Valderrama, F., Cordeiro, J.V., Schleich, S., Frischknecht, F., and Way, M. (2006). Vaccinia virus-induced cell motility requires F11L-mediated inhibition of RhoA signaling. *Science* **311**, 377-381.

Van der Merwe, P.A. (2001). Surface Plasmon Resonance <http://users.path.ox.ac.uk/~vdmerwe/internal/spr.pdf>.

Van Eerde, A., Hamiaux, C., Perez, J., Parsot, C., and Dijkstra, B.W. (2004). Structure of Spa15, a type III secretion chaperone from Shigella flexneri with broad specificity. *Embo Rep* **5**, 477-483.

Vassilyev, D.G., and Symersky, J. (2007). Crystal structure of pyruvate dehydrogenase phosphatase 1 and its functional implications. *Journal of Molecular Biology* **370**, 417-426.

Veenendaal, A.K.J., Hodgkinson, J.L., Schwarzer, L., Stabat, D., Zenk, S.F., and Blocker, A.J. (2007). The type III secretion system needle tip complex mediates host cell sensing and translocon insertion. *Mol Microbiol* **63**, 1719-1730.

- Vershinina, O.A., and Znamenskaya, L.V. (2002). The Pho regulons of bacteria. *Microbiology* 71, 497-511.
- Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). Automated structure solution with autoSHARP. *Methods Mol Biol* 364, 215-230.
- Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011). Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* 67, 293-302.
- Wallace, A.C., Laskowski, R.A., and Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* 8, 127-134.
- Walter, T.S., Meier, C., Assenberg, R., Au, K.F., Ren, J.S., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I., and Grimes, J.M. (2006). Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* 14, 1617-1622.
- Walther, T.H., Grage, S.L., Roth, N., and Ulrich, A.S. (2010). Membrane Alignment of the Pore-Forming Component TatA(d) of the Twin-Arginine Translocase from *Bacillus subtilis* Resolved by Solid-State NMR Spectroscopy. *J Am Chem Soc* 132, 15945-15956.
- Watarai, M., Funato, S., and Sasakawa, C. (1996). Interaction of Ipa proteins of *Shigella flexneri* with alpha5beta1 integrin promotes entry of the bacteria into mammalian cells. *J Exp Med* 183, 991-999.
- Wexler, M., Sargent, F., Jack, R.L., Stanley, N.R., Bogsch, E.G., Robinson, C., Berks, B.C., and Palmer, T. (2000). TatD is a cytoplasmic protein with DNase activity - No requirement for TatD family proteins in Sec-independent protein export. *Journal of Biological Chemistry* 275, 16717-16722.
- White, G.F., Schermann, S.M., Bradley, J., Roberts, A., Greene, N.P., Thomson, A.J., and Berks, B.C. (2010). Subunit Organization in the TatA Complex of the Twin Arginine Protein Translocase A SITE-DIRECTED EPR SPIN LABELING STUDY. *Journal of Biological Chemistry* 285, 2294-2301.
- Widdick, D.A., Eijlander, R.T., van Dijl, J.M., Kuipers, O.P., and Palmer, T. (2008). A facile reporter system for the experimental identification of twin-arginine translocation (Tat) signal peptides from all kingdoms of life. *Journal of Molecular Biology* 375, 595-603.
- Wilharm, G., Dittmann, S., Schmid, A., and Heesemann, J. (2007). On the role of specific chaperones, the specific ATPase, and the proton motive force in type III secretion. *International Journal of Medical Microbiology* 297, 27-36.
- Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* 43, 186-190.
- Woestyn, S., Sory, M.P., Boland, A., Lequenne, O., and Cornelis, G.R. (1996). The cytosolic SycE and SycH chaperones of *Yersinia* protect the region of YopE and YopH involved in translocation across eukaryotic cell membranes. *Mol Microbiol* 20, 1261-1271.
- Wood, R.W. (1902). On a Remarkable Case of Uneven Distribution of Light in a Diffraction Grating Spectrum. *Proc. Phys. Soc. London* 18.

Wooldridge, K. (2009). Bacterial secreted proteins : secretory mechanisms and role in pathogenesis ([Wyomondham]: Caister Academic Press).

Xia, Y., Zhao, J., Chen, H., Liu, X., Wang, Y., Tian, F., Zhang, H.P., Zhang, H., and Chen, W. (2010). Extracellular secretion in *Bacillus subtilis* of a cytoplasmic thermostable beta-galactosidase from *Geobacillus stearothermophilus*. *J Dairy Sci* 93, 2838-2845.

Xie, T., Penelle, J., and Verraver, M. (2002). Experimental investigation on the reliability of routine SEC-MALLS for the determination of absolute molecular weights in the oligomeric range. *Polymer* 43, 3973-3977.

Yahr, T.L., and Wickner, W.T. (2001). Functional reconstitution of bacterial Tat translocation in vitro. *Embo Journal* 20, 2472-2479.

Yamane, K., and Maruo, B. (1978a). Alkaline-Phosphatase Possessing Alkaline Phosphodiesterase Activity and Other Phosphodiesterases in *Bacillus-Subtilis*. *Journal of Bacteriology* 134, 108-114.

Yamane, K., and Maruo, B. (1978b). Purification and Characterization of Extracellular Soluble and Membrane-Bound Insoluble Alkaline-Phosphatases Possessing Phosphodiesterase Activities in *Bacillus-Subtilis*. *Journal of Bacteriology* 134, 100-107.

Yang, H.Y., and Roberts, M.F. (2002). Cloning, overexpression, and characterization of a bacterial Ca²⁺-dependent phospholipase D. *Protein Science* 11, 2958-2968.

Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* 21, 3369-3376.

Yasmin, S., Andrews, S.C., Moore, G.R., and Le Brun, N.E. (2011). A New Role for Heme, Facilitating Release of Iron from the Bacterioferritin Iron Biomineral. *Journal of Biological Chemistry* 286, 3473-3483.

Yip, C.K., Finlay, B.B., and Strynadka, N.C.J. (2005). Structural characterization of a type III secretion system filament protein in complex with its chaperone. *Nature Structural & Molecular Biology* 12, 75-81.

Yong, S.C., Roversi, P., Lillington, J.E.D., Zeldin, O.B., Garman, E.F., Lea, S.M., Berks, B.C. ((to be published)). Recycling phosphorous: structure and mechanism of an iron and calcium-dependent alkaline phosphatase.

Yu, X.J., McGourty, K., Liu, M., Unsworth, K.E., and Holden, D.W. (2010). pH Sensing by Intracellular *Salmonella* Induces Effector Translocation. *Science* 328, 1040-1043.

Yuan, J.J., Zweers, J.C., van Dijl, J.M., and Dalbey, R.E. (2010). Protein transport across and into cell membranes in bacteria and archaea. *Cellular and Molecular Life Sciences* 67, 179-199.

Zambonelli, C., Casali, M., and Roberts, M.F. (2003). Mutagenesis of putative catalytic and regulatory residues of *Streptomyces chromofuscus* phospholipase D differentially modifies phosphatase and phosphodiesterase activities. *J Biol Chem* 278, 52282-52289.

Zambonelli, C., and Roberts, M.F. (2003). An iron-dependent bacterial phospholipase D reminiscent of purple acid phosphatases. *J Biol Chem* 278, 13706-13711.

Zhang, X., Cekan, P., Sigurdsson, P., and Qin, P.Z. (2009). Studying RNA Using Site-Directed Spin-Labeling and Continuous-Wave Electron Paramagnetic Resonance Spectroscopy Methods in *Enzymology* 469, 303-328.

Zielke, V., Eickmeier, H., Hideg, K., Reuter, H., and Steinhoff, H.J. (2008). A commonly used spin label: S-(2,2,5,5-tetramethyl-1-oxyl-Delta(3)-pyrrolin-3-ylmethyl) methanethiosulfonate. *Acta Crystallographica Section C-Crystal Structure Communications* 64, O586-589.

Appendices.

A. CW-EPR experimental setup.

i. Spectrometer tuning.

Coarse coupling of the spectrometer is carried out initially: The gunn diode microwave source is used to pass a range of frequencies so that microwave resonance with the cavity, the dip, may be found. The detector records the microwave power reflected back from the cavity. This is a minimum at resonance, and so a dip is observed.

The sample is inserted into the cavity, and the microwave attenuation decreased to 25 dB. The high power allows the dip to be easily found. Tuning is performed without the reference arm current on. The dip is centred, maximised and symmetrised, before the bias/reference arm is reactivated.

Following coarse tuning, the spectrometer is fine tuned. In operate mode, only power at the resonant frequency is transferred to the cavity. The lock offset of the automatic frequency control is first centred. At high attenuation to protect the diode (45 dB), the bias is adjusted, establishing the current at the diode (200 μ A). The relationship between power and resulting current is linear if the power levels are kept high. As a result, the diode currents are kept at 200 μ A and extra power is provided to the detector by splitting some of the source power into a reference arm. This does not irradiate the sample and goes straight to the detector diode.

The cavity and waveguide are critically coupled/matched, achieved by decreasing the attenuation while adjusting the iris to maintain diode current. The iris admits the microwaves into the cavity and is screwed more/less tightly while simultaneously increasing the power, such that the diode current recorded does not change with the power. Critical coupling is indicated by independence of diode current with microwave power.

ii. Spectrometer cooling.

The cryostat outer vacuum chamber is evacuated using the vacuum pump after isolation and evacuation of the pump line. Pumping is completed at 10^{-4} torr.

The system is connected, with the gas flow controller VC41 coupled to the cryostat and to the helium pump for temperature control. The transfer tube leg is inserted into the helium dewar, venting if pressure rises upon insertion above 0.5 bar. The head end is inserted into the cryostat, but only connected when the flow rate reaches 1.5 l/hr. Upon a decrease in temperature, the desired level is set on the temperature controller, while reducing the helium flow to a minimum.

Samples are changed with care taken to ensure that no air sucked into the cooled apparatus. With the pump removed and pressure at one atmosphere, the heater is turned off and VC needle valve isolated. The sample is quickly withdrawn and replaced with a new sample or cap, before reopening the needle valve and setting automatic temperature control.

iii. Easyspin.

CW data were analysed using the program EasySpin. An example m.file is given in Table A.i for plotting the data with respect to both magnetic field and g-value.

Easyspin/Matab m.file.
<pre>%Easyspin reading spectrometer data. close all; clear all; [B1,data1,pars1]=eprload('Spa15.DTA'); nu_1 = pars1.MWFQ; g_1 = ((6.636*10^-34)*nu_1)/((9.274*10^-24)*B1*10^(-4)); [B2,data2,pars2]=eprload('Tempol.DTA'); nu_2 = pars2.MWFQ; g_2 = ((6.636*10^-34)*nu_2)/((9.274*10^-24)*B2*10^(-4)); figure(1); plot(B1,data1,B2,data2);axis 'tight' figure(2); plot(g_1,data1,g_2,data2);axis 'tight'; xlabel('magnetic field [mT]');ylabel('intensity');title('EPRspectrum')</pre>

Table A.i An m.file for the plotting of data through Easyspin.

B. DEER experimental setup.

The apparatus is cooled and the pre-cooled sample in a capillary tube inserted. The spectrometer is tuned in continuous wave mode with the reference arm off - attenuation is decreased (20dB) to see the dip, which is centred. The attenuation is increased again (60 dB) and the reference arm switched on.

In pulse mode, an echo experiment is performed. Both real and imaginary components of the echo are detected, most easily seen if attenuation is decreased. The real part of the echo is maximised and symmetrised (with phase) for all channels.

A field sweep experiment locates the EPR spectrum. The centre of the field is found which will be coincident with the pump pulse field/frequency. Flipping of the pump pulse is modified via the inverted echo sequence, and the pump power channel increased until the echo inversion is maximised.

An Eldor experiment is next carried out. The pump frequency remains invariant and the second frequency implemented: For nitroxides, the frequency is increased 65 MHz from the main frequency. Both intensity and phase of the probe π pulse channel are altered. The observer $\pi/2$ pulse is phase cycled relative to this for removal of receiver artifacts, thus each of the channels is set up.

A standing DEER experiment is performed. This is where the pump pulse occurs at the time of the undetected echo ($t=0$), which allows adjusting of the signal phase. For a deuterated solvent, $\tau_1 = 400$ ns and $\tau_2 = 800$ ns are suitable parameters for the standing DEER.

Finally, the DEER experiment is run with the parameters determined above. This is often repeated to increase signal:noise ratio.

C. SPR experimental setup.

The CM5 chip is docked, primed and sensorgram started	The running buffer is flowed over the chip at a suitable temperature for the interaction. The flow rate is kept low for activation (around 10 $\mu\text{l}/\text{min}$).
Activation	Ethyl(dimethylaminopropyl) carbodiimide (EDC – 0.4 M) and N-hydroxysuccinimide (NHS - 0.1 M) are mixed in 1:1 ratio. 70 – 140 μl are injected onto chip lanes. This activates the carboxymethyl groups on the dextran. Greater EDC/NHS exposure on a lane will lead to greater protein binding.
Ligand Coupling	Protein is diluted (from stock > 0.5 mg/ml) to 50 $\mu\text{g}/\text{ml}$ in an amine free buffer with low ionic strength, and a pH below the pI of the protein. This drives the interaction between the negatively charged carboxylated dextran and positive protein. 500 RU is initially coupled using ca. 70 μl sample. (Desirable response is 500 – 1500RU). A typical chip: L1:Ligand 500 RU L2:Ligand 1000 RU L3:Analyte 500 RU (self association control) L4:Blank/Non-interacting protein (negative binding control).
Deactivation	70 μl ethanolamine. This step blocks the remaining activated carboxymethyl groups.
Initial coupling	Initial detection of an interaction using conditions of 20 $\mu\text{l}/\text{min}$ flowrate, 100 μl analyte at concentration $0.01 - 100 \times K_d$. A non binding protein is used as a negative binding control.
Regeneration	Removal of ligand from analyte. This may be achieved by a high ionic strength solution, low pH, or high pH.
Affinity	K_d may be determined through equilibrium binding analysis. A slower flow rate than for rate measurements can be used.
Kinetic measurements	k_{on} (association rate constant) and k_{off} (disassociation rate constant) do not require equilibrium and may be used to estimate K_d . A series of concentrations of analyte are flowed over the ligand. To prevent mass transport (where the rate limiting step becomes the movement of analyte to the surface rather than the binding of ligand to analyte), the flow rate is kept fast (40 $\mu\text{l}/\text{min}$, 150 μl).
Analysis	BIAevaluation 3.2RC1

Table C.i SPR protocol.