Structural studies of bacterial biofilm matrix fibres

Jan Böhning
Lincoln College
University of Oxford

Trinity Term 2022

Submitted for the degree of
Doctor of Philosophy

Abstract

Many bacteria live in multicellular communities called biofilms. Within the biofilm, cells are embedded in a self-secreted extracellular biofilm matrix, which provides numerous benefits to bacteria inside. Biofilms also form during human infection, where the matrix provides protection against antibiotic treatment and the immune system. A major component of the biofilm matrix are protein fibres, which provide stability to the biofilm and scaffold its formation. Little is known about the structure of the biofilm matrix in general, and matrix fibres specifically. Here, structural and functional studies on biofilm matrix fibres are presented. An electron cryomicroscopy (cryo-EM) structure of the biofilm-promoting archaic Chaperone-Usher pilus CupE from *Pseudomonas aeruginosa* was solved, showing a zig-zag subunit architecture, where the majority of inter-subunit interactions are mediated by an N-terminal donor strand that extends into the following subunit and completes its Ig-like fold. Electron cryotomography (cryo-ET) imaging shows that the CupE pilus can adopt significant curvature *in situ*, which may aid in promoting cohesion between cells. Structural studies on the biofilm matrix fibre Fap from *P. aeruginosa* are presented, supporting previous work showing the fibres are amyloid. Moreover, a cryo-EM fibre structure of the major protein component of the *Bacillus subtilis* biofilm matrix, TasA, was solved. The structure shows that TasA fibres are not, like previously proposed, amyloid but instead polymerise through a donor strand, similar to Chaperone-Usher pili, but employing a different, previously undescribed assembly machinery. Finally, a compressed sensing algorithm is trialled for reconstruction of cryo-ET data of biological specimens, leading to improved direct visualisation and subtomogram averaging of macromolecules from small datasets,

which may aid future studies in visualising the biofilm matrix *in situ*. Together, these findings will advance our structural understanding of biofilms.

Certification

I declare that this document has not already been submitted for any degree and is not being submitted as part of candidature for any other degree. I declare that the document has been written solely by me, and that all sources and materials presented but not authored by me have been acknowledged. I furthermore certify that the word count of this thesis does not exceed 50,000 words.

Signature of candidate:

Publications pertaining to this thesis

Peer-reviewed publications

Böhning, J., Bharat, T.A., Collins, S.M. (2022). Compressed sensing for electron cryotomography and high-resolution subtomogram averaging from biological specimens. ***Structure**, 30(3), 408-417.*

Melia, C., Bolla, J.R., Katharios-Lanwermeyer, S., Mihaylov, D., Hoffmann, P., Huo, J., Wozny, M., Elfari, L., **Böhning, J.**, Morgan, A.N., Hitchman, C.J., Owens, R., Robinson, C.V., O'Toole, G.A., Bharat, T.A. (2021). Architecture of cell-cell junctions in situ reveals a mechanism for bacterial biofilm inhibition. ***PNAS**,* 118 (31) e2109940118.

Böhning, J. and Bharat, T.A. (2021). Towards high-throughput *in situ* structural biology using electron cryotomography. ***Progress in Biophysics and Molecular Biology***, *160*, pp.97-103.

Pre-print publications

Böhning, J., Dobbelstein, A., Sulkowski, N., Eilers, K., von Kügelgen, A., Tarafder, A. K., Alva, V., Filloux, A., & Bharat, T.A. (2022). Architecture of the biofilm-associated archaic CupE pilus from *Pseudomonas aeruginosa*. ***bioRxiv*** (in revision at *PLOS Pathogens*).

Böhning, J., Ghrayeb, M., Pedebos, C., Abbas, D., Khalid, S., Chai, L., & Bharat, T.A. (2022). Molecular architecture of the TasA biofilm scaffold in *Bacillus subtilis*. ***bioRxiv*** (in revision at *Nature Communications*).

News & Views articles

Böhning, J. and Bharat, T.A. (2021). Impressions of expression: bringing structure to the cell. ***Nature Reviews Microbiology**, 19*(6), pp.346-346.

International oral presentations

Böhning, J., Ghrayeb, M., Pedebos, C., Abbas, D., Khalid, S., Chai, L., & Bharat, T.A. (2022). Molecular architecture of the TasA biofilm scaffold in *Bacillus subtilis*.

EMBO Bacterial Networks (BacNet22), Sant Feliu de Guíxols, Spain

Acknowledgments

Firstly, I would like to thank my supervisor Dr. Tanmay Bharat, whose boundless curiosity and love of research have been inspiring and motivating throughout my degree, and whose advice has been greatly helpful in learning to navigate the world of research. My thanks also go to Prof. Jordan Raff, my graduate advisor, who has been providing helpful advice on science and career along the way. I would further like to thank the scientists and staff at the Sir William Dunn School of Pathology, who have been constantly striving to - and succeeding in - making the Department an inclusive and enjoyable place to work, and the Oxford-MRC Doctoral Training Partnership, which has broadened my horizons and helped inform my career. I would furthermore be remiss to not also thank both the Department and the Oxford-MRC DTP for generously providing funding for my studentship.

My thanks also go to our collaborators, including Dr. Sean Collins, Dr. Kira Glatzel, Dr. Thomas Clamens, Prof. Alain Filloux, Adrian Dobbelstein, Dr. Vikram Alva, Prof. Morten Dueholm, Mnar Ghrayeb, Prof. Liraz Chai, Dr. Conrado Pedebos and Prof. Syma Khalid, which not only helped to advance the projects but also helped to widen my perspective. I am also indebted to Daniel Abbas, who, as an undergraduate research student, has been greatly helpful and skilful in advancing the Fap project.

A significant aspect contributing to my enjoyment of the last four years have been the members of the Bharat lab, including but not limited to Dr. Abul Tarafder, Dr. Charlotte Melia, Andriko von Kügelgen, Dr. Nina Sulkowski, Dr. Charlotte Guyomar, Zo Ford, Suzi Letham, Matthew Herdman, Louis Elfari, Miles Graham, Dr. Kazune Tamura and Dr. Ido Caspy. I would also like to specifically thank Dr. Abul Tarafder and Andriko von Kügelgen for excellent technical and experimental support and advice over the years. I am deeply grateful for my parents, who have always been

loving and supportive; my brother, who has been invaluable in providing guidance and advice; and my friends and my housemates, who have continuously been making this journey such an enjoyable one. Finally, I am deeply grateful to and for my partner Heather, who has been keeping me going with her unconditional love, warmth, and support.

Abbreviations

AFM	Atomic force microscopy
Amp	Ampicillin
ART	Algebraic reconstruction technique
CD	Circular dichroism
c-di-GMP	Cyclic di-guanosine monophosphate
CUP	Chaperone-Usher Pathway
cryo-EM	Electron cryomicroscopy
cryo-ET	Electron cryotomography
CS	Compressed sensing
CS-TV^2	Second-order total variation compressed sensing
EMDB	Electron Microscopy Data Bank
EMPIAR	Electron Microscopy Public Image Archive
EPS	Extracellular polymeric substance
FIB	Focused ion beam
FSC	Fourier shell correlation
FFT	Fast Fourier transform
HBV	Hepatitis B virus
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
Ig	Immunoglobulin
IM	Inner membrane
IMAC	Immobilised metal ion chromatography
LB	Lysogeny broth
MD	Molecular dynamics
MRE	Molar residue ellipticity
MS	Mass spectrometry
MSA	Multiple sequence alignment
MSgg	Minimal salt glycerol glutamate
OD	Optical density
OM	Outer membrane
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PDE	Phosphodiesterase
px	Pixel
RMSD	Root mean square deviation
RT	Room temperature
SDS	Sodium dodecyl sulphate
SEM	Scanning electron microscope/microscopy
SIRT	Simultaneous iterative reconstruction technique
SPA	Single particle analysis
STA	Subtomogram averaging
TEM	Transmission electron microscope/microscopy
ThT	Thioflavin T
Tris	Tris (hydroxymethyl) aminomethane

TV^{2}	Second-order total variation
WBP	Weighted back-projection
WT	Wild-type
2D	Two-dimensional
3D	Three-dimensional

Units

In addition to SI units, the following units are used:

bp	Base pair
Da	Dalton
h	Hour(s)
kb	Kilobase
kDa	Kilodalton
min	Minute(s)
pH	$-\log^{10} c$, with $c = [H+]$ in mol/l
rcf	Relative centrifugal force
s	Seconds

Table of Contents

List of Figures

List of Tables

“Ye cannot live for yourselves;
a thousand fibres connect you
with your fellow-men […]”

Henry Melvill

1. Introduction

1.1 Bacterial biofilms

The evolution of multicellularity has been one of the major evolutionary transitions in life (Leigh Jr, 1995). While a seemingly unlikely event, it has become increasingly clear over recent years that multicellular organisms have evolved independently up to 25 times (Grosberg & Strathmann, 2007), underlining how a community lifestyle of cells may confer a significant survival advantage onto its members that allows them to thrive where single cells cannot. While multicellular life is often thought of as eukaryotic, multicellular communities embedded into an extracellular matrix exist no less in bacteria and archaea, where such assemblies are referred to as biofilms (Vert *et al.*, 2012). By all accounts, the biofilm lifestyle is an ancient one: Fossil records from more than 3 billion years ago confirm the presence of bacterial communities during those times (Hall-Stoodley *et al.*, 2004; Westall *et al.*, 2001), indicating that a biofilm lifestyle has aided them in persisting through the harsh environmental conditions of the early Earth. Indeed, the formation of biofilm communities was found to be a key characteristic of some of the most ancient 'living fossil' bacteria within the phylogenetic tree of life (Jahnke *et al.*, 2001; Reysenbach *et al.*, 2000).

A biofilm lifestyle can provide numerous advantages, providing protection from extracellular stresses, including UV light, chemical stresses, and predators, and allows the accumulation and sharing of resources (Dragoš *et al.*, 2018; Elasri & Miller, 1999; Matz *et al.*, 2008; Olson *et al.*, 2002; Sharma *et al.*, 2019) (Figure 1.1). Cells within the biofilms undergo complex three-dimensional arrangements, enabling division of labour that is not possible in a single-celled lifestyle (Dragoš *et* *al.*, 2018). Thus, biofilms play a major role in bacterial ecology – indeed, they are thought to represent the majority of the bacterial biomass and the dominant form of bacterial life (Flemming & Wuertz, 2019).

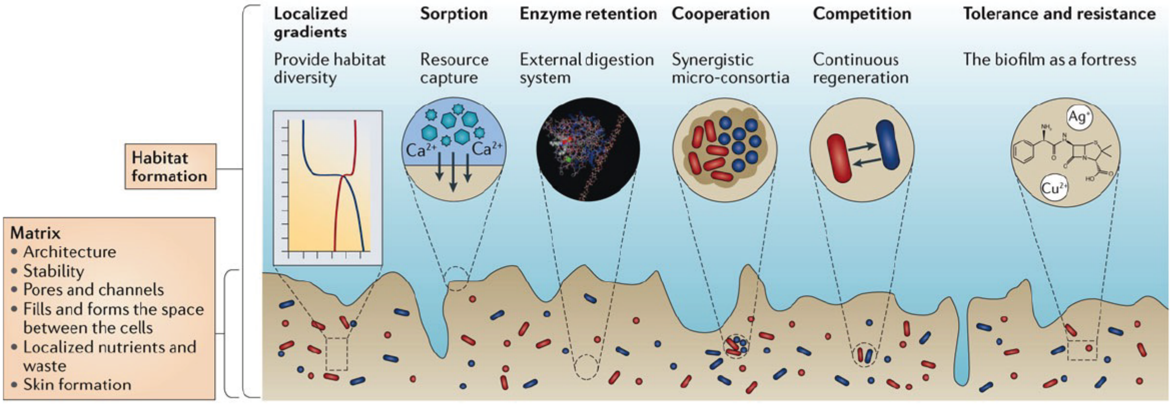


Figure 1.1: Emergent properties of biofilms. Bacteria in biofilms form their own microenvironments by secreting, and living in, a biofilm matrix made of extracellular polymeric substance (EPS). This results in several survival advantages for biofilm-embedded cells compared to planktonic cells that are termed 'emergent properties' (Flemming *et al.*, 2016). Such include localised gradients of cells that allow complex three-dimensional arrangements; capture and retention of nutrients and resources; extracellular digestion systems and retention of digestive enzymes; social interactions between cells, including cooperation and/or competition that may cause increased fitness; and increased survival against antimicrobial compounds, including antibiotics, through tolerance and resistance mechanisms. Adapted with permission from Springer Nature, *Nature Reviews Microbiology* from Flemming *et al.* (2016), ©2016 Springer Nature.

1.2 The biofilm lifecycle

Bacteria can live in two distinct states - as single planktonic cells, or as multi-cellular biofilms, embedded into extracellular polymeric substance (EPS) (Chua *et al.*, 2014). Which mode is employed at a certain place and time heavily depends on environmental conditions and varies significantly between different species of bacteria and growth conditions (O'Toole *et al.*, 2000a). A commonly cited general framework of biofilm formation is a five-step model (Figure 1.2) that was proposed based on the observation of *P. aeruginosa* under laboratory conditions (Sauer *et al.*, 2002; Sauer *et al.*, 2022). According to this model, the first step of biofilm formation is attachment of free-floating, planktonic cells to a surface. This first step is reversible, and cells may detach from the surface to return to a planktonic state. In a second step, the attachment becomes irreversible, and cells begin to secrete an extracellular matrix. Two maturation steps follow, in which both the cell mass and EPS production increase, resulting in the formation of cell clusters (stage 3) and microcolonies (stage 4), respectively. Since environmental conditions can grow unsuitable and nutrients can deplete, biofilms can release cells back into the environment in order to not trap its members under deteriorating conditions. Such dispersion of planktonic cells from the biofilm into the environment constitutes the final step and closes the cycle, allowing those planktonic cells to form biofilms once new suitable conditions are found.

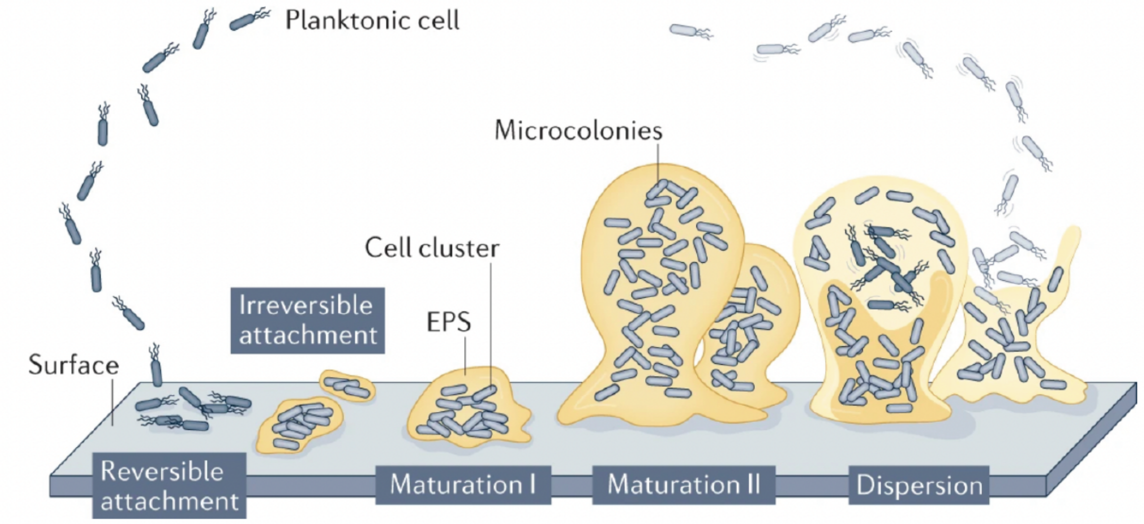


Figure 1.2: Five-step model of biofilm formation, based on observations made in *P. aeruginosa* biofilms under laboratory conditions. Biofilm formation begins by reversible attachment, during which cells attach through their flagellum or cell pole, followed by attachment of their longitudinal axis. Transition to irreversible attachment coincides through downregulation of flagella-encoding genes and upregulation of EPS components. Irreversible attachment correlates with increased susceptibility to antibiotic treatment. This is followed by creation of a cell cluster that is several cells thick, which then matures into a microcolony. Upon dispersion, matrix components are down-regulated and degraded, and susceptibility of dispersed cells increases markedly compared to biofilm cells. Figure adapted with permission from Springer Nature, *Nature Reviews Microbiology* from Sauer *et al.* (2022), ©2022 Springer Nature.

1.3 Regulation of biofilm formation

The major regulator of biofilm formation in numerous Gram-negative and Gram-positive species is cyclic di-guanylate (c-di-GMP) (Jenal *et al.*, 2017). The secondary messenger molecule c-di-GMP is synthesised from two GTP molecules by guanylate cyclases and degraded by phosphodiesterases. In *P. aeruginosa*, for example, more than 40 enzymes were found to be encoded that regulate c-di-GMP metabolism (Kulesekara *et al.*, 2006). Elevated c-di-GMP levels result in secretion of EPS and cell aggregation, caused by expression of adhesins and EPS components that promote biofilm maturation (Bordeleau *et al.*, 2015; Ross *et al.*, 1987; Serra *et al.*, 2013). Phosphodiesterase activity and reduction of c-di-GMP levels, contrarily, results in an induction of swarming motility (Kuchma *et al.*, 2007). Often, phosphodiesterases contain sensory domains that respond to environmental cues (Fazli *et al.*, 2014).

Under which conditions exactly biofilm formation is triggered varies from species to species: *P. aeruginosa* and *P. fluorescens*, for example, readily form biofilms under most conditions that enable growth (O'Toole & Kolter, 1998). Other species, such as *E. coli* K12 and *Vibrio cholerae*, do not form biofilms in minimal media unless specific nutrients are supplied (Pratt & Kolter, 1999; Watnick *et al.*, 1999). In addition, numerous other environmental factors influence biofilm formation, including but not limited to temperature, pH, osmotic pressure, flow rates, and oxygen availability (Ahn & Burne, 2007; Brading *et al.*, 1995; Chang *et al.*, 2015; Else *et al.*, 2003; Linnes *et al.*, 2013; Nostro *et al.*, 2012; O'Toole *et al.*, 2000b; O'Toole & Kolter, 1998; Percival *et al.*, 1999; Pratt & Kolter, 1999; Stanley, 1983; Stepanović *et al.*, 2003; Stoodley *et al.*, 1998).

1.4 Molecular composition of the biofilm matrix

Embedding into a matrix made of EPS is a defining feature of biofilm formation and responsible for transferring many of its protective effects upon bacteria (Stoodley *et al.*, 2002). The EPS acronym was initially used to refer to ‘extracellular polysaccharides’ that were thought to be the main matrix component (Allison & Sutherland, 1987; Sutherland, 2001); the acronym later obtained a new meaning upon discovering significant compositional complexity beyond its original namesake (Flemming & Wingender, 2010). While its composition differs from organism to organism, the EPS generally consists of a variety of polymeric molecules such as various proteins, including fibrous proteins such as pili (Vallet *et al.*, 2001) and flagella (Barken *et al.*, 2008), lipids (Mann & Wozniak, 2012), polysaccharides (Limoli *et al.*, 2015), and extracellular DNA (Buzzo *et al.*, 2021; Whitchurch *et al.*, 2002). This complex mixture of macromolecules not only provides structure and stability to the biofilm and promotes cohesion between cells, but can also modulate the physicochemical character of the biofilm, thus altering how bacteria collectively interact with their environment, i.e., through 'skin' formation by hydrophobic matrix components (Hobley *et al.*, 2013; Kobayashi & Iwano, 2012). The EPS matrix has been dubbed the 'dark matter of biofilms' (Flemming *et al.*, 2007) – on one hand because of its compositional complexity, on the other hand because of its inaccessibility to commonly used analysis methodology.

1.5 Archaic Chaperone-Usher Pathway pili

1.5.1 The EPS of *P. aeruginosa*

Biofilms formed by *P. aeruginosa* have come into particular focus due to their involvement in infectious disease, including lung infections that are the leading cause of mortality and morbidity in cystic fibrosis patients (Høiby *et al.*, 2010) and drug-resistant nosocomial infections (Obritsch *et al.*, 2005). Combined with its ready formation of biofilms under laboratory conditions, *P. aeruginosa* has been established as a model organism for biofilm formation in Gram-negative bacteria (O'Toole *et al.*, 2000a). In the infection context, its biofilms promote persistence and enable evasion of the immune system as well as tolerance to antibiotic treatment (Alhede *et al.*, 2014; Mah & O'Toole, 2001; Rybtke *et al.*, 2015). Indeed, biofilm formation has been shown to lower the susceptibility of bacterial cells to antibiotics by several orders of magnitude (Ceri *et al.*, 1999), and *P. aeruginosa* - along with several other biofilm-forming organisms - is among the most antibiotic-resistant bacteria (Tacconelli *et al.*, 2018).

The *P. aeruginosa* EPS is comparatively well-characterised, with the function of numerous polysaccharide and protein components described in the literature (Wei & Ma, 2013). *P. aeruginosa* employs at least three polysaccharides (Psl, Pel and alginate) that play a role in structural maintenance of the biofilm and have been described to play a crucial role in antibiotic resistance (Franklin *et al.*, 2011; Hentzer *et al.*, 2001; Wei & Ma, 2013). As commonly found in most biofilms, a further component of the EPS is extracellular DNA, which appears to be mostly chromosomal and may leak from dead cells within the biofilm yet adopts a structural role in maintaining biofilm stability and scaffolding (Buzzo *et al.*, 2021; Whitchurch *et al.*, 2002). The protein composition of *P. aeruginosa* biofilm EPS includes flagella (Bouteiller *et al.*, 2021), which are well-characterised for their role in swarming and swimming motility (O'Toole & Kolter, 1998) but may also adopt a structural role (Bouteiller *et al.*, 2021), and Type IV pili, which are crucial for twitching motility (Talà *et al.*, 2019) and may hence be important factors in the initial attachment and formation of microcolonies within the early biofilm (Barken *et al.*, 2008). Type IV pili specifically were found to also interact with DNA (Van Schaik *et al.*, 2005), and may function as a structural cross-link between matrix components. Soluble but membrane-tethered lectins were further found to bind Psl polysaccharides (Passos da Silva *et al.*, 2019), coordinating its location within the biofilm and providing another element of cross-linking between matrix components that provides stability to the matrix. The matchstick-like CdrA protein, which extends from the *P. aeruginosa* cell surface, also contains a lectin-like domain that interacts with Psl polysaccharides and provides further cohesion between matrix components (Melia *et al.*, 2021; Reichhardt *et al.*, 2018). Other proteins that have been described within the *P. aeruginosa* matrix are fibrous in nature and include Fap amyloid-like fibres that are proposed to be abundant in biofilms in *Pseudomonas* and fulfil a structuring role (Dueholm *et al.*, 2010; Zeng *et al.*, 2015); the bacteriophage Pf4 that is encoded as a prophage and highly upregulated within biofilms, forming liquid crystalline structures within the matrix that contribute to antibiotic tolerance, possibly by performing a diffusion barrier (Secor *et al.*, 2015; Tarafder *et al.*, 2020); and so-called Chaperone-Usher Pathway (CUP) pili, which are specialised adhesins with various attributions to biofilm cohesion and development (Giraud *et al.*, 2011; Mikkelsen *et al.*, 2009; Vallet *et al.*, 2004; Vallet *et al.*, 2001).

1.5.2 Assembly and structure of CUP pili

Chaperone-Usher Pathway pili are among the best-characterised adhesins in Gram-negative bacteria, being responsible for adhesion to both biotic and abiotic surfaces (Sauer *et al.*, 2004; Waksman & Hultgren, 2009). Among the best-characterised CUP pili systems are the P pilus (Kuehn *et al.*, 1992) and Type I pilus (Schilling *et al.*, 2001) systems in *E. coli*, which are both associated with uropathogenic phenotypes by providing adhesion to host cell receptors, which enables bacteria to persist in the urinary tract (Kau *et al.*, 2005). CUP systems are usually encoded within single operons that are readily transferred horizontally between bacterial species (Nuccio & Bäumler, 2007), and consist of three main components: A main pilus-forming subunit (pilin) that is the major protein component of the pilus; a chaperone that stabilises the pilin prior to assembly within the periplasm; and an outer membrane usher protein which both plays an important role in pilus assembly as well as provides the translocation pore from which the pilus extends (Nuccio & Bäumler, 2007; Waksman & Hultgren, 2009). A key principle of pilus assembly and stability is the fact that each pilin subunit has an immunoglobulin- (Ig-)like fold that consists of a β-sandwich, but lacks the final, antiparallel β-strand that would complete the fold. Instead, each subunit extends an N-terminal so-called donor strand that completes the fold of the following subunit along the fibre (Barnhart *et al.*, 2000; Choudhury *et al.*, 1999; Hospenthal *et al.*, 2016; Remaut *et al.*, 2006) (Figure 1.3). Prior to assembly, a chaperone provides the missing β-strand to the subunit, playing an important role in assembling the filament at the usher pore (Barnhart *et al.*, 2000). Interactions between the chaperone and the usher's N- and C-terminal domains are proposed (Remaut *et al.*, 2008), although a molecular model of interaction is yet to be suggested. In addition to these three core components that exist in all CUP systems (Nuccio & Bäumler, 2007), several other minor pilin subunits are often encoded. These include adhesin subunits that form the tip of the pilus (Hultgren *et al.*, 1989; Jones *et al.*, 1995; Pakharukova *et al.*, 2018), separate structural elements (Kuehn *et al.*, 1992), or termination subunits that end the pilus elongation reaction (Båga *et al.*, 1987). The subunit architecture for Type I and P pili is shown in Figure 1.3; the main pili-forming subunits (PapA/FimA) in these classical CUP pili arrange into a distinct rod-like shape that can uncoil to provide further reach and elasticity to the pilus (Hospenthal *et al.*, 2016; Hospenthal *et al.*, 2017).

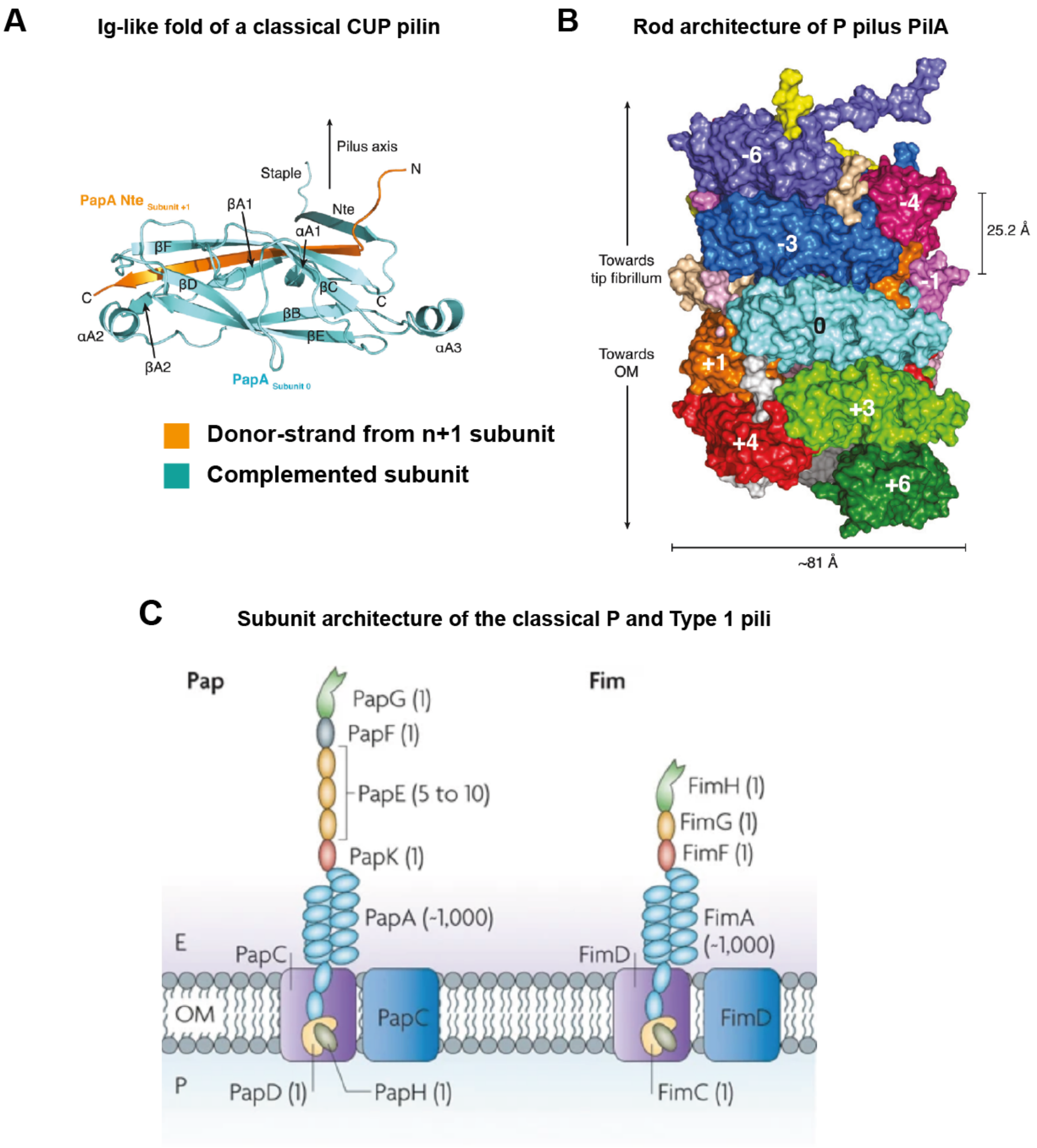
1.5.3 Phylogenetic classification of CUP pili

CUP pili can be phylogenetically divided into three main classes: classical, archaic, and alternate CUP pili (Nuccio & Bäumler, 2007). Previous research has heavily focused on the P and type I pilus systems in *E. coli*, which both belong to the classical type of CUP pili. Little structural information is known about alternate and archaic-type pili. While classical and alternate CUP pili only exist in beta- and gammaproteobacteria, archaic CUP systems were found in alpha- and deltaproteobacteria, and even in members of cyanobacteria and *Deinococcus*

(Nuccio & Bäumler, 2007), suggesting they may be the most ancient type of CUP pili from which other types originate. Of the five CUP systems described in *P. aeruginosa,* four are classical-type CUP systems (CupA-D), and one (CupE) represents an archaic-type system (Giraud & de Bentzmann, 2012).

1.5.4 Classical CUP systems in *P. aeruginosa*

The role of CupA was the first to be discovered through a screen for factors inhibiting early biofilm formation on abiotic surfaces (Vallet *et al.*, 2001), controlled by the master regulator protein MvaT (Vallet *et al.*, 2004). Two similar CUP systems were detected within the same study (*cupB* and *cupC*); later studies found they are controlled by a different two-component system termed 'regulator of *cup'* (Roc) (Kulasekara *et al.*, 2005) which triggers expression of *cupB* and *cupC* along with microcolony formation during biofilm development. These triggers, which appear to differ from *cupA*, are complex and not fully understood (Giraud & de Bentzmann, 2012). Interestingly, *cupB* and *cupC* seem to cooperate and are not interchangeable (Kulasekara *et al.*, 2005) within this process. The fourth CUP operon in *P. aeruginosa*, *cupD,* is present in only a few strains of *P. aeruginosa* such as PA14, where it is located on the PAPI-I pathogenicity island, which may be partially responsible for the increased pathogenicity of this strain (Mikkelsen *et al.*, 2009). Expression of *cupD* also correlates with microcolony and biofilm formation and a sessile lifestyle, and the overall operon structure suggests similarity with and redundance to *cupA* (Giraud & de Bentzmann, 2012); indeed, bioinformatic analysis confirms that *cupA* and *cupD* rarely co-exist within a given *P. aeruginosa* strain (Böhning *et al.*, 2022b).



A Ig-like fold of a classical CUP pilin

B Rod architecture of P pilus PilA

C Subunit architecture of the classical P and Type 1 pili

Figure 1.3: Structural features of classical CUP pili in *E. coli*. A) Ribbon depiction of the main pilin of the P pilus, PapA. The incomplete Ig-like fold is complemented by an N-terminal donor strand of the n+1 subunit (orange). Adapted from Hospenthal *et al.* (2016). B) Rod architecture of the P pilus, showing numerous inter-subunit contacts between subunits of the main pilin PilA. Adapted from Hospenthal *et al.* (2016). C) Pilus architecture of the P and Type 1 pilus. Adapted with permission from Springer Nature, *Nature Reviews Microbiology* from Waksman & Hultgren (2009), ©2009 Springer Nature.

1.5.5 The archaic CupE operon in *P. aeruginosa*

The *cupA-D* operons encode classical CUP systems (Giraud & de Bentzmann, 2012) and are hence more closely related to the well-characterised rod-like P and type I pili in *E. coli*. A fifth characterised *cup* operon, *cupE*, belongs to the evolutionarily divergent σ-clade, representing an archaic type of CUP pilus (Giraud *et al.*, 2011; Nuccio & Bäumler, 2007). Based on previous data, archaic CUP pili are expected to show a thinner, more hair-like architecture that is unknown at the molecular level (Tomaras *et al.*, 2003). CupE is regulated by the two-component system PprA-PprB (Giraud *et al.*, 2011), activation of which is associated with an adhesive signature corresponding to a hyper-biofilm phenotype, including expression of CupE pili, type IVb pili, CupE fimbriae, BapA adhesin, and eDNA, without involvement of exopolysaccharides (De Bentzmann *et al.*, 2012). CupE was found to be involved in both micro- and macrocolony formation, but also played a crucial role in late-stage biofilms by maintaining the three-dimensional mushroom-like shape of the *P. aeruginosa* biofilm (Giraud *et al.*, 2011). The CupE operon consists of six proteins: the main fibre-forming protein CupE1; two structurally related minor fibre components CupE2 and CupE3; a chaperone CupE4; an Usher protein CupE5; and a putative adhesin subunit CupE6 (Giraud *et al.*, 2011; Nuccio & Bäumler, 2007) (Figure 1.4).

The *cup* operon closely resembles the *csu* operon in *Acinetobacter baumannii* both in architecture and in sequence (Giraud *et al.*, 2011), suggesting it may have been acquired through horizontal gene transfer. The *csu* operon in *A. baumannii* is well-characterised and a major factor in *A. baumannii* biofilm formation (Tomaras *et al.*, 2003), and closely correlates with pathogenicity (Colquhoun & Rather, 2020). Indeed, in a study evaluating 32 isolates from patients with multi-drug resistant *A.*

baumannii biofilm infections, the most prevalently expressed gene found in 100% of clinical isolates was *csuE* encoding for the Csu pilus tip adhesin subunit (Zeighami *et al.*, 2019). Several proteins of the *csu* operon have been structurally characterised: The structure of the main pilin CsuA/B in complex with its chaperone CsuC was solved using X-ray diffraction, showing that chaperone-bound CsuA/B subunits adopt a substantially disordered structure prior to assembly (Pakharukova *et al.*, 2015). In a later study, an X-ray structure of the tip domain CsuE could also be solved, revealing that it allows the pilus to interact with its environment through hydrophobic fingers on its N-terminal domain (Pakharukova *et al.*, 2018), which are suggested to enable bacterial persistence on plastic surfaces within hospital environments and thus contribute to hospital-acquired infections. Indeed, inhibition of CsuE through an antibody was found to significantly decrease the ability of *A. baumannii* to form biofilms (Pakharukova *et al.*, 2018). Csu pili were also found, in combination with other pilins, to induce a protective response against *A. baumannii* infection and may thus be suitable as vaccines (Ramezanalizadeh *et al.*, 2020), underlining the importance of understanding CUP pili-mediated biofilm formation. Structural information on intact archaic CUP pili, however, is scarce.

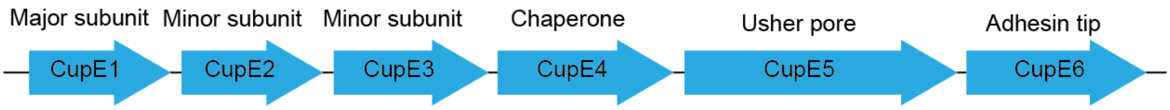


Figure 1.4: Organisation of the *cupE* operon (PA4648-4653). The operon encodes three homologous pilin subunits, of which CupE was determined to be the major subunit (Giraud *et al.*, 2011). A chaperone (CupE4) is predicted to stabilise the pilins (CupE1-3) prior to assembly and shuttle them to the usher pore (CupE5), through which subunits are secreted through the OM. A specialised subunit is found at the fibre tip (CupE6).

1.5.6 Project aims

While archaic CUP pili are phylogenetically wide-spread and important for biofilm formation, no fibre structure of any archaic CUP pilus exists. The aim of this project was to structurally analyse the CupE pilus in *P. aeruginosa*, a model biofilm-forming organism. Therefore, I built on previous work in the Bharat laboratory by Nina Sulkowski (Sulkowski, 2021), during which the expression and isolation of CupE was established and initial cryo-EM characterisation was conducted. The goal of this study was to obtain an atomic model of the CupE pilus by cryo-EM. Additionally, electron cryotomography was performed to image the CupE pilus in a cellular environment, elucidating the structural properties of the pilus as it interacts with cells.

1.6 Functional amyloid in *Pseudomonas*

1.6.1 Biophysical characteristics of amyloids

Amyloids are a specialised type of protein fibre where constituent subunits form β-sheets that are perpendicular arranged to the fibre axis (Eisenberg & Jucker, 2012). This architecture means that each subunit interacts with its preceding and following subunit through extensive β-sheet interactions, and provides amyloids with unique properties: Given their highly ordered nature, amyloids are extraordinarily stable, including against protease and detergent action (Fowler *et al.*, 2007). Amyloid fibres can be stained with the dye Congo Red, which leads to green birefringence when examined under cross-polarised light (Sipe *et al.*, 2010), and with the fluorescent dye Thioflavin T (Khurana *et al.*, 2005). While such secondary hallmarks have long been employed to identify amyloids, more modern definitions assign amyloid character through a more direct structural method, X-ray diffraction (Eisenberg & Jucker, 2012): Amyloid fibres result in distinct diffraction at ~4.75 Å, corresponding to the distance between the stacked β-sheet making up the amyloid core.

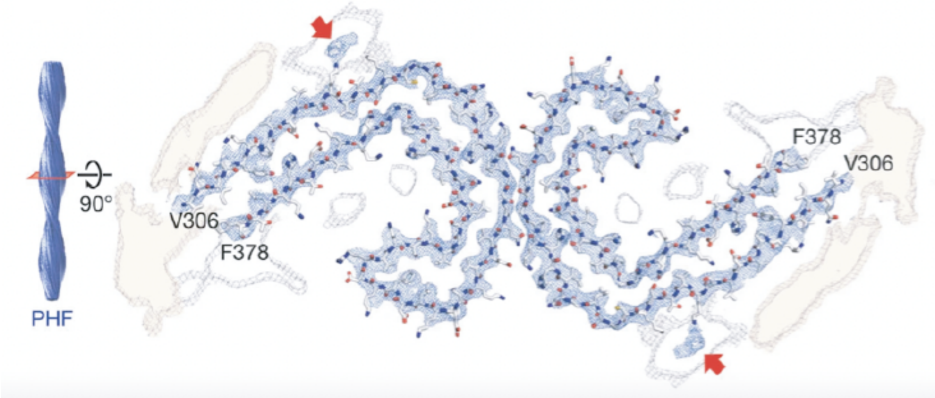


Figure 1.5: Example of previous studies on amyloid structure via cryo-EM. Cross-section of cryo-EM structure of Tau filaments isolated from patients' brains at 3.4 Å resolution. Side chains are clearly visible. A schematic on the left shows the overall architecture of the paired helical filament (PHF), which shows the so-called super-twist of the amyloid protofilament. Adapted with permission from Springer Nature, *Nature* from Fitzpatrick *et al.* (2017), ©2017 Springer Nature.

1.6.2 Disease-related and functional amyloids in eukaryotes

Amyloids are highly studied assemblies because of their involvement in neurodegenerative diseases such as Alzheimer's disease or Parkinson's disease, where amyloid aggregates within patient neurons were found to correlate with disease progression (Hardy, 1997). It is thought that such amyloid aggregates represent protein that misfolds but cannot be appropriately cleared within the cell, ultimately leading to cell death (Chiti & Dobson, 2006). Indeed, a given amyloid protein can adopt a wide variety of different amyloid confirmations (Scheres *et al.*, 2020; Zhang *et al.*, 2019), prompting the hypothesis that formation of insoluble amyloid aggregates is accidental in nature. Given the difficulties that cells face in disassembling amyloid aggregates, which are often protease-resistant, researcher believed for several decades that all amyloids in nature are assembled by accident. In recent years, however, it became increasingly clear that amyloid formation can occur on purpose, and indeed is employed by both bacteria and eukaryotes (Fowler *et al.*, 2007; Perrett *et al.*, 2014). For example, in mammals, amyloid fibres formed by the mature form of the protein Pmel17 appear to scaffold the biosynthesis of melanin (Berson *et al.*, 2001). In *Saccharomyces cerevisiae*, the translation termination factor Sup35p can undergo reversible aggregation into amyloids (King *et al.*, 1997). Since Sup35p prevents read-through translation at stop codons, its aggregation – and hence inactivation - results in an altered proteome consisting of proteins with C-terminal extension, which results in phenotypic diversity that can then be inherited by daughter cells in an epigenetic-like, non-Mendelian manner (True & Lindquist, 2000).

1.6.3 Amyloid fibres in bacteria

Amyloid fibres are widespread in bacteria, where they predominantly function as a structuring agent (Figure 1.6) (Fowler *et al.*, 2007; Taglialegna *et al.*, 2016). Amyloid-like fibres may be particularly suitable for such a task given their extraordinarily high strength, which has previously been compared to steel but maintains flexibility (Smith *et al.*, 2006). While the exact cause of amyloid assembly in neurodegenerative disease is shrouded in mystery and appears to depend on a variety of biochemical and environmental factors (Lövestam *et al.*, 2022), numerous bacteria have dedicated assembly machineries for the production, secretion and assembly of amyloid-like fibres that often form an abundant structural component within biofilms (Taglialegna *et al.*, 2016). The first discovered amyloid fibres in bacteria are so-called curli fibres in *E. coli* (Chapman *et al.*, 2002), which are an abundant structural component of *E. coli* biofilms and represent one of the major classes of cell surface fimbriae (Nuccio & Bäumler, 2007). Consistent with human amyloids, curli fibres bind to the amyloid-binding dye Congo Red, inducing a spectral shift, and further bind the fluorescent dye Thioflavin T, resulting in fluorescence emission at ~480 nm, consistent with human amyloid proteins (Chapman *et al.*, 2002). Beyond the role of curli fibres in forming a structural scaffold, they also modulate adhesion to biotic and abiotic surfaces, presumably by modulating the biophysical properties of the biofilm (Boyer *et al.*, 2007; Gophna *et al.*, 2002; Goulter-Thorsen *et al.*, 2011; Pawar *et al.*, 2005; Saldaña *et al.*, 2009). On a genetic level, curli fibres are encoded within two operons, *csgABC* and *csgDEFG*. The *csgABC* operon encodes the main fibre-forming protein CsgA and the minor subunit CsgB, which was found to act as a nucleator of CsgA assembly into fibres (Bian & Normark, 1997; Hammer *et al.*, 2007). The periplasmic CsgC is an accessory factor regulating amyloid subunit stability and assembly (Taylor *et al.*, 2011). The *csgDEFG* operon encodes accessory factors for fibre assembly. CsgD is a master regulator that influences transcription of the *csgABC* operon as well as cellulose, another component of the biofilm (Brombacher *et al.*, 2006). Its regulation is complex and depends on a range of factors, including temperature, osmolarity

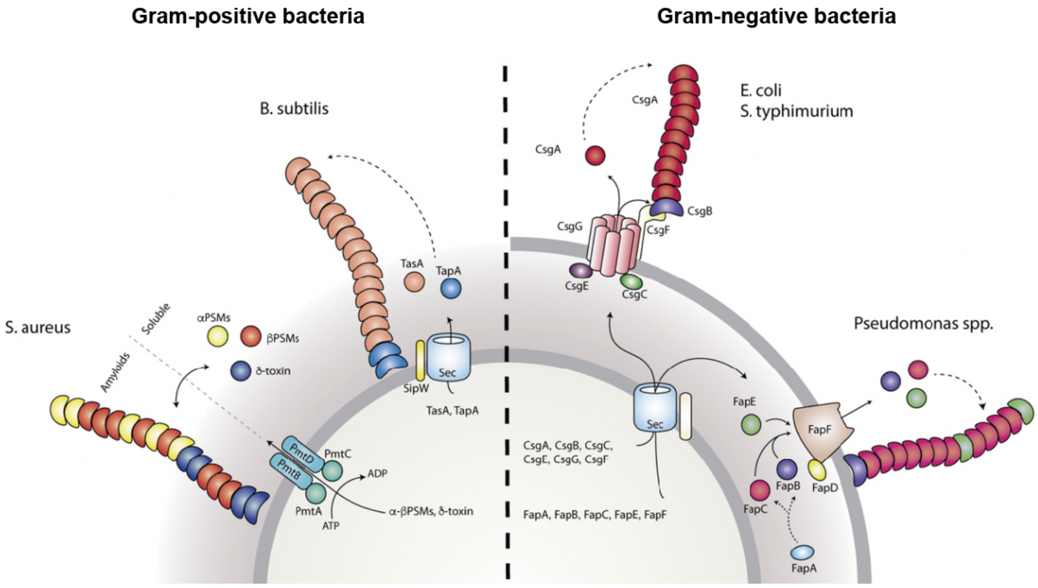


Figure 1.6: Proposed amyloid-like systems functioning as biofilm scaffolds in Gram-positive and Gram-negative bacteria. Left: Gram-positive systems include the TasA/TapA system in *B. subtilis*, where the main fibre-forming subunit TasA is exported through the Sec complex, cleaved by the signal peptidase SipW, and assembled and possibly anchored by the accessory protein TapA through an unknown mechanism. In *S. aureus* biofilms, toxins termed 'phenol-soluble modulins' (PSM), a family that also includes δ-toxin, are thought to assemble into amyloid-like fibres. Right: Gram-negative systems. The curli system in *E. coli* and *Salmonella typhimurium* encodes a main fibre-forming subunit, CsgA, that is secreted through the outer membrane pore CsgG and nucleated on the cell surface by CsgB. Accessory proteins CsgC and CsgE aid export through CsgG. CsgF assists nucleation of fibres. The Fap system in *Pseudomonas* encodes a main fibre-forming component, FapC, that is secreted through the outer membrane pore FapF, nucleated by FapB and containing FapE as a minor component. FapA is a regulator of FapB and FapC assembly. FapD is predicted to act as a signal protease. Components of both described systems are secreted through the Sec channel into the periplasm. Adapted with permission by the American Society for Microbiology from Taglialegna *et al.* (2016), ©2016 American Society for Microbiology.

and pH (Prigent-Combaret *et al.*, 2001). CsgG is a nonameric translocation pore (Goyal *et al.*, 2014) that is thought to function through a diffusion-based, entropy-driven mechanism and has been famously engineered for so-called nanopore sequencing (Van der Verren *et al.*, 2020), enabled through the fact that transport through CsgG does not require an ion gradient or an energy source (Goyal *et al.*, 2014). CsgE and CsgF are accessory factors associated with the secretion pore CsgG, enabling correct amyloid assembly, and preventing aggregation of the amyloid subunits (Goyal *et al.*, 2014; Nenninger *et al.*, 2011).

After the discovery of the amyloid character of curli fibres (Chapman *et al.*, 2002), numerous other so-called functional amyloids were detected in various other bacteria and indeed in many of the most well-studied bacterial organisms. Among the most well-studied examples are phenol-soluble modulins (Schwartz *et al.*, 2012) and Bap (Cucarella *et al.*, 2001; Cucarella *et al.*, 2004) in *Staphylococcus aureus*, Fap fibres in *Pseudomonas* spp. (Dueholm *et al.*, 2010), and TasA fibres in *B. subtilis* (Romero *et al.*, 2010). The latter two are the particular subject of this thesis and will be further described in the following sections.

1.6.4 Fap fibres in *Pseudomonas* biofilms

Functional amyloid in *Pseudomonas* (Fap) was first discovered in the *P. fluorescens* group strain UK4 during a search for amyloid-like fibres in environmental strains (Dueholm *et al.*, 2010). Subsequently isolated and heterologously expressed in *E. coli*, the *fap* operon leads to characteristic aggregation of cells into a biofilm-like phenotype (Dueholm *et al.*, 2010). It was later shown that *fap* deletions can be complemented with *fap* overexpression plasmids, leading to increased aggregation (Dueholm *et al.*, 2013b), and that the expression of the *fap* operon greatly increases biofilm stiffness and hydrophobicity, allowing it to withstand stronger mechanical forces and modulating interactions with the environment (Zeng *et al.*, 2015). The presence of *fap* operons was indeed also detected in various other *Pseudomonas* strains, including *P. putida* and the important pathogen *P. aeruginosa* (Dueholm *et al.*, 2013b). Fap fibres can be isolated from *Pseudomonas* strains by employing extraordinarily harsh extraction conditions: After enzymatic and detergent-enabled cell lysis, lysates are treated with – and boiled in – 2% SDS, which solubilises all cellular components except for Fap fibres, which remain stable in the procedure (Dueholm *et al.*, 2013b), showcasing the remarkable biochemical features and high stability of these amyloid-like fibres. The fibres can then be collected by centrifugation, as they remain insoluble throughout.

Compared to curli in *E. coli*, Fap fibres are expressed in a single operon, *fapABCDEF* (Dueholm *et al.*, 2010; Dueholm *et al.*, 2013b). Fap fibres consist of a

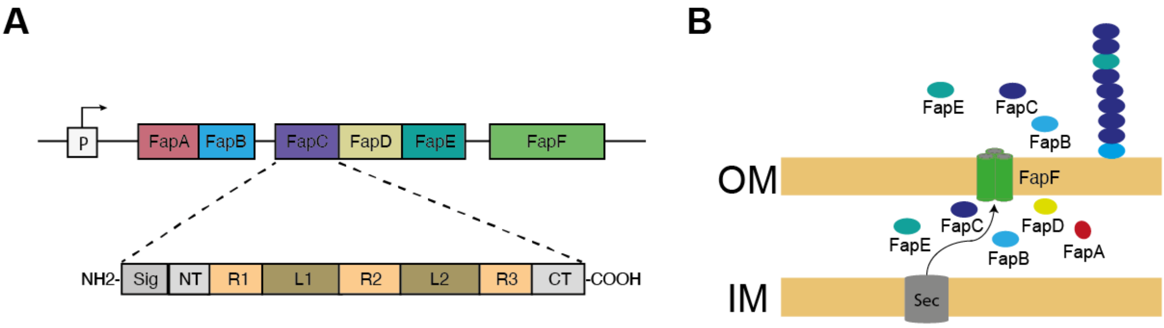


Figure 1.7: Organisation and function of the *fap* operon. A) Organisation and components of the *fap* operon, and predicted organisation of the main fibre-forming component FapC (Dueholm *et al.*, 2010). FapC consists of a signal peptide (Sig), an N- and C-terminal region (NT and CT), three repeat regions (R1-R3) and two linker regions (L1-2). B) Localisation of Fap proteins in the periplasm and on the cell surface. FapA and FapD are localised in the periplasm, potentially having chaperone (FapA) and proteolytic functions (FapD). FapC, FapB and FapE are translocated through a specialised translocation pore (FapF) to the extracellular space, where they assemble into fibres predominantly consisting of FapC, with minor amounts of FapB and FapE present. Proteins are shown in the context of their location relative to the outer membrane (OM) and inner membrane (IM).

main fibre-forming component (FapC), accompanied by a sequence-related but smaller protein thought to act as a nucleator of fibre assembly (FapB) (Dueholm *et al.*, 2010). After secretion into the periplasmic space by the Sec complex, FapC is thought to be proteolytically processed by the protease FapD (Dueholm *et al.*, 2010), and then secreted through a trimeric secretion pore termed FapF (Rouse *et al.*, 2017), aided by the FapE accessory protein, which is thought to be incorporated into fibres in sub-stoichiometric amounts (Dueholm *et al.*, 2013b). The role of FapA is unknown but assumed to be regulatory; it constitutes the only component whose deletion does not abolish the structural function of the operon or causes loss of amyloid protein (Zeng *et al.*, 2015), and loss of FapA is indeed relatively frequent across species (Dueholm *et al.*, 2013a). Interestingly, deletion of *fapA* causes fibres to predominantly consist of FapB instead of FapC (Dueholm *et al.*, 2013b), suggesting FapA may influence subunit composition. Consistent with previous approaches targeting biofilm formation as a means to combat infection (Chen *et al.*, 2013), targeting FapC amyloid assembly has been furthermore proposed for the treatment of bacterial biofilm infections, and nanosilver-based therapeutics have been shown to be able to inhibit biofilm formation by specifically targeting Fap amyloid (Huma *et al.*, 2020). Such treatments build on previous successes with battling infection in mouse models with compounds targeting curli fibres in *E. coli*, termed ‘curlicides’ (Cegelski *et al.*, 2009).

Early bioinformatic analysis (Dueholm *et al.*, 2010), combined with structural modelling (Rouse *et al.*, 2018) and deletion analysis (Rasmussen *et al.*, 2019) shows that FapC consists of three repeat regions (R1-3) separated by two linker regions (L1-2). Within these repeat regions, there is a significant conservation of serine, glutamine and asparagine residues that are connected by flexible alanine and glycine linkers (Dueholm *et al.*, 2013b; Rasmussen *et al.*, 2019). While not identical, the repeat regions are heavily sequence-related and deletion of one or two delete regions can still result in amyloid formation. Interestingly, even deletion of all three repeat regions can result into the formation of some fibres, although they easily disassemble (Rasmussen *et al.*, 2019). Modelling of FapC structure suggests that it may be arranged similar to a β-helix, with each repeat region constituting a strand-turn-strand motif, so that a single FapC subunit measures three turns, spanning 3 x 4.75 Å = 14.25 Å along the fibre axis. It is predicted that FapC subunits would then stack laterally, i.e., the repeat region R3 of one subunit would interact with the repeat region R1 of the following subunit. This arrangement – i.e., that FapC subunits are repeated every 14.25 Å along the fibre axis - would somewhat stand in contrast to amyloids in human neurodegenerative disease, where a single subunit is repeated every 4.75 Å along the fibre axis, but is plausible as repeat regions are highly similar to each other, with each FapC subunit emulating a single classical amyloidogenic unit repeated three times.

While the amyloid-like core appears conserved between *Pseudomonas* species (Dueholm *et al.*, 2013b), the length of the repeat region remains constant in *Pseudomonas* spp., the length of the linker region L2 can vary significantly: While L2 is 36 residues in the *P. fluorescens* group UK4 strain, it is 122 residues in *P. aeruginosa*, and even 275 residues in *P. entomophila* L48. The exact function of this linker region is poorly understood; it is unclear whether the linker regions or the amyloid core itself influence how Fap fibres interact with each other to form the large aggregates isolated from biofilms (Dueholm *et al.*, 2010), and how fibres convey stability to the biofilm and modulate its biophysical properties (Zeng *et al.*, 2015).

1.6.5 Project aims

While numerous functional amyloid fibres have been described to scaffold biofilm formation in bacteria, there is little direct structural knowledge on subunit arrangement within bacterial functional amyloids. The aim of this project was to produce cryo-EM samples of functional amyloid in *Pseudomonas* (Fap) by recombinant expression of the main fibre-forming component FapC in *E. coli* and subsequent reconstitution into fibres. After cryo-EM data acquisition, insights into subunit arrangement of FapC within Fap fibres and its structure were thus obtained.

1.7 Molecular architecture of the biofilm scaffold protein TasA

1.7.1 Biofilms of the Gram-positive bacterium *B. subtilis*

One of the main model organisms for biofilm formation in Gram-positive bacteria is the hay bacillus, *B. subtilis* (Azulay *et al.*, 2022; Branda *et al.*, 2006; Earl *et al.*, 2008; Kearns *et al.*, 2005; Serrano *et al.*, 1999; Vlamakis *et al.*, 2013). Compared to the opportunistic pathogen *P. aeruginosa*, *B. subtilis* is non-pathogenic, and its biofilms are most commonly known for colonising plant roots (Earl *et al.*, 2008). Still, closely related species such as *Bacillus cereus* group are well-studied as food pathogens and include the causative agent of anthrax, *B. anthracis* (Ehling-Schulz *et al.*, 2019). Biofilms of *B. subtilis* have a particularly characteristic morphology that includes the formation of complex three-dimensional structures termed wrinkles (Branda *et al.*, 2001), which appear to be a result of EPS formation and localised cell death within the biofilm (Asally *et al.*, 2012).

Compared to the complex composition of the biofilm matrix of *P. aeruginosa*, there are comparatively few matrix components in *B. subtilis*, although it is not clear whether this is the result of a simpler organisation of *B. subtilis* biofilms, or whether *B. subtilis* EPS is just less studied than Gram-negative biofilm model organisms. The matrix predominantly consists of three components, each of which was shown to be essential for biofilm formation: EPS polysaccharides encoded by a fifteen-protein operon (Branda *et al.*, 2001); a fibrous protein termed TasA that is abundant in the matrix (Branda *et al.*, 2006; Romero *et al.*, 2010); and the BslA surface protein which forms a hydrophobic layer on the biofilm (Hobley *et al.*, 2013; Kobayashi & Iwano, 2012). Deletion of any of these components results in defective biofilm formation on agar plates or on the air-water surface (‘pellicle biofilms’) (Branda *et al.*, 2006; Ostrowski *et al.*, 2011). Interestingly, Δ*eps* strains can be combined with

Δ*tasA* strains to result in fully functional biofilm formation (Branda *et al.*, 2006); equally, biofilm formation of Δ*bslA* strains can be restored by combination with Δ*eps* Δ*tasA* strains (Ostrowski *et al.*, 2011). Cells within the biofilm may perform division of labour to produce either TasA or eps (Dragoš *et al.*, 2018).

1.7.2 TasA in *B. subtilis* biofilms

TasA was originally discovered as a spore coat protein with antimicrobial properties and was accordingly named as a 'translocation-dependent antimicrobial spore component'. Later studies discovered that TasA makes up a significant amount of the *B. subtilis* biofilm matrix, and that its deletion results in defective biofilm formation (Branda *et al.*, 2006). It was found that fibres formed by TasA are the functional form of TasA within biofilms (Romero *et al.*, 2010).

TasA is encoded as part of the three-component *tapA*-*sipW*-*tasA* operon under the control of the biofilm master regulator SinR (Kearns *et al.*, 2005). The TasA operon encodes, apart from the fibre-forming protein TasA itself, two accessory proteins that are essential for TasA fibre assembly. The first, TapA (named as TasA-associated protein (Romero *et al.*, 2011)), is also secreted into the extracellular matrix, where it appears to form foci that are associated with the cell wall through an unknown mechanism that are thought to anchor TasA fibres to the cell wall (Romero *et al.*, 2011). Deletion of TapA results in biofilm defects with a phenotype similar to TasA deletion *in vivo* (Romero *et al.*, 2011). A further study has shown that TapA accelerates the assembly of TasA fibres *in vitro* (Romero *et al.*, 2014), suggesting a role as a nucleator similar to CsgB (Bian & Normark, 1997). The second accessory protein, SipW, is a membrane-associated signal peptidase that cleaves the N-terminal signal peptides after secretion through the Sec-complex. In addition to protease activity, SipW was furthermore proposed to have a secondary role by stimulating eps expression by an unknown mechanism (Terra *et al.*, 2012).

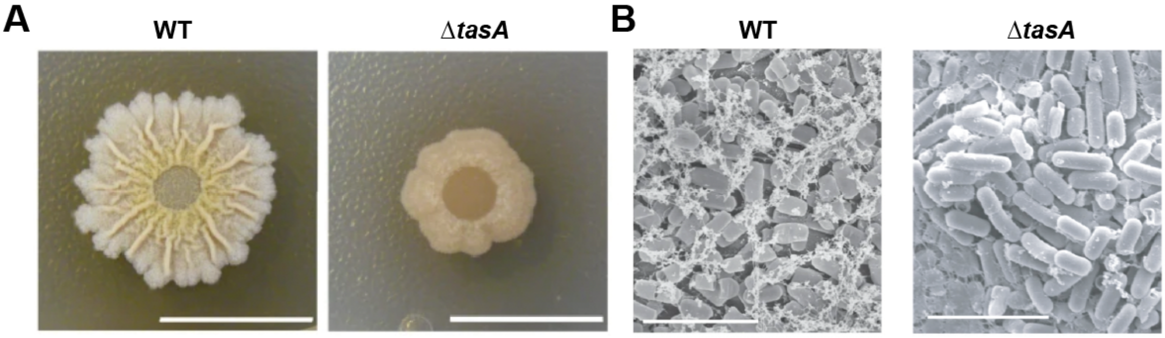


Figure 1.8: TasA is an integral part of *B. subtilis* biofilms. A) WT versus Δ*tasA* *B. subtilis* biofilm on agar plates. Scale bars are 1 cm. B) WT versus Δ*tasA* biofilms in the melon phylloplane imaged using SEM, indicating that *tasA* expression correlates with the presence of large fibre aggregates. Scale bars are 25 μm. Adapted from Cámara-Almirón *et al.* (2020).

1.7.3 The molecular architecture of TasA

TasA itself is a 26 kDa protein that – similar to *E. coli* curli and *Pseudomonas* Fap – has been suggested to form amyloid fibres that scaffold the biofilm (Romero *et al.*, 2010). This assessment is based on biophysical characterisation of TasA fibres, including positive Congo Red staining of *B. subtilis* biofilms that is absent in Δ*tasA* strains (Romero *et al.*, 2010) and differential Congo Red absorbance of purified fibres (Romero *et al.*, 2010), increased Thioflavin T fluorescence upon polymerisation of TasA into fibres (Romero *et al.*, 2010), harsh conditions (10% formic acid treatment) required for depolymerisation of fibres (Romero *et al.*, 2010), binding to the amyloid-specific A11 antibody (Romero *et al.*, 2010), a diffraction pattern interpreted as consistent with cross-β arrangement (Diehl *et al.*, 2018), and protease resistance (Diehl *et al.*, 2018). Compared to curli and Fap fibres, however, the monomeric form of the fibre-forming protein TasA adopts a folded state, and a crystal structure could be solved, revealing a Jellyroll fold with a β-sandwich core and some α-helical content (Diehl *et al.*, 2018) (Figure 1.9 A). The N- and C-termini were further found to be disordered in the monomeric state by NMR studies (Diehl *et al.*, 2018). The same study that solved the monomer structure, however, still suggests transitioning of the monomer to an amyloid fibre in the presence of EPS components (Diehl *et al.*, 2018) (Figure 1.9 B). Indeed, TasA has been widely cited as a functional amyloid in literature (Balistreri *et al.*, 2020; Otzen & Riek, 2019; Perrett *et al.*, 2014; Taglialegna *et al.*, 2016).

A notable dissenting study has concluded that, while TasA does form fibres, they are not amyloid, i.e., do not have a cross-β architecture (Erskine *et al.*, 2018). This is because the authors found that Congo Red staining did not depend on TasA

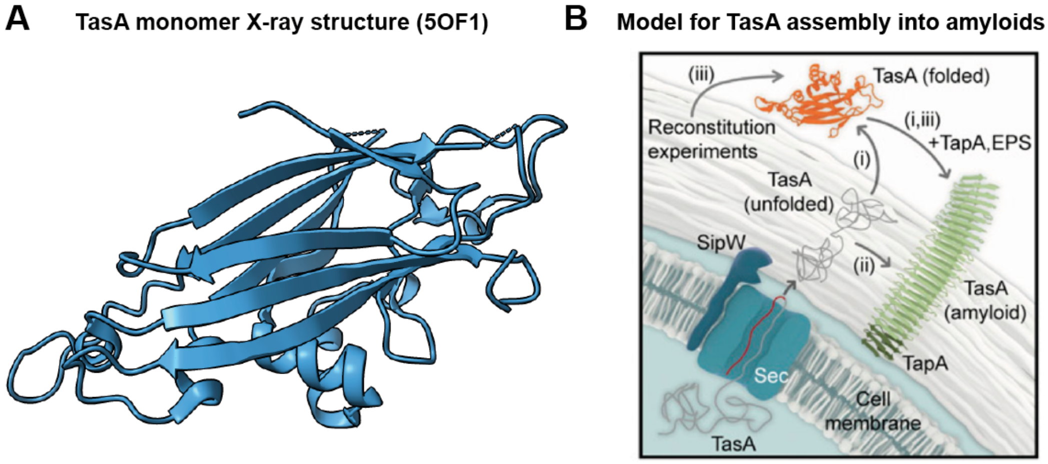


Figure 1.9: TasA is thought to transition from a folded, soluble monomer to an amyloid fibre in the biofilm matrix. A) Monomeric X-ray diffraction structure of TasA (PDB 50F1) , adopting a β-sandwich Jellyroll-fold. B) TasA is thought to undergo amyloid formation in the biofilm matrix in the presence of TapA and other matrix components. Adapted from Diehl *et al.* (2018).

expression, and because circular dichroism (CD) measurements of monomeric and fibrous recombinant TasA show similar spectra, leading the authors to assume there are no changes in secondary structure; thus, the authors propose a head-to-tail arrangement of globular subunits within fibres. Furthermore, a helical repeat is detected around 45 Å (Erskine *et al.*, 2018). In principle, such a helical repeat could arise even in an amyloid protein by residues outside the amyloid core; however, the arrangement of TasA in fibres remains unclear. Apart from a mere structural biology standpoint, there is further relevance to this question since bacterial amyloids are used as models for amyloid generation in neurodegenerative disease, and even used in screens for anti-amyloid compounds (Jain *et al.*, 2017; Romero *et al.*, 2013). It has furthermore been shown that TasA expression results in large fibre bundles visible using scanning electron microscopy (SEM). Indeed, further studies have shown that TasA fibres spontaneously associate into bundles, depending on fibre and salt concentration (Ghrayeb *et al.*, 2021). It is unclear, however, how these interactions are mediated on a molecular level, which is an important prerequisite to understanding how TasA scaffolds the biofilm.

1.7.4 Project aims

TasA is of significant importance for biofilm formation in *B. subtilis*. While it has been established that TasA forms fibres which mediate biofilm scaffolding (Romero *et al.*, 2010), it is unclear how TasA subunits are arranged into fibres. Two opposing ideas have been proposed: One suggests that TasA fibres are amyloid, with TasA monomers assembling into a cross-β fibre (Diehl *et al.*, 2018; Romero *et al.*, 2010); another suggests that TasA are not amyloid in character, but that fibres form through head-to-tail arrangements of monomeric TasA subunits (Erskine *et al.*, 2018). This project aimed to perform helical-reconstruction cryo-EM on TasA fibres, allowing to determine a high-resolution cryo-EM density and enabled *de novo* model building. As TasA fibres arrange into higher-order bundles in order to mediate their function within the biofilm, we furthermore performed electron cryotomography to analyse fibre arrangement within TasA bundles *in vitro* and image fibres in the context of a biofilm, thus providing the structural basis of the assembly of TasA monomers into fibres, and fibres into bundles that stabilise *B. subtilis* biofilms.

1.8 CS-TV^2 reconstruction for *in situ* structural biology

1.8.1 Structural solution of macromolecules using electron cryomicroscopy

Over the recent decade, electron cryomicroscopy (cryo-EM) has risen from a relatively obscure technique for nanometre-resolution determination of biological assemblies to a high-resolution structural technique that has displaced X-ray crystallography as a go-to method for structure determination (Kühlbrandt, 2014). Cryo-EM is based on the imaging of biological specimens applied to a sample grid that has been flash-frozen in liquid ethane, leading to vitrification of samples, and allowing imaging in an electron cryomicroscope, where radiation damage and sample movement can be minimised (Dubochet *et al.*, 1988). The most popular cryo-EM method is single particle analysis, a technique in which three-dimensional densities are reconstructed from 2D projection images of a given object, as each projection image corresponds to a slice in Fourier space of the original 3D object (Frank, 2006). By determining how the 2D projection image relates towards the 3D structure of the object, the object can be reconstructed if particles within the 2D images are randomly oriented (Frank *et al.*, 1981; Naydenova & Russo, 2017). Among the ground-breaking developments that have improved the power of the technique in the last decade are so-called direct electron detectors that can detect single electrons with higher spatial resolution and enable rapid acquisition of image sequences (Kühlbrandt, 2014), thus allowing to correct beam-induced molecular motion that has been one of the limiting factor in determining high-resolution structures (Brilot *et al.*, 2012; Zheng *et al.*, 2017). The development of more and more sophisticated image analysis procedures has accompanied and aided these developments (Punjani *et al.*, 2017; Scheres, 2012; Tegunov & Cramer, 2019; Tegunov *et al.*, 2021; Zhong *et al.*, 2021). Compared to X-ray crystallography, cryo-

EM requires only small amounts of specimen, and extensive crystallisation trials can be avoided, leading to ever-increasing popularity of the method (Bai *et al.*, 2015). Obtaining ever-increasing resolutions, SPA was shown to be able to reach true atomic resolution where individual atoms within proteins are resolved with ideal samples (Nakane *et al.*, 2020; Yip *et al.*, 2020).

1.8.2 Electron cryotomography and subtomogram averaging

While SPA remains the most popular flavour of cryo-EM, an alternative cryo-EM method termed electron cryotomography (cryo-ET) has also steadily gained in popularity (Lučić *et al.*, 2005). In cryo-ET, a series of cryo-EM images of an object of interest are acquired whilst the sample stage is tilted. Based on prior knowledge (and precise determination) of the applied stage tilt upon acquisition, the images can be used to reconstruct a 3D volume of the sample termed a tomogram. Uniquely, this allows the reconstruction of pleomorphic objects, such as viruses or cells, as it does not require any averaging of the object of interest in order to create a 3D volume (Baumeister *et al.*, 1999). Additionally, compared to SPA, it does not require purification of the macromolecule of interest, and macromolecules can thus be imaged in their native cellular environment. This way, native conformations and interaction partners can be resolved that may be lost upon purification (Briggs, 2013; Medalia *et al.*, 2002).

Cryo-EM and cryo-ET require thin samples, as samples thicker than ~200 nm result in lowered signal-to-noise ratios due to increasing inelastic scattering of electrons at conventionally used voltages (McIntosh *et al.*, 2005). To allow imaging of thick prokaryotic cells and eukaryotic cells, samples were initially thinned by cryo-sectioning, using diamond knifes to slice cells into thin sections that are amenable to cryo-EM studies (Al-Amoudi *et al.*, 2004). This technique, however, is labour-intensive and typically results in sample deformations in the direction of the diamond knife's path, which make it unsuitable for structure determination. A more advanced technique for sample thinning that has rapidly increased in popularity is focused ion beam (FIB) milling (Villa *et al.*, 2013). Here, a targeted beam of, typically, gallium ions is used to ablate areas of the target cell above and below the area of interest, resulting in the creation of a thin cellular slab that is amenable to high-resolution structural studies. Among many more success stories, this technique enabled studies in the cellular environment showing proteasomes are tethered to the nuclear pore complex (Albert *et al.*, 2017), the architecture of the Parkinson's Disease protein LRRK2 on microtubules (Watanabe *et al.*, 2020), and the interaction of transcription and translation complexes in bacterial cells (O'Reilly *et al.*, 2020).

Cryo-ET data of biological specimens typically suffers from inherently low contrast. This is because only limited dose can be applied to the specimen, as radiation damage can result in the destruction of molecular bonds, hindering structure determination (Baker & Rubinstein, 2010). In order to improve upon the signal-to-noise ratio – and hence resolution – of a macromolecule of interest, a technique called subtomogram averaging (STA) can be applied (Briggs, 2013). Similar to SPA, this technique employs the averaging of copies of individual copies of the macromolecule of interest within the tomogram (so-called subtomograms) to obtain a reconstruction with a higher signal-to-noise ratio and, thus, higher resolution. This way near-atomic resolution of ribosomes in *Mycoplasma pneunomiae* could recently be achieved for this ideal sample (Tegunov *et al.*, 2021). Bottlenecks in sample preparation, target abundance and identification, data acquisition and processing, however, typically result in resolutions in the nanometre range between 10 Å and 50 Å for most specimens (Böhning & Bharat, 2021).

1.8.3 Reconstruction algorithms for cryo-ET

In order to obtain a 3D volume from acquired 2D tilt images, the 2D images need to be reconstructed into a 3D volume. On a mathematical level, each 2D image corresponds to a slice of the 3D volume in Fourier space (De Rosier & Klug, 1968). Since the tilt angles with which each image was acquired are known, Fourier transforms of the 2D images can be arranged accordingly in 3D Fourier space to reconstruct the volume (Baumeister *et al.*, 1999). In the most popular method, weighted back-projection (WBP), reconstruction is performed purely based on such geometrical considerations, usually after tilt angles are refined by tracking of so-called fiducials throughout the tilt series (Kremer *et al.*, 1996). The central (near-zero) components of the Fourier space are down-weighted as they are over-represented in the reconstruction. This reconstruction type, however, suffers from low signal-to-noise ratios and is hence not the preferred methodology for generating tomograms with high contrast. A wide range of alternative reconstruction algorithms have hence been created with a goal of producing higher-contrast reconstructions. A wide-spread option is the simultaneous iterations reconstruction technique (SIRT), which is based on iteratively minimising differences between the reconstructed tomogram and the tilt series (Gilbert, 1972). To assess and minimise differences, the reconstructed tomogram is projected into a simulated tilt series, compared to the original tilt series, and reconstruction iteratively adjusted according to differences between the two. While low-resolution features – and, thus, contrast - are thusly enhanced and the visibility of macromolecular features in the tomogram is increased, high-resolution features below the noise levels of tomograms are not well retained (Wan & Briggs, 2016). This loss of high-resolution information is the main reason why WBP is the commonly used reconstruction technique for subsequent STA, where (sub)nanometre-resolution features are to be resolved. Some algorithms have been developed in an effort to retain high-resolution information within the reconstruction. Examples of these are the iterative nonuniform fast Fourier transform reconstruction (INFR) (Chen & Förster, 2014), super-sampling simultaneous algebraic reconstruction technique (super-sampling SART) (Kunz & Frangakis, 2014), and the progressive stochastic reconstruction technique (PSRT) (Turoňová *et al.*, 2015), which show increased resolution in STA at 20-35 Å as compared to WBP. Additionally, these reconstruction techniques demonstrated benefits in reducing tomographic data requirements: Because the sample can only be feasibly tilted in a limited range (usually between -60° and +60°), the areas of Fourier space outside this range are not filled, leading to a so-called missing wedge, resulting in missing information and limited resolution in the Z direction of the tomogram (Schmid & Booth, 2008). Furthermore, tilt images are typically acquired only every 1-3°, and information in between tilt images may be missing.

1.8.4 Compressed sensing reconstruction

Compressed sensing is a reconstruction method that is commonly employed in electron tomography of non-biological materials to increase sample contrast within the reconstruction (Leary *et al.*, 2013; Saghi *et al.*, 2011). On a mathematical level, compressed sensing exploits the fact that tomographic reconstructions - while ‘undersampled’, meaning only a certain amount of tilt images can be taken - can be approximated in a mathematically more compact form since the object that it depicts is typically highly structured and can hence be recovered using limited coefficients

(Donoho, 2006). In order for compressed sensing to be able to recover the object of interest with high fidelity, the object itself needs to be ‘sparse’, meaning it is amenable to being recovered with less coefficients. If this is the case, the object can be approximated using significantly less data than required by typical sampling theorems for tomographic reconstruction. This means that fewer tilt images can be acquired, increasing data acquisition throughput, and decreasing dose sensitivity requirements (Leary *et al.*, 2013).

For tomographic reconstruction of electron tomography data, compressed sensing algorithms can be formulated as the regularisation reconstruction problem

$$\boldsymbol{f}(\boldsymbol{r}) = \arg\min_{\boldsymbol{r}} \left\{ ||\widehat{P}f(\boldsymbol{r}) - \mathit{\Gamma}^{exp}\,||_{\ell_2}^{2} + \lambda||\psi\{\boldsymbol{f}(\boldsymbol{r})\}\,||_{\ell_1} \right\}$$

with f(r) representing the reconstructed volume with 3D coordinates r, P the projection operator, r^{exp} the experimental data and $||\ \ ||_{lp}$ denoting an lp-norm. The first part of the equation produces a ‘least squares’ term to account for fidelity to the experimental data. The second part of the equation $\lambda||\psi(f(r)\}||_{l1}$ accounts for regularisation, introducing sparsity constraints intended to optimally account for the experimental data.

In order to be able to reproduce an object using limited signal, compressed sensing involves identifying the sparse representation of the object of interest. A particular set of algorithms recently employed in materials science exploits second-order total variation (Sanders *et al.*, 2017), which uses second-order gradients as the transform domain (Figure 1.10). This requires that the represented object is dense – i.e., has many non-zero intensities – and can assume various intensities while assuming a piece-wise linear structure in the reconstructed object. Since these characteristics typically apply to high-resolution molecular structures, this algorithm may be suitable for reconstruction of biological macromolecules.

While CS algorithms have found ample application in material science, there are few applications for cryo-ET of biological samples. Notably, one study has shown compressed sensing reconstruction can increase sub-tomogram averaging resolutions up to ~15 Å (Deng *et al.*, 2016), but it is unclear whether CS approaches can improve resolution at even higher resolutions; particularly, it is unclear whether sub-nanometre level resolution information, with which secondary structure can be resolved, can be retained upon reconstruction. Furthermore, while it has been shown that CS can reduce the number of tilt images required to reconstruct a 3D volume (Guay *et al.*, 2016; Leary *et al.*, 2013), it is unclear whether this translates into less data being needed to obtain an STA average of a certain resolution.

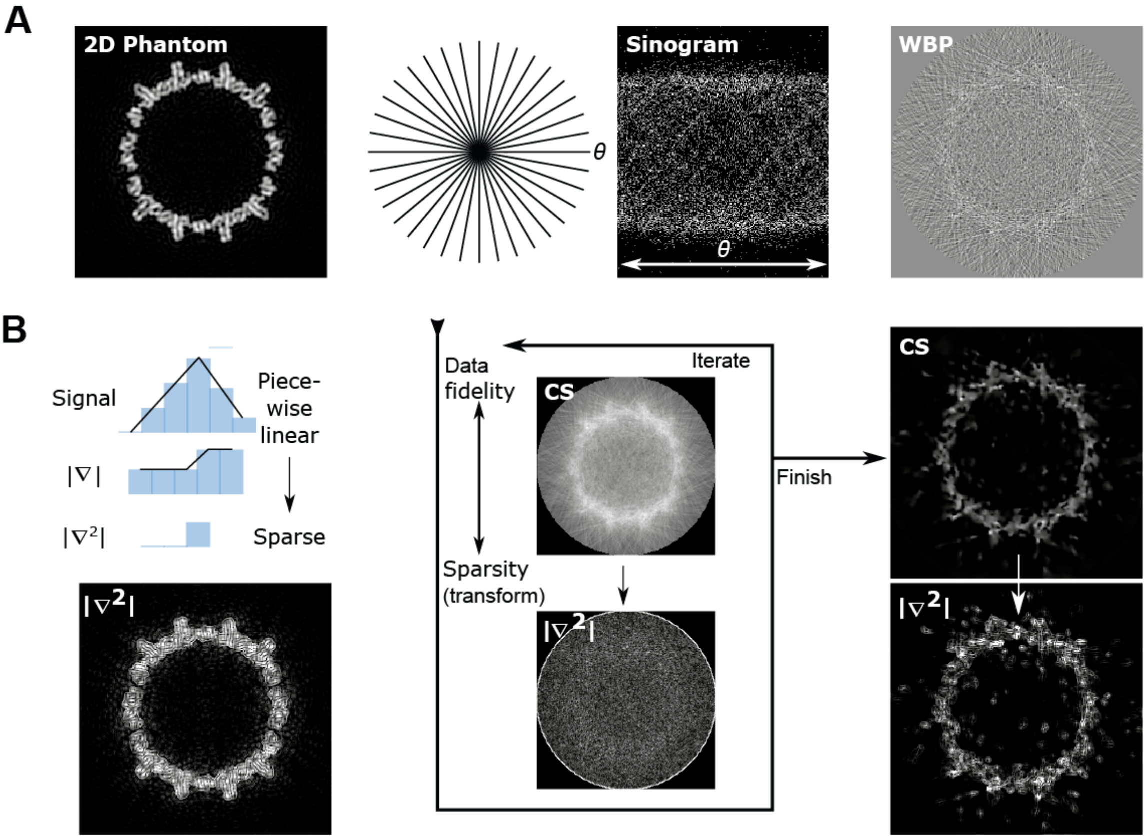


Figure 1.10: Schematic depiction of the steps of 3D second-order total variation (CS-TV^2) tomographic reconstruction. A) A ground-truth density of an example object (here: HBV capsid) is projected to create simulated tilt images that are arranged into a sinogram, which is used to calculate a WBP density. B) CS-TV^2 aims to find a sparse representation of non-zero elements (shown in blue) in the second-order gradient. The algorithm iteratively balances data fidelity with sparsity in the second-order gradient. Created by Sean Collins; published in Böhning *et al.* (2022a).

1.8.5 Project aims

Previous work has shown that compressed sensing algorithms are suitable for reconstruction of electron cryotomograms in materials (Leary *et al.*, 2013) and biological science (Deng *et al.*, 2016; Guay *et al.*, 2016). This work aimed to establish whether a second-order total variation CS algorithm (CS-TV^2), which has recently been successfully applied in materials science (Collins *et al.*, 2019), could also be applicable to biological data. Compared to previous applications of CS algorithms to biological samples (Deng *et al.*, 2016), which apply regularisation in 2D, this algorithm performs regularisation in 3D and may hence be more suitable for the reconstruction of biological specimens which are inherently 3D objects. This study hence assessed whether CS-TV^2 is suitable for increasing the contrast of biological structures in tomograms. Furthermore, it is unclear whether CS-TV^2 can retain high-resolution information in biological samples, ET data of which is typically noisier than in materials science. In particular, no previous reconstruction algorithm other than WBP has been shown to allow the resolution of subnanometre features of biological macromolecules using STA. This study hence aimed to establish whether CS-TV^2 allows the resolution of such features. Furthermore, copy number of target macromolecules within cryo-ET data is often a limiting factor, as throughput limitations in sample preparation, localisation, acquisition, and processing mean only limited data can sometimes be produced; this is a particular bottleneck for *in situ* structural biology, when electron cryotomography is applied to cellular specimens to resolve macromolecules in their cellular environment (Böhning & Bharat, 2021). This study hence aimed to determine whether CS-TV^2 reduces the subtomogram number needed to reconstruct an STA average at a certain resolution, and whether it may thus alleviate data requirements in STA.

2. Methods

2.1 Used strains, plasmids, and oligonucleotides

Table 1: Bacterial strains used in this study.

Strain	Description	Source
E. coli		
BL21 (DE3)	Genotype: B F^- *dcm ompT hsdS* (r_B^- m_B^-) *gal* λ(DE3)	Agilent
XL-1	Genotype: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB* $lacI^q$ *ZΔM15* Tn*10* (Tet^r)]	Agilent
P. aeruginosa		
PAO1 Δ*pilA* Δ*fliC* Δ*mvaT* Δ*cupA6* (TBS060)	PAO1; chromosomal deletion of *pilA* (PA4525), *fliC* (PA1092), *mvaT* (PA4315) and *cupA6* (PA2133)	Kira Glatzel, Published in Böhning *et al.* (2022b)
PAO1 Δ*pilA* Δ*fliC* Δ*mvaT* Δ*cupA6* CupE1 $111\text{-}113_{AGA}$ (TBS234)	PAO1; chromosomal deletion of *pilA* (PA4525), *fliC* (PA1092), *mvaT* (PA4315) and *cupA6* (PA2133) with residues 111-113 of CupE1 substituted with Ala-Gly-Ala	Thomas Clamens, published in (Böhning *et al.*, 2022b)
B. subtilis		
NCIB 3610	Wild-type, undomesticated strain	Liraz Chai
ZK4363	NCIB 3610 Δ*sinR* Δ*eps yve-yvf*::*tet* (*epsA-O*::*tet*) *sinR*::*spc*	Liraz Chai
ZK3657	NCIB 3610 Δ*tasA tasA*::*kan*	Liraz Chai
MG1	NCIB 3610 Δ*eps* Δ*sinR* encoding TasA Δ28-38 *yve-yvf*::*tet* (*epsA-O*::*tet*) *sinR*::*spc tasA*Δ28-38-kan	Mnar Ghrayeb Published in Böhning et al. (2022c)
MG2	NCIB 3610 Δ*eps* Δ*sinR* encoding TasA $174\text{-}177_{AAAA}$ *yve-yvf*::*tet* (*epsA-O*::*tet*) *sinR*::*spc tasA*-D174A, G175A, K176A, T177A-*kan*	Mnar Ghrayeb Published in Böhning *et al.* (2022c)
MG3	NCIB 3610 WT encoding TasA Δ28-38 *yve-yvf*::*tet* (*epsA-O*::tet) *sinR*::*spc tasA*Δ28-38-*kan*	Mnar Ghrayeb Published in Böhning *et al.* (2022c)
MG4	NCIB 3610 WT encoding TasA $174\text{-}177_{AAAA}$ *yve-yvf*::*tet* (*epsA-O*::*tet*) *sinR*::*spc tasA*-D174A, G175A, K176A, T177A-kan	Mnar Ghrayeb Published in Böhning *et al.* (2022c)

Table 2: Plasmids used in this study.

Plasmid	Description	Source
pKNG-(Δ*cupA6*)	Suicide vector for chromosomal deletion of *cupA6* from *P. aeruginosa* PAO1, Sm^{R}	Kira Glatzel, published in Böhning *et al.* (2022b)
pKNG-(Δ*cupE1-2*)	Suicide vector for chromosomal deletion of *cupE1-2* from *P. aeruginosa* PAO1, Sm^{R}	Kira Glatzel, published in Böhning *et al.* (2022b)
pKNG-(CupE1-2 AGATSST)	Introduction of CupE1-2-encoding sequence with 111-113$_{\mathrm{AGA}}$ substitution into Δ*cupE1-2* background	Thomas Clamens, Published in Böhning *et al.* (2022b)
pKNG101	Non-replicative suicide vector for *P. aeruginosa* chromosomal mutagenesis. ori6K, mobRK2, *sacB* for sucrose sensitivity, Sm^{R}	Kaniga & Davison (1991)
pET31b-FapC-His	Expression vector for *P. aeruginosa* PAO1 FapC-His	Morten Dueholm

Table 3: Oligonucleotides used in this study.

Oligonucleotide	Oligonucleotide name	Sequence (5'-3')
pKNG-(Δ*cupA6*)	Δ*cupA6* P1	AACCGGATACCGCGCTGGC
	Δ*cupA6* P2	TCACCCCTGACCGTTCACTGTAGAGC
	Δ*cupA6* P3	GTGAACGGTCAGGGGTGACCGGGGAGG
	Δ*cupA6* P4	CCTGGGTGAACAGTTCCACG
	Δ*cupA6* P5	GAGAAACTCGCGCTCGGCAG
	Δ*cupA6* P6	AGTACCAGCAACCACAGGCTG
pKNG-(Δ*cupE1-2*)	Δ*cupE1-2* P1	CTGCTGTTCGGCACCATCGCCTC
	Δ*cupE1-2* P2	GGCGGTGTTCAACAAGACTGCCGGAAT
	Δ*cupE1-2* P3	GTCTTGTTGAACACCGCCCTCACCCTG
	Δ*cupE1-2* P4	CTGCTGCTGCCGCCATTGA
	Δ*cupE1-2* P5	GCCATCACCCTGGCCTTC
	Δ*cupE1-2* P6	TTGCCCTCCCAGCGGACAT
pKNG-(*cupE1-2* CupE1 111-113$_{\mathrm{AGA}}$)	CupE1 111-113$_{\mathrm{AGA}}$ Fw	AGCGGCGGCCAGCGCTACATGCAGAACGCCGGCGCCGGCGCCGGCGCCATCGCCTACAACATCTATTCGGACG
	CupE1 111-113$_{\mathrm{AGA}}$ Rv	CGTCCGAATAGATGTTGTAGGCGATGGCGCCGGCGCCGGCGCCGGCGTTCTGCATGTAGCGCTGGCCGCCGCT
Chromosomal deletion of TasA residues 28-38	T2838P1 Fw	GTATATAAACCGGCAGGCTACCCG
	T2838P2 Rv	CGTACCTGATGCAAAAGTTGCCCATGTTCCTCCTCCAAC
	T2838P3 Fw	ACTTTTGCATCAGGTACGCTTG
	T2838P3 Fw	ACTTTTGCATCAGGTACGCTTG
	T2838P4 Rv	CTGAGCGAGGGAGCAGAATTAATTTTTATCCTCGCTATGCGCTTTTTC
	T2838P5 Fw	GTTGACCAGTGCTCCCTGTAACAGCAAAAAAAAGAGACGGCC
	T2838P6 Rv	AGTTGAAATTGAAAATGGCGGATTGTTT

Substitution of residues 174-177 with AAAA	T174AP1 Fw	`GACGTTATTACATTTATGCAGGATGCAAAT`
	T174AP2 Rv	`AGGAGCTGCAGCTGCAGCAATTGTTGCTACATTGACTTTACCGCTTG`
	T174AP3 Fw	`GCTGCAGCTGCAGCTCCTGAATATGATGGTGTTCCA`
Sequencing primers	PseqF Fw	`ACAAAGGACAGCACCATGTC`
	PseqR Rv	`CTGAGCGAGGGAGCAGAATTAATTTTTATCCTCGCTATGCGCTTTTTC`

2.2 Concentration measurements of proteins and nucleic acids

Unless differently stated, protein and DNA concentrations were measured using a NanoDrop spectrophotometer (Thermo Scientific). The concentration of double-stranded DNA was determined based on the assumption that an absorbance at 260 nm corresponds to DNA concentrations of 50 ng/µl. Protein concentrations were determined according to the Lambert-Beer law from their absorbance at 280 nm and the Molar extinction coefficient of the respective target protein as calculated by the Benchling software. TasA concentrations were measured by Mnar Ghrayeb using a BCA protein assay kit (ThermoFisher Scientific) (Böhning *et al.*, 2022c).

2.3 Creation of mutant strains in *P. aeruginosa*

This method was performed by Dr. Kira Glatzel and Dr. Thomas Clamens in the group of Prof. Alain Filloux (Imperial College, London) and is described in the corresponding study (Böhning et al., 2022b).

2.4 Creation of genomic TasA mutants in *B. subtilis*

This method was performed by Mnar Ghrayeb and Prof. Liraz Chai (Hebrew University Jerusalem) and is described in the corresponding study (Böhning et al., 2022c).

2.5 Recombinant FapC protein expression and purification

For recombinant expression of FapC, 50-100 ng of an expression vector encoding the mature form of FapC with a C-terminal His-tag (pET31b-FapC-His, kindly provided by Prof. Morten Dueholm, Aalborg University) were transformed into *E. coli* BL21 cells and grown to an OD of 0.4-0.8 at 37°C at 180 rpm agitation. Expression was induced by adding 1 mM IPTG and further incubation for 3 h at 37°C at 180 rpm agitation. Cells were harvested by centrifugation for 20 min at 5,000 rcf and 4°C. Cell pellets were stored at -80°C until further use. For purification of His-tagged FapC, cell pellets were thawed and resuspended in denaturing lysis buffer (50 mM HEPES pH 7.4, 500 mM NaCl, 8 M urea, 20 mM imidazole). Cells were lysed using a sonicator at 10 µM amplitude for 1-2 min. The lysate was subjected to centrifugation for 30 min at 35,000 rcf at 4°C to separate insoluble cell debris from soluble protein. The resulting supernatant was filtered using a 0.45 µm syringe filter and subjected to immobilised metal ion chromatography using a HisTrap FF 5 ml column (Cytiva) mounted on an AEKTA Pure system in lysis buffer. Protein was eluted using a linear gradient from 0 to 250 mM imidazole. For analysis of FapC-His purification, protein samples were loaded onto a 4-12% SDS-PAGE gel and separated using a voltage of 200 V. Gels were stained using Coomassie Blue to visualise protein migration. AllBlue protein marker (BioRad) was used to facilitate assignment of approximate protein sizes. The resulting fractions were concentrated using a 10 kDa cut-off spin concentrator (Millipore) and subjected to desalting using a HiPrep 26/10 Desalting column (Cytiva). Resulting fractions containing FapC protein were frozen in liquid nitrogen and stored at -80°C until further use.

To assemble fibres, fibre formation was seeded by substoichiometric addition of 1:100 (w/w) sonicated FapC fibres to FapC-His (1-2 g/l), followed by incubation at

37°C for 3 d in 20 mM HEPES (pH 7.4) and 10 mM CHAPS. Polymerised FapC-His fibres were kindly provided by Prof. Morten Dueholm for initial experiments; further seeds were created by polymerisation and sonication of thusly created FapC fibres.

2.6 Purification of CupE pili from *P. aeruginosa*

For purification of CupE pili, cells of the strains TBS060 or TBS234 were grown in LB liquid culture overnight at 37°C and 180 rpm agitation, plated onto LB agar plates, and incubated at 37°C overnight. Resulting lawn cultures were harvested by scraping off the bacterial lawn from the plate and resuspended in 1x phosphate-buffered saline (PBS). The resulting suspension was vortexed for 90 s to promote dissociation of pili from the cell surface. Cells were then submitted to centrifugation at 4,500 rcf for 20 min at 4°C. The resulting supernatant was submitted to 3-4 centrifugation steps at 16,000 rcf at RT to remove remaining cells and cellular debris from the solution. Pili were precipitated by adjusting the buffer to a final concentration of 500 mM NaCl and 3% (w/v) PEG 6000 and incubation on ice for 1 h. Precipitated pili were collected using centrifugation for 30 min at 16,000 rcf at RT and resuspended in 1x PBS. For cryo-EM, precipitation was repeated once in the same manner to produce the final sample.

2.7 Purification of TasA fibres from *B. subtilis*

This experiment was performed by Mnar Ghrayeb and is described in Böhning et al. (2022c).

2.8 Circular Dichroism

This method was performed by Mnar Ghrayeb and Liraz Chai and is described in Böhning et al. (2022c).

2.9 Cryo-EM grid preparation

All cryo-EM samples were prepared at the Central Oxford Structural and Molecular Imaging Centre. For grid preparation of purified CupE pili and Fap fibres, 2.5 μl of isolated CupE pili (1-3 g/l) or Fap fibres (2 g/l) were applied to a freshly glow-discharged Quantifoil R2/2 Cu/Rh 200 mesh grid and plunge-frozen in liquid ethane using a Vitrobot Mark IV. For tomography sample preparation of *P. aeruginosa* PAO1 Δ*fliC* Δ*pilA* Δ*mvaT* Δ*cupA6*, cells from an overnight lawn culture on an LB Agar plate grown at 37°C were resuspended in PBS and applied to a glow-discharged Quantifoil R3.5/1 Cu/Rh 200 cryo-EM grid and plunge-frozen.

For the preparation of cryo-EM grids of purified TasA fibres, TasA isolates were diluted to 10 μM TasA in 12.5 mM Tris pH 8 and 25 mM NaCl, and 2.5 μl of sample were applied to a freshly glow-discharged Quantifoil R 2/2 Cu/Rh 200 mesh grid and plunge-frozen into liquid ethane using a Vitrobot Mark IV (ThermoFisher) at 100% humidity at 10 °C. For tomographic data acquisition, 10 nm Protein-A-gold beads (CMC Utrecht) were added as fiducials prior to plunge-freezing. For imaging of disassembled biofilms, *B. subtilis* Δ*eps* Δ*sinR* biofilm was scraped from a plate and resuspended in PBS immediately prior to sample preparation.

2.10 Cryo-EM data acquisition

Cryo-EM data was collected in a Titan Krios G3 microscope (ThermoFisher) with an acceleration voltage of 300 kV and a Quantum energy filter (slit width 20 eV) and a K3 direct electron detector (Gatan) at the Central Oxford Structural and Molecular Imaging Centre. Images were collected in super-resolution counting mode using a pixel size of 1.092 Å/px for helical reconstruction of TasA and CupE fibres, 1.382 Å/px for Fap, and 3.402 Å/px for imaging of all cellular samples and mutant TasA (see also Table 4 and Table 5 for cryo-EM acquisition details and statistics of TasA and CupE), generally using electron doses of 50-60 e/Å^2.

For imaging of *B. subtilis* biofilm ECM components, a total dose of 47.5 e/Å^2 distributed over 80 frames and defoci between -3 to -7 µm were used. For the acquisiton of FapC-His, defoci between -1 and -3.5 µm were used with a dose of 52 e/Å^2, distributed over 40 frames. Cryo-ET tilt series data was collected on the same microscope using SerialEM (Mastronarde, 2005), generally with a total dose of 120-180 e/Å^2 and defoci of -8 to -12 µm, collected between ±60° tilt of the stage at 1° tilt increments.

2.11 Helical reconstruction image processing

Cryo-EM processing of CupE, FapC and TasA fibres was performed in RELION (He & Scheres, 2017; Scheres, 2012). CTF parameters were estimated using CTFFIND4 (Rohou & Grigorieff, 2015). Initial helical symmetry for both CupE and TasA filaments was estimated from the observed periodicity of subunits within 2D class averages. For TasA fibres, an initial model was created from 2D class averages using the *relion_helix_inimodel2d* programme. For CupE fibres, an initial map representing the low-resolution features of the filament was used as previously created by Nina Sulkowski (Sulkowski, 2021). 3D classification was used to identify a subset of particles that supported refinement to high resolution. For final refinement, CTF multiplication was used for the final polished set of particles. Symmetry searches were used during reconstruction. Gold-standard resolution FSC measurements and local resolution measurements were performed using RELION.

2.12 Model building and refinement

Model building for CupE1 and TasA filaments was performed in Coot. For TasA, a previously published crystal structure of the soluble TasA monomer was used as a starting model (PDB 5OF1). For CupE1, a homology model was calculated based on the structure of CsuA/B from *Acinetobacter baumannii* by Dr. Vikram Alva using MODELLER (Webb & Sali, 2017). Generally, residues of the starting model inconsistent with the experimental cryo-EM density were manually deleted and rebuilt. Built structures were subjected to real-space refinement against the cryo-EM map within the Phenix package (Adams *et al.*, 2010). Three subunits of TasA were built and used for final refinement, and five subunits of CupE1. Non-crystallographic symmetry between individual TasA subunits was applied for all refinement runs. Model validation including map-vs-model resolution estimation was performed in Phenix.

Table 4: TasA cryo-EM data acquisition and processing statistics.

Data collection and processing	TasA
Magnification	81,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	48.5
Defocus range (μm)	-1 to -2
Pixel size (Å)	0.546 (super-resolution)
Symmetry imposed	Helical, final: -55.18°, 48.36 Å
Initial particle images (no.)	1,478,370
Final particle images (no.)	103,009
Map resolution (Å) FSC threshold	3.5 (0.143)
Map local resolution range (Å)	3.6 – 4.1
Refinement	
Initial model used (PDB code)	5OF1
Model resolution (Å) FSC threshold	3.7 (0.5)
Model resolution range (Å)	n/a
Map sharpening *B* factor (Å^2)	-67.9
Model composition	
Non-hydrogen atoms	5,409
Protein residues	699
Ligands	0
B factors (Å^2)	
Protein	
Ligand	
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.544
Validation	
MolProbity score	1.73
Clashscore	6.34
Poor rotamers (%)	0
Ramachandran plot	
Favored (%)	94.37
Allowed (%)	5.63
Disallowed (%)	0

Table 5: CupE1 cryo-EM data acquisition and processing statistics.

Data collection and processing	CupE1 WT	CupE1 111-113$_{AGA}$
Microscope	Krios Titan G3	Krios Titan G3
Magnification	81,000	81,000
Voltage (kV)	300	300
Electron exposure (e^{-}/Å^{2})	46	45
Defocus range (μm)	-1 to -2.5	-1 to -2.5
Pixel size (Å)	0.546 (super-resolution) 1.092 (physical, final)	0.546 (super-resolution) 1.092 (physical, final)
Symmetry imposed	Helical, final: 214.56° 33.12 Å	Helical, final: 214.67°, 33.04 Å
Initial particle images (no.)	3,766,858	1,259,369
Final particle images (no.)	274,457	88,886
Map resolution (Å)	3.5	4.1
FSC threshold	0.143	0.143
Map local resolution range (Å)	3.4-4.2	3.8-4.7
Refinement		No atomic model built
Initial model used (PDB code)	n/a	
Model resolution (Å)	2.9/3.1/3.9	
FSC threshold	0/0.143/0.5	
Model resolution range (Å)	n/a	
Map sharpening *B* factor (Å^{2})	-97	-152
Model composition		
Non-hydrogen atoms	5655	
Protein residues	795	
Ligands	0	
B factors (Å^{2})		
Protein	66.88	
Ligand	n/a	
R.m.s. deviations		
Bond lengths (Å)	0.002	
Bond angles (°)	0.547	
Validation		
MolProbity score	1.95	
Clashscore	8.03	
Poor rotamers (%)	0	
Ramachandran plot		
Favored (%)	91.08	
Allowed (%)	8.92	
Disallowed (%)	0	

2.13 Data visualisation

All cryo-EM images were visualised in IMOD. For raw micrographs, Fiji (Schindelin *et al.*, 2012) was used for bandpass and Gaussian-filtering to increase contrast. All atomic structures were visualised in ChimeraX (Goddard *et al.*, 2018). Segmentation of tomograms was performed in IMOD. Hydrophobic surfaces were calculated in ChimeraX using the in-built *mlp* function. Difference maps were calculated using EMDA (Warshamanage *et al.*, 2022), with maps lowpass-filtered to the lowest common resolution. Fourier Shell correlation plots and bar diagrams were plotted in MATLAB (MathWorks). HBV STA maps were lowpass-filtered to their gold-standard resolution if secondary structure was visible; otherwise, they were lowpass-filtered to 20 Å resolution.

2.14 Quantification of TasA mutant fibre formation

To determine the fibre formation propensity of TasA Δ28-38, a random subset of 20 micrographs was used and fibres were counted manually. To determine fibre formation propensity and bundling in the TasA $174\text{-}177_{\text{AAAA}}$ mutant, 100 fibres were counted in random micrographs. Fibres were considered bundling if at least two visible periodic interactions were visible.

2.15 Sequence alignment

This procedure was performed by Adrian Dobbelstein and Vikram Alva.

The multiple sequence alignment (MSA) of archaic CUP pilins was performed by calculating initial alignments using PROMALS3D (Pei *et al.*, 2008) using default settings and subsequent manual curation. Sequence alignment of TasA with TapA was performed using the ClustalW web service from EMBL-EBI (McWilliam *et al.*, 2013).

2.16 Structural modelling

A homology model of CalY was calculated based on the cryo-EM model of TasA obtained in this study using SWISSMODEL (Schwede *et al.*, 2003). AlphaFold Multimer (Evans *et al.*, 2021) predictions of CupE filaments were conducted by Adrian Dobbelstein and Vikram Alva using AlphaFold version 2.1.1. Signal peptides were predicted using SignalP-6.0 (Teufel *et al.*, 2022) and excluded from the query sequence. Template modelling was used, and structures inferred using one iteration of MSA recycling and all five different model parameter sets. Predictions were ranked by pTM score and the highest-ranking model was respectively selected. Predictions were performed using the 'Raven' high-performance computer run by the Max-Planck Computing & Data facility in Garching, Germany. AlphaFold Multimer predictions of the TasA-TapA complex and FapC were calculated using the Google Colab implementation of Alphafold 2.1.0 using default settings.

2.17 Molecular dynamics simulations

This method was performed by Conrado Pedebos and Syma Khalid and is published as part of Böhning et al. (2022c).

Models of fibre doublets were obtained by fitting fibre models made from the SPA structure into the tomographic density of fibre doublets. This model was then submitted to the Martini Maker tool (Qi *et al.*, 2015) within the CHARM-GUI server (Jo *et al.*, 2008) to produce an initial coarse grain model. The system was solvated with explicit water, and counter ions were added to neutralise charges, with 1 M KCl added to all simulations. The GROMACS simulation suite (version 2020.4) (Van Der Spoel *et al.*, 2005) was used for MD simulations along with the Martini 2.2 force field (de Jong *et al.*, 2013). To coarse-grain the model, the ElNeDyn model was used with elastic network cut-off of 0.9 nm and a force constant of 500 kJ $mol^{-1}nm^{-2}$.

Filament starting orientations were oriented along the Y axis in such a way that extremities of each filament interacted non-covalently with their periodic images to form a continuous and stable structure. Two stages of equilibration were performed, employing NVT and NPT ensembles for 10 ns each, keeping protein beads restrained but enabling ions and water to relax. Additional equilibration times were used for anti-parallel fibre interactions. Three production simulations were performed for antiparallel and parallel simulations, respectively, with durations of 500 ns each at 313 K using the velocity rescale thermostat with a coupling constant of $\tau = 1$. Pressure coupling was implemented using a time constant of 12 ps with the Parrinello-Rahman barostat. The reaction-field method was used to treat electrostatics, using dielectric constants of 15 for short-term regimes, and infinity for long-range regimes. For non-bonded and electrostatic interactions in the short range, a cut-off of 1.1 nm was used. In order to improve conformational sampling, starting random velocities were modified at the beginning of each replicate. Manipulation, visualisation and analysis of molecules for MD was performed using VMD (Humphrey *et al.*, 1996) and Pymol (DeLano, 2002). Interactions were quantified by calculating how much time a residue from a filament was in a distance of 0.6 nm to a residue of the other filament. Interaction time of residue pairs was converted to a scale from 0 to 1, where 0 means 0% of the interactions and 1 means 100% of the time within the cut-off was spent interacting.

2.18 Cryo-ET reconstruction

Alignment of all tilt series was performed by tracking of gold fiducials using the etomo package in IMOD (Kremer *et al.*, 1996). Tomograms reconstruction was performed either using the WBP method as implemented in IMOD or SIRT in tomo3d (Agulleiro & Fernandez, 2011).

2.19 Subtomogram averaging of CupE pili and TasA fibres

STA of CupE pili on cells and TasA fibres in bundles was performed using RELION4 (Zivanov *et al.*, 2022), employing the in-built helical reconstruction algorithm (He & Scheres, 2017). In order to prevent any reference bias, a cylindrical density created using *relion_helix_toolbox* was employed as a reference. Helical symmetry was applied to improve signal-to-noise and alignment of the particles. Consistent with practice in helical reconstruction (He & Scheres, 2017), the final map is respectively unsymmetrised.

2.20 Compressed sensing (CS-TV^2) reconstruction

Reconstruction of CS-TV^2 tomograms was performed by Sean Collins.

Due to the significant computational requirements and the possibility of parallelisation, tomograms containing HBV capsid particles (dataset described in Bharat *et al.* (2015)) or ribosomes (EMPIAR-10045) were not reconstructed in full, but each particle was reconstructed separately. For this, 300x300 px^2 tilt series centred on each particle were created. To this end, 3D coordinates within the tomogram were used to centre particle positions in the tilt image at subpixel precision using the HyperSpy Python package. Reconstructions were then performed using box sizes of 300x300x300 px^3 at a pixel size of 2.17 Å. The background in each particle was determined by selecting an area outside the capsid; intensities were inverted relative to the background value and scaled to a maximum intensity of one. 200 iterations were performed for each reconstruction. Reconstructions took approximately 20 minutes on a local workstation with one GPU. Reconstructions were also performed in a fully parallelised manner on the ARC3 cluster at the University of Leeds, where a P100 GPU node took ~35 minutes for one reconstruction. To remove artifacts of CS-TV2 reconstruction near the particle edge, boxes were cropped to 216x216x216 px^3 in RELION3 (Scheres, 2012; Zivanov *et al.*, 2018).

Cellular tomograms cannot be reconstructed in full due to memory requirements.For the *C. crescentus* stalk, the tomogram was hence split up into boxes of 600x600x100 px^3 or 800x800x100 px^3 that each overlapped by 25 pixels, and were stitched back together into full tomograms using the HyperSpy Python package. For the *C. crescentus* cell, box sizes of 1770x3058x16 px^3 were used.

2.21 Subtomogram averaging of CS-TV^2-reconstructed subtomograms

As previous work on the employed HBV capsid dataset had shown that α-helical densities could be obtained from particles from three tilt series, we performed STA using particles manually picked in IMOD (Kremer *et al.*, 1996) from these tomograms (n=188). Subsets of particles were created at random using the *shuf* functionality of GNU by selecting lines of the respective star files. STA was performed in RELION3 (Bharat *et al.*, 2015; Scheres, 2012). While RELION3 typically creates a per-particle CTF model which accounts for the defocus of each particle separately, CS-TV2 reconstruction was performed using already CTF-phase-flipped tilt series, as otherwise missing or negative contrast transfer may be amplified by the reconstruction algorithm. Thus, a wedge-shaped CTF model was manually created and used for STA of HBV particles in RELION. WBP particles from tomograms reconstructed in IMOD (Kremer *et al.*, 1996) were implemented into the RELION3 pipeline the same manner. Parameters remained unchanged between refinements of CS-TV^2 and WBP subtomograms. As a reference, an atomic model of the HBV capsid (PDB 6HTX) was converted into an EM density using the UCSF Chimera (Pettersen *et al.*, 2004) *molmap* command and heavily lowpass filtered during refinement runs. A hollow sphere with a soft edge was used a mask. STA maps showing secondary structure were filtered to their respective gold-standard resolution, according to a 0.143 cut-off criterion of independent half-maps. All other maps were filtered to 20 Å resolution to enable the visualisation of map quality differences. B-factors for map sharpening were determined manually as estimation of the B-factor based on the Guinier plot (Rosenthal & Henderson, 2003) was found not to be suitable. B-factors were hence chosen to result in increased features within the density while preventing the appearance of high-frequency noise within the structure. CS reconstructions required higher B-factors, likely due to the down-weighing of high frequency information within the reconstruction. Model-vs-map FSC calculations were performed using RELION against a density map generated from an atomic model (PDB 6HTX).

3. Cryo-EM structure of the archaic Chaperone-Usher pilus CupE

3.1 Acknowledgments

This work is published as part of a preprint (Böhning *et al.*, 2022b) and is, at the time of submission of this thesis, in revision at *PLOS Pathogens*. This project is a collaborative effort with the groups of Dr. Vikram Alva (Max Planck Institute of Biology, Tübingen, Germany) and Prof. Alain Filloux (Imperial College London, United Kingdom). The work described was performed by the candidate unless otherwise indicated.

3.2 Introduction

Chaperone-Usher Pathway pili are among the most common adhesins in Gram-negative bacteria (Waksman & Hultgren, 2009). Phylogenetically, they can be divided into three groups: Classical, alternate, and archaic CUP pili (Nuccio & Bäumler, 2007). While most studies have focused on classical CUP pili, there is little known about archaic CUP pili, which are important for biofilm formation in several human pathogens (Giraud *et al.*, 2011; Tomaras *et al.*, 2003). The archaic CupE pilus in *P. aeruginosa* plays a structural role in maintaining the three-dimensional 'mushroom' shape of biofilms.

Previous work in the laboratory by Dr. Nina Sulkowski (Sulkowski, 2021) had shown that the deletion of the phosphodiesterase CupA6 encoded in the *cupA* operon (gene PA2133) leads to overexpression of the *cupE* operon (Figure S1, Appendix). This may be due to the deletion of the phosphodiesterase leading to increases in cyclic di-GMP levels (Kulesekara *et al.*, 2006). This work also demonstrated that a *P. aeruginosa* PAO1 strain with additional knockouts for the genes encoding type IV pilin (*pilA*), flagellin (*fliC*), and the MvaT transcriptional regulator protein that represses CUP expression (Vallet *et al.*, 2004) (overall phenotype of ∆*fliC* ∆*pilA*

Δ*mvaT* Δ*cupA6*) can be used for high-yield purification of CupE pili. Here, the structure of the CupE pilus is presented.

3.3 Results

3.3.1 Cryo-EM structure of the CupE pilus

To purify CupE pili, *P. aeruginosa* PAO1 Δ*fliC* Δ*pilA* Δ*mvaT* Δ*cupA6* was grown on LB agar plates overnight, pili were sheared from the cell surface and collected via PEG precipitation (Methods), and subsequently plunge-frozen for cryo-EM imaging. A cryo-EM dataset was acquired, depicting long, characteristically zig-zag shaped

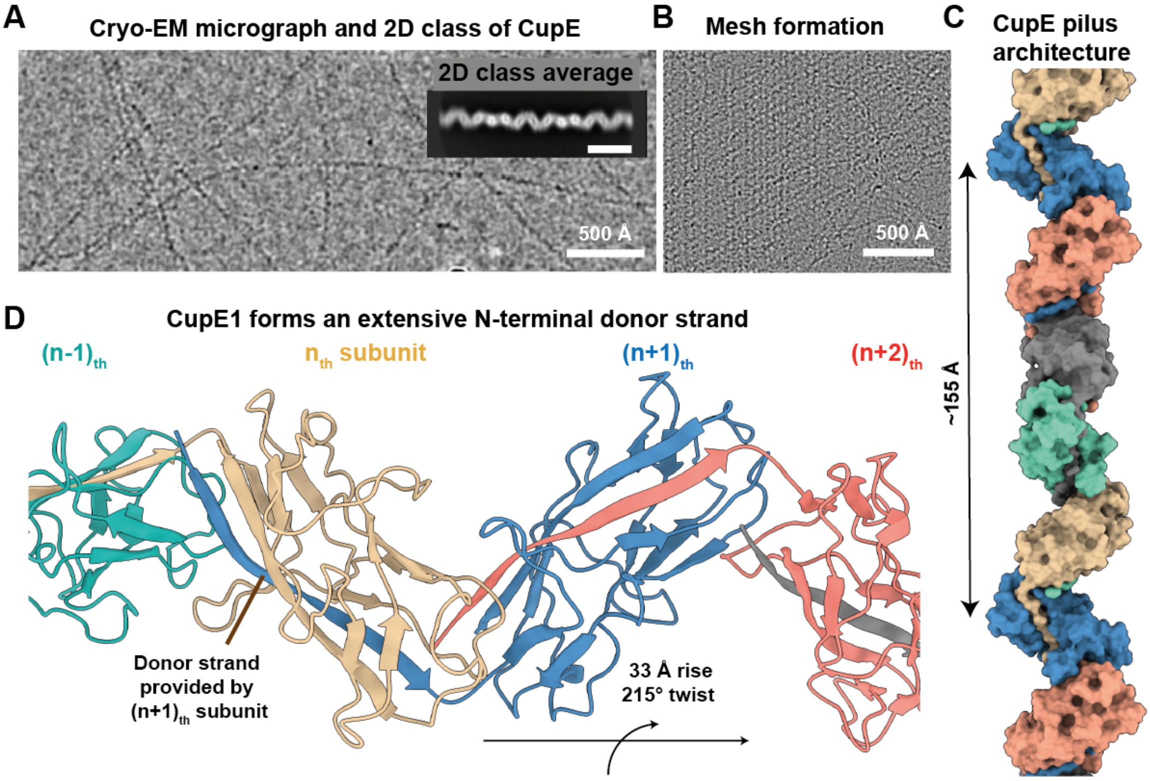


Figure 3.1: The cryo-EM structure of the CupE pilus. A) Cryo-EM micrograph and 2D class average (scale bar: 100 Å) of isolated CupE pili. B) CupE pili were found to be able to form lattice-like arrangements on cryo-EM grids. C) CupE pilus architecture demonstrating the arrangement of CupE1 subunits. A repeat of 155 Å is marked. D) Ribbon depiction of CupE1 subunits within the CupE pilus, showing an extensive donor strand extending into the respectively N-terminal subunit. This data is published (Böhning *et al.*, 2022b).

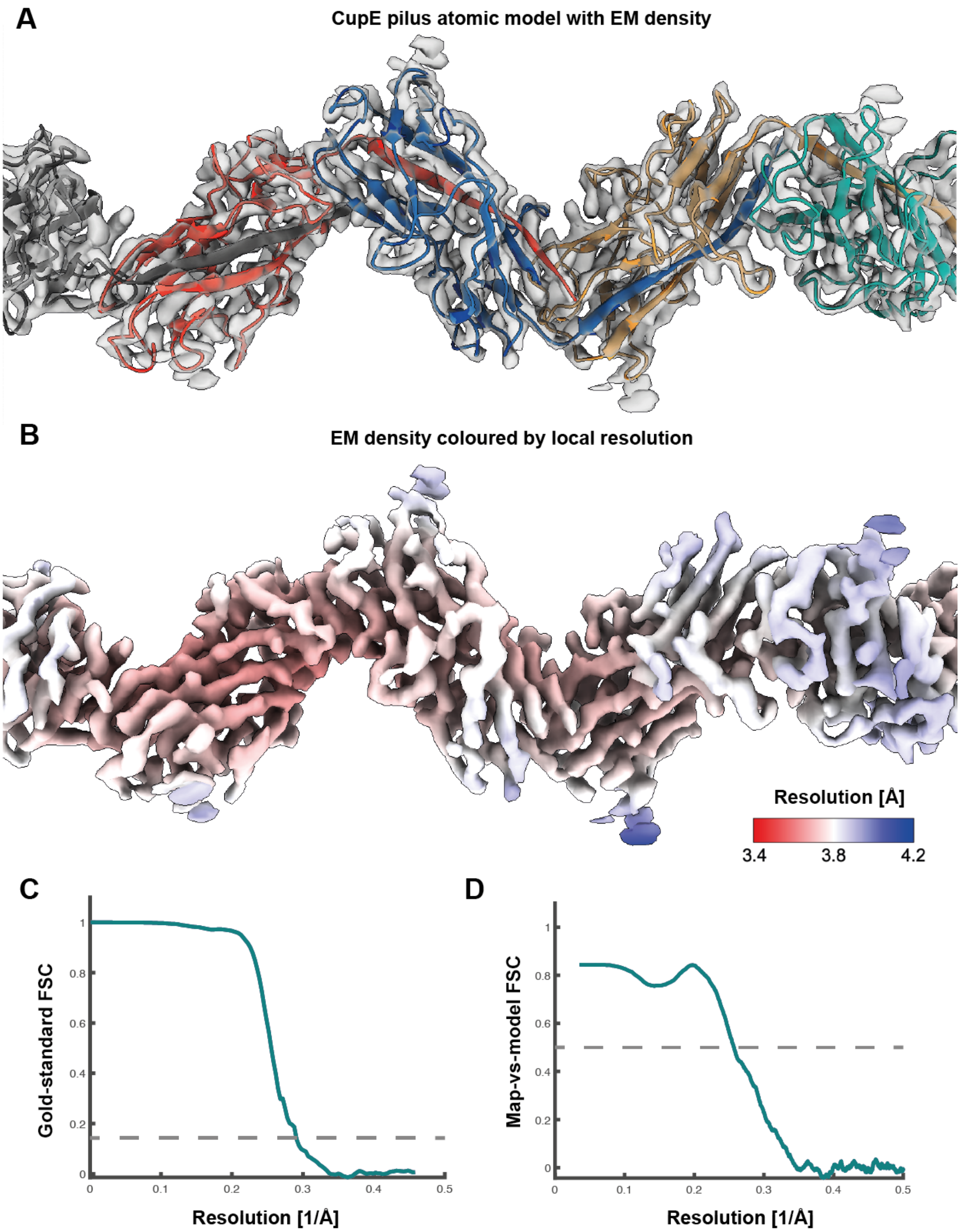


Figure 3.2: Cryo-EM density of the CupE pilus. A) CupE ribbon depiction in cryo-EM density at 15 σ contour level. B) Cryo-EM density coloured by local resolution. C) Masked gold-standard half-map Fourier shell correlation (FSC). The 0.143 criterion is indicated by a dashed line. D) Map-vs-model FSC as an output of Phenix. The 0.5 criterion is indicated by a dashed line. This data is published (Böhning *et al.*, 2022b).

filaments (Figure 3.1 A). Interestingly, CupE was also found to occasionally be forming mesh-like lattice structures on cryo-EM grids (Figure 3.1 B). After deducing the helical symmetry of the pilus from 2D class averages of single fibres (Figure 3.1 A), a cryo-EM density of the CupE pilus was solved via helical reconstruction to 3.5 Å resolution as estimated by Fourier shell correlation (FSC) (Figures 3.1 C-D and 3.2, Table 4, Methods). An atomic model was built into the cryo-EM density, and reveals a zig-zag arrangement of CupE1 subunits (Figure 3.1 C-D), in agreement with previous studies showing CupE1 to be the major subunit of the CupE pilus (Giraud *et al.*, 2011). Similar to previously solved cryo-EM structures of classical CUP pili (Hospenthal *et al.*, 2016; Hospenthal *et al.*, 2017), an N-terminal β-donor strand can be seen extending from each subunit, completing the Ig-like fold of the respectively following subunit (Figure 3.1 C-D, Figure 3.3 A-B). This confirms that donor-strand complementation is also a key feature of the archaic CUP pilus. Apart from the hydrogen interactions mediated by the donor strand with the complemented subunit, there are also significant hydrophobic interactions between hydrophobic residues within the donor strand and a hydrophobic groove of the complemented subunit (Figure 3.3 C, Figure 3.4 A). This way, the donor strand mediates extensive subunit-subunit interactions. The β-sheet from which the donor strand extends contains two cysteine residues (Figure 3.3 D) that form a disulphide bond. Multi-sequence alignment with minor CupE pilins and homologues in other species (performed by Adrian Dobbelstein and Vikram Alva, MPI for Biology) shows that these two cysteine residues are highly conserved (Figure 3.5), suggesting a potential role in maintaining the structural integrity of the pilin subunit.

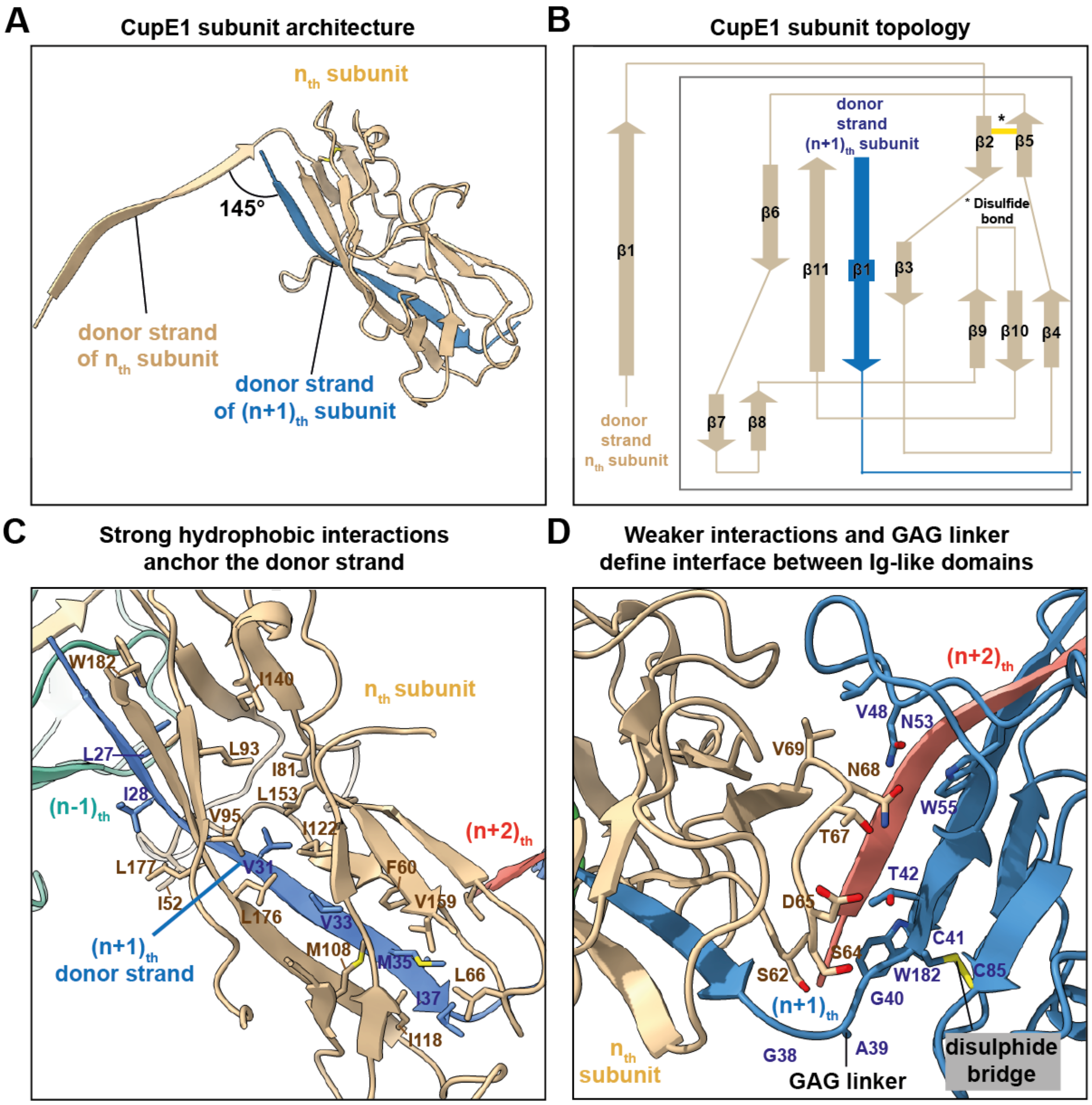


Figure 3.3: Interactions between CupE1 subunits within the CupE pilus. A) Ribbon model depicting the 13-residue donor strand complementing the Ig-like fold of CupE1. B) β-sheet topology of the CupE1 subunit. C) Hydrophobic interactions anchor the donor strand. Hydrophobic side chains in the groove embedding the donor strand are shown. D) Comparatively weaker interactions exist at the interface between Ig-like CupE1 subunits. Two loops containing valine residues V48 and V69 provide the main interaction at the interface between the Ig-like subunits. A Gly-Ala-Gly (GAG) linker motif connects the donor strand with the main part of the subunit. A disulphide bridge can be seen in the β-sheet from which the donor strand extends. This data was published (Böhning *et al.*, 2022b).

Within the pilus, the donor strand provides the majority of inter-subunit interactions, and the interface between the globular CupE1 Ig-like domains has only few strong interactions (Figure 3.3 D); this is particularly in contrast to rod-like classical pilins, which interact with numerous subunits simultaneously (Figure 1.3 B). The main contact at the interface between the Ig-like subunits within the

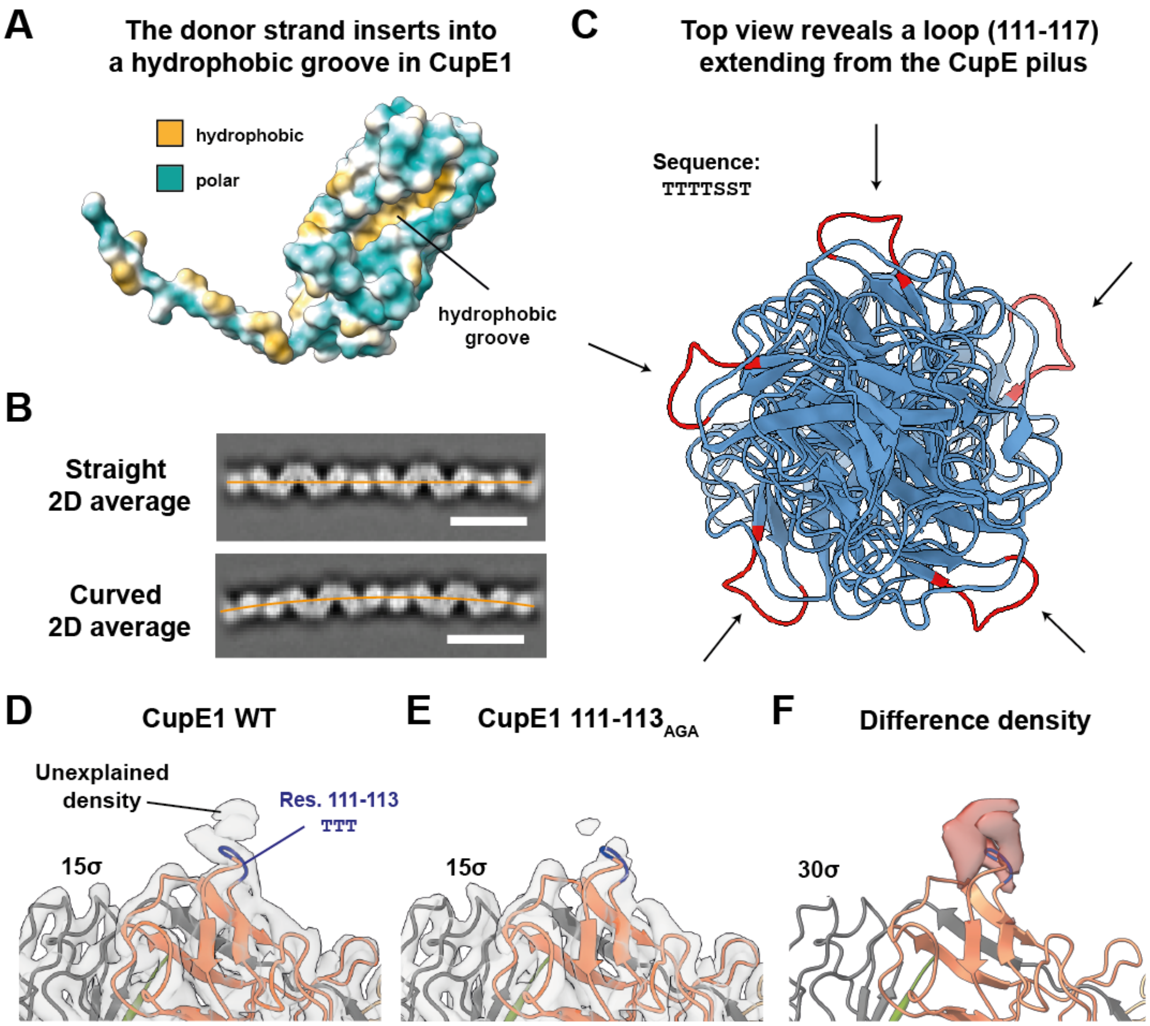
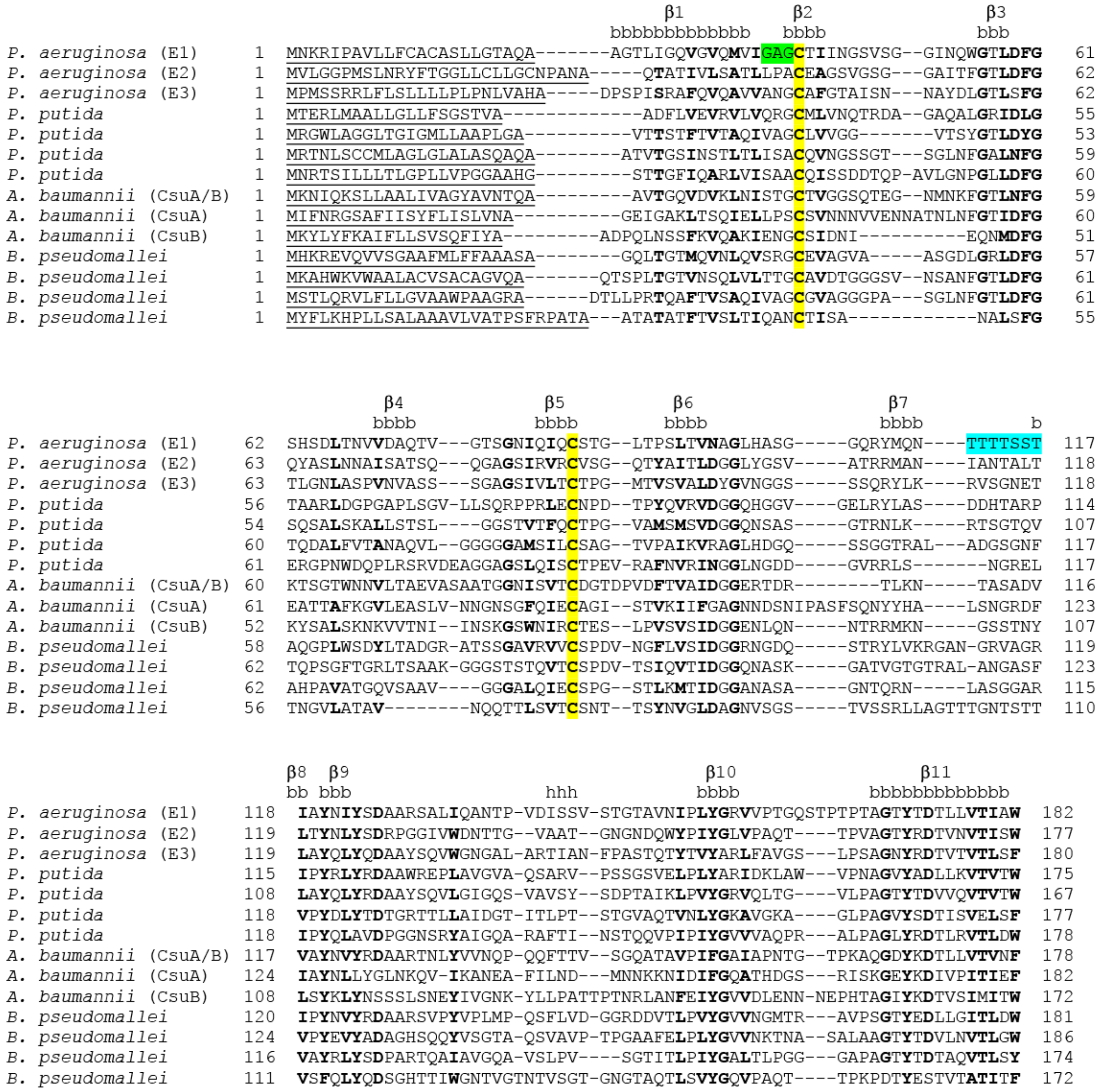


Figure 3.4: Structural features of the CupE pilus. A) Hydrophobic surface depiction of an uncomplemented CupE1 subunit reveals a hydrophobic groove into which the donor strand inserts. B) Two different 2D class averages reveal varying degrees of curvature within the pilus. C) Orthographic top view of a five-subunit model of the CupE pilus reveals a loop (residues 111-117) extending from the pilus. D-F) Comparison of the cryo-EM density of WT CupE1 and 111-113$_{\text{AGA}}$ CupE1 reveals a potential post-translational modification of residues 111-113. This data was published (Böhning *et al.*, 2022b).



```
                                                                 β1          β2                β3
                                                            bbbbbbbbbbbbb   bbbb              bbb
P. aeruginosa (E1)        1   MNKRIPAVLLFCACASLLGTAQA-------AGTLIGQVGVQMVIGAGCTIINGSVSG--GINQWGTLDFG   61
P. aeruginosa (E2)        1   MVLGGPMSLNRYFTGGLLCLLGCNPANA-----QTATIVLSATLLPACEAGSVGSG---GAITFGTLDFG   62
P. aeruginosa (E3)        1   MPMSSRRLFLSLLLLPLPNLVAHA-----DPSPISRAFQVQAVVANGCAFGTAISN---NAYDLGTLSFG   62
P. putida                 1   MTERLMAALLGLLFSGSTVA-------------ADFLVEVRVLVQRGCMLVNQTRDA--GAQALGRIDLG   55
P. putida                 1   MRGWLAGGLTGIGMLLAAPLGA----------VTTSTFTVTAQIVAGCLVVGG-------VTSYGTLDYG   53
P. putida                 1   MRTNLSCCMLAGLGLALASQAQA--------ATVTGSINSTLTLISACQVNGSSGT---SGLNFGALNFG   59
P. putida                 1   MNRTSILLLTLGPLLVPGGAAHG---------STTGFIQARLVISAACQISSDDTQP-AVLGNPGLLDFG   60
A. baumannii (CsuA/B)     1   MKNIQKSLLAALIVAGYAVNTQA---------AVTGQVDVKLNISTGCTVGGSQTEG--NMNKFGTLNFG   59
A. baumannii (CsuA)       1   MIFNRGSAFIISYFLISLVNA----------GEIGAKLTSQIELLPSCSVNNNVVENNATNLNFGTIDFG   60
A. baumannii (CsuB)       1   MKYLYFKAIFLLSVSQFIYA---------ADPQLNSSFKVQAKIENGCSIDNI----------EQNMDFG   51
B. pseudomallei           1   MHKREVQVVSGAAFMLFFAAASA--------GQLTGTMQVNLQVSRGCEVAGVA-----ASGDLGRLDFG   57
B. pseudomallei           1   MKAHWKVWAALACVSACAGVQA-------QTSPLTGTVNSQLVLTTGCAVDTGGGSV--NSANFGTLDFG   61
B. pseudomallei           1   MSTLQRVLFLLGVAAWPAAGRA------DTLLPRTQAFTVSAQIVAGCGVAGGGPA---SGLNFGTLDFG   61
B. pseudomallei           1   MYFLKHPLLSALAAAVLVATPSFRPATA---ATATATFTVSLTIQANCTISA------------NALSFG   55

                                       β4             β5          β6                  β7
                                      bbbb           bbbb        bbbb                bbbb          b
P. aeruginosa (E1)        62  SHSDLTNVVDAQTV---GTSGNIQIQCSTG--LTPSLTVNAGLHASG-----GQRYMQN----TTTTSST  117
P. aeruginosa (E2)        63  QYASLNNAISATSQ---QGAGSIRVRCVSG--QTYAITLDGGLYGSV-----ATRRMAN----IANTALT  118
P. aeruginosa (E3)        63  TLGNLASPVNVASS---SGAGSIVLTCTPG--MTVSVALDYGVNGGS-----SSQRYLK----RVSGNET  118
P. putida                 56  TAARLDGPGAPLSGV-LLSQRPPRLECNPD--TPYQVRVDGGQHGGV----GELRYLAS----DDHTARP  114
P. putida                 54  SQSALSKALLSTSL----GGSTVTFQCTPG--VAMSMSVDGGQNSAS-----GTRNLK-----RTSGTQV  107
P. putida                 60  TQDALFVTANAQVL--GGGGGAMSILCSAG--TVPAIKVRAGLHDGQ-----SSGGTRAL---ADGSGNF  117
P. putida                 61  ERGPNWDQPLRSRVDEAGGAGSLQISCTPEV-RAFNVRINGGLNGDD-----GVRRLS-------NGREL  117
A. baumannii (CsuA/B)     60  KTSGTWNNVLTAEVASAATGGNISVTCDGTDPVDFTVAIDGGERTDR--------TLKN-----TASADV  116
A. baumannii (CsuA)       61  EATTAFKGVLEASLV-NNGNSGFQIECAGI--STVKIIFGAGNNDSNIPASFSQNYYHA----LSNGRDF  123
A. baumannii (CsuB)       52  KYSALSKNKVVTNI--INSKGSWNIRCTES--LPVSVSIDGGENLQN-----NTRRMKN-----GSSTNY  107
B. pseudomallei           58  AQGPLWSDYLTADGR-ATSSGAVRVVCSPDV-NGFLVSIDGGRNGDQ-----STRYLVKRGAN-GRVAGR  119
B. pseudomallei           62  TQPSGFTGRLTSAAK-GGGSTSTQVTCSPDV-TSIQVTIDGGQNASK-----GATVGTGTRAL-ANGASF  123
B. pseudomallei           62  AHPAVATGQVSAAV----GGGALQIECSPG--STLKMTIDGGANASA-----GNTQRN-----LASGGAR  115
B. pseudomallei           56  TNGVLATAV--------NQQTTLSVTCSNT--TSYNVGLDAGNVSGS-----TVSSRLLAGTTTGNTSTT  110

                              β8  β9                                 β10                 β11
                              bb bbb                  hhh           bbbb            bbbbbbbbbbbbb
P. aeruginosa (E1)        118  IAYNIYSDAARSALIQANTP-VDISSV-STGTAVNIPLYGRVVPTGQSTPTPTAGTYTDTLLVTIAW  182
P. aeruginosa (E2)        119  LTYNLYSDRPGGIVWDNTTG--VAAT--GNGNDQWYPIYGLVPAQT----TPVAGTYRDTVNVTISW  177
P. aeruginosa (E3)        119  LAYQLYQDAAYSQVWGNGAL-ARTIAN-FPASTQTYTVYARLFAVGS---LPSAGNYRDTVTVTLSF  180
P. putida                 115  IPYRLYRDAAWREPLAVGVA-QSARV--PSSGSVELPLYARIDKLAW---VPNAGVYADLLKVTVTW  175
P. putida                 108  LAYQLYRDAAYSQVLGIGQS-VAVSY--SDPTAIKLPVYGRVQLTG----VLPAGTYTDVVQVTVTW  167
P. putida                 118  VPYDLYTDTGRTTLLAIDGT-ITLPT--STGVAQTVNLYGKAVGKA----GLPAGVYSDTISVELSF  177
P. putida                 118  IPYQLAVDPGGNSRYAIGQA-RAFTI--NSTQQVPIPIYGVVVAQPR---ALPAGLYRDTLRVTLDW  178
A. baumannii (CsuA/B)     117  VAYNVYRDAARTNLYVVNQP-QQFTTV--SGQATAVPIFGAIAPNTG--TPKAQGDYKDTLLVTVNF  178
A. baumannii (CsuA)       124  IAYNLLYGLNKQV-IKANEA-FILND---MNNKKNIDIFGQATHDGS---RISKGEYKDIVPITIEF  182
A. baumannii (CsuB)       108  LSYKLYNSSSLSNEYIVGNK-YLLPATTPTNRLANFEIYGVVDLENN-NEPHTAGIYKDTVSIMITW  172
B. pseudomallei           120  IPYNVYRDAARSVPYVPLMP-QSFLVD-GGRDDVTLPVYGVVNGMTR---AVPSGTYEDLLGITLDW  181
B. pseudomallei           124  VPYEVYADAGHSQQYVSGTA-QSVAVP-TPGAAFELPLYGVVNKTNA--SALAAGTYTDVLNVTLGW  186
B. pseudomallei           116  VAYRLYSDPARTQAIAVGQA-VSLPV----SGTITLPIYGALTLPGG---GAPAGTYTDTAQVTLSY  174
B. pseudomallei           111  VSFQLYQDSGHTTIWGNTVGTNTVSGT-GNGTAQTLSVYGQVPAQT----TPKPDTYESTVTATITF  172
```

Figure 3.5: Multiple sequence alignment (MSA) of CupE1, CupE2, CupE3 and homologues encoded in *cupE*-like operons in *P. putida*, *A. baumannii*, and *Burkholderia pseudomallei*. Homologous proteins in *P. putida* correspond to UniProt IDs WP_041167793.1, WP_016500513.1, WP_016500512.1, and WP_016500511.1; homologous proteins in *B. pseudomallei* correspond to UniProt IDs WP_004526781.1, WP_004526782.1, WP_004526783.1, and WP_004526328.1. Conserved cysteine residues are marked in yellow. Conserved hydrophobic residues are indicated in bold. The GAG linker motif and threonine- and serine-rich loop are marked in CupE1. Secondary structure is indicated above the sequences. This analysis was performed and visualised by Adrian Dobbelstein and Vikram Alva; published in Böhning *et al.* (2022b).

CupE pilus is mediated by two loop regions (in recent literature termed ‘clinch’ (Pakharukova *et al.*, 2022)) that mainly interact through two valine residues (V48 and V69; Figure 3.3 D). The donor strand is connected to the rest of the subunit by a GAG motif (Figure 3.3 D). The lack of strong interactions between subunits, such as extensive hydrophobic interactions or salt bridges, thus may indicate that comparatively weak contacts at the subunit-subunit interface could be a key feature of the archaic CUP pilus. Indeed, in several 2D class averages, curvature can be seen, indicating some level of flexibility at the subunit-subunit interface (Figure 3.4 B).

CupE1 contains a serine- and threonine-rich loop with an unusual sequence (`TTTTSST`) that extends away from the fibre centre, representing the most outward part of the pilus (Figure 3.4 C). Sequence alignment with other pilins, including the pilins of the related Csu pilus (Nuccio & Bäumler, 2007; Tomaras *et al.*, 2003), reveals that this sequence is unique to CupE1 (Figure 3.5). Significant density adjacent to this unusual sequence could be detected in the cryo-EM map that is not explained by the atomic model of CupE1 (Figure 3.4 D). Threonine and serine residues can be O-glycosylated, a modification seen in several other adhesins (Iwashkiw *et al.*, 2013; Nothaft & Szymanski, 2010). Indeed, substitution of three residues within the loop to AGA (resulting in the sequence 111-`AGATSST`-117; mutant strain generated by Dr. Clamens Thomas and Dr. Kira Glatzel, Imperial College London) followed by structural solution by helical reconstruction (Table 4, Methods) showed a reduced density near this loop (Figure 3.4 D-F), indicating an additional density that depends on the presence of these residues. Mass spectrometric analysis could, in the future, confirm a possible modification of such residues.

3.3.2 Imaging of CupE pili *in situ* using electron cryotomography

CupE pili were previously shown to be essential for biofilm architecture, with knockout of pilins resulting in a significant reduction of biofilm formation, resulting in defects in the mushroom shape of the biofilm during late-stage biofilm development

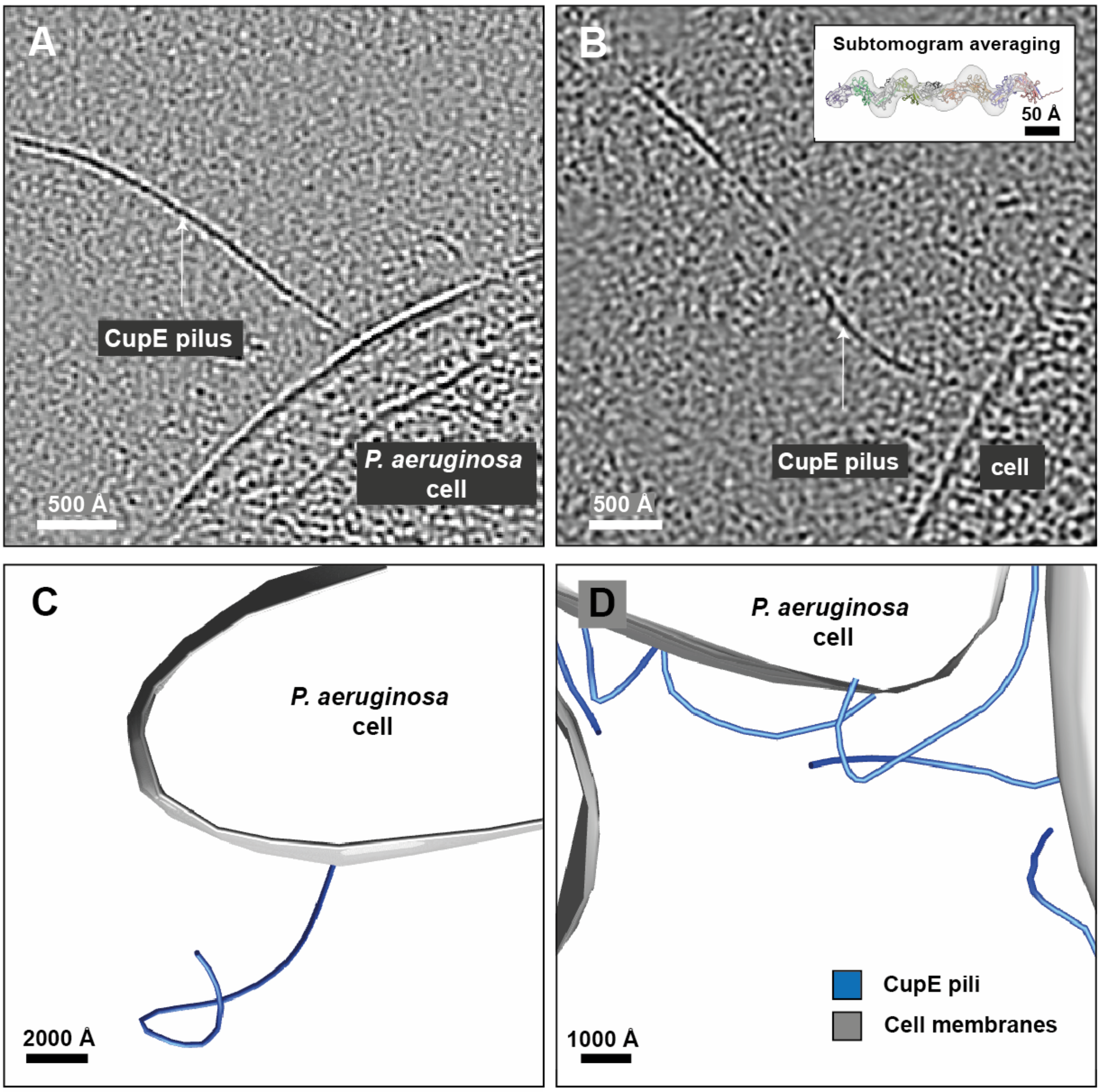


Figure 3.6: Cryo-ET imaging of CupE pili on *P. aeruginosa* cells. A-B) Examples of pili extending from the *P. aeruginosa* cell surface. Inset: Sub-tomogram averaging reveals the pili shown have a characteristic zig-zag architecture consistent with the density of purified CupE. C-D) Segmentation of tomograms depicting CupE pili locally adopting significant curvature and folding back onto cells. This data is published (Böhning *et al.*, 2022b).

(Giraud *et al.*, 2011). In order to determine how the CupE pilus interacts with *P. aeruginosa* cells, we performed electron cryotomography on plunge-frozen *P. aeruginosa* cells, using the same Δ*cupA6* strain with increased CupE production as employed for pilus isolation. In the resulting electron cryotomograms, *P. aeruginosa* cells can be seen with pili extending from the cell surface that are morphologically similar to CupE as observed in previous 2D cryo-EM images (Figure 3.6 A-B). To confirm that the observed pili are indeed CupE pili, we performed STA on the pili using RELION4 (Zivanov *et al.*, 2022) (Figure 3.6 B, inset), confirming that the pili on cells indeed have the same zig-zag appearance observed in our atomic model of the CupE pilus, suggesting they are equivalent.

Beyond previous results showing that the CupE pilus adopted increased curvature in some 2D class averages, CupE pili were found to adopt significant curvature *in situ*, extending away from and back onto cells (Figure 3.6). This ability to undergo significant curvature is contrary to classical rod-like CUP pili, which were previously shown to have a tube-like appearance that can uncoil in the direction of the pilus but is stiff laterally (Hospenthal *et al.*, 2016). Locally enabling curvature may be advantageous for promoting cohesion of biofilms: Given that archaic CUP pili have a length of several micrometres (Pakharukova *et al.*, 2022), a completely stiff pilus would only allow attachment to substrates at a set distance - corresponding to the pilus length - away from the cell; flexible pili, however, can adjust to the complex three-dimensional matrix environment of the biofilm, potentially facilitating adhesion.

3.3.3 Structural modelling of minor CupE subunits

The cryo-EM map of the CupE pilus only resolved the main pilus-forming subunit CupE1. The *cupE* operon encodes two minor pilin subunits, CupE2 and CupE3, that are homologous to the main pilus-forming subunit CupE1 (Figure 3.5) but have an unknown function. Using AlphaFold-Multimer prediction (Evans *et al.*, 2021), we predicted tetramers formed by CupE2 and CupE3 (Figure 3.7) and compared them with predictions for CupE1 (performed by Adrian Dobbelstein and Vikram Alva, Max Planck Institute for Biology). The CupE1 structure solved by cryo-EM was

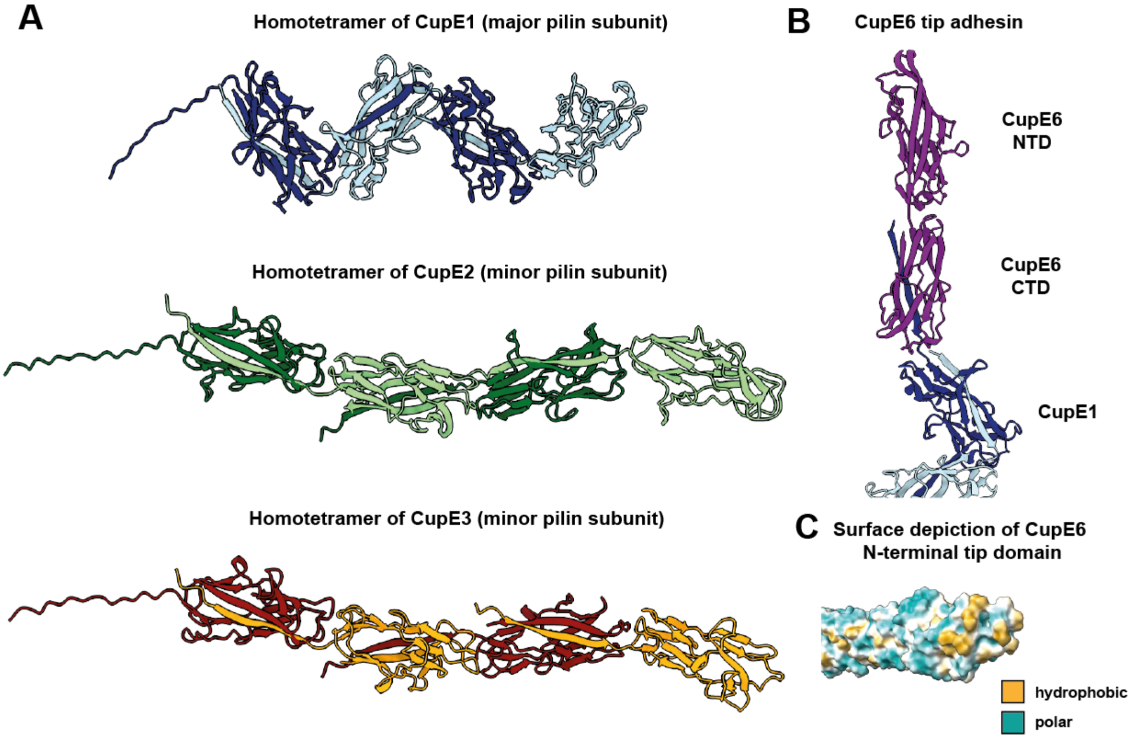


Figure 3.7: AlphaFold-Multimer predictions for minor pilins of the CupE pilus. A) Comparison of filaments formed by four subunits of CupE1, CupE2 and CupE3. All subunits are predicted to form donor-strand-exchanged filaments. B) AlphaFold-Multimer prediction of the CupE6 tip adhesin subunit consisting of a C-terminal domain (CTD) and N-terminal domain (NTD). C) Hydrophobic surface depiction of the CupE6 N-terminal tip domain demonstrating hydrophobic patches. Predictions were created using AlphaFold-Multimer (Evans *et al.*, 2021) by Adrian Dobbelstein and Vikram Alva. This data is published (Böhning *et al.*, 2022b).

accurately predicted by AlphaFold (Figure 3.7 A, RMSD 0.96 Å). Interestingly, both CupE2 and CupE3 were predicted to form filaments through donor-strand complementation, suggesting they may be incorporated into the pilus. While for CupE1, clinch contacts exist that hold the pilus in a zig-zag arrangement, predictions for filaments formed by CupE2 and CupE3 represent a more conventional linear arrangement. The exact role and arrangement of the minor subunits of the archaic CUP pilus remains to be determined. We furthermore predicted the structure of the adhesive tip domain CupE6 (Figure 3.7 B-C); in agreement with predictions, the domain consists of an N-terminal domain (NTD) and C-terminal domain (CTD), whereas the N-terminal domain extends a number of so-called hydrophobic fingers that may interact with hydrophobic substrates, in agreement with previous literature (Pakharukova *et al.*, 2022)

3.4 Discussion

While classical CUP pili are well-characterised, little is known about the structure of archaic CUP pili, with no pilus structures available, despite being the phylogenetically most ancient CUP system (Nuccio & Bäumler, 2007). In this study, a 3.5 Å cryo-EM structure of the archaic Chaperone-Usher Pathway pilus CupE from *P. aeruginosa* is presented. The structural model reveals that the CupE pilus consists of a zig-zag arrangement of CupE1 subunits that are connected by an extensive donor strand that completes the fold of the following, N-terminal subunit next along the filament. Apart from the donor strand, few interactions are found at the interface between the Ig-like domains, predominantly mediated by a region interacting mainly through two valine residues (Figure 3.3 D) that was termed ‘clinch’ in recent literature (Pakharukova *et al.*, 2022). Given that this is the main subunit-

subunit interface apart from the inserted donor strand, it is plausible that this clinch contact maintains the zig-zag architecture of the pilus (Figure 3.1). In electron cryotomograms of *P. aeruginosa* cells expressing CupE, the pilus appears to mostly retain its zig-zag architecture, as shown by STA, but can locally adopt significant curvature (Figure 3.6). High curvature may be enabled by breaking the clinch contact, which, according to our structure, would likely allow the interface between the two subunits to be fully flexible. Thus, this suggests a model where the pilus can exist in two states: In the straight zig-zag confirmation (as solved by SPA and helical reconstruction), where inter-subunit clinch contacts hold the pilus in a zig-zag helical architecture, and which represents the thermodynamically favoured form; or in a state where mechanical stress causes the clinch (V48-V69) interaction to be broken, locally allowing high levels of flexibility. This model can be verified in future efforts by mutation of residues at the interface and observation of the resulting fibre architecture; the before-mentioned valine residues may constitute good targets for this. If the residues contribute significantly to maintain the zig-zag architecture of the pilus, mutation should result in a loss of the zig-zag architecture and reduce stiffness *in vitro* and *in situ*.

Of considerable interest, a high-resolution structure of another model archaic CUP pilus, the homologous Csu pilus in *A. baumannii*, was published very recently (Pakharukova *et al.*, 2022). Consistent with our data, the structure shows a zig-zag architecture similar to CupE, indicating that zig-zag architecture – and the clinch contact – is conserved between the two systems. Most notably, the study includes optical tweezer experiments that show that the Csu pilus can extend up to twice its length if sufficient pulling forces are applied. While we did not perform similar experiments, the inference is similar to our study: it suggests that the clinch interface can be broken, allowing for extension (Pakharukova *et al.*, 2022) or adopting curvature (this study). Notably, the study by the Zavialov group describes the Csu pilus as ‘stiff’, both by appearance in cryo-EM and its behaviour during optical tweezer experiments, suggesting it cannot bend. This is contrary to our cryo-ET imaging of CupE pili on cells, where pili can adopt significant curvature. The Csu and CupE pili are sequence-related (Figure 3.5), suggesting the pili may have similar properties. Compared to our study, however, there is no data on how the Csu pilus behaves on cells, which requires three-dimensional imaging methods to resolve. This leads to two possibilities to resolve this discrepancy: Either the Csu pilus and CupE pilus have different structural properties, and only the CupE pilus can adopt significant curvature; or both pili have similar properties, in which case the Csu pilus behaves differently *in vitro* than in a cellular context, *in situ*. It is also noteworthy that the Csu pilus was studied using a recombinant expression system in *E. coli*, while we use a more native *P. aeruginosa* expression system, and it is not established whether pilus properties are the same when heterologously expressed in *E. coli* as compared to native expression in *A. baumannii*. In both cases, it is furthermore unclear how minor pilin subunits are incorporated: Both the *cupE* and the *csu* operon encode two minor pilin subunits each, but only the main pilin is resolved in each structure. AlphaFold-Multimer predictions suggest CupE2 and CupE3 are filament-forming, suggesting they can be incorporated into the CupE pilus. It hence remains unclear what the role of the minor subunits is, and whether they modulate the structural properties of the pilus itself (i.e., induce curvature, modulate interactions, etc.) or whether they might have a different role in pilus assembly, or termination of such, similar to the PapH subunit in P pili (Båga *et al.*, 1987). Furthermore, while our cryo-EM map shows additional density outside a serine- and threonine-rich loop in CupE1, mass spectrometry will need to be employed to confirm whether this sequence may be post-translationally modified, and what the exact nature of this modification is.

A fundamental question that cannot be addressed currently is how archaic CUP pili interact with the biofilm matrix to provide cohesion to the cells, and how this role differs from classical CUP pili (CupA-C) in *P. aeruginosa*, which were found to co-occur in most *P. aeruginosa* strains (Figure S2, Appendix).

CupE pili are expressed during later-stage biofilm formation, where they were found to maintain the three-dimensional mushroom-shape of the biofilm (Giraud *et al.*, 2011). We (Figure 3.7 C) and previous studies (Pakharukova *et al.*, 2018) provide evidence that the CupE6 adhesin tip subunit, much like the CsuE adhesin subunit (Pakharukova *et al.*, 2018), has hydrophobic patches that are thought to interact with hydrophobic substrates. It is, however, unclear what this hydrophobic substrate may be in the biofilm context. Interestingly, we find some evidence that CupE pili can form mesh-like superstructures through lateral interactions (Figure 3.1 B), suggesting pili can also interact with each other. How these interactions are mediated, and whether they are physiologically relevant, will be the subject of future studies. Ideally, cryo-ET and focused ion beam milling will reveal the molecular architecture of CupE pili within the intact biofilm context and reveal the interaction partners to unravel how exactly adhesion and the maintenance of biofilm architecture is mediated by CupE pili. This would also shed light on how pili utilise curvature within biofilms and whether (and under which conditions) uncoiling of the pilus - a general property shown for Csu - could occur in the biofilm.

4. The structure of Fap fibres in *Pseudomonas*

4.1 Acknowledgment

The plasmid for heterologous expression of *P.* aeruginosa PAO1 FapC-His in *E. coli* (pET31b-FapC-His) was kindly provided by Prof. Morten Dueholm (Aalborg University). AlphaFold predictions shown were performed by Daniel K. Abbas, a Part II student under supervision of the candidate.

4.2 Introduction

Biofilms in the Gram-negative *Pseudomonas* spp. are rich in amyloid-like fibres termed Fap. Numerous so-called functional amyloids had previously been proposed to function in biofilm scaffolding in various bacteria (Taglialegna *et al.*, 2016), but the assignment as amyloids is based purely on secondary hallmarks, such as binding of Thioflavin T (Dueholm *et al.*, 2010). Despite efforts in the field, no structures of bacterial functional amyloids assembling from monomeric proteins have been solved to date. The aim of this study is to advance the structural understanding of Fap fibres from *P. aeruginosa* in order to better understand the role of the EPS of *P. aeruginosa* biofilms.

4.3 Results

4.3.1 Cryo-EM of reconstituted FapC fibres

FapC samples were produced by recombinant expression of the main fibre-forming component FapC from *P. aeruginosa* PAO1 in *E. coli* and purification by immobilised metal ion chromatography (IMAC) under denaturing conditions, producing sufficiently pure protein in the elution fraction (Figure 4.1 A). Fibres were assembled according to previously established methods (Rasmussen *et al.*, 2019) by seeding with fibre fragments (Methods), and cryo-EM samples were prepared therefrom. A resulting cryo-EM dataset depicts fibres with a diameter of ~7-10 nm (Figure 4.1 B-C). Binned 2D averages reveal some features parallel to the fibre axis but little detail on subunit arrangement (Figure 4.1 C). Focusing on the fibre core by reducing the mask size and by using unbinned particles for 2D classification, cross-β-like secondary structure features with a spacing of 4.75 Å can be resolved, although no

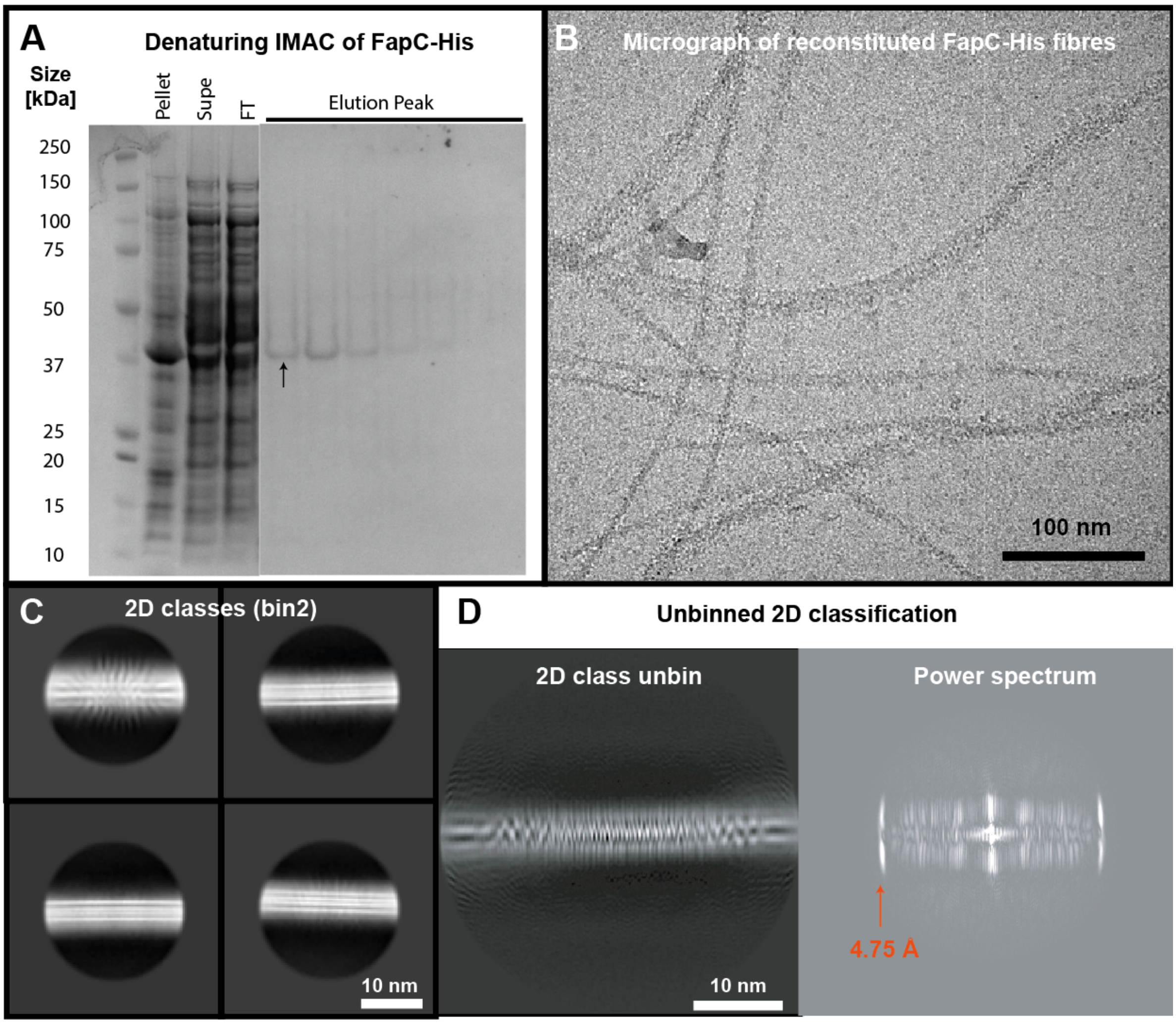


Figure 4.1: Cryo-EM of reconstituted FapC fibres. A) SDS-PAGE of PAO1 FapC-His IMAC, showing the insoluble pellet after cell lysis and centrifugation at 35,000 g, supernatant ('supe') of the same step, unbound protein ('FT') and elution peak of denaturing HisTrap IMAC. One band corresponding to FapC-His is highlighted by an arrow. Protein identity was confirmed using MS. B) Cryo-EM micrograph of reconstituted FapC fibres in 10 mM CHAPS. C) 2D class averages with 2x binning. D) Un-binned class average and amplitude spectrum showing β-sheet features, including a layer line at 4.75 1/Å, but limited other features. No symmetry obtained through Fourier-Bessel methods allowed for structural solution.

distinct other layer lines can be detected that may facilitate determination of symmetry, instead producing a wide range of signals in the power spectrum (Figure 4.1 D). Previous studies have predicted three repeat regions within the protein (Dueholm *et al.*, 2010; Dueholm *et al.*, 2013b), which together with structural modelling (Rouse *et al.*, 2018) was suggested to produce three β-strand layers to the amyloid-like core, i.e., producing a rise per subunit of 3 * 4.75 Å = 14.25 Å. Assuming an amyloid-like character of the fibre, it can furthermore be assumed that the twist per subunit is close to 0°, consistent with previous cryo-EM structures of human amyloids (Scheres, 2020). Performing helical reconstruction with helical symmetries with or near those parameters, however, did not yield a high-resolution map.

4.3.2 Structural prediction of FapC fibres

As greatly improved structural prediction by deep learning methods became available throughout the course of this study, we performed AlphaFold2 and AlphaFold-Multimer (Evans *et al.*, 2021; Jumper *et al.*, 2021) predictions of FapC. The resulting prediction of PAO1 FapC shows an M-shaped amyloid-like core (Figure 4.2 A), predicting each subunit indeed provides three β-sheet ‘layers’ to the fibre. A pronounced L2 linker region is predicted outside the fold for PAO1 FapC (Figure 4.2 A). We also performed predictions for *P. fluorescens* UK4 FapC and *P. putida* F1 FapC (Figure 4.2 A). Interestingly, the predicted amyloid-like core shows a similar architecture, but the L2 region is markedly different, being either very short (UK4) or extremely large (F1) compared to *P. aeruginosa* PAO1.

Within the core, residues that are identical or similar between the repeat regions stack vertically, including hydrophobic, glutamine and asparagine residues, as the repeat regions have high sequence similarity with each other. Using AlphaFold

Multimer, the amyloid-like core of two subunits is predicted to stack vertically (Figure 4.2 B), suggesting this may be the mechanism of fibre formation.

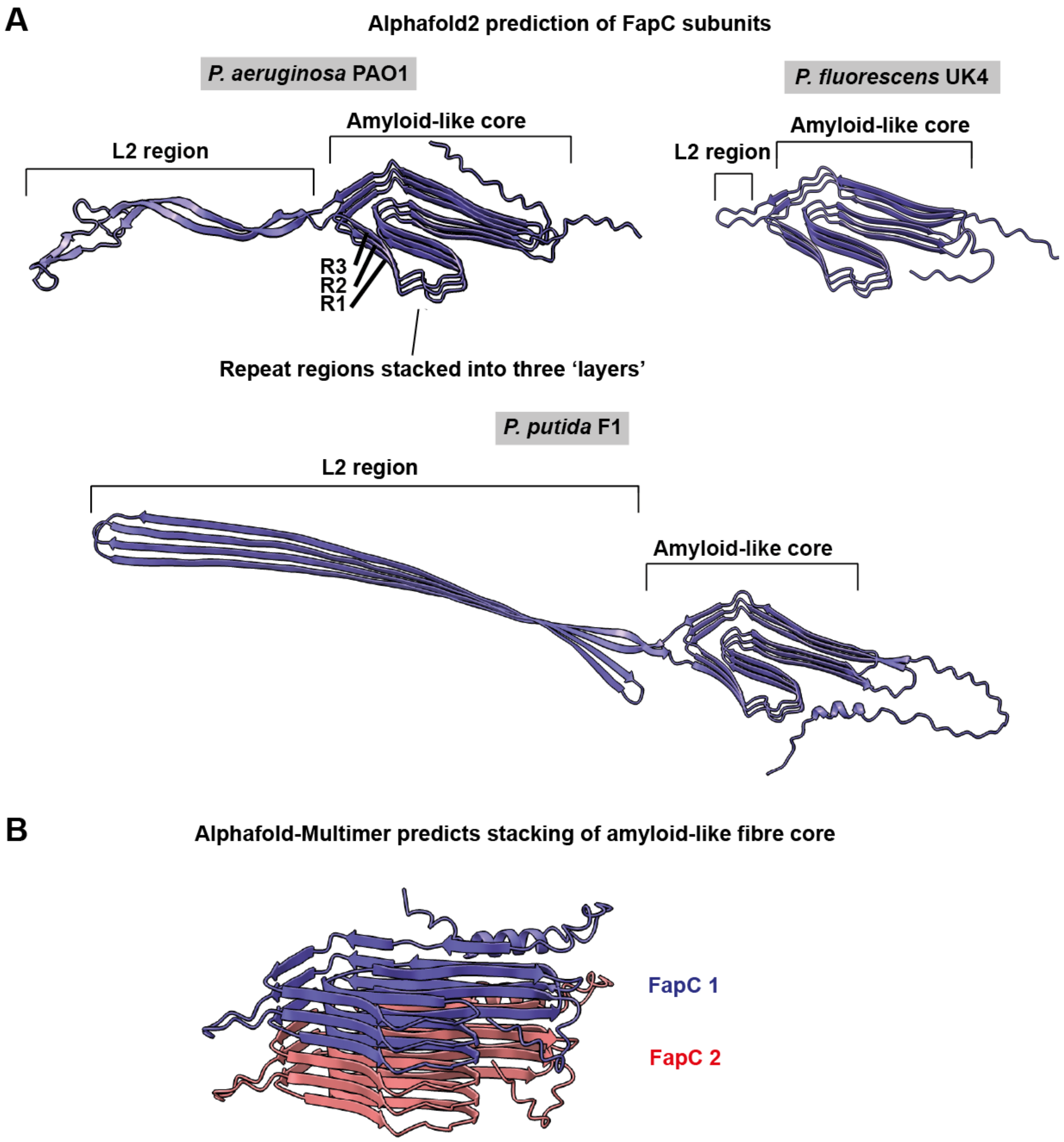


Figure 4.2: AlphaFold prediction of FapC. A) Ribbon diagram for *P. aeruginosa* PAO1 FapC, *P. fluorescens* UK4 FapC and *P. putida* F1 FapC show an amyloid-like core in each case, with the three repeat regions (R1-R3) each forming one layer; the L2 linker region is predicted to extend from the amyloid-like core and varies significantly in size from species to species. B) AlphaFold-Multimer of UK4 FapC suggests that stacking of the amyloid-like core mediates fibre formation.

4.4 Discussion

Fap fibres are a major biofilm matrix component in *Pseudomonas* spp.; however, no high-resolution structures of Fap fibres exist. We performed cryo-EM on reconstituted Fap fibres. 2D class averages resolved β-sheet-like features, supporting the hypothesis that Fap fibres are indeed amyloid (Dueholm *et al.*, 2010); however, no high-resolution structure could be solved. These results similar to a recently published study on the amyloid-like biofilm matrix fibre curli in *E. coli* (Sleutel *et al.*, 2022), where, equally, 2D class averages resolved β-sheets but no high-resolution structures of the fibre could be obtained. AlphaFold modelling predicts an amyloid-like core within subunits where each subunit provides three β-strand repeats perpendicular to the fibre axis. Consistent with current practice in the structural solution of amyloid fibres using cryo-EM (Scheres, 2020), this suggests a rise of ~14.25 Å and ~0° twist per subunit. Why could a high-resolution structure of the fibres not be solved? Firstly, wrong symmetry could preclude structural solution, if the symmetry is not as predicted in literature (Rouse *et al.*, 2018) and by AlphaFold (Figure 4.2), i.e., close to 14.25 Å rise and ~0° rotation per subunit. Assuming this predicted symmetry is correct, a possible other reason may be structural heterogeneity within the amyloid core; the fibres may not be sufficiently locally ordered to be amenable to helical reconstruction cryo-EM, which requires a consistent, symmetric arrangement of subunit (Egelman, 2000). A reason why fibres were not found to be amenable to helical reconstruction may also be the lack of structural features in the β-strand-rich amyloid core, which may prevent successful alignment: When performing cryo-EM on human amyloids, a so-called cross-over can typically be detected readily in micrographs and is used to estimate the initial symmetry (Lövestam *et al.*, 2022; Scheres, 2020); if it cannot be detected, this is usually taken as a sign that the fibre is unamenable to structural solution. Recombinant Fap (Figure 4.1) and curli (Sleutel *et al.*, 2022) fibres do not show such a cross-over. Not having any such distinct low-resolution features may also increase the chance of misalignment by 3D structural alignment and refinement programmes; a small, approximately round fibre is more readily misaligned than a large fibre with strong low-resolution features such as Tau amyloids (Fitzpatrick *et al.*, 2017; Lövestam *et al.*, 2022). A unique problem is also given by the fact that FapC has three very similar but not identical repeats, potentially causing misalignment of the register. A previous study struggled with a similar problem but found a solution to this issue: In the case of bactofilin, a β-helical fibre, nanobodies were used that provided additional lower-resolution features within the 2D class averages (Deng *et al.*, 2019), allowing to discern the helical symmetry more readily, and also providing additional density that facilitates alignment. If structural flexibility indeed is a limiting factor, nanobodies may also help to produce a less flexible, more ordered conformation (Uchański *et al.*, 2020). Understanding how Fap fibres assemble in bacteria will provide significant advances in understanding amyloid-like systems in other bacteria, and may help understanding assembly of disease-related amyloids in human systems.

Finally, the linker domain L2 extends significantly from the amyloid-like core region in PAO1 FapC (Figure 4.1 A) and may thus contribute to structural flexibility precluding structural solution. In this case, it could be considered to employ *P. fluorescens* UK4 FapC, where this linker region is significantly shorter (Dueholm *et al.*, 2010). Why linker regions have different lengths in different *Pseudomonas* species is unclear as their structural function is unelucidated; interestingly, the *E.*

coli amyloid curli fibre protein CsgA, also involved in biofilm scaffolding, does not appear to have any similar region (Sleutel *et al.*, 2022).

Another possibility is that the recombinant fibres employed do not resemble the native fibres accurately, which may introduce artefacts. Isolation of Fap fibres from *P. aeruginosa* could thus be performed, but Fap fibres tend to aggregate heavily and are insoluble (Dueholm *et al.*, 2013b), and it may be difficult to obtain single fibres on cryo-EM grids that are a prerequisite for structural solution by helical reconstruction.

Beyond the fibre structure, in order to fully understand Fap function in *Pseudomonas* biofilms, it will be important to determine how Fap fibres are structurally arranged within the native biofilm and how they interact with the EPS. Fibres were previously found to be aggregative, suggesting they stack to form larger bundles that may play a role in providing stability and hydrophobicity to the biofilm (Dueholm *et al.*, 2010; Zeng *et al.*, 2015). Understanding fibre bundling on a structural level – i.e., how fibres interact to form bundles – as well as how these bundles are arranged in the biofilm will allow for a more complete structural understanding of Fap function in *Pseudomonas* biofilms.

5. Cryo-EM structure of the biofilm scaffold protein TasA

5.1 Acknowledgment

This work is published as part of a preprint (Böhning *et al.*, 2022c) and is, at the time of this submission, in revision at *Nature Communications*. The publication is a collaboration with the groups of Prof. Liraz Chai (Hebrew University of Jerusalem, Israel) and Prof. Syma Khalid (University of Oxford, United Kingdom). The work shown was performed by the candidate unless otherwise indicated.

5.2 Introduction

Fibres formed by the soluble protein TasA are the major component of the *B. subtilis* biofilm matrix, and deletion of *tasA* abolishes the ability to form surface-attached biofilms (Branda *et al.*, 2006). While TasA had been previously shown to form fibrous structures (Romero *et al.*, 2010), it is unclear how TasA subunits arrange into fibres. The most widely cited theory in the field is that TasA forms amyloid fibres, arranging into a cross-β structure (Balistreri *et al.*, 2020; Perrett *et al.*, 2014; Romero *et al.*, 2010; Taglialegna *et al.*, 2016). On the other hand, a dissenting study suggests that TasA subunits instead retain the folded, globular character of the monomeric subunit (Erskine *et al.*, 2018), the structure of which was previously solved (Diehl *et al.*, 2018), and form fibres through head-to-tail arrangement of this globular subunit without structural rearrangement of the subunit itself. Additionally, recent X-ray diffraction studies on *B. subtilis* biofilms *in situ* have only detected weak cross-β patterns (Azulay *et al.*, 2022). Without high-resolution structural information on TasA fibres, it is unclear which of these models is true, precluding a structural understanding of biofilm scaffolding in *B. subtilis*.

5.3 Results

5.3.1 Cryo-EM structure of TasA fibres

To study the structure of TasA, we performed cryo-EM on TasA fibres purified from *B. subtilis* Δ*eps* Δ*sinR* according to previously established procedures (Romero *et al.*, 2010) (obtained from Prof. Liraz Chai, Hebrew University of Jerusalem). The

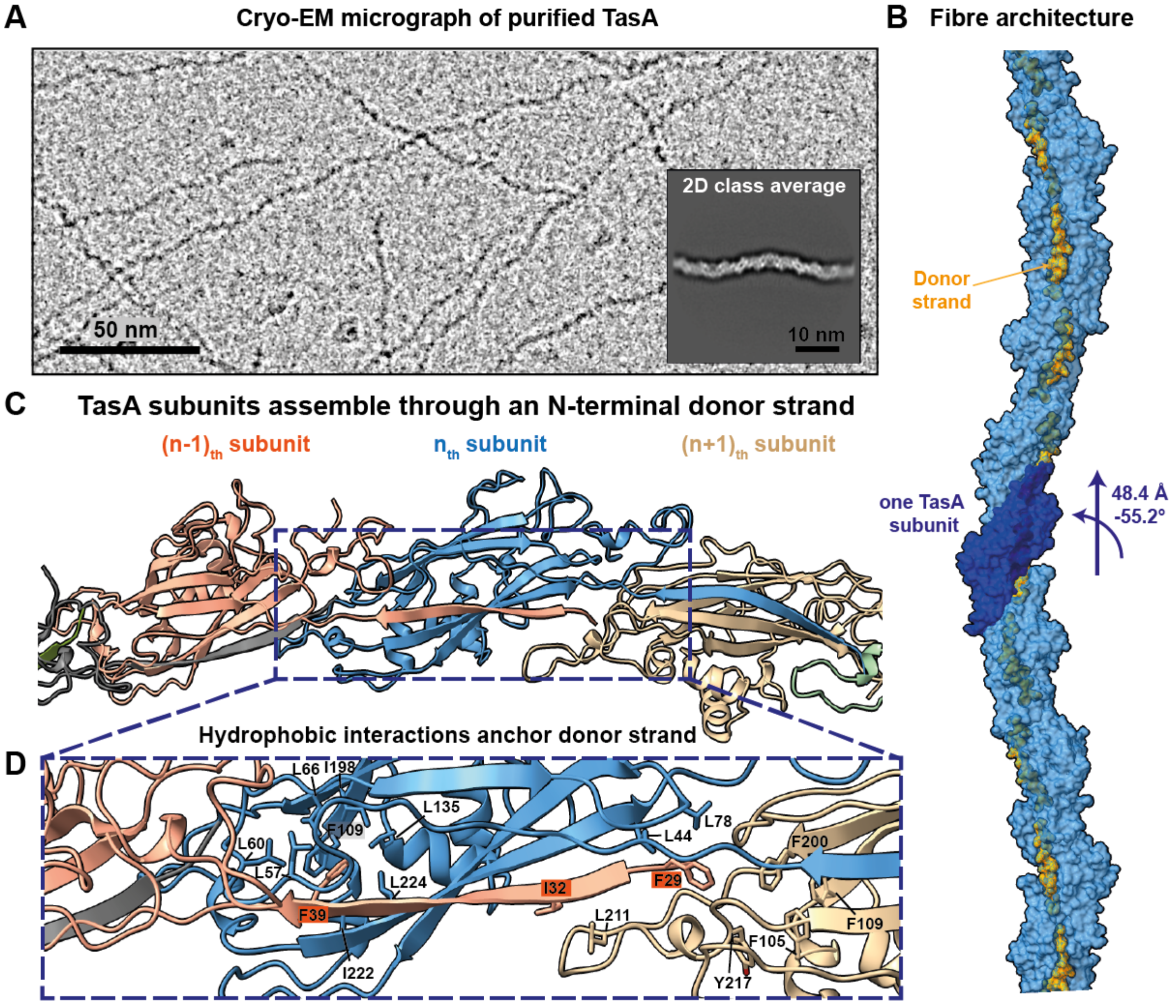


Figure 5.1: TasA cryo-EM structure reveals donor-strand complementation between subunits rather than amyloid formation. A) Cryo-EM micrograph of purified TasA reveals a wave-like appearance. Inset: 2D class average. B) Surface depiction of the TasA fibre, with donor strands marked in orange and one TasA subunit marked in dark blue. C) Ribbon depiction of three TasA subunits, illustrating each subunit extending an N-terminal donor-strand into the respectively $(n+1)_{th}$ subunit. This data is published (Böhning *et al.*, 2022c).

resulting micrographs show a wave-like pattern in individual fibres (Figure 5.1 A), consistent with previous studies (Ghrayeb *et al.*, 2021). 2D averages suggest the presence of individual subunits, showing a ~48 Å repeat that was also detected in previous studies on TasA fibres (Erskine *et al.*, 2018; Ghrayeb *et al.*, 2021). Deducing the symmetry of the fibre from the 2D class averages (Figure 5.1 A), helical reconstruction cryo-EM was performed and resulted in a cryo-EM map at 3.5 Å resolution, and an unambiguous atomic model was built into the density (Figures 5.1., 5.2 and 5.3; Methods; Table 5). Surprisingly, the model reveals that TasA fibres indeed do not have a cross-β structure; instead, subunits interact through an N-terminal donor strand that extends from each subunit into the next, completing a β-

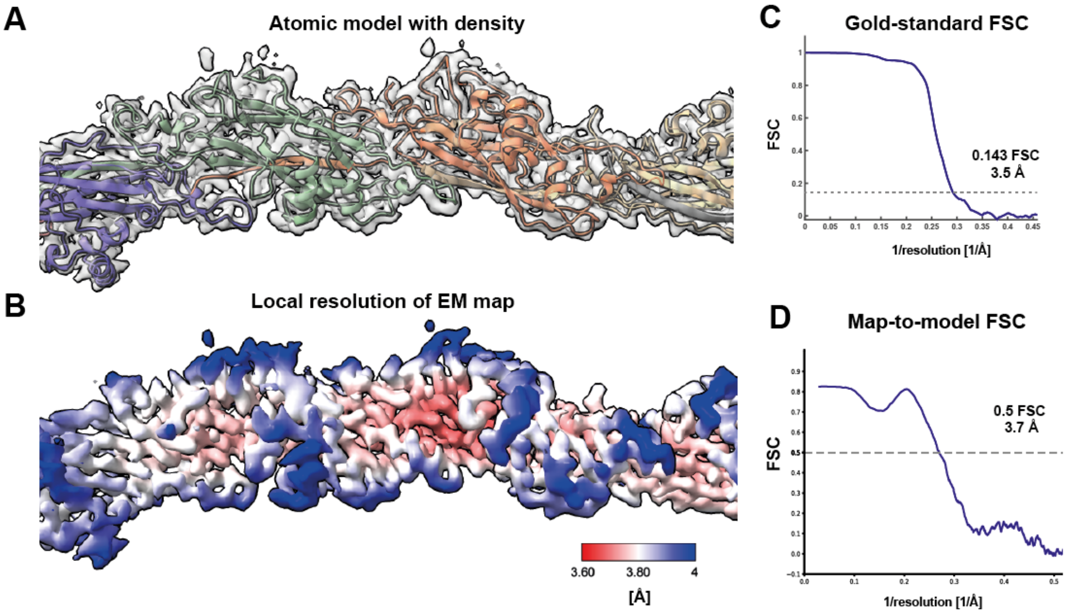


Figure 5.2: Cryo-EM density of TasA fibres. A) Atomic model of TasA fibres shown in cryo-EM density at 12 σ isosurface contour level. B) Cryo-EM density coloured by local resolution estimate. C) Gold-standard half-map FSC plot, with the 0.143 criterion indicated by dashed lines. D) Map-to-model FSC as provided by Phenix. The 0.5 criterion is indicated by a dotted line. This data is published (Böhning *et al.*, 2022c).

sheet within the following subunit, a process previously described as 'donor-strand complementation' (Sauer *et al.*, 2004). The donor strand interacts with the complemented subunit through significant hydrophobic interactions, which are mediated by three highly hydrophobic residues within the donor strand (F29, I32, F39) and embed into a hydrophobic groove within the following subunit (Figure 5.1 D). Beyond interacting with only the following (n+1) subunit, F29 also interacts with hydrophobic residues within the (n+2) subunit (Figure 5.1 D). Such a high degree of subunit interactions may enhance fibre stability, likely an important characteristic for a scaffolding fibre. The C-terminus of the TasA subunit, which was previously

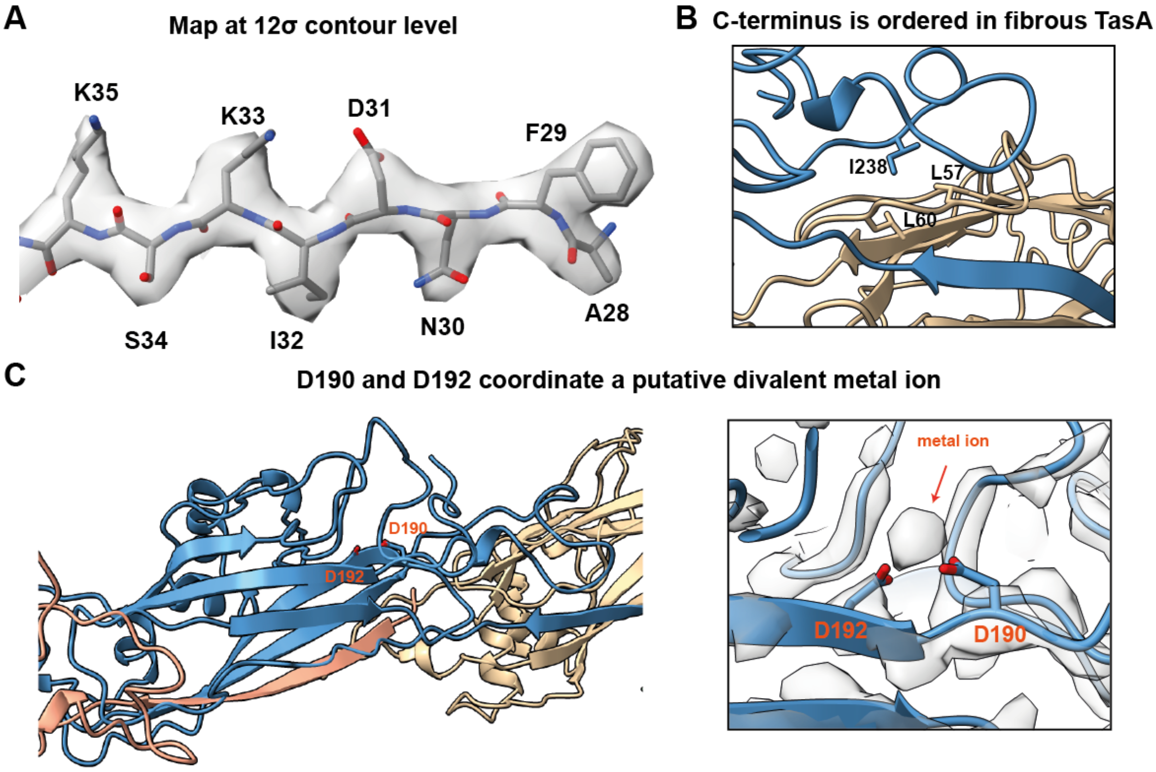


Figure 5.3: Structural features of the TasA fibre model. A) Cryo-EM density of the donor-strand shown in the cryo-EM density at 12 σ. B) While the C-terminus was previously shown to be disordered in monomeric TasA (Diehl *et al.*, 2018). C) The cryo-EM map shows a density for an ion coordinated by D190 and D192. Adapted from (Böhning *et al.*, 2022c).

found to be unstructured in NMR studies in its monomeric state (Diehl *et al.*, 2018), is ordered in the fibre structure and provides another subunit interface (Figure 5.3 B). This is similar to the N-terminal donor strand (residues 28-39), for which no density was found in an X-ray diffraction structure (Diehl *et al.*, 2018), but which mediates subunit interactions in the fibre form (Figure 5.3 B). An ion coordinates by D190 and D192 can furthermore be seen in the structure (Figure 5.3 C).

The structure shows that TasA fibres are neither amyloids, as they do not possess any cross-β structure, nor are they head-to-tail arrangements of globular subunits. Instead, the β-sandwich fold of the TasA subunit stays largely intact, but a donor strand extends to provide β-strand interactions between the individual subunits. Inter-subunit β-strand interactions are the main interaction in amyloid fibres and thought to be causative of their high stability (Eisenberg & Jucker, 2012). The fibre structure does reveal inter-subunit β-strand interactions as in the amyloid model, but shows that the globular fold of the monomeric TasA subunit is largely retained. Given these structural features, TasA fibres are structurally more similar to Chaperone-Usher Pathway (Type I) pili as shown in Chapter 3 than to functional amyloid fibres (such as Fap or curli) in Gram-negative bacteria. Structural comparison with bacterial pilins undergoing donor-strand complementation, which also include the recently characterised Type V pili (Shibata *et al.*, 2020), shows that β-sandwich folds are a prominent feature in subunits undergoing donor-strand exchange (Figure 5.4 A). Eukaryotic proteins shown to undergo β-strand exchange also feature a β-sandwich architecture, including in the Mcg IgG protein involved in light-chain amyloidosis (Brumshtein *et al.*, 2014) (Figure 5.4 B) and uromodulin (Stanisich *et al.*, 2020). This indicates that β-sandwich architectures may be especially amenable for donor-strand exchange to form fibrous assemblies.

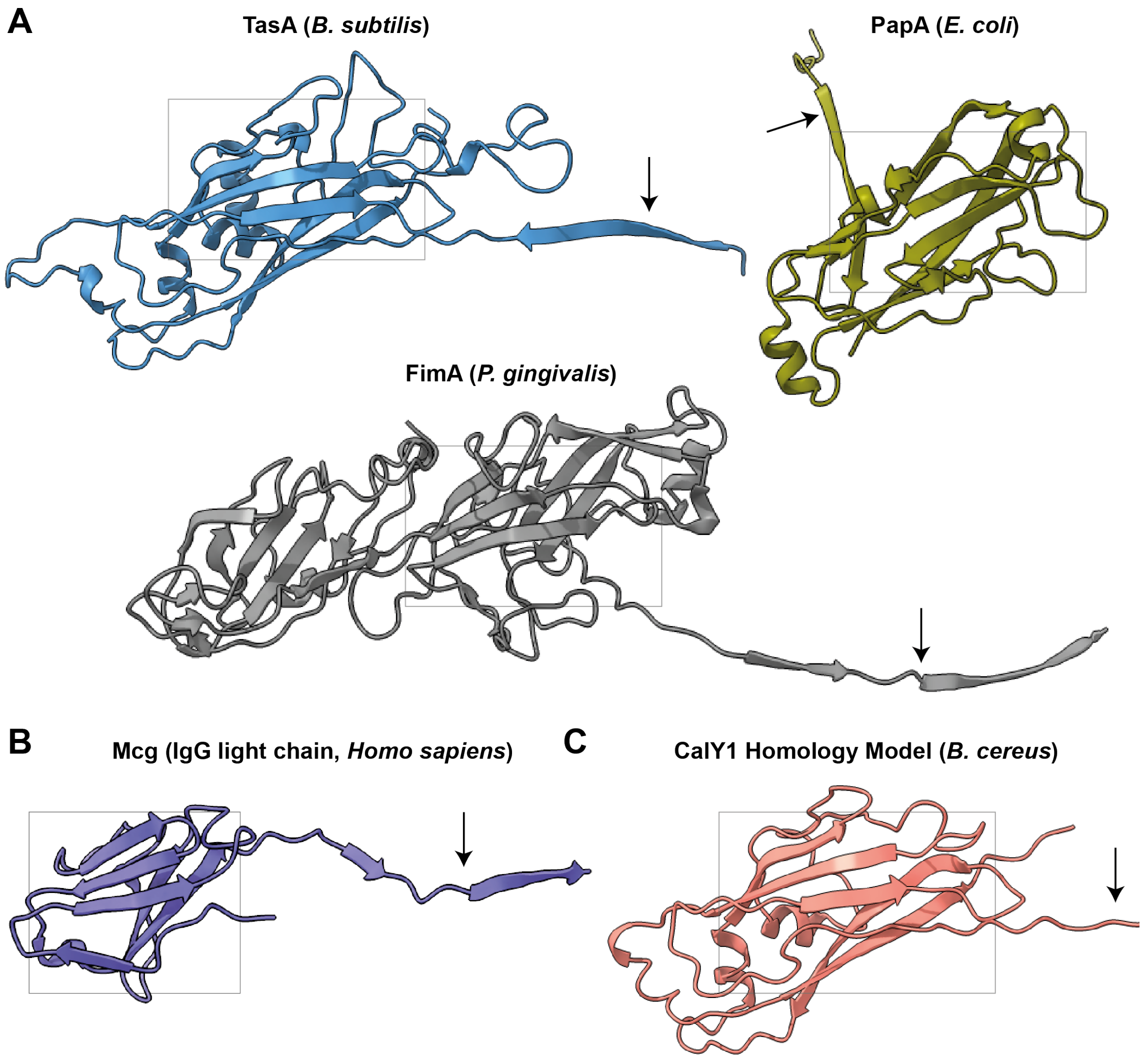


Figure 5.4: Comparison of fibrous TasA with other proteins undergoing donor-strand complementation. A) Comparison of ribbon depictions of TasA with the type I pilin PapA of the P pilus (PDB 5FLU) and the FimA pilin of a type V pilus from *Porphyromonas gingivalis* (PDB 6KMF). The β-sandwich fold is respectively framed. B) Donor-strand-complemented structure of Mcg IgG light chain protein involved in light-chain amyloidosis (PDB 4UNT). C) Homology model of *B. cereus* CalY1 based on the structure of TasA solved in this study. Donor strands are marked with an arrow. This data was published (Böhning *et al.*, 2022c).

5.3.2 TasA undergoes structural transitions upon fibre formation

Donor-strand-complemented fibres in Gram-negative bacteria have different modes of assembly, owing to the challenges of maintaining stability of the unassembled, uncomplemented subunit. In Gram-negative bacteria, operons belonging to the Chaperone-Usher Pathway encode a significant assembly machinery that also includes a chaperone, which complements and stabilises the incomplete subunit by providing a donor strand prior to assembly (Nuccio & Bäumler, 2007; Sauer *et al.*, 2004). The chaperone then shuttles the subunit to the Usher pore, where it is transferred onto a donor strand of the preceding subunit in the growing fibre. In type V pili, protease cleavage within the monomeric protein triggers structural re-arrangement and donor-strand exchange (Shibata *et al.*, 2020). The mechanism for the assembly of TasA into donor-strand-exchanged fibres, however, is unclear. Apart from its signal protease SipW (Terra *et al.*, 2012), only one accessory protein – TapA - is encoded with TasA. TapA's molecular function has not been previously elucidated but appears to involve nucleating fibre assembly both *in vitro* and *in vivo* (Romero *et al.*, 2011, 2014). TapA was furthermore found to form foci outside the cell wall (Romero *et al.*, 2011), suggesting it might provide connectivity of fibres to the cell wall.

The donor strand covers a hydrophobic groove in the TasA fold that would be exposed if not complemented by a donor strand (Figure 5.1 D). The monomeric subunit, however, was previously found to be monomeric and stable without any accessory protein (Diehl *et al.*, 2018); contrary to CUP proteins, there are also no chaperones encoded with TasA. This raises the possibility that TasA undergoes structural rearrangement when assembling from monomers into fibres. In order to analyse whether this is the case, we compared a previously solved crystal structure of monomeric TasA (PDB 5OF2) (Diehl *et al.*, 2018)with the cryo-EM fibre structure (Figure 5.5). Interestingly, the hydrophobic groove that would be exposed in absence of a donor strand is self-complemented in the monomeric form by a different β-strand (Figure 5.5 A). In fibre form, this self-complementing β-strand is displaced by the donor strand of the preceding subunit, which causes the previously self-complementing β-strand to fold over the other β-sheet. This also results in extension of residues 39-47, and presents the donor strand, allowing for assembly of the next subunit.

5.3.3 Structural predictions suggest a role for TapA in fibre nucleation

While this chain-rection mechanism of assembly allows to rationalise how TasA fibres grow, it does not explain how fibre assembly is nucleated. The fact that

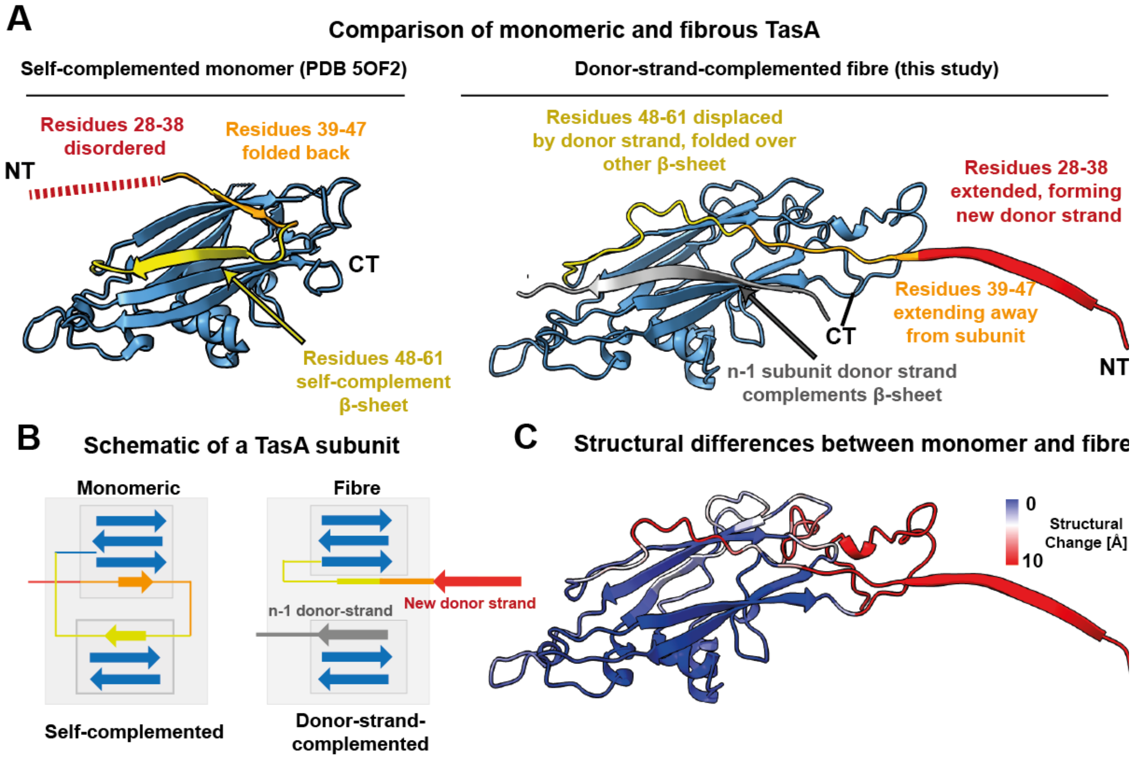


Figure 5.5: Structural re-arrangement of TasA upon fibre formation. A) Comparison of soluble, monomeric TasA (PDB 5OF2, left) and fibrous TasA (this study, right) showing monomeric TasA is self-complemented in its monomeric state. To facilitate visualisation of the re-arrangement of the N-terminus, segments are coloured in yellow (residues 48-61), orange (residues 39-47) and red (residues 28-38). The C-terminus (residues 239-261) in monomeric TasA is disordered (Diehl *et al.*, 2018) and is not present in the crystal structure of the monomer. B) Schematic showing β-strand architecture in the monomeric versus fibrous subunit. C) Structural differences between monomeric and fibre form marked by colour in ribbon depiction of fibrous TasA. This data was published (Böhning *et al.*, 2022c).

extension of the donor strand requires the subunit to be complemented first as it otherwise results in exposure of a hydrophobic groove (Figure 5.6 A) raises the question how the first donor strand in a fibre is provided. In previous studies, TapA was shown to be required for TasA fibre assembly *in vivo* (Romero *et al.*, 2011) and also accelerates *in vitro* assembly of TasA into fibres (Romero *et al.*, 2014). Interestingly, there is significant homology between the donor strand of TasA and

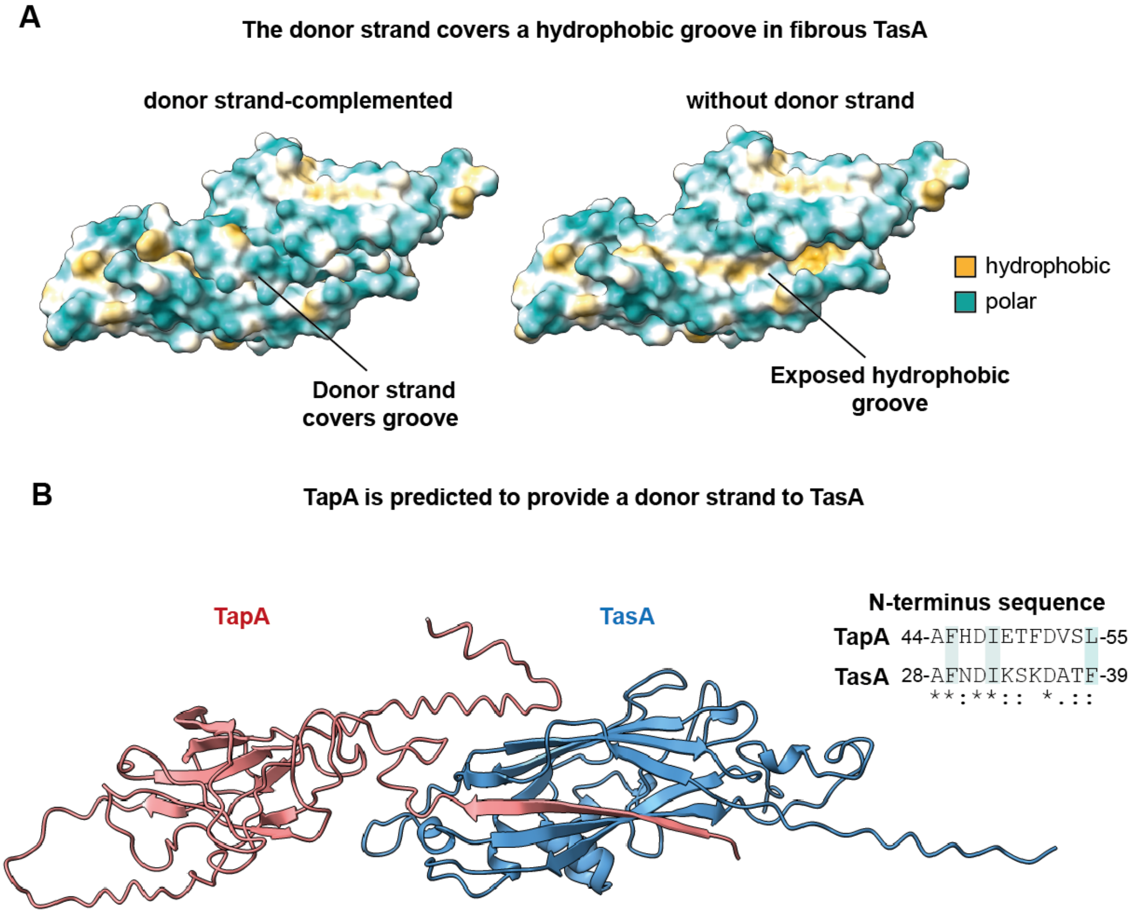


Figure 5.6: TapA may trigger fibre assembly by providing a donor strand to the first TasA subunit. A) If not complemented by another subunit's donor strand, extension of the donor strand by TasA would expose a hydrophobic groove, indicating that TasA may not be able to start forming a fibre without a co-factor. B) ClustalW sequence alignment and AlphaFold-Multimer prediction suggest that TapA extends a donor strand that can complement TasA, consistent with a previously proposed role of TapA in triggering fibre assembly (Romero *et al.*, 2011, 2014). A version of this figure was published (Böhning *et al.*, 2022c).

the N-terminus of TapA (Figure 5.6 B), including conservation of the hydrophobic residues shown to mediate key interactions between subunits (Figure 5.1 D). Since fibre ends are not amenable to helical reconstruction cryo-EM due to their lack of symmetry, we performed AlphaFold-Multimer modelling (Evans *et al.*, 2021; Jumper *et al.*, 2021) to test whether TapA and TasA are predicted to interact. Indeed, the resulting structural model (Figure 5.6 B) shows the N-terminus of TapA forming a donor strand that complements the TasA subunit, which then extends its own donor strand. In addition to previous studies showing a role of TapA in the nucleation of fibre assembly (Romero *et al.*, 2011, 2014), it appears plausible that TapA may trigger the donor-strand exchange reaction of TasA by providing the first donor strand to the fibre.

Indeed, this would be consistent with a previous study that demonstrates that only the N-terminal 57 amino acid residues of the protein – which include the donor strand – are important for maintaining biofilm architecture (Earl *et al.*, 2020).

5.3.4 Cryo-ET of TasA bundles

TasA fibres stack laterally to form bundles that scaffold biofilms. Such bundles were shown to spontaneously form by incubating purified TasA protein *in vitro* (Ghrayeb *et al.*, 2021). In order to obtain a structural view of how TasA fibres interact to form the bundles that are likely important for biofilm scaffolding, we performed electron cryotomography of TasA bundles reconstituted *in vitro* from purified protein (Figure 5.7 A-B). While large bundles could be readily detected, we also detected numerous instances of single fibres and smaller bundles (Figure 5.7 B), indicating that fibre-

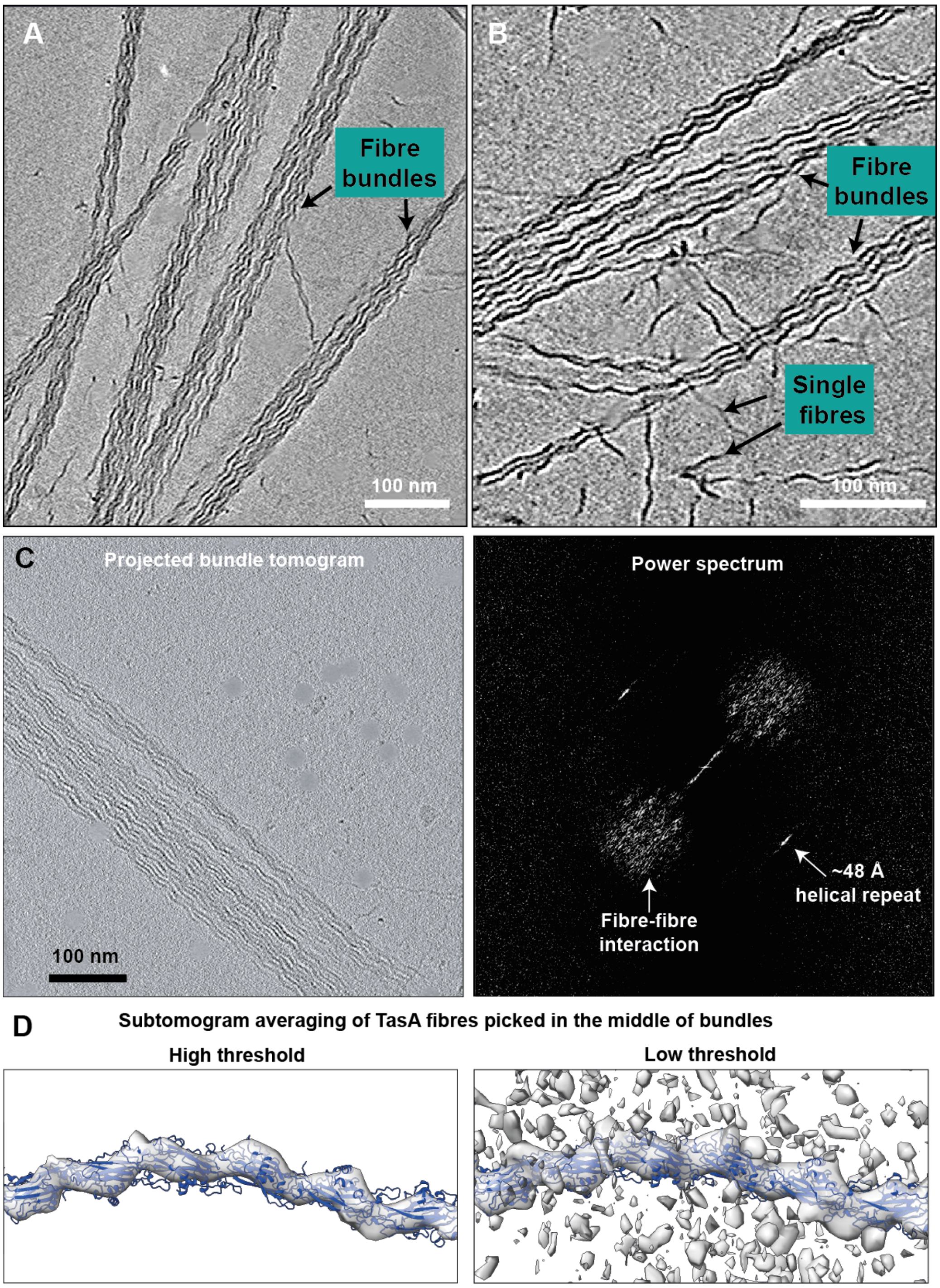


Figure 5.7: Bundle formation by TasA. A-B) Cryo-ET slices of purified TasA, depicting the formation of TasA fibre bundles, with bundles of various sizes and single fibres visible. C) Power spectrum of a projected TasA bundle tomogram shows a distinct layer line for the helical repeat, but no crystalline-like lattice spots for the interaction between fibres, indicating the interaction is not highly ordered. D) Subtomogram averaging of fibres picked in the centre of bundles resolved a single fibre, but no ordered interactions around the fibre. Parts of this data are not currently published but part of the revision for a manuscript currently available as a preprint (Böhning *et al.*, 2022c).

fibre interactions are not high-affinity, consistent with a previous study showing that bundle formation is concentration-dependent (Ghrayeb *et al.*, 2021). The arrangement of fibres is reminiscent of bundle formation of Pf4 bacteriophage, a major matrix component in *Pseudomonas aeruginosa* biofilms (Tarafder *et al.*, 2020). There, Pf4 forms a liquid crystalline matrix where fibres are orientationally aligned but positionally random (Secor *et al.*, 2015; Tarafder *et al.*, 2020). A Fourier transform of TasA bundles suggests, due to a lack of distinct crystalline-like spots for interactions between fibres (Figure 5.7 C), that the interaction between TasA fibres may, equally, not be rigid. Indeed, subtomogram averaging specifically within bundles was able to resolve a single fibre but did not resolve any interacting fibres (Figure 5.7 D), suggesting a non-regular, possibly dynamic, interaction. While we were not able to obtain an atomised model of the interaction, we fitted our single-particle structure into the tomographic density formed by small bundles formed by two TasA fibres ('doublets') (Figure 5.8 A-C). The fit shows that fibres do not interact continuously along the long axis of the fibre, instead periodically interacting every ~32 nm (Figure 5.8 C). A loop which extends away from the fibre centre and contains residues 174-177 may be responsible for fibre-fibre interactions as it represents the element reaching out furthest into this interface (Figure 5.8 D). To determine whether this loop forms part of the interaction, we performed molecular dynamics (MD) simulations on a two-fibre system (performed by Conrado Pedebos

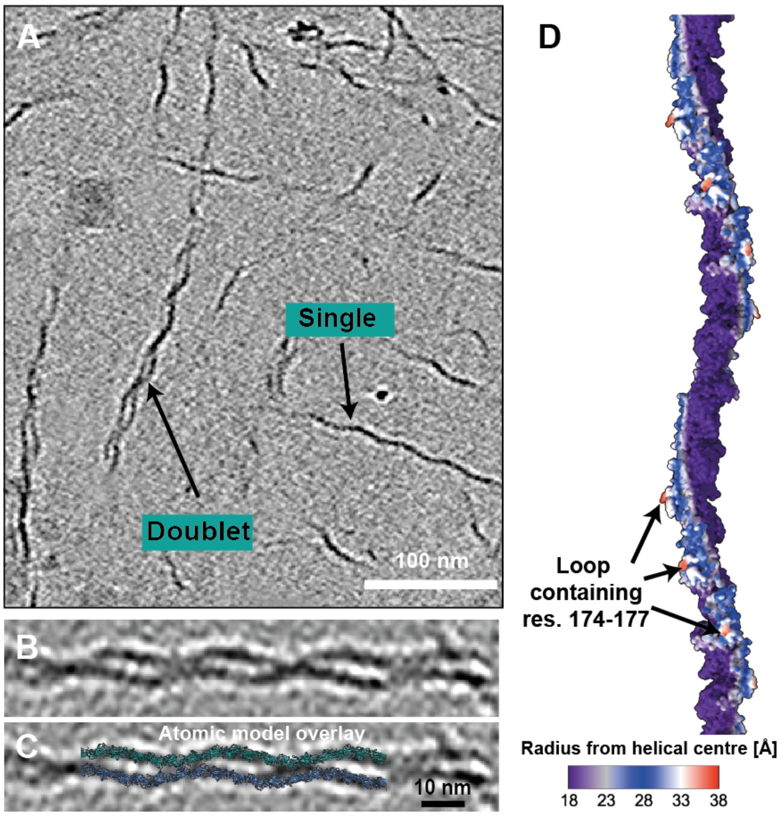


Figure 5.8: TasA fibres interact periodically, with a loop containing residues 174-177 exposed towards the fibre-fibre interface. A) Cryo-ET slice depicting a fibre doublet. B) Closer view of a fibre doublet density. C) The atomic cryo-EM model of a TasA fibre is overlayed with the tomographic density, showing a periodic interaction between fibres. D) Colouring the fibre surface by cylinder radius shows that a loop containing residues 174-177 is the most distal part of the fibre, positioned towards the fibre-fibre interface. Adapted from Böhning *et al.* (2022c).

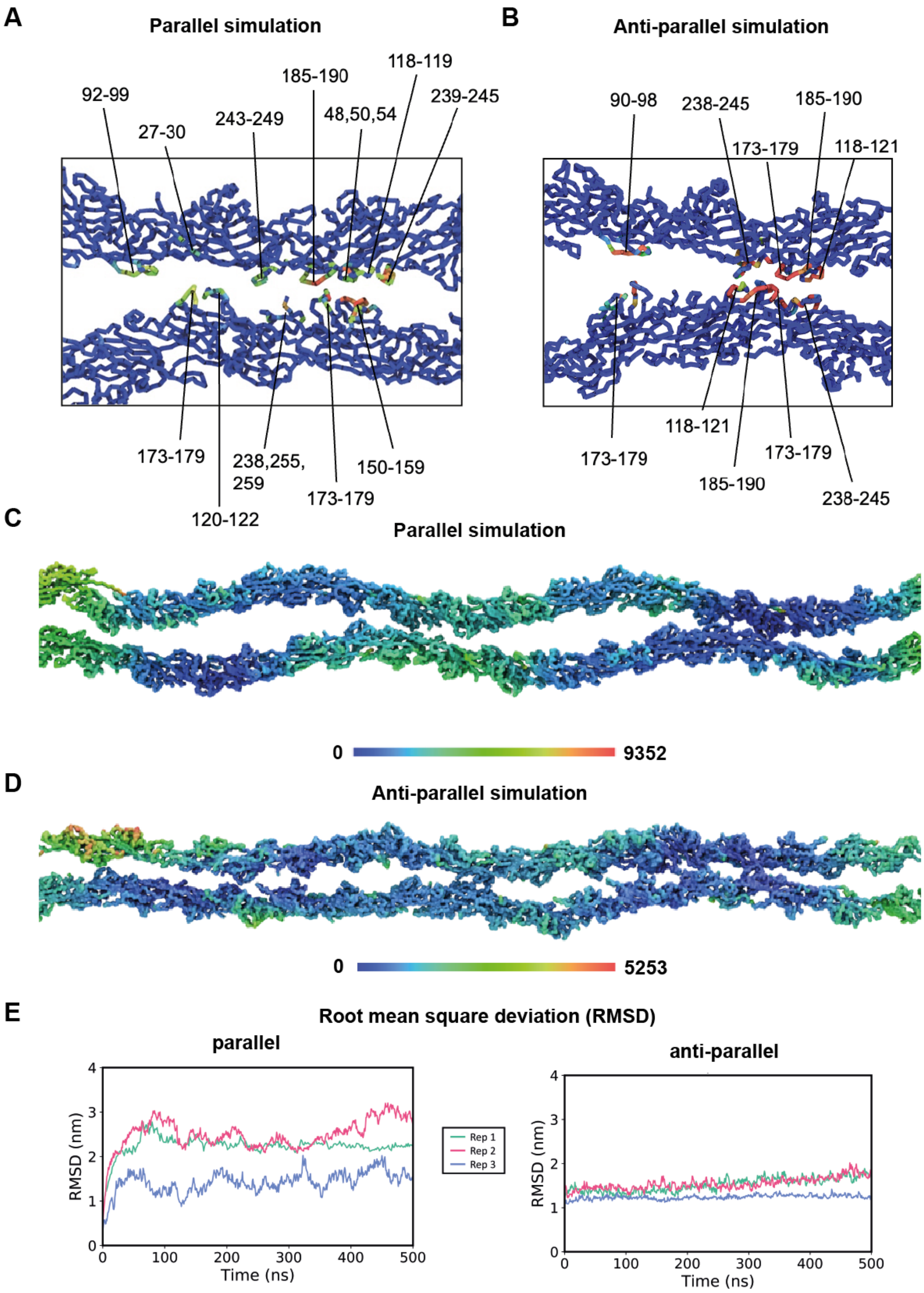


Figure 5.9: Molecular Dynamics simulations of TasA fibre bundles. A-B) Stable solution of MD simulations of an A) parallel and B) anti-parallel TasA fibre doublet. Interactions shown are protein-protein interactions between fibres. Red – strong interactions; green – weak interactions. C-D) MD solution coloured by RMSD. E) RMSD plot for parallel versus anti-parallel simulations. This experiment was performed by Conrado Pedebos and Syma Khalid. This data is not published but part of the revision of a study currently available as a preprint (Böhning *et al.*, 2022c).

and Syma Khalid, Department of Biochemistry; Figure 5.9). The resulting simulations of a parallel and antiparallel doublet of fibres show a stable interaction that consistently includes residues 173-179 (IDGKTAP) (Figure 5.9).

Genomic mutation of residues 174-177 DGKT to AAAA (174-177$_{\text{AAAA}}$; performed by Mnar Ghrayeb and Prof. Liraz Chai, Hebrew University of Jerusalem) in *B.* subtilis resulted in a Δ*tasA* phenotype in pellicle and agar-attached biofilms (Figure 5.10), suggesting that bundling may have been disrupted. Cryo-EM of isolated fibres shows severely reduced bundling; however, this phenotype correlates with a reduced number of fibres visible despite using the same concentration of TasA (Figure 5.10 D-E). This may hint that fibres formed by 174-177$_{\text{AAAA}}$ are less stable, although the reason for this is unclear.

Additionally, we also generated donor strand-defective *tasA* lacking residues 28-38 (Δ28-38, performed by Mnar Ghrayeb and Prof. Liraz Chai, Hebrew University of Jerusalem), which equally shows a defective pellicle and biofilm phenotype, consistent with the role of the donor strand in fibre assembly (Figure 5.10; performed by Mnar Ghrayeb and Prof. Liraz Chai). Indeed, cryo-EM of isolated fibres shows severely reduced fibre formation, with only smaller contaminant fibres being visible in the cryo-EM images (Figure 5.10 D).

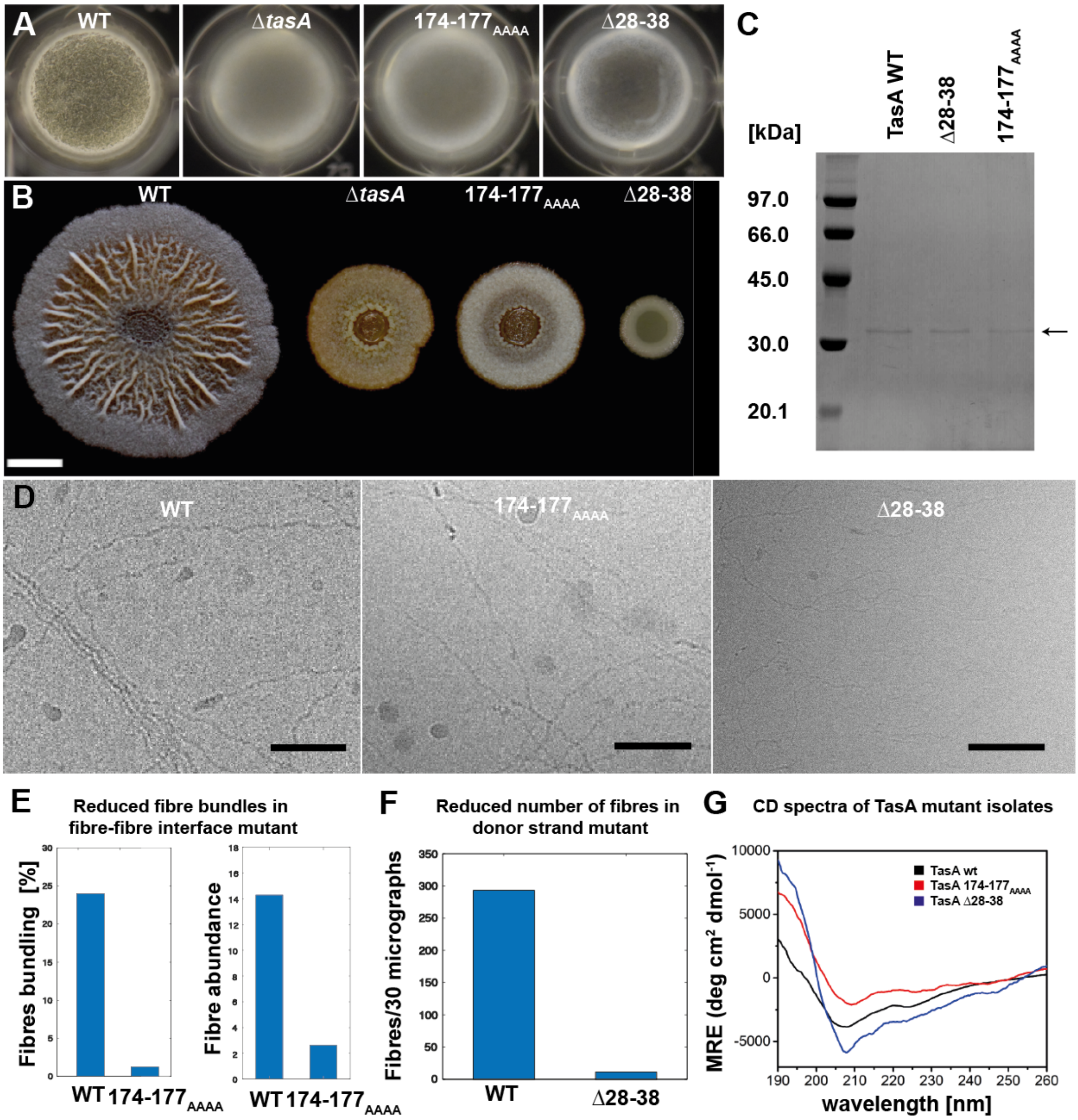


Figure 5.10: Biofilm phenotypes of donor strand (Δ28-38) or fibre-fibre interface (174-177$_{AAAA}$) mutants. A) Pellicle (water surface) biofilms of *B. subtilis* strains WT, ΔtasA, loop mutant 174-177$_{AAAA}$, and donor strand mutant Δ28-38. Well size 0.75 cm. B) Biofilm as in A) but grown on agar. Scale bar is 0.5 cm. C) SDS-PAGE of TasA WT and mutant strains purified from the same number of cells, indicating similar expression levels. The band for TasA is marked with an arrow. D) Cryo-EM of TasA fibres isolated from WT versus mutant strains. Only smaller, contaminant fibres can be seen for the donor strand mutant Δ28-38 strain. Scale bar is 50 nm. E) Quantitation of fibre number and bundling behaviour of loop mutant TasA. Fibre abundance is fibres per micrograph. F) Quantitation of fibre number in Δ28-38 donor strand mutant strain. G) Circular dichroism spectra showing Molar Residue Elipiticity (MRE) of WT and mutant TasA. A-C and E-G were performed by Mnar Ghrayeb and Prof. Liraz Chai. Adapted from Böhning *et al.* (2022c).

5.3.5 Imaging of TasA in disassembled biofilms

We furthermore went on to test whether donor-strand-complemented TasA is the major form of TasA in biofilms. Therefore, we employed an exopolysaccharide-deficient strain with deletion of the biofilm repressor sinR (Δ*eps* Δ*sinR*). Thus, we grew Δ*eps* Δ*sinR* biofilms, resuspended it and plunge-froze it for subsequent cryo-EM. Cryo-EM imaging reveals that the Δ*eps* Δ*sinR* biofilm readily disassembled into cells and bundles of fibres (Figure 5.11). The fibre bundles show the same characteristics as those formed by purified TasA, including the same wave-like appearance of the TasA fibre. 2D class averages of this dataset confirm that the helical arrangement of this biofilm sample is identical to donor-strand-complemented TasA (Figure 5.11 B, inset), strongly suggesting that it is the main component of *B. subtilis* biofilms.

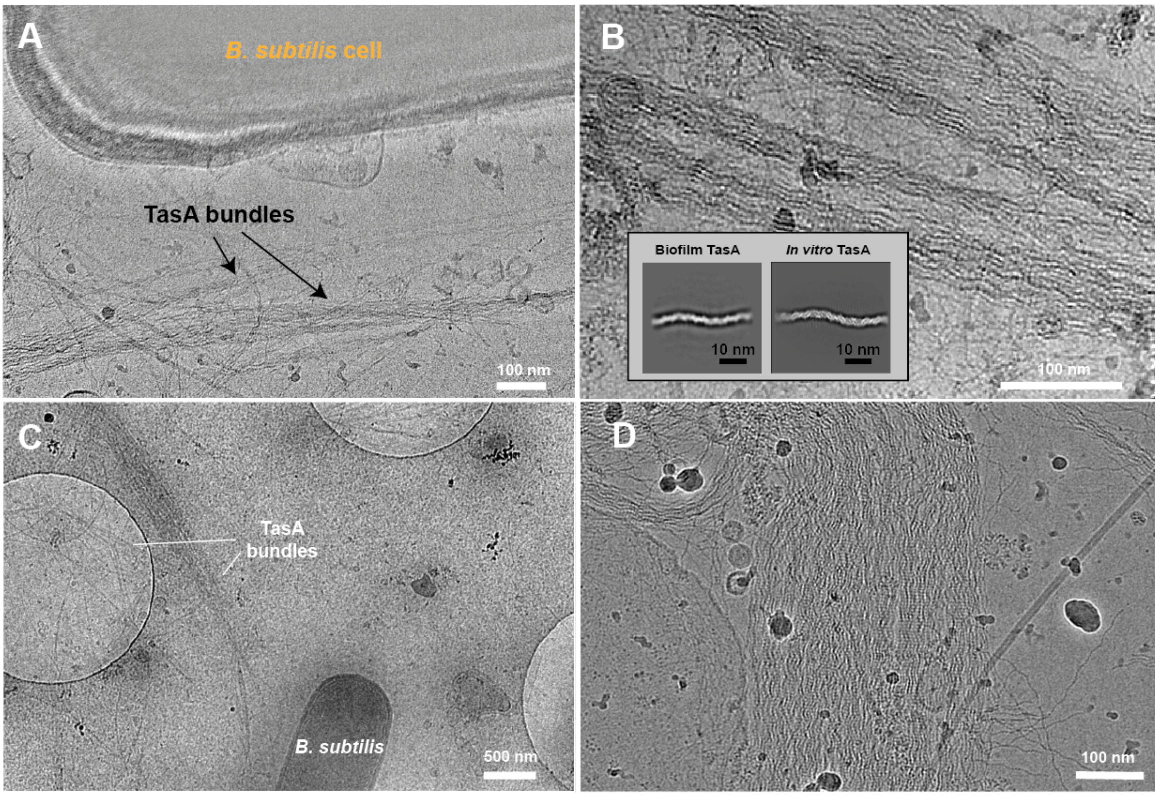


Figure 5.11: TasA fibres in *B. subtilis* ∆*eps* ∆*sinR* biofilms are indistinguishable from purified donor-strand-exchanged TasA. A-B) Cryo-EM images of resuspended *B. subtilis* ∆*eps* ∆*sinR* biofilm. Compared to WT, biofilms of this strain readily disassemble into cells and TasA bundles. Inset: 2D averages of TasA *in vitro* are indistinguishable from 2D averages from this sample, suggesting donor-strand-exchanged TasA is the dominant form of TasA in biofilms. C-D) Cryo-EM images showing the formation of large bundles. Adapted from Böhning *et al.* (2022c).

5.4 Discussion

In the model biofilm-forming Gram-positive bacterium *B. subtilis*, TasA is the major protein component of the biofilm matrix. Previous studies suggested differing architectures for the fibres: One highly cited study suggests TasA fibres are amyloid (Romero *et al.*, 2010), with a cross-β architecture; another study provides evidence for a model where subunits join head-to-tail to form fibres without structural rearrangement (Erskine *et al.*, 2018). In this study, a 3.5 Å structure of TasA fibres was presented, showing that, instead, structural rearrangement of the N-terminus results in the extension of an N-terminal donor-β-strand into the respectively next subunit, completing its fold and thus forming a fibre. The C-terminus, unstructured in its monomeric form, is ordered in the fibre structure and forms another subunit interface. Hence, β-strands provide interaction between subunits, but the fold of the monomeric TasA subunit stays intact.

To date, the original publication stating that TasA is an amyloid has been cited over 700 times (Romero *et al.*, 2010) and had significant influence on the field, resulting in TasA being cited as an amyloid in numerous reviews and research publication (Balistreri *et al.*, 2020; Cámara-Almirón *et al.*, 2020; Malishev *et al.*, 2018; Oli *et al.*, 2012; Perrett *et al.*, 2014; Romero *et al.*, 2014; Taglialegna *et al.*, 2016). This assessment was based on increased Congo Red absorbance of fibrous TasA compared to monomeric protein, and increased ThT fluorescence upon the polymerisation of fibres (Romero *et al.*, 2010). Furthermore, fibres were found to bind to the A11 antibody that was thought to bind specifically to amyloid oligomers (Romero *et al.*, 2010). This is in contrast to a different study which, using circular dichroism measurements, found that monomeric and fibre form have identical secondary structure and finds no increased ThT and Congo Red fluorescence

(Erskine *et al.*, 2018). From the cryo-EM structure solved in this study, it is clear that TasA is, by its definition, no amyloid, and that structural re-arrangement occurs upon fibre formation. This suggests that indirect structural methods should always be used with caution and underlines the importance of direct methods for protein structural solution. The field of so-called functional amyloids is steadily growing and includes tens of proteins that have not been characterised structurally; indeed, it has been proposed by some in the field that amyloid structure 'is the norm, not the exception' (Balistreri *et al.*, 2020). Our study urges caution about such generalisations based on data that does not include direct structural methods.

This study provides the first example of a donor-strand exchange system in Gram-positive bacteria. Previously characterised systems in Gram-negative bacteria include Chaperone-Usher Pathway (Type I) pili (Sauer *et al.*, 2004) and Type V pili (Shibata *et al.*, 2020); interestingly, subunit folds of the donor-strand-exchanged subunits discovered so far appear to always be β-sandwich-like (Figure 5.4). Gram-negative systems employ different mechanisms to enable donor-strand exchange, using Chaperone-Usher systems (Remaut *et al.*, 2006) and protease-mediated assembly (Shibata *et al.*, 2020), respectively. Given a lack of structural and sequence homology of TasA with Gram-negative systems, this suggests that the donor strand motif has evolved independently. Indeed, donor strands provide extensive inter-subunit contacts, both through β-strand hydrogen bonding and hydrophobic interactions of donor strand residues with the subunit core (see also Chapter 3), and this may be the reason for why fibres were found to be particularly stable. Given that TasA makes up a large part of the *B. subtilis* biofilm matrix (Branda *et al.*, 2006), the stability of the fibre may correlate with stability of the biofilm itself. The hypothesis that donor-strand-exchanged fibres are especially stable is given additional evidence by two recent studies in extremophile archaea (Gaines *et al.*, 2022; Wang *et al.*, 2022); fibres found under such hyperthermic conditions were also found to be donor-strand-exchanged, underlining the success and wide-spread nature of such systems throughout the realms of life. One of these fibres, so-called archaeal bundling pili (ABP) in *Pyrobaculum calidifontis*, show significant homology to TasA, and appear to be involved in archaeal biofilm formation. A mechanism of how biofilm formation by these TasA-like fibres could be mediated is furthermore suggested in the study: fibres extending from different cells bundle with each other, thus providing cohesion between cells (Wang *et al.*, 2022). It has not been conclusively shown whether TasA fibres in *B. subtilis* are anchored to the cell, but it is possible that a similar mechanism exists in *B. subtilis*. Interestingly, even in human filaments that are produced to clot bacteria in the urinary tract (Weiss *et al.*, 2020), subunits were found to polymerise through donor-strand complementation (Stanisich *et al.*, 2020).

TasA fibres form bundles in order to scaffold biofilms. Using cryo-ET, we show that fibres within small bundles interact periodically (Figure 5.8). MD simulations and mutational studies suggest that an exposed loop containing residues 174-177 mediates part of this interaction (Figure 5.9). We, however, were not able to determine an atomic structure of the interaction, and subtomogram averaging failed to recapture any interaction between fibres. This is consistent with a Fourier transform of a bundle showing a wide range of diffuse signal, but no crystalline-like spots, for interactions between fibres (Figure 5.7), indicating there may not be a rigid mode of interaction between TasA fibres. Interestingly, the recent study on a TasA-like system in the archaeon *P. calidifontis* is able to resolve such an interaction through helical reconstruction of bundles (Wang *et al.*, 2022). In this system, small

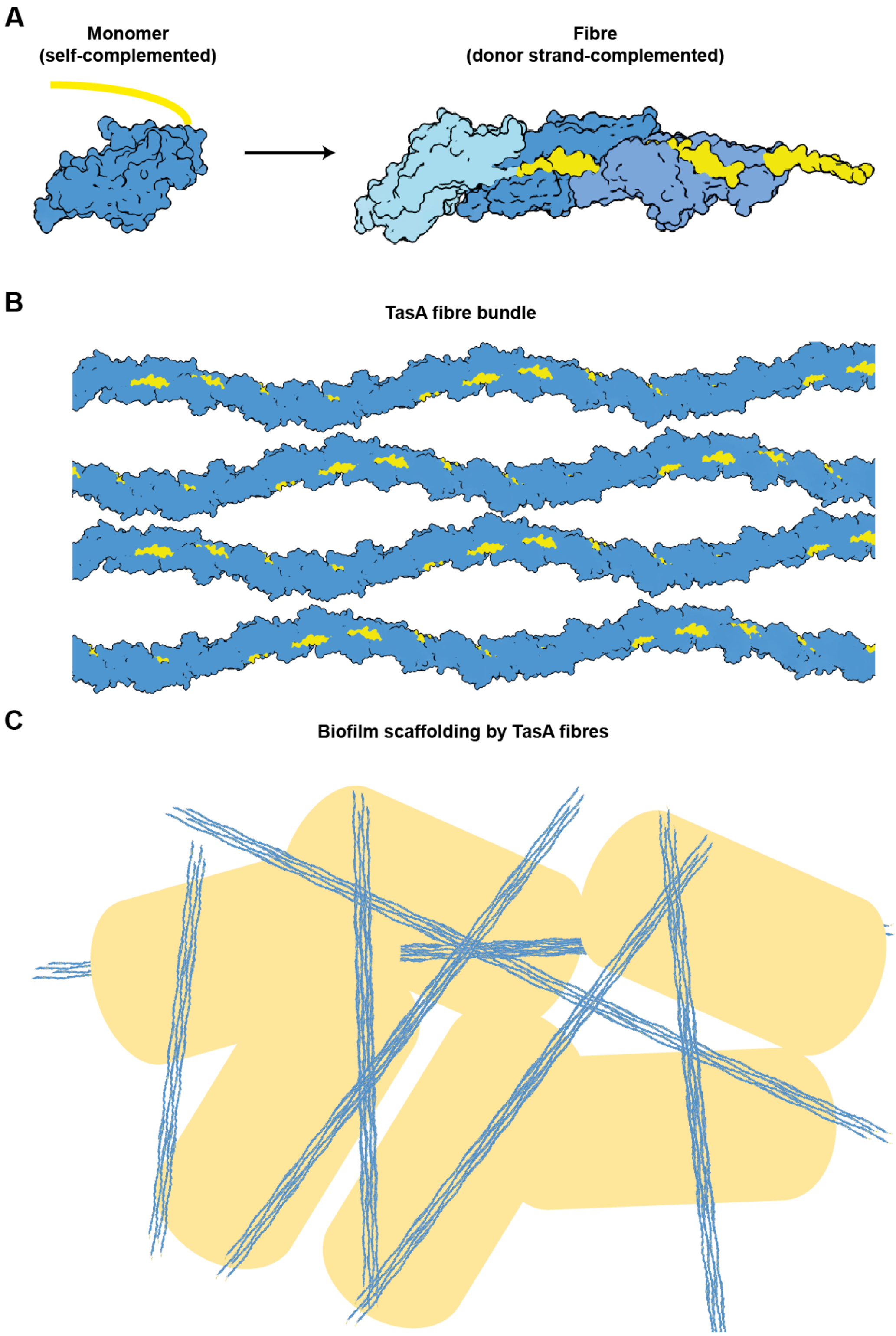


Figure 5.12: Model of biofilm scaffolding by TasA. A) Monomeric subunits are self-complemented, with the donor strand (residues 28-39, yellow) disordered. Upon fibre formation, the donor strand inserts into subsequent subunits along the filament. B) Fibres form bundles through periodic interactions. C) Biofilm scaffolding by TasA fibre bundles as suggested by AFM imaging of *B. subtilis* biofilms (Figure S3, Appendix). Adapted from Böhning *et al.* (2022c).

bundles are formed by five outer fibres interacting with one inner, antiparallel fibre. It is possible that such antiparallel interactions also exist for TasA, but this cannot be conclusively stated at this point; TasA bundles in *B. subtilis* appear to vary significantly more in size. Future studies will hopefully shed further light on the exact mode of interaction.

Interestingly, this study indicates a role for the accessory protein TapA, which in structural models is predicted to complement the TasA subunit (Figure 5.6). Together with previous data (Romero *et al.*, 2011, 2014), this suggests a nucleating function of TapA by providing the first donor strand to the donor-strand exchange reaction. This is consistent with a study showing only the TapA N-terminus is important for biofilm scaffolding (Earl *et al.*, 2020). The role of the TapA donor strand could be experimentally confirmed in future studies through mutation of the TapA donor strand itself, for example, by mutating the conserved hydrophobic residues within. On a functional level, it is still unclear whether TapA anchors TasA fibres to the cell wall as suggested by fluorescence imaging (Romero *et al.*, 2011). Such questions may be answered in the future through FIB-milling and cryo-ET of intact biofilms, which could visualise anchoring to components of the cell wall as shown previously (Melia *et al.*, 2021). It may furthermore elucidate a role of the unstructured C-terminus of TapA, as it is unclear which role – if any – it fulfils in TapA function (Earl *et al.*, 2020). Additionally, insights into the size and arrangement of TasA bundles within the biofilm could be made, allowing to better understand the architectural role of this important structural component.

6. Second-order total variation compressed sensing for cryo-ET

6.1 Acknowledgment

This work is published in *Structure* (Böhning *et al.*, 2022a) and was performed in collaboration with Dr. Sean M. Collins (University of Leeds). The shown work was performed by the candidate unless otherwise indicated.

6.2 Introduction

Electron cryotomography is a popular tool for the visualisation of macromolecular complexes in their native environment. Biological specimens are dose-sensitive, meaning only limited dose can be applied before the sample becomes damaged; thus, electron cryotomograms thus often suffer from low signal-to-noise ratios, hindering direct visualisation and STA. Alternative reconstruction algorithms exist that increase the detail visible by the naked eye within the tomogram but are not amenable to resolving higher-resolution details below the noise level of the tomogram (~3-4 nm) (Wan & Briggs, 2016). Here, a reconstruction algorithm based on compressed sensing second-order total variation (CS-TV^2) is applied, which had previously been successfully employed in cryo-ET in materials science (Collins *et al.*, 2019). We aimed to determine whether CS-TV^2 can reduce noise within subtomograms of biological specimens and whether, despite its denoising effect, it retains high-resolution information.

6.3 Results

6.3.1 CS-TV^2 reconstruction of HBV capsid particles

In order to determine whether CS-TV^2 retains subnanometre information, we aimed to perform STA, a method that is often applied to increase the signal-to-noise ratio and hence retrieve higher-resolution information, on a dataset of HBV capsids that had previously been shown to produce subnanometre resolution maps (Bharat *et al.*, 2015). HBV Particles from three tilt series (n=188) were used as these yielded STA maps with resolved α-helical features in previous studies within the laboratory. For reconstruction, we employed varying regularisation (λ) parameters, which determine the strength of the regularisation term (Figure 1.10). Overall, a higher regularisation parameter would be expected to produce smoother reconstructions; indeed, upon visual inspection of the uncropped reconstructed volumes for three λ values, this is the case (Figure 6.1; reconstruction performed by Sean Collins; all visualisation, STA and analysis performed by the candidate). For the remainder of the study, we chose λ=0.05, as this produced the best results in subsequent experiments. As CS-TV^2 reconstruction produces fringe artifacts near the volume

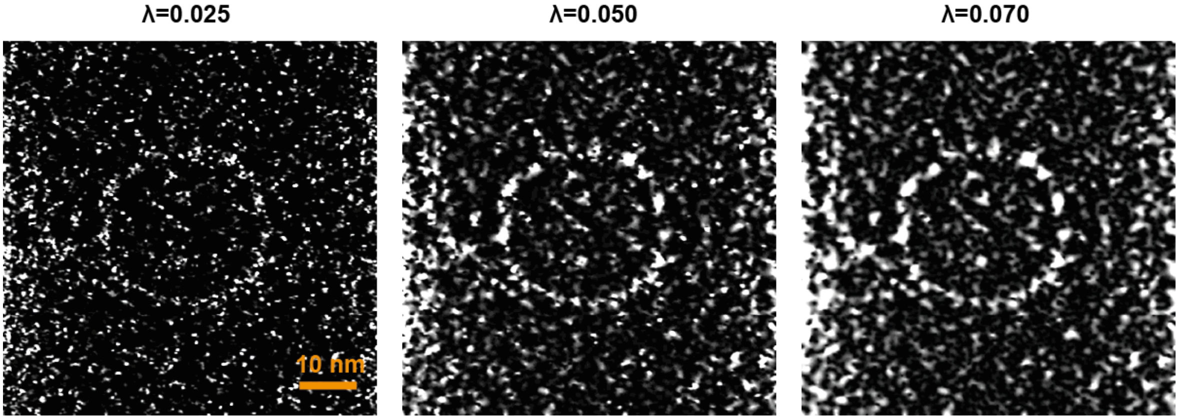


Figure 6.1: HBV capsid particle reconstructed with CS-TV^2 using varying regularisation (λ) parameters. Central XY slices are shown. The particles show reconstruction artefacts at the particle edge, which is why particles were cropped for further processing. This data is published (Böhning *et al.*, 2022a).

edge, particles were cropped in RELION (Bharat *et al.*, 2015) and a spherical mask was applied. Upon comparison of the reconstruction to WBP, the CS-TV^2 reconstruction appears significantly less noisy, and a spherical outline of the HBV capsid can be detected (Figure 6.2 A). This is even more evident upon applying the internal icosahedral symmetry (I2) of the HBV capsid (Figure 6.2 B-C), where the CS-TV^2 reconstruction already resembles the shape of the HBV capsid both in XY-slices and isosurface views, whereas the noise levels in WBP remain very high.

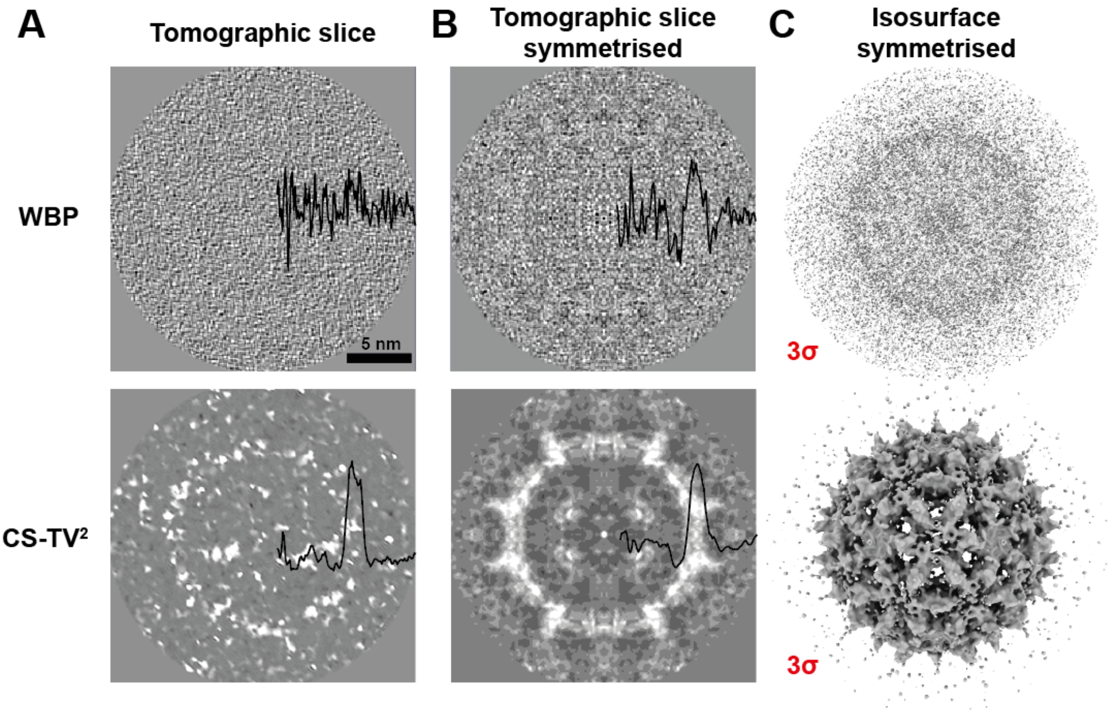


Figure 6.2: CS-TV^2 versus WBP reconstruction of Hepatitis B virus (HBV) T=4 capsids. A) Cryo-ET slice of an HBV capsid reconstructed using WBP (top) and CS-TV^2 (bottom). A normalised radial average of intensities within the tomogram is depicted. B) Internal I2-symmetry of the HBV capsid is applied (240 capsid protein subunits per capsid). Radial intensity average is shown. C) Isosurface of the symmetrised reconstruction shown in B). This data is published (Böhning *et al.*, 2022a).

6.3.2 STA of HBV capsid particles reconstructed with CS-TV^2

Next, we performed STA of the dataset in RELION3, comparing CS-TV^2 reconstructions with WBP (Figure 6.3 A-B). STA maps from the CS-TV^2 dataset showed α-helical densities after B-factor sharpening (Figure 6.3 B). In order to

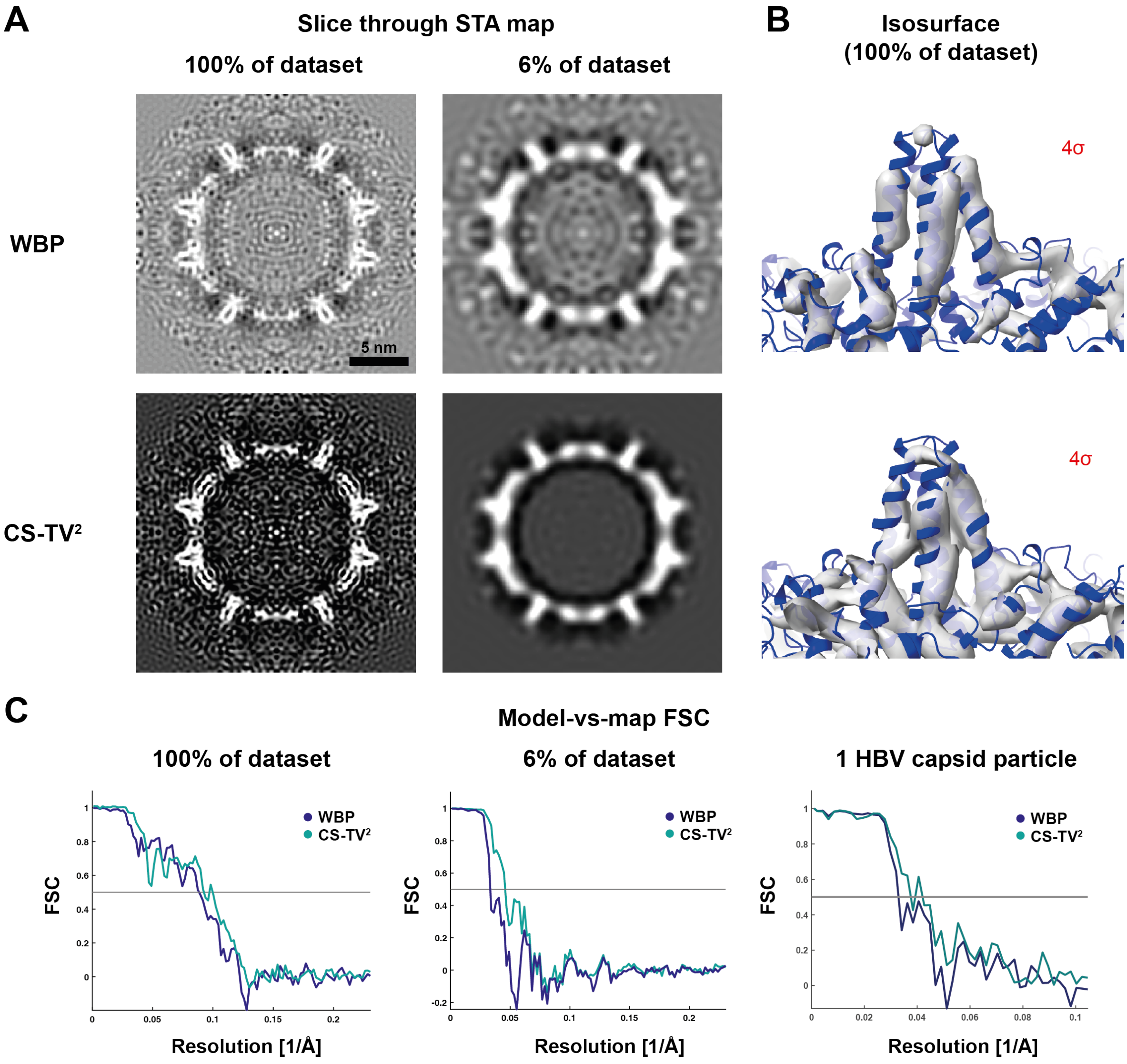


Figure 6.3: STA of CS-TV^2-reconstructed HBV capsids. A) STA map from a dataset reconstructed with CS-TV^2 versus WBP, using the full dataset (188 capsids) and ~6% of the dataset (12 capsids) visualised as XY-slices. B) Isosurface of the STA map of the full dataset (4 σ contour level) with atomic model (PDB 6HTX) fitted into the density. C) Model-vs-map FSC plot against an atomic model (PDB 6HTX) for a dataset of 188 particles (100%), 12 particles (6%), and 1 HBV particle (0.5%). This data is published (Böhning *et al.*, 2022a).

assess the quality of the reconstruction, we compared it to a high-resolution atomic model (PDB 6HTX) via model-versus-map FSC. Using a 0.5 cut-off criterion, this resulted in a measured model-vs-map FSC resolution of 10.9 Å for CS-TV^2 and 11.4 Å for WBP (Figure 6.3 C), suggesting CS-TV^2 performs equally, if not slightly better, for this small dataset.

As a main effect of compressed sensing reconstruction is denoising (Leary *et al.*, 2013), it can be reasoned that CS-TV^2 reconstruction provides the biggest improvements compared to WBP when the data amount is limited, i.e., when noise is the main factor limiting the resolution of the STA map. We therefore reduced the size of the dataset and repeated STA (Figure 6.3). Resolutions of STA of WBP and CS-TV^2 reconstructed particles were comparable when using 50% of each dataset (96 particles, model-vs-map resolution 11.2 Å for CS-TV^2 versus 11.4 Å for WBP; Figure S4, Appendix), but significant differences could be detected when only using ~6% of the dataset (12 particles), with significantly higher correlation in the 2-3 nanometre resolution range for CS-TV^2, with 20.8 Å model-vs-map resolution measured for CS-TV^2 versus 31.2 Å for WBP (Figure 6.3 and Figure S4, Appendix). This suggests that employing CS-TV^2 for STA may be particularly beneficial for small datasets; improving map quality in the low nanometre resolution range may be particularly helpful for integrative modelling for *in situ* structural biology, where subnanometer resolutions have been rarely achieved (Böhning & Bharat, 2021).

6.3.3 CS-TV^2 reconstruction and STA of 80S ribosomes

A major disadvantage of CS-TV^2, compared to other commonly used reconstruction techniques such as WBP and SIRT, is that several regularisation (λ) parameters may need to be trialled before an optimal reconstruction is obtained. To determine whether regularisation parameters are approximately robust over different dataset, we reconstructed a tomographic ribosome dataset (EMPIAR-10045; Figure 6.4) with CS-TV^2 using the same λ parameter as employed for HBV capsids (Figure 6.3). In projections of particles within the dataset, the envelope of the ribosome can more clearly be distinguished for CS-TV^2, as quantified by radial intensity line diagrams (Figure 6.4 A). In a slice through the 3D reconstruction, approximately equal detail can be seen (Figure 6.4 B), and map-vs-map FSCs versus a single-particle 3.7 Å ribosome reconstruction from the same sample (kindly provided by Dr. Israel Fernandez) show comparable map-vs-map FSCs (Figure 6.4 C). This indicates that the choice of λ, while dependent on tilt series parameters and the intrinsic sparsity of the depicted object, is approximately robust between the two shown samples acquired using the same magnification and acquisition- and microscope conditions.

6.3.4 CS-TV^2 reconstruction of *C. crescentus* stalks and cells

As CS-TV^2 appeared to improve the SNR in subtomograms of purified specimens even without STA, the algorithm was applied to more complex cellular specimen. First, we applied the algorithm to tomographic data of a *C. crescentus* cell stalk, a cellular appendage that is covered by a surface layer made of pseudo-hexameric

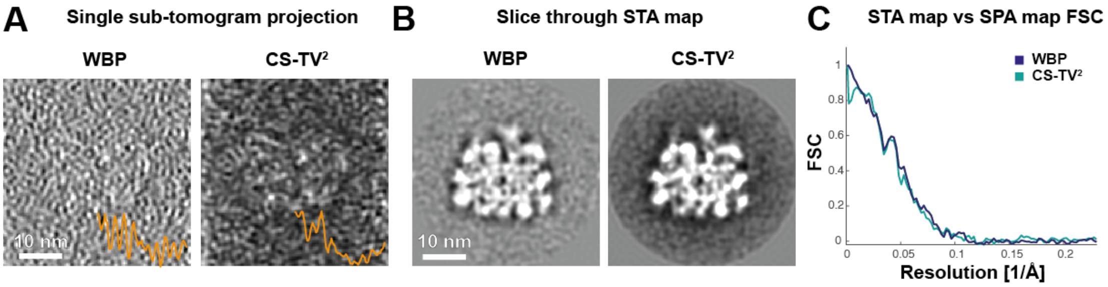


Figure 6.4: CS-TV^2 reconstruction and STA of 80S ribosomes (EMPIAR-10045). A) Projected CS-TV^2 versus WBP subtomogram with radially averaged intensity plotted onto the image. B) STA map of all particles of the dataset. C) Map-vs-map FSC against a cryo-EM map obtained using SPA of the same sample (kindly provided by Israel Fernandez). This data is published (Böhning *et al.*, 2022a).

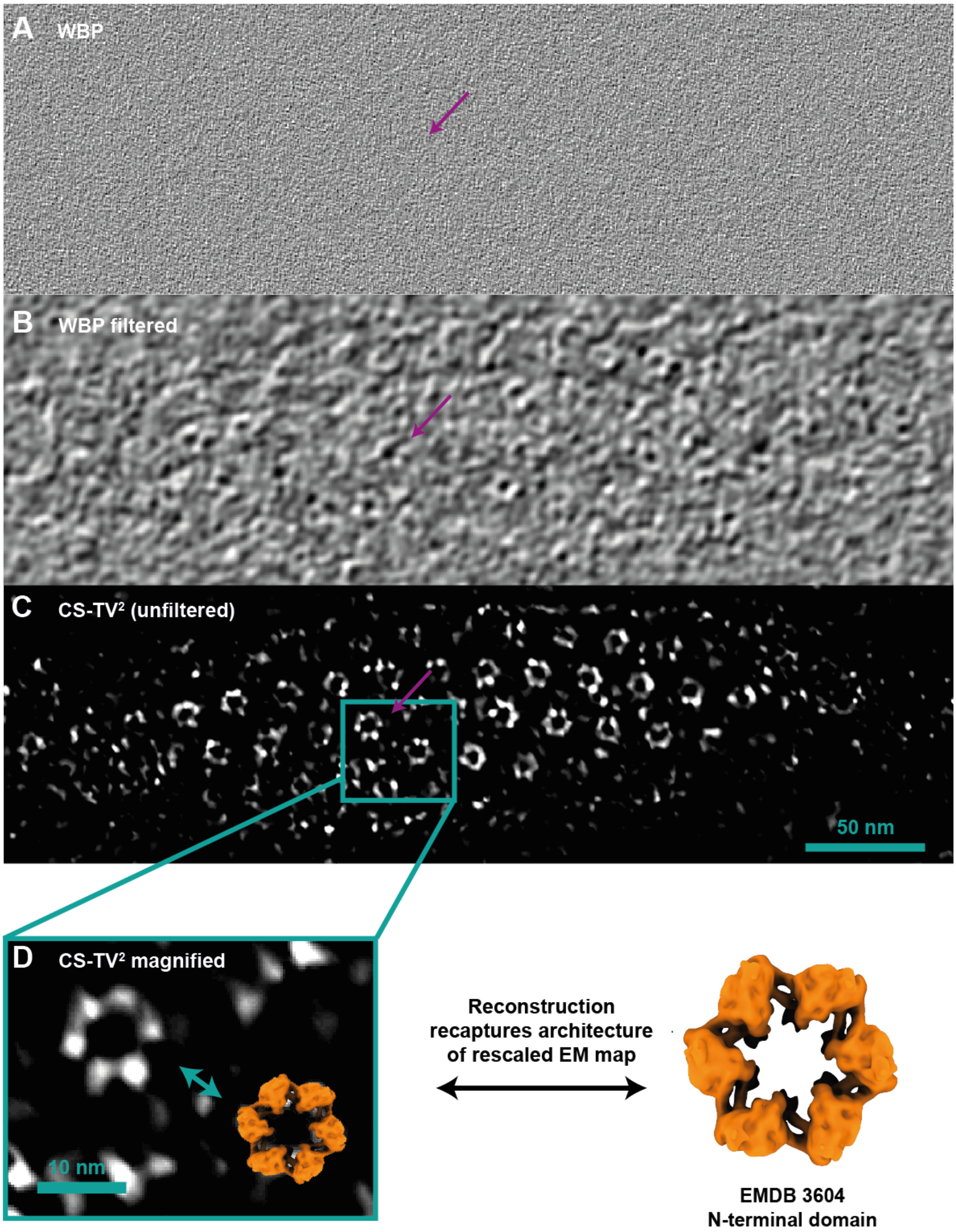


Figure 6.5: CS-TV^2 reconstruction of an S-layer on a *C. crescentus* stalk. A) Unfiltered WBP reconstruction. B) WBP reconstruction filtered with a 3-σ Gaussian and a ~45 Å lowpass filter. C) Unfiltered CS-TV^2 reconstruction ($\lambda = 0.005$). D) Comparison of CS-TV^2 reconstruction with an appropriately rescaled STA map of the N-terminal domain of the RsaA hexamer. This data is published (Böhning *et al.*, 2022a).

RsaA subunits (Figure 6.5). As no STA was performed, the regularisation parameter (λ) was chosen according to visual inspection (λ=0.005). In the resulting tomogram, the unfiltered WBP reconstruction is extremely noisy and does not show any distinct detail of the S-layer itself (Figure 6.5 A). Heavy lowpass- (45 Å) and Gaussian-filtering of the WBP reconstruction reveals the outline of the RsaA hexamer (Figure 6.5 B), while the CS-TV^2 reconstruction resolves individual subunits within the RsaA complex that recapture the features of an STA map of the RsaA N-terminal domain (EMDB 3604; Figure 6.5 C-D), suggesting CS-TV^2 may be applicable for resolving molecular detail in cellular cryo-ET data. We furthermore applied the algorithm to whole *C. crescentus* cells (Figure 6.6). In an unfiltered WBP reconstruction, visual detection of any cellular features except for membranes and the S-layer is difficult (Figure 6.6 A). After 3-σ Gaussian filtering and lowpass-filtering to 45 Å, features within the cell become slightly clearer (Figure 6.6 B), but macromolecular complexes within the cell are still difficult to discern. In the CS-TV^2 reconstruction (Figure 6.6 C), ribosomes become clearly discernible over the cellular background with overall increased contrast. This suggests that CS-TV^2 may be useful for the

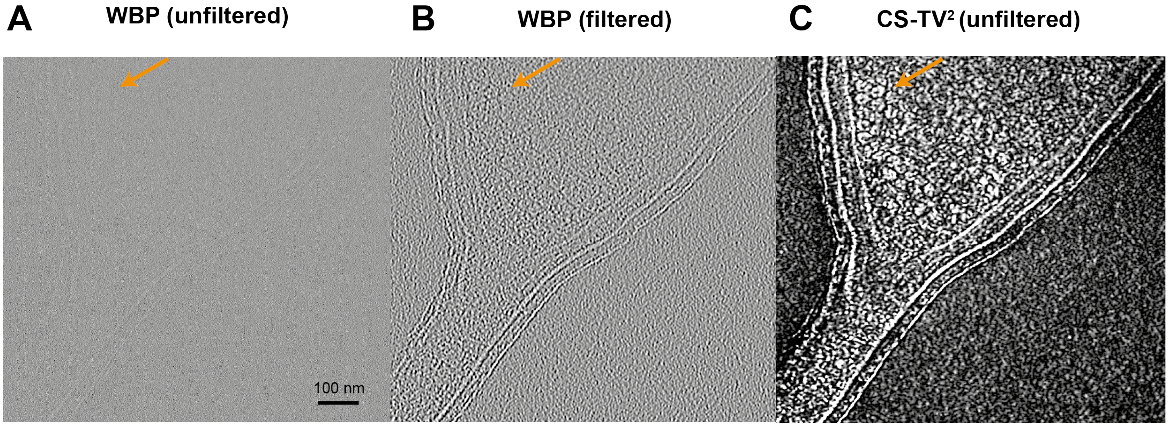


Figure 6.6: CS-TV^2 reconstruction of a *C. crescentus* cell. A) Unfiltered WBP reconstruction, B) 3-σ Gaussian- and ~45 Å lowpass-filtered WBP reconstruction, and C) unfiltered CS-TV^2 (λ = 0.0015) reconstruction of a *C. crescentus* cell body. A putative polysome can be seen within the reconstruction (marked by arrow). Published in Böhning *et al.* (2022a).

6.4 Discussion

A key problem for tomographic reconstruction algorithms is that conventionally used algorithms either retain high-resolution information but produce reconstructions with high noise levels and low contrast, such as WBP; or, that contrast is enhanced but high-resolution information is lost, as in SIRT.

In this chapter, a compressed sensing second-order total variation algorithm (CS-TV^2), as previously applied for tomographic reconstruction in materials science (Collins *et al.*, 2019), was applied to a biological macromolecular dataset and it was demonstrated it may be useful for STA of small datasets of biological specimens, and may improve visualisation of cellular specimen. Using an STA dataset of 188 HBV particles from three tomograms, slightly improved resolution could be achieved using CS-TV^2 compared to WBP, whilst allowing to resolve α-helical densities in the CS-TV^2 map (Figure 6.3). Significant improvements in STA resolution could be obtained compared to WBP when submitting just a subset of particles for STA (6% of the dataset); then, CS-TV^2 can outperform WBP by almost 10 Å resolution (Figure 6.3). In cellular tomograms, CS-TV^2 allows to resolve domain features of the RsaA inner hexamer forming the S-layer on *C. crescentus* cells (Figure 6.5). Together, these results indicate that CS-TV^2 significantly denoises reconstructions whilst retaining higher-resolution features. The fact that secondary structure can be resolved is significant, as alternative reconstruction algorithms that enhance contrast, such as ART and SIRT, tend to lose high-resolution features below the noise level of the tomograms (Wan & Briggs, 2016), and the resolution of secondary structure has, to our knowledge, not been previously demonstrated with reconstruction algorithms other than WBP. Interestingly, we also found tomograms of whole cells can be denoised (Figure 6.6), facilitating the detection of macromolecules within the tomogram. This may also be advantageous for detecting target macromolecules in tomograms – which can be difficult, particularly for smaller complexes - using template matching algorithms (Ortiz *et al.*, 2006), as lower frequency features are increased in the reconstruction; further studies will be required to confirm whether this is the case.

A major drawback of the CS-TV^2 algorithm is reconstruction time. For reconstruction of a 300x300x300 px^3 box, approximately 20 minutes were required a conventional, 8 CPU core workstation with a single GPU. In the case of HBV, the highly symmetrical nature of the icosahedral virus capsid was advantageous as less than 200 particles were required to resolve secondary structure, representing an ideal test case; for less symmetric molecules, however, there are significantly increased reconstruction times. Additional problems exist for cellular reconstruction: Since reconstructions can only be performed in small volumes in computationally reasonable timeframes, for cellular tomograms, the volume must be split into overlapping parts, separately reconstructed, and added back together to form the full tomogram; some scripting is currently required to do this.

Another issue with all compressed sensing reconstruction techniques is that the regularisation parameter λ must be manually optimised. The λ parameter varies depending on its molecular structure (or, more accurately, the intrinsic sparsity of the sample), the signal-to-noise ratio of the reconstruction – in its turn depending on imaging parameters, including magnification and defocus – and errors in tilt series alignment. In practice, this typically means that several reconstructions must be trialled, increasing the reconstruction time for a dataset by a multitude. This study, however, provides some indication that no major re-probing of the λ parameter may be required for similar samples acquired at equivalent imaging conditions (i.e., ribosomes and HBV capsids). Furthermore, whilst the λ parameter is typically probed by visual inspection in materials science, measuring the FSC of the final map against an atomic model provide an objective target measure for optimising the λ parameter. While reconstruction on an 8-CPU workstation for a dataset of 10,000 particles would approximately take 3-4 months, high-performance clusters may be employed to alleviate this problem. The parallelisable nature of the iterative reconstruction may also mean that more efficient GPU parallelisation is, in principle, feasible.

Given these difficulties in the reconstruction process and modest resolution gains for large datasets, in which situations is CS-TV^2 actually useful? To answer this question, it is a good idea to ask when CS-TV^2 is likely not going to be very useful: wherever large datasets can be acquired, so that signal-to-noise is not an issue in the resulting reconstruction. Given that WBP reconstruction simply arranges the aligned tilt images in Fourier space (while downweighing over-represented low frequencies) to obtain a real-space reconstruction, it retains high-resolution information ideally, assuming tilt series alignment has been successful. Some powerful evidence for this is given by the fact that, to the author’s best knowledge, all previously recorded subnanometre STA maps were created using WBP reconstructions. It can be suspected that CS-TV^2 will mainly be appropriate in situations where so few particles are present that the iterative reference in the alignment process is noisy itself, resulting in subtomogram misalignment. Given the possibility to produce more data during *in vitro* cryo-ET without bottlenecks in sample thinning, this is less likely to be an issue for *in vitro* studies. For *in situ* cryo-ET, however, significant throughput issues exist (reviewed in Böhning & Bharat (2021)), and it can be difficult to acquire large datasets of macromolecules that are not highly abundant. In this case, it could be useful to employ CS-TV^2 to ‘make the best’ of the collected data. Notably, our current study quantified resolution on purified, *in vitro* specimen; further studies on actual *in situ* datasets will be required to confirm its suitability.

CS-TV^2 may also be helpful to resolve features of macromolecules below the noise level of tomograms, as shown at the example of the S-layer on a *C. crescentus* stalk. While lowpass-filters can help in WBP reconstruction to obtain higher contrast by eliminating high-resolution noise, by definition, high-resolution features are eliminated. Further studies will be required to compare the performance of CS-TV^2 with a wide range of available post-reconstruction filters and denoising algorithms to determine which produces better results for visualisation of smaller macromolecular features, including visualisation of protein domains and *in situ* structural biology. Such methodology will, for example, be important to appropriately visualise the native biofilm matrix by cryo-ET, where constituent fibres are often difficult to visualise due to the small size of their subunits.

7. Conclusion and outlook

This thesis presents structural data for three biofilm matrix fibres, including a cryo-EM structure of the CupE pilus in *P. aeruginosa*, 2D class averages and structural predictions of FapC in *P. aeruginosa,* and the cryo-EM structure of TasA fibres in *B. subtilis*. The structure of the CupE pilus shows an Ig-like β-sandwich fold of the pilin subunit that polymerises through donor-strand exchange, providing extensive inter-subunit contacts through the donor strand but few at the interface between globular Ig-like subunits, which may enable flexibility. While no cryo-EM structure of Fap fibres could thus far be obtained, AlphaFold predictions for FapC multimers suggest an amyloid-like arrangement of subunits in fibres where each subunit may provide three β-strand repeats perpendicular to the fibre axis; these predictions are still to be confirmed using experimental data. In the case of *B. subtilis* TasA, its fibres were thought to be amyloid based on biophysical evidence – similar to Fap - but our cryo-EM structure presented here shows that the *tasA/tapA* operon is rather a Gram-positive donor-strand exchange system, and thus fibres are structurally more similar to the CupE pilus. The three studied fibres are rich in β-sheets and form β-sheet contacts between subunits. For biofilm matrix fibres, it can be assumed that a primary driver for the evolution of the fibre is stability; a scaffold must not collapse easily, and the strength of interactions between subunits within a fibre are a primary determinant of fibre stability. β-sheet interactions, by providing numerous hydrogen bonds connecting subunits, are likely one of the most effective ways of providing strong interactions between subunits. Notably, β-sheet interactions are respectively supplemented by hydrophobic interactions in all three cases.

This thesis describes the first donor strand system in Gram-positive bacteria, and very recent literature that was published after our preprint on the structure of TasA found similar systems in archaea (Gaines *et al.*, 2022; Wang *et al.*, 2022), suggesting that donor strands are phylogenetically wide-spread. Indeed, given that many donor strand systems lack homology to others (e.g., TasA to CUP), donor-strand complementation seems to have evolved numerous times, underlining the success of this mechanism for fibre assembly.

Interestingly, all previously shown fibres undergoing donor-strand exchange were found to be outside the cell. Why do we not find such systems inside the cell? Stability can be a double-edged sword: While high stability can be advantageous, it suppresses structural dynamics, and not being able to effectively disassemble fibrous structures within the cell may lead to aggregates and cell damage, as for example seen in the case of amyloids in neurodegenerative disease (Eisenberg & Jucker, 2012). Fap and TasA fibres are both protease resistant (Diehl *et al.*, 2018; Dueholm *et al.*, 2013a), and Fap is not disassembled even when boiled in the presence of 2% SDS (Dueholm *et al.*, 2013b). Such stability may be useful when competing with other bacteria, which may secrete proteases, effectively acting as a structural 'shield'. Difficulties in turnover may not be problematic since cells were shown to be able to simply detach from the matrix (Kolodkin-Gal *et al.*, 2010); if it is no longer favourable to live in a biofilm state, planktonic cells can disperse into the environment from the edge of the biofilm (Sauer *et al.*, 2002). Donor-strand-complemented fibres have not been seen in eukaryotes, with the notable exception of uromodulin, a donor-strand-complemented protein fibre abundant in human urine (Stanisich *et al.*, 2020; Weiss *et al.*, 2020). Uromodulin, however, avoids degradation issues as it is excreted through urination and thus does not need to be turned over.

Important for understanding biofilm matrix fibre function is understanding the formation of fibre bundles. Bundle formation is documented for Fap (Dueholm *et al.*, 2013b) and TasA (Ghrayeb *et al.*, 2021) and we show evidence in this thesis that CupE can form lattice-like arrangements (Figure 3.1 B), suggesting bundling of pili may be possible. In the case of TasA, we show evidence for a loop (containing residues 174-177) being involved in the fibre-fibre interaction, but no conclusive, atomistic model could be obtained. Understanding how bundles form will be of interest to fully understand fibre function. How these bundles are arranged within the biofilm to provide stability and cohesion to the biofilm will be of further importance. A technique that is ideally suited to resolve fibres in their native biofilm environment is FIB-milling combined with cryo-ET (Villa *et al.*, 2013). The sample-thinning step will be crucial as biofilms are thick objects that are not intrinsically amenable to cryo-ET (McIntosh *et al.*, 2005). Once a thinned sample is created, cryo-ET can then be used to determine the arrangement of fibres and bundles within intact biofilms, and may allow to resolve connections of biofilm matrix fibres to cells, which have, for example, been proposed in the case of *B. subtilis* TasA (Romero *et al.*, 2011). Indeed, there are numerous advantages for performing cryo-EM of native biofilms compared to isolated or reconstituted systems. More native interactions are obtained, and interactions with the numerous other components of the biofilm matrix can potentially be resolved. So why not do cryo-ET, all the time? Firstly, FIB-milling is time-intensive and, despite of efforts in automating parts of the milling process (Klumpe *et al.*, 2021; Zachs *et al.*, 2020), is still time-consuming and expensive. Further challenges have been limited cryo-ET acquisition speed and computational limitations in image processing (reviewed in Böhning & Bharat (2021)). Still, techniques are constantly evolving and *in situ* structural biology may one day be routine, hopefully allowing native high-resolution imaging and structural solution of biofilm components to better understand their function. This way, the arrangement of protein components can be better understood, particularly fibres, that are more readily located and resolved due to their size. Better image processing methods, such as CS-TV^2 reconstruction, may further help to resolve macromolecules in their native environment. It is, however, important to note that a significant part of the biofilm EPS consists of polysaccharide and DNA components that are not amenable to cryo-EM structural solution and complementing methodology will be required to identify the exact composition of and interactions between EPS components.

Apart from the aspects of advancing foundational biofilm biology, a structural understanding of the biofilm matrix will be crucial to find better therapeutics against bacterial infections involving biofilm formations. Indeed, vaccines targeted against the Csu pilus in *A. baumannii*, which is highly related to the CupE pilus in *P. aeruginosa*, have been proposed to be effective against infection in mice (Ramezanalizadeh *et al.*, 2020). Other agents such as antibodies or nanobodies could be used to break up matrix components or prevent their function (Cegelski *et al.*, 2009; Jain *et al.*, 2017). High concentrations of the anti-biofilm agents may be required if abundant components are targeted, although this can be alleviated by targeting, for example, adhesive tip subunits of fibres (Pakharukova *et al.*, 2018) or adhesins with lower abundances (Melia *et al.*, 2021). Future studies will determine whether targeting biofilm matrix components is a viable strategy.

Finally, a question arising is why *Pseudomonas* and *Bacillus* biofilms have differing apparent compositional complexities. Biofilms in *Pseudomonas* encode numerous fibre components (Fap, classical CUP pili CupA-D, archaic CUP pili CupE, Pf4), adhesins (CdrA, Bap, LecA) and at least three polysaccharides (Psl, Pel, alginate), while *Bacillus* biofilms appear to just require one fibre component and one polysaccharide. It is plausible that *Pseudomonas* can, this way, adopt to more environments, employing an adhesive toolkit that fits the current situation best. What the exact triggers for the expression of each adhesive signature are will be the goal of future research to better understand the structural basis of bacterial multicellular life.

8. References

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* 66: 213-221

Agulleiro J, Fernandez J-J (2011) Fast tomographic reconstruction on multicore computers. *Bioinformatics* 27: 582-583

Ahn S-J, Burne RA (2007) Effects of oxygen on biofilm formation and the AtlA autolysin of Streptococcus mutans. *Journal of Bacteriology* 189: 6293-6302

Al-Amoudi A, Chang JJ, Leforestier A, McDowall A, Salamin LM, Norlén LP, Richter K, Blanc NS, Studer D, Dubochet J (2004) Cryo-electron microscopy of vitreous sections. *The EMBO Journal* 23: 3583-3588

Albert S, Schaffer M, Beck F, Mosalaganti S, Asano S, Thomas HF, Plitzko JM, Beck M, Baumeister W, Engel BD (2017) Proteasomes tether to two distinct sites at the nuclear pore complex. *Proceedings of the National Academy of Sciences* 114: 13726-13731

Alhede M, Bjarnsholt T, Givskov M, Alhede M (2014) Pseudomonas aeruginosa biofilms: mechanisms of immune evasion. *Advances in Applied Microbiology* 86: 1-40

Allison DG, Sutherland IW (1987) The role of exopolysaccharides in adhesion of freshwater bacteria. *Microbiology* 133: 1319-1327

Asally M, Kittisopikul M, Rué P, Du Y, Hu Z, Çağatay T, Robinson AB, Lu H, Garcia-Ojalvo J, Süel GM (2012) Localized cell death focuses mechanical forces during 3D patterning in a biofilm. *Proceedings of the National Academy of Sciences* 109: 18891-18896

Azulay DN, Spaeker O, Ghrayeb M, Wilsch-Bräuninger M, Scoppola E, Burghammer M, Zizak I, Bertinetti L, Politi Y, Chai L (2022) Multiscale X-ray study of Bacillus subtilis biofilms reveals interlinked structural hierarchy and elemental heterogeneity. *Proceedings of the National Academy of Sciences* 119: e2118107119

Båga M, Norgren M, Normark S (1987) Biogenesis of E. coli Pap pili: PapH, a minor pilin subunit involved in cell anchoring and length modulation. *Cell* 49: 241-251

Bai X-C, McMullan G, Scheres SH (2015) How cryo-EM is revolutionizing structural biology. *Trends in biochemical sciences* 40: 49-57

Baker LA, Rubinstein JL (2010) Radiation damage in electron cryomicroscopy. In: *Methods in enzymology*, pp. 371-388. Elsevier:

Balistreri A, Goetzler E, Chapman M (2020) Functional amyloids are the rule rather than the exception in cellular biology. *Microorganisms* 8: 1951

Barken KB, Pamp SJ, Yang L, Gjermansen M, Bertrand JJ, Klausen M, Givskov M, Whitchurch CB, Engel JN, Tolker-Nielsen T (2008) Roles of type IV pili, flagellum-mediated motility and extracellular DNA in the formation of mature multicellular structures in Pseudomonas aeruginosa biofilms. *Environmental Microbiology* 10: 2331-2343

Barnhart MM, Pinkner JS, Soto GE, Sauer FG, Langermann S, Waksman G, Frieden C, Hultgren SJ (2000) PapD-like chaperones provide the missing information for folding of pilin proteins. *Proceedings of the National Academy of Sciences* 97: 7709-7714

Baumeister W, Grimm R, Walz J (1999) Electron tomography of molecules and cells. *Trends in Cell Biology* 9: 81-85

Berson JF, Harper DC, Tenza D, Raposo G, Marks MS (2001) Pmel17 initiates premelanosome morphogenesis within multivesicular bodies. *Molecular Biology of the Cell* 12: 3451-3464

Bharat TA, Russo CJ, Löwe J, Passmore LA, Scheres SH (2015) Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. *Structure* 23: 1743-1753

Bian Z, Normark S (1997) Nucleator function of CsgB for the assembly of adhesive surface organelles in Escherichia coli. *The EMBO Journal* 16: 5827-5836

Böhning J, Bharat TA (2021) Towards high-throughput in situ structural biology using electron cryotomography. *Progress in Biophysics and Molecular Biology* 160: 97-103

Böhning J, Bharat TA, Collins SM (2022a) Compressed sensing for electron cryotomography and high-resolution subtomogram averaging of biological specimens. *Structure*

Böhning J, Dobbelstein A, Sulkowski N, Eilers K, von Kugelgen A, Tarafder AK, Alva V, Filloux A, Bharat TA (2022b) Architecture of the biofilm-associated archaic CupE pilus from Pseudomonas aeruginosa. *bioRxiv*

Böhning J, Ghrayeb M, Pedebos C, Abbas D, Khalid S, Chai L, Bharat T (2022c) Molecular architecture of the TasA biofilm scaffold in Bacillus subtilis. *bioRxiv*

Bordeleau E, Purcell EB, Lafontaine DA, Fortier L-C, Tamayo R, Burrus V (2015) Cyclic di-GMP riboswitch-regulated type IV pili contribute to aggregation of Clostridium difficile. *Journal of Bacteriology* 197: 819-832

Bouteiller M, Dupont C, Bourigault Y, Latour X, Barbey C, Konto-Ghiorghi Y, Merieau A (2021) Pseudomonas flagella: Generalities and specificities. *International Journal of Molecular Sciences* 22: 3337

Boyer RR, Sumner SS, Williams RC, Pierson MD, Popham DL, Kniel KE (2007) Influence of curli expression by Escherichia coli O157: H7 on the cell's overall hydrophobicity, charge, and ability to attach to lettuce. *Journal of food protection* 70: 1339-1345

Brading M, Boyle J, Lappin-Scott H (1995) Biofilm formation in laminar flow using Pseudomonas fluorescens EX101. *Journal of Industrial Microbiology and Biotechnology* 15: 297-304

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

Branda SS, González-Pastor JE, Ben-Yehuda S, Losick R, Kolter R (2001) Fruiting body formation by Bacillus subtilis. *Proceedings of the National Academy of Sciences* 98: 11621-11626

Briggs JA (2013) Structural biology in situ—the potential of subtomogram averaging. *Current Opinion in Structural Biology* 23: 261-267

Brilot AF, Chen JZ, Cheng A, Pan J, Harrison SC, Potter CS, Carragher B, Henderson R, Grigorieff N (2012) Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology* 177: 630-637

Brombacher E, Baratto A, Dorel C, Landini P (2006) Gene expression regulation by the curli activator CsgD protein: modulation of cellulose biosynthesis and control of negative determinants for microbial adhesion. *Journal of Bacteriology* 188: 2027-2037

Brumshtein B, Esswein SR, Landau M, Ryan CM, Whitelegge JP, Phillips ML, Cascio D, Sawaya MR, Eisenberg DS (2014) Formation of amyloid fibers by monomeric light chain variable domains. *Journal of Biological Chemistry* 289: 27513-27525

Buzzo JR, Devaraj A, Gloag ES, Jurcisek JA, Robledo-Avila F, Kesler T, Wilbanks K, Mashburn-Warren L, Balu S, Wickham J (2021) Z-form extracellular DNA is a structural component of the bacterial biofilm matrix. *Cell* 184: 5740-5758. e5717

Cámara-Almirón J, Navarro Y, Díaz-Martínez L, Magno-Pérez-Bryan MC, Molina-Santiago C, Pearson JR, de Vicente A, Pérez-García A, Romero D (2020) Dual functionality of the amyloid protein TasA in Bacillus physiology and fitness on the phylloplane. *Nature communications* 11: 1-21

Cegelski L, Pinkner JS, Hammer ND, Cusumano CK, Hung CS, Chorell E, Åberg V, Walker JN, Seed PC, Almqvist F (2009) Small-molecule inhibitors target Escherichia coli amyloid biogenesis and biofilm formation. *Nature Chemical Biology* 5: 913-919

Ceri H, Olson ME, Stremick C, Read R, Morck D, Buret A (1999) The Calgary Biofilm Device: new technology for rapid determination of antibiotic susceptibilities of bacterial biofilms. *Journal of Clinical Microbiology* 37: 1771-1776

Chang Y-W, Fragkopoulos AA, Marquez SM, Kim HD, Angelini TE, Fernández-Nieves A (2015) Biofilm formation in geometries with different surface curvature and oxygen availability. *New Journal of Physics* 17: 033017

Chapman MR, Robinson LS, Pinkner JS, Roth R, Heuser J, Hammar M, Normark S, Hultgren SJ (2002) Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science* 295: 851-855

Chen M, Yu Q, Sun H (2013) Novel strategies for the prevention and treatment of biofilm related infections. *International Journal of Molecular Sciences* 14: 18488-18501

Chen Y, Förster F (2014) Iterative reconstruction of cryo-electron tomograms using nonuniform fast Fourier transforms. *Journal of Structural Biology* 185: 309-316

Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry* 75: 333-366

Choudhury D, Thompson A, Stojanoff V, Langermann S, Pinkner J, Hultgren SJ, Knight SD (1999) X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic Escherichia coli. *Science* 285: 1061-1066

Chua SL, Liu Y, Yam JKH, Chen Y, Vejborg RM, Tan BGC, Kjelleberg S, Tolker-Nielsen T, Givskov M, Yang L (2014) Dispersed cells represent a distinct stage in the transition from bacterial biofilm to planktonic lifestyles. *Nature communications* 5: 1-12

Collins SM, MacArthur KE, Longley L, Tovey R, Benning M, Schönlieb C-B, Bennett TD, Midgley PA (2019) Phase diagrams of liquid-phase mixing in multi-component metal-organic framework glasses constructed by quantitative elemental nano-tomography. *APL materials* 7: 091111

Colquhoun JM, Rather PN (2020) Insights into mechanisms of biofilm formation in Acinetobacter baumannii and implications for uropathogenesis. *Frontiers in Cellular and Infection Microbiology* 10: 253

Cucarella C, Solano C, Valle J, Amorena B, Lasa Ini, Penadés JR (2001) Bap, a Staphylococcus aureus surface protein involved in biofilm formation. *Journal of Bacteriology* 183: 2888-2896

Cucarella C, Tormo MA, Ubeda C, Trotonda MP, Monzón M, Peris C, Amorena B, Lasa Í, Penadés JR (2004) Role of biofilm-associated protein bap in the pathogenesis of bovine Staphylococcus aureus. *Infection and immunity* 72: 2177-2185

De Bentzmann S, Giraud C, Bernard CS, Calderon V, Ewald F, Plésiat P, Nguyen C, Grunwald D, Attree I, Jeannot K (2012) Unique biofilm signature, drug susceptibility and decreased virulence in Drosophila through the Pseudomonas aeruginosa two-component system PprAB. *PLoS Pathogens* 8: e1003052

de Jong DH, Singh G, Bennett WD, Arnarez C, Wassenaar TA, Schafer LV, Periole X, Tieleman DP, Marrink SJ (2013) Improved parameters for the martini coarse-grained protein force field. *Journal of chemical theory and computation* 9: 687-697

De Rosier D, Klug A (1968) Reconstruction of three dimensional structures from electron micrographs. *Nature* 217: 130-134

DeLano WL (2002) Pymol: An open-source molecular graphics tool. *CCP4 Newsl Protein Crystallogr* 40: 82-92

Deng X, Gonzalez Llamazares A, Wagstaff JM, Hale VL, Cannone G, McLaughlin SH, Kureisaite-Ciziene D, Löwe J (2019) The structure of bactofilin filaments reveals their mode of membrane binding and lack of polarity. *Nature Microbiology* 4: 2357-2368

Deng Y, Chen Y, Zhang Y, Wang S, Zhang F, Sun F (2016) ICON: 3D reconstruction with 'missing-information'restoration in biological electron tomography. *Journal of Structural Biology* 195: 100-112

Diehl A, Roske Y, Ball L, Chowdhury A, Hiller M, Molière N, Kramer R, Stöppler D, Worth CL, Schlegel B (2018) Structural changes of TasA in biofilm formation of Bacillus subtilis. *Proceedings of the National Academy of Sciences* 115: 3237-3242

Donoho DL (2006) Compressed sensing. *IEEE Transactions on information theory* 52: 1289-1306

Dragoš A, Kiesewalter H, Martin M, Hsu C-Y, Hartmann R, Wechsler T, Eriksen C, Brix S, Drescher K, Stanley-Wall N (2018) Division of labor during biofilm matrix production. *Current Biology* 28: 1903-1913. e1905

Dubochet J, Adrian M, Chang J-J, Homo J-C, Lepault J, McDowall AW, Schultz P (1988) Cryo-electron microscopy of vitrified specimens. *Quarterly reviews of biophysics* 21: 129-228

Dueholm MS, Otzen D, Nielsen PH (2013a) Evolutionary insight into the functional amyloids of the pseudomonads. *PLoS One* 8: e76630

Dueholm MS, Petersen SV, Sønderkær M, Larsen P, Christiansen G, Hein KL, Enghild JJ, Nielsen JL, Nielsen KL, Nielsen PH (2010) Functional amyloid in Pseudomonas. *Molecular Microbiology* 77: 1009-1020

Dueholm MS, Søndergaard MT, Nilsson M, Christiansen G, Stensballe A, Overgaard MT, Givskov M, Tolker-Nielsen T, Otzen DE, Nielsen PH (2013b) Expression of Fap amyloids in P seudomonas aeruginosa, P. fluorescens, and P. putida results in aggregation and increased biofilm formation. *Microbiology open* 2: 365-382

Earl AM, Losick R, Kolter R (2008) Ecology and genomics of Bacillus subtilis. *Trends in Microbiology* 16: 269-275

Earl C, Arnaouteli S, Bamford NC, Porter M, Sukhodub T, MacPhee CE, Stanley-Wall NR (2020) The majority of the matrix protein TapA is dispensable for Bacillus subtilis colony biofilm architecture. *Molecular Microbiology* 114: 920-933

Egelman EH (2000) A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy* 85: 225-234

Ehling-Schulz M, Lereclus D, Koehler TM (2019) The Bacillus cereus group: Bacillus species with pathogenic potential. *Microbiology spectrum* 7: 7.3. 6

Eisenberg D, Jucker M (2012) The amyloid state of proteins in human diseases. *Cell* 148: 1188-1203

Elasri MO, Miller RV (1999) Study of the response of a biofilm bacterial community to UV radiation. *Applied and Environmental Microbiology* 65: 2025-2031

Else TA, Pantle CR, Amy PS (2003) Boundaries for biofilm formation: humidity and temperature. *Applied and Environmental Microbiology* 69: 5006-5010

Erskine E, Morris RJ, Schor M, Earl C, Gillespie RM, Bromley KM, Sukhodub T, Clark L, Fyfe PK, Serpell LC (2018) Formation of functional, non-amyloidogenic fibres by recombinant Bacillus subtilis TasA. *Molecular Microbiology* 110: 897-913

Evans R, O'Neill M, Pritzel A, Antropova N, Senior AW, Green T, Žídek A, Bates R, Blackwell S, Yim J (2021) Protein complex prediction with AlphaFold-Multimer. *bioRxiv*

Fazli M, Almblad H, Rybtke ML, Givskov M, Eberl L, Tolker-Nielsen T (2014) Regulation of biofilm formation in P seudomonas and B urkholderia species. *Environmental Microbiology* 16: 1961-1981

Fitzpatrick AW, Falcon B, He S, Murzin AG, Murshudov G, Garringer HJ, Crowther RA, Ghetti B, Goedert M, Scheres SH (2017) Cryo-EM structures of tau filaments from Alzheimer’s disease. *Nature* 547: 185-190

Flemming H-C, Neu TR, Wozniak DJ (2007) The EPS matrix: the “house of biofilm cells”. *Journal of Bacteriology* 189: 7945-7947

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

Flemming H-C, Wingender J, Szewzyk U, Steinberg P, Rice SA, Kjelleberg S (2016) Biofilms: an emergent form of bacterial life. *Nature Reviews Microbiology* 14: 563-575

Flemming H-C, Wuertz S (2019) Bacteria and archaea on Earth and their abundance in biofilms. *Nature Reviews Microbiology* 17: 247-260

Fowler DM, Koulov AV, Balch WE, Kelly JW (2007) Functional amyloid–from bacteria to humans. *Trends in biochemical sciences* 32: 217-224

Frank J (2006) *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*. Oxford university press

Frank J, Shimkin B, Dowse H (1981) SPIDER–a modular software system for electron image processing. In: *Single-Particle Cryo-Electron Microscopy: The Path Toward Atomic Resolution: Selected Papers of Joachim Frank with Commentaries*, pp. 94-108. World Scientific:

Franklin MJ, Nivens DE, Weadge JT, Howell PL (2011) Biosynthesis of the Pseudomonas aeruginosa extracellular polysaccharides, alginate, Pel, and Psl. *Frontiers in Microbiology* 2: 167

Gaines M, Isupov MN, Sivabalasarma S, Haque RU, McLaren M, Tripp P, Neuhaus A, Gold V, Albers S-V, Daum B (2022) Donor strand complementation, isopeptide bonds and glycosylation stabilise highly resilient archaeal thread filaments. *bioRxiv*

Ghrayeb M, Hayet S, Lester-Zer N, Levi-Kalisman Y, Chai L (2021) Fibrilar polymorphism of the bacterial extracellular matrix protein TasA. *Microorganisms* 9: 529

Gilbert P (1972) Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology* 36: 105-117

Giraud C, Bernard CS, Calderon V, Yang L, Filloux A, Molin S, Fichant G, Bordi C, de Bentzmann S (2011) The PprA–PprB two-component system activates CupE, the first non-archetypal Pseudomonas aeruginosa chaperone–usher pathway system assembling fimbriae. *Environmental Microbiology* 13: 666-683

Giraud C, de Bentzmann S (2012) Inside the complex regulation of Pseudomonas aeruginosa chaperone usher systems. *Environmental Microbiology* 14: 1805-1816

Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science* 27: 14-25

Gophna U, Oelschlaeger TA, Hacker J, Ron EZ (2002) Role of fibronectin in curli-mediated internalization. *FEMS microbiology letters* 212: 55-58

Goulter-Thorsen R, Taran E, Gentle I, Gobius K, Dykes G (2011) CsgA production by Escherichia coli O157: H7 alters attachment to abiotic surfaces in some growth environments. *Applied and Environmental Microbiology* 77: 7339-7344

Goyal P, Krasteva PV, Van Gerven N, Gubellini F, Van den Broeck I, Troupiotis-Tsaïlaki A, Jonckheere W, Péhau-Arnaudet G, Pinkner JS, Chapman MR (2014) Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature* 516: 250-253

Grosberg RK, Strathmann RR (2007) The evolution of multicellularity: a minor major transition? *Annual Review of Ecology, Evolution, and Systematics*: 621-654

Guay MD, Czaja W, Aronova MA, Leapman RD (2016) Compressed sensing electron tomography for determining biological structure. *Scientific Reports* 6: 1-14

Hall-Stoodley L, Costerton JW, Stoodley P (2004) Bacterial biofilms: from the natural environment to infectious diseases. *Nature Reviews Microbiology* 2: 95-108

Hammer ND, Schmidt JC, Chapman MR (2007) The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. *Proceedings of the National Academy of Sciences* 104: 12494-12499

Hardy J (1997) Amyloid, the presenilins and Alzheimer's disease. *Trends in Neurosciences* 20: 154-159

He S, Scheres SH (2017) Helical reconstruction in RELION. *Journal of Structural Biology* 198: 163-176

Hentzer M, Teitzel GM, Balzer GJ, Heydorn A, Molin S, Givskov M, Parsek MR (2001) Alginate overproduction affects Pseudomonas aeruginosa biofilm structure and function. *Journal of Bacteriology* 183: 5395-5401

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

Høiby N, Ciofu O, Bjarnsholt T (2010) Pseudomonas aeruginosa biofilms in cystic fibrosis. *Future Microbiology* 5: 1663-1674

Hospenthal MK, Redzej A, Dodson K, Ukleja M, Frenz B, Rodrigues C, Hultgren SJ, DiMaio F, Egelman EH, Waksman G (2016) Structure of a chaperone-usher pilus reveals the molecular basis of rod uncoiling. *Cell* 164: 269-278

Hospenthal MK, Zyla D, Costa TR, Redzej A, Giese C, Lillington J, Glockshuber R, Waksman G (2017) The cryoelectron microscopy structure of the type 1 chaperone-usher pilus rod. *Structure* 25: 1829-1838. e1824

Hultgren SJ, Lindberg F, Magnusson G, Kihlberg J, Tennent JM, Normark S (1989) The PapG adhesin of uropathogenic Escherichia coli contains separate regions for receptor binding and for the incorporation into the pilus. *Proceedings of the National Academy of Sciences* 86: 4357-4361

Huma Ze, Javed I, Zhang Z, Bilal H, Sun Y, Hussain SZ, Davis TP, Otzen DE, Landersdorfer CB, Ding F (2020) Nanosilver mitigates biofilm formation via FapC amyloidosis inhibition. *Small* 16: 1906674

Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *Journal of molecular graphics* 14: 33-38

Iwashkiw JA, Vozza NF, Kinsella RL, Feldman MF (2013) Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Molecular Microbiology* 89: 14-28

Jahnke LL, Eder W, Huber R, Hope JM, Hinrichs K-U, Hayes JM, Des Marais DJ, Cady SL, Summons RE (2001) Signature lipids and stable carbon isotope analyses of Octopus Spring hyperthermophilic communities compared with those of Aquificales representatives. *Applied and Environmental Microbiology* 67: 5179-5189

Jain N, Ådén J, Nagamatsu K, Evans ML, Li X, McMichael B, Ivanova MI, Almqvist F, Buxbaum JN, Chapman MR (2017) Inhibition of curli assembly and Escherichia coli biofilm formation by the human systemic amyloid precursor transthyretin. *Proceedings of the National Academy of Sciences* 114: 12184-12189

Jenal U, Reinders A, Lori C (2017) Cyclic di-GMP: second messenger extraordinaire. *Nature Reviews Microbiology* 15: 271-284

Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry* 29: 1859-1865

Jones CH, Pinkner JS, Roth R, Heuser J, Nicholes AV, Abraham SN, Hultgren SJ (1995) FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the Enterobacteriaceae. *Proceedings of the National Academy of Sciences* 92: 2081-2085

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596: 583-589

Kaniga K, Davison J (1991) Transposon vectors for stable chromosomal integration of cloned genes in rhizosphere bacteria. *Gene* 100: 201-205

Kau AL, Hunstad DA, Hultgren SJ (2005) Interaction of uropathogenic Escherichia coli with host uroepithelium. *Current Opinion in Microbiology* 8: 54-59

Kearns DB, Chu F, Branda SS, Kolter R, Losick R (2005) A master regulator for biofilm formation by Bacillus subtilis. *Molecular Microbiology* 55: 739-749

Khurana R, Coleman C, Ionescu-Zanetti C, Carter SA, Krishna V, Grover RK, Roy R, Singh S (2005) Mechanism of thioflavin T binding to amyloid fibrils. *Journal of Structural Biology* 151: 229-238

King C-Y, Tittmann P, Gross H, Gebert R, Aebi M, Wüthrich K (1997) Prion-inducing domain 2–114 of yeast Sup35 protein transforms in vitro into amyloid-like filaments. *Proceedings of the National Academy of Sciences* 94: 6618-6622

Klumpe S, Fung HK, Goetz SK, Zagoriy I, Hampoelz B, Zhang X, Erdmann PS, Baumbach J, Müller CW, Beck M (2021) A modular platform for automated cryo-FIB workflows. *eLife* 10

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

Kolodkin-Gal I, Romero D, Cao S, Clardy J, Kolter R, Losick R (2010) D-amino acids trigger biofilm disassembly. *Science* 328: 627-629

Kremer JR, Mastronarde DN, McIntosh JR (1996) Computer visualization of three-dimensional image data using IMOD. *Journal of Structural Biology* 116: 71-76

Kuchma SL, Brothers KM, Merritt JH, Liberati NT, Ausubel FM, O'Toole GA, 2007. BifA, a cyclic-Di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by Pseudomonas aeruginosa PA14. Am Soc Microbiol.

Kuehn MJ, Heuser J, Normark S, Hultgren SJ (1992) P pili in uropathogenic E. coli are composite fibres with distinct fibrillar adhesive tips. *Nature* 356: 252-255

Kühlbrandt W (2014) The resolution revolution. *Science* 343: 1443-1444

Kulasekara HD, Ventre I, Kulasekara BR, Lazdunski A, Filloux A, Lory S (2005) A novel two-component system controls the expression of Pseudomonas aeruginosa fimbrial cup genes. *Molecular Microbiology* 55: 368-380

Kulesekara H, Lee V, Brencic A, Liberati N, Urbach J, Miyata S, Lee DG, Neely AN, Hyodo M, Hayakawa Y (2006) Analysis of Pseudomonas aeruginosa diguanylate cyclases and phosphodiesterases reveals a role for bis-(3′-5′)-cyclic-GMP in virulence. *Proceedings of the National Academy of Sciences* 103: 2839-2844

Kunz M, Frangakis AS (2014) Super-sampling SART with ordered subsets. *Journal of Structural Biology* 188: 107-115

Leary R, Saghi Z, Midgley PA, Holland DJ (2013) Compressed sensing electron tomography. *Ultramicroscopy* 131: 70-91

Leigh Jr EG, 1995. The Major Transitions of Evolution. JSTOR.

Limoli DH, Jones CJ, Wozniak DJ (2015) Bacterial extracellular polysaccharides in biofilm formation and function. *Microbiology spectrum* 3: 3.3. 29

Linnes JC, Ma H, Bryers JD (2013) Giant extracellular matrix binding protein expression in Staphylococcus epidermidis is regulated by biofilm formation and osmotic pressure. *Current Microbiology* 66: 627-633

Lövestam S, Koh FA, van Knippenberg B, Kotecha A, Murzin AG, Goedert M, Scheres SH (2022) Assembly of recombinant tau into filaments identical to those of Alzheimer's disease and chronic traumatic encephalopathy. *eLife* 11: e76494

Lučić V, Förster F, Baumeister W (2005) Structural studies by electron tomography: from cells to molecules. *Annual Review of Biochemistry* 74: 833-865

Mah T-FC, O'Toole GA (2001) Mechanisms of biofilm resistance to antimicrobial agents. *Trends in Microbiology* 9: 34-39

Malishev R, Abbasi R, Jelinek R, Chai L (2018) Bacterial model membranes reshape fibrillation of a functional amyloid protein. *Biochemistry* 57: 5230-5238

Mann EE, Wozniak DJ (2012) Pseudomonas biofilm matrix composition and niche biology. *FEMS microbiology reviews* 36: 893-916

Mastronarde DN (2005) Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology* 152: 36-51

Matz C, Webb JS, Schupp PJ, Phang SY, Penesyan A, Egan S, Steinberg P, Kjelleberg S (2008) Marine biofilm bacteria evade eukaryotic predation by targeted chemical defense. *PloS One* 3: e2744

McIntosh R, Nicastro D, Mastronarde D (2005) New views of cells in 3D: an introduction to electron tomography. *Trends in Cell Biology* 15: 43-51

McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R (2013) Analysis tool web services from the EMBL-EBI. *Nucleic Acids Research* 41: W597-W600

Medalia O, Weber I, Frangakis AS, Nicastro D, Gerisch G, Baumeister W (2002) Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science* 298: 1209-1213

Melia CE, Bolla JR, Katharios-Lanwermeyer S, Mihaylov DB, Hoffmann PC, Huo J, Wozny MR, Elfari LM, Böhning J, Morgan AN (2021) Architecture of cell–cell junctions in situ reveals a mechanism for bacterial biofilm inhibition. *Proceedings of the National Academy of Sciences* 118: e2109940118

Mikkelsen H, Ball G, Giraud C, Filloux A (2009) Expression of Pseudomonas aeruginosa CupD fimbrial genes is antagonistically controlled by RcsB and the EAL-containing PvrR response regulators. *PloS One* 4: e6018

Nakane T, Kotecha A, Sente A, McMullan G, Masiulis S, Brown PM, Grigoras IT, Malinauskaite L, Malinauskas T, Miehling J (2020) Single-particle cryo-EM at atomic resolution. *Nature* 587: 152-156

Naydenova K, Russo CJ (2017) Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nature communications* 8: 1-5

Nenninger AA, Robinson LS, Hammer ND, Epstein EA, Badtke MP, Hultgren SJ, Chapman MR (2011) CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation. *Molecular Microbiology* 81: 486-499

Nostro A, Cellini L, Di Giulio M, D'Arrigo M, Marino A, Blanco AR, Favaloro A, Cutroneo G, Bisignano G (2012) Effect of alkaline pH on staphylococcal biofilm formation. *Apmis* 120: 733-742

Nothaft H, Szymanski CM (2010) Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews Microbiology* 8: 765-778

Nuccio S-P, Bäumler AJ (2007) Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiology and Molecular Biology Reviews* 71: 551-575

O'Toole G, Kaplan HB, Kolter R (2000a) Biofilm formation as microbial development. *Annual review of microbiology* 54: 49

O'Toole GA, Gibbs KA, Hager PW, Phibbs Jr PV, Kolter R (2000b) The global carbon metabolism regulator Crc is a component of a signal transduction pathway required for biofilm development by Pseudomonas aeruginosa. *Journal of Bacteriology* 182: 425-431

O'Toole GA, Kolter R (1998) Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. *Molecular Microbiology* 30: 295-304

O'Reilly FJ, Xue L, Graziadei A, Sinn L, Lenz S, Tegunov D, Blötz C, Singh N, Hagen WJ, Cramer P (2020) In-cell architecture of an actively transcribing-translating expressome. *Science* 369: 554-557

Obritsch MD, Fish DN, MacLaren R, Jung R (2005) Nosocomial infections due to multidrug-resistant Pseudomonas aeruginosa: epidemiology and treatment options. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 25: 1353-1364

Oli M, Otoo H, Crowley P, Heim K, Nascimento M, Ramsook C, Lipke P, Brady L (2012) Functional amyloid formation by Streptococcus mutans. *Microbiology* 158: 2903

Olson ME, Ceri H, Morck DW, Buret AG, Read RR (2002) Biofilm bacteria: formation and comparative susceptibility to antibiotics. *Canadian journal of veterinary research* 66: 86

Ortiz JO, Förster F, Kürner J, Linaroudis AA, Baumeister W (2006) Mapping 70S ribosomes in intact cells by cryoelectron tomography and pattern recognition. *Journal of Structural Biology* 156: 334-341

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

Otzen D, Riek R (2019) Functional amyloids. *Cold Spring Harbor perspectives in biology* 11: a033860

Pakharukova N, Garnett JA, Tuittila M, Paavilainen S, Diallo M, Xu Y, Matthews SJ, Zavialov AV (2015) Structural insight into archaic and alternative chaperone-usher pathways reveals a novel mechanism of pilus biogenesis. *PLoS Pathogens* 11: e1005269

Pakharukova N, Malmi H, Tuittila M, Dahlberg T, Ghosal D, Chang Y-W, Myint SL, Paavilainen S, Knight SD, Lamminmäki U *et al* (2022) Archaic chaperone-usher pili self-secrete into superelastic zigzag springs. *Nature*: 1-2

Pakharukova N, Tuittila M, Paavilainen S, Malmi H, Parilova O, Teneberg S, Knight SD, Zavialov AV (2018) Structural basis for Acinetobacter baumannii biofilm formation. *Proceedings of the National Academy of Sciences* 115: 5558-5563

Passos da Silva D, Matwichuk ML, Townsend DO, Reichhardt C, Lamba D, Wozniak DJ, Parsek MR (2019) The Pseudomonas aeruginosa lectin LecB binds to the exopolysaccharide Psl and stabilizes the biofilm matrix. *Nature communications* 10: 1-11

Pawar D, Rossman M, Chen J (2005) Role of curli fimbriae in mediating the cells of enterohaemorrhagic Escherichia coli to attach to abiotic surfaces. *Journal of Applied Microbiology* 99: 418-425

Pei J, Kim B-H, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Research* 36: 2295-2300

Percival S, Knapp J, Wales D, Edyvean R (1999) The effect of turbulent flow and surface roughness on biofilm formation in drinking water. *Journal of Industrial Microbiology and Biotechnology* 22: 152-159

Perrett S, Pham CL, Kwan AH, Sunde M (2014) Functional amyloid: widespread in Nature, diverse in purpose. *Essays in biochemistry* 56: 207-219

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

Pratt LA, Kolter R (1999) Genetic analyses of bacterial biofilm formation. *Current Opinion in Microbiology* 2: 598-603

Prigent-Combaret C, Brombacher E, Vidal O, Ambert A, Lejeune P, Landini P, Dorel C (2001) Complex regulatory network controls initial adhesion and biofilm formation in Escherichia coli via regulation of the csgD gene. *Journal of Bacteriology* 183: 7213-7223

Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* 14: 290-296

Qi Y, Ingólfsson HI, Cheng X, Lee J, Marrink SJ, Im W (2015) CHARMM-GUI martini maker for coarse-grained simulations with the martini force field. *Journal of chemical theory and computation* 11: 4486-4494

Ramezanalizadeh F, Owlia P, Rasooli I (2020) Type I pili, CsuA/B and FimA induce a protective immune response against Acinetobacter baumannii. *Vaccine* 38: 5436-5446

Rasmussen CB, Christiansen G, Vad BS, Lynggaard C, Enghild JJ, Andreasen M, Otzen D (2019) Imperfect repeats in the functional amyloid protein FapC reduce the tendency to fragment during fibrillation. *Protein Science* 28: 633-642

Reichhardt C, Wong C, Passos da Silva D, Wozniak DJ, Parsek MR (2018) CdrA interactions within the Pseudomonas aeruginosa biofilm matrix safeguard it from proteolysis and promote cellular packing. *mBio* 9: e01376-01318

Remaut H, Rose RJ, Hannan TJ, Hultgren SJ, Radford SE, Ashcroft AE, Waksman G (2006) Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted β strand displacement mechanism. *Molecular Cell* 22: 831-842

Remaut H, Tang C, Henderson NS, Pinkner JS, Wang T, Hultgren SJ, Thanassi DG, Waksman G, Li H (2008) Fiber formation across the bacterial outer membrane by the chaperone/usher pathway. *Cell* 133: 640-652

Reysenbach A-L, Ehringer M, Hershberger K (2000) Microbial diversity at 83 C in Calcite Springs, Yellowstone National Park: another environment where the Aquificales and" Korarchaeota" coexist. *Extremophiles* 4: 61-67

Rohou A, Grigorieff N (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* 192: 216-221

Romero D, Aguilar C, Losick R, Kolter R (2010) Amyloid fibers provide structural integrity to Bacillus subtilis biofilms. *Proceedings of the National Academy of Sciences* 107: 2230-2234

Romero D, Sanabria-Valentín E, Vlamakis H, Kolter R (2013) Biofilm inhibitors that target amyloid proteins. *Chemistry & Biology* 20: 102-110

Romero D, Vlamakis H, Losick R, Kolter R (2011) An accessory protein required for anchoring and assembly of amyloid fibres in B. subtilis biofilms. *Molecular Microbiology* 80: 1155-1168

Romero D, Vlamakis H, Losick R, Kolter R (2014) Functional analysis of the accessory protein TapA in Bacillus subtilis amyloid fiber assembly. *Journal of Bacteriology* 196: 1505-1513

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

Ross P, Weinhouse H, Aloni Y, Michaeli D, Weinberger-Ohana P, Mayer R, Braun S, De Vroom E, Van der Marel G, Van Boom J (1987) Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. *Nature* 325: 279-281

Rouse SL, Hawthorne WJ, Berry J-L, Chorev DS, Ionescu SA, Lambert S, Stylianou F, Ewert W, Mackie U, Morgan RML (2017) A new class of hybrid secretion system is employed in Pseudomonas amyloid biogenesis. *Nature communications* 8: 1-13

Rouse SL, Matthews SJ, Dueholm MS (2018) Ecology and biogenesis of functional amyloids in Pseudomonas. *Journal of Molecular Biology* 430: 3685-3695

Rybtke M, Hultqvist LD, Givskov M, Tolker-Nielsen T (2015) Pseudomonas aeruginosa biofilm infections: community structure, antimicrobial tolerance and immune response. *Journal of Molecular Biology* 427: 3628-3645

Saghi Z, Holland DJ, Leary R, Falqui A, Bertoni G, Sederman AJ, Gladden LF, Midgley PA (2011) Three-dimensional morphology of iron oxide nanoparticles with reactive concave surfaces. A compressed sensing-electron tomography (CS-ET) approach. *Nano letters* 11: 4666-4673

Saldaña Z, Xicohtencatl-Cortes J, Avelino F, Phillips AD, Kaper JB, Puente JL, Girón JA (2009) Synergistic role of curli and cellulose in cell adherence and biofilm formation of attaching and effacing Escherichia coli and identification of Fis as a negative regulator of curli. *Environmental Microbiology* 11: 992-1006

Sanders T, Gelb A, Platte RB, Arslan I, Landskron K (2017) Recovering fine details from under-resolved electron tomography data using higher order total variation ℓ1 regularization. *Ultramicroscopy* 174: 97-105

Sauer FG, Remaut H, Hultgren SJ, Waksman G (2004) Fiber assembly by the chaperone–usher pathway. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1694: 259-267

Sauer K, Camper AK, Ehrlich GD, Costerton JW, Davies DG, 2002. Pseudomonas aeruginosa displays multiple phenotypes during development as a biofilm. Am Soc Microbiol.

Sauer K, Stoodley P, Goeres DM, Hall-Stoodley L, Burmølle M, Stewart PS, Bjarnsholt T (2022) The biofilm life cycle: expanding the conceptual model of biofilm formation. *Nature Reviews Microbiology*: 1-13

Scheres SH (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* 180: 519-530

Scheres SH (2020) Amyloid structure determination in RELION-3.1. *Acta Crystallographica Section D: Structural Biology* 76: 94-101

Scheres SH, Zhang W, Falcon B, Goedert M (2020) Cryo-EM structures of tau filaments. *Current Opinion in Structural Biology* 64: 17-25

Schilling JD, Mulvey MA, Hultgren SJ (2001) Structure and function of Escherichia coli type 1 pili: new insight into the pathogenesis of urinary tract infections. *The Journal of infectious diseases* 183: S36-S40

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9: 676-682

Schmid MF, Booth CR (2008) Methods for aligning and for averaging 3D volumes with missing data. *Journal of Structural Biology* 161: 243-248

Schwartz K, Syed AK, Stephenson RE, Rickard AH, Boles BR (2012) Functional amyloids composed of phenol soluble modulins stabilize Staphylococcus aureus biofilms. *PLoS Pathogens* 8: e1002744

Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* 31: 3381-3385

Secor PR, Sweere JM, Michaels LA, Malkovskiy AV, Lazzareschi D, Katznelson E, Rajadas J, Birnbaum ME, Arrigoni A, Braun KR (2015) Filamentous bacteriophage promote biofilm assembly and function. *Cell Host & Microbe* 18: 549-559

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: e00103-00113

Serrano Mn, Zilhão R, Ricca E, Ozin AJ, Moran Jr CP, Henriques AO (1999) A Bacillus subtilis secreted protein with a role in endospore coat assembly and function. *Journal of Bacteriology* 181: 3632-3643

Sharma D, Misba L, Khan AU (2019) Antibiotics versus biofilm: an emerging battleground in microbial communities. *Antimicrobial Resistance & Infection Control* 8: 1-10

Shibata S, Shoji M, Okada K, Matsunami H, Matthews MM, Imada K, Nakayama K, Wolf M (2020) Structure of polymerized type V pilin reveals assembly mechanism involving protease-mediated strand exchange. *Nature Microbiology* 5: 830-837

Sipe JD, Benson MD, Buxbaum JN, Ikeda S-I, Merlini G, Saraiva MJ, Westermark P (2010) Amyloid fibril protein nomenclature: 2010 recommendations from the nomenclature committee of the International Society of Amyloidosis. *Amyloid* 17: 101-104

Sleutel M, Pradhan B, Remaut H (2022) Structural analysis of the bacterial amyloid curli. *bioRxiv*

Smith JF, Knowles TP, Dobson CM, MacPhee CE, Welland ME (2006) Characterization of the nanoscale properties of individual amyloid fibrils. *Proceedings of the National Academy of Sciences* 103: 15806-15811

Stanisich JJ, Zyla DS, Afanasyev P, Xu J, Kipp A, Olinger E, Devuyst O, Pilhofer M, Boehringer D, Glockshuber R (2020) The cryo-EM structure of the human uromodulin filament core reveals a unique assembly mechanism. *eLife* 9: e60265

Stanley PM (1983) Factors affecting the irreversible attachment of Pseudomonas aeruginosa to stainless steel. *Canadian journal of microbiology* 29: 1493-1499

Stepanović S, Ćirković I, Mijač V, Švabić-Vlahović M (2003) Influence of the incubation temperature, atmosphere and dynamic conditions on biofilm formation by Salmonella spp. *Food Microbiology* 20: 339-343

Stoodley P, Dodds I, Boyle JD, Lappin-Scott H (1998) Influence of hydrodynamics and nutrients on biofilm structure. *Journal of Applied Microbiology* 85: 19S-28S

Stoodley P, Sauer K, Davies DG, Costerton JW (2002) Biofilms as complex differentiated communities. *Annual Review of Microbiology* 56: 187-209

Sulkowski N, 2021. Structural studies of bacterial adhesion. University of Oxford.

Sutherland IW (2001) Biofilm exopolysaccharides: a strong and sticky framework. *Microbiology* 147: 3-9

Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, Pulcini C, Kahlmeter G, Kluytmans J, Carmeli Y (2018) Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases* 18: 318-327

Taglialegna A, Lasa I, Valle J (2016) Amyloid structures as biofilm matrix scaffolds. *Journal of Bacteriology* 198: 2579-2588

Talà L, Fineberg A, Kukura P, Persat A (2019) Pseudomonas aeruginosa orchestrates twitching motility by sequential control of type IV pili movements. *Nature Microbiology* 4: 774-780

Tarafder AK, von Kügelgen A, Mellul AJ, Schulze U, Aarts DG, Bharat TA (2020) Phage liquid crystalline droplets form occlusive sheaths that encapsulate and protect infectious rod-shaped bacteria. *Proceedings of the National Academy of Sciences* 117: 4724-4731

Taylor JD, Zhou Y, Salgado PS, Patwardhan A, McGuffie M, Pape T, Grabe G, Ashman E, Constable SC, Simpson PJ (2011) Atomic resolution insights into curli fiber biogenesis. *Structure* 19: 1307-1316

Tegunov D, Cramer P (2019) Real-time cryo-electron microscopy data preprocessing with Warp. *Nature Methods* 16: 1146-1152

Tegunov D, Xue L, Dienemann C, Cramer P, Mahamid J (2021) Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nature Methods* 18: 186-193

Terra R, Stanley-Wall NR, Cao G, Lazazzera BA (2012) Identification of Bacillus subtilis SipW as a bifunctional signal peptidase that controls surface-adhered biofilm formation. *Journal of Bacteriology* 194: 2781-2790

Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H (2022) SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*: 1-3

Tomaras AP, Dorsey CW, Edelmann RE, Actis LA (2003) Attachment to and biofilm formation on abiotic surfaces by Acinetobacter baumannii: involvement of a novel chaperone-usher pili assembly system. *Microbiology* 149: 3473-3484

True HL, Lindquist SL (2000) A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* 407: 477-483

Turoňová B, Marsalek L, Davidovič T, Slusallek P (2015) Progressive stochastic reconstruction technique (PSRT) for cryo electron tomography. *Journal of Structural Biology* 189: 195-206

Uchański T, Pardon E, Steyaert J (2020) Nanobodies to study protein conformational states. *Current Opinion in Structural Biology* 60: 117-123

Vallet I, Diggle SP, Stacey RE, Cámara M, Ventre I, Lory S, Lazdunski A, Williams P, Filloux A (2004) Biofilm formation in Pseudomonas aeruginosa: fimbrial cup gene clusters are controlled by the transcriptional regulator MvaT. *Journal of Bacteriology* 186: 2880-2890

Vallet I, Olson JW, Lory S, Lazdunski A, Filloux A (2001) The chaperone/usher pathways of Pseudomonas aeruginosa: identification of fimbrial gene clusters (cup) and their involvement in biofilm formation. *Proceedings of the National Academy of Sciences* 98: 6911-6916

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. *Journal of computational chemistry* 26: 1701-1718

Van der Verren SE, Van Gerven N, Jonckheere W, Hambley R, Singh P, Kilgour J, Jordan M, Wallace EJ, Jayasinghe L, Remaut H (2020) A dual-constriction biological nanopore resolves homonucleotide sequences with high fidelity. *Nature Biotechnology* 38: 1415-1420

Van Schaik EJ, Giltner CL, Audette GF, Keizer DW, Bautista DL, Slupsky CM, Sykes BD, Irvin RT (2005) DNA binding: a novel function of Pseudomonas aeruginosa type IV pili. *Journal of Bacteriology* 187: 1455-1464

Vert M, Doi Y, Hellwich K-H, Hess M, Hodge P, Kubisa P, Rinaudo M, Schué F (2012) Terminology for biorelated polymers and applications (IUPAC Recommendations 2012). *Pure and Applied Chemistry* 84: 377-410

Villa E, Schaffer M, Plitzko JM, Baumeister W (2013) Opening windows into the cell: focused-ion-beam milling for cryo-electron tomography. *Current Opinion in Structural Biology* 23: 771-777

Vlamakis H, Chai Y, Beauregard P, Losick R, Kolter R (2013) Sticking together: building a biofilm the Bacillus subtilis way. *Nature Reviews Microbiology* 11: 157-168

Waksman G, Hultgren SJ (2009) Structural biology of the chaperone–usher pathway of pilus biogenesis. *Nature Reviews Microbiology* 7: 765-774

Wan W, Briggs JA (2016) Cryo-electron tomography and subtomogram averaging. *Methods in Enzymology* 579: 329-367

Wang F, Cvirkaite-Krupovic V, Krupovic M, Egelman EH (2022) Archaeal bundling pili of Pyrobaculum calidifontis reveal similarities between archaeal and bacterial biofilms. *Proceedings of the National Academy of Sciences* 119: e2207037119

Warshamanage R, Yamashita K, Murshudov GN (2022) EMDA: a python package for electron microscopy data analysis. *Journal of Structural Biology* 214: 107826

Watanabe R, Buschauer R, Böhning J, Audagnotto M, Lasker K, Lu T-W, Boassa D, Taylor S, Villa E (2020) The in situ structure of Parkinson's disease-linked LRRK2. *Cell* 182: 1508-1518. e1516

Watnick PI, Fullner KJ, Kolter R (1999) A role for the mannose-sensitive hemagglutinin in biofilm formation by Vibrio cholerae El Tor. *Journal of Bacteriology* 181: 3606-3609

Webb B, Sali A (2017) Protein structure modeling with MODELLER. In: *Functional genomics*, pp. 39-54. Springer:

Wei Q, Ma LZ (2013) Biofilm matrix and its regulation in Pseudomonas aeruginosa. *International Journal of Molecular Sciences* 14: 20983-21005

Weiss GL, Stanisich JJ, Sauer MM, Lin C-W, Eras J, Zyla DS, Trück J, Devuyst O, Aebi M, Pilhofer M (2020) Architecture and function of human uromodulin filaments in urinary tract infections. *Science* 369: 1005-1010

Westall F, de Wit MJ, Dann J, van der Gaast S, de Ronde CE, Gerneke D (2001) Early Archean fossil bacteria and biofilms in hydrothermally-influenced sediments from the Barberton greenstone belt, South Africa. *Precambrian Research* 106: 93-116

Whitchurch CB, Tolker-Nielsen T, Ragas PC, Mattick JS (2002) Extracellular DNA required for bacterial biofilm formation. *Science* 295: 1487-1487

Yip KM, Fischer N, Paknia E, Chari A, Stark H (2020) Atomic-resolution protein structure determination by cryo-EM. *Nature* 587: 157-161

Zachs T, Schertel A, Medeiros J, Weiss GL, Hugener J, Matos J, Pilhofer M (2020) Fully automated, sequential focused ion beam milling for cryo-electron tomography. *eLife* 9: e52286

Zeighami H, Valadkhani F, Shapouri R, Samadi E, Haghi F (2019) Virulence characteristics of multidrug resistant biofilm forming Acinetobacter baumannii isolated from intensive care unit patients. *BMC infectious diseases* 19: 1-9

Zeng G, Vad BS, Dueholm MS, Christiansen G, Nilsson M, Tolker-Nielsen T, Nielsen PH, Meyer RL, Otzen DE (2015) Functional bacterial amyloid increases Pseudomonas biofilm hydrophobicity and stiffness. *Frontiers in Microbiology* 6: 1099

Zhang W, Falcon B, Murzin AG, Fan J, Crowther RA, Goedert M, Scheres SH (2019) Heparin-induced tau filaments are polymorphic and differ from those in Alzheimer’s and Pick’s diseases. *eLife* 8: e43584

Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* 14: 331-332

Zhong ED, Bepler T, Berger B, Davis JH (2021) CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods* 18: 176-185

Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, Scheres SH (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* 7: e42166

Zivanov J, Otón J, Ke Z, Qu K, Morado D, Castaño-Díez D, von Kügelgen A, Bharat T, Briggs J, Scheres S (2022) A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *bioRxiv*: 2022.2002.2028.482229

9. Appendix

Figures S1-4

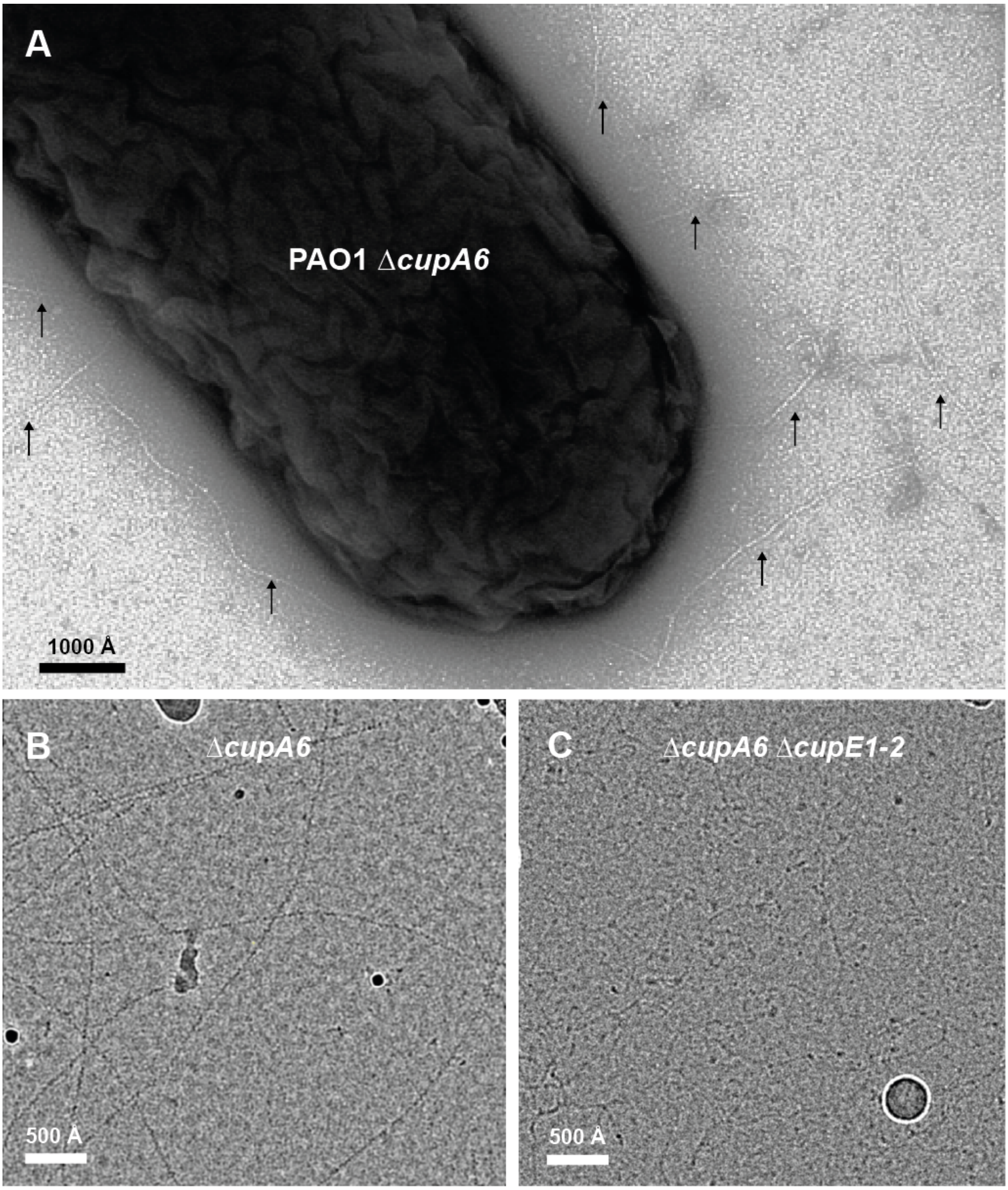


Figure S1: Previous work on CupE1. Images were acquired by Nina Sulkowski (Sulkowski, 2021). A) A *∆cupA6* strain of *P. aeruginosa* PAO1 (TBS060) over-expresses pili that can be sheared of the cell surface. B-C) Purified pili were found to be corresponding to the product of the *cupE* operon, as their presence depends on *cupE1-2*. Adapted from (Böhning *et al.*, 2022b).

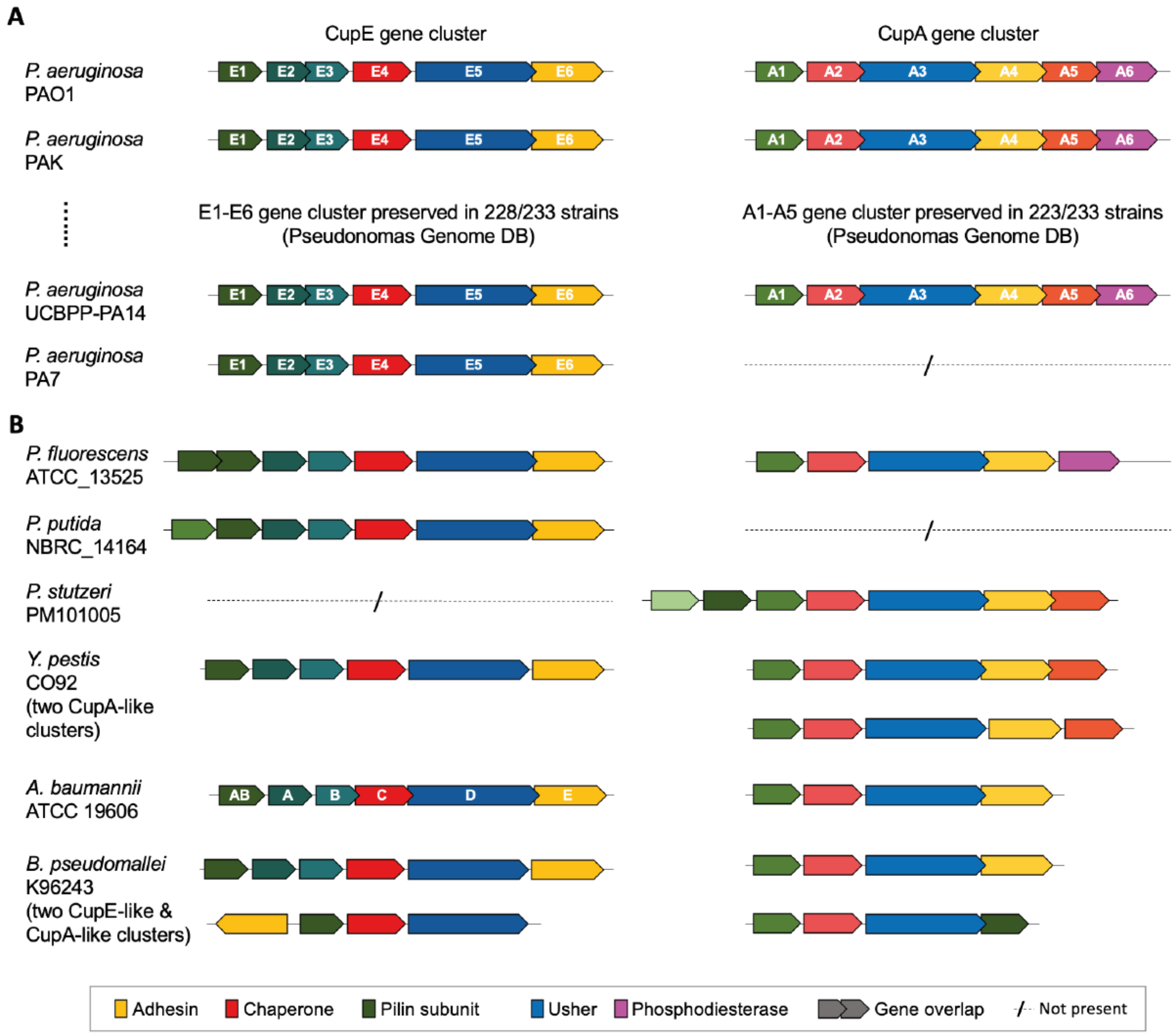


Figure S2: CupE and CupA clusters typically co-occur in genomes. A) Bioinformatic analysis suggests CupE and CupA co-occur in most *P. aeruginosa* strains. Overlapping sequences may indicate translational coupling. B) CupA und CupE homologs co-occur in numerous proteobacteria, including *P. fluorescens*, *Y. pestis, A. baumannii* and *B. pseudomallei*, but gene order is not preserved outside *P. aeruginosa*. This analysis was performed by Adrian Dobbelstein and Vikram Alva and is published (Böhning *et al.*, 2022b).

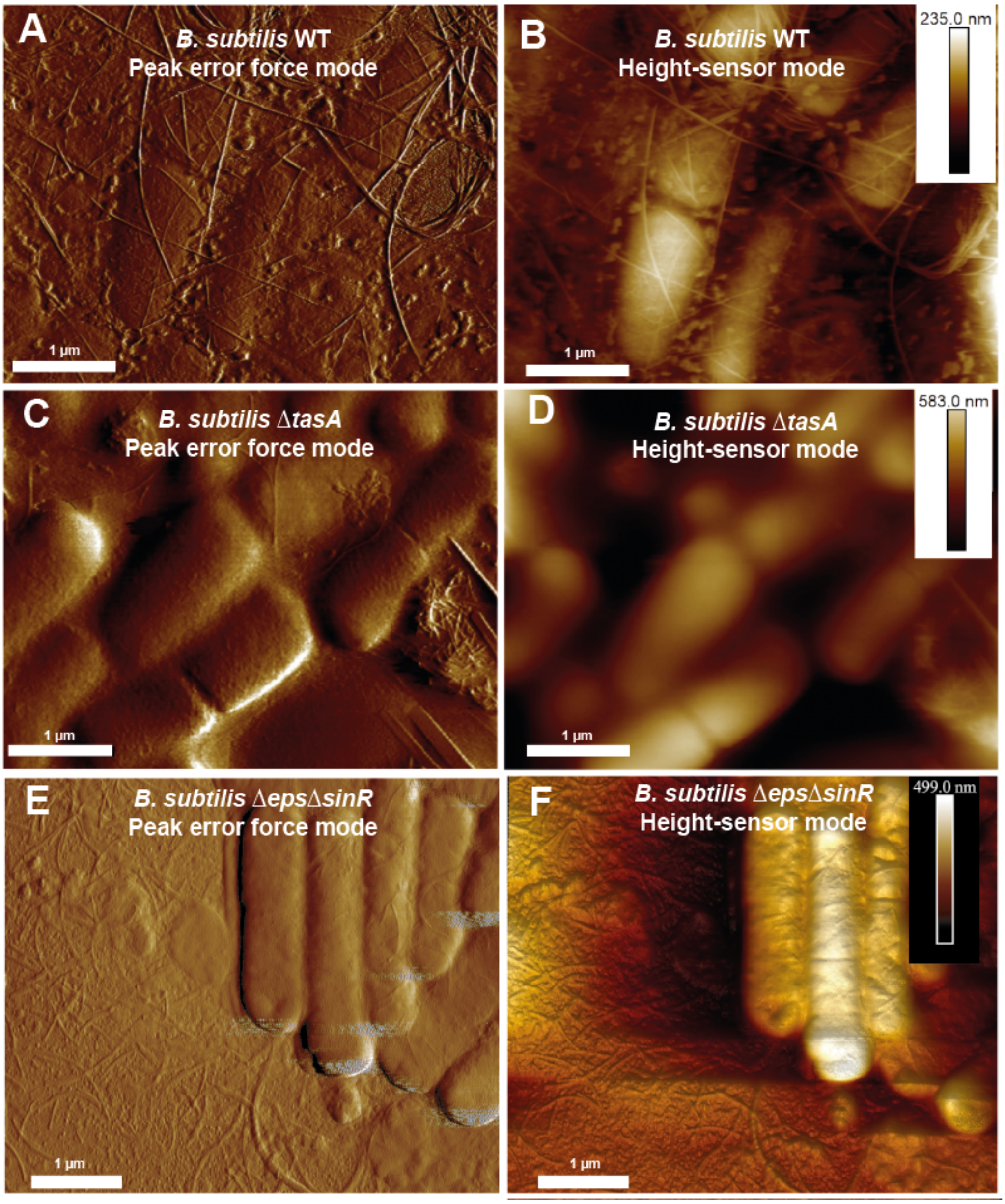


Figure S3: AFM imaging of *B. subtilis* biofilms. Peak force error and height sensor mode images of A-B) WT (ZK0541, expressing TasA-mCherry), C-D) Δ*tasA* (ZK3657) and E-F) Δ*eps* Δ*sinR* pellicles. This experiment was conducted by Mnar Ghrayeb and Prof. Liraz Chai. Panels A-D are published (Böhning *et al.*, 2022c).

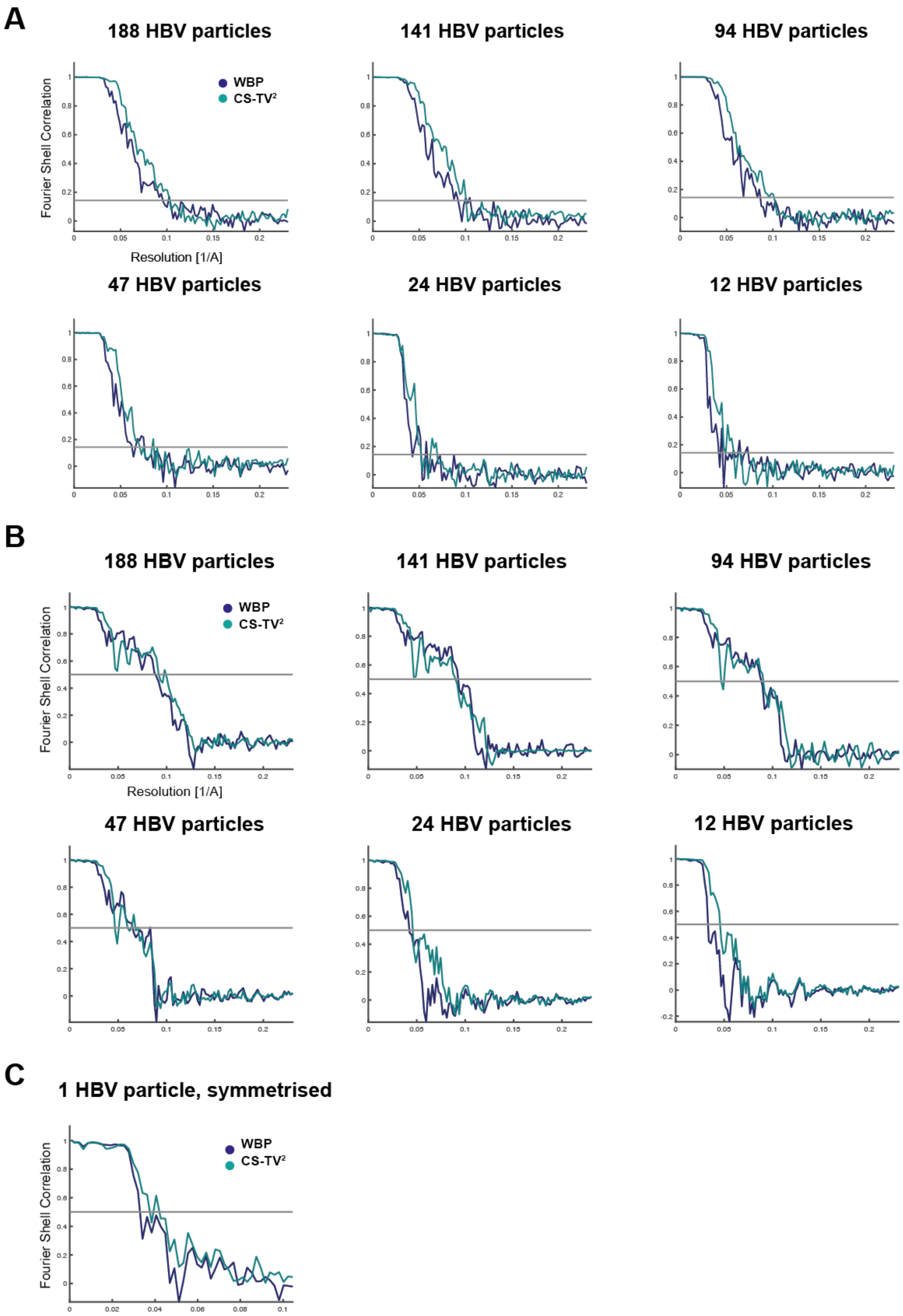


Figure S4: Fourier Shell Correlation (FSC) plots of STA of HBV capsids. A)

Gold-standard FSC plots (0.143 cut-off, FSC of independent half-maps) of STA generated using varying particle numbers. B) Model-vs-map FSC plots (0.5 cut-off) versus PDB 6HTX for STA maps using varying particle numbers. C) Model-vs-map FSC plot as in B) for a single, symmetrised HBV capsid particle. This data is published in Böhning *et al.* (2022a).