Structural studies of the *Clostridium difficile* surface layer

A thesis submitted by

Oishik Banerji

For the degree of Doctor of Philosophy

Department of Molecular Biology and Biotechnology

The University of Sheffield, UK

July 2017

Abstract

Clostridium difficile is an important nosocomial pathogen, affecting patients with compromised intestinal microbiota, causing antibiotic-associated diarrhoea and severe, sometimes fatal inflammatory complications, such as pseudomembranous colitis. *C. difficile*, like many other bacteria, possesses an outermost paracrystalline surface layer (S-layer), containing the essential cell surface protein – SlpA. Following expression, SlpA undergoes post-translational cleavage, resulting in a high molecular weight (HMW) and a low molecular weight (LMW) surface layer protein (SLP). The SLPs form a complex that self-assembles to form the S-layer. This layer is critical for pathogenicity and may be involved in gut colonisation. The two-dimensional, paracrystalline nature of this layer makes it ideal for analysis using electron crystallography.

The native structure of the S-layer, along with S-layer mutants was resolved, using high-resolution electron microscopy. Together the structures shed light on some fundamental functional aspects of this layer. This was made possible due to techniques developed to isolate and observe the intact S-layer. A cryo-EM projection structure of this layer was resolved to 7.5 Å, along with a three-dimensional structure of the *C. difficile* S-layer at 20 Å. A novel SlpA$^-$ *C. difficile* strain was used to express a truncated SlpA, lacking a major domain of the LMW SLP, which self-assembled into a structurally variant S-layer. The deleted region forms a unique, highly variable, surface-exposed motif in the native SlpA. This region along with important domains were identified and localised from this mutant, using cryo-EM projections and 3D modelling.

The S-layer is nearly ubiquitous in Bacteria and Archaea. Despite its importance very little is known about the structure of bacterial S-layers. In an inevitable future of antibiotic resistance, an understanding of such a layer might lead to development of novel therapeutics and colonisation inhibitors.

Another self-assembling paracrystalline protein system is the *Bacillus subtilis* biofilm protein BslA. The biofilm produced by *B. subtilis* has been shown to be highly hydrophobic in nature, a feature attributed to the surface exposed BslA, which forms a 'raincoat' on the biofilm. Upon integrating into the biofilm-air interface, the protein self-assembles to form a 2D paracrystalline layer. This layer was analysed using electron crystallography, to obtain a 3D model, offering a look at the process of self-assembly and its hydrophobicity.

Publications

Kirk, J.A.*, Banerji, O.* & Fagan, R.P., 2017. Characteristics of the Clostridium difficile cell envelope and its importance in therapeutics. Microbial biotechnology, 10(1), pp.76–90.

*First co-authors

Acknowledgements

This thesis is the result of enterprising endeavour, perspiration and lots of very nice coffee, thanks to our office of espresso-connoisseurs. I would like to thank my supervisors Prof. Per Bullough and Dr. Robert Fagan for guiding me through this fascinating PhD project, for the last three years. I would also like to thank Dr. Svetomir Tzokov, our EM manager, for his indispensable assistance. I also appreciate the help and suggestions from Dr. Paula Salgado (University of Newcastle, UK) and Prof. Neil Fairweather (Imperial College London, UK).

Much of my cryo-EM work, including grid making, cryo-transfers and imaging, were carried out in MRC-LMB. I am very grateful for the help of Dr. Guiseppe Cannone (MRC LMB, Cambridge, UK) and Prof. emer. Richard Henderson (MRC LMB, Cambridge, UK) during this phase.

I would like to acknowledge our collaborators Prof. Nicola Stanley-Wall (University of Dundee, UK), Prof. Cait MacPhee (University of Edinburgh, UK), and Dr. Marieke Schor (University of Edinburgh, UK) for the work I did on the BslA 2D crystals.

On a personal note, I would like to thank all my close friends, including my colleagues, Caitlin Brumby, Ainhoa Dafis-Sagarmendi and Alexander Taylor, for their encouragement and help in easing the experiences of the challenging but fulfilling doctoral studies. Finally, I would like to thank my parents for their continuous and exceptional support. This PhD is the result of their consistent faith in whatever I do.

I am grateful to the interdisciplinary Imagine: Imaging Life research initiative, of the University of Sheffield, for their funding for my project.

Abbreviations

2D	Two-dimension
3D	Three-dimension
∞	Infinity
$\times$ g	Relative centrifugal force - measured in multiples of acceleration due to Earth's gravity (g)
aa	Amino acids
ATCC	American type culture collection
bR	Bacteriorhodopsin
CCP4	Collaborative computational programming 4
CTF	Contrast transfer function
CWB	Cell wall binding
CWP	Cell wall protein
DQE	Detective quantum efficiency
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscope
$\mathcal{F}$ and $\mathcal{F}^{-1}$	Fourier transform and inverse Fourier transform
FFT	Fast Fourier transform
gDNA	Genomic DNA
h	hour(s)
H/L	HMW/LMW complex
HMW	High molecular weight
IQ	Value representing the signal-to-noise ratio of reflection amplitude given by 7/grade (Henderson et al, 1986)
kDa	Kilodalton
LMW	Low molecular weight
keV	kiloelectron volts
min	minute(s)
PCR	Polymerase chain reaction
pdb	Protein data bank
pfam	Protein family
MRC	Medical Research Council
MTF	Modulation transfer function
OD	Optical density
rms	Root mean square

s	seconds
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SLPs	S-layer proteins
SlpA	S-layer protein A
SNR	Signal-to-noise
TEM	Transmission electron microscope
Tris	Tris(hydroxymethyl)aminomethane
w/v	Weight/volume
v/v	Volume/volume

Contents

Table of tables

Table of figures

Chapter 1 - Introduction

Imaging as a science has seen tremendous improvements in recent times. Interactions between macromolecular complexes are the basis for all biochemical processes sustaining life. The study of the structure, behaviour, and interactions among the different components of a macromolecular complex has been made possible using electron microscopy (EM). The high resolving power of EM is attributed to the fact that the wavelength of electrons, accelerated to 300 keV in a transmission electron microscope (TEM), is less than 0.02 Å. The electrons interact more strongly with matter than X-rays, making it possible to image thin biological material and small molecules.

In addition to observing structures directly visible under TEM, Fourier optics allows the use of high-resolution images of thin crystals to generate structural information. Its potential was revealed when the first structure of a membrane protein was resolved to 3.5 Å using cryo-electron crystallography (Henderson et al, 1986). The last decade has seen EM emerge as an unrivalled tool, allowing both structure determination at atomic resolution and the imaging of molecular complexes at the moment of biochemical reactions (Schmeing & Ramakrishnan, 2009; Radics et al, 2014; Bartesaghi et al, 2015; Zhu et al, 2016).

This thesis exploits the developments in EM techniques and expands on the structure of the two-dimensional paracrystalline surface layer (S-layer) of *Clostridium difficile.* The human pathogen, *C. difficile*, is a major healthcare concern, causing several major outbreaks in recent years (Rupnik et al, 2009). The S-layer covers the entirety of the vegetative cell and is essential for the organism, forming the basis for its interaction with the environment and the host. Little is known about the structure of the S-layer. However, the crystalline order of the layer makes it ideal for structural and functional analysis using electron crystallography (EC). The following sections provide an introduction to the organism, the S-layer and details on the techniques used.

1.1. Clostridium difficile

A Gram-positive, spore-forming, anaerobic bacterium, *Clostridium difficile*, is the leading cause of antibiotic-associated, hospital-acquired diarrhoea. Symptoms range from mild diarrhoea to life-threatening inflammation of the bowel. Complications from *C. difficile* infection (CDI) can include pseudomembranous colitis, toxic megacolon, perforation of the colon, and sepsis (Smits et al, 2016). Antibiotic use, although life saving, is also the main risk factor in acquiring CDI. Nosocomial patients are frequently treated with broad-spectrum antibiotics, thereby resulting in catastrophic damage to their gut microbiota (Dethlefsen et al, 2008). This can lead to colonisation and proliferation of *C. difficile* in the gut (Lawley et al, 2009). In the elderly population the 30-day mortality exceeds 30% (McGowan et al, 2011).

C. difficile infection is triggered by gut dysbiosis. The process of infection starts with germination of indigenous or ingested spores, triggering the colonisation phase. This phase may require the attachment of the bacterium to the host intestinal epithelium. This is followed by a virulence phase, where toxins are released into the lumen of the gut, resulting in onset of disease (Smits et al, 2016). Diarrhoea, triggered by the release of toxins, causes the discharge of spores to the environment. Most clinical isolates produce two related toxins, TcdA and TcdB (Jank & Aktories, 2008).

Two hyper-virulent strains (PCR ribotypes 027 and 078) have been identified in previous years, linked to several global outbreaks, with poor clinical outcomes (McDonald et al, 2005; Goorhuis et al, 2008). These strains produce the *C. difficile* binary toxin, CDT, in addition to TcdA and TcdB toxins. The emergence of these epidemic strains since 2001 and improvements in genetic tools has resulted in enhanced interest and understanding of this bacterium (Ng et al, 2013; Dembek et al, 2015). The crucial host-pathogen interaction is mediated by the cell surface components of *C. difficile*, the understanding of which is still in its infancy (Kirk et al, 2017a). The external most of these surface components is the essential paracrystalline S-layer. In this project we look in detail at the S-layer of a hypervirulent, toxin $A^{+}/B^{+}/CDT^{+}$ strain of ribotype 027, identifying key structural components via mutational studies. In addition, we resolve the structure of the S-layer in a toxin A^{-}/B^{+} ribotype 017 strain, highlighting notable similarities and differences in significant regions.

1.2. Surface layers

Surface layers or S-layers are a monolayer formed of proteins or glycoproteins, known as SLPs, arranged in a two-dimensional paracrystalline array coating the entire outer surface of a cell. They are highly prevalent in archaea and bacteria (Sara & Sleytr, 2000). S-layers were reported as early as 1952 in *Spirillum* sp. (Houwink & Le Poole, 1952; Houwink, 1953), but their absence in model bacterial organisms such as *Staphylococcus aureus*, *Escherichia coli* and *Bacillus subtilis*, meant that although explored in depth among archaea (Sara & Sleytr, 2000), their role in the disease processes of infection, immunity and resistance has been largely under-studied.

A typical *Clostridium difficile* cell has a surface area of 18.85 μm^2 (assuming it is a perfect capsule). As this thesis will demonstrate, the *C. difficile* S-layer consists of two S-layer proteins within a unit cell of 64 nm^2. Therefore, there are approximately 590,000 S-layer subunits on the cell surface. This necessitates synthesis, export and assembly of 164 subunits per second, per cell during exponential growth. Such a large production volume makes SLPs among the most physically abundant and metabolically expensive proteins in those cells. In fact, SLPs are estimated to be the most abundant biopolymers on Earth, accounting for around 10% of the biomass of all prokaryotic organisms (which account for two-thirds of the total biomass on Earth) (Sara & Sleytr, 2000).

1.2.1. Binding to the cell envelope

The S-layers are attached to different external structures, based on the organism's cell envelope. In archaea, the S-layers are linked via a hydrophobic anchor directly to the lipid bilayer of the cell membrane (Sleytr & Beveridge, 1999). In Gram-negative bacteria, the often glycosylated SLPs attach to the underlying lipopolysaccharide via ionic interactions (Bingle et al, 1997). Many Gram-positive bacteria possess secondary cell wall polymers (SCWPs), which act as anchors for the SLPs via ionic interactions (Sara & Sleytr, 2000).

Previous work done in *Bacillus anthracis* (Kern & Schneewind, 2008) identified three tandem surface layer homology (SLH) motifs, which were subsequently used to predict 24 *B. anthracis* cell wall proteins. Similar non-covalent attachment domains (NCAD) were found in *Lactobacilli* (Boot et al, 1995). In *C. difficile*, three tandem cell wall binding 2 (CWB2) domains anchor the S-layer protein to the cell wall underneath (Wright et al, 2005). The CWB2 repeats bind the SLPs to secondary cell-wall polysaccharide, PS-II (Willing et al, 2015). CWB2-containing surface proteins have also been identified in *Clostridium botulinum* and *Clostridium tetani* (Sebaihia et al, 2007; Bruggemann et al, 2003).

1.2.2. Overall Structure

SLPs undergo self-assembly to form a regular array, showing diverse lattice symmetries, such as oblique (*p*1, *p*2), square (*p*4) or hexagonal (*p*6), with unit cell dimensions in the range of 2.5 to 35 nm, as revealed by freeze-etching studies (Sara & Sleytr, 2000; Sleytr & Glauert, 1976; Sleytr & Beveridge, 1999). Each structural unit forming the lattice may consist of one or more identical SLPs. Hexagonal symmetry is predominant in archaea. The archaeal S-layers exhibit a thickness of up to 70 nm, whereas bacterial S-layers are generally 5 – 20 nm thick showing more diversity in symmetry (Sleytr et al, 2007).

Interestingly, the lattice of certain S-layers can alter under varying environmental conditions, e.g. *Geobacillus stearothermophilus* NRS 2004/3a strain can change from a wild type oblique (*p*2) lattice to a square (*p*4) array, with oxygen stress; while strain PV72 change to a *p*2 lattice from *p*6 under oxygen stress (Sleytr & Beveridge, 1999).

SLPs display an anchoring domain, and some show the presence of a crystallisation domain where SLP-SLP interactions take place. Crystallisation domains have been previously identified in *G. stearothermophilus* S-layer protein SbsB, where a 3D crystal structure revealed atomic contact points among 6 SbsB SLPs forming a p6 hexagonal lattice, coordinated by Ca^{2+} ions (Baranova et al, 2012). Similar Ca^{2+} ion-mediated crystallisation has been reported in *Caulobacter crescentus* S-layer (Bharat et al, 2017). Although, crystal structures of truncated *G. stearothermophilus* S-layer proteins (Baranova et al, 2012; Pavkov et al, 2008) and truncated regions of *C. crescentus* S-layer protein (Bharat et al, 2017) have been reported, native structure of a complete S-layer is yet to be published.

1.2.3. Function

The functions of S-layers are poorly understood, but their importance to the organism is evident in the fact that natural selection has retained such a metabolically expensive structure. These structures are poorly conserved across organisms and even within a single species (Fagan & Fairweather, 2014), perhaps reflecting their role in evolution against pathogens, and in case of infectious organisms, the host defences.

Various functions have been attributed to the S-layer, with varying evidence, including maintenance of cell shape, protection of cells, virulence factors and adhesion (Wang et al, 2000). The S-layer in archaea is the only cell envelope structure outside the plasma membrane. This layer is responsible for mechanical stability, integrity of the cell shape and cell division (Wildhaber & Baumeister, 1987; Messner & Sleytr, 1992). In *C. difficile*, the S-layer proteins has been implicated in binding to host intestinal cells (Calabi et al, 2002; Merrigan et al, 2013), in addition to providing resistance against bacteriocins, lysozyme and antimicrobial peptides (Kirk et al, 2017b). The S-layer also has a role in acting as a molecular sieve, providing protection against environmental changes, bacteriophages and bacteriolytic enzymes (Sara & Sleytr, 2000). An absence of the *C. difficile* S-layer protein SlpA results in morphological changes in cell shape and size (Joseph Kirk, personal communication). Similar findings have been observed here and are discussed further in Chapter 4.

1.3. *C. difficile* S-layer

In some isolates of *C. difficile*, electron microscopy analysis of the cell surface showed presence of an S-layer with square-ordered lattice (Kawata et al, 1984; Cerquetti et al, 2000). Low pH glycine buffer solubilises proteins in this layer, revealing two distinct proteins that appear to constitute the S-layer (Takeoka et al, 1991; Cerquetti et al, 2000). The molecular weight of the two proteins varied among the different strains, with high-molecular weight (HMW) SLP of 42 – 50 kDa and a low molecular weight (LMW) SLP of 32 – 38 kDa. The LMW SLP, which shows remarkable sequence diversity (Ni Eidhin et al, 2008; Calabi & Fairweather, 2002), was found to be immunodominant in patients with antibiotic-associated diarrhoea (Pantosti et al, 1989). Also, rabbit anti-sera raised against *C. difficile* were found to bind to the LMW SLP (Cerquetti et al, 2000). These raise the possibility of it being surface exposed.

The two SLPs were found to be post-translation cleavage products of a single pre-protein SlpA (Calabi et al, 2001; Calabi & Fairweather, 2002). SlpA appears to be essential, as evidenced from an inability to generate gene knockouts and transposon-mediated insertional mutants in the *slpA* region (Dembek et al, 2012). SlpA shows three distinct domains (Figure 1.1) – (1) a signal peptide, (2) the highly variable LMW region, and (3) the HMW region containing three tandem cell wall binding-2 repeats, which anchors it to the cell wall (Fagan et al, 2009). The inter-strain sequence variability seen among the LMW SLPs is, however, absent in the HMW SLP, which shows high sequence identity among strains. *slpA* is strongly transcribed during the entire growth phase of *C. difficile* (Savariau-Lacomme et al, 2003). The secretion of the SlpA pre-protein involves the accessory Sec secretion system (Fagan & Fairweather, 2011). The signal peptide region directs its translocation across the cell membrane. The cleavage of the pre-protein into the mature form is mediated by Cwp84, another cell wall protein (Kirby et al, 2009; Dang et al, 2010).

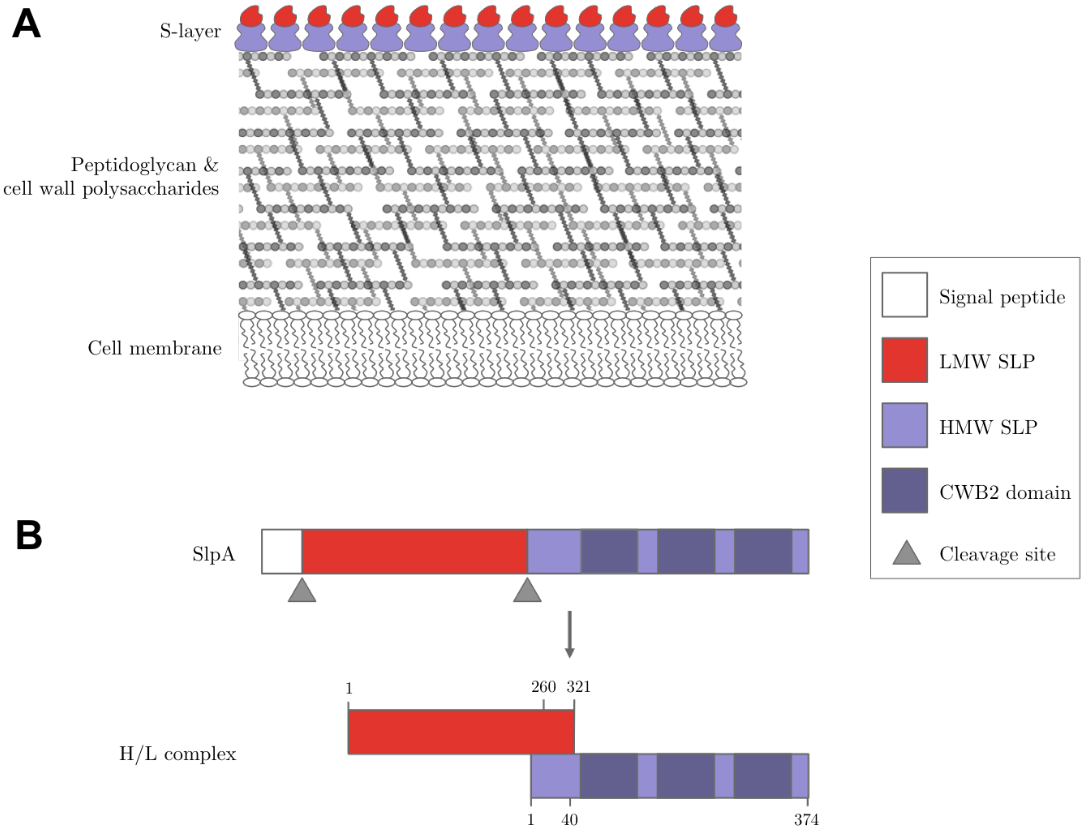


Figure 1.1 The cell envelope of *C. difficile*

A: The cell envelope with the outermost S-layer, shown as blue and red structures, with repeating units of H/L complex, below which is the cell wall containing peptidoglycan and secondary cell wall polysaccharides and cell membrane. **B**: A cartoon representation of the S-layer precursor protein SlpA and the mature H/L complex protein without the signal peptide. Arrows indicate the two cleavage sites. The CWB2 (pfam PF04122) motifs are shown as darkers boxes. The H/L interaction region is also indicated in numbers, as identified in strain 630 (Fagan et al, 2009).

C. difficile S-layers are not known to be glycosylated (Qazi et al, 2009). Although, a strain of *C. difficile* have been shown to contain a glycosylation locus in the S-layer cluster. In addition this strain contains a distinct S-layer cassette type (see below) producing SlpA of a much shorter LMW domain (Dingle et al, 2013). Some work on the projection structure of the S-layer of the strain Ox247 has been done as part of this thesis (Appendix 7). Structure of the S-layer of a derived mutant, Ox247Δorf2, containing an insertional mutation in glycosylation cluster (orf2 in the putative glycosylation locus) (Prof. Neil Fairweather, personal communication), has also been described in this thesis (Appendix 7).

1.3.1. The H/L complex

The two SLPs have been found to form a stable HMW and LMW (H/L) complex (Fagan et al, 2009). Conserved regions of the HMW SLP (N-terminal residues 1 – 40) and LMW SLP (C-terminal residues 260 – 321) are essential for the formation of the H/L complex (Fagan et al, 2009) (Figure 1.1). Also, a small-angle X-ray scattering (SAXS) model and a crystal structure of a truncated LMW SLP, reported in the same paper, revealed an extended conformation of the H/L complex, where the HMW and LMW SLPs interact non-covalently along the N- and C-terminal regions respectively (Fagan et al, 2009). The structure of the H/L complex in *C. difficile* and how it organises and self-assembles on the cell wall is yet to be reported.

1.3.2. Cell wall binding (CWB2) region

In *C. difficile*, three tandem cell wall binding-2 (CWB2) domains (pfam 04122) were identified, which anchor the cell wall protein to the cell wall underneath (Willing et al, 2015). The CWB2 motif was first identified in CwlB – an autolysin that cleaves the peptidoglycan in the cell wall of *B. subtilis* (Kuroda & Sekiguchi, 1990; Kuroda et al, 1992). Although, *B. subtilis* lacks an S-layer, the CWB2 motif is required for the retention of CwlB on the cell wall. In *C. difficile* the three CWB2 domains were identified in SlpA and 28 other proteins, termed as cell wall proteins (CWPs)

(Fagan et al, 2011). Based on sequence-derived putative functions, these proteins include autolysins, proteases and adhesins (Figure 1.2).

The three tandem SLH repeats, in *B. anthracis*, have been shown to assume a pseudo-trimeric arrangement of alpha helices forming a three-pronged spindle in its crystal structure, which bind non-covalently to secondary cell wall polysaccharides (Kern et al, 2011). In *C. difficile*, Cwp6 and Cwp8 from strain 630 were crystallised and their structure was reported (Usenik et al, 2017). Cwp8 and Cwp6 contain CWB2 pseudo-trimers in the C-terminal and middle of the polypeptide, respectively. The trimers form a disc-shaped, triangular trimer of Rossmann-folds (Figure 1.3), with a diameter of 60 Å and thickness of 30Å. Each of the three domains is formed of two α-helices, flanking four-stranded parallel β-sheet. Three α-helices, one from each domain, form a helical bundle in the middle. The structures of the SLH repeats in *B. anthracis* and the *C. difficile* CWB2 domains show remarkable similarity (structures compared in Appendix 1).

The teichoic acid-like secondary cell wall polysaccharide, PS-II, is known to be a ligand of CWB2 (Ganeshapillai et al, 2008; Willing et al, 2015). Docking experiments revealed interactions between the conserved regions of CWB2 (Usenik et al, 2017) and PS-II, indicating this region is bound to the cell wall.

1.3.3. LMW SLP

The crystal structure of the truncated region of the LMW SLP (residues 1 – 262), lacking 59 C-terminal residues, was reported (Fagan et al, 2009) (Figure 1.4). The structure shows two distinct domains, namely domain 1 and 2. Domain 1 contains residues that are conserved between various *C. difficile* strains. A loop-rich domain 2, which shows high sequence variability among different strains, follows this domain, and is most likely exposed to the surface. The high content of loops might allow sequence variability while retaining the overall domain fold. The domain 1 crystal structure (residues 1 – 88 and 239 – 249) shows a two-layer sandwich fold (Figure 1.4).

Figure 1.2 ***C. difficile*** **cell surface protein family**

Family of 29 proteins have been identified as *C. difficile* cell wall anchoring proteins. All these cell wall proteins (CWPs) contain the three tandem CWB2 domains (either C- or N-terminal or middle of the protein). The proteins are arranged with N-terminal from left to right, with signal sequence shown as a black box. Some domains have known functions, indicated in the box on the right. The proteins are to scale with scale bar indicating 200 amino acids. Figure reproduced from (Fagan & Fairweather, 2014), with permission from the Nature Publishing Group.

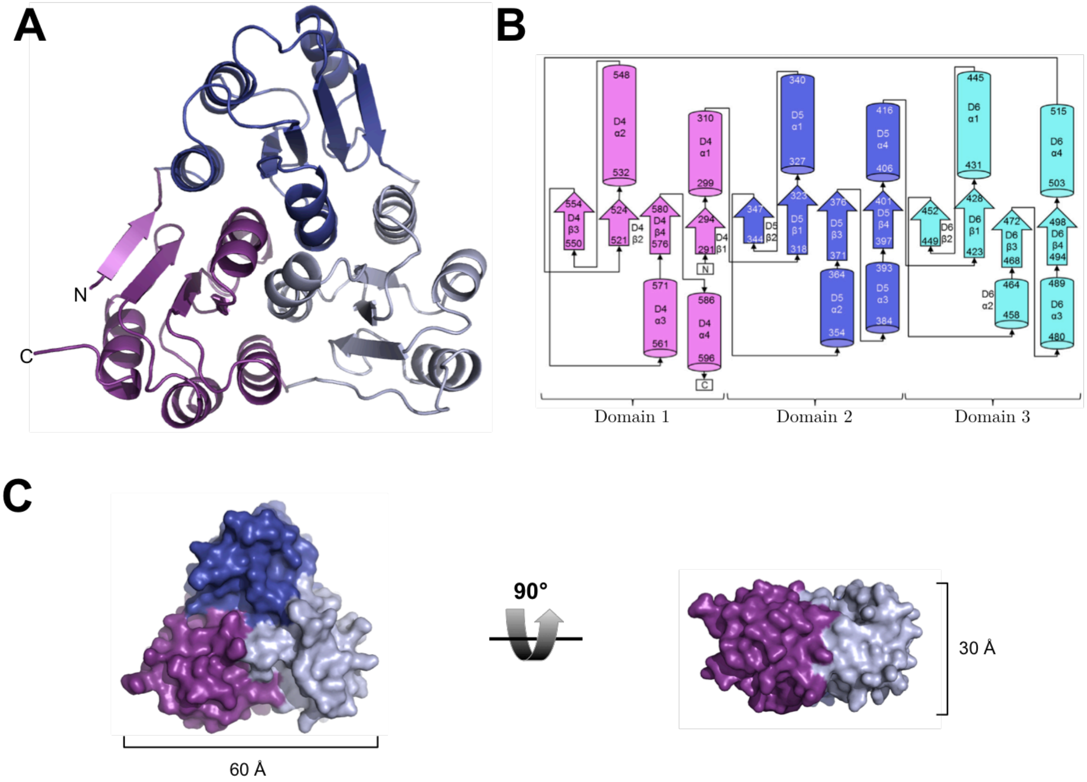


Figure 1.3 Crystal structure of the CWB2 motifs

A: Ribbon representation of the three CWB2 domains from the crystal structure of Cwp8 in *C. difficile* strain 630 (pdb code 5j6q) (Usenik et al, 2017). The three domains of Cwp8 are highlighted in shades of blue and purple, showing the pseudo-trimer of Rossmann folds in the structure. **B**: The topographical map of the structure. **C**: The surface representation of the same structure as in **A**, along with a side view, indicating the width and height of the structure. Figure adapted from (Usenik et al, 2017), with permission from Elsevier Ltd.

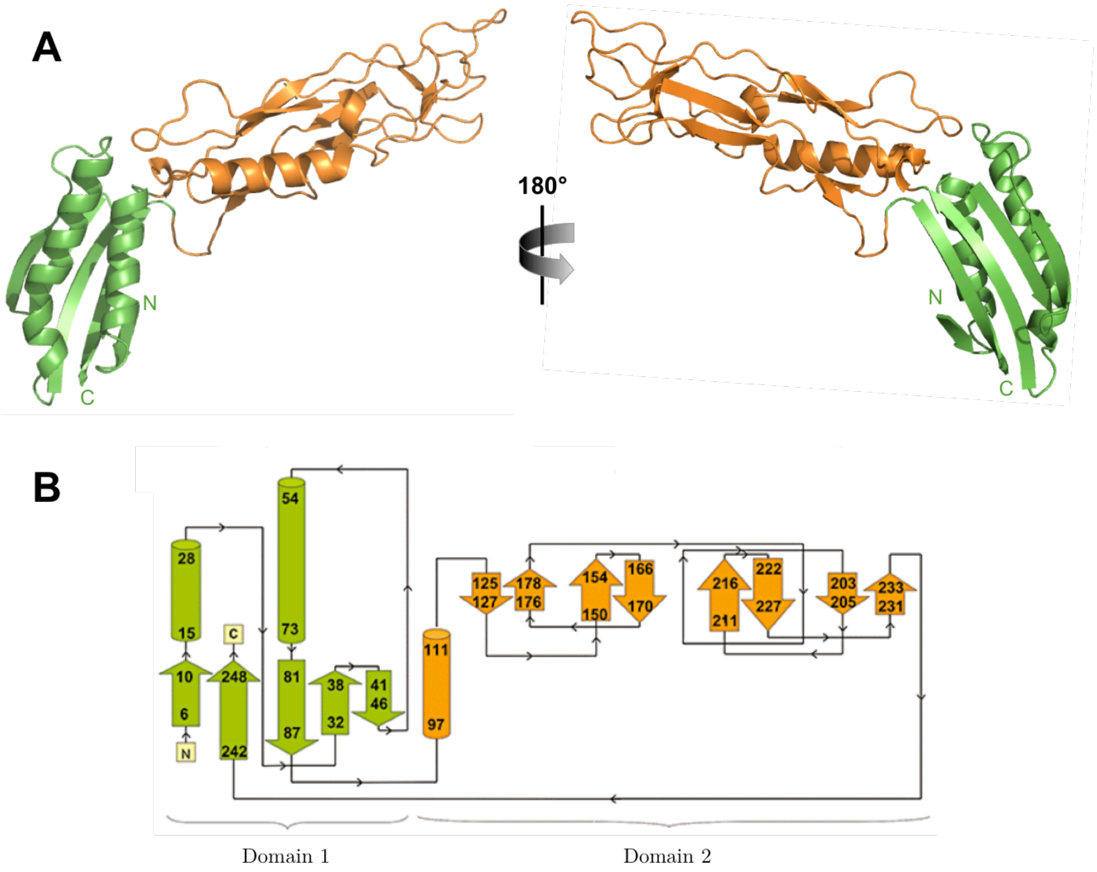


Figure 1.4 Crystal structure of LMW SLP$_{1\text{–}262}$

A: Two views of the truncated LMW SLP$_{1\text{-}262}$ from *C. difficile* strain 630 (pdb code 3cvz) (Fagan et al, 2009). The conserved domain 1 is in green. The hypervariable domain 2, showing extensive loops in the structure, is highlighted in orange. The 263 – 321 residues were not resolved in the structure and extend from domain 1. Residues 260 – 321 form the interaction region with the HMW SLP. **B**: The topological diagram of the two domains in the structure shown in **A**. Domain 1 contains both the C- and N-termini forming two-layer sandwich, while domain 2 has a novel fold. The structure in **A** was generated using Mac Pymol software from pdb coordinates. **B** is from (Fagan et al, 2009) with permission from John Wiley & Sons, Inc.

ELISA binding assays indicated 62 C-terminal residues (260 – 321) were essential for the binding with the HMW SLP, most of which are not in the crystal structure (Fagan et al, 2009). This domain also contains the majority of conserved region among the strains, and located towards the interior of the protein, most likely contributing to the stability of the domains. A sequence comparison of the LMW and HMW SLP, among 12 different cassettes types (see below) of *C. difficile* is shown in Appendix 2.

1.3.4. S-layer gene cluster

A ~37 kb cassette was revealed by sequencing of several *C. difficile* clinical strains, containing 17 genes, 11 of which encode CWB2 motifs (Dingle et al, 2013). This gene cluster encodes SlpA*,* ten other CWPs, and SecA2 – the essential SlpA translocase (Dingle et al, 2013) (Figure 1.5). Several of these CWPs have been biochemically characterized.

Cwp66, a heat-shock inducible adhesin, mediates adherence to epithelial cells *in vitro*, and might be involved in binding to the host gut (Calabi et al, 2002; Waligora et al, 2001). Cwp84 is a cysteine protease and cleaves the SlpA pre-protein into HMW and LMW SLPs (Kirby et al, 2009). CwpV, not present in the cassette, has been shown to promote auto-aggregation of cells in solid and liquid media (Reynolds et al, 2011), shows antigenic variability, and provide protection against some bacteriophages (Sekulovic et al, 2015). The other CWPs are less well characterised, however, functional sequence similarities indicates various enzymatic functions – ranging from β-lactamase, α-amylase, cellulase, glucosaminidase, etc. (Fagan & Fairweather, 2014) (Figure 1.2).

Interestingly, 12 variants of this S-layer cluster have been identified in *C. difficile*, referred to as S-layer cassette types 1 – 12 (Dingle et al, 2013). These cassettes resemble polysaccharide capsular switching cluster or S-layer glycosylation gene cluster seen in other bacterial species – causing antigenic shifts, while the rest of the genome remains unchanged (Ristl et al, 2011).

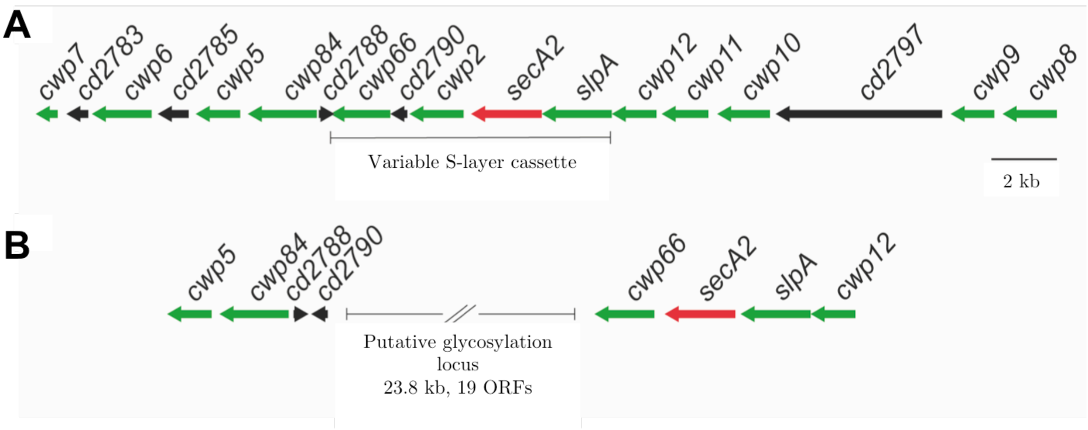


Figure 1.5 The S-layer gene cluster with the variable S-layer cassette

A: Eleven of the 28 CWPs are encoded in a cluster along with the gene for the SlpA translocase, SecA2 (red arrow). The 12 genes, which include *slpA*, are highlighted as green arrows, along with five unrelated genes. There is a glycan synthesis cluster downstream from *cwp7*, implicated in the synthesis of secondary cell wall polysaccharide PS-II (Willing et al, 2015) responsible for the binding of CWB2 domain. The variable cassette region is highlighted. **B**: One of the 12 known cassettes (cassette 11), which has a 23.8 kb insertion that includes 19 ORFs with predicated functions of glycan synthesis and transfer. Figure taken from (Kirk et al, 2017a). Note *cwp2* is absent in **B** along with rearrangement of the two flanking ORFs, *cd2790* and *cwp66*.

Selective pressure to avoid bacteriophage and host immune responses may have contributed to this diversity. Also, one cassette was found to contain the putative glycosylation cluster (Dingle et al, 2013).

S-layer structures described within this thesis are from several toxin positive and clinically significant strains (Table 2.2). Strain R7404, which is toxin A negative and toxin B positive, belongs to PCR ribotype 017 with S-layer cassette type 7. The hypervirulent strain, R20291, produces toxin A, toxin B and CDT, and is one of two epidemic strains from the 2004 Stoke Mandeville Hospital, Buckinghamshire, UK outbreak (Stabler et al, 2009). The S-layer structure of R20291 along with derived mutants is described later on. R20291 belongs to PCR ribotype 027 and S-layer cassette type 4. The glycosylated S-layer containing strain Ox247 and derived mutant Ox247Δorf2 belong to PCR ribotype 005 and cassette type 11. Ox247 is unique in producing a glycosylated S-layer along with a much smaller LMW SLP (Prof. Neil Fairweather, personal communication). The glycosylation locus is part of the S-layer gene cluster (Figure 1.5).

1.4. High resolution electron microscopy

Images observed in transmission electron microscopy (TEM) are two-dimensional projections of a three-dimensional biological object. There are various requirements that need to be considered before the structure of an object can be interpreted from its projection: (1) the object itself must be thin, (2) the electron dose must be optimal to reduce damage to biological samples, (3) the specimen preparation must be done carefully, to avoid artefacts, (4) images need to have sufficient contrast usually obtained by under-focusing to achieve phase contrast, (5) the effects of defocusing and objective lens aberrations need to be corrected carefully, (6) several images of the object need to be averaged to minimise the limitations imposed by radiation damage, and finally, (7) the projections need to be brought together to reconstruct a 3D model. The following sections give a brief description of these issues. Excellent overviews of the process can be found in (Amos et al, 1982; Henderson et al, 1990; Rosenthal & Henderson, 2003; Glaeser, 2007; Orlova & Saibil, 2011; Rodriguez & Gonen, 2016; Glaeser, 2016).

1.4.1. Kinematic scattering

Electrons incident on a sample, placed in a TEM, interact directly with the potential energy of the atomic nuclei in the sample. The electrons then scatter either elastically or inelastically, resulting in different outcomes.

1.1.1.1. Elastic scattering

In elastic scattering the kinetic energy of the electron is conserved while the direction of propagation of the particle changes. Elastic scattering increases with increasing atomic number as a function of the *Coulomb potential.* Electrons with large scattering angles get blocked as they hit the inside of the microscope column or one of the apertures. The high scattering angle may produce some contrast at low or medium resolution as the electrons passing through high-scattering regions of the specimen get blocked from reaching the detector, producing *amplitude contrast.* At high resolution, *phase contrast*, however, is the predominant means of image formation. All elastically scattered electrons retain their energy, but suffer an approximate phase-shift of $\pi/2$ (Amos et al, 1982). The electron beam passing through a periodic specimen, such as a 2D crystal, obeys *Bragg's law* and produces a diffraction pattern in the back-focal plane of the objective lens.

1.1.1.2. Inelastic scattering

Electrons may interact directly with the orbital electrons in the specimen. This interaction results in a transfer of around 25 eV of energy to the specimen. The inelastically scattered beams are poorly focussed and contribute to an incoherent background to the image, which can reduce the signal-to-noise ratio (SNR) in the image. The ratio of elastic (ε_e) to inelastically scattered (ε_i) electrons is dependent on the atomic number (Z) and is given by

$$\frac{\varepsilon_e}{\varepsilon_i} = \frac{Z}{19} \qquad 1.1$$

(Crewe et al, 1970). Hence, biological specimens, with small Z numbers, predominantly suffer inelastic scattering, causing extensive sample damage. The damage can be greatly reduced by cooling the sample to cryogenic temperatures and imaging with very low electron dose. Averaging several low-dose images of multiple identical molecules, either as single particles or arranged in a 2D crystal, can generate an image, overcoming this barrier.

Inelastic electrons can also be removed from the image, before the electrons reach the detector, by using an energy filter such a GIF Quantum Energy filter (Gatan).

1.1.2. The weak phase object approximation

The incident electrons are a wave function that is affected directly by the object. In case of elastic scattering, the object wave following such an interaction can be represented by $\psi_o(x,y)$,

$$\psi_o(x,y) = e^{-i\phi(x,y)} \qquad 1.2$$

where $\phi(x,y)$ is the phase-shift relative to incident wave, in radians, in the specimen plane x,y. The 2D projection of the sample's potential distribution in 3D, $V(x,y,z)$, can be considered to be an integration of $V(x,y,z)$ over the sample thickness t.

$$\phi(x,y) = \int_0^t V(x,y,z)\, dz \qquad 1.3$$

In the case of a thin sample, the incident electron is not expected to undergo more than one elastic collision, thereby the phase-shift $\phi(x,y)$ is small enough so that $e^{-i\phi(x,y)}$ in Equation 1.2 can be replaced with the linear relationship $1-i\phi(x,y)$, neglecting higher order terms of ϕ^2, ϕ^3, etc., from the expansion of the exponential term.

$$\psi_{image}(x,y) \approx 1 - i\sigma\phi(x,y) \qquad 1.4$$

[Here σ is a constant $= \frac{2\pi me\lambda}{h^2} = \pi/\lambda E_0$ for electrons of wavelength λ, Energy E_0, mass m and charge e.]

This linear relationship in Equation 1.4 allows the direct interpretation of the object wave ψ_o from the image wave ψ_{image}. This is possible due to the object being thin, resulting in a weak phase shift, and is termed as the *weak phase object approximation*. Fortunately, most biological crystals are weak phase objects and allow direct interpretation of object density from the image.

1.1.3. Fourier optics and diffraction

The image formation by the objective lens can be assumed to occur in two stages (Figure 1.6). Firstly, the lens produces a diffraction pattern at the back focal plane after it collects the radiation that has been scattered from the object. The unscattered electrons, which are parallel to the optical axis, will focus on to a point in the centre of the diffraction pattern. In the second step, the diffracted waves at the back focal plane recombine to form the image, which is captured by the detector.

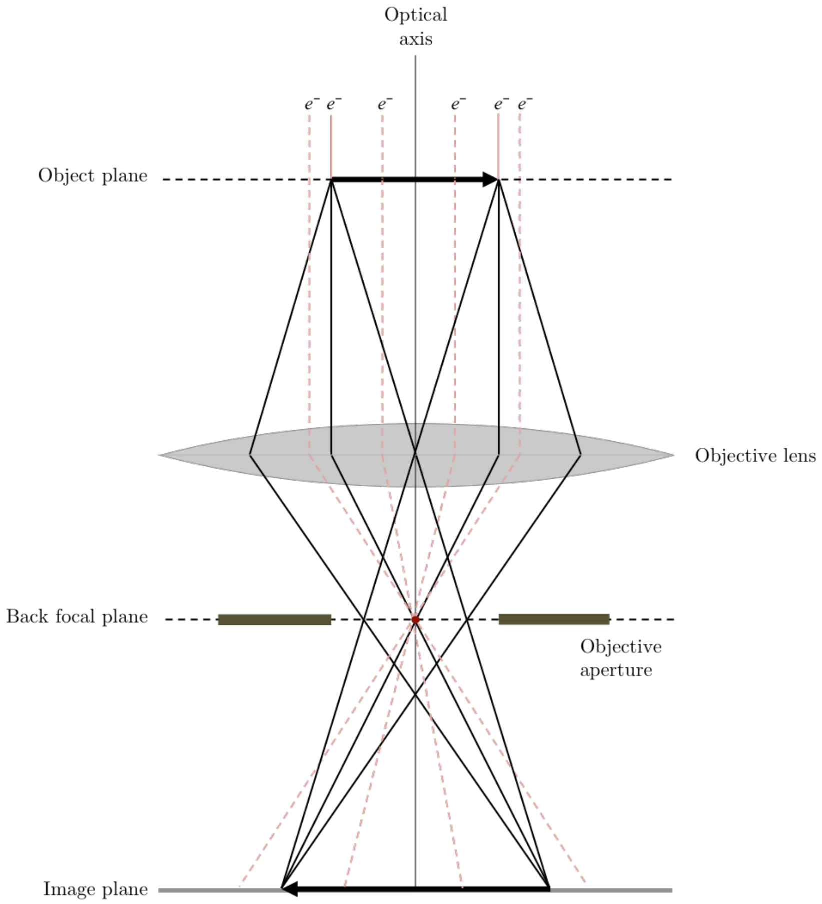


Figure 1.6 Schematic showing two stages of image formation by the objective lens

The image formation of an object (black arrow) by the objective lens occurs in two stages. The unscattered beam (dashed red lines) passes through the centre (red circle) of the diffraction plane, also known as the back focal plane. The electrons that do scatter (black lines from the object) interfere and produce diffraction pattern on the back-focal plane. The image formation takes place when the diffracted beams recombine at the image plane and interfere with the unscattered electrons, producing amplitude and phase contrast. The column itself, the edge of the objective lens, or the objective aperture prevents the high-angle scattering electrons from contributing to the image. This results in more amplitude contrast. The objective aperture is shown at the back focal plane. Weak-phase objects, which include most bio-molecules, do not scatter at such high angles, only producing phase contrast.

Mathematically, this represents two Fourier transformations, one resulting in a diffraction pattern, while the next is the Fourier synthesis or *inverse transform*, resulting in reformation of the image, but this time with modifications made by the objective lens.

The limited aperture of the objective lens, when using the objective aperture, does not allow the electrons from high angle scattering to enter the lens and thus are excluded, reducing the resolution.

However, further factors also limit the resolution, namely the lens aberrations and defocus. These cause a *phase distortion*, γ, causing the diffracted wave to be multiplied by $e^{i\gamma}$ before the Fourier synthesis is performed. As described later (Section 1.1.4 below), this has a significant effect on the image formation.

The electron diffraction pattern in the back focal plane of the objective lens can be written as a Fourier transform of Equation 1.4,

$$\psi_d(u,v) = \delta(0,0) - i\sigma\boldsymbol{\mathcal{F}}[\phi(x,y)] = \delta(0,0) - i\sigma\mathbf{F}(\mathbf{s}) \qquad 1.5$$

ψ_d is the wave function at the diffraction plane, which is the Fourier transform of ψ_o. $\mathbf{F}(\mathbf{s})$ is the Fourier transform of $\phi(x,y)$, where $\mathbf{s}$ is the two-dimensional spatial frequency vector in Fourier space. $\delta(0,0)$ represents the direct unscattered beam of electrons at zero angle of scattering, located at the origin of the transform. (u,v) are the reciprocal space coordinates.

For a 2D crystal, the transform is only sampled at reciprocal lattice points (h,k), which are the Miller indices of the Bragg plane. The indices are related to the real-space unit cell length of the crystal, a and b, by

$$\frac{h^2}{a^2}+\frac{k^2}{b^2}=u^2+v^2 \qquad 1.6$$

The diffraction *intensities* are given by the square of the amplitude of the diffraction spots. In this case $|\psi_d|^2=\psi_d\,\psi_d^*$, where ψ_d^* is the complex conjugate of ψ_d. The intensity is hence a real number.

Taking into consideration the objective lens aberrations, such as the limiting aperture and the phase distortion γ, one can rewrite Equation 1.5 as follows,

$$\psi_d(u,v) = [\delta(0,0) - i\sigma\mathbf{F}(\mathbf{s})] \cdot A(\mathbf{s}) \cdot e^{i\,\gamma(\mathbf{s})} \qquad 1.7$$

where A is the aperture function, a mathematical representation of the aperture of the objective lens.

The next step in image formation is the inverse transform of Equation 1.7

$$\psi_{image}(x,y) = 1 - \mathrm{i}\sigma\ \phi(x,y) * \mathcal{F}^{-1}[A(s) \cdot e^{i\,\gamma(s)}] \qquad 1.8$$

The $*$ represents the *convolution* product, since the Fourier transform of a real-space multiplication is a convolution product in Fourier space.

The image intensity, I_{image}, in turn is $\left|\psi_{image}\right|^2 = \psi_{image}\ \psi^*_{image}$, which after ignoring the quadratic terms (since it is a weak phase object) and replacing $e^{i\,\gamma(\mathbf{s})}$ with *Euler relationship* ($e^{i\,\gamma(\mathbf{s})} = \cos\gamma(\mathbf{s}) + i\sin\gamma(\mathbf{s})$), Equation 1.8 becomes

$$I_{image} = 1 + 2\sigma\ \phi(x,y) * \mathcal{F}^{-1}[A(\mathbf{s}) \cdot \sin\gamma(\mathbf{s})\,] \qquad 1.9$$

The significance of this equation is clear when a Fourier transform of a micrograph is performed, as follows

$$\mathcal{F}^{-1}[I_{image}] = \delta(0,0) + 2\sigma\ \mathbf{F}(\mathbf{s}) \cdot A(\mathbf{s}) \cdot \sin\gamma(\mathbf{s})\,] \qquad 1.10$$

where $\delta(0,0)$ represents the direct unscattered electron beam producing a central peak in the transform (indexed $u = 0$, $v = 0$).

During subsequent image processing, the aperture function, $A(\mathbf{s})$, and the *sine* of the phase distortion, $\sin\gamma(\mathbf{s})$, modulate the outcome. *Hence, a computer generated Fourier transform of a micrograph does not generate the diffraction pattern from the back focal plane of the objective lens. The effects of the objective lens itself modify the image and hence the transform.* For the sake of simplicity from here onwards, the peaks generated in the Fourier transform of the image will be referred to as diffraction spots (or the crystallographic term, *reflections*).

1.1.4. Phase Contrast and the CTF

From Equation 1.10 one can say that *the transform of an image of a thin specimen is the product of the transform of the specimen structure and the microscope's transfer function.* The aperture function, $A(\mathbf{s})$, and the phase distortion, $\gamma(\mathbf{s})$, are components of the microscope's transfer function, independent of the specimen structure. Whether the specimen is introduced in the beam or not, the point spread function (PSF) of the microscope will be modulated by the above components. The effect of the modulation, however, is felt in the final image only when the specimen causes phase distortions.

The function $\sin\gamma(\mathbf{s})$, also called the contrast transfer function (CTF), depends on the lens' spherical aberration coefficient C_s, the defocus Δf, and the axial astigmatism. Equation 1.11 shows the dependence of $\gamma(\mathbf{s})$ on the lens aberrations. Depending on the values of C_s and Δf, the $\sin\gamma(\mathbf{s})$ alters the image transform, causing rapid oscillations between positive and negative values (Figure 1.7). As the function oscillates, a negative value of $\sin\gamma(\mathbf{s})$ indicates a complete reversal of contrast for the respective spatial frequencies in the image. Even for a positive number, the contrast will be less than unity at most spatial frequencies and the corresponding Fourier coefficients $\mathbf{F}(\mathbf{s})$will have attenuated values (Figure 1.7).

$$\gamma(\mathbf{s})=2\pi\left\{\frac{C_s}{4}\lambda^3\,\nu^4-\frac{\Delta f}{2}\lambda\,\nu^2\right\} \qquad 1.11$$

where ν stands for spatial frequency, λ is the wavelength of the electrons. An additional term for axial astigmatism has been ignored for simplicity.

The need to defocus becomes clear when one looks at the CTF *at focus*, i.e. $\Delta f = 0$ (Figure 1.7). The values of $\sin\gamma(\mathbf{s})$ at lower spatial frequencies are near zero, providing very little contrast at lower resolution.

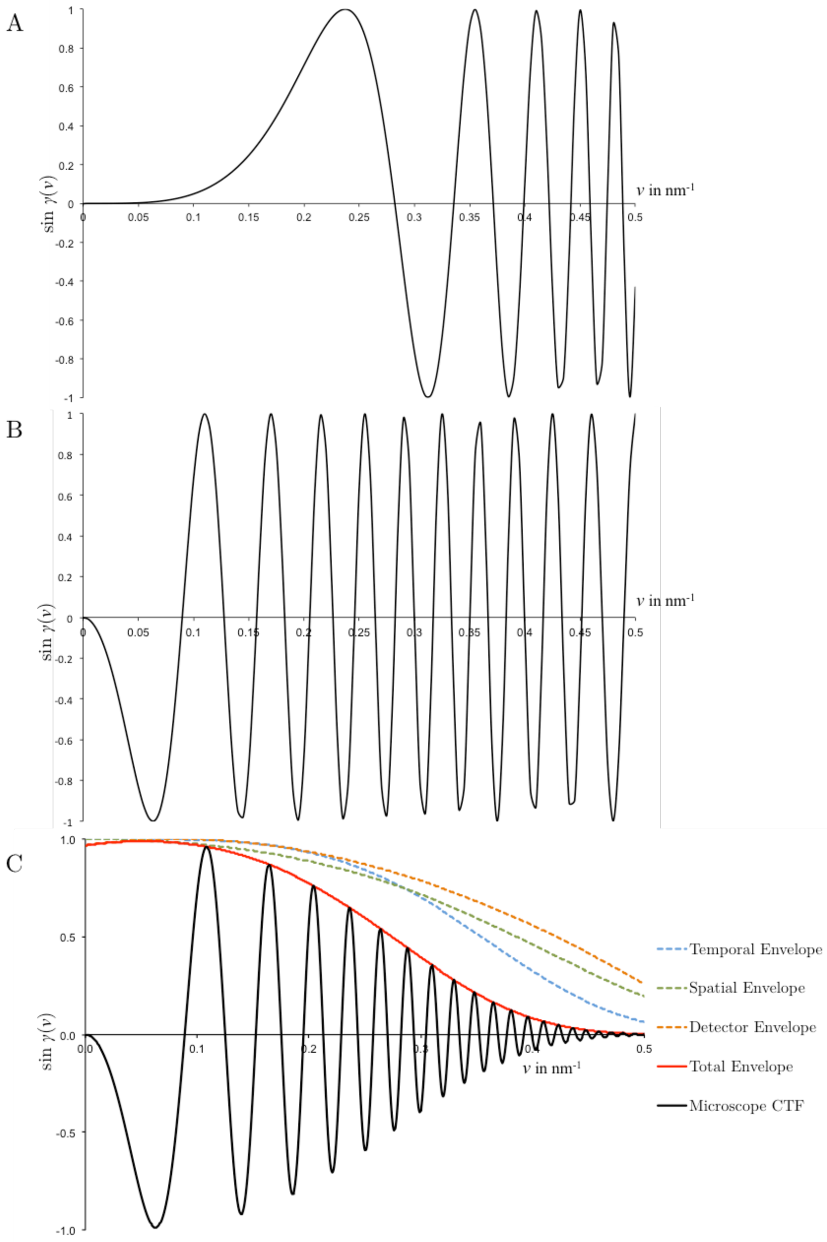


Figure 1.7 The phase contrast transfer function over spatial frequency

Plots in **A** and **B** show the CTF $\sin\gamma(\mathbf{s})$ plotted against spatial frequency ν in nm^{-1} for a microscope operating at 200 keV and spherical aberration constant C_s of 2 mm, without any dampening of the contrast at higher spatial frequencies. **A** shows the defocus Δf at 0 nm (i.e. at focus). **B** shows the defocus at 500 nm (underfocus). **C** shows defocus at 500 nm (underfocus) with all the envelope functions, which dampen the contrast at higher resolution. The temporal envelope includes chromatic aberrations, energy spread, focal spread and scattering of electrons. The spatial envelope includes factors such as an incoherent beam of electrons and instabilities in the lens. Detector envelope is the modulation transfer function of the detector. CTF plots generated using a web-based application (Jiang & Chiu, 2001).

However, as can be seen from Figure 1.7, there are caveats to taking images at a high defocus. Although the contrast does increase, so does the number of times the $\sin\gamma(\mathbf{s})$ crossing zero, giving rise to missing information in the Fourier components at those spatial frequencies. An optimum defocus thus needs to be reached to obtain maximum information with minimum data loss.

The fluctuations in the transform are clearly visible in the Fourier transform of the final image (Figure 1.8). The oscillations appear as bands of concentric rings, which show smoothly alternating bright and dark regions corresponding to the $\sin\gamma(\mathbf{s})$ fluctuations. These are called *Thon rings.*

The dark rings in between the bands are when the function crosses zero. The contrast from Fourier components at these points is lost from the image. Therefore, data need to be collected at various defoci and then merged, in order to make sure the lost information is recovered. The effects of the CTF are sometimes evident in the image, as reversals in contrast, around objects contributing to a lower spatial frequency.

The values of the oscillating CTF are dampened at higher spatial frequencies by an *envelope function* (Figure 1.7 and Figure 1.8). This dampening arises due to chromatic aberrations, energy spread, focal spread, scattering of electrons, spatial aberrations due to incoherent beam of electrons, and electrical instabilities in the lenses. The envelope function depends on the environment and microscope in use.

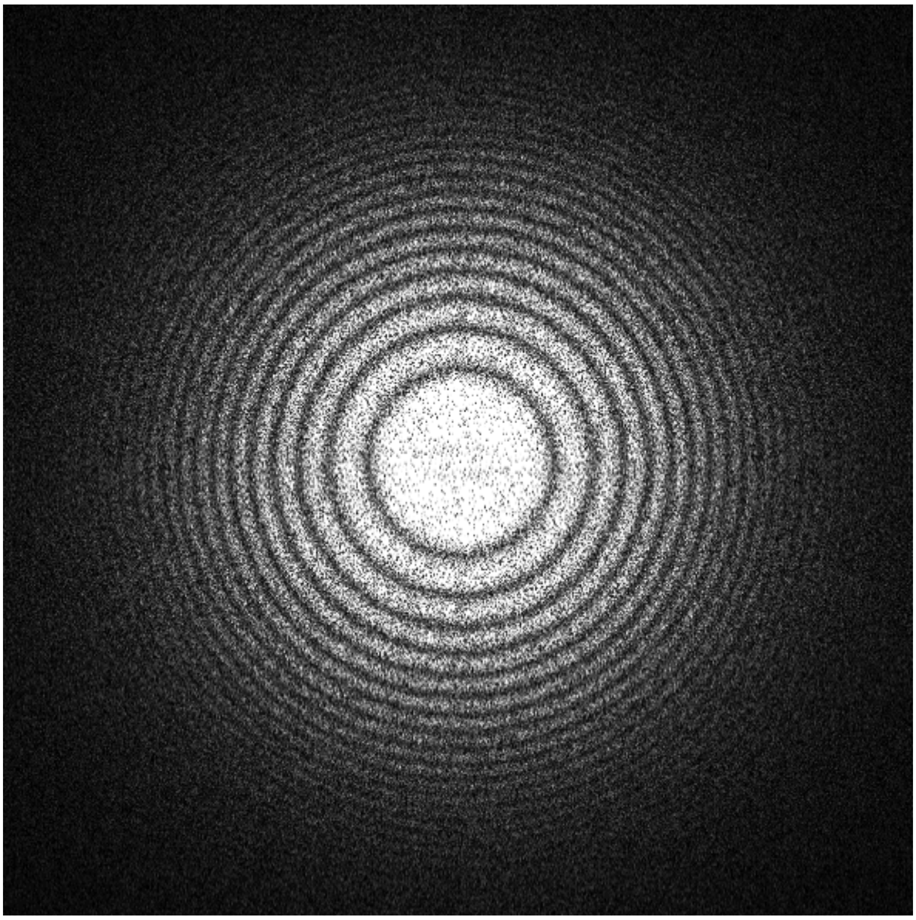


Figure 1.8 CTF fluctuations as seen in the FFT of a micrograph of amorphous carbon

The Fourier Transform of a micrograph of catalase crystals on amorphous carbon film, collected on a FEI Tecnai F20 G2 microscope operating at 200 keV at ~8600 nm underfocus, showing the *Thon rings* clearly. The oscillations of CTF, with dampening effects at higher spatial frequency, are clearly visible. The image was recorded at ~68,000 × magnification with a Falcon II direct electron detector (FEI).

1.1.5. Electron Crystallography

Similar to an arbitrary function, the *electron density* in a crystal can be represented as the sum of sines and cosines. To simplify, one can consider a one-dimensional crystal, such as these series of N number of dots.

………………………………………………

Equation 1.12 below represents the electron density $\rho(x)$ of this 1D crystal. The frequency or periodic repeats of the crystal is represented by the unit cell periodicity, L, and the two other sinusoidal wave characteristics, the amplitude and phase for each unit cell, n, is given by $F(n)$ and $\phi(n)$.

$$\rho(x)=\sum_{n=1}^{N} F(n)\, cos\left\{-2\pi\frac{n}{L}x+\phi(n)\right\} \qquad 1.12$$

The above equation is also called the Fourier series. Since, n/L is the repeating period of the cosine wave, its reciprocal is the spatial frequency (s). For a 3D crystal the repeating units in reciprocal space are represented by integers h, k, and l also known as *miller indices.*

Using *Euler identity* on the above and introducing miller indices in place of n/L we can write,

$$\rho(x,y,z)=\sum_{h,k,l}|F_{hkl}|\, e^{-2\pi i(hx+ky+lz-\phi_{hkl})} \qquad 1.13$$

This forms the basis for reconstruction of electron density from the amplitude, phase and h,k,l indices. The trio of amplitude, phase and miller index is known as the *structure factors* for a reflection. The miller indices are related to the real-space unit cell dimensions of the crystal using the relationship described in Equation 1.6.

1.4.1.1. 2D crystals

The Fourier transform of a periodic function is a set of discreet regularly spaced points in Fourier space. Consequently, the Fourier transform of a 3D crystal (which is a periodic electron density function) produce a 3D set of points in Fourier space, corresponding to its diffraction spots, with each spot having the miller indices *h,k,l*. The structure factor information from each spot can then be used to reconstruct the electron density as shown in Equation 1.13. However, in electron crystallography one is looking at thin crystals that are a single molecule thick. The periodicity, in these crystals, is along the *x,y* plane. The diffraction pattern consequently, in the corresponding reciprocal axis, *c**, assumes a different pattern. Below is a brief overview of what happens.

Mathematically, a 2D crystal is simply a product of a 3D-crystal and a rectangle function, which extends to infinity, and assumes the thickness of a single unit cell of the crystal (Figure 1.9). In the equation below, the rectangular slab, $rect(z)$, has a value of 1 inside the slab and 0 outside it. Now, a Fourier transform of this mathematical product, becomes a convolution product in reciprocal space (Equations 1.14).

$$\begin{aligned} \rho_{2D}(x,y) &= \rho_{3D}(x,y,z) \cdot rect(z) \\ \mathcal{F}\{\rho_{2D}(x,y)\} &= \mathcal{F}\{\rho_{3D}(x,y,z) \cdot rect(z)\} \qquad 1.14 \\ F_{2D}(s) &= F_{3D}(s) * \sigma(z^*) \end{aligned}$$

The *x,y* is replaced with the special frequency vector, **s**, in reciprocal space. σ is the Fourier transform of the rectangular function, which is a *sinc* function. F_{3D} represents the 3D reciprocal lattice space.

The convolution product of the 3D lattice points with a one-dimensional sinc function leaves the sampling unaffected in two-dimensions, therefore one can still get discreet diffraction spots, identified by (h,k) indices. However, *the convolution product generates a continuous function in the third-dimension with the variable z*.*

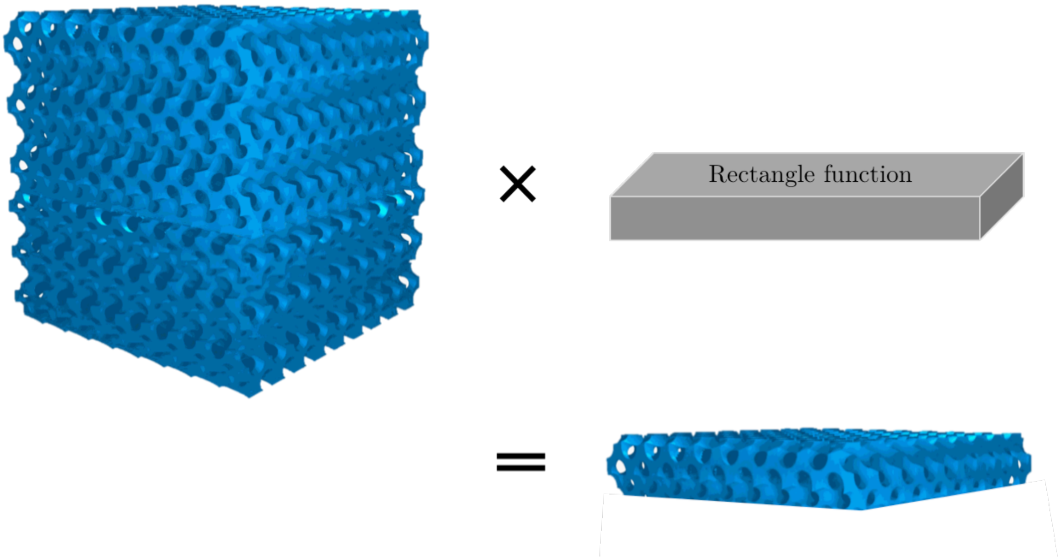


Figure 1.9 A 2D crystal is the product of a 3D crystal and a rectangle function

A 2D crystal can be represented as a mathematical product between a 3D crystal and a rectangle function, which is one unit cell thick. This multiplication becomes a convolution in Fourier space between the 3D diffraction pattern, which is the transform of a 3D crystal, and a *sinc* function, which is the Fourier transform of a rectangle function. Image adapted from (Chutinan et al, 2003), copyrighted by the American Physical Society.

This continuous reciprocal lattice line, sometimes called lattice rods, shows structure factor values varying continuously along the line (Figure 1.10). Therefore each diffraction spot is identified with two, rather than three, miller indices, *h,k*. When a 2D crystal is tilted however, the lattice line is sampled at a distinct point, indicated by (h,k,z^*).

One can imagine a situation where layers of 2D crystals are stacked on top of each other to eventually make it a 3D crystal. As the periodicity increases in the c^* direction, the lattice lines assume a more distinct shape, slowly converging towards a spot at an index (h,k,l). The Fourier transform of a 3D-crystal and a transform of a 2D crystal, is therefore, identical at position where $z^* = lc^*$.

In case of 2D crystals the l index is imposed later, to sample the values from the lattice lines at particular points, before reconstructing the density map using Equation 1.13. This is done by estimating a specific value of c, which in most cases is 1.5 to 2-times the best estimate of the crystal thickness.

1.4.1.2. Projection theorem and sampling tilted structures

The relationship between the diffraction pattern of a weak-phase object, produced in the back focal plane, and the Fourier transform of the image was discussed earlier in Section 1.1.3. This relationship also requires the assumption that the *Ewald sphere* closely approximates a plane, in reciprocal space.

This approximation is valid in the case of electron crystallography of proteins (except at very high resolution or for very thick crystals), where the wavelength of high-energy incident electrons is very small, resulting in a very large radius of the Ewald sphere (which is 1/wavelength). The surface of the sphere thus approximates a plane. The Fourier transform of a 2D projection therefore resembles the ‘Ewald plane’ sampling the reciprocal lattice lines.

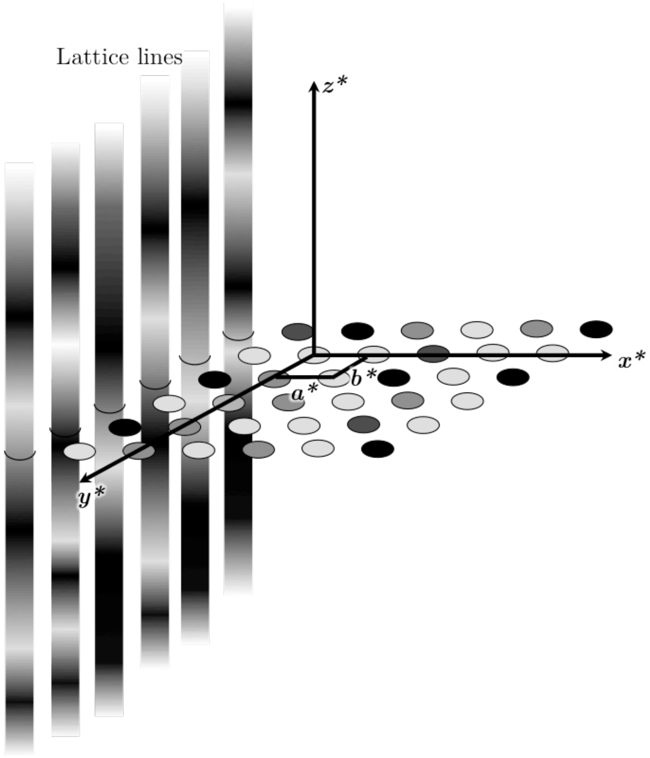


Figure 1.10 Cartoon representation of the lattice lines in *z**

Each spot in the transform of a 2D crystal extends, from the untilted *x*,y** plane, into the third dimension of Fourier space as lattice lines. The parallel lattice lines are shown as a cartoon representation, extending in the *z** direction. Only six lattice lines are shown. The others are represented as shaded discs intersecting the *x*,y** plane. The differential shading indicates how the amplitude and phase values vary continuously along the lattice lines. The *a** and *b**, shown are the lattice dimensions in reciprocal space. All the unit cell angles are shown as 90° for convenience.

This 2D plane in reciprocal space is a section through the 3D Fourier transform of the object, passing through the origin of the Fourier transform. The direction of the projection, in real-space, is parallel to direction of the incident electrons. The central section in reciprocal space is in the same direction as the projection but passes through the origin of the reciprocal space. This relationship of a 2D projection and its Fourier transform is known as the *central section theorem* or the *projection theorem.*

As the sample being observed is tilted, the direction of incident electrons changes and so does the 2D projection, relative to the sample. The Fourier transform of this tilted projection is a central section through the origin of the transform, with the same tilt parameters as the sample. The central section (Ewald plane) intersects the continuous reciprocal lattice lines in z^* direction at particular points depending on the tilt parameters. This sampling results in (h, k, z^*) spots on the 2D transform. For an untilted specimen, the central section is parallel to the a^*,b^* plane, therefore there is no sampling of the lattice lines and the indices for each spot is ($h,k,0$). As the tilt angle of the sample in real space is changed so does the sampling in reciprocal space along z^*. The structure factors vary with each h,k spots along the z^*.

Images can be collected at various tilt angles to sample different regions of the 3D transform. This builds up an overall 3D pattern that can be later used to reconstruct the 3D density of the crystal. However, instrumental limitations make it difficult to tilt a sample to angles higher than 60°. The result is two inverted cones of information that cannot be sampled. This is commonly known as the *missing cone problem* (Figure 1.11). Consequentially, the inverse transform yields a 3D model that appears somewhat distorted and stretched in the z direction.

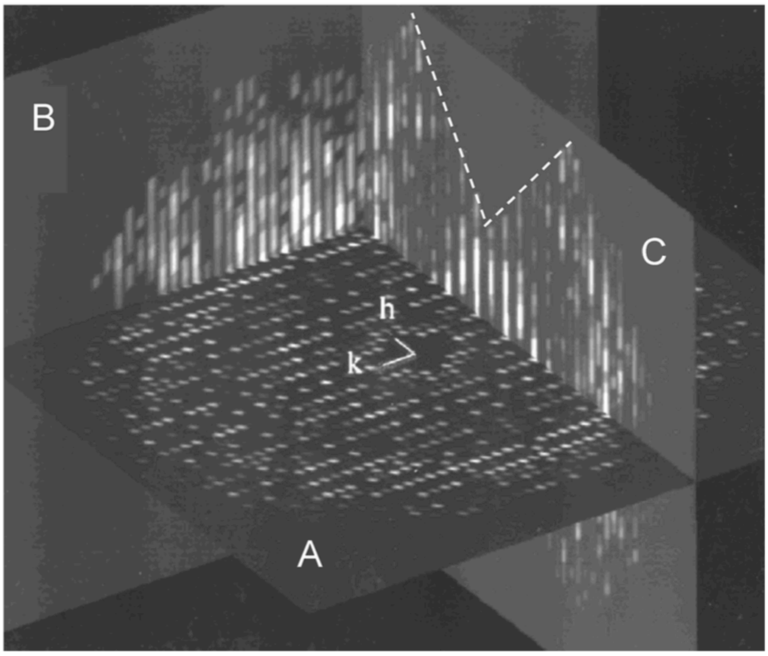


Figure 1.11 Missing wedge of data in a 3D transform

The above image represents the full 3D transform of a tubulin crystal displayed using three planes. Plane A, B and C are perpendicular to each other and parallel to the x^*y^*, y^*z^* and x^*z^* planes, respectively. These intersect the 3D transform, with the intensity of the transform made to become visible where the planes intersect the transform. The missing wedge can be seen in plane C, indicated with dashed lines. In three-dimensions this wedge becomes a cone. There is another wedge underneath the plane A. h and k are the unit vectors indicating the origin of the transform. Image reproduced from (Nogales et al, 1997) with permission, copyrighted by Elsevier Ltd.

1.5. Cryo-electron microscopy

Imaging unstained biological samples, which have low atomic number, results in immense damage due to inelastic scattering (Section 1.1.1.2). Unwin and Henderson reported the first TEM structure of purple membrane and catalase, up to a resolution of 7 and 9 Å respectively, by embedding the crystalline sample in glucose (Unwin & Henderson, 1975), using low-dose techniques.

Imaging the cold sample, at cryogenic temperatures of around −190°C or less, has been found to reduce sample damage significantly. Furthermore, with the advent of DDDs, several exposures can be carried out in multiple frames per second, allowing users to select the initial frames with the least radiation damage and the highest resolution information, while discarding the later frames. Any drifts and sample movement can also be corrected by cross-correlation among the frames, further improving the resolution. Remarkably, Unwin's and Henderson's forecast that improvement in technology will result in structures reported up to 3 Å has so far been surpassed and atomic resolution models are regularly being reported using cryo-electron microscopy and DDDs (Bartesaghi et al, 2015).

In cryo-EM, the sample is usually embedded in a thin layer of vitreous ice. This ice behaves as a glass-like amorphous solid within which the sample remains embedded in its native hydrated state. Rapid freezing of the sample in water or in any aqueous solution generates this ice, while also preventing formation of cubic or hexagonal ice crystals that cause sample damage and electron scattering (Dubochet et al, 1982). The sample can then be kept at very low temperatures, which reduces radiation damage. Such low temperatures allow imaging at higher dose of electrons for better SNR. For some low resolutions, better SNR is achieved when samples are over-irradiated rather than under-irradiated (Baker et al, 2010), but electron dose can change the structure of the sample at higher resolution.

Therefore a compromise needs to be achieved between SNR and resolution, so that low-resolution structures can be observed which allow sample identification and alignment, but preserves higher resolution information as well. Most images in cryo-EM are taken at low SNR, aiming to maintain the high-resolution information. Figure 1.12 shows the loss of SNR as a function of increasing electron dose for various resolutions.

1.5.1. Radiation damage and temperature factor

Figure 1.13 shows the exponential decrease in intensities of diffraction spots with increasing electron dose for a purple membrane crystal in glucose (Unwin & Henderson, 1975). The intensities of spots at higher spatial frequencies drop quicker than the rest.

Other than radiation damage, the disorder in crystals can contribute to a drop in intensities at higher spatial frequencies. This disorder can be attributed to thermal motion and imperfections in crystals such as small structures, which do not repeat perfectly from one unit cell to the next. The fall-off in intensities in a cryo-EM sample can be visualised by plotting the natural logarithm of the intensities against the square of spatial frequency (s^2) in a *Guinier plot.* The falling intensities can be corrected, based on a reference structure, by multiplying a negative *temperature factor* or $B_{restore}$ to the measured intensities, in the form of Gaussian function $e^{-(\frac{B_{restore}}{2}s^2)}$ (Havelka et al, 1995).

1.6. Image recording

1.6.1. CCD and Direct detectors

Digital detectors brought about a change in the way imaging is done on TEM, with ease of use. Charge-coupled device (CCD) sensors convert electrons to photons using a scintillator. The photons then travel to a peltier-cooled CCD array, via fibre optics, which convert the photons to electric signals. CCDs are very sensitive and show a very linear response.

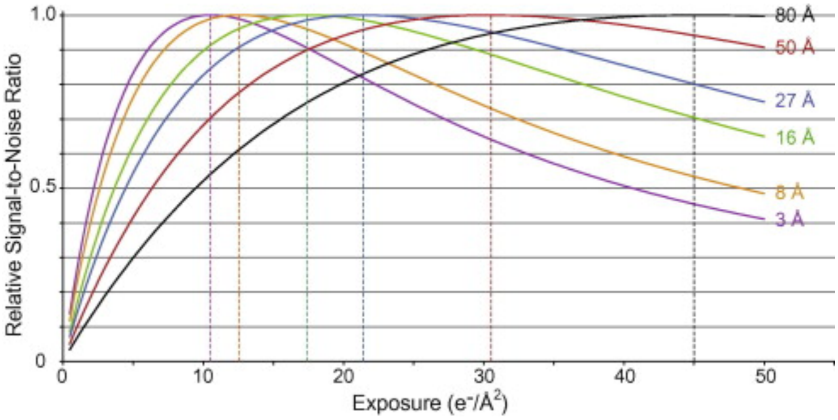


Figure 1.12 Relative signal-to-noise ratio as a function of electron dose

The relative SNR from 0 to 1 are plotted against electron dose in electrons per Å^2, for six different resolution shells. Dotted lines indicate the doses at which SNR is 1 for the respective resolution shells. Some low-resolution information has better SNR at high doses, but at such doses the SNR decreases dramatically for the high-resolution data. Thus, a best exposure requires a compromise between resolution and exposure. Image obtained from (Baker et al, 2010) with permission from Elsevier Ltd.

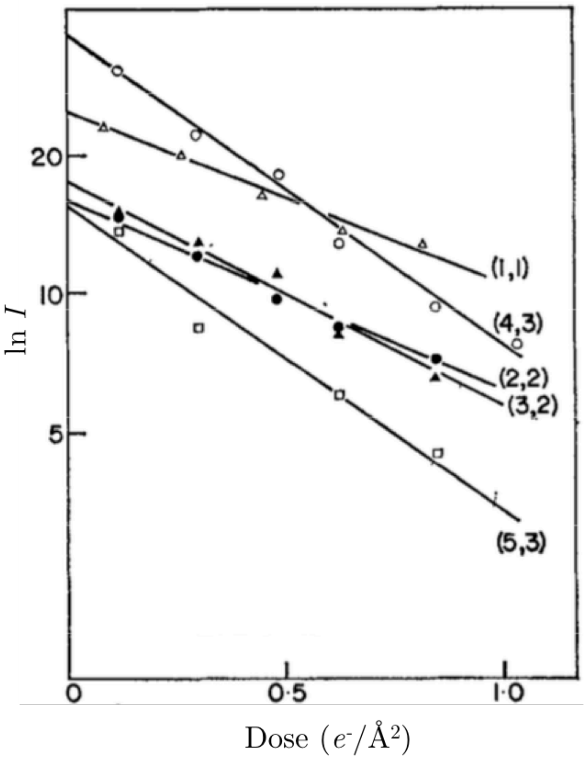


Figure 1.13 Intensity decline due to radiation damage

Figure from (Unwin & Henderson, 1975) showing the effects of radiation damage on certain diffraction spots of the purple membrane 2D crystals embedded in glucose. Logarithms of diffraction intensities are plotted against electron dose in number of electrons per square angstroms. Intensities of high-resolution spots fall-off quicker compared to low resolution spots. Figure reproduced with permission, copyrighted by Elsevier ltd.

Common CCD chips used measure 5 × 5 cm^2, with pixel size of 15 µm, giving 4k × 4k pixels in one image. Although this resolution is closer to what is achieved after the digitisation of photographic films, the gold-standard medium in the past for EM imaging, it does impose certain magnification restrictions (Chen et al, 2008). The modulation transfer function (MTF) of the device influences the resolution of CCD detectors.

Direct detection devices (DDDs) have recently replaced the use of films completely (McMullan et al, 2016). These devices work using pixels with complementary metal-oxide semiconductor (CMOS), which amplifies signals that are directly received by monolithic active-pixel sensors (MAPS) that track charged particles. This results in an instant conversion of incident electrons to voltage, without the need to convert to light energy first. This direct exposure improves the signal-to-noise (SNR) ratio. Furthermore, the silicon pixel sensors are ‘back-thinned’, resulting in reduced track for the electrons and a dramatic improvement in SNR. Improvements using *hybrid* pixels have been made, where each pixel can ‘count’ electrons rather than simply producing a signal at the incident region, further increasing the resolution (Faruqi & Henderson, 2007).

Other than the MTF, the detective quantum efficiency or DQE is an important factor in image detection. It is related to SNR by the following relationship

$$\mathrm{DQE} = \frac{\mathrm{SNR}^2{}_{\mathrm{in}}}{\mathrm{SNR}^2{}_{\mathrm{out}}} \qquad 1.15$$

SNR is the square root of the ratio of number of incident electrons producing signal at each pixel and the total number of electrons at that pixel. Since at low electron doses, the number electron incident on each pixel is small, hence the noise from the detector cannot be ignored. In case of $\mathrm{SNR}_{\mathrm{out}}$ the mean square value of noise produced by the detector is added to the denominator of the ratio. Figure 1.14 plots the DQE against spatial frequency between film and three types of DDDs.

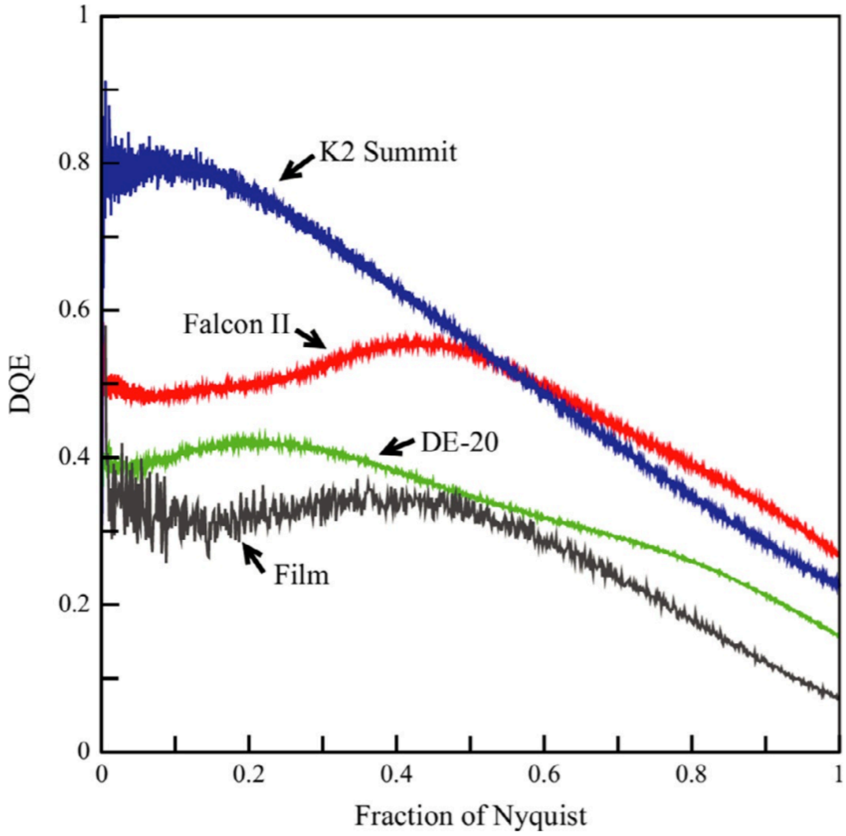


Figure 1.14 Detective quantum efficiency (DQE) of film and DDDs

The graph compares the detective quantum efficiency (DQE) between commonly used electron microscope film (Kodak SO-163) and three of the currently used common DDDs – DE-20 (Direct Electron), Falcon II (FEI) and K2 Summit (Gatan). The DQE is plotted against fraction of Nyquist frequency. The fall of in DQE represents the reduction in SNR in the detectors at increasing spatial frequency. Image taken with permission from (McMullan et al, 2014), copyrighted by Elsevier Ltd.

Chapter 2 - Materials and Methods

2.1. General Methods

2.1.1. Bacterial strains and media

2.1.1.1. Bacterial strains

Table 2.1 *Escherichia coli* strains

Strain name	Source/Reference	Notes	Genotype
NEB5-alpha	New England Biolabs	Modified DH5-α strain	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
CA434	(Purdy et al, 2002)	Strain HB101 carrying the IncPβ conjugative plasmid, R702, Dam^{+}, Tra^{+}, Mob^{+}	thi-1 hsdS20 (r-B, m-B) *supE44 recAB ara-14 leuB5proA2 lacY1 galK rpsL20* (strR) *xyl-5 mtl-1*

Table 2.2 *Clostridium difficile* strains

Strain name	Ribotype	Source	Notes
R20291	027	Stoke Mandeville Hospital, Buckinghamshire, UK, outbreak 2004 (Stabler et al, 2009)	Hyper-virulent strain; Toxin A^{+}B^{+}Cdt^{+}
Ox247	005	University of Oxford 2007	Glycosylated SlpA LMW SLP, putative glycosylation genes in S-layer cassette (Dingle et al, 2013)
Ox247Δorf2	005	Prof. Neil Fairweather	Clostron mutant of Ox247 (orf2) lacking glycosylation of SlpA, nonsense mutation in open reading frame 2 in the putative glycosylation gene cluster in the S-layer cassette
630	012	Isolated in Zurich outbreak, 1982 (Wüst et al, 1982)	General laboratory stain
R7404	017	Prof. Neil Fairweather	UK and Ireland strain; Serotype F; Toxin A^{-}B^{+}
FM2.5	027	Avidbiotics Ltd.	Avidocin resistant derivative of R20291, *slpA*$^{-}$, frame-shift mutation ('A' insertion at residue 283) in *slpA*
SlpA$^{\Delta Ld2}$	027		FM2.5 strain with LMW domain 2 deleted *slpA* from ribotype 027 inserted in to wild type 027 *slpA* locus

2.1.1.2. Bacterial growth media

All media were prepared using Millipore 18 MΩcm^{-1} grade water. Media were sterilised by autoclaving at 121°C for 20 min prior to use. Additives not suitable for autoclaving were sterilised by filtration through 0.2 µm Millex filters (Milipore).

Table 2.3 Table of growth media

Media	Ingredients (Source)	Amount required for 1 litre
Luria-Bertani (LB) broth	LB Broth, Miller (Fisher Scientific)	25 g
Luria-Bertani (LB) agar	LB Broth, Miller (Fisher Scientific)	25 g
	Difco Bacto Agar (BD)	15 g
TY Broth	Bacto Tryptone (BD)	30 g
	Bacto Yeast Extract (BD)	20 g
Brain Heart Infusion (BHI) broth	Brain Heart Infusion Broth (Oxoid)	37 g
Brain Heart Infusion (BHI) agar	Brain Heart Infusion Broth (Oxoid)	37 g
	Difco Bacto Agar (BD)	15 g
SOC Medium	SOC Outgrowth Medium (New England Biolabs)	

All antibiotics stocks were filter sterilised through 0.2 µm Millex filters (Milipore) and stored at −20°C. Chloramphenicol was prepared as a 30 mg/ml stock in 70% (v/v) ethanol and used at a working concentration of 30 or 15 µg/ml. Similarly, thiamphenicol was prepared as a 15 mg/ml stock in 70% (v/v) ethanol and used at a concentration of 15 µg/ml. Cycloserine solution at 10 mg/ml stock was prepared fresh in water and used at the working concentration of 250 µg/ml.

2.1.1.3. Culture conditions

E. coli were routinely grown in LB agar and in LB broth, aerobically at 37°C. *C. difficile* was cultured in an anaerobic cabinet (Don Whitley Scientific) supplied with anaerobic gas mixture (BOC) of 80% N_2, 10% H_2 and 10% CO_2, maintained at 37°C and 70% humidity. Media were depleted of oxygen by introduction into the cabinet for an h or more, prior to use, depending on the volume of media. For mid-logarithmic phase cultures, fresh media were inoculated from overnight cultures to an optical density at 600 nm (OD_{600}) of 0.05. The culture was then grown to the required OD_{600}.

2.1.2. Buffers and stock solutions

Table 2.4 Buffers and stock solutions used

Buffers	**Composition**
Lysis buffer	200 mM NaCl
	50 mM EDTA
	20 mM Tris-HCl, pH 8.0
Pronase	20 mg/ml Pronase from *Streptomyces griseus* (Sigma-Aldrich)
Coomassie brilliant blue gel staining solution	0.25 % (w/v) Coomassie brilliant blue R-250
	45% (v/v) Methanol
	10% (v/v) Acetic acid
Destaining solution	45% (v/v) Methanol
	10% (v/v) Acetic acid
Resolving gel (final concentration 12% Acrylamide) – 10 ml	4 ml 30% Acrylamide/Bis
	2.5 ml 1.5 M Tris-HCl, pH 8.8
	2.4 ml deionised water
	100 µl 10% SDS
	50 µl 10% APS
	10µl TEMED
Stacking gel (final concentration 5% Acrylamide) – 5 ml	833 µl 30% Acrylamide/Bis
	1.25 ml 0.5 M Tris-HCl, pH 6.8
	2.87 ml deionised water
	50 µl 10% SDS
	25 µl 10% APS
	5µl TEMED
SDS PAGE buffer	25 mM Tris base, pH 8.3
	192 mM glycine
	0.1 % SDS
1 × Laemmli sample loading buffer	277.8 mM Tris-HCl, pH 6.8
	4.4% SDS
	0.02% Bromophenol blue
	20% (v/v) Glycerol
TAE buffer	40 mM Tris, pH 7.6
	20 mM Acetic acid
	1 mM EDTA

DNA sample loading dye	10 mM Tris-HCl, pH 7.6 (adjusted with NaOH) 0.03% Bromophenol blue 0.03% Xylene cyanol FF 60% Glycerol 60 mM EDTA
HEPES buffered saline	50 mM HEPES (Sigma-Aldrich), pH 7.5 150 mM NaCl

2.1.3. Plasmids and primers

2.1.3.1. Plasmids

Table 2.5 Table of Plasmids

Plasmid name	**Size**	**Vector Backbone**	**Markers**	**Notes**
pRPF233	9.5 kbp	pMTL960	Chloramphenicol (15µg/ml) & Thiamphenicol (15µg/ml)	pRPF185 (from pMTL960; Ptet, gusA double terminator) with R20291 *slpA*
pOB001	9.1 kbp	pMTL960	Chloramphenicol & Thiamphenicol resistance	pRPF233 with 027 *slpA* with LMW domain 2 (4,496 – 4,931 bp) deleted and replaced with GGA GGT (Gly-Gly residue)

2.1.3.2. Primers

Table 2.6 Table of Primers

Primer name	**Sequence 5’ → 3’**	**Annealing temperature**
NF1323	`CTGGACTTCATGAAAAACTAAAAAAAAATATTG`	59.2°C
NF2035	`CCTGTGTTTGCAGCAGAAGATATG`	61.0°C
RF4	`CTATATCTATCTGGAGATTTACC`	55.3°C
RF25	`gtttttgttaccctaagtttTTCCTGTTAAAGCATTTAGATATAC`	58.8°C
RF33	`GGCTTCTATAGCTTGGTGAA`	55.3°C
RF102	`ggtTCTGGAAGCCCAGTAATAACTAAAC`	59.7°C
RF103	`TCCAGAGCTTATTAAGAAATCTACATAAtcc`	57.8°C
RF793	`GCTAAAGGTTTAGATACATAATATACC`	57.4°C
RF794	`GGAGAATAATGACTGAGAGAATGGC`	61.3°C

2.2. S-layer extraction

The S-layer from *C. difficile* was extracted in two different procedures for EM imaging. The first method described, isolates the S-layer intact as a capsule retaining the original shape. The cell, along with its membrane, ruptures, releasing its contents into the suspension. On fixing this sample on an EM grid with negative stain, the capsule appears as a flattened double-layered tubular structure. This technique uses a peptidoglycan digesting endolysin enzyme from the *C. difficile* phage CD27 (Mayer et al, 2011). The His-tagged enzyme was expressed from plasmid pHAS042 in *E. coli*, and purified using HisTrap HP column (GE Healthcare) on AKTA Prime Plus FPLC system (GE healthcare).

The second method for S-layer extraction was adapted from (Kawata et al, 1984) This method breaks the cell mechanically using glass beads of a particular size and density. The fragments were stored in 2% (v/v) Triton X-100. After flash freezing in liquid nitrogen, the sample can be stored indefinitely at −80°C.

2.2.1. Endolysin treatment

10 ml TY broth was inoculated with an overnight culture of *C. diffcile* to an OD_{600} of 0.05. This was incubated anaerobically for 3-4 h at 37°C. The OD_{600} was measured to confirm a mid-log phase. 1 ml of the sample was then harvested at 10,000 × g for 10 min and the pellet resuspended in 50 mM HEPES buffered saline, to an OD_{600} of 10. Endolysin (10% v/v) was added to this sample and the sample was incubated at 37°C for 30 min to 1 h. Longer incubation reduces sample quality dramatically, possibly due to the complete disintegration of the cell wall, resulting in the loss of an ordered lattice. 5 µl of the sample was immediately transferred to a freshly glow-discharged carbon-coated TEM grid and stained with 2% uranyl acetate or 0.75% uranyl formate for TEM imaging.

2.2.2. Cell surface fragmentation

2 ml aliquots from 5 ml overnight culture of *C. difficile* was inoculated in 60 ml TY

broth and incubated anaerobically for 3-4 h at 37°C. When the OD_{600} reached 0.6 – 0.8, the culture was centrifuged at 2,000 × g for 15 min at 4°C. From this step onwards, all procedures were carried out at 4°C or on ice. The cell pellets were resuspended in 30 ml cold deionised water, by gentle pipetting. The cells were harvested and washed again, resuspending in 15 ml cold deionised water.

An OD_{600} reading of around 2.4 – 3.2 is expected at this stage. The resuspension was combined together and added to a homogenizing flask with equal volume of pre-cooled 212 – 300 µm acid-washed glass beads (Sigma G1277). The cold resuspension and glass bead mixture was then homogenized in a Braun MSK (953030) homogenizer for 30 s and immediately put back on ice. The homogenate was allowed to cool for 5 min and then carefully decanted, leaving the glass beads behind. The homogenate was centrifuged at 800 × g for 10 min at 4°C to remove unbroken cells. The supernatant was then centrifuged at 3000 × g for 10 min at 4°C. The resulting pellet, containing cell wall/S-layer fragments with the cell membrane, was washed with 400 µl cold 1 M NaCl and centrifuged at 3000 × g for 10 min at 4°C. Finally the pellet was resuspended in 200 µl cold 2% Triton X-100 to remove the cell membrane and aliquoted into 1.5 ml centrifuge tubes.

The sample was checked immediately on freshly glow-discharged carbon-coated TEM grid by negative staining. Fragments were more stable in 2% Triton X-100. Further washing steps with cold deionized distilled water can be performed. However, the fragments roll up and undergo phase-separation over time, so immediate flash freezing in liquid nitrogen is recommended. The fragments can be stored in 1 M NaCl as well, but best results were obtained when samples are in 2% Triton X-100, where they show well-ordered crystals after thawing.

2.3. Nucleic acid methodologies

2.3.1. Transformation of *E. coli* strains

Transformation was done using chemically competent *E. coli* cells. Cells were made competent using the $CaCl_2$ method, as described below. For high-efficiency transformations, NEB5-alpha cells (New England Biolabs) were purchased directly from the commercial source.

2.3.1.1. Transformation of chemically competent cells

E. coli cells were inoculated in LB broth and incubated overnight. The overnight culture was then used to inoculate a fresh 100 ml LB broth and grown to an OD_{600} between 0.4 and 0.6. The cultures were incubated on ice for 20 min and the cells were harvested by centrifugation at 6,000 × g for 8 min at 4°C. The pellet was resuspended in 20 ml of cold $CaCl_2$ solution (60 mM $CaCl_2$, 15% (v/v) glycerol, 10 mM PIPES, pH 7) and harvested as before. The pellet was resuspended in 20 ml cold $CaCl_2$ solution and incubated on ice for 30 min before harvesting again. Finally, the cells were resuspended in 4 ml cold $CaCl_2$ and incubated overnight on ice and stored at −80°C in 100 µl aliquots.

The competent cells were transformed by gently mixing freshly thawed aliquots of competent cells from −80°C with 0.01 – 1 µg of plasmid DNA. The mixture was then incubated on ice for 30 min. The cells were heat shocked at 42°C for 30 s and immediately placed on ice for 5 min. 950 µl of SOC medium was then added and placed at 37°C for 60 min with vigorous shaking or rotating. The cells were serially diluted to several 10-fold dilutions in SOC medium and 50 – 100 µl of each dilution was pipetted onto pre-warmed selection plates. The plates were then incubated overnight at 37°C.

2.3.2. Purification of plasmids from *E. coli*

Plasmids were purified using GeneJET Plasmid Miniprep Kit (Thermo-Fisher) following the manufacturer's protocol. All centrifugation was done in a microcentrifuge at 20,000 × g at room temperature, unless otherwise stated. 5 ml *E. coli* cells were grown at 37°C overnight in LB media with constant shaking. The cells were then harvested by centrifuging at 6,000 × g for 2 min. The supernatant was decanted and the pellet was resuspended in 250 µl Resupension Solution and vortexed. 250 µl of Lysis Solution was added to this and the tube was inverted 4-6 times. 350 µl of Neutralising Solution was added and the tube was inverted 4-6 times. The tube was then centrifuged for 5 min. The supernatant was transferred to the GeneJET Spin Column and centrifuged for 1 min. The flow-through was discarded. The column was washed twice with 500 µl of Wash Solution and centrifuged for 1 min. The flow-through was discarded each time. Finally the column was centrifuged empty for 1 min. The column was then transferred to a fresh 1.5 ml centrifuge tube. The DNA was eluted from the column by added 50 µl of nuclease-free water (New England Biolabs), incubated for 2 min and centrifuged for 1 min. The concentration of the eluted plasmid DNA was measured at 260 nm wavelength using a nanodrop machine and stored at −20°C.

2.3.3. Purification of chromosomal DNA from *C. difficile*

1 ml of overnight culture of *C. difficile* cells in TY broth was harvested by centrifuging at 20,000 × g. The pellet was resuspended in 200 µl of lysis buffer (Table 2.4) and transferred to a fresh 1.5 ml centrifuge tube. 4 µl of endolysin was added to this and incubated in a 37°C water bath for 60 min. To the lysate, 10 µl of 20 mg/ml Pronase from *Streptomyces griseus* (Sigma-Aldrich) was added and incubated in a 55°C water bath for 60 min, to digest cellular proteins. 80 µl of 10% N-laurylsarcosine (Sigma-Aldrich) was added and incubated in a 37°C water bath for 60 min. 200 µl of nuclease-free water (New England Biolabs) was added to the above along with 0.2 mg/ml RNase (New England Biolabs) and incubated in a 37°C water bath for 60 min. This sample was transferred to a 1.5 ml Phase Lock Gel (PLG) tube (VWR), along with an equal volume of phenol:chloroform:isoamyl alcohol in a 25:24:1 ratio, and mixed by inversion. The sample was centrifuged at 16,100 × g for 2 min and the aqueous phase from the top of the gel was pipetted and transferred to fresh 1.5 ml PLG tube, where equal volume of 25:24:1 ratio of phenol:chloroform:isoamyl alcohol was added and mixed by inversion and centrifuged again at 16,100 × g for 2 min. The aqueous phase was extracted again and transferred to fresh 1.5 ml PLG tube with an equal volume of 24:1 ratio of chloroform:isoamyl alcohol and centrifuged at 16,100 × g for 2 min. The aqueous

phase was now transferred to new 1.5 ml centrifuge tube. The DNA was precipitated by adding an equal volume of ice-cold 100% isopropanol and left overnight at −20°C. The sample was spun at 4,000 × g at 4°C to pellet the precipitated DNA. The pellet was washed with 70% ethanol and air-dried. The dry pellet was finally resuspended by adding 50 µl nuclease-free water and allowing the DNA to dissolve overnight at 4°C. The concentration of DNA was measured using NanoDrop Lite (Thermo Scientific).

2.3.4. Polymerase chain reaction

Polymerase chain reactions (PCRs) were carried out to amplify DNA for sequencing, to visualise DNA on agarose gel, or to carry out various downstream processes. PCR was performed on plasmid as well as chromosomal DNA templates. Two different enzymes were used and reactions were carried out in a T100 Thermal Cycler (Bio-Rad).

RedTaq (Sigma-Aldrich) enzyme mix was used which contains 5 units/µl 98 kDa Taq polymerase from *Thermus aquaticus,* which has an optimum activity at 74°C and stable up to 95°C. It has a 5'→3' DNA polymerase activity, and comes along with a red dye to visualise the dye front on 1% agarose gel electrophoresis and can be loaded directly without the need of any loading dye. One unit of the enzyme is capable of polymerising 10 nmol of deoxyribonucleoside triphosphates (dNTPs) into acid insoluble DNA in 30 min at 74°C. The 2 × RedTaq ReadyMix PCR Reaction Mix (Sigma-Aldrich) was used, which reduced pipetting steps and included 25 mM $MgCl_2$ and 10 mM of each dNTPs (dATP, dTTP, dGTP, and dCTP). PCR was carried out in thin-walled 200 µl PCR tubes. Following are typical reaction components and conditions.

Component	**Amount**	**Final Concentration**
2 × Readymix	25 µl	1 ×
Forward primer (~15-30 bases in length)	2.5 µl from 10 µM stock	0.5 µM
Reverse primer	2.5 µl from 10 µM	0.5 µM

(~15-30 bases in length)	stock	
Template DNA (<1.5 kb)	0.5 µl	~200 pg/µl
Water	19.5 µl	-
Total volume	**50 µl**	

The reaction conditions set on the thermal cycler are as follows.

Step	**Temperature**	**Time**	
Initial Denaturation	95°C	3 min	
Denaturation	95°C	30 s	× 35 cycles
Annealing	Appropriate annealing temperature (45 – 72°C)	30 s	
Extension	72°C	1 min/kb DNA template	
Final Extension	72°C	5 min	
Hold	12°C	∞	

For longer amplicons (>1.5 kb) 2 × Phusion High-Fidelity PCR Master Mixes (New England Biolabs) were used. This mix contains a recombinant *Pyrococcus furiosus* DNA Polymerase with an error rate 50-fold lower than that of *Taq* DNA Polymerase. The master mix contains 20 units/ml Phusion DNA polymerase, 0.2 mM of each dNTP, 1 × Phusion HF Buffer. Phusion DNA polymerase has 5'→3' exonuclease activity and is stable up to 98°C.

A typical reaction mixture contains the following.

Component	**Amount**	**Final Concentration**
2 × Phusion® High-Fidelity PCR Master Mix	25 µl	1 ×
Forward primer (~15-30 bases in length)	2.5 µl from 10 µM stock	0.5 µM
Reverse primer (~15-30 bases in length)	2.5 µl from 10 µM stock	0.5 µM

Template DNA (<1.5 kb)	0.5 µl	~200 pg/µl
Water	19.5 µl	-
Total volume	**50 µl**	

The reaction conditions set on the thermal cycler are as follows.

Step	**Temperature**	**Time**	
Initial Denaturation	98°C	3 min	
Denaturation	98°C	30 s	× 35 cycles
Annealing	Appropriate annealing temperature (45 – 72°C)	30 s	
Extension	72°C	15 s/kb DNA template	
Final Extension	72°C	5 min	
Hold	12°C	∞	

2.3.5. Agarose gel electrophoresis

DNA samples were visualised following electrophoresis separation on 1% agarose gel electrophoresis. 1 g of agarose (Sigma-Aldrich) was dissolved in 100 ml of 1 × TAE buffer and heated in a microwave for 1-2 min to dissolve the agarose. Ethidium bromide was added to the molten agarose to a final concentration of 1 µg/ml after the solution was cooled down to around 60°C. The solution was poured into a gel mould with combs, to create wells where the sample was loaded. After the gel solidified, the comb was removed and the gel submerged in 1 × TAE buffer. The DNA sample was loaded into the wells with DNA loading dye and the gel was electrophoresed in a Mini-PROTEAN Tetra Cell System (Bio-Rad) at 120 V, in 1 × TAE buffer, for 40 – 50 min or until the dye front reaches the bottom of the gel.

2.3.6. Cloning by inverse PCR to mutate *slpA*

Inverse PCR was used to delete a region on the *slpA* gene located on plasmid pRPF233 (see Chapter 4). 5’ phosphorylated primers (RF102 and RF103) were designed flanking the region of interest to be deleted, diverging from the region of interest.

PCR was performed in 50 µl volume, as described in Section 2.3.4, using 2 × Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB) with 0.5 µM forward and reverse primers and plasmid DNA. The PCR result was confirmed using 1% Agarose gel electrophoresis

The PCR product was subjected to DpnI digestion to remove the methylated bacterial plasmid from *E. coli.* The digestion was performed in Phusion buffer where 2 µl of 20,000 U/ml DpnI (New England Biolabs) was added to 45 µl of the PCR product and incubated at 37°C for 2 h. The PCR product was purified using the GeneJet PCR Purification kit (Thermo-Scientific). 50 ng of the PCR product was ligated using QuickStick Ligase (Bioline). 4 µl of 4 × QuickStick Ligase buffer was added to 50 ng of the PCR product in 14 µl volume. 1 µl QuickStick Ligase was added to the above mixture and the volume made up to 20 µl using nuclease-free deionized water. The mixture was incubated at room temperature for 15 min. 2 µl of the ligation product was transformed into 22 µl aliquot of NEB5-alpha highly-competent cells (NEB), using the protocol described in Section 2.3.1, along with an unligated PCR product as control.

2.3.7. Restriction digestion

Restriction digestion was performed in 10 × CutSmart Buffer (NEB) using the following reagent volumes and the digest from each reaction was visualized using 1% Agarose gel electrophoresis.

Deionised water	7 µl
CutSmart buffer	2 µl
Restriction enzyme 1	0.5 µl
Restriction enzyme 2	0.5 µl
Plasmid DNA	10 µl
Total volume	20 µl

Digestion was done at 37°C for 1 h and the 20 µl was loaded on to a 1% agarose gel for electrophoresis.

2.3.8. Conjugative transfer of plasmid DNA from *E. coli* cells to *C. difficile*

The conjugative transfer of plasmid from *E. coli* cells to *C. difficile* was performed using a protocol modified from (Purdy et al, 2002). Plasmid DNA was transformed into chemically competent CA434 *E. coli* cells (see Section 2.3.1 above), and grown on LB agar plates with 15 µg/ml chloramphenicol. Transformed colonies were picked from the plate and grown in LB broth with 15 µg/ml chloramphenicol, overnight. At the same time an overnight culture of *C. difficile* cells was grown anaerobically at 37°C in TY Broth.

1 ml of the overnight culture of transformed *E. coli* was centrifuged at 4000 × g for 2 min and most of the supernatant was aspirated out gently. The pellet, along with the remaining supernatant was brought into the anaerobic chamber and mixed gently with 200 µl of overnight *C. difficile* cells. This suspension was then deposited in 20 µl drops on a pre-reduced BHI agar plate. The plate was incubated undisturbed for 24 h in an anaerobic chamber at 37°C. 600 µl of TY broth was then added to the plate and cells were gently scraped off the surface of BHI agar using a sterile spreader. 100 µl of these cells were then spread on BHI agar plates containing 15 µg/ml thiamphenicol and 250 µg/ml cycloserine, to inhibit growth of unconjugated *C. difficile* cells and *E. coli* cells, respectively. After overnight growth at 37°C, a few colonies from these plates were then streaked on to individual BHI agar plates containing 15 µg/ml thiamphenicol and 250 µg/ml cycloserine, to make sure growth of all unconjugated *C. difficile* cells and *E. coli* cells had been inhibited.

2.4. Analysis of proteins

2.4.1. Low pH glycine extraction of surface proteins

Non-covalently linked surface proteins on the cell wall of *C. difficile* were extracted and solubilised using 0.2 M glycine (pH 2.2). Cells were harvested, either at the mid-logarithmic phase or stationary phase, by centrifuging at 20,000 × g for 2 min. The pellet was then resuspended to an OD_{600} of 50 in 0.2 M glycine (pH 2.2). The suspension was mixed at room temperature for 15-20 min followed by centrifugation at 20,000 × g for 2 min. The supernatant, containing the solubilised cell surface proteins, was retained. The supernatant was neutralised using 2 M Tris (Sigma) and either visualised using SDS PAGE or stored at −20°C.

2.4.2. SDS PAGE

Proteins were separated in a polyacrylamide gel and visualised by coomassie-blue staining (Laemmli, 1970). The proteins were treated using SDS and β-mercaptoethanol and separated on a 12% polyacrylamide gel. A discontinuous gel was formed out of buffers made up of two compositions (see Table 2.4). The gel was placed in SDS-PAGE running buffer (Table 2.4). Electrophoresis was carried out in a Mini-PROTEAN Tetra Cell System (Bio-Rad) at 200 V for 45 min. Gels were stained using coomassie blue gel staining solution (Table 2.4).

2.5. Electron microscopy

2.5.1. EM grid preparation

For negative staining and sample optimisation, a 300-mesh copper grid (EM Resolutions) was coated with a thin layer of amorphous carbon film. For cryo-EM, 300-mesh copper Quantifoil 2/2 grids (EM Resolutions) were used. Quantifoil 2/2 grids are pre-coated with 10 – 12 nm thick amorphous carbon support film containing 2 µm diameter holes, 2 µm apart. However, due to residues from the manufacturing process, the carbon support film is about 20 nm thick. The residuals did not result in any detrimental effects on image quality. Optimisations were also performed on Holey and Lacey carbon 300-mesh copper grids (Agar Scientific), with best results obtained from Quantifoil grids.

2.5.1.1. Evaporating carbon and preparing support film

For a continuous carbon support film, a sharpened 6 mm high-purity carbon rod (Agar Scientific) was placed between two high-voltage electrodes in the evacuation chamber of a Cressington 208 Carbon Coater. A freshly cleaved piece of mica, is placed under the carbon rod, on a fresh piece of white filter paper. The carbon is evaporated under high vacuum ($< 10^{-5}$ mbar or lower) by increasing the high-tension (HT) slowly until the carbon rod's sharpened end starts glowing white. The HT is increased slowly just shy of the point when the carbon rod starts to sputter, and then is turned down. This process is repeated in short bursts of a few seconds. The thickness of carbon being deposited can be measured using a thickness monitor or visually estimated based on the darkening of the filter paper. The carbon was aged for 48 h before transferring it to copper grids.

A Pyrex tank was filled to the brim with distilled water and a piece of filter paper (Whatman no. 1) is placed at the bottom on a raised platform. The copper grids were sonicated for 10 s in acetone, followed by 10 s in ethanol, and 10 s in distilled water. The grids were dried on a filter paper and then placed on the submerged filter paper. The carbon-coated mica sheet was then dipped through the surface of the distilled water by inserting the sheet gently at a 20 – 30° angle to the surface of the water. The water was then drained out from the tank by using a syringe and the carbon film was guided along so it gets deposited on the copper grids. The grids were then dried at room temperature before using it for negative staining.

2.5.2. Negative staining for microscopic examination

2.5.2.1. Preparing 2% (w/v) Uranyl acetate

0.1 g of uranyl acetate powder was added to 5 ml boiling-hot distilled water and stirred for 10 min. The solution was filtered through a 0.22 µm Minisart filter unit (GE Healthcare) into an aluminium foil-covered tube. The solution is stored in the dark at 4°C and is stable for up to a week.

2.5.2.2. Preparing 0.75% (w/v) Uranyl formate

0.035 g of uranyl formate (Polysciences Inc.) was added to 5 ml of boiling distilled water and stirred for 5 min. A few drops (~8 µl) of 5 M NaOH were slowly added until a colour change occured and the residual powder dissolved completely. The solution was stirred further for 5 min and then filtered using a 0.22 µm Minisart filter unit (GE Healthcare). The solution is stored in the dark at 4°C and is stable for up to a week.

2.5.2.3. Negative staining

Carbon-coated grids are glow-discharged, (carbon-side facing towards the plasma) in a glow-discharge unit, in air at ~10^{-2} mbar with atmospheric air, at 35 mA current in a Cressington 208 Carbon Coater, for 30 – 60 s. 5 µl of sample is then added to the carbon side of the grid and allowed to adsorb for 1 – 2 min. The excess sample is blotted away from the edge of the grid using filter paper (Whatman no. 1). The grid is then rinsed in a 50 µl drop of distilled water on a clean surface of Parafilm M (Bemis Company, Inc.) and blotted from the edge. The grid is again rinsed, this time in 50 µl drop of negative stain and blotted dry using filter paper. Finally the grid is floated on a fresh 50 µl drop of negative stain and stained for 20 s and blotted dry using filter paper and a vacuum drier.

2.5.3. Plunge freezing and vitrification

The freezing process was done using a Vitrobot Mark III (FEI), which partially automates the process of cryo-grid preparation, following the protocol of (Iancu et al, 2006). The ethane was condensed into liquid in a pre-cooled Vitrobot cup with liquid nitrogen. The liquid ethane was cooled further until the temperature was below −170°C, as measured using a temperature probe. The climate controlled chamber was adjusted to 22°C and 100% humidity, while fresh blotting paper (Whatman no. 1) was left in the chamber for at least 15 min to equilibrate. The EM grid was then glow-discharged, with the carbon film side towards the plasma, for 1 min at 10 mA and between 10^{-2} and 10^{-4} mbar air pressure. The grid was loaded on to the Vitrobot tweezers and fixed on the plunger hook. 2.5 µl sample was added on to the glow-discharged side of the grid, the tweezers were withdrawn into the climate controlled chamber and left to adsorb in the climate controlled chamber for at least 1 min. The following setting was used, in general, but was adjusted to specific sample needs.

Vitrobot (Mark III)			
Wait time	0 s	Blot offset	-3 mm
Blot time	30 s	Drain time	1 s

The sample was plunge-frozen and transferred into a cryo-grid holder and stored in liquid nitrogen.

2.5.4. Electron microscopy

2.5.4.1. Microscopes used

Samples were screened in either a Philips CM100 TEM operating at 100 keV with a LaB6 crystal thermionic electron source, Philips CM200 TEM operating at 200 keV with a tungsten FEG emission gun, or F12 Tecnai Spirit (FEI) TEM operating at 120 keV with a tungsten thermionic source. Images were recorded on a Gatan 1k ×

1k Multiscan 794 CCD camera. Images for crystallography were recorded at 32,000× magnification with 1.5 – 4 µm underfocus. For 3D structure determination the grid holder was tilted using a goniometer to collect tilt series between −60° to 60° with 10 – 20° steps.

For cryo-EM, FEI Tecnai F20 microscope operating at 200 keV with tungsten FEG emission source was used. The sample was scanned using a Gatan Orius 1k × 1k CCD camera and images were recorded on 4k × 4k Falcon 2 direct electron detector (FEI).

2.5.4.2. Low-dose imaging

The lens parameters were set up to shift between the Search, Focus and Exposure modes. In Search mode the grid was viewed at a magnification around 500× to keep electron dose very low, and at a high defocus to increase contrast. Search mode was used to scan the grid and choose areas of interest to take images. When an area of interest was chosen, the mode was shifted to focus, where the beam is shifted to a defined location off the sample and magnification switches to around 100,000×. This mode is used to primarily adjust the focus and set a defined defocus, while avoiding exposing the area of interest to electrons. Finally when the focus has been defined, the final image is taken in the Exposure mode, where the beam shifts to the area of interest and exposes it, while at the same time recording the image on the detector. The magnification and electron dose at this mode is pre-defined and images were taken at around 52,000×. Low-dose imaging improves contrast in images significantly when used on negatively stained samples (described in Appendix 9)

2.5.4.3. Cryo-EM of ice-embedded samples

The Gatan 626 cryo-transfer holder was pre-pumped and dried in a pumping station. The holder was then loaded in to a Gatan C1002 cryostage workstation and cooled to less than −185°C with liquid nitrogen. The temperature was monitored using an attached probe. Grids were prepared using the protocol described in Section 2.5.3 and transferred from the cryo-storage box onto the pre-cooled cryo-transfer holder on the cryo-stage. The grid was covered with the cryo-shield contamination protector sleeve and swiftly taken out of the cryo-stage and inserted into the pre-turbo pumped microscope compustage. Swift insertion into the microscope prevents the contamination of the grid and the holder with ice and ensures the grid does not warm up to temperatures above −165°C, when the ice starts to devitrify (Dubochet et al, 1982).

2.6. Data processing for 2D crystals

Image processing for electron crystallography was done using MRC suite of programmes (Crowther et al, 1996; Unwin & Henderson, 1975; Henderson et al, 1990; Valpuesta et al, 1994). Software system called *2dx* utilises MRC programmes and programmes designed specifically for *2dx*, providing a user-friendly, platform-independent GUI for processing images taken of 2D crystals (Gipson et al, 2007b). A brief description of the general process employed is given here. The MRC and CCP4 programmes in the description have been italicised in capitals.

2.6.1. Fast Fourier transform (FFT)

In the first step a Fourier transform of the digital image is performed. In case of a 2D crystal the repeating units produce a sharp amplitude peak in the FFT. The FFT in *2dx* was generated using *FFTTRANS.* The cut-off frequency in the FFT, which represents the resolution limit of the digital image is called the *Nyquist frequency* and is given by 1/(2 × pixel spacing). Thus if each pixel on the image represents 2 Å, the Nyquist frequency is ¼ $Å^{-1}$, thereby giving resolution limit of image at 4 Å.

2.6.2. Indexing and extraction of amplitude and phase

In *2dx* the FFT can be visualised and various interactive steps can be performed. The visible *Thon* rings are used to define the defocus using Define Defocus mode. A Spot Selection mode and a Lattice Refinement mode allows for the selection of the lattice vectors and spot list for later *unbending.* The peaks in the FFT are assigned miller indices *h* and *k.* To define the reciprocal lattice vectors, *0,1* and *1,0* indices are manually selected. The rest of the indexing is generated automatically by *FINDLAT*. All the above interactive steps on the FFT can also be performed using independent scripts. For example, *CTFFIND3* is used to calculate the defocus and astigmatism, *GETDEFTILT* calculates the tilt in the image by drawing a plane through the defoci in 7 × 7 regions in the image, *GETLATTICE* calculates the lattice vectors, whereas *GETSPOT1* generates a spot list which is used for subsequent unbending.

The unbending process was carried out iteratively and it involved the following steps (Figure 2.1). In the first step, noise and amorphous information in the image were removed by masking the FFT to isolate only the diffraction spots, which were then reverse Fourier transformed. The resulting image of the crystalline information was then used to unbend the image. By selecting a small reference motif that appears to have a satisfactory crystallinity, and cross-correlating that to the entire image, a displacement map was generated showing the distortions in the original image, compared to the selected motif. These distortions were then corrected by interpolation. A new reference motif was then used to unbend again, generating an improved image. Such iterative unbending improves the image quality dramatically. The MRC programmes involved in the process of iterative unbending and their descriptions are detailed in Table 2.7.

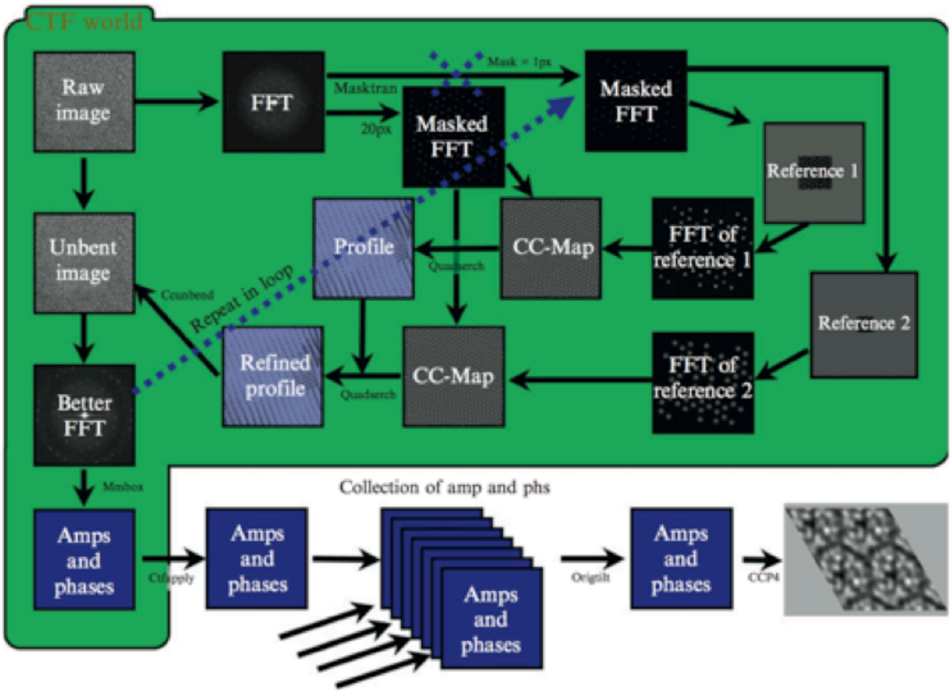


Figure 2.1 Unbending

The process of unbending performed in the 2dx imaging suite is shown above, beginning with the raw image of the crystal on the top left. The FFT is generated from the image and masked to remove the amorphous information and noise in the image. The improved image is used to generate a reference for unbending, generating a better image of the crystal. CC-map refers to the cross-correlation map (Gipson et al, 2007b). This process is repeated several times, with appropriate parameters, to generate a better FFT, giving the amplitude and phase values for each *h*,*k* spot. The MRC programs performing each process are represented on the arrows. The description of each process is in Table 2.7. The microscope's CTF modulates the whole green area. The amplitude and phase obtained at the end of the process is finally corrected, following which several other amplitude and phase values are merged along a defined phase origin. The crystallographic programme *CCP4* is used to generate the final map from the merged amplitude and phase values. Figure with permission from (Gipson et al, 2007a).

Table 2.7 Table showing the sequence of steps that are carried out by the MRC programmes to unbend an image

	Unbending step	**MRC Programme**
(a)	FFT of the original image is calculated	*FFTRANS*
(b)	A suitable mask is placed on the FFT, with appropriate hole size to isolate the peaks in the FFT selected by *getspot1*	*MASKTRAN*
(c)	A filtered image is then generated from the masked FFT by inverse Fourier transform of the masked FFT. This eliminates objects in the image that are not crystalline	*FFTRANS*
(d)	A small reference motif (reference box) is then boxed out in the centre of the filtered image in (c) or another region deemed to have best crystalline order.	*BOXIMAGE*
(e)	A FFT of this reference motif is generated. This FFT is taken as the "ideal" FFT.	*FFTRANS*
(f)	The product of the FFTs in (a) and (e) is taken, which is a cross-correlation function between the reference motif and the original image. i.e. (a × e)	*TWOFILE*
(g)	The deviation positions in the cross-correlation peaks from their expected positions are represented using displacement vectors. This shows the "bent nature" of the crystal compared to the reference box selected in (d)	*QUADSERCH*
(h)	The image is "unbent" by applying local corrections determined from the displacement vectors generated in (g) to the original image before (a).	*CCUNBEND*
(i)	The FFT is generated of the unbent image. This FFT is sharper and shows improved SNR.	*FFTRANS*
(j)	A mask is applied to (i) and the process is repeated from step (b) to (i) in Unbend II.	**Repeat steps (b) to (i)**
(k)	List of amplitude and phase values for each spot is generated which is subsequently used to generate a map.	*MMBOX*

2.6.3. CTF Correction

Other systematic image defects such as the CTF needs correcting as well. As discussed in Section 1.1.4, the Fourier components within adjacent *Thon rings* correspond to the spatial frequencies where the $sin\gamma$ of the CTF crosses zero and reverses in sign. These will have alternate contrasts and the phases flipped by 180°. Also the amplitudes drop for the Fourier components in the vicinity of the zeros of the CTF. The phase values of each Fourier components are corrected in the region of the FFT where the sign of $sin\gamma$ changes, by multiplying the FFT with the FFT of the microscope's point spread function (PSF), i.e. the CTF (Equation 2.1).

$$\text{Image}_{\text{corrected}} = \mathcal{F}^{-1}\{\mathcal{F}\{\text{Image}\}\mathcal{F}\{PSF\}\} = \mathcal{F}^{-1}\{\mathcal{F}\{\text{Image}\}\,CTF\} \qquad 2.1$$

When several images of different defoci are merged together, the information lost in the vicinity of the CTF minima are restored by redundancy. However, dividing the above equation with $CTF^2 + 1/SNR$ (Equation 2.2) corrects the amplitude contrast, while a *Weiner filter* takes into account the SNR (see also Figure 2.2).

$$\text{Image}_{\text{corrected}} = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{\text{Image}\}\,CTF}{CTF^2 + \frac{1}{SNR}}\right\} \qquad 2.2$$

Programme *CTFAPPLYK* is used to correct the CTF at the end of the unbending, where data directly from *MMBOX* is fed into *ORIGTILT* (Section 2.6.5) and processed. This can be corrected by phase flipping, or CTF multiplication, or applying a *Weiner filter*. For the current project phase flipping and amplitude correction was performed.

2.6.4. Space group and 2D density maps

The symmetry and phase origin for the images were determined using *ALLSPACE,* which uses the lattice parameters and final amplitude and phase values to determine which of the 17 plane groups for a 2D crystal is possible for the given image. The plane groups were assigned based on relationships between the phases of symmetry-related diffraction spots. The difference in the phases of symmetry-related spots in a given image is listed as the phase residual. The smaller overall phase residual value for a particular symmetry operation indicates the possible space group of the 2D crystal.

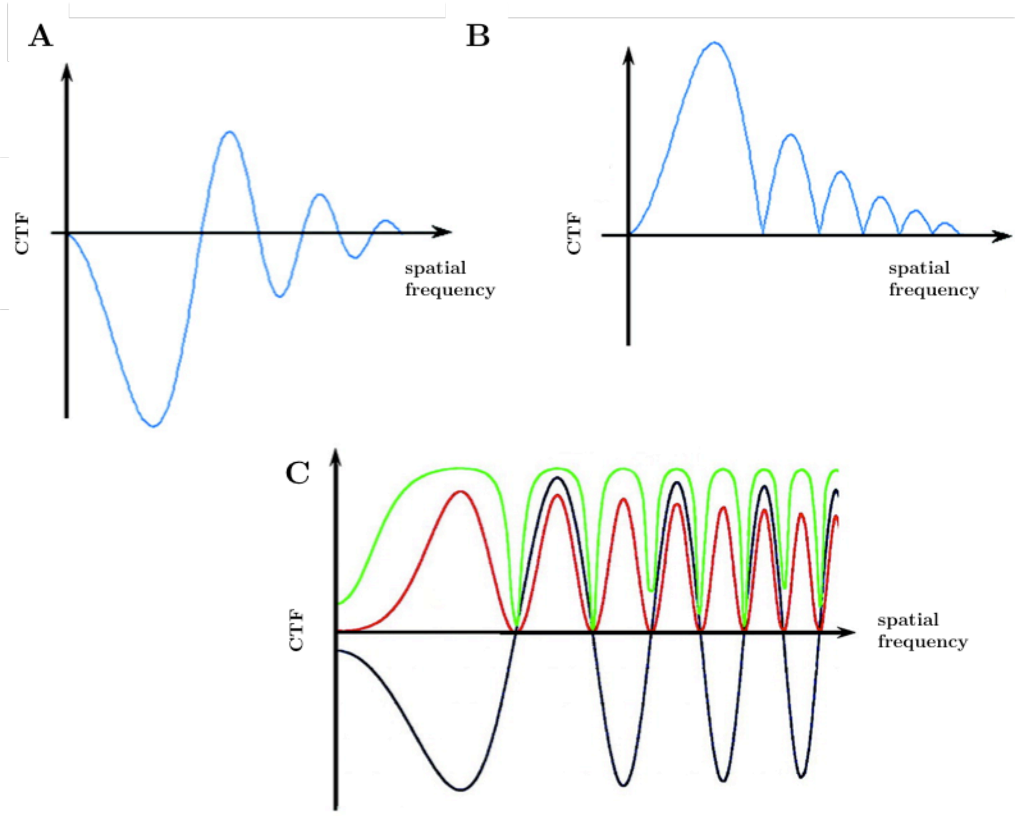


Figure 2.2 CTF correction

A to **C**: The effects of CTF correction on amplitude and phase. **A** shows the uncorrected CTF. **B** represents the square of the CTF after rescaling. **C**: The black line represents the uncorrected CTF, which contributed by the amplitude and phase. The squared and scaled value of the phase contrast is shown in the red line. The green line represents the CTF after the amplitude and phase is corrected. Figure taken with permission from (Orlova & Saibil, 2011), copyrighted by the American Chemical Society.

The extracted amplitude and phases from several images of the same sample are then merged together to those of a reference image. An average amplitude and phase file is created. The SNR of the images is improved by taking into account the redundancy of symmetry related spots.

For centro-symmetric plane groups, phases are constrained to 0° or 180° by *CENTRIC2*. A final map is generated taking into account space group symmetry and *Friedel relationship*. The crystallographic programme *CCP4* then plots the EM density ρ in x,y coordinates, from the structure factors, F_{hk} for each h and k indices, by inverse Fourier transform (Equation 2.3).

$$\rho(x,y) = \sum_{h,k} |F_{hk}|\, e^{-2\pi i(hx+ky-\phi_{hk})} \qquad 2.3$$

2.6.5. Common phase origin and *ORIGTILT*

After extracting the amplitude and phase information the ultimate goal is to create a 3D data set that can be used to create a 3D density map in real space. In order to achieve this objective, two problems need to be solved. The first problem is to establish correctly the exact tilt angle and orientation of the tilt axis of the sample, relative to the crystal axes of the untilted sample. The second problem is to determine a common phase origin.

The amplitudes of the individual images are independent of each other, however the phases are not. To establish a correct relationship between the phases extracted from the different images, one needs to calculate the Fourier transform relative to a common coordinate system with an origin called the phase origin. The calculated phases from the images change based on the origin selected. This is due to the *shift theorem* in Fourier transforms.

For an arbitrary image $I(\mathbf{x})$, when the phase origin (given by vector $\mathbf{x}$) is shifted by $\mathbf{a}$, the phases of the image effectively shifts by $2\pi\, \mathbf{s}\cdot\mathbf{a}$ (Equation 2.4). Therefore, the phases depend not just on the shift, but also upon the spatial frequency, $\mathbf{s}$, at which the Fourier transform is being calculated.

$$\mathcal{F}\{I(\mathbf{x}\text{-}\mathbf{a})\}(\mathbf{s})=\mathcal{F}\{I(\mathbf{x})\}(\mathbf{s})\,e^{-i\,2\pi\,\mathbf{s}\cdot\mathbf{a}} \qquad 2.4$$

Often, one of the symmetry axes of the crystal is taken as the phase origin. This helps in later 3D merging, as described below.

Central section theorem (Section 1.4.1.2) defines the Fourier transform of a tilted projection, as a section through the 3D Fourier transform of the same sample. Therefore two Fourier transforms at different tilt angles form two central sections that intersect along a common line (Figure 2.3). The common phase origin can thus be defined by requiring all respective spots within a defined z^* window, along this line, to have the same phase. Each central section is refined and aligned in this manner.

As is clear from Figure 2.3, the geometry of the lattice is distorted at higher tilt angles. The tilt parameters can be calculated from the geometry of the lattice distortions in the FFT of individual images relative to the untilted reference lattice. Since the geometric distortions of the lattice will be very small at low tilt angles, the accuracy of tilt parameter estimations increases with higher tilt angles. At lower tilts however, the tilt parameters can be calculated using C*TFFIND3*, which looks at the defocus variation across the image in 7 × 7 locations and fits a plane across the values. The tilt angle of this plane is an approximate tilt angle of the grid in the microscope. The defocus value stays constant across the tilt axis of the image, which remains perpendicular to the electron beam.

For a tilted sample, the defocus varies across the image. As is clear from Figure 2.4, the part of the sample that is lower down in the microscope is more underfocussed in the image, compared to the part of the sample higher up. The z^* signs for structure factors change, based on the side of the tilt axis the structure factors are located.

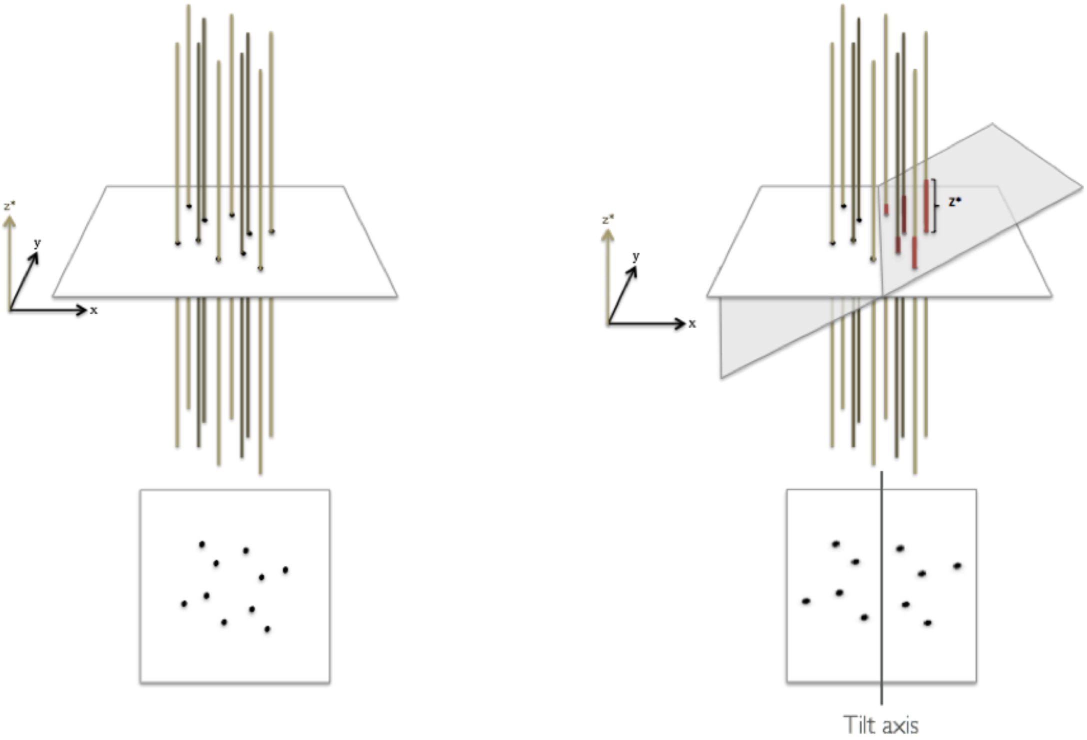


Figure 2.3 Central section through lattice lines

The intersecting planes (Ewald planes) represent the central sections, which pass through the centre of the 3D transform. The Fourier transform of the projection of a 2D crystal is a central section, with the same tilt parameters, through the 3D transform of the 2D crystal. The untilted central section is shown on the left panel with lattice lines (in shades of ochre) extending into the *z** direction. The diffraction pattern seen on the plane is shown below. The panel on the right shows a second plane intersecting the first at a high tilt angle. The two planes intersect along a common line. The lattice distortion seen in the tilted plane is shown below. The distortion can be used to estimate the tilt parameters of the crystal. The distance (shown in red), along the lattice line for each identical *h.k* spot, between the tilted plane and the untilted plane is given by the third index *z**. Several tilt angles are used to sample different regions along the lattice lines, generating spots with *h*,*k*,*z** indices.

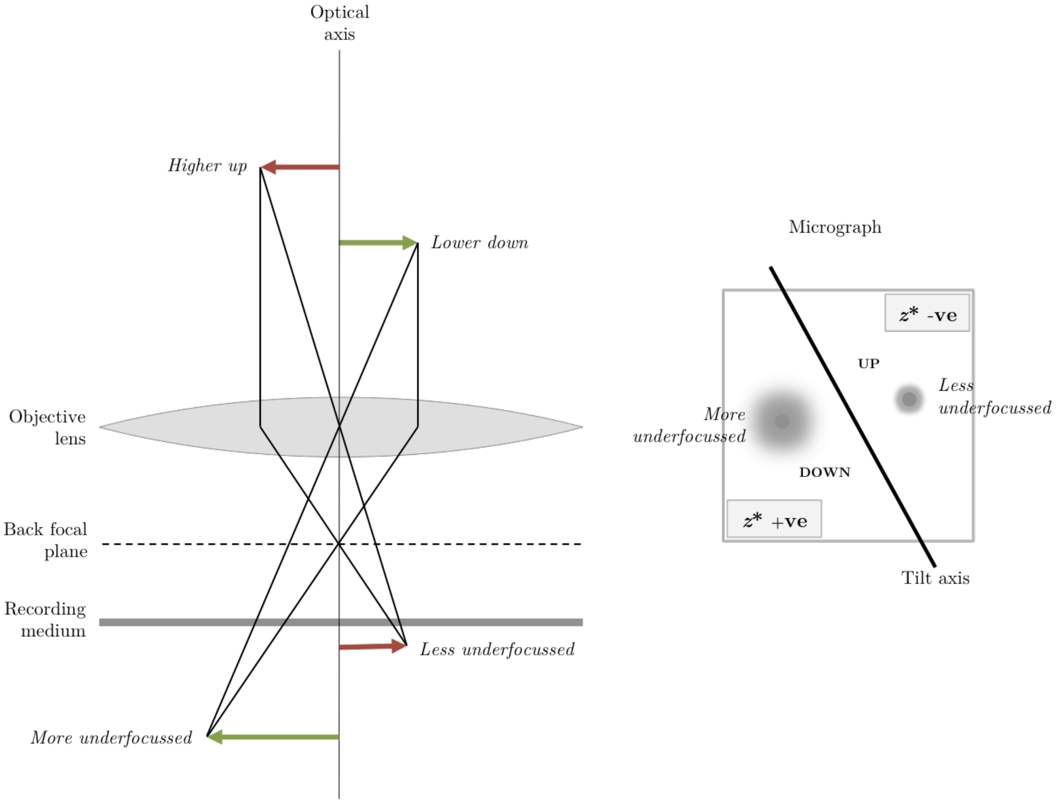


Figure 2.4 The underfocus and z^* sign is affected by the sample height in the microscope

Tilting the sample results in differing sample heights in the electron beam, which manifests as different levels of defocus in the image. In the figure on the left, one sample (red arrow) is higher up in the microscope than the other (green arrow). The green arrow is therefore more underfocussed in the final image detected on the recording medium. The micrograph appears as shown on the right, with the two samples projected on the same plane, but having different defocus values based on their respective height in the microscope. The sign of the z^* index from the micrograph follows the MRC convention of negative sign for higher up spots in the central section intersecting the lattice lines in the 3D transform.

In order to maintain the correct handedness of the final structure, the signs assigned to z^* need to correspond to the correct sample height in the microscope. The convention used in MRC programmes puts the signs of z^* as positive for spots on the central sections located below the untilted plane – corresponding to a lower sample height in the microscope (Figure 2.4).

As more data is merged by adding central sections in the reciprocal space at increasing tilt angles, a pattern in the amplitude and phase distribution along the *lattice lines* begins to emerge. Smooth curves are fitted along the amplitude and phase values with increasing z^*. As discussed earlier in Section 1.4.1.2, the intensity distribution along the lattice lines should theoretically vary smoothly. The average amplitude and phase values are plotted against increasing z^*. The initial estimates of tilt angle and phase origin are further refined using these fitted curves. The tilt parameters are altered by *ORIGTILT*, within the given z^* window, to fit the data closer to the extrapolated phase values along the lattice lines. The refined values can be used to establish better fitting curves and the process is repeated iteratively. A final value of overall phase residual provides an indication of the quality of the merged dataset.

2.6.6. 3D density map and structural interpretation

The complex structure factors $\boldsymbol{F}_{hkl}$ are obtained by sampling the continuous lattice lines at regular intervals in the z^* direction. In other words, the lattice line indicated by index *h,k* is sampled computationally for the amplitude and phase value at a pre-set interval of c^*, where $c^* = 1/c$ and c is the real-space unit cell length in z-axis. When the experimental value of c is not known, it is approximated to 1.5 to 2-times the best estimate of crystal thickness. This gives rise to a complete set of *hkl* data from the hkz^*. The *hkl* dataset was then converted to a binary *mtz* file, compatible with *CCP4*.

Now the discreet set of complex structure factor values $\boldsymbol{F}_{hkl}$ in an *mtz* file were used to calculate the real-space EM density $\rho(x,y,z)$ of the sample (using *CCP4* *programme*) by plotting the inverse Fourier transform in three dimensions, shown in Equation 2.5; multiplying a weighting factor (figure of merit = cosine of phase error) for each structure factors (*FFTBIG*). The final volume map was extended to show 2 × 2 unit cells by *EXTEND*. The contour map showing the 2D projections was plotted using *NPO*.

$$\rho(x,y,z) = \sum_{h,k,l} |F_{hkl}|\, e^{-2\pi i(hx+ky+lz-\phi_{hkl})} \tag{2.5}$$

2.6.6.1. 0,0,*l* estimation

Missing structure factors along 0,0,*l* can be estimated by calculating the Fourier transform of the real-space density sampled at discreet planes as it varies in Z-sections of the calculated volume map. The density at each section is plotted against the distance from the centre of the structure. The plot showing density as a function of unit cell height is then subjected to a Fourier transform. The amplitude values of the transform were scaled against the existing amplitude values of corresponding *h*,*k*,0 indices. The scaled amplitude and phase values of the 1st, 2nd, and 3rd Fourier components were added to the data from *MMBOX* at (0,0,1), (0,0,2), and (0,0,3) indices. The volume map was re-calculated.

2.6.6.2. Threshold estimation for 3D volume

The volume of the 3D model from negatively stained data was approximately calculated by calculating the theoretical volume expected from the proteins forming the unit cells. The volume of a single protein was calculated based on an average protein density of 1.35 g cm^{-3} and its mass calculated from the size of the protein. A protein size of 1 kDa equates to 1.66 × 10^{-21} grams. The volume was calculated as the mass over density. Since the generated 3D model contains 2 × 2 unit cells the volume was multiplied based on the number of asymmetric units present in the 3D model. This volume was displayed in software *Chimera* (Pettersen et al, 2004).

2.6.6.3. Estimating temperature factor or B-factor

As discussed in Section 1.5.1, for cryo-EM images the intensity of structure factors is expected to drop at higher resolution. A temperature factor or B-factor was added to the existing data to correct the fading of amplitudes as a function of resolution. The amplitudes were corrected using *SCALIMAMP3D* for individual images before merging the images and producing the final map (Havelka et al, 1995). The B-factor was estimated by using the reference structure of bacteriorhodopsin (bR) (Henderson et al, 1990).

Applying an inverse of the temperature factor B_{xy} to minimise the sum in the following, rescales data in each image.

$$\sum F_{ref} - F_{obs} e^{0.25 B_{xy}(X^2+Y^2)} \qquad 2.6$$

where F_{ref} is the average amplitude of the reference data (in this case, bR).

The values less than 0 and greater than 1000 were truncated to 0 and 1000, since it is unlikely that the data was more ordered than the reference and more disordered than a $B > 1000$.

2.6.7. Difference maps

Difference maps were calculated to highlight the structural difference between two EM maps. The structure factors from *SCALIMPAMP3D*, with correct temperature factors, were used to calculate the difference in Fourier space (Figure 2.5). The amplitude and phase values of the subtracted vectors were used to calculate a new density map, this time highlighting the difference in 2D and 3D. The data, before calculating the difference, were scaled against each other to get a root mean square value, which was then multiplied to the amplitude and phase values (*SCALEMAP*).

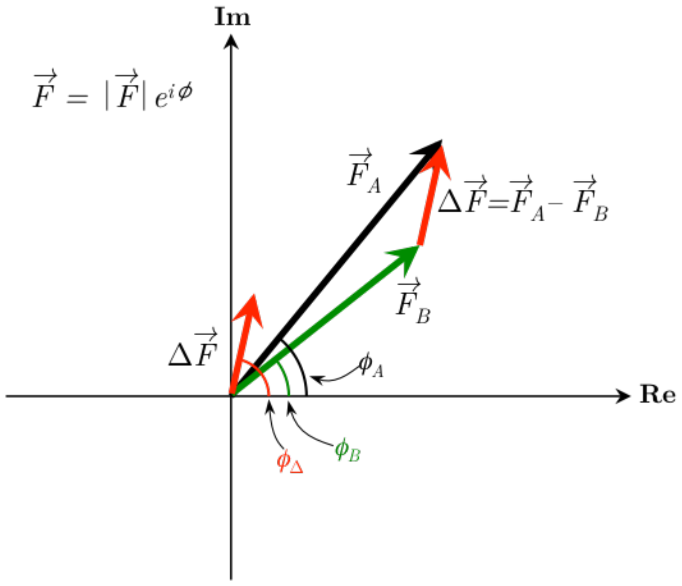


Figure 2.5 Difference between structure factors used to calculate phase errors and difference map

To calculate a difference map, the structure factors were subtracted in Fourier space after the two datasets were scaled together. Here the structure factor vector $\boldsymbol{F}_A$ (thick black arrow) is subtracted from $\boldsymbol{F}_B$ (green arrow). $\Delta \boldsymbol{F}$ (red arrow) is the resulting structure factor with amplitude $|\Delta \boldsymbol{F}|$ and phase ϕ_Δ (in red), which has been translated to the origin to show the phase. The difference in phase between $\boldsymbol{F}_A$ and $\boldsymbol{F}_B$ is the phase of $\Delta \boldsymbol{F}$ (red arrow), which has a relatively large value compared to ϕ_{A} and ϕ_{B}. Due to the relationship between amplitude and phase in the complex plane, small changes in amplitude can alter residual phases rather dramatically.

Chapter 3 - Native S-layer structure of *C. difficile* by electron crystallography

3.1. Introduction

The intact S-layer on the surface of cells is difficult to image by TEM, as most bacterial cells are dense structures with a thickness ranging in microns. Structural analysis of cell surface structures so far has involved imaging structures by electron tomography on thin sections or on structures present along the edges of cells, where the thickness of the cell body is substantially reduced, e.g. imaging done to visualize flagella motor proteins (Chen et al, 2011). Structural analysis of S-layers has been predominantly done by freeze-fracturing and metal shadowing techniques (Wildhaber & Baumeister, 1987; Cerquetti et al, 2000), or more recently on thin cell envelope extensions in *Caulobacter crescentus* stalks (Bharat et al, 2017). Scanning probe microscopy has also been used to image the surface in low resolution (Cheong et al, 1993; Günther et al, 2014).

In contrast, high-resolution structures reported using electron crystallography are mainly from 2D crystals of purified proteins reconstituted *in vitro* on lipids (Kühlbrandt et al, 1994). Bacterial S-layers, which are not known to assemble *in vivo* on lipid membranes, do not readily crystallise on lipid bilayers. For the same reason, crystallising S-layer proteins into 3D crystals for X-ray crystallography is difficult. Various attempts at circumventing the problem have been made, including the use of nanobodies (Baranova et al, 2012) and structural modifications, to allow crystallisation in the third dimension (Fagan et al, 2009; Usenik et al, 2017). However, for structures like the S-layer, *in vitro* 3D crystals may not resemble their *in situ* 2D organisation.

A method was therefore needed to isolate the intact S-layer, retaining its native state, but lacking the bulk of the cell material. The technique developed for this project is described below.

3.2. Results

3.2.1. S-layer extraction by endolysin treatment

Several attempts were made to extract and isolate the S-layer from *C. difficile*, by sonication and French press, with little success. Enzyme-induced lysis is difficult in *C. difficile* owing to its resistance against lysozyme (Ho et al, 2014). Enzyme-induced cell lysis was carried out, instead, using prolonged exposure to a bacteriophage-derived endolysin enzyme. *C. difficile* phage ϕCD27 produces cell wall digesting endolysin (Mayer et al, 2011), which was recombinantly expressed and purified from *E. coli*. A novel method was devised to use this enzyme to partially digest the cell wall, in order to separate the cell protoplasm from the S-layer-cell wall complex. Considerable effort was put into improving and optimising this technique.

The *C. difficile* cells were harvested and resuspended in HEPES buffered saline before treating with endolysin to digest the cell wall partially (detailed methods described in Section 2.2.1). This dissociated the protoplasm along with the cell membrane (the *spheroplast*) from the S-layer/cell wall complex. The reaction when carried out in an osmotically buffered solution containing 500 mM sucrose, resulted in intact spheroplasts, which could then be harvested by centrifuging at a low speed of around 2000 × g. The S-layer/cell wall complex remained separated from the rest of the dense cell material. SDS PAGE analysis of the osmotically buffered reaction supernatant showed the presence of cell surface proteins, where the spheroplasts were separated after centrifugation (Figure 3.1). Incubation with endolysin for longer than an hour, however, resulted in the disintegration of the cell wall and loss of crystallinity in the S-layer. The incubation time was critical in order to obtain a well-ordered sample.

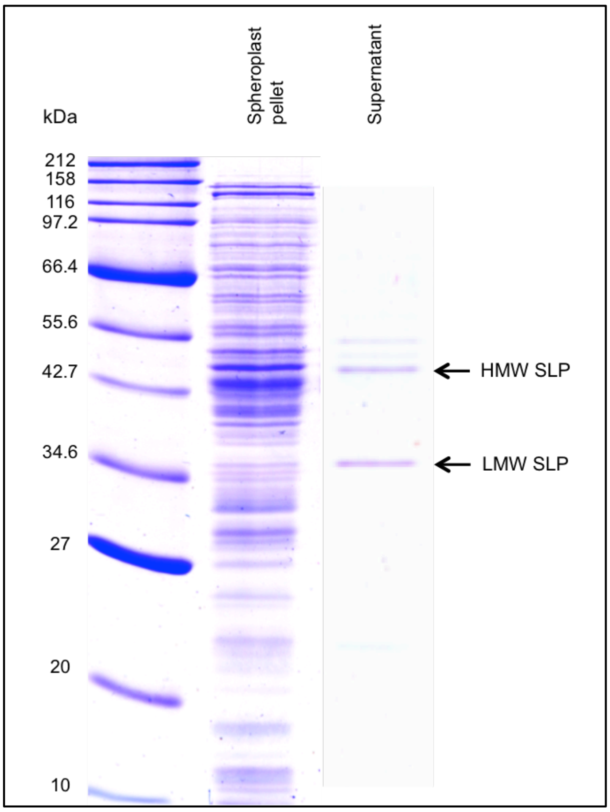


Figure 3.1 Endolysin extraction technique separates the S-layer from the protoplasm

SDS PAGE analysis of supernatant and pellet fractions of endolysin treated extract of R7404 cells from exponential phase, after centrifuging at 2000 × g for 2 min. The denser spheroplasts, generated with the treatment in 0.5 M sucrose containing buffer, separates after centrifugation to form the pellet. The spheroplast pellet shows the separated intracellular proteins on the gel. The supernatant containing the less dense S-layer/cell wall extracts show the presence of the S-layer proteins (HMW and LMW SLPs) and other CWPs.

Phase contrast light microscopy showed formation of spheroplasts in the endolysin reaction mixture (Figure 3.2). Figure 3.2 shows *C. difficile* cells in logarithmic phase of growth immediately after endolysin treatment and after incubation at 37°C for 30 min. Phase-dark spheroplasts were visible, along with intact whole cells of *C. difficile*. The S-layer/cell wall extract was not visible under light microscope, but was shown to be present using TEM (Figure 3.3B).

This protocol effectively generated a capsular structure, whose outline resembled that of a whole cell, which we named a '*ghost*'. The *ghosts* were around 4-6 µm long. They were, however, wider than a whole cell due to a lack of cellular material and flattening of the cylindrical structure upon drying during negative staining (Figure 3.3). Most importantly, these structures were thin, allowing phase contrast under TEM, and subsequent processing. This protocol also retained the crystalline order in the S-layer.

When imaged at high magnification, the extracts displayed two superimposed S-layer lattices (Figure 3.4). The two lattices rarely overlapped and were easily indexed from the FFT. The primary lattice chosen was that which stained the best and appeared to show sharp spots to the highest order. The projection maps were calculated (without applying symmetry) from both the primary and secondary lattices (Figure 3.4C & D) to check for consistency.

This technique was used to image several different S-layer types in different *C. difficile* strains, starting with R7404 as a proof of principle. EM image processing techniques described in Section 2.6 was used to generate the projection structure from micrographs obtained, throughout the project.

3.2.2. Ribotype 017

3.2.2.1. S-layer projection at 20 Å

Strain R7404 (ribotype 017) was used to obtain the 2D projection and 3D structure of the S-layer, using the endolysin extraction protocol.

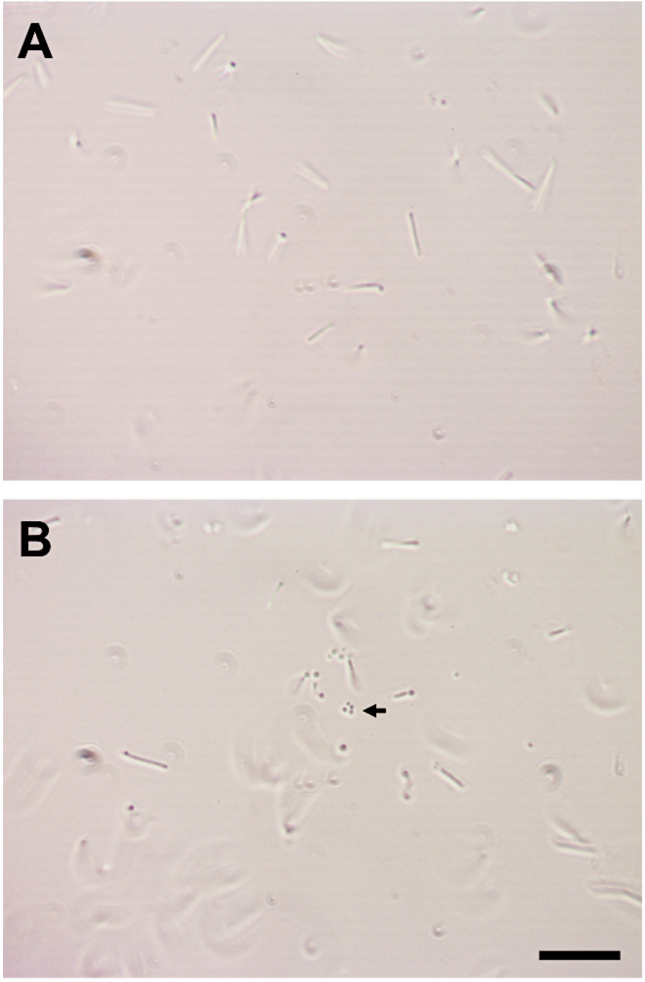


Figure 3.2 Endolysin extraction produces spheroplasts visible under phase contrast light microscope

Phase contrast light micrograph showing live *C. difficile* cells at 0 min in **A** and 30 min in **B**, after treatment with endolysin. The solution was osmotically buffered to preserve the spheroplasts, which appear as dense circles (indicated by arrow). The S-layer/cell wall extracts were not dense enough to produce phase contrast under a light microscope and only visible in TEM. Scale bar in **B** is 10 µm and applies to both panels.

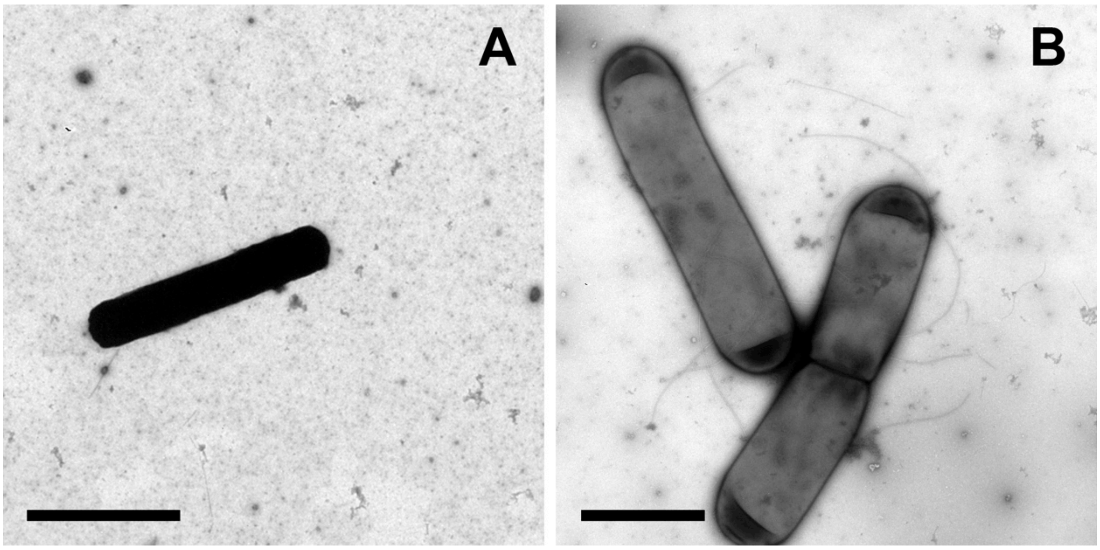


Figure 3.3 Whole *C. difficile* cells along with endolysin-treated extracts

A: A negatively stained untreated whole cell of strain R20291, which appears very electron dense. Scale bar: 5 µm. **B**: Endolysin treated *ghosts* of R20291 strain, which retain their overall cell shape but appear less dense due to the lack of intra-cellular components. Scale bar in A and B: 2 µm.

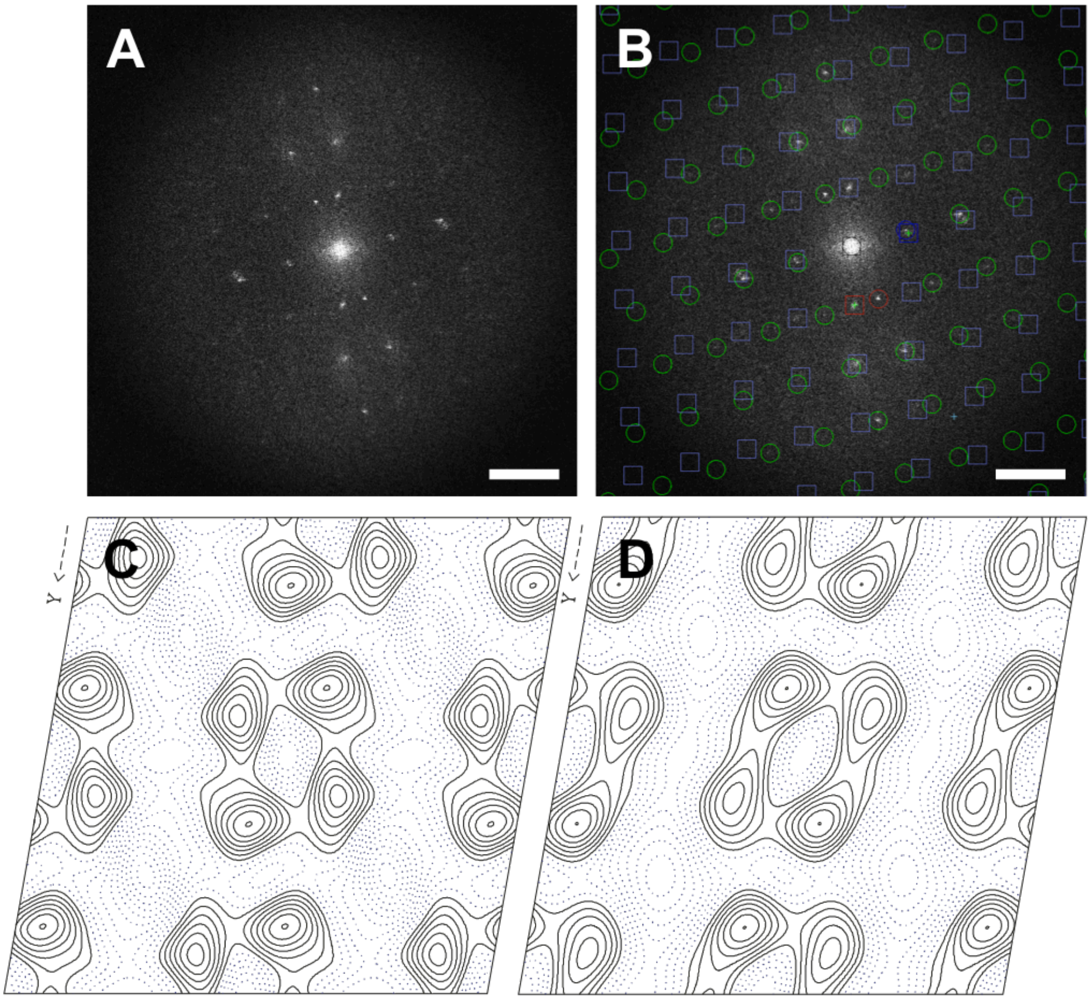


Figure 3.4 FFT and reconstructed projection map of an S-layer extract from R7404

A: The FFT of an untilted image of R7404 S-layer extract obtained by endolysin treatment, displaying the two superimposed lattices. **B**: The two lattices are highlighted. The primary lattice is shown in green circles and the secondary lattice is in blue squares. **C** and **D**: The projection map of 2 × 2 unit cells, reconstructed from the primary and secondary lattice, respectively. They both show opposite handedness as is clear when both maps are compared. Unit cell dimensions for both maps are $a = b = 80$ Å and $\gamma = 100°$ with $p2$ symmetry applied. The Y-axis is indicated. The scale bars in **A** and **B** represent 0.125 nm^{-1}.

The untilted 2D projection revealed unit cell dimensions of $a = b = 80$ Å ± 1 Å and unit cell angle of 100° ± 0.6°. Based on the symmetrised phase residuals calculated using programme ALLSPACE, an unambiguous $p2$-symmetry was suggested (Table 3.1). Figure 3.5 shows the resolution plot showing all reflections and a 2D projection map of 2 × 2 unit cells up to a resolution of 20 Å, from a single image, without any symmetry imposed. Figure 3.6 shows the resolution plot and the $p2$-symmetrised map after the merging of data from six crystals.

3.2.2.2. 3D model

2D projection images at various tilt angles ranging from +50° to −50° were merged to generate a 3D volume map. Total of 36 micrographs of individual crystals were used, generating 146 unique reflections. Resolution was truncated to 20 Å, as being the limit of negative stain. *ORIGTILT* was used to search for a common phase origin, and merge tilted and untilted data, using vertical window in z^* direction of 0.05 $Å^{-1}$.

A summary of statistics from the 3D data set is shown in Table 3.2. The overall phase residual for the dataset was 22.3°. Figure 3.7A to D shows the 3D volume map of 2 × 2 unit cells with $p2$ symmetry, with symmetry axis parallel to Z-axis. The surface of the volume was thresholded to a volume encompassing two subunits of SlpA protein in each unit cell. The volume was calculated from an average protein density (1.35 g/cm^3), and the molar mass of 2 × 86 kDa H/L complex. The z^* lattice line plot, up to a vertical resolution of 20 Å, are shown in Appendix 4 for this data set.

3.2.3. Ribotype 027

After establishing the structure of the S-layer from the ribotype 017, the structure of the S-layer of 027 was analysed. Strain R20291 (ribotype 027) as discussed in Chapter 1 is a hyper-virulent, clinically significant strain.

Table 3.1 Phase residual table for different plane groups for a representative untilted S-layer crystal of R7404

Plane Group[1]	**Phase residual (random = 90°)**	**Target residual**[2]	**Plane Group**[1]	**Phase residual (random = 90°)**	**Target residual**[2]
p1	20.8°		p22121	44.7°	25°
p2	27.9°*	30.1°	c222	56.7°	25°
p12b	74°	22.7°	p4	45.3°	24.7°
p12a	66.3°	23°	p422	55.4°	22.8°
p121b	58.2°	22.7°	p4212	54.6°	22.8°
p121a	56.8°	23°	p3	46.5°	20.8°
c12b	74°	22.7°	p312	57.5°	21.4°
c12a	66.3°	23°	p321	59.7°	22°
p222	56.7°	25°	p6	51.1°	23.9°
p2221b	71.2°	25°	p622	56.8°	22.3°
p2221a	55.8°	25°			

[1] a and b represent the respective symmetry axis for the plane group

[2] Target residual indicates the expected phase residual of each symmetry group based on the SNR of the spots from the concerned image

Symbols for possible plane groups: * = acceptable, ! = should be considered, ` = possibility

Table 3.2 Statistical summary of 3D data from R7404 crystals

3D reconstruction parameters	
Crystal plane group symmetry	*p*2
Crystal unit cell parameters	a = 80.0 Å, b = 80.0 Å, γ = 100.0°, c = 160.0 Å
Number of individual crystals	36
Range of defocus	7020 to 16200 nm
IQ range used for 3D refining	1 to 5
Tilt range used	0 to 51.2°
In-plane resolution cut-off	20 Å
Vertical resolution cut-off	20 Å
Number of observed reflections	828
Number of observed unique reflections	695
Overall phase residual	22.3°

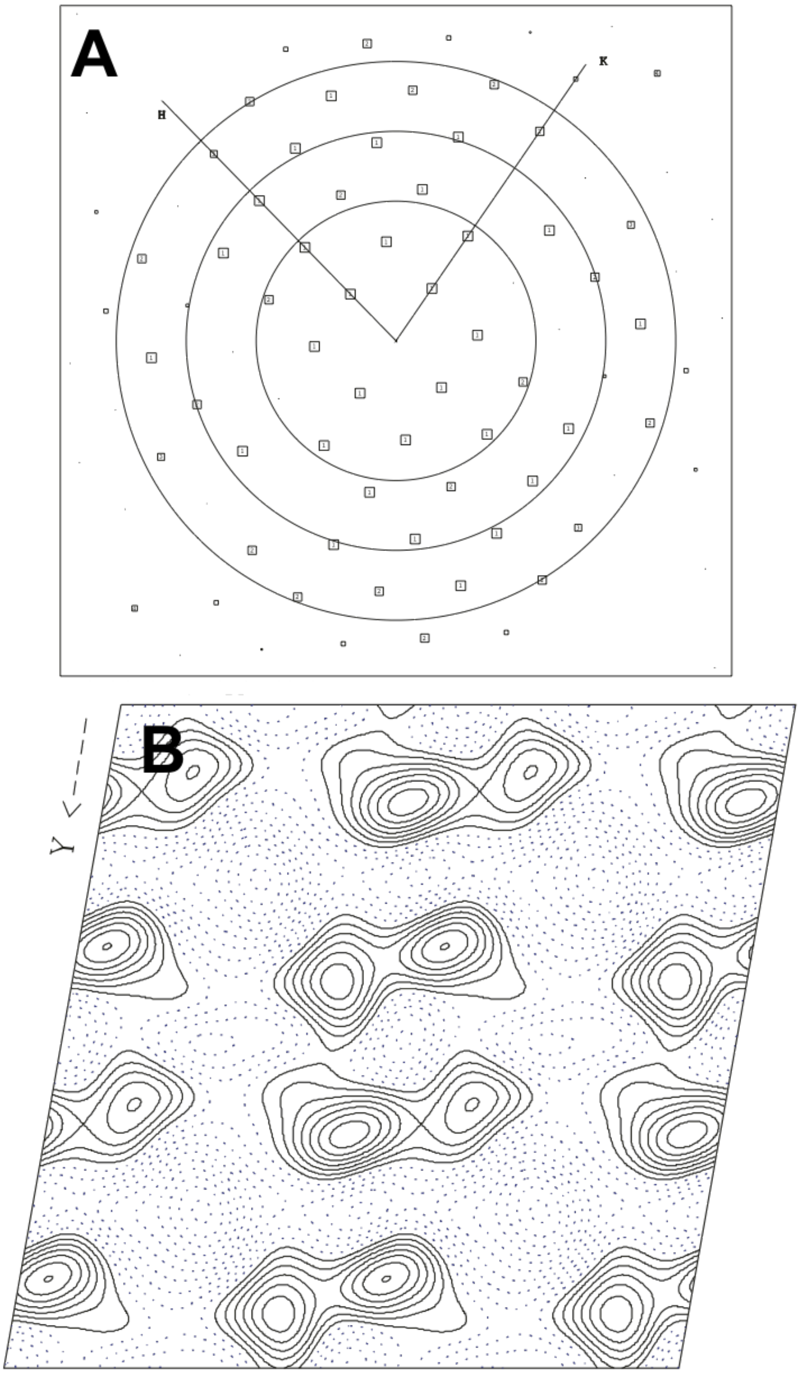


Figure 3.5 Resolution plot and unsymmetrised 2D projection of the S-layer of R7404

A: Resolution circle plot following image processing of an untilted R7404 S-layer extract with reflections indicated in squares. The number in each square represents the IQ of the respective reflection, with spots of highest signal-to-noise having an IQ of 1 (Henderson et al, 1986). The H and K axis are indicated. The resolution rings indicate 36, 24 and 18 Å (outward from centre of image) with Nyquist frequency of the plot at 15 Å. **B**: The projection structure from **A**, without any imposed symmetry. 2 × 2 unit cells are represented. Dimensions of the unit cell are $a = b = 80$ Å and $\gamma = 100°$ and the Y axis is indicated.

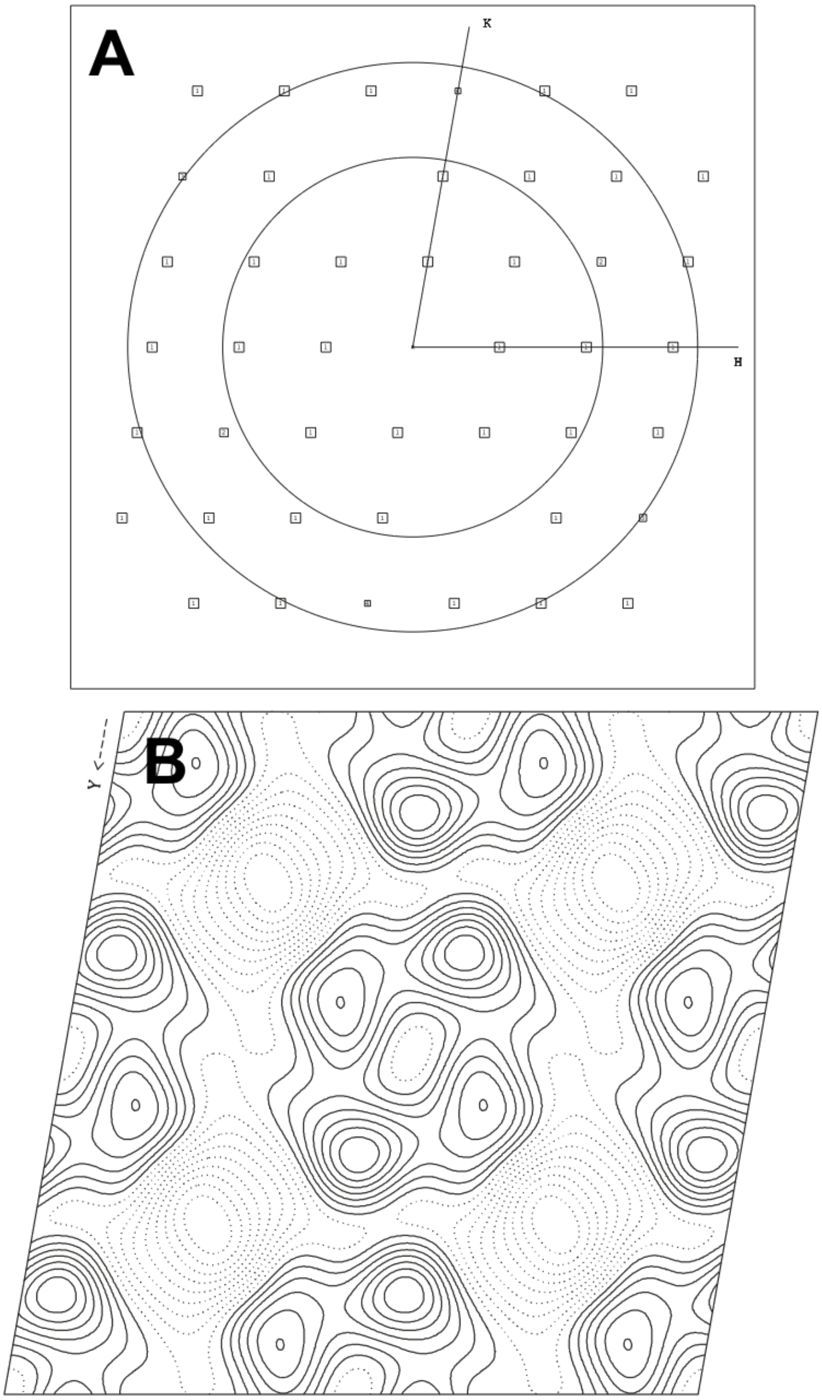


Figure 3.6 Resolution plot and *p*2 symmetrised projection of merged data of R7404 S-layer

A: Canonical resolution circle plot of merged data from six untilted projection images of R7404 S-layer extracts with reflections indicated in squares, as detailed above. The resolution rings indicate 36, and 24 Å (outward from centre of image) with Nyquist frequency of the plot at 20 Å. **B**: Projection structure from **A**, with imposed *p*2 symmetry. 2 × 2 unit cells are represented. Dimensions of the unit cell are $a = b = 80$ Å and $\gamma = 100°$ and the Y axis is indicated.

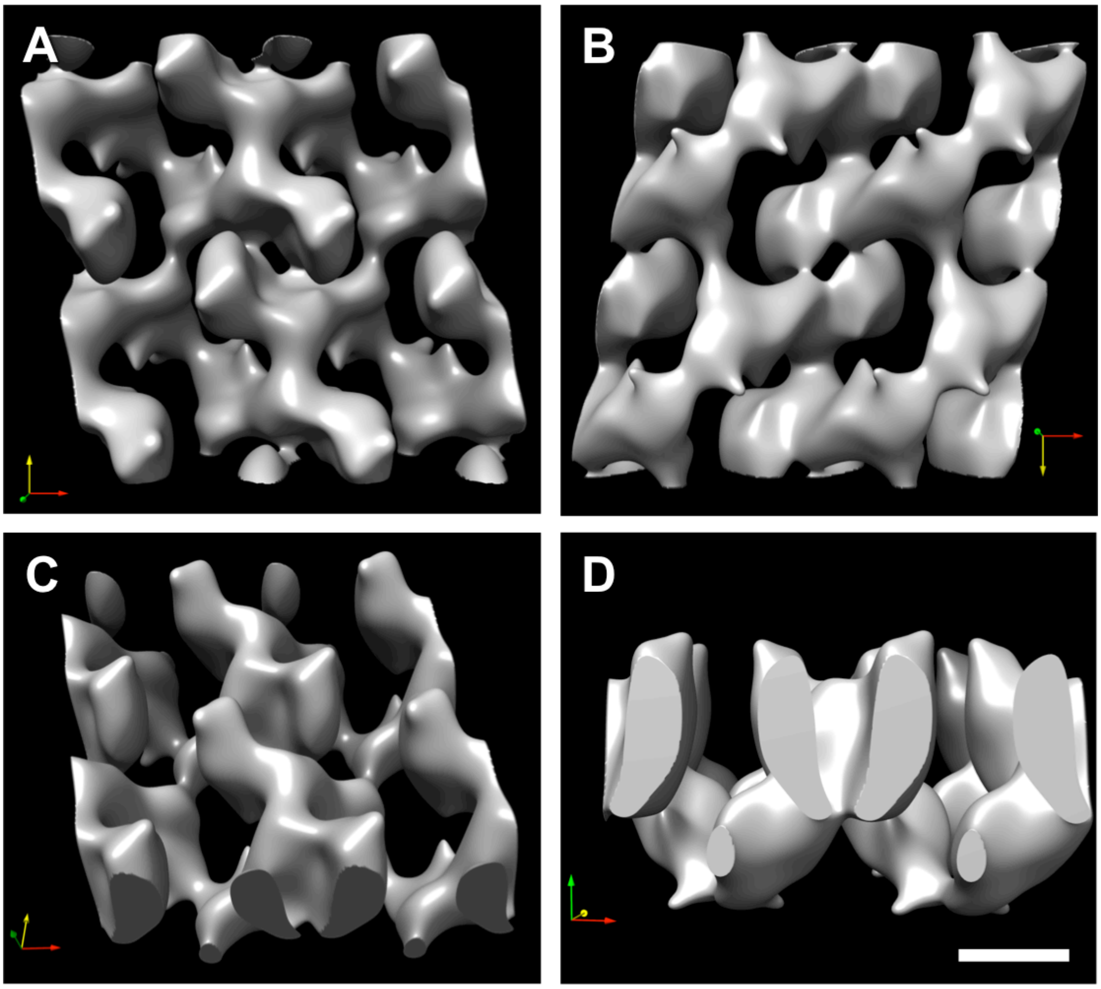


Figure 3.7 3D Volume map of the R7404 S-layer

A to **D:** Various orientations of the R7404 S-layer shown in 3D volume. The scale bar indicates 40 Å. **A** and **B** are viewed perpendicular to the S-layer surfaces. The threshold chosen to render the surface was determined by calculating the volume of 2 × 2 unit cells containing total eight H/L complexes. The X, Y and Z-axes are represented as red, yellow and green arrows respectively. *p*2 symmetry was applied, with symmetry axis along Z. Scale bar 4 nm.

3.2.3.1. S-layer projection structure at 20 Å

TEM images showed the presence of an orthorhombic space group with mean unit cell lengths, $a = b = 80$ Å ± 1 Å, and cell angle, $\gamma = 100^\circ \pm 1^\circ$. Figure 3.8 shows a *p*1 (unsymmetrised) projection map. Based on symmetrised phase residuals given by programme ALLSPACE, an unambiguous *p*2-symmetry was suggested (Table 3.3). Figure 3.9 shows a *p*2-symmetry imposed merged map from merged data set of six different untilted images of the crystal. The maps are truncated to 20 Å resolution and displays 2 × 2 unit cells with symmetry axis perpendicular to the surface of the paper.

3.2.3.2. 3D model

Thirty-five micrographs of individual crystals, with tilt angles ranging from +55 to −55°, and 710 unique reflections, were used to calculate the final 3D map. Appendix 5 shows the distribution of amplitudes and phases of some spots along z^*. The overall phase residual (as described in Section 2.6.5), indicating the overall phase error in the 3D data was 22.3°. A summary of statistics from the 3D data set is shown in Table 3.4.

The final volume map was calculated based on the observed $(h,k,0)$ data and the modelled $(0,0,l)$ values (see Section 2.6.6.1). A surface representation of the calculated 3D volume is shown in Figure 3.10. The threshold for the volume map was set at a level where the surface encompasses a volume corresponding to two subunits of R20291 ribotype 027 SlpA (2 × 88 kDa) arranged in an unit cell.

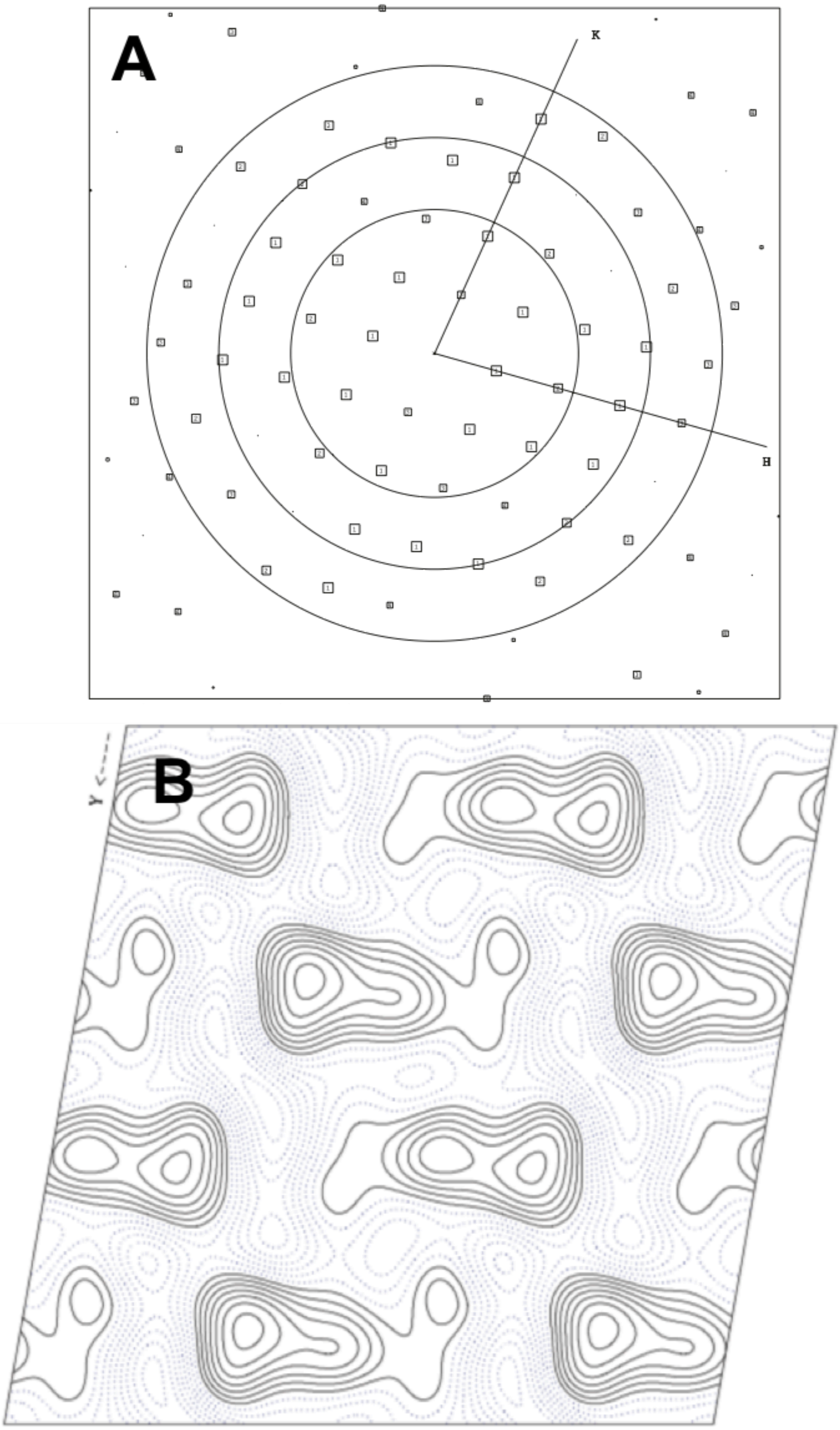


Figure 3.8 Resolution plot and unsymmetrised map of R20291 S-layer

A: Resolution circle plot following image processing of an untilted R20291 S-layer extract with reflections indicated in squares. The number in each square represents the figure of merit of the respective reflection with the H and K axis indicated. The resolution rings indicate 36, 24 and 18 Å (outward from centre of image) with Nyquist frequency of the plot at 15 Å. **B**: The projection structure from **A**, without any imposed symmetry. 2 × 2 unit cells are represented. Dimensions of the unit cell are $a = b = 80$ Å and $\gamma = 100°$.

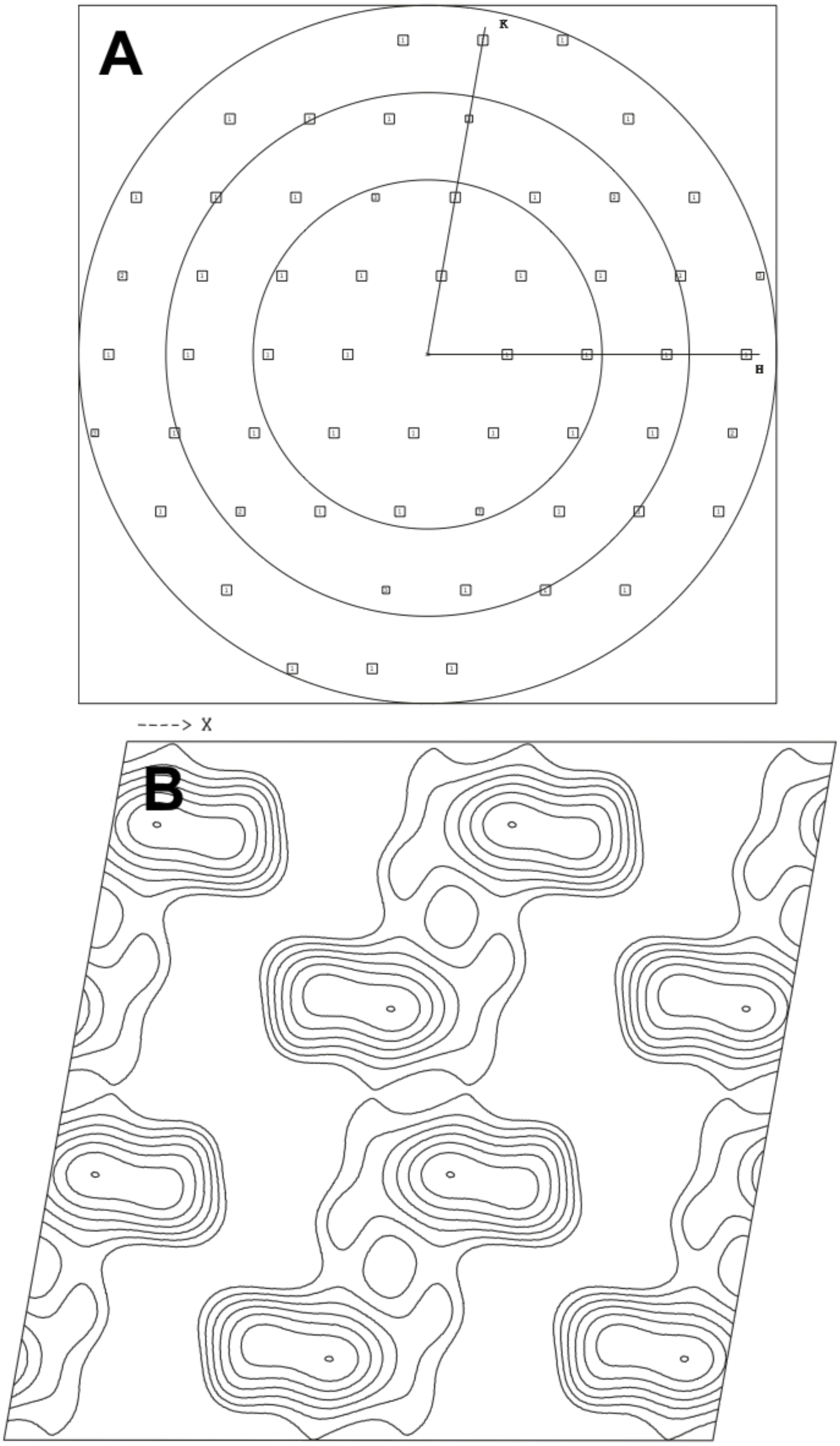


Figure 3.9 Resolution plot and *p*2 symmetrised projection of merged data of R20291 S-layer

A: Canonical resolution circle plot of the merged data set from six R20291 S-layer images with reflections indicated in squares, as detailed in Figure 3.8. The resolution rings indicate 36, 24 and 18 Å with Nyquist frequency of the plot at 18 Å. **B**: The projection structure from **A**, with imposed *p*2 symmetry. 2 × 2 unit cells are represented. Dimensions of the unit cell are $a = b = 80$ Å and $\gamma = 100°$.

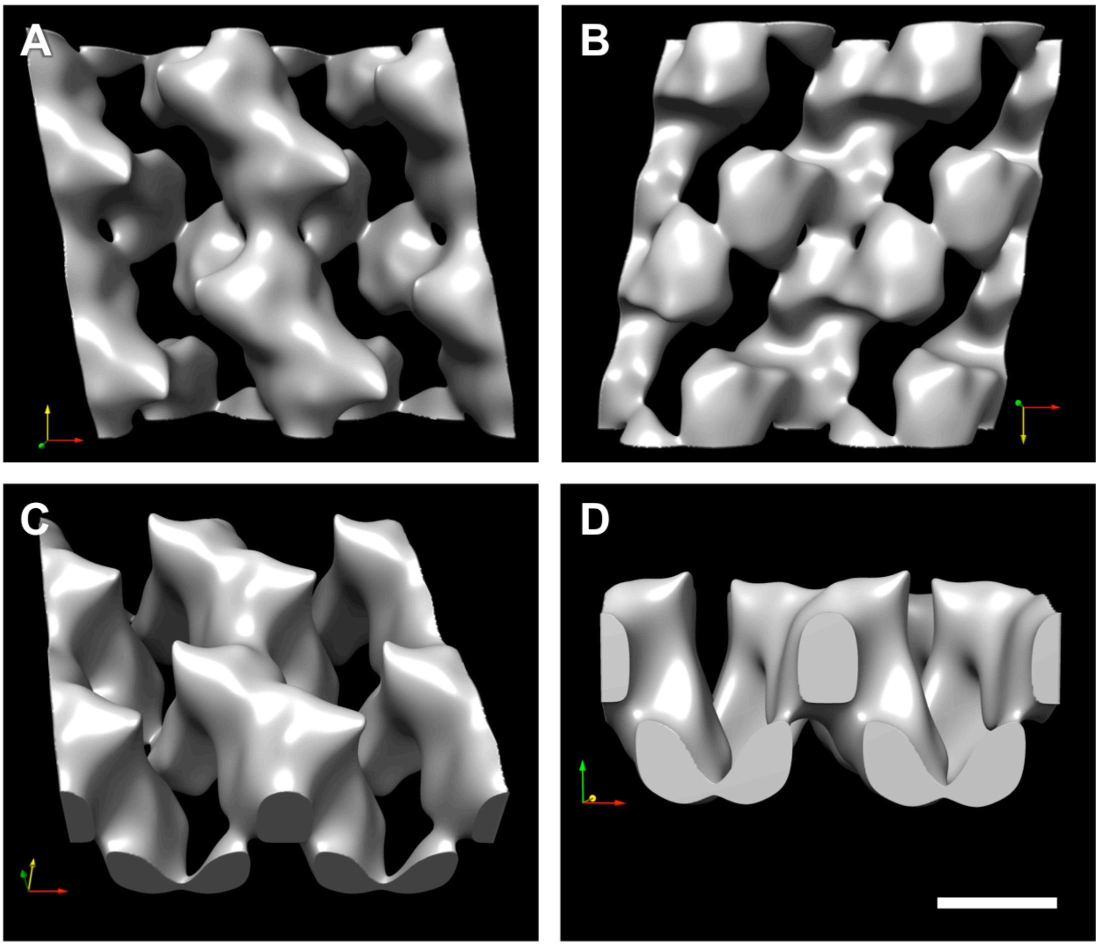


Figure 3.10 3D volume map of R20291 S-layer

A to **D:** Various orientations of the R20291 S-layer shown in 3D volume. The scale bar indicates 40 Å. **A** and **B** are viewed perpendicular to the S-layer surfaces. The threshold chosen to render the surface was determined by calculating the volume of 2 × 2 unit cells containing total eight H/L complexes. The X, Y and Z-axes are represented as red, yellow and green arrows respectively. $p2$ symmetry was applied, with symmetry axis along Z. Scale bar 4 nm.

Table 3.3 Phase residual table for different plane groups for a representative untilted S-layer crystal of R20291

Plane Group[1]	**Phase residual (random = 90°)**	**Target residual**[2]	**Plane Group**[1]	**Phase residual (random = 90°)**	**Target residual**[2]
p1	24.7°		p22121	58.5°	30°
p2	27.5°*	36.1°	c222	69°	30°
p12b	58.8°	28.1°	p4	47.4°°	29.8°
p12a	58.2°	27.6°	p422	68.6°	27.2°
p121b	45.9°	28.1°	p4212	57.6°	27.2°
p121a	40°	27.6°	p3	54.7°	24.7°
c12b	58.8°	28.1°	p312	66°	25.4°
c12a	58.2°	27.6°	p321	61.6°	26.2°
p222	69°	30°	p6	48.7°	28.3°
p2221b	39.3°`	30°	p622	58.7°	26.4°
p2221a	43.4°	30°			

[1] a and b represent the respective symmetry axis for the plane group

[2] Target residual indicates the expected phase residual of each symmetry group based on the SNR ratio of the spots from the concerned image

Symbols for possible plane groups: * = acceptable, ! = should be considered, ` = possibility

Table 3.4 Statistical summary of 3D data from R20291 crystals

3D reconstruction parameters	
Crystal plane group symmetry	$p2$
Crystal unit cell parameters	a = 80.0 Å, b = 80.0 Å, γ = 100.0°, c = 160.0 Å
Number of individual crystals	35
Range of defocus	7760 to 16020 nm
IQ range used for 3D refining	1 to 5
Tilt range used	0 to 55.2°
In-plane resolution cut-off	20 Å
Vertical resolution cut-off	20 Å
Number of observed reflections	1085
Number of observed unique reflections	710
Overall phase residual	22.3°

3.3. Discussion

3.3.1. S-layer extraction technique

Cryo-electron tomography on thin sections or edges of the cells is the predominant technique used in generating high-resolution structural analysis of cell surface assemblies (Chen et al, 2011; Bharat et al, 2017; Engel et al, 2015). Structural analyses of S-layers were largely done by freeze-fracturing and metal shadowing techniques (Wildhaber & Baumeister, 1987; Cerquetti et al, 2000). Scanning probe microscopy has also been used to image the surface in low resolution (Cheong et al, 1993; Günther et al, 2014).

High-resolution structural analysis in EM usually involves either imaging single particles in extensive quantities and orientations, or imaging proteins purified into 2D crystalline arrays, *in vitro*. Looking at high-resolution structures of native S-layers is especially challenging. S-layer proteins are difficult to crystallise and assemble *in vitro*, presumably due to absence of the underlying substrate, such as the cell wall or membrane, which provides anchoring structures. The structure of S-layer protein SbsB in the archean *Geobacillus stereothermophilus* was solved by crystallising the protein using 'nanobodies', while its native assembly state speculated on using cryo-EM projections of *in vitro* 2D crystallised protein (Baranova et al, 2012).

The technique developed and used here for this project is so far the only described method that allows imaging of an intact native S-layers on the cell wall, providing unique biological significance to the data.

3.3.2. The structure

Previous EM analyses on S-layers of six different *C. difficile* strains by freeze-etching revealed an ordered lattice with the ratio of unit cell vectors b/a close to 1, indicating a square ordered lattice (Cerquetti et al, 2000). In the same paper, analysis of sonicated fragments in phosphotungstate, however, revealed a 'hexagonal' lattice, with lattice spacing of 10 nm. The FFT displayed in the paper shows a complex arrangement of spots, which are not consistent with a basic hexagonal lattice arrangement (Cerquetti et al, 2000). The 'hexagonal' reflections in the FFT may be attributed to either overlapping S-layers from fragments retaining both sides of the cell surface (as shown in this thesis, Figure 3.4), or contributed by multiple domains of lattices in slightly different orientations (as seen in Figure 5.1). Damage to the crystalline order due to the process of sonication is another possibility that cannot be ruled out. Such damage has been observed in our earlier S-layer isolation trials (data not shown). As such, the findings in that paper of a 'hexagonal' lattice may actually correspond to a complex arrangement of p2-symmetrised lattices.

In this thesis, the resolution limit due to the use of negative stain meant structural details up to 20 Å are reliable. The S-layer structure, in case of R7404 and R20291 strains, and as will be seen later, of other strains as well, show an oblique lattice with p2 symmetry.

Crystals of purified R7404 SlpA have yielded X-ray diffraction data to 2.8 Å (Dr Paula Salgado, personal communication). This unpublished dataset has not been phased at the time of this thesis. The unit cell dimensions were a = 78 Å, b = 80 Å and c = 134 Å, with unit cell angle γ = 101° and $\alpha = \beta$ = 90°. The crystal packed into $P2_111$ or P211 symmetry. Discounting the packing in the third dimension, this data shows remarkable similarity with the structure obtained from the EM analysis of native S-layers. Both strains shown here show $p2$ symmetry with symmetry axis along Z, and unit cell dimensions $a = b$ = 80 Å and γ = 100°. It is very likely that the native S-layer represents a single 2D layer of the 3D crystal in the a,b plane.

Although, the S-layer of the two strains show similar unit cell dimensions and symmetry, the overall structure shows significant difference. $p2$ symmetry generates two-asymmetric units in each unit cell. The HMW and LMW SLPs are known to form a complex (Fagan et al, 2009), which indicates the presence of two of these complexes in a single unit cell. Volume analysis shows that to be the case. In the EM model of R20291, the size of the HMW/LMW complex corresponds to ~88 kDa.

The surface rendered within a single unit cell is consistent with 88 kDa per asymmetric unit for the maps shown in Figure 3.7 and Figure 3.10.

SlpA in *C. difficile* is known to interact, non-covalently, with the cell wall associated carbohydrate, PS-II. The HMW SLP contains the three tandem repeats of cell wall binding (CWB2) domains (Reynolds et al, 2011) which must therefore be oriented towards the cell wall, while the hyper-variable immunogenic domain of the LMW SLP is expected to be exposed to the outer surface of the cell. It is difficult to obtain relative orientations from EM models. However, a structural modification was used to obtain the relative orientation of the map, to the cell wall. This result is described in the following chapter.

Chapter 4 - Localisation of the hyper-variable domain of SlpA

4.1. Introduction

One way of identifying the orientation of a structure is to introduce a mutation, such as a deletion, to the structure. If the deletion does not result in a dramatic conformational change in the final structure, the original model can then be compared with the mutated protein model to reveal the location of that deletion. If the deleted region is known to have certain features, revealing its likely spatial localisation, then the orientation of the EM model can be established.

slpA appears to be an essential gene and mutation studies have so far not been possible (Section 1.1). Transposon mediated mutagenesis studies show a complete absence of any *slpA* mutants (Dembek et al, 2015). However, when *C. difficile* cells were exposed to S-layer-targeting bactericidal protein complexes, termed “Avidocin-CDs”, the selective pressure made it possible to isolate novel Avidocin-resistant mutants *in vitro* that completely lack SlpA (Kirk et al, 2017b).

Bacteriocins targeting *C. difficile* (Avidocins) were recently isolated from *C. difficile* cells (Gebhart et al, 2012). The cells are themselves insensitive to their own Avidocins, suggesting that a strain-specific variable structure is the binding receptor. The Avidocins are analogous to the R-type bacteriocins from *Pseudomonas aeruginosa*, called “pyocins”, containing a contractile myophage-like sheath and receptor binding proteins (RBPs) at the tail end (Nakayama et al, 2000). Upon binding to the cell surface receptors, the sheath contracts, driving a needle-like core through the target bacterial cell envelope. This dissipates the membrane potential killing the bacterium (Scholl et al, 2009). Further analyses revealed that Avidocins specifically target the S-layer. Avidocin Av-D291.2 was able to successfully kill ribotype 027 strains (Kirk et al, 2017b).

Under selective pressure, spontaneous R20291 mutants, resistant to Av-D291.2, appeared at a frequency of less than 10^{-9}. Independent mutants, FM2.5 and FM2.6, were isolated and characterised to understand the mechanism of resistance. A frame-shift and nonsense mutation was located, respectively, in FM2.5 and FM2.6, one codon apart in the *slpA* region encoding the LMW SLP domain 1 (Figure 4.1A) (Kirk et al, 2017b). As expected, analysis of surface proteins revealed a lack of both HMW and LMW SLPs that occur in wild type R20291 (Figure 4.1B). Complementation of the two strains with wild type *slpA* from R20291 was performed by homologous recombination. This restored the wild type *slpA* in place of FM2.5 and FM2.6 *slpA,* making the mutants susceptible to Av-D291.2 (Kirk et al, 2017b). The "revertants", FM2.5R and FM2.6R produce both HMW and LMW SLPs (Figure 4.1B). The *slpA* mutants formed small dense colonies, whereas the revertant strain showed restored colony morphology, identical to the wildtype (Kirk et al, 2017b). The nonsense mutant FM2.5 *slpA* gene is designated as *slpA** hereafter.

The LMW domain localisation experiment described here relied on mutation of *slpA** to remove the SlpA LMW domain 2 coding sequence, followed by homologous recombination to place the modified allele on the FM2.5 chromosome. Analysis of the S-layer structure of a mutant lacking a LMW domain enabled us to compare it with the wild type S-layer structure of R20291 described in Chapter 3 and subsequently localise the LMW domain.

4.2. Results

4.2.1. Generation of *slpA*$^{\Delta Ld2}$ mutant

The *slpA*$^{\Delta Ld2}$ mutant was generated by deleting the hyper-variable domain (domain 2) of the LMW SLP region (Figure 4.2), followed by introduction of two glycine residues in its place. This specific domain was identified in a previously published crystal structure of LMW SLP of *C. difficile* strain 630 (Figure 4.2B) (Fagan et al, 2009). An *in silico* model of this region in R20291 was generated using the PHYRE2 Protein Fold Recognition Server (Kelley et al, 2015) (Figure 4.2C).

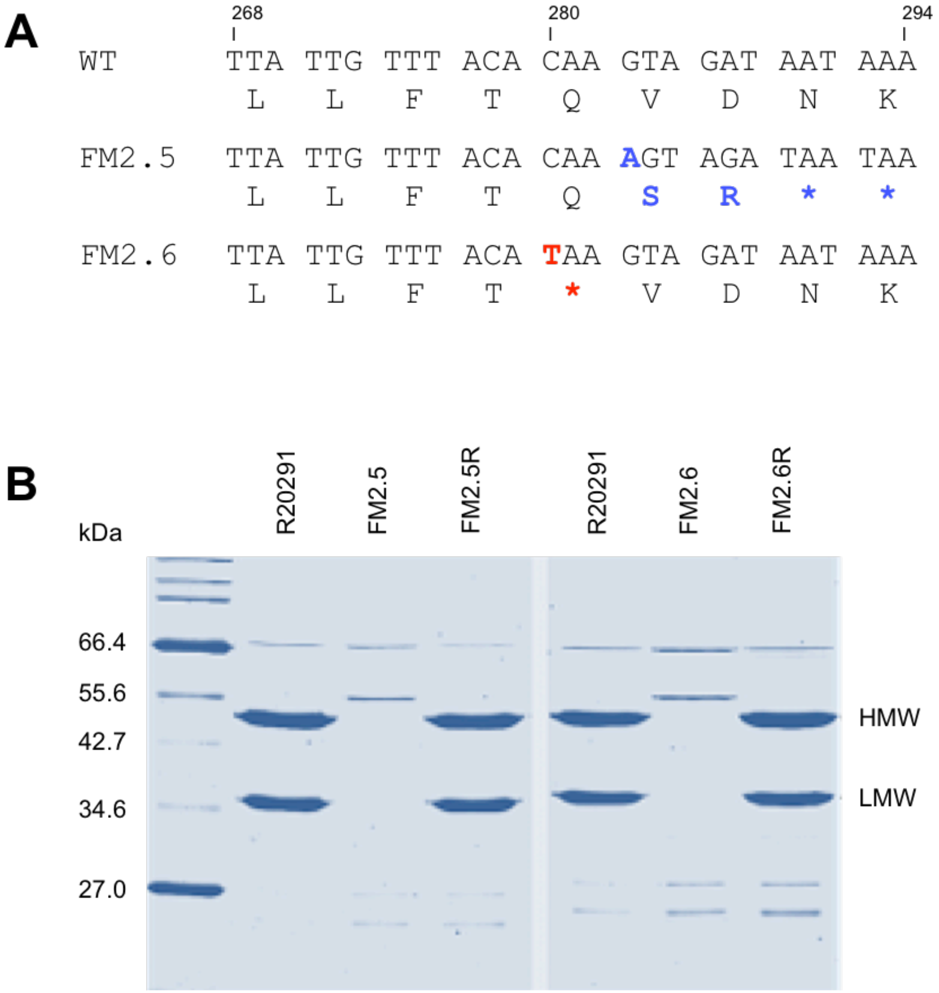


Figure 4.1 S-layer deficient mutants FM2.5 and FM2.6

A: *slpA* sequence alignments from R20291 (wild type) and Av-D291.2-resistant mutants, FM2.5 and FM2.6, in the 268 – 294 nucleotide region. The region encodes part of the LMW SLP domain 1. A nucleotide insertion at position 283 results in a frame shift mutation in FM2.5 (highlighted in blue), whereas a nucleotide substitution at position 280 for FM2.6 results in a nonsense mutation (highlighted in red). **B:** Coomassie-stained SDS-PAGE analysis of the low-pH glycine extraction from R20291, FM2.5, FM2.6 and the respective *slpA*-substituted "revertants", FM2.5R and FM2.6R (**B** image from (Kirk et al, 2017b), taken with permission from the American Association for the Advancement of Science (AAAS)). The positions of the HMW and LMW SLPs are indicated. The Av-D291.2-resistant mutants show an absence of the HMW and LMW SLPs.

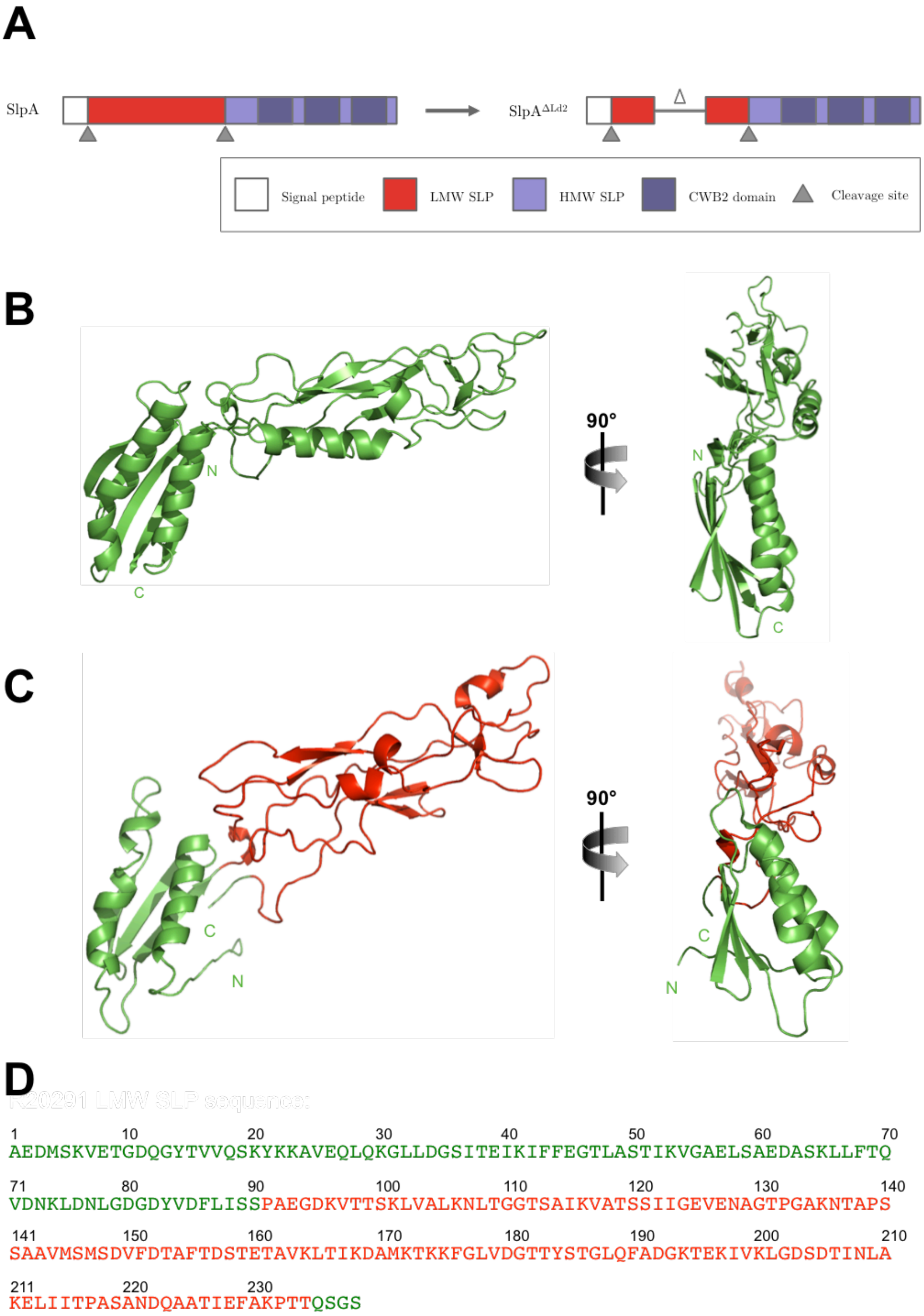


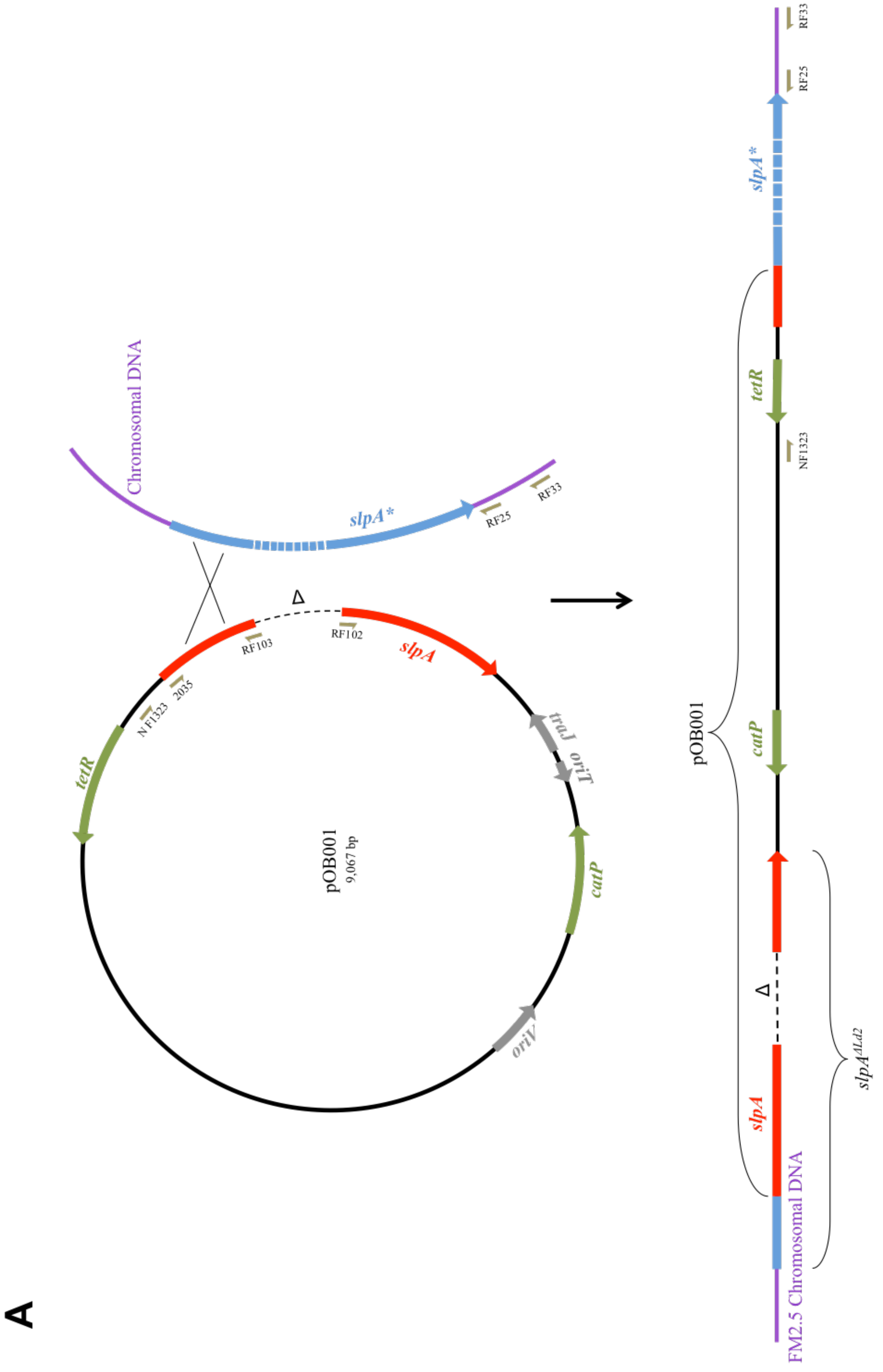
Figure 4.2 The LMW domain 2 deletion

A: A schematic diagram showing the domain 2 deletion in the SlpA pre-protein. The 'Δ' represents the deletion region, with the legend below indicating the various regions of interest in SlpA. **B**: Ribbon diagram showing two views of the LMW SLP crystal structure from *C. difficile* 630 strain (Fagan et al, 2009) (pdb no. 3cvz). **C**: Ribbon diagram showing the *in silico* model of the R20291 LMW SLP, the domain 2 is highlighted in red. Following deletion of domain 2, the outgoing and return strands from domain 1 (green) were joined with Gly-Gly hairpin. This model was generated using the web based Phyre 2 protein fold prediction service (Kelley et al, 2015), from the known structure of the strain 630 LMW SLP. D: The protein sequence of the R20291 LMW SLP is shown below the ribbon diagram, with the deleted domain highlighted in red. The sequence is numbered from the first residue of the LMW SLP after the signal peptide. The deletion region represents residues 91 – 234.

The DNA sequence encoding the domain 2 of the LMW SLP was identified in a plasmid (pRPF233) borne copy of the R20291 *slpA* gene and deleted using inverse PCR, replacing it with additional sequences coding for Gly-Gly residues.

The plasmid (named pOB001, Figure 4.3A) was conjugated into *C. difficile* FM2.5 cells (Section 2.3.8). The gene was allowed to recombine into the *slpA** locus in the chromosome of the *slpA*$^{-}$ strain FM2.5. Following conjugation, colonies that appeared morphologically different to FM2.5 were likely candidates where homologous recombination might have occurred between the *slpA*$^{\Delta Ld2}$ and *slpA** gene. Such colonies were selected and tested for the presence of the *slpA*$^{\Delta Ld2}$ gene in the native locus. Single crossover was confirmed by PCR (Figure 4.3B). Antibiotic selection pressure was removed to allow plasmid loss, following a second recombination event. Colonies that retained altered morphology following plasmid loss were further screened to ascertain retention of the *slpA*$^{\Delta Ld2}$ allele in the native locus (Figure 4.3C).

Finally, the colonies that retained *slpA*$^{\Delta Ld2}$ were grown overnight and the cell surface proteins were extracted by low-pH glycine (described in Section 2.4.1) and visualized following SDS PAGE (Figure 4.4). The SlpA$^{\Delta \mathrm{Ld2}}$ cells express the S-layer proteins with a truncated LMW SLP of 19 kD molecular weight, compared to the wild type 36 kD LMW SLP. SlpA^{-} strain shows an absence of both SLPs. The colonies show very distinct morphologies, which aided in the selection of recombinants (Figure 4.4B). The SlpA$^{\Delta \mathrm{Ld2}}$ recombinants selected were DNA sequenced to confirm deletion of domain 2 and addition of Gly-Gly residues.


A
Chromosomal DNA
slpA*
RF25
RF33
Δ
RF103
RF102
slpA
2035
NF1323
tetR
pOB001
9,067 bp
traJ
oriT
catP
oriV
RF33
RF25
slpA*
tetR
NF1323
pOB001
catP
Δ
slpA
slpA^ΔLd2
FM2.5 Chromosomal DNA

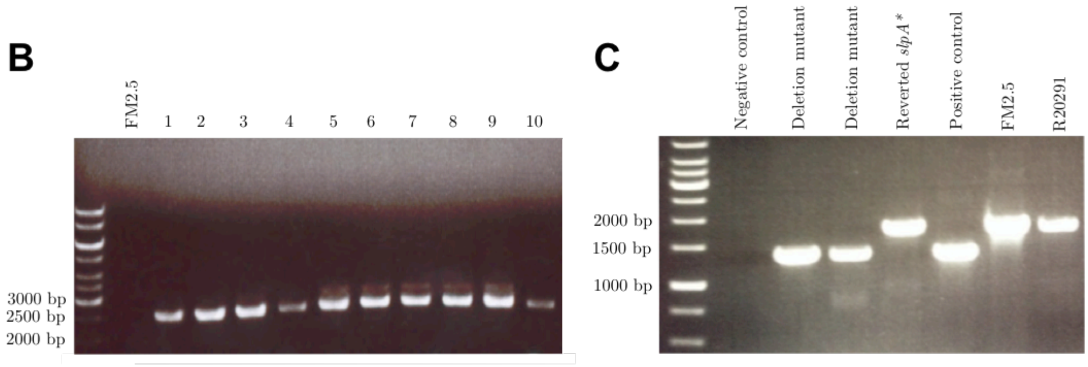


Figure 4.3 Homologous recombination and insertion of pOB001 into the native *slpA locus in FM2.5**

A: Plasmid pOB001 used to construct the SlpA$^{\Delta Ld2}$ mutant along with a schematic showing a recombination event with *slpA** in FM2.5. Large arrows represent genes, with the direction indicating transcription. Smaller half arrows indicate the primers used in various steps of the protocol. *tetR* is the tetracycline repressor which binds to the *tet* promoter. *catP* is the chloramphenicol acetyl transferase gene. *oriV* is the *E. coli* origin of replication. The *oriT* and *traJ* both are involved in conjugative transfer of the plasmid. Delta (Δ) indicates the region in the *slpA* gene that has been deleted (dashed lines) from the original pRPF233 plasmid.

A single recombination event downstream of the deletion site is depicted below the plasmid. The chromosomal DNA of FM2.5 is shown in purple with the *slpA** gene in blue. The dashed blue region in the *slpA** gene indicates the equivalent region which was deleted in the *slpA*$^{\Delta Ld2}$ gene in plasmid pOB001. The *catP* and *tetR* (green) are now integrated into the FM2.5 chromosome. A second recombination event occurs when the antibiotic selection pressure is removed and plasmid DNA is lost. This second recombination event along with loss of plasmid DNA may leave behind a functional *slpA*$^{\Delta Ld2}$ gene or result in reversion to *slpA**. The primers (half arrows) used to test for these events have been labeled (refer to primers in Table 2.6).

B: Agarose gel of PCR with primers NF1323 and RF25, on gDNA extracted from selected colonies following the first recombination event. Successful amplification indicates integration of the plasmid at the *slpA* locus. **C**: Agarose gel of PCR with gDNA, using RF33 and 2035 flanking the deleted region, following the second recombination event. The reverted phenotype can be seen with a larger sized band (1,793 bp), similar to FM2.5 and R20291, compared to the deletion mutant (1,362 bp).

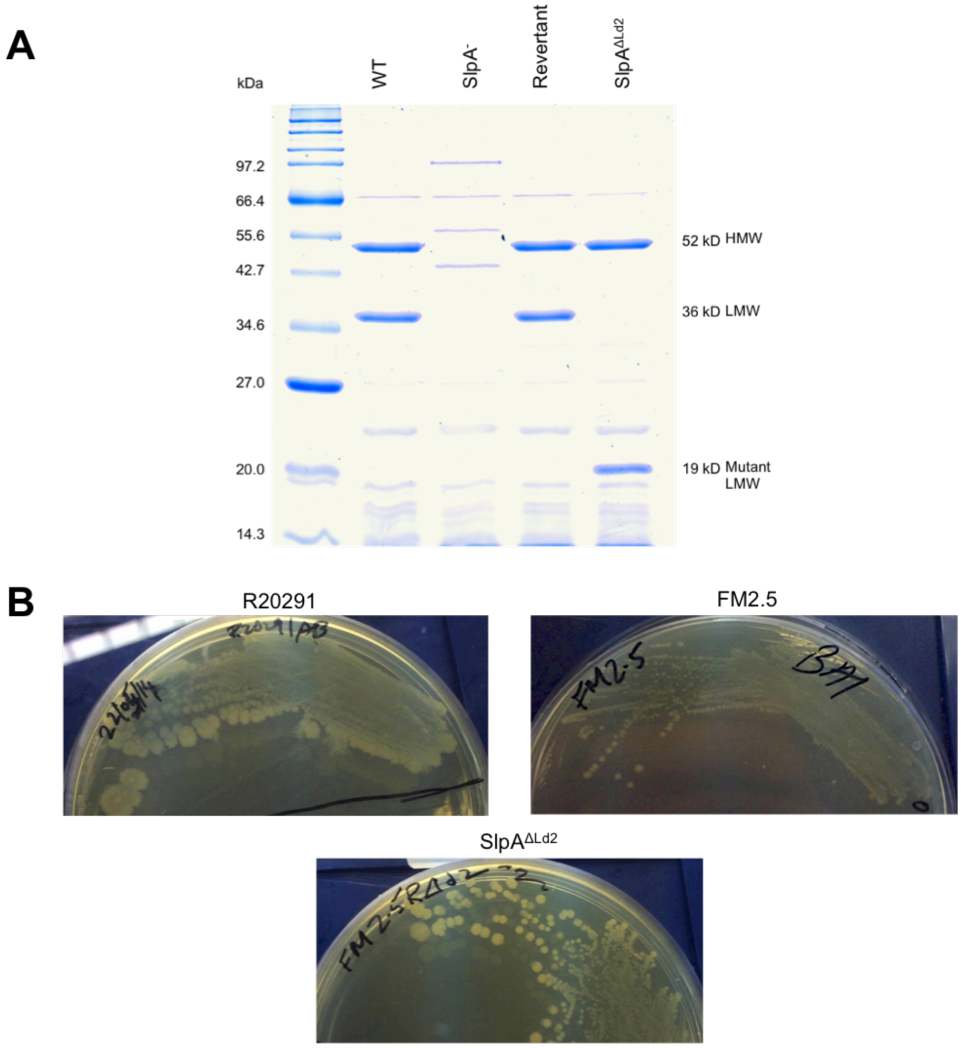


Figure 4.4 Cell surface proteins and colony morphology for wild type R20291, FM2.5 mutant and the SlpA$^{\Delta Ld2}$

A: SDS PAGE gel showing the non-covalently bound cell surface proteins. Clearly visible are the two major bands for SlpA in the R20291 wild type (WT), FM2.5 (SlpA^{-}), revertant strain expressing wild type SlpA (Revertant) and SlpA$^{\Delta Ld2}$ strain expressing truncated LMW SLP. The revertant strain was generated using the exact same protocol as discussed in this section (work by Joseph A Kirk), but with FM2.5 conjugated with plasmid containing wild type *slpA* from R20291 (Kirk et al, 2017b). The HMW SLP appears as a band at 52 kD and the wild type LMW SLP appears as a 36 kD band. The SlpA^{-} strain shows a noticeable absence of HMW and LMW SLP. SlpA$^{\Delta Ld2}$ strain shows expression of a truncated smaller LMW SLP at 19 kD. **B**: R20291, FM2.5 and SlpA$^{\Delta Ld2}$ colonies grown on BHI agar at 37°C anaerobically for 48 h. The FM2.5 *slpA*$^{-}$ mutant shows small dense colonies, while the SlpA$^{\Delta Ld2}$ strain shows larger sized colonies, intermediate between the FM2.5 and R20291. The distinctive colony morphologies between the mutants aided the selection of recombinants.

4.2.2. Structural analysis of the SlpA$^{\Delta Ld2}$ mutant

SlpA$^{\Delta Ld2}$ was grown in reduced TY broth to exponential phase. The cells were then harvested and treated with endolysin extracted from *C. difficile* phage CD27, following the S-layer protocol discussed in Section 2.2.1. The S-layer extracts were then loaded on to EM grids and negatively stained using method described in Section 2.5.2 and observed under the TEM.

4.2.2.1. S-layer projection structure at 20 Å

The S-layer extracts formed *ghosts*, similar to the ones discussed in Chapter 3, which appeared flattened on the carbon film. Figure 4.5 compares the appearance of SlpA$^{\Delta Ld2}$ S-layer extract to its wild type. Some of the mutant S-layer extracts appeared to have a curved morphology and asymmetric division septum. Both are reminiscent of the FM2.5 morphology. However unlike FM2.5, these SlpA$^{\Delta Ld2}$ curved cells do possess an S-layer. The SlpA$^{\Delta Ld2}$ cells are normally shorter in size by 1-2 µm. High-magnification images reveal the presence of two superimposed ordered lattices (Figure 4.6).

The S-layer capsules for the SlpA$^{\Delta Ld2}$ showed identical unit cell dimensions, $a = b = 80$ Å $\pm$ 1 Å, $\gamma = 100° \pm 2°$. The symmetry, as calculated from the Fourier terms of the reflections using *ALLSPACE*, suggested the most likely space group as $p2$, similar to the wild type 027 S-layer. Phase residual table from *ALLSPACE* of one representative crystal is shown in Table 4.1. The 2D map, revealing the untilted projection density was calculated, once again, using the method described in Section 2.6. Figure 4.7 shows the resolution circle plot, representing the reflections obtained after the image processing of one single crystal, and the projection map calculated from that image, without any symmetry imposed. Figure 4.8 shows a merged projection map from six different images with $p2$ symmetry applied.

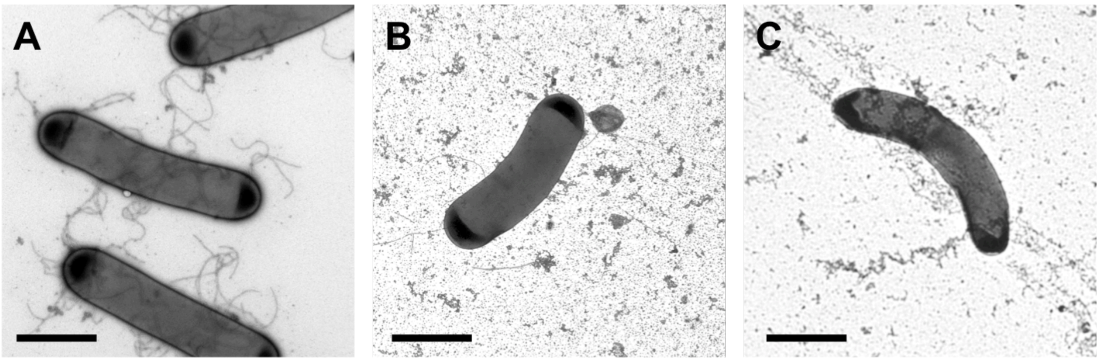


Figure 4.5 Comparing the morphologies of wild type R20291 and mutant SlpA$^{\Delta Ld2}$

A: The negatively stained EM image of endolysin-treated S-layer extract from wild type R20291. **B** and **C**: Negatively stained SlpA$^{\Delta Ld2}$ S-layer extracts. Both display gentle curvatures with the extract in **C** also displaying an asymmetric septum. All images are at the same magnifications and scale bars represent 2 µm.

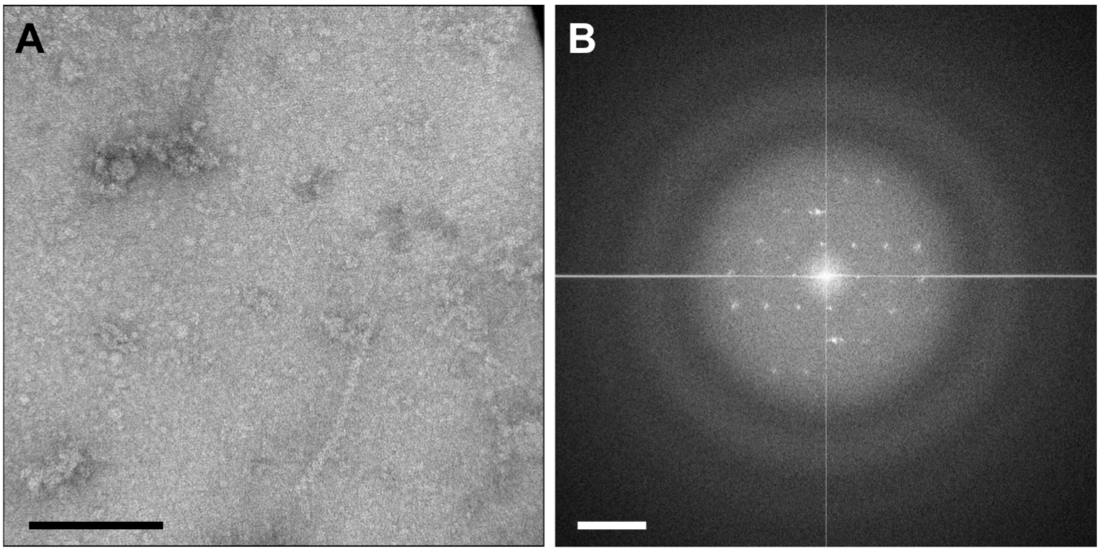


Figure 4.6 Magnified image of the S-layer extract from SlpA$^{\Delta Ld2}$ and its corresponding FFT

A: Magnified image of the endolysin-treated S-layer extract of mutant SlpA$^{\Delta Ld2}$. **B**: The FFT from A, clearly indicating the two overlapping lattices. Scale bars in **A** represents 200 nm and **B** 0.0625 nm^{-1}.

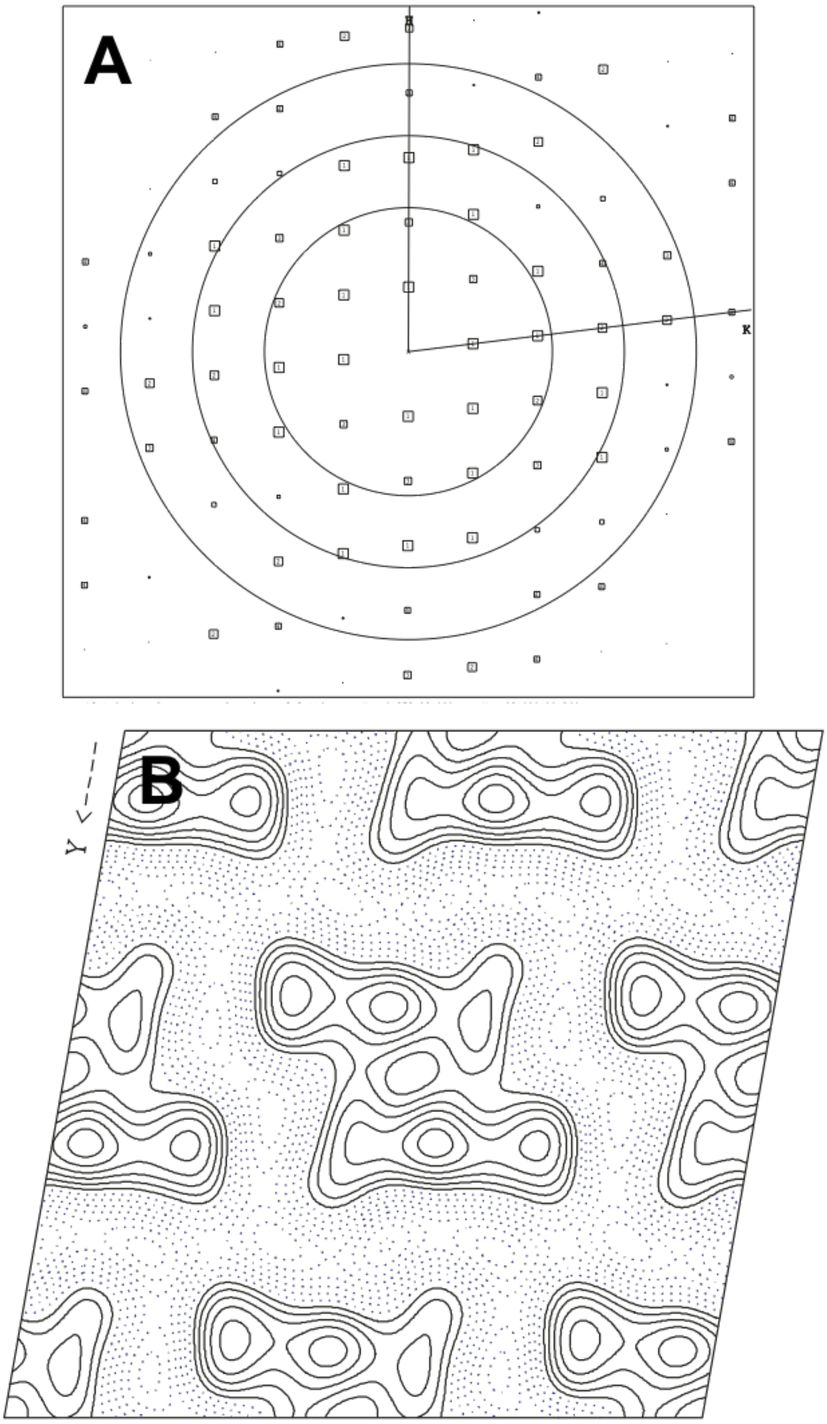


Figure 4.7 Resolution plot and unsymmetrised untilted projection map of a single SlpA$^{\Delta Ld2}$ S-layer crystal

A: Resolution circle plot following image processing of an untilted SlpA$^{\Delta Ld2}$ S-layer with reflections indicated in squares. The number in each square represents the figure of merit of the respective reflection, with the H and K axis indicated. The resolution rings indicate 36, 24 and 18 Å (outward from centre of image) with Nyquist frequency of the plot at 15 Å. **B**: The projection structure without any imposed symmetry. 2 × 2 unit cells are represented. Dimensions of the unit cell are $a = b = 80$ Å and $\gamma = 100°$ and the Y-axis is indicated.

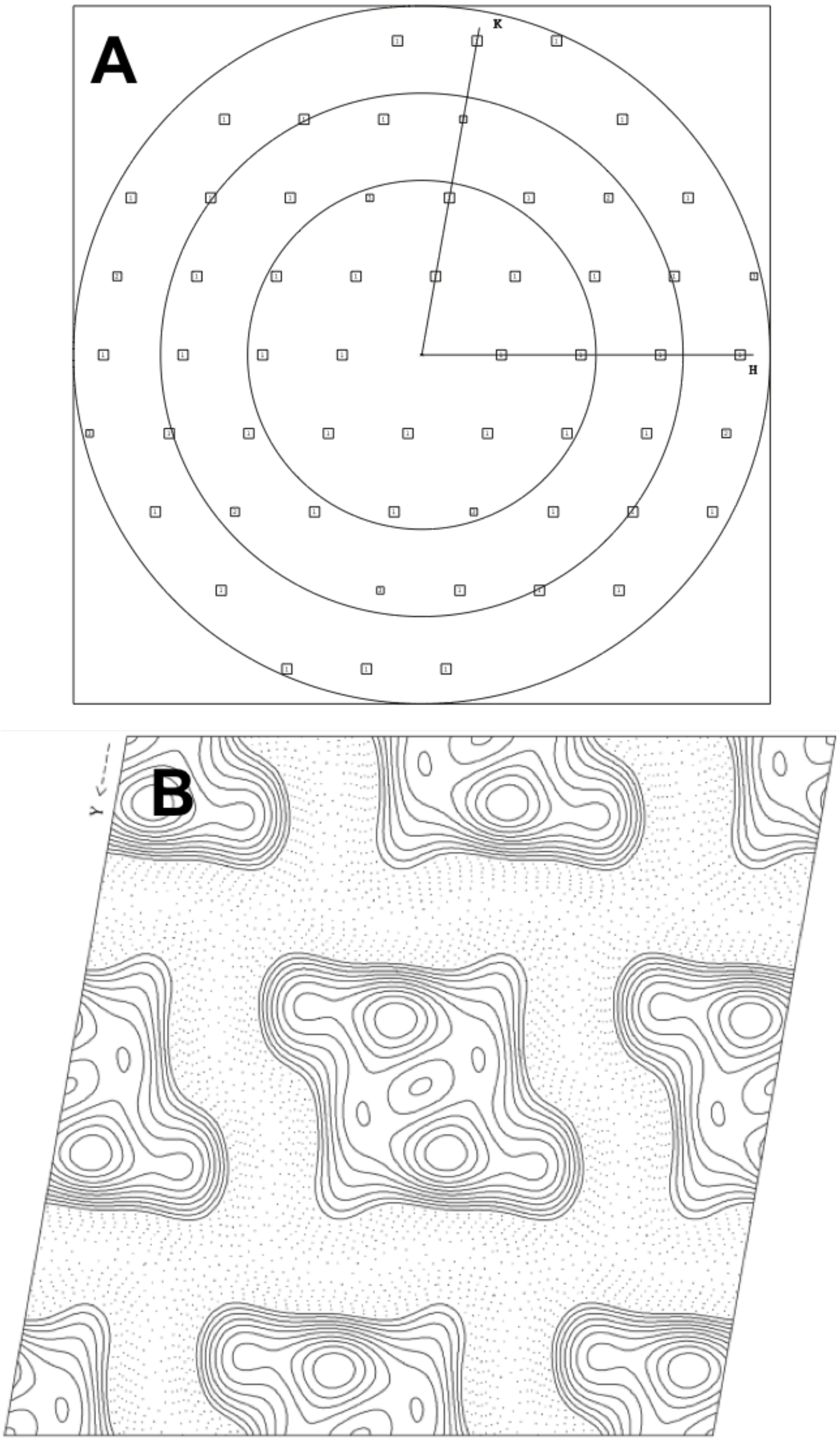


Figure 4.8 Resolution plot and merged symmetrised projection map of mutant

A: Canonical resolution circle plot of merged data set from untilted images of six SlpA$^{\Delta\text{Ld2}}$ S-layer extracts with reflections indicated in squares, as detailed before. The resolution rings indicate 36, 24 and 18 Å (outward from centre of image) with Nyquist frequency of the plot at 18 Å. **B**: The projection structure with imposed *p*2 symmetry. 2 × 2 unit cells are represented. Dimensions of the unit cell are $a = b = 80$ Å and $\gamma = 100°$ and the Y axis is indicated.

4.2.2.2. 3D volume map

Tilting the sample under the electron beam, a series of images were recorded between +45 to –45°. From a total of 29 micrographs, 542 unique reflections were obtained, and were used to calculate tilted projections. The structure factors of the tilted projections were subjected to three-dimensional analysis. The overall phase residual was 13.9°. Appendix 6 shows the distribution of amplitudes and phases of some *h*,*k* spots along *z** for this dataset. The overall phase residual in the 3D dataset was 13.9°. The estimated values for (0,0,*l*) indices were modelled into the 3D transform, using the method described in Section 2.6.6.1, and a surface representation of the calculated 3D volume is shown in Figure 4.9. The threshold for the volume map was set at a level where the surface encompasses a volume corresponding to two subunits of SlpA$^{\Delta Ld2}$ (2 × 71 kDa) arranged in an unit cell. A summary of statistics from the 3D data set is shown in Table 4.2.

Table 4.1 Phase residual table for different plane groups for a representative untilted S-layer crystal of SlpA$^{\Delta Ld2}$

Plane Group[1]	Phase residual (random = 90°)	Target residual[2]	Plane Group[1]	Phase residual (random = 90°)	Target residual[2]
p1	19.8°		p22121	54°	23.7°
p2	19.0°*	28.5°	c222	69.3°	23.7°
p12b	52.1°	21.8°	p4	26.0°!	23.3°
p12a	67.3°	22.3°	p422	71°	21.4°
p121b	63.7°	21.8°	p4212	51.9°	21.4°
p121a	36.2°	22.3°	p3	51.5°	19.8°
c12b	52.1°	21.8°	p312	55.4°	20.1°
c12a	67.3°	22.3°	p321	58.5°	20.8°
p222	69.3°	23.7°	p6	52.4°	22.1°
p2221b	33.6°`	23.7°	p622	64.6°	20.9°
p2221a	48.7°	23.7°			

[1] a and b represent the respective symmetry axis for the plane group

[2] Target residual indicates the expected phase residual of each symmetry group based on the SNR of the spots from the concerned image

Symbols for possible plane groups: * = acceptable, ! = should be considered, ` = possibility

Table 4.2 Statistical summary of 3D data from SlpA$^{\Delta Ld2}$ crystals

3D reconstruction parameters	
Crystal plane group symmetry	$p2$
Crystal unit cell parameters	a = 80.0 Å, b = 80.0 Å, γ = 100.0°, c = 160.0 Å
Number of individual crystals	29
Range of defocus	8520 to 22630 nm
IQ range used for 3D refining	1 to 5
Tilt range used	0 to 41.1°
In-plane resolution cut-off	20 Å
Vertical resolution cut-off	20 Å
Number of observed reflections	667
Number of observed unique reflections	542
Overall phase residual in IQ 1 to 8	13.9°

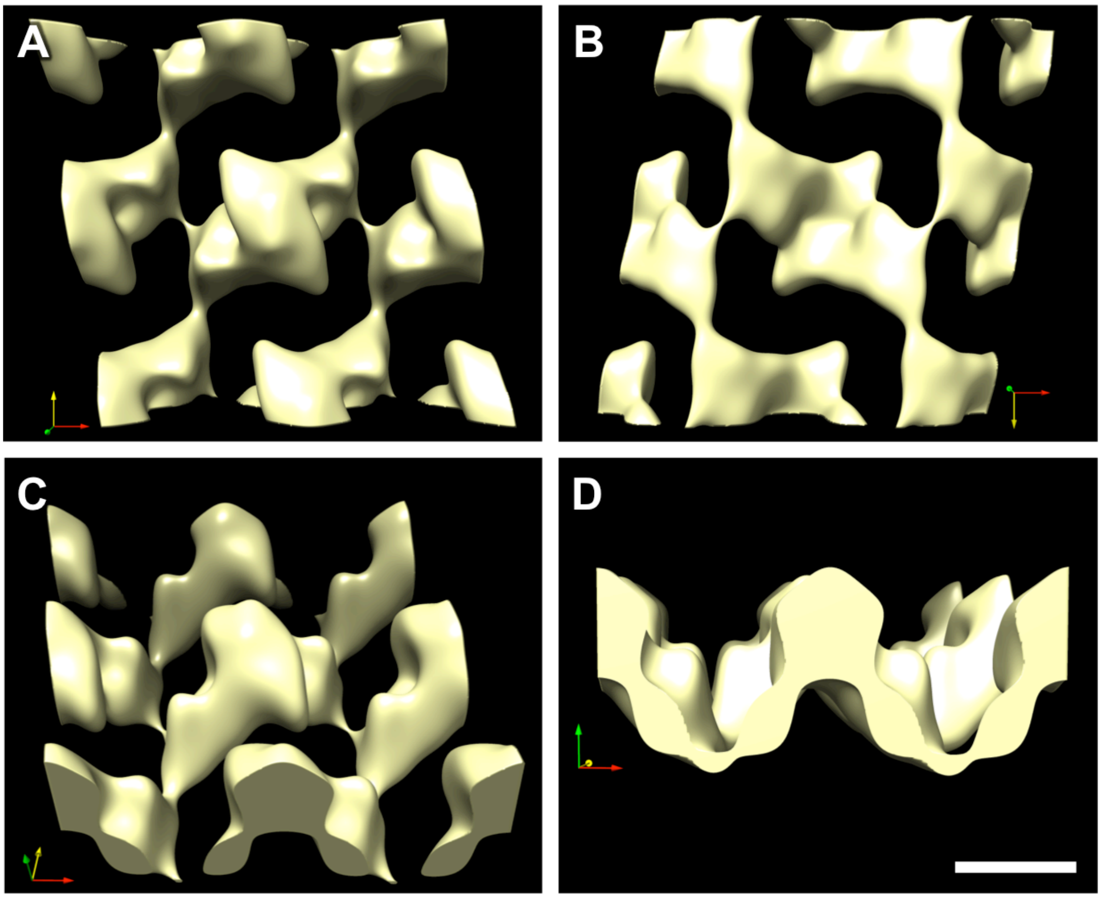


Figure 4.9 3D volume map of mutant SlpA$^{\Delta Ld2}$ S-layer

A to **D**: Surface representations of the 3D volume map of the SlpA$^{\Delta Ld2}$ S-layer from various orientations. The scale bar indicates 40 Å. **A** and **B** are either the top or bottom views of the layer. The X, Y and Z-axes are represented as red, yellow and green arrows, respectively. *p*2 symmetry was applied, with symmetry axis along Z. Scale bar is 4 nm and applies to all panels.

4.2.2.3. Overlaying R20291 S-layer

Visually the two structures display strong similarities. The wild type (R20291) and mutant (SlpA$^{\Delta Ld2}$) untilted projection maps, at 20 Å resolution, were overlaid based on structural similarities. This set the phase origin based on which all maps were generated, allowing direct visual comparison between the two types of maps. The structural variation was quantified by generating a difference map, where the Fourier terms were scaled followed by a real space difference calculation. This showed a statistically significant difference in EM density between the two maps, indicated by blue contour lines in Figure 4.10. Contour lines were set at 0.5 rms value of the average between the WT and mutant EM densities.

Furthermore, a real-space difference map was generated using the 3D volume maps after scaling the individual maps. An overlapped 3D surface rendering of both maps is shown in Figure 4.11. The grey wireframe map is the wild type surface, whereas, the yellow surface represents the mutant volume. A clear void of absent density is visible in the overlay. In Figure 4.12 the real-space difference is shown as a blue density overlaid with the wild type and in Figure 4.12C the blue difference density is overlaid on the mutant volume. The location of the density coincides with the 2D projection difference and indicates the region of structural deviation from the wild type.

4.3. Discussion

4.3.1. Selection of the LMW domain

Fagan et al. (2009) demonstrated that the LMW domain shows the highest sequence variability among strains of *C. difficile*, and is therefore likely to be involved in bacteriophage binding or host immune evasion. The signal sequence and the Cwp84 cleavage sequences remain unaffected following deletion of this domain. The domain 2 also does not interact with the HMW SLP, and therefore would be unlikely to interfere with HMW/LMW complex assembly.

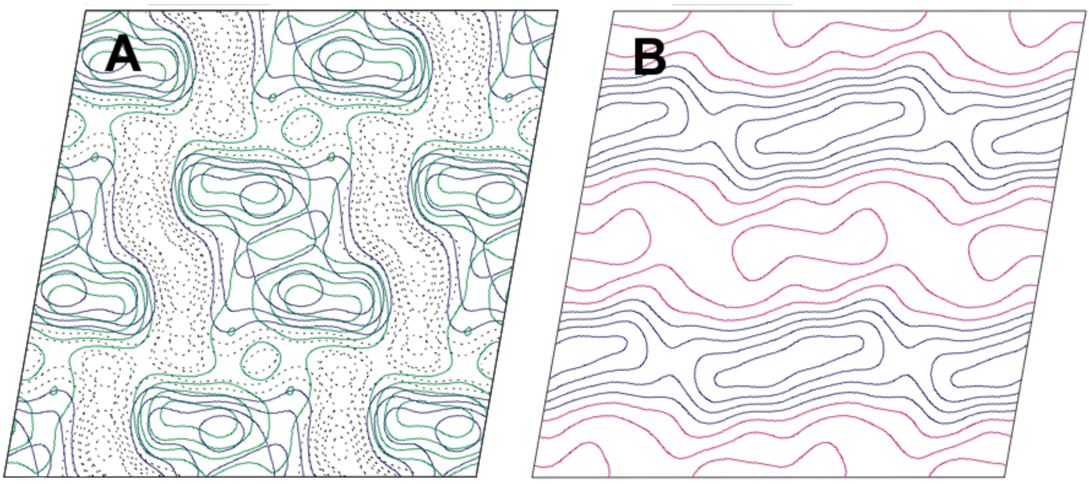


Figure 4.10 Overlaid map and difference map of untilted 2D projections of mutant and wild type S-layer

A: The wild type R20291 map (in green) overlaid on SlpA$^{\Delta Ld2}$ map (in blue) showing structural similarities. **B**: The difference map calculated as detailed in the text, with positive difference indicated by blue contour line. Red contours indicate negative difference. The contour level was set at 0.5 of the rms value of the average between the EM densities of the wild type and mutant maps. The maps show 2 × 2 unit cells with each unit cell having *p*2 symmetry and a = b = 80 Å and γ = 80°.

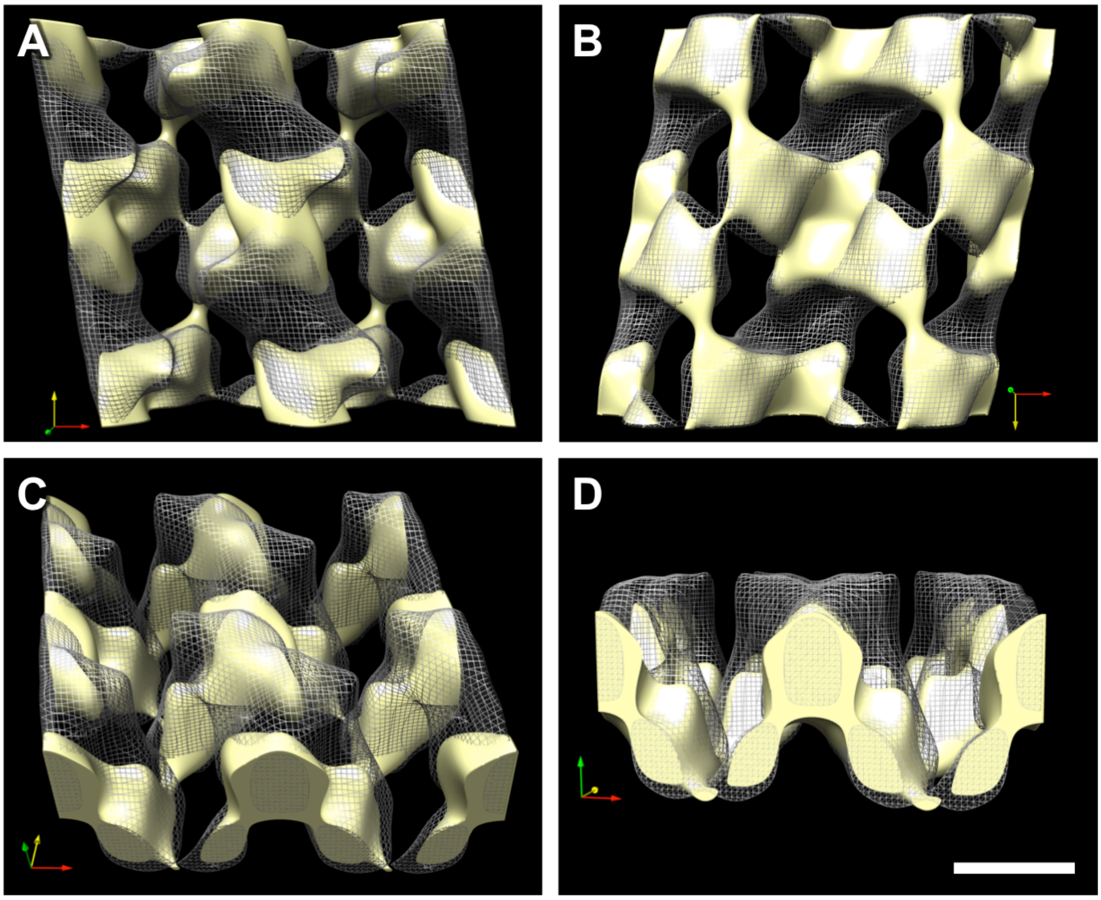


Figure 4.11 Overlaid scaled 3D volume maps of wild type and mutant S-layer

A to **D**: The wild type R20291 volume map with the mutant SlpA$^{\Delta Ld2}$ volume overlaid. The wild type map's surface is represented by grey wireframe and the mutant volume is represented by the yellow surface. A clear void of absent density represented by the wireframe without associated yellow surface is visible in the overlay. In **A**, the location of the absent density corresponds entirely to the density in the difference map calculated in Figure 4.10B. The scale bar indicates 4 nm and applies to all panels. **A** and **B** are either the top or bottom views of the layer. The X, Y and Z-axes are represented as red, yellow and green arrows, respectively. *p*2 symmetry was applied, with symmetry axis along Z.

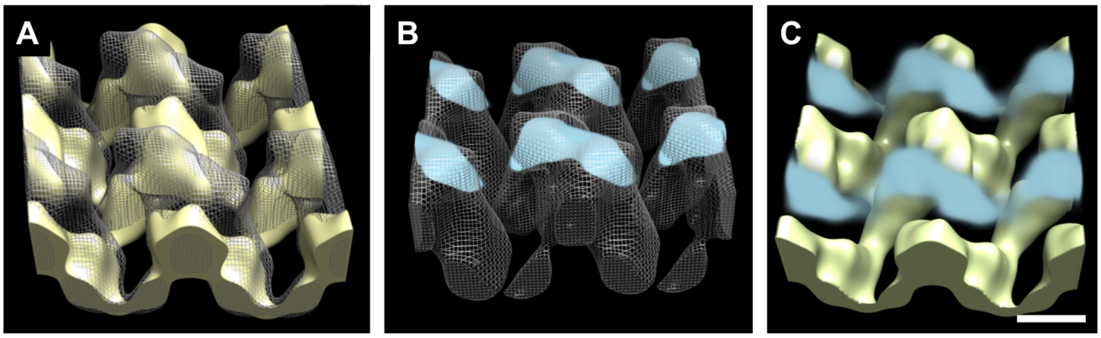


Figure 4.12 Overlaid 3D volume maps and real-space difference in volume shown superimposed on wild type and mutant volumes

A: The wild type R20291 volume map with the mutant SlpA$^{\Delta Ld2}$ volume overlaid. The wild type map surface is represented by grey wireframe and the mutant volume is represented by the yellow surface. A clear void of absent density is visible in the overlay. **B**: The wild type volume map in grey wireframe overlaid with blue density, which represents the real-space difference between the wild type and mutant volumes in **A**. **C**: The difference density is overlaid now on the yellow surface representation of mutant S-layer volume map. The scale bar in **C** is 4 nm and applies to all the panels.

The crystal structure of the truncated LMW SLP of strain 630 provides an indication of the location of the domain 2. The LMW SLP structure in R20291 was simulated *in silico* by using the *PHYRE 2* protein fold recognition tool (Kelley et al, 2015). The predicted *PHYRE* structure showed similar organisation of the domain 2, as in the strain 630 LMW structure. Introducing a Gly-Gly residue closed the open ends, following the deletion of domain 2. Whether self-assembly would be prevented by this deletion was not known at this point.

4.3.2. Homologous recombination and mechanism

The mechanism of homologous recombination is conserved across all three domains of life. Because of the mechanism involved and the nature of the universal recombinase enzyme complex, recombination can only occur between identical sequences of DNA in close proximity. Recombination replaces the recipient gene, at the site of recombination, with the donor gene. This process can be manipulated *in vitro*, by introducing an antibiotic-resistance marker containing plasmid, with the donor gene. With the selective pressure, the successfully recombined bacteria would survive in the antibiotic containing media. In a plasmid-induced recombination, however, the recipient gene remains on either side of the recombined (now linear) plasmid, which integrates into the bacterial chromosome at any site along the two identical sequences between the donor and recipient gene.

The experiment to localise domain 2 was made possible by introducing an antibiotic-resistance marker containing plasmid (pOB001), with a wild type *slpA* gene lacking LMW domain 2 (*slpA*$^{\Delta Ld2}$), via conjugation with *E. coli*. The cells in which homologous recombination did occur, within the *slpA* gene, survived the antibiotic stress that was present in the media. Also, the presence of a full-length, functional *slpA*$^{\Delta Ld2}$ gene in the cytoplasm of the *slpA*$^{-}$ mutants meant they produced cells which were phenotypically and morphologically different, often showing a better growth rate, forming visibly larger colonies. However, because of the deletion in LMW domain 2, this recombination gave rise to two potential outcomes, with recombination occurring either upstream or downstream to the deletion sequence.

In the case of recombinant FM2.5 cells, due to the residual *slpA** gene on either side of the newly recombined pOB001 DNA, an additional intra-chromosomal homologous recombination occurred. The similarity in sequence between the *slpA** and $slpA^{\Delta Ld2}$ genes, and the significant distance between the two the sequences, made a second recombination possible.

In the case of recombined FM2.5 with pOB001, the identical sequences are present within the recipient gene and donor gene on the linearised plasmid (Figure 4.3). There are, however, two potential outcomes of this second intra-chromosomal recombination. One possible outcome results in the ejection of the complete pOB001 plasmid and a return to the $slpA^{-}$ genotype. The other possibility results in the ejection of the *slpA** gene, along with the rest of pOB001, and a retention of the functional $slpA^{\Delta Ld2}$ gene. Upon removal of antibiotic selection pressure, the recombined colonies are allowed to survive. After several rounds of division, the plasmid DNA is effectively lost. Colonies with the functional $slpA^{\Delta Ld2}$ gene were expected to have a different colony morphology when compared to those of FM2.5. Several larger colonies were selected and tested by PCR to eliminate those showing the first outcome of intra-chromosomal recombination. This process allowed for successful manipulation of the *slpA* gene, with mutated but still functioning $slpA^{\Delta Ld2}$ gene.

4.3.3. Self-assembly

The deleted region was chosen, not only because of its biological significance, but also because of its lack of interaction with the HMW domain and separation from the signal peptide and HMW cleavage sequence. These considerations meant the overall structure of the HMW/LMW$^{\Delta\text{Ld2}}$ complex formation was least likely to be affected.

As is clear from the results, the deletion did not have any detrimental effects on self-assembly and gave rise to a paracrystalline S-layer complex, with significant similarity of structure to the wild type. Hence, the LMW domain 2 does not appear to possess any crystallisation domains necessary for self-assembly. The self-assembly, as is clear from the overlaid structure of the wild type S-layer and the mutant, is contributed by, either or both the HMW SLP and the remaining LMW domain 1. Further mutational studies might reveal further insight into this process.

4.3.4. Morphology of the SlpA$^{\Delta Ld2}$ mutant

Previous publications have not reported any link between the S-layer and cell morphology. However, the SlpA^{-} FM2.5 cells appear short and curved (Kirk et al, 2017b). It is not known why the absence of the S-layer protein, SlpA, results in such morphological defects. The "revertant" FM2.5R and FM2.6R show a return to the wild type morphology (Kirk et al, 2017b). Interestingly, the SlpA$^{\Delta Ld2}$ mutant shows a morphology that lies between that of FM2.5 and the wild type. The cells are longer than FM2.5, but still shorter than the wild type, and slightly curved. This observation seems to suggest that the lack of LMW domain 2 does not allow the size and shapes of the cells to fully 'recover' to its native state. Why that should be the case is not clear, unless the full length S-layer protein does have a role in cellular morphology. Further work might be needed to delve deeper into the relationship between the S-layer and cell shape, and understand if such a correlation exists, not only in *C. difficile,* but also in other bacteria.

4.3.5. The structure

The unit cell dimensions of the 2D projection map of the SlpA$^{\Delta Ld2}$ mutant was found to be $a = b = 80$ Å and cell angle, $\gamma = 100°$. This shows remarkable similarity with the unit cell dimensions seen in the wild type projection map. The two maps also have the same $p2$ plane group symmetry. The two maps were overlaid such that common density peaks in the two maps overlapped. This overlay was a basis for establishing an origin, based on which the 3D models were calculated. The overlaid map is shown in Figure 4.10.

For the SlpA$^{\Delta Ld2}$ mutant, a $p2$ symmetry suggested the presence of two-asymmetric units in each unit cell – each asymmetric unit containing the HMW and LMW domain 1 forming a complex. The size of the HMW/LMW$^{\Delta Ld2}$ complex is ~71 kDa.

Using the calculation performed in Section 3.3.2, the volume of the complex can be estimated to be 87.3 nm^3. Since the volume map displays 4 unit cells, each with two asymmetric units, the final expected volume of the 2 × 2 unit cell map is 698400 Å^3. This volume fitted very well within the 3D map generated from the data. The volume maps shown in Figure 4.9 are a representation of a surface stretched over the EM density displaying the topographic features visible at this resolution. The volume encompassed by this surface was thresholded to enclose the calculated volume.

Figure 4.11 shows the 3D volume map of the SlpA$^{\Delta Ld2}$ mutant S-layer overlapped with that of the R20291. The two maps show several matching structural features when overlapped. The deletion of the LMW domain 2 did not generate significant structural deformation. However, as shown in Figure 4.11 panel D, the similarity extends only to about half of the structure, towards one face, and disappears further along the axis, towards the other face. The two data produced maps with the same handedness, and the identical structural features were present on similar sides of the volume maps. This validated the correct alignment of the two models and allowed the difference in the volume to be calculated from the two maps.

The overlapped volume maps show a clear missing void of density, visible from the absence of an underlying surface, under the grey wireframe representation of the wild type in Figure 4.11. In panel A the missing density agrees entirely with the additional density as calculated from the difference map (Figure 4.10). In the projection map, the difference is represented as blue contour lines. In 3D, the missing density is represented by the blue volume, which is calculated after real-space subtraction of the scaled volumes of the wild type and mutant maps in Figure 4.12.

4.3.6. Orientating the EM structure

From both the 2D and 3D representations (Figure 4.10 and Figure 4.12), the substantial similarities evident on one face of the overlapped structures provide an orientation to the model. It is clear that the deletion of the LMW domain 2 results in a significant missing density in the mutant S-layer on the other face of the structure. The assignment of the obverse and reverse face to structure depends on the known biological significance of the deleted domain.

Out of the three identifiable structural motifs in SlpA (the signal peptide region, the LMW SLP, and the HMW SLP), the domain 2 of the LMW SLP contains the hyper-variable region which shows remarkable sequence diversity among strains (Calabi & Fairweather, 2002). Immunoblot analysis using serum from patients with antibiotic associated diarrhoea indicated the presence of IgG against the LMW SLP among several of the patients tested (Pantosti et al, 1989). Hyper-variability of amino acid sequences generally indicates either resistance against pathogens like phages, or the subversion of host immune response, or both. The inconstancy of the region and its immunodominance indicates it is exposed to the environment.

In the structure described here, the deleted region is therefore the side facing the external environment, and located away from the underlying cell wall. Conversely, the reverse side of the structure is therefore adjacent to the cell wall and is non-covalently anchored to it. The three cell wall binding domains, CWB2, located tandemly in the HMW SLP, are hence present on this reverse side. The crystal structure of the three CWB2 domains was reported for *C. difficile* cell wall proteins Cwp6 and Cwp8, which share homology to the HMW SLP (Usenik et al, 2017). This binding region, is therefore located on the opposite side of the structure (facing the reader in Figure 4.11, panel B). A figure showing the 3D model of R20291 S-layer complemented with the three CWB2 domains from Cwp8 is shown in Figure 7.3 and discussed in Section 7.1.3.

Chapter 5 - Cryo-electron microscopy on S-layer fragments

5.1. Introduction

The structural analysis made possible using genetic modifications and electron crystallography, described in Chapters 3 and 4, lead to the possibility of structural work on these *in situ* S-layer assemblies using cryo-EM. This has the potential to give a higher resolution image of the S-layer in a native and hydrated state. After imaging the S-layer using negative stain, the identical sample was used to prepare cryo-EM grids. However, the method developed to image the S-layer 'ghosts' for negative stain was found to be unsuitable for cryo-EM. In negative staining the sample is dehydrated following the staining procedure. When frozen in vitreous ice, the 'ghosts' appeared as dense objects, presumably due to retention of the hydrated state. A different sample preparation technique was developed to get thin fragments of S-layer, suitable for imaging in ice, without damaging its structure.

5.2. Results

5.2.1. Sample preparation and vitrification

Several attempts were made using different sample preparation techniques to obtain suitable S-layers fragments for imaging in vitreous ice. A modified protocol from (Kawata et al, 1984) was found to produce fragments of S-layers with the underlying cell wall. The method developed, after considerable optimisations to suit cryo-EM, involved mechanical fragmentation and purification of the cell envelope. The complete procedure, described in Section 2.2.2, involved the use of a non-ionic detergent to solubilise the underlying cell membrane. The S-layer/cell wall fragments were left suspended in the detergent, reducing the surface tension and allowing the suspension to form a thin layer on the EM grid. This was followed by immediate cryo-fixation on a cryo-EM grid (described in Section 2.5.3). This protocol was used to isolate, purify and image the S-layer fragments from both

R20291 and the SlpA$^{\Delta Ld2}$ mutant. The frozen grid was stored in a cryo-storage box in liquid nitrogen for later observation. Images described below were obtained on a Falcon II direct electron detector (FEI) using Tecnai F20 (FEI) TEM microscope operating at 200 keV.

5.2.2. Microscopy and image processing

The sample was screened under the microscope initially using negative stain at room temperature. The R20291 and SlpA$^{\Delta Ld2}$ S-layer fragments were similar in appearance. Figure 5.1 shows the SlpA$^{\Delta Ld2}$ S-layer fragments in negative stain at low magnification, next to a higher magnification image with inset FFT. The figure also shows a single-layered fragment, indicated by a distinct square-ordered lattice visible in the FFT. Figure 5.2 shows the low magnification micrograph of SlpA$^{\Delta Ld2}$ S-layer fragments in ice and a high magnification micrograph imaged using low-dose technique of a double-layered fragment with the FFT. Images of S-layer fragments were taken suspended in ice, among holes of the Quantifoil carbon film. The sample appeared to form single-layered or double-layered tubular structures in thin ice. The two could be differentiated based on the presence of overlapping lattices visible in the FFT and the higher contrast of the double-layered fragments in the images (Figure 5.2B and Figure 5.3A).

Crystalline order was well preserved for most samples. After unbending, reflections in the FFT of the best image extended to ~5 Å resolution for the wild type S-layer fragment, and ~7 Å for the mutant S-layer fragment. The resolution limit for the micrograph, given by the Nyquist frequency, is 4.16 Å. An image of a single-layered S-layer fragment, in a hole on the Quantifoil grid, is shown in Figure 5.3. The single-layered fragments appeared to have very low contrast compared to the double-layered structure (Figure 5.2).

Image processing (Section 2.6) was carried out using the best images of 8 individual crystals of the wild type and 9 individual crystals of the mutant S-layer. The quality was determined based on the intensity of the high-frequency spots in the FFT.

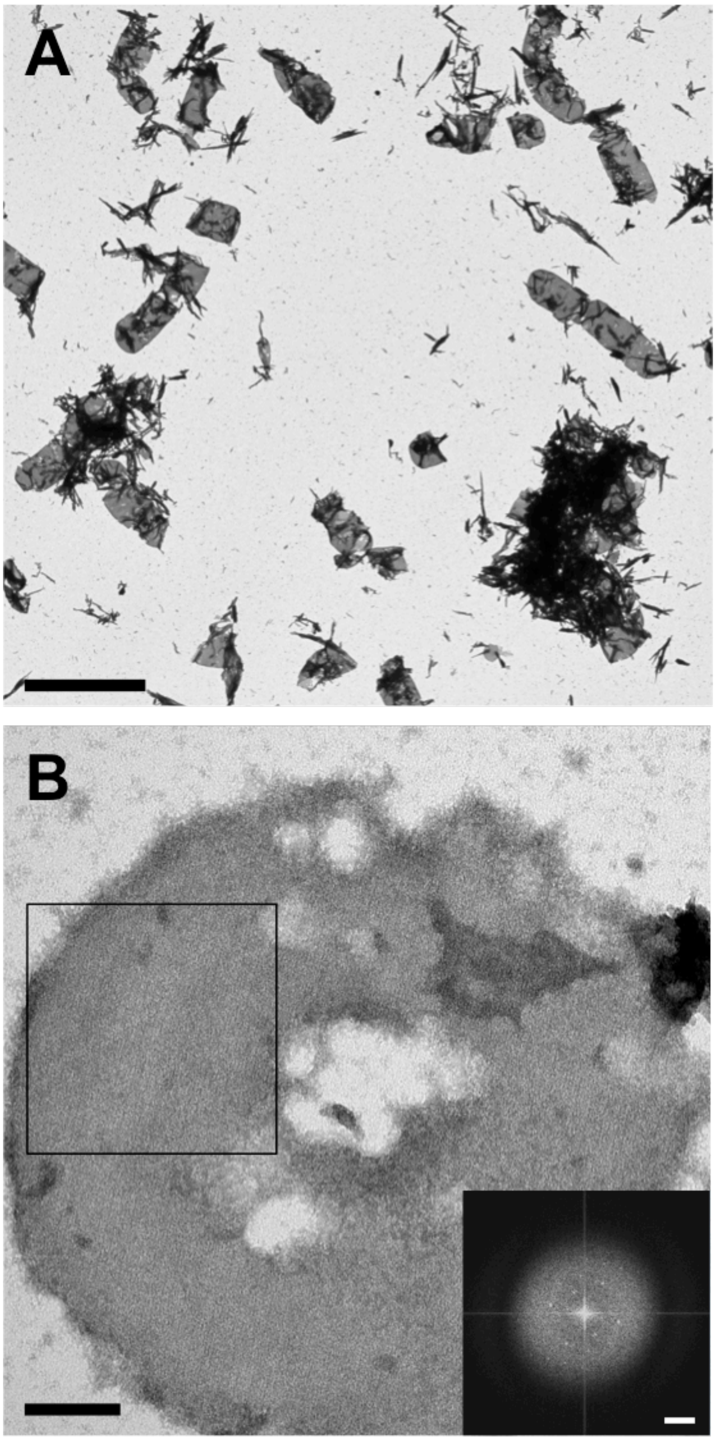


Figure 5.1 S-layer fragments from SlpA$^{\Delta Ld2}$ S-layer mutant in negative stain

A: Low magnification image of negatively stained S-layer fragments obtained by cell fragmentation technique from SlpA$^{\Delta Ld2}$ whole cells. **B**: Single layered fragment at high magnification, with the inset FFT of the square region, indicated by the black outlined box. The spots visible in the FFT indicate a single-layered lattice. Scale bars are 4 µm for panel A and 200 nm for panel B. Scale bar in inset FFT indicates 0.25 nm^{-1}.

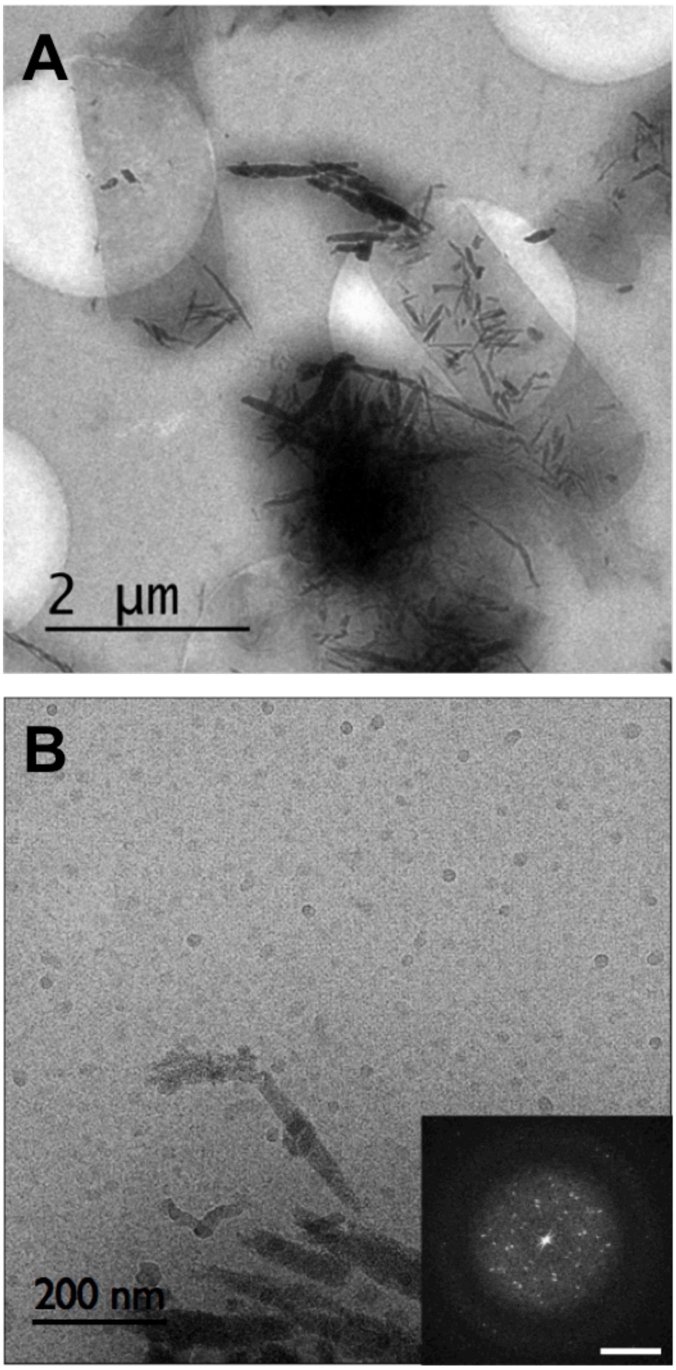


Figure 5.2 S-layer fragments in vitreous ice

A: The S-layer fragments extracted by homogenisation from SlpA$^{\Delta Ld2}$ mutant cells in vitreous ice on a Quantifoil® 2/2 grid. The capsular structures are fragments of S-layer, which retain their original shape. The dark objects are smaller pieces of fragments that are folded tubes. **B**: A high magnification image of a double-layered fragment with a FFT showing reflections in the central region. The FFT was binned and the brightness adjusted to show the low frequency spots. Scale bar in FFT is 0.25 nm^{-1}.

The density maps of 2 × 2 unit cells were calculated from merged structure factors from the FFT of the individual images, after restricting phases to either 0 or 180°. The resolution of the final map was determined by the average phase errors of the reflections in set resolution bands of between 40 and 50 reflections. Owing to the centro-symmetry, spots showing phase of around 90° would reflect noise, hence a phase error of 45° or more would indicate the phases from noise were merged in that resolution band. Based on this criterion, calculations were terminated at the resolution indicated by good phase residuals.

5.2.3. Projection structure of R20291 S-layer fragments in ice

Images of R20291 S-layer fragments were taken for both double-layered and single-layered fragments. Image processing, described in Section 2.6, was used to obtain information on the S–layer structure. The data obtained from the best 8 images showed unit cell dimensions remained similar to those from negatively stained fragments. Average unit cell dimensions were a = 85 ± 2 Å, b = 86 Å ± 1 Å and γ = 98° ± 1°. Table 5.1 shows the lattice dimensions for individual images. Figure 5.3A shows one of the micrographs displaying a single-layered crystal, located in a hole in the Quantifoil grid. Figure 5.3B shows an IQ plot showing the reflections up to 5 Å with the Thon rings indicating where the phase contrast transfer function crosses zero. The H and K axes of the reciprocal lattice are shown in the figure. The reflections are shown by a number and colour denoting its IQ value (Henderson et al, 1986), representing the SNR of the reflection amplitude. Lower value of IQ represents spots with higher SNR.

Figure 5.3C shows the projection map generated from the crystal in panel A as a contour plot, without any symmetry applied. The respective temperature factor calculated from *SCALIMAMP3D* is applied.

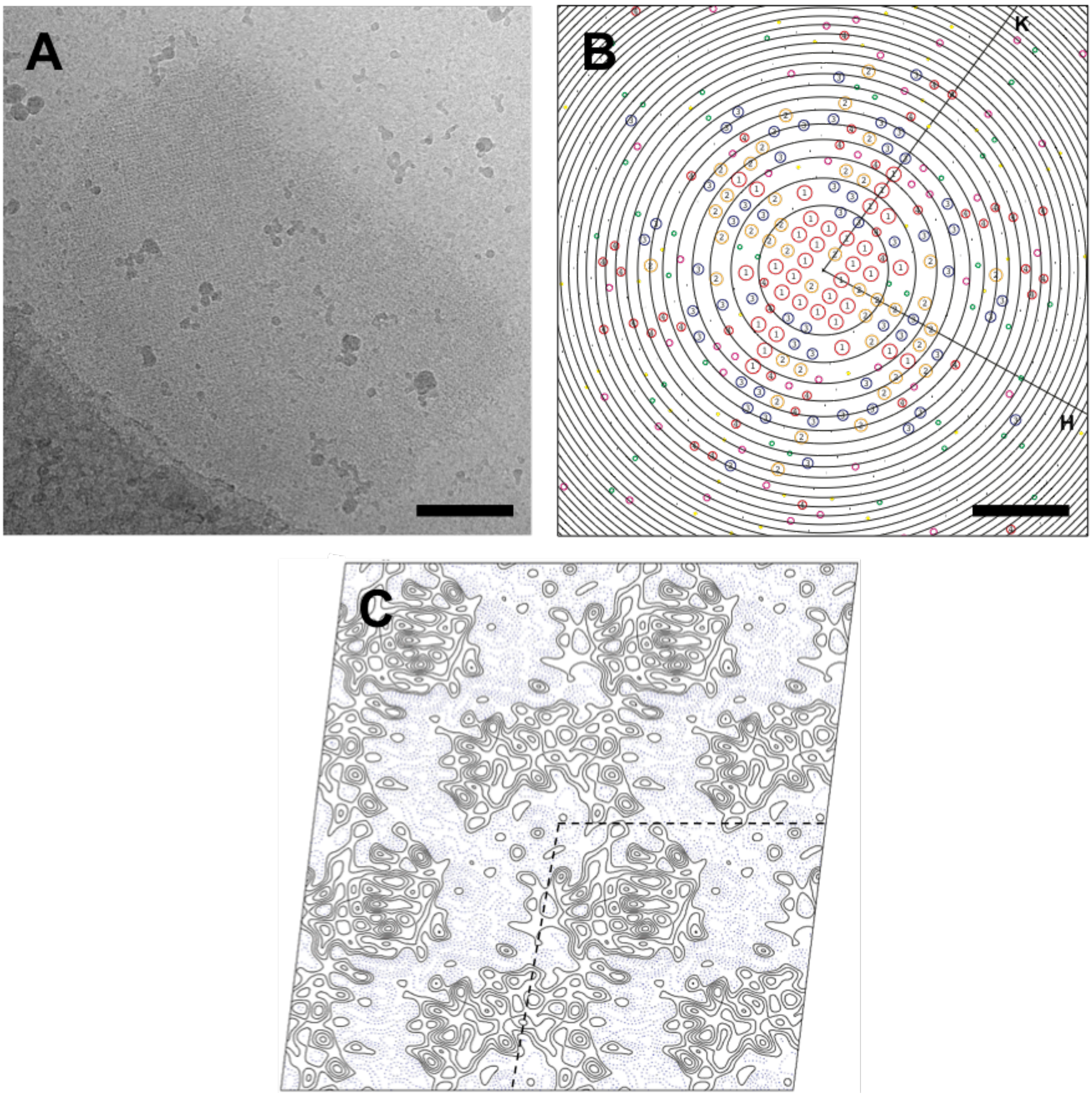


Figure 5.3 Image of a single-layered S-layer fragment with corresponding IQ plot and contour plot of final map in p1

A: A high magnification image of a single-layered S-layer fragment of R20291 in a hole in ice (scale bar: 200 nm). **B**: The corresponding IQ plot in reciprocal space obtained after unbending **A**, with nyquist of plot at 5 Å. The circles in the image represent the Thon rings, where the value of the phase contrast transfer function crosses zero. The colour and size of each circle represents the IQ of the spot, as defined by 7/grade, with the grade representing the SNR of the spot amplitude. Lower IQ values have higher amplitude compared to the noise. The H and K represent the two axes of the reciprocal lattice (scale bar 0.625 nm^{-1}). **C**: The contour plot of the final map from the image in **A**, without any symmetry applied (p1) to it. The plot shows 2 × 2 unit cells, with the dashed lines representing a single unit cell of 85 × 87 Å cell length and 97° cell angle. Dotted blue lines show the negative contours.

The intensity of Fourier components decreases with increasing resolution as shown in a Guinier plot of the merged intensities from the data (Appendix 8A). The amplitudes at high spatial frequencies were sharpened, in each image, by applying a negative temperature factor or B-factor, using the programme *SCALIMAMP3D*. A maximum B-factor value of 1000 was used. The B-factor values for each image are listed in Table 5.1.

An examination of the symmetrised phase residuals of several images, using *ALLSPACE*, indicated an unequivocal two-fold (*p*2) symmetry. The symmetrised phase residuals for a single representative crystal are shown in Table 5.2. The images were merged with *p*2 symmetry. Table 5.3 shows the phase error of the averaged data from the 8 images (with phases set to either 0 or 180°) and the number of reflections, in resolution bands of 40 – 50 reflections each. The mean phase error for resolution bands above 7.5 Å was higher than 45°, and therefore had contributions from random noise. The calculation was therefore terminated at 7.5 Å.

A merged symmetrised map was calculated from 176 structure factors, with phases set to either 0 or 180°, truncated to 7.5 Å resolution (Figure 5.4). Figure 5.4A shows the density plot with EM density variation represented in shades of grey over black background. Figure 5.4B shows the map in contour plot where each contour line connects X,Y coordinates with the same EM density. The *p*2 symmetry axis is perpendicular to the paper towards the reader. The map generated has unit cell dimensions of 85 × 85 Å and 98°.

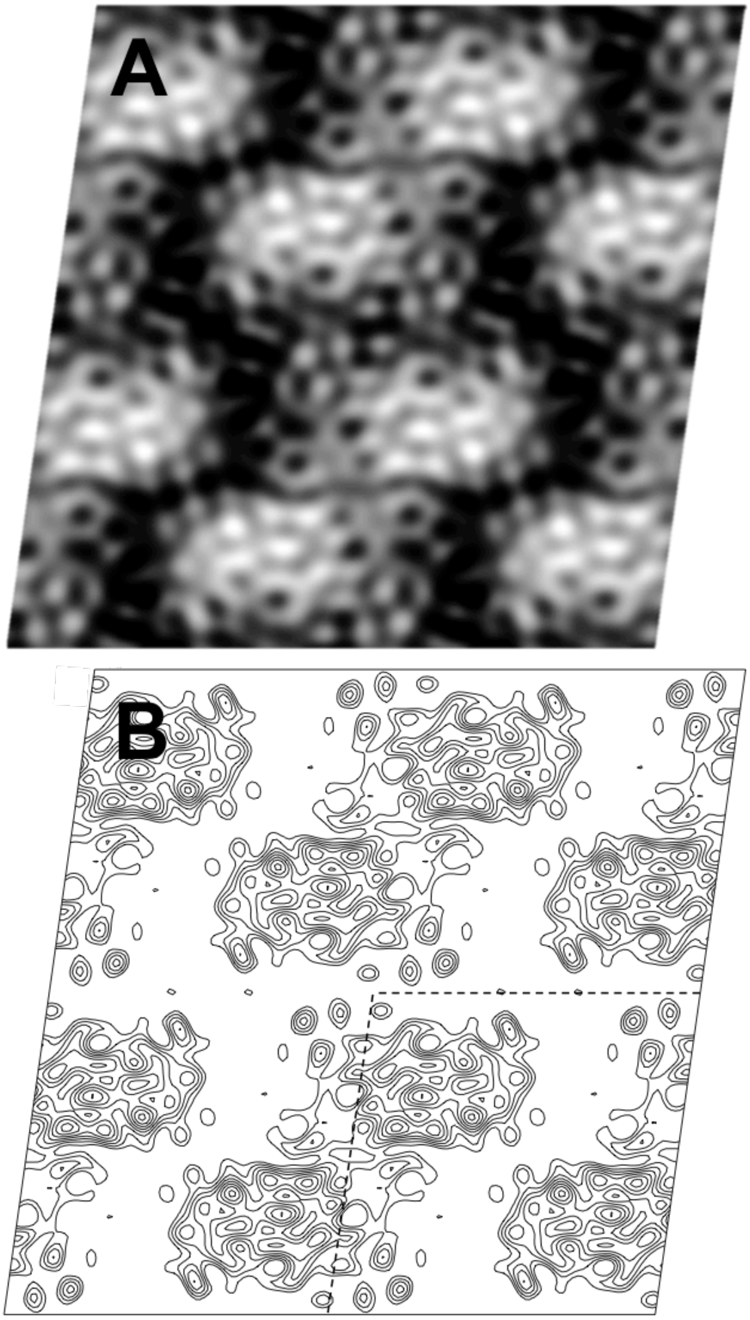


Figure 5.4 Merged projection map in *p*2 for R20291 S-layer at 7.5 Å resolution

Final projection map generated from merged dataset of R20291 (WT) S-layer fragment images with $p2$ symmetry imposed. Resolution was truncated at 7.5 Å and temperature factor was applied using *SCALIMAMP3D*. **A** shows 2 × 2 unit cells with the EM density in grey scale, with proteins appearing as white over a black background. Higher density appears as lighter shades of grey. **B** shows 2 × 2 unit cells as a contour plot. The contour line connects X,Y coordinates where EM density value is the same. Dashed line represents one unit cell of length $a = b = 85$ Å and 98° cell angle. The symmetry axis is perpendicular to the page.

Table 5.1 Summary of data for each image of R20291 S-layer fragments merged to resolution of 7.5 Å

Image number	a (Å)	b (Å)	γ (degrees)	Overall phase residual (degrees)	Defocus* (nm)			Temperature Factor**
					x	y	degrees	B_{xy}
300	87.7	84.3	98.8	32.8	38877	38611	37	86.5
800	85.3	87.6	97.7	22.8	20150	20037	35	869.5
900	85.6	86.1	96.9	31.99	22189	22094	0	1011.0
1000	85.8	86.7	98.0	38.59	30577	30197	30	887.3
1100	84.9	87.6	97.4	38.85	27553	27361	10	910.1
1200	84	88.3	99.6	32.4	25114	24531	25	813.6
1300	82.2	87.0	99.2	31.71	23448	23176	36	854.5
1600	87.5	85.1	98.9	38.27	25337	24834	16	955.3

Note: *All images were underfocussed. **Maximum B-factor used was 1000, any values above that were curtailed to 1000.

Table 5.2 Phase residual table for different plane groups for a representative untilted crystal of R20291

Plane Group[1]	Phase residual (random = 90°)	Target residual[2]	Plane Group[1]	Phase residual (random = 90°)	Target residual[2]
p1	27°		p22121	62.2°	33.6°
p2	37.5°*	39.7°	c222	67.7°	33.6°
p12b	77.6°	29.8°	p4	46.8°	33.6°
p12a	55.1°	29.8°	p422	61.9°	30.1°
p121b	57.2°	29.8°	p4212	63.5°	30.1°
p121a	51°	29.8°	p3	52.5°	27°
c12b	77.6°	29.8°	p312	60.4°	27.7°
c12a	55.1°	29.8°	p321	63°	28.5°
p222	67.7°	33.6°	p6	50.4°	31.6°
p2221b	52.7°	33.6°	p622	58.1°	29.3°
p2221a	47.5°	33.6°			

[1] a and b represent the respective symmetry axis for the plane group

[2] Target residual indicates the expected phase residual of each symmetry group based on the SNR of the spots from the concerned image

Symbols for possible plane groups: * = acceptable, ! = should be considered, ` = possibility

Table 5.3 Comparing average error of centro-symmetric phases in resolution bands for R20291 merged data

Resolution range (Å)	Number of phases	Mean phase residual (degrees)	Standard error of mean phase residual	Mean figure of merit
∞ – 15	43	15.3	2.4	0.9
15 – 11	42	22.9	2.8	0.8
11 – 8.7	47	32.2	3.6	0.6
8.7 – 7.5	44	36.8	3.9	0.5
7.5 – 6.7	42	50.7	4.3	0.4
∞ – 7.5	176	27.0	1.7	0.7

Note: Resolution bands were chosen to include between 40 and 50 phases. Spots up to IQ 8 were used. Mean phase residuals of 45° and above are from random phases.

5.2.4. Projection structure of SlpA$^{\Delta Ld2}$ S-layer fragments in ice

The SlpA$^{\Delta Ld2}$ mutant was treated with the same fragmentation technique as described in Section 2.2.2. Imaging was done at liquid nitrogen temperatures, at different defoci, using an electron dose less than 20 $e^-/Å^2$. Images of both single and double layered fragments were collected, and projection maps were calculated for the SlpA$^{\Delta Ld2}$ S-layer fragments, using techniques described in Section 2.6.

Reflections in the best image extended to about 7 Å after unbending. A total of 9 micrographs were selected for image processing. Average unit cell dimensions were a = 85 ± 2 Å, b = 86 Å ± 1 Å and γ = 98° ± 1°. Lattice dimensions and defoci for each individual image are summarised in Table 5.4. An examination of symmetrised phase residuals of the images, using *ALLSPACE*, indicated a two-fold ($p2$) symmetry. For a representative crystal, the symmetrised phase residuals for different plane groups are shown in Table 5.5.

A Guinier plot, showing the intensity of merged reflections from the data, as a function of resolution, is depicted in Appendix 8B. The amplitudes at high spatial frequencies were sharpened, in each image, by applying a B-factor in *SCALIMAMP3D*, using bR as a reference. A maximum B-factor of 1000 was used.

The B-factor values for each image are listed in Table 5.4.

The best 8 images were merged with *p*2 symmetry. Table 5.6 shows the mean phase errors (after setting phases to either 0 or 180°) and the number of reflections obtained, for the merged data, averaged from the 9 micrographs of individual crystals, in resolution bands containing 40 – 50 reflections. Phase errors of 45° or above indicate random phases from noise.

From the mean phase error data a maximum resolution of 8.7 Å was used and calculation terminated at that resolution. A *p*2-symmetrised merged map was calculated from a total of 130 structure factors with (phases set to either 0 or 180°), truncated to 8.7 Å resolution (Figure 5.5). Panel A shows the density plot and Figure 5.5B shows the map in contour plot. The *p*2 symmetry axis is perpendicular to the paper towards the reader. The map generated has unit cell dimensions of 85 × 85 Å and 98° and the dashed line indicates one unit cell.

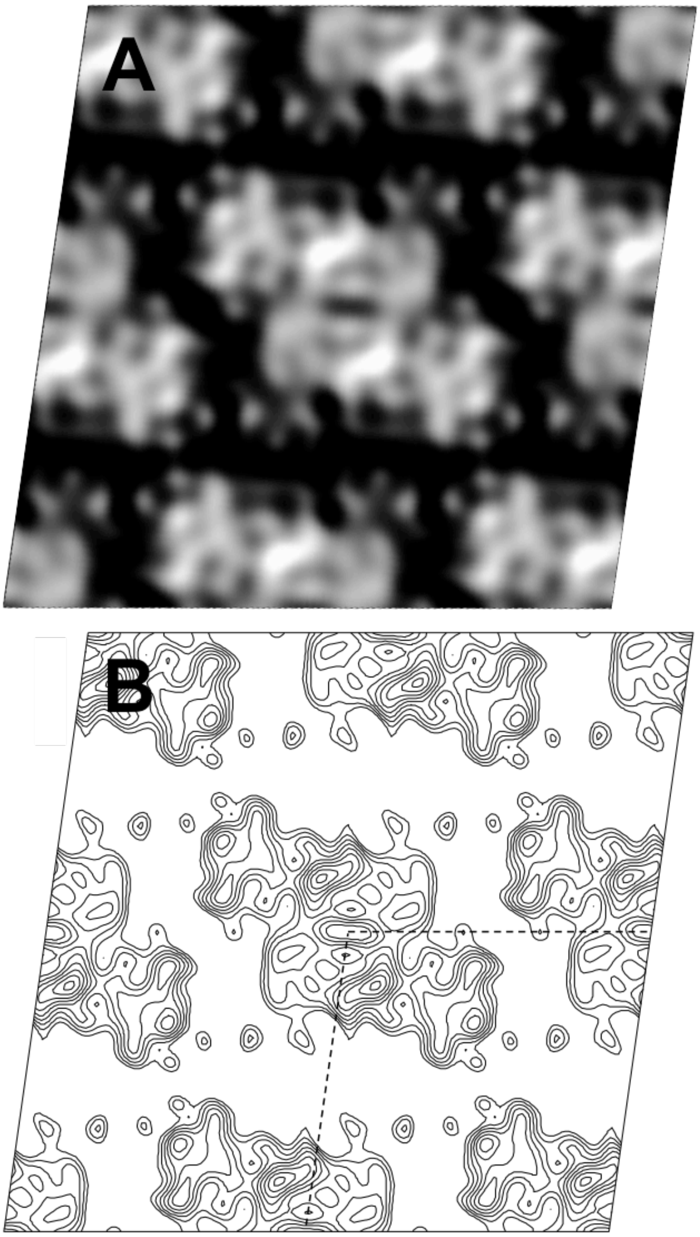


Figure 5.5 Merged projection map in *p*2 for SlpA$^{\Delta Ld2}$ S-layer at 8.7 Å resolution

Final projection map generated from merged dataset of SlpA$^{\Delta Ld2}$ S-layer fragment images with *p*2 symmetry imposed. Resolution was truncated at 8.7 Å and temperature factor was applied using *SCALIMAMP3D*. **A** shows 2 × 2 unit cells with the EM density in grey scale, with proteins appearing as white over a black background. Higher density appears as lighter shades of grey. **B** shows 2 × 2 unit cells as a contour plot. The contour line connects X,Y coordinates where EM density value is the same. The dashed line represents one unit cell of length $a = b = 85$ Å and 98° cell angle. The symmetry axis is perpendicular to the page.

Table 5.4 Summary of data for each image of SlpA$^{\Delta Ld2}$ S-layer fragments merged to 8.7 Å resolution.

Image number	a (Å)	b (Å)	γ (degrees)	Overall phase residual (degrees)	Defocus* (nm)			Temperature Factor**
					x	y	degrees	B_{xy}
200	87.7	84.6	97.8	36.74	22478	23061	45	1087.8
300	84.4	86.2	98.9	22.62	30331	30425	31	1217.1
400	86.5	86.0	96.9	18.92	29815	29589	15	886.0
600	84.0	87.0	99.6	38.98	20265	19859	25	1249.8
700	83.0	87.6	98.4	28.03	20655	21228	65	1212.3
900	87.5	84.8	98.0	21.72	25367	25978	70	1001.8
1000	83.2	86.1	97.0	17.41	33936	34266	64	1160.8
1100	84.1	85.9	98.2	21.22	32368	32726	50	1313.3
1200	82.8	87.2	98.3	26.00	26916	27299	47	1166.4

Note: *All images were underfocussed. **Maximum B-factor used was 1000, any values above that were curtailed to 1000.

Table 5.5 Phase residual table for different plane groups for a representative untilted crystal of SlpA$^{\Delta Ld2}$ S-layer

Plane Group[1]	Phase residual (random = 90°)	Target residual[2]	Plane Group[1]	Phase residual (random = 90°)	Target residual[2]
p1	31.8°		p22121	63.9	40.1°
p2	47.7°*	47.3°	c222	65.7	40.1°
p12b	71.6°	36.3°	p4	57.7	39.7°
p12a	56.9°	35.6°	p422	67.4	35.8°
p121b	49.7°	36.3°	p4212	64.1	35.8°
p121a	62.4°	35.6°	p3	51.7	31.8°
c12b	71.6°	36.3°	p312	60.6	32.7°
c12a	56.9°	35.6°	p321	59.7	33.9°
p222	65.7°	40.1°	p6	57.2	37.3°
p2221b	62.8°	40.1°	p622	64.2	34.6°
p2221a	52.1°	40.1°			

[1] a and b represent the respective symmetry axis for the plane group

[2] Target residual indicates the expected phase residual of each symmetry group based on the SNR of the spots from the concerned image

Symbols for possible plane groups: * = acceptable, ! = should be considered, ` = possibility

Table 5.6 Comparing average error of centro-symmetric phases in set resolution bands for SlpA$^{\Delta Ld2}$ merged data

Resolution range (Å)	Number of phases	Mean phase residual (degrees)	Standard error of mean phase residuals	Mean figure of merit
∞ – 15	43	21.5	3.5	0.8
15 – 11	42	28.7	4.1	0.7
11 – 8.7	45	33.8	3.6	0.6
8.7 – 7.5	43	45.2	3.8	0.4
∞ – 8.7	130	28.0	2.2	0.7

Note: Resolution bands were chosen to include between 40 and 50 phases. Spots up to IQ 8 were used. Mean phase residuals of 45° and above are from random phases.

5.2.5. Difference map

The R20291 projection map was truncated to 8.7 Å resolution and scaled in Fourier space with the SlpA$^{\Delta Ld2}$ projection. The scaling was performed in reciprocal space, where the structure factor amplitudes for each of the two maps were multiplied by a common scaling factor. The two maps were then subtracted in real-space to display the difference. Figure 5.6A shows the contour plots of the scaled R20291 and SlpA$^{\Delta Ld2}$ maps overlapped. The wild type map is shown in green while the mutant map is in blue. A region showing the wild type density (in green) is visible in between the overlapped regions. The calculated real-space difference in EM density is shown in red contour lines in Figure 5.6B along with the overlapped mutant map. The difference map is overlaid with the R20291 and SlpA$^{\Delta Ld2}$ maps in Figure 5.6C and D. A distinct region of difference is seen in the map located on the axis of symmetry. This region coincides with the missing volume seen in negative stain data (Figure 4.11).

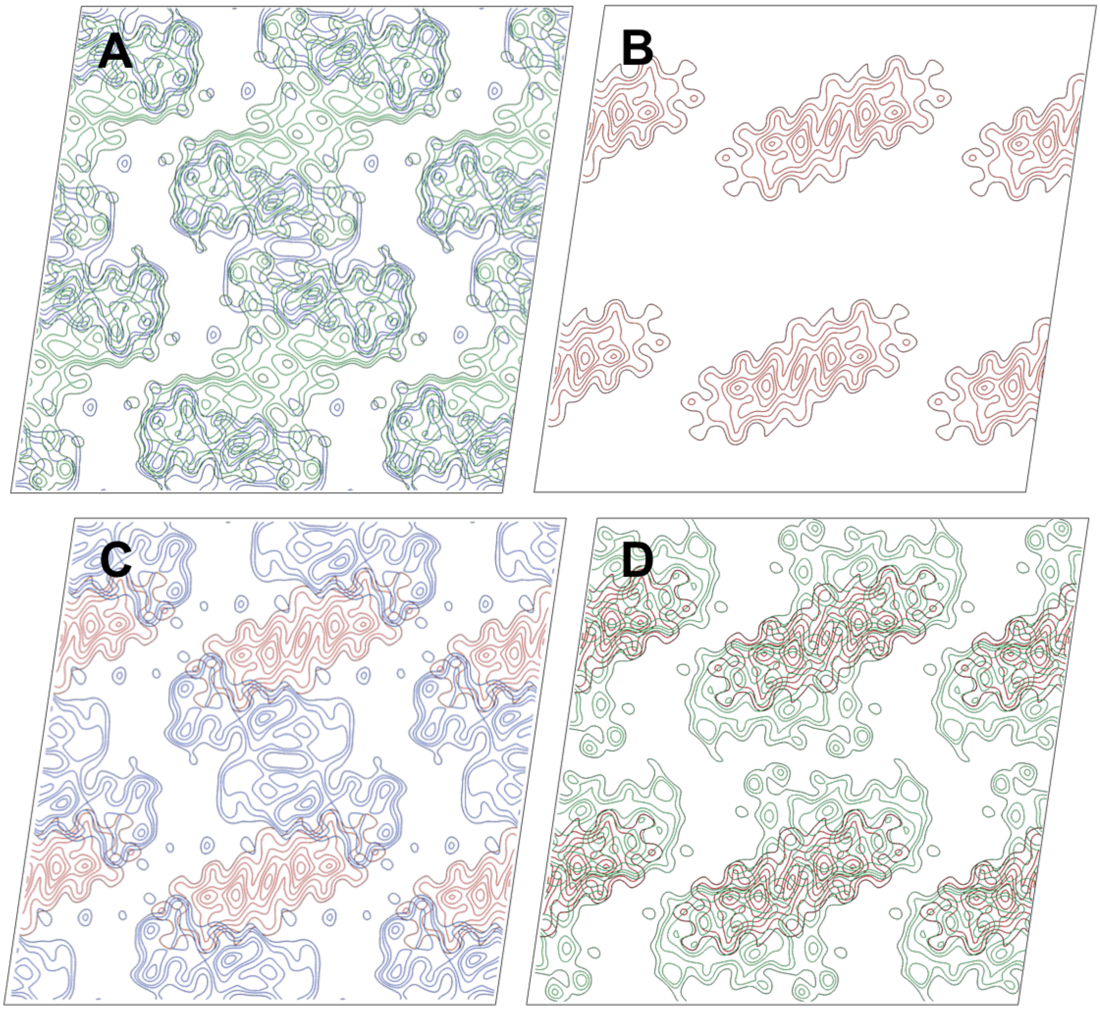


Figure 5.6 Overlaid scaled map and difference map between R20291 and SlpA$^{\Delta Ld2}$ S-layer Merged 2D projection maps of R20291 and SlpA$^{\Delta Ld2}$ S-layers were scaled in reciprocal space. **A**: The two *p*2 symmetrised scaled maps overlaid (R20291 in green and SlpA$^{\Delta Ld2}$ in blue contour lines). Contour lines represent density in X,Y with the same value. **B**: The real-space difference in density (red lines) between the two maps in **A**. The SlpA$^{\Delta Ld2}$ map is overlaid with the difference map in **C** and the R20291 map and difference maps are overlaid in **D**. All maps are 2 × 2 unit cells of a = b = 85 Å and $\gamma = 98°$. The axis of symmetry is perpendicular to the page.

5.3. Discussion

5.3.1. Sample preparation and Microscopy

The S-layer extracts from endolysin treated whole cells were difficult to image in cryo-EM. The sample appeared dense in the vitreous ice. This problem persisted even after prolonged blotting time on the EM grid. It may be the result of hydration of the tubular extracts due to being preserved in a hydrated state. Negative staining, on the other hand, involves drying the sample after staining. This drying process results in collapse of the tubular extracts resulting in a thin sample, suitable for our imaging. The cryo-EM sample preparation does not involve such a dehydration step.

A modified sample preparation was therefore required that could produce fragments of S-layer, without damaging the crystalline nature of the sample. A modified method from (Kawata et al, 1984) was used. The sample preparation was optimised to obtain reasonable distribution of the fragments in a grid area with least damage to the structure (Figure 5.1). Negatively stained samples revealed well-distributed fragments, showing a tubular structure that appeared double-layered and fragments that were single-layered. The FFTs of images of the fragments showed reflections extending to 20 – 25 Å, the expected resolution limit with negative stain. Fragments treated with detergent, to remove the cell membrane, appeared to show better contrast in images. The same protocol was used to obtain S-layer fragments from wild type and mutant *C. difficile* cells. Both types of samples showed similar encouraging results.

For cryo-EM, the blotting and glow-discharge time on Quantifoil grids was adjusted to get a thin layer of ice with good sample distribution (Figure 5.2). The presence of detergent in the sample also helped in producing a uniform layer of ice across the grid. The images obtained of the sample in ice appeared to show good crystallinity with visible spots in the FFT of the raw image extending to 8 – 10 Å. Image processing and unbending steps further improved the information obtainable from the images (Figure 5.3).

Both single and double-layered fragments were imaged. However, double-layered fragments were easier to visualise in 'search mode' during low-dose imaging. At low magnification the single layers were not visible, due to low contrast, even with high underfocus. A longer exposure time was required to achieve enough cumulative dosage to obtain contrast to spot such fragments. Also, single-layered fragments were only visible suspended in ice, within the holes in the Quantifoil grid, where the background due to the carbon film was absent. This difficulty in identifying single-layered fragments can be appreciated when one looks at the crystal in Figure 5.3, which would not have been observable if it was situated on the surrounding carbon film. These single layers, although difficult to locate, provided the best images.

5.3.2. Structure

The merged maps produced from averaged phases appeared to show identical unit cell dimensions for both wild type and mutant fragments. The structures are very much comparable to that of the negative stain projection maps, showing similar unit cell dimensions (Figure 3.9 and Figure 4.8). The unit cell lengths were on an average 5 Å longer than the average lengths of the negative stain data in Sections 3.2 and 4.2, and also dimensions from the negative stained fragments obtained using the fragmentation protocol. This may be due to contraction of structures when dehydrating following negative staining. The cryo-EM data, however, revealed identical unit cell angle and symmetry. A *p*2 plane group was obtained, unambiguously, by comparing the symmetrised phase residuals in individual images, using *ALLSPACE*.

The resolution limit imposed can be indicative of a low SNR at higher resolutions or a lack of symmetry beyond that range. The lack of symmetry can be partly due to the fragmentation process used to obtain the sample. Equally likely, however, is the possibility that the native S-layers lack symmetry beyond a certain limit. Nevertheless, one crystal did show reflections up to a resolution of 5 Å, after unbending, with amplitudes above statistical noise.

5.3.3. Difference map

Both the wild type and mutant maps were calculated with the same unit cell dimensions. The two maps were scaled in reciprocal space, after truncating the wild type map to 8.7 Å, and overlaid. The overlay shows similarity in structural detail, as seen in Figure 5.6. The overlaid maps also show a clear region of density missing in the mutant.

The scaled wild type and mutant maps were then subtracted. The calculated difference map showed the density that is missing in the mutant, indicating the location of the deleted domain 2 of the LMW SLP, in projection. This is consistent with the missing region observed in the 3D volume maps obtained, as shown in Chapter 4 (Figure 4.12).

5.3.4. Significance

This is the first reported structure of an S-layer in its native state, studied and analysed using cryo-electron crystallography. Interestingly, the S-layer structures appear to assemble in domains on the cell wall. During the growth phase, several nucleation points for crystallisation of the S-layer proteins may result in such domain-based assembly (Figure 5.1B).

This method allows us to obtain single-layered crystals. These crystals not only provide much improved data quality, but also make it possible to obtain data from high tilt angles. The geometric distortions at higher tilts make it difficult to correctly index a lattice in the case of double-layered crystals. Single-layered crystals are easier to index and are also thinner, reducing the chances of spot splitting at higher tilts. This raises the possibility of obtaining a 3D cryo-EM structure of the S-layer to a high-resolution. However, challenges of locating the single layers at low-magnification, during the search mode, must be overcome.

Chapter 6 - Electron crystallography on *Bacillus subtilis* biofilm protein, BslA

6.1. Introduction

Communities of bacteria in their natural environments encase themselves in self-produced biofilms (Costerton et al, 1995; Flemming & Wingender, 2010). Biofilms are associated in the majority of infections (Costerton et al, 1999) and play a critical role in the organism's survival in natural environments such as soil and water (Costerton et al, 1987). Biofilms are composed of proteins, exopolysaccharides and nucleic acids (Flemming & Wingender, 2010). Recent examination of *E. coli* biofilms have revealed the microanatomy of the matrix in their biofilm, showing cellular differentiation into different zones with varying putative functions (Serra et al, 2013). However, not much is known about the 3D architecture of biofilms, their interaction with the bacteria producing them, and the environment.

Bacillus subtilis, produces biofilms *in vivo* and *in vitro*, linked to its survival in the natural environment of the rhizosphere (Morikawa, 2006). The bulk of the *B. subtilis* biofilm is composed of an exopolysaccharide (EPS) and an amyloid fibre-like protein, TasA. The matrix assembles in the presence of a small protein called BslA (formerly known as YuaB) (Branda et al, 2006; Kobayashi & Iwano, 2012; Ostrowski et al, 2011). The biofilm produced by *B. subtilis* has been shown to be highly hydrophobic in nature, attributed to EPS (Epstein et al, 2011) and BslA (Kobayashi & Iwano, 2012). The surface hydrophobicity of the biofilms protects the colony on the root surfaces. Confocal microscopy revealed that BslA forms an elastic film at the surface interface of the biofilm (Hobley et al, 2013). Purified BslA spontaneously self-assemble at air-surface interfaces, *in vitro*, producing an elastic film that is highly hydrophobic at this interface.

The crystal structure of BslA (181 amino acids) residues 48 – 172 (pdb ID code 4bhu) was reported, with the crystals comprised of a 10 molecule asymmetric unit assembling in space group $P2_12_12_1$ (Hobley et al, 2013). The structure consists of 3 helices and 13 β-strands forming two β-sheets (Figure 6.1). The β-strands A, B, E, D, and C, F, G form the two sheets which give rise to a convex and concave surface, respectively, on opposite faces of the structure – resembling an immunoglobulin-like fold. An additional three-stranded β-sheet, on one end of the structure, displays numerous hydrophobic amino acids that are exposed outward (Figure 6.1C). This hydrophobic cap is protected from the solvent by virtue of the decameric crystal packing. The total surface area of BslA is ~6,670 $Å^2$ with the cap forming 24% of this. These surface exposed hydrophobic residues are essential to the water-excluding properties of BslA (Hobley et al, 2013).

We obtained purified truncated BslA, containing residues 42 – 181 (15.3 kDa), with CEC to AEA mutation in amino acids 163 to 165 (termed as BslA-AxA hereafter). These mutations were shown to produce self-assembled 2D crystals when observed under TEM. The sample was prepared by Dr Marieke Schor (University of Edinburgh, UK), following which the assembly and EM grid preparation and microscopy was done in Sheffield. The aims were to observe these crystals using negative stain and generate an untilted projection structure along with a 3D EM model. An *ab initio* attempt at fitting the crystal structure of residues 48 – 172 into the EM model was made, revealing the possible packing in these 2D crystals.

6.2. Sample preparation

The purified 15.3 kDa BslA-AxA protein was resuspended from freeze-dried state to dilutions of 0.025 and 0.05 mg/ml in double-distilled water and phosphate buffer (pH 7.5). Sample handling was important in maintaining solubility of protein; agitation or repeated use of a pipette tip causes protein to assemble into a thick film. 5 µl of each dilution was carefully added to carbon-coated copper EM grids, glow-discharged for 30 s, prior to use. The samples were incubated on the grid for 2 and 5 min.

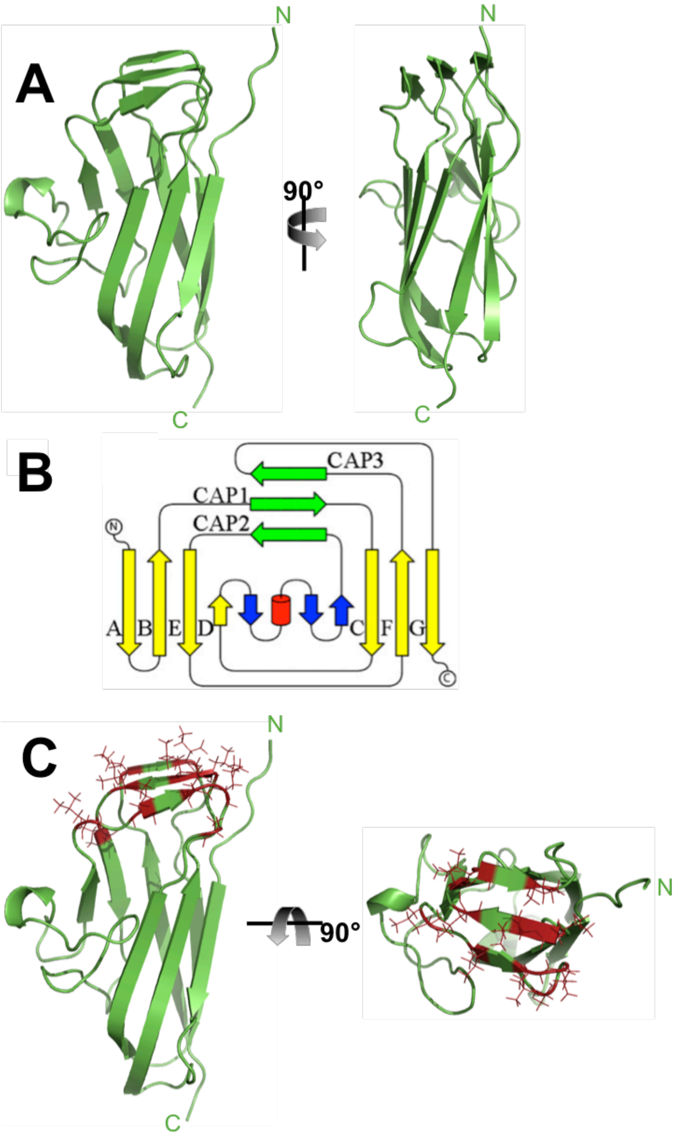


Figure 6.1 The crystal structure of BslA, showing a topological representation, and hydrophobic residues in the cap region

A: Cartoon representation of the crystal structure of BslA (pdb ID: 4bhu) showing the β-strands forming the Ig-like β-sheets on either face of the structure. The cap consists of the three β-strands at the top. **B**: A topological representation of the structure generated using TopDraw (Bond, 2003). The yellow and blue β-strands are common among immunoglobulins; the exception being the green β-strands making the hydrophobic cap region and the α-strand in red. **C**: The crystal structure in **A**, with the surface-exposed, hydrophobic residues in the cap region shown in red; the side chains of these are displayed as red lines.

The 0.025 and 0.05 mg/ml protein samples in double-distilled water were added on hydrophobic EM grids (coated with more than 48-h-old carbon film). 5 µl of the each dilution of sample was incubated for 2 min and 5 min on EM grids, at room temperature, before blotting and staining with 0.75% uranyl formate. The 0.05 mg/ml protein dissolved in double-distilled water was also incubated for 60 min, on a hydrophobic carbon film. This sample was placed on a glass slide in a covered humid chamber, to prevent evaporation, followed by blotting and negative staining, as above. Microscopy was done with CM100 TEM (Philips), operating at 100 keV. Images were recorded on a Gatan 1 k × 1 k Multiscan 794 CCD camera. The images were collected and processed using method described in Section 2.6.

6.3. Results

On glow-discharged hydrophilic grids, BslA-AxA of both concentrations appeared to form well-distributed, 'spherical' assemblies of ~50 nm diameter, in phosphate buffer (Figure 6.2A). In distilled water, the protein produced sheet-like structures of varying sizes (Figure 6.2B). These sheets did not appear in phosphate buffer, nor did the spherical structures appear in distilled water. These sheets, on hydrophilic grids, did not show any diffraction, indicating a lack of crystalline nature.

The hydrophobic grid yielded large 2D crystals of BslA-AxA, spanning several microns. Figure 6.3A shows the 2D crystals from the sample of 0.05 mg/ml BslA-AxA in distilled water, on a hydrophobic carbon film. Similar crystals, but fewer in number, were seen for the 0.025 mg/ml concentration. These crystals at high magnification showed sharp peaks in the FFT, indicating a square lattice (Figure 6.3B).

The 0.05 mg/ml sample in water was also incubated on the hydrophobic carbon film coated grids for 60 min. This longer incubation did not yield any better results compared to the 5 min incubation, at least at the resolution limit imposed due to negative stain.

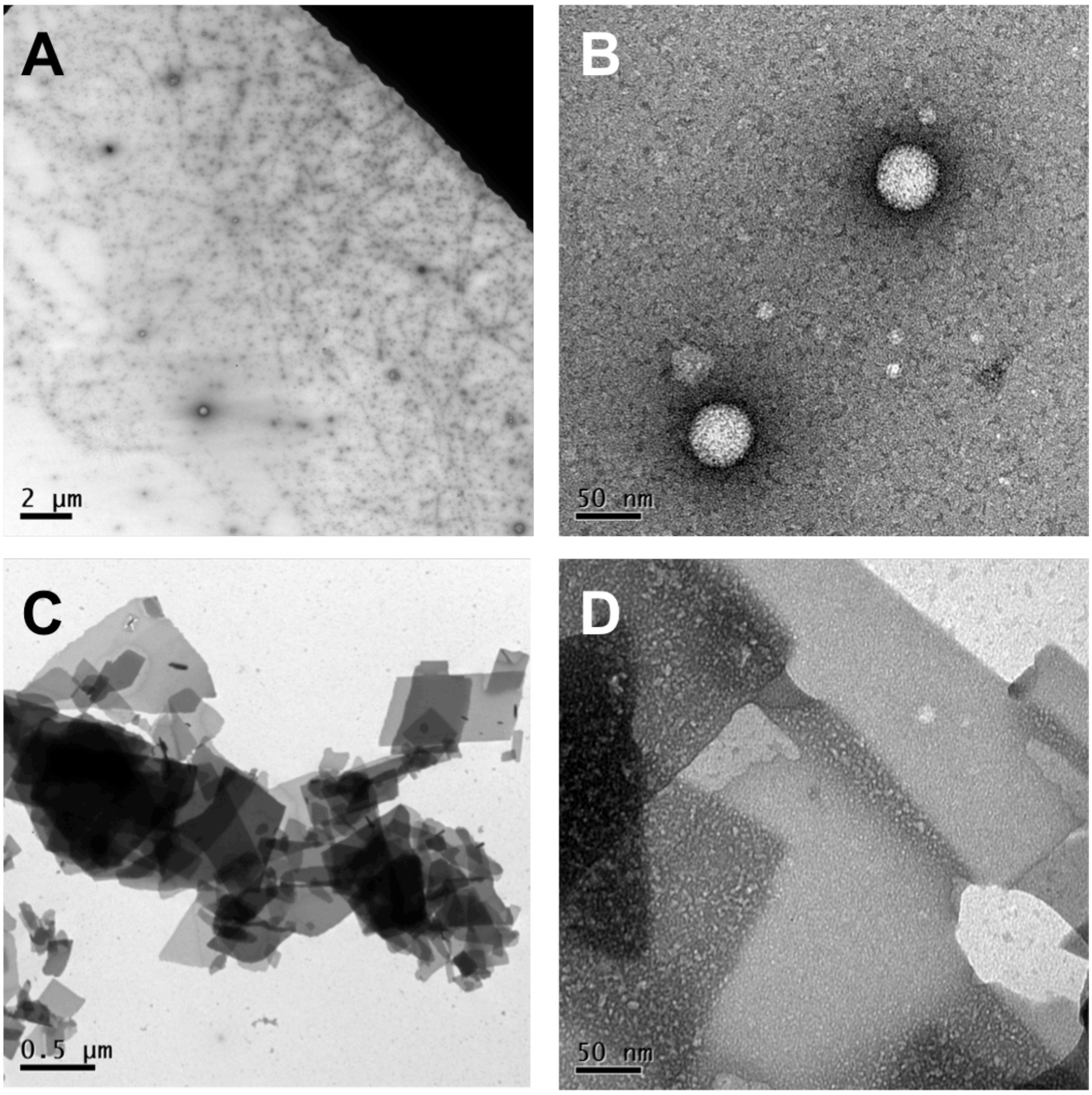


Figure 6.2 EM images of BslA-AxA protein in phosphate buffer and distilled water, on hydrophilic grid

A to **D**: BslA-AxA in 0.025 mg/ml concentration on hydrophilic carbon-coated grid. **A**: A low magnification image of BslA dissolved in phosphate buffer showing the numerous 'spherical' structures. **B**: A magnified view of the 'spherical' structures. **C**: Fragments of varying sizes seen with BslA dissolved in water. Magnified image of the fragments that are amorphous is shown in **D**.

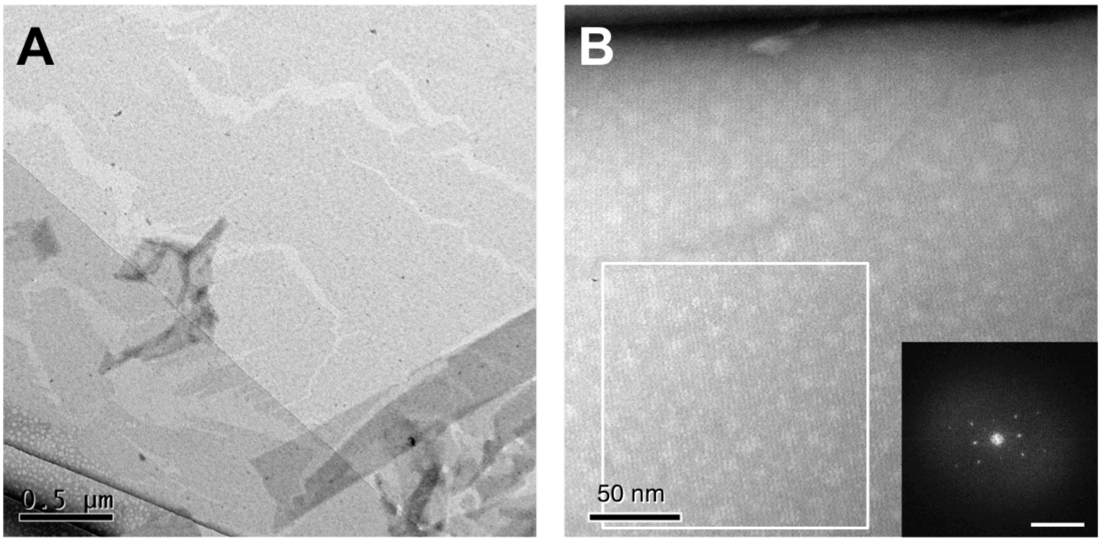


Figure 6.3 2D crystals of BslA-AxA on hydrophobic EM grid

A: EM image, at low magnification, showing large sheets of self-assembled BslA-AxA proteins in phosphate buffer on a hydrophobic carbon film. The sheets tend to aggregate in large numbers towards the edge of damaged or folded carbon film. **B**: A magnified image of such a 2D crystalline sheet. The FFT of the white-boxed region shows sharp peaks. Scale bar of FFT in **B** shows 0.125 nm^{-1}.

6.3.1. Projection structure of BslA

FFT from the TEM images indicated an orthorhombic space group. Images of 11 untilted crystals showed mean unit cell lengths, a = 38 Å ± 1 Å, b = 44 Å ± 2 Å, and cell angle, γ = 90° ± 3°. Figure 6.4 shows an unsymmetrised projection map (p1), and the resolution circle plot showing the merged reflections in resolution rings of 26, 19 and 15 Å. Images were taken with low exposure to preserve phases around 15 Å. Phase residuals for different plane groups generated by ALLSPACE for a typical untilted BslA-AxA crystal are shown in Table 6.1. Figure 6.5 shows the *p*2-symmetry imposed merged map from merged data set of 11 different untilted crystals. The maps display 2 × 2 unit cells with symmetry axis perpendicular to the surface of the page and truncated to 14 Å resolution.

Table 6.1 Phase residual table for different plane groups for a representative untilted crystal of BslA-AxA

Plane Group[1]	**Phase residual (random = 90°)**	**Target residual**[2]	**Plane Group**[1]	**Phase residual (random = 90°)**	**Target residual**[2]
p1	16.6°		p22121	55.4°	19.3°
p2	17.0°*	23.8°	c222	58°	19.3°
p12b	63.6°	18.4°	p4	58.2°	19.2°
p12a	27.8°	17.6°	p422	54.3°	17.8°
p121b	18.8°*	18.4°	p4212	57.8°	17.8°
p121a	63.1°	17.6°	p3	35.4°	16.6°
c12b	63.6°	18.4°	p312	53.4°	17.3°
c12a	27.8°	17.6°	p321	36.7°	17.5°
p222	58°	19.3°	p6	47.7°	18.5°
p2221b	70.1°	19.3°	p622	46.8°	17.5°
p2221a	21.0°!	19.3°			

[1] a and b represent the respective symmetry axis for the plane group

[2] Target residual indicates the expected phase residual of each symmetry group based on the SNR of the spots from the concerned image

Symbols for possible plane groups: * = acceptable, ! = should be considered, ` = possibility

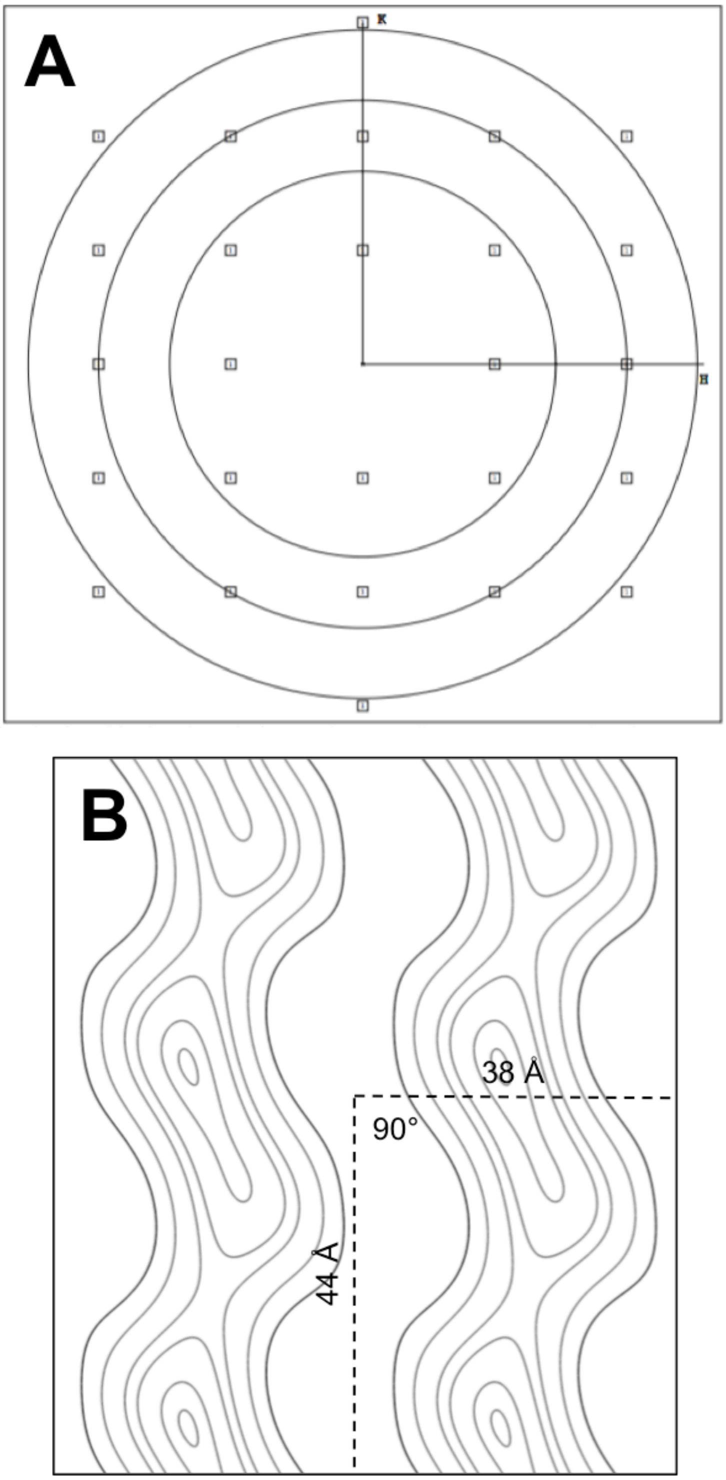


Figure 6.4 Resolution plot and projection map of a BslA-AxA 2D crystal in p1

A: Resolution circle plot with the reflections indicated in squares with numbers representing the IQ values. The circles are placed at 26 Å, 19 Å, and 15 Å spatial frequencies, with the H and K axis indicated. **B**: Contour plot showing the 2D projection map from one crystal in p1 symmetry. A single unit cell is represented by the dashed line with dimensions of a = 38 Å, b = 44 Å, and cell angle, γ = 90°.

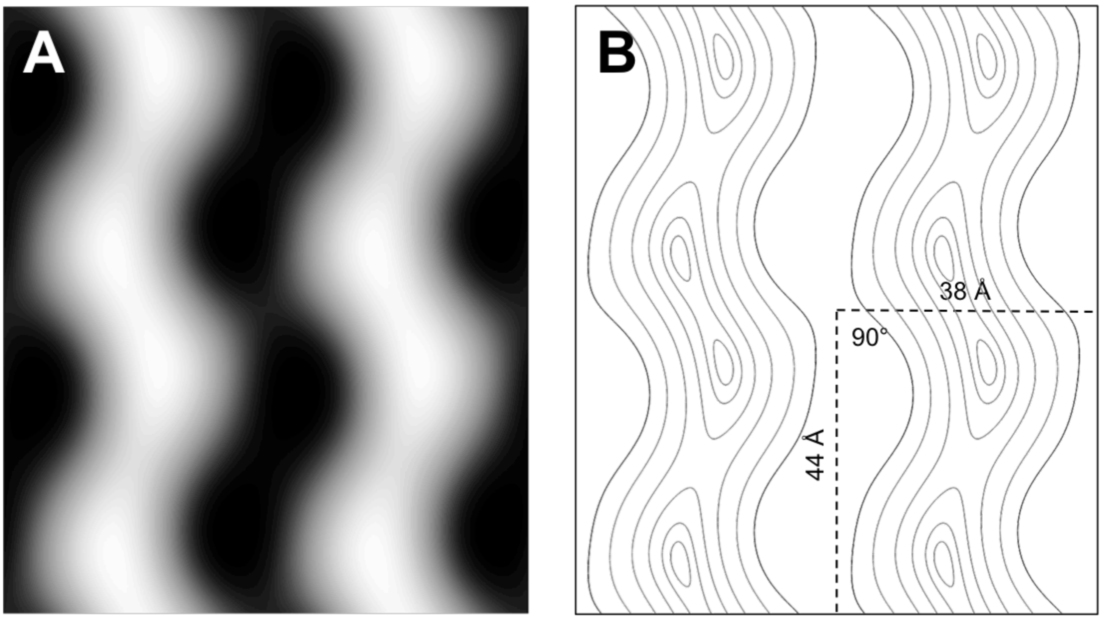


Figure 6.5 Merged projection map of BslA-AxA 2D crystals showing density and contour representations in *p*2 symmetry

A: Density representation of 2 × 2 unit cells of BslA-AxA, merged from 11 untilted crystals, with *p*2 symmetry imposed. The EM density is in grey scale, with proteins appearing as white over a black background. Higher density appears as lighter shades of grey. **B**: The same map in A shown as a contour plot. Each contour line connects X,Y coordinates where EM density value is the same. Dashed line represents one unit cell of length a = 38 Å, b = 44 Å, and γ = 90°. The symmetry axis is perpendicular to the paper.

6.3.2. 3D volume

Forty micrographs of crystals, with tilt angles ranging from +45 to –40°, and 414 unique reflections, were used to calculate the final 3D map. Appendix 10 shows the distribution of amplitudes and phases of spots along z^*. The overall phase residual (as described in Section 2.6.5), indicating the overall phase error in the 3D data was 22.0°. A summary of statistics from the 3D dataset is shown in Table 6.2.

A surface representation of the calculated 3D volume is shown in Figure 6.6. The threshold for the volume map was set at a level where the surface of each unit cell encompasses a volume corresponding to two BslA molecules (2 × 15.3 kDa).

Table 6.2 Summary statistics of 3D data of BslA-AxA crystals

3D reconstruction parameters	
Crystal plane group symmetry	$p2$
Crystal unit cell parameters	a = 38.0 Å, b = 44.0 Å, γ = 90.0°, c = 120.0 Å
Number of micrographs	40
Range of defocus	250 to 508 nm
IQ range used for 3D refining	1 to 5
Tilt range used	0 to 42.4°
In-plane resolution cut-off	14 Å
Vertical resolution cut-off	20 Å
Number of observed reflections	520
Number of observed unique reflections	414
Overall phase residual	22.0°

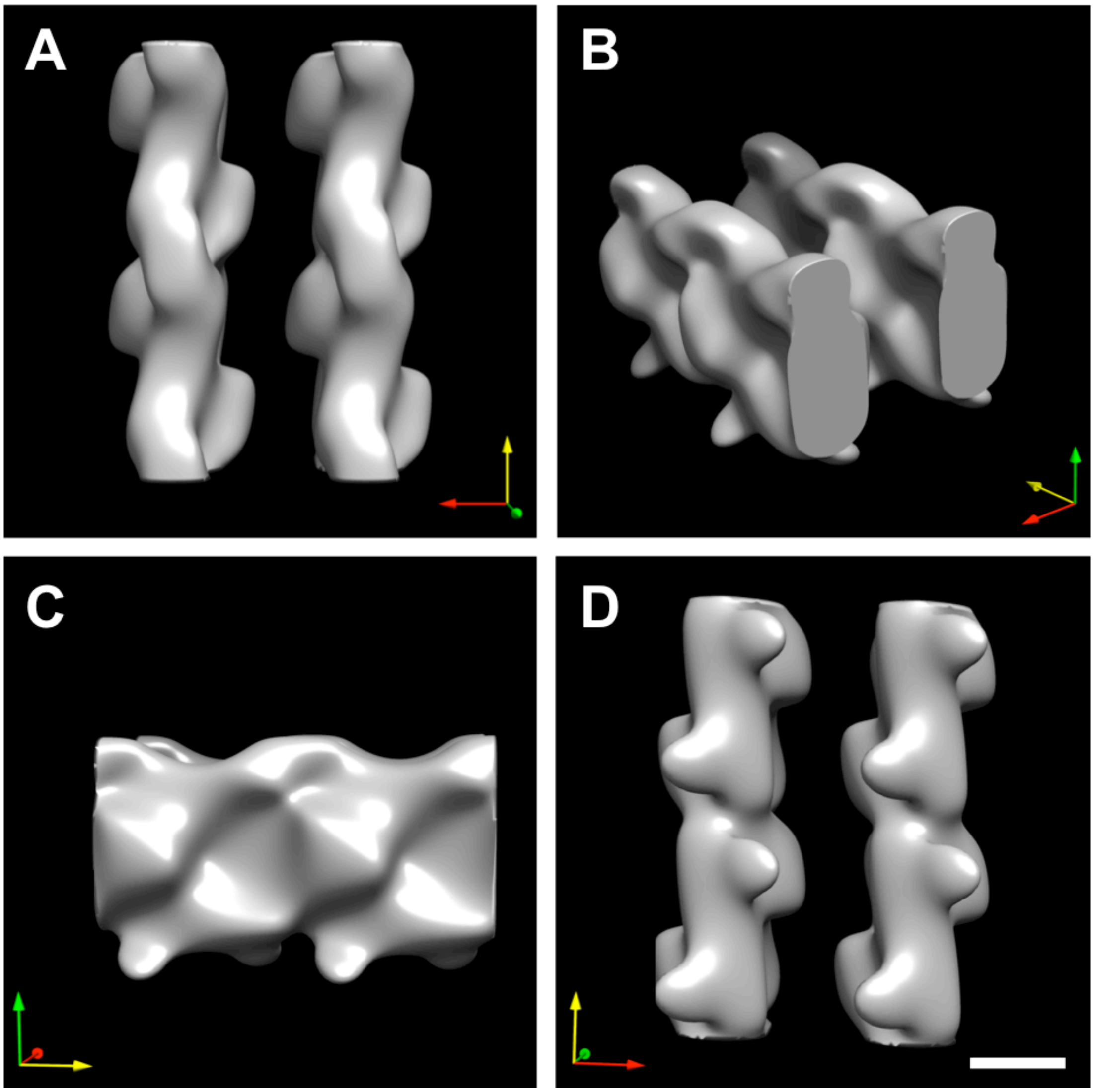


Figure 6.6 3D volume map of BslA-AxA crystals

A to **D**: Different views of surface representation of the 3D volume map of BslA-AxA crystals, self-assembled on hydrophobic carbon film. 2 × 2 unit cells are shown here. The X, Y and Z-axes are represented as red, yellow and green arrows, respectively. *p*2 symmetry axis is along Z-axis. Scale bar in **D** is 2 nm and applies to all panels.

6.4. Discussion

6.4.1. Variation in form based on environment

As seen in Figure 6.2, BslA-AxA forms structures that resemble flakes, on hydrophilic carbon film, when the sample is in distilled water. The size of the flakes ranged from 100 nm to 2 µm. The grid was heavily populated by such structures. Interestingly, these only appeared in the case of BslA-AxA solubilised in water when deposited on a freshly glow-discharged hydrophilic carbon film. Such flakes, resembling crystals, did not show any diffraction spots in the FFT, indicating an amorphous nature.

When the protein was dissolved in phosphate buffer, however, such flakes were replaced with what might be spherical structures of around 50 to 100 nm diameter (Figure 6.2). The carbon film was glow discharged before loading the sample. These structures resembled micelles formed by long-chain polar molecules. Although the protein does have a polar nature, orientation and organisation of the molecule in the spherical structure is not clear. The formation of spherical structure in the presence of phosphate buffer (pH 7.5) may be due to charge alteration in the protein-buffer interfaces.

The BslA-AxA, on the other hand, formed very clear self-assembled sheets of 2D crystals on a hydrophobic carbon film. Such structures appeared most numerous near the edge of a damaged carbon film or the copper grid bar. This location can indicate that these structures might be the residual material left behind after the excess sample is blotted away during the staining procedure. The 2D crystals assembled at several nucleation points, as evidenced by the presence of domains with lattices of the same unit cell dimensions in different orientations. The presence of crystals on hydrophobic grid is indicative of an interaction between the hydrophobic cap and the grid, which might aid the protein self-assembly process.

6.4.2. The structure & future work

The untilted projection indicated unit cell lengths of a = 38 Å, b = 44 Å, and cell angle, γ = 90°. Owing to such small unit cell dimensions and resolution limit imposed due to negative staining, reflections higher than the second order were not expected. However, some crystals did show (2,2) reflection, up to 15 Å. The calculation of phases was, as a result, terminated at 14 Å, in-plane.

The crystal structure of residues 48 – 172 was fitted into the 3D volume map, manually (Figure 6.7). Potential clashes between residues in adjacent subunits were corrected and checked using UCSF Chimera (Pettersen et al, 2004). The hydrophobic caps of the four subunits in the volume map face the same direction. The BslA-AxA volume map comprises of residues 42 – 181, 15 more than the crystal structure. The overlaid volume map displays regions not ccupied by the crystal structure. This preliminary fit needs to be further refined using putative inter-subunit interactions based on analysis of mutated variants of BslA (on-going work, Prof. Cait MacPhee, University of Edinburgh, personal communications).

The large numbers of readily assembled crystals on the grid are good candidates for electron crystallographic analysis using cryo-EM. These would provide better structural information, which is currently limited due to the small size of the repeating unit in the crystals, reaching the limit of negative stain.

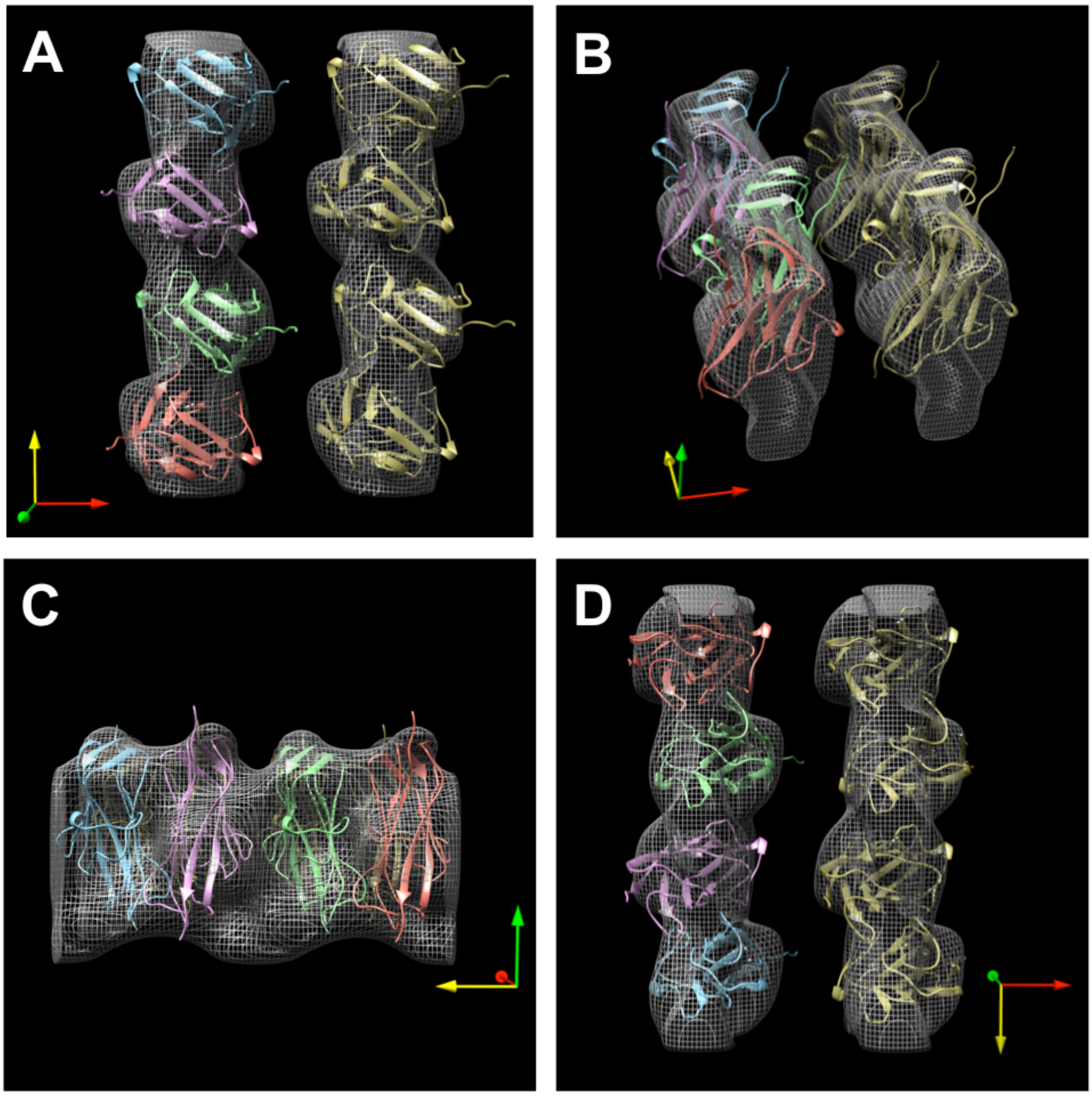


Figure 6.7 Eight crystal structures of BslA$_{48-172}$ fitted into BslA-AxA volume map

A to **D**: Crystal structure of BslA$_{48-172}$ (pdb ID: 4bhu) fitted in the BslA-AxA volume map, after manual refinements. The best fit was estimated based on overall encapsulation in the volume and absence of steric clashes, analysed using ‘contact and clashes’ tool in UCSF Chimera (Pettersen et al, 2004). Note that 6 and 9 amino acids are absent from the N and C-terminals in the crystal structure. The N-termini in the ribbon structures are facing towards the top of the page in **B** and **C**. The volume map represents 2 × 2 unit cells of a = 38 Å, b = 44 Å, γ = 90°, and p2-symmetrised. The red, yellow and green arrows indicate the X, Y and Z real-space axes, respectively.

Chapter 7 - General discussions

Surface layers are nearly ubiquitous structures found in two domains of life. In addition to being the outermost layer, directly in contact with the environment, this layer is also one of the most abundant proteins expressed by the cell. The unique paracrystalline form of this layer has stimulated many investigations in the past (Sara & Sleytr, 2000; Messner & Sleytr, 1992; Sleytr et al, 2014). However, the understanding of the structure of this layer remains in its infancy.

The two-dimensional nature of the S-layer proteins poses a difficulty in crystallisation for X-ray structure determination. Attempts at circumventing this problem includes the use of nanobodies (Baranova et al, 2012) and structural modifications (Fagan et al, 2009; Usenik et al, 2017). Such *in vitro* crystal structures, however, may not resemble the *in situ* nature of the S-layer complex. In addition, this layer resides on the surface of cells, which are dense bodies, making it difficult to image using EM. So far structural analysis has involved either freeze-fracture and metal-shadowing techniques (Wildhaber & Baumeister, 1987; Cerquetti et al, 2000), or cryo-electron tomography on thin sections and structures present along the edges of the cells (Chen et al, 2011; Bharat et al, 2017).

This thesis has described the methods that have enabled us to image native S-layers on the cell envelope, in a near *in situ* state. These paracrystals can be isolated and analysed directly from the parent structure, providing exclusive biological insight not obtained by *in vitro* 2D or 3D crystallisation methods. Electron crystallography, as a technique has provided us with a powerful tool to resolve the high-resolution structure of self-assembling molecules.

7.1. Structure of the S-layer components in *C. difficile*

The *C. difficile* S-layer is a complex comprising the HMW and LMW S-layer proteins. The complex forms a 2D crystal on the surface of the bacterial cell with an oblique lattice and *p*2 plane group symmetry. The unit cell dimensions are $a = b =$

80 Å with $\gamma = 100°$ (Chapter 3). From the 3D volume map generated, we know that one unit cell in this structure contains two H/L complexes (Chapter 3).

7.1.1. Localisation of LMW SLP

The domain 2 of the LMW SLP shows remarkable sequence diversity among *C. difficile* strains (Ni Eidhin et al, 2008; Calabi & Fairweather, 2002), indicative of a role in pathogen and host immune evasion. The LMW SLP was also found to be immunodominant in patients with antibiotic-associated diarrhoea (Pantosti et al, 1989). Additionally, LMW SLP binds anti-sera from rabbits raised against *C. difficile* (Cerquetti et al, 2000). These suggest the possibility of it being surface exposed. We have localised the LMW SLP by deleting the hyper-variable domain 2 (Chapter 4). This development has enabled us to identify the orientation of the S-layer volume and allowed us to fit an *in silico* model of the R20291 LMW SLP, based on the known crystal structure of the truncated LMW SLP$_{1\text{–}262}$ from strain 630 (Fagan et al, 2009), into the 3D volume (Figure 7.1). The structure is fitted into the volume based on the location of the deleted region (Chapter 4). The model was generated using the PHYRE2 Protein Fold Recognition Server (Kelley et al, 2015). The SlpA$^{\Delta Ld2}$ S-layer volume is displayed overlaid on the wild type S-layer volume with the PHYRE model of the LMW SLP fitted in (Figure 7.1). The crystal structures fit very well within the wild type volume. Overlaying the mutant SlpA$^{\Delta Ld2}$ S-layer volume exposes the deleted LMW domain 2 region, unoccupied by the mutant volume.

The LMW domain 2 shows very little similarity in sequence between related strains. The domain 2 folds in the PHYRE model of the R20291 LMW SLP therefore are not accurately modelled due to lack of homology with other proteins in the structure database. The domain 1 however, shows sequence similarities among different *C. difficile* strains and the predicted model is based on the 630 LMW SLP$_{1\text{–}262}$ structure.

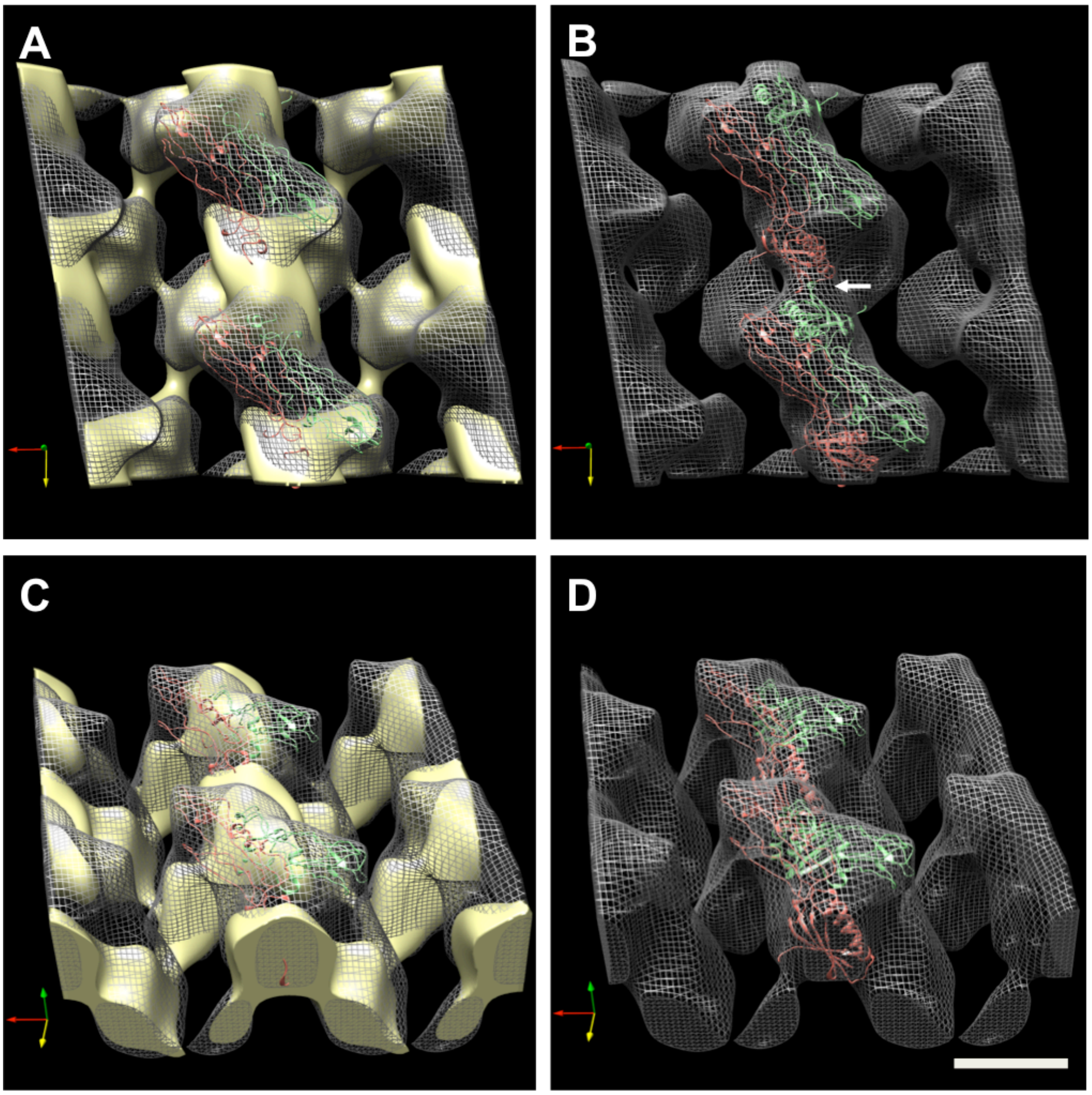


Figure 7.1 Overlaid wild type R20291 S-layer volume with SlpA$^{\Delta Ld2}$ mutant, fitted with *in silico* model of LMW SLP

The wild type R20291 S-layer volume map (grey wireframe) is overlaid on the SlpA$^{\Delta Ld2}$ mutant S-layer map (yellow surface) along with four *in silico* models of the R20291 LMW SLP fitted in. The green and salmon coloured cartoon models indicate a pair of LMW domain 2 interacting laterally. **A** and **C**: The LMW SLP domain 2 model occupying the empty void, not occupied by the mutant, representing the deleted domain. **B** and **D:** The model with the mutant indicating the position of the LMW SLP. 2 × 2 unit cells of the S-layer are shown in *p*2 symmetry. Note the adjacent LMW SLP domain 1 interaction region (white arrow). The red, yellow and green arrows indicate the X, Y and Z axes, respectively. Scale bar: 4 nm.

From our data we believe this to be a true representation of the position of the surface-exposed LMW SLP in the S-layer (Figure 7.1). This location also makes it ideal for the variable domain to evade the host immune system and provide resistance against phages.

7.1.2. LMW – LMW interaction domains

Interestingly, as can be seen from Figure 7.1, the fitted model of the LMW complex, two adjacent LMW SLPs interact with each other laterally in the domain 2 region. This lateral 'domain 2 – domain 2' interaction seems to be present in the crystal packing of the 630 truncated LMW SLP (Fagan et al, 2009). Each asymmetric unit in the $P6_4$–packed crystal contained four LMW SLPs (Figure 7.2). Two of the subunits show similar lateral interactions to the observed fitted model. The interacting residues (VTkeNSlNek, with residues facing the opposite domain 2 in capitals) are highlighted in the figure. This interaction however does not appear to be responsible for the structural integrity of the S-layer, as the deletion of the LMW domain 2 does not prevent S-layer assembly (Chapter 4).

Intriguingly, there appears to be an interface between two adjacent LMW SLP domain 1 regions (highlighted with white arrow in Figure 7.1 and Figure 7.2). This region, unlike domain 2, shows similarity in sequence across *C. difficile* strains and might be a very good candidate for future mutational studies to understand any putative roles in crystallisation and self-assembly.

7.1.3. Cell wall binding region

The HMW SLP is known to bind to the secondary cell surface polysaccharide, PS-II (Ganeshapillai et al, 2008; Willing et al, 2015). The position of the surface exposed LMW SLP in the 3D volume allows us to localise the HMW SLP. The crystal structure of the HMW SLP is not known for SlpA. However, the structure of Cwp6 and Cwp8 have been reported (Usenik et al, 2017), which contain the three CWB2 domains. The CWB2 region forms three Rossmann folds, producing a disc-like region, which binds to PS-II (Figure 1.3).

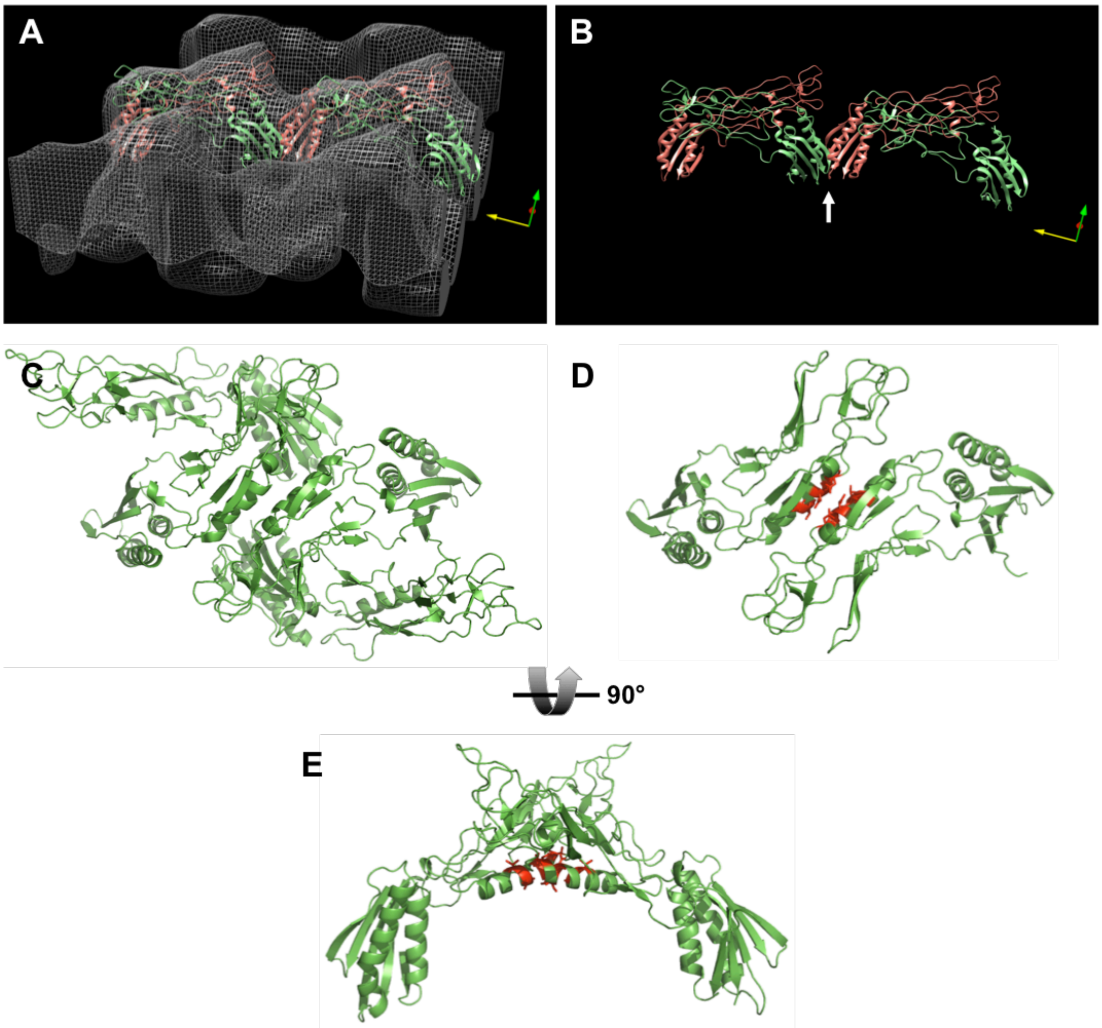


Figure 7.2 Lateral interactions between adjacent LMW SLPs also evident in the crystal structure of the strain 630 LMW SLP

The *in silico* models of the R20291 LMW SLPs adjacent to each other interact with each other laterally along domain 2, as seen in **A** with the R20291 volume map (grey wireframe) overlaid and **B** where the volume map is removed. The green and salmon coloured cartoon models indicate a pair of LMW domain 2 interacting laterally. The domain 1 interaction indicated with a white arrow. The volume map represents 2 × 2 unit cells of $a = b = 80$ Å length, and $p2$-symmetrised. The red, yellow and green arrows indicate the X, Y and Z axes, respectively. The interaction results in the formation of a bridge-like structure. Similar structure is also evident in the crystal packing of the LMW SLP$_{1\text{–}262}$. **C**: One asymmetric unit consisting of four subunits of the LMW SLP$_{1\text{–}262}$, that packs in P6$_4$ symmetry. **D** shows the packing with two subunits removed and **E** shows the two remaining LMW SLP$_{1\text{–}262}$ rotated 90° along Y-axis highlighting the bridge-like structure, comparable to the arrangement seen in **B**. The laterally interacting residues are highlighted in red in **D** and **E**.

Extensive homology between the sequence of the CWB2 domains among cell wall proteins in *C. difficile* means the structure of this region in the HMW SLP can be modelled into the EM volume map (Figure 7.3). Four CWB2 crystal structures from Cwp8 are fitted manually (using UCSF Chimera (Pettersen et al, 2004)) into the four complete HMW SLP volumes in the EM map. The domains fit very well in our map, revealing the spatial organisation of the HMW SLP in the H/L complex on the cell surface. It should be noted that the EM model is likely to represent the CWB2 regions in a PS-II-bound state on the cell wall, which the crystal structure does not.

The CWB2-fitted volume map can be complemented with the PHYRE model of the R20291 LMW SLP, depicting the potential structural organisation of the entire H/L complex. Shown here is a single H/L complex fitted into the EM volume (Figure 7.4). It should be noted that the N-terminal residues (1 – 40) of the HMW SLP are not homologous to any region in Cwp8 and is absent in the CWB2 domain structure fitted here. This is also the region that interacts with the LMW SLP to form the H/L complex (Fagan et al, 2009).

7.2. S-layer and cell shape

As discussed in Chapter 4, the lack of the S-layer or indeed any structural modifications to the S-layer results in a major change in growth and morphology of *C. difficile* cells (Figure 4.5). This author was unable to find a link between cell shape and growth, and the S-layer in the current literature. The change in growth rate is unlikely to be due to the result of unobserved mutation(s) in the *slpA*⁻ genome of FM2.5 and 2.6, since complementing the genome with a functional *slpA* gene restores the wild type phenotype (Kirk et al, 2017b). However, the S-layer might provide anchoring points for receptors required for cellular growth and multiplication. The slow growth might be a result of the loss of those anchoring points. In addition, the SlpA$^{\Delta Ld2}$ S-layer might provide reduced anchoring points due to its altered structure, showing an intermediate cellular size, when compared to wild type and *slpA*⁻ phenotypes.

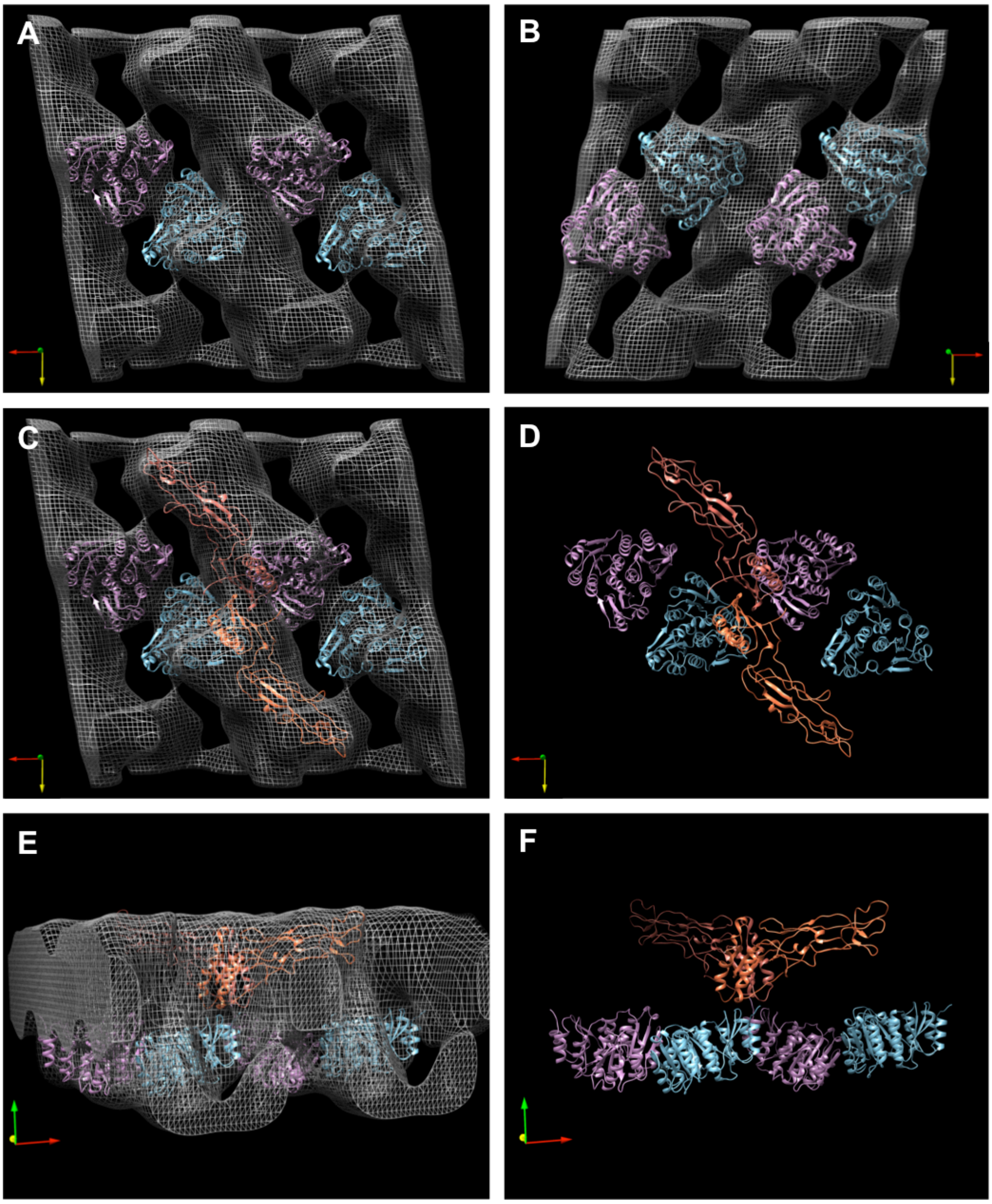


Figure 7.3 R20291 volume map fitted with CWB2 domain from Cwp8 cell wall protein and LMW SLP model

A and **B**: Two views of the volume map of R20291 S-layer (grey wireframe) fitted with four CWB2 domains from Cwp8 protein in strain 630 (in blue and pink colours). This region is conserved among the different strains and the cell wall protein family. **C** and **E**: Two different views of the four CWB2 domains along with two units of the *in silico* model of R20291 LMW SLP (in orange colour) displaying how the SLPs are likely to be located in the crystal. Note that the LMW SLP interaction region of the HMW SLP (residues 1 – 40) is not present in Cwp8 and hence the interface between the H/L complex cannot be visualised. **D** and **F:** The same views in **C** and **E**, respectively, with the S-layer volume removed. The red, yellow and green arrows indicate the X, Y and Z real-space axes, respectively, in all panels. The volume map represents 2 × 2 unit cells of $a = b = 80$ Å length, and $p2$-symmetrised.

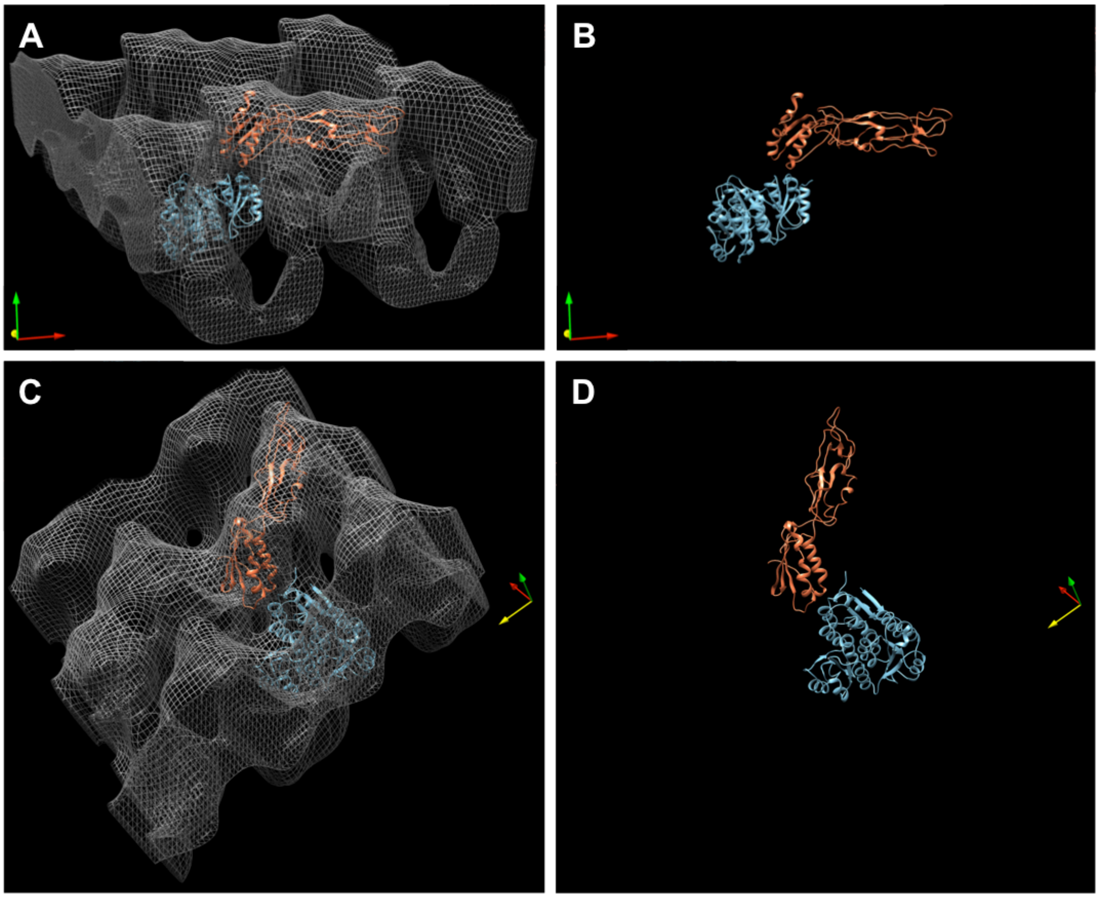


Figure 7.4 Model of one CWB2 domain and LMW SLP model fitted in the R20291 S-layer volume map

A and **C**: Two views of the volume map of R20291 S-layer (grey wireframe) fitted with one CWB2 domain (in blue colour) from Cwp8 protein in strain 630 and *in silico* model of R20291 LMW SLP (in orange colour) displaying how the SLPs are likely to be located in the crystal. **B** and **D**: The same views as **A** and **C**, respectively, without the volume map. Note that the LMW SLP interaction region of the HMW SLP (residues 1 – 40) is not present in Cwp8 and hence the interface between the H/L complex cannot be visualised. The red, yellow and green arrows indicate the X, Y and Z real-space axes, respectively, in all panels. The volume map represents 2 × 2 unit cells of $a = b = 80$ Å length, and $p2$-symmetrised.

In addition, the cytosolic proteins, MreB and its homologues, have been shown to maintain cell shape in rod-shaped bacteria, with mutational studies indicating formation of abnormal curved morphologies (Figge et al, 2004; Shiomi et al, 2008; Errington, 2015). The curved shape of S-layer mutant cells may be due to the S-layer being associated directly or indirectly to the cytoskeletal protein. The SlpA may also be linked to other unknown cytoskeletal complexes.

Furthermore, the SlpA mutants also produce cells showing asymmetric septa during cell division (Figure 4.5). The septum localisation system involves the Min group of proteins in rod-shaped cells, which have been linked to cellular curvature (Levin et al, 1992). An altered curvature of the bacterium, seen for example in the SlpA$^{\Delta Ld2}$ mutant, could alter the location of the septum as seen in this case (Figure 4.5). The relationship between SlpA and cell shape is an exciting dimension to the S-layer story that merits further work.

7.3. Organisation of the *C. difficile* cell envelope *in situ*

Imaging near the poles of the cell, where the least cell material is in the way of the electron beam has been used as strategy to image cell envelope architecture, such as the flagella motor complexes *in situ* in ice (Chen et al, 2011). The *C. difficile* cell surface was similarly imaged using cryo-EM (Figure 7.5). The figure shows clear layers close to the cell surface, several of which can be identified and measured in the image intensity profile attached, including the S-layer and the underlying cell wall. The cell membrane can be seen as a dense layer. There is a noticeable dense, sometimes discontinuous, intermediate layer present between the cell wall and cell membrane, distinguishable from the less dense cell wall. The identity of this layer remains undetermined (Figure 7.5). It will be fascinating to understand the role of this layer and whether it shows any resemblance to the Gram-negative inner membrane and periplasmic space. Using cryo-EM, vitrified sections of some Gram-positive cells have been shown to contain such a layer in *Bacillus subtilis* and *Staphylococcus aureus* (Matias & Beveridge, 2005; 2006).

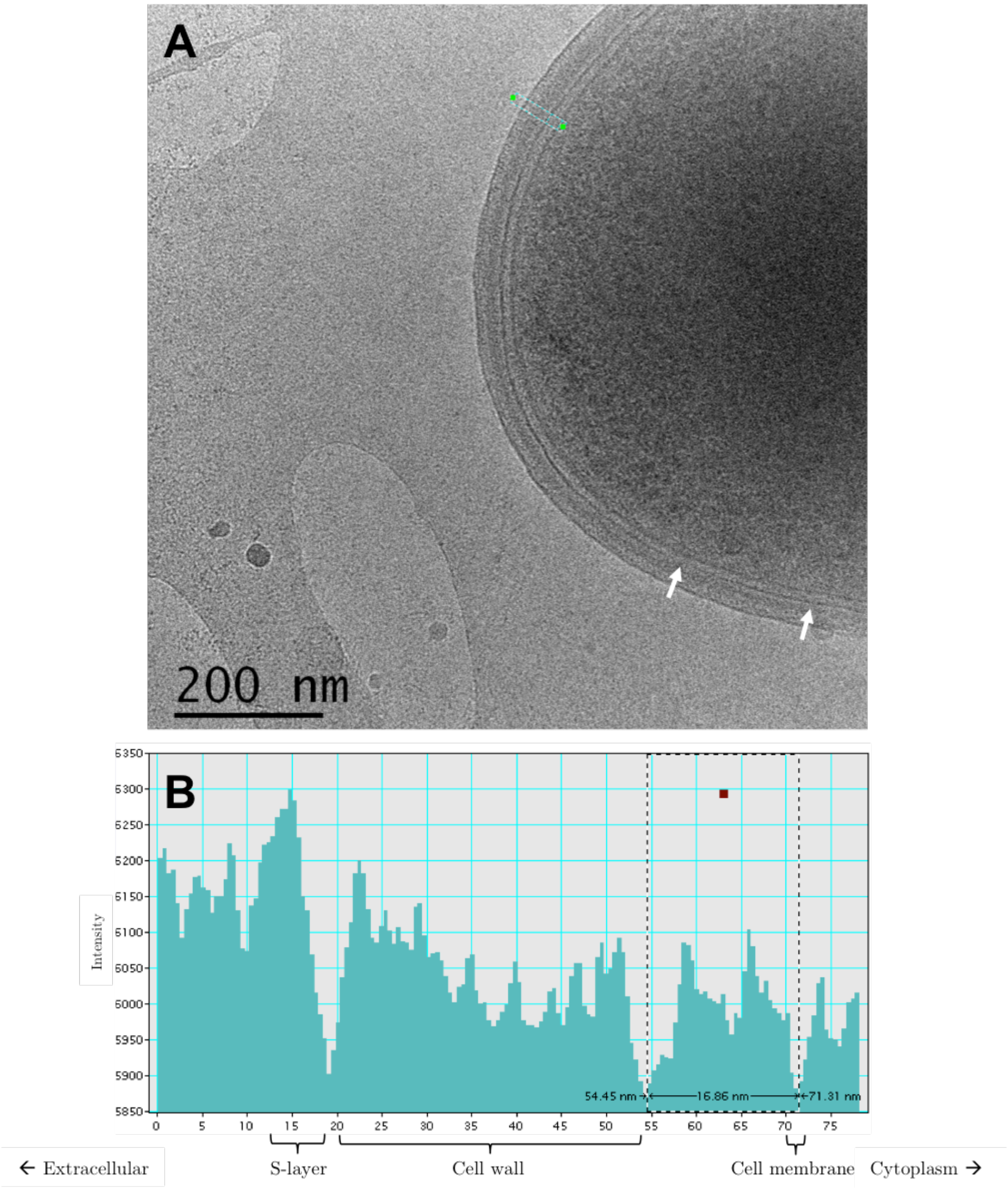


Figure 7.5 Several layers of the cell envelope are seen in the polar end of a whole *C. difficile* cell imaged embedded ice

A: The polar end of an untreated *C. difficile* cell imaged using cryo-EM. The layers of the cell envelope are visible clearly. Image was taken using a 4k × 4k CCD camera (Gatan) on a 200 keV CM200 (Philips) microscope. A green boxed region in the image represents the area shown as an intensity profile in **B**. This profile indicates the pixel intensity, across the cell envelope (extracellular space to cytoplasm from left to right), as a function of distance in nm. The darker regions show low intensity. Clear distinction can be made of the cell membrane, cell wall and S-layer, which are labelled below the intensity profile. The ~17 nm region above the cytoplasmic membrane is unidentified. The additional layer separating the region is discontinuous, as seen in **A**, with the gaps indicated by white arrows.

The fact that these layers can be distinctly visualised in ice raises the possibility of future experiments, using cryo-electron tomography, to understand the organisation of the cell envelope. The contrast of such images can be greatly improved with the use of Volta phase plates, direct electron detectors and energy filters.

Interestingly, the S-layer assembles on the cell surface in distinct domains (Figure 7.6). This distinct mosaic arrangement of the S-layer has also been observed in previous freeze-fracture EM studies of archaea and bacteria (Sara & Sleytr, 2000), and atomic force microscopy of immobilized bacteria (Günther et al, 2014). These domains may result from distinct *nucleation points* during the process of self-assembly. This form of 'mosaic' assembly might allow the S-layer to follow the cellular curvature – encapsulating the cylindrical cell body as well as the hemispherical poles. The domains may also allow for the presence of appendages and other cell surface structures.

The nucleation points for S-layer self-assembly may represent the sites where the machinery required for the secretion of SlpA are co-localised. The accessory Sec secretion system is responsible for the transport and secretion of SlpA (Fagan & Fairweather, 2011). In addition, cell wall protein Cwp84 is known to cleave the SlpA into the HMW and LMW SLPs (Kirby et al, 2009). Preliminary light microscopy work has indicated that the SecA2 protein, part of the accessory Sec secretion system, may be localized in distinct regions across the cell surface (Dr Robert Fagan, personal communication). It will be interesting to understand further the mechanism of maturation and S-layer organization. Further work using light microscopy and EM may reveal the relationship between the secretion machinery and maturation of the S-layer.

7.4. Resolution of cryo-projections and future work

The highest resolution obtained from cryo-EM projection maps of the *C. difficile* S-layer fragments was 7.5 Å (Chapter 5). The images were taken on a 200 keV Tecnai F20 (FEI) microscope with a Falcon II (FEI) direct electron detector.

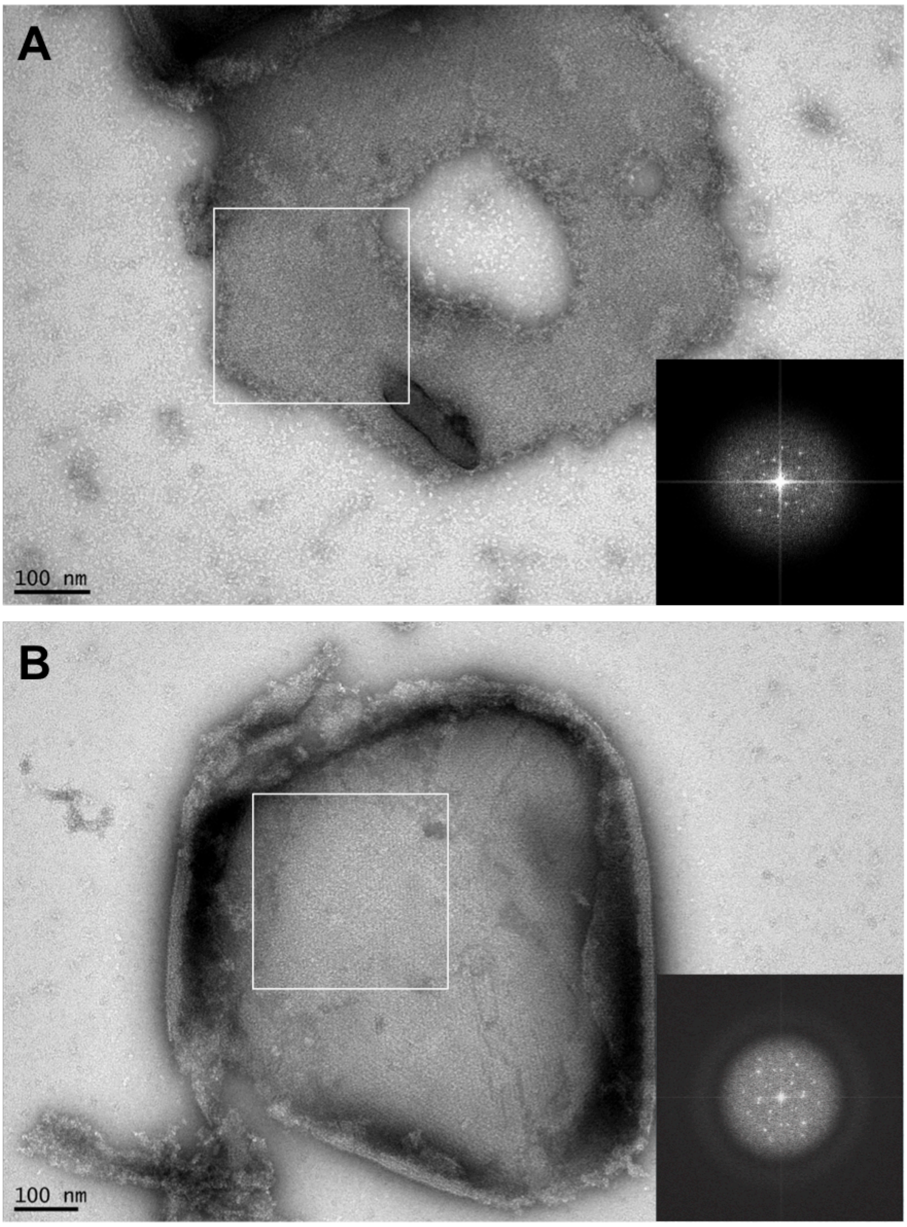


Figure 7.6 The S-layer organises itself into a mosaic assembly on the cell surface

The R20291 S-layer fragments, isolated using the fragmentation technique (Chapter 5), show clear 'mosaic' pattern of self-assembly. **A** and **B** show two of such fragments displaying S-layer, with the inset FFT of the region highlighted in the white box. The images were taken on a 4k × 2k Orius CCD camera (Gatan), using a T12 Tecnai Spirit microscope (FEI) operating at 80 keV. The fragments were negatively stained after isolation.

Further improvements in resolution might be achieved using a more stable stage and environmental control. Collecting data in 'movie mode', where several frames of images are collected over a single exposure, followed by motion correction to correct for unstable sample movement can also improve the resolution (Scheres, 2014). However, this resolution limit may be a direct consequence of damage to the S-layer fragments during the experimental procedures or simply intrinsic disorder in the S-layer.

As discussed in Chapter 5, it becomes challenging to index diffraction spots obtained from tubular double-layered crystals, which are more abundant in the sample. The problem is exacerbated by the fact that even a single layer of crystal appears to have multiple domains, with the resulting FFT containing overlapping lattices in different alignments (Figure 7.6). A double layer compounds this problem. When the sample is tilted to collect projections at higher tilt angles, the lattice distortion as a result may make it more challenging to identify the correct lattice in the resulting FFT. Single layers therefore are preferable for data collection at high tilts. However, the single layers are more difficult to identify on a low magnification, high noise image, in 'search mode' (Figure 5.3). A suitable method needs to be developed which allows the identification of these structures at low magnification, making data collection simpler.

Single particle cryo-EM technique has transformed the way the structures of proteins are resolved. Until recently the low contrast in cryo-EM images of particles smaller than 100 kDa had been a limiting step in identification and processing of images of frozen, hydrated samples of small proteins. The ~90 kDa H/L complex of *C. difficile* is one such structure. However, current improvements in the use of electron phase plate, as a method of increasing contrast of lower spatial frequency features in the structures of proteins, have enabled scientists to localise and computationally process images of proteins previously undetectable in a cryo-EM image (Danev & Baumeister, 2016). Single particle derived 3.2 Å structure of haemoglobin (64 kDa) has been reported, determined using the Volta phase plate

(Khoshouei et al, 2017), indicating a breakthrough in structural biology using cryo-EM. These developments make it possible to use single particle analysis as a technique to determine the near-atomic resolution structure of the H/L complex, in addition to an *in situ* S-layer structure using electron crystallography.

7.5. Perspectives

C. difficile is a worldwide healthcare concern. Broad-spectrum antibiotic use, followed by gut dysbiosis, is the predominant factor resulting in *C. difficile* infection. The rising use of antibiotics in the community poses a real danger of the pathogen transmitting outside hospital settings, where safety and containment measures are absent. Developing countries, where the use of broad-spectrum antibiotics is rife, poses an even greater challenge. The surface layer is the outermost layer in *C. difficile* and interacts directly with the host environment. The essential nature of the S-layer in *C. difficile* provides us with a unique opportunity to tackle this pathogen by targeting this layer. S-layer targeting anti-microbials are already being investigated as a therapeutic measure (Gebhart et al, 2012; 2015; Kirk et al, 2017b). Further understanding of this layer and its structure will no doubt provide us with innovative solutions to prevent and cure infections.

This thesis presents the first glimpse into the structure of the S-layer of a very important pathogen. The unique structural features of the paracrystalline S-layer have been revealed in detail using mutagenesis and electron crystallography. The structural features include – the characteristics and location of the domain involved in anti-microbial resistance and immune evasion, the cell wall binding domain, and a domain putatively involved in self-assembly. We have been able to complement the EM volume with predicted and known crystal structures to reveal the overall architecture of the S-layer. Comparing the 3D structure of S-layers from two strains of *C. difficile* we can visualise the highly variable nature of the LMW SLP. This layer is uniquely exposed to the cell surface and allows the use of targeted anti-microbials. We have also shown that this technique can be used to image and determine the structure of glycosylated S-layer of *C. difficile* Ox247 (Appendix 7).

Increasing antibiotic resistance among important pathogens calls for further research into alternative strategies for disease prevention and cure. The ubiquitous nature of the S-layer makes it an important therapeutic target among many microorganisms. The techniques specifically developed for this project, combined with the powerful technique of electron crystallography, will help improve our understanding of self-assembling 2D complexes and S-layers on other organisms and open up further avenues for tackling diseases.

References

Amos LA, Henderson R & Unwin PN (1982) Three-dimensional structure determination by electron microscopy of two-dimensional crystals. Progress in Biophysics and Molecular Biology 39: 183–231

Baker LA, Smith EA, Bueler SA & Rubinstein JL (2010) The resolution dependence of optimal exposures in liquid nitrogen temperature electron cryomicroscopy of catalase crystals. Journal of Structural Biology 169: 431–437

Baranova E, Fronzes R, Fronzes R, Garcia-Pino A, Garcia-Pino A, Van Gerven N, Van Gerven N, Papapostolou D, Papapostolou D, Péhau-Arnaudet G, Péhau-Arnaudet G, Pardon E, Pardon E, Steyaert J, Steyaert J, Howorka S, Howorka S, Remaut H & Remaut H (2012) SbsB structure and lattice reconstruction unveil Ca2+ triggered S-layer assembly. Nature 487: 119–122

Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JLS & Subramaniam S (2015) 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. Science 348: 1147–1151

Bharat TAM, Kureisaite-Ciziene D, Hardy GG, Yu EW, Devant JM, Hagen WJH, Brun YV, Briggs JAG & Löwe J (2017) Structure of the hexagonal surface layer on *Caulobacter crescentus* cells. Nature Microbiology 2: 17059

Bingle WH, Nomellini JF & Smit J (1997) Linker mutagenesis of the *Caulobacter crescentus* S-layer protein: toward a definition of an N-terminal anchoring region and a C-terminal secretion signal and the potential for heterologous protein secretion. Journal of Bacteriology 179: 601–611

Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. Bioinformatics 19: 311–312

Boot HJ, Kolen CP & Pouwels PH (1995) Identification, cloning, and nucleotide sequence of a silent S-layer protein gene of *Lactobacillus acidophilus* ATCC 4356 which has extensive similarity with the S-layer protein gene of this species. Journal of Bacteriology. 177: 7222–7230

Branda SS, Chu F, Kearns DB, Losick R & Kolter R (2006) A major protein component of the *Bacillus subtilis* biofilm matrix. Molecular Microbiology 59: 1229–1238

Bruggemann H, Baumer S, Fricke WF, Wiezer A, Liesegang H, Decker I, Herzberg C, Martinez-Arias R, Merkl R, Henne A & Gottschalk G (2003) The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. Proceedings of the National Academy of Sciences U.S.A. 100: 1316–1321

Calabi E & Fairweather N (2002) Patterns of sequence conservation in the S-Layer proteins and related sequences in *Clostridium difficile*. Journal of Bacteriology. 184: 3886–3897

Calabi E, Calabi F, Phillips AD & Fairweather NF (2002) Binding of Clostridium difficile surface layer proteins to gastrointestinal tissues. Infection and Immunity 70: 5770–5778

Calabi E, Ward S, Wren B, Paxton T, Panico M, Morris H, Dell A, Dougan G & Fairweather N (2001) Molecular characterization of the surface layer proteins from *Clostridium difficile.* Molecular Microbiology 40: 1187–1199

Cerquetti M, Molinari A, Sebastianelli A, Diociaiuti M, Petruzzelli R, Capo C & Mastrantonio P (2000) Characterization of surface layer proteins from different *Clostridium difficile* clinical isolates. Microbial pathogenesis 28: 363–372

Chen DH, Jakana J, Liu X & Schmid MF (2008) Achievable resolution from images of biological specimens acquired from a 4k× 4k CCD camera in a 300-kV electron cryomicroscope. Journal of Structural Biology 63(1): 45–52

Chen S, Beeby M, Murphy GE, Leadbetter JR, Hendrixson DR, Briegel A, Li Z, Shi J, Tocheva EI, Müller A, Dobro MJ & Jensen GJ (2011) Structural diversity of bacterial flagellar motors. EMBO Journal. 30: 2972–2981

Cheong GW, Guckenberger R & Fuchs KH (1993) The structure of the surface layer of *Methanoplanus limicola* obtained by a combined electron microscopy and scanning tunneling microscopy approach. Journal of Structural Biology 111: 125–134

Chutinan A, John S & Toader O (2003) Diffractionless flow of light in all-optical microchips. Physical Reviews Letters 90: 123901

Costerton JW, Cheng KJ, Geesey GG, Ladd TI, Nickel JC, Dasgupta M & Marrie TJ (1987) Bacterial biofilms in nature and disease. Annual Review of Microbiology 41: 435–464

Costerton JW, Lewandowski Z, Caldwell DE, Korber DR & Lappin-Scott HM (1995) Microbial biofilms. Annual Review of Microbiology 49: 711–745

Costerton JW, Stewart PS & Greenberg EP (1999) Bacterial biofilms: a common cause of persistent infections. Science 284: 1318–1322

Crewe AV, Wall J & Langmore J (1970) Visibility of single atoms. Science 168: 1338–1340

Crowther RA, Henderson R & Smith JM (1996) MRC image processing programs. Journal of Structural Biology 116: 9–16

Danev R & Baumeister W (2016) Cryo-EM single particle analysis with the Volta phase plate. eLife 5: 439

Dang THT, Dang THT, Riva L de L, la Riva de L, Fagan RP, Storck EM, Storck EM, Heal WP, Heal WP, Janoir C, Janoir C, Fairweather NF, Tate EW & Tate EW (2010) Chemical probes of surface layer biogenesis in *Clostridium difficile.* ACS Chemical Biology 5: 279–285

Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, Fairweather NF & Fagan RP (2015) High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile.* mBio 6: e02383

Dembek M, Dembek M, Reynolds CB, Reynolds CB, Fairweather NF & Fairweather NF (2012) Clostridium difficile cell wall protein CwpV undergoes enzyme-independent intramolecular autoproteolysis. Journal of Biological Chemistry 287: 1538–1544

Dethlefsen L, Huse S, Sogin ML & Relman DA (2008) The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS biology 6: e280

Dingle KE, Didelot X, Ansari MA, Eyre DW, Vaughan A, Griffiths D, Ip CLC, Batty EM, Golubchik T, Bowden R, Jolley KA, Hood DW, Fawley WN, Walker AS, Peto TE, Wilcox MH & Crook DW (2013) Recombinational switching of the *Clostridium difficile* S-layer and a novel glycosylation gene cluster revealed by large-scale whole-genome sequencing. Journal of Infectious Diseases 207: 675–686

Dubochet J, Lepault J, Freeman R, Berriman JA & Homo JC (1982) Electron microscopy of frozen water and aqueous solutions. Journal of Microscopy 128: 219–237

Engel BD, Schaffer M, Kuhn Cuellar L, Villa E, Plitzko JM & Baumeister W (2015) Native architecture of the Chlamydomonas chloroplast revealed by in situ cryo-electron tomography. eLife 4: 3583–29

Epstein AK, Pokroy B, Seminara A & Aizenberg J (2011) Bacterial biofilm shows persistent resistance to liquid wetting and gas penetration. Proceedings of the National Academy of Science U.S.A. 108: 995–1000

Errington J (2015) Bacterial morphogenesis and the enigmatic MreB helix. Nature Reviews Microbiology 13: 241–248

Fagan RP & Fairweather NF (2011) *Clostridium difficile* has two parallel and essential Sec secretion systems. Journal of Biological Chemistry 286: 27483–27493

Fagan RP & Fairweather NF (2014) Biogenesis and functions of bacterial S-layers. Nature Reviews Microbiology 12: 211–222

Fagan RP, Albesa-Jové D, Albesa-Jové D, Qazi O, Qazi O, Svergun DI, Svergun DI, Brown KA, Brown KA & Fairweather NF (2009) Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. Molecular Microbiology. 71: 1308–1322

Fagan RP, Janoir C, Collignon A, Mastrantonio P, Poxton IR & Fairweather NF (2011) A proposed nomenclature for cell wall proteins of *Clostridium difficile*. Journal of Medical Microbiology 60: 1225–1228

Faruqi AR & Henderson R (2007) Electronic detectors for electron microscopy. Current opinion in structural biology 17(5): 549–555

Figge RM, Divakaruni AV & Gober JW (2004) MreB, the cell shape-determining bacterial actin homologue, co-ordinates cell wall morphogenesis in *Caulobacter crescentus*. Molecular Microbiology 51: 1321–1332

Flemming H-C & Wingender J (2010) The biofilm matrix. Nature Reviews Microbiology 8: 623–633

Ganeshapillai J, Vinogradov E, Rousseau J, Weese JS & Monteiro MA (2008) *Clostridium difficile* cell-surface polysaccharides composed of pentaglycosyl and hexaglycosyl phosphate repeating units. Carbohydrate Research 343: 703–710

Gebhart D, Lok S, Clare S, Tomas M, Stares M, Scholl D, Donskey CJ, Lawley TD & Govoni GR (2015) A Modified R-Type Bacteriocin Specifically Targeting *Clostridium difficile* Prevents Colonization of Mice without Affecting Gut Microbiota Diversity. mBio 6: e02368–14–14–13

Gebhart D, Williams SR, Bishop-Lilly KA, Govoni GR, Willner KM, Butani A, Sozhamannan S, Martin D, Fortier L-C & Scholl D (2012) Novel high-molecular-weight, R-type bacteriocins of *Clostridium difficile*. Journal of Bacteriology 194: 6240–6247

Gipson B, Zeng X & Stahlberg H (2007a) 2dx_merge: Data management and merging for 2D crystal images. Journal of Structural Biology 160: 375–384

Gipson B, Zeng X, Zhang ZY & Stahlberg H (2007b) 2dx--user-friendly image processing for 2D crystals. Journal of Structural Biology 157: 64–72

Glaeser RM (2007) Electron crystallography of biological macromolecules Oxford; New York: Oxford University Press

Glaeser RM (2016) Specimen Behavior in the Electron Beam. Methods in Enzymology 579: 19–50

Goorhuis A, Bakker D, Corver J, Debast SB, Harmanus C, Notermans DW, Bergwerff AA, Dekker FW & Kuijper EJ (2008) Emergence of Clostridium difficile infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 47: 1162–1170

Günther TJ, Suhr M, Raff J & Pollmann K (2014) Immobilization of microorganisms for AFM studies in liquids. RSC Advances 4: 51156–51164

Havelka WA, Havelka WA, Henderson R, Oesterhelt D & Oesterhelt D (1995) Three-dimensional structure of halorhodopsin at 7 A resolution. Journal of Molecular Biology 247: 726–738

Henderson R, Baldwin JM, Ceska TA, Zemlin F, Beckmann E & Downing KH (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. Journal of Molecular Biology 213: 899–929

Henderson R, Baldwin JM, Downing KH & Lepault J (1986) Structure of purple membrane from *Halobacterium halobium*: recording, measurement and evaluation of electron micrographs at 3.5 Å resolution. Ultramicroscopy 19: 147–178

Ho TD, Williams KB, Chen Y, Helm RF, Popham DL & Ellermeier CD (2014) *Clostridium difficile* extracytoplasmic function sigma factor sigmaV regulates lysozyme resistance and is necessary for pathogenesis in the hamster model of infection. Infection and Immunity 82: 2345–2355

Hobley L, Ostrowski A, Rao FV, Bromley KM, Porter M, Prescott AR, MacPhee CE, van Aalten DMF & Stanley-Wall NR (2013) BslA is a self-assembling bacterial hydrophobin that coats the *Bacillus subtilis* biofilm. Proceedings of the National Academy of Science U.S.A. 110: 13600–13605

Houwink AL (1953) A macromolecular mono-layer in the cell wall of Spirillum spec. Biochimica et Biophysica Acta 10: 360–366

Houwink AL & Le Poole JB (1952) Eine Struktur in der Zellmembran einer Bakterie. Physikalische Verhandlungen 3, 98

Iancu CV, Tivol WF, Schooler JB, Dias DP, Henderson GP, Murphy GE, Wright ER, Li Z, Yu Z, Briegel A, Gan L, He Y & Jensen GJ (2006) Electron cryotomography sample preparation using the Vitrobot. Nature Protocol 1: 2813–2819

Jank T & Aktories K (2008) Structure and mode of action of clostridial glucosylating toxins: the ABCD model. Trends in Microbiology 16: 222–229

Jiang W & Chiu W (2001) Web-based Simulation for Contrast Transfer Function and Envelope Functions. Microscopy and Microanalysis 7: 329–334

Kawata T, Takeoka A, Takumi K & Masuda K (1984) Demonstration and preliminary characterization of a regular array in the cell wall of *Clostridium difficile.* FEMS Microbiology 24: 323–328

Kelley LA, Mezulis S, Yates CM, Wass MN & Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 10: 845–858

Kern J, Wilton R, Zhang R, Binkowski TA, Joachimiak A & Schneewind O (2011) Structure of Surface Layer Homology (SLH) Domains from *Bacillus anthracis* Surface Array Protein. Journal of Biological Chemistry 286: 26042–26049

Kern JW & Schneewind O (2008) BslA, a pXO1-encoded adhesin of B*acillus anthracis.* Molecular Microbiology 68: 504–515

Khoshouei M, Radjainia M, Baumeister W & Danev R (2017) Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. Nature communications 8:16099

Kirby JM, Kirby JM, Ahern H, Ahern H, Roberts AK, Roberts AK, Kumar V, Kumar V, Freeman Z, Freeman Z, Acharya KR, Acharya KR, Shone CC & Shone CC (2009) Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of *Clostridium difficile.* Journal of Biological Chemistry 284: 34666–34673

Kirk JA, Banerji O & Fagan RP (2017a) Characteristics of the *Clostridium difficile* cell envelope and its importance in therapeutics. Microbial Biotechnology 10: 76–90

Kirk JA, Gebhart D, Buckley AM, Lok S, Scholl D, Douce GR, Govoni GR & Fagan RP (2017b) New Class of Precision Antimicrobials Redefines Role of *Clostridium difficile* S-layer in Virulence and Viability. Science Translational Medicine 9(406): eaah6813

Kobayashi K & Iwano M (2012) BslA(YuaB) forms a hydrophobic layer on the surface of *Bacillus subtilis* biofilms. Molecular Microbiology 85: 51–66

Kuroda A & Sekiguchi J (1990) Cloning, sequencing and genetic mapping of a *Bacillus subtilis* cell wall hydrolase gene. Journal of General Microbiology 136: 2209–2216

Kuroda A, Rashid MH & Sekiguchi J (1992) Molecular cloning and sequencing of the upstream region of the major *Bacillus subtilis* autolysin gene: a modifier protein exhibiting sequence homology to the major autolysin and thespoIIDproduct. Journal of General Microbiology 138: 1067–1076

Kühlbrandt W, Wang DN & Fujiyoshi Y (1994) Atomic model of plant light-harvesting complex by electron crystallography. Nature 367: 614–621

Laemmli UK (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature 227: 680–685

Lawley TD, Clare S, Walker AW, Goulding D, Stabler RA, Croucher N, Mastroeni P, Scott P, Raisen C, Mottram L, Fairweather NF, Wren BW, Parkhill J & Dougan G (2009) Antibiotic treatment of *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. Infection and Immunity 77: 3661–3669

Levin PA, Margolis PS, Setlow P, Losick R & Sun D (1992) Identification of *Bacillus subtilis* genes for septum placement and shape determination. Journal of Bacteriology 174: 6717–6728

Matias VRF & Beveridge TJ (2005) Cryo-electron microscopy reveals native polymeric cell wall structure in *Bacillus subtilis* 168 and the existence of a periplasmic space. Molecular Microbiology 56: 240–251

Matias VRF & Beveridge TJ (2006) Native cell wall organization shown by cryo-electron microscopy confirms the existence of a periplasmic space in *Staphylococcus aureus*. Journal of Bacteriology. 188: 1011–1021

Mayer MJ, Garefalaki V, Spoerl R, Narbad A & Meijers R (2011) Structure-Based Modification of a *Clostridium difficile*-Targeting Endolysin Affects Activity and Host Range. Journal of Bacteriology 193: 5477–5486

McDonald LC, Killgore GE, Thompson A, Owens RCJ, Kazakova SV, Sambol SP, Johnson S & Gerding DN (2005) An epidemic, toxin gene-variant strain of *Clostridium difficile*. The New England Journal of Medicine 353: 2433–2441

McGowan AP, Lalayiannis LC, Sarma JB, Marshall B, Martin KE & Welfare MR (2011) Thirty-day mortality of *Clostridium difficile* infection in a UK National Health Service Foundation Trust between 2002 and 2008. The Journal of Hospital Infection 77: 11–15

McMullan G, Faruqi AR & Henderson R (2016) Direct Electron Detectors. Methods in Enzymology 579: 1–17

McMullan G, Faruqi AR, Clare D & Henderson R (2014) Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. Ultramicroscopy 147: 156–63

Merrigan MM, Merrigan MM, Venugopal A, Venugopal A, Roxas JL, Roxas JL, Anwar F, Anwar F, Mallozzi MJ, Mallozzi MJ, Roxas BAP, Roxas BAP, Gerding DN, Gerding DN, Viswanathan VK, Viswanathan VK, Vedantam G & Vedantam G (2013) Surface-layer protein A (SlpA) is a major contributor to host-cell adherence of *Clostridium difficile*. PLoS ONE 8: e78404

Messner P & Sleytr UB (1992) Crystalline bacterial cell-surface layers. Advances in Microbial Physiology 33: 213–275

Morikawa M (2006) Beneficial biofilm formation by industrial bacteria *Bacillus subtilis* and related species. Journal of Bioscience and Bioengineering 101: 1–8

Nakayama K, Takashima K, Ishihara H, Shinomiya T, Kageyama M, Kanaya S, Ohnishi M, Murata T, Mori H & Hayashi T (2000) The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. Molecular Microbiology 38: 213–231

Ng YK, Ehsaan M, Philip S, Collery MM, Janoir C, Collignon A, Cartman ST & Minton NP (2013) Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using pyrE alleles. PLoS ONE 8: e56051

Ni Eidhin DB, O'Brien JB, McCabe MS, Athie-Morales V & Kelleher DP (2008) Active immunization of hamsters against *Clostridium difficile* infection using surface-layer protein. FEMS Immunology and Medical Microbiology 52: 207–218

Nogales E, Wolf SG & Downing KH (1997) Visualizing the secondary structure of tubulin: three-dimensional map at 4 Å. Journal of Structural Biology 118: 119–127

Orlova EV & Saibil HR (2011) Structural analysis of macromolecular assemblies by electron microscopy. Chemical Reviews 111: 7710–7748

Ostrowski A, Mehert A, Prescott A, Kiley TB & Stanley-Wall NR (2011) YuaB Functions Synergistically with the Exopolysaccharide and TasA Amyloid Fibers To Allow Biofilm Formation by *Bacillus subtilis*. Journal of Bacteriology. 193: 4821–4831

Pantosti A, Cerquetti M, Viti F, Ortisi G & Mastrantonio P (1989) Immunoblot analysis of serum immunoglobulin G response to surface proteins of *Clostridium difficile* in patients with antibiotic-associated diarrhea. Journal of Clinical Microbiology 27: 2594–2597

Pavkov T, Egelseer EM, Tesarz M, Svergun DI, Sleytr UB & Keller W (2008) The Structure and Binding Behavior of the Bacterial Cell Surface Layer Protein SbsC. Structure 16: 1226–1237

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry 25: 1605–1612

Purdy D, O'Keeffe TAT, Elmore M, Herbert M, McLeod A, Bokori-Brown M, Ostrowski A & Minton NP (2002) Conjugative transfer of clostridial shuttle vectors from Escherichia coli to *Clostridium difficile* through circumvention of the restriction barrier. Molecular Microbiology 46: 439–452

Qazi O, Hitchen P, Tissot B, Panico M, Morris HR, Dell A & Fairweather N (2009) Mass spectrometric analysis of the S-layer proteins from *Clostridium difficile* demonstrates the absence of glycosylation. Journal of Mass Spectrometry. JMS 44: 368–374

Radics J, Koenigsmaier L & Marlovits TC (2014) Structure of a pathogenic type 3 secretion system in action. Nature Structural and Molecular Biology 21(1): 82–7

Reynolds CB, Reynolds CB, Emerson JE, Emerson JE, la Riva de L, la Riva de L, Fagan RP & Fairweather NF (2011) The *Clostridium difficile* cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function. PLoS Pathogens 7: e1002024

Ristl R, Steiner K, Zarschler K, Zayni S, Messner P & Schaffer C (2011) The S-layer glycome-adding to the sugar coat of bacteria. International journal of microbiology 2011: 1–16

Rodriguez JA & Gonen T (2016) High-Resolution Macromolecular Structure Determination by MicroED, a cryo-EM. Methods in Enzymology 579: 369–392

Rosenthal PB & Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. Journal of Molecular Biology 333: 721–745

Rupnik M, Wilcox MH & Gerding DN (2009) Clostridium difficile infection: new developments in epidemiology and pathogenesis. Nature Reviews Microbiology 7: 526–536

Sara M & Sleytr UB (2000) S-Layer proteins. Journal of Bacteriology. 182: 859–868

Savariau-Lacomme MP, Lebarbier C, Karjalainen T, Collignon A & Janoir C (2003) Transcription and Analysis of Polymorphism in a Cluster of Genes Encoding Surface-Associated Proteins of *Clostridium difficile.* Journal of Bacteriology 185: 4461–4470

Scheres SHW (2014) Beam-induced motion correction for sub-megadalton cryo-EM particles. eLife 3: e03665

Schmeing TM & Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. Nature 461: 1234–1242

Scholl D, Cooley M, Williams SR, Gebhart D, Martin D, Bates A & Mandrell R (2009) An engineered R-type pyocin is a highly specific and sensitive bactericidal agent for the food-borne pathogen *Escherichia coli* O157:H7. Antimicrobial Agents and Chemotherapy 53: 3074–3080

Sebaihia M, Peck MW, Minton NP, Thomson NR, Holden MT, Mitchell WJ, Carter AT, Bentley SD, Mason DR, Crossman L, Paul CJ, Ivens A, Wells-Bennik MH, Davis IJ, Cerdeno-Tarraga AM, Churcher C, Quail MA, Chillingworth T, Feltwell T, Fraser A, et al (2007) Genome sequence of a proteolytic (Group I) *Clostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. Genome Research 17: 1082–1092

Sekulovic O, Ospina Bedoya M, Fivian-Hughes AS, Fairweather NF & Fortier L-C (2015) The *Clostridium difficile* cell wall protein CwpV confers phase-variable phage resistance. Molecular Microbiology 98: 329–342

Serra DO, Richter AM, Klauck G, Mika F & Hengge R (2013) Microanatomy at cellular resolution and spatial order of physiological differentiation in a bacterial biofilm. mBio 4: e00103–13

Shiomi D, Sakai M & Niki H (2008) Determination of bacterial rod shape by a novel cytoskeletal membrane protein. EMBO Journal 27: 3081–3091

Sleytr UB & Beveridge TJ (1999) Bacterial S-layers. Trends in Microbiology 7: 253–260

Sleytr UB & Glauert AM (1976) Ultrastructure of the cell walls of two closely related clostridia that possess different regular arrays of surface subunits. Journal of Bacteriology 126: 869–882

Sleytr UB, Schuster B, Egelseer E-M & Pum D (2014) S-layers: principles and applications. FEMS Microbiology Reviews 38: 823–864

Sleytr UB, Sleytr UB, Huber C, Huber C, Ilk N, Ilk N, Pum D, Pum D, Schuster B & Egelseer EM (2007) S-layers as a tool kit for nanobiotechnological applications. FEMS Microbiology Letters 267: 131–144

Smits WK, Lyras D, Lacy DB, Wilcox MH & Kuijper EJ (2016) *Clostridium difficile* infection. Nature Reviews. Disease primers 2: 16020

Stabler RA, He M, Dawson L, Martin M, Valiente E, Corton C, Lawley TD, Sebaihia M, Quail MA, Rose G, Gerding DN, Gibert M, Popoff MR, Parkhill J, Dougan G & Wren BW (2009) Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biology 10: R102

Takeoka A, Takumi K, Koga T & Kawata T (1991) Purification and characterization of S layer proteins from *Clostridium difficile* GAI 0714. Journal of General Microbiology 137: 261–267

Unwin P & Henderson R (1975) Molecular-Structure Determination by Electron-Microscopy of Unstained Crystalline Specimens. Journal of Molecular Biology 94: 425–&

Usenik A, Renko M, Mihelič M, Lindič N, Borišek J, Perdih A, Pretnar G, Müller U & Turk D (2017) The CWB2 Cell Wall-Anchoring Module Is Revealed by the Crystal Structures of the *Clostridium difficile* Cell Wall Proteins Cwp8 and Cwp6. Structure 25: 514–521

Valpuesta JM, Carrascosa JL & Henderson R (1994) Analysis of Electron-Microscope Images and Electron-Diffraction Patterns of Thin-Crystals of Φ29-Connectors in Ice. Journal of Molecular Biology 240: 281–287

Waligora AJ, Hennequin C, Mullany P, Bourlioux P, Collignon A & Karjalainen T (2001) Characterization of a Cell Surface Protein of *Clostridium difficile* with Adhesive Properties. Infection and Immunity 69: 2144–2153

Wang B, Kraig E & Kolodrubetz D (2000) Use of defined mutants to assess the role of the *Campylobacter rectus* S-layer in bacterium-epithelial cell interactions. Infection and Immunity 68: 1465–1473

Wildhaber I & Baumeister W (1987) The cell envelope of *Thermoproteus tenax*: three-dimensional structure of the surface layer and its role in shape maintenance. EMBO Journal 6-5: 1475–1480

Willing SE, Candela T, Shaw HA, Seager Z, Mesnage S, Fagan RP & Fairweather NF (2015)

Clostridium difficile surface proteins are anchored to the cell wall using CWB2 motifs that recognise the anionic polymer PSII. Molecular Microbiology 96: 596–608

Wright A, Wait R, Begum S, Crossett B, Nagy J, Brown K & Fairweather N (2005) Proteomic analysis of cell surface proteins from *Clostridium difficile*. Proteomics 5: 2443–2452

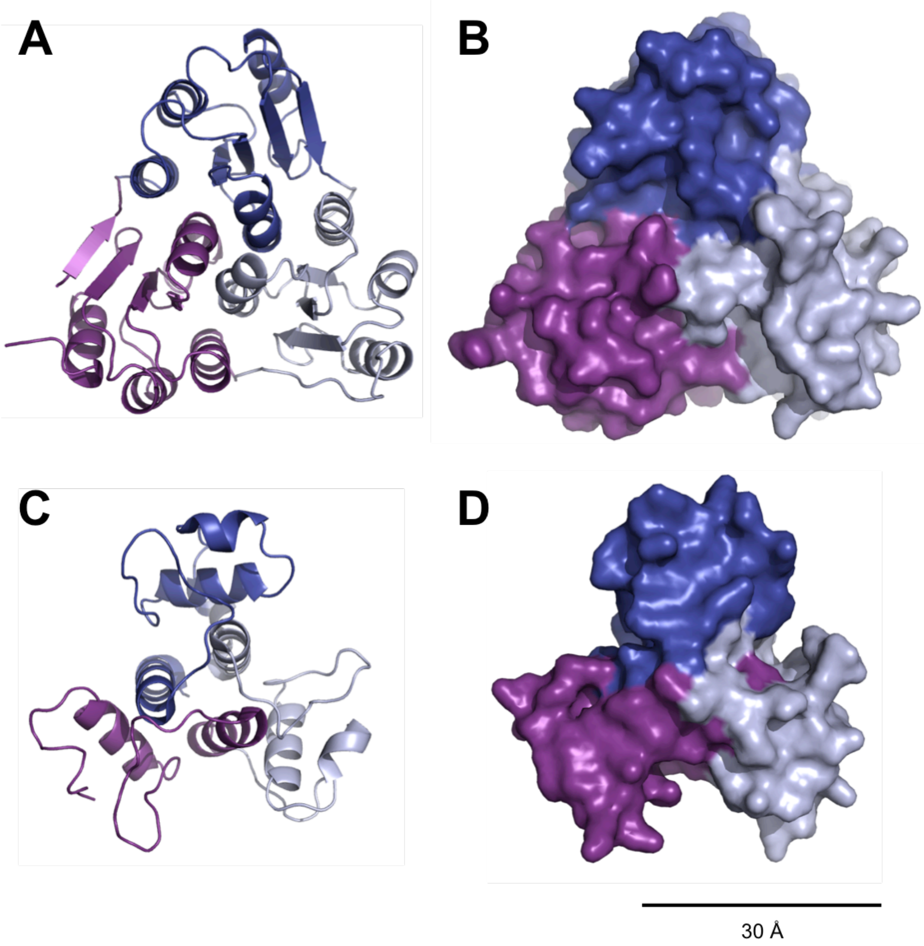
Wüst J, Sullivan NM, Hardegger U & Wilkins TD (1982) Investigation of an outbreak of antibiotic-associated colitis by various typing methods. Journal of Clinical Microbiology 16: 1096–1101

Zhu J, Vinothkumar KR & Hirst J (2016) Structure of mammalian respiratory complex I. Nature 536: 354–358

Appendix

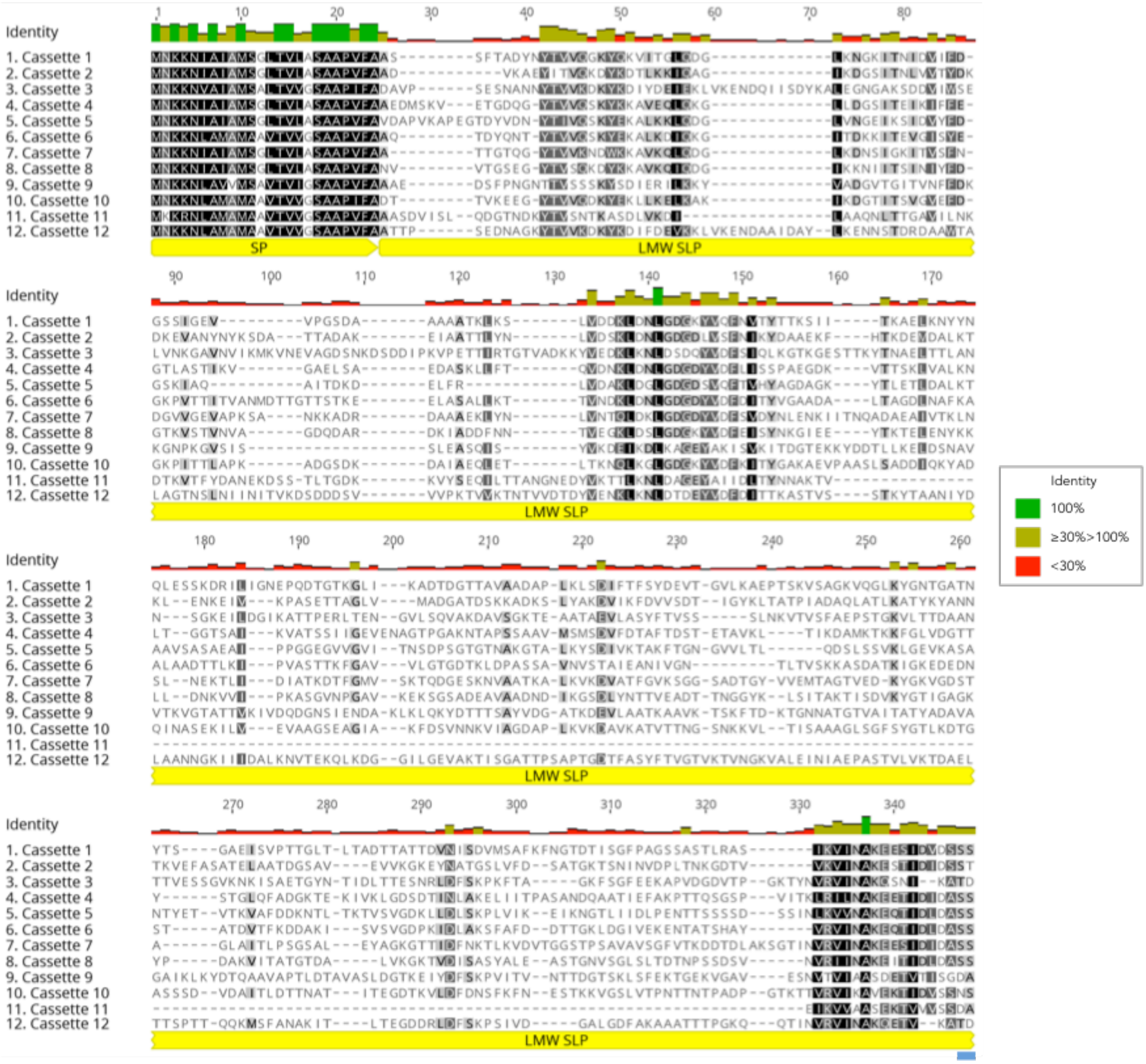
Appendix 1: Comparing structures of CWB2 domains and SLH domains

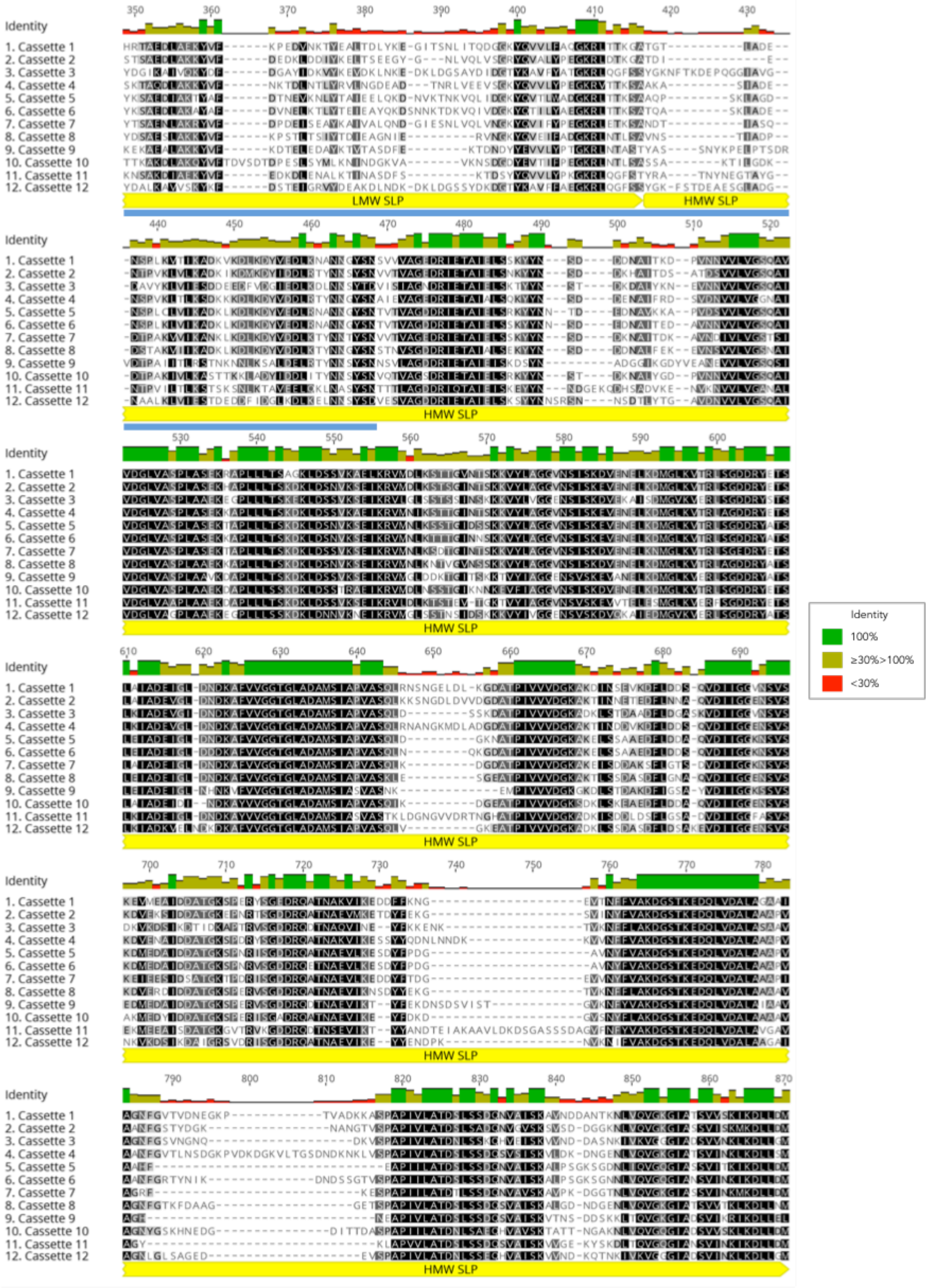
The crystal structures of CWB2 domains and SLH domains compared to show the pseudo-trimeric arrangement of alpha-helices forming the cell wall binding region of cell surface proteins in *C. difficile* and *Bacillus anthracis*. The domains 1, 2 and 3 have been coloured magenta, blue and grey, respectively for both the structures. **A** and **B**: Ribbon and surface representations of the structure of the CWB2 domains from CWP8 in *C. difficile* strain 630 (Usenik et al, 2017) (pdb no. 5J6Q). **C** and **D**: Ribbon and surface representations of the structure of the SLH domains from *Bacillus anthracis* (Kern et al, 2011) (pdb no. 3PYW).



Appendix 2: SlpA sequences aligned from 12 S-layer cassette types of *C. difficile* aligned

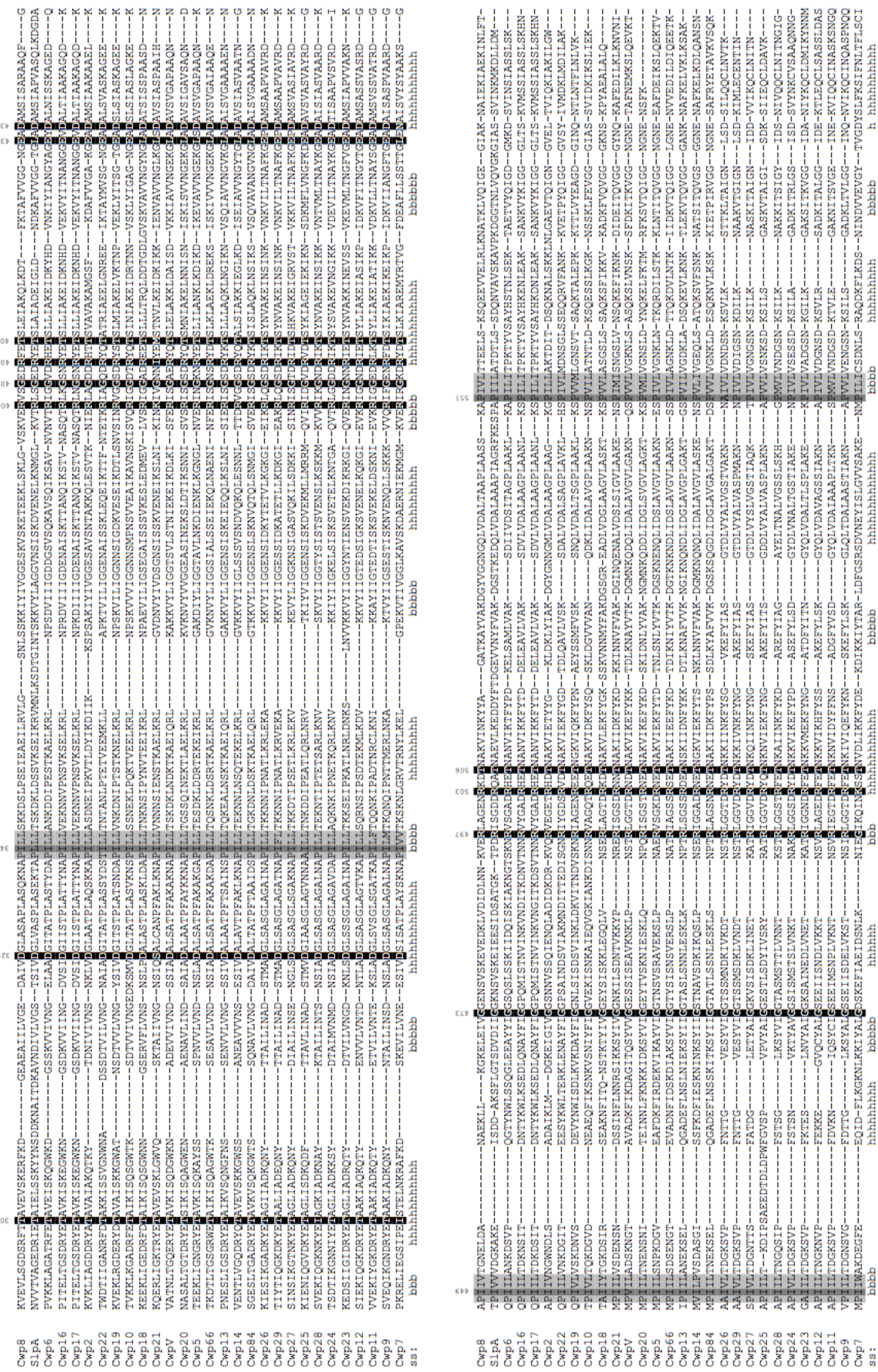
Alignment of 12 cassettes types of SlpA, from 12 different strains of *C. difficile,* showing conservation indicated by shading (black shading, white text, 100%; dark grey shading, white text, 80 – 100%; light grey shading, black text, 80% – 60%; white shading, grey text, less than 60%). Green, olive and red bars, above the sequences, also indicate the identity among the sequences. The signal peptide (SP), LMW SLP, and HMW SLP regions are indicated using thick yellow arrows, based on known structural information (Fagan et al, 2009). LMW/HMW SLP interaction region, annotated according to strain 630 (Fagan et al, 2009) is indicated by blue line below the residues. The LMW SLP sequence ends with the last residue of mature LMW SLP based on homology to known N-terminal sequences of HMW SLP. Note the high levels of sequence identity in the HMW region. Sequences are aligned with *CLUSTALW* algorithm using *Geneious* version 10.1.




Identity
1. Cassette 1
2. Cassette 2
3. Cassette 3
4. Cassette 4
5. Cassette 5
6. Cassette 6
7. Cassette 7
8. Cassette 8
9. Cassette 9
10. Cassette 10
11. Cassette 11
12. Cassette 12
LMW SLP
HMW SLP
350
360
370
380
390
400
410
420
430
440
450
460
470
480
490
500
510
520
530
540
550
560
570
580
590
600
610
620
630
640
650
660
670
680
690
700
710
720
730
740
750
760
770
780
790
800
810
820
830
840
850
860
870
Identity
🟩 100%
≥30%>100%
🟥 <30%

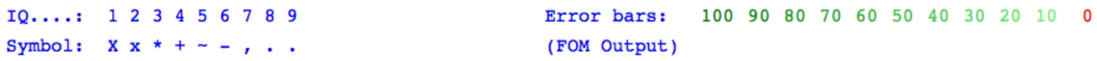
Appendix 3: Sequence alignment of CWB2 sequences from 29 Cwp proteins in *C. difficile* 630

Amino acid sequences for the CWB2 trimer are aligned, showing high level of sequence identity. The black and grey highlights indicate the conserved surface residues based on the crystal structure of Cwp6 and 8 (Usenik et al, 2017). Consensus predicted secondary structures are indicated as h for α-helix and b for β-strand. Image taken with permission from (Usenik et al, 2017), copyrighted by Elsevier Ltd.



Appendix 4: Amplitude and phase variations along lattice lines as a function of increasing *z** for indicated merged *h*,*k* from the 3D transform of R7404 S-layer

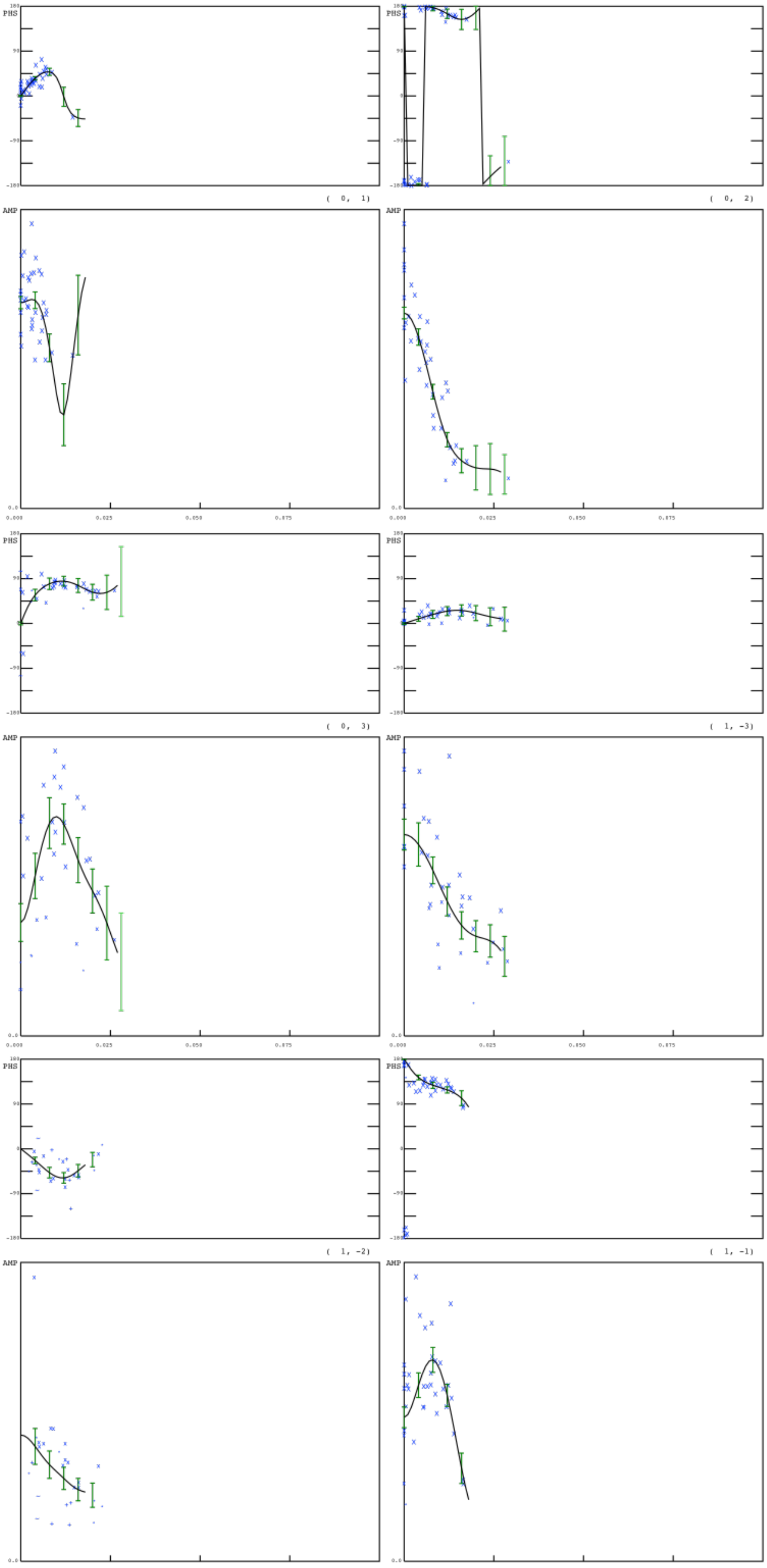
Key:



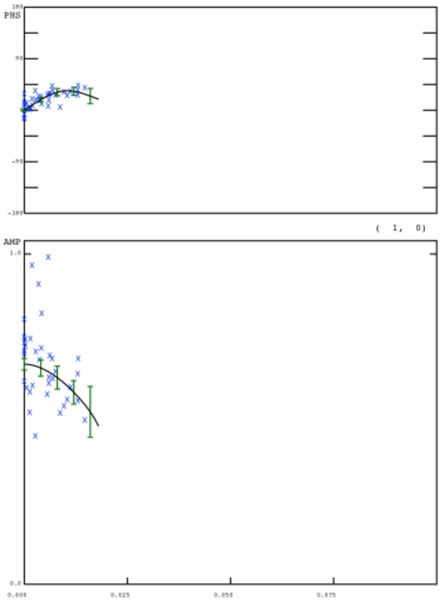
```
IQ....:  1 2 3 4 5 6 7 8 9                 Error bars:   100 90 80 70 60 50 40 30 20 10  0
Symbol:  X x * + ~ - , . .                 (FOM Output)
```

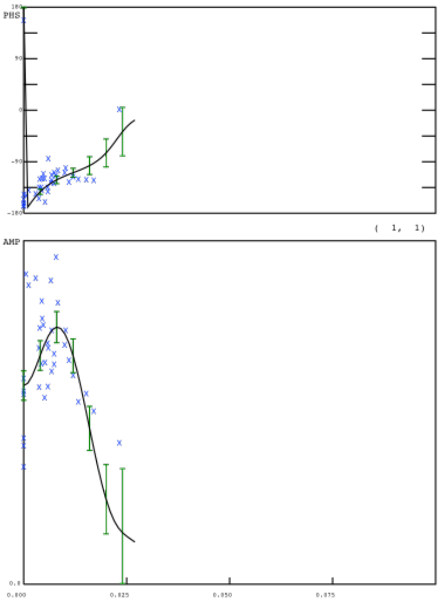
Appendices 4.1 to 4.4 shows the lattice lines from the merged data of R20291 S-layer. Smaller panel (labelled PHS) shows phase variation in degrees and larger panel below it (labelled AMP) shows amplitude variation along the latttice lines, for each indicated *h*,*k* indexed lines. Horozontal axis shows increasing z* value (in reciprocal angstroms) from the *x*y** plane. The green lines indicate error bars, with darker green colour indicating lower standard error, in lattice line fitting. The blue symbols indicate IQ values for the experimental spots as shown in the key at the bottom.

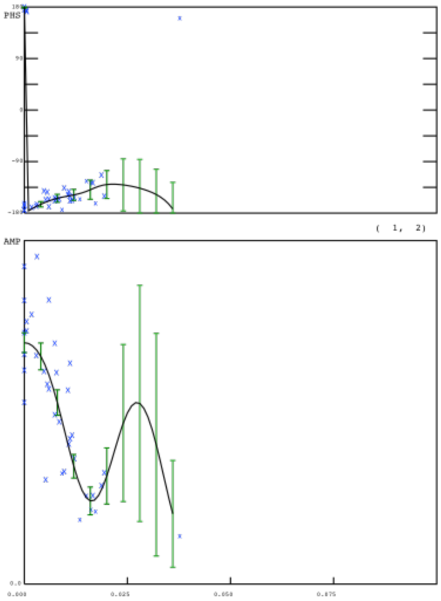
Appendix 4.1 *Lattice lines R7404 S-layer*

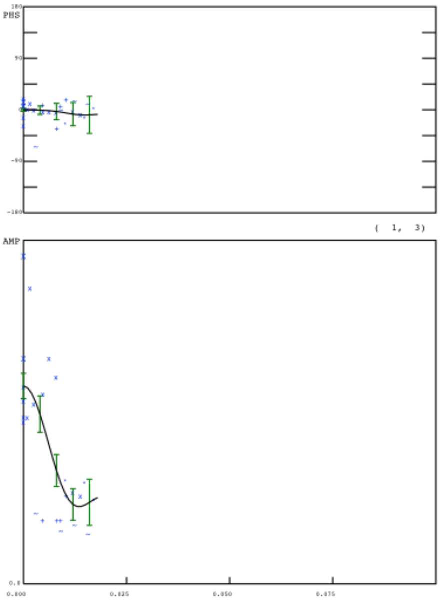


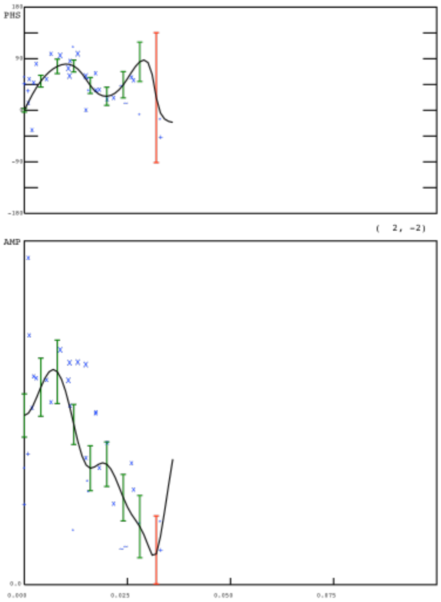
Appendix 4.2 *Lattice lines R7404 S-layer*

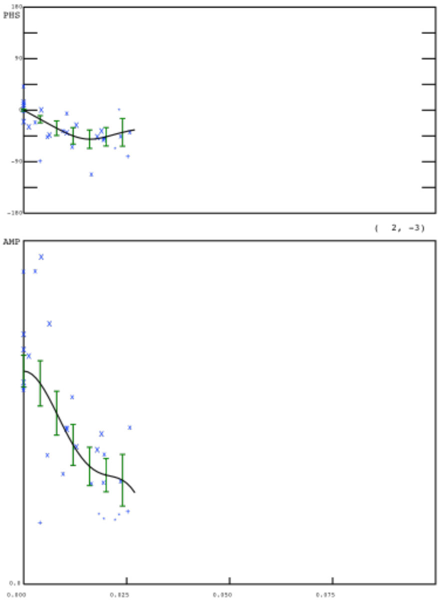




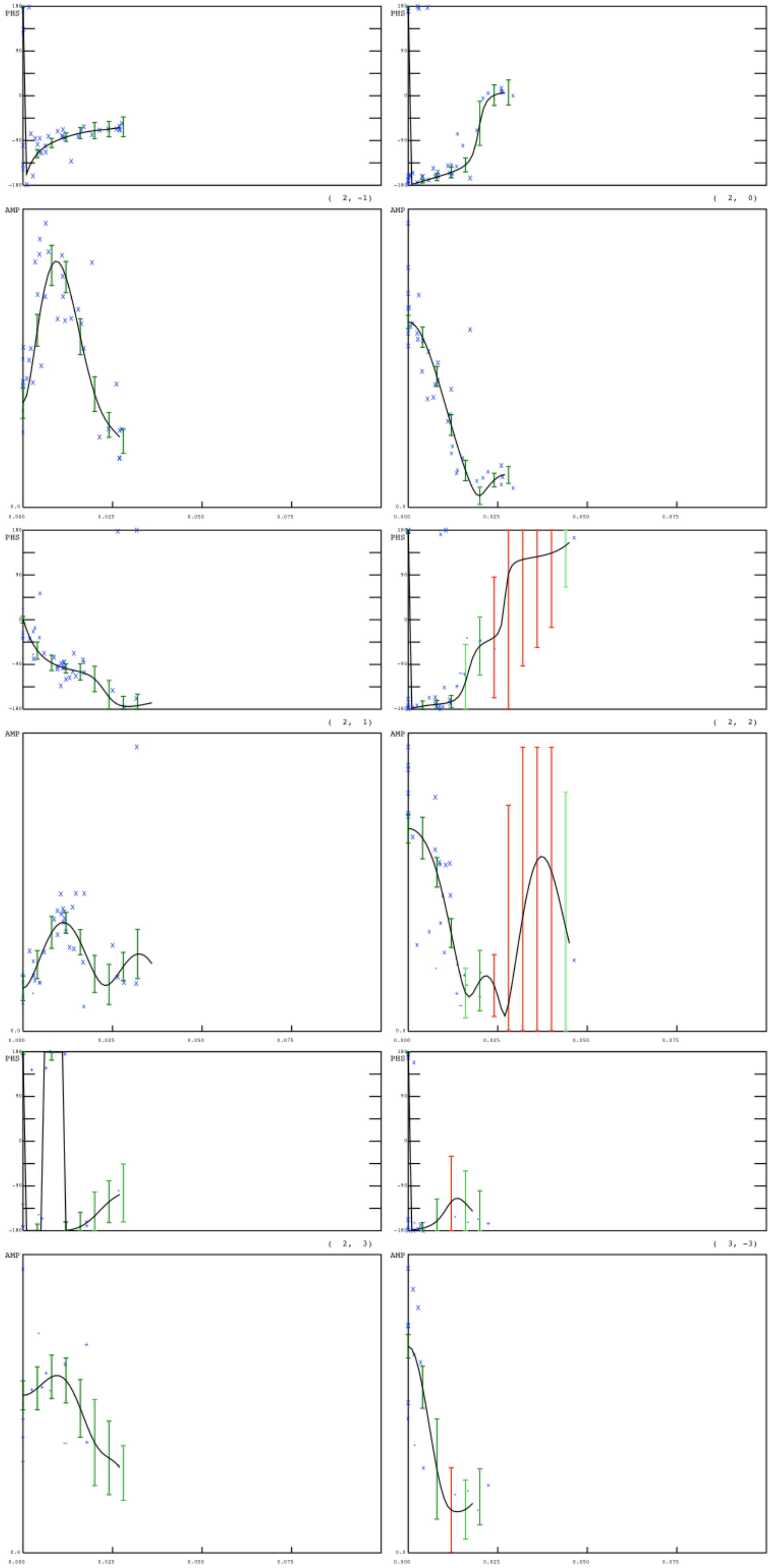




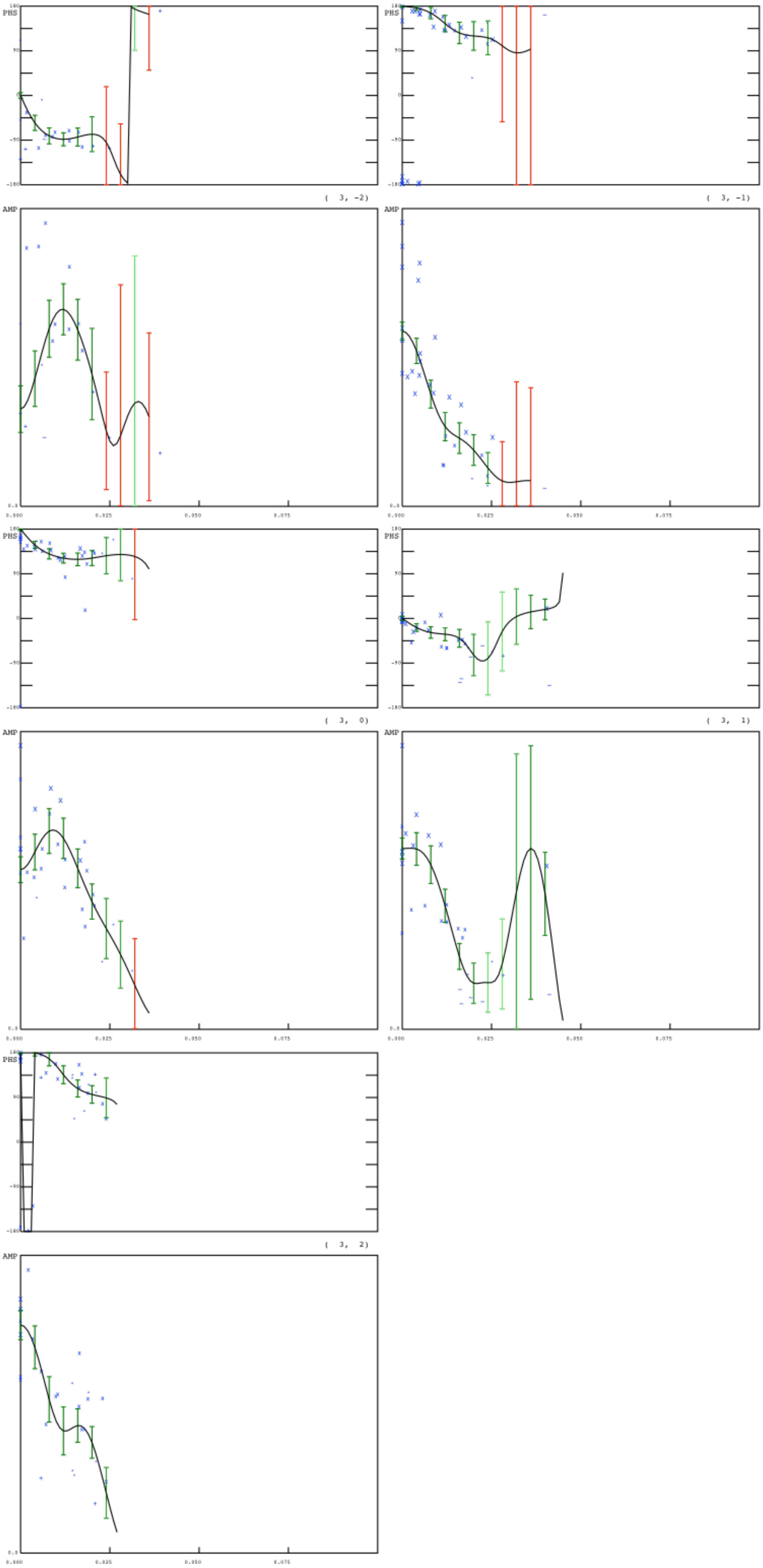




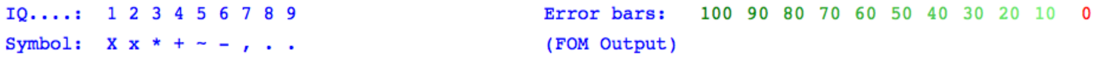
Appendix 4.3 *Lattice lines R7404 S-layer*



Appendix 4.4 *Lattice lines R7404 S-layer*



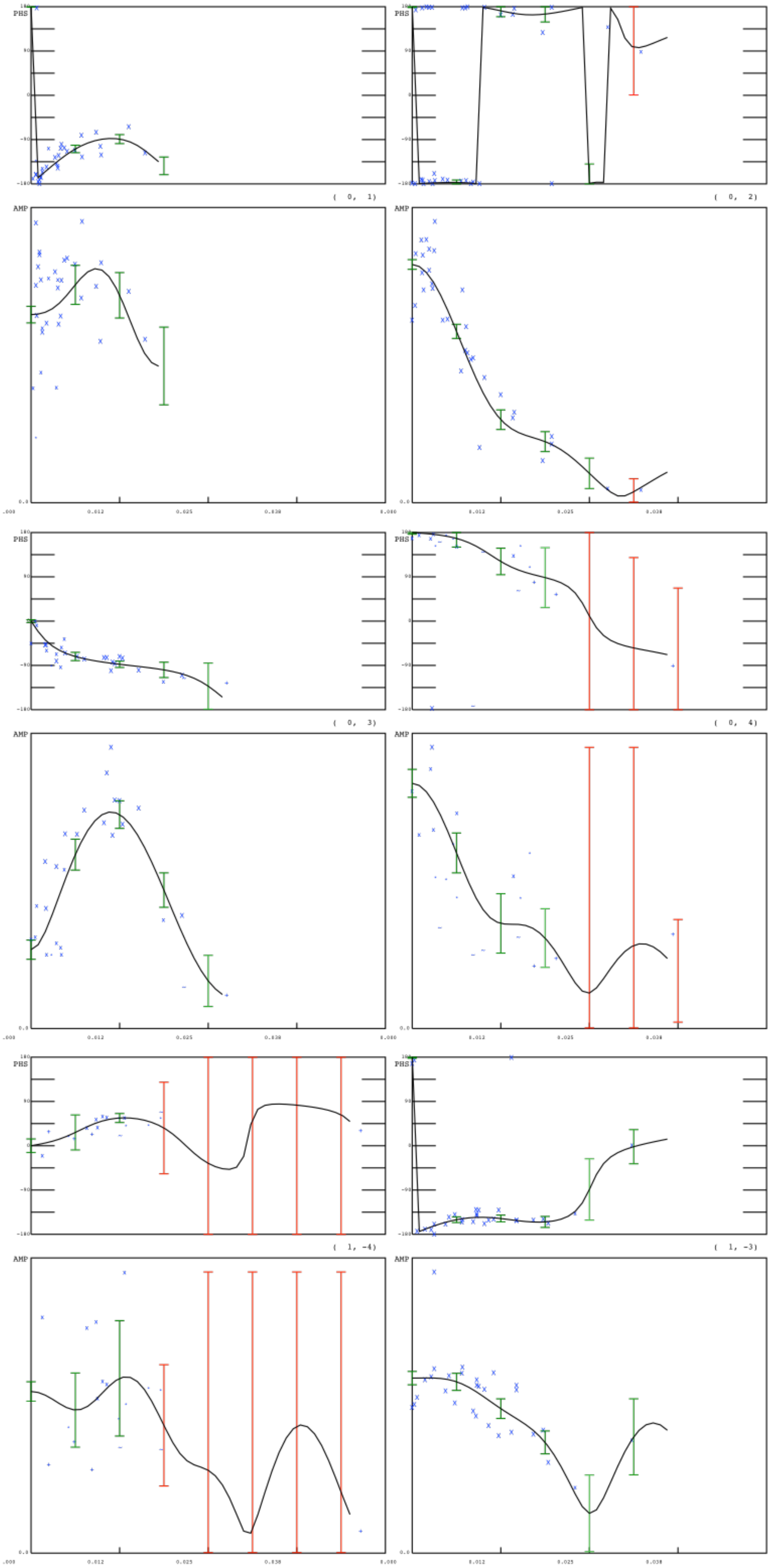
Appendix 5: Amplitude and phase variations along lattice lines as a function of increasing *z** for indicated merged *h*,*k* from the 3D transform of R20291 S-layer



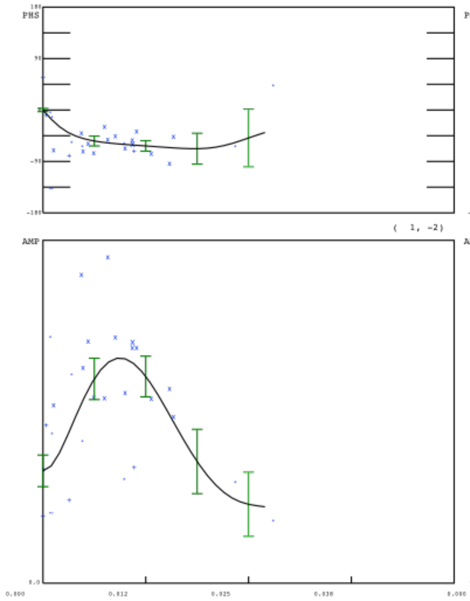
```
IQ....:  1 2 3 4 5 6 7 8 9                    Error bars:   100 90 80 70 60 50 40 30 20 10  0
Symbol:  X x * + ~ - , . .                    (FOM Output)
```

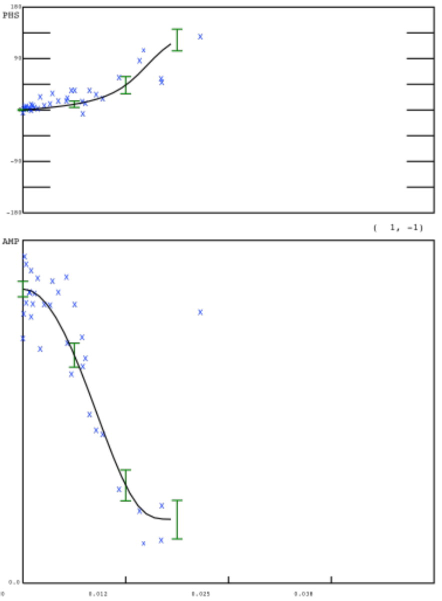
Appendices 5.1 to 5.4 shows the lattice lines from the merged data of R20291 S-layer. Smaller panel (labelled PHS) shows phase variation in degrees and larger panel below it (labelled AMP) shows amplitude variation along the latttice lines, for each indicated *h*,*k* indexed lines. Horozontal axis shows increasing z* value (in reciprocal angstroms) from the *x*y** plane. The green lines indicate error bars, with darker green colour indicating lower standard error, in lattice line fitting. The blue symbols indicate IQ values for the experimental spots as shown in the key at the bottom.

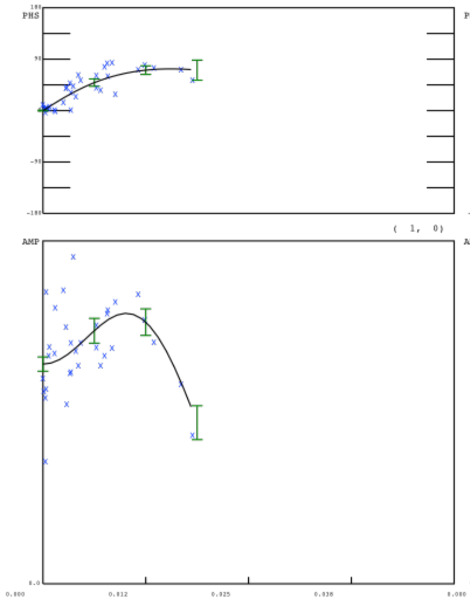
Appendix 5.1 *Lattice lines R20291 S-layer*

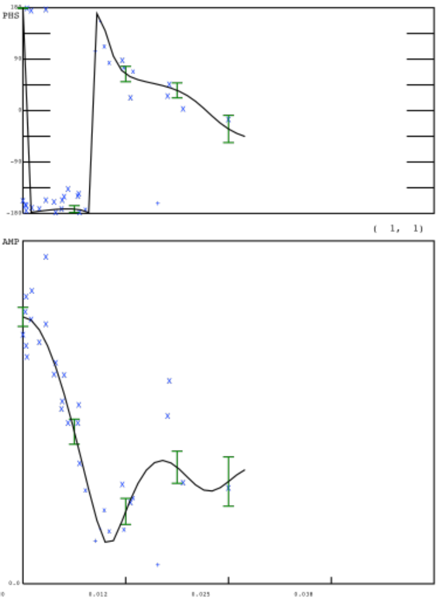


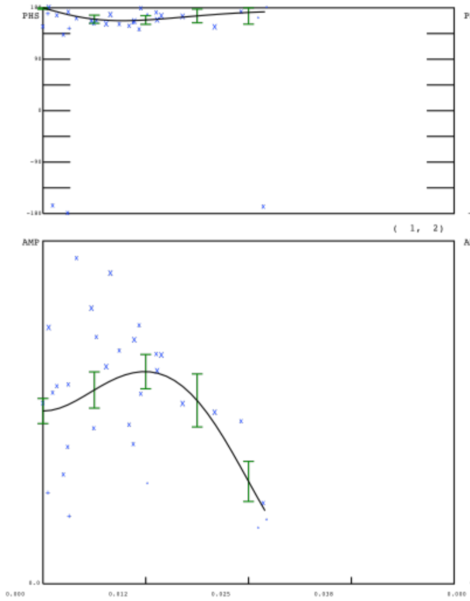
Appendix 5.2 *Lattice lines R20291 S-layer*

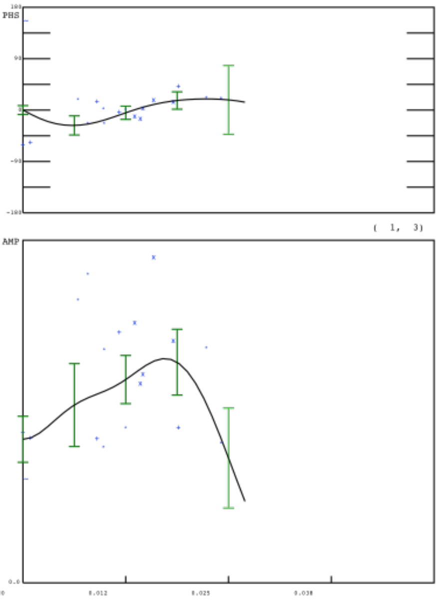




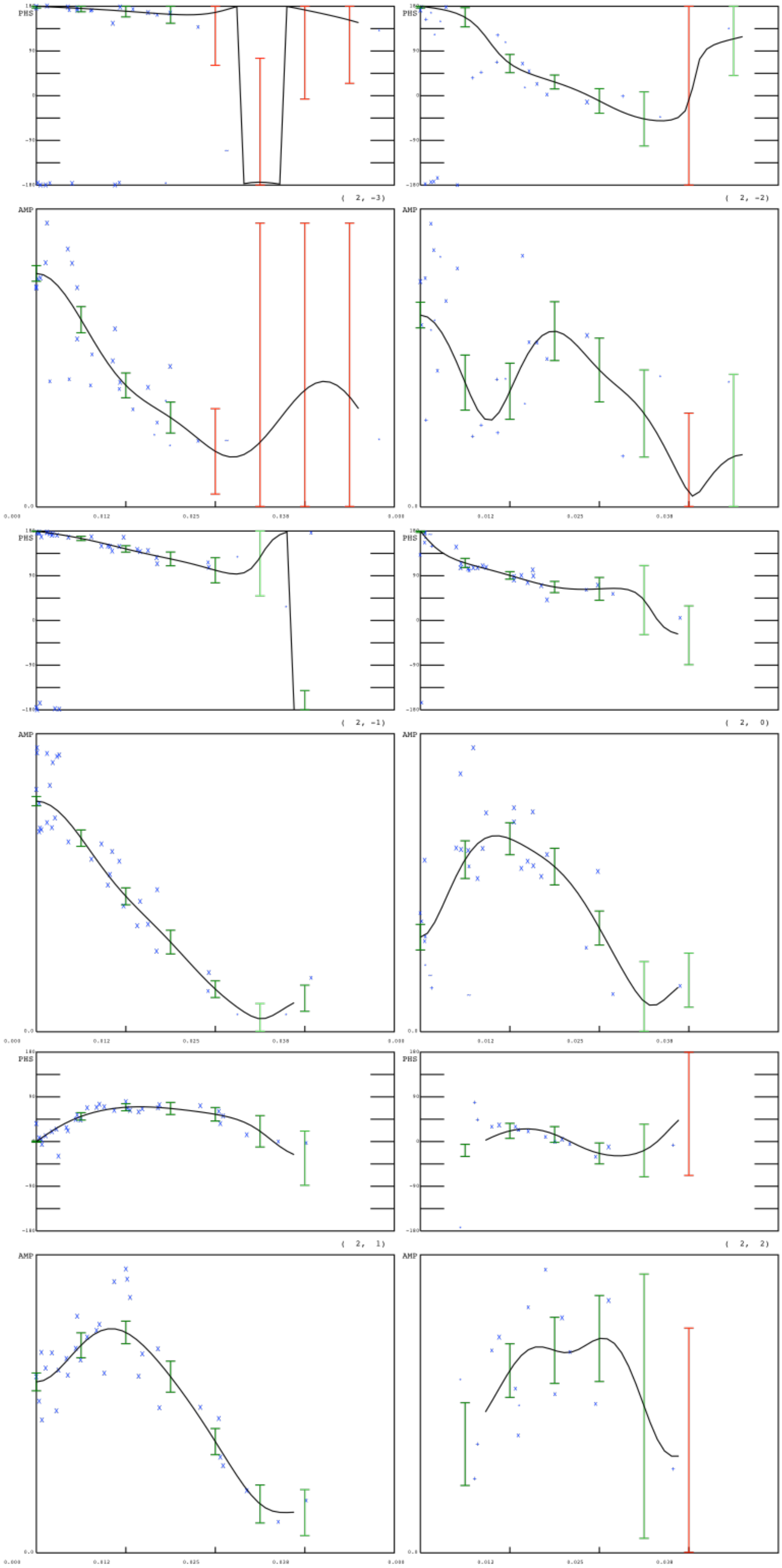




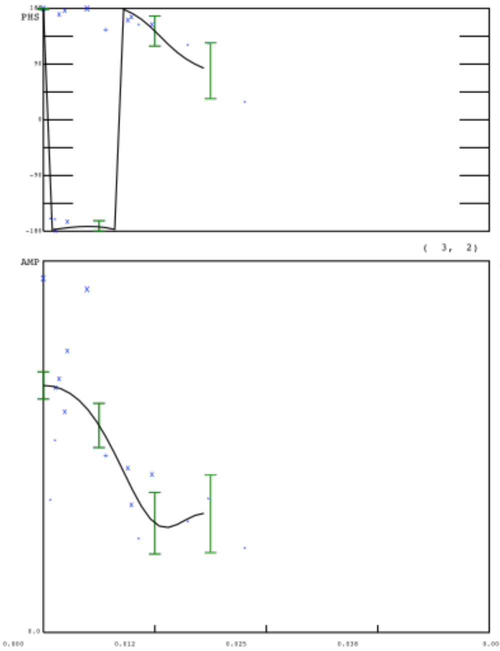


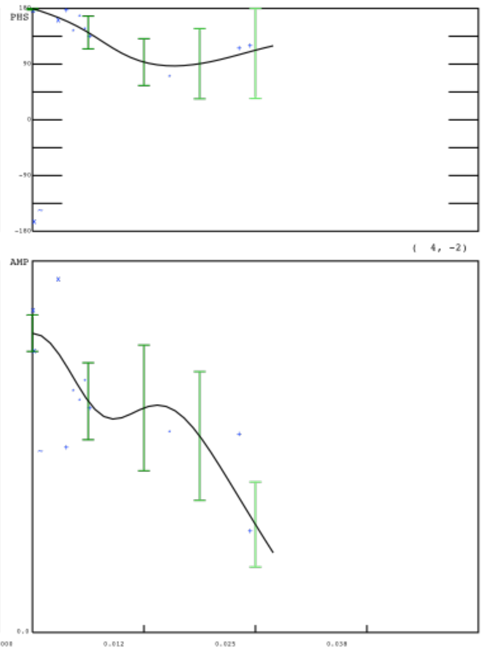


Appendix 5.3 *Lattice lines R20291 S-layer*



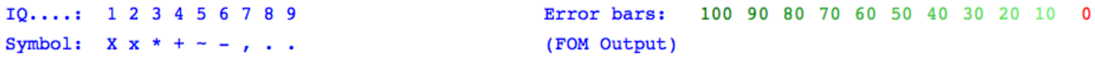
Appendix 5.4 *Lattice lines R20291 S-layer*





Appendix 6: Amplitude and phase variations along lattice lines as a function of increasing *z** for indicated merged *h,k* from the 3D transform of SlpA$^{\Delta Ld2}$ S-layer

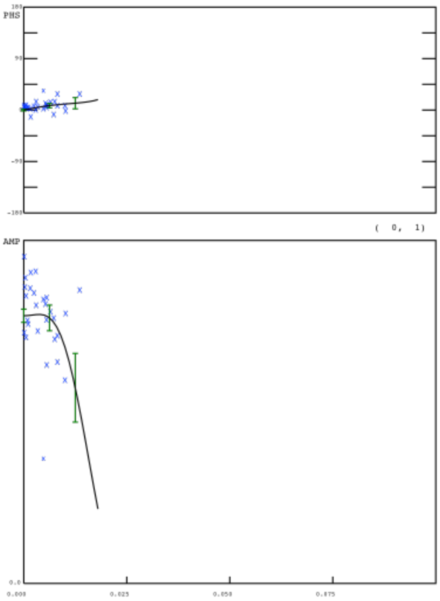
Key:

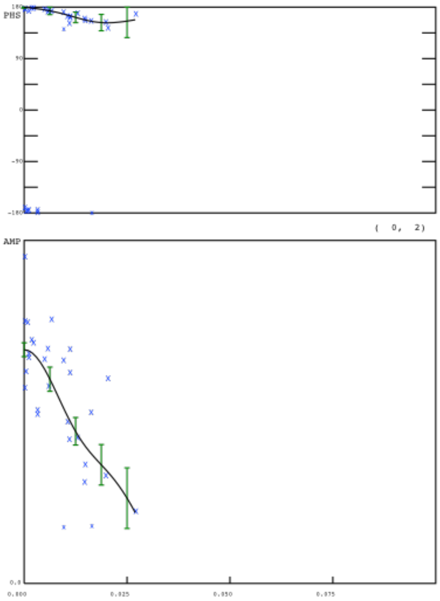


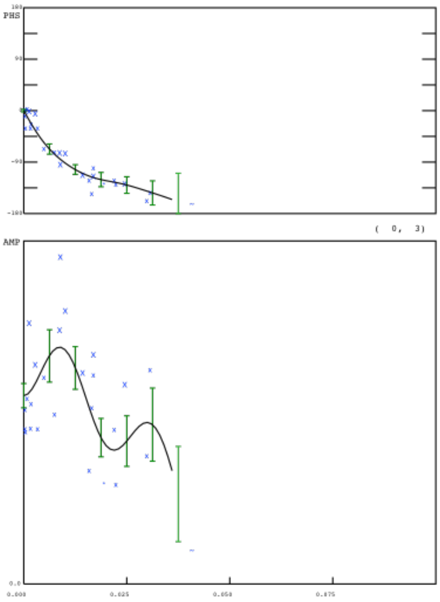
```
IQ....:  1 2 3 4 5 6 7 8 9                     Error bars:   100 90 80 70 60 50 40 30 20 10  0
Symbol:  X x * + ~ - , . .                     (FOM Output)
```

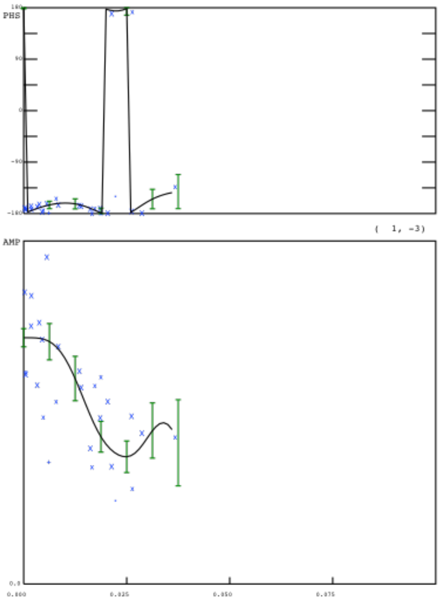
Appendices 6.1 to 6.4 shows the lattice lines from the merged data of R20291 S-layer. Smaller panel (labelled PHS) shows phase variation in degrees and larger panel below it (labelled AMP) shows amplitude variation along the latttice lines, for each indicated *h*,*k* indexed lines. Horozontal axis shows increasing z* value (in reciprocal angstroms) from the *x*y** plane. The green lines indicate error bars, with darker green colour indicating lower standard error, in lattice line fitting. The blue symbols indicate IQ values for the experimental spots as shown in the key at the bottom.

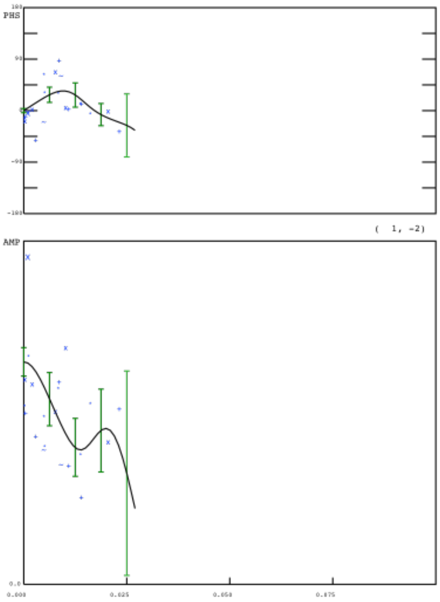
Appendix 6.1 *Lattice lines SlpA$^{\Delta Ld2}$ S-layer*

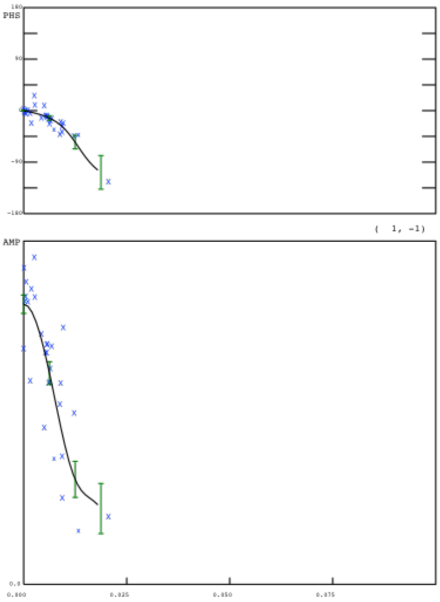




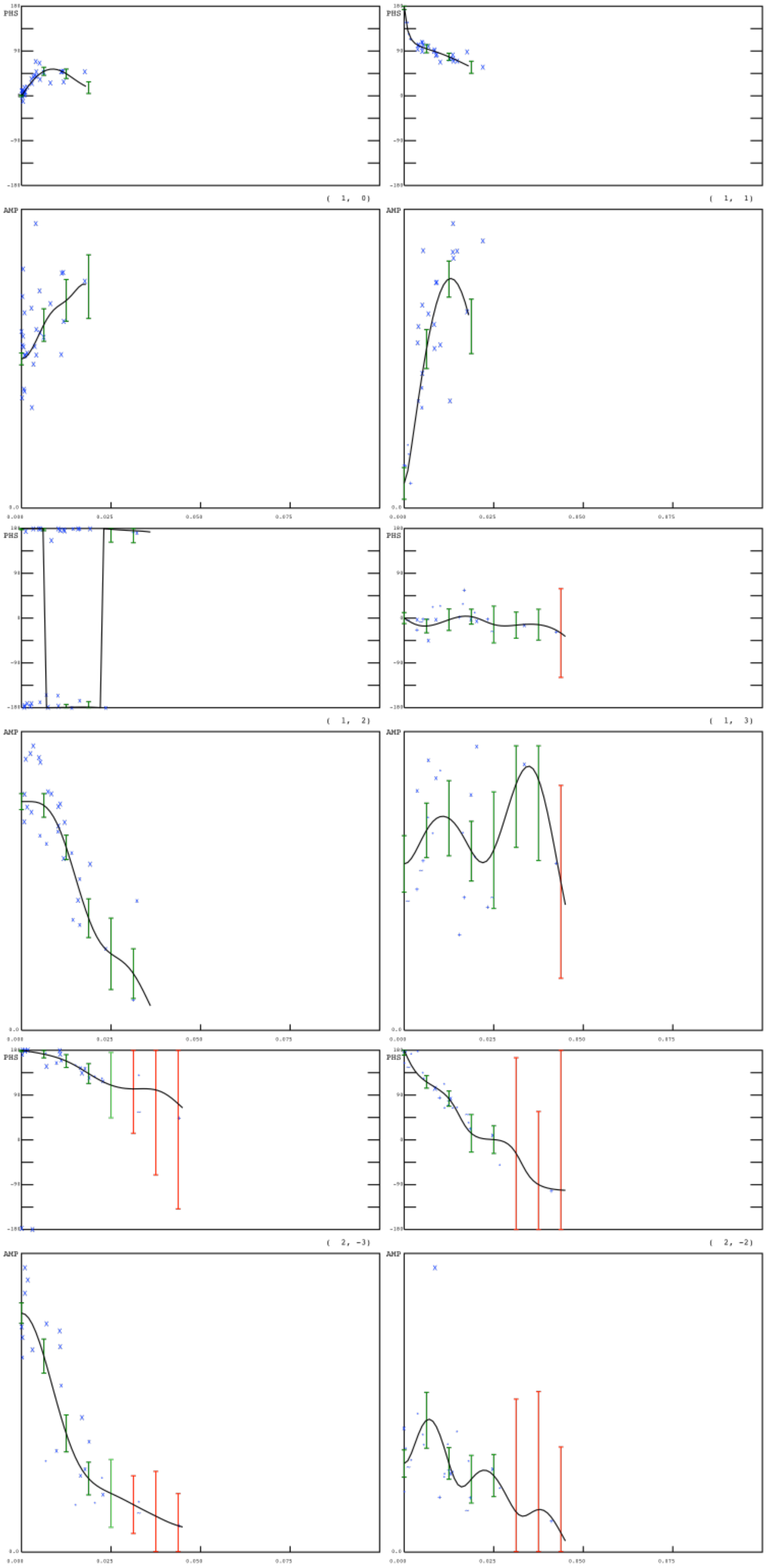




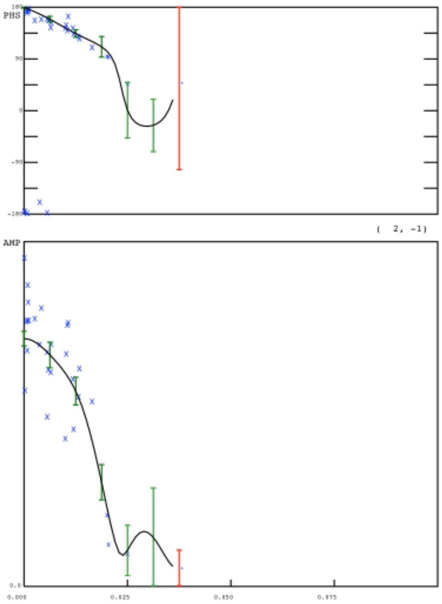


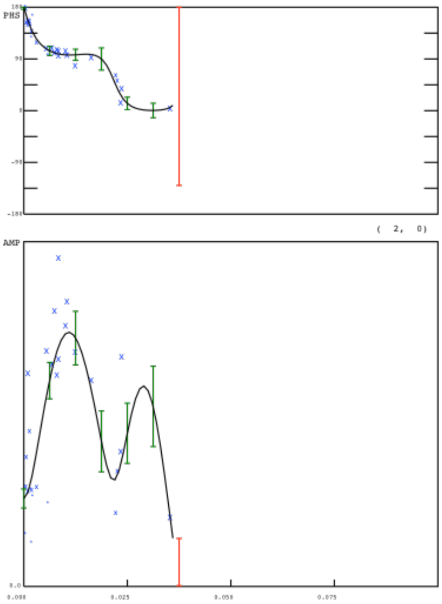


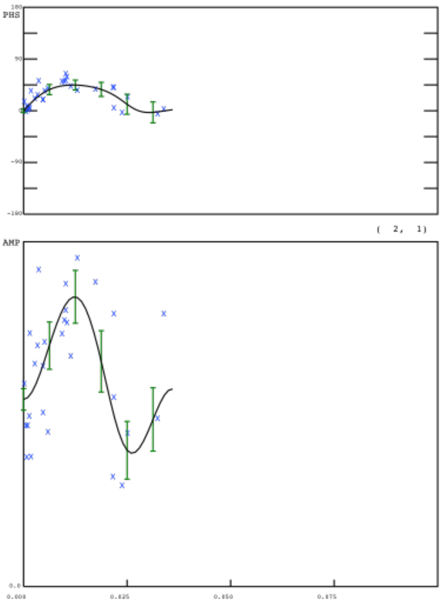
Appendix 6.2 *Lattice lines SlpA$^{\Delta Ld2}$ S-layer*

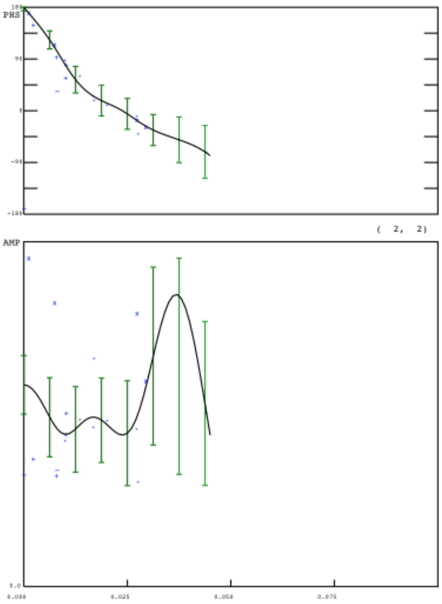


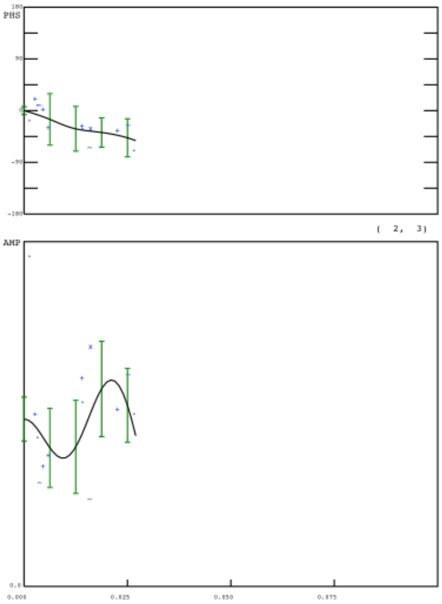
Appendix 6.3 *Lattice lines SlpA$^{\Delta Ld2}$ S-layer*

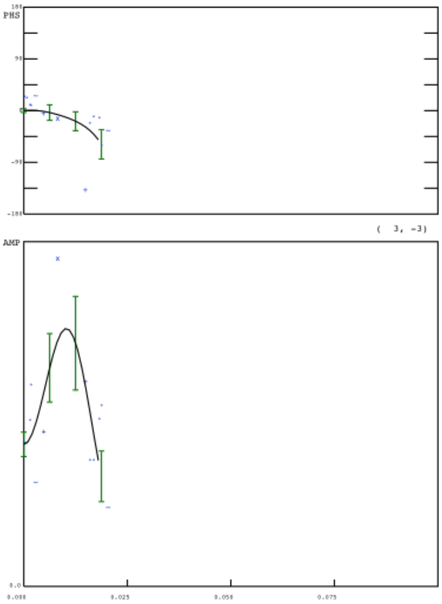




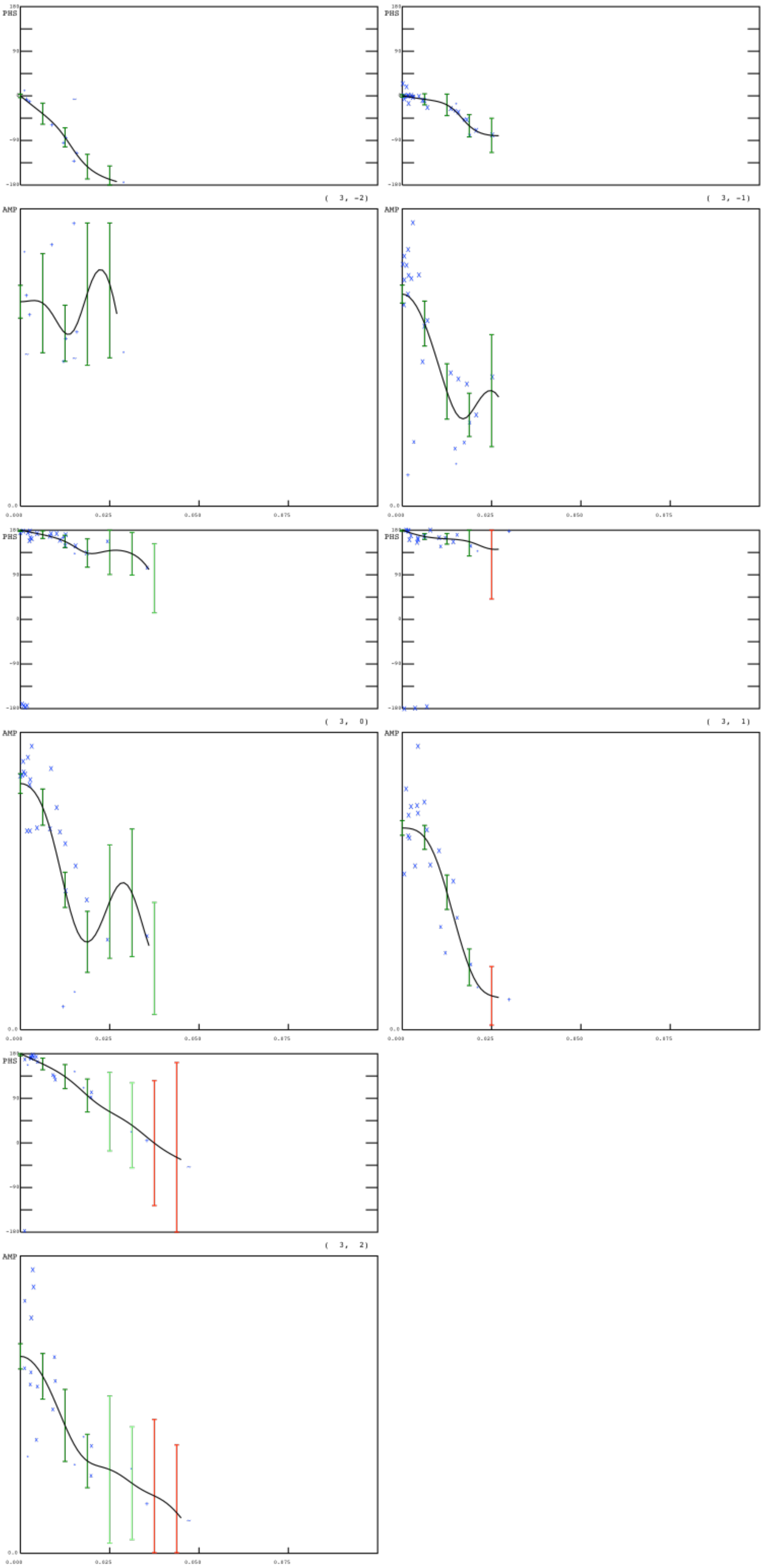






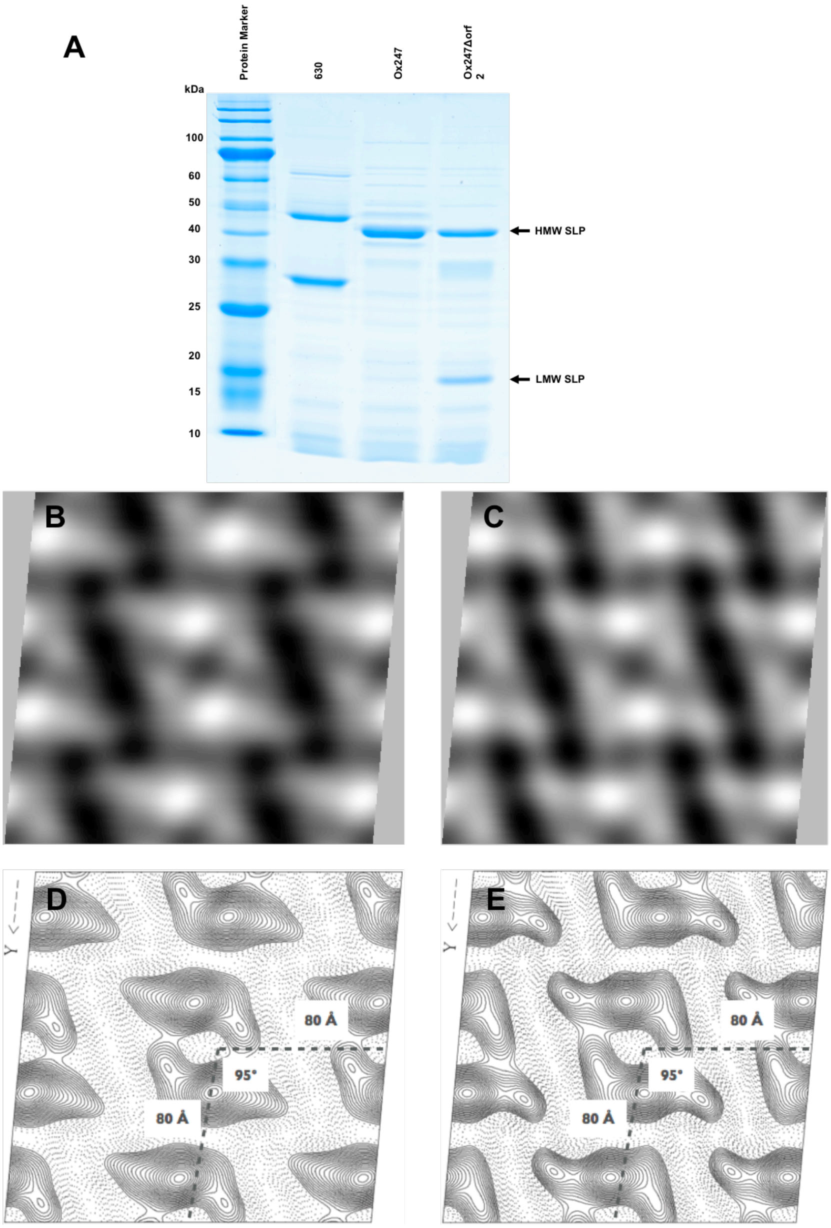


Appendix 6.4 *Lattice lines SlpA*$^{\Delta Ld2}$ *S-layer*



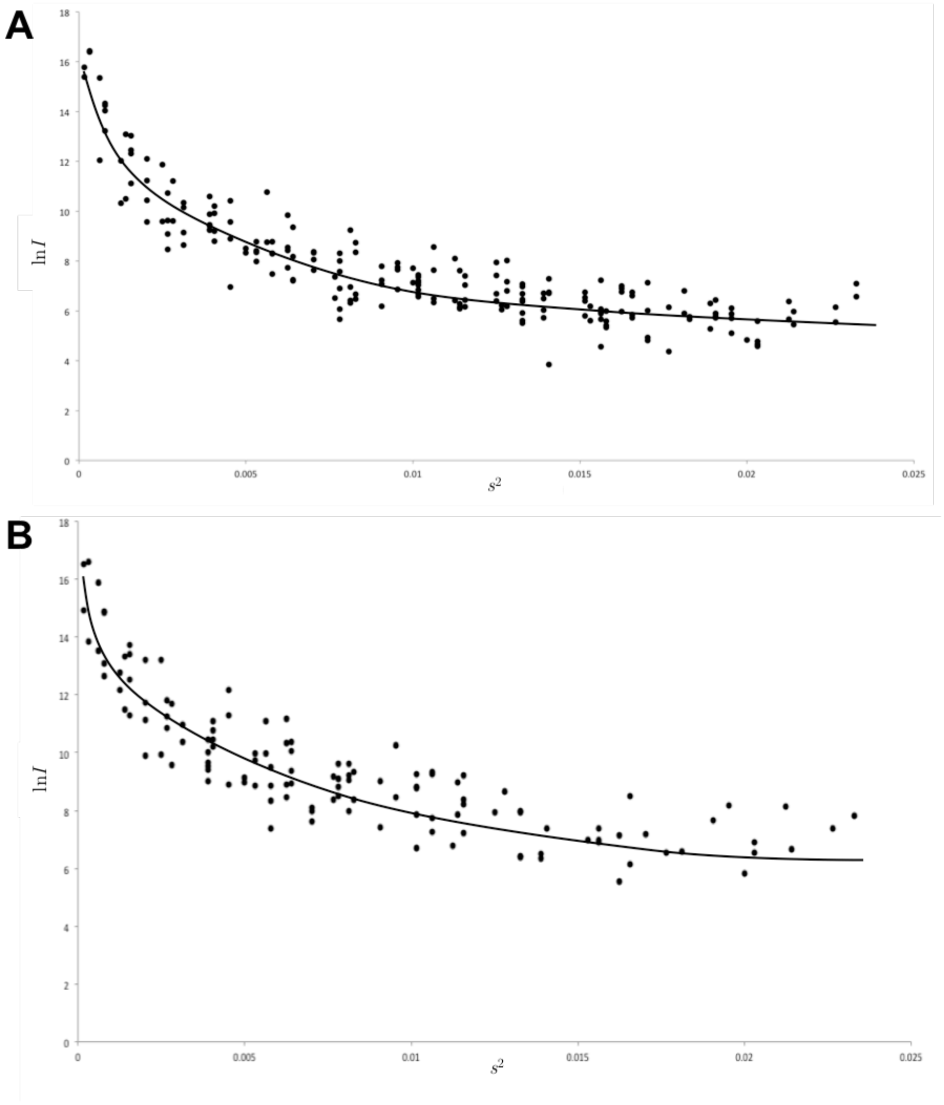
Appendix 7: Projection structure of Ox247 glycosylated S-layer and Ox247Δorf2 mutant

C. difficile strain Ox247 has a putative glycosylation locus in the S-layer cassette (Dingle et al, 2013) and shows a glycosylated LMW SLP in the S-layer. Ox247 Δ orf2 has a Clostron mutation in the glycosylation cluster (ORF-2 in the glycosylation cluster with putative function of glycan synthesis initiation) in the S-layer cassette expressing SlpA without glycosylation. The SDS PAGE gel (**A**) shows a low pH glycine extraction of the two strains compared with strain 630. The HMW and LMW SLP are indicated. The LMW SLP for Ox247 is unstained due to the presence of glycosylation. The projection structure of the S-layer of Ox247 is shown as a density (**B**) and contour plot (**D**). The projection structure of the S-layer of Ox247 Δ orf2 mutant is shown as a density (**C**) and contour plot (**E**). The unit cell dimensions are indicated along dashed lines representing a single unit cell. The symmetry is p2 with the symmetry axis perpendicular to the page. The EM density difference between the two projections was not significant.



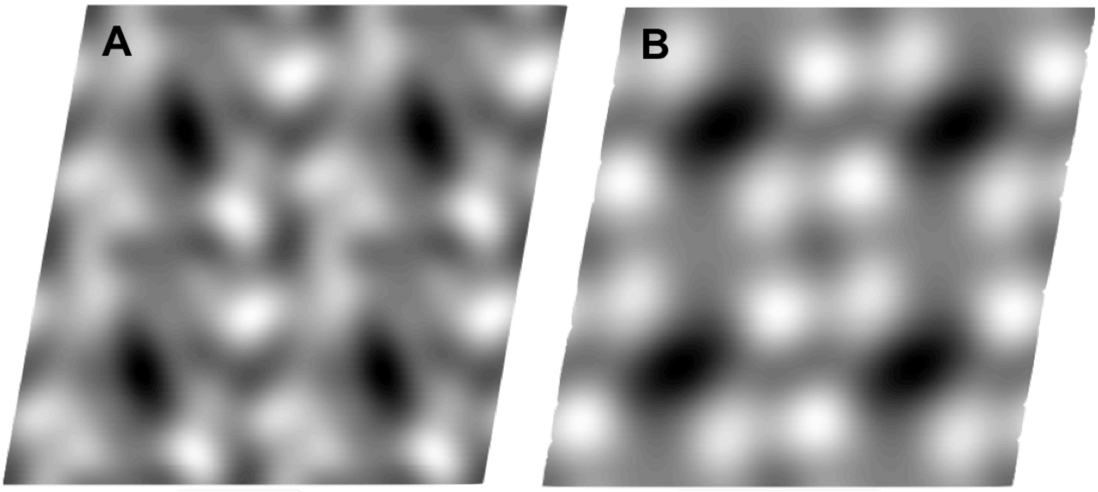
Appendix 8: Guinier plots showing decreasing image intensity with increasing resolution in cryo-EM data

The natural logarithm of the recorded intensities from S-layer merged projection data of R20291 (**A**) and SlpA$^{\Delta Ld2}$ (**B**) S-layer fragment (data described in Chapter 5) is plotted against the square of spatial frequency, *s*. The fall-off in recorded intensity is clearly visible. This is corrected using a negative temperature factor (B-factor). The resolution is truncated to 7.5 Å and 8.7 Å for R20291 and SlpA$^{\Delta Ld2}$ S-layer projection data.



Appendix 9: Improved phases using low electron dose with negative stain

Projections structures of merged unsymmetrised projections of 2 × 2 unit cells of untilted R7404 S-layer show the detrimental effect of electron dose on structural information, on a negative stained crystal. **A** shows more detailed structural information obtained from the identical sample as in **B**. Enhanced structural information is a result of better phases obtained from low-dose imaging techniques used in **A**. Electron dose was kept around 20 $e^-/\text{Å}^2$. **B** was imaged using conventional high electron dose imaging with dose exceeding 80,000 $e^-/\text{Å}^2$.

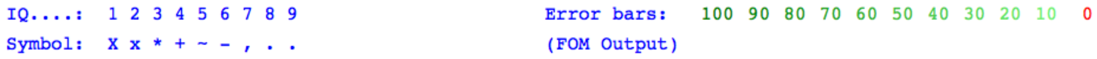


$20\ e^-/\text{Å}^2$

$> 80{,}000\ e^-/\text{Å}^2$

Appendix 10: Amplitude and phase variations along lattice lines as a function of increasing *z** for indicated merged *h*,*k* from the 3D transform of BslA-AxA S-layer

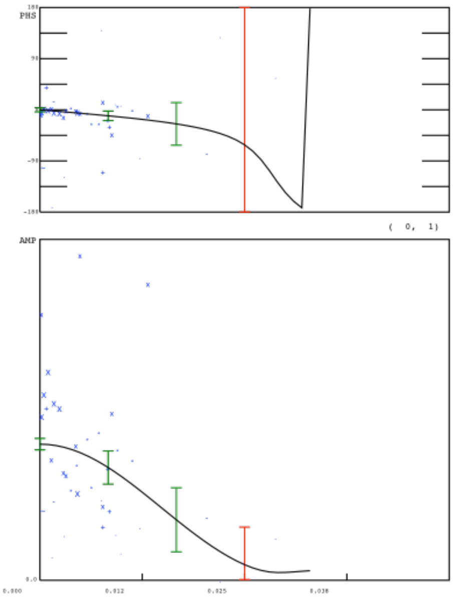
Key:

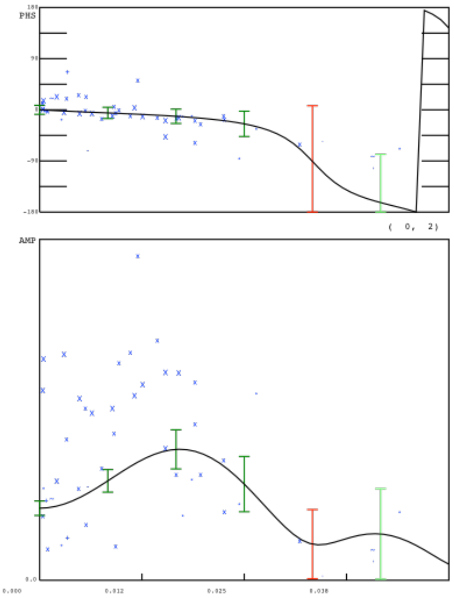


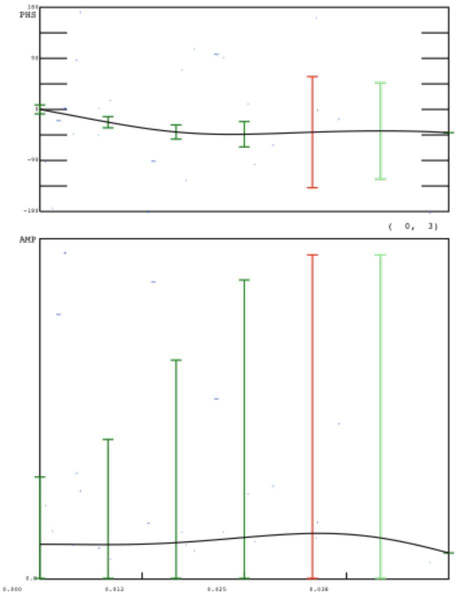
```
IQ....:  1 2 3 4 5 6 7 8 9                       Error bars:   100 90 80 70 60 50 40 30 20 10  0
Symbol:  X x * + ~ - , . .                       (FOM Output)
```

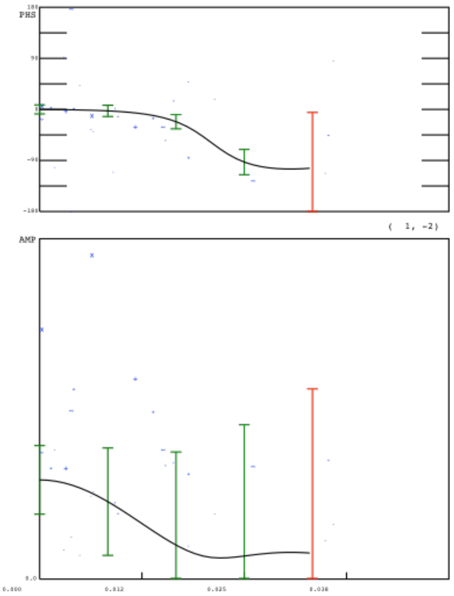
Appendices 9.1 to 9.3 shows the lattice lines from the merged data of R20291 S-layer. Smaller panel (labelled PHS) shows phase variation in degrees and larger panel below it (labelled AMP) shows amplitude variation along the latttice lines, for each indicated *h*,*k* indexed lines. Horozontal axis shows increasing z* value (in reciprocal angstroms) from the *x*y** plane. The green lines indicate error bars, with darker green colour indicating lower standard error, in lattice line fitting. The blue symbols indicate IQ values for the experimental spots as shown in the key at the bottom.

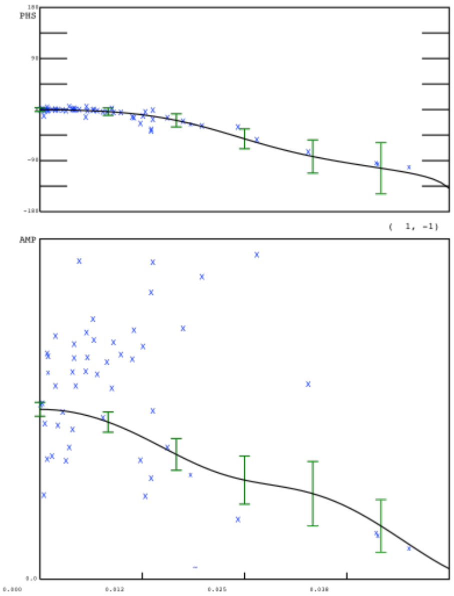
Appendix 10.1 *Lattice lines BslA-AxA S-layer*

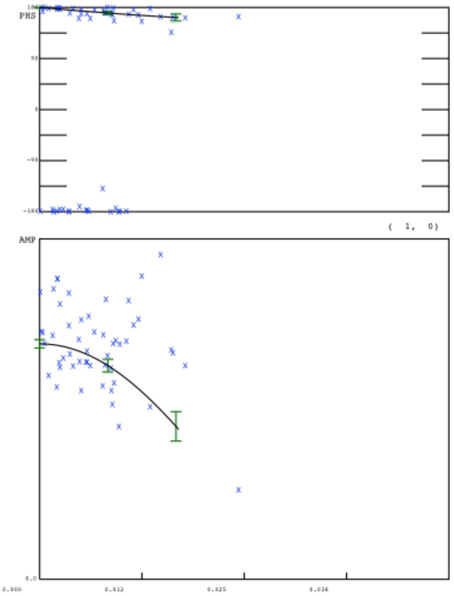




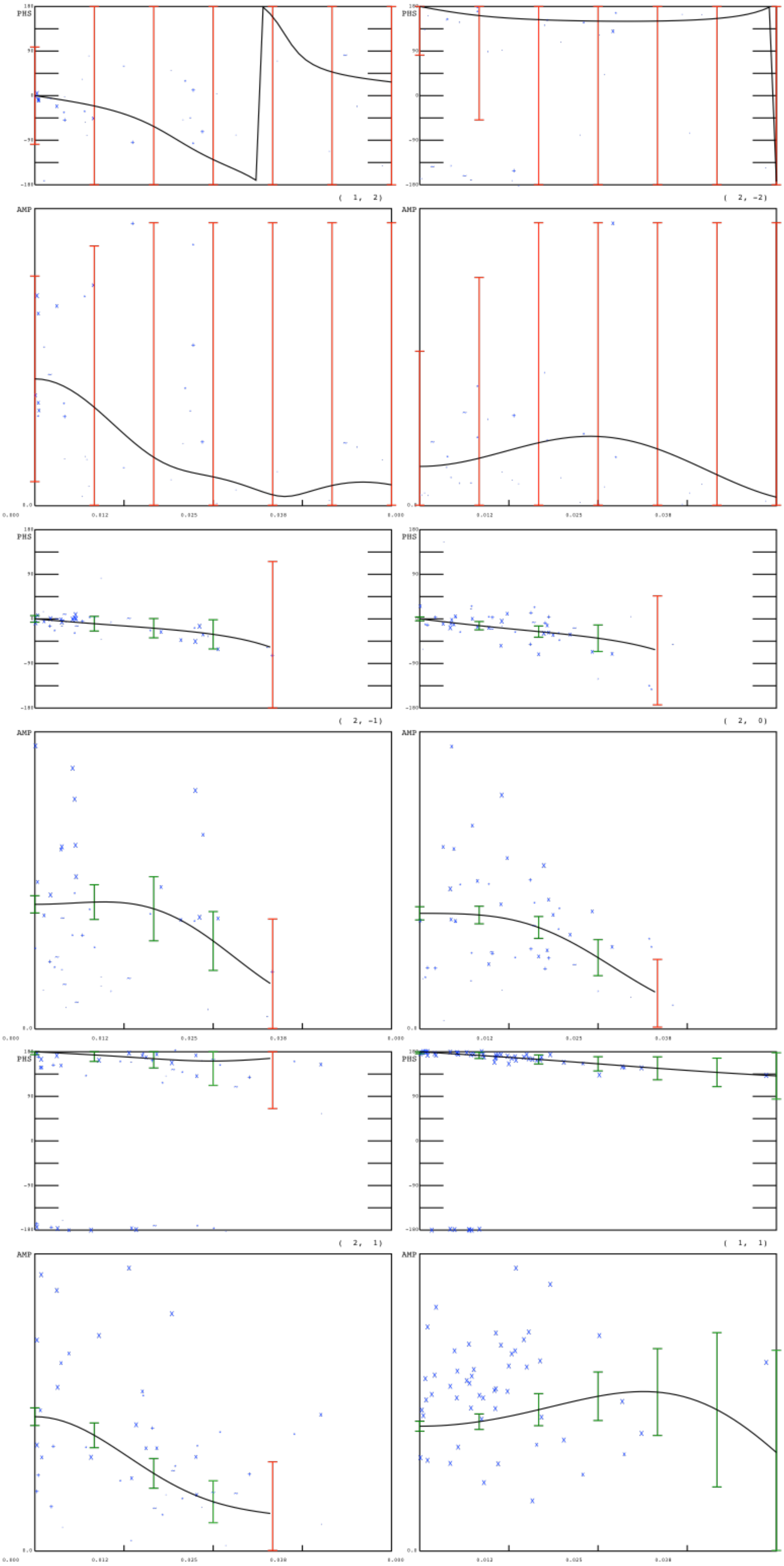








Appendix 10.2 *Lattice lines BslA-AxA S-layer*



Appendix 10.3 *Lattice lines BslA-AxA S-layer*