THE UNIVERSITY OF SHEFFIELD

DOCTORAL THESIS

Structural studies of the TssA and TssK subunits from the type VI secretion system

---

Author:
Samuel R. DIX

Supervisor:
Prof. David W. RICE

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

in the

X-ray Crystallography group
Department of Molecular Biology and Biotechnology

February 14, 2019

Declaration of Authorship

I, Samuel R. DIX, declare that this thesis titled, "Structural studies of the TssA and TssK subunits from the type VI secretion system" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date:

_____

Abstract

The type VI secretion system (T6SS) is a protein assembly involved in the injection of effector proteins into target cells. The T6SS is critical for inter-bacterial competition, and also play a role in eukaryotic virulence. The T6SS assembly consists of a cell membrane complex to which a bacteriophage-like contractile tail is attached, mediated through a baseplate complex. Currently, very little information is known about the structural arrangement of the T6SS baseplate, in particular the components TssA and TssK. This thesis discusses the structural and complementary interaction studies carried out for both components from various bacterial orthologues.

TssA has been classified into four sub-clades (TssA1^{A}, TssA1^{B}, TssA2^{A} and TssA2^{B}). Subsequent structural studies have provided a model for the conserved ImpA_N region, found in the Nt1 domain of all TssA's. In addition, the structure for the CTD oligomer from a TssA1^{B} representative (*B. cenocepacia*) and TssA2^{A} representative (*A. hydrophila*) were determined, demonstrating the broad range of symmetry and mobility that can be exhibited by TssA. Further interaction studies indicate that despite the radically different symmetries displayed by the various TssA orthologue CTD oligomers, they maintain similar interactions with other T6SS subunits.

The second component, TssK, has also been structurally characterised from both *B. cenocepacia* and *B. pseudomallei*. The X-ray structure of *B. cenocepacia* TssK indicates the subunits assemble as a trimer and are formed of three domains the shoulder, neck and head, displaying structural similarity with *Lactococcal* siphophage receptor binding proteins, consistent with *E. coli* TssK. In addition, interaction studies of *B. cenocepacia* TssK with other T6SS components suggests a similar pattern of interaction as *E. coli* TssK. Furthermore, structural comparisons of *B. cenocepacia* TssK with *E. coli* TssK identifies distinct differences in the orientation of the head domain, suggesting that mobility is an important feature in the role of TssK.

Acknowledgements

During this body of work, many people have been critical in helping me complete this PhD. Firstly, I would like to thank my supervisor Prof. David Rice for the opportunity to work within the crystallography group and for multiple inspiring conversations over the years. In addition, his guidance and patience with me, especially through the writing process of the manuscripts and this thesis. I would also like to thank Dr Patrick Baker, for the numerous times he has helped my understanding and many fun evenings climbing! Many thanks must go to Dr John Rafferty, for not only being my advisor throughout the project but also for being the one who first introduced me to the lab and helped me cut my teeth in crystallography. I would like to thank all of those that have helped contribute to the work on the type VI secretion system project, especially Dr Mark Thomas, Dr Hayley Owen, Dr Svetomir Tzokov and Prof. Per Bullough. A lot of blood, sweat and tears have gone into the work on TssA and TssK over the years, but hopefully, it was all worth it. Furthermore, I would like to thank Dr Svetlana Sedelnikova for her extensive knowledge of protein purification and guidance. In addition, many thanks to Fiona Rodgers whose tireless efforts keep the lab running and suitably caffeinated, as well as everyone else in the crystallography group, including Adli Aziz, Alicia Churchill-Angus, Azura Mohd Noor, in addition, our previous members Dr Jason Wilson and Dr Claudine Bisson. You guys have truly made it a great group to work with; there is a certain bond you share after many sleep deprived diamond trips. Furthermore, I would like to thank the project students I have had over the years, notably Matthew Harris, Amy Milburn and Carmen Apostol, I wish you the best in your PhDs. I would also like to thank Angus Robertson and Henry Wood of the NMR group for their timely distractions, especially when it has been a tough day. To my friends who have put up with me over the years, thank you all, you know who you are. I'd like to give a massive thank you to my girlfriend, Amy, who has put up with a lot throughout this time but has always been patient and caring. I hope I can offer the same support throughout the rest of her PhD as she has for me. Finally, to my parents and grandmother, who have supported me tirelessly over the years, and have been far too neglected during this process, which I endeavour to rectify.

Contents

List of Figures

List of Tables

List of Abbreviations

APS	Ammonium Persulphate
BHI	Brain Heart Infusion
BlaM	Beta-lactamase
CC	Correlation Coefficient
CFE	Cell Free Extract
CTD	C-terminal Domain
CV	Column Volume
DEAE	Diethylaminoethyl
DLS	Diamond Light Source
DNA	Deoxyribonucleic Acid
EAEC	Enteroaggregative *E. coli*
ECT	Electron Cryotomography
EDTA	EthyleneDiamineTetraAcetic acid
EMP	Ethylmercury Phosphate
FOM	Figure Of Merit
gDNA	Genomic DNA
Gsp	General Secretion Proteins
hcp	Haemolysin Coregulated Protein
idr	Rhs-related Proteins
ids	Self-identity Proteins
IM	Inner Membrane
IMAC	Immobilized Metal Affinity Chromatography
IMC	Inner Membrane Complex
IMP	Inner Membrane Platform
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny Broth

LLG	Log-likelihood Gain
MCS	Multiple Cloning Site
NCS	Non-Crystallographic Symmetry
NEB	New England Biolabs
Nt1	N-terminal Domain 1
Nt2	N-terminal Domain 2
OD	Optical Density
OM	Outer Membrane
OMC	Outer Membrane Complex (T2SS)
PAAR	Proline-Alanine-Alanine-Proline
PAP	Periplasmic Adapter Protein
PCR	Polymerase Chain Reaction
PMSF	Phenylmethane Sulfonyl Fluoride
PPP	Periplasmic Pseudopilus
RBP	Receptor Binding Protein
Rhs	Recombinant Hotspot
RMSD	Root Mean Square Deviation
RND	Resistance Nodulation Division
RPM	Revolutions Per Minute
SAD	Single Wavelength Anomalous Dispersion
SAX	Small Angle Scattering X-ray
SEC	Size Exclusion Chromatography
SEC-MALLS	Size Exclusion Chromatography - Multi-Angle Laser Light Scattering
SDS-PAGE	Sodium Dodecyl Sulfate-polyacrylamide Gel Electrophoresis
SLS	Sodium Lauryl Sulfate
SOC	Super Optimal Broth with Catabolite Repression
T1SS	Type One Secretion System
T2SS	Type Two Secretion System
T3SS	Type Three Secretion System
T4SS	Type Four Secretion System
T5SS	Type Five Secretion System

T6SS	Type Six Secretion System
TAE	Tris Acetic Acid EDTA
Tae	Amidase Effector Proteins
TEM	Transmission electron microscopy
TEMED	N,N,N′,N′-tetramethylethane-1,2-diamine
TFZ	Translation Function
Tle	Lipase Effector Proteins
Tm	Melting Temperature
TolC	Outer Membrane Component (T1SS)
tss	Type Six Secretion
UV	Ultra Violet
YFP	Yellow Fluorescent Protein

Publications

1. S. R. Dix, R. Sun, M. J. Harris, S. L. Batters, S. E. Sedelnikova, P. J. Baker, M. S. Thomas and D. W. Rice. TssA from *Aeromonas hydrophila*: Expression, purification and crystallographic studies. Acta Crystallogr. Sect. F Struct. Biol. Commun. 74, 578–582 (2018). (Dix et al., 2018a).

2. S. R. Dix, H. J. Owen, R. Sun, A. Ahmad, S. Shastri, H. L. Spiewak, D. J. Mosby, M. J. Harris, S. L. Batters, T. A. Brooker, S. B. Tzokov, S. E. Sedelnikova, P. J. Baker, P. A. Bullough, D. W. Rice, M. S. Thomas. Structural insights into the function of type VI secretion system TssA subunits. Nat. Commun. 9, 4765 (2018). (Dix et al., 2018b).

3. S. R. Dix, H. J. Owen, R. Sun, A. E. Milburn, S. E. Sedelnikova, P. J. Baker, M. S. Thomas and D. W. Rice. TssK from *Burkholderia cenocepacia*: Expression, purification and preliminary crystallographic studies, **prepared for submission**.

4. S. R. Dix, R. Sun, A. Ahmad, A. E. Milburn, H. J. Owen, S. Nathan, F. Raih, S. E. Sedelnikova, P. J. Baker, M. S. Thomas, and D. W. Rice. The role of mobility and sequence conservation in the structure/function relationships of *Burkholderia* TssK, **prepared for submission**.

Papers that I have co-authored on during my studies - Not presented in this thesis

---

5. Rosa, L. T., Dix, S. R., Rafferty, J. B. Kelly, D. J. Structural basis for high-affinity adipate binding to AdpC (RPA4515), an orphan periplasmic-binding protein from the tripartite tricarboxylate transporter (TTT) family in *Rhodopseudomonas palustris*. FEBS J. 284, 4262–4277 (2017). (Rosa et al., 2017).

6. Rosa, L. T., Dix, S. R., Rafferty, J. B. Kelly, D. J. A New Mechanism for High-Affinity Uptake of C4-Dicarboxylates in Bacteria Revealed by the Structure of Rhodopseudomonas palustris MatC (RPA3494), a Periplasmic Binding Protein of the Tripartite Tricarboxylate Transporter (TTT) Family. J. Mol. Biol. 431, 351–367 (2019). (Rosa et al., 2019).

Chapter 1

Introduction

1.1 Thesis introduction

Bacteria have evolved various mechanisms for the transport of biomolecules across the membrane. These take the form of secretion systems and, as will be described, there are six established systems that can be found in Gram-negative bacteria, with the most recently discovered: the type VI secretion system (T6SS). This introductory chapter discusses the current understanding for the T6SS, in particular its architecture, relationship with phage injection machinery and its role in pathogenicity. The latter part of this chapter describes, in detail, the protein components of the T6SS, and discusses the work carried out on one particular sub-complex, the baseplate, of which two components are the topic of this thesis.

1.2 Gram-negative bacterial secretion systems

1.2.1 General overview of bacterial secretion systems

Gram-negative bacteria are comprised of two phospholipid membranes (inner and outer) along with a peptidoglycan cell wall present between them in the periplasm, generating a multi-layered cell envelope with a thickness of $\sim$ 40 nm (400 Å) (Matias et al., 2003). Consequently, this provides a significant barrier for the passage of small molecules, DNA and proteins, which are substrates for critical mechanisms within the cell, including: bacterial survival, adaptation, surface adhesion and pathogenicity (Costa et al., 2015). Therefore, bacteria have evolved multiple secretion systems to allow for their transport.

Currently, in addition to the widely distributed Sec and Tat protein export systems (Palmer and Berks, 2012; Lycklama a Nijeholt and Driessen, 2012), six major secretion systems have

been discovered in Gram-negative bacteria (T1SS - T6SS). These can be categorised as systems which span both lipid membranes, as is the case for T1SS, T2SS, T3SS, T4SS and T6SS, or span only the outer membrane, as in the T5SS.

The type I secretion system (T1SS)

The T1SS has been shown to be involved in the secretion of virulence components and nutrient acquisition proteins (Kanonenberg, Schwarz, and Schmitt, 2013). They have a similar architecture to resistance-nodulation-division (RND) multi-drug efflux pumps, whereby the T1SS is formed as a three-part system, consisting of the inner membrane component (IMC), periplasmic adapter protein (PAP), and outer membrane component (TolC), which together span the bacterial double-membrane (figure 1.1) (Piddock, 2006). The IMC is a member of the ABC transporter family and is involved in substrate recognition (Kanonenberg, Schwarz, and Schmitt, 2013). This is connected to the PAP, which enables the handoff of substrate from the IMC to TolC (Balakrishnan, Hughes, and Koronakis, 2001; Delepelaire, 2004). The structure of TolC is comprised of an α-barrel which interacts with the PAP and a β-barrel to allow transport of the substrate to the extracellular environment, through the outer-membrane (Koronakis, Eswaran, and Hughes, 2004; Pei et al., 2011).

The type II secretion system (T2SS)

The T2SS system is widely found in both pathogenic and non-pathogenic Gram-negative bacteria(Nivaskumar and Francetic, 2014). It is able to secrete folded proteins, often associated with cellular survival in hosts, from the periplasm to the extracellular environment (Nivaskumar and Francetic, 2014). The T2SS is comprised of between 12-15 general secretion proteins (Gsp), which are organised into four distinct sub-complexes: the outer-membrane complex (OMC), periplasmic pseudopilus (PPP), the inner membrane platform (IMP) and the cytoplasmic ATPase (figure 1.1) (Korotkov, Sandkvist, and Hol, 2012). The OMC is comprised of a dodecameric assembly of GspD, which forms the channel through the outer membrane (Korotkov, Sandkvist, and Hol, 2012). This is able to interact with the second sub-complex, the IMP, which is formed from four membrane proteins: GspC, GspF, GspL and GspM (Abendroth et al., 2004; Abendroth et al., 2005; Abendroth et al., 2009), where GspC mediates the interaction with GspD, hence links the OMC with the IMP (Korotkov

et al., 2011). The third sub-complex is the pseudopilus, which extends from the inner membrane into the periplasm, and is formed from GspGHIJK (Campos et al., 2010; Cisneros, Pehau-Arnaudet, and Francetic, 2012). The interaction of the pseudopilus with GspC and GspD enables substrate to be pushed through the OMC by the extension of the pseudopilus (Korotkov et al., 2011; Nivaskumar et al., 2014). The fourth and final sub-complex is the ATPase (GspE) of the T2SS, which is located within the cytoplasm associated to the IMP through interactions with GspF and GspL, where it provides energy for the system (Py, Loiseau, and Barras, 2001; Korotkov et al., 2011).

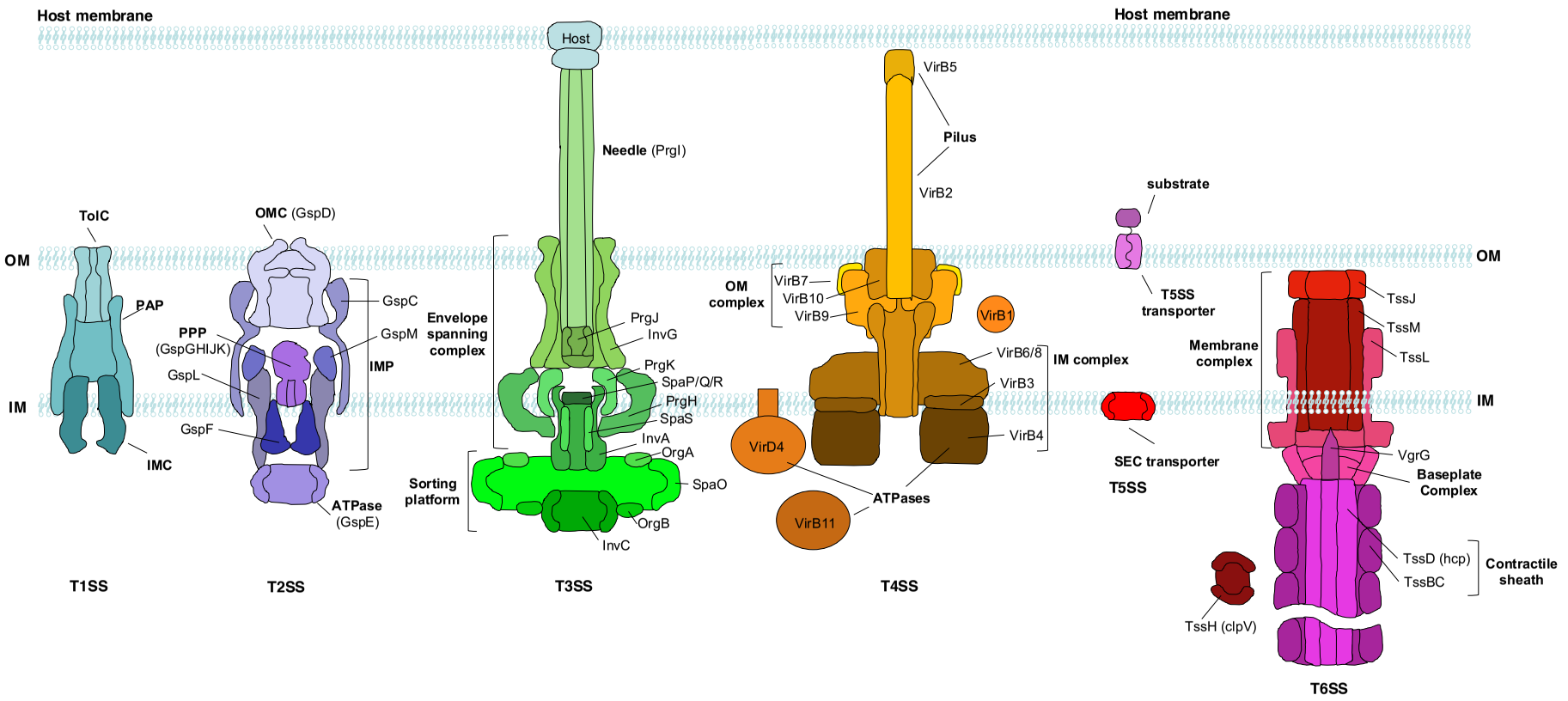


FIGURE 1.1: **The T1SS, T2SS, T3SS, T4SS, T5SS and T6SS of Gram-negative bacteria. T1SS)** IMC - Inner membrane component, PAP - periplasmic adapter protein, TolC - outer membrane component. **T2SS)** OMC - outer membrane complex (GspD), PPP - periplasmic pseudopilus (GspGHIJK), IMP - inner membrane platform (GspCFLM), ATPase (GspE). **T3SS)** Needle - PrgI, Envelope spanning complex (PrgHJK, InvAG, SpaPQRS), sorting platform (InvAC, OrgAB, SpaO). **T4SS)** Pilus (VirB2/5), OM complex - Outer membrane complex (VirB7/9/10), IM complex - inner membrane complex (VirB3/6/8/10), ATPases (VirB1/11, VirD4). **T5SS)** transporter - outer membrane domain, SEC transporter - general secretion pathway. **T6SS)** Membrane complex - TssJLM, Baseplate complex - TssEFGK and VgrG, Contractile sheath - TssBC, AAATpase - TssH. Figure modified from Costa et al., 2015.

The type III secretion system (T3SS)

The T3SS is often involved in the transfer of bacterial effector proteins to the cytoplasm of eukaryotic host cells; consequently, this system is often found in pathogenic Gram-negative bacteria (Galan and Wolf-watz, 2006; Cornelis, 2006). The T3SS is comprised of $\sim$ 25 proteins, which are arranged in three sub-complexes: the needle filament, cell envelope spanning assembly and the sorting platform, forming an overall structure termed the injectisome (figure 1.1) (Schraidt and Marlovits, 2011; Marlovits et al., 2004; Marlovits et al., 2006). PrgK and PrgH form a two-layered complex, through the inner membrane, to which SpaP, SpaQ, SpaR, SpaS and InvA associate to the inner channel formed by the two proteins (Kubori et al., 1998; Worrall, Lameignere, and Strynadka, 2011). InvG sits above the PrgK/PrgH inner membrane complex, where it spans the cell envelope, before inserting into the outer membrane forming the outer membrane channel (Schraidt and Marlovits, 2011). The needle filament is then able to pass through this channel and extend into the extracellular space to interact with target cells (Kubori et al., 2000; Kimbrough and Miller, 2000). The filament is formed of repeating units of PrgI and the minor component PrgJ (Zhong et al., 2012), which together form a helical-like structure with a lumen of $\sim$ 25 Å in diameter, indicating that substrates are translocated in an unfolded state (Kimbrough and Miller, 2000; Loquet et al., 2012; Radics, Königsmaier, and Marlovits, 2014). To enable efficient translocation of unfolded substrate, the T3SS cell envelope spanning assembly has an associated sorting platform, the final sub-complex of the T3SS. This platform is formed from SpaO, OrgA, OrgB and InvC and aids in localising chaperone-substrate complexes to the T3SS for their subsequent secretion (Lara-Tejero et al., 2011; Hu et al., 2015).

The type IV secretion system (T4SS)

The T4SS is found throughout both Gram-negative and Gram-positive bacteria and, unlike the other secretion systems, it can secrete not only protein substrates but also DNA, allowing for their translocation into target bacterial and eukaryotic cells. Consequently, the T4SS is a common mechanism for the conjugation of plasmid DNA between bacteria, for example, factors that encode for antibiotic resistance genes (Alvarez-Martinez and Christie, 2009). The T4SS consists of 12 core components VirB1-11 and VirD4, which are organised into two sub-complexes, the translocation assembly and the pilus (figure 1.1).

The translocation assembly contains VirB3 and VirB6-10 (Christie, Whitaker, and González-Rivera, 2014), which can be further sub-divided into two modules. The first, the outer membrane complex, consists of 14 copies of VirB7, VirB9 and the C-terminal domains of VirB10, with VirB10 generating the pore for substrate translocation (Chandran et al., 2009). The second, the inner membrane complex, consists of 14 copies of the N-terminal domains of VirB10, 24 copies of VirB6, possibly VirB3, 12 copies of VirB8 and 12 copies of VirB4; however, VirB4 is present on the cytoplasmic face of the inner membrane (Low et al., 2014). VirB10 is the only protein in the translocation assembly which spans the whole bacterial envelope (Cascales and Christie, 2004); however, this connection has yet to be structurally determined. The second sub-complex, the pilus, is formed by VirB2 and VirB5 and extends out from the cell into the extracellular space (Trokter et al., 2014). The system also contains three associated ATPases to power it: VirD4, VirB4 and VirB11 (Trokter et al., 2014). The final core member of the T4SS is VirB1, a trans-glycosylase located in the periplasm responsible for degrading the local peptidoglycan layer to allow the formation of the pilus (Trokter et al., 2014).

The type V secretion system (T5SS)

The T5SS is often associated with biofilm formation and cell-to-cell adhesion (Leyton, Rossiter, and Henderson, 2012). Unlike the other systems, the T5SS is not a dedicated secretion system *per se*. The secretion apparatus, in the form of a transporter domain, is fused to the substrate and hence forms a single polypeptide; as a consequence. The members of this group are named the autotransporters (Leo, Grin, and Linke, 2012). T5SS proteins utilise the SEC secretion pathway to enter the periplasm unfolded, whereby the transporter domain folds to form a β-barrel for insertion into the outer membrane and the substrate passes through it (figure 1.1) (Junker, Besingi, and Clark, 2009).

1.2.2 Discovery of the T6SS

A putative, novel, bacterial secretion system was first described in 2003 when Bladergroen and co-workers discovered a cluster of 'impaired in nitrogen fixation' (imp) genes, which, when mutated, disrupted the symbiotic relationship *Rhizobium leguminosarum* had with its host. In addition, this imp-like gene cluster was widely distributed within Gram-negative bacterial species and was proposed to be involved in the secretion of proteins (Bladergroen,

Badelt, and Spaink, 2003). Subsequently, these genes were shown to be involved in the secretion of virulence factors from *Edwardsiella tarda* through a mass spectrometry screen (Srinivasa Rao et al., 2004). Finally, these studies were unified and the gene cluster was classified as the sixth bacterial secretion system (Pukatzki et al., 2006).

1.3 Overall architecture of the T6SS

1.3.1 Gene nomenclature and genetic distribution

As mentioned, the T6SS gene cluster is broadly distributed throughout Gram-negative bacteria. The gene cluster is particularly prevalent in pathogenic species including *E. coli*, *S. enterica*, *V. cholerae*, *P. aeruginosa*, *A. tumefaciens*, *Y. pestis*, *F. tularensis* and *B. mallei*, where it was found to be highly conserved (Pukatzki et al., 2007). The T6SS gene cluster is comprised of 13 core proteins with additional copies of some components found distributed in the bacterial genome near known virulence factor proteins (Shalom, Shaw, and Thomas, 2007; Boyer et al., 2009; Barret et al., 2011; De Maayer et al., 2011). Over the past $\sim$ 10 years, the names of these genes have changed many times; however, the accepted nomenclature currently in use is Type Six Secretion (Tss) A-M, except for TssD, TssI and TssH which are often referred to as hcp, VgrG and ClpV, respectively. Early bioinformatic analysis of the T6SS protein sequences identified that they have sequence homology with proteins in other bacterial secretion systems, suggesting their involvement in secretion (Boyer et al., 2009). Depending on the organism, the number of T6SS gene clusters that are found can vary, and thus while *P. aeruginosa* contains three T6SS gene clusters, *B. pseudomallei* contains six as seen in figure 1.2. This has been suggested as either a redundancy mechanism or the distinct clusters could be involved in different biological processes (Shalom, Shaw, and Thomas, 2007).

1.3.2 General structure

The T6SS utilises a contractile tail for the secretion of substrate and penetration of neighbouring target membranes. Through the use of electron cryotomography (ECT), it was shown that the T6SS assembly could be broadly sub-divided into two major complexes. The membrane complex, which spans the bacterial envelope, and the cytoplasmic injection machinery (figure 1.1, 1.3a) (Basler et al., 2012). The membrane complex consists of three protein members the outer membrane protein TssJ, and inner membrane proteins TssL and TssM (figure

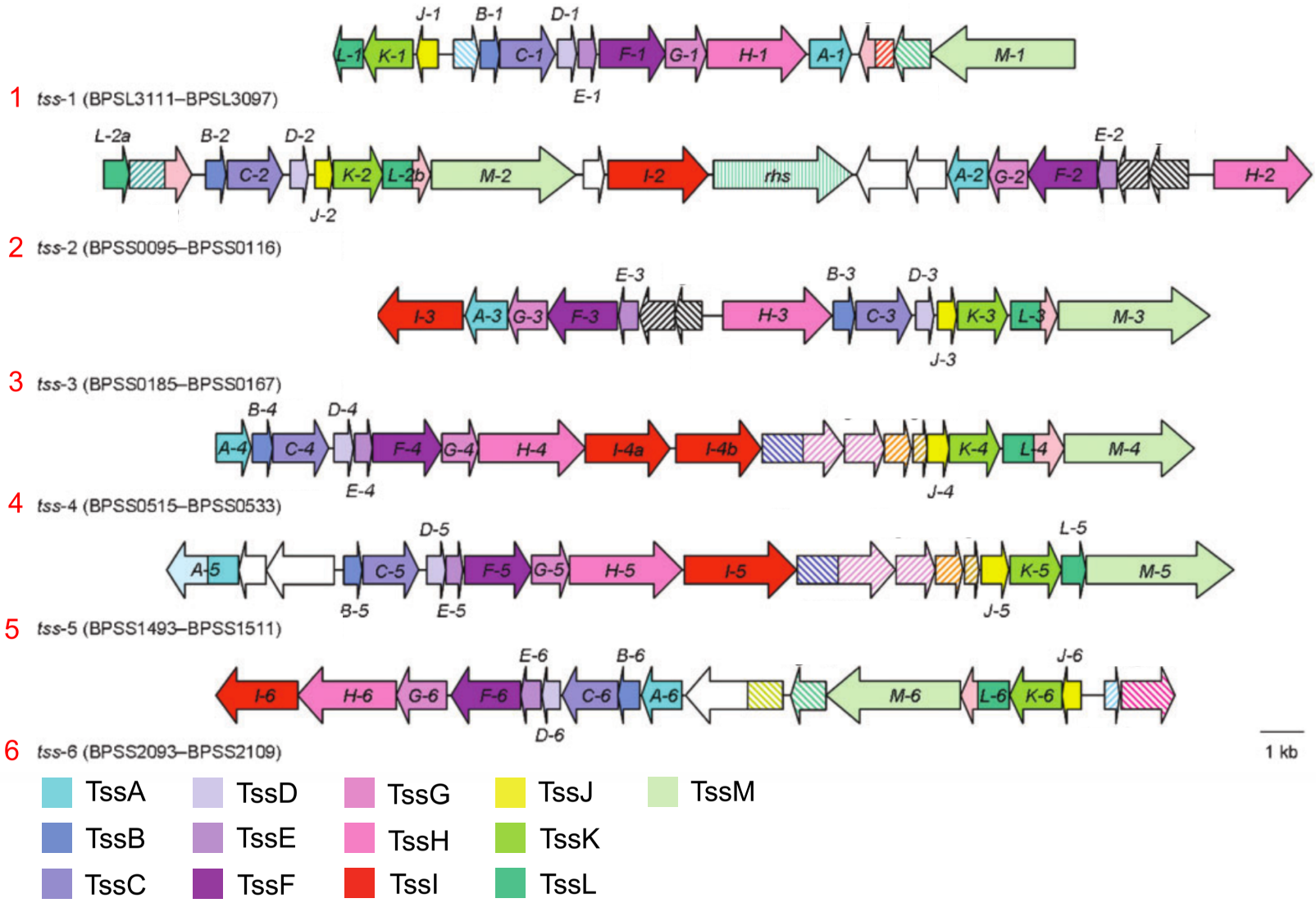


FIGURE 1.2: **T6SS gene cluster organisation in *B. pseudomallei*.** *B. pseudomallei* contains six T6SS gene clusters (tss1-6), each core component is coloured according to the included key. Figure modified from Shalom, Shaw, and Thomas, 2007.

1.4a) (Ho, Dong, and Mekalanos, 2014). The second complex, the injection machinery, can be further divided into two sub-complexes, the contractile sheath and the baseplate (figure 1.3a) (Basler et al., 2012). The baseplate is anchored to the membrane complex, where it acts to attach the contractile sheath to the membrane complex. The baseplate is proposed to contain the core protein components TssE, TssF, TssG, TssK and possibly TssA (figure 1.4a) (Brunet et al., 2015).

The final sub-complex, the contractile tail, is assembled perpendicular to the membrane using the baseplate as an assembly platform (Basler et al., 2012; Zoued et al., 2016). The contractile tail assembly occurs through the polymerisation of a TssB/C sheath surrounding a TssD (hcp) inner tube. The hcp tube is further capped by TssI (VgrG) (figure 1.4a). Once tail assembly is initiated the polymerised sheath rapidly extends within the bacterial cell. Through the use of fluorescence microscopy, it has been shown that the T6SS sheath can grow from the inner membrane to lengths of approximately 0.75 - 1 μm, which is approaching the width of the cell (Basler et al., 2012). In addition, wild-type *V. cholerae* cells have been shown to form between 1-5 of these assemblies simultaneously (Basler et al., 2012).

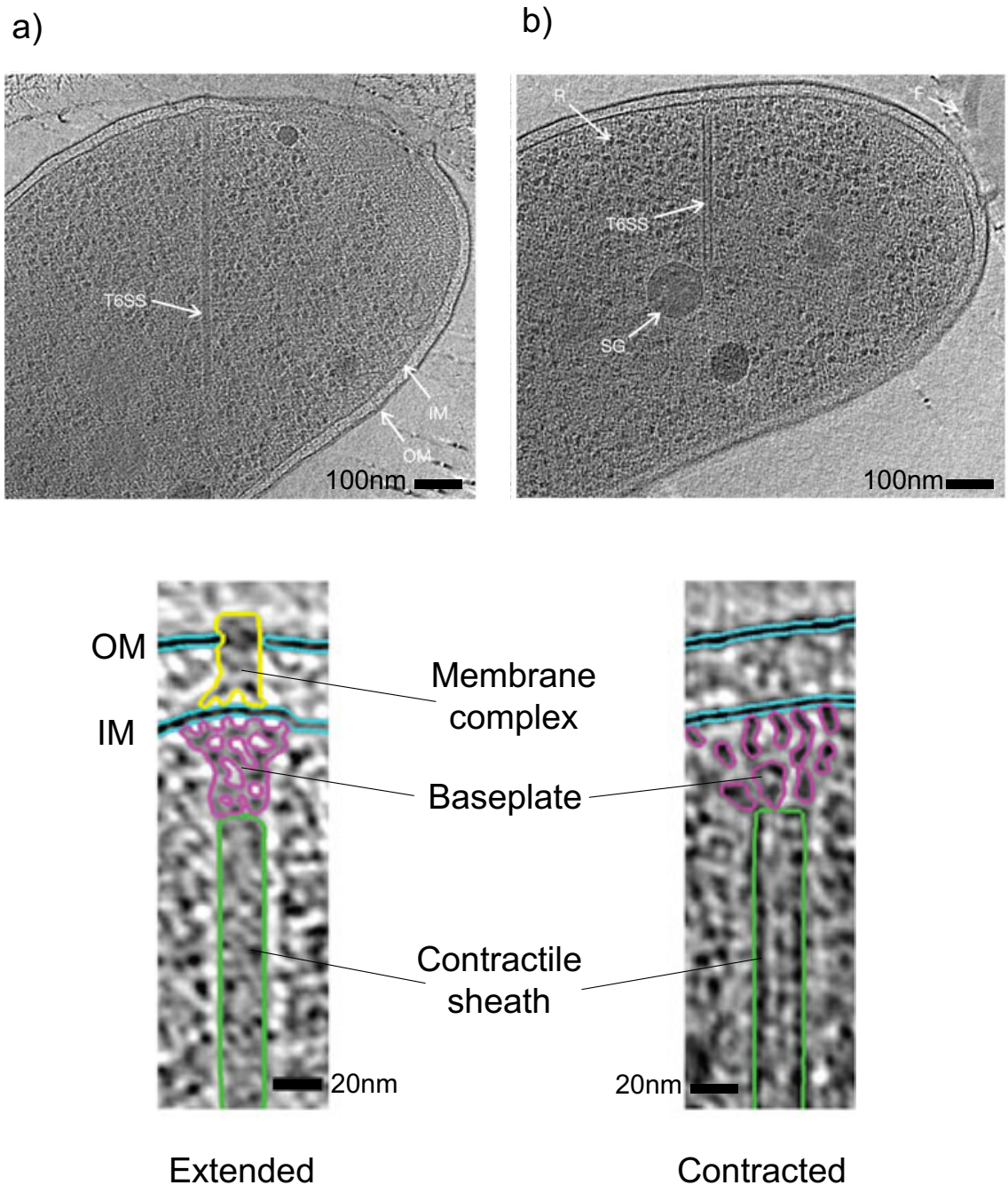


FIGURE 1.3: **Electron cryotomography of the T6SS in *V. cholerae*. a)** The extended form of the T6SS. The T6SS contractile sheath can be seen spanning the width of the bacterial cell. **b)** The contracted form of the T6SS. The sheath is distinctly shorter, and the baseplate has undergone a structural rearrangement. SG - storage granule, F - flagella, R - possible ribosome. OM - outer membrane, IM - inner membrane. Figure adapted from Basler et al., 2012

.

1.3.3 Mechanism of function

As mentioned, the T6SS uses a contractile tail for the secretion of substrate. This action utilises a contractile mechanism for the ejection of the inner tube through the membrane complex, such that it can penetrate the target membrane. The process of contraction was observed using fluorescence microscopy, indicating that it is very fast, occurring in under 5 ms, compared to the polymerisation and extension of the sheath which takes 10s of seconds (Basler et al., 2012). During contraction, a structural rearrangement of the T6SS subcomplexes occurs (figure 1.3b), with the most dramatic involving the compression of the contractile sheath against the baseplate (figure 1.3b, 1.4b). During which the sheath is reduced in length by $\sim$ 50 % compared to its extended form (Basler et al., 2012).

This 'firing mechanism' and overall architecture have led to comparisons being made between the T6SS and the T4 contractile tail (Leiman et al., 2009). In addition, a number of individual phage protein components share a strong structural relationship to those in the T6SS, such that it has been proposed that phage contractile assemblies might be the evolutionary origin of the T6SS, albeit inverted (Leiman et al., 2009). However, unlike the phage T4, the T6SS can disassemble once contraction has occurred, allowing for the components to be recycled and reused in an ATPase (TssH)-dependent process (figure 1.4c), highlighting the dynamic nature of the contractile tail (Basler et al., 2012; Bönemann et al., 2009). Furthermore, the scale of the contractile apparatus is distinctly different for both systems. The T4 phage contractile tail is only $\sim$ 100 nm in length, compared with the $\sim 1\mu$m T6SS sheath, a 10x difference in potential penetration distance (Duda et al., 1985; Basler et al., 2012).

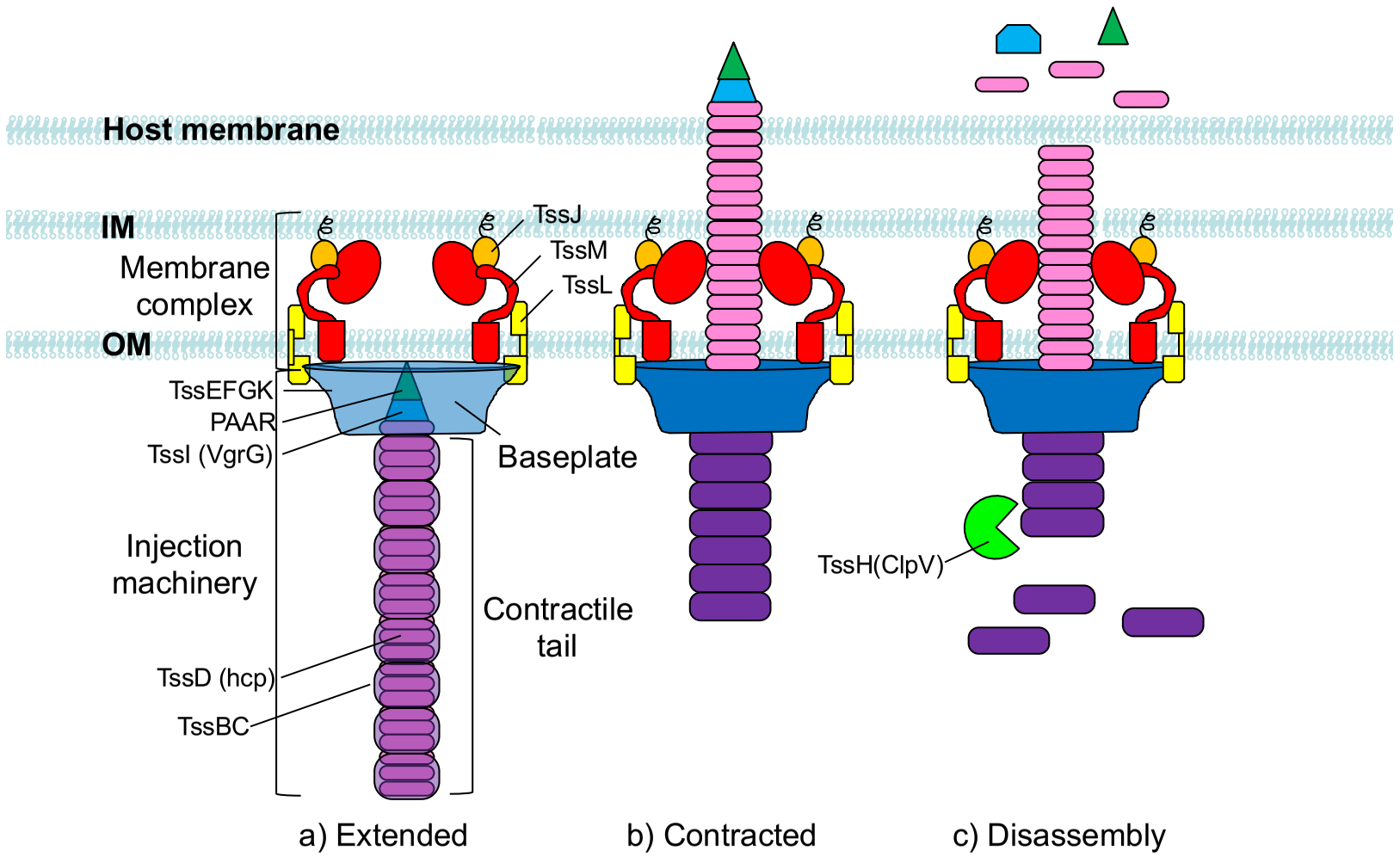


FIGURE 1.4: **Structural overview of the T6SS.** TssJ - orange, TssL - yellow, TssM - red, TssEFGK - dark blue, PAAR - dark green, TssI - light blue, TssD - pink, TssBC - purple, TssH - light green **a)** Extended form of the T6SS, before it has contracted. The contractile tail propagates inside the bacterial cytoplasm. **b)** Contracted form of the T6SS. The sheath is compressed and the tail tube with associated cargo is expelled through the baseplate and membrane complex and able to penetrate the target membrane **c)** Once contracted the sheath and T6SS assembly is recycled for future firing events. Figure adapted from Cianfanelli, Monlezun, and Coulthurst, 2016

1.4 Related systems to the T6SS

The architecture of a contractile tail is seen in a variety of other macromolecular machines, in particular, the bacteriophage T4, but also R-type pyocins (Ge et al., 2015), *serratia* antifeeding prophage (Heymann et al., 2013), rhapidosomes (Yamamoto, 1967), Photorhabdus virulence cassette (Yang et al., 2006) and metamorphosis-associated contractile arrays (Shikuma et al., 2014).

1.4.1 Bacteriophage T4 contractile tail

The bacteriophage T4 is a member of the *myoviridae* family, which posses a contractile tail for the injection of viral DNA into the host (Leiman et al., 2010). The T4 phage tail assembly consists of a contractile sheath, inner tail tube and baseplate complex, which enable host recognition, penetration and DNA secretion (Leiman et al., 2010; Leiman and Shneider, 2012a). Together, they form a large assembly consisting of at least 15 unique polypeptides (Leiman et al., 2010). Within the assembly, the contractile sheath is thought to consist of over 100 subunits of gp18 surrounding the rigid inner tube formed from gp19, which, when assembled, adopt a six-start helix architecture (Leiman et al., 2010). During contraction, a structural rearrangement of the gp18 subunits occurs, resulting in the compression of the sheath and the expulsion of the gp19 tube (Leiman and Shneider, 2012b). Attached to the contractile sheath/inner tube is the baseplate assembly, which consists of two complexes the wedge and hub (figure 1.5e). The hub is located at the core of the baseplate and forms the apex of the gp19 tube. Recent cryo-EM studies of the complete T4 baseplate suggested that the hub consists of six copies of the adapter proteins gp54 and gp48, three copies of the tip proteins gp27 and three copies of gp5 which enable penetration into the target cell (figure 1.5a) (Taylor et al., 2016). The second complex, the wedge, surrounds the hub and displays a distinct six-fold symmetry (Leiman et al., 2010). These studies further suggest that the wedge can be sub-divided into three discrete layers, the inner, intermediate and outer. The inner layer of the wedge (figure 1.5b) generates the pore opening for the tail tube to pass through. It consists of 12 subunits of gp6 arranged as two hexameric rings (gp6a and gp6b), six copies of domain two from gp7, six copies of gp25 and six copies of gp53, where gp25 and gp53 interact with the hub before contraction (Leiman et al., 2009; Taylor et al., 2016). The intermediate layer (figure 1.5c) consists of six copies of the remaining domains

of gp7 and six dimers of gp8. Finally, the outer layer of the wedge (figure 1.5d) consists of six trimers of gp9, gp10, gp11 and gp12, respectively, where gp12 forms the extended arms allowing for interaction and recognition of the host surface (Taylor et al., 2016).

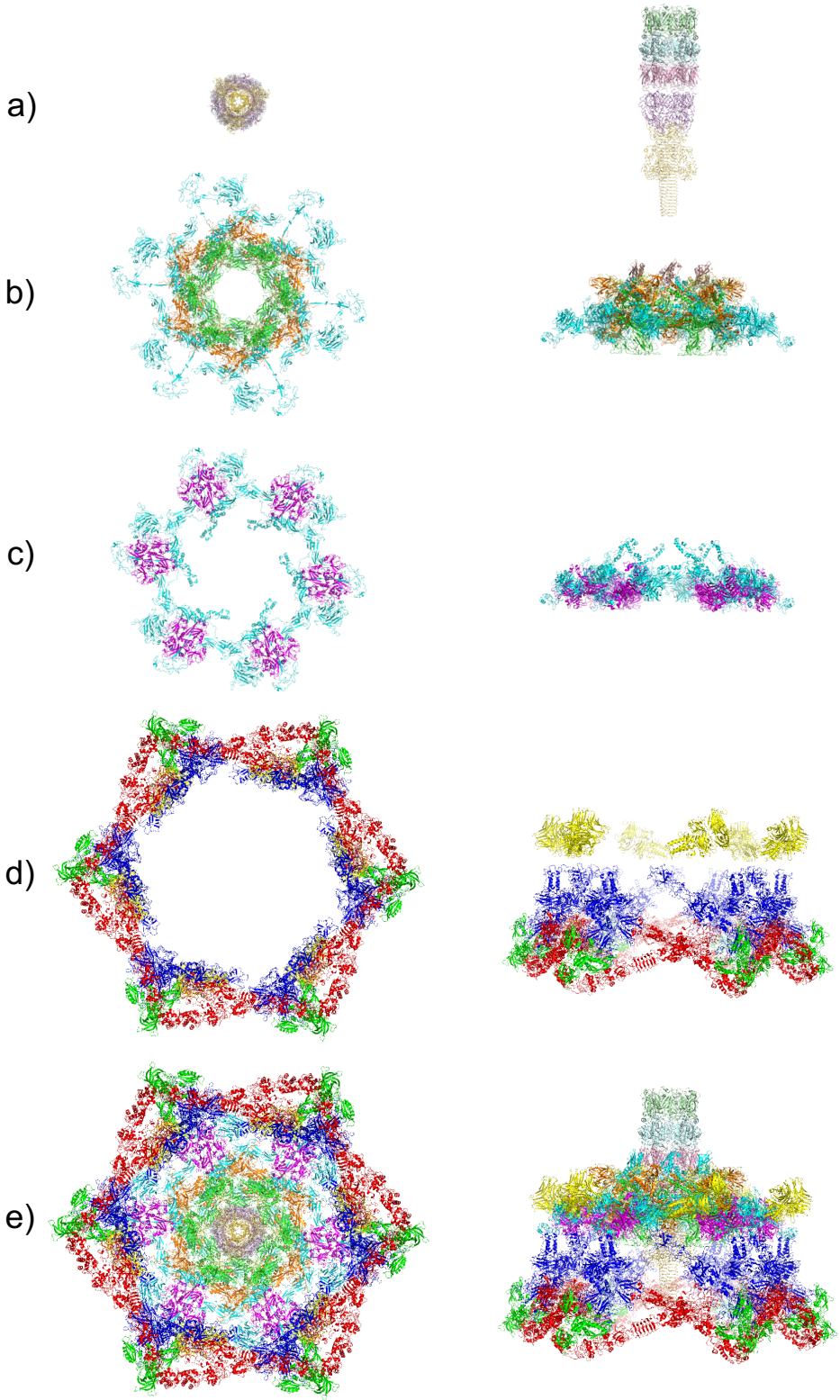


FIGURE 1.5: **Structural overview of the T4 bacteriophage baseplate.** Vertical (Left - view down the six-fold axis of the baseplate) and horizontal (Right - view perpendicular to the six-fold axis of the baseplate) view of the T4 contractile tail. **a)** The hub contains gp5 (light yellow), gp27 (like purple), gp48 (light red), gp54 (light blue) and then gp19 (light green) which forms the tube. **b)** The inner baseplate contains gp6a (green), gp6b (orange), domain two of gp7 (cyan), gp25 (brown), gp53 (wheat). **c)** The intermediate baseplate contains domains 1,3 and 4 of gp7 (cyan) and gp8 (magenta). **d)** The outer baseplate contains gp9 (yellow), gp10 (blue), gp11 (green) and gp12 (red). **e)** Complete T4 baseplate assembly. Figure created from PDB ID: 5IV5.

1.5 The function of the T6SS

The T6SS is capable of penetrating and translocating substrate to within a target cell through a contractile mechanism. This section will detail the biological function of the T6SS and show how the targets can be both prokaryotic and eukaryotic cells (Ho, Dong, and Mekalanos, 2014).

1.5.1 Bacterial competition

Bacteria often live in highly competitive environments and have evolved to utilise secretion systems as a mechanism to compete in inter-bacterial conflict (Russell, Peterson, and Mougous, 2014). The first indication that the T6SS was involved in bacterial competition was the discovery of three known bacterial virulence factors (effector proteins) in *P. aeruginosa*, which showed they were under the control of the same operon that makes up the nearby T6SS gene cluster (Hood et al., 2010). This finding was subsequently supported by mutational studies, whereby knock-out of the T6SS gene cluster from *V. cholerae* eliminated killing of prey Gram-negative bacteria, thus confirming that the T6SS confers bactericidal activity (MacIntyre et al., 2010). An additional function that the T6SS has been linked to is aiding in the formation of biofilms (Aubert, Flannagan, and Valvano, 2008). When the T6SS belonging to gene cluster 1 of *B. thailandensis* was knocked out, it reduced the ability of the organism to compete in a mixed bacterial biofilm with *P. putida* (Schwarz et al., 2010). This suggests that the T6SS is involved in bacterial fitness within mixed-species biofilms. Furthermore, the T6SS has been linked to the ability of bacterial populations to recognise self vs non-self. This is evident in *P. mirabilis*, whereby an interaction of the T6SS and the *ids*/*idr* self-recognition systems is identified. When transposon-mutants of core T6SS components were generated they inhibited self-recognition, suggesting that the T6SS is responsible for the secretion of *ids*/*idr* components required for self-recognition (Wenren et al., 2013).

1.5.2 Eukaryotic virulence

In addition to bacterial competition, the T6SS plays a crucial role in eukaryotic virulence. During infection of J774 macrophage and *D. discoideum* amoebae by *V. cholerae*, the eukaryotic cells undergo actin cross-linking, causing inhibition of phagocytosis, and ultimately cell death. However, when VgrG of the T6SS is knocked out this process is eliminated.

This indicates that the T6SS is directly involved in cell cytotoxicity within eukaryotic cells (Pukatzki et al., 2007; Ma et al., 2009). Furthermore, knock-out studies of the T6SS in *H. hepaticus* promotes larger colonisation of the gastrointestinal tract and an increase in pro-inflammatory responses within animal models. Therefore, it has been suggested that the T6SS aids in immunomodulation, such that it can maintain a symbiotic relationship with the host (Chow and Mazmanian, 2010). This agrees with the observation that, as previously discussed (1.2.2), knock-out of the T6SS in *Rhizobium leguminosarum* prevents infection of pea plants, indicating that the T6SS is involved in the symbiotic relationship between the two organisms (Bladergroen, Badelt, and Spaink, 2003).

1.6 The membrane complex

The T6SS membrane complex was initially proposed to comprise proteins TssJLM, organised such that they can link both of the bacterial membranes (Aschtgen et al., 2010; Ma, Lin, and Lai, 2009; Felisberto-Rodrigues et al., 2011). When the first view of a putative membrane complex was observed through the use of ECT, it identified a structure that can span the entire membrane envelope (figure 1.3a), confirming previous theories (Basler et al., 2012). Subsequently, negative stain EM of isolated membrane complexes from Enteroaggregative *E. coli* (EAEC) revealed it had a structure with dimensions of $\sim$ 300 Å in height and $\sim$ 205 Å diameter (figure 1.6a) (Durand et al., 2015). The membrane complex displays four distinct segments, the base, arches, tip complex and cap, generating a cavernous structure with a 15 Å pore (figure 1.6b,c) suggesting it provides a channel for the passage of the tail tube after contraction (Ho, Dong, and Mekalanos, 2014; Durand et al., 2015). Furthermore, this structure provides sufficient volume for the placement of 10 copies of TssJ, L and M (Durand et al., 2015).

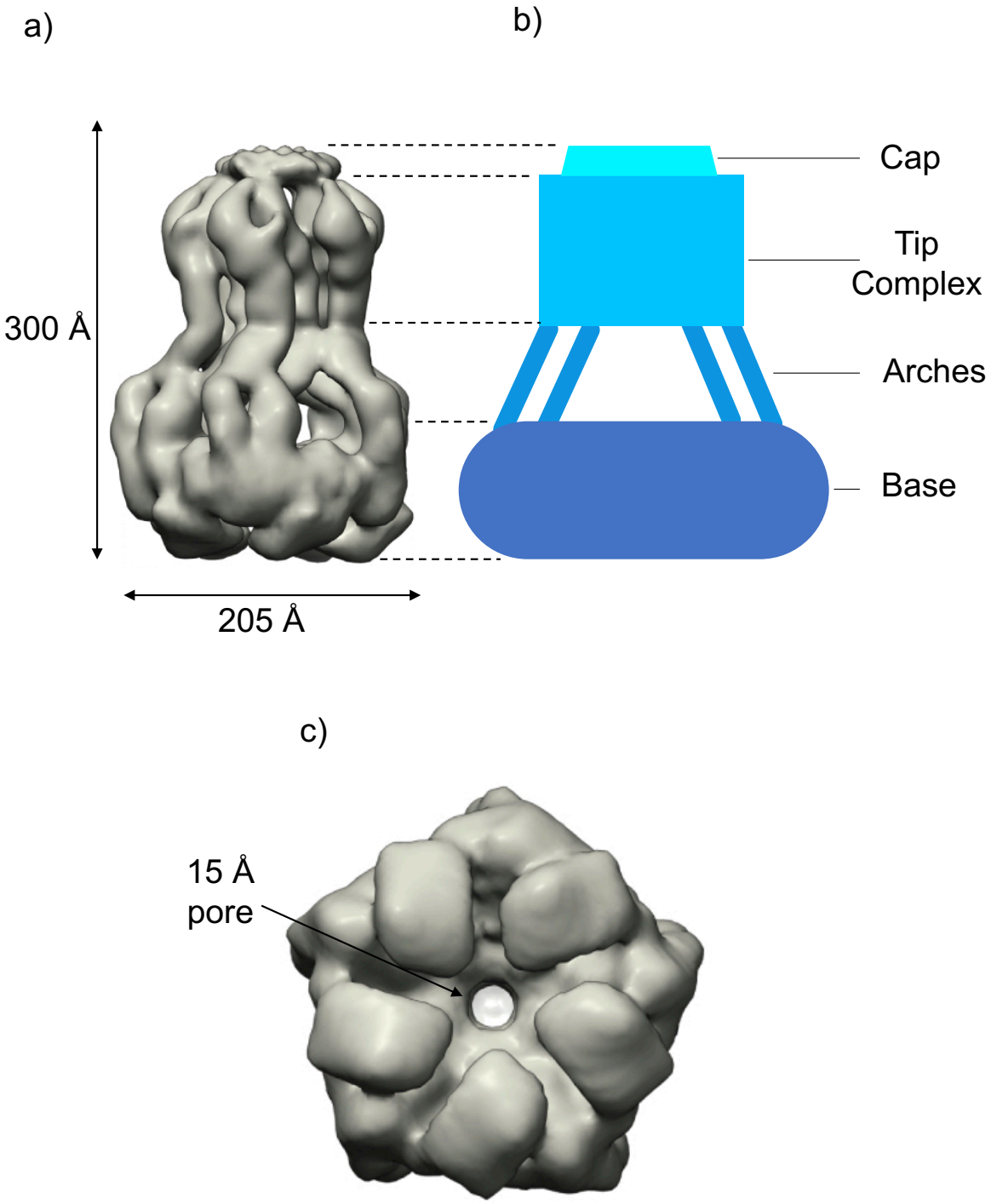


FIGURE 1.6: **Structure of the T6SS membrane complex.** Negative stain EM model of the isolated membrane complex from EAEC. **a)** Side view of the membrane complex. **b)** Diagram illustrating the four different segments of the membrane complex (Base, Arches, Tip complex and Cap) and how they relate to the EM model. **c)** View of the membrane complex from the cytoplasmic base. This shows the 15 Å pore that passes through the complex, in addition, highlights the five-fold symmetry that the model displays. Figure adapted from Durand et al., 2015.

1.6.1 TssJ

TssJ is the first T6SS component of the membrane complex to be discussed. It is a $\sim$ 18 kDa lipo-protein which, through the use of SLS-selective detergent solubilisation and proteinase K, was identified to be anchored to the outer-membrane and facing the periplasm (Aschtgen et al., 2008). TssJ was established as a critical member of the T6SS gene cluster, as TssJ knock-out mutants are unable to secrete hcp; hence it is required for the correct function of the T6SS (Aschtgen et al., 2008). The structure of TssJ was determined through X-ray crystallography and showed that TssJ folds into a single β-sandwich formed by two β-sheets consisting of four β-strands each, with a loop formed of three short α-helices (figure 1.7a). In addition, the protein is present as a monomer in the asymmetric unit with no indication of a biological interface to other symmetry-related molecules, suggesting TssJ does not form any oligomeric assembly by itself (Felisberto-Rodrigues et al., 2011). Further to this, complementary SEC-MALLS analysis of the solubilised TssJ supported this observation that TssJ is a monomer (Felisberto-Rodrigues et al., 2011).

To identify the role of TssJ interaction assays were carried out. Using immunoprecipitation with other T6SS components, TssJ was identified to interact with TssM within the periplasm (Felisberto-Rodrigues et al., 2011). In agreement with this observation, a structure of TssJ in complex with TssM was determined. This identified that TssJ interacts with the C-terminal domain of TssM in a 1:1 ratio, hence anchoring the membrane complex to the outer membrane (figure 1.7d) (Durand et al., 2015).

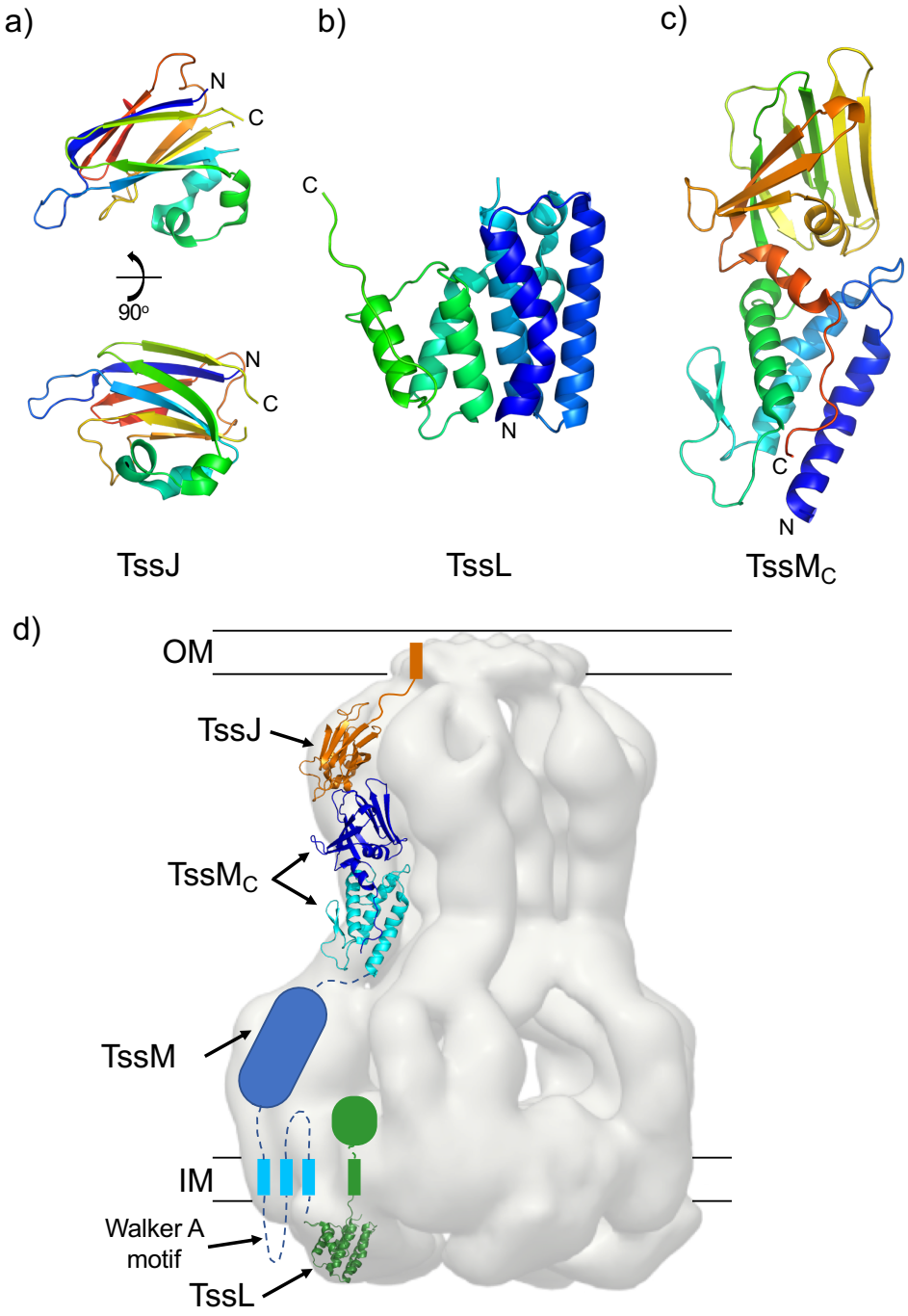


FIGURE 1.7: **Protein components of the T6SS membrane complex.** X-ray models of TssJ, L and M from the T6SS membrane complex. **a)** Structure of the periplasmic domain of TssJ from EAEC (PDB ID: 4Y7O). TssJ folds into a β-sandwich with three short α-helices. The N-terminal outer membrane lipid anchor is absent in the final model. **b)** Structure of the cytoplasmic domain of TssL from EAEC (PDB ID: 3U66). The structure is comprised of eight α-helices and the C-terminus in the structure would transition into the inner membrane-spanning domain. **c)** Structure of the C-terminal domains of TssM (TssM$_C$) from EAEC (PDB ID: 4Y7O). The structure shows this region is comprised of two domains. **d)**. Overlay of the TssJLM structures docked onto the EM model of the membrane complex. TssJ (orange) is present near the outer membrane where it interacts with the C-terminal domains of TssM (blue). Both of these are present in the tip complex region of the membrane complex. TssL (green) is shown in density on the cytoplasmic face of the complex, transitioning through the inner membrane. Figure adapted from Durand et al., 2015.

1.6.2 TssL

TssL is a $\sim$ 55 kDa protein, which shares similarities with IcmH of the T4SS, suggesting that TssL is located in the inner membrane (Ma, Lin, and Lai, 2009; Aschtgen et al., 2008). Bioinformatic analysis identified that TssL contains a single trans-membrane region near the C-terminal end of the protein and that the C-terminal region contains an OmpA/Pal/MotB motif. This suggested that the C-terminal end of the protein is directed towards the periplasm, and the N-terminus towards the cytoplasm (Aschtgen et al., 2010; Aschtgen et al., 2012). Through the use of bacterial two-hybrid and co-immunoprecipitation assays, the N-terminal domain of TssL was shown to be cytoplasmic, such that it interacted with the cytoplasmic TssK of the baseplate complex. This also indicated that TssL plays a role in anchoring the baseplate complex to the membrane complex (Zoued et al., 2013). In addition, gel filtration analysis of solubilised TssL and TssM suggests that they can interact in a 1:1 ratio, like that of TssM and TssJ (Durand et al., 2015).

When the structure of the N-terminal, cytoplasmic, domain of TssL was determined to 2.6 Å through X-ray crystallography it showed the domain consists of eight α-helices arranged as two bundles, with one involving helices 1-5, and the other helices 6-8. The structure was present as a monomer within the asymmetric unit, with no indication of a biologically relevant interface. This is consistent with observations by gel filtration suggesting TssL does not self-associate (Durand et al., 2012; Robb, Nano, and Boraston, 2012).

1.6.3 TssM

TssM is the final membrane complex component and is $\sim$ 130 kDa, which, through bioinformatics, is suggested to be similar to IcmF of the T4SS and therefore an inner membrane protein. Through solubilisation and fractionation of the membrane component, TssM was confirmed as being located in the inner membrane (Ma, Lin, and Lai, 2009). In addition, TssM was predicted to contain multiple trans-membrane regions and a putative Walker A motif. Through the use of beta-lactamase (BlaM)-TssM fusion proteins, (BlaM confers ampicillin resistance when it is present in the periplasm) the membrane topology of TssM was determined to contain three trans-membrane domains, with a Walker A motif present between the second and third domain. This results in the C-terminal region being orientated towards the periplasm, and the N-terminus being predicted to be cytoplasmic (Ma, Lin, and

Lai, 2009). While the Walker A motif has been shown to carry out ATP hydrolysis, it is still unclear if this is biologically relevant for the function of the mechanism of the T6SS (Zheng and Leung, 2007; Ma, Lin, and Lai, 2009).

Structural analysis of the C-terminal region (residues 869–1107) identified it is comprised of two domains. The N-terminal domain consists of a bundle of four α-helices attached to a C-terminal domain consisting of a β sandwich comprised of nine β-strands. This second domain is responsible for the interaction with TssJ as described above (1.6.1) (Durand et al., 2015). Furthermore, the N-terminal cytoplasmic domain of TssM, like that of TssL, has been shown to interact with TssK (1.8.3) (Zoued et al., 2013).

1.7 The contractile tail

The T6SS contains a contractile tail that can be subdivided into the contractile sheath and the inner tube, which together have similarities to bacteriophage T4. This has raised the question if the T6SS contractile tail components are functionally equivalent to those in the bacteriophage T4. The proteins described below have been shown to be critical for the formation and function of the T6SS contractile tail, and detail of their relationship to phage proteins are included.

1.7.1 TssD/ (hcp)

Hcp (haemolysin coregulated protein) is a $\sim$ 17 kDa protein, which has been shown to be secreted during the firing of the T6SS apparatus. This has led to the development of hcp secretion assays as being a key functional test for T6SS functionality (Pukatzki et al., 2006; Pukatzki, McAuley, and Miyata, 2009). Structural studies of Hcp have shown that it forms a doughnut-like hexamer *in vitro* with internal and external diameters of 40 and 85 Å, respectively (figure 1.8a). Structural alignment indicates that hcp shares a strong similarity to that of the T4 phage gp19 protein (figure 1.8b), and like that of gp19, hcp hexamers can form extended nanotubes (Mougous et al., 2006; Ballister et al., 2008; Leiman et al., 2009). These tubes of hcp hexamers assemble head-tail with a gentle twist of 24 ^{o}, the same as the surrounding contractile sheath (Ballister et al., 2008; Wang et al., 2017).

1.7.2 TssI/ (VgrG)

VgrG is a further T6SS component, which is often, but not always, found adjacent to hcp in the T6SS gene cluster (Shalom, Shaw, and Thomas, 2007; Pukatzki et al., 2007). VgrG knock-out mutants are unable to secrete hcp and VgrG and are therefore classed as a core component of the T6SS (Brunet et al., 2015). The protein forms a homotrimeric assembly (figure 1.8c), and shares a striking structural similarity to the gp27-gp5 tail spike complex in T4 bacteriophage (figure 1.8d), despite, in the case of *E. coli* CFT073, only sharing 13 % sequence identity (Leiman et al., 2009). In phage, gp27 and gp5 exist as separate polypeptides, while in the T6SS, their function is fused into a single VgrG polypeptide, such that VgrG exhibits a repeating β-helix (like gp5), which functions as a puncturing tip sat atop a trimeric platform (like gp27) (figure 1.8c) (Pukatzki et al., 2007). Through the use of yeast two-hybrid and co-immunoprecipitation experiments, it was shown that VgrG interacts with hcp. Further to this, a structural similarity between the gp27-like domain of VgrG and the hcp β-tube domain was found. Taken together, this suggests that VgrG sits atop of the hcp tube, in an analogous interaction as gp27/5 with the hub complex to form the membrane-puncturing apparatus (Pukatzki et al., 2007; Leiman et al., 2009; Lin, Ma, and Lai, 2013).

While VgrGs do not contain the OB-fold and lysozyme domains of the phage tip protein, they sometimes encode for a further, extended, C-terminal domain and these forms are collectively named 'evolved VgrGs' (Pukatzki et al., 2007). These C-terminal domains have been shown to have enzymatic activity involved in both prokaryotic, and eukaryotic toxicity and have subsequently been called effector proteins (1.7.2) (Pukatzki et al., 2007). The question as to whether VgrG is a contractile tail or baseplate component (as will be discussed) has still not been resolved. However, within this thesis, it is presented in the context of the contractile tail, due to its close functional relationship with hcp.

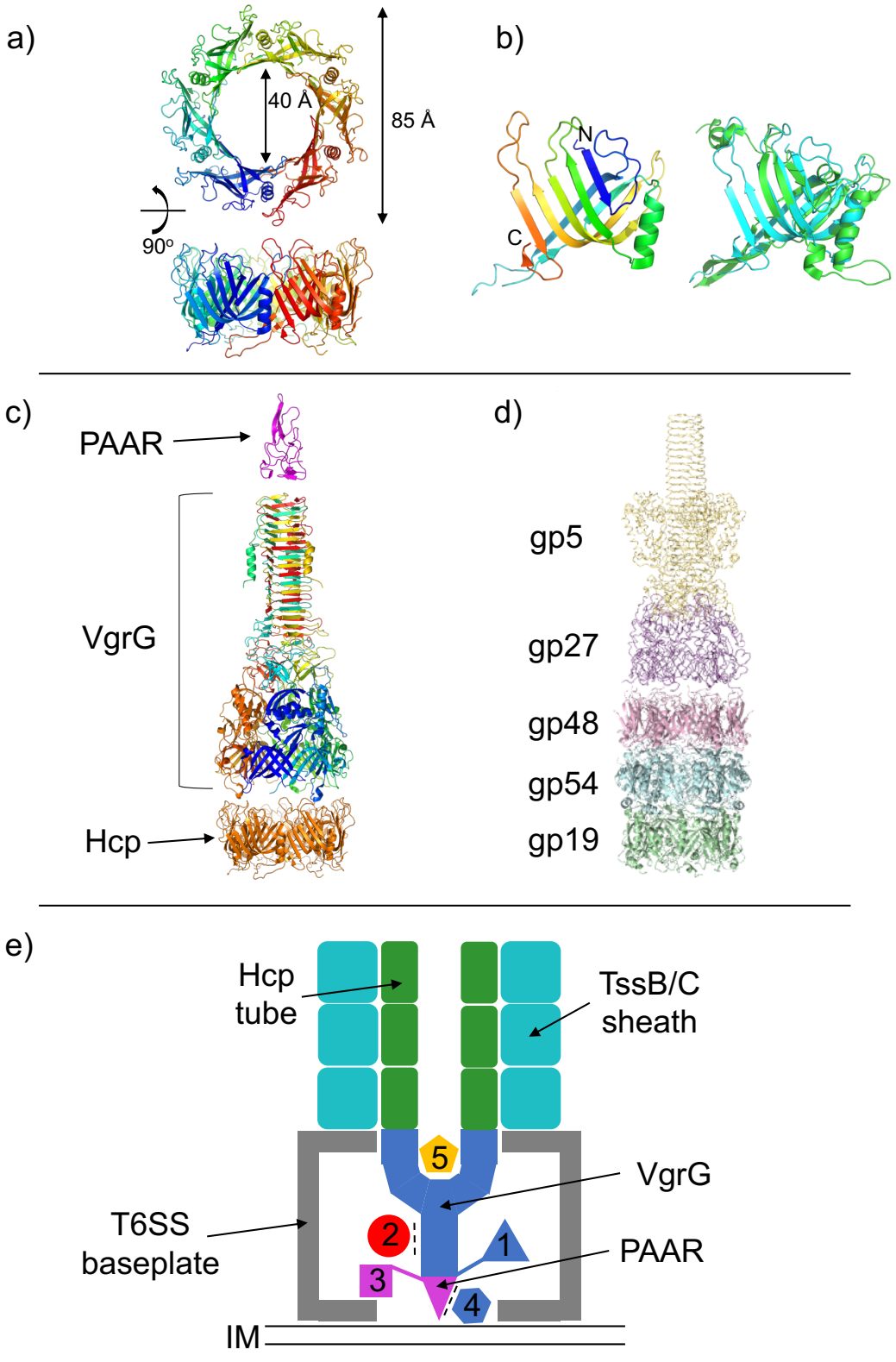


FIGURE 1.8: **Protein components of the T6SS tube. a)** Structure of the hcp hexamer from *P. aeruginosa* (PDB ID: 1Y12). **b)** Structure of the hcp monomer (left) from *P. aeruginosa* (PDB ID: 1Y12). Structural alignment of hcp (blue) with gp19 (green) from T4 phage (PDB ID: 5W5F) (right). **c)** Assembly of hcp hexamer (orange) (PDB ID: 1Y12)/VgrG trimer (chainbow) (PDB ID: 4MTK) /PAAR monomer (magenta) (PDB ID: 4JIW). **d)**. The hub complex taken from the bacteriophage T4 (PDB ID: 5IV5). **e)** Diagram showing the various methods effector proteins can be associated with the T6SS ejection apparatus. 1 - VgrG C-terminal domain, 2 - bind to VgrG, 3 - PAAR extension, 4 - bind to PAAR, and 5 - hcp interaction. Figure adapted from Shneider et al., 2013.

PAAR proteins

While VgrGs have been shown to be equipped with functional C-terminal domains, a separate protein, PAAR (proline-alanine-alanine-arginine) can associate to the tip of VgrG (figure 1.8c) (Shneider et al., 2013). PAAR protein-encoding genes are often found immediately downstream of VgrG genes, indicating their close association with one another (Pukatzki et al., 2007). While the core of the PAAR protein folds into a symmetrical cone with a sharpened tip and a β-helix base which allow for interaction with VgrG, the PAAR polypeptide may be extended at both the N and C-terminus. This permits for the fusion of effector domains to the PAAR polypeptide without disturbing the PAAR/VgrG interaction (Shneider et al., 2013).

Effector proteins

As discussed above, VgrG and PAAR can be either fused or decorated with effector proteins as the payload of the T6SS (Hood et al., 2010; Brooks et al., 2013). Effector proteins show a large range of functions, typically targeting features which are both highly conserved and essential. They typically fall into three distinct categories with regards to function; cell wall degrading, e.g. the cell wall amidase Tse1 which is part of the amidase effector family (Tae) (Russell et al., 2011); cell membrane targeting, e.g the lipase effector family (Tle) (Russell et al., 2013; Russell, Peterson, and Mougous, 2014); and nucleic acid targeting, such as the recombinant hotspot (Rhs) endonuclease family (Poole et al., 2011; Koskiniemi et al., 2013; Russell, Peterson, and Mougous, 2014). However, other groups of effector proteins are known, whose function can lead to ADP ribosylation, actin cross-linking (particularly important for eukaryotic virulence (Pukatzki et al., 2006)) and pore-formation (Salomon et al., 2014).

Five mechanisms have been proposed for how effector proteins may associate with the T6SS (figure 1.8e), these are as follows:

1. Associate with the C-terminal domain of VgrG (Pukatzki et al., 2007).
2. Bind to the surface of VgrG (Dong et al., 2013).
3. Form an N or C-terminal extension of PAAR (Koskiniemi et al., 2013).
4. Bind to additional domains of PAAR (Shneider et al., 2013).

5. Interact with the hcp tube directly (Silverman et al., 2013).

Immunity proteins

Immunity proteins have been identified as being encoded adjacent to effector proteins. This has been proposed as a measure to protect the bacterium against self during T6SS firing events, but also to protect against effector proteins delivered from neighbouring cells. (Benz et al., 2012; Zhang et al., 2013; Russell et al., 2012; Russell, Peterson, and Mougous, 2014). Some examples of these include the Tai immunity proteins that are located adjacent to the tae effector proteins in the genome. In addition, the Tli immunity proteins which directly inactivate the Tle lipase effector proteins (Russell, Peterson, and Mougous, 2014).

1.7.3 TssBC - The T6SS contractile sheath

TssB and TssC are two further core components of the T6SS, which were shown to be critical for correct T6SS function (Bönemann et al., 2009). Through co-purification TssB and TssC were shown to form a strong interaction with one another. In addition, negative stain EM identified that TssBC form large extended fibres in solution which displayed 12-fold symmetry along the cross-section. This led to the hypothesis that TssBC were the functional equivalents of gp18 from phage, which is arranged as hexameric rings, such that they form the T6SS contractile sheath (Bönemann et al., 2009; Basler et al., 2012). As previously described, these fibres associate with the baseplate, perpendicular to the membrane, and can span great distances in the cell (Basler et al., 2012).

Recent cryo-EM studies of the T6SS from *V. cholera* identified the first atomic structure of the contractile sheath, in its contracted form. This confirmed that TssBC interacted with one another and identified they are present as a heterodimer in the contractile sheath (Kudryashev et al., 2015). TssB is organised as four β-strands and six α-helices, whereas TssC consists of 18 α-helices and 12 β-strands. The structure of the TssBC heterodimer revealed they are organised 'head to tail' of one another such that the N-terminus of TssC is adjacent to the C-terminus of TssB (figure 1.9a). The heterodimer can be divided into three domains, where domain one is orientated towards the core of the sheath, and domain three is presented to the periphery of the TssBC sheath (Kudryashev et al., 2015). To generate the sheath, one TssBC heterodimer mediates specific interactions with neighbouring heterodimers on either

side (figure 1.9b). This produces two inter-penetrating layers of six TssBC heterodimers generating a six-start helix (figure 1.9c), which is responsible for the apparent 12-fold symmetry seen in earlier negative stain EM work (Bönemann et al., 2009; Kudryashev et al., 2015).

Due to the lack of a structure of the contractile sheath in the extended form, the mechanism that promoted disassembly remained unclear. This was resolved when the cryo-EM model of an extended sheath was determined and identified the structural differences responsible for disassembly. The most obvious difference was the change in inner diameter, rise and twist of the sheath (figure 1.9d). In the extended form, it has an inner diameter of 80 Å, a rise of 38.0 Å and a 23.6 º twist (Wang et al., 2017). Whereas, in the contracted form, the sheath displays an inner diameter of 110 Å, a rise of 21.8 Å and a 29.4 º twist (Kudryashev et al., 2015). Overall, during contraction, the sheath TssBC subunit rotates $\sim$ 34 º and expands radially by about 15 Å, thus compressing and releasing the hcp tube in the process; hence hcp is absent in figure 1.9c. This mechanism for sheath re-arrangement is elegantly modelled in the video linked to figure 1.9e. During the structural rearrangement of the sheath, it can be seen that domain three is exposed during contraction of the sheath, suggesting that this is the site that binds ClpV triggering sheath disassembly (Wang et al., 2017).

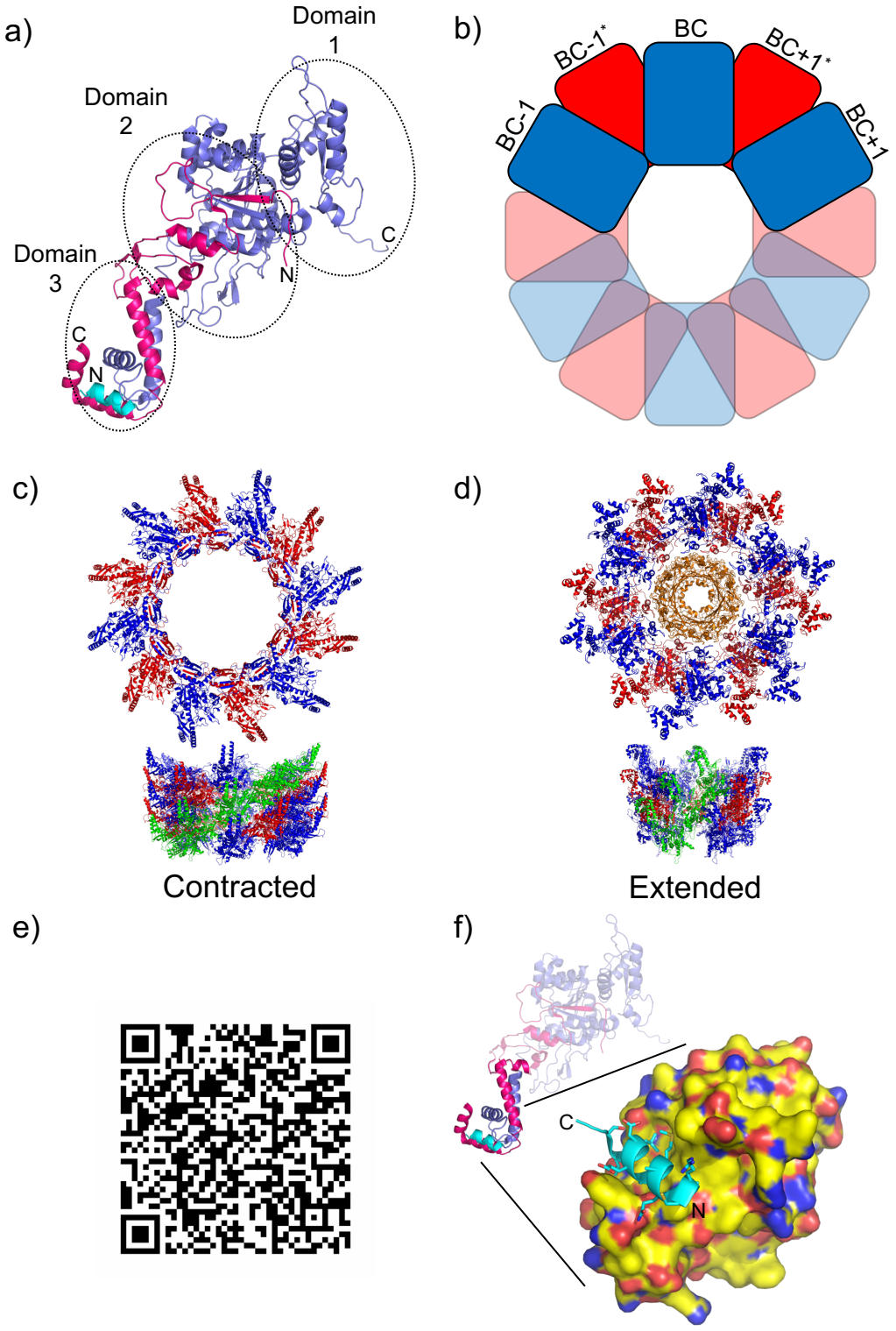


FIGURE 1.9: **Components and assembly of the T6SS sheath. a)** Structure of the TssBC heterodimer from *V. cholera* (PDB ID: 5OJQ). TssB in pink, and TssC in lilac, with the N-terminal helix of TssC shown in cyan. **b)** Schematic of TssBC sheath assembly. Four interfaces are generated: 1) TssB0 with TssC+1. 2) TssC0 with TssC+1*. 3) TssC0 and TssB-1*. 4) TssC0 with TssB-1*. **c)** Cryo-EM model of the contractile sheath from *V. cholera* (PDB ID: 3j9g) in the contracted form. The six-start helix is indicated in green. **d)**. Cryo-EM model of the contractile sheath from *V. cholera* (PDB ID: 5OJQ) in the extended form. The hcp inner tube is in brown **e)** QR code linked to sheath contraction video (https://static-content.springer.com/esm/art%3A10.1038%2Fs41564-017-0020-7/MediaObjects/41564_2017_20_MOESM4_ESM.mp4). **f)** N-terminal helix of TssC (cyan) bound to ClpV (PDB ID: 3ZRJ) from K12 *E. coli*.

1.7.4 TssH/ ClpV

Unlike the firing mechanism found in T4 bacteriophage, which is essentially an irreversible event, the T6SS can re-assemble to allow for successive firing. This implies that a source of energy must be used to enable the restructuring of the contractile tail into the extended state. TssH (ClpV) has been identified as a member of the AAA+ ATPase protein family (Schlieker et al., 2005; Bönemann et al., 2009), and the generation of ClpV knock-out mutants, in *Vibrio cholerae*, exhibited either severely reduced or the total elimination of hcp/ VgrG secretion, indicating a critical role in the T6SS mechanism (Zheng, Ho, and Mekalanos, 2011; Bönemann et al., 2009). In addition, YFP-labelled ClpV has been identified to co-localise to contracted TssBC sheaths, but not the extended form (Basler et al., 2012; Kapitein et al., 2013). Taken together, it has been proposed that ClpV provides the energy to re-model the TssBC sheath once it is in the contracted state (Bönemann et al., 2009; Kapitein et al., 2013). Furthermore, structural studies have shown that ClpV interacts with the N-terminal helix of TssC, which is present in domain three of the TssBC heterodimer (figure 1.9f). This is exposed in the contracted state of the sheath, thus triggering the rapid disassembly of the sheath (figure 1.4c) and explains why it cannot disassemble extended sheaths (Bönemann et al., 2009; Pietrosiuk et al., 2011; Wang et al., 2017).

1.8 The baseplate complex

Like that of contractile bacteriophage (and R-type pyocins), it was proposed that the T6SS would utilise a baseplate complex as part of its assembly (Leiman and Shneider, 2012a; Brunet et al., 2015). ECT studies on the T6SS from *V. cholerae* identified its general architecture. This included a 'flared bell-shaped base' present between the membrane and needle/sheath complex which was proposed to be a T6SS baseplate complex (Basler et al., 2012) (figure 1.3a,b). Through the use of *in vivo* cross-linking assays knock-out mutants for TssE, F, G, K, and A, exclusively, were able to block the secretion of hcp, indicating their critical importance in the T6SS mechanism, and as putative components for the T6SS baseplate (Lossi et al., 2011; Brunet et al., 2014; Brunet et al., 2015). However, to date, no high-resolution structure for the T6SS baseplate complex has been determined.

1.8.1 TssE

TssE is a $\sim$ 15 kDa protein, that shares sequence homology to that of the highly conserved T4 bacteriophage baseplate wedge protein gp25 (17 % sequence identity and 39 % sequence similarity) (Leiman et al., 2009) suggesting an evolutionary relationship. TssE has been shown to be required for sheath assembly and is proposed to bind to exposed β-strands donated by TssBC, therefore allowing the tail sheath to be anchored to the baseplate during polymerisation, suggesting a role similar to gp25 in contractile bacteriophage (Lossi et al., 2011; Basler et al., 2012). Along with hcp and VgrG, TssE is another example of a T6SS protein component that has a shared common ancestor with that of contractile bacteriophage. Currently, no high-resolution structure of TssE has been determined; however, it is not expected to be radically different from that of gp25 (figure 1.10a).

1.8.2 TssF and G

Two further members of the T6SS gene cluster are TssF and TssG, which, for this introduction, shall be considered under a single entity. As previously discussed, both TssF and G are critical for hcp secretion and are hence core components of the T6SS (English et al., 2014; Brunet et al., 2015). What was still unknown was the functional role of the proteins. While limited information is available on TssF and G, their respective genes are often found adjacent to one another, or even overlapping in the T6SS gene cluster (figure 1.2) (Boyer et al., 2009; English et al., 2014); suggesting a combined function for the proteins. This potential intimate relationship was further investigated through bacterial two-hybrid and *in vitro* His-IMAC purification, which identified that TssF and G can stabilise each other by forming a complex (English et al., 2014; Brunet et al., 2015). In addition, TssFG were shown to interact with TssE, TssK, VgrG, Hcp and TssC through bacterial two-hybrid and co-immunoprecipitation assays and form a stable complex with TssK, such that they can form a complex > 670 kDa, with a possible stoichiometry of 12:6:3 TssK, TssF, TssG (English et al., 2014). Taken together, this shows that TssFG can interact with the previously established baseplate and tail proteins. Furthermore, co-localisation experiments identified that TssFG are able to associate with the membrane complex, before the formation of the tail sheath (Brunet et al., 2015). In total, these results suggest that TssFG are intimately related proteins that can form a stable complex with other baseplate components, in addition to

localising to the membrane complex to act as a platform for the polymerisation of the tail sheath. While no high-resolution structure of TssF and TssG has yet been reported, bioinformatics and secondary structure prediction suggest that TssF and TssG are structurally related to the core T4 baseplate components gp6 and gp53, respectively (figure 1.5) (Brunet et al., 2015).

1.8.3 TssK

TssK is a $\sim$ 50 kDa, cytoplasmic protein, which, through TEM and SAX, has been shown to assemble as a homotrimer (Zoued et al., 2013) (figure 1.10b). As discussed above, TssK can interact with the cytoplasmic domains of TssL and TssM, suggesting it is involved in anchoring the baseplate to the membrane complex (Zoued et al., 2013). This finding is supported such that in *B. pseudomallei* TssK is often found adjacent to TssL within the T6SS gene clusters (figure 1.2) (Shalom, Shaw, and Thomas, 2007). Furthermore, TssK can interact with the Hcp, TssA, TssC, TssF and TssG components of the T6SS; this provides more evidence that TssK is a member of the T6SS baseplate. Taken together, these results suggest that TssK aids in the assembly of the T6SS baseplate to the membrane complex (Zoued et al., 2013; English et al., 2014). While little information is known about the structure of TssK, it has been suggested that it resembles gp8 of the T4 baseplate, based on low-level sequence homology (11 %) (Planamente et al., 2016). Due to the limited structural information available for TssK, it was selected as a suitable candidate for structural studies to understand its role in the baseplate of the T6SS.

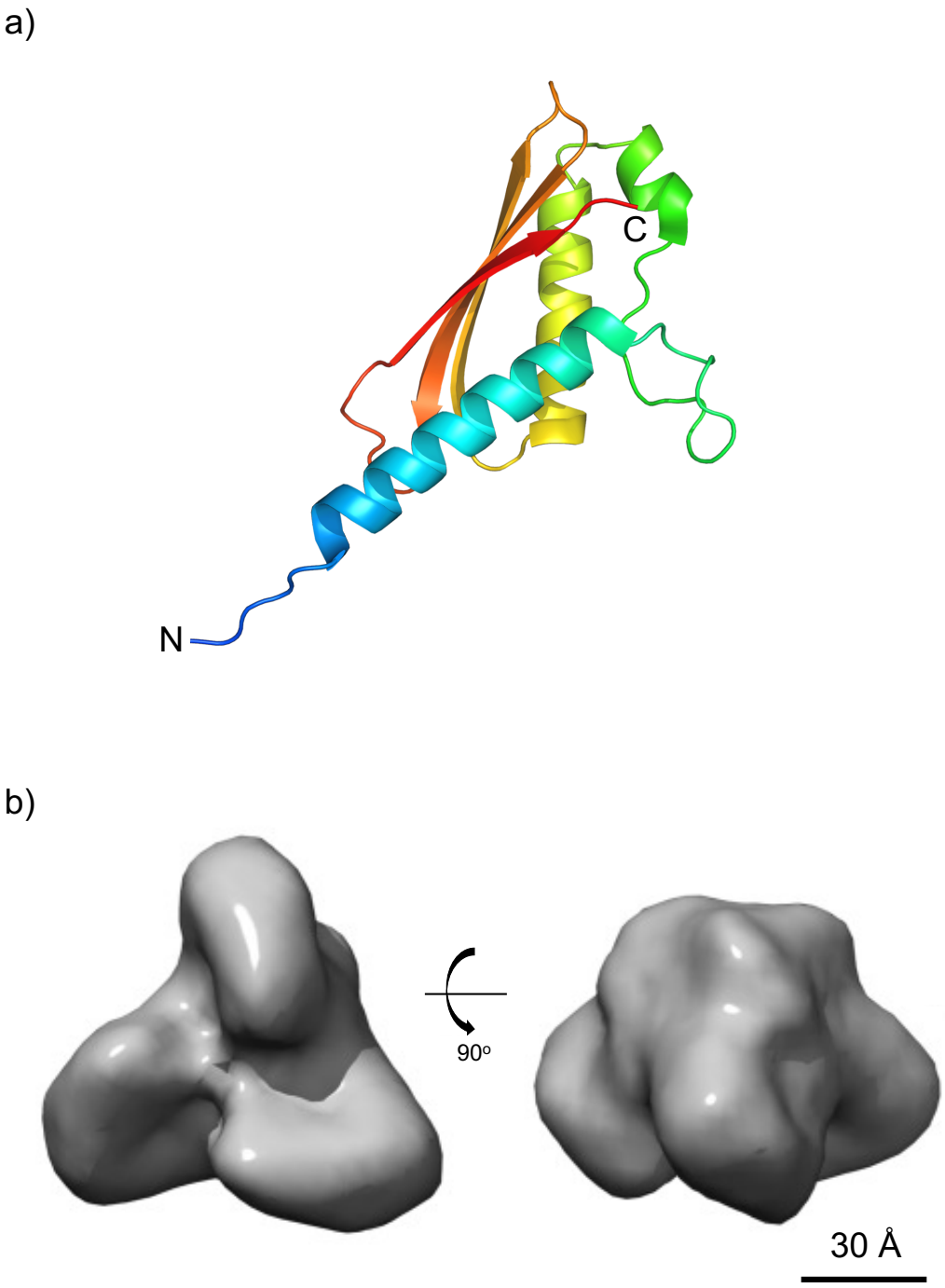


FIGURE 1.10: **gp25 from the T4 phage baseplate and TssK from the T6SS baseplate. a)** Structure of gp25 from the T4 baseplate (PDB ID: 5iw9), structure is represented in chainbow colour. **b)** EM reconstruction of TssK, indicating a molecule $\sim$ 100 Å across. EM adapted from Zoued et al., 2013.

1.8.4 TssA

At the start of this project very little information was known about TssA, except that it was believed to be a core component of the T6SS, consisting of a highly conserved N-terminal domain containing the ImpA_N; PF06812 motif and a more divergent C-terminal domain (CTD) (Boyer et al., 2009; Shalom, Shaw, and Thomas, 2007; Zoued et al., 2017). Further phylogenetic analysis suggested that TssA can be sub-divided into three clades 1, 2 and 3 (Planamente et al., 2016). For example, *P. aeruginosa* (PA0082) TssA is a member of clade 1, *A. hydrophila* (AHA1844) TssA is a member of clade 2 and *B. cenocepacia* (BCAL0348) is a member of clade 3 (figure 1.11a). More recently, TssA knock-out mutants identified it to be critical for the function of the T6SS, such that mutants were unable to form stable contractile sheaths, suggesting that one of TssA's roles is stabilising the polymerisation of the contractile sheath (Planamente et al., 2016). Biochemically, TssA has been shown to be a $\sim$ 50 kDa, cytoplasmic, protein, which through the CTD can self-assemble to form high molecular weight oligomers, as was shown by negative stain EM of TssA from *P. aeruginosa* (figure 1.11b) (Planamente et al., 2016; Zoued et al., 2016). However, structural details on TssA remain sparse and little was known of the relationship between the structure and the function of TssA and how the different clades of TssA relate to each other.

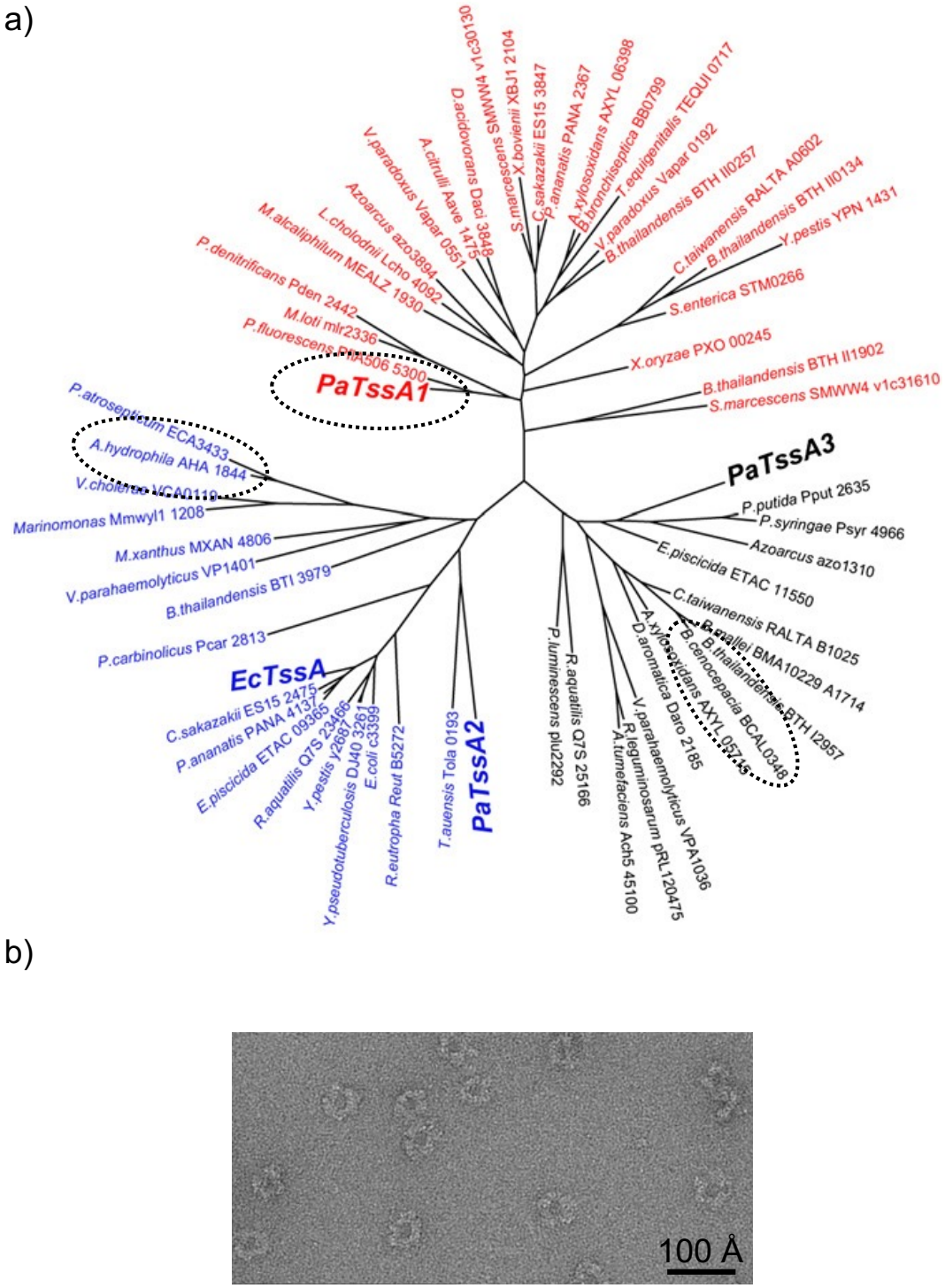


FIGURE 1.11: **Phylogenetic and EM analysis of *P. aeruginosa* TssA. a)** Phylogenetic analysis from Gram-negative bacteria. Those highlighted are representatives from each TssA clade. **b)** Negative stain EM of the *P. aeruginosa* (PA0082) TssA C-terminal domain. This indicates the C-terminal domain is able to form large ring-like complexes. Figures adapted from Planamente et al., 2016.

1.9 Aims of the project

Given that the T6SS is the least well understood of the major secretion systems, in particular, the shortage of knowledge on the structure and function of the baseplate components TssA and TssK, this identified them as suitable proteins for structural studies. Therefore, the main aim of this study was to determine the structure of TssA and TssK from various bacterial systems to try and expand the understanding of their respective roles within the T6SS. As part of this project, T6SS components were selected from various Gram-negative bacteria that are known to confer human virulence as species varients during structural studies. These included *Aeromonas hydrophila*, *Burkholderia cenocepacia* and *Burkholderia pseudomallei*. *Aeromonas hydrophila* TssA was chosen as it represents a member of the TssA2 clade (discussed in 3.2). This would complement the work being carried out in parallel on a TssA1 representative from *Burkholderia cenocepacia*. Furthermore, *Aeromonas hydrophila* TssA and *Burkholderia cenocepacia* TssK constructs suitable for structural studies had been produced by our collaborators in the medical school to complement the two-hybrid protein interaction studies they were carrying out. Therefore, these were utilised first.

1.9.1 *Aeromonas hydrophila*

Aeromonas hydrophila is a human pathogen which is often associated with gastroenteritis, soft tissue infection and septicaemia (Janda and Abbott, 1998). While it is generally considered to be an opportunistic bacterium, numerous cases of infection in healthy individuals indicate it could be more pathogenic than previously thought (Chopra and Houston, 1999). *A. hydrophila* contains a single T6SS gene cluster, which is under the control of the virulence-associated gene *vasH* (Suarez et al., 2008). This suggests that *A. hydrophila* utilises the T6SS as part of its pathogenic mechanism. This is supported by knock-out studies which showed that disruption of the T6SS gene cluster blocked *A. hydrophila*'s ability to evade phagocytosis, indicating that the T6SS is critical for innate immunity evasion and subsequent spread throughout the organism (Suarez et al., 2010). Example T6SS effector proteins found in *Aeromonas hydrophila* are TseC a colicin-like domain, and TseI a putative toxin known to target insect and mammalian cells. Both of which have been shown to be under the control of the *Aeromonas hydrophila* T6SS (Liang et al., 2015).

1.9.2 *Burkholderia cenocepacia*

Burkholderia cenocepacia is a member of the *Burkholderia cepacia* complex (bcc), which is a group of related opportunistic human pathogens (Lyczak, Cannon, and Pier, 2002; Aubert, Flannagan, and Valvano, 2008; Flannagan et al., 2012). It has been shown to be a causative agent of severe lung and respiratory tract infections in patients with cystic fibrosis (Aubert, Hu, and Valvano, 2015). In addition, like *A. hydrophila*, *B. cenocepacia* contains a single T6SS gene cluster, which plays a critical role in intracellular survival for the bacterium (Rosales-Reyes et al., 2012; Aubert, Hu, and Valvano, 2015). An example T6SS effector protein found in *B. cenocepacia* is TecA, an asparagine deamidase shown to aid in the survival of *B. cenocepacia* within host macrophages. This acts through actin disruption mediated through inactivation of Rho GTPases (Aubert et al., 2016).

1.9.3 *Burkholderia pseudomallei*

The final bacterium as part of this study, *Burkholderia pseudomallei*, is the causative agent of the infectious human disease melioidosis (Whitmore, 1912). While most T6SS encoding organisms contain a single T6SS gene cluster, as in *Aeromonas hydrophila* and *Burkholderia cenocepacia*, *Burkholderia pseudomallei* stands out as it contains six (Shalom, Shaw, and Thomas, 2007) (figure 1.2). These T6SS gene clusters are proposed to aid in intracellular survival in macrophages and aid in the secretion of toxic factors which cause actin polymerisation and the formation of multi-nucleated giant cells (Shalom, Shaw, and Thomas, 2007; Chen et al., 2011). Except for cluster 5 (T6SS-5), which is most likely to contribute to macrophage intracellular survival and pathogenicity, the specific contribution of each of the gene clusters is currently unknown (Shalom, Shaw, and Thomas, 2007; Chen et al., 2011).

This thesis aims to:

- Expand the structural knowledge of both TssA and TssK from the T6SS through X-ray crystallographic studies.
- Explore the structural differences between species variants of TssA and TssK.
- Investigate the quaternary assembly of TssA and TssK and how this may relate to their respective biological function within the T6SS assembly.

- Identify any structural homologs with known functions to suggest a possible role for TssA and TssK.

Chapter 2

Material and Methods

2.1 General Methods and recipes

2.1.1 Antibiotics

The following antibiotics were used as selection markers for plasmid maintenance, and made up in H_20:

TABLE 2.1: Antibiotics

Antibiotic	Stock concentration (mg/ mL)
Ampicillin	100
Chloramphenicol	25*

*Dissolved in 100 % ethanol.

2.1.2 LB medium

Lysogeny Broth (LB) (Miller, 1972) was created by mixing 10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl in distilled H_20 before auto-claving at 121 °C for 15 minutes.

2.1.3 LB Agar

LB Agar was prepared by mixing 10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl in distilled H_20, once all ingredients thoroughly dissolved 15 g/L Bacto-Agar added to the solution before auto-claving at 121 °C for 15 minutes.

2.1.4 SOC media

Super optimal broth with catabolite repression (SOC) media (Hanahan, 1983) was prepared by mixing 20 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 0.186 g/L KCl, 0.952 g/L

$MgCl_2$, 2.47 g/L $MgSO_4$ and 3.6 g/L glucose in distilled H_20 before auto-claving at 121 °C for 15 minutes.

2.1.5 BHI media

Brain Heart Infusion (BHI) media (Rosenow, 1919; Haden, 1923) was prepared by adding 37 g/L of Oxoid pre-prepared BHI media powder to distilled H_20 before auto-claving at 121 °C for 15 minutes.

2.1.6 Agarose gel

Agarose gels were used to optically analyse DNA products during cloning. 1 % agarose gels were prepared by dissolving 0.4 g Iberose in 40 mL TAE buffer (40 mM Tris pH 7.6, 20 mM acetic acid, and 1 mM EDTA) with the addition of 4 μL gelRed (1:10000 dilution) the solidified gel would then be submerged in further TAE buffer for the samples to be loaded. 5 μL of hyper ladder 1 kb was used as the DNA ladder reference and gels were run for 45 minutes at 100 V before visualisation under UV light.

2.1.7 Miniprep, PCR cleanup and gel extraction

All DNA purification stages were carried out using the relevant NEB Monarch Nucleic Acid Purification Kits (Monarch Plasmid Miniprep Kit, Monach PCR & DNA Cleanup Kit, and Monarch DNA Gel Extraction Kit) with the included protocols with the alteration that DNA samples were incubated for 10 minutes on ice after the addition of 35 μL autoclaved MilliQ H_20 before elution into a 0.5 mL tube.

2.1.8 Transformations – Cloning and expression strains

For plasmid propagation and transformation of ligation products the *E. coli* strains DH5α (Bioline), and silver efficiency DH5α (Bioline) were used respectively. For protein over expression the *E. coli* strain BL21 DE3 (Agilent) was utilised. All strains were chemically competent and pre-aliquoted into 50 μL volumes. For transformation cell aliquots were defrosted on ice for 5 minutes before the addition of 2-3 μL of > 2 ng/ μL plasmid/ DNA samples to the cell suspension. Cells remained on ice for a further 30 minutes before the cell suspension was heat shocked at 42 °C for 30 seconds. The cell suspension was then

placed back onto the ice for a further 5 minutes before the addition of 100 μL pre-warmed SOC media. This suspension was then incubated at 250 rpm at 37 °C before being plated onto LB agar plates with the relevant antibiotic. Plates were subsequently incubated at 37 °C overnight to allow colony growth.

2.1.9 Plasmid propagation

To produce suitable yields of plasmids for down-stream cloning or protein expression, the plasmid of interest was transformed into 50 μL of DH5α *E. coli* as described in 2.1.8. Colonies containing the plasmid of interest were added to 5 mL LB including the required antibiotic marker (2.1), these cultures were then incubated at 250 rpm at 37 °C overnight. After the cultures had grown the cells were pelleted at 4000 g for 20 minutes and the LB discarded. The plasmids were then liberated via miniprep.

2.1.10 Colony PCR

To validate if the correct PCR product from the ligation reactions had been inserted correctly into the expression vector of choice colony PCR was utilised. This technique is used to amplify the multiple cloning site found in the chosen expression vector and any possible insert found within in via PCR. For this, colonies that have grown on LB agar selection plates, which hopefully contain the ligation product of choice, were picked, re-streaked on LB Agar + antibiotic plates and then placed in 5 μL autoclaved MilliQ H_20. These samples were then boiled for 15 minutes on a dry heat block (105 °C) before the cell debris was removed through pulse centrifugation. The colony PCR reaction was set up as follows:

TABLE 2.2: PCR reaction

PCR reaction	**Volume: (μL)**
T7 (F) (10 pmol/μL)	0.5
T7 (R) (10 pmol/μL)	0.5
DNA template (released from cells)	4
Biomix PCR master mix (Bioline)	5
Total	10

The finished reactions were then run on a 1 % agarose gel to determine if successful ligation had occurred by the presence of a band at the inserts corresponding base pair length.

TABLE 2.3: PCR protocol

Segment	Cycles	Temperature	Time
1	1	95 °C	2 minutes
2	28	95 °C	30 seconds
		58 °C	30 seconds
		72 °C	1 minute
3	1	72 °C	2 minutes

2.1.11 DNA sequencing

Colonies confirmed to contain the correct construct through colony PCR were grown in 5 mL LB + antibiotic overnight at 37 °C. These cultures were then pelleted at 4500 rpm (3644 g) before plasmid liberation via miniprep before elution in 35 μL autoclaved MilliQ H_20. 5 μL of plasmid DNA at concentration 100 ng/μL or greater was added to 5 μL T7 forward primer (10 nM), the 10 μL sample was then sent to GATC Biotech, Germany for Sanger sequencing. Sequencing results determined if the generated construct contained any mutations in the sequence.

2.1.12 SDS-PAGE gels (cast)

SDS-PAGE (Sodium dodecyl sulfate-polyacrylamide gel electrophoresis) (Laemmli, 1970) was utilised to analyse protein over-expression, protein purification and any other technique requiring protein mass quantification. All SDS-PAGE gels were created to a 1 mm thickness in a BioRad Mini-PROTEAN casting glass plate with a 12 % w/v acrylamide resolve and a 6 % w/v acrylamide stack. The recipe for each is as follow:

TABLE 2.4: SDS-PAGE (resolve)

12 % resolve	Volume per 1 mm gel
30 % acrylamide	2.5 mL
1 M Tris pH 8.8	2.35 mL
MilliQ H_20	1.28 mL
10 % SDS	62.5 μL
TEMED	6.25 μL
10 % APS*	62.5 μL

*APS - 10 % w/v solution prepared by adding 0.1 g APS to 1 mL MilliQ H_20.

TABLE 2.5: SDS-PAGE (stack)

6 % stack	**Volume per 1 mm gel**
30 % acrylamide	0.75 mL
1 M Tris pH 8.8	0.47 mL
MilliQ H_20	2.46 mL
10 % SDS	37.5 μL
TEMED	3.75 μL
10 % APS*	37.5 μL

*APS - 10 % solution prepared by adding 0.1 g APS to 1 mL MilliQ H_20.

The 12 % resolve solution was prepared first, with the addition of the TEMED (N,N,N',N'-tetramethylethane-1,2-diamine) and 10 % APS being added last. The complete solution was poured into a casting plate, with a layer of isopropanol added to generate a smooth finish devoid of air bubbles while the gel set. Once the resolve had solidified, the isopropanol was removed and the 6 % stack solution was created in the same way as the resolve. The stack solution was then poured in and a sample-loading comb added. When the gel had fully set it was placed in a Mini-PROTEAN Tetra Electrophoresis system, along with a Mini cell buffer dam, and the sample-loading comb was removed. The entire system was then flooded with 1x SDS-PAGE running buffer (3 g/L Tris, 14.4 g/L glycine, 1 g/L SDS) before samples were applied to the wells.

Protein samples for SDS-PAGE were typically cells that had over-expressed protein or protein samples that had been purified. In either case, $\sim$ 15 μg of protein would be obtained in 15 μL of 50 mM Tris pH 8.0. To this 2 μL of reducing agent (10x NuPAGE sample reducing agent (Thermo Fisher Scientific)) and 5 μL of sample buffer (NuPAGE SDS Sample Buffer (Thermo Fisher Scientific)) would be added, before being boiled for 15 minutes on a dry heat block (105 °C) and the debris being removed through pulse centrifugation.

SDS-PAGE samples were loaded into the gel using a Hamilton needle along with 5 μL of the protein ladder Mark12 (Thermo Fisher Scientific). The gel was then run at 100 V for 10 minutes followed by 200 V for 34 minutes. The gel was next removed from the casting plate and submerged in Coomassie blue stain for 5-10 minutes before being de-stained and the gel visualised.

2.1.13 SDS-PAGE gels (pre-cast)

Novex NuPage Pre-cast SDS-PAGE gels were also used for the analysis of protein over expression, purification and general protein work. Unlike the cast SDS-PAGE gels, the pre-cast SDS-PAGE gradient gels consisted of 4-12 % w/v acrylamide Bis-Tris and utilised Novex NuPage x20 MES SDS running buffer. Protein sample preparation was as described above (2.1.12).

SDS-PAGE samples were loaded into the gel using a Hamilton needle along with 5 μL of the protein ladder Mark12 (Thermo Fisher Scientific). The gel was then run at 100 V for 10 minutes followed by 200 V for 34 minutes. The gel was next removed from the pre-cast plastic cartridge and submerged in Instant Blue (Expedeon) for 5-10 minutes before being washed with H_20 and visualised.

2.1.14 Bradford assay

The protein concentration of whole cell extract was estimated using a WPA lightwave II 7120 photo-spectrometer. Reactions were set up in 1 mL cuvettes containing 1 μL protein sample, 0.2 mL Bradford reagent solution (Bio-Rad), and 0.8 mL H_20. A_{595} nm measurements were then taken and the protein concentration calculated by:

$$\text{Concentration (mg/mL)} = \frac{\text{Absorbance}_{(595\text{ nm})} \text{ x } 15}{\text{volume } (\mu\text{L})} \tag{2.1}$$

2.1.15 Nanodrop

DNA and protein concentrations were estimated utilising an Implen Nanophotometer 7122 nanodrop. This allowed, in the instance of double-stranded DNA, for the absorbance at 260 nm to be measured and the concentration in ng/μL to be determined. In the case of proteins, which contained sufficient aromatic residues, the concentration could be estimated through absorbance at 280 nm this is calculated by:

$$\text{Concentration (mg/mL)} = \frac{\text{Absorbance}_{(280\text{ nm})} \text{ x lid factor}}{\text{extinction coefficient } (\text{M}^{-1}\text{ cm}^{-1})} \tag{2.2}$$

The extinction co-efficient was calculated using Protpram (Gasteiger et al., 2005).

2.1.16 gDNA template generation

Concentrated genomic DNA (gDNA) from *B.pseudomallei* D286 (Lee et al., 2007) and *B. cenocepacia* H111 (Romling et al., 1994) was kindly provided by Sheila Nathan, Malaysia and Mark Thomas, Department of Infection and Immunity, Sheffield, respectively. gDNA was subsequently diluted down to 200 ng/μL for use in cloning.

2.2 Cloning

2.2.1 Primer design

Primers were designed for the incorporation of the desired gene to be inserted within the pET-21a(+) expression vector (figure 2.1). NEBcutter and the NEB Q5 primer Tm calculator were utilised to check for restriction enzyme compatibility and check the melting temperatures of the primer pairs such that they were sufficiently balanced for the polymerase reaction, respectively. The final primer designs were ordered from Eurofins, and lyophilised primers were made up to 100 pmol/μL with MilliQ H_20, however, the working concentrations of primers in PCR was 10 pmol/μL.

2.2.2 PCR of target DNA

Amplified DNA of the desired genes was achieved through PCR. Previously purified gDNA (2.1.16) or an expression vector containing the gene of interest was used as the template source. The PCR reaction was set up as in table 2.6. Due to the high GC content of the gDNA, high melting temperatures (Tm's) were used during primer design, consequently, the following PCR protocol (2.7) was used almost exclusively, with alterations stated otherwise. All PCR reactions were carried out on a Techne 3Primer thermocycler. 4 μL of the finished reaction was mixed with 1 μL Bioline 5x DNA loading buffer and analysed using a 1 % agarose gel, if successful the remaining PCR product was purified (2.1.7).

2.2.3 Restriction digest and Alkaline phosphotase treatment

Once the gene of interest had been successfully amplified through PCR. The amplicon and stock pET21a(+) vector were digested, separately, with compatible restriction enzymes as selected during primer design. This generated compatible DNA 'sticky' ends to the linear

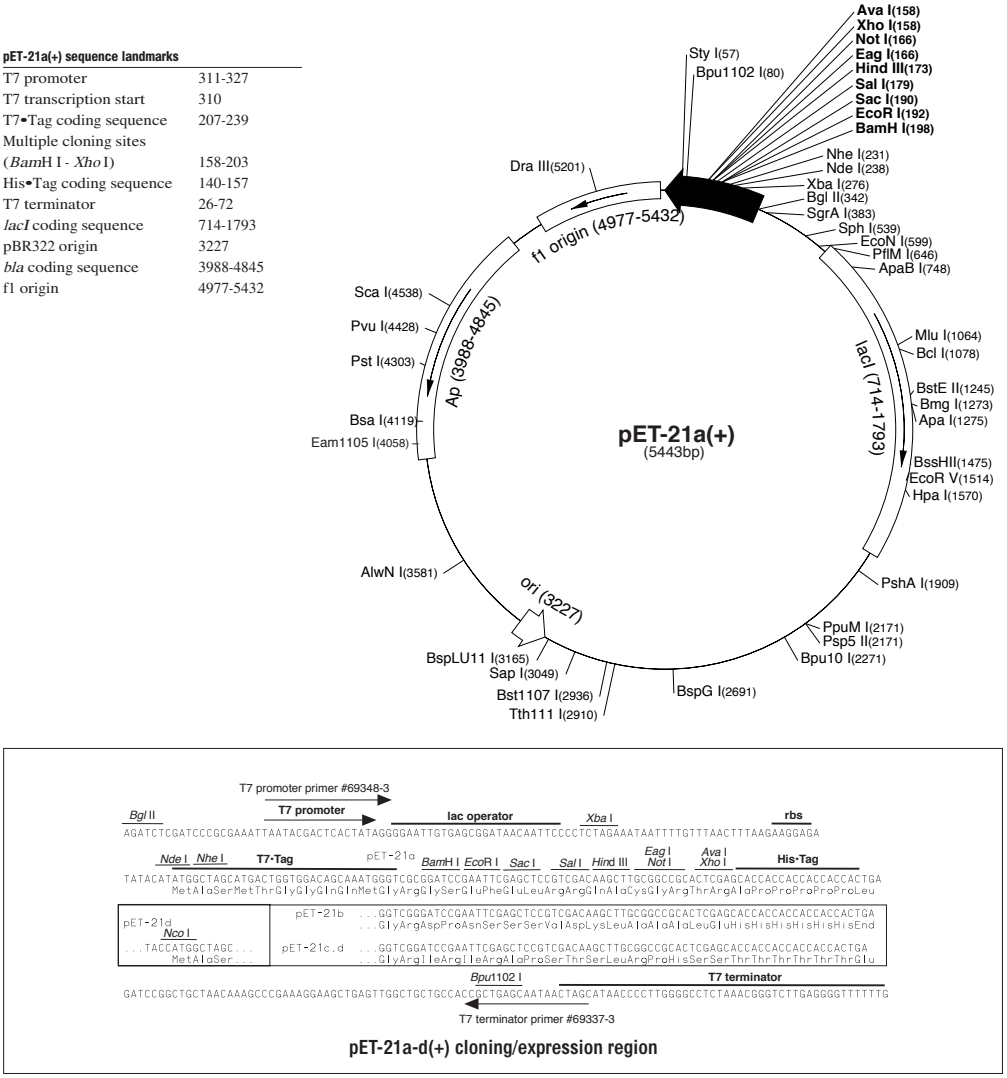


FIGURE 2.1: **pET-21a(+) Vector Map (Novagen).** pET-21a(+) map indicating the resistance marker and suitable restriction sites within the multiple cloning site (MCS).

TABLE 2.6: PCR reaction

PCR reaction	**Volume: (μL)**
Primer F (10 pmol/μL)	1.25
Primer R (10 pmol/μL)	1.25
Q5 mastermix	12.5
gDNA/template DNA (200 ng)	1-5
ddH_20	5-9
Total	25

TABLE 2.7: PCR protocol

Segment	**Cycles**	**Temperature**	**Time**
1	1	98 °C	30 seconds
2	25	98 °C	10 seconds
		72 °C	25 seconds
		72 °C	30 seconds
3	1	72 °C	2 minutes

DNA fragments for their subsequent joining. The reactions were set up as in table 2.8 and incubated at 37 °C for one hour. Reactions containing restricted vector DNA had 1 μL Shrimp Alkaline Phosphotase (NEB) added before a further 1 hour incubation at 37 °C. All reactions were subsequently stopped with an incubation of 30 minutes at 65 °C followed by a DNA purification step (2.1.7) to remove non-DNA components.

TABLE 2.8: Restriction digest

Digest reaction	**Volume: (μL)**
Restriction Enzyme 1	2
Restriction Enzyme 2	2
x10 CutSmart buffer (NEB)*	5
PCR product/ vector	35
ddH_20	6
Total	50

* Where appropriate x10 Cutsmart buffer was replaced with x10 NEBuffer 1-4

2.2.4 DNA ligation

Following restriction digest the gene amplicons and linearised vectors contain compatible sticky ends, the ligation reaction utilises the T4 ligase enzyme to join the termini of the two

DNA fragments and form a phosphodiester bond, in an ATP-dependent reaction. Thus incorporating the gene amplicon into the expression vector. The ligation reactions were often set up as in table 2.9 resulting in a molar excess of gene product:vector as to promote gene incorporation. The reaction was incubated at room temperature for 2 hours before 5 μL was placed overnight at 4 °C, whilst the remainder of the reaction was used to transform silver efficiency DH5α cells. If the transformation was unsuccessful, then the remaining ligation mixture was transformed into further silver efficiency DH5α cells.

TABLE 2.9: Ligation reaction

Ligation reaction	**Volume: (μL)**
T4 ligase buffer	1
T4 ligase	0.5
restricted PCR product	4
restricted vector	1
ddH_20	4.5
Total	10

2.3 Protein over expression

Protein over expression of all protein constructs was carried out in transformed BL21 DE3 *E. coli*. Over expression was achieved through the exploitation of the lac operon present in the expression plasmids used for molecular cloning. IPTG was used as a non-hydrolysable analogue to lactose; so continuous over expression of the desired protein occurred.

2.3.1 Trial expression protocol

Constructs provided by Dr Mark Thomas had an established expression protocol, so they did not require expression trials, however, those cloned during this project required testing to see if the constructs produced suitable quantities of protein for downstream processes. Trial expressions were carried out by inoculating 50 mL LB or BHI media and relevant antibiotic with a single colony of *E. coli* transformant, before incubating the cultures at 37 °C 250 rpm overnight. 50 mL of LB/BHI and antibiotic was then inoculated with 2 % (v/v) of the overnight culture and incubated at 37 °C 250 rpm until the OD_{600} had reached between 0.6 – 0.7, after which 1 mM IPTG was added before a further 4 hour incubation at 37 °C 250

rpm. Protein expression was determined through SDS-PAGE; temperature conditions were then adjusted if necessary to optimise protein expression/solubility.

2.3.2 Large scale over expression

To produce sufficient quantities of protein for protein crystallography large-scale expression cultures had to be produced. To do so 500 mL LB/BHI (construct dependent) and suitable resistance markers were inoculated with 1 % (v/v) overnight culture. Cultures were then incubated at 37 °C 200 rpm until an OD_{600} of between 0.6 and 0.7 was achieved. 1 mM IPTG was then added, before further incubation at either 37 °C or 30 °C (construct dependent) for an additional 4 hours. Cells were then pelleted by centrifugation at 13000 *g* for 20 minutes, 4 °C before the cell paste was recovered and frozen at -20 °C until required.

2.3.3 Selenomethionine over expression

Often in crystallography, the incorporation of selenomethionine into the protein target is utilised to provide a source of phasing in downstream data-collection from protein crystals. To do so a separate expression protocol is required. Minimal media supplemented with selenomethionine was made by adding 10.5 g/L K_2HPO_4, 4.5 g/L KH_2PO_4, 1.0 g/L $(NH_4)SO_4$, 0.5 g/L Tri-sodium citrate, 5 g/L glycerol, and 0.5 g/L Adenine, Thymine, Guanosine and Uracil to distilled water before autoclaving, producing the minimal media base. An amino acid solution was also created, this contained: 1 g/L $MgSO_4.7H_20$, 4 mg/L Thiamine, 50 mg/L of L-valine, L-leucine, L-isoleucine, and 100 mg/L of L- threonine, L-lysine and L-phenylalanine all added to 20 mL distilled water creating a 50x stock and stored at – 20 °C. Final selenomethionine minimal media was created by the addition of 10 mL amino acid solution, antibiotic selection marker and 20 mg selenomethionine to 500 mL volumes of minimal media base. Over expression cultures were initiated as described previously and grown to an OD_{600} of 0.6 - 0.7. Cells were then spun down at 8000 *g*, 20 °C for 20 minutes, before being washed twice with selenomethionine minimal media to remove any contaminating LB media. Cells were then resuspended in 500 mL selenomethionine minimal media and incubated at 37 °C 200 rpm for 1 hour to allow the cells to recover. After this cells were induced with 1 or 0.1 mM IPTG (construct dependent) and over expression carried on as previously described.

2.3.4 Extraction of cell free extract

To liberate the over expressed protein for further work the *E. coli* cells were fractured using sonication as the source of lysis before the cell debris was removed from the cell free extract (CFE). To achieve this frozen cell pellets from over expression were re-suspended in lysis buffer (50 mM Tris (pH 8.0)), typically 5 mL per 1 g of cell paste. The solution was then subjected to sonication using a Soniprep 150 probe set to 18 microns. Sonication was carried out using 2 x 20 s bursts on ice to keep the solution from warming up. Alternatively, cell lysate was provided in 50 mM Tris (pH 8.0), 200 mM NaCl, 10 % glycerol and PMSF (25 mg/mL). The resulting lysate was then centrifuged at 35000 *g*, 4 °C for 30 minutes to remove the cell debris from the CFE containing the over expressed protein. CFE was then kept at 4 °C for protein purification.

2.3.5 Protein over expression analysis

To determine if protein expression had been successful, 1 mL samples were taken at required time points (e.g. before and after IPTG induction) and spun down at 13000 rpm (20000 *g*). Subsequently, samples were lysed as in 2.3.4, before the level of protein expression was analysed utilising SDS-PAGE gels as described in 2.1.12.

2.4 Protein Purification

Protein purification was used to isolate the desired over expressed protein product from the remaining clarified *E. coli* cell lysate such that it was suitable for subsequent experiments. All chromatography steps utilised an ÄKTA pure or ÄKTA purifier maintained at room temperature.

2.4.1 Nickel affinity chromatography

A number of the protein constructs used during this work contained a poly-histidine N or C-terminal tag. This allowed for them to be purified through IMAC purification. CFE was passed through a 5 mL Ni-NTA column (GE Healthcare), before passing 3 CV (column volumes) of 500 mM NaCl, 50 mM Tris (pH 8.0), 20 mM Imidazole (buffer A) through the column at 4 mL/min to remove unbound material. A gradient of buffer B (500 mM NaCl, 50

mM Tris (pH 8.0), 500 mM Imidazole) was then applied over a linear gradient of 0 – 100 % for 10 CV, the fractions eluted were then analysed by UV and SDS-PAGE. Suitable fractions were then pooled together.

2.4.2 Anion exchange chromatography

Anion exchange chromatography allows for the separation of proteins based on their charge, such that the positive charge of the resin allows it to interact with negatively charged groups present on protein residues or DNA. Typically, CFE diluted in 0/50 mM NaCl, 50 mM Tris (pH 8.0) is applied to a 5 mL HiLoad Diethylaminoethyl (DEAE) FF weak anion exchanger (GE Healthcare) column with a flow rate of 5 mL/min, before the bound fraction is eluted over a gradient of 0-500 mM NaCl. Fractions are collected throughout the purification and analysed by UV and SDS-PAGE. An additional 'polishing' step can be applied to the protein sample utilising a 6 mL resourceQ column (GE healthcare) with Source 15Q resin, which is a strong anion exchanger. Like the DEAE column, the ResourceQ column utilised a 5 mL/min flow rate and the bound sample is eluted over a gradient of 0-500 mM NaCl. Fractions are collected throughout the purification and analysed by UV and SDS-PAGE.

2.4.3 Ammonium sulphate precipitation

Proteins have different solubility in solution which can be exploited to remove contaminating proteins within the protein prep by altering the salt concentration of the protein solution. 4 M ammonium sulphate is slowly added to the CFE until the protein of interest precipitates. The sample is then centrifuged at 50000 g for 10 minutes, after which the remaining supernatant is removed and the precipitated protein is re-suspended in 50 mM Tris (pH 8.0).

2.4.4 Gel filtration chromatography

Gel filtration (size exclusion) chromatography allows for the separation of proteins based on size. This step was often applied as a final purification stage if impurities were still present in the protein preps. Protein samples were concentrated to a maximum volume of 2 mL before being applied to a 16/60 column (GE Healthcare) packed with 200 pg superdex (GE Healthcare) resin, which had been pre-equilibrated with 500 mM NaCL, 50 mM Tris (pH 8.0). The column was run at 1.2 mL/min and fractions were collected after a delay of 0.3 CV had

passed. The eluted fractions were then analysed by UV and SDS-PAGE. Protein standards were used to generate a calibration plot so that the molecular weight of the protein could be estimated from the elution volume from the column.

2.5 Protein Crystallisation

A general overview of the practical methods of protein crystallography used during this body of work are detailed below. The theory outlining these techniques are described in the textbooks, Crystallography Made Crystal Clear (Rhodes, 2006) and Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology (Rupp, 2009).

2.5.1 Preparation for protein crystallisation

Proteins from purification columns frequently eluted over many fractions, and as a result would be dilute and not suitable for either the next stage of purification or to be taken forward for downstream experiments. The protein, therefore, needed to be concentrated. Protein samples would be applied to a VIVAspin concentrator with a molecular weight cut off smaller than that of the desired protein. This would then be centrifuged at room temperature at 4500 g until the desired volume or protein concentration was achieved. Proteins were often eluted in buffers with high salt concentrations, therefore the buffer solution needed to be exchanged prior to crystallisation work into low salt conditions. A Zeba spin column would be equilibrated by passing 3 x 300 μL of the desired buffer through the column by centrifugation at 1200 g for 1 minute. Up to 120 μL of purified protein would then be applied to the column before centrifugation at 1500 g for 2 minutes which exchanged the buffer.

2.5.2 Initial crystallisation trials

Initial screening to identify suitable conditions that would promote crystallisation was carried out by subjecting the purified protein constructs to sparse-matrix crystallisation screens. A number of 96-well sparse-matrix screens are commercially available; these include PACT, JCSG+, ProPlex, AmS0_4 screen, Morpheus, pH clear, MPD, and Classics (Molecular dimensions and Quiagen). Initial screening was carried out in 96-well vapour diffusion crystallisation tray, where each well consisted of a large well for mother liquor and two smaller wells for crystallisation reactions. Small-volume liquid handling was carried out either by

a Matrix Hydra II PlusOne crystallisation robot or a TTP Mosquito. To generate each tray, 50 μL of crystallisation solution was dispensed into each of the large wells. Crystallisation reactions were then set up in a 1:1 ratio of 100 nL protein solution and 100 nL crystallisation solution in the small wells as sitting drops, the trays were sealed with clear plastic film before being centrifuged at 1200 rpm for two minutes. The trays were then incubated at either 17 °C or 7 °C to allow for equilibration of the solutions and vapour diffusion, conditions were frequently checked to see if any crystals had successfully grown.

2.5.3 Hanging drop optimisation

Once a condition had been found that generated protein crystals, these conditions would often undergo optimisation in an attempt to improve both the size of the crystals and their diffraction properties. Protein crystal optimisation utilised hanging drop vapour diffusion set up in 24-well tissue cultures trays containing 500 μL mother liquor of crystallisation solution in each well. Hanging drops were typically created by mixing 1 μL protein sample with 1 μL mother liquor, producing a 1:1 ratio, and placing it on a siliconised coverslip, before inverting the coverslip and sealing it onto the well through the use of immersion oil. The optimisation would utilise a 2D matrix of crystallisation conditions generated around the initial crystallisation condition. This often involved varying the pH and precipitant concentrations in 0.5 pH and 2 % w/v steps, respectively.

2.5.4 Cryo-protectant generation

Protein crystals are cryo-cooled to 100 K to reduce radiation damage during data collection. However during the cryo-cooling process, ice crystals can form disrupting the protein crystal lattice. In addition, ice rings can compromise the quality of the X-ray data. Cryo-protectants are required if little/no low molecular weight 'anti-freeze' molecules are present in the crystallisation conditions, such as ethylene glycol, low molecular weight PEGs, glycerol, MPD, and monosaccharide sugars. To generate cryo-protectants, low molecular weight 'anti-freeze' solutions, typically either ethylene glycol or glycerol were added to 1 mL of mother liquor from the condition the crystal grew in. Usually, up to 25 % of cryo-protectant was added by replacing the water in the solution.

2.5.5 Heavy metal soaking

An alternative source of experimental phasing to that of selenomethionine incorporated protein is to soak heavy metals into protein crystals to generate a derivative. For example, mercury can react with exposed cysteine residues present in the protein. Therefore, $\sim 20\mu$g of ethylmercury phosphate (Hampton research) powder was added to the crystal optimisation drop and left to diffuse into the crystal for 5, 10, 30 and 60 minutes before looping.

2.5.6 Looping and cryo-cooling

Prior to data collection, crystals were looped in 0.1 – 0.5 micron litholoops (Molecular Dimension) pre-mounted on magnetic cryo-bases (Molecular Dimension), where the loop size was selected to try to match that of the crystal. Crystals were moved from their mother liquor to suitable cryo-protectant before subsequently being cryo-cooled in liquid nitrogen. They were then placed in Diamond Light Source (DLS) standard sample containers compatible with DLS sample handlers.

2.5.7 Mass spectrometry of protein crystals

To validate if a particular crystal that had grown was the intended protein, they were sent for mass spectrometry. Crystals of interest were looped and then passed through multiple drops of H_20 to wash away contaminating well solution and finally the crystal was dissolved in 5 μL H_20 before being sent to the biological mass spectrometry facility, Sheffield for molecular weight determination.

2.6 Crystal screening and data collection

All testing and data collection of protein crystals was carried out on the MX beamlines at 100 K at DLS, Oxfordshire, United Kingdom. The MX beamlines, in general, have tunable wavelength X-ray sources, which allow for various collection strategies to be employed, for example Single-wavelength anomalous dispersion (SAD). In addition, the beamlines are equipped with Pilatus (DECTRIS) detectors which allow for rapid data collection. Specific details of the data collection strategies are described below.

2.6.1 Testing

To test the diffraction potential of protein crystals a standard beamline strategy was used. This consisted of 0.2 s exposure, 0.2 ° oscillation, 40 % transmission, a delta of 45 ° and 5 images collected per sample. This provided suitable evidence for the potential resolution to which data could be collected. Either imosflm (Battye et al., 2011) or EDNA (Leal et al., 2011) were used to determine preliminary characteristics of the crystal including the unit cell parameters and spacegroup.

2.6.2 Data collection

Once a protein crystal was determined to be of suitable quality for data collection a data collection strategy was calculated. For a native dataset, data would be collected based on the rotation range the reciprocal space required, based on the estimated space group. Typically, between 200 ° and 360 ° of data would be collected, with an image collected over an oscillation of 0.1 °, an exposure time of 0.05 s, and a beam intensity between 40-100 % depending on the stability of the crystal in the X-ray beam. Native datasets were typically collected at a wavelength of 0.9795 Å (12658 eV). However, this was dependent on the beamline the experiment was carried out on. For example, I04-1 is fixed at 0.9159 Å. Experimental phasing took advantage of SAD data collection strategy. For selenomethionine incorporated protein, a fluorescence scan was carried out near the K edge of selenium, this was analysed by COOCH (Evans and Pettifer, 2001) to identify the peak wavelength to use in data collection, in addition, provide values for f' and f''. Furthermore, derivative crystals containing mercury were collected at a wavelength of 1.0088 Å. This wavelength is near the L-III absorption edge for mercury. High multiplicity datasets were collected at the peak wavelength for the respective heavy atom depending on which atom was incorporated. To generate high multiplicity data, the beam intensity would be attenuated to typically 5-30 %, to avoid excess radiation damage. In addition, multiple isomorphous datasets would be collected, typically from the same crystal, allowing for the images to be processed together. All images were collected on a Dectris Pilatus 6M detector, irrespective of the beamline used. Specific details of each data collection can be found in the relevant manuscripts or appendices.

2.6.3 Data Processing

Data processing utilises many programs which are often packaged into automated pipelines for immediate processing once the data has been collected. Fundamentally, each pipeline consists of programs that can carry out spot detection on the diffraction images, indexing of the reflections, deciding the spacegroup, scaling the data, merging the data and data reduction. For the work presented in this thesis, the FastDP and Xia2 pipelines were utilised. FastDP utilises XDS (Evans, 2006; Kabsch, 2010) to carry out spot detection and reflection indexing, subsequently XDS and POINTLESS (Evans, 2011) are used for indexing and spacegroup determination and finally, XDS, POINTLESS and Scala (Winn et al., 2011) are used for scaling, merging and data reduction. Xia2 has two optional setting that were used. The first of them is DIALS, which utilises the DIALS (Winter, 2010; Winter and McAuley, 2011) spot finding and indexing software, with POINTLESS being used to determine the space group and Aimless (Evans and Murshudov, 2013) is used for scaling, merging and data reduction. The second option is 3dii, which utilises the XDS spot finding software before using downstream processes identical to DIALS. CC $_{1/2}$ was used to determine the resolution cutoff in the data sets (Karplus and Diederichs, 2015).

2.6.4 Phasing and structure determination

The phase problem

The amplitude of each of the reflections can be measured by data collection, however the phase for each reflection is lost. Therefore, a source of phasing is required to perform a reverse Fourier transform to generate an electron density map. Currently, the phase information can be obtained either through experimental phasing or from a homologous protein structure by molecular replacement.

Experimental phasing - Single wavelength anomalous dispersion (SAD)

Derivative crystals generated through the incorporation of selenomethionine (2.3.3), or through heavy-atom soaking (2.5.5) were used as sources of experimental phasing during the project. SAD takes advantage of the anomalous scattering property exhibited by heavy atoms at X-ray wavelengths that correspond to their absorption edge. The SHELX suite (Sheldrick,

2008) was utilised to generate a heavy-atom substructure, hand determination, solvent flattening and produce an initial electron density map and model. The software package consists of three programs SHELXC, SHELXD and SHELXE, which were run using the default settings. SHELXC determines the amount of anomalous signal available in the data over the full resolution range of the data. SHELXD was used to identify the location of the heavy atoms within the asymmetric unit. This process is guided based on the number of expected heavy atoms based on the protein sequence and the compatible heavy atom used, e.g. the number of selenium atoms will be related to the number of methionines in the protein sequence. The final program is SHELXE, which utilises the heavy atom sites determined by SHELXD to calculate an initial phase set for the whole data. This involves initial hand determination and iterative density modification. Both the original and inverted hands are scored based on mean FOM and pseudo-free CC, where the one with higher values is selected, and the resultant electron density map is subjected to rounds of automated model building. The electron density map from both hands were visualised. If the map contained connecting density indicative of protein secondary structure, that hand was selected. Alternatively, the CRANK2 software package was used for experimental phasing (Skubák and Pannu, 2013). CRANK2 utilises SHELX. However, it can optionally use SOLOMON (Abrahams and Leslie, 1996) and Multicomb (Skubák, Waterreus, and Pannu, 2010) as alternatives to SHELXE for density modification and phase combination, respectively. The details of the software configuration can be found in the appendices. The resultant electron density map and initial model is then used to guide model building.

Molecular replacement

Molecular replacement, using Phaser (McCoy et al., 2007), was utilised as a source of phasing if a suitably similar homologue structure was available in the protein database. The software employs a rotation and translation search function, using maximum likelihood statistics. The phase set generated by placing the search model is then used to generate an initial electron density map from the target structure factors. The success of the solution is then determined based on the refined TFZ (Translation function) score, the refined LLG (log-likelihood gain) and any potential packing clashes. Ultimately, the resultant electron density is analysed to identify the success of the solution.

2.6.5 Model building and refinement

Once a suitable initial map had been generated, model building was initiated. This process often began with automatic model building software, e.g. BUCCANEER (Cowtan, 2006). After an initial model was built, manual model building and map visualisation would be carried out using COOT (Emsley et al., 2010). Density modification using PARROT (Cowtan, 2010) was often used improve the map if multiple copies of the molecule existed within the asymmetric unit. Refinement of the model was carried out using REFMAC5 (Murshudov, Vagin, and Dodson, 1997; Murshudov et al., 2011) with the use of bias checks throughout iterative model building and refinement cycles. The model was deemed finished once the major features in the electron density map had been accounted for, and the R_{work} and R_{free} values had converged.

2.6.6 Model validation and analysis

Once the significant features in the electron density map were accounted for, and the refinement had converged, the model would be validated to check for correct bond angles and lengths, Ramachandran angles, favoured side chain rotamers and any packing clashes. Validation took advantage of the MolProbity server (Chen et al., 2010). If any major problems with the model were flagged, they would be assessed and the model manually rebuilt and refined to improve these issues. The final model is provided with a score determined by its various validation parameters compared to those already available in the PDB. The finished model is then suitable for structure deposition to the PBD. Figures containing protein models used PyMOL (The PyMOL Molecular Graphics System, version 1.8/2.0 Schrödinger, LLC). Figures with electron density were generated using either COOT (Emsley et al., 2010) or FFT (Read and Schierbeek, 1988). Structural alignment of the protein model was often used to identify the similarity between related protein structures. This used GESAMT (Krissinel, 2012) which calculated the RMSD between two models based on the alignment of the Cα backbone over equivalenced regions of residues.

Chapter 3

Publications

1. TssA from *Aeromonas hydrophila*: Expression, purification and crystallographic studies, (Dix et al., 2018a).

2. Structural studies provide a unified framework for the function of type VI secretion system TssA subunits, (Dix et al., 2018b).

3. TssK from *Burkholderia cenocepacia*: Expression, purification and preliminary crystallographic studies, **prepared for submission**.

4. The role of mobility and sequence conservation in the structure/function relationships of *Burkholderia* TssK, **prepared for submission**.

3.1 Paper 1

3.1.1 Summary

The first paper presented in this thesis is a crystallisation article which discusses the generation, expression and purification of protein constructs generated from *tssA* (AHA1844) of *A. hydrophila* (Ah-TssA). In addition, the paper details suitable conditions that generated diffraction quality crystals and subsequent preliminary data collection attempts. During this body of work, structural studies on TssA (Ec042_4540) from EAEC (Ec-TssA) suggested that TssA is comprised of three domains (Nt1, Nt2 and CTD) (Zoued et al., 2016). A sequence alignment of Ah-TssA and Ec-TssA led to the generation of five constructs. These include the full-length TssA protein, and four domain constructs (Nt1, Nt2, Nt2-CTD and CTD). Successful crystallisation conditions were found for constructs Nt2 and Nt2-CTD, and details of preliminary crystal analysis are discussed.

3.1.2 Protein summary

TABLE 3.1: Summary of TssA proteins discussed

Paper I			
Construct ID	**Residue range**	**Resolution**	**Space group**
Ah_TssA2^{A}-FL	(1-478)	N/A	N/A
Ah_TssA2^{A}-Nt1	(1-229)	N/A	N/A
Ah_TssA2^{A}-Nt2	(223-387)	2.05 Å	$P2_1$
Ah_TssA2^{A}-Nt2 (Hg)	(223-387)	2.28 Å	$P2_1$
Ah_TssA2^{A}-CTD	(381-478)	N/A	N/A
Ah_TssA2^{A}-Nt2-CTD	(223-478)	3.75 Å	$P2_1$

3.1.3 Contribution

M.S.T. conceived the investigation; S.R.D., R.S., M.S.T. and D.W.R. designed the experiments; S.R.D., R.S., M.J.H., S.E.S. and S.L.B. performed the experiments; S.R.D., R.S., P.J.B., D.W.R and M.S.T. interpreted the data and S.R.D., M.S.T and D.W.R. wrote the manuscript. Experimentally, R.S. and M.S.T designed, constructed and established and expression protocol for all constructs within the manuscript. S.E.S carried out the initial protein purification. Whereas, I carried out (with the assistance of M.J.H) the repeat protein production, purification and crystallisation of the TssA constructs (Nt2, CTD and Nt2-CTD). In addition, I carried out the X-ray data collection for the TssA crystals (Nt2 and Nt2-CTD) and subsequent data analysis.

research communications

STRUCTURAL BIOLOGY COMMUNICATIONS
ISSN 2053-230X

TssA from *Aeromonas hydrophila*: expression, purification and crystallographic studies

Samuel R. Dix,[a]‡ Ruyue Sun,[b]‡ Matthew J. Harris,[c] Sarah L. Batters,[b] Svetlana E. Sedelnikova,[a] Patrick J. Baker,[a] Mark S. Thomas[b]* and David W. Rice[a]*

[a]Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, England, [b]Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield Medical School, Beech Hill Road, Sheffield S10 2RX, England, and [c]Department of Chemistry, King's College London, Britannia House, London SE1 1DB, England. *Correspondence e-mail: m.s.thomas@sheffield.ac.uk, d.rice@sheffield.ac.uk

Received 30 April 2018
Accepted 18 July 2018

Edited by R. L. Stanfield, The Scripps Research Institute, USA

‡ These authors contributed equally.

Keywords: type VI secretion system; TssA subunit; *Aeromonas hydrophila*.

Supporting information: this article has supporting information at journals.iucr.org/f

TssA is a core subunit of the type VI secretion system, which is a major player in interspecies competition in Gram-negative bacteria. Previous studies on enteroaggregative *Escherichia coli* TssA suggested that it is comprised of three putative domains: a conserved N-terminal domain, a middle domain and a ring-forming C-terminal domain. X-ray studies of the latter two domains have identified their respective structures. Here, the results of the expression and purification of full-length and domain constructs of TssA from *Aeromonas hydrophila* are reported, resulting in diffraction-quality crystals for the middle domain (Nt2) and a construct including the middle and C-terminal domains (Nt2-CTD).

1. Introduction

The type VI secretion system (T6SS) is a multiprotein complex found in Gram-negative bacteria that is involved in the delivery of various effector proteins into target cells. It is thought to be evolutionarily related to the contractile tail assembly of bacteriophages (for example T4) and to R-type pyocins, which all share a similar contraction mechanism (Leiman *et al.*, 2009; Basler *et al.*, 2012; Ge *et al.*, 2015). These nanomachines utilize a platform known as the baseplate, onto which a tube, surrounded by a contractile sheath, is polymerized. The inner tube is capped at the baseplate end with proteins that provide a tapered end, which facilitates the penetration of target cells upon contraction of the sheath against the baseplate (Leiman *et al.*, 2009; Shneider *et al.*, 2013; Ge *et al.*, 2015). The T6SS is comprised of multiple copies of 13 core subunits, and a single PAAR tip protein, that collectively can be categorized into two major components. The first is the membrane complex, consisting of ten copies each of TssJ, TssL and TssM. Together, these form a transmembrane assembly which acts as an 'anchor' for the other component, the 'injection machinery'. In addition, it allows the passage of the contractile tube and the associated effector proteins out of the bacterium (Zheng & Leung, 2007; Boyer *et al.*, 2009; Durand *et al.*, 2015; Cianfanelli *et al.*, 2016). The injection machinery consists of two subcomplexes. One of these contains the tube, which is comprised of stacked hexamers of TssD (Hcp) and is capped by a trimer of TssI (VgrG) and a single PAAR subunit. The tube is surrounded by polymerized heterodimers of TssBC, arranged in a six-start helix, that form the contractile sheath (Basler *et al.*, 2012; Shneider *et al.*, 2013; Brunet *et al.*,

2014; Kube *et al.*, 2014; Wang *et al.*, 2017). The tube and sheath are assembled on the second subcomplex, the baseplate, which contains TssE, TssF, TssG and TssK. During contraction of the sheath against the baseplate, the inner tube and associated effector proteins are driven through an opening in the centre of the baseplate (Pukatzki *et al.*, 2007; Brunet *et al.*, 2015; Taylor *et al.*, 2016). Post-construction, the sheath is then recycled by the AAA+ ATPase TssH (ClpV; Bönemann *et al.*, 2009; Basler *et al.*, 2012).

Until recently, very little was known regarding the role and location of the TssA subunit within the T6SS complex. It has been shown that all TssAs comprise a conserved N-terminal region containing a domain of unknown function termed ImpA_N (Pfam PF0681224; Finn *et al.*, 2016) in addition to a divergent C-terminal region (Shalom *et al.*, 2007; Zoued *et al.*, 2017). Recent fluorescence microscopy and cryo-EM structural studies on enteroaggregative *Escherichia coli* (EAEC) and *Vibrio cholerae* T6SS, respectively, have shown that TssA is located and maintained at the baseplate distal end of the growing TssBCD complex during tube and sheath polymerization (Zoued *et al.*, 2016; Nazarov *et al.*, 2018). Furthermore, proteolytic and structural studies of domains from EAEC TssA (Ec042_4540; Ec-TssA) have identified a putative three-domain organization corresponding to an N-terminal domain, a middle domain and a C-terminal domain, and have provided X-ray structures for the latter two. This revealed that the C-terminal domain is organized into a six-pointed star consisting of 12 subunits with D_6 symmetry (Zoued *et al.*, 2016).

We have observed that some TssA homologues are ∼40 amino acids shorter at the C-terminus compared with Ec-TssA. This prompted an investigation into one such TssA subunit, AHA1844, that is present in *Aeromonas hydrophila* ATCC 7966. In this paper, we report the generation of suitable constructs for the crystallization of *A. hydrophila* TssA (Ah-TssA) and its constituent domains (N-terminal domain, Nt1; middle domain, Nt2; C-terminal domain, CTD). Furthermore, we present the first crystallization trials and X-ray diffraction data for proteins derived from constructs of both the Ah-TssA Nt2 and Nt2-CTD domains as initial steps to determining the structures of these domains and their relationship to EAEC TssA.

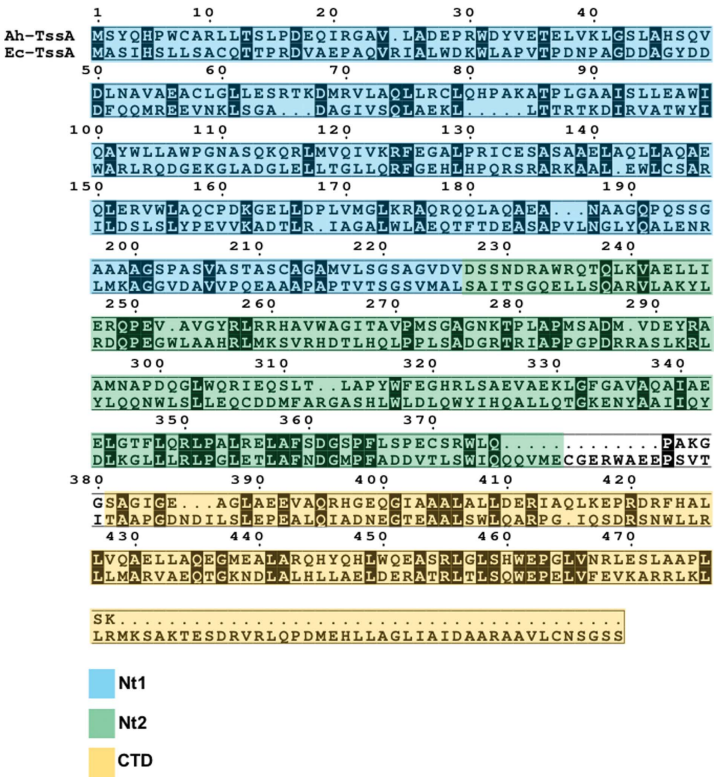


Figure 1
Putative domain boundaries of Ah-TssA and Ec-TssA. Proteolytic and structural studies of Ec-TssA identified putative domain boundaries for Nt1, Nt2 and CTD (Zoued *et al.*, 2016). This information was used along with a sequence alignment between Ah-TssA and Ec-TssA to design suitable Ah-TssA domain constructs for crystallization. The key indicates the colour assigned to each putative domain.

Table 1
Primers and construct design.

Bold sequences indicate restriction sites.

Construct	His6.Ah-TssA	His6.Ah-TssA Nt1	His6.Ah-TssA Nt2	Ah-TssA CTD	His6.Ah-TssA Nt2-CTD
Forward primer	GCGC**AGATCT**AATGAGCT ATCAACACCCCTG	GCGC**AGATCT**AATGAGCT ATCAACACCCCTG	GCGC**GGATCC**AGGCGTCG ACGTCGACAGTTC	GCGC**GGATCC**GAGTGCGG GCATTGGCGAGGC	GCGC**GGATCC**AGGCGTCG ACGTCGACAGTTC
Reverse primer	GCGC**AAGCTT**TCATTTCG ACAACGGCGCCG	GCGC**GGTACC**TTAGCTGG AACTGTCGACGTCGACG	GCGC**AAGCTT**TTAAGCCT CGCCAATGCCCGCACT	GCGC**AAGCTT**TCATTTCG ACAACGGCGCCG	GCGC**AAGCTT**TCATTTCG ACAACGGCGCCG
Cloning/expression vector	pACYCDuet-1	pACYCDuet-1	pACYCDuet-1	pACYCDuet-1	pACYCDuet-1
Restriction sites	BamHI and HindIII	BamHI and KpnI	BamHI and HindIII	BamHI and HindIII	BamHI and HindIII
Expression host	*E. coli* BL21(DE3)	*E. coli* BL21(DE3)	*E. coli* BL21(DE3)	*E. coli* BL21(DE3)	*E. coli* BL21(DE3)
Amino acids	His_6.(1–478) (52.6 kDa)	His_6.(1–229) (26.0 kDa)	His_6.(223–387) (19.7 kDa)	381–478 (10.7 kDa)	His_6.(223–478) (29.8 kDa)

2. Materials and methods

2.1. Macromolecule production

Domain boundaries for Nt1, Nt2 and CTD were defined using an amino-acid sequence alignment of Ec-TssA and Ah-TssA (Fig. 1). DNAs encoding Ah-TssA or Ah-TssA domains (UniProt A0KJC7) were amplified from *A. hydrophila* ATCC 7966 (Seshadri *et al.*, 2006) by the polymerase chain reaction (PCR) utilizing the primer pairs described in Table 1. The products were then ligated into the pACYCDuet-1 expression vector (Novagen) downstream of the first T7 promoter using compatible restriction sites. The recombinant expression plasmids containing *tssa* DNA were transformed into *E. coli* strain JM83 (Yanisch-Perron *et al.*, 1985), verified by nucleotide sequencing and then transformed into *E. coli* BL21(DE3) cells (Novagen; Studier & Moffatt, 1986) for protein overproduction (Hanahan, 1983).

Ah-TssA constructs representing the full-length protein, the three identified domains (Nt1, Nt2 and CTD) and a combination of two of them (Nt2-CTD) were grown in 2 l BHI broth (Oxoid) at 37°C with 25 µg ml^{-1} chloramphenicol. Once the cultures had reached an OD_{600} of 0.5–0.7, IPTG was added to a final concentration of 1 m*M* (0.1 m*M* for His6.Ah-TssA Nt2) for the induction of the T7 promoter present within each expression vector (Table 1). Cultures were incubated for a further 2–3 h at 37°C (30°C was used to induce the synthesis of His6.Ah-TssA, His6.Ah-TssA Nt1 and His6.Ah-TssA Nt2) before the cells were harvested. The harvested cells were resuspended in 50 m*M* Tris–HCl pH 8.0 containing 2 m*M* EDTA, 200 m*M* NaCl, 10% glycerol (5 ml per gram of cell paste), treated with lysozyme (0.2 mg ml^{-1}) for 30 min at 4°C followed by sodium deoxycholate (0.5 mg ml^{-1}) and PMSF (25 µg ml^{-1}) for a further 30 min, and lysed by sonication [EDTA was omitted and imidazole (10 m*M* final concentration) was included if the protein was to be purified by nickel-affinity chromatography, and the lysozyme and sodium deoxycholate steps were omitted if the cell paste was frozen before the sonication step]. The lysate was then cleared by centrifugation for 30 min at 35 000*g*.

His-tagged proteins (*i.e.* all constructs except Ah-TssA CTD) were applied onto a 5 ml HisTrap HP 5 ml column (GE Healthcare Life Sciences) and eluted with a linear gradient of imidazole (10–500 m*M*). His6.Ah-TssA was further purified by SEC (Superose 6 10/300 GL, GE Healthcare Life Sciences) in 50 m*M* Tris–HCl pH 8.0 containing 500 m*M* NaCl. The Ah-TssA CTD lysate was first precipitated with 4 *M* ammonium sulfate and the pellet was subsequently resuspended in 50 m*M* Tris–HCl pH 8.0 and applied onto an anion-exchange column (Resource Q 6 ml, GE Healthcare Life Sciences) before being eluted with a linear gradient of NaCl (0–500 m*M*). Ah-TssA CTD was further purified by SEC (HiLoad Superdex 200 16/600, GE Healthcare Life Sciences) in 50 m*M* Tris–HCl pH 8.0 containing 500 m*M* NaCl. Protein purity was analysed by SDS–PAGE and subsequent staining.

2.2. Crystallization

Crystallization trials were performed on all purified Ah-TssA constructs using the protein concentration and protein buffer compositions shown in Table 2. Initial screening was carried out using a Matrix Hydra II Plus One crystallization robot dispensing into 96-well MRC2 sitting-drop crystallization trays, whereby a 1:1 ratio of precipitant:protein was created generating 400 nl drops, which were allowed to equilibrate through vapour diffusion at 290 K (280 K for Ah-TssA Nt2-CTD). Commercially available sparse-matrix screens (pH Clear, Morpheus, PACT, JCSG+, MPD and ProPlex) were used to identify conditions that formed crystals. Successful crystallization conditions for each construct, if applicable, can be found in Table 2. Optimization of conditions was carried out when required in 24-well trays with a reservoir volume of 500 µl and a drop size of 2 µl. A mercury derivative of the His6.Ah-TssA Nt2 crystals was generated by the addition of powdered ethylmercury phosphate (EMP; ∼20 µg) to a drop containing crystals to provide a derivative.

2.3. Data collection and processing

Crystals were harvested in cryoloops (Hampton Research) and immediately soaked in a solution consisting of mother liquor with an additional 25%(*v*/*v*) ethylene glycol [supplemented with an additional 10%(*v*/*v*) glycerol for crystals of His6.Ah-TssA Nt2] for about 10 s before being flash-cooled to 100 K in liquid nitrogen. Native data were collected for crystals of His6.Ah-TssA Nt2 and His6.Ah-TssA CTD-Nt2 on MX beamlines I03 and I04 at Diamond Light Source, Oxford, respectively. In addition, data for a mercury derivative of crystals of His6.Ah-TssA Nt2 were collected on beamline I03. Diffraction data were collected as shown in Table 3 and the

Table 2
Crystallization.

N/A: constructs that did not crystallize.

Construct	His6.Ah-TssA	His6.Ah-TssA Nt1	His6.Ah-TssA Nt2	Ah-TssA CTD	His6.Ah-TssA Nt2-CTD
Method	Vapour diffusion	Vapour diffusion	Vapour diffusion	Vapour diffusion	Vapour diffusion
Plate type	96-well MRC2	96-well MRC2	24-well optimization tray	96-well MRC2	24-well optimization tray
Temperature (K)	290	290	290	290	280
Protein concentration (mg ml^{-1})	10	10	10	18	11
Buffer composition of protein solution	50 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl	50 m*M* Tris–HCl pH 8.0, 50 m*M* NaCl	10 m*M* Tris–HCl pH 8.0	10 m*M* Tris–HCl pH 8.0	10 m*M* Tris–HCl pH 8.0
Composition of reservoir solution	N/A	N/A	0.04 *M* potassium phosphate monobasic, 16%(*w*/*v*) PEG 8000, 20%(*v*/*v*) glycerol	N/A	0.2 *M* sodium acetate, 0.1 *M* sodium citrate pH 5.5, 10%(*w*/*v*) PEG 4000

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

Construct	His6.Ah-TssA Nt2	His6.Ah-TssA Nt2	His6.Ah-TssA Nt2-CTD
Crystal	Native	EMP derivative	Native
Diffraction source	I03	I03	I04
Wavelength (Å)	0.9763	1.0088	0.9795
Temperature (K)	100	100	100
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Rotation range per image (°)	0.1	0.1	0.1
Total rotation range (°)	200	360	200
Exposure time per image (s)	0.05	0.05	0.1
Data-processing package	*xia*2 -3dii	*xia*2 -3dii	*DIALS*
Space group	$P2_1$	$P2_1$	$P2_1$
a, *b*, *c* (Å)	46.0, 40.1, 101.0	46.1, 40.1, 100.8	73.0, 202.7, 137.7
α, β, γ (°)	90.0, 103.2, 90.0	90, 102.7, 90	90.0, 92.3, 90.0
Cell volume (Å^3)	181631	182001	2036168
Resolution range (Å)	2.05 (2.10–2.05)	2.28 (2.34–2.28)	3.75 (3.85–3.75)
R_{meas}	0.142 (1.315)	0.176 (1.488)	0.117 (0.950)
$CC_{1/2}$	0.995 (0.574)	0.996 (0.532)	0.862 (0.757)
$\langle I/\sigma(I)\rangle$	6.2 (1.2)	7.3 (1.0)	6.8 (1.5)
Completeness (%)	98.8 (99.2)	99.3 (95.1)	100.0 (100.0)
Multiplicity	3.6 (3.7)	6.4 (5.0)	3.7 (3.7)
Total reflections	81939 (6224)	107275 (5874)	153004 (11196)
Unique reflections	22646 (1700)	16664 (1183)	41019 (3066)
Overall *B* factor from Wilson plot (Å^2)	31	42	170
Anomalous completeness (%)	—	98.9 (93.2)	—
Anomalous multiplicity	—	3.3 (2.6)	—
Anomalous correlation	—	0.217 (−0.022)	—
Anomalous slope	—	1.105	—

data were processed utilizing the *xia*2 software pipeline (Winter, 2010; Winter & McAuley, 2011) with/without running the optional *XDS* package (Kabsch, 2010) as indicated. All details of data processing are shown in Table 3. The estimated content of the asymmetric unit and the self-rotation function were determined using the *CCP*4 suite (Winn *et al.*, 2011).

3. Results and discussion

The full-length TssA construct, His6.Ah-TssA, and the four different domain constructs were all successfully over-produced and purified as described. The final protein yields after purification were 14 mg for His6.Ah-TssA, 5.6 mg for His6.Ah-TssA Nt1, 2.7 mg for His6.Ah-TssA Nt2, 7.2 mg for Ah-TssA CTD and 16 mg for His6Ah-TssA Nt2-CTD. Analysis by gel filtration indicates that His6.Ah-TssA Nt1 is monomeric and that His6.Ah-TssA Nt2 is present as a dimer, whereas gel filtration of His6.Ah-TssA, Ah-TssA CTD and His6.Ah-TssA Nt2-CTD indicated approximate molecular weights of 0.6, 0.2 and 0.4 MDa, respectively, indicating the presence of larger complexes.

The full-length protein and the four domain constructs were subjected to crystallization trials, and two (His6.Ah-TssA Nt2 and His6.Ah-TssA Nt2-CTD) produced crystals suitable for data collection at Diamond Light Source, Oxford (Fig. 2). SDS–PAGE analysis of the two purified protein constructs (His6.Ah-TssA Nt2 and His6.Ah-TssA Nt2-CTD) indicated initial purities of approximately >95% (Supplementary Fig. S1). The crystals diffracted to 2.05 and 3.75 Å resolution, respectively. The asymmetric unit of the His6.Ah-TssA Nt2 crystals is estimated to contain two molecules, with a probable Matthews coefficient of 2.53 Å^3 Da^{-1}. Calculating the probable Matthews coefficient for the data from the His6.Ah-TssA Nt2-CTD crystals suggested an asymmetric unit containing

(*a*) (*b*)

Figure 2
Crystals of Ah-TssA domains. (*a*) His6.Ah-TssA Nt2 crystals grown in 0.04 *M* potassium phosphate monobasic, 16%(*w*/*v*) PEG 8000, 20%(*v*/*v*) glycerol. (*b*) His6.Ah-TssA Nt2-CTD crystals grown optimally in 0.2 *M* sodium acetate, 0.1 *M* sodium citrate pH 5.5, 10%(*w*/*v*) PEG 4000.

between 10 and 18 subunits. Analysis of a self-rotation function based on all data for 4–16 Å resolution with a range of integration radii gave no indication as to the symmetry of His6.Ah-TssA Nt2-CTD within the asymmetric unit. Preliminary data to 2.3 Å resolution from the mercury-soaked His6.Ah-TssA Nt2 crystals provided an initial map to which the sequence could be fitted, indicating that the crystals were of the Nt2 domain (Supplementary Fig. S2). Refinement of the structure and attempts to collect data to higher resolution are currently ongoing.

Acknowledgements

We would like to thank the I03 and I04 beamline staff at Diamond Light Source, Oxford for their assistance during data-collection.

Funding information

This work was supported by a University of Sheffield scholarship awarded to SRD.

References

Basler, M., Pilhofer, M., Henderson, G. P., Jensen, G. J. & Mekalanos, J. J. (2012). *Nature (London)*, **483**, 182–186.
Bönemann, G., Pietrosiuk, A., Diemand, A., Zentgraf, H. & Mogk, A. (2009). *EMBO J.* **28**, 315–325.
Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. & Attree, I. (2009). *BMC Genomics*, **10**, 1–14.
Brunet, Y. R., Hénin, J., Celia, H. & Cascales, E. (2014). *EMBO Rep.* **15**, 315–321.
Brunet, Y. R., Zoued, A., Boyer, F., Douzi, B. & Cascales, E. (2015). *PLoS Genet.* **11**, 1–21.
Cianfanelli, F. R., Alcoforado Diniz, J., Guo, M., De Cesare, V., Trost, M. & Coulthurst, S. J. (2016). *PLoS Pathog.* **12**, e1005735.
Durand, E., Nguyen, V. S., Zoued, A., Logger, L., Péhau-Arnaudet, G., Aschtgen, M. S., Spinelli, S., Desmyter, A., Bardiaux, B., Dujeancourt, A., Roussel, A., Cambillau, C., Cascales, E. & Fronzes, R. (2015). *Nature (London)*, **523**, 555–560.
Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J. & Bateman, A. (2016). *Nucleic Acids Res.* **44**, D279–D285.
Ge, P., Scholl, D., Leiman, P. G., Yu, X., Miller, J. F. & Zhou, Z. H. (2015). *Nature Struct. Mol. Biol.* **22**, 377–383.
Hanahan, D. (1983). *J. Mol. Biol.* **166**, 557–580.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kube, S., Kapitein, N., Zimniak, T., Herzog, F., Mogk, A. & Wendler, P. (2014). *Cell Rep.* **8**, 20–30.
Leiman, P. G., Basler, M., Ramagopal, U. A., Bonanno, J. B., Sauder, J. M., Pukatzki, S., Burley, S. K., Almo, S. C. & Mekalanos, J. J. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 4154–4159.
Nazarov, S., Schneider, J. P., Brackmann, M., Goldie, K. N., Stahlberg, H. & Basler, M. (2018). *EMBO J.* **37**, e97103.
Pukatzki, S., Ma, A. T., Revel, A. T., Sturtevant, D. & Mekalanos, J. J. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 15508–15513.
Seshadri, R. *et al.* (2006). *J. Bacteriol.* **188**, 8272–8282.
Shalom, G., Shaw, J. G. & Thomas, M. S. (2007). *Microbiology*, **153**, 2689–2699.
Shneider, M. M., Buth, S. A., Ho, B. T., Basler, M., Mekalanos, J. J. & Leiman, P. G. (2013). *Nature (London)*, **500**, 350–353.
Studier, F. W. & Moffatt, B. A. (1986). *J. Mol. Biol.* **189**, 113–130.
Taylor, N. M. I., Prokhorov, N. S., Guerrero-Ferreira, R. C., Shneider, M. M., Browning, C., Goldie, K. N., Stahlberg, H. & Leiman, P. G. (2016). *Nature (London)*, **533**, 346–352.
Wang, J., Brackmann, M., Castaño-Díez, D., Kudryashev, M., Goldie, K. N., Maier, T., Stahlberg, H. & Basler, M. (2017). *Nature Microbiol.* **2**, 1507–1512.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
Winter, G. & McAuley, K. E. (2011). *Methods*, **55**, 81–93.
Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). *Gene*, **33**, 103–119.
Zheng, J. & Leung, K. Y. (2007). *Mol. Microbiol.* **66**, 1192–1206.
Zoued, A., Durand, E., Brunet, Y. R., Spinelli, S., Douzi, B., Guzzo, M., Flaugnatti, N., Legrand, P., Journet, L., Fronzes, R., Mignot, T., Cambillau, C. & Cascales, E. (2016). *Nature (London)*, **531**, 59–63.
Zoued, A., Durand, E., Santin, Y. G., Journet, L., Roussel, A., Cambillau, C. & Cascales, E. (2017). *Bioessays*, **39**, 1600262.

supporting information

Supplementary Figure 1. SDS-PAGE analysis of protein purification for *A. hydrophila* TssA His_6.Ah-TssA Nt2 and His_6.Ah-TssA Nt2-CTD constructs. (*a*) Purification analysis of the His_6.Ah-TssA Nt2 construct. (*b*) Purification analysis of the His_6.Ah-TssA Nt2-CTD construct. Lane 1 - Mark12 ladder, Lane 2 - Cell debris, Lane 3 - Cell free extract, Lane 4 - Unbound fraction, Lane 5 - Final preparation of construct. Gels indicate successful purification of their respective constructs, protein of interest indicated by arrow, protein appears &

3.2 Paper II

3.2.1 Summary

The second paper in this thesis discusses the bioinformatic, interaction and structural studies carried out on TssA. The work has led to the identification of two major clades of TssA (1 and 2), which can both be divided into sub-clades (A and B). In addition, the first interaction study for a TssA1^{B} sub-clade member with other T6SS subunits is presented, indicating that all TssA's, independent of clade confer a similar pattern of interaction with other T6SS components. The paper also describes the first high-resolution structures for the constituent domains of a TssA1 representative (*B. cenocepacia*), presenting the structure of the Nt1 domain, which contains the highly conserved ImpA_N region, in addition, the CTD oligomer which consists of 32 subunits displaying D_{16} symmetry. Furthermore, the paper discusses the structure determination of an Nt2-CTD construct from a TssA2^{A} representative (*A. hydrophila*), illustrating that the molecule consists of 10 subunits exhibiting D_5 symmetry. It further discusses the inherent domain mobility TssA confers as a result of the presence of interdomain linkers and the possible impact this has on the biological function of the molecule. These studies, taken together, identify that despite the radically different symmetries displayed by the various TssA sub-clades, they all mediate similar patterns of interactions with other T6SS subunits indicating a common function.

3.2.2 Protein summary

TABLE 3.2: Summary of TssA proteins discussed

Paper II				
Construct ID	**Residue range**	**Resolution**	**Space group**	**PDB ID**
Bc_TssA1^{B}-Nt1	(1-255)	2.04 Å	$P2_12_12$	6HS5
Bc_TssA1^{B}-CTD	(303-373)	3.03 Å	I222	6HS6
Bc_TssA1^{B}-CTD	(303-358)	2.35 Å	$P6_2$	6H8E
Bc_TssA1^{B}-CTD	(303-347)	1.87 Å	I222	6H8F
Ah_TssA2^{A}-Nt1-Nt2	(1-374) (231-376)	1.76 Å	$P2_1$	6G7B
Ah_TssA2^{A}-Nt2-CTD	(223-478)	3.13 Å	$P2_1$	6G7C
Ec_TssA2^{B}-Nt2	(221-377)	3.37 Å	$P2_12_12$	4YO3
Ec_TssA2^{B}-CTD	(399-529)	3.35 Å	$P2_1$	4YO5

3.2.3 Contribution

M.S.T. conceived the investigation. S.R.D., H.J.O., R.S., A.A., S.S., H.L.S., S.B.T. and M.S.T. designed the experiments; S.R.D., H.J.O., R.S., A.A., S.S., M.J.H., S.L.B., H.L.S., D.J.M., S.B.T. and S.E.S. performed the experiments; S.R.D., H.J.O., R.S., A.A., S.S., H.L.S., S.B.T., P.J.B., P.A.B., D.W.R and M.S.T. interpreted the data and S.R.D., H.J.O., R.S., H.L.S., D.W.R. and M.S.T. wrote the manuscript. Experimentally, R.S., A.A., S.S., H.L.S., S.B.T, D.J.M, and M.S.T carried out the bioinformatic analysis to identify the possible TssA sub-clades, the two-hybrid analysis, T6SS inactivation assay and negative stain EM. S.E.S carried out the initial purification of the TssA components for structural studies. H.J.O carried out the crystalisation, data collection and structural analysis of the *B. cenocepacia* TssA1^{B} constructs. I carried out the design, expression, purification and crystallisation of the TssA Nt1-Nt2 construct. In addition, I carried out the data collection, X-ray data analysis, structure determination and structural analysis for all *A. hydrophila* TssA2^{A} constructs.

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-07247-1 OPEN

Structural insights into the function of type VI secretion system TssA subunits

Samuel R. Dix[1], Hayley J. Owen[1], Ruyue Sun[2], Asma Ahmad[2], Sravanthi Shastri[2], Helena L. Spiewak[2,4], Daniel J. Mosby[2], Matthew J. Harris[1,3], Sarah L. Batters[2], Thomas A. Brooker[2], Svetomir B. Tzokov[1], Svetlana E. Sedelnikova[1], Patrick J. Baker[1], Per A. Bullough[1], David W. Rice[1] & Mark S. Thomas[2]

The type VI secretion system (T6SS) is a multi-protein complex that injects bacterial effector proteins into target cells. It is composed of a cell membrane complex anchored to a contractile bacteriophage tail-like apparatus consisting of a sharpened tube that is ejected by the contraction of a sheath against a baseplate. We present structural and biochemical studies on TssA subunits from two different T6SSs that reveal radically different quaternary structures in comparison to the dodecameric *E. coli* TssA that arise from differences in their C-terminal sequences. Despite this, the different TssAs retain equivalent interactions with other components of the complex and position their highly conserved N-terminal ImpA_N domain at the same radius from the centre of the sheath as a result of their distinct domain architectures, which includes additional spacer domains and highly mobile interdomain linkers. Together, these variations allow these distinct TssAs to perform a similar function in the complex.

[1] Department of Molecular Biology and Biotechnology, Krebs Institute, University of Sheffield, Sheffield S10 2TN, UK. [2] Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield Medical School, Beech Hill Road, Sheffield S10 2RX, UK. [3] Present address: Department of Chemistry, King's College London, Britannia House, London SE1 1DB, UK. [4] Present address: Northern Genetics Service, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Institute of Genetic Medicine, International Centre for Life, Newcastle upon Tyne NE1 3BZ, UK. These authors contributed equally: Samuel R. Dix, Hayley J. Owen, Ruyue Sun, Asma Ahmad. Correspondence and requests for materials should be addressed to D.W.R. (email: d.rice@sheffield.ac.uk) or to M.S.T. (email: m.s.thomas@sheffield.ac.uk)

Contractile bacteriophages of the family *Myoviridae* (i.e. T4), R-type pyocins and the type VI secretion system (T6SS) of Gram-negative bacteria are evolutionarily related nano-scale injection machines that puncture target cell membranes using a shared contraction mechanism[1–3]. These injection devices are comprised of an inner tube, surrounded by a contractile sheath, that are both assembled on a platform known as the baseplate. The inner tube is sharpened with spike proteins at the baseplate proximal end, which facilitates its penetration of target cells upon contraction of the sheath against the baseplate[2–5].

The T6SS secretion machinery is formed from multiple copies of 12 core subunits (TssA-TssG, TssI-TssM) and a single PAAR tip protein[6–9] and can be subdivided into two main components. One of these, the membrane complex, consists of 10 subunits each of TssJ, TssL, and TssM that assemble into a chamber-like structure with five-fold symmetry which serves to anchor the injection machinery at the cell envelope as well as providing an exit channel for translocated subunits and effectors[10–15]. The other component, the injection machinery, consists of two sub-complexes. One sub-complex consists of the inner tube, which is comprised of stacked hexameric rings of TssD (Hcp), capped by the trimeric hub protein, TssI (VgrG), and sharpened by the PAAR subunit, surrounded by repeating TssBC heterodimers that form the contractile sheath[1,3,5,16,17]. The latter consists of a six-start helix that possesses six-fold symmetry, giving a cogwheel-like appearance when viewed end-on[1,18–21]. Both the inner tube and sheath exhibit the same degree of helical twist thereby ensuring a six-fold symmetry match along the entire length of the tube-sheath complex[21]. The other sub-complex is the baseplate, which consists of TssE, TssF, TssG and TssK, and contains a central channel through which the sharpened inner tube passes upon contraction of the sheath[3,17,22–24]. The sheath is subsequently recycled by the AAA+ ATPase, TssH (ClpV)[1,18,25].

Until recently, relatively little was known about the location and role of the TssA subunit within the T6SS complex. TssA subunits are enigmatic as they possess a conserved N-terminal region of unknown function, previously identified as ImpA_N (PFAM: PF06812[26]), whereas sequences located C-terminal to this region are highly divergent[6,27,28]. Consistent with this, phylogenetic analysis has suggested that the TssA family can be subdivided into three clades (TssA1, TssA2 and TssA3)[28]. The C-terminal regions of TssA1 and TssA2 have been shown to be required for assembly of these TssA subunits into higher order oligomers and both subunits are required for T6SS function[27,28]. However, the TssA3 subunit has not been previously investigated.

Recent studies on the TssA2 subunit of enteroaggregative *Escherichia coli* (EAEC), Ec042_4540, have provided structures for two of its putative three domains (the middle (Nt2) and the C-terminal domain (CTD)), leaving the structure of the highly conserved N-terminal domain (Nt1), yet to be determined. These structural studies showed that the CTD assembles into a dodecamer that resembles a six-pointed star. Further analysis showed that TssA2 interacts with components of the baseplate, inner tube, sheath and the T6SS membrane complex[27]. This led to the proposal of a capping model whereby TssA2 initially interacts with the core TssJLM membrane complex, thereby triggering baseplate recruitment. According to the model, TssA2 subsequently serves to coordinate the assembly of the inner tube and contractile sheath, during which it migrates away from the baseplate complex, remaining in contact with the distal end of the polymerising tube[27,29]. In a separate study, the TssA1 subunit of *Pseudomonas aeruginosa*, PA0082, was shown to form a ring-like complex of indeterminate stoichiometry and symmetry, and in contrast to TssA2, was proposed to serve as the T6SS counterpart of the phage T4 baseplate protein gp6, in a baseplate-associated model of TssA function[28]. However, in the absence of a high-resolution structure for TssA1 or TssA3, any relationship between their CTDs and the Nt2 and CTDs of TssA2 subunits are, as yet, unclear.

In this paper, we present evidence that there are only two main TssA clades (TssA1 and TssA2) and that each clade can be subdivided into two distinct sub-clades, referred to here as A and B. We provide the first structure for any TssA1 subunit (TssA1^{B}) to show that it is composed of two domains instead of three: the conserved N-terminal domain (Nt1), containing the ImpA_N region, that is tethered by a long linker to a distinct CTD that assembles into a ring containing 32 subunits with 16-fold symmetry. Furthermore, we present the first T6SS subunit interaction analysis of a TssA1^{B} member. We also provide the structure of an Nt2-CTD TssA2^{A} construct, demonstrating that 10 subunits assemble into a complex with distinct five-fold symmetry rather than the six-fold symmetry seen in EAEC TssA2^{B}. The structure further reveals a striking range of conformational mobility of the TssA2^{A} Nt2 domain with respect to the CTD oligomer. Despite the differences in symmetry of TssAs belonging to the different clades, the similar subunit interaction network and disposition of their conserved domains suggests that they function by a related mechanism to coordinate inner tube/sheath assembly.

Results

Bioinformatic and proteolytic analysis of TssA family proteins. Bioinformatic analysis suggests that all TssA subunits contain an N-terminal conserved region of ~120 amino acids (ImpA_N) composed of three shorter motifs of significant sequence similarity (ImpA_N1–3) (Fig. 1a and Supplementary Fig. 1). However, these proteins possess one of two distinct C-terminal regions based on amino acid sequence conservation, and are referred to here as TssA1 and TssA2 (Fig. 1a and Supplementary Fig. 2). In members of the TssA1 clade, ImpA_N forms part of a large N-terminal domain (Nt1) of ~250 amino acids which is connected by a central region of variable length (40–60 amino acids) and sequence to a conserved C-terminal region of 60–75 amino acids containing a signature EPxxP motif (Supplementary Fig. 2a). Phylogenetic analysis shows that TssA1 subunits fall into one or other of two sub-clades, referred to here as A and B (Fig. 1b). In TssA1^{A} members, such as *P. aeruginosa* PA0082 (Pa TssA1^{A}), the EPxxP motif is EPSxP, whereas in TssA1^{B} members, such as *B. cenocepacia* I35_RS01755 (Bc TssA1^{B}), an alternative EP(H/Q)SP motif is present (Supplementary Fig. 2a). TssA1^{A} corresponds to TssA1 in the previously proposed nomenclature[28], whereas TssA1^{B} corresponds to TssA3. In comparison, TssA2 orthologues are longer (480–540 amino acids) and are predicted to contain three domains: Nt1 (including the ImpA_N region), a middle domain (Nt2) and CTD, each separated by regions of variable length and sequence (Fig. 1a and Supplementary Fig. 1 and 2b). However, our phylogenetic analysis reveals that TssA2 proteins are also subdivided into two sub-clades (Fig. 1b) with members of sub-clade B, represented by the EAEC TssA2 subunit (EAEC TssA2^{B}) having an extension of 20–40 amino acids at the C-terminus compared to sub-clade A members, such as *A. hydrophila* AHA1844 (Ah TssA2^{A}).

To provide support for the bioinformatics, analysis of protein fragments derived by either limited or adventitious proteolysis was used to identify putative domain boundaries. Analysis of Bc His_6.TssA1^{B} degradation products suggested that TssA1^{B} possesses a large folded Nt1 domain (~31 kDa) connected to a small CTD (~8.5 kDa) by an interdomain linker of ~40 amino acids (Supplementary Fig. 3a, c, and d, Supplementary Table 1).

Products of limited proteolysis carried out on purified Ah His_6.TssA2^{A} showed that residues in regions corresponding to positions R178-G190 and K378-E399, two regions of high

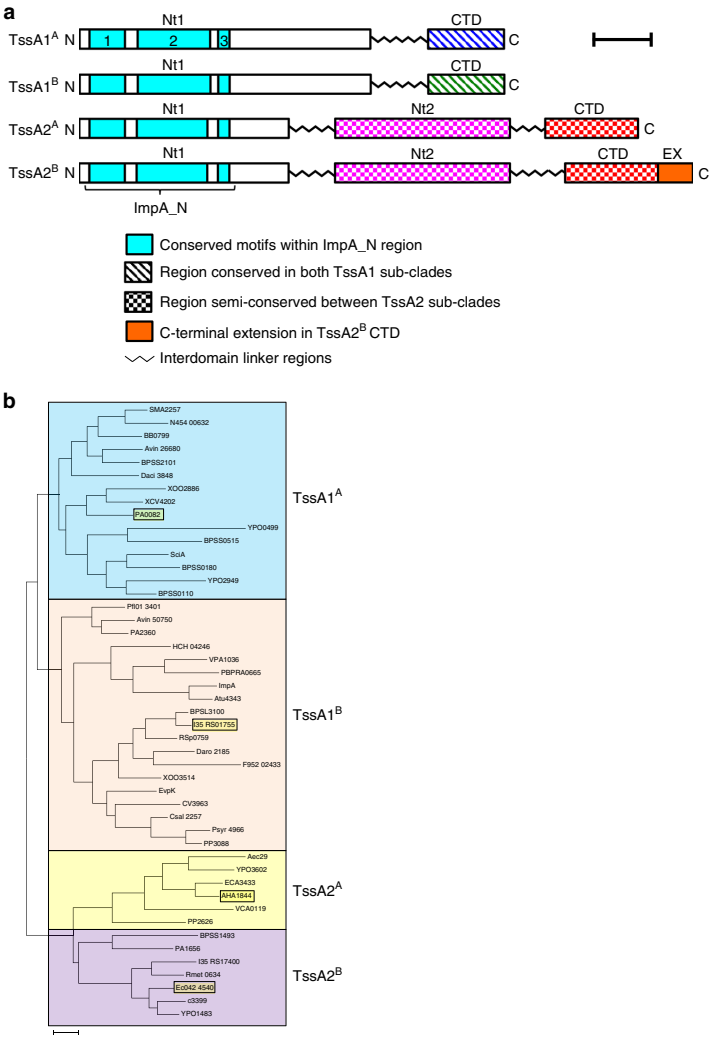


Fig. 1 Phylogeny and domain organisation of TssA family proteins. **a** Proposed domain organisation of TssA1^{A}, TssA1^{B}, TssA2^{A} and TssA2^{B} subunits based on structural studies, secondary structure prediction and amino acid sequence alignment. The interdomain linkers comprise regions of variable length and amino acid sequence. The N-terminal conserved region, ImpA_N (comprised of three major regions of conservation, ImpA_N1-3), Nt1, Nt2, CTD and EX (C-terminal extension present in TssA2^{B} proteins) are indicated. Scale bar corresponds to 50 amino acids. **b** Maximum-likelihood phylogenetic tree from amino acid sequences of 47 TssA family members, showing the subdivision into three major clades as observed previously [28]but with a revised nomenclature. The TssA2 clade is further divided into sub-clades A and B, based on patterns of sequence identity. TssA members discussed in detail in this study are highlighted in dark yellow. Scale bar represents 0.2 substitutions per site

sequence variability, were particularly susceptible to protease cleavage. This is consistent with the predicted three-domain architecture with approximate boundaries of M1-R178 for Nt1, D231-L374 for Nt2, and I402-K478 for the CTD (Supplementary Fig. 3b and e).

Interactions of TssA1^{B} and its domains with other T6SS components. The interactions of a TssA1^{B} subunit with other components of the T6SS have not been previously investigated. To address this, a two-hybrid analysis was performed. Complementation of CyaA function was assayed on maltose MacConkey indicator plates and by performing β-galactosidase assays, both of which indicated that TssA1^{B} may interact with TssC, Hcp, TssE, TssF, the conserved core region of VgrG (VgrG$_{C}$), TssK, TssL and the N-terminal cytoplasmic region of TssM (TssM$_{N}$) (Fig. 2a, b). To validate these results, co-IP experiments were performed using FLAG-tagged TssA1^{B} in pairwise combinations with other epitope-tagged subunits. The results supported the in vivo observations. Thus, epitope-tagged TssC, Hcp, TssF, VgrG$_{C}$, TssK, TssL and

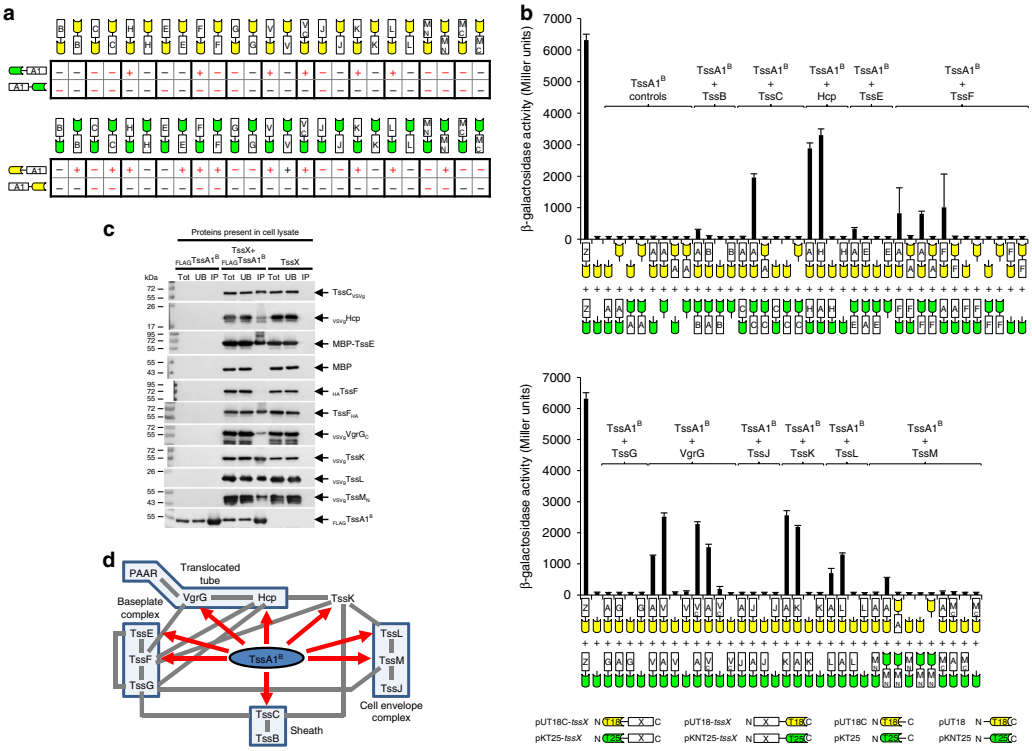


Fig. 2 Interaction of TssA1^B with other T6SS subunits. **a** Two-hybrid analysis (maltose phenotypes). Hybrid proteins are represented by a green or yellow coloured motif representing the CyaA fragment (T25 and T18, respectively) linked to a white rectangle labelled according to the fused T6SS subunit, as shown in the key at the bottom of (**b**). T6SS subunits are indicated by a single letter corresponding to the suffix used in the Tss nomenclature (i.e. A1 corresponds to TssA1^B, B corresponds to TssB etc) except for H (Hcp), V (VgrG) and V_C (VgrG core region). M_N and M_C represent the N-terminal cytoplasmic and C-terminal periplasmic regions of TssM, respectively. The efficiency of complementation of representative combinations (phenotypes shown in red font) were determined by β-galactosidase assay. **b** Two-hybrid analysis (β-galactosidase activity). Data is representative of three independent experiments ($n = 3$) performed in duplicate and values correspond to the mean ± standard deviation. Nomenclature as in (**a**). Z represents the Zip control. Values are presented in Supplementary Data 1. **c** Co-immunoprecipitation analysis. FLAG-tagged Bc TssA1^B and potential interacting T6SS subunits (TssX) tagged with the VSV-g or HA epitope tags, or with MBP, were co-expressed in *E. coli*. FLAG.TssA1^B was immunoprecipitated from cell lysates and recovered prey proteins (IP) were detected with the appropriate epitope antibody by western blotting. MBP was included as a control and was detected with MBP pAb. Proteins present in the cell lysate (Tot) and the unbound material (UB) were also analysed. Blots were also probed with FLAG mAb (an example of such a blot following co-expression of FLAG.TssA1^B and VSVg.TssM$_N$ is shown in the bottom panel). Uncropped images of the blots are shown in Supplementary Fig. 4. **d** Bc TssA1^B interaction network. Interactions between Bc TssA1^B and other T6SS subunits are indicated by red arrows. Previously reported interactions are indicated by grey connectors. Note that VgrG and TssK have been shown to interact with the TssF-TssG complex[22] but for simplicity are linked to TssF in this figure

TssM$_N$ were all co-immunoprecipitated with FLAG.TssA1^B (Fig. 2c). Demonstration of an interaction between TssA1^B and TssE required solubilisation of TssE by fusion to the MBP solubility tag. In summary, the protein-protein interaction analysis suggests that TssA1^B makes contact with components of the baseplate, the membrane-anchoring complex, the inner tube and its cap, and the surrounding sheath (Fig. 2d). Two-hybrid analysis was then performed to investigate the role of the N- and C-terminal regions of TssA1^B in interactions with other T6SS subunits. This showed that TssA1^B CTD interacts with Hcp, TssF and VgrG (and also with VgrG$_C$) but interactions with other subunits were not observed for the Nt1 domain (Supplementary Fig. 6).

TssA1^B is an essential component of the T6SS. P.a TssA1^A, *V. cholerae* TssA2^A (VCA0119) and EAEC TssA2^B have been shown to be essential for T6SS activity[27,28,30]. Therefore, the requirement for TssA1^B for a fully functional T6SS was explored by assaying the ability of a *B. cenocepacia tssA1^B* mutant to secrete Hcp—an indicator of T6SS activity. The results showed that, unlike the wild-type strain, Hcp was absent from the culture supernatant of the *tssA1^B* mutant, but it was secreted when *tssA1^B* mutant bacteria contained a plasmid expressing *tssA1^B* (Fig. 3), thereby confirming the requirement for this subunit for a functional T6SS[31].

Structure determination of the TssA1^B Nt1 domain. SEC and two-hybrid analysis of the overexpressed Bc TssA1^B Nt1 domain indicated that it is monomeric (Supplementary Fig. 7a, b). Analysis of the crystal structure of the TssA1^B Nt1 domain, determined to 1.80 Å (Supplementary Table 2), revealed a monomer composed of 11 α-helices that are organised into two subdomains, an N-terminal subdomain 1 (Sd1) connected to a larger

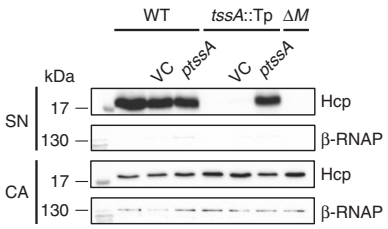


Fig. 3 Effect of TssA1^B inactivation on T6SS secretion activity. Cell associated protein (CA) and spent supernatants (SN) from cultures of *B. cenocepacia* H111 (WT) and the isogenic *tssA1^B* mutant (*tssA*::Tp), each with or without pBBR1MCS (VC) or pBBR1MCS-tssA1^B (*ptssA*), were fractionated by SDS-PAGE, blotted onto PVDF membrane and probed with a Hcp pAb or a mAb specific for the RNAP β subunit (lysis control). Detection of bound antibodies was with HRP-conjugated secondary antibodies. The *tssM* mutant (Δ*M*) was included as a control as TssM has been shown to be an essential component of the *B. cenocepacia* T6SS[31]. Uncropped images of the blots are shown in Supplementary Fig. 5

subdomain 2 (Sd2) (Fig. 4a, b). Sd1 (M1-P113) is comprised of five antiparallel α-helices (α1–α5) and contains the ImpA_N1 (L8-I31) and the ImpA_N2 (D55–P113) motifs (Fig. 4a–c).

Sd2 of Bc TssA1^B (L114-Q251) contains six α-helices (α6–α11), of which the three longest (α6, α10 and α11) form a five-helix bundle with α4–α5 of Sd1. Residues from ImpA_N3 (R123–R133) (Fig. 4a–c) form the start of α6, a helix which includes a sharp kink between L124 and G125. The C-terminal helix of Sd2 (α11) terminates at Q251 and is thereafter followed by the connecting interdomain linker to the CTD. The sequence conservation in the ImpA_N region across the TssA clades (Supplementary Fig. 1) implies that all Nt1 domains are closely related and adopt a similar fold, which includes the entirety of Sd1, and the N-terminal part of the first helix of Sd2 of Bc TssA1^B. Mapping of the 34 conserved residues in the ImpA_N region onto the structure of Nt1, shows that approximately two-thirds are involved in packing interactions in the core, with the remainder located predominantly at one end of the molecule in the loops between helices α1–α2, α3–α4, and α5–α6 (Fig. 4a, c).

TssA1^B CTD forms a double-layered ring with D_{16} symmetry. SEC-MALLS analysis of native Bc TssA1^B indicated an apparent molecular mass of 1.24 MDa (Supplementary Fig. 7c), suggesting that the TssA1^B oligomer contains ~30 subunits, and negative stain EM revealed rosette-like structures in which globular domains were irregularly distributed around a central ring (Fig. 5a). Two-hybrid analysis demonstrated that oligomerisation of TssA1^B required the CTD (Supplementary Fig. 7b), suggesting that the inner ring is composed of polymerised CTDs. This was confirmed by negative stain EM of the isolated CTD which was observed to form rings without peripherally located globular domains (Supplementary Fig. 7e).

The structure of a Bc TssA1^B CTD domain construct (I303-S373) was determined to 3.08 Å (Fig. 5b and Supplementary Table 3) and the asymmetric unit was observed to contain eight CTD monomers (chains A–H) (Supplementary Fig. 8a–c). Each chain folds into four α-helices (α12-α15) that lie in approximately the same plane, and form two antiparallel pairs (Fig. 5e). The eight monomers in the asymmetric unit further assemble into a higher order oligomer of 32 subunits, consistent with the SEC-MALLS analysis of TssA1^B and TssA1^B CTD oligomers (Supplementary Fig. 7c, d), with D_{16} symmetry using the crystallographic two-fold axes to make up a double-layered ring with 16 subunits in each layer (Fig. 5b and Supplementary Figure 8a). In the assembled Bc TssA1^B CTD ring, analysis by PISA[32] revealed an extensive dimer interface (~950 Å^2) between monomers in the different layers, which principally involves the interaction of residues located in helices α12–α13 with their two-fold related subunits forming a region of hydrophobic packing with hydrophilic interactions at the periphery (Fig. 5c, d). Further less extensive interactions (~570 Å^2) were identified between dimers that facilitate formation of the ring, with each pair of helices α14-α15 in one dimer, together with a small number of residues from α13, interacting with their equivalents in adjacent 16-fold related dimers on either side (Fig. 5e, f). This type of interlocking series of interactions is repeated in a cyclic arrangement to form a double-layered ring (Supplementary Fig. 8d). Nested deletion of consecutive helical segments of TssA1^B CTD revealed that loss of α15 resulted in the failure of the protein to assemble into a high molecular weight ring, and structure determination of the construct H12–H14 (helices α12–α14) and of that consisting of α12 and α13 alone (residues 303–347), revealed them both to be dimers (Supplementary Figure 8e–h and Supplementary Tables 4 and 5).

In the crystal structure of the TssA1^B CTD oligomer, the double ring is ~25 Å thick with a lumenal diameter of ~110 Å and an outer diameter of ~200 Å (Fig. 5b), in agreement with the dimensions of Bc TssA1^B and purified CTD oligomers determined by negative stain EM analysis (Supplementary Fig. 7e, f). The C-termini of the TssA1^B CTD monomers are located at the inner face of the ring, while their N-termini are orientated on the outside, consistent with EM analysis, which places the Nt1 domain at the periphery of the ring in the intact complex (Fig. 5a). The irregular arrangement of the globular domains and the poor resolution of the region linking them to the inner ring are consistent with the presence of a flexible linker connecting the Nt1 domain to the CTD. Analysis of the CTD sequences of Bc TssA1^B and Pa TssA1^A identified 17 conserved residues, of which 11 are involved in interactions around the subunit interfaces, including the highly conserved EPxxP motif (Fig. 5g and Supplementary Fig. 2a). The Bc TssA1^B CTD structure shows that this motif defines a loop that connects α12 and α13, and includes a salt bridge between E324 from this motif and a conserved arginine, R306, from a symmetry-related subunit (Fig. 5d). The conservation of these residues suggests that the quaternary structures of TssA1^A and TssA1^B are related. Moreover, although it has been previously proposed that the C-terminal region of Pa TssA1^A is structurally similar to the C-terminal part of the gp6 subunit of the T4 phage baseplate on the basis of low level sequence similarity[28], the similarity in sequence between TssA1^A and TssA1^B and the difference in structure of the TssA1^B CTD to gp6 indicates that this is not the case.

TssA2^A Nt2 is structurally similar to the TssA1^B Nt1 domain. The structure of an Ah TssA2^A Nt2 domain construct (D231-L374) was solved to 1.76 Å with four subunits (chains A–D) in the asymmetric unit (Supplementary Table 6) arranged as two independent dimers, consistent with previous SEC analysis[33], that showed that Ah TssA2^A Nt2 exists as dimers in solution. Each monomer folds into a cluster of seven α-helices (α1–α7), the arrangement of which is very similar to that of the EAEC TssA2^B Nt2 domain (RMSD = 2.0 Å) as determined by structural alignment[27,34]. Subsequent comparison using the Dali structural alignment server[35] identified a match between the Sd1 domain of Bc TssA1^B Nt1 and the Nt2 domains of both Ah TssA2^A and EAEC TssA2^B (RMSD = 2.7 and 2.6 Å, respectively) despite there being essentially no amino acid sequence similarity between them. Specifically, the related regions include helices α1-α5 of Bc TssA1^B Sd1 (i.e. most of the ImpA_N region) and α2–α6 of Ah

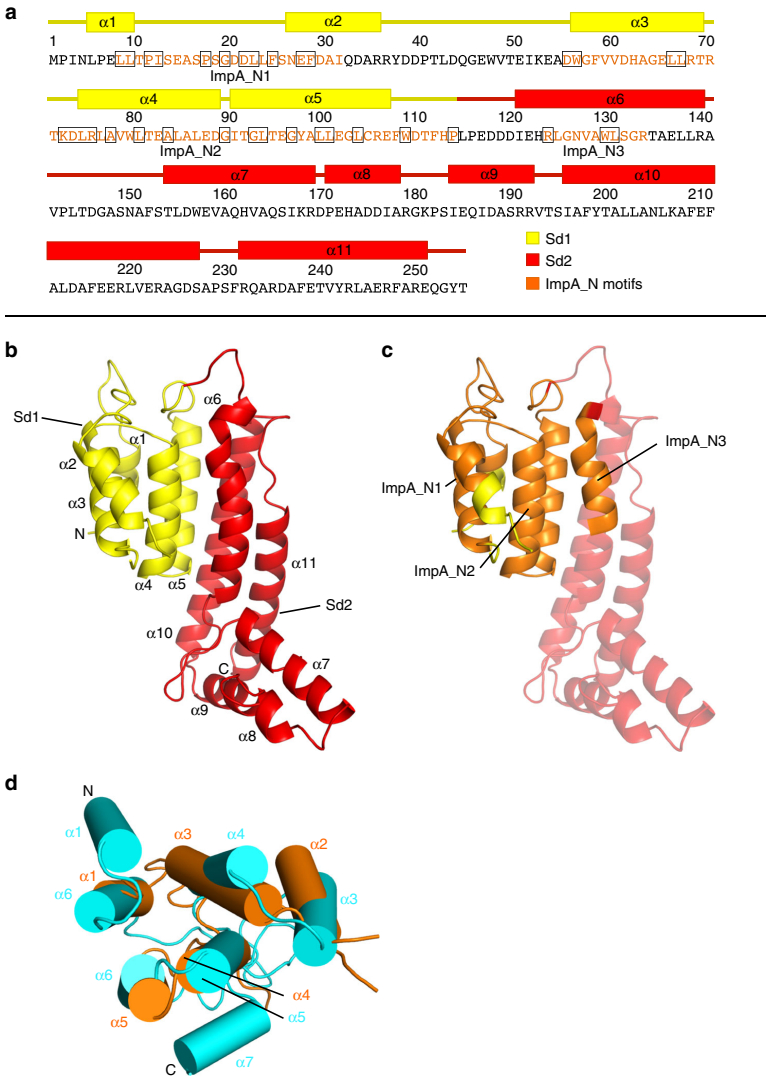


Fig. 4 High-resolution structure of Bc TssA1^B Nt1 and its similarity to Ah TssA2 Nt2. **a** Bc TssA1^B Nt1 sequence annotated to show the elements of secondary structure. Location of α-helices are shown by coloured rectangles above the amino acid sequence with the region corresponding to Sd1 highlighted in yellow and Sd2 highlighted in red. The three ImpA_N motifs are indicated in orange font and conserved ImpA_N domain residues (see Supplementary Fig. 1) are shown in black boxes. **b** X-ray structure of Bc TssA1^B Nt1 (PDB: 6HS5 https://www.ebi.ac.uk/pdbe/entry/pdb/6hs5). The two subdomains, Sd1 and Sd2, are shown in yellow and red, respectively. **c** X-ray structure of Bc TssA1^B Nt1 represented as in (b) showing location of ImpA_N1-ImpA_N3 regions in orange. For clarity, the helices of Sd2 (α6-α11) located C-terminal to ImpA_N3 are made more transparent than those in Sd1. **d** Bc TssA1^B Sd1 (orange) superimposed on the Ah TssA2^A Nt2 domain (PDB: 6G7B https://www.ebi.ac.uk/pdbe/entry/pdb/6g7b; cyan) showing the close structural similarity of helices α1-α5 of Bc TssA1^B Nt1 (majority of ImpA_N) to helices α2-α6 of Ah TssA2^A Nt2[34]. α-helices are shown as cylinders

TssA2^A Nt2 (Fig. 4d). This similarity suggests that an ancient gene duplication event may have played a role in the evolution of the Nt1 and Nt2 domains of TssA2 subunits.

TssA2^A CTD assembles into an oligomer with D_5 symmetry. Two-hybrid and SEC MALLS analysis indicated that Ah TssA2^A CTD self-associates and dictates the overall stoichiometry of the TssA2^A complex (Supplementary Fig. 9a, b, c, d). This is also a feature of EAEC TssA2^B [27]. However, unlike EAEC TssA2^B, the molecular weight estimations of Ah TssA2^A, Ah TssA2^A CTD and an Ah TssA2^A Nt2-CTD fusion construct corresponded to a subunit stoichiometry of 10 rather than 12. Consistent with this, negative stain EM showed that Ah TssA2^A CTD and Ah TssA2^A Nt2-CTD assembled into complexes that resembled a five-

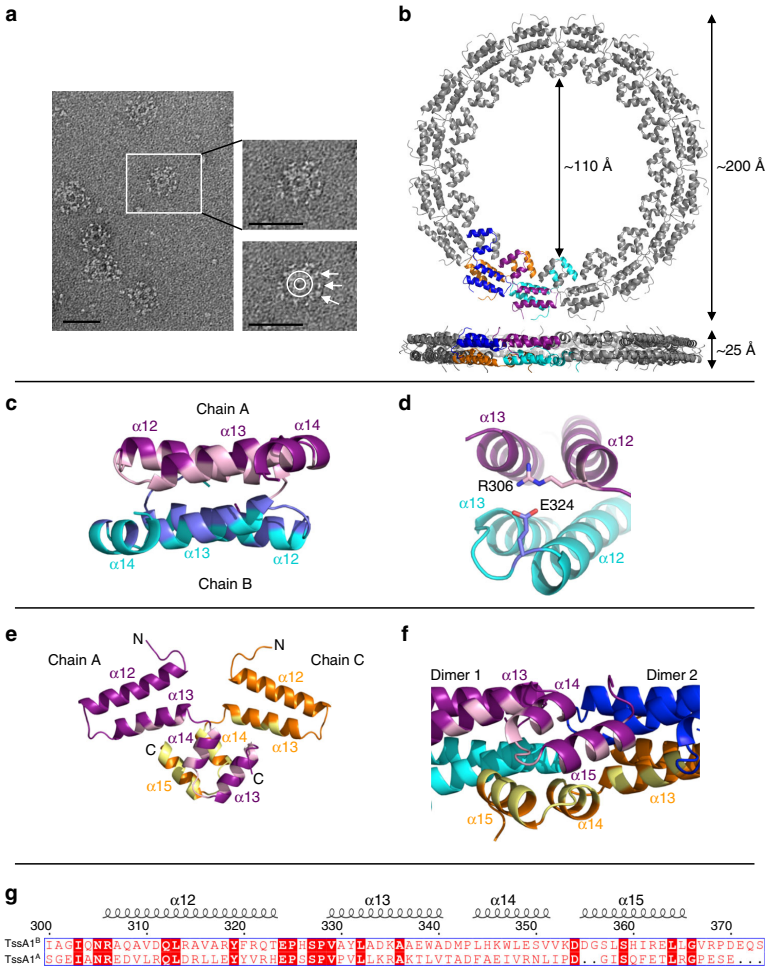
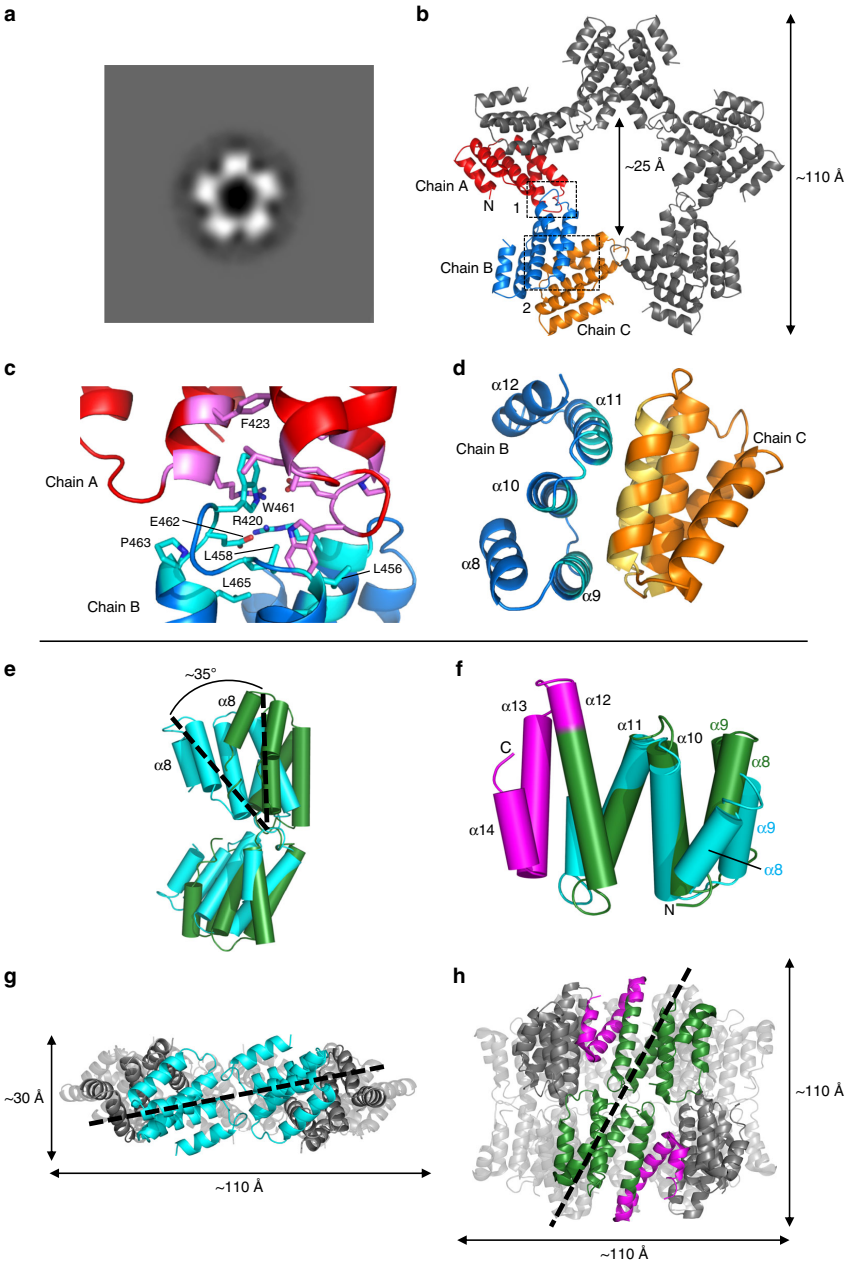


Fig. 5 Structure of the Bc TssA1^B CTD ring. **a** Negative stain EM of TssA1^B particles. The magnified view indicates the location of the Nt1 domains (arrows) presented around the CTD oligomeric ring (circle). Scale bar corresponds to 50 nm. **b** Bc TssA1^B CTD monomers assemble to form a double-layered ring containing 32 subunits exhibiting D_{16} symmetry (PDB: 6HS6 [https://www.ebi.ac.uk/pdbe/entry/pdb/6hs6]). The views shown correspond to a top view down the 16-fold axis and a side view perpendicular to the 16-fold axis, with the dimensions of the ring shown. **c**, **d** Dimerisation of Bc TssA1^B CTD monomers. **c** Side view from inside the Bc TssA1^B CTD ring showing interaction of residues from helices α12, α13 and α14 within a dimer (chain A, purple; chain B, cyan). Interacting surfaces are highlighted (chain A, light pink; chain B, blue). **d** Formation of the salt bridge between R306 and E324 that caps both ends of the dimer interface. **e**, **f** Interactions between dimers to form the Bc TssA1^B CTD ring. **e** Expanded view of the two chains shown in purple and orange in (**b**), to show interactions leading to ring assembly. **f** Side view from inside the Bc TssA1^B CTD ring showing the interlocking interaction between helices α13, α14 and α15 from two dimers which are critical to ring formation around the 16-fold axis: dimer 1, purple/cyan; dimer 2, blue/orange. The interfacing regions of the purple and orange monomers are shown in light pink and yellow, respectively. **g** Structure-based sequence alignment of the Pa TssA1^A and Bc TssA1^B CTD domains. The secondary structure of TssA1^B CTD is highlighted above the sequence[75]

pointed star (Fig. 6a and Supplementary Fig. 9e, f). The structure of crystals of an Ah His_6.TssA2^A Nt2-CTD construct (R232-K478) was successfully determined to 3.13 Å through molecular replacement using the coordinates of the Ah TssA2^A Nt2 dimer as a search model (Supplementary Table 7). This identified the position of five dimers of the TssA2^A Nt2 domain (R232 to L374, helices α1–α7) in the asymmetric unit, which did not appear to be arranged with any obvious symmetry with respect to each other. Following refinement, these dimers could be seen to be essentially identical in structure to that of the dimer formed by the isolated TssA2^A Nt2 domain (RMSD = 0.4 Å). The remaining C-terminal residues of each of the 10 subunits (chains A-J) were organised into a separate domain (the CTD) consisting of five α-helices (G388-L472, α8–α12). The 10 CTDs were, in turn, assembled into a flat ring resembling a star with D_5 symmetry (Fig. 6b), consistent with the negative stain EM and the SEC-MALLS analysis.

In the assembled TssA2^A CTD decamer two distinct two-fold interfaces can be identified (Fig. 6b). In the first, a two-fold related interface (shown by PISA[32] to be 530 Å^2) is formed primarily through interactions between residues from the



extended loop linking α11 and α12 (Fig. 6b, c) and the N-terminal end of α10 and their symmetry-related partners. This loop includes a WEP motif (W461, E462 and P463) that is conserved across both TssA2 sub-clades, where W461 is partially buried within a hydrophobic pocket formed by the conserved residues L456, L458 and L465, and an interaction with F423, a non-conserved residue, in the two-fold related monomer. The entrance to this pocket is capped by a salt bridge between E462 of the motif and R420, two residues that are again conserved (Fig. 6c and Supplementary Fig. 2b). In the second, more extensive interface (950 Å^2), the WEP-mediated dimers assemble around the five-fold axis through interactions between the surface of helices α9, α10 and α11 (Fig. 6b, d) and their two-fold related symmetry mates to form a mixed interface of hydrophobic/hydrophilic contacts. The TssA2^A CTD (S381-K478) ring has a thickness of ~30 Å, an outer diameter of ~110 Å, and a lumenal

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-07247-1 ARTICLE

Fig. 6 Structure of the Ah TssA2^A CTD ring. **a** Negative stain EM particle averaging of Ah TssA2^A CTD indicating five-fold symmetry. **b** Structure of 10 Ah TssA2^A CTD monomers assembled to form a decameric oligomer exhibiting D_5 symmetry (PDB: 6G7C https://www.ebi.ac.uk/pdbe/entry/pdb/6g7c). Boxed region 1 (Interface 1), two-fold axis formed by dimerisation via the WEP motif (red and blue chains). Boxed region 2 (Interface 2), two-fold axis generated through packing of helices α9, α10 and α11 (blue and orange chains). The N-terminus of chain A is indicated and the dimensions of the ring are shown. **c** Interface 1 with key interacting residues indicated (chain A, pink: chain B, light blue). **d** Side view of interface 2 with interacting surfaces highlighted in light blue and sand. **e** Superposition of WEP-mediated dimers from Ah TssA2^A (cyan) and EAEC TssA2^B (PDB: 4YO5, green) about the WEP interface. The monomers exhibit a different relative orientation of ~35° about the WEP interface. View corresponds to chains A and B shown in (**b**). **f** Structural similarity of the CTD monomers of Ah TssA2^A (cyan) and EAEC TssA2^B (PDB: 4YO5, green). Helices α13 and α14 are unique to EAEC TssA2^B and form the C-terminal extension (magenta), which is important for the different quaternary structures assembled by TssA2^A and TssA2^B. **g**, **h** Side by side views of the Ah TssA2^A (**g**) and EAEC TssA2^B (**h**) CTD oligomers with their respective five-fold and six-fold axes vertical. Assembly of Ah TssA2^A WEP-mediated dimers (cyan) around the five-fold axis involves a rotation relative to EAEC TssA2^B (green) of ~50° around the dimer axis as illustrated by the dotted black line. The neighbouring subunits with which each monomer within a WEP-mediated dimer interacts are shown in grey

diameter of 25 Å (Fig. 6b, g). As in the structure of the Bc TssA1^B CTD ring, the N-terminus of each TssA2^A CTD monomer is orientated on the outside of the CTD oligomer resulting in the presentation of the Nt2-CTD interdomain linker at the periphery of the molecule (Fig. 6b).

Despite the differences in symmetry between TssA2 CTD oligomers assembled by members of sub-clades 2^A and 2^B, the fold of the Ah TssA2^A CTD monomer is closely related to that of its counterpart in EAEC TssA2^B (PDB = 4YO5, RMSD = 1.6 Å). Moreover, close inspection shows that the conserved WEP motif is maintained at the dimer interface in members of both TssA2 sub-clades, albeit that the two monomers of Ah TssA2^A are arranged in a slightly different orientation with respect to each other, corresponding to a rotation of ~35° compared to their TssA2^B counterparts (Fig. 6e).

The way in which the dimers assemble around their respective five-fold and six-fold symmetry axes in the Ah and EAEC TssA2 CTDs are strikingly different, in contrast to the similarity seen around their two-fold axes. In Ah TssA2^A, two monomers, related by a WEP-mediated two-fold axis, pack adjacent to one another, parallel to the five-fold axis, resulting in the formation of an oligomer which resembles two interpenetrating rings (Fig. 6g). The face of Ah TssA2^A, used in subunit assembly around the five-fold axis (residues from α9, α10, and α11), interacts with a symmetry related equivalent (Fig. 6g). In comparison, EAEC TssA2^B CTD contains a 41 residue C-terminal extension which folds into two additional helices (α13 and α14) (Fig. 6f). These dominate the interactions around the six-fold axis, packing against the face of helices α9, α10 and α11 of an adjacent subunit (Fig. 6h). Consequently, this causes an incline of the WEP-mediated dimer by ~50°. Therefore, in EAEC TssA2^B, two monomers, related by a WEP-mediated two-fold axis, stack on top of one another, parallel to the six-fold axis, with the dimers then assembling around the latter to form a double-layered ring (Fig. 6h). Interestingly, despite the difference in symmetry we note that the lumenal diameter and outer diameter of the TssA2^A and TssA2^B CTD oligomers are similar (Fig. 6g, h), thereby placing the Nt2-CTD interdomain linker at the same distance from the centre of the complex.

A flexible linker in TssA2^A Nt2-CTD mediates domain mobility. Despite the electron density for the Nt2 and CTD domains being clear, density was observed for only part of the Nt2-CTD interdomain linker, and, depending on the chain, between 6 and 13 residues (Q375-A387) could not be identified within this region. However, restrictions imposed by the maximum length of the absent linker residues and the relative orientations of the Nt2 and CTD domains allowed the unambiguous assignment of the Nt2 domains to their corresponding CTD (i.e. in the same polypeptide chain). This revealed that in the decameric Ah TssA2^A Nt2-CTD complex, the Nt2 domains are not related by the five-fold symmetry exhibited by their corresponding CTDs (Fig. 7a). Instead, dimers of Nt2 make no contacts with their own CTDs but rather contact either TssA2^A Nt2 or CTD domains of other decamers in the crystal lattice. Moreover, as a result of the flexible linker, the Nt2 dimers have quite different orientations with respect to the five-fold axis, with angular rotations of up to ~60° about the vertical axis (Fig. 7b), and ~45° and ~60° about the horizontal and torsional axes, respectively. Although, the observed Nt2 domain orientations described above are believed to be an artefact of crystal packing, the range of motion seen is believed to reflect the inherent flexibility between CTD and Nt2 domains. This is consistent with negative stain EM analysis that revealed a lower calculated correlation coefficient for strict five-fold symmetry when comparing an Nt2-CTD construct to a CTD construct (Supplementary Fig. 9e–h).

Sequence alignments suggest that the Nt2-CTD interdomain linker is longer in TssA2^B subunits than in TssA2^A subunits (Supplementary Fig. 2b). For example, the linker in Ah TssA2^A is composed of ~21 residues whereas it contains ~37 residues in EAEC TssA2^B. The most likely explanation for the presence of a longer linker in TssA2^B is to overcome a restricted range of motion of Nt2 dimers that would be imposed by a shorter linker due to the possible physical interference between the Nt2 dimer and the 'taller' double-layered ring arrangement formed by EAEC TssA2^B CTD compared to the 'squat' interpenetrating ring assembled by Ah TssA2^A CTD (Fig. 6g, h).

The quite different orientation of the CTD WEP-mediated dimers in TssA2^A and TssA2^B oligomers has one further consequence, which is that the orientation of the interdomain linker between Nt2 and CTD is necessarily rotated by ~50° (Fig. 6g, h). We assume that the flexibility of the linker allows this structural difference to be maintained without affecting the biological role of the protein. The finding that in the Ah TssA2^A Nt2-CTD structure the Nt2 domains consistently occur as dimers further suggests that the Nt2 domains remain associated as part of the biological role of TssA, in contrast to the proposal by Zoued and co-workers[27], who interpreted low resolution negative stain EM data to suggest that they separate to form monomers[27]. A rigorous test will require a high-resolution EM reconstruction from TssA particles embedded in vitreous ice, although this could be challenging given the apparent flexibility of domain linkers.

Discussion

The structural studies presented here show that TssA1^B, TssA2^A and TssA2^B possess structurally diverse CTDs that lead to distinct overall architectures for their respective complexes. Nevertheless, all these subunits contain a conserved N-terminal ImpA_N region that is peripherally located in the assembled complexes. The question that arises from this is whether these TssAs play the same role in the function of the T6SS? The presence of the ImpA_N region in the Nt1 domain of subunits from all TssA

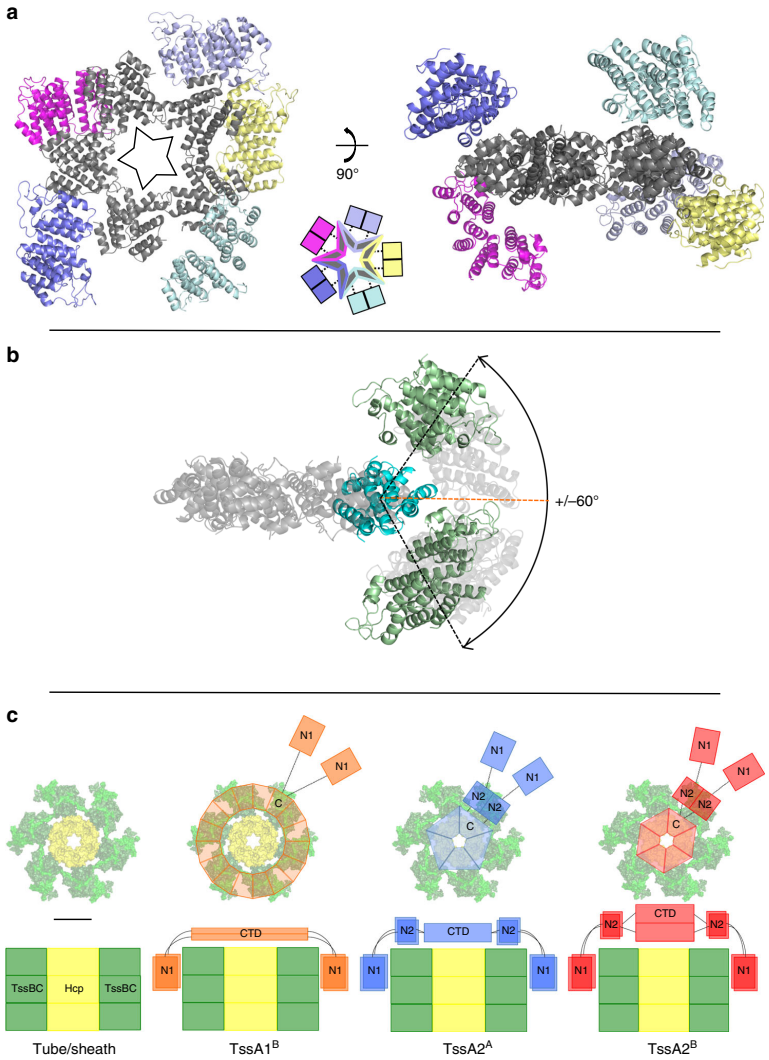


Fig. 7 Schematic representation of TssA ring dimensions and domain flexibility. **a** Arrangement of the Ah TssA2^{A} Nt2 domains around the CTD oligomer as observed in the asymmetric unit. The Nt2 domains do not follow strict five-fold symmetry. The decameric CTD ring is shown in dark grey perpendicular to the five-fold (left) and two-fold (right) axes with each peripheral Nt2 dimer shown in a different colour. The diagram between the two structures shows the relationship between dimer-forming Nt2 domains (represented by coloured squares) and their corresponding C-terminal domains in the TssA2^{A} Nt2-CTD decamer. **b** Displacement of the Ah TssA2^{A} Nt2 dimers perpendicular to the five-fold axis of the CTD oligomer. The CTD ring is shown in light grey with one WEP-mediated CTD dimer indicated in cyan to which is tethered an Nt2 dimer. The Nt2 dimer (light green) is shown at its maximum displacement of ~60° either side of the plane of the CTD ring (orange line). Other observed positions of Nt2 are indicated by translucent shading. **c** Top and side view representations showing the domain architecture of TssA subunits from clades 1^{B}, 2^{A} and 2^{B} modelled on the baseplate-distal end of the T6SS tube/sheath assembly during polymerisation. The *V. cholerae* TssBC extended sheath (green) surrounding a Hcp hexamer (yellow) (PDB: 5MXN) is shown on the left. Onto this, is modelled the TssA1^{B} CTD oligomer (orange) with 16-fold symmetry, showing one stacked CTD dimer attached to a pair of Nt1 domains via flexible linkers, the TssA2^{A} CTD oligomer (blue) with five-fold symmetry, showing a WEP-mediated CTD dimer connected to the two monomers of an Nt2 dimer via flexible linkers, each of which are in turn connected to an Nt1 domain, and the TssA2^{B} CTD oligomer (red) displaying six-fold symmetry, showing a CTD dimer connected to the two monomers of an Nt2 dimer via flexible linkers, each of which are in turn connected to an Nt1 domain. Labels are: C, CTD dimers (top view); N1, Nt1; N2, Nt2; CTD, CTD oligomer (side view). Scale bar corresponds to 65 Å

clades strongly suggests that they have a common function that is of critical importance. This is further supported by the studies described here which reveal that TssA1^B interacts with the baseplate components TssE, TssF and TssK, the inner tube subunit, Hcp, the tip protein, VgrG, the sheath subunit TssC and cytoplasmic components of the cell envelope-anchoring complex. This is strikingly similar to the TssA2^B subunit that was shown to interact with the same components except for TssF[27]. Indeed the TssA1^B and TssA2^B CTD oligomers were both shown to interact with Hcp and VgrG despite their quite different structures. This would argue for a similarity of function despite the divergent architecture of the two proteins. Given the sequence similarity and predicted structural relationship between TssA1^A and TssA1^B, this suggests that all three clades of TssA perform related roles within the T6SS. If the role of these quite different TssA subunits is similar, then how is this achieved given the structural variations that are observed?

To date, two distinct models for the role of TssA have been proposed. In the first, baseplate-associated model, it was proposed that TssA is a component of the T6SS baseplate required for priming the inner tube and sheath for assembly, but not migrating during polymerisation[28]. This model was based partly on predicted similarities in structure for the CTDs of TssA1^A and the phage T4 baseplate protein gp6, to which it was believed that this TssA subunit was related. The structure we have obtained for TssA1^B, which is closely related to TssA1^A, reveals that it has a quite different structure to gp6. This suggests that TssA1^A is very unlikely to be a gp6 orthologue. Due to the absence of structural similarity to baseplate proteins and the conservation of the Nt1 domain among all TssA subunits, it is therefore likely that TssA1 subunits fulfil a similar role to TssA2 subunits.

In the alternative model proposed for a TssA subunit that corresponds to TssA2^B, TssA acts as a coordinator of inner tube and sheath assembly, in which the CTD oligomer of TssA2^B would lie close to the centre of the tail assembly, above the position occupied by Hcp where it would coordinate polymerisation of the inner tube and sheath[27,29,36,37]. In one version of this proposal (the cap model) eversion of CTD monomers, from a closed form of the CTD oligomer to an open form, is then proposed to permit the passage of Hcp through the lumen of the open ring. This would allow sequential stacking of Hcp hexamers, with the N-terminal domains of TssA2^B interacting with the polymerising sheath. However, as yet, an open form of TssA2^B has not been observed and the proposed reorientation of monomers in the CTD oligomer remains speculative.

The studies described here show that the fold of TssA2^A retains clear structural similarity to TssA2^B yet generates a decamer possessing five-fold symmetry in contrast to the six-fold symmetry of the TssA2^B dodecamer[27]. Therefore, while it could be argued that it might lie in the same position in the complex as TssA2^B, the difference in symmetry implies there must be changes to the complementary surfaces of surrounding TssB/-C/-D subunits to allow pseudo equivalent interactions. This situation becomes even more extreme given the quite different organization of the TssA1^B ring oligomer with its sixteen-fold symmetry. One possibility is that the structure of TssA1^B we have determined represents an open form of the TssA CTD oligomer, leaving a closed form of the TssA1^B CTD oligomer, which might pack on top of a Hcp hexamer, still to be identified. Alternatively, if some features of the cap model are correct, and TssA1^A/TssA1^B and TssA2^A also serve to coordinate tube and sheath assembly, then, given their quite different symmetries and architectures, rationalising how they function in an equivalent way to that proposed for TssA2^B is problematic. One way for this dilemma to be resolved is if the different structures of these distinct TssAs represent different ways in which a common function can be maintained.

Comparing the TssAs of clades 1 and 2, these structures indicate that the Nt1 domain is linked either to a large diameter CTD ring (200 Å) of high symmetry, via a single flexible segment, as seen in sub-clade TssA1^B, or to a smaller (100–110 Å) diameter CTD of lower symmetry, through a middle domain (Nt2) flanked by two flexible linkers, as seen in sub-clades 2^A and 2^B. Despite these two alternative architectures, the radial position of the Nt1 domain from the centre of the complex is similar in both TssA1 and TssA2 clades (Fig. 7c). This suggests that a role of the Nt2 domain is to act as a spacer connecting the Nt1 domain of TssA2^A and TssA2^B to their corresponding CTDs.

In light of the above, and given the conservation of the Nt1 domain, we suggest that the common feature shared by all these TssAs is that the Nt1 domain is positioned around the periphery of the ring-forming CTD oligomers, furthest away from the symmetry axis but at an equivalent distance. Given the apparent symmetry mismatch between the tube/sheath assembly and the TssA CTD oligomers of some TssA species, it is important to consider how equivalent interactions between these components can be maintained. The simplest explanation is to consider a model whereby the interactions between TssA subunits and the tube/sheath assembly are solely the property of the conserved Nt1 domains, which have been suggested to participate in assembly of the latter[27]. This would mean that the role of the CTD or Nt2-CTD oligomers of clades 1 or 2, respectively, are to provide a platform for the presentation of the Nt1 domains above the growing tube/sheath complex at a similar distance from the centre of each TssA complex (Fig. 7c). In such a model it would not be necessary for the CTDs to interact directly with the tube/sheath assembly except perhaps in a transient manner as indicated by interaction studies. This agrees with surface plasmon resonance data for EAEC TssA2^B where the interaction of the Nt1-Nt2 region with the TssBC sheath is significantly greater than that of the corresponding CTD with Hcp[27].

This proposal in which one domain or component of a protein subunit (CTD/Nt2-CTD) associates into a platform that facilitates the role of a second domain (Nt1) in subunit incorporation by a polymerising protein assembly, is a theme that occurs in the mechanism of flagellum assembly by the HAP2 cap protein[27,29,38,39]. An important aspect of the flagellin-capping model is the observation of domain mobility, a feature clearly shown here for both the relative motion of Nt1 with respect to the CTD in TssA1^B and between the CTD and Nt2 in TssA2^A and TssA2^B, and the predicted flexibility of the linker between Nt2 and Nt1 in TssA2^A/TssA2^B. This would provide the necessary movement of the Nt1 domains for dynamic interactions with the TssBC subunits during sheath biogenesis. This might involve the displacement of Nt1 domains and their linkers, consequently allowing passage of Hcp, in addition to further TssB/C subunits to be incorporated into the growing sheath. This model does not necessarily require eversion of the CTD rings. However, we do not rule out the possibility that TssA2^A/TssA2^B CTD oligomers can re-orientate to allow access of Hcp hexamers to the growing tube. In this regard, we note that the TssA1^B CTD ring oligomer can adequately accommodate Hcp hexamers within its lumen.

A question arising from the dihedral symmetry shown by all TssA structures that have been determined to date, is how this leads to a uni-directional assembly of the tube/sheath complex. While any initial interaction between TssA and other subunits of the complex can use either two-fold related face of the TssA oligomer, once contact has been made with T6SS components, the symmetry is broken. This would occur when TssA interacts with cytoplasmic components of the TssJLM membrane complex that occurs during the initial steps in T6SS assembly[27], as one

surface of the TssA ring faces the membrane while the other faces the cytosol. According to the current model, this asymmetry is reinforced once the baseplate is recruited to the membrane complex, whereupon one face of the TssA oligomer is located adjacent to the cytoplasmic face of the baseplate[27]. Here TssA is ready to coordinate inner tube and sheath assembly once the first Hcp ring is recruited to the base of the VgrG trimer located at the hub of the baseplate[40]. Although the precise mode of action of TssA subunits remains unknown, they are likely to lock the inner tube/sheath structure in the high energy form prior to firing[29]. Thus, TssA may possess chaperone-like activity.

A final question is to consider the mismatch in symmetry between the assemblies formed by members of different TssA clades, on the one hand, and other components of the T6SS, on the other, and how this impacts on the mechanism by which TssA executes its role. For example, the six-fold symmetry of the TssA2^{B} oligomer is consistent with the symmetry of the tube-sheath structure but not with the five-fold symmetry proposed for the TssJLM membrane complex[10,27]. In contrast, the symmetry of the TssA2^{A} oligomer matches that of the membrane complex but not the contractile machinery. In this regard it is interesting to note that, as seen with the different symmetries of TssA2^{A}/TssA2^{B}, both five-fold and six-fold symmetrical variants of the HAP2 capping protein can interact with the common 11-fold symmetrical flagellin filament[41,42] suggesting that the apparent symmetry mismatch between oligomers formed by some TssA species and the tube/sheath assembly is not necessarily problematic. Based on these considerations we suggest that a flagellin-like capping model is a plausible mechanism for the role of all TssA subunits. Ultimately, these proposals need to be confirmed by the structural analysis of other complexes of secretion systems from each of the different TssA clades. Once this has been achieved, the manner in which the difference in structure and symmetries among TssA subunits resolves to provide equivalent function will be revealed.

Methods

Bacterial strains and growth conditions. Bacterial strain details are given in Supplementary Table 8. *E. coli* strains JM83 and DH5α were used for routine cloning steps. *B. cenocepacia* strain H111 and *A. hydrophila* ATCC 7966 were the source of the TssA1^{B} and TssA2^{A} subunit genes, respectively. LB was routinely used for growing bacteria except where indicated[43]. Lennox medium was used with glucose included at 0.2%[44]. D-BHI broth was made by dissolving 9.25 g brain-heart infusion (BBL) in 25 ml water and dialysed (12,000–14,000 MWCO) against 500 ml water overnight[45]. Antibiotics were used at the following concentrations for *E. coli*: ampicillin, 100 μg/ml; kanamycin, 50 μg/ml; chloramphenicol, 25 μg/ml; trimethoprim, 25 μg/ml (selection for trimethoprim resistance was applied in/on isosensitest (IST) broth/agar (Oxoid)). Bacto MacConkey agar base used for two-hybrid assays was obtained from Becton, Dickinson and Co.

Plasmids and primers. All oligonucleotides and plasmids used in this study are listed in Supplementary Tables 9,10. Construction of plasmids is also described in Supplementary Table 10. Plasmids were transferred to *E. coli* by transformation[46] and to *B. cenocepacia* by conjugation[47,48].

Construction of a *B. cenocepacia tssA1*^{B} mutant. To construct the *B. cenocepacia* H111 *tssA*::Tp mutant, a segment of the *tssA1*^{B} ORF was amplified with primers iotAfor3 and iotArev and cloned between the *Hin*dIII and *Bam*HI sites of pBBR1MCS (to give pBBR1MCS-'tssA1^{B}) whereupon it was disrupted by insertion of the *dfrB2* (Tp^{R}) cassette from p34E-Tp (generating pBBR1MCS-'tssA1^{B}::Tp). The *tssA1*^{B}::Tp fragment was then transferred to the suicide vector pSHAFT2 and the resultant plasmid (pSHAFT2-'tssA1^{B}::Tp) conjugally mobilised into H111 using *E. coli* S17–1(λpir). Ex-conjugants were selected on M9-glucose agar containing trimethoprim (25 μg/ml). Mutants were identified by PCR screening chloramphenicol-sensitive ex-conjugants using primers iotAfor and iotArev2 that anneal to sequences located outside the genomic region contained on the suicide vector. Construction of the *tssM* deletion mutant was performed by conjugally mobilising the allelic replacement plasmid, pSNUFF-ΔtssM, containing a *tssM* deletion allele (codons 506–1216 removed) into strain H111 using *E. coli* SM10 (λpir). Ex-conjugants containing the integrated allelic replacement vector were selected on M9-glucose agar containing trimethoprim (25 μg/ml) and were then plated on M9-glucose agar containing *p*-chlorophenylalanine to identify recombinants in which a second crossover had resulted in excision of the integrated plasmid[49]. Mutants were identified by diagnostic PCR using primers tssM-OPfor and tssM-OPrev2.

Hcp secretion assay. D-BHI broth was inoculated with overnight cultures of *B. cenocepacia* strains to give a starting OD_{600} of 0.03, and grown to an OD_{600} of 1.0 at 37 °C. Cultures were centrifuged to collect the cells and the supernatant was filter-sterilised. To precipitate secreted proteins, sodium deoxycholate was added to the cleared supernatant to a final concentration of 0.2 mg/ml followed by incubation on ice for 30 min. Trichloroacetic acid was then added to a final concentration of 10% (v/v), and the suspension was incubated overnight at −20 °C. Precipitated proteins were recovered by centrifugation for 15 min at 14,000 × *g* at 4 °C and washed with ice cold acetone. Pellets were air-dried and resuspended in 0.001 of the original volume of SDS-PAGE sample buffer. Cell-associated proteins were isolated by resuspending the initial bacterial pellet in a volume of SDS-PAGE sample buffer that was 20-fold less than the starting culture volume and transferred to a microcentrifuge tube. All protein samples were boiled for 5 min, centrifuged for 10 min at 14,000 × *g* to remove cell debris and then fractionated by electrophoresis in a 15% discontinuous SDS-PA gel followed by electroblotting onto a PVDF membrane (Millipore). Membranes were blocked with 5% (w/v) non-fat skimmed milk in TBS containing Tween 20 (0.05% (v/v)) solution for 1 h. Hcp was detected by probing with a custom polyclonal rat antibody raised to purified His$_6$.Hcp (1:1000) and goat anti-rat HRP secondary antibody (1:5000). Antibody to the cytoplasmically located β subunit of *E. coli* RNA polymerase (1:2000) that cross-reacted to the *Burkholderia* spp. subunit was also used as a lysis control, with a rabbit anti-mouse HRP secondary antibody (1:5000). Detection of chemiluminescence was performed using an EZ-ECL detection kit and a XRS+ imaging system (BioRad). Uncropped blots are shown in Supplementary Fig. 5.

Two-hybrid analysis. The bacterial adenylate cyclase two-hybrid (BACTH) system was employed[50]. *B. cenocepacia* T6SS subunit genes (*tssA-tssC*, *hcp*, *tssE-tssG*, *vgrG*, *tssJ-tssM*) or domain coding sequences were fused to the N- and C-terminal coding sequences of the *B. pertussis* CyaA T25 or T18 fragments present in plasmids pKT25, pKNT25, pUT18 and pUT18C. To avoid membrane or periplasmic targeting of CyaA fusion proteins, signal sequences and TMDs were omitted from the TssJ, TssL and TssM hybrid proteins. As a consequence, the cytoplasmic N-terminal region of TssM (TssM$_N$) and the periplasmic C-terminal region (TssM$_C$) were fused to the CyaA domains separately. Hybrid proteins were generated with full-length VgrG and also with the conserved core region corresponding to the phage T4 gp27 and gp5 proteins (VgrG$_C$). DNA encoding full-length *A. hydrophila* TssA2^{A} and the TssA2^{A} Nt1 domain were also cloned into all four vectors while DNA encoding the Nt2 and CTDs were cloned into pKT25 and pUT18C only. Further details on plasmid constructions are given in Supplementary Table 10. The *E. coli cya*$^{-}$ strain BTH101 was transformed with the two compatible plasmids expressing fusion proteins and plated on MacConkey indicator agar containing 1% (w/v) maltose, ampicillin (100 μg/ml) and kanamycin (50 μg/ml). Colony phenotypes were scored after 3 and 5 nights incubation at 30 °C (Mal^{+} (Cya^{+}) colonies are purple and Mal^{-} (Cya^{-}) colonies are white). The appearance of purple colonies after 3 days incubation indicated a strong Cya^{+} phenotype. The positive control combination (Zip control), where each plasmid encoded the leucine zipper segment of the yeast transcriptional activator protein, GCN4, fused to T25 and T18[50] was always included as were all negative control combinations (two compatible empty vectors and one empty vector in combination with a plasmid specifying a hybrid protein).

The efficiency of CyaA complementation was quantitated by measurement of the β-galactosidase activity in broth cultures. Cells were grown in LB broth containing antibiotics and 0.5 mM IPTG at 30 °C with aeration for 14–16 h. The cultures were then chilled on ice before being diluted in LB medium to OD_{600} 0.4–0.6. 50 μl of each diluted culture was added to test tubes containing 950 μl of Z buffer (0.06 M Na_2HPO_4, 0.04 M NaH_2PO_4, 0.01 M KCl, 0.001 M $MgSO_4$ and 0.27% (v/v) β-mercaptoethanol), chloroform (30 μl) and 0.1% SDS (30 μl). The mixtures were vortexed for 10 s and equilibrated at 30 °C for 15 min whereupon 200 μl of ONPG (4 mg/ml in Z buffer) was added to initiate the reaction. Reactions were stopped by addition of 500 μl 1 M Na_2CO_3 and the absorbance at 420 and 550 nm was measured. The β-galactosidase activity (in Miller units) was calculated using the equation 1000 (OD_{420}−1.75 × OD_{550}/*t* x *v* × OD_{600}) where *t* corresponds to the duration of the reaction in minutes and *v* corresponds to the volume of culture used in the assay[43]. The background activity measured in the negative controls (a pair of empty vectors or one empty vector in combination with a plasmid encoding a hybrid protein) was 75–90 Miller units (Mu). Each assay reaction was performed in duplicate on three biological replicates ($n = 3$), and the mean, along with standard deviation, was calculated.

Co-immunoprecipitation assay. Co-IP analysis was performed on cell lysates following co-expression of N-terminal FLAG-tagged TssA1^{B} as bait and an epitope-tagged T6SS subunit (or domain) as prey in *E. coli*. FLAG.TssA1^{B} was expressed from the second T7 RNAP-dependent promoter contained on pACYCDuet-1 while the prey protein was expressed from the upstream promoter

on the same plasmid (except where indicated below). Prey proteins were generally tagged at their N-terminus with the vesicular stomatitis virus glycoprotein (VSV-g) epitope tag (TssC was C-terminally tagged). TssF was tagged at the N- or C-terminus with the human influenza haemagglutinin (HA) epitope tag. In two cases, a separate plasmid was used to express the prey protein: N-terminal HA-tagged TssF was expressed from the downstream T7 promoter of pETDuetΔO whereas MBP-TssE was expressed from the *tac* promoter present on pMAL-c5X. Further details on plasmid constructions are given Supplementary Table 10.

Expression of all proteins was induced by addition of 0.5–1 mM IPTG (0.1 mM for the TssC co-IP) to *E. coli* BL21(DE3) cells grown to OD_{600} 0.5–0.7 in 50 ml BHI broth at 37 °C (for the Hcp, TssK and TssL co-IPs), 30 °C (for the TssE, TssF and $VgrG_C$ co-IPs) or 22 °C (for the TssC and $TssM_N$ co-IPs), according to the solubility of the prey protein, and incubation was continued at the same temperature for 2–3 h. Bacterial cells were harvested and stored frozen at −20 °C overnight. Defrosted *E. coli* cell pellets containing co-expressed proteins were resuspended in TBS (50 mM Tris-Cl (pH 7.4), 150 mM NaCl (or 500 mM NaCl for $VgrG_C$ and $TssM_N$)) with EDTA-free protease inhibitor tablets (Roche) at a ratio of 5 ml TBS to 1 g cell pellet. Following sonication, the soluble fraction was cleared by centrifugation at 20,000 *g* for 15 min at 4 °C, and $MgCl_2$ was added to a final concentration of 2.5 mM. After gently mixing by inversion, the mixture was centrifuged at 13,000 *g* for 10 min. The supernatant was then incubated with Tween 20 (0.2%) for one hour with gentle mixing on a rotating wheel at room temperature, and followed by centrifugation at 15,000 *g* for 10 min. The supernatant was used for co-immunoprecipitation with the anti-FLAG affinity gel. Fifty microlitres of the anti-FLAG M2 affinity resin suspension equilibrated in TBS was added to 200–1000 µl cell lysate (depending on the abundance of FLAG-tagged $TssA1^B$ in the lysate) and the final volume was brought to 1 ml by adding TBS. After 2–3 h gentle mixing at 4 °C to allow for precipitation, the mixture was centrifuged at 7000 × *g* for 30 s, and the supernatant was removed (unbound material). The resin was then washed three times with 1 ml TBS with gentle mixing for 5–15 min at 4 °C between each wash followed by centrifugation at 7000 × *g* for 30 s. After discarding the supernatant of the final wash, the immunoprecipitated material was dissociated from the resin by boiling with 25 µl of SDS-PAGE sample buffer with or without 5% β-mercaptoethanol for 3 min. Finally, the mixture was centrifuged at 7000 × *g* for 30 s to clear the suspension. The supernatant (immunoprecipitated material) was subjected to SDS-PAGE alongside a pre-stained protein ladder, blotted onto PVDF membrane and probed with tag-specific antibody (anti-VSVg, anti-HA or anti-FLAG, each at 1:5000–1:10,000 dilution) and detected with HRP-conjugated secondary antibody (1:5000–1:6666 dilution). MBP fusion proteins were detected with HRP-conjugated anti-MBP (1:5000 dilution). The sizes of the tagged *B. cenocepacia* T6SS subunits in kDa (excluding the N-terminal formyl-methionine) are: $TssA1^B$, 42.6; TssC, 56.1; Hcp, 19.8; TssE, 61.6; TssF, 69.8 (both versions); $VgrG_C$, 60.7; TssK, 51.3; TssL, 24.4; $TssM_N$, 44.5; MBP, 45.5 kDa. Uncropped blots are shown in Supplementary Fig. 4.

Antibodies. Polyclonal antibody to *B. cenocepacia* Hcp was raised in rats at the University of Sheffield Biological Services unit using N-terminal His-tagged Hcp which was purified following its overproduction in *E. coli* as described below. Rabbit anti-FLAG mAb and anti-VSVg pAb were obtained from Sigma-Aldrich (cat. no. F7425 and V4888, respectively), whereas rabbit anti-HA mAb was purchased from Cell Signaling Technology (cat. no. C29F4). HRP-conjugated mouse anti-MBP polyclonal antibodies were obtained from New England Biolabs (cat. no. E8038S). The mouse mAb to the β subunit of *E. coli* RNA polymerase was obtained from Neoclone (cat. no. W0023). Proteins harbouring a polyhistidine tag were detected using the HisProbe-HRP conjugate (Thermo Fisher, cat. no. 15165) or mouse anti-KLH-conjugated His-tag peptide (BioLegend, 652501). HRP-linked goat-derived (secondary) antibodies that cross-react to rat and rabbit IgG were obtained from SouthernBiotech (cat. no. 3050–05) and Vector Laboratories Ltd (cat. no. PI-1000), respectively. HRP-conjugated rabbit anti-mouse IgG secondary antibody were purchased from Thermo Fisher (cat. no. 31450).

Expression and purification of Hcp for antibody production. His_6.Hcp was overproduced following a 4 h induction of the T7 promoter on pET14b-His_6.TssD2 carried by *E. coli* BL21(DE3) cells growing in BHI at 37 °C. His_6.Hcp, which lacks the first three native Hcp amino acids, was purified by sequential IMAC on nickel-sepharose (HisTrap HP, GE Healthcare Life Sciences) and SEC using a Superose-6 10/300 GL column equilibrated in PBS. Prior to the SEC step, the protein was concentrated using a 10,000 MWCO centrifugal concentrator.

Expression and purification of TssA subunits and domains. Expression and purification of His_6.$TssA2^A$ (1–478), His_6.$TssA2^A$ Nt2 (223–387), His_6.$TssA2^A$ Nt2-CTD (223–478) and $TssA2^A$ CTD (381–478) was carried out as previously described[33]. Both MBP.$TssA1^B$ CTD (303–373) and His_6.$TssA1^B$ Nt1 (1–255) were overexpressed and purified, with $TssA1^B$ CTD (303–373) being separated from the MBP solubility tag by Factor Xa cleavage[51]. MBP.$TssA1^B$ CTD (294–373) was produced and purified in a similar manner as MBP.$TssA1^B$ CTD (303–373). However, following Factor Xa cleavage of MBP.$TssA1^B$ CTD (294–373), $TssA1^B$ CTD oligomers were separated from the MBP solubility tag for EM analysis using a 100 kDa MWCO concentrator (Amicon) rather than SEC and the additional amylose column purification step was not included. Truncated $TssA1^B$ CTD derivatives were overproduced with an N-terminal MBP-His_6 tag in a similar manner to that described for MBP.$TssA1^B$ CTD (303–373) but were purified by IMAC on a HisTrap HP column with imidazole gradient elution rather than by amylose affinity chromatography. For crystallisation, MBP.His_6.$TssA1^B$ CTD H12–H14 (amino acids 303–358) was purified by amylose affinity chromatography, and eluted with 10 mM maltose. His_6.$TssA1^B$ CTD H12–H14 was proteolytically removed from the MBP tag by treatment with Factor Xa, followed by purification on a nickel column with imidazole gradient elution and SEC on a Superdex 200 column prior to concentration and final buffer exchange[51].

$TssA2^A$ Nt1-Nt2.His_6 (1–374), His_6.$TssA2^A$ CTD (381–478) and the $TssA1^B$ variants (Native $TssA1^B$ (1–373), His_6.$TssA1^B$ (1–373) and linkerHis_6.$TssA1^B$ (1–373)) were overproduced by inducing the T7 promoter present on the appropriate expression vector derivative in *E. coli* BL21(DE3) cells growing in BHI broth at 37 °C by addition of IPTG (final concentration 1 mM) to the culture at OD_{600} 0.5–0.7 and incubation was continued for a further 2–3 h (30 °C was used for inducing the linker His-tagged derivative). Harvested cells were resuspended in 50 mM Tris-Cl (pH 8.0), 100 mM NaCl (5 ml/g cell paste) and lysed by sonication. If the construct contained a His-tag, the lysate was cleared by centrifugation at 35,000 × *g* for 30 min. The clarified lysate was applied to nickel-sepharose (HisTrap HP, GE Healthcare Life Sciences) and eluted with a linear gradient of imidazole (10–500 mM).

For purification of native $TssA1^B$ (1–373), nucleic acids were removed from the clarified supernatant by precipitation with polyethyleneimine (0.15% (v/v)) at pH 8.0. After centrifugation at 40,000 *g* for 30 min at 4 °C, ammonium sulphate was added to the supernatant to a final concentration of 30% to precipitate TssA. The precipitated protein was recovered by centrifugation at 40,000 × *g* for 30 min at 4 °C and resuspended in a small volume of 50 mM Tris-Cl (pH 8.0) containing 50 mM NaCl, which was then applied onto a Superose 6 column equilibrated with the same buffer. SEC was performed at 0.5 ml/min flow-rate. $TssA1^B$ eluted from the column as a single protein peak at ~13.75 min corresponding to a size of ~1.9 MDa.

SEC. Size estimation of His_6.$TssA1^B$ Nt1 (1–255) was performed on a Superose 12 GL column (GE Healthcare) in 50 mM Tris-Cl (pH 8.0) containing 500 mM NaCl. The following standards were used (molecular weights in kDa): thyroglobulin, 669; apoferritin, 443; alcohol dehydrogenase, 149.5; conalbumin, 75; ovalbumin, 43; carbonic anhydrase, 29; ribonuclease, 13.7.

SEC-MALLS. For size exclusion chromatography-multiangle laser light scattering (SEC-MALLS), native $TssA1^B$ (1–373) and $TssA1^B$ CTD (294–373) were purified as described above and analysed on a Superose-6 10/300 column in 10 mM phosphate buffer (pH 7.4) containing 2.7 mM KCl and 137 mM NaCl, whilst purified His_6.$TssA2^A$ (4.5 mg/ml), His_6.$TssA2^A$ Nt2-CTD (6.8 mg/ml) and His_6.$TssA2^A$ CTD (6.0 mg/ml), were filtered (0.2 µm pore size) and analysed on an S200 10/300 column (GE Healthcare) in 50 mM Tris.HCl (pH 8.0), 500 mM NaCl. These were carried out at ~20 °C using a system comprising a Dawn Heleos-II multi-angle light scattering detector, an Optilab rEX refractive index detector and a QELS Plus module (all items by Wyatt Technology) linked to a Shimadzu HPLC system (SPD-20A UV detector, LC20-AD isocratic pump system, DGU-20A3 degasser and SIL-20A autosampler) at the Molecular Interactions Lab, University of York, UK. Peak areas chosen for analysis were defined by Astra software (Wyatt Technology).

Limited proteolysis. Twenty microlitres of purified His_6.$TssA2^A$ at 3.7 mg/ml in 16.7 mM Tris-HCl (pH 8.0), 175 mM NaCl was mixed with 1 µl of glutamyl endopeptidase (V8 protease or Glu-C, Sigma), glycyl endopeptidase (Gly-C, Protogen) and trypsin type I (Sigma) prepared at 1 mg/ml in water. The reaction was allowed to proceed for 90 min at room temperature and then stopped by addition of 20 µl 10% acetic acid. Digestion products were fractionated by 4–20% SDS-PAGE and blotted on to a PVDF membrane for N-terminal sequence analysis or analysed by mass spectrometry.

Mass spectrometry and N-terminal analysis. Mass spectrometry (MS) analysis of $TssA1^B$ and $TssA2^A$ cleavage products in ultrapure water was carried out on an electrospray Q-TOF mass spectrometer at the Astbury Centre for Structural Molecular Biology, University of Leeds, UK. Samples were diluted into 50% acetonitrile/0.1% formic acid, and analysed by Z-spray nanoelectrospray ionisation MS using a quadrupole-IMS-orthogonal time-of-flight MS (Synapt HDMS, Waters UK Ltd., Manchester, U.K.) and in-house fabricated gold/palladium coated nanospray capillaries. The MS was operated in positive TOF mode using a capillary voltage of 1.2 kV, cone voltage of 20 V, nanoelectrospray nitrogen gas pressure of 0.1 bar, backing pressure of 2.47 mbar and a trap bias of 4 V. The source and desolvation temperatures were set at 80 °C and 150 °C, respectively. During TOF-MS acquisition, argon was used as the buffer gas, at a pressure of 4.0×10^{-3} mbar in the trap and transfer regions. Mass calibration was performed by a separate injection of sodium iodide at a concentration of 2 µg/µl. Data processing was performed using MassLynx v4.1 software.

Sequencing of $TssA1^B$ and $TssA2^A$ cleavage products was performed using an ABI Procise Sequencer operating standard Edman degradation procedures. After electrophoresis, proteins were electroblotted from the gel on to PVDF membrane

(ABI ProBlott) using CAPS buffer at pH 11.0. Blotting was performed for 1–2 h at 450 mA with water cooling. Proteins were located on the membrane by staining with 1% Coomassie blue for 1 min, followed by destaining in 50% methanol. Transfer efficiency was verified by staining the gel after the transfer. The bands of interest were excised from the blot using a scalpel blade and transferred to a blot cartridge for sequence determination.

Electron microscopy (EM). 0.75% (w/v) uranyl formate was used as a negative stain. Carbon-coated grids were glow discharged for 30 s in a Cressington 208 carbon coater system. Purified protein samples (~5 µl of 5–50 µg/ml in 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10% glycerol (LinkerHis$_6$.TssA1^B (1–373), His$_6$.TssA2^A Nt2-CTD (223–478) and His$_6$.TssA2^A CTD (381–478)) or 20 mM Tris-HCl (pH 7.4), 200 mM NaCl, 1 mM EDTA (TssA1^B CTD (294–373))) were incubated on the grids for 1 min. Excess protein sample was removed by blotting and the grid was then washed twice in distilled water and once in negative stain (1 s for each wash) before staining for 20 s. After final blotting to remove excess stain, the grids were vacuum dried and imaged on Philips CM100 transmission electron microscope equipped with a Gatan Mulitiscan 794 1k × 1k charge-coupled device camera, at nominal magnification 28500 and defocus between 500 and 900 nm or Philips CM200 FEG transmission electron microscope using a Gatan UltraScan 890 (US4000SP) 2k × 2k charge-coupled device camera, at a nominal magnification of 66,000× and defocus between −200 and −800 nm. Apart from the initial screening of His$_6$.TssA2^A CTD and His$_6$.TssA2^A Nt2-CTD, all other images were captured on the CM200 FEG. Alignments and averaging for EM analysis were carried out using the IMAGIC-5 software package[52,53].

X-ray crystallography. Native data on crystals of purified His$_6$.TssA1^B Nt1 (1–255)[51] were collected on beamline I02 at the Diamond Light Source, and 150° of data were processed in xia2–3d[54–57] to 1.8 Å in spacegroup $P2_12_12$ (Supplementary Table 2). Crystals from the same conditions were subjected to sublimation of elemental iodine crystals overnight. Crystals were cryoprotected in 0.16 M calcium acetate, 0.08 M sodium cacodylate buffer (pH 6.5), 16.4% (w/v) PEG 8000 and 30% glycerol. Iodine SAD data were collected at 1.7 Å wavelength on beamline I03 at the Diamond Light Source, and data were processed in FastDP[58] to 2.04 Å in space-group $P2_12_12_1$ (Supplementary Table 2). The structure was solved using SHELX[59] alone and as part of FastEP based on the identification of seven potential iodine sites. The subsequent poly-alanine model was partially refined and used to solve the structure of the 1.8 Å native data. This model then underwent rounds of model building in Coot[60] and refinement using Refmac5[61,62] to produce the final model (Supplementary Table 2). The asymmetric unit contains one molecule of His$_6$.TssA1^B Nt1 (1–255), with eight N-terminal tag residues, 119 water molecules, two calcium ions and an ethylene glycol molecule.

Native data on crystals of purified TssA1^B CTD (303–373)[51] and the selenomethionine derivative were collected on beamline I24 at the Diamond Light Source. Data were processed in xia2–3da[54–56,63,64] (Supplementary Table 3) to 3.08 Å, in spacegroup I222. Selenomethionine peak SAD data were collected at 0.97922 Å wavelength on beamline I03 at the Diamond Light Source. Data were processed in xia2–3d[54–57] (Supplementary Table 3) to 3.03 Å, in spacegroup I222. The SHELX program suite, SHELXC, D and E[59], was used to identify 15 selenium sites and produce an initial electron density map. An initial model was built using Coot[60], Buccaneer[65] and employing a structure of a shorter Bc TssA1^B CTD derivative comprising residues 303–358 which had been solved separately, as a guide. This model was used as input to Phenix[66,67], and subsequent output was subjected to rounds of building in Coot[60], NCS averaging and solvent flattening in DM[68]. Refinement to the native data at 3.08 Å was performed using Refmac5[61,62]. Between 4 and 6 residues of the protein remain unmodelled at the C-terminus of each chain.

Purified His$_6$.TssA1^B CTD H12–H14 (303–358), together with 12 additional residues from the tag, was set down in sitting-drop crystallisation trials and produced crystals in conditions 0.18 M tri-ammonium citrate, 20% (w/v) PEG 3350, which were cryoprotected in 0.2 M di-ammonium citrate, 22% (w/v) PEG 3350, 25% (v/v) ethylene glycol. Data were collected on beamline I02 at the Diamond Light Source, and processed in xia2 3dii[54–57,63,64] to 2.35 Å in spacegroup P6 (Supplementary Table 4). The structure was solved in spacegroup $P6_2$ by molecular replacement, in Phaser[69], using the previously solved structure of TssA1^B CTD containing helices α12 and α13. The model was built and refined to native data at 2.35 Å in Coot[60] and Refmac[61,62], with two molecules in the asymmetric unit.

The structure of a TssA1^B CTD fragment corresponding to residues 303–347 was determined as follows. TssA1^B CTD (303–373), together with 4 additional residues from the tag, also produced crystals in condition 0.1 M citric acid (pH 5.0), 20% (w/v) PEG 6000 which were cryoprotected in 0.1 M citric acid (pH 5.0), 22% (w/v) PEG 6000, 25% (v/v) ethylene glycol. Native data were collected on beamline I04–1 at the Diamond Light Source, and processed in xia2 3daii[54–57,63,64] to 1.78 Å in spacegroup I222 (Supplementary Table 5). Selenomethionine-incorporated MBP.TssA1^B CTD (303–373) was overexpressed and purified using a similar strategy to that described previously[51]. Crystals were produced in condition 0.1 M citric acid (pH 5.0), 20% (w/v) PEG 6000, and were cryoprotected in 0.1 M citric acid (pH 4.0), 22% (w/v) PEG 6000, 25% (v/v) glycerol. Selenomethionine MAD peak, inflection and high energy remote data were collected on beamline I03 at the Diamond Light Source and data were processed in xia2 3da or 3daii[54–57,63,64] (Supplementary Table 5). The structure was solved by MAD using the SHELX program suite[59], followed by building in Coot[60] and refinement in Refmac5[61,62] to the native data at 1.78 Å. Analysis of the electron density map strongly suggested that the protein had been clipped after W347.

To generate a derivative for crystals produced from purified His$_6$.TssA2^A Nt2 (223–387)[33], flakes of ethyl mercury phosphate powder were added to a drop containing crystals for 5 min, before subsequent cryoprotection in 0.04 M potassium phosphate monobasic, 16% (w/v) PEG 8000, 30% (v/v) glycerol. Hg SAD data were collected at 1.0088 Å wavelength on beamline I03 at the Diamond Light Source, and data were processed in xia2-DIALS[54–58] to 2.16 Å in spacegroup $P2_1$ (Supplementary Table 6). The structure of mercury bound Nt2 was solved using SHELX[59] based on the identification of one mercury site for each of the two monomers in the asymmetric unit and built using Buccaneer and Coot[60,65]. Refinement was performed using Refmac5[61,62]

Purified TssA2^A Nt1-Nt2.His$_6$ (1–374) containing a C-terminal LEHHHHHH tag crystallised in 0.2 M sodium chloride, 0.1 M HEPES buffer (pH 7.0), and 20% (w/v) PEG 6000. Crystals were cryoprotected in 0.2 M sodium chloride, 0.1 M HEPES buffer (pH 7.0), and 20% (w/v) PEG 6000 and 25% ethylene glycol. Native data were collected on beamline I04 at the Diamond Light Source, and data were processed in xia2-DIALS[54–58] to 1.76 Å in spacegroup $P2_1$. The structure of these crystals was solved by molecular replacement[69] using the coordinates of the mercury bound TssA2^A Nt2 dimer as a search model. The model was iteratively built using Coot and refined to 1.8 Å using Refmac5[61,62] to produce the final model (Supplementary Table 6). The asymmetric unit contains 4 molecules of TssA2^A Nt1-Nt2.His$_6$ (231–374), and 380 water molecules. Despite the fact that the construct covered both the Nt1 and Nt2 domains the purified protein had undergone proteolytic cleavage leading to the loss of the Nt1 domain. The model therefore contains the entire Nt2 domain (residues 231–374).

Native data on crystals of His$_6$.TssA2^A Nt2-CTD (223–478)[33] were collected on beamline I03 at the Diamond Light Source. Data were processed in xia2-DIALS[54–56,58] to 3.13 Å, in spacegroup $P2_1$ (Supplementary Table 7). Molecular replacement[69] using a search model of the TssA2^A Nt2 dimer identified the position of five copies of the TssA2^A Nt2 dimer in the asymmetric unit with additional density being seen for 10 copies of the CTD. An initial model was built using Coot[60] and subsequently was subjected to rounds of re-building[60], and NCS averaging with solvent flattening in Parrot[70]. Refinement to the native data at 3.13 Å was performed using Refmac5[61,62] (Supplementary Table 7). Representative electron density for each deposited model is shown in Supplementary Fig. 10. All structures were validated using Molprobity[71], with average B-factors calculated using Baverage[72]. Structural alignment was carried out with LSQKAB using secondary structure matching using Cα of the main chain[34]. PISA was used to calculate interface areas between molecules[32]. Structural homologues were discovered using the Dali structural recognition server[35]. Diagrams shown in this paper were generated using the PyMOL Molecular Graphics System, (Schrödinger, LLC).

Bioinformatics. Phylogenetic analysis was carried out using MEGA 6[73] and sequence alignment utilised Clustal Omega[74]. ESpript 3 was used for secondary structure mapping of conserved residues[75].

Data availability

The X-ray crystal structures of TssA1^B Nt1, TssA1^B CTD, His$_6$.TssA1^B CTD H12–H14, TssA1^B CTD fragment (303–347), TssA2^A Nt2, and TssA2^A Nt2-CTD and associated data have been deposited in the Protein Data Bank under the ID codes 6HS5 https://www.ebi.ac.uk/pdbe/entry/pdb/6hs5, 6HS6 https://www.ebi.ac.uk/pdbe/entry/pdb/6hs6, 6H8E https://www.ebi.ac.uk/pdbe/entry/pdb/6h8e, 6H8F https://www.ebi.ac.uk/pdbe/entry/pdb/6h8f, 6G7B https://www.ebi.ac.uk/pdbe/entry/pdb/6g7b and 6G7C https://www.ebi.ac.uk/pdbe/entry/pdb/6g7c, respectively. Other data that support the findings of the study are available from the corresponding author(s) upon request.

Received: 11 May 2018 Accepted: 23 October 2018
Published online: 12 November 2018

References

1. Basler, M., Pilhofer, M., Henderson, G. P., Jensen, G. J. & Mekalanos, J. J. Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**, 182–186 (2012).
2. Ge, P. et al. Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. *Nat. Struct. Mol. Biol.* **22**, 377–382 (2015).
3. Leiman, P. G. et al. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl Acad. Sci. USA* **106**, 4154–4159 (2009).
4. Kanamaru, S. et al. Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**, 553–557 (2002).

5. Shneider, M. M. et al. PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* **500**, 350–353 (2013).
6. Shalom, G., Shaw, J. G. & Thomas, M. S. In vivo expression technology identifies a type VI secretion system locus in Burkholderia pseudomallei that is induced upon invasion of macrophages. *Microbiology* **153**, 2689–2699 (2007).
7. Zheng, J. & Leung, K. Y. Dissection of a type VI secretion system in Edwardsiella tarda. *Mol. Microbiol.* **66**, 1192–1206 (2007).
8. Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. & Attree, I. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genom.* **10**, 104 (2009).
9. Cianfanelli, F. R. et al. VgrG and PAAR Proteins Define Distinct Versions of a Functional Type VI Secretion System. *PLoS Pathog.* **12**, 1–27 (2016).
10. Durand, E. et al. Biogenesis and structure of a type VI secretion membrane core complex. *Nature* **523**, 555–560 (2015).
11. Aschtgen, M. S., Bernard, C. S., De Bentzmann, S., Lloubes, R. & Cascales, E. SciN is an outer membrane lipoprotein required for Type VI secretion in Enteroaggregative Escherichia coli. *J. Bacteriol.* **190**, 7523–7531 (2008).
12. Ma, L. S., Lin, J. S. & Lai, E. M. An IcmF family protein, ImpLM, is an integral inner membrane protein interacting with ImpKL, and its Walker a motif is required for type VI secretion system-mediated Hcp secretion in Agrobacterium tumefaciens. *J. Bacteriol.* **191**, 4316–4329 (2009).
13. Aschtgen, M. S., Gavioli, M., Dessen, A., Lloubès, R. & Cascales, E. The SciZ protein anchors the enteroaggregative Escherichia coli Type VI secretion system to the cell wall. *Mol. Microbiol.* **75**, 886–899 (2010).
14. Aschtgen, M.-S., Zoued, A., Lloubès, R., Journet, L. & Cascales, E. The C-tail anchored TssL subunit, an essential protein of the enteroaggregative Escherichia coli Sci-1 Type VI secretion system, is inserted by YidC. *Microbiologyopen* **1**, 71–82 (2012).
15. Durand, E. et al. Structural characterization and oligomerization of the TssL protein, a component shared by bacterial type VI and type IVb secretion systems. *J. Biol. Chem.* **287**, 14157–14168 (2012).
16. Brunet, Y. R., Hénin, J., Celia, H. & Cascales, E. Type VI secretion and bacteriophage tail tubes share a common assembly pathway. *EMBO Rep.* **15**, 315–321 (2014).
17. Pukatzki, S., Ma, A. T., Revel, A. T., Sturtevant, D. & Mekalanos, J. J. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc. Natl Acad. Sci.* **104**, 15508–15513 (2007).
18. Bönemann, G., Pietrosiuk, A., Diemand, A., Zentgraf, H. & Mogk, A. Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *EMBO J.* **28**, 315–325 (2009).
19. Lossi, N. S. et al. The HsiB1C1 (TssB-TssC) complex of the Pseudomonas aeruginosa type VI secretion system forms a bacteriophage tail sheathlike structure. *J. Biol. Chem.* **288**, 7536–7548 (2013).
20. Kube, S. et al. Structure of the VipA/B type VI secretion complex suggests a contraction-state-specific recycling mechanism. *Cell Rep.* **8**, 20–30 (2014).
21. Wang, J. et al. Cryo-EM structure of the extended type VI secretion system sheath-tube complex. *Nat. Microbiol.* **2**, 1507–1512 (2017).
22. Brunet, Y. R., Zoued, A., Boyer, F., Douzi, B. & Cascales, E. The Type VI secretion TssEFGK-VgrG phage-like baseplate is recruited to the TssJLM membrane complex via multiple contacts and serves as assembly platform for tail tube/sheath polymerization. *PLoS Genet.* **11**, 1–21 (2015).
23. Taylor, N. M. I. et al. Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **533**, 346–352 (2016).
24. English, G., Byron, O., Cianfanelli, F. R., Prescott, A. R. & Coulthurst, S. J. Biochemical analysis of TssK, a core component of the bacterial Type VI secretion system, reveals distinct oligomeric states of TssK and identifies a TssK–TssFG subcomplex. *Biochem. J.* **461**, 291–304 (2014).
25. Kapitein, N. et al. ClpV recycles VipA/VipB tubules and prevents non-productive tubule formation to ensure efficient type VI protein secretion. *Mol. Microbiol.* **87**, 1013–1028 (2013).
26. Finn, R. D. et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).
27. Zoued, A. et al. Priming and polymerization of a bacterial contractile tail structure. *Nature* **531**, 59–63 (2016).
28. Planamente, S. et al. TssA forms a gp6-like ring attached to the type VI secretion sheath. *EMBO J.* **35**, 1613–1627 (2016).
29. Zoued, A. et al. TssA: The cap protein of the Type VI secretion system tail. *Bioessays* **39**, 1600262 (2017).
30. Zheng, J., Ho, B. & Mekalanos, J. J. Genetic Analysis of Anti-Amoebae and Anti-Bacterial Activities of the Type VI Secretion System in Vibrio cholerae. *PLoS ONE* **6**, e23876 (2011).
31. Aubert, D. F., Hu, S. & Valvano, M. A. Quantification of type VI secretion system activity in macrophages infected with Burkholderia cenocepacia. *Microbiology* **161**, 2161–2173 (2015).
32. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
33. Dix, S. R. et al. TssA from Aeromonas hydrophila: expression, purification and crystallographic studies. *Acta Crystallogr. Sect. F. Struct. Biol. Commun.* **74**, 578–582 (2018).
34. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sect. A* **32**, 922–923 (1976).
35. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
36. Vettiger, A., Winter, J., Lin, L. & Basler, M. The type VI secretion system sheath assembles at the end distal from the membrane anchor. *Nat. Commun.* **8**, 1–9 (2017).
37. Nazarov, S. et al. Cryo-EM reconstruction of Type VI secretion system baseplate and sheath distal end. *EMBO J.* **37**, e97103 (2018).
38. Yonekura, K. et al. The bacterial flagellar cap as the rotary promoter of flagellin self-assembly. *Science* **290**, 2148–2152 (2000).
39. Yonekura, K., Maki-Yonekura, S. & Namba, K. Growth mechanism of the bacterial flagellar filament. *Res. Microbiol.* **153**, 191–197 (2002).
40. Renault et al. The gp27-like Hub of VgrG Serves as Adaptor to Promote Hcp Tube Assembly. *J. Mol. Biol.* **430**, 3143–3156 (2018).
41. Song, W. S., Cho, S. Y., Hong, H. J., Park, S. C. & Yoon, S. Self-oligomerizing structure of the Flagellar cap protein FliD and its implication in filament assembly. *J. Mol. Biol.* **429**, 847–857 (2017).
42. Yonekura, K., Maki-Yonekura, S. & Namba, K. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424**, 643–650 (2003).
43. Miller, J. Experiments in Molecular Genetics. (Cold Spring Harb. Lab. Press, Cold Spring Harb, NY, 1972).
44. Lennox, E. S. Transduction of linked genetic characters of the host by bacteriophage P1. *Virology* **1**, 190–206 (1955).
45. Sokol, P. A., Ohman, D. E. & Iglewski, B. H. A more sensitive plate assay for detection of protease production by Pseudomonas aeruginosa. *J. Clin. Microbiol.* **9**, 538–540 (1979).
46. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).
47. Herrero, M., De Lorenzo, V. & Timmis, K. N. Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. *J. Bacteriol.* **172**, 6557–6567 (1990).
48. de Lorenzo, V. & Timmis, K. N. Analysis and construction of stable phenotypes in gram-negative bacteria with Tn5- and Tn10-derived minitransposons. *Methods Enzymol.* **235**, 386–405 (1994).
49. Barrett et al. Genetic Tools for Allelic Replacement in Burkholderia Species. *Appl. Environ. Microbiol.* **74**, 4498–4508 (2008).
50. Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl Acad. Sci.* **95**, 5752–5756 (1998).
51. Owen, H. J. et al. TssA from Burkholderia cenocepacia: expression, purification, crystallization and crystallographic analysis. *Acta Crystallogr. Sect. F. Struct. Biol. Commun.* **74**, 536–542 (2018).
52. van Heel, M. & Keegstra, W. IMAGIC: a fast, flexible and friendly image analysis software system. *Ultramicroscopy* **7**, 113–129 (1981).
53. Van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
54. Nonaka, T. The CCP4 suite-computer programs for protein crystallography. *Nihon Kessho Gakkaishi* **36**, 224–227 (1994).
55. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **62**, 72–82 (2006).
56. Kabsch, W. Xds. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 125–132 (2010).
57. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
58. Winter, G. & McAuley, K. E. Automated data collection for macromolecular crystallography. *Methods* **55**, 81–93 (2011).
59. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. Sect. A Found. Crystallogr.* **64**, 112–122 (2008).
60. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 486–501 (2010).
61. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **67**, 355–367 (2011).
62. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **53**, 240–255 (1997).
63. Sauter, N. K., Grosse-Kunstleve, R. W. & Adams, P. D. Robust indexing for automatic data collection. *J. Appl. Crystallogr.* **37**, 399–409 (2004).
64. Zhang, Z., Sauter, N. K., Van Den Bedem, H., Snell, G. & Deacon, A. M. Automated diffraction image analysis and spot searching for high-throughput crystal screening. *J. Appl. Crystallogr.* **39**, 112–119 (2006).

65. Cowtan, K. Fitting molecular fragments into electron density. *Acta Crystallogr. Sect. D.-Biol. Crystallogr.* **64**, 83–89 (2008).
66. Adams, P. D. et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, (213–221 (2010).
67. Echols, N. et al. Graphical tools for macromolecular crystallography in PHENIX. *J. Appl. Crystallogr.* **45**, 581–586 (2012).
68. Cowtan, K. 'Dm': an automated procedure for phase improvement by density modification. *Jt. CCP4 ESF-EACBM Newsl. Protein Crystallogr.* **31**, 34–38 (1994).
69. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
70. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 470–478 (2010).
71. Chen, V. B. et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 12–21 (2010).
72. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D.-Biol. Crystallogr.* **67**, 235–242 (2011).
73. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
74. Sievers, F. & Higgins, D. G. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.* **1079**, 105–116 (2014).
75. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

Acknowledgements

We would like to acknowledge James Ault, Andrew Leech and Arthur Moir for carrying out the mass spectrometry, SEC-MALLS and protein sequencing analysis, and for helpful discussions. We would like to thank the Diamond Light Source for access to beamlines I02, I03, I04, I04–1 and I24 that contributed to these results, and the beamline scientists for their assistance with data collection. This work was supported by BBSRC Doctoral Training Grants BB/J014443/1 and BB/F016840/1 awarded to H.J.O. and H.L.S., respectively, a scholarship from the Libyan Ministry of Higher Education awarded to A. A. and a University of Sheffield scholarship awarded to S.R.D. H.J.O was also supported by Newton Fund grants R/151699 and R/152968.

Author contributions

M.S.T. conceived the investigation. S.R.D., H.J.O., R.S., A.A., S.S., H.L.S., S.B.T. and M.S.T. designed the experiments; S.R.D., H.J.O., R.S., A.A., S.S., M.J.H., S.L.B., H.L.S., D.J.M., S.B.T., T.A.B. and S.E.S. performed the experiments; S.R.D., H.J.O., R.S., A.A., S.S., H.L.S., S.B.T., P.J.B., P.A.B., D.W.R and M.S.T. interpreted the data and S.R.D., H.J.O., R.S., H.L.S., D.W.R. and M.S.T. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-07247-1.

Competing interests: The authors declare no competing interests.

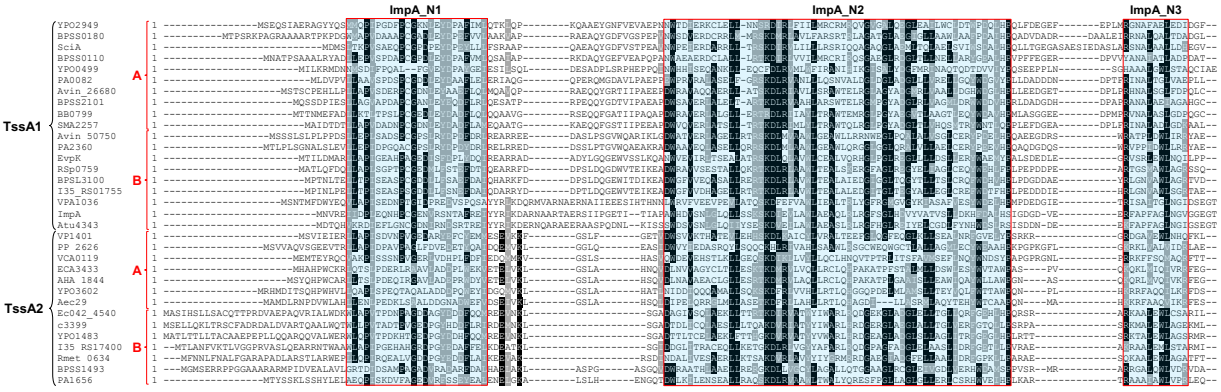
Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

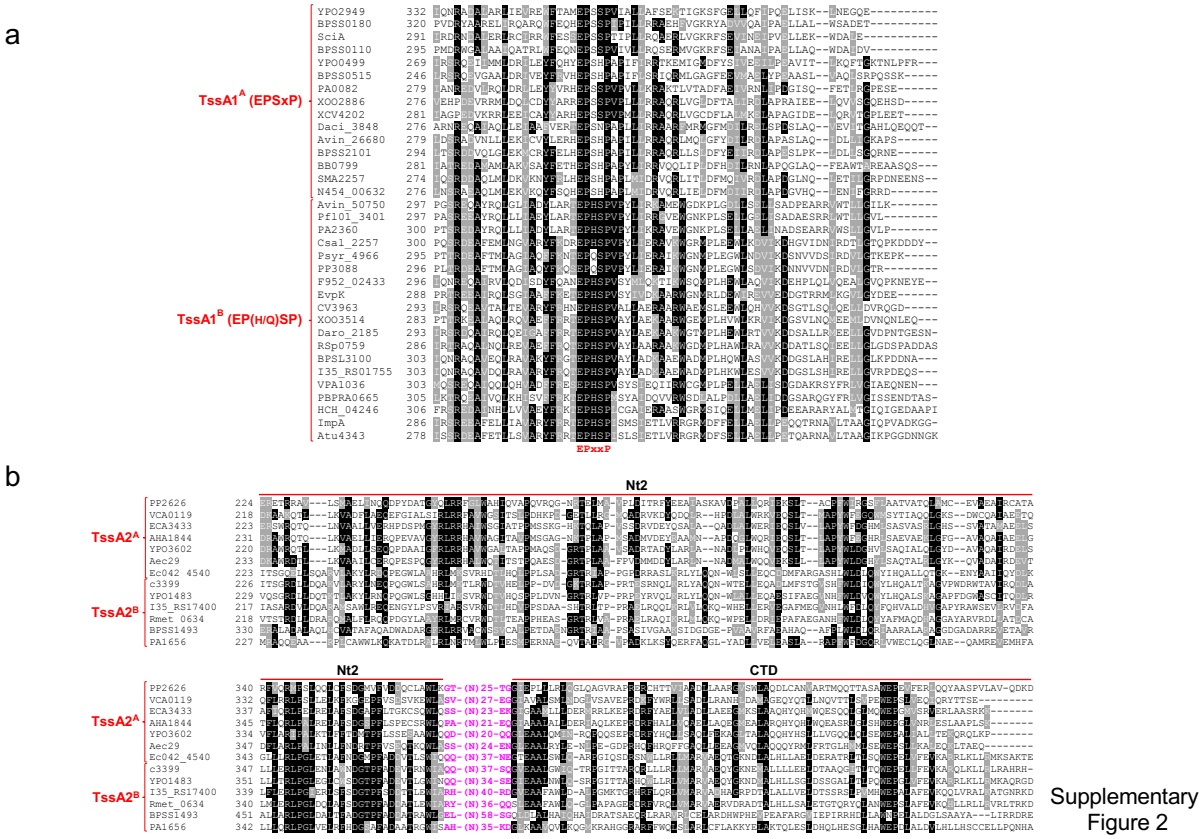
Supplementary Information

Structural insights into the function of type VI secretion system TssA subunits

Dix *et al*.

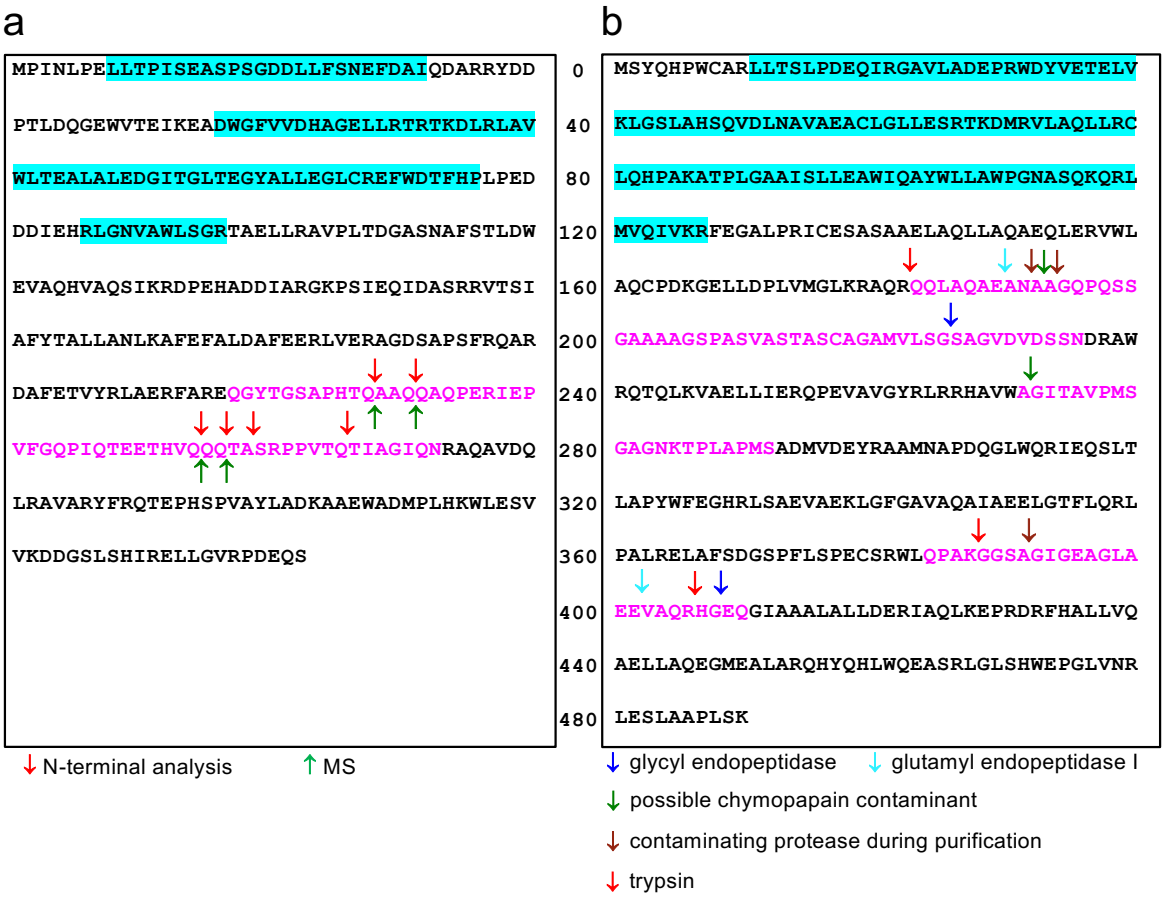


Supplementary Figure 1. Alignment of the ImpA_N region of TssA subunits. The N-terminal region of representative TssA1 and TssA2 subunits, corresponding to M1-E136 of *B. cenocepacia* TssA1^{B} (I35_RS01755), were aligned using Clustal Omega[1] and conserved residues were highlighted using Boxshade. Amino acids in white font with black shading are identical at the corresponding position in ≥ 50% of sequences aligned, whereas those highlighted in grey are similar. The three conserved motifs (ImpA_N1-ImpA_N3) that define the ImpA_N region are enclosed in red boxes with blue coloured fill. Members of the A and B sub-clades of each TssA family are indicated on the left of the alignment in red font. Gaps have been removed between non-alignable amino acids N-terminal to the ImpA_N1 motif to accommodate TssA sequences in a single line.



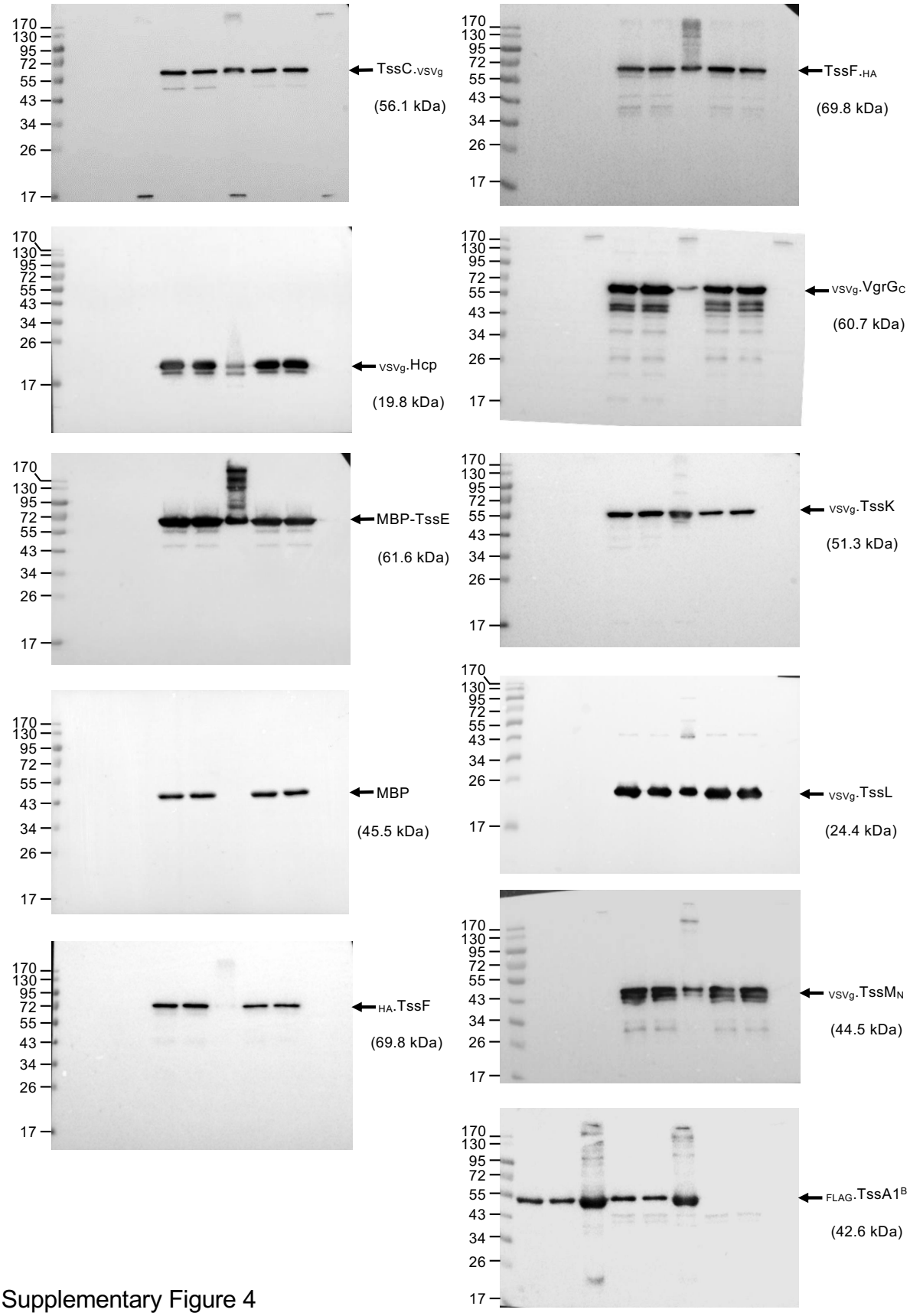
Supplementary Figure 2

Supplementary Figure 2. Alignment of C-terminal amino acid sequences of TssA1 and TssA2 subunits. Amino acid sequences of the C-terminal regions of representative TssA subunits were aligned using Clustal Omega[1] and conserved residues were highlighted using Boxshade. Amino acids in white font with black shading are identical at the corresponding position in ≥ 50% of sequences aligned, whereas those highlighted in grey are similar. **(a)** Alignment of TssA1 C-terminal domains. The CTDs of representative TssA1^{A} and TssA1^{B} orthologues are shown along with the last three residues of the predicted interdomain linker. The location of the conserved EPxxP motif is shown below the alignment. (**b**) Alignment of TssA2 Nt2 and C-terminal domains. Representative members of the A and B sub-clades, based on a phylogenetic analysis of full-length TssA2 subunits, are shown. Sequences shown in magenta are non-conserved and are predicted to constitute an interdomain linker connecting Nt2 to the CTD (the number of linker amino acids that are not shown is indicated). For clarity, the amino acid sequences at the extreme C-terminus of eight orthologues are not shown.



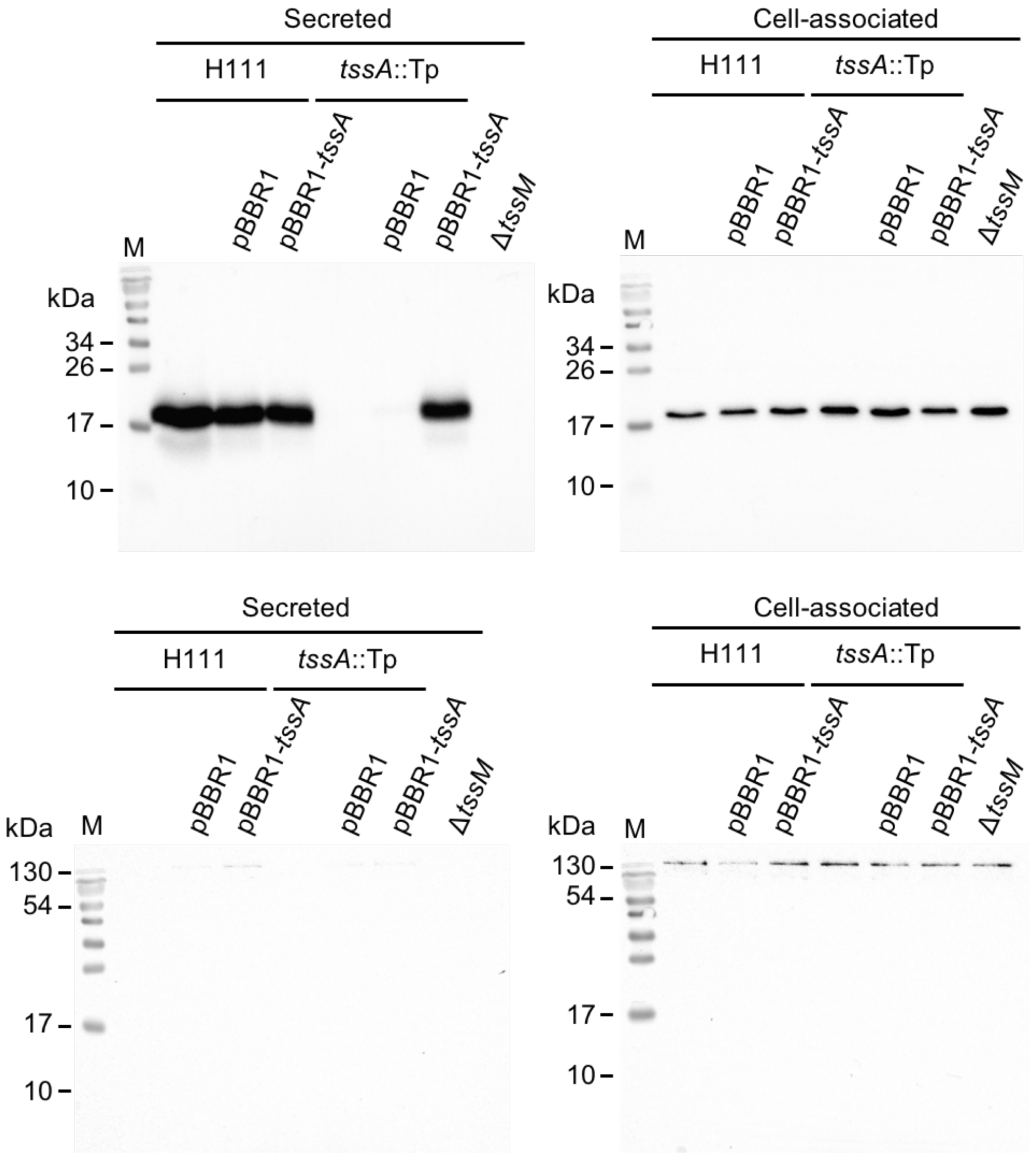
Supplementary Figure 3

Supplementary Figure 3. Domain analysis of TssA1^{B} and TssA2^{A}. Sequences of Bc TssA1^{B} (**a**) and Ah TssA2^{A} (**b**) with interdomain linkers shown in magenta font. (**a**) Adventitious protease cleavage sites utilised in interdomain linker identification are indicated by up and down arrows, for mass spectrometry and N-terminal analysis, respectively. (**b**) Arrows indicate protease specific cut sites within Ah TssA2^{A} identified by mass spectrometry that were used to identify the interdomain linker. (**c**) Analysis of large proteolytic degradation products of Bc TssA1^{B}. 12% SDS-PAGE analysis of His_6.TssA1^{B} following IMAC purification, dialysis and concentration, and application to a Q sepharose column in 20 mM Tris-HCl (pH 8.0) with elution by a NaCl step gradient, revealed two major species (~31 kDa and ~34 kDa), that corresponded to polypeptides of 31.3, 31.5, 34.2 and 34.4 kDa by mass spectrometry. Protein molecular weight markers are shown in kDa. (**d**) Analysis of small proteolytic degradation products of Bc TssA1^{B}. 15% SDS-PAGE analysis of His_6.TssA1^{B} degradation products before application to the Q column (left hand gel) and following elution (right hand gel), identified a series of low molecular weight peptides ~8, 10, 12, 15 and 23 kDa, these peptides were subjected to N-terminal amino acid sequence analysis. Protein molecular weight markers are shown in kDa. (**e**) Analysis of Ah TssA2^{A} proteolytic digestion products. SDS-PAGE analysis of His_6.TssA2^{A} following proteolytic digestion with contaminating protease (lane 2), glutamyl endopeptidase (Glu-C) (lane 3), glycyl endopeptidase (Gly-C) (lane 4), and trypsin (lane 5). Protein molecular weight markers are shown in kDa.

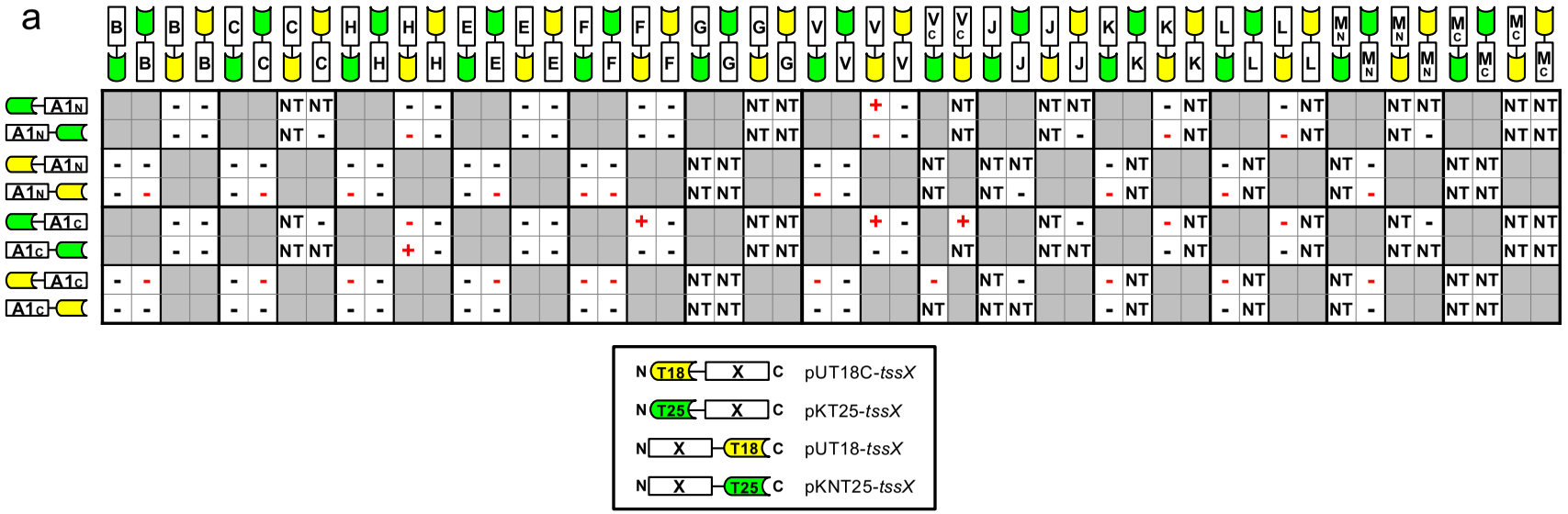


Supplementary Figure 4

Supplementary Figure 4. Co-immunoprecipitation analysis of FLAG-tagged Bc TssA1^{B} with other T6SS subunits. Uncropped images of western blots presented in **Fig. 2c** are shown. Molecular weights of the detected T6SS subunits (and control MBP) are indicated (which include any epitope, affinity or solubility tags) and markers are indicated in kDa. See legend to **Fig. 2** for more details.

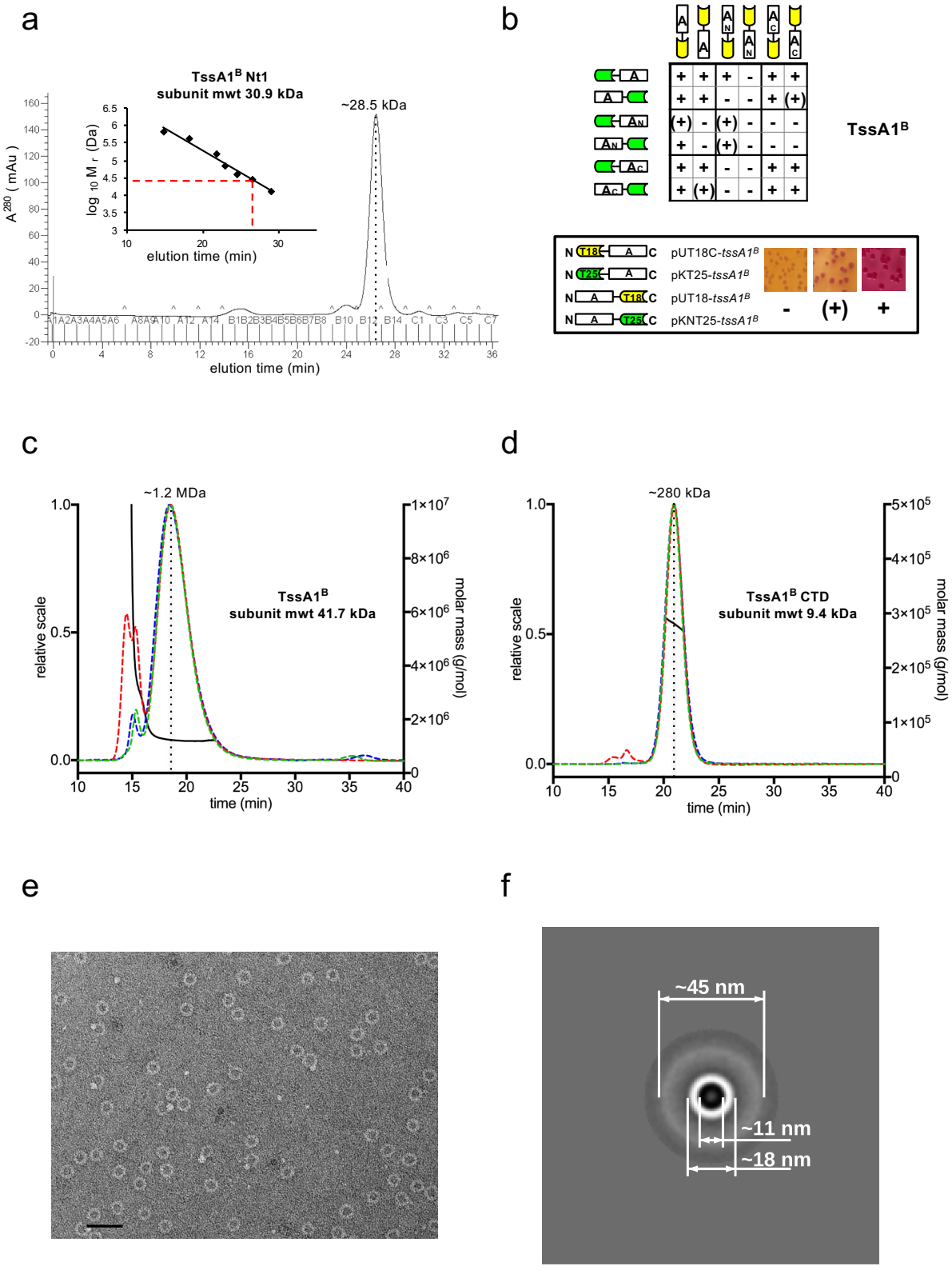


Supplementary Figure 5. Effect of TssA1^{B} inactivation on T6SS secretion activity. Uncropped images of western blots presented in **Fig. 3** are shown. See legend to **Fig. 3** for more details.



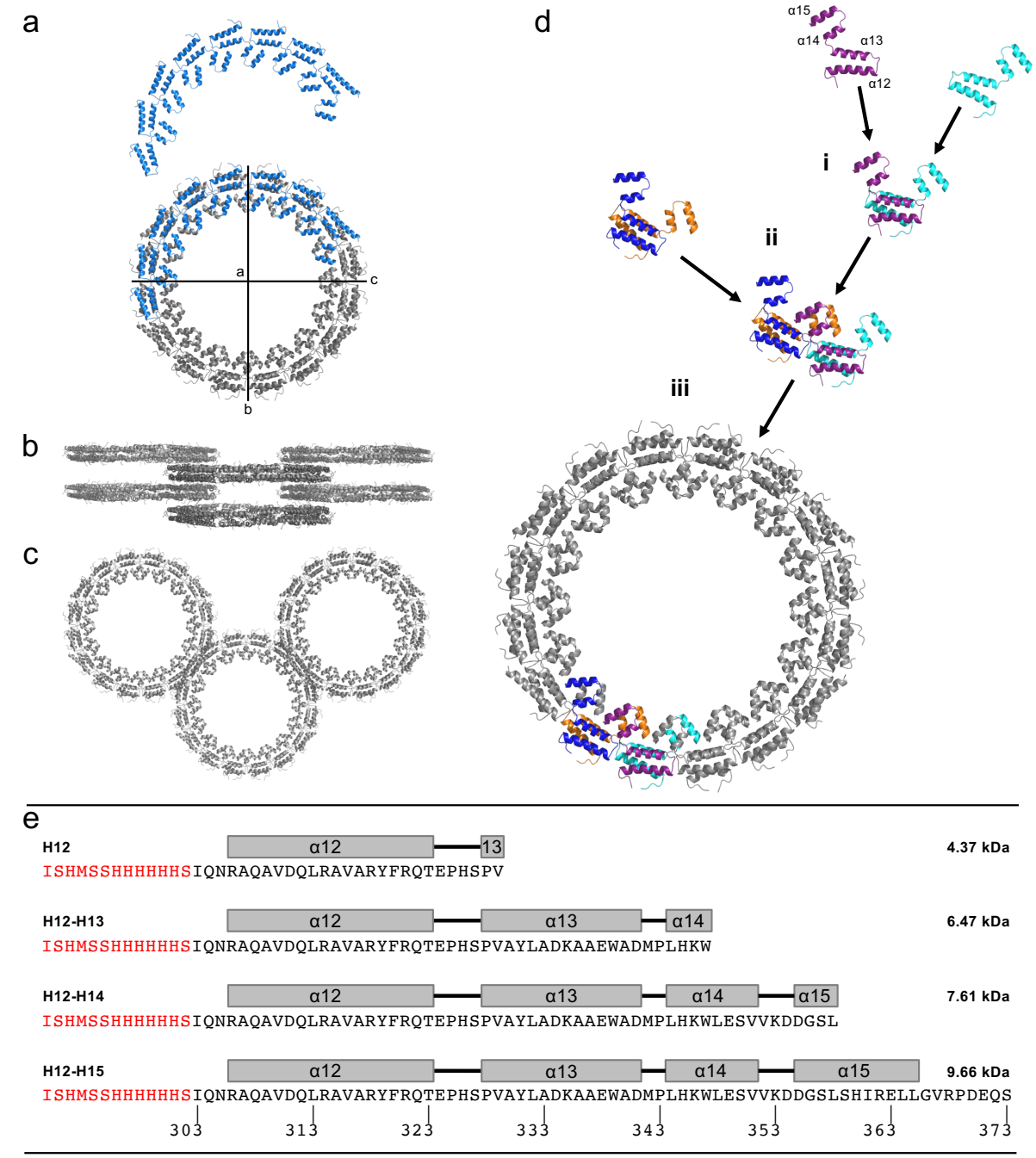
Supplementary Figure 6

Supplementary Figure 6. Two-hybrid analysis of TssA1^B domain interactions. (**a**) Each hybrid protein is represented by a green or yellow coloured motif indicating the CyaA fragment (T25 and T18, respectively) linked to a white rectangle labelled according to the T6SS subunit or TssA1^B domain. T6SS subunits are indicated by a single letter corresponding to the suffix used in the Tss nomenclature (i.e. B corresponds to TssB) except for H (Hcp), V (VgrG) and V_C (the VgrG core region, although only two hybrid proteins were constructed in this case). $A1_N$ and $A1_C$ correspond to TssA1^B NTD (Nt1) and CTD, respectively. Interactions of the cytoplasmic N- and periplasmic C-terminal regions of TssM (M_N and M_C, respectively) were investigated separately. Compatible pairwise combinations of BACTH plasmids encoding the indicated hybrid proteins were introduced into *E. coli* BTH101 and the resulting Mal phenotypes were scored on maltose-MacConkey agar. Colonies were recorded as +, positive or -, negative after 120 h incubation at 30 °C. Combinations involving TssG and $TssM_C$ that gave rise to a maltose-negative phenotype in combination with TssA1^B were not tested (NT). In most of the other cases only representative combinations were tested. Results shown in red font are from plasmid combinations that were quantitated by β-galactosidase assay and shown in (b). (**b**) β-galactosidase activities in *E. coli* BTH101 containing representative pairwise combinations of TssA1^B domain and T6SS subunit hybrid proteins were assayed. A_N, TssA1^B Nt1 (N-terminal 256 aa); A_C, TssA1^B CTD (C-terminal 80 aa); Z represents the Zip positive control. Otherwise, nomenclature is as in (a). Combinations involving TssG, TssJ and $TssM_C$ were not assayed as they do not appear to interact with TssA1^B. Data is representative of three independent experiments (n = 3) performed in duplicate and values correspond to the mean ± standard deviation. Values presented in **Supplementary Data 1**.



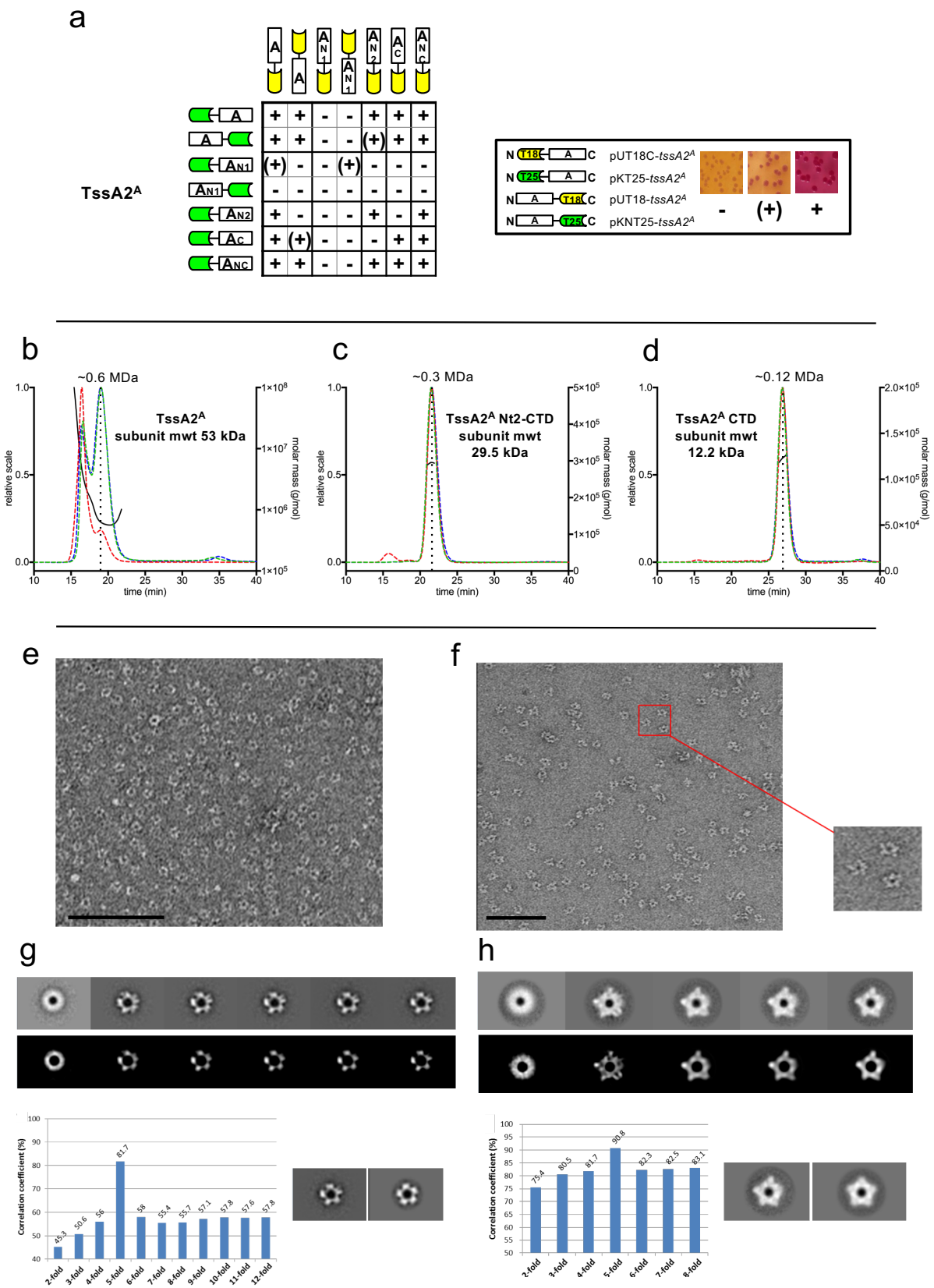
Supplementary Figure 7

Supplementary Figure 7. TssA1^{B} self-associates into a high molecular weight ring-like oligomer. (**a**) SEC profile of the Bc His_6.TssA1^{B} Nt1 domain indicating a monomer of 30.6 kDa is present. (**b**) Self-interaction analysis of Bc TssA1^{B} and its domains. Each hybrid protein is represented by a green or yellow coloured motif representing the CyaA fragment (T25 and T18, respectively) linked to a white rectangle labelled according to the fused region of TssA1^{B} (see key). All four possible hybrid proteins were constructed and analysed for TssA1^{B} and its domains in pairwise combinations (N, N-terminal domain (Nt1); C, C-terminal domain (CTD)). Colonies were recorded as +, positive; (+), weak positive; or -, negative (see key). (**c**) SEC-MALLS elution profile, peak positions, and molar mass plot of TssA1^{B} (~1.2 MDa). The normalised signals for light scattering (LS) are shown in red, refractive index (RI) in blue, UV absorption (280 nm) in green, and molecular weight estimate curve is shown in black. The molecular weight was determined by a Zimm fit procedure at each point. The X-axis is the elution time in minutes. (**d**) SEC-MALLS elution profile, peak positions, and molar mass plot of TssA1^{B} CTD (~280 kDa). Details as for (c). (**e**) Negative stain EM of single particle views of TssA1^{B} CTD. Scale bar is 50 nm. (**f**) Negative stain EM average of TssA1^{B}. Measurements indicate the inner diameter of the CTD ring (110 Å), the CTD ring outer diameter (180 Å), and diameter of the whole particle (450 Å).



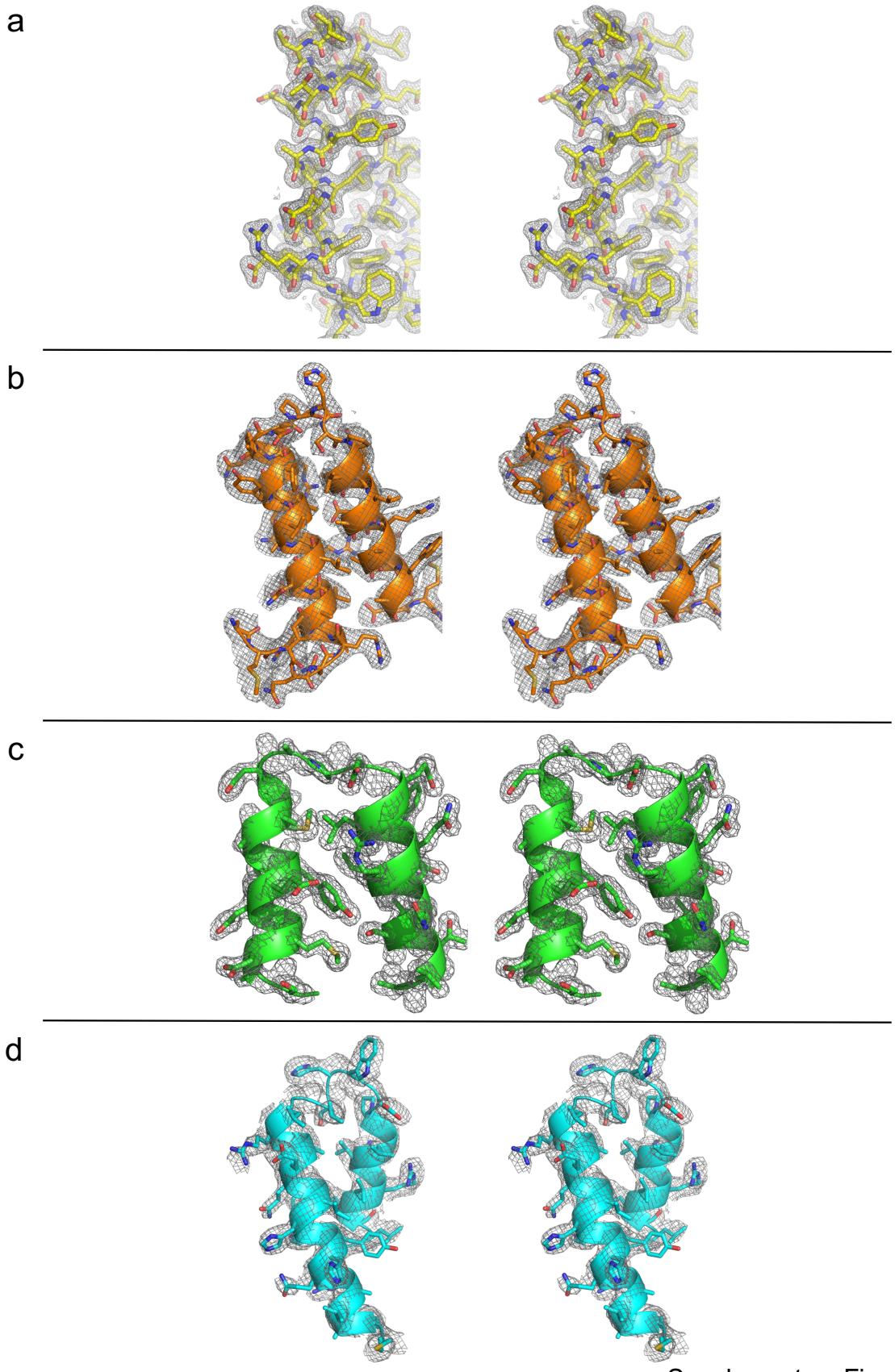
Supplementary Figure 8

Supplementary Figure 8. Assembly of the TssA1^{B} CTD ring oligomer. (a) Formation of the TssA1^{B} (I35_RS01755) CTD ring from the 8 subunits of the asymmetric unit. The contents of the asymmetric unit (blue) are related to the remaining sections of the ring (grey) by the 2-fold crystallographic symmetry axes around a, b and c. View is down the crystallographic axis a. **(b)** and **(c)** Packing of TssA1^{B} CTD ring oligomers in the crystal lattice. (b) View perpendicular to the 16-fold axis, (c) view parallel to the 16-fold axis. **(d)** Proposed assembly pathway for the TssA1^{B} CTD ring oligomer. (i) Two monomers (purple and cyan) dimerise, via helices α12 and α13; (ii) two dimers (purple/cyan and orange/blue) assemble with the majority of interactions occurring between helices α14 and α15 (purple and orange); (iii) sequential addition of further dimers (or higher order oligomers) results in assembly of a ring containing 32 subunits that exhibits D_{16} symmetry. **(e)** Nested deletion of TssA1^{B} CTD, containing an N-terminal MBP.His_6 tag. The sequences shown are released upon Factor Xa cleavage, with remaining His_6-tag residues highlighted in red. Construct H12-H15 corresponds to full-length TssA1^{B} CTD and was shown to form ring oligomers by EM, whereas the truncated derivatives form lower molecular weight complexes. **(f)** SEC profile of the H12-H14 construct on Superdex 200 (following removal of the MBP tag) to show failure to assemble ring oligomers. The elution peak is seen to shift from a high molecular weight structure ~355 kDa for selenomethionine-incorporated TssA1^{B} CTD[2], to a much lower molecular weight of ~30 kDa for the H12-H14 construct. SEC also showed that the MBP-tagged H12 construct formed monomers while MBP-tagged H12-H13 formed dimers. **(g)** Crystal structure of H12-H14 determined to 2.35 Å showing dimer formation. Residues 303-358 and 303-354 are resolved in chains A (wheat) and B (teal), respectively. **(h)** Crystal structure showing dimer formation by a truncated TssA1^{B} CTD derivative containing residues 303-347 (corresponding to helices α12 and α13) that arose from adventitious clipping of TssA1^{B} CTD following its release from MBP.TssA1^{B} CTD (303-373). Structure determined to 1.78 Å resolution.



Supplementary Figure 9

Supplementary Figure 9. TssA2^{A} self-associates into a high molecular weight ring-like oligomer. (**a**) Self-interaction analysis of Ah TssA2^{A} and its domains. Hybrid proteins are represented as in Suppl. Fig. 6b All possible hybrid proteins were constructed and analysed for TssA2^{A} and its domains (N1, N-terminal domain (Nt1); N2, middle domain (Nt2); C, C-terminal domain (CTD); NC, Nt2-CTD region). (**b**)-(**d**) SEC-MALLS elution profile, peak positions, and molar mass plot of TssA2^{A}, TssA2^{A} Nt2-CTD and TssA2^{A} CTD, respectively. Details as for Supplementary Fig. 7c. (**e**) Negative stain EM of Ah TssA2^{A} CTD single particles. Scale bar is 100 nm. (**f**) Negative stain EM of Ah TssA2^{A} Nt2-CTD single particles, with a sample of individual particles selected in the magnified insert. Scale bar is 100 nm. (**g**)-(**h**) Class averaging and correlation co-efficient calculation for Ah TssA2^{A} CTD and Ah TssA2^{A} Nt2-CTD particles, respectively. The x-axis shows the specific symmetry that was tested, and the y-axis shows the correlation coefficient (%) of all the single particles rotated at a specific angle according to the fold of symmetry, then correlated to their unrotated self.



Supplementary Figure 10

Supplementary Figure 10. Electron density maps for TssA1^{B} Nt1, TssA1^{B} CTD, TssA2^{A} Nt2 and TssA2^{A} CTD. (**a**) Electron density visible for residues K72-P115 of helices α4-α5 of TssA1^{B} Nt1. (**b**) Electron density visible for tag residues H301-M302 and residues I303-W347 of helices α12-α13 of TssA1^{B} CTD. (**c**) Electron density visible for residues M284-P315 of helices α3-α4 of the TssA2^{A} Nt2 structure**.** (**d**) Electron density visible for residues M438-A474 of helices α11-α12 of the TssA2^{A} CTD structure. All 2Fo-Fc maps contoured to 1.0 σ RMSD.

Supplementary Table 1. N-terminal amino acid sequence of low mol. wt. His_6.$TssA1^B$ degradation products

Peptide mol. wt.[a]	N-terminal sequence
23 kDa	G-S-S-H-H-H-H-H-[b] H-A-D-X-X-X-
15 kDa	A-A-Q-Q-A-Q-P-E Q-A-Q-P-E-X-I-E-
12 kDa	K-R-D-P-E-H-A-D-D- R-D-P-E-H-A-D-D-
10 kDa	Q-Q-T-A-S-R-P- T-A-S-R-P-P-V- S-R-P-P-V-T-
8 kDa	T-I-A-G-I-Q-N-

[a]Sizes were estimated based on migration in SDS PA gels.
[b]Corresponds to the N-terminus of His_6.$TssA1^B$ following removal of the N-terminal methionine.

Supplementary Table 2. Data collection and refinement statistics for His_6.$TssA1^B$ Nt1 (1-255)

	His_6.$TssA1^B$ Nt1 (1-255) Native	His_6.$TssA1^B$ Nt1 (1-255) Peak - Iodine
PDB code	**6HS5**	
Data collection[a]		
Wavelength (Å)	1.04434	1.70000
Space group	$P2_12_12$	$P2_12_12$
Cell dimensions		
a, *b*, *c* (Å)	49.7, 125.7, 45.6	49.8, 125.5, 45.6
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	49.67-1.80 (1.85-1.80)	29.64-2.04 (2.09-2.04)
R_{merge}	0.043 (0.344)	0.169 (0.865)
I / σI	21.2 (4.4)	19.3 (2.8)
Completeness (%)	99.8 (99.0)	98.1 (74.2)
Redundancy	5.4 (5.3)	23.0 (15.8)
No. total reflections	146307 (10493)	427376 (15999)
No. unique reflections	27244 (1966)	18554 (1011)
Refinement		
Resolution range (Å)	46.2-1.80	
No. reflections (Work/Free)	25817/1381	
R_{work} / R_{free}	0.177/0.226	
No. atoms		
Protein	1991	
Ligand/ion[b]	6	
Water	119	
B-factors (Å^2)		
Protein	30.6	
Ligand/ion[b]	22.5/20.5	
Water	29.4	
R.m.s. deviations		
Bond lengths (Å)	0.0157	
Bond angles (°)	1.5665	
Ramachandran favoured (%)[c]	98.79	
Ramachandran outliers (%)[c]	0.00	

The number of crystals for the His_6.$TssA1^B$ Nt1 (1-255) structure is 1 per data set
[a]Values in parentheses are for highest-resolution shell.
[b]1 x ethylene glycol/2 x calcium ion
[c]Calculated using MolProbity[3]

Supplementary Table 3. Data collection and refinement statistics for TssA1^{B} CTD (303-373)

	TssA1^{B} CTD (303-373) Native	TssA1^{B} CTD (303-373) Peak - Selenomethionine
PDB code	**6HS6**	
Data collection[a]		
Wavelength (Å)	1.70001	0.97922
Space group	I222	I222
Cell dimensions		
a, *b*, *c* (Å)	46.3, 201.7, 263.7	46.0, 200.3, 263.0
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	41.49-3.08 (3.16-3.08)	100.12-3.03 (3.11-3.03)
R_{merge}	0.131 (0.786)	0.220 (2.144)
I / σI	17.5 (3.7)	9.8 (1.4)
Completeness (%)	99.9 (100.0)	99.9 (100.0)
Redundancy	12.6 (11.9)	12.9 (13.4)
No. total reflections	297122 (20394)	315386 (23629)
No. unique reflections	23602 (1709)	24385 (1759)
Refinement		
Resolution range (Å)	41.49-3.08	
No. reflections (Work/Free)	22387/1215	
R_{work} / R_{free}	0.198/0.242	
No. atoms		
Protein	4401	
Ligand/ion	0	
Water	0	
B-factors (Å^2)		
Protein	70.2	
Ligand/ion	-	
Water	-	
R.m.s. deviations		
Bond lengths (Å)	0.0114	
Bond angles (°)	1.6210	
Ramachandran favoured (%)[b]	96.05	
Ramachandran outliers (%)[b]	0.00	

The number of crystals for the TssA1^{B} CTD (303-373) structure is 1 per data set
[a]Values in parentheses are for highest-resolution shell.
[b]Calculated using MolProbity[3]

Supplementary Table 4. Data collection and refinement statistics for His_6.$TssA1^B$ CTD H12-H14 (303-358)

	His_6.$TssA1^B$ CTD H12-H14 (303-358)
PDB code	**6H8E**
Data collection[a]	
Wavelength (Å)	0.97943
Space group	$P6_2$
Cell dimensions	
a, *b*, *c* (Å)	65.3, 65.3, 66.1
α, β, γ (°)	90.0, 90.0, 120.0
Resolution range (Å)	42.97-2.35 (2.41-2.35)
R_{merge}	0.094 (0.713)
$I / \sigma I$	17.1 (3.1)
Completeness (%)	99.9 (99.4)
Redundancy	8.5 (8.1)
No. total reflections	57299 (4002)
No. unique reflections	6747 (492)
Refinement	
Resolution range (Å)	42.97-2.35
No. reflections (Work/Free)	6372/357
R_{work} / R_{free}	0.193/0.237
No. atoms	
Protein	897
Ligand/ion	0
Water	0
B-factors (Å^2)	
Protein	45.1
Ligand/ion	-
Water	-
R.m.s. deviations	
Bond lengths (Å)	0.0129
Bond angles (°)	1.4320
Ramachandran favoured (%)[b]	99.06
Ramachandran outliers (%)[b]	0.00

The number of crystals for the His_6.$TssA1^B$ CTD (303-358) structure is 1
[a]Values in parentheses are for highest-resolution shell.
[b]Calculated using MolProbity[3]

Supplementary Table 5. Data collection, phasing and refinement statistics for TssA1^{B} CTD (303-373) fragment

	TssA1^{B} CTD (303-373) fragment		TssA1^{B} CTD (303-373) fragment - Selenomethionine	
PDB code	**6H8F**			
Data collection[a]				
		Peak	*Inflection*	*Remote*
Wavelength (Å)	0.92000	0.97934	0.97949	0.96411
Space group	I222	I222	I222	I222
Cell dimensions				
a, *b*, *c* (Å)	48.3, 62.7, 65.7	47.8, 62.6, 65.6	47.8, 62.8, 65.8	47.7, 62.9, 65.7
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	38.90-1.78 (1.83-1.78)	16.79-1.71 (1.75-1.71)	38.64-1.87 (1.92-1.87)	38.61-2.04 (2.09-2.04)
R_{merge}	0.068 (0.697)	0.059 (0.512)	0.065 (0.739)	0.072 (0.703)
I / σI	13.2 (2.0)	21.4 (2.7)	22.0 (2.9)	21.7 (4.0)
Completeness (%)	97.0 (99.3)	97.9 (83.9)	100.0 (100.0)	100.0 (100.0)
Redundancy	4.5 (4.6)	11.2 (5.6)	12.5 (9.6)	12.7 (13.1)
No. total reflections	42707 (3238)	120101 (3596)	105843 (5954)	83404 (6305)
No. unique reflections	9502 (700)	10692 (644)	8487 (618)	6577 (482)
Refinement				
Resolution (Å)	38.90-1.78			
No. reflections (Work/Free)	9048/454			
R_{work} / R_{free}	0.189/0.237			
No. atoms				
Protein	752			
Ligand/ion	0			
Water	39			
B-factors (Å^2)				
Protein	27.6			
Ligand/ion	-			
Water	27.6			
R.m.s deviations				
Bond lengths (Å)	0.0117			
Bond angles (°)	1.3164			
Ramachandran favoured (%)[b]	100.00			
Ramachandran outliers (%)[b]	0.00			

The number of crystals for the TssA1^{B} CTD (303-373) fragment structure is 1 per data set
[a]Values in parentheses are for highest-resolution shell.
[b]Calculated using MolProbity[3]

Supplementary Table 6. Data collection and refinement statistics for TssA2^{A} Nt1-Nt2.His$_6$ (1-374) and TssA2^{A} His$_6$.Nt2 (231-374)

	TssA2^{A} Nt1-Nt2.His$_6$ (1-374)	TssA2^{A} His$_6$.Nt2 (231-374) - Mercury
PDB code	**6G7B**	
Data collection[a]		
Wavelength (Å)	0.9795	1.0088
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	39.8, 101.3, 78.4	45.8, 40.0, 100.5
α, β, γ (°)	90.0, 104.6, 90.0	90.0, 102.5, 90.0
Resolution range (Å)	42.13-1.76 (1.79-1.76)	44.71-2.16 (2.23-2.16)
R_{merge}	0.065 (0.611)	0.162 (1.492)
I / σI	7 (1.2)	9.4 (1.2)
Completeness (%)	99.9 (98.7)	95.9 (71.9)
Redundancy	3.6 (3.3)	12.4 (8.6)
No. total reflections	214716 (9711)	231080 (8786)
No. unique reflections	59629 (2971)	18671 (1026)
Refinement		
Resolution range (Å)	101.32-1.76	
No. reflections (Work/Free)	56717/2805	
R_{work} / R_{free}	0.272/0.306	
No. atoms		
Protein	4584	
Ligand/ion	0	
Water	380	
B-factors (Å^2)		
Protein	22.9	
Ligand/ion	-	
Water	29.3	
R.m.s. deviations		
Bond lengths (Å)	0.0175	
Bond angles (°)	1.702	
Ramachandran favoured (%)[b]	98.8	
Ramachandran outliers (%)[b]	0.00	

The number of crystals for the TssA2^{A} Nt1-Nt2.His$_6$ (1-374) structure is 1 per data set
[a]Values in parentheses are for highest-resolution shell.
[b]Calculated using MolProbity[3]

Supplementary Table 7. Data collection and refinement statistics for TssA2^{A} His$_6$.Nt2-CTD (223-478)

	TssA2^{A} His$_6$.Nt2-CTD (223-478)
PDB code	**6G7C**
Data collection[a]	
Wavelength (Å)	0.97623
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	70.0, 202.1, 139.0
α, β, γ (°)	90.0, 90.7, 90.0
Resolution range (Å)	139.00-3.13 (3.21-3.13)
R_{merge}	0.113 (1.368)
I / σI	5.8 (0.9)
Completeness (%)	99.9 (100.0)
Redundancy	3.7 (3.6)
No. total reflections	253133 (18358)
No. unique reflections	67977 (5030)
Refinement	
Resolution range (Å)	202.08-3.13
No. reflections (Work/Free)	64474/3380
R_{work} / R_{free}	0.255/0.328
No. atoms	
Protein	18363
Ligand/ion	0
Water	0
B-factors (Å^2)	
Protein	99.3
Ligand/ion	-
Water	-
R.m.s. deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.60
Ramachandran favoured (%)[b]	90.5
Ramachandran outliers (%)[b]	1.35

The number of crystals for the TssA2^{A} His$_6$.Nt2-CTD (223-478) structure is 1
[a]Values in parentheses are for highest-resolution shell.
[b]Calculated using MolProbity[3]

Supplementary Table 8. Bacterial strains

Strain	Genotype/Description[a]	Source or reference
B. cenocepacia		
H111	CF isolate, prototroph	(4)
H111ΔtssM	H111 containing an in-frame deletion within *tssM*	This study
H111-tssA::Tp	H111 containing a Tp^R cassette insertion within *tssA*	This study
A. hydrophila		
ATCC 7966	Wild type. Prototroph.	(5)
E. coli		
JM83	F^- *ara* Δ(*lac-proAB*) *rpsL* ϕ80d*lacZΔM15* (Sm^R) F^- *recA1 endA1 gyrA96 thi-1 hsdR17* ($r_K^- m_K^+$)	(6)
DH5α	*supE44 relA1 Δ(lacZYA-argF)U169 deoR phoA* λ^- ϕ80d*lacZΔM15*	(7)
S17-1(λpir)	*thi proA hsdR recA* RP4-2-*tet*::Mu-1 *kan*::Tn*7* integrant λpir (Tp^R Sm^R)	(8)
SM10(λpir)	*thi-1 thr leu tonA lacY supE recA* RP4-2-Tc::Mu λpir (Km^R)	(8)
BTH101	F^-, *cya-99 araD139 galE15 galK16 rpsL1* (Sm^R) *hsdR2 mcrA1 mcrB1*	(9)
BL21(DE3)	F^- *ompT* $hsdS_B$ ($r_B^- m_B^+$) *dcm gal* λDE3	(10)
ER2523	*fhuA2* [*lon*] *ompT gal sulA11 (mcr-73*::mini-Tn*10)2* (Tc^S) [*dcm*] *zgb-210*::Tn*10* (Tc^S) *endA1 Δ(mcrC-mrr)114*::IS*10*	New England Biolabs

[a]Km^R, kanamycin resistant; Rf^R, rifampicin resistant; Sm^R, streptomycin resistant; Tc^S, tetracycline sensitive; Tp^R, trimethoprim resistant.

Supplementary Table 9. Oligonucleotides

Oligo/Primer name	Sequence (5'-3')[a,b,c]
Vector modification	
delacO1for	GATCACGGGATCTCGACGCTCTCCCTATTAGGAAATTAATACG
delacO1for2	ACTCACTATAGGGACCTGTAGAAATAATTTTGTTTAACTTTAATA AGGAGATATAC
delacO1rev	CATGGTATATCTCCTTATTAAAGTTAAACAAAATTATTTCTACAG GTCC
delacO1rev2	CTATAGTGAGTCGTATTAATTTCCTAATAGGGAGAGCGTCGAGA TCCCGT
delacO2for	GTACACGGCCGCATAATCGAAATTAATACGACT
delacO2for2	CACTATAGGGACCATCTTAGTATATTAGTTAAGTATAAGAAGGA GATATACA
delacO2rev	TATGTATATCTCCTTCTTATACTTAACTAATAT
delacO2rev2	ACTAAGATGGTCCCTATAGTGAGTCGTATTAATTTCGATTATGCG GCCGT
pMALHisFor	GCGCCAATTGACCAACAAGGACCATAGATTATGAAAATC*CATCACCAT CACCATCAC*GAAGAAGGTAAACTGG TAATCT
pMALHisRev	TGGCGGGTTCGGCAGCAGATCT
Construction of *tssA1*^{B} mutant	
iotAfor	GCGCAAGCTTCACGCGACATCTCATGCATC
iotAfor3	GCGCAAGCTTCCGCGGAAATTCGACGCGATCCAGGAC
iotArev	GCGCGGATCCCCCGTCTTGCGTTTACGACT
iotArev2	ATCACGAAGAGCATTCCGCC
Construction of *tssM* mutant	
tssMfor	GCGCTCTAGAGGAACCTGAACGTCCTATGC
tssMrev2	GCGCGGTACCTCATTGCGCCCTCTGTGCAT
tssM-OPfor	TCATCCCGTTTGACAGCATG
tssM-OPrev2	ATCTTGCCGAAGTAGGCGATTT
Subunit expression for purification	
TssA1^{B}	

tssAnonHisFor	GCGCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGC CGATCAATCTCCCCGA
pET14b-iotAfor	GCGCCATATGCCGATCAATCTCCCCGA
pET14b-iotArev	GCGCGGATCCTGCGTTTACGACTGCTCGTC
TssAlinkerHis1for	GCG*CATCACCATCACCATCAC*GCGCAGCAGGCCCAGCCCGA
TssAlinkerHis1rev	CGC*GTGATGGTGATGGTGATG*CGCCTGCGTGTGCGGCGCGC
pET14b-ACTDfor	GCGCCATATGCCTCCGGTGACGCAGACGAT
pACYC-tssA.CTDfor	GCGCCATATGAGCAGC*CATCATCATCATCATCAC*AGCATCCAGAACC GTGCGCAGGC
pACYC-tssA.CTDH1rev2	GCGCGGATCCTTAGACCGGGCTGTGCGGCTCGG
pACYC-tssA.CTDH2rev2	GCGCGGATCCTTACCACTTGTGCAGCGGCATGT
pACYC-tssA.CTDH3rev2	GCGCGGATCCTTACAGCGAGCCGTCGTCCTTCA
TssA2[A]	
pET21a-AHA1844.Nt1Nt2.for	ATATATACATATGAGCTATCAACACCCCTGGTGTGCAC
pET21a-AHA1844.Nt1Nt2.rev	TATATTATATACTCGAGCAGCCAGCGGCTGCA
pACYC-AHA1844.CTDSD2for	GCGCGGATCCGAGTGCGGGCATTGGCGAGGC
pACYC-AHA1844.rev	GCGCAAGCTTTCATTTCGACAACGGCGCCG
TssD	
pET-His-iotDfor2	TATGCGAAGGGCCATGTTACAT
pET14b-iotDrev	GCGCGGATCCTTAGACCGCGTAGGTCTT
Subunit expression for co-IP	
TssA.NdeI.FLAG.for	GCGCCATATG*GATTATAAAGACGACGATGATAAA*CCGATCAATCTCCC CGAGCT
TssA.BglII.rev	GCGCAGATCTTGCGTTTACGACTGCTCGTC
tssC.NcoI.for	GCGCCCATGGCAAACCAGCAAACGGCTGC
tssC.VSVg.BamHI.rev	GCGCGGATCCTTA*CTTACCCAGGCGGTTCATTTCGATATCAGTGTA*TG AATTTGCCGGTTTGG
tssD.forpACYC.NtermVSVgTag	GCGCCCATGGCA*TACACTGATATCGAAATGAACCGCCTGGGTAAG*AT GTTACATATGCACTTGCAGTTTGGTAGT
tssD. BamHI.rev	GCGCGGATCCTTAGACCGCGTAGGTCTTGT

pMALtssEfor	GCGCCATATGAAACGATTCGAACCCAG
pMALtssErev	GCGCGGATCCCGGCTTGATTTACTGCATGC
tssF.NcoI.for	GCGCCCATGGAAGAATTGCTGCCGTATTA
tssF.HA.HindIII.rev	GCGCAAGCTTCTA*AGCGTAGTCTGGGACGTCGTATGGGTA*CGCCAG GATCGATTCGC
tssF.NdeI.HA.for	GCGCCATATG*TACCCATACGACGTCCCAGACTACGCT*GAAGAATTGC TGCCGTATTA
C-term-iotFrev	GCGCGGTACCATCGACTACGCCAGGATCGA
tssI.BspHI.VSVgtag. For	GCGCTCATGAGC*TACACTGATATCGAAATGAACCGCCTGGGTAAG*AT GGATGCGCACAGCATGAT
BCAM0148gp5.BglII. rev	GCGCAGATCTTTAAAACTCCTGGGACCGGTAGC
TssKforpET.Nterm VSVgTag	GCGCCCATGGCA*TACACTGATATCGAAATGAACCGCCTGGGTAAG*AT GAGTTATTCGGCCAAGGT
TssK.BglII.rev	GCGCAGATCTTCATGATGTGACCGCGATCA
TssLforpET.Nterm VSVgTag	GCGCCCATGGCA*TACACTGATATCGAAATGAACCGCCTGGGTAAG*AT GAGCTACGCGCCTTCCTT
TssL.Rev	GCGCAGATCTTTACTTCAAGCGGTGCGCGAT
TssM1.VSVg.NcoI.for	GCGCCCATGGCA*TACACTGATATCGAAATGAACCGCCTGGGTAAG*GT GAAGCGTGCGAACCAGCA
TssM1.HindIII.rev	GCGCAAGCTTTTAGAAGGTCGCGTAGCGCAGCC
Two-hybrid system	
TssA1[B]	
Cterm-iotAfor	GCGCCTGCAGGGATGCCGATCAATCTCCCCGA
Cterm-iotArev	GCGCTCTAGATGCGTTTACGACTGCTCGTC
Nterm-iotAfor	GCGCAAGCTTGATGCCGATCAATCTCCCCGA
N-iotAfullrev	GCGCGGATCCGACGACTGCTCGTCGGGCCG
C-iotAlongNTDrev	GCGCTCTAGAGGTTAGCTGCCGGTATAGCCCTGTT
N-iotAlongNTDrev	GCGCGGATCCTCGCTGCCGGTATAGCCCTGTT
C-term-iotAfor	GCGCCTGCAGGGATGCCGATCAATCTCCCCGA
C-iotAshortCTDfor	GCGCCTGCAGGTCCTCCGGTGACGCAGACGAT
N-iotAshortCTDfor	GCGCAAGCTTTCCTCCGGTGACGCAGACGAT
TssA2[A]	
AHA1844fullfor	GCGCTCTAGAGATGAGCTATCAACACCCCT
AHA1844-Crev2	GCGCGGTACCGGCATTCATTTCGACAACGG
AHA1844-Nrev	GCGCGGTACCCATTTCGACAACGGCGCCGC

AHA1844NTDrev	GCGCGGTACCTTAGCTGGAACTGTCGACGTCGACG
AHA1844NTDrev2	GCGCGGTACCGTGGAACTGTCGACGTCGACG
AHA1844for	GCGCTCTAGACGTCGACGTCGACAGTTCCA
AHA1844CTDrev	GCGCGGTACCTTAGGCCTCGCCAATGCCCGCACT
AHA1844for2	GCGCTCTAGAAAGTGCGGGCATTGGCGAGGC
Other T6SS subunits	
Cterm-iotBfor	GCGCTCTAGACGTCGAGATGGCCAAGAAAGA
Cterm-iotBrev	GCGCGAATTCTTCGTCATTCGCCGTCTTTGT
Nterm-iotBfor	GCGCGCATGCCACCGCAACTGACGCGATAAC
N-iotBfullrev	GCGCTCTAGAGATTCGCCGTCTTTGTCG
Cterm-iotCfor	GCGCCTGCAGAGATGAACCAGCAAACGGCT
Cterm-iotCrev	GCGCTCTAGACTCGCTTATGAATTTGCCGG
Nterm-iotCfor	GCGCAAGCTTGCAGATGAACCAGCAAACGGCT
Nterm-tssCrev.new	GCGCGGATCCTTGAATTTGCCGGTTTGGGCA
Cterm-tssCfor.new	GCGCCTGCAGAATGAACCAGCAAACGGCT
pUT18C-iotDfor	GCGCCTGCAGGAGTCGAAGGGCCATGTTACA
Cterm-iotDrev	GCGCTCTAGAGTTTAGACCGCGTAGGTCTTGT
Nterm-iotDfor	GCGCAAGCTTGCACTTGCAGTTTGGTAGTCC
Nterm-iotDrev	GCGCGGATCCGCGTAGGTCTTGTCGTTCTT
Cterm-iotEfor	GCGCTCTAGAGGCGATGAAACGATTCGAACCC
Cterm-iotErev	GCGCGAATTCTCGGCTTGATTTACTGCATGC
Nterm-iotEfor	GCGCAAGCTTGACGATGAAACGATTCGAACCC
N-iotEfullrev	GCGCGGATCCGACTGCATGCGTGC CGC GCG T
Cterm-iotFfor	GCGCTCTAGAGACGATGGAAGAATTGCTGCCG
Cterm-iotFrev	GCGCGGTACCATCGACTACGCCAGGATCGAT
Nterm-iotFfor	GCGCAAGCTTGCAGATGGAAGAATTGCTGCCGT
N-iotFfullrev	GCGCTCTAGAGACGCCAGGATCGATTCGCC
C-iotGfrag1for	GCGCCTCTAGAGATGCAAGCCCCGAACCGGCG
Cterm-iotGrev	GCGCGAATTCTTGGGTCTCAGTGAATCACGTG
N-iotGfrag1for	GCGCAAGCTTGATGCAAGCCCCGAACCGGCG
Nterm-iotGrev2	GCGCGGATCCTCGTGAATCACGTGCAGTTCGTA
TssIfor	GCGCTCTAGAGATGGATGCGCACAGCATGAT
C-tssIrev	GCGCCCCGGGAGATCTAGTCTATGCTTCCGACGAAG
N-tssIrev	GCGCCCCGGGAAAGATCTCCTGCTTCCGACGAAGGACTGA
Rev.gp5	GCGCAGATCTTTAAAACTCCTGGGACCGGTAG
For1tssj	GCGCCTGCAGAACTTCTGGCCGGATGCGCGGC

C-tssJrev	GCGCGGATCCTTAACCGCAGGAGACCGAAG
For2tssj	GCGCCTGCAGACTTCTGGCCGGATGCGCGGC
N-tssJrev	GCGCGGATCCAAACCGCAGGAGACCGAAGACA
tssKfor	GCGCGGATCCAATGAGTTATTCGGCCAAGGT
C-tssKrev	GCGCGGTACCTCATGATGTGACCGCGATCA
N-tssKrev	GGTACCAATGATGTGACCGCGATCAGTT
tssLfor	CGCGTCTAGAAATGAGCTACGCGCCTTCCTT
C-tssLrev	GCGCGGTACCTTACTTCAAGCGGTGCGCGATCT
N-tssLrev	GCGCGGTACCACTTCAAGCGGTGCGCGATCT
tssM1for	CGCTCTAGAAGTGAAGCGTGCGAACCAGCA
C-tssM1rev	CGCGGTACCTTAGAAGAAGGTCGCGTAGCGCA
N-tssM1rev	CGCGGTACCAAAAGAAGGTCGCGTAGCGCAG
tssM2for	CGCGGATCCCACGATCGGCAACCAGCAGCT
C-tssM2rev	GCGGGTACCGTCATTGCGCCCTCTGTGCA
N-tssM2rev	GCGGGTACCAATTGCGCCCTCTGTGCATTCG

[a]Sequences specifying restriction endonuclease cleavage sites are underlined.

[b]Sequences specifying affinity or epitope tags are shown in italic font.

[c]Stop codons are shown in red font.

Supplementary Table 10. Plasmids

Plasmid	Description[a]	Source or reference
pBBR1MCS	Mobilisable BHR cloning vector (Cm^R)	(11)
pBBR1MCS-TssA1^B	Complementation plasmid. Constructed by amplifying *B. cenocepacia tssA* with primers iotAfor and iotArev, restricting with *Hin*dIII and *Bam*HI, and ligating the amplicon to the same sites of pBBR1MCS.	This study
pBBR1MCS-'tssA1^B	The final 347 codons of *B. cenocepacia tssA* were amplified with primers iotAfor3 and iotArev, the amplicon was cut with *Hin*dIII and *Bam*HI, and ligated between the same sites of pBBR1MCS.	This study
p34E-Tp	*dfrB2* (Tp^R) cassette vector (Ap^R, Tp^R)	(12)
pBBR1MCS-'tssA1^B::Tp	pBBR1MCS-'tssA3 with Tp^R cassette from p34E-Tp inserted into the *Eco*RI site within the *tssA* coding sequence.	This study
pSHAFT2	R6K-derived suicide vector for allelic replacement (Ap^R, Cm^R)	(13)
pSHAFT2-'tssA1^B::Tp	Disrupted *tssA1^B* allele was removed from pBBR1MCS-'tssA1^B::Tp with *Xho*I and *Not*I, and ligated to the *Sal*I and *Not*I sites of pSHAFT2.	This study
p34E-TpTer	p34E containing *dfrB2* gene fused to *rrnB* T1T2 terminators (Ap^R, Tp^R)	(13)
pEX18Tp-pheS	Gene replacement vector with ColE1-derived replicon, and mutated phenylalanyl tRNA synthase α-subunit gene, *pheS**, for cPhe counter- selection (Tp^R)	(14)
pEX18Tp-ΔrrnB-pheS	pEX18Tp-pheS derivative with *Afe*I-*Bsa*I restriction fragment removed (Tp^R)	This study
pSNUFF	pEX18Tp-ΔrrnB-pheS derivative containing *dfrB2-rrnB* T1T2 cassette in replacement of *Spe*I-*Eco*RV restriction fragment (Tp^R)	This study
pBBR1-tssM(-)	pBBR1MCS containing *tssM* amplified as a 3,943 bp DNA fragment with primers tssMfor and	This study

	tssMrev2, and ligated between the *Xba*I and *Kpn*I sites	
pBBR1-ΔtssM	pBBR1-tssM(-) with a 2133 bp in-frame deletion introduced into the *tssM* ORF by restriction with *Xho*I followed by religation	This study
pSNUFF-ΔtssM	pSNUFF containing Δ*tssM* allele from pBBR1-ΔtssM ligated between the *Xba*I and *Kpn*I sites	This study
Protein expression		
pET14b	T7 promoter expression vector allowing for incorporation of N-terminal His_6 tag, pMB1 origin (Ap^R)	Novagen
pET-21a	T7 promoter expression plasmid for incorporation of C-terminal His_6 tag, pMB1 origin (Ap^R)	Novagen
pACYCDuet-1	Dual T7-lac promoter expression vector, *lacI*, p15A origin (Cm^R)	Novagen
pETDuet-1	Dual T7-lac promoter expression vector, *lacI*, pMB1 origin (Ap^R)	Novagen
pETDuetΔO	pETDuet-1 lacking the *lac* operators at the two T7 promoters and containing a partial deletion of the *lacI* gene. Deletion of *lacO* at first T7 promoter and part of the *lacI* gene involved replacement of the 572 bp region between the *Bcl*I-*Not*I sites by annealed oligonucleotides delacO1for, delacO1for2, delacO1rev and delacO1rev2. Deletion of the *lacO* at the downstream T7 promoter was carried out by replacing the 101 bp region between the *Bsr*GI-*Nde*I sites by annealed oligonucleotides delacO2for, delacO2for2, delacO2rev and delacO2rev2. For both manipulations, the 'for2' and 'rev2' primers were kinased to facilitate ligation of the assembled double-stranded oligonucleotide to the vector.	This study
pMAL-c5X	Plasmid for generating fusions of MalE (MBP) to the N-terminus of target proteins. *malE* is under P_{tac} control. Contains *lacI* under P_{lacIQ} control (Ap^R)	New England Biolabs

pMAL-c5X-His_6	pMAL-c5X encoding MBP with a hexa-histidine tag located between amino acids 3 and 4. Constructed by replacing the 390 bp *Mfe*I-*Bgl*II fragment (that includes the *malE* Shine-Dalgarno sequence and first ~360 bp of *malE*) with a *Mfe*I-*Bgl*II DNA fragment generated by PCR using pMAL-c5X as template and primers pMALHisFor and pMALHisRev.	This study
pET14b-$TssA1^B$	pET14b encoding native $TssA1^B$. *B. cenocepacia tssA* was amplified with primers tssAnonHisFor and pET14b-iotArev, cut with *Xba*I and *Bam*HI, and ligated between the same sites of pET14b	This study
pET14b-His_6.$TssA1^B$	pET14b encoding N-terminal His-tagged $TssA1^B$. *B. cenocepacia tssA* was amplified with primers pET14b-iotAfor and pET14b-iotArev, cut with *Nde*I and *Bam*HI, and ligated between the same sites of pET14b. The encoded product contains a 20 amino acid N-terminal tag that harbours a hexahistidine tag and a thrombin cleavage site.	This study
pACYCDuet-linkerHis1.$TssA1^B$	pACYCDuet-1 encoding *B. cenocepacia* TssA with 6 consecutive histidine codons located between codons 262-263. Constructed by SOE PCR using two pairs of primers (pET14b-iotAfor and TssAlinkerHis1rev, and TssAlinkerHis1for and pET14b-iotArev) in which the products from the first pair of PCRs were combined and spliced in a second PCR using pET14b-iotAfor and pET14b-iotArev. The resulting amplicon was cut with *Nde*I and *Bgl*II, and ligated to the same sites of pACYCDuet-1.	This study
pMAL-c5X-$TssA1^B$.CTD	pMAL-c5X encoding a MBP-$TssA1^B$ CTD fusion. DNA encoding amino acids 294-373 of *B. cenocepacia tssA* was amplified with primers pET14b-ACTDfor and pET14b-iotArev, the product was restricted with *Nde*I and *Bam*HI, and ligated between the same sites of pMAL-c5X.	This study

	Following cleavage of the MBP fusion with Factor Xa, the released $TssA_{CTD}$ contains four additional non-native amino acids (ISHM) at the N-terminus.	
pMAL-c5X-TssA1^{B}.CTD3	pMAL-c5X encoding MBP fused to amino acids 303-373 of *B. cenocepacia* TssA1^{B}.	(2)
pMAL-c5X-His$_6$.TssA1^{B}.CTD H12	pMAL-c5X encoding MBP fused to His$_6$-tagged amino acids 303-329 of Bc TssA1^{B} (includes α-helix H12). *B. cenocepacia* DNA was amplified with primers pACYC-tssA.CTDfor and pACYC-tssA.CTDH1rev2, the product was restricted with *Nde*I and *Bam*HI, and ligated between the same sites of pMAL-c5X.	This study
pMAL-c5X-His$_6$.TssA1^{B}.CTD H12-H13	pMAL-c5X encoding MBP fused to His$_6$-tagged amino acids 303-347 of Bc TssA1^{B} (includes α-helices H12-H13). *B. cenocepacia* DNA was amplified with primers pACYC-tssA.CTDfor and pACYC-tssA.CTDH2rev2, the product was restricted with *Nde*I and *Bam*HI, and ligated between the same sites of pMAL-c5X.	This study
pMAL-c5X-His$_6$.TssA1^{B}.CTD H12-H14	pMAL-c5X encoding MBP fused to His$_6$-tagged amino acids 303-358 of Bc TssA1^{B} (includes α-helices H12-H14). *B. cenocepacia* DNA was amplified with primers pACYC-tssA.CTDfor and pACYC-tssA.CTDH3rev2, the product was restricted with *Nde*I and *Bam*HI, and ligated between the same sites of pMAL-c5X.	This study
pMAL-c5X-His$_6$.TssA1^{B}.CTD	pMAL-c5X encoding MBP fused to His$_6$-tagged amino acids 303-373 of Bc TssA1^{B}, i.e. the entire CTD (includes H12-H15). *B. cenocepacia* DNA was amplified with primers pACYC-tssA.CTDfor and pET14b-iotArev, the product was restricted with *Nde*I and *Bam*HI, and ligated between the same sites of pMAL-c5X.	This study
pET21a-TssA2^{A}.Nt1-Nt2.His$_6$	pET21a encoding Nt1-Nt2 region of TssA2^{A} with His-tag at C-terminus of Nt2. *A. hydrophila tssA* codons (1-374) amplified with primers pET21a-	This study

	AHA1844.Nt1Nt2.for and pET21a-AHA1844.Nt1Nt2.rev, and ligated between the *Nde*I and *Xho*I sites of pET21a	
pACYCDuet-His_6.$TssA2^A$.CTD	pACYCDuet-1 encoding N-terminal His-tagged $TssA2^A$ CTD. *A. hydrophila tssA* codons 381-478 amplified with primers pACYC-AHA1844.CTDSD2for and pACYC-AHA1844.rev, and ligated between the *Bam*HI and *Hin*dIII sites of pACYCDuet-1	This study
pET14b-His_6.TssD2	pET14b encoding N-terminal His-tagged TssD. *B. cenocepacia tssD* amplified with primers pET-His-iotDfor2 and pET14b-iotDrev, cut with *Nde*I and *Bam*HI, and ligated to the same sites of pET14b. Expresses TssD lacking the first three amino acids of TssD and containing a 20 amino acid N-terminal tag that harbours a hexahistidine tag and a thrombin cleavage site.	This study
Protein expression for co-IP		
pACYCDuet-FLAG.tssA	*B. cenocepacia tssA* amplified with primers TssA.NdeI.FLAG.for and TssA.BglII.rev, and inserted between *Nde*I and *Bgl*II sites of pACYCDuet-1. Expresses $_{FLAG}$.$TssA1^B$.	This study
pACYCDuet-tssC.VSVg	*B. cenocepacia tssC* amplified with primers tssC.NcoI.for and tssC.VSVg.BamHI.rev, and inserted between *Nco*I and *Bam*HI sites of pACYCDuet-1. Expresses TssC.$_{VSVg}$.	This study
pACYCDuet-tssC.VSVg-FLAG.tssA	*B. cenocepacia tssC* amplified with primers tssC.NcoI.for and tssC.VSVg.BamHI.rev, and inserted between *Nco*I and *Bam*HI sites of pACYCDuet-FLAG.tssA. Expresses TssC.$_{VSVg}$ and $_{FLAG}$.$TssA1^B$.	This study
pACYCDuet-VSVg.tssD	*B. cenocepacia tssD* amplified with tssD.forpACYC.NtermVSVgTag and tssD. BamHI.rev, and inserted between *Nco*I and	This study

	*Bam*HI sites of pACYCDuet-1. Expresses $_{VSVg}$.TssD.	
pACYCDuet-VSVg.tssD-FLAG.tssA	*B. cenocepacia tssD* amplified with tssD.forpACYC.NtermVSVgTag and tssD. BamHI.rev, and inserted between *Nco*I and *Bam*HI sites of pACYCDuet-FLAG.tssA. Expresses $_{VSVg}$.TssD and $_{FLAG}$.TssA1^{B}.	This study
pMAL-c5X-His$_6$.tssE	*B. cenocepacia tssE* amplified with pMALtssEfor and pMALtssErev, and inserted between *Nde*I and *Bam*HI sites of pMAL-c5X-His$_6$. Expresses $_{His6}$.MBP-TssE.	This study
pACYCDuet-tssF.HA	*B. cenocepacia tssF* amplified with tssF.NcoI.for and tssF.HA.HindIII.rev, and inserted between *Nco*I and *Hin*dIII sites of pACYCDuet-1. Expresses TssF.$_{HA}$.	This study
pACYCDuet-tssF.HA-FLAG.tssA	*B. cenocepacia tssF* amplified with tssF.NcoI.for and tssF.HA.HindIII.rev, and inserted between *Nco*I and *Hin*dIII sites of pACYCDuet-FLAG.TssA. Expresses TssF.$_{HA}$ and $_{FLAG}$.TssA1^{B}.	This study
pETDuetΔO-HA.tssF	*B. cenocepacia tssF* amplified with tssF.NdeI.HA.for and C-term-iotFrev, and inserted between *Nde*I and *Acc*65I sites of pETDuetΔO. Expresses $_{HA}$.TssF.	This study
pACYCDuet-VSVg.tssI$_{gp27gp5}$	*B. cenocepacia tssI* conserved core region amplified with tssI.BspHI.VSVgtag.For and BCAM0148gp5.BglII.rev, and inserted between *Nco*I and *Bam*HI sites of pACYCDuet-1. Expresses $_{VSVg}$.TssI$_{C}$ corresponding to the gp27gp5 core region (N-terminal 526 amino acids).	This study
pACYCDuet-VSVg.tssI$_{gp27gp5}$-FLAG.tssA	*B. cenocepacia tssA* amplified with TssA.NdeI.FLAG.for and TssA.BglII.rev, and inserted between *Nde*I and *Bgl*II sites of pACYCDuet-VSVg. tssI$_{gp27gp5}$. Expresses $_{VSVg}$.TssI$_{C}$ and $_{FLAG}$.TssA1^{B}.	This study

pACYCDuet-VSVg.tssK	*B. cenocepacia tssK* amplified with TssKforpET.NtermVSVgTag and TssK.BglII.rev, and inserted between *Nco*I and *Bam*HI sites of pACYCDuet-1. Expresses $_{VSVg}$.TssK.	This study
pACYCDuet-VSVg.tssK-FLAG.tssA	*B. cenocepacia tssA* amplified with TssA.NdeI.FLAG.for and TssA.BglII.rev, and inserted between *Nde*I and *Bgl*II sites of pACYCDuet-VSVg.tssK. Expresses $_{VSVg}$.TssK and $_{FLAG}$.TssA1^{B}.	This study
pACYCDuet-VSVg.tssL	*B. cenocepacia tssL* amplified with TssLforpET.NtermVSVgTag and TssL.Rev, and inserted between *Nco*I and *Bam*HI sites of pACYCDuet-1. Expresses $_{VSVg}$.TssL.	This study
pACYCDuet-VSVg.tssL-FLAG.tssA	*B. cenocepacia tssL* amplified with TssLforpET.NtermVSVgTag and TssL.Rev, and inserted between *Nco*I and *Bam*HI sites of pACYCDuet-FLAG.TssA. Expresses $_{VSVg}$.TssL and $_{FLAG}$.TssA1^{B}.	This study
pACYCDuet-VSVg.tssM$_{NTR}$	*B. cenocepacia tssM* codons 62-449 amplified with TssM1.VSVg.NcoI.for and TssM1.HindIII.rev, and inserted between *Nco*I and *Hin*dIII sites of pACYCDuet-1. Expresses $_{VSVg}$.TssM$_{NTR}$ (TssM cytoplasmic component).	This study
pACYCDuet-VSVg.tssM$_{NTR}$-FLAG.tssA	*B. cenocepacia tssM* codons 62-449 amplified with TssM1.VSVg.NcoI.for and TssM1.HindIII.rev, and inserted between *Nco*I and *Hin*dIII sites of pACYCDuet-FLAG.tssA. Expresses $_{VSVg}$.TssM$_{NTR}$ and $_{FLAG}$.TssA1^{B}.	This study
Two-hybrid system[b]		
pKT25	Vector for generating fusions to C-terminus of CyaA T25 fragment, *ori*$_{p15A}$ (Km^{R})	(15)
pKT25-zip	pKT25 containing the coding sequence of the yeast GCN4 leucine zipper fused to the T25 coding sequence	(9)
pKNT25	Vector for generating fusions to N-terminus of CyaA T25 fragment, *ori*$_{p15A}$ (Km^{R})	(16)

pUT18	Vector for generating fusions to N-terminus of CyaA T18 fragment, ori_{pMB1} (Ap^R)	(15)
pUT18C	Vector for generating fusions to C-terminus of CyaA T18 fragment, ori_{pMB1} (Ap^R)	(15)
pUT18C-zip	pUT18C containing the coding sequence of the yeast GCN4 leucine zipper fused to the T18 coding sequence	(15)
pKT25-tssA1^B	*B. cenocepacia tssA* amplified with Cterm-iotAfor and Cterm-iotArev, and inserted between the *Pst*I and *Xba*I sites of pKT25.	This study
pKNT25-tssA1^B	*B. cenocepacia tssA* amplified with Nterm-iotAfor and N-iotAfullrev, and inserted between the *Hin*dIII and *Bam*HI sites of pKNT25	This study
pUT18-tssA1^B	*B. cenocepacia tssA* amplified with Nterm-iotAfor and N-iotAfullrev, and inserted between the *Hin*dIII and *Bam*HI sites of pUT18	This study
pUT18C-tssA1^B	*B. cenocepacia tssA* amplified with Cterm-iotAfor and Cterm-iotArev, and inserted between the *Pst*I and *Xba*I sites of pUT18C	This study
pKT25-tssA1$^B_{Nt1}$	*B. cenocepacia tssA* codons 1-256 amplified with C-term-iotAfor and C-iotAlongNTDrev, and inserted between the *Pst*I and *Xba*I sites of pKT25	This study
pKNT25-tssA1$^B_{Nt1}$	*B. cenocepacia tssA* codons 1-256 amplified with Nterm-iotAfor and N-iotAlongNTDrev, and inserted between the *Hin*dIII and *Bam*HI sites of pKNT25	This study
pUT18-tssA1$^B_{Nt1}$	*B. cenocepacia tssA* codons 1-256 amplified with Nterm-iotAfor and N-iotAlongNTDrev, and inserted between the *Hin*dIII and *Bam*HI sites of pUT18	This study
pUT18C-tssA1$^B_{Nt1}$	*B. cenocepacia tssA* codons 1-256 amplified with C-term-iotAfor and C-iotAlongNTDrev, and inserted between the *Pst*I and *Xba*I sites of pUT18C	This study

pKT25-tssA1$^{B}_{CTD}$	*B. cenocepacia tssA* codons 294-374 amplified with C-iotAshortCTDfor and Cterm-iotArev, and inserted between the *Pst*I and *Xba*I sites of pKT25	This study
pKNT25-tssA1$^{B}_{CTD}$	*B. cenocepacia tssA* codons 294-373 amplified with N-iotAshortCTDfor and N-iotAfullrev, and inserted between the *Hin*dIII and *Bam*HI sites of pKNT25	This study
pUT18-tssA1$^{B}_{CTD}$	*B. cenocepacia tssA* codons 294-373 amplified with N-iotAshortCTDfor and N-iotAfullrev, and inserted between the *Hin*dIII and *Bam*HI sites of pUT18	This study
pUT18C-tssA1$^{B}_{CTD}$	*B. cenocepacia tssA* codons 294-374 amplified with C-iotAshortCTDfor and Cterm-iotArev, and inserted between the *Pst*I and *Xba*I sites of pUT18C	This study
pKT25-tssA2^{A}	*A. hydrophila tssA* amplified with primers AHA1844fullfor and AHA1844-Crev2, and inserted between the *Xba*I and *Kpn*I sites of pKT25.	This study
pKNT25-tssA2^{A}	*A. hydrophila tssA* amplified with primers AHA1844fullfor and AHA1844-Nrev, and inserted between the *Xba*I and *Kpn*I sites of pKNT25	This study
pUT18-tssA2^{A}	*A. hydrophila tssA* amplified with primers AHA1844fullfor and AHA1844-Nrev, and inserted between the *Xba*I and *Kpn*I sites of pUT18	This study
pUT18C-tssA2^{A}	*A. hydrophila tssA* amplified with primers AHA1844fullfor and AHA1844-Crev2, and inserted between the *Xba*I and *Kpn*I sites of pUT18C.	This study
pKT25-tssA2^{A} Nt1	*A. hydrophila tssA* codons 1-229 amplified with primers AHA1844fullfor and AHA1844NTDrev, and inserted between the *Xba*I and *Kpn*I sites of pKT25	This study
pKNT25-tssA2^{A} Nt1	*A. hydrophila tssA* codons 1-229 amplified with primers AHA1844fullfor and AHA1844NTDrev2,	This study

	and inserted between the *Xba*I and *Kpn*I sites of pKNT25	
pUT18-tssA2^{A} Nt1	*A. hydrophila tssA* codons 1-229 amplified with primers AHA1844fullfor and AHA1844NTDrev2, and inserted between the *Xba*I and *Kpn*I sites of pUT18	This study
pUT18C-tssA2^{A} Nt1	*A. hydrophila tssA* codons 1-229 amplified with primers AHA1844fullfor and AHA1844NTDrev, and inserted between the *Xba*I and *Kpn*I sites of pUT18C	This study
pKT25-tssA2^{A} Nt2	*A. hydrophila tssA* codons 244-386 amplified with primers AHA1844for and AHA1844CTDrev, and inserted between the *Xba*I and *Kpn*I sites of pKT25	This study
pUT18C-tssA2^{A} Nt2	*A. hydrophila tssA* codons 244-386 amplified with primers AHA1844for and AHA1844CTDrev, and inserted between the *Xba*I and *Kpn*I sites of pUT18C	This study
pKT25-tssA2^{A} CTD	*A. hydrophila tssA* codons 381-478 amplified with primers AHA1844for2 and AHA1844-Crev2, and inserted between the *Xba*I and *Kpn*I sites of pKT25	This study
pUT18C-tssA2^{A} CTD	*A. hydrophila tssA* codons 381-478 amplified with primers AHA1844for2 and AHA1844-Crev2, and inserted between the *Xba*I and *Kpn*I sites of pUT18C	This study
pKT25-tssA2^{A} Nt2-CTD	*A. hydrophila tssA* codons 244-478 amplified with primers AHA1844for and AHA1844-Crev2, and inserted between the *Xba*I and *Kpn*I sites of pKT25	This study
pUT18C-tssA2^{A} Nt2-CTD	*A. hydrophila tssA* codons 244-478 amplified with primers AHA1844for and AHA1844-Crev2, and inserted between the *Xba*I and *Kpn*I sites of pUT18C	This study

pKT25-tssB	*B. cenocepacia tssB* amplified with Cterm-iotBfor and Cterm-iotBrev, and inserted between the *Xba*I and *Eco*RI sites of pKT25	This study
pKNT25-tssB	*B. cenocepacia tssB* amplified with Nterm-iotBfor and N-iotBfullrev, and inserted between the *Sph*I and *Xba*I sites of pKNT25	This study
pUT18-tssB	*B. cenocepacia tssB* amplified with Nterm-iotBfor and N-iotBfullrev, and inserted between *Sph*I and *Xba*I sites of pUT18	This study
pUT18C-tssB	*B. cenocepacia tssB* amplified with Cterm-iotBfor and Cterm-iotBrev, and inserted between the *Xba*I and *Eco*RI sites of pUT18C	This study
pKT25-tssC	*B. cenocepacia tssC* amplified with Cterm-iotCfor and Cterm-iotCrev, and inserted between the *Pst*I and *Xba*I sites of pKT25	This study
pKNT25-tssC	*B. cenocepacia tssC* amplified with Nterm-iotCfor and Nterm-tssCrev.new, and inserted between the *Hin*dIII and *Bam*HI sites of pKNT25, followed by *Acc*65I digestion, end filling and self-ligation.	This study
pUT18-tssC	*B. cenocepacia tssC* amplified with Nterm-iotCfor and Nterm-tssCrev.new, and inserted between *Hin*dIII and *Bam*HI sites of pUT18	This study
pUT18C-tssC	*B. cenocepacia tssC* amplified with Cterm-tssCfor.new and Cterm-iotCrev, and inserted between the *Pst*I and *Xba*I sites of pUT18C	This study
pKT25-tssD	*B. cenocepacia tssD* amplified with pUT18C-iotDfor and Cterm-iotDrev, and inserted between the *Pst*I and *Xba*I sites of pKT25	This study
pKNT25-tssD	*B. cenocepacia tssD* amplified with Nterm-iotDfor and Nterm-iotDrev, and inserted between the *Hin*dIII and *Bam*HI sites of pKNT25	This study
pUT18-tssD	*B. cenocepacia tssD* amplified with Nterm-iotDfor and Nterm-iotDrev, and inserted between *Hin*dIII and *Bam*HI sites of pUT18	This study

pUT18C-tssD	*B. cenocepacia tssD* amplified with pUT18C-iotDfor and Cterm-iotDrev, and inserted between the *Pst*I and *Xba*I sites of pUT18C	This study
pKT25-tssE	*B. cenocepacia tssE* amplified with Cterm-iotEfor and Cterm-iotErev, and inserted between the *Xba*I and *Eco*RI sites of pKT25	This study
pKNT25-tssE	*B. cenocepacia tssE* amplified with Nterm-iotEfor and N-iotEfullrev, and inserted between the *Hin*dIII and *Bam*HI sites of pKNT25	This study
pUT18-tssE	*B. cenocepacia tssE* amplified with Nterm-iotEfor and N-iotEfullrev, and inserted between *Hin*dIII and *Bam*HI sites of pUT18	This study
pUT18C-tssE	*B. cenocepacia tssE* amplified with Cterm-iotEfor and Cterm-iotErev, and inserted between the *Xba*I and *Eco*RI sites of pUT18C	This study
pKT25-tssF	*B. cenocepacia tssF* amplified with Cterm-iotFfor and Cterm-iotFrev, and inserted between the *Xba*I and *Acc*65I sites of pKT25	This study
pKNT25-tssF	*B. cenocepacia tssF* amplified with Nterm-iotFfor and N-iotFfullrev, and inserted between the *Hin*dIII and *Xba*I sites of pKNT25	This study
pUT18-tssF	*B. cenocepacia tssF* amplified with Nterm-iotFfor and N-iotFfullrev, and inserted between *Hin*dIII and *Xba*I sites of pUT18	This study
pUT18C-tssF	*B. cenocepacia tssF* amplified with Cterm-iotFfor and Cterm-iotFrev, and inserted between the *Xba*I and *Acc*65I sites of pUT18C	This study
pKT25-tssG	*B. cenocepacia tssG* amplified with C-iotGfrag1for and Cterm-iotGrev, and inserted between the *Xba*I and *Eco*RI sites of pKT25	This study
pKNT25-tssG	*B. cenocepacia tssG* amplified with N-iotGfrag1for and Nterm-iotGrev2, and inserted between the *Hin*dIII and *Bam*HI sites of pKNT25	This study
pUT18-tssG	*B. cenocepacia tssG* amplified with N-iotGfrag1for and Nterm-iotGrev2, and inserted between *Hin*dIII and *Bam*HI sites of pUT18	This study

pUT18C-tssG	*B. cenocepacia tssG* amplified with C-iotGfrag1for and Cterm-iotGrev, and inserted between the *Xba*I and *Eco*RI sites of pUT18C	This study
pKT25-tssI	*B. cenocepacia tssI* amplified with TssIfor and C-tssIrev, and inserted between the *Xba*I and *Bam*HI sites of pKT25 following cleavage of the amplicon with *Xba*I and *Bgl*II	This study
pKNT25-tssI	*B. cenocepacia tssI* amplified with TssIfor and N-tssIrev, and inserted between the *Xba*I and *Bam*HI of pKNT25 following cleavage of the amplicon with *Xba*I and *Bgl*II	This study
pUT18-tssI	*B. cenocepacia tssI* amplified with TssIfor and N-tssIrev, and inserted between the *Xba*I and *Bam*HI sites of pUT18 following cleavage of the amplicon with *Xba*I and *Bgl*II	This study
pUT18C-tssI	*B. cenocepacia tssI* amplified with TssIfor and C-tssIrev, and inserted between the *Xba*I and *Bam*HI sites of pUT18C following cleavage of the amplicon with *Xba*I and *Bgl*II	This study
pKT25-tssI$_C$	*B. cenocepacia tssI* codons 1-526 amplified with TssIfor and Rev.gp5, and inserted between the *Xba*I and *Bam*HI sites of pKT25 following cleavage of the amplicon with *Xba*I and *Bgl*II	This study
pUT18C-tssI$_C$	*B. cenocepacia tssI* codons 1-526 amplified with TssIfor and Rev.gp5, and inserted between the *Xba*I and *Bam*HI sites of pUT18C following cleavage of the amplicon with *Xba*I and *Bgl*II	This study
pKT25-tssJ	*B. cenocepacia tssJ* codons 13-200 amplified with For1tssj and C-tssJrev, and inserted between the *Pst*I and *Bam*HI sites of pKT25	This study
pKNT25-tssJ	*B. cenocepacia tssJ* codons 13-200 amplified with For2tssj and N-tssJrev, and inserted between the *Pst*I and *Bam*HI sites of pKNT25	This study
pUT18-tssJ	*B. cenocepacia tssJ* codons 13-200 amplified with For2tssj and N-tssJrev, and inserted between *Pst*I and *Bam*HI sites of pUT18	This study

pUT18C-tssJ	*B. cenocepacia tssJ* codons 13-200 amplified with For2tssj and C-tssJrev, and inserted between the *Pst*I and *Bam*HI sites of pUT18C	This study
pKT25-tssK	*B. cenocepacia tssK* amplified with tssKfor and C-tssKrev, and inserted between the *Bam*HI and *Acc*65I sites of pKT25	This study
pKNT25-tssK	*B. cenocepacia tssK* amplified with tssKfor and N-tssKrev, and inserted between the *Hin*dIII and *Bam*HI sites of pKNT25	This study
pUT18-tssK	*B. cenocepacia tssK* amplified with tssKfor and N-tssKrev, and inserted between *Bam*HI and *Acc*65I sites of pUT18	This study
pUT18C-tssK	*B. cenocepacia tssK* amplified with tssKfor and C-tssKrev, and inserted between the *Bam*HI and *Acc*65I sites of pUT18C	This study
pKT25-tssL	*B. cenocepacia tssL* codons 1-204 amplified with tssLfor and C-tssLrev, and inserted between the *Xba*I and *Acc*65I sites of pKT25	This study
pKNT25-tssL	*B. cenocepacia tssL* codons 1-204 amplified with tssLfor and N-tssLrev, and inserted between the *Xba*I and *Acc*65I sites of pKNT25	This study
pUT18-tssL	*B. cenocepacia tssL* codons 1-204 amplified with tssLfor and N-tssLrev, and inserted between *Xba*I and *Acc*65I sites of pUT18	This study
pUT18C-tssL	*B. cenocepacia tssL* codons 1-204 amplified with tssLfor and C-tssLrev, and inserted between the *Xba*I and *Acc*65I sites of pUT18C	This study
pKT25-tssM$_{NTR}$	*B. cenocepacia tssM* codons 62-450 amplified with tssM1for and C-tssM1rev, and inserted between the *Xba*I and *Acc*65I sites of pKT25	This study
pKNT25-tssM$_{NTR}$	*B. cenocepacia tssM* codons 62-450 amplified with tssM1for and N-tssM1rev, inserted between the *Xba*I and *Acc*65I sites of pKNT25, followed by *Acc*65I digestion, end filling and self-ligation.	This study
pUT18-tssM$_{NTR}$	*B. cenocepacia tssM* codons 62-450 amplified with tssM1for and N-tssM1rev, and inserted	This study

	between *Xba*I and *Acc*65I sites of pUT18, followed by *Acc*65I digestion, DNA end filling and self-ligation.	
pUT18C-tssM$_{NTR}$	*B. cenocepacia tssM* codons 62-450 amplified with tssM1for and C-tssM1rev, and inserted between the *Xba*I and *Acc*65I sites of pUT18C	This study
pKT25-tssM$_{CTR}$	*B. cenocepacia tssM* codons 467-1309 amplified with tssM2for and C-tssM2rev, and inserted between the *Bam*HI and *Acc*65I sites of pKT25	This study
pKNT25-tssM$_{CTR}$	*B. cenocepacia tssM* codons 467-1308 amplified with tssM2for and N-tssM2rev, and inserted between the *Bam*HI and *Acc*65I sites of pKNT25	This study
pUT18-tssM$_{CTR}$	*B. cenocepacia tssM* codons 467-1308 amplified with tssM2for and N-tssM2rev, and inserted between the *Bam*HI and *Acc*65I sites of pUT18	This study
pUT18C-tssM$_{CTR}$	*B. cenocepacia tssM* codons 467-1309 amplified with tssM2for and C-tssM2rev, and inserted between the *Bam*HI and *Acc*65I sites of pUT18C	This study

[a] Ap^R, ampicillin resistance; Cm^R, chloramphenicol resistance; Km^R, kanamycin resistance; Tp^R, trimethoprim resistance; BHR, broad host-range

[b] T6SS subunit or domain coding sequences included the native or an added translation termination codon (as appropriate) when ligated to pKT25 or pUT18C but were omitted for pKNT25 and pUT18.

Supplementary references

1. Sievers, F. & Higgins, D. G. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.* **1079,** 105–16 (2014).
2. Owen, H. J. *et al.* TssA from Burkholderia cenocepacia : expression, purification, crystallization and crystallographic analysis. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **74,** 536–542 (2018).
3. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 12–21 (2010).
4. Romling, U. *et al.* Epidemiology of chronic Pseudomonas aeruginosa infections in cystic fibrosis. *J Infect Dis* **170,** 1616–1621 (1994).
5. Seshadri, R. *et al.* Genome sequence of Aeromonas hydrophila ATCC 7966T: Jack of all trades. *J. Bacteriol.* **188,** 8272–8282 (2006).
6. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* **33,** 103–119 (1985).
7. Jessee, J. New subcloning efficiency competent cells: >1 × 106 transformants/μg. *Focus Bethesda Res. Lab.* **8,** 9–10 (1986).
8. Simon, R., Priefer, U. & Pühler, A. A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Biotechnology* **1,** 784–791 (1983).
9. Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci.* **95,** 5752–5756 (1998).
10. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189,** 113–130 (1986).
11. Kovach, M. E., Phillips, R. W., Elzer, P. H., Roop, R. M. & Peterson, K. M. pBBR1MCS: a broad-host-range cloning vector. *Biotechniques* **16,** 800–802 (1994).
12. DeShazer, D. & Woods, D. E. Broad-Host-Range Cloning and Cassette Vectors

Based on the R388 Trimethoprim Resistance Gene. *Biotechniques* **20,** 762–764 (1996).

13. Shastri, S. *et al.* An efficient system for the generation of marked genetic mutants in members of the genus Burkholderia. *Plasmid* **89,** 49–56 (2017).
14. Barrett, A. R. *et al.* Genetic Tools for Allelic Replacement in Burkholderia Species. *Appl. Environ. Microbiol.* **74,** 4498–4508 (2008).
15. Karimova, G., Ullmann, A. & Ladant, D. Protein-protein interaction between Bacillus stearothermophilus tyrosyl-tRNA synthetase subdomains revealed by a bacterial two-hybrid system. *J. Mol. Microbiol. Biotechnol.* **3,** 73–82 (2001).
16. Karimova, G., Dautin, N. & Ladant, D. Interaction network among Escherichia coli membrane proteins involved in cell division as revealed by bacterial two-hybrid analysis. *J. Bacteriol.* **187,** 2233–2243 (2005).

3.3 Paper III

3.3.1 Summary

The third paper in this thesis describes the construct generation, protein production, purification, crystallisation, preliminary X-ray data collection and X-ray data analysis of TssK from *B. cenocepacia* (Bc-TssK) (I35_RS01705). The manuscript identifies the successful production of both native and selenomethionine incorporated Bc-TssK protein, in addition, the generation of diffraction quality crystals, which provided suitable preliminary X-ray data such that an initial estimate of the asymmetric content and unit cell parameters could be determined. The Bc-TssK crystals have been determined to contain the full-length protein through mass spectrometry, and the asymmetric unit is estimated to contain two molecules of Bc-TssK.

3.3.2 Protein summary

TABLE 3.3: Summary of TssK proteins discussed

Paper III			
Construct ID	**Residue range**	**Resolution**	**Space Group**
Bc_TssK	(1-448)	2.5 Å	P63
Bc_TssK (Se)	(1-448)	3.6 Å	P63

3.3.3 Contribution

M.S.T. conceived the investigation; S.R.D., R.S., M.S.T. and D.W.R. designed the experiments; S.R.D., R.S., A.A., A.E.M., S.E.S. and H.J.O. performed the experiments; S.R.D., R.S., P.J.B., D.W.R and M.S.T. interpreted the data and S.R.D., M.S.T and D.W.R. wrote the manuscript. Experimentally, R.S. and A.A carried out the construct design and established the expression protocol for the protein. S.E.S. carried out the initial protein purification. R.S. and M.S.T. prepared samples for SEC-MALLS. H.J.O. carried out preliminary crystallisation trials of the Bc-TssK native protein. I carried out the repeat protein expression and purification of the Bc-TssK construct including the semet incorporated protein. In addition, I carried out the crystallisation (with the assistance of A.E.M.), X-ray data collection and data analysis.

publBio research communications

TssK from *Burkholderia cenocepacia:* expression, purification and crystallographic studies.

Samuel R. Dix[a], Hayley J. Owen[a], Ruyue Sun[b], Amy E. Milburn[a], Svetlana E. Sedelnikova[a], Patrick J. Baker[a], Mark S. Thomas[b]* and David W. Rice[a]*

[a]Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield, South Yorkshire, S10 2TN, United Kingdom

[b]Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield Medical School, Beech Hill Road, Sheffield, S. Yorkshire, S10 2RX, UK

Correspondence email: m.s.thomas@sheffield.ac.uk; d.rice@sheffield.ac.uk

Keywords: Type VI secretion system, TssK subunit, *Burkholderia cenocepacia*

Synopsis

The TssK subunit from *Burkholderia cenocepacia* have been crystallised in a form suitable for X-ray analysis.

Abstract

TssK is a core subunit of the type VI secretion system, a major participant in interspecies competition between Gram-negative bacteria. Previous biochemical and structural studies of enteroaggregative *Escherichia coli* TssK have demonstrated that it is a trimeric molecule with each subunit organised into three domains, one of which, the shoulder domain, shares structural similarities with phage receptor binding proteins, suggesting that TssK acts as an adaptor protein for correct T6SS assembly. In addition, the head domain displays a high degree of mobility compared to the rest of the molecule. Here we present the expression and purification of TssK from *Burkholderia cenocepacia* and a preliminary analysis of diffraction quality crystals that were produced following extensive crystallisation trials, as the first step to explore the TssK species variant and investigate the nature of the head domain mobility.

1. Introduction

The type VI secretion system (T6SS) is a multi-protein complex, found in Gram-negative bacteria, that transports effector proteins to target cells (Leiman *et al*., 2009). It is comprised of 13 different core subunits (TssA-TssM) and an additional PAAR protein, which together make up two distinct assemblies within the complex. One of them, the membrane complex, comprised of TssJLM, is a chamber-like complex that is required for the docking of the contractile machinery to the cell envelope and allows for the passage of the inner tube and effector proteins during firing (Zheng & Leung, 2007; Durand *et al*., 2015; Cianfanelli *et al*., 2016). The second assembly is the contractile machinery, which is comprised of two sub-complexes. The first, the contractile sheath, consists of the inner tube that is formed from stacked hexamers of TssD (hcp), capped by the trimeric TssI (VgrG) protein and a PAAR subunit. The inner tube is surrounded by a six-start helix comprised of repeating heterodimers of TssBC (Basler *et al*., 2012; Shneider *et al*., 2013; Brunet *et al*., 2014; Kube *et al*., 2014; Wang *et al*., 2017). The second sub-complex is the baseplate assembly formed by TssEFG and K. Once the T6SS is triggered, the sheath contracts against the baseplate which results in ejection of the inner tube and effector proteins through the lumen of the baseplate (Pukatzki *et al*., 2007; Brunet *et al*., 2015; Taylor *et al*., 2016). After firing, the contracted sheath is recycled by TssH (ClpV), an AAA^+ ATPase (Bönemann *et al*., 2009; Basler *et al*., 2012). The T6SS is evolutionarily related to R-type pyocins and contractile bacteriophage tails, such that they share a similar 'load and fire′ mechanism (Leiman *et al*., 2009; Shneider *et al*., 2013; Ge *et al*., 2015).

Until recently little was known about the structure and role of the TssK subunit within the T6SS complex. Knock-out studies demonstrated that the presence of this subunit is critical for the correct function of the T6SS and it was shown to interact with the T6SS membrane complex and baseplate (Zoued *et al*., 2013; English *et al*., 2014). Negative stain electron microscopy studies indicated that *E. coli* TssK adopts an overall homotrimeric structure (Zoued *et al*., 2013), a finding confirmed by the recent crystallographic studies of TssK from enteroaggregative *E. coli* (EAEC) which provided the first high-resolution model (Nguyen *et al*., 2017). The latter studies identified that TssK is comprised of three domains referred to as the shoulder, neck and head domains and comparisons of the fold of TssK to that of other proteins in the PDB revealed that the three-dimensional architecture of its shoulder domain displays similarity to that of receptor binding proteins from siphophage viruses (Spinelli *et al*., 2006; Nguyen *et al*., 2017).

Interaction studies with TssK domains have shown that the neck and shoulder domains are involved in the association of TssK with the baseplate, whereas the head domain interacts with TssL and possibly TssM within the membrane complex. These interactions are in agreement with recent cryo-EM studies whereby trimers of TssK, associated with the T6SS baseplate, can be identified and in which the orientation is such that it would allow for interaction of the head domain with the membrane complex (Nguyen *et al*., 2017; Nazarov *et al*., 2017). However, in the cryo-EM structure of TssK, the head domain appears to be mobile and in the X-ray structure it is less well ordered than the associated neck and shoulder domains (Nguyen *et al*., 2017; Nazarov *et al*., 2017), giving rise to unanswered questions concerning the orientation of the head domain in the trimer.

In order to contribute to our understanding of the mobility and orientation of the head domain in the trimer, we have initiated the study of TssK from *Burkholderia cenocepacia*, an opportunistic human pathogen, which contains a single T6SS, and is notorious for causing potentially lethal infections in cystic fibrosis patients (Lyczak *et al*., 2002; Aubert *et al*., 2008). In this paper, we report the over expression and purification of recombinant *B. cenocepacia* TssK (′Bc-TssK′), together with the result of crystallisation trials and preliminary X-ray analysis.

2. Materials and methods

2.1. Macromolecule production

To produce native Bc-TssK, DNA encoding Bc-TssK (I35_RS01705) was amplified from *B. cenocepacia* H111 (Römling et al., 1994) genomic DNA, by the polymerase chain reaction (PCR), using primers listed in Table 1. The amplicon was digested with *Nde*I and *Bgl*II and then ligated into the corresponding restriction sites of the pACYCDuet-1 (Novagen) expression vector downstream of the T7 promoter, generating pACYCDuet-tssK. In addition, a plasmid expressing an N-terminal hexahistidine-tagged derivative, Bc-His6.TssK, was constructed through the digestion of pUT18C-tssK (Dix et al., 2018) with *Bam*HI and *Acc*65I, and ligation of the released ~ 1.35 kb DNA fragment to the *Bam*HI and *Bsr*GI sites of pACYCDuet-1 to generate pACYCDuet-His_6.tssK. Both plasmid constructs were validated through DNA sequencing and subsequently transformed into *E. coli* BL21(DE3) (Novagen) for protein over expression.

E. coli BL21(DE3) containing pACYCDuet-Bc-TssK was grown, in a shaker incubator, in 1 l (2 x 500 ml in 2 l flasks) of BHI broth (Oxoid) at 37 °C supplemented with 100 µg/µl ampicillin. Once the cultures had reached an OD_{600} between 0.5-0.7, 1 mM IPTG was added for the induction of the recombinant Bc-TssK protein expression via the second T7 promoter present within the pACYCDuet-1 expression vector (Table 1). Cultures were incubated for a further 2-3 h at 30 °C before the cells were harvested, producing 4 g of cell paste. In parallel, Bc-His_6.TssK protein was produced from pACYDuet-His_6.tssK using the same protocol.

Selenomethionine incorporated Bc-TssK protein (SeMet-Bc-TssK) was produced by growing 2 l (4 x 500 ml in 2 l flasks) of *E. coli* BL21(DE3) cells containing pACYDuet-tssK in BHI medium to an OD_{600} of 0.6-0.7, whereupon cells

were harvested and resuspended in the same volume of selenomethionine minimal medium (10.5 g/l K_2HPO_4, 1.0 g/l $(NH_4)_2SO_4$, 4.5 g/l KH_2PO_4, 0.5 g/l tri-sodium citrate.$2H_2O$, 5.0 g/l glycerol and 0.5 g/l each of adenine, guanosine, thymine and uracil) supplemented with 1 g/l $MgSO_4.7H_2O$, 4 mg/l thiamine, 100 mg/l each of L-lysine, L-phenylalanine and L-threonine, 50 mg/l each of L-isoleucine, L-leucine and L-valine, and 40 mg/l seleno-L-methionine. Growth was continued for 1 h at 37 °C followed by the addition of 1 mM IPTG to induce Bc-TssK over expression, and a further 12 h of growth at 25 °C, whereupon cells were harvested, producing 8 g of cell paste.

Harvested cells were resuspended in 5 ml/g 50 mM Tris (pH 8.0) and lysed by sonication. The lysate was then cleared by centrifugation for 30 min at 35,000 *g*. Bc-TssK lysate was applied to an anion exchange column (DEAE FF 10 ml, GE Healthcare Life Sciences) and eluted over a linear gradient of NaCl (0-500 mM), 50 mM Tris (pH 8.0), followed by addition of 4.0 M ammonium sulphate to a final concentration of 1.4 M to precipitate Bc-TssK and remove soluble contaminants. Bc-TssK was re-suspended in 50 mM Tris (pH 8.0) before being further purified by gel filtration (HiLoad Superdex 200 16/600, GE Healthcare Life Sciences) in 50 mM Tris (pH 8.0) containing 500 mM NaCl. Bc-His_6.TssK lysate was applied to a 5 mL HisTrap HP column (GE Healthcare Life Science) and eluted with a linear gradient of imidazole (10-500 mM). Protein purity was analysed via SDS-PAGE and subsequent staining with Coomassie blue. The protein was concentrated for subsequent crystallisation trials utilising an Amicon protein concentrator with a molecular weight cut off of 30 kDa. To determine the oligomerisation status of Bc-His_6.TssK (subunit mwt. 51.5 kDa), the purified protein was subjected to SEC-MALLS analysis following elution from a Superose-6 10/300 column (GE Healthcare Life Sciences) at the Molecular Interactions Lab, University of York, UK (Fig. 1b).

2.2. Crystallisation

In order to carry out crystallisation trials, native and SeMet-Bc-TssK were concentrated to 10 mg/ml. A Matrix Hydra II PlusOne crystallisation robot was utilised to carry out initial screening, generating 400 nl drops with a 1:1 ratio of precipitant:protein in 96-well MRC2 sitting-drop crystallisation trays. These were then incubated at 290 °K to allow for equilibration through vapour diffusion. Commercially available sparse-matrix screens from Nextal (pH Clear, Morpheus, PACT, $JCSG^+$, MPD, ProPlex and $AmSO_4$) were used to identify conditions that formed crystals. Diffraction quality protein crystals were identified for both the native and selenomethionine proteins in 0.1 M magnesium chloride, 0.1 M sodium HEPES (pH 7.5) and 10 % (w/v) PEG 4000 and were optimised in 24-well trays with a reservoir volume of 500 µl. Crystals of Bc-TssK were sent for mass spectrometry analysis to confirm if the identity of the crystallised protein.

2.3. Data collection and processing

Crystals were harvested in cryoloops (Hampton Research) before being cryo-protected in a solution of mother liquor with an additional 25 % ethylene glycol. The crystals were then flash-cooled in liquid nitrogen to 100 °K. Native data were collected at Diamond Light Source, Oxford, for crystals of Bc-TssK on MX beamline i03, whereas selenomethione peak data were collected for crystals of SeMet Bc-TssK on the i04-1 MX beamline. Data were processed utilising the Xia2 software pipeline (Winter, 2010; Winter & McAuley, 2011). Details of data collection and processing statistics are shown in Table 3. Contents of the asymmetric unit, were estimated using the Matthews coefficient and self-rotation functions were calculated using Polarrfn as part of the CCP4 software package (Winn *et al*., 2011).

3. Results and discussion

Native and SeMet-Bc-TssK were successfully over expressed and purified (Fig. 1a) using ion exchange chromatography followed by ammonium sulphate precipitation. Further purification was achieved by gel-filtration chromatography which indicated that the protein has an apparent molecular weight of 165 kDa. The final yield of native and SeMet-Bc-TssK from 4 g and 8 g of cell paste was 14 mg and 6.4 mg, respectively. Further analysis of Bc-His6.TssK by SEC-MALLS to

determine the oligomeric nature of Bc-TssK suggested a molecular weight of 155 kDa (Fig. 1b), consistent with a trimeric quarternary structure.

Subsequently, both native and SeMet-Bc-TssK protein preparations were subjected to crystallisation trials, producing crystals with a rod-like morphology suitable for data collection (Fig. 2). Full details of crystallisation conditions can be found in Table 2. Mass spectrometry analysis of Bc-TssK crystals were consistent with the expected molecular weight of full-length TssK (49.9 kDa) (Sup. fig. 1), confirming the correct protein had been crystallised.

Crystals of Bc-TssK and SeMet-Bc-TssK diffract to 2.50 Å and 3.60 Å, respectively and belong to space group P63. Analysis of a self-rotation function calculated on data 86.6 Å-2.50 Å, using an integration radius of 30 Å indicates the presence of non-crystallographic two-fold axes every 30° (Fig. 3) in the a,b plane with a peak height equal to over 90 % of the origin. This would suggest that the asymmetric unit contains two molecules (49.9 kDa) of TssK, with a Matthews coefficient of 3.07 $Å^3$ Da^{-1}.

Attempts to improve the diffraction quality of the crystals and solve the structures of Bc-TssK by experimental phasing are currently underway. This work provides the initial step to determining the structure of Bc-TssK and identifying the differences between EAEC *E. coli* and *B. cenocepacia* TssK, in addition to shedding further light on the role of the TssK head domain.

Table 1

Macromolecule production information

Construct ID	pACYDuet-tssK (Bc-TssK)	pACYDuet-His_6.tssK (Bc-His_6.TssK)
Source organism	*B. cenocepacia* H111	*B. cenocepacia* H111
Gene ID	I35_RS01705	I35_RS01705
Forward primer	GCGCCATATGAGTTATTCGGCCAAGGT	N/A
Reverse primer	GCGCAGATCTTCATGATGTGACCGCGATCA	N/A
Restriction sites	*Nde*I and *Bgl*II	*Bam*HI + *Acc*65I/*Bsr*GI
Expression vector	pACYCDuet-1	pACYCDuet-1
Expression host	*E. coli* BL21 (DE3)	*E. coli* BL21 (DE3)
Amino acids	(1-448) (49.9 kDa)	MGSS$(H)_6$SQDP.(1-448) (51.5 kDa)

Underlined sequence indicated restriction sites.

Table 2

Crystallisation

Construct	Bc-TssK (+ SeMet)
Method	Vapour diffusion (hanging drop)
Plate type	24-well tissue culture tray
Temperature (K)	290
Protein concentration (mg/ml)	10
Buffer composition of protein solution	50 mM Tris (pH 8.0)
Composition of reservoir solution	0.1 M magnesium chloride, 0.1 M sodium HEPES (pH 7.5) and 10 % (w/v) PEG 4000
Volume and ratio of drop	2 μl drop, 1:1 ratio
Volume of reservoir (μl)	500

Table 3

Data collection and processing

Values for the outer shell are given in parentheses.

Crystal form	Bc-TssK	SeMet-Bc-TssK
Diffraction source	i03	i04-1
Wavelength (Å)	0.97622	0.92820
Temperature (K)	100	100
Detector	DECTRIS PILATUS 6M-F	DECTRIS PILATUS 6M-F
Rotation range per image (°)	0.1	0.1
Total rotation range (°)	200	360
Exposure time per image (s)	0.05	0.04
Data-processing package	Xia2-DIALS	Xia2-DIALS
Space group	$P6_3$	$P6_3$
a, *b*, *c* (Å)	100.0, 100.0, 212.7	99.9, 99.9, 214.9
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Cell volume (Å^3)	1843475	1856556
Resolution range (Å)	86.64-2.50 (2.54-2.5)	214.9-3.60 (3.66-3.60)
Total No. of reflections	464498 (21861)	289942 (15082)
No. of unique reflections	41618 (2052)	14109 (704)
Completeness (%)	100.0 (100.0)	100.0 (99.3)
Redundancy	11.2 (10.7)	20.6 (21.4)
$\langle I/\sigma(I) \rangle$	9.6 (0.9)	12.9 (4.6)
R_{merge}	0.171 (2.967)	0.210 (0.828)
R_{pim}	0.054 (0.958)	0.048 (0.182)
CC-half	0.998 (0.420)	0.983 (0.933)
Overall *B* factor from Wilson plot (Å^2)	46	52
Anomalous completeness	N/A	100.0 (5.1)
Anomalous multiplicity	N/A	10.4 (10.8)
Anomalous correlation	N/A	-0.061 (0.007)
Anomalous slope	N/A	1.002

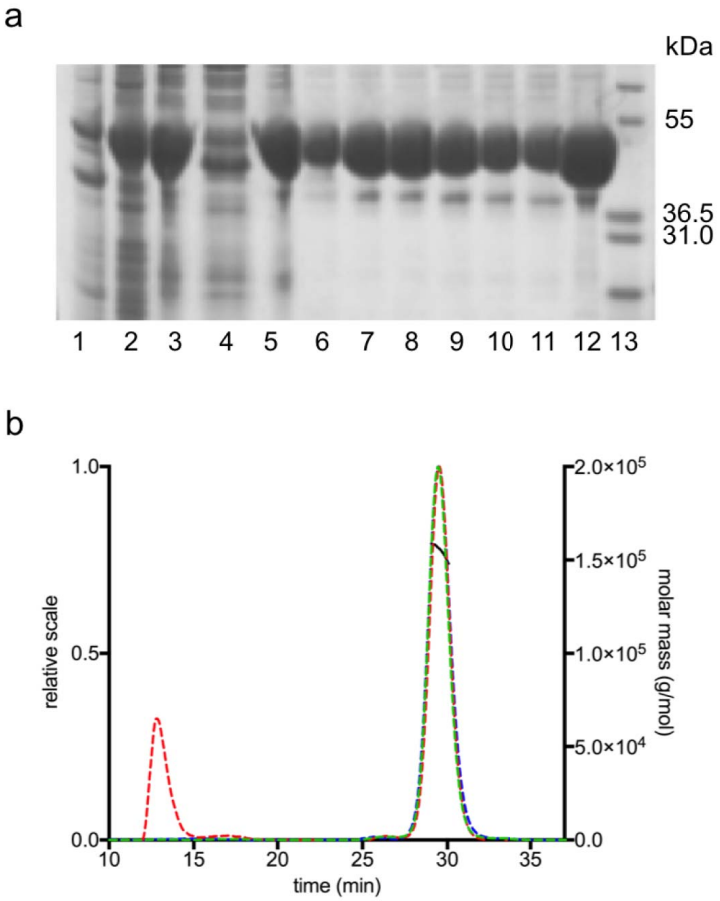


Figure 1

Purification and analysis of Bc-TssK. (*a*) Typical SDS-PAGE analysis of protein samples taken at different stages throughout the protein purification of Bc-TssK. Lane 1 - Cell debris. Lane 2 - Cell lysate. Lane 3 - Sample after DEAE ion exchange. Lane 4 - Soluble fraction after $AmSO_4$ precipitation. Lane 5 - Re-dissolved precipitate after $AmSO_4$ step. Lanes 6-11 - Fractions taken from gel filtration. Lane 12 - Final sample. Lane 13 - Mark12 protein ladder. The minor protein species at ~ 36 kDa is present throughout purification and is thought to represent a degradation product of Bc-TssK. (*b*) SEC-MALLS elution profile, peak positions, and molar mass plot of Bc-His_6.TssK (~155 kDa). The normalised signals for light scattering (LS) are shown in red, refractive index (RI) in blue, UV absorption (280 nm) in green, and molecular weight estimate curve shown in black. The molecular weight was determined by a Zimm fit procedure at each point. The X-axis represents the elution time in minutes.

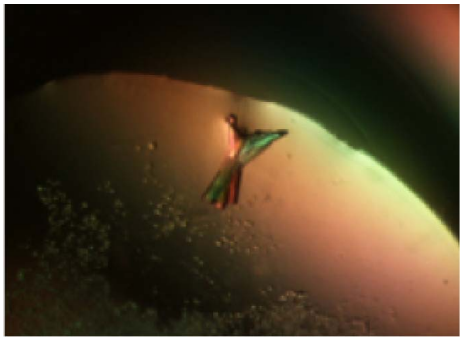


Figure 2

Crystals of Bc-TssK. Bc-TssK crystals grown optimally in 0.1 M magnesium chloride, 0.1 M sodium HEPES (pH 7.5) and 10 % (w/v) PEG 4000.

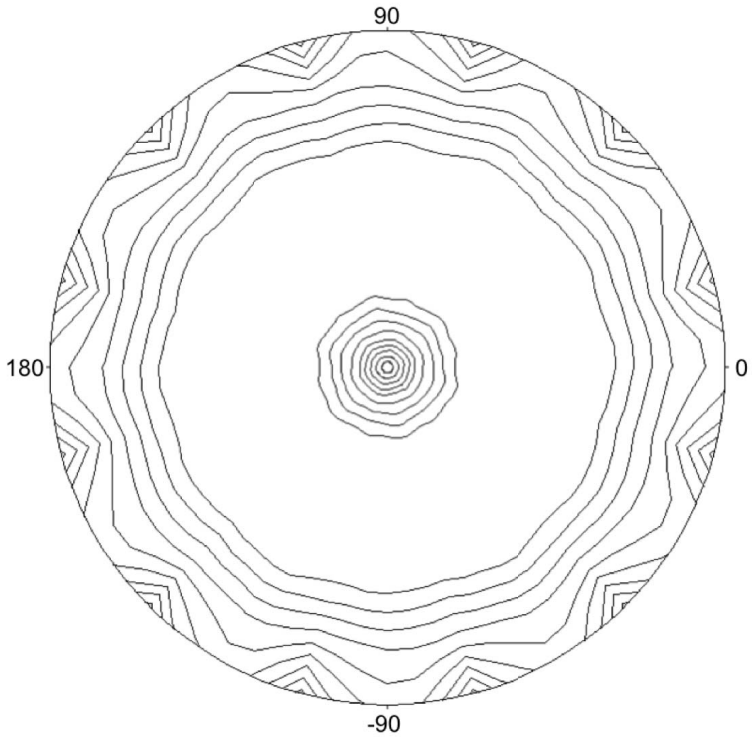


Figure 3

Self-rotation function plot of the Bc-TssK data. Graphical representation of a self-rotation function calculated on the kappa = 180° section on all data to 2.50 Å with an integration radius of 30 Å. The plot identifies the presence of non-crystallographic two-fold axes every 30°, at a peak height equal to over 90 % of the origin.

Acknowledgements

We would like to thank the beamline scientists on i03 and i04-1 at the Diamond Light Source, Oxford for their assistance during data collection and Andrew Leech, University of York, for carrying out the SEC-MALLS analysis.

Funding information:

This work was supported by a University of Sheffield scholarship and a BBSRC Doctoral Training Grant (BB/J014443/1) awarded to S.R.D. and H.J.O., respectively.

References

Aubert, D. F., Flannagan, R. S. & Valvano, M. A. (2008). Infection and Immunity, **76**, 1979–1991.

Basler, M., Pilhofer, M., Henderson, G. P., Jensen, G. J. & Mekalanos, J. J. (2012). Nature, **483**, 182–186.

Bönemann, G., Pietrosiuk, A., Diemand, A., Zentgraf, H. & Mogk, A. (2009). EMBO J. **28**, 315–325.

Brunet, Y. R., Henin, J., Celia, H. & Cascales, E. (2014). EMBO Reports, **15**, 315–321.

Brunet, Y. R., Zoued, A., Boyer, F., Douzi, B. & Cascales, E. (2015). PLoS Genetics, **11**, 21.

Cianfanelli, F. R., Alcoforado Diniz, J., Guo, M., De Cesare, V., Trost, M. & Coulthurst, S. J. (2016). PLoS Pathogens, **12**, 1–27.

Dix, S. R. *et al.* (2018). Structural insights into the function of type VI secretion system TssA subunits. *Nat. Commun.*, **9**, 4765.

Durand, E., Nguyen, V. S., Zoued, A., Logger, L., Pehau-Arnaudet, G., Aschtgen, M. S., Spinelli, S., Desmyter, A., Bardiaux, B., Dujeancourt, A., Roussel, A., Cambillau, C., Cascales, E. & Fronzes, R. (2015). Nature, **523**, 555–560.

English, G., Byron, O., Cianfanelli, F., Prescott, A. & Coulthurst, S. (2014). Biochemical Journal, **461**, 291–304.

Kube, S., Kapitein, N., Zimniak, T., Herzog, F., Mogk, A. & Wendler, P. (2014). Cell Reports, **8**, 20–30.

Leiman, P. G., Basler, M., Ramagopal, U. A., Bonanno, J. B., Sauder, J. M., Pukatzki, S., Burley, S. K., Almo, S. C. & Mekalanos, J. J. (2009). Proceedings of the National Academy of Sciences of the United States of America, **106**, 4154–4159.

Lyczak, J. B., Cannon, C. L. & Pier, G. B. (2002). Clinical Microbiology Reviews, **15**, 194–222.

Nazarov, S., Schneider, J. P., Brackmann, M., Goldie, K. N., Stahlberg, H. & Basler, M. (2018). EMBO J. p. e201797103.

Nguyen, V. S., Logger, L., Spinelli, S., Legrand, P., Huyen Pham, T. T., Nhung Trinh, T. T., Cherrak, Y., Zoued, A., Desmyter, A., Durand, E., Roussel, A., Kellenberger, C., Cascales, E. & Cambillau, C. (2017). Nature Microbiology, **2**, 1–9.

Pukatzki, S., Ma, A. T., Revel, A. T., Sturtevant, D. & Mekalanos, J. J. (2007). Proceedings of the National Academy of Sciences of the United States of America, **104**, 15508– 15513.

Römling, U., Fiedler, B., Bosshammer, J., Grothues, D., Greipel, J., von der Hardt, H. & Tummler, B. (1994). The Journal of Infectious Diseases, **170**, 1616–1621.

Shneider, M. M., Buth, S. A., Ho, B. T., Basler, M., Mekalanos, J. J. & Leiman, P. G. (2013). Nature, **500**, 350–353.

Spinelli, S., Desmyter, A., Verrips, C. T., de Haard, H. J. W., Moineau, S. & Cambillau, C. (2006). Nature Structural & Molecular Biology, **13**, 85–89.

Taylor, N. M. I., Prokhorov, N. S., Guerrero-Ferreira, R. C., Shneider, M. M., Browning, C., Goldie, K. N., Stahlberg, H. & Leiman, P. G. (2016). Nature, **533**, 346–352.

Wang, J., Brackmann, M., Castaño-Díez, D., Kudryashev, M., Goldie, K. N., Maier, T., Stahlberg, H. & Basler, M. (2017). Nature Microbiology, **2**, 1507–1512.

Winn, M. D. *et al*. (2011). Acta Cryst. D**67**, 235-242.

Winter, G. (2010). Journal of Applied Crystallography, **43**, 186–190.

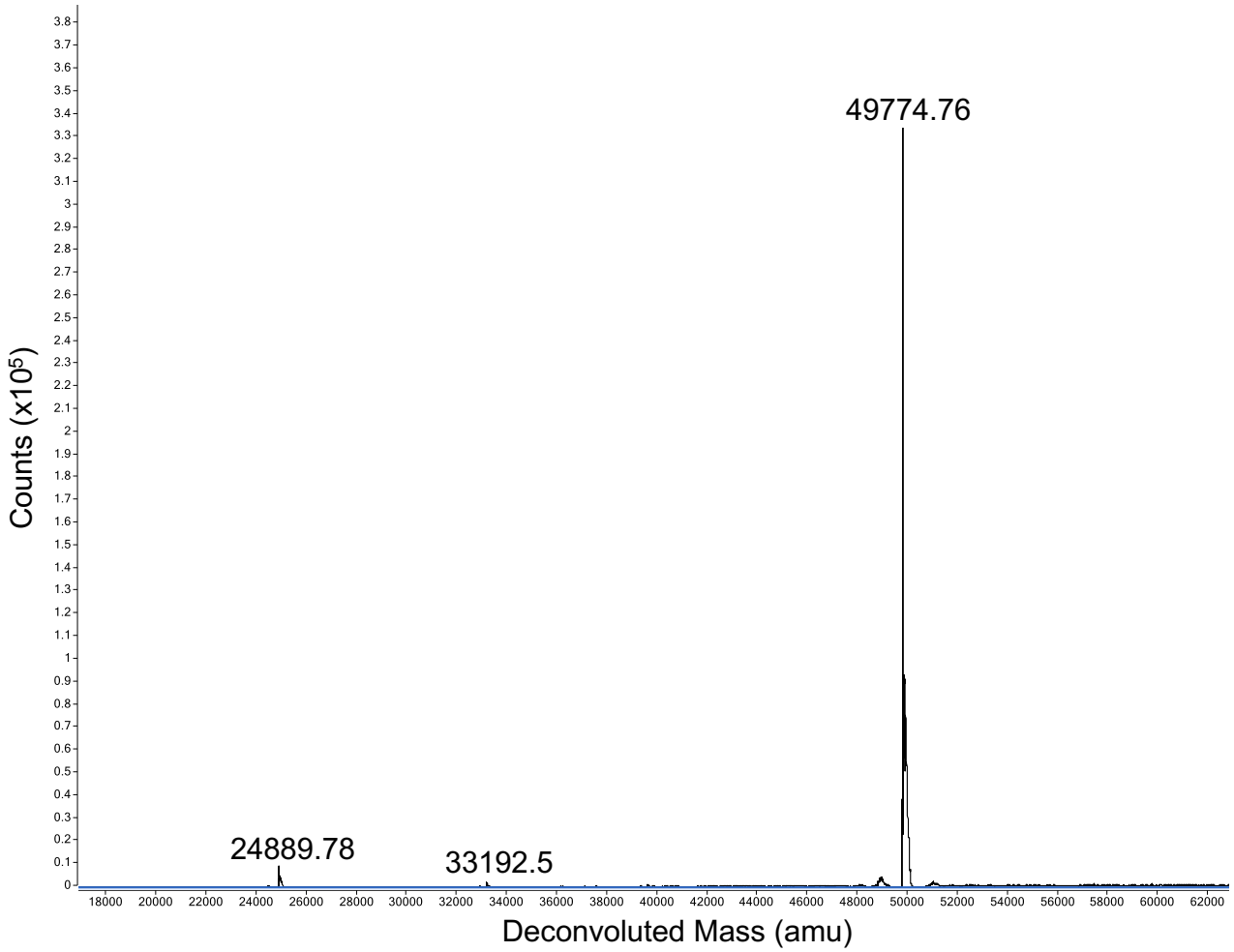
Winter, G. & McAuley, K. E. (2011). Methods, **55**, 81–93.

Zheng, J. & Leung, K. Y. (2007). Molecular Microbiology, **66**, 1192–1206.

Zoued, A., Durand, E., Bebeacua, C., Brunet, Y. R., Douzi, B., Cambillau, C., Cascales, E. & Journet, L. (2013). Journal of Biological Chemistry, **288**, 27031–27041.

supporting information

Mass spectrometry of Bc-TssK crystals. Mass spectrometry identified a peak at 49774.76 Da which is consistent with the full-length Bc-TssK protein, without the N-terminal methionine.



3.4 Paper IV

3.4.1 Summary

The final paper in this thesis discusses the X-ray structure determination and interaction studies carried out on TssK from *B. cenocepacia* (Bc-TssK) and *B. pseudomallei* (Bp-TssK). The paper describes the interaction studies between Bc-TssK and its constituent domains with that of other T6SS subunits. These results indicate a broadly similar pattern of interaction as that of EAEC TssK (Ec-TssK), despite the radically different structure and symmetry of Bc-TssA compared to Ec-TssA. The paper also discusses the structure determination of Bc-TssK and Bp-TssK, indicating a strong structural similarity between one another, and to the Ec-TssK orthologue. Furthermore, structural comparisons of the constituent head domain between Bc-TssK and Ec-TssK identifies a significant change in orientation and position between the species variants, suggesting that domain mobility is a key feature of TssK. Lastly, the paper describes the location of conserved residues within each domain of TssK, drawing comparisons with the structurally similar receptor binding proteins of *Lactococcal* siphophage, which could indicate specific roles for each domain within TssK.

3.4.2 Protein summary

TABLE 3.4: Summary of TssK proteins discussed

Paper IV				
Construct ID	**Residue range**	**Resolution**	**Space group**	**PDB ID**
Bc_TssK	(1-478)	2.2 Å	P63	N/A
Bc_TssK (Se)	(1-478)	3.0 Å	P63	N/A
Bc_TssK$_{SN}$	(1-317)	1.6 Å	H32	N/A
Bp_TssK$_{SN}$	(1-328)	2.2 Å	H32	N/A
Ec_TssK	(1-445)	2.6 Å	$P2_12_12_1$	5M30

3.4.3 Contribution

M.S.T. conceived the investigation; S.R.D., R.S., M.S.T. and D.W.R. designed them experiments; S.R.D., R.S., A.A., A.E.M., S.E.S. and H.J.O. performed the experiments; S.R.D., R.S., P.J.B., D.W.R and M.S.T. interpreted the data and S.R.D., M.S.T and D.W.R. wrote

the manuscript. Experimentally, R.S., A.A., and M.S.T. carried out the two-hybrid analysis. H.J.O crystalised the truncated fragment of Bc-TssK. I carried out the generation of the Bc/Bp-TssK domain constructs, their expression, purification and crystallisation. In addition, I carried out the X-ray data collection, structure determination and structural analysis of all Bc/Bp-TssK protein crystals and subsequent models.

The role of mobility and sequence conservation in the structure/function relationships of *Burkholderia* TssK

Samuel R. Dix[1], Ruyue Sun[2], Asma Ahmad[2], Amy E. Milburn[1], Hayley J. Owen[1], Sheila Nathan[3], Firdaus Raih[3], Svetlana E. Sedelnikova[1], Patrick J. Baker[1], Mark S. Thomas[2,*] and David W. Rice[1,*]

[1]Department of Molecular Biology and Biotechnology, Krebs Institute, University of Sheffield, Sheffield S10 2TN, UK.

[2]Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield Medical School, Beech Hill Road, Sheffield S10 2RX, UK.

[3]School of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia.

[*]Corresponding authors

Correspondence: d.rice@sheffield.ac.uk (D.W.R.); m.s.thomas@sheffield.ac.uk (M.S.T.).

ABSTRACT

TssK is a core subunit of the type six secretion system and has been proposed to interact with subunits of the T6SS baseplate and membrane complex, acting as an adapter between the two. Interaction studies of *Burkholderia cenocepacia* TssK suggest a similar pattern of interactions are maintained as previously discovered, despite differences in symmetry of core components found within related T6SS systems. The structure determination of TssK from *B. cenocepacia* and *B. pseudomallei* shows that the head

domain is held in a completely different orientation with respect to *E. coli* TssK signifying mobility is a critical function. Furthermore, analysis of contacts at the trimer three-fold axis indicate order/disorder transitions of packing residues may be important for the role of TssK.

INTRODUCTION

The type VI secretion system (T6SS) is a multi-subunit protein complex anchored to the cytoplasmic membrane of many species of Gram-negative bacteria, which serves to deliver a variety of effector proteins into target cells. The T6SS can be described as an inverted contractile tail comparable to that of bacteriophage and R type pyocins[1–3]. These molecular machines utilise a baseplate platform for the assembly of a contractile tail, which once 'fired', expel a sharpened tube to penetrate target cells[1,3,4].

The T6SS secretion machinery comprises 14 core subunits organised into two distinct components. The first, the membrane complex, forms an integral membrane structure that spans the bacterial cell envelope and contains ten copies each of the outer membrane lipoprotein TssJ, and the inner membrane proteins TssL and TssM. Together these form a 1.7 MDa complex with five-fold symmetry which allows the outward passage of the inner tube and associated effector proteins and acts as an anchor for the second component, the injection machinery[5–8]. The injection machinery is comprised of two sub-complexes. One of these is the contractile tail, which consists of an inner tube formed from stacked, hexameric rings of TssD (Hcp) subunits, capped by a trimer of TssI (VgrG) and a single PAAR subunit. This in turn is surrounded by a contractile sheath consisting of polymerised heterodimers of TssB/C arranged as a six-start helix[2,4,9–11]. The contractile tail is attached to the second sub-complex, the baseplate, which contains multiple copies of the TssE, TssF, TssG and TssK subunits. During ‘firing’ the contractile sheath undergoes

a significant conformational rearrangement allowing for the inner tube and associated effector proteins to be driven out of the bacterial cell through the membrane complex[12–14].

Recently, the structure of an enteroaggregative *E. coli* (EAEC) TssK subunit (Ec-TssK) has shown that it is comprised of three domains: the shoulder (N-terminal domain (S)), neck (Mid-domain (N)), and head (C-terminal domain (H)), which assemble to form a trimer, where the quaternary structure is primarily driven by contacts between symmetry-related shoulder and neck domains[15,16]. Structural comparisons have revealed that previous suggestions of similarity between TssK and gp8 of the T4 phage were incorrect[17], but noted a similarity between the fold of the shoulder domain of TssK and that of one of the three domains, the shoulder domain, of receptor binding proteins (RPBs) from siphophage baseplates[16]. RBPs utilise their shoulder domains to interact with additional phage baseplate proteins, while the head domain mediates the attachment of the siphophage particle with the host membrane receptors[15,18]. These data, coupled with other biophysical and biochemical studies on TssK, led to the suggestion that TssK acts as an adapter protein between the baseplate and membrane complexes of T6SS, with the shoulder domain interacting with the former and the head domain the latter[19–21].

In this paper, we show that *Burkholderia cenocepacia* TssK (Bc-TssK) interacts with the same T6SS subunits as Ec-TssK, including TssA, despite the radical difference in structure and symmetry of the Bc-TssA oligomer in comparison to that of Ec-TssA. In addition, we describe the X-ray structure of full-length TssK from *Burkholderia cenocepacia* (Bc-TssK) and the corresponding shoulder and neck domains of *Burkholderia pseudomallei* TssK

(Bp-TssK$_{SN}$), to expand the structural repertoire of this key component of T6SS. These studies have allowed us to identify conserved clusters of residues on the surface of TssK, highlighting their role in the formation of the protein's quaternary structure. Critically, the structures presented here provide new insights into the conformational flexibility of the TssK head domain with regards to the rest of the molecule and the impact of this on the functional role of TssK in T6SS assembly.

RESULTS AND DISCUSSION

Interaction of Bc-TssK with components of the T6SS

Previous work, from a range of species, have shown has shown that TssK interacts with subunits within the baseplate (the TssF-TssG complex), the contractile tail (Hcp, TssC and TssA) and the membrane complex (TssL and TssM[13,17,19,20,22]). However, given the diverse symmetry identified in TssA orthologues from different species which display 5, 6 and 16-fold symmetry (TssA2^{A}, TssA2^{B} and TssA1^{B}, respectively[22]), it was important to confirm the similarity in the pattern of interaction of *B. cenocepacia* TssK (Bc-TssK) with major components of the *B. cenocepacia* T6SS. Two-hybrid assays indicate that Bc-TssK interacts with TssA1^{B}, -D, -F, -I and -L, suggesting that it is in intimate contact with the membrane complex, the baseplate and the inner tube in the assembled T6SS (Figure 1a). A separate two-hybrid analysis of individual regions of Bc-TssK (the shoulder domain, a combination of the shoulder and neck domains, and the head domain) with other T6SS subunits revealed that, while no interactions were identified for the Bc-TssK shoulder domain on its own, the combined shoulder-neck domains interacted with TssA and TssI (Figure 1b). In addition, the head domain interacts with the cytoplasmic domain of TssL

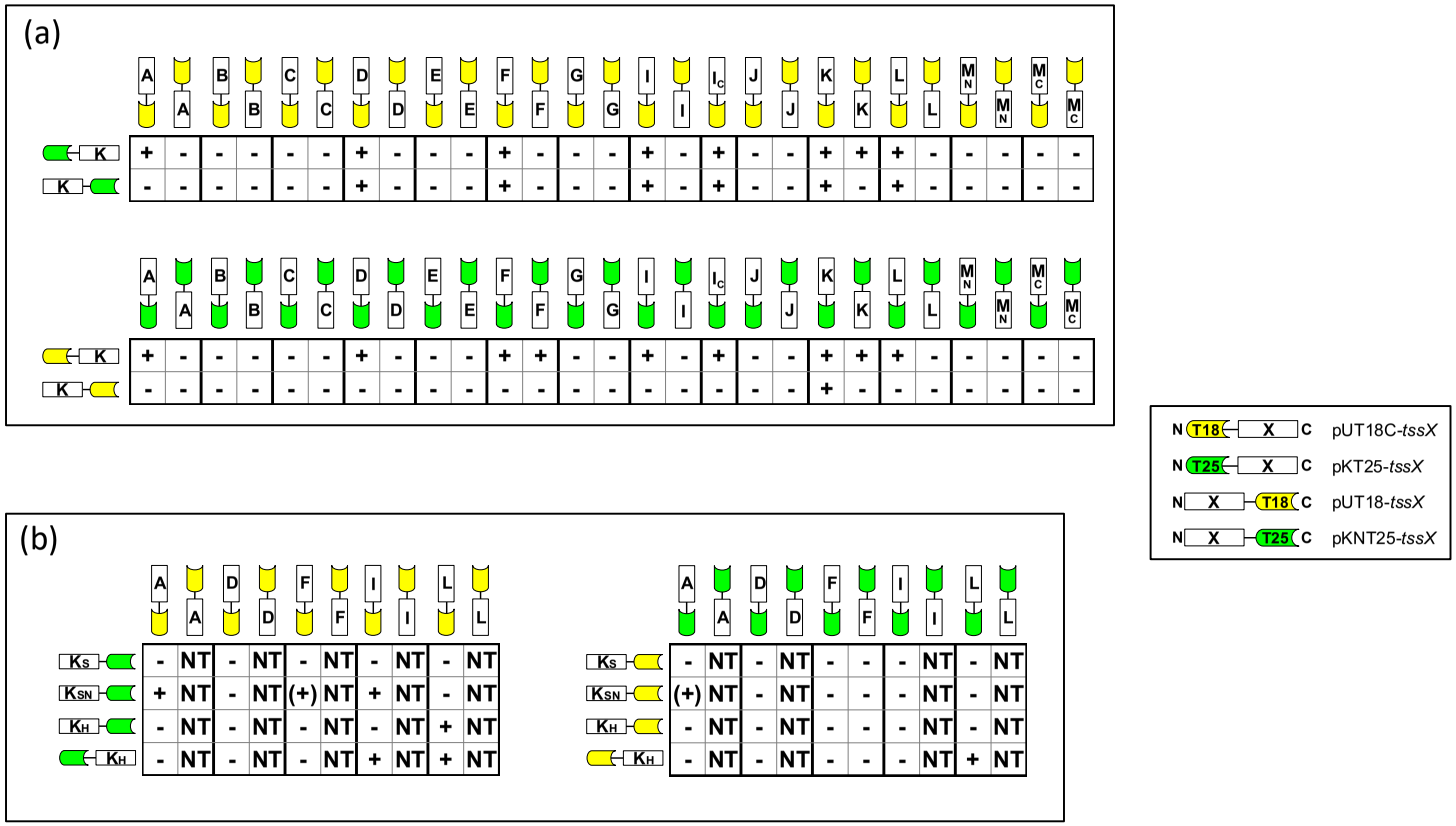


Figure 1 Two-hybrid analysis of the interaction of Bc-TssK and its domains with other T6SS subunits. Each hybrid protein is represented by a green or yellow coloured motif representing the CyaA fragment (T25 and T18, respectively) linked to a white rectangle labelled according to the fused T6SS subunit, as shown in the key. Interactions of the cytoplasmic N- and periplasmic C-terminal regions of TssM (M_N and M_C, respectively) with TssK were investigated separately. The interaction of TssK with the gp27.gp5-like core region of TssI (I_C) was also investigated. Interactions of TssK with TssH (ClpV) and PAAR were not investigated. All four possible hybrid proteins were analysed for each T6SS subunit except for $TssI_C$ where only two of the four possible hybrid proteins were constructed. Maltose colony phenotypes are indicated by '+', Mal^+; '(+)', intermediate; '-', Mal^-. The positive control gave rise to a strong Mal^+ phenotype and all negative control combinations (two compatible 'empty' vectors or one empty vector in combination with a plasmid specifying a hybrid protein) gave rise to Mal^- colonies (results not shown). **(a.)** Plasmids encoding Bc-TssK hybrid proteins tested in pairwise combinations with plasmids expressing T6SS subunit hybrid proteins. The data for the interaction between TssA and TssK has been published previously (Dix et al., submitted) and is presented here for comparison. **(b.)** Pairwise combinations of plasmids encoding the Bc-TssK shoulder (K_S), head (K_H) or shoulder-neck region (K_{SN}) fused to the CyaA T18 and T25 components in combination with hybrid proteins containing TssA, -D, -F, -I and -L as indicated.

in three of the four possible combinations of hybrid proteins tested. These results support the hypothesis that the combined Bc-TssK shoulder and neck domains can interact with T6SS baseplate components, whereas the C-terminal head domain interacts with the membrane complex[21]. Taken together, this indicates a very similar pattern of interaction for Bc-TssK to that observed elsewhere, despite the difference in symmetry of TssA. The only difference noted was the additional interaction of Bc-TssK with the core region of TssI, an association not detected in previous studies[19].

Structure determination of Bc-TssK and Bp-TssK$_{SN}$

The structure of crystals of a full-length construct of Bc-TssK (crystal form A) was initially determined using SAD phasing to 2.9 Å. Examination of the electron density map identified two molecules in the asymmetric unit, and good density could be observed for the shoulder and neck domains of each Bc-TssK molecule. While analysis by mass spectrometry confirmed that crystals contained full-length Bc-TssK (Dix et al., submitted), at that stage, there was essentially no density for either head domain, despite considerable space for their domains suggesting that they might be disordered in the crystal.

A model derived from the form A structure was then used in molecular replacement to solve and subsequently refine the structure of a second crystal form (form B) of Bc-TssK using data to 2.3 Å. The structure solution indicated that the latter crystals contain two molecules in the asymmetric unit, which had been clipped at or around residue V316 such that the resultant structure represented that of the shoulder and neck domains alone. The discovery of the nature of the clipping led to the design of a new construct corresponding to just the shoulder and neck domains of Bc-TssK (Bc-TssK$_{SN}$) (1-317), and the purified protein produced a new crystal form with improved diffraction qualities and with a single

123 molecule of Bc-TssK$_{SN}$ in the asymmetric unit. This structure was solved by molecular
124 replacement and refined to 1.6 Å (Table 1), and subsequently, the refined model was used
125 to improve the refinement of a higher-resolution (2.2 Å,) dataset of the form A crystals
126 following crystal optimisation (Table 1). Examination of the new electron density map of
127 the latter structure led to the observation of density for the head domain for one of the
128 two molecules in the asymmetric unit (subunit A), enabling the similarity of the fold of
129 the chain to that of the head domain of Ec-TssK to be recognised. This permitted the
130 structure of the latter to be superposed and a preliminary model of the head domain
131 constructed. In the final electron density map of Bc-TssK, good side chain density was
132 observed for the majority of the residues in the shoulder and neck domains of both
133 subunits, but the density associated with the head domain is generally poorer, with only
134 ~ 50 % of the sidechains identifiable and with discontinuous density in three of the inter-
135 connecting loops. The partial disorder in the head domain is reflected in higher B-factors
136 for this part of the structure (122 Å^2) compared to that of the two shoulder (52 Å^2) and
137 neck domains (60 Å^2) (Supp. Fig. 1c). In addition, this partial disorder is similar to that
138 observed for the single head domain in crystals of full-length Ec-TssK. As a result of the
139 poor electron density for the head domain, in the final Bc-TssK model, we have elected to
140 model this region as a poly-alanine backbone. The final model of form A Bc-TssK at 2.2 Å
141 has two molecules of Bc-TssK (chains A + B) in the asymmetric unit, one of which is full-
142 length (chain A), while the other (chain B) consists of residues from the shoulder and
143 neck domains alone (Bc-TssK$_{SN}$).
144 In parallel with the structural analysis of Bc-TssK, a construct representing the shoulder
145 and neck domains of TssK from one of the six independent T6SSs in *B. pseudomallei*
146 (T6SS-1) (residues 1-328), that corresponds to the sole *B. cenocepacia* T6SS^{23}(Spiewak et
147 al., submitted), was produced and over-expressed in *E. coli* (BPSL3110; Bp-TssK$_{SN}$). This

148 led to the production of crystals that enabled the structure (S4-T315) to be determined
149 to 2.25 Å by molecular replacement (Table 1). The structures of Bc-TssK$_{SN}$ and Bp-TssK$_{SN}$
150 are closely related to that of full-length Bc-TssK (0.57 and 0.54 Å RMSD over 292 and 291
151 equivalenced residues, respectively) (Supp. Fig. 1b). Therefore, the remainder of the
152 manuscript focuses on the analysis of the 2.2 A structure of the full-length Bc-TssK, with
153 comparisons to other structures where appropriate.
154
155 **The fold and quaternary structure of Bc-TssK**
156
157 The structure determination of Bc-TssK shows the fold of each of the three domains is
158 closely related to that of the Ec-TssK (PDB: 5M30) (2.5 Å RMSD over 264 equivalenced
159 residues (full length) and 1.8 Å RMSD over 250 equivalenced residues from the shoulder
160 and neck domains) to which it shares 33 % sequence identity (Figure 2a, b). The N-
161 terminal shoulder domain (Y3-P190) contains an extended loop region which precedes
162 the first helix, α1. This is then connected to a β-sandwich motif (β1-4/6/7), which
163 contains an elongated loop comprising β5/6 (Figure 2a) and the short α2 helix. Lastly,
164 β8 interacts with α2 and β7 as the polypeptide chain transitions into the neck domain.
165 An interdomain linker (C191-P198) connects the shoulder domain to the start of the neck
166 domain, which comprises a four α-helical bundle (α3-6; I199-V316) (Figure 2a). In the
167 final electron density map, residues Y219-G235, which form the link between α3 and α4,
168 cannot be identified and are assumed to be disordered. The last visible residue of the neck
169 domain, in subunit A, is V316, after which the density ends abruptly and does not
170 reappear until the first residue of the head domain, Y321. The head domain of subunit A
171 (Y321-T447) folds into a structure consisting of eight β-strands (β9-16) and three α-

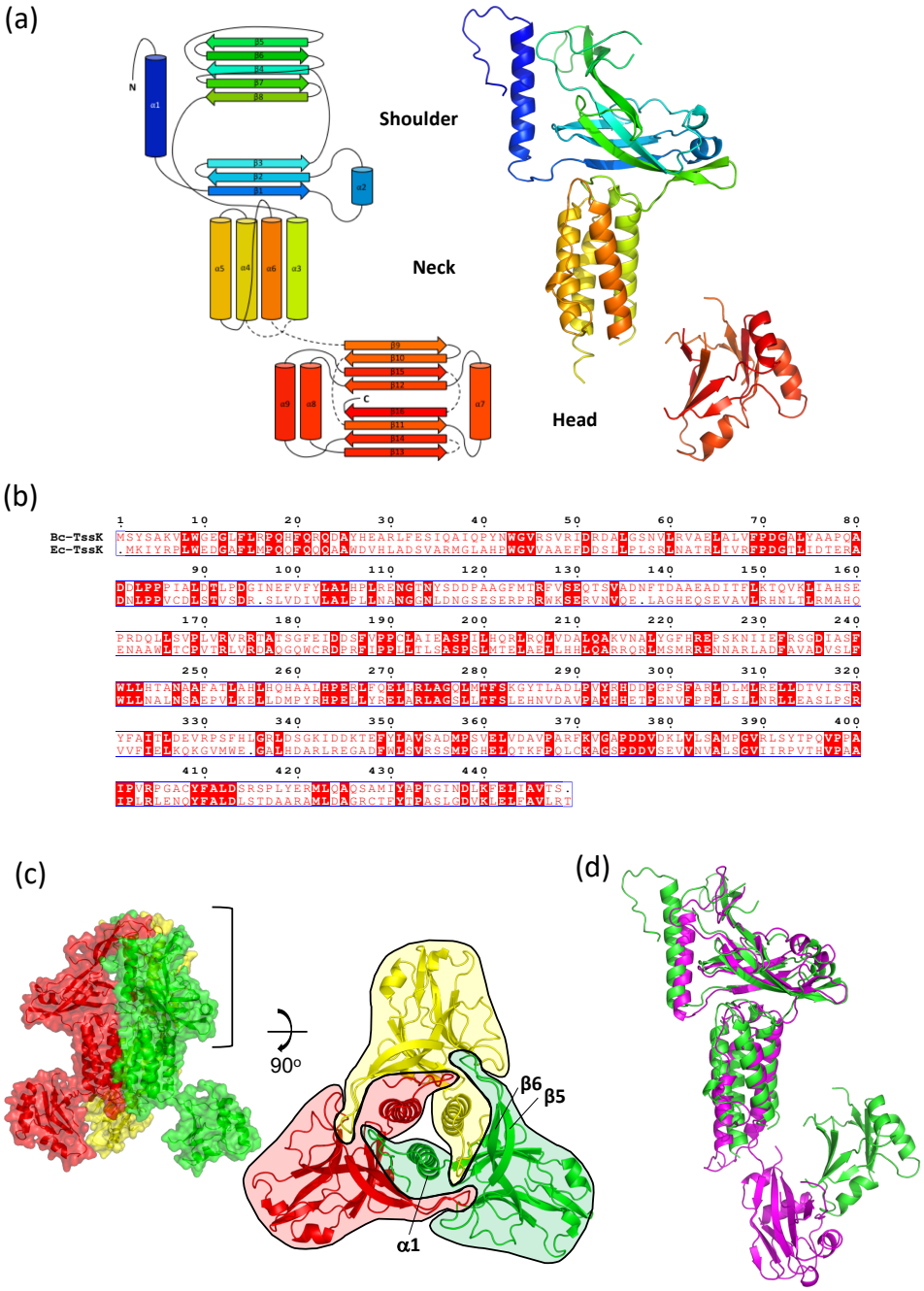


Figure 2 – Structural analysis of Bc-TssK (**a.**) (Left) 2D topology map of Bc-TssK, indicating the position in the structure of nine α-helices and 16 β-strands. (Right) 3D model of the Bc-TssK monomer showing the relative position of the shoulder, neck and head domain in subunit A. Both models are coloured in PyMOL chainbow, with the N-terminus shown in blue and the C-terminus red. (**b.**) Amino acid sequence alignment of Bc-TssK and Ec-TssK. Amino acid residues that are identical at the corresponding position in both proteins (33 % of the total) are highlighted in dark red. (**c.**) (Left) Surface representation of the Bc-TssK trimer formed by subunit A. (Right) View of the shoulder domains within the Bc-TssK trimer down the three-fold symmetry axis, to illustrate the interdigitated contacts between the subunits. This interface involves the N-terminal residues and α1, also β5 and β6 and their inter-connecting loop. (**d.**) Structural alignment between the shoulder and neck domains of Bc-TssK (green) and Ec-TssK (magenta). This shows the large displacement of the head domain between the respective models.

helices (α7-9) (Figure 2a), but with density for three interconnecting loops (G341-K346, P395-A407 and T434-L439) being weak and uninterpretable.

The two monomers in the asymmetric unit of the cell of the form A crystals each form independent trimers with three-fold symmetry related molecules in the cell. This produces trimers with an overall hourglass shape (Figure 2c) and dimensions of ~ 100 Å along the direction of the three-fold axis and with an approximate radius of ~50, 20, and 60 Å at the shoulder, neck, and head domains, respectively. This resembles the quaternary structure displayed by the Ec-TssK homologue. Trimer assembly is driven by extensive interactions that occur exclusively between the shoulder and neck domains which buries ~ 7 % (~2600 Å^2) of the solvent accessible area per monomer[24]. Trimerisation between the shoulder domains is achieved through the formation of a hydrophobic core about the three-fold symmetry related axis of the trimer. This is mediated through residues contributed from α1 and its symmetry-related partners. These are then surrounded by a highly intertwined, largely hydrophilic, interface created where the β5/6 loop of one subunit interdigitates and mediates interactions, on one face, with residues N137 and F138 of the elongated loop between β4/5 of the adjacent subunit. The other face of the β5/6 loop interacts with the N-terminal end of α1 (Q18), and the N-terminal loop preceding α1 (K6 & L8) from the third subunit of the trimer (Figure 2c). The highly interlocked and complex nature of this interface implies that conformational flexibility of the TssK domains, specifically the N-terminal loop and α1, is important for the assembly of the trimer. Interactions within the trimer also involve residues of the neck domain and are formed by the packing of the four-helical bundles (α3-6) against each other, maintained through a largely hydrophobic interface involving helices α4 and α5.

196

197 **Comparison of the Bc-TssK and Ec-TssK structures reveals mobility of the head**
198 **domain.**

199

200 Superposition of the Bc-TssK and Ec-TssK structures, clearly shows that the relative
201 orientation of the shoulder and neck domains are very similar whereas the position of
202 the head domain is quite different, such that a maximal displacement of ~ 45 Å can be
203 observed between them. In addition, the Bc-TssK and Ec-TssK head domains are
204 orientated quite differently as a result of a further rotation of ~ 120° about an axis
205 perpendicular to the three-fold axis of the trimer, near the C-terminus of the neck domain
206 (Figure 2d)[25].
207 In the crystal structures of both Bc-TssK and Ec-TssK, there are no contacts between the
208 shoulder/neck region and their respective head domains, so the positions of the latter
209 are maintained only via non-equivalent adventitious crystal contacts. In the Bc-TssK head
210 domain, the residues involved in forming these contacts are primarily from one face of
211 the head domain including residues, T325, D327, R337 and D339. (~ 5 % of the solvent
212 accessible area) The electron density for these residues is well defined and the
213 temperature factors for residues on this face of the head domain are generally
214 comparable with those in the shoulder and neck domain. In contrast, residues in the head
215 domain distant from the crystal contacts the electron density is poorer and the associated
216 temperature factors are considerably higher. This suggests that there may be some
217 residual mobility of the domain in the crystal consistent with the limited direct contacts
218 made between the molecules (Supp. Fig 1c). These data, together with the absence of
219 density for the linker between the neck and head domains in subunit A of Bc-TssK, for
220 both the linker and head domains in subunit B of Bc-TssK and the head domains for two

of the three molecules in the asymmetric unit for Ec-TssK suggests that in the full-length protein the head domain is mobile. This is consistent with the absence of the head domain in the ~ 26 Å EM structure of Ec-TssK[19]. In light of this finding, we suggest that the head domain acts like a 'ball on the end of a tether' such that it can sample surrounding space to mediate its attachment to the membrane complex, with dramatically different conformations being possible.

Sequence conservation and structural similarities of Bc-TssK to siphophage proteins

To examine the role of conserved residues in Bc-TssK a multiple sequence alignment of 150 related TssK homologues from UNIREF90 was generated using MAFFT[26], and subsequently mapped onto the Bc-TssK structure using Consurf[27–29]. This identified three 3D clusters of conserved residues (clusters 1, 2 and 3), one in each of the domains of the Bc-TssK structure (Figure 3a, Supp. Fig. 1a). Conservation in the shoulder domain (cluster 1) involves 12 conserved residues found within the N-terminal loop (W9-R16) and the N-terminal end of α1 (P17-Q23) (Figure 3a). While these residues fall within a region which is structurally conserved between TssK and *lactococcal* siphophage RBPs, including phage 1358 (Figure 3b) and p2[16], they are not conserved. The N-terminal loop and N-terminus of α1 in the RBP proteins have been shown to mediate an interaction with the 'arm and hand' motif of additional baseplate proteins, as seen in the p2 baseplate structure[18]. Moreover, structural studies on the RBP of Lactococcal phage TP901-1 (P335-species)[30] where the shoulder domain is a minimal unit consisting of solely of the N-terminal loop and α1 (Figure 3b) also shows that it makes interactions with the BppU baseplate protein while being devoid of the β-sandwich motif. Furthermore, antibody-

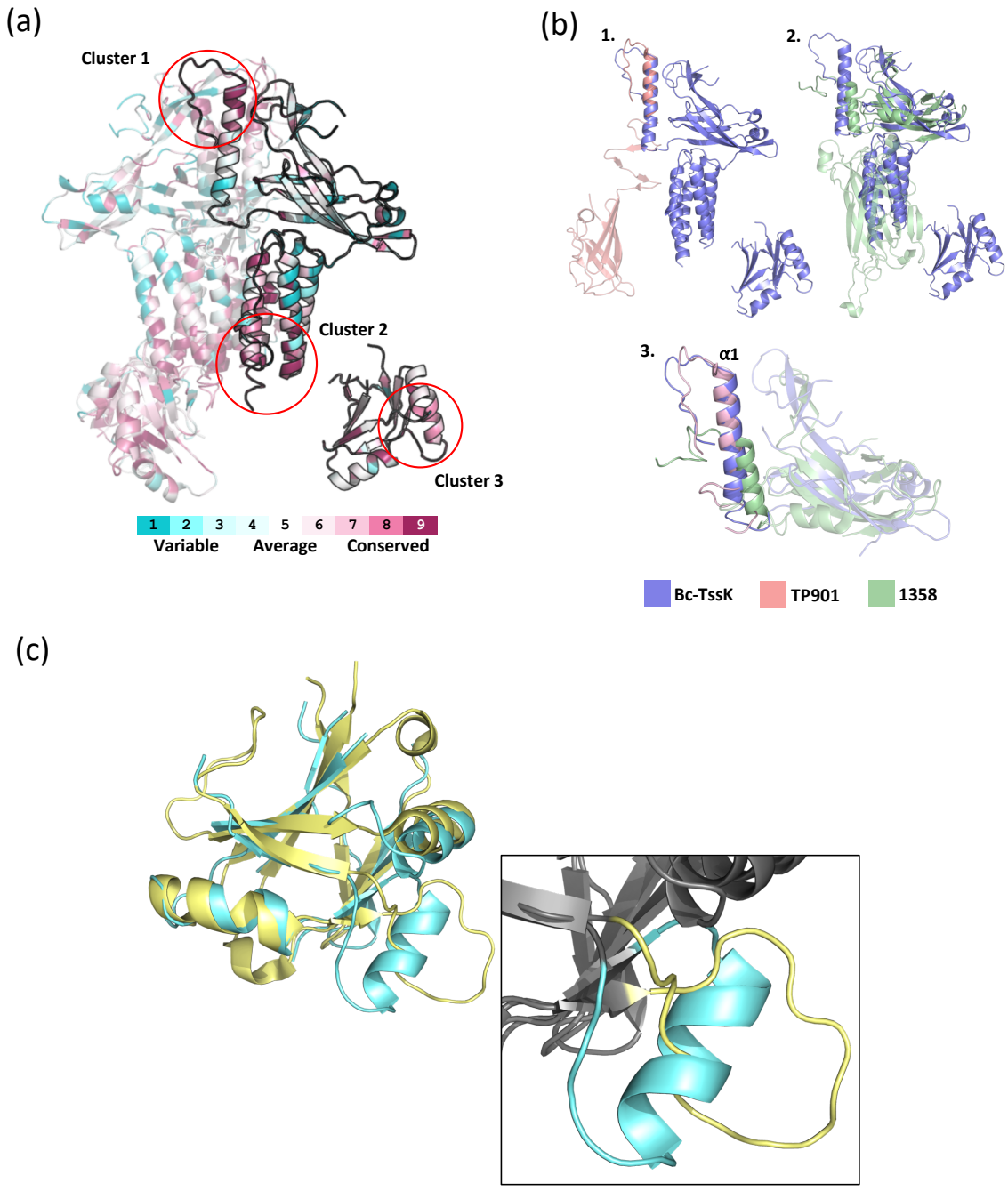


Figure 3 - 3D mapping of conserved residues within TssK. (**a.**) Representation of the Bc-TssK trimer coloured based on the level of conservation for each residue in the protein model using the Consurf colour scheme indicated by the key. Red circles indicate the three clusters of conservation (Clusters 1, 2 and 3). (**b.**) (1) Structural alignment of the Bc-TssK shoulder domain (blue) with RBP TP901 (pink). (2) Structural alignment of the Bc-TssK shoulder domain (blue) with RBP 1358 (green). (3) Structural alignment of the Bc-TssK shoulder domain (blue), RBP 1358 (green) and RBP TP901 (pink). Structural alignments of Bc-TssK with RBPs indicate a strong structural conservation of the N-terminal loop and α1, even with with diverse neck and head domains. (**c.**) Structural alignment of the Bc-TssK and Ec-TssK head domains (blue and yellow, respectively). The inset shows that residues forming α8 (P375-L383) in Bc-TssK adopt an extended loop in Ec-TssK despite this region containing five conserved residues.

246 binding experiments have shown that the TssK shoulder domain interacts with TssFG
247 during baseplate assembly[16], and this interaction is believed to mirror that of RBP
248 shoulder domains to siphophage baseplate proteins. It is, therefore, proposed that TssK
249 is utilising the conserved residues of cluster 1 to interact with an 'arm and hand' motif
250 equivalent in TssFG.
251
252 Conserved residues in the neck domain (cluster 2) are located at its distal end furthest
253 from the shoulder domain. These include seven residues within D236-T245, from the N-
254 terminal end of α4, and three residues, from the C-terminal end of α5 (T280-S282) which
255 together form part of the face that is buried during trimer formation. In addition, residues
256 K214/L218 at the C-terminal end of α3 and R310/L312/T315/V316 at the C-terminal
257 end of α6 form contacts that mediate the helical neck bundle. While good electron density
258 can be observed for the main chain of residues within the neck domain, such that the
259 location of the main chain and many of the sidechains is unambiguous, the density for the
260 connecting loop between α3 and α4 is absent, a feature shared by Bp-TssK and two of the
261 three monomers in the full-length structure of Ec-TssK. However, in both Bc-TssK and
262 Bp-TssK, poor density can be seen for the conserved sidechains of W241 and F281,
263 despite good density for their associated main chain and for the flanking residues. The
264 main chain of residue F241 in α4 lies close to that for W281 in α5 and both are close to
265 the three-fold symmetry axis. Nearby there is a large electron density feature, that is also
266 found in Bp-TssK (Supp. Fig. 2b, c), and a strong electron density feature on the three-fold
267 symmetry axis. Neither of these features can be accounted for by either residue alone,
268 and adverse steric clashes would be generated if both sidechains occupied either of these
269 features. The interpretation for the positions of side chains W241 and F281 was therefore
270 unclear and could not be resolved by the use of omit maps. One possibility is that W241

271 and F281 occur in multiple conformations in the crystal, such that the density is averaged
272 (Supp. Fig. 2a.). We note that the corresponding two conserved residues in Ec-TssK, have
273 been modelled with deviations from three-fold symmetry which overcome this issue
274 (Supp. Fig. 2c). The conservation of residues in cluster 2, the absence of the loop between
275 α3 and α4, and the ambiguity with modelling residues W241 and F281 may suggest that
276 order/disorder transitions within this region are important in the role of TssK.
277
278 The conserved residues in the head domain (cluster 3) consist of residues K371 within
279 β11, three residues within G373-G388 of α8 and Y431 in β15. Preliminary comparisons
280 of this region in the Ec-TssK and Bc-TssK head domain structures suggest that there may
281 be some structural differences between these regions, despite the sequence conservation
282 (Figure 3c) such that residues P375-L383 are present as α8 in the head domain of Bc-
283 TssK, whereas the equivalent residue range in Ec-TssK head domain adopts an extended
284 loop. However, a detailed comparison is not justified at this stage, due to the uncertainties
285 associated with the current Bc-TssK head domain structure. Attempts to clarify the fold
286 of this domain are underway. If such a change in fold remains once the structure has been
287 clarified it may well be biologically significant, and speculatively, the conservation found
288 in this region of the TssK head domain could be attributed with the interaction with the
289 membrane complex; however further study would be required.
290
291 The structural studies on Bc-TssK and Bp-TssK presented here have shed further light on
292 the structure/function relationships for this molecule. Interactions studies of Bc-TssK
293 and its constituent domains have indicated a similar pattern of interaction with other
294 T6SS subunits as previously seen, despite the radically different symmetries displayed by
295 Ec-TssA and Bc-TssA. Furthermore, the disorder associated with conserved residues

296 found within the neck domain at the three-fold symmetry axis of TssK possibly indicates
297 the presence of multiple conformations for the sidechains which could be associated with
298 order/disorder transactions within this region as part of the function of TssK. In addition,
299 these studies have revealed the quite different positions that the TssK head domain can
300 occupy suggesting TssK is a highly dynamic molecule. Taken together, the studies on Bc-
301 TssK and Bp-TssK support the hypothesis that TssK acts as an adapter protein between
302 the membrane and baseplate complexes and suggest that the inherent mobility displayed
303 by TssK may complement a structural rearrangement of the baseplate complex upon a
304 firing event of the T6SS.
305
306 **METHODS**
307
308 **Bacterial strains and growth conditions**
309 Details of bacterial strains used in this work are given in Table 2. LB[31] was routinely used
310 for maintenance and propagation of bacterial strains. The T6SS subunit genes used for
311 construction of the two-hybrid plasmids and constructs for crystallisation were amplified
312 from *B. cenocepacia* H111 and *B. pseudomallei* D286.
313
314 **Plasmids and primers**
315
316 All plasmids employed for the bacterial two-hybrid assay have been previously
317 described[22] except for those expressing TssK domains. Plasmids constructed during this
318 investigation are listed in Table 3 and corresponding primers are shown in Table 4.
319 Routine cloning procedures were carried out in *E. coli* strains JM83 and DH5α.
320

321 **Bacterial two-hybrid assay**

322 The bacterial adenylate cyclase two-hybrid (BACTH) system was used, in which pairs of
323 compatible plasmids expressing the T25 and T18 components of *B. pertussis* CyaA fused
324 to potentially interacting proteins are tested for their ability to complement an *E. coli cya*
325 mutant[32]. *B. cenocepacia* T6SS subunit genes or domain coding sequences were fused to
326 the N- and C-terminal coding sequences of the T25 or T18 fragments present in plasmids
327 pKT25, pKNT25, pUT18 and pUT18C. Pairwise combinations of compatible plasmids
328 encoding a T18 and a T25 hybrid protein were introduced into the *E. coli* indicator strain
329 BTH101 and the resulting maltose (Mal) phenotypes were scored following incubation
330 on MacConkey agar containing maltose (1%), ampicillin (100 μg/ml) and kanamycin (50
331 μg/ml) for 120 h at 30 °C (colony phenotypes were recorded as Mal^+, dark red/deep
332 purple colonies; weak Mal^+ (i.e. intermediate phenotype), pink or 'patchy' red colonies;
333 Mal^-, white colonies). The positive control ('Zip') comprised a pair of compatible plasmids
334 encoding the leucine zipper of the yeast transcriptional activator protein, GCN4, fused to
335 T18 and T25 (Karimova et al., 1998). For each T6SS hybrid protein tested, a
336 corresponding negative control was also carried out using the plasmid encoding the
337 hybrid protein in combination with the empty vector used to express the potential
338 interacting protein.

339

340 **Expression and purification TssK constructs for crystallisation**

341 The pACYCDuet-tssK construct was generated, expressed and purified as previously
342 described in Dix et al., submitted, producing diffraction quality crystals of both native and
343 Selenomethionine incorporated protein. Both pET21-$tssK_{SN}$.His6 and pET21-
344 $BptssK_{SN}.His_6$ constructs were independently transformed into *E. coli* BL21 (DE3) and
345 grown in 1 L brain-heart infusion (BHI) broth at 30 °C prior to induction by addition of 1

346 mM IPTG. Cells were lysed via sonication in 50 mM Tris (pH 8.0) and Bc-TssK$_{SN}$ was
347 purified from cell lysate via nickel affinity due to the presence of a His$_6$-tag. In contrast,
348 Bp-TssK$_{SN}$ was purified from cell lysate by initial application onto a 5 ml DEAE FF
349 cartridge (GE Healthcare) and eluted over a NaCl gradient (0 - 0.5 M). After this, the
350 sample was precipitated using 4 M ammonium sulphate solution before being
351 resolubilised and applied on a 1.6x60 HiLoadSuperdex200 gel filtration column (GE
352 Healthcare).
353
354 **X-ray crystallography**
355 Samples of Bc-TssK$_{SN}$ and Bp-TssK$_{SN}$ were concentrated to 10 mg/ml prior to
356 crystallisation trials using sitting-drop vapour diffusion, whereby 100 nL protein solution
357 were equilibrating with an equal volume precipitant. Crystals of Bc-TssK were grown and
358 cryo-protected as previously described (Dix *et al.*, 2018, submitted) and data from native
359 and Selenomethionine incorporated Bc-TssK crystals were collected to 2.2 Å and 2.9 Å on
360 Diamond Light Source (DLS) MX beamlines I04 and I04-1, respectively and belong to
361 spacegroup P6$_3$ with two molecules in the asymmetric unit. Crystals of Bc-TssK$_{SN}$ were
362 produced in 0.2 M magnesium chloride hexahydrate, 0.1 M Tris (pH 7.0) and 10 % PEG
363 8000. Data from these crystals were collected to 1.6 Å on DLS MX beamline I04-1 and
364 belong to spacegroup H32 with one molecule in the asymmetric unit. Crystals of Bp-
365 TssK$_{SN}$ were grown in conditions 0.1 M Na acetate pH 5.0, and 2.0 M Na formate. Data
366 from these crystals were collected to 2.2 Å on DLS MX beamline I04 and belong to
367 spacegroup H32 with one molecule in the asymmetric unit. All datasets were processed
368 using the Xia2-DIALS (Bc-TssK selenomethionine data was processed using 3dii)
369 software package[33–36] and full collection statistics are given in Table 1. The structure of
370 Bc-TssK was solved using experimental phases obtained from Selenomethionine peak

SAD data using the software pipeline CRANK2[37]. Automatic model building was carried out using Buccaneer[38], before refinement against native Bc-TssK data at 2.2 Å. The structures of Bc-TssK$_{SN}$ and Bp-TssK$_{SN}$ were solved via molecular replacement using Phaser MR[39] using co-ordinates of the shoulder and neck domain of Bc-TssK as a search model. All models underwent successive rounds of manual model building in COOT[40] and refinement using REFMAC5[41,42]. Final model validation was carried out using Molprobity[43]. Structural alignment was carried out using GESAMT[44] based on the Cα's of the protein models. Structural conservation utilised the CONSURF servers[27–29].

REFERENCES

1. Leiman, P. G. *et al.* Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 4154–9 (2009).
2. Basler, M., Pilhofer, M., Henderson, G. P., Jensen, G. J. & Mekalanos, J. J. Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483,** 182–186 (2012).
3. Ge, P. *et al.* Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. *Nat. Struct. Mol. Biol.* **22,** 377–82 (2015).
4. Shneider, M. M. *et al.* PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* **500,** 350–353 (2013).
5. Zheng, J. & Leung, K. Y. Dissection of a type VI secretion system in Edwardsiella tarda. *Mol. Microbiol.* **66,** 1192–1206 (2007).
6. Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. & Attree, I. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics* **10,** 104

(2009).

7. Durand, E. *et al.* Biogenesis and structure of a type VI secretion membrane core complex. *Nature* **523,** 555–560 (2015).

8. Cianfanelli, F. R. *et al.* VgrG and PAAR Proteins Define Distinct Versions of a Functional Type VI Secretion System. *PLoS Pathog.* **12,** 1–27 (2016).

9. Brunet, Y. R., Hénin, J., Celia, H. & Cascales, E. Type VI secretion and bacteriophage tail tubes share a common assembly pathway. *EMBO Rep.* **15,** 315–321 (2014).

10. Kube, S. *et al.* Structure of the VipA/B type VI secretion complex suggests a contraction-state-specific recycling mechanism. *Cell Rep.* **8,** 20–30 (2014).

11. Wang, J. *et al.* Cryo-EM structure of the extended type VI secretion system sheath-tube complex. *Nat. Microbiol.* **2,** 1507–1512 (2017).

12. Pukatzki, S., Ma, A. T., Revel, A. T., Sturtevant, D. & Mekalanos, J. J. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc. Natl. Acad. Sci.* **104,** 15508–15513 (2007).

13. Brunet, Y. R., Zoued, A., Boyer, F., Douzi, B. & Cascales, E. The Type VI Secretion TssEFGK-VgrG Phage-Like Baseplate Is Recruited to the TssJLM Membrane Complex via Multiple Contacts and Serves As Assembly Platform for Tail Tube/Sheath Polymerization. *PLoS Genet.* **11,** e1005545 (2015).

14. Taylor, N. M. I. *et al.* Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **533,** 346–352 (2016).

15. Spinelli, S. *et al.* Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. *Nat. Struct. Mol. Biol.* **13,** 85–89 (2006).

16. Nguyen, V. S. *et al.* Type VI secretion TssK baseplate protein exhibits structural similarity with phage receptor-binding proteins and evolved to bind the

membrane complex. *Nat. Microbiol.* **2,** 1–9 (2017).

17. Planamente, S. *et al.* TssA forms a gp6-like ring attached to the type VI secretion sheath. *EMBO J.* **35,** 1613–1627 (2016).

18. Sciara, G. *et al.* Structure of lactococcal phage p2 baseplate and its mechanism of activation. *Proc. Natl. Acad. Sci.* **107,** 6852–6857 (2010).

19. Zoued, A. *et al.* TssK is a trimeric cytoplasmic protein interacting with components of both phage-like and membrane anchoring complexes of the type VI secretion system. *J. Biol. Chem.* **288,** 27031–41 (2013).

20. English, G., Byron, O., Cianfanelli, F. R., Prescott, A. R. & Coulthurst, S. J. Biochemical analysis of TssK, a core component of the bacterial Type VI secretion system, reveals distinct oligomeric states of TssK and identifies a TssK-TssFG subcomplex. *Biochem. J.* **461,** 291–304 (2014).

21. Nazarov, S. *et al.* Cryo-EM reconstruction of Type VI secretion system baseplate and sheath distal end. *EMBO J.* **37,** e97103 (2018).

22. Dix, S. R. *et al.* Structural insights into the function of type VI secretion system TssA subunits. *Nat. Commun.* **9,** 4765 (2018).

23. Shalom, G., Shaw, J. G. & Thomas, M. S. In vivo expression technology identifies a type VI secretion system locus in Burkholderia pseudomallei that is induced upon invasion of macrophages. *Microbiology* **153,** 2689–2699 (2007).

24. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372,** 774–797 (2007).

25. Hayward, S. & Berendsen, H. J. C. Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins Struct. Funct. Genet.* **30,** 144–154 (1998).

26. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence

446 alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 1–7
447 (2017). doi:10.1093/bib/bbx108
448 27. Landau, M. *et al.* ConSurf 2005: The projection of evolutionary conservation
449 scores of residues on protein structures. *Nucleic Acids Res.* **33,** 299–302 (2005).
450 28. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010:
451 Calculating evolutionary conservation in sequence and structure of proteins and
452 nucleic acids. *Nucleic Acids Res.* **38,** 529–533 (2010).
453 29. Celniker, G. *et al.* ConSurf: Using evolutionary data to raise testable hypotheses
454 about protein function. *Isr. J. Chem.* **53,** 199–206 (2013).
455 30. Veesler, D. *et al.* Suggests an Alternative Host Adhesion Mechanism. *Proc. Natl.*
456 *Acad. Sci.* **109,** 8954–8 (2012).
457 31. Miller, J. Experiments in Molecular Genetics. *Cold Spring Harb. Lab. Press. Cold*
458 *Spring Harb. NY* (1972).
459 32. Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system
460 based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci.* **95,**
461 5752–5756 (1998).
462 33. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol.*
463 *Crystallogr.* **62,** 72–82 (2006).
464 34. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).
465 35. Winter, G. Xia2: An expert system for macromolecular crystallography data
466 reduction. *J. Appl. Crystallogr.* **43,** 186–190 (2010).
467 36. Winter, G. & McAuley, K. E. Automated data collection for macromolecular
468 crystallography. *Methods* **55,** 81–93 (2011).
469 37. Skubák, P. & Pannu, N. S. Automatic protein structure solution from weak X-ray
470 data. *Nat. Commun.* **4,** 2777 (2013).

38. Cowtan, K. The Buccaneer software for automated model building. Tracing protein chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62,** 1002–1011 (2006).

39. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

40. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

41. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53,** 240–255 (1997).

42. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 355–367 (2011).

43. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 12–21 (2010).

44. Krissinel, E. Enhanced fold recognition using efficient short fragment clustering. *J. Mol. Biochem.* **1,** 76–85 (2012).

45. Romling, U. *et al.* Epidemiology of chronic Pseudomonas aeruginosa infections in cystic fibrosis. *J Infect Dis* **170,** 1616–1621 (1994).

46. Lee, S. *et al.* Burkholderia pseudomallei animal and human isolates from Malaysia exhibit different phenotypic characteristics. *Diagn. Microbiol. Infect. Dis.* **58,** 263–270 (2007).

47. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* **33,** 103–119 (1985).

48. Jessee, J. New subcloning efficiency competent cells: >1 × 106 transformants/μg. *Focus (Bethesda Res. Lab.* **8,** 9–10 (1986).

496 49. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct
497 selective high-level expression of cloned genes. *J. Mol. Biol.* **189,** 113–130 (1986).
498 50. Karimova, G., Ullmann, A. & Ladant, D. Protein-protein interaction between
499 Bacillus stearothermophilus tyrosyl-tRNA synthetase subdomains revealed by a
500 bacterial two-hybrid system. *J. Mol. Microbiol. Biotechnol.* **3,** 73–82 (2001).
501 51. Karimova, G., Dautin, N. & Ladant, D. Interaction network among Escherichia coli
502 membrane proteins involved in cell division as revealed by bacterial two-hybrid
503 analysis. *J. Bacteriol.* **187,** 2233–2243 (2005).

504

505 **ACKNOWLEDGEMENTS**

506 We would like to thank the Diamond Light Source for access to beamlines I04 and I04-1
507 that contributed to these results, and beamline scientists for their assistance with data
508 collection. This work was supported by a Royal Society International collaboration grant
509 IC170306, a BBSRC Doctoral Training Grant BB/J014443/1 awarded to H.J.O. and a
510 University of Sheffield scholarship awarded to S.R.D.

511

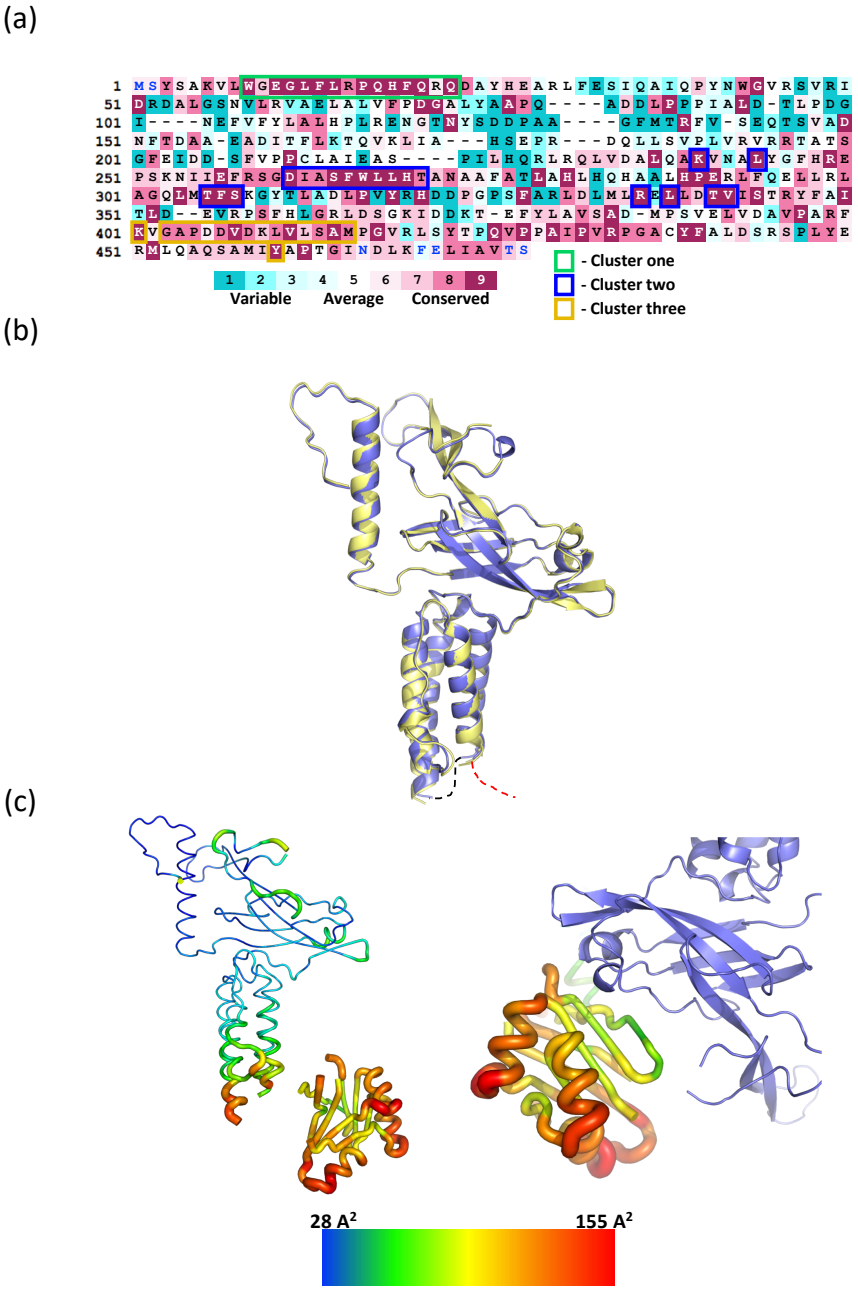
512

513

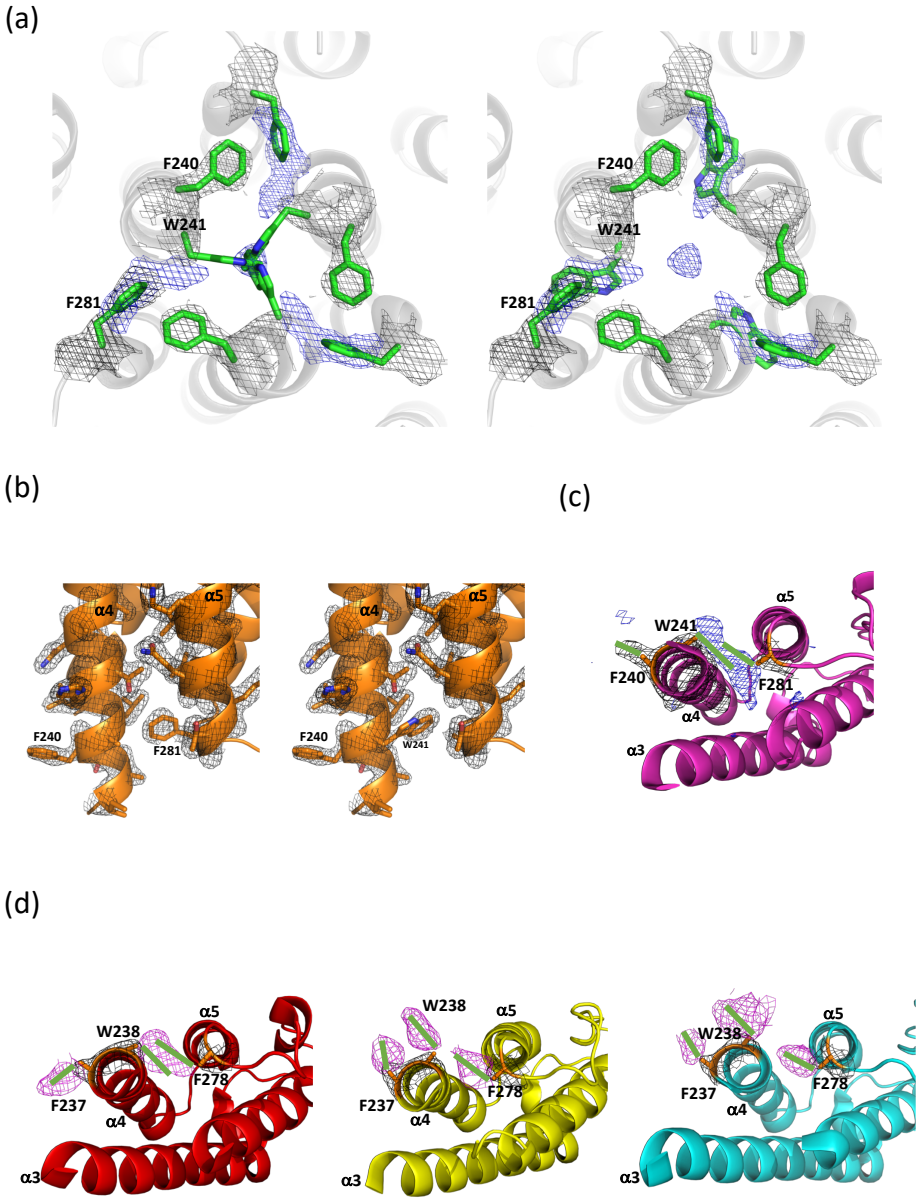
514

515

516



Supplementary figure 1 – Sequence conservation analysis in TssK. (**a.**) Consurf analysis of 150 representative sequences of TssK. Mapping of conserved residues onto the Bc-TssK sequence using the Consurf conservation colour scheme (see included legend). Residues shown in black and white are included in the alignment, whereas those in blue had insufficient data to be incorporated. The sequences of conserved clusters 1, 2 and 3 are highlighted in green, dark blue and mustard, respectively. (**b.**) Structural alignment of Bc-TssK$_{SN}$ (yellow) and Bp-TssK$_{SN}$ (blue). The structures share a very similar architecture. The black dotted line represents the absent loop between α3 and α4, while the red dotted line indicates the linker to the head domain. (**c.**) B-factor analysis of Bc-TssK. (Left) B-factors for Cα's of Bc-TssK residues indicate a greater degree of disorder within the head domain compared to the rest of the molecule. (Right) The head domain generates crystal contacts with a neighbouring symmetry related shoulder domain (blue) within the crystal lattice stabilising residues at the contact interface. Consequently, this face has lower B-factors then the rest of the head domain. Colour and thickness of the chain trace indicate the magnitude of the B-factor for each residue (see key).



Supplementary figure 2 - Conformational assessment of hydrophobic residues at the trimer interface for the TssK neck domain. (**a.**) Alternative conformations of W241 in the Bc-TssK neck domain. (Left), W241 modelled towards density present at the crystallographic three-fold symmetry axis with F281 modelled in the ambiguous density. (Right), W241 modelled away from the crystallographic three-fold symmetry axis where it is present in the density previously occupied by F281. Maps contoured as 2Fo-Fc (black) = 1.3, Fo-Fc (blue) = 3.0 RMSD (**b.**) 2Fo-Fc maps surrounding W241 and F281 of Bc-TssK. The maps indicate that either W241 or F281 can accommodate the side-chain density. Map contoured to 1.0 RMSD. (**c.**) Fo-Fc and 2Fo-Fc maps for Bp-$TssK_{SN}$ surrounding F240, W241 and F281. Density indicates a similar situation to Bc-TssK, such that good density for the main chain can be observed. However, as indicated in the 2Fo-Fc map, density for the side chains W240 and W281 remains ambiguous. Maps contoured 2Fo-Fc (black) = 1.0, Fo-Fc (blue) = 3.0 RMSD. Green lines indicate potential orientations of the missing side chains. (**d.**) Fo-Fc and 2Fo-Fc maps for Ec-TssK around F237, W238 and F278 for chains A (red), B (yellow) and C (blue). 2Fo-Fc maps indicate good density for the main chain, whereas the Fo-Fc maps suggest that these hydrophobic residues exist in alternative conformations. Maps contoured 2Fo-Fc (black) = 1.0, Fo-Fc (magenta) = 3.0 RMSD. Green lines indicate potential orientations of the missing side chains.

Table 1. Data collection and processing statistics

Data collection –	**Bc-TssK (1-448) - Native**	**Bc-TssK (1-448)-SeMet Peak**	**Bc-TssK$_{SN}$ (1-317) - Native**	**Bp-TssK$_{SN}$ (1-328) - Native**
Wavelength (Å)	0.97951	0.92819	0.92819	0.9795
Resolution range (Å)	49.96-2.24 (2.30-2.24)	107.2- 2.91 (2.99- 2.91)	53.60-1.61 (1.65-1.61)	62.9-2.21 (2.28-2.21)
Space group	$P6_3$	$P6_3$	H32	H32
Unit cell (a,b,c) (Å)	99.9, 99.9, 212.2	100.2, 100.2, 214.4	107.2 107.2 191.1	106.5, 106.5, 188.8
(α,β,γ) (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90, 90, 120	90, 90, 120
Total reflections	394046 (29158)	465669 (13422)	1200301 (89301)	221106 (14226)
Unique reflections	57392 (4250)	26759 (1983)	54870 (4034)	20952 (1533)
Multiplicity	6.9 (6.9)	17.4 (6.8)	21.9 (22.1)	10.6 (9.3)
Completeness (%)	99.9 (99.7)	100.0 (99.9)	100.0 (100.0)	100.0 (100.0)
Mean I/σ	12.0 (1.3)	15.4 (1.2)	14.3 (0.6)	8.1 (1.1)
Wilson B factor (Å^2)	45	64	27	45
R_{merge}	0.074 (1.296)	0.178 (1.574)	0.061 (2.969)	0.129 (1.525)
R_{pim} (I)	0.046 (0.811)	0.062 (0.992)	0.019 (0.923)	0.061 (0.781)
CC 1/2	0.999 (0.587)	0.999 (0.536)	1.00 (0.765)	0.996 (0.886)
Anomalous completeness (%)		99.7 (98.6)		
Anomalous multiplicity		8.5 (3.3)		
Anomalous correlation		0.174 (0.063)		
Anomalous slope		1.142		
Figures in parentheses refer to data in the outer shell.				
Refinement -				
Rfactor	0.234		0.197	0.215
Rfree	0.274		0.218	0.247
No. of non-H atoms				
Protein	5171		2324	2310
Ligands	12		16	8
Water	99		173	54
Protein residues	674		287	291
RMSD (bonds) (Å)	0.01		0.0133	0.009
RMSD (angles) (°)	1.44		1.47	1.40
Ramachandran favoured (%)	94.4		97.2	92.8
Ramachandran outliers (%)	0.7		0.0	0.0
Favoured rotamers (%)	97.4		97.6	95.4
Poor rotamers (%)	0.2		1.2	0.4
Molprobity score	1.43 (99th percentile, 2.24Å ± 0.25Å)		0.86 (100th percentile, 1.61Å ± 0.25Å)	1.55 (98th percentile, 2.21Å ± 0.25Å)
Average B factors				
Main chain (Å^2)	61.51		34.09	68.6
Side chains (Å^2)	58.24		38.51	72.8
Ligands (Å^2)	44.49		35.64	65.3
Water (Å^2)	45.84		39.71	60.4
PDB ID	**XXX**		**XXX**	**XXX**

Table 2. Bacterial strains

Strain	Genotype/Description[a]	Source or reference
B. cenocepacia		
H111	CF isolate	(44)
B. pseudomallei		
D286	Human melioidosis isolate	(45)
E. coli		
JM83	F⁻ *ara* Δ(*lac-proAB*) *rpsL* ϕ80d*lacZΔM15* (Sm^R)	(46)
DH5α	F- *recA1 endA1 gyrA96 thi-1 hsdR17* ($r_K^- m_K^+$) *supE44 relA1 Δ(lacZYA-argF)U169 deoR phoA* λ- ϕ80d*lacZΔM15*	(47)
BTH101	F-, *cya-99 araD139 galE15 galK16 rpsL1* (Sm^R) *hsdR2 mcrA1 mcrB1*	(31)
BL21(DE3)	F- *ompT hsdS_B* ($r_B^- m_B^+$) *dcm gal* λDE3	(48)

[a]Sm^R, streptomycin resistant.

Table 3. Plasmids

Plasmid	Description[a]	Source or reference
TssK expression		
pACYCDuet-1	Dual T7-lac promoter expression vector, *lacI*, p15A origin (Cm^R)	Novagen
pET21a(+)	T7-lac promoter expression vector, *lacI*, ColE1 origin (Ap^R)	Novagen
pACYCDuet-tssK	pACYCDuet-1 expressing native Bc-TssK	Dix et al., submitted
pACYCDuet-His_6.tssK	pACYCDuet-1 expressing N-terminal His-tagged Bc-TssK	Dix et al., submitted
pET21-$tssK_{SN}$.His_6	*B. cenocepacia tssK* codons 1-317 amplified with Bc-$TssK_{SN}$.for and Bc-$TssK_{SN}$.rev, and inserted between the *Nde*I and *Xho*1 sites of pET21a(+) with a C-terminal His-tagged	This study
pET21-Bp$tssK_{SN}$.His_6	*B. pseudomallei tssK* (BPSL3110) codons 1-448 amplified with Bp-TssK.for and Bp-TssK.rev, and inserted between the *Nde*I and *Xho*1 sites of pET21a(+). The final construct consists of codons 1-328 with a C-terminal His-tag.	This study
Two-hybrid system[b]		
pKT25	Vector for generating fusions to C-terminus of CyaA T25 fragment, ori_{p15A} (Km^R)	(49)
pKT25-zip	pKT25 containing the coding sequence of the yeast GCN4 leucine zipper fused to the T25 coding sequence	(31)
pKNT25	Vector for generating fusions to N-terminus of CyaA T25 fragment, ori_{p15A} (Km^R)	(50)
pUT18	Vector for generating fusions to N-terminus of CyaA T18 fragment, ori_{pMB1} (Ap^R)	(49)

pUT18C	Vector for generating fusions to C-terminus of CyaA T18 fragment, ori_{pMB1} (Ap^R)	(49)
pUT18C-zip	pUT18C containing the coding sequence of the yeast GCN4 leucine zipper fused to the T18 coding sequence	(49)
pKNT25-$tssK_S$	*B. cenocepacia tssK* codons 1-193 amplified with Nterm-tssK.for and Nterm-tssK.NTD.rev, and inserted between the *Xba*I and *Acc*65I sites of pKNT25	This study
pUT18-$tssK_S$	*B. cenocepacia tssK* codons 1-193 amplified with Nterm-tssK.for and Nterm-tssK.NTD.rev, and inserted between the *Xba*I and *Acc*65I sites of pUT18	This study
pKNT25-$tssK_{SN}$	*B. cenocepacia tssK* codons 1-317 amplified with Nterm-tssK.for and Nterm-tssK.NTD-MD.rev, and inserted between the *Xba*I and *Acc*65I sites of pKNT25	This study
pUT18-$tssK_{SN}$	*B. cenocepacia tssK* codons 1-317 amplified with Nterm-tssK.for and Nterm-tssK.NTD-MD.rev, and inserted between the *Xba*I and *Acc*65I sites of pUT18	This study
pKT25-$tssK_H$	*B. cenocepacia tssK* codons 318-449 amplified with Nterm-tssK.CTD.for and Cterm-tssK.CTD.rev, and inserted between the *Xba*I and *Acc*65I sites of pKT25	This study
pKNT25-$tssK_H$	*B. cenocepacia tssK* codons 318-448 amplified with Nterm-tssK.CTD.for and Nterm-tssK.CTD.rev, and inserted between the *Xba*I and *Acc*65I sites of pKNT25	This study
pUT18-$tssK_H$	*B. cenocepacia tssK* codons 318-448 amplified with Nterm-tssK.CTD.for and Nterm-tssK.CTD.rev, and inserted between the *Xba*I and *Acc*65I sites of pUT18	This study

pUT18C-tssK$_H$	*B. cenocepacia tssK* codons 318-449 amplified with Nterm-tssK.CTD.for and Cterm-tssK.CTD.rev, and inserted between the *Xba*I and *Acc*65I sites of pUT18C	This study

[a]Ap^R, ampicillin resistance; Cm^R, chloramphenicol resistance; Km^R, kanamycin resistance.

[b]Codon 449 of *tssK* refers to the native translation termination codon.

Table 4. Oligonucleotides

Primer name	Sequence (5'-3')[a,b,c]
TssK subunit expression	
Bc-TssK$_{SN}$.for	ATTAATCATATGAGTTATTCGGCCAAGGTGCTG
Bc-TssK$_{SN}$.rev	ATTAATCTCGAGGATCACGGTGTCGAGCAGTTCGCGC
Bp-TssK.for	ATTAATCATATGAGTTATTCGGCCAAGGTGCTC
Bp-TssK.rev	ATTAATGAGCTCTCATGACGTGACGGCGAT
Two-hybrid analysis	
Nterm-tssK.for	GCGCTCTAGAGATGAGTTATTCGGCCAAGGT
Nterm-tssK.NTD.rev	GCGCGGTACCCGCGCGAGGCACGGCGGCACGA
Nterm-tssK.NTD-MD.rev	GCGCGGTACCCGGATCACGGTGTCGAGCAGTT
Nterm-tssK.CTD.for	GCGCTCTAGAGATGTCGACGCGCTATTTCGCGAT
Nterm-tssK.CTD.rev	GCGCGGTACCCGTGATGTGACCGCGATCAGTT
Cterm-tssK.CTD.rev	GCGCGGTACCTCATGATGTGACCGCGATCA

[a]Sequences specifying restriction endonuclease cleavage sites are underlined.

[b]Sequences specifying affinity tags are shown in italic font.

[c]Stop codons are shown in red font.

Chapter 4

Discussion and future work

4.1 Discussion

This thesis aimed to structurally explore the T6SS baseplate components TssA and TssK due to the distinct lack of information available for both their structure, location and their respective role within the T6SS. As has been discussed in papers two and four, the structure determination of TssA and TssK from various bacterial systems has been successful and has allowed for a thorough structural analysis for both components to help in determining their respective role in the T6SS assembly and function.

4.1.1 TssA

Through the use of bioinformatics and proteolytic studies presented in this thesis and elsewhere (Zoued et al., 2016), the TssA family has been sub-divided into four sub-clades (TssA1^{A}, TssA1^{B}, TssA2^{A} and TssA2^{B}) based on domain organisation and sequence identity. Members of the TssA1 clade consist of a single N-terminal domain (Nt1) and a C-terminal domain (CTD), whereas TssA2 members consist of two N-terminal domains (Nt1 and Nt2) associated to an apparently unrelated CTD. The structure determination of TssA from *E. coli* (a TssA2^{B} member) provided detailed information on the fold and validated the three-domain architecture of the TssA2 clade, specifically providing the first high-resolution structure of the Nt2 and CTD domains (Zoued et al., 2016). As detailed in papers one and two, the work carried out at the same time as part of this thesis, resulted in the structure determination of TssA from *A. hydrophila* (TssA2^{A}) and *B. cenocepacia* (TssA1^{B}) enabling a direct structural comparison of three out of the four TssA sub-clades (TssA1^{B}, TssA2^{A} and TssA2^{B}).

Comparison of the TssA2^{A} and TssA2^{B}TssA CTD indicates that the $\sim$ 40 residue C-terminal extension found in TssA2^{B} sub-clade members promotes the radically different assembly of the CTD monomers, such that the *E. coli* TssA2^{B} and *A. hydrophilia* TssA2^{A} CTD's generate a D_6 and D_5 symmetric oligomer, respectively. In contrast, the structure determination of the CTD from *B. cenocepacia* (TssA1^{B}) identified it displays D_{16} symmetry. Despite the distinctly different CTD symmetries generated by the three sub-clades, interaction assays shown in paper two and elsewhere suggest TssA mediates contacts with broadly similar T6SS components independent of the sub-clade (Zoued et al., 2016; Planamente et al., 2016). Taken together, this indicates that all TssAs possess a conserved function despite their structural diversity and apparent symmetry mismatch with other T6SS components. One question that remains unanswered that arises as a result of the symmetry diversity is how such diversity can lead to a similar biological function and how the different symmetry seen in the TssAs can be rationalised with further differences in symmetry of the other components of the T6SS.

Currently, there are two proposed models for the function of TssA. In the first, TssA is associated with, and maintained at, the baseplate during polymerisation of the contractile sheath. This model has been suggested based on secondary structure weighted sequence alignment between TssA1^{A} from *P. aeruginosa* and the CTD of gp6 (Planamente et al., 2016). However, the structure determination of the *B. cenocepacia* TssA1^{B} CTD, and the high sequence identity between both the TssA1^{A} and TssA1^{B} sub-clades indicates that TssA1^{A} does not look like gp6, therefore undermining this model. In contrast, model two suggests that TssA is initially located at the membrane complex, before being displaced by the T6SS baseplate and maintained at the growing end of the contractile tail (Zoued et al., 2016). This 'cap' model is supported by the recent cryo-EM reconstruction of the *V. cholerae* T6SS assembly, whereby density potentially associated with TssA occurs at the distal end of the contractile tail (Nazarov et al., 2018). One variant of the cap model suggests that eversion of the CTD subunits generates a lumen of suitable size for hcp to pass through and become incorporated into the growing inner tube (Zoued et al., 2017). However, the work on TssA in this thesis suggests that eversion is not a feature of the TssA CTD due to the structural differences between the various sub-clades, negating the possibility of a common mechanism. Furthermore, the lack of structural evidence for a TssA CTD sub-clade existing in its alternative conformation, indicates this model, in its current form, is inadequate.

A further observation is the conservation of the ImpA_N motif within the Nt1 domain throughout all TssA clades. Despite the differences in CTD symmetry, the Nt1 is not only the most conserved domain but is presented at a similar position at the periphery of the TssA complex in all TssA sub-clades. This presentation is achieved either through a large diameter CTD oligomer (TssA1) or with a CTD with a smaller diameter and the presence of the Nt2 domain (TssA2), this suggests that the Nt2 domain of TssA2 are used as spacers to position the TssA Nt1 domains correctly. The most straightforward mechanism would, therefore, be an interaction maintained through the Nt1 with other T6SS components, while the CTD is an oligomerisation module positioning and tethering the Nt1 domains. This hypothesis is consistent with bacterial two-hybrid, and SPR analysis of Ec-TssA which indicates that interactions with the sheath are mediated through the N-terminal domains while association to the hcp tube is carried out by the CTD (Zoued et al., 2016).

One further observation as part of this study is the inherent mobility exhibited by TssA. As seen in the Ah-TssA2^{A} Nt2-CTD structure, the Nt2 domains can sample a large range of space about the plane of the CTD ring, as a result of the Nt2-CTD interdomain linker. Given the identification of a further linker between Nt1 and Nt2, it can be assumed this domain (Nt1) is also highly mobile with respect to the CTD oligomer. The mobility of the constituent TssA domains is also seen in the EM reconstructions of full-length TssA from both EAEC and *V. cholerae* due to a lack of density present for the Nt1 domain (Zoued et al., 2016; Nazarov et al., 2018). This mobility exhibited by the TssA N-terminal domains could allow for the passage of the hcp subunits, without the need for CTD eversion.

The final feature of TssA, which has already been alluded to in this discussion, is the apparent symmetry mismatch displayed between the various TssA sub-clades and that of the membrane complex (5-fold) and contractile sheath (six-start helix). As has been discussed in paper two, the role of symmetry mismatch within TssA could be analogous to that of the flagella capping protein, HAP2, which uses symmetry mismatch to promote systematic subunit incorporation into the growing flagella. The parallel in TssA would be that the Nt1 domains are inserted into the polymerising contractile sheath and subsequently displaced upon movement of the CTD oligomer; however, further work is required to test this hypothesis.

4.2 TssK

As presented in papers three and four, the structure of full-length domain constructs of TssK has been determined from *B. cenocepacia* and of a domain construct of the *B. pseudomallei* homologue. The structure identified that TssK assembles as a trimer, whereby each subunit consists of three domains (shoulder, neck and head). This finding is consistent with the structure of TssK from *E. coli* (Ec-TssK), which was determined in parallel work elsewhere (Nguyen et al., 2017). Structural comparisons carried out between Ec-TssK with proteins in the protein database has provided evidence that the shoulder domain of TssK has a strong structural similarity with the equivalent domain of receptor binding proteins (RBP)s as part of the phage tail assembly complex from *lactococcal* siphophage, specifically p2 and 1358 (Spinelli et al., 2006a; Spinelli et al., 2006b; Nguyen et al., 2017). This similarity was then shown to extend to structural conservation across other siphophage baseplate RBPs including TP901-1. Given that RBPs are baseplate components, where they are used to bind phage to receptors on the cell surface of targets cells, this relationship between components with an analogous function is interesting.

The interaction of Bc-TssK, and its constituent domains, with other components of the T6SS, are presented in paper four. These identified a broadly similar pattern of interaction to that shown for Ec-TssK, including that the C-terminal (head) domain is responsible for TssK interacting with the cytoplasmic regions of TssL (Nguyen et al., 2017). These observations are supported by cryo-EM reconstructions of non-contractile sheaths from *V. cholerae* which displays trimers of TssK with the shoulder domains interacting with the baseplate, while the TssK head domains are orientated towards the membrane complex (Nazarov et al., 2018).

On inspection of the TssK shoulder domain, the N-terminal loop and N-terminus of α1 contain multiple highly conserved residues. While these structural motifs are conserved within RBP shoulder domains, they feature no sequence identity with TssK. This region within RBPs is responsible for the interaction with the additional baseplate components within phage. A well-studied example of this is the p2 baseplate (figure B.21a) whereby the baseplate protein ORF15 utilises an 'arm and hand' motif to insert into the N-terminal loops generated by the P2 RBP (ORF18) trimer (figure B.21b) (Spinelli et al., 2006a; Spinelli et al., 2006b; Sciara et al., 2010; Farenc et al., 2014). Therefore, it has been speculated that the TssFG complex could possess an equivalent 'arm and hand' motif for interaction with the

TssK shoulder domain (Nguyen et al., 2017).

Structural analysis within the neck domain identified that the side chains of conserved residues W240 and F241 located at the three-fold symmetry axis of the TssK trimer are disordered. Initial attempts to model their respective side chains resulted in significant packing clashes. It appears these residues are inherently dynamic, such that the Bc-TssK$_{SN}$ structure possibly represents an average of the different positions the side chains can occupy. This observation is supported in the Ec-TssK structure, whereby the equivalent residues deviate from three-fold symmetry. In addition, the connecting loop between α3 and α4 displays varying degrees of disorder between the structures of Bc-TssK, Bp-TssK and Ec-TssK suggesting that this region of the neck domain is involved in some form of conformational change. Taken together, the sequence conservation and inherent disorder within the distal end of the neck domain suggest that order/disorder transitions within this region may be important to the role of TssK.

Furthermore, the structural comparison between Bc-TssK and Ec-TssK identified a distinctly different position and orientation between their respective head domains. This suggests mobility is an inherent property of the head domain relative to the rest of the molecule. Inspection of the secondary structure within the head domain identified an alternative conformation for α8 between the Bc-TssK and Ec-TssK orthologues, despite consisting of a number of conserved residues. However, the inability to improve the diffraction quality of crystals grown from Bc-TssK$_{H}$.his$_{6}$ protein did not allow for a high-resolution model of this domain to be built. Therefore, a thorough comparison of the fold within the head domain could not be carried out.

During the construction of this thesis the structure of TssK from Serratia sp. (PDB ID: 5Y0E, not yet published) was released. The protein possesses a similar fold to that of Bc_TssK and Ec_TssK discussed in this thesis which is unsurprising considering it contains 30 % sequence identity to Bc_TssK (figure 4.1). Furthermore, the *Serratia* TssK structure indicates distinctly different positions of the head domain for each respective chain in the TssK trimer. Further indicating the inherent mobility of this domain with respect to the rest of the molecule.

Taken together, these findings support the hypothesis that TssK acts as a connector helping to link the baseplate and the membrane complex within the T6SS assembly. Furthermore, the inherent mobility of the head domain, and the disorder associated with the distal

end of the neck domain, seen between the various structures of TssK, promotes the idea that TssK could undergo a conformational change. Given this, it seems likely that this rearrangement is related to the contraction, and thus 'firing' of the T6SS, in a mechanism not too dissimilar to phage contraction upon binding to host (English et al., 2014; Hu et al., 2015; Nazarov et al., 2018).

The data presented for both TssA and TssK in this thesis and that introduced elsewhere during the project vastly expand the previous structural and functional knowledge for these components. Consequently, a more detailed picture of the T6SS apparatus can be constructed. The TssA component, due to its location at the distal end of the growing contractile sheath, is now proposed to aid in the polymerisation and stabilisation of the contractile tail during sheath assembly in the high-energy, extended form. Furthermore, it may be that the symmetry mismatch between the TssA CTD and the contractile sheath facilitate transient interactions between subunits allowing for subsequent sheath subunit incorporation through the displacement of Nt1 domains (figure B.21c). These studies have shown that TssK is structurally related to *Lactoccocal* siphophage RBPs, suggesting that the shoulder domain can interact with the baseplate, while the head domain interacts with the membrane complex, promoting the hypothesis that TssK acts as an adapter protein linking the baseplate to the membrane complex (figure B.21c). Therefore, while it was initially proposed that the T6SS represents an inverted phage tail, it appears that the T6SS machinery shares evolutionary relationships between multiple phage systems. In addition, the mobility of the TssK head domain, and the apparent dynamic arrangement of residues within the neck domain may promote a conformational change within TssK resulting in T6SS contraction.

After the submission of this thesis, the first high-resolution (4.6 Å) cryo-EM structure of the *E. coli* T6SS baseplate was released (Cherrak et al., 2018). This structure identified that the baseplate assembly is comprised of six units each consisting of TssFGKE in a 2:1:6:1 stoichiometry, generating a particle with six-fold symmetry. The model not only provided the first structures for TssF and TssG but also provided the first evidence of the interaction between the TssK trimer and the baseplate. This interaction showed that TssG contains two extended loops which interact with the conserved residues found within the N-terminus of the TssK shoulder domains, mirroring that of the 'hand and arm' motif with RBPs. In addition, the TssK trimers are orientated, such that the head domains are presented towards the membrane complex, therefore sterically permitting an interaction between them both.

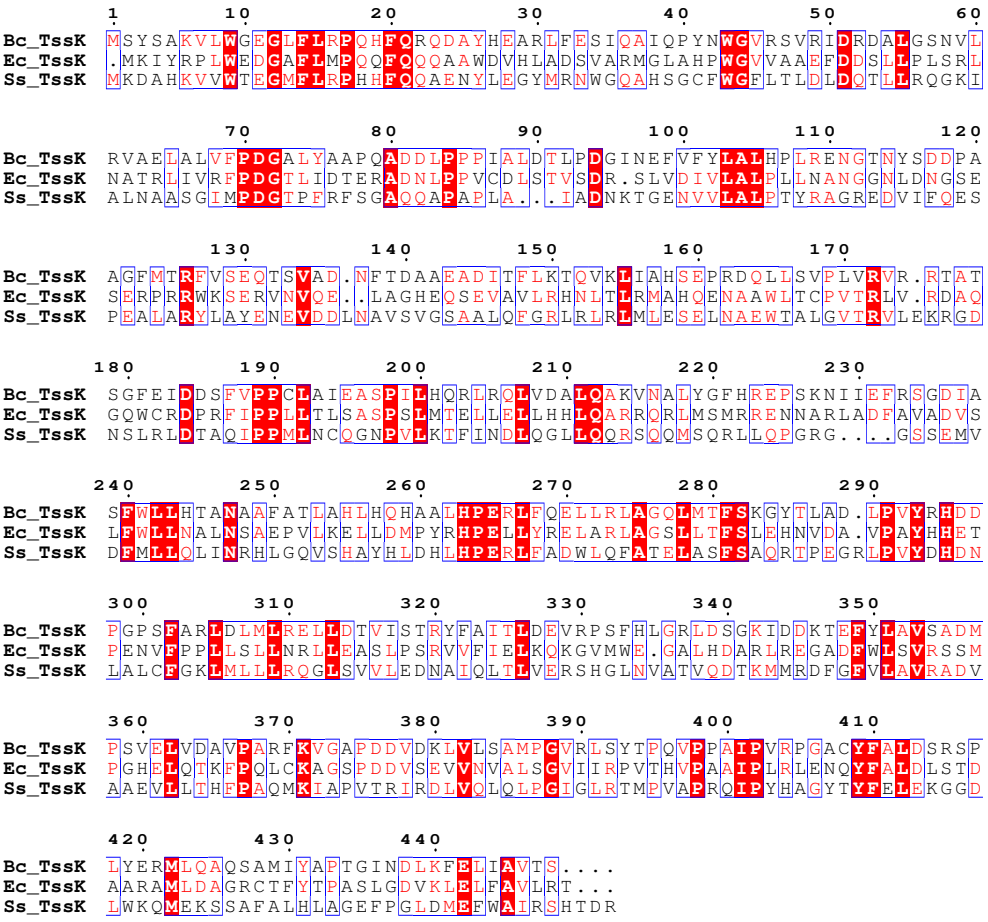


FIGURE 4.1: **Sequence alignment of *B. cenocepacia*, *E. coli* and *S. sp* TssK**. A multiple sequence alignment generated by ESPript (Robert and Gouet, 2014) which indicates the strong sequence similarity of *S. sp* TssK (Ss_TssK) with that of *B. cenocepacia* (Bc_TssK) and *E. coli* (Ec_TssK).

Both these observations have been proposed in this thesis.

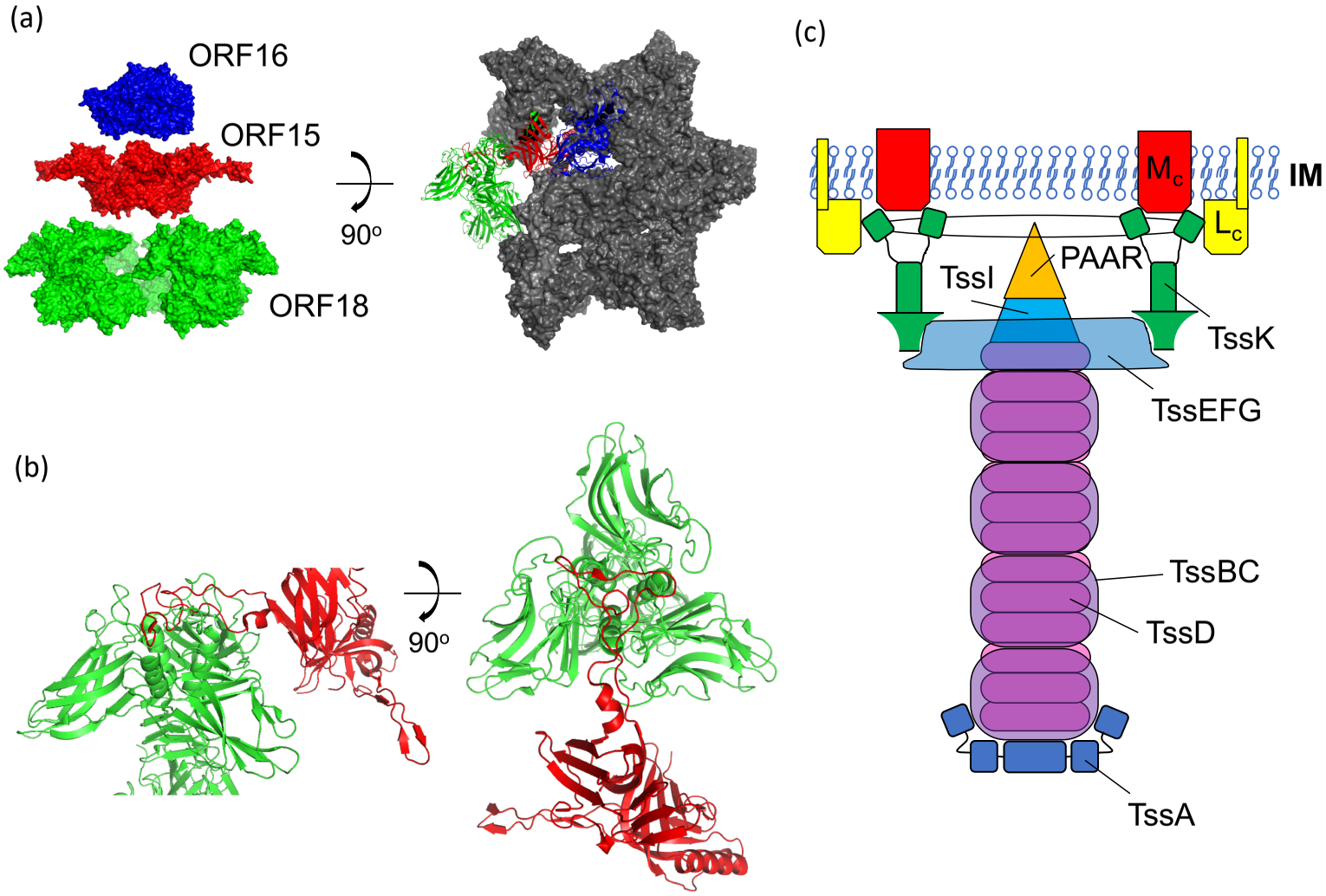


FIGURE 4.2: **TssA and TssK - updated roles.** **a)** Cartoon representation of the P2 phage baseplate components. blue - baseplate component ORF16, red - Dit-like baseplate component ORF15, green - Trimers of p2 RBPs ORF18. **b)** View of the interaction between the shoulder domain of the p2 RBP (green) with the 'arm and hand' motif of ORF15. panel a and b generated using PBD ID 2WZP. **c)** Updated model of the T6SS assembly. TssK is shown in green interacting with the membrane complex and baseplate. TssA is shown in dark blue maintained at the distal end of the growing contractile tail. Panel adapted from Cianfanelli, Monlezun, and Coulthurst, 2016.

4.3 Future work

This thesis presents a number of structural discoveries with regards to the protein components TssA and TssK from the T6SS. However, it does raise immediate questions that future work may be able to resolve.

4.3.1 Structure determination of the Ah-TssA Nt1 domain

A theme discussed in paper two was the potentially conserved role of the ImpA_N region within the TssA Nt1 domain. Although the structure of Nt1 was determined for Bc-TssA1^{B}, a member of clade 1, the equivalent domain has not yet been structurally determined for a member of clade 2. While the ImpA_N region of Nt1 is highly conserved, the remainder of the Nt1 domain shows a greater sequence diversity between clade 1 and 2. Therefore, it is possible that the structure of Nt1 from clade 1 does not best represent the structure of Nt1 from clade 2. Consequently, the use of species variants within clade 2 species may provide a suitable domain which crystallises, and the similarities/difference can be determined.

4.3.2 High-resolution structure determination of a member of the remaining TssA sub-clade (TssA1^{A})

X-ray studies on TssA have now provided structures for the CTD domain of three of the four proposed TssA sub-clades (TssA1^{B}, TssA2^{A} and TssA2^{B}); however, a high-resolution model of the CTD domain from a TssA1^{A} representative remains unknown. Therefore the structure determination of a TssA1^{A} CTD could identify the relationship between sub-clades TssA1^{A} and TssA1^{B}.

4.3.3 Resolve the consequences of the difference in symmetry displayed by TssA within alternative bacterial systems

As has been shown in paper two, TssA from different bacterial species can display differences in symmetry yet maintain similar interactions with other T6SS protein components. Therefore, it warrants further studies into the apparent symmetry mismatch exhibited by T6SS components. This would require the isolation and structure determination of complexes involving different T6SS subunits with TssA.

4.3.4 Structure determination of the TssK head domain

Although a model of the Bc-TssK head domain was built as part of the full-length Bc-TssK structure, due to the inherent mobility of the domain and the resultant poor electron density, the domain has only been modelled as a poly-alanine chain. Therefore, a detailed comparison with the Ec-TssK head domain was not possible. This warrants the production of a new Bc-TssK head domain construct without the C-terminal His-tag, in an attempt to improve the crystallisation ability of this domain and obtain a high-resolution structure.

4.3.5 Investigate the binding of TssK shoulder and neck domains with the baseplate proteins

The interaction of the TssK shoulder domains with the T6SS baseplate components, indicates the potential of carrying out co-crystallisation studies in an attempt to determine the structure of the TssK shoulder and neck domains in complex with other baseplate components, particularly TssFG.

4.3.6 Investigate the binding of TssK head domain with TssL/M

As has been suggested through interaction studies of TssK with other T6SS, the head domain can interact with the membrane complex components TssL and TssM. Therefore, it is worth carrying out co-crystallisation trials between the TssK head domain and TssLM cytoplasmic domains, in an attempt to obtain a complex structure between the components.

4.3.7 The generation of TssK domain knock out mutants

If TssK acts as an adapter that aids the association of the baseplate with the membrane complex as proposed, this would suggest that knocking out domains of TssK will inhibit T6SS assembly and therefore function. This hypothesis could be tested either by the removal of either the shoulder or head domain or, potentially, through site-directed mutagenesis of the conserved residues in TssK.

4.3.8 Structure determination of sub-complexes

Whilst it has previously been possible to generate TssFGK complexes in isolation (English et al., 2014), structural studies have yet to be published. One factor that could be causing

issues is the inherent mobility of the TssK head domain. Therefore, generating $TssFGK_{SN}$ complexes may produce more suitable assemblies for structural studies.

Appendix A

Extended methodology - TssA

A.1 TssA phylogenetic analysis

As can be seen in 3.2, Dr Mark Thomas carried out a phylogenetic analysis of TssA representatives from various Gram-negative bacteria. This was presented as a maximum-likelihood phylogenetic tree as seen in 3.2, figure 1, calculated through the use of MEGA (Tamura et al., 2013). This tree identified two major clades within the TssA family, which were classified as TssA1 and TssA2. A clade in this instance represents a group of proteins thought to contain the evolutionary descendants that arise from a common ancestor. These clades could both be further sub-divided into two sub-clades (A/B). These divisions have been proposed as an alternative to the three clades identified in Planamente et al., 2016. For this project it was of interest to structurally compare representatives from the four sub-clades.

A.2 Construct design

A number of DNA constructs of *tssA* from *A. hydrophila* (Ah) ATCC 7966 were utilised throughout this project; this section provides some extended detail on their acquisition or how they were produced as part of this body of work.

A.2.1 Ah-TssA constructs provided by Mark Thomas

A selection of Ah-TssA constructs were provided by Dr Mark Thomas, Department of Infection and Immunity, Sheffield Medical School. These consisted of either full-length or domain constructs generated from *tssA* (AHA 1844) from *A. hydrophila* ATCC 7966, and the different domain constructs are shown in figure A.1. Full details of the cloning can be found in 3.1, table 1.

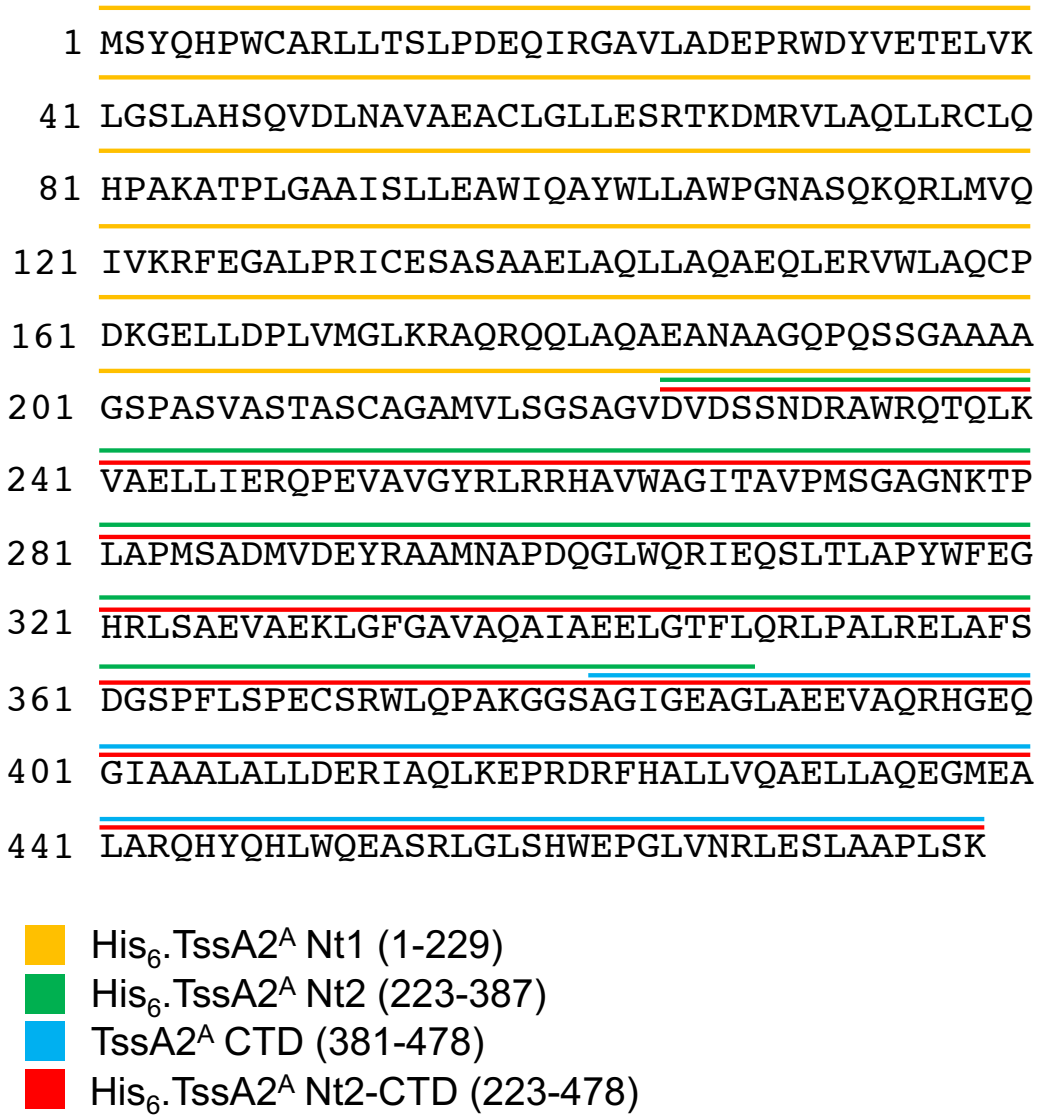


FIGURE A.1: **Sequence details of the Ah-TssA domain constructs that were provided by Dr Mark Thomas.** The constructs that were provided were used for either cloning or protein expression for subsequent crystallisation trials. The protein sequence for each construct is indicated in different colours, according to the included key. The full-length Ah-TssA construct has been omitted as it spans the full gene.

A.2.2 Generation of the Ah-TssA Nt1-Nt2.his$_6$ construct

The generation of the Ah-TssA Nt1-Nt2.his$_6$ construct was carried out as an alternative attempt to crystallise the Nt1 domain from *A. hydrophila* TssA as preliminary crystallisation of this domain, in isolation, had proven unsuccessful (see 3.1 for more details). Primers were generated using the NEB primer design tool suitable for use with the NEB Q5 polymerase kit. The primers used are shown in 3.2, table 9. The PCR reaction mixture was made up as in table 2.6 using 2 μL at 100 ng/μL pACYduet-1/his$_6$.Ah-TssA (1-478) plasmid as the DNA template. The PCR reaction was carried out in the thermocycler as in table 2.7, the reaction was run in duplicate to improve the chances of success. The amplified DNA products of the reaction were subsequently analysed on a 1 % agarose gel, and identified good amplification of the $\sim$ 1100 bp gene amplicon (figure A.2a).

The Ah-TssA Nt1-Nt2 PCR products were then purified as in 2.1.7 resulting in 2x 35 μL at $\sim$ 45 ng/μL for each product. These PCR products, in addition to 30 μL of 100 ng/μL purified pET21a vector, were then digested separately, using restriction enzymes *Xho*1 and *Nde*1 according to 2.2.3. This allowed for the incorporation of a C-terminal 6xhis affinity tag present in the pET21a vector (2.1) to the Ah-TssA Nt1-Nt2 protein product. Following restriction digest, 35 μL at $\sim$ 20 ng/μL Ah-TssA2 Nt1-Nt2 PCR product and 35 μL at $\sim$ 50 ng/μL pET21a were recovered, and a ligation reaction was set up to generate the final construct as in 2.2.4.

The successful incorporation of the Ah-TssA Nt1-Nt2 gene fragment into the pET21a vector was initially determined through colony PCR, set up as described in 2.1.10. After the PCR reaction had finished the PCR products were run on a 1 % agarose gel and can be seen in figure A.2b. The successful colonies 1 and 4 were then propagated, and their plasmids liberated through miniprep, these were then sent for sequencing which identified successful incorporation of the Ah-TssA Nt1-Nt2 gene fragment into the pET21a expression vector with no errors in the sequence. The final construct is shown in figure A.3.

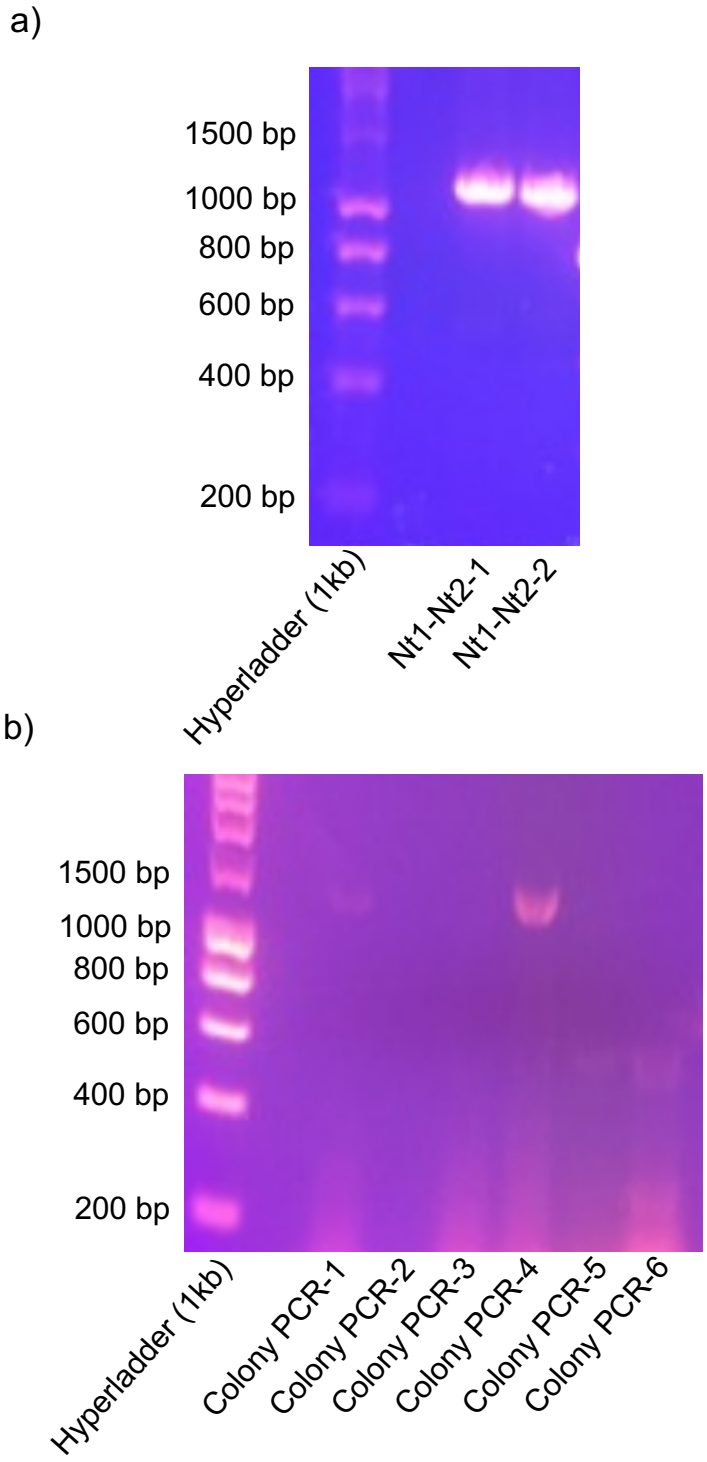


FIGURE A.2: **Agarose gels during the production of the Ah-TssA Nt1-Nt2.his$_6$ construct.** Each sample lane consists of 4 μL DNA product, mixed with 1 μL loading buffer. **a)** Once the PCR reaction had finished the PCR product was tested to determine if Ah-TssA Nt1-Nt2 had been successfully amplified. The marked lanes indicate successful amplification at $\sim$ 1100 bp, the expected size for the Ah-TssA Nt1-Nt2 gene fragment. **b)** Amplified PCR products from colonies expected to contain the desired Ah-TssA Nt1-Nt2 gene product in the pET21a vector. Reactions 1 and 4 can be seen to have worked, due to the presence of a band at $\sim$ 1300 bp, (use of T7 primers add $\sim$ 150 bp to the final PCR product).

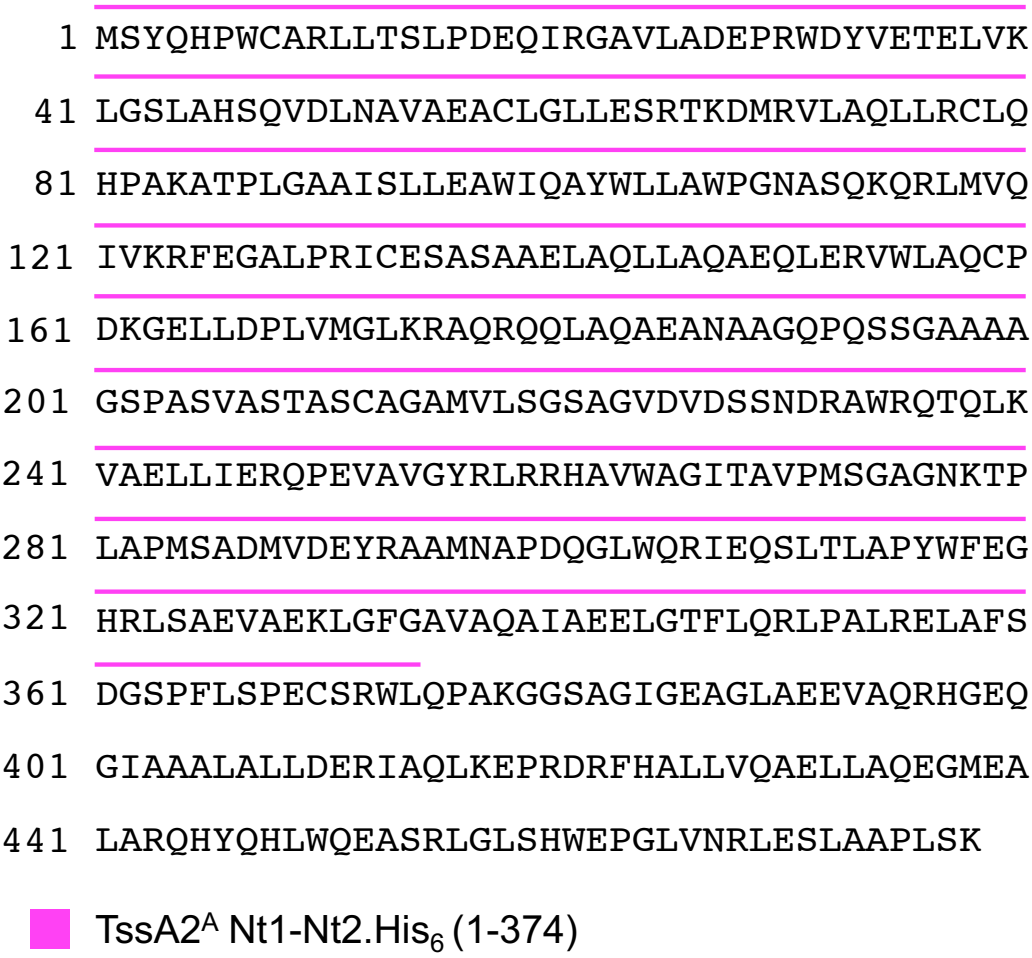


FIGURE A.3: **Sequence of the Ah-TssA Nt1-Nt2.his$_6$ construct.** The protein sequence shown below the pink lines indicate the residues encoded by the Ah-TssA Nt1-Nt2.his$_6$ construct.

A.3 Protein production of Ah-TssA constructs

While the Ah-TssA2^{A} constructs provided by Mark Thomas already had an established expression protocol (see paper 3.1 for details), the expression of the Ah-TssA Nt1-Nt2.His$_6$ construct was unknown and therefore initial expression trials were required.

A.3.1 Test expression of Ah-TssA Nt1-Nt2.His$_6$ (1-374)

To test if the Ah-TssA Nt1-Nt2.His$_6$ construct would produce sufficient protein with suitable solubility when expressed in *E. coli* small-scale test expressions were set up as described in 2.3.1 using BHI media supplemented with ampicillin (100 μg/μL). The result of the test expressions were successful in producing soluble protein at the expected molecular weight for Ah-TssA Nt1-Nt2.His$_6$ (41.6 kDa) and the associated SDS-PAGE gel can be seen in figure A.4.

A.3.2 Large-scale protein production of Ah-TssA constructs for crystallisation

To produce sufficient protein for downstream structural studies large-scale protein expression of the Ah-TssA Nt1-Nt2.His$_6$, Ah-TssA His$_6$.Nt2, Ah-TssA CTD and Ah-TssA His$_6$.Nt2-CTD was carried out. To achieve this 2 L of BHI media supplemented with 25 μg/μL chloramphenicol (Ah-TssA His$_6$.Nt2, Ah-TssA CTD and Ah-TssA His$_6$.Nt2-CTD) or 100 μg/μL ampicillin (Ah-TssA Nt1-Nt2.His$_6$) was used as the expression media. Large-scale over expression was carried out as in 2.3.2 with 37 °C (30 °C for Ah-TssA His$_6$.Nt2) used as the incubation temperature after induction with 1 mM (0.1 mM for Ah-TssA His$_6$.Nt2) IPTG. Due to the heavy over expression of the constructs, incubation of the cultures after induction was left for 2-3 hours, the resultant over expression of the Ah-TssA His$_6$.Nt2, Ah-TssA CTD and Ah-TssA His$_6$.Nt2-CTD domain constructs can be seen in figure A.5, large scale over expression of Ah-TssA Nt1-Nt2.His$_6$ did not differ in expression profile to that of A.4.

A.4 Protein purification of the Ah-TssA constructs

This section will detail the purification process for the Ah-TssA proteins used in crystallisation trials as part of this project.

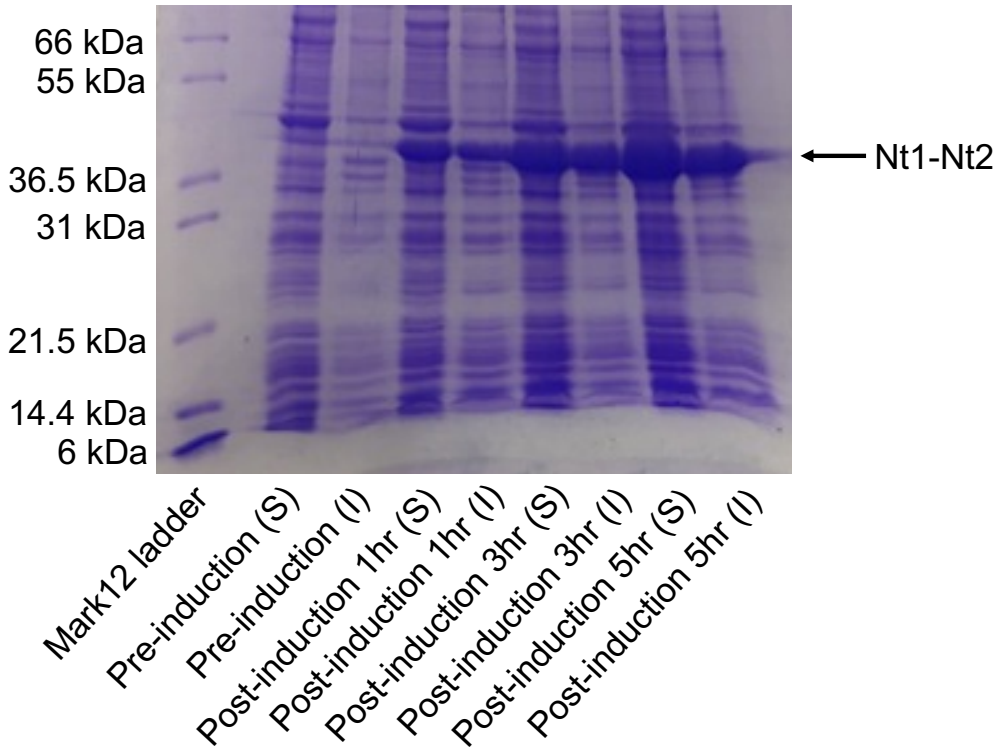


FIGURE A.4: **Test expression of the Ah-TssA Nt1-Nt2.His$_6$ construct.** 6/12 % SDS-PAGE gel of the pre-induction and post-induction samples. Successful over production of the protein product can be identified through the bands indicated by the arrow. This corresponds with the expected molecular weight of the construct (41.6 kDa).

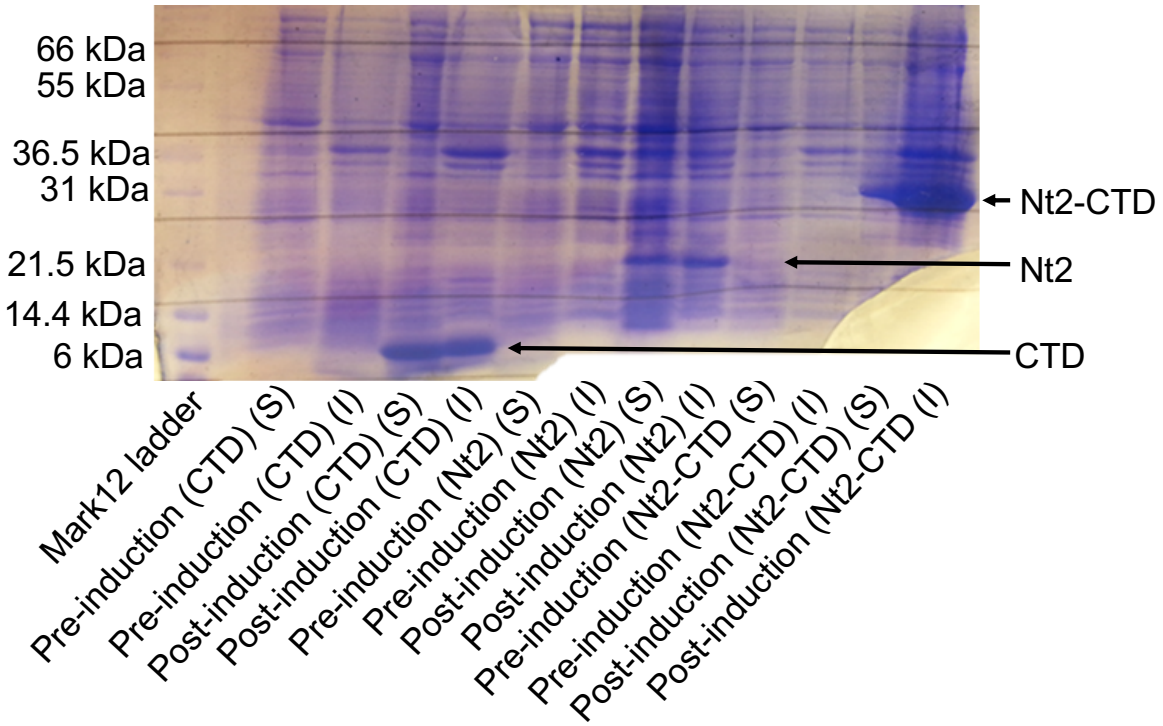


FIGURE A.5: **Over expression of Ah-TssA His$_6$.Nt2, Ah-TssA CTD and Ah-TssA His$_6$.Nt2-CTD domain constructs.** 6/12 % SDS-PAGE gel for the over expression for the Ah-TssA His$_6$.Nt2, Ah-TssA CTD and Ah-TssA His$_6$.Nt2-CTD protein. The arrows indicate the proteins run close to the expected molecular weights of $\sim$ 19.7, 10.7 and 29.8 kDa respectively.

A.4.1 Protein purification of Ah-TssA his_6.Nt2

From 2 l of BHI media, 4 g of *E. coli* cell paste was recovered. As described in 2.3.4, the cell free extract was liberated and clarified yielding 27 mL of protein solution at an approximate concentration of 9.5 mg/mL. This was then purified via nickel affinity chromatography as described in 2.4.1 and the chromatogram from the purification can be seen A.6a. After purification, fractions 11-13 were pooled together to generate the final sample, SDS-PAGE analysis of the purification can be seen in figure A.6b. The final sample was concentrated down to 100 μL with a protein concentration of 10 mg/mL in 10 mM Tris (pH 8.0) for use in crystallisation trials as in 2.5.1.

A.4.2 Protein purification of Ah-TssA CTD

From 2 l of BHI media, 4 g of *E. coli* cell paste was recovered. This was subsequently re-suspended and lysed as described in 2.3.4 before being centrifuged to separate the CFE from the cell debris. This produced 30 mL CFE at 11 mg/mL. The sample was initially purified through the addition of 4 M ammonium sulphate as described in 2.4.3. The resultant 32 mL of re-suspended solution was applied to a ResourceQ anion exchange column, and chromatography was carried out as in 2.4.2 (figure A.7a (left)). Fractions 15-17 which equate to the peak found at $\sim$ 75 mL were taken and concentrated to a volume of 2 mL before being applied to the gel filtration column as in 2.4.4. The resultant chromatogram can be seen in figure A.7a (right). Fractions 5 and 6 corresponding to $\sim$ 61 mL elution volume, were pooled and concentrated to produce the final sample. The final sample was 22.5 mg/mL in a volume of 300 μL 10 mM Tris (pH 8.0), before being diluted down to 18 mg/mL for subsequent crystallisation trials.

A.4.3 Protein purification of Ah-TssA his_6.Nt2-CTD

From 2 l of BHI media, 4 g of cell paste was recovered from the over expression of Ah-TssA his_6.Nt2-CTD. The cells were re-suspended and subsequently lysed as described in section 2.3.4, which, after centrifugation and separating from the cell debris, yielded 21.5 mL CFE at 4 mg/mL. The sample was applied to a Ni-NTA column as described in 2.4.1 and the corresponding chromatogram is shown in figure A.8a. Fractions 11-13 which equated to volumes around the peak found at 46 mL were taken and combined to form the final sample.

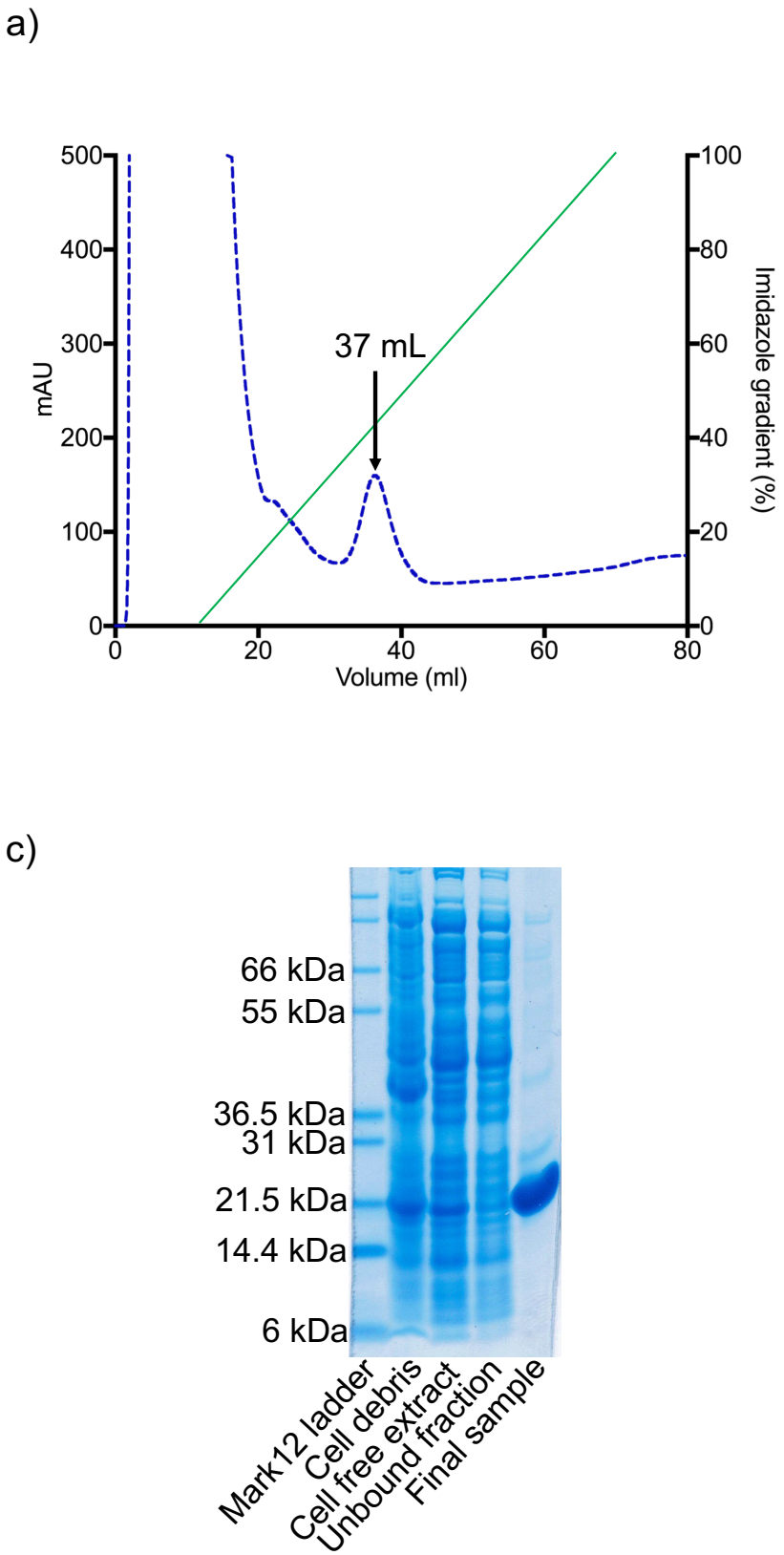


FIGURE A.6: **Purification of the Ah-TssA his$_6$.Nt2 construct. a)** Ni-NTA chromatography purification trace of Ah-TssA his$_6$.Nt2. The green line indicates the percentage of imidazole being applied to the column. The graph shows a peak at $\sim$ 37 mL indicated by the arrow, which is the elution peak corresponding to Ah-TssA his$_6$.Nt2. **b)** 4/12 % gradient SDS-PAGE gel of the purification of Ah-TssA his$_6$.Nt2. The SDS-PAGE gel indicates successful purification showing that the final sample of Ah-TssA his$_6$.Nt2 is at the correct molecular weight ($\sim$ 19.7 kDa) and is > 95 % pure.

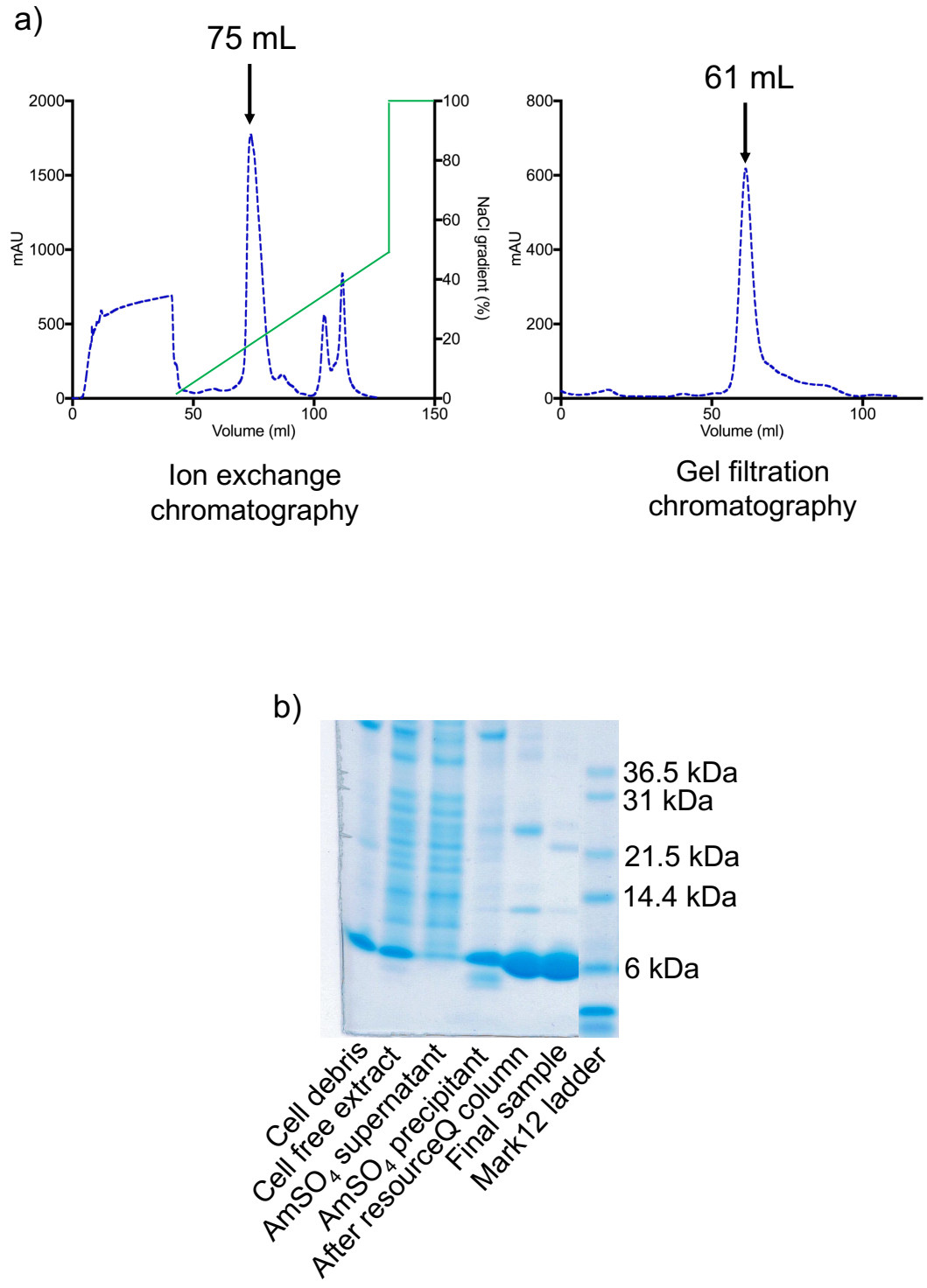


FIGURE A.7: **Protein purification of Ah-TssA CTD. a)** (Left) Ion exchange chromatography purification trace of Ah-TssA CTD, the trace indicates a peak at $\sim$ 75 mL indicated by the arrow. (Right) Analytical gel filtration chromatography purification trace, where the trace indicates a peak at $\sim$ 61 mL, this corresponds with an apparent molecular weight of $\sim$ 0.2 MDa. **b)** 4/12 % gradient SDS-PAGE gel of the purification of Ah-TssA CTD. The SDS-PAGE gel indicates successful purification of Ah-TssA CTD as seen by the final sample (> 95 % pure). However, the protein appears to migrate slightly smaller on SDS-PAGE gel than the expected molecular mass (10.7 kDa).

Analysis of the purification can be seen by the SDS-PAGE in figure A.8b. The final sample was concentrated to 40 mg/mL in 300 μL of 10 mM Tris (pH 8.0), before being diluted down to 11 mg/mL for subsequent crystallisation trials.

A.4.4 Protein purification of Ah-TssA Nt1-Nt2.his$_6$

From 2 l of BHI media, 4 g of cell paste was recovered from the over expression of Ah-TssA Nt1-Nt2.his$_6$. The cells were then re-suspended and lysed as in section 2.3.4 before being clarified and the cell-free extract separated from the cell debris yielding 30 mL at 9 mg/mL. Ni-NTA chromatography was subsequently carried out as in 2.4.1 and the associated chromatogram can be seen in figure A.9a. After purification fractions 12 - 17 were taken for analysis via SDS-PAGE and consequently fractions 14 and 15 were pooled together to form the final product. SDS-PAGE analysis of the purification can be seen in figure A.9b. The final product showed good purity of > 95 % and was subsequently concentrated and buffer exchanged as in section 2.5.1 yielding 1.2 mL at 32 mg/mL Ah-TssA Nt1-Nt2.his$_6$ in 50 mM Tris (pH 8.0). Before being diluted down to 10 and 20 mg/mL for subsequent crystallisation trials.

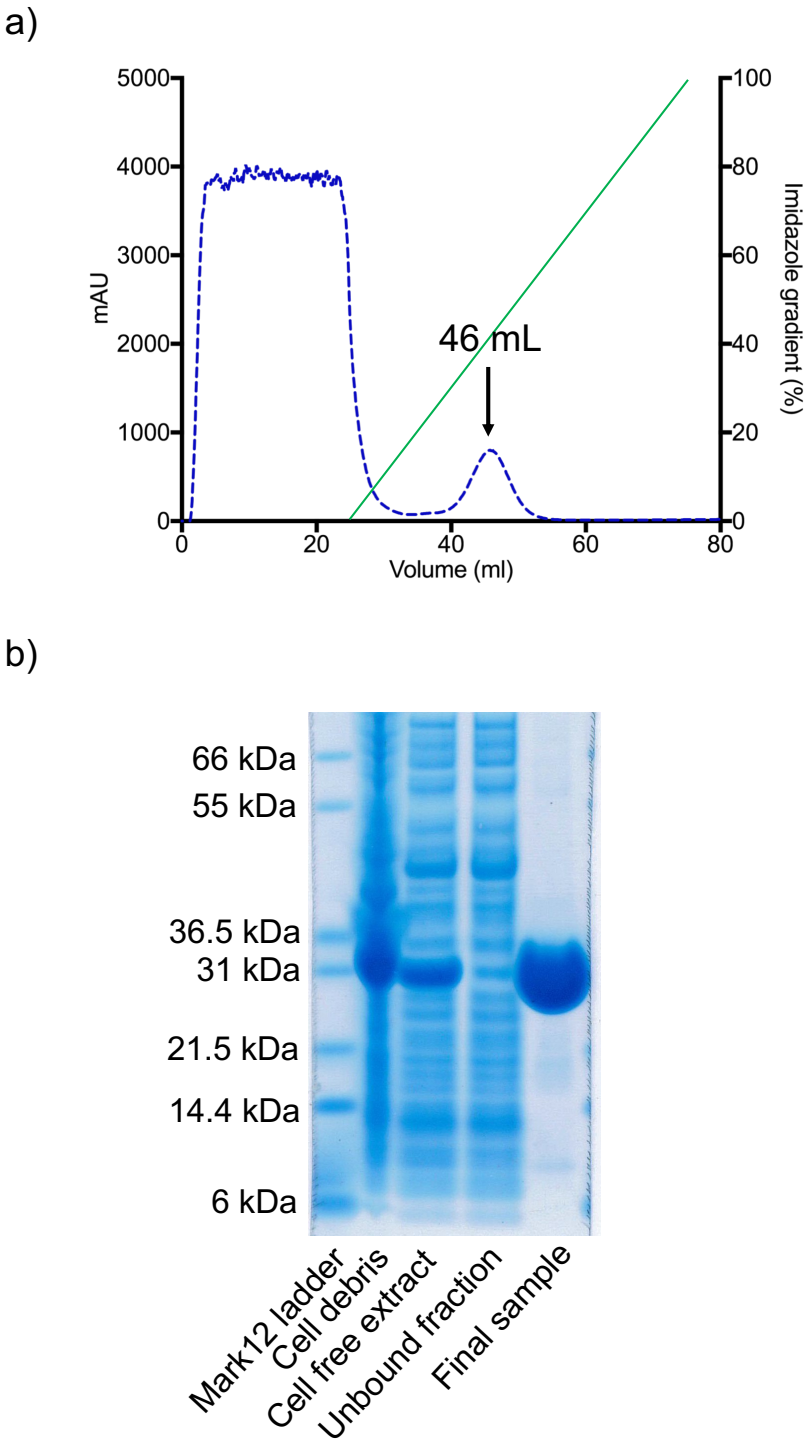


FIGURE A.8: **Protein purification of Ah-TssA his$_6$.Nt2-CTD. a)** Ni-NTA chromatography purification trace of Ah-TssA his$_6$.Nt2-CTD. The green line indicates the percentage of imidazole being applied to the column. The trace indicates a peak at $\sim$ 46 mL indicated by the arrow, which is the elution peak corresponding to Ah-TssA his$_6$.Nt2-CTD. **b)** 4/12% gradient SDS-PAGE gel of the purification of Ah-TssA2 His$_6$.Nt2-CTD. The SDS-PAGE gel shows successful purification indicating that the final sample of Ah-TssA His$_6$.Nt2-CTD is at the expected molecular weight ($\sim$ 29.8 kDa) and is > 95 % pure.

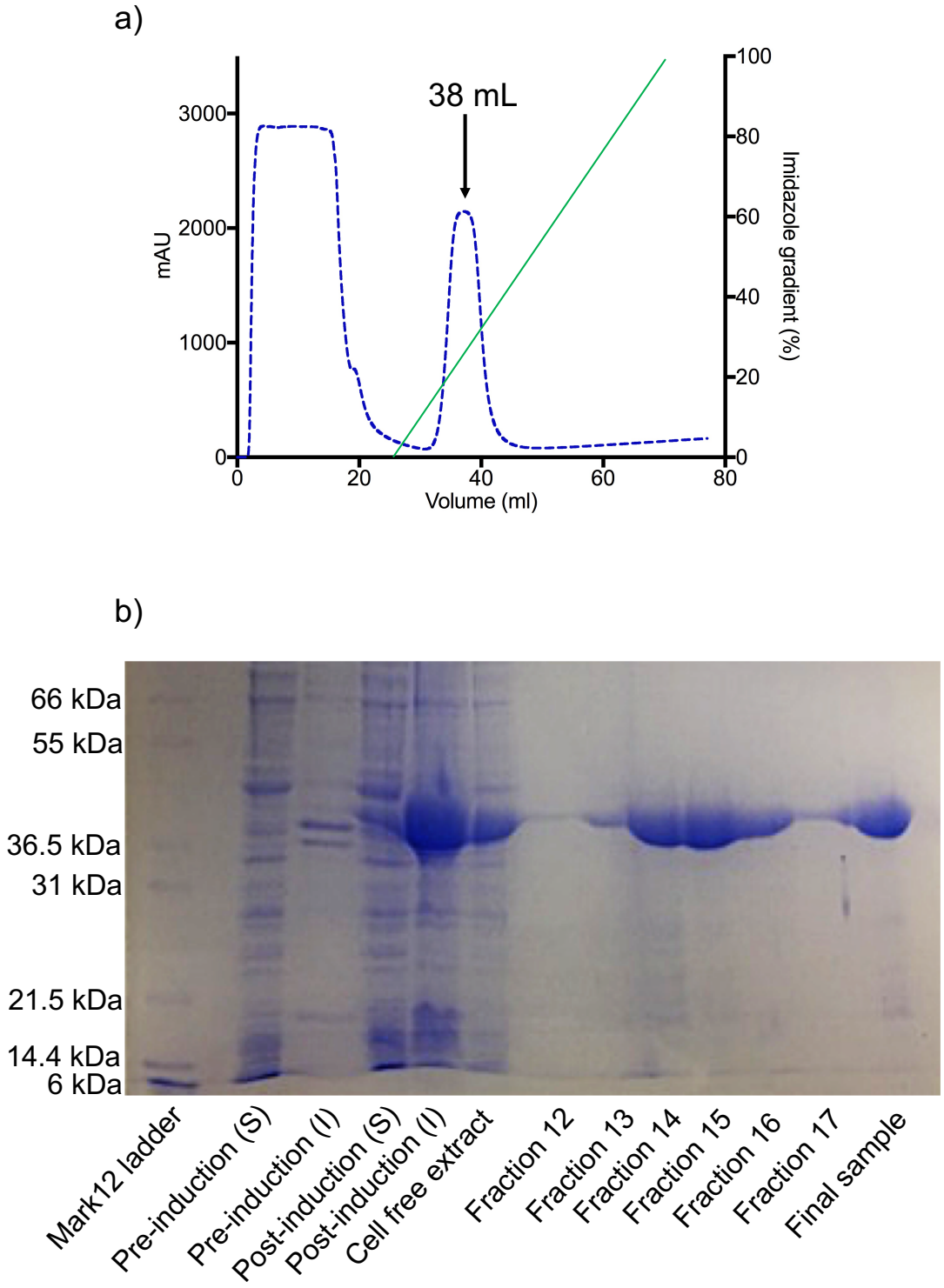


FIGURE A.9: **Protein purification of Ah-TssA Nt1-Nt2.his$_6$. a)** Ni-NTA chromatography purification trace of Ah-TssA Nt1-Nt2.his$_6$. The green line indicates the percentage of imidazole being applied to the column. The trace shows a peak at $\sim$ 38 mL indicated by the arrow, which is the elution peak corresponding to Ah-TssA Nt1-Nt2.his$_6$. **b)** 6/12% SDS-PAGE gel of the purification of Ah-TssA Nt1-Nt2.his$_6$. The SDS-PAGE gel shows successful purification indicating that the final sample of Ah-TssA Nt1-Nt2.his$_6$ is at the correct molecular weight ($\sim$ 41.6 kDa) and is > 95 % pure.

A.5 Crystallisation trials of Ah-TssA domains.

Previous crystallisation trials of the full-length Ah-TssA protein and an Ah-TssA construct containing only the Nt1 domain were unsuccessful in producing protein crystals. As part of this project purified protein samples of Ah-TssA his_6.Nt2, Ah-TssA Nt1-Nt2.his_6, Ah-TssA CTD and Ah-TssA his_6.Nt2-CTD were concentrated as described in 2.5.1 and subjected to crystallisation trials. The details of the crystallisation for Ah-TssA his_6.Nt2, Ah-TssA CTD and Ah-TssA his_6.Nt2-CTD was the subject of 3.1, whereas the detail of Ah-TssA Nt1-Nt2.his_6 was significantly condensed and shall, therefore, be described below.

A.5.1 Crystallisation of Ah-TssA Nt1-Nt2.his_6

Initial crystallisation trials of Ah-TssA Nt1-Nt2.his_6 were carried out as in 2.5.2. To improve the chance of crystallisation two concentrations were used, 10 mg/mL and 20 mg/mL, as mentioned in A.4.4. The plates were kept at 17 °C, and after a month of incubation, crystals had appeared. A single condition was found to generate crystals (0.2 M Sodium chloride, 0.1 M HEPES (pH 7.0) and 20 % w/v PEG 6000) with crystals being identified at both 10 mg/mL and 20 mg/mL protein concentrations. As shown in figure A.10 three different crystal morphologies were identified, and all crystals were of suitable size for subsequent data collection tests. Therefore they were looped, cryo-protected and cryo-cooled according to 2.5.6. Attempts to optimise these crystallisation conditions were unsuccessful in generating crystals.

A.6 X-ray data collection of Ah-TssA2^{A} constructs

A.6.1 Data collection of Ah-TssA2^{A} Nt2

As discussed in 3.1, $\sim$ 20 μg of powdered Ethylmercury phosphate (EMP) was added to crystals of Ah-TssA his_6.Nt2 to generate a heavy atom incorporated derivative crystal. This was intended to provide initial phases taking advantage of the single cysteine residue in the Ah-TssA his_6.Nt2 protein, to which mercury can covalently bond. The heavy atom soaking was carried out as in 2.5.5.

Diffraction quality of the Ah-TssA his_6.Nt2 protein crystals with and without EMP were initially tested as described in 2.6. As discussed in 3.1, a preliminary derivative and native

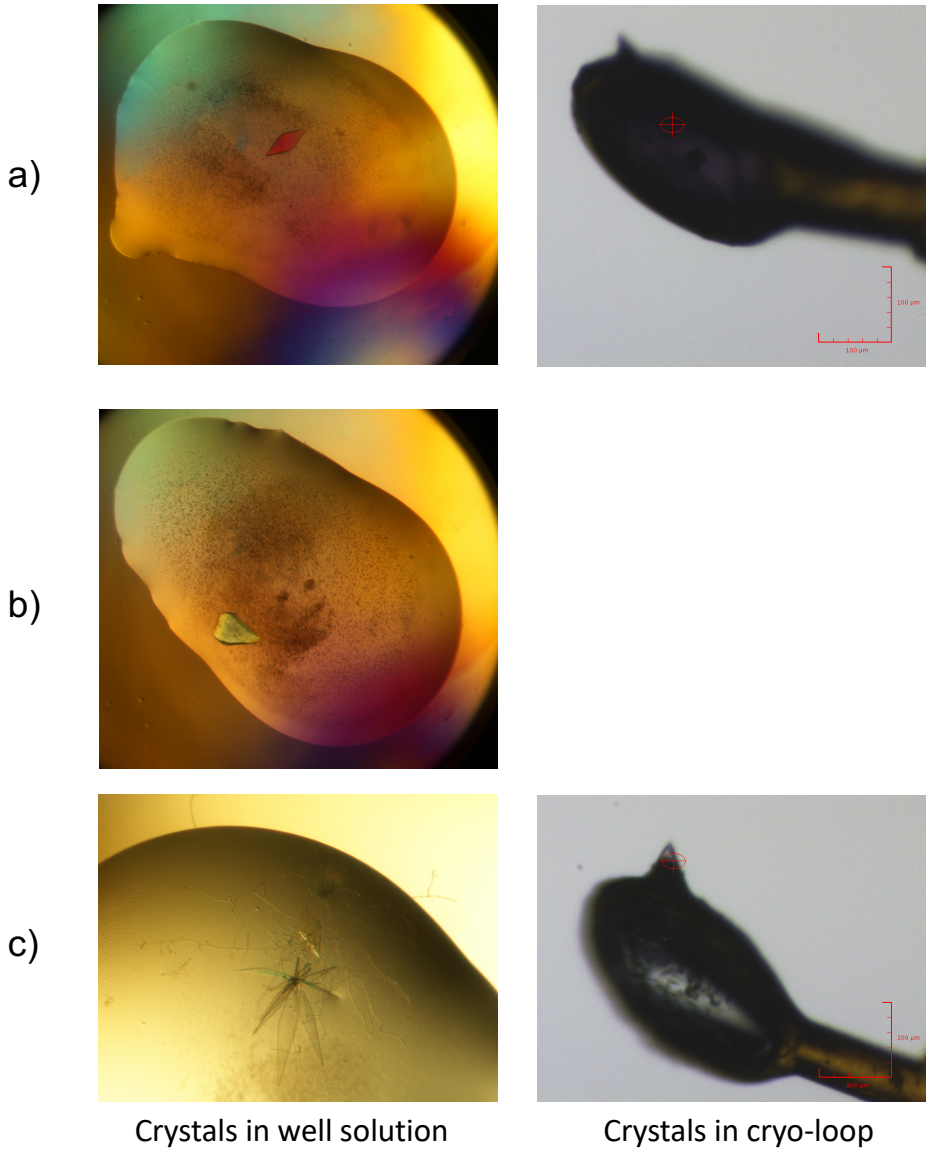


FIGURE A.10: **Ah-TssA Nt1-Nt2.his$_6$ crystals used for data collection.** (left) Images of the crystals taken while they are still growing, (right) images of the crystals once they had been mounted onto the beamline. **a)** Crystal grown in PACT C7 (0.2 M sodium chloride, 0.1 M HEPES (pH 7.0) 20 % w/v PEG 6000) with a protein concentration of 10 mg/mL used, the crystal has a bi-pyramidal morphology. **b)** Crystal grown in PACT C7 with a protein concentration of 20 mg/mL used. **c)** Crystals grown in pH clear D4 (0.2 M sodium chloride, 0.1 M HEPES buffer (pH 7.0), and 20 % (w/v) PEG 6,000) with a protein concentration of 10 mg/mL used. Crystals grew as bi-pyramidal plates emanating from a single nucleation site.

dataset for crystals of Ah-TssA his$_6$.Nt2 were collected and are presented in 3.1, table 3. Subsequently, a higher-resolution dataset of a mercury derivative and native crystal were collected, and data processing statistics can be found in 3.2, table 6. The cell parameters, suggested that the crystals contain two molecules within the asymmetric unit, as determined through the Matthews calculation (figure A.11). This indicated there were two cysteines present within the asymmetric unit, and therefore two mercury atoms to search for during the initial sub-structure determination.

A.6.2 Data collection of Ah-TssA his$_6$.Nt2-CTD

Like that of Ah-TssA his$_6$.Nt2, a preliminary native dataset for Ah-TssA his$_6$.Nt2-CTD was collected (3.1, table 3). Subsequently, a higher resolution dataset was collected, and data processing statistics can be found in 3.2, table 7. While an initial estimate for Ah-TssA his$_6$.Nt2 confidently identified two molecules in the asymmetric unit, the equivalent calculation for Ah-TssA his$_6$.Nt2-CTD was more ambiguous. Although the most probable number of molecules within the asymmetric unit was 14 (A.12), there is a distinct lack of confidence with this value (probability < 0.2). However, this does still indicate a large number of molecules within the asymmetric unit, which correlates with the large unit cell parameters.

A.6.3 Data collection of Ah-TssA Nt1-Nt2.his$_6$

Native data for crystals of Ah-TssA Nt1-Nt2.his$_6$ were collected on DLS MX beamline I04 using an X-ray beam with a wavelength of 0.9795 Å, a start phi of 0 º, 0 º delta, 0.1 º oscillation, an exposure time of 0.05 s exposure and transmission of 100 %. 3600 images were collected and processed using the Xia2 automatic pipelines (Winter, 2010), see 2.6.3 for details. The resultant dataset can be seen in 3.2, supplementary table 6. Initial prediction of the asymmetric content using the Matthews coefficient estimated either one or two molecules to be present. However, these values both resulted in an unfavourable amount of solvent (66 % and 33 % respectively) (figure A.13).

a)

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0218	0.0692	5.02	75.49	17920.00
2	0.9564	0.9149	2.51	50.98	35840.00
3	0.0218	0.0159	1.67	26.46	53760.00

b)

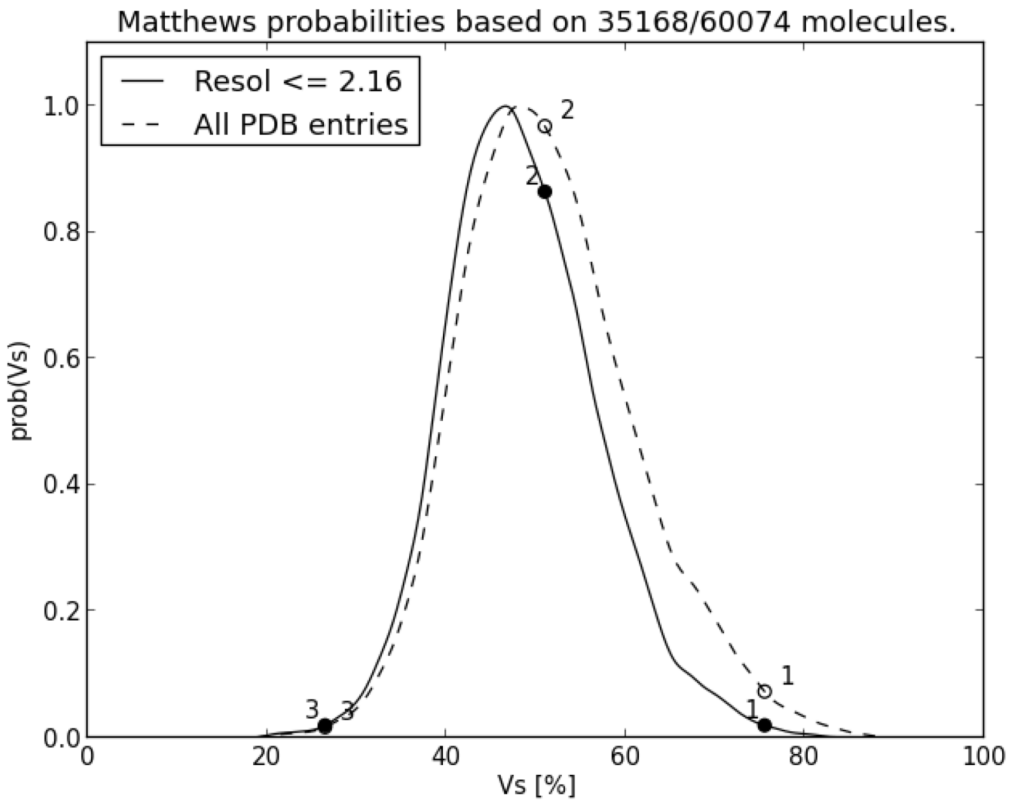


FIGURE A.11: **Ah-TssA his$_6$.Nt2 Matthews calculation. a)** Table of the possible number of molecules within the asymmetric unit with their respective probability and corresponding percentage solvent content. This would suggest two molecules in the asymmetric unit. **b)** A graphical representation of the data presented in a. Figure created using http://www.ruppweb.org/mattprob/default.html (Matthews, 1968)

a)

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0000	0.0000	33.21	96.30	29600.00
2	0.0000	0.0000	16.61	92.59	59200.00
3	0.0000	0.0001	11.07	88.89	88800.00
4	0.0006	0.0013	8.30	85.19	118400.00
5	0.0025	0.0037	6.64	81.48	148000.00
6	0.0058	0.0076	5.54	77.78	177600.00
7	0.0134	0.0157	4.74	74.08	207200.00
8	0.0250	0.0282	4.15	70.37	236800.00
9	0.0374	0.0407	3.69	66.67	266400.00
10	0.0598	0.0618	3.32	62.97	296000.00
11	0.0897	0.0907	3.02	59.26	325600.00
12	0.1253	0.1252	2.77	55.56	355200.00
13	0.1522	0.1498	2.55	51.86	384800.00
14	0.1626	0.1586	2.37	48.15	414400.00
15	0.1450	0.1411	2.21	44.45	444000.00
16	0.1020	0.0990	2.08	40.75	473600.00
17	0.0480	0.0467	1.95	37.05	503200.00
18	0.0202	0.0196	1.85	33.34	532800.00
19	0.0069	0.0067	1.75	29.64	562400.00
20	0.0022	0.0022	1.66	25.94	592000.00
21	0.0011	0.0011	1.58	22.23	621600.00
22	0.0001	0.0001	1.51	18.53	651200.00
23	0.0000	0.0000	1.44	14.83	680800.00
24	0.0000	0.0000	1.38	11.12	710400.00

b)

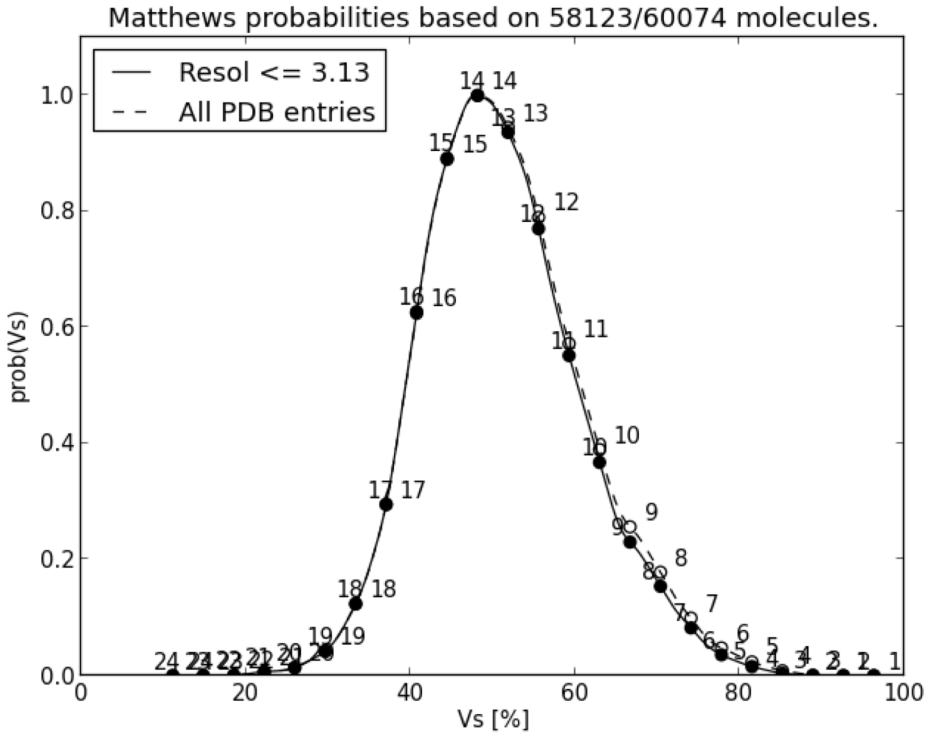


FIGURE A.12: **Ah-TssA his$_6$.Nt2-CTD Matthews calculation. a)** Table of the possible number of molecules within the asymmetric unit with their respective probability and corresponding percentage solvent content. This would suggest about 14 molecules in the asymmetric unit as indicated by the red arrow. **b)** A graphical representation of the data presented in a. Figure created using http://www.ruppweb.org/mattprob/default.html (Matthews, 1968)

a)

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.2136	0.6900	3.68	66.54	41600.00
2	0.7864	0.3100	1.84	33.09	83200.00

b)

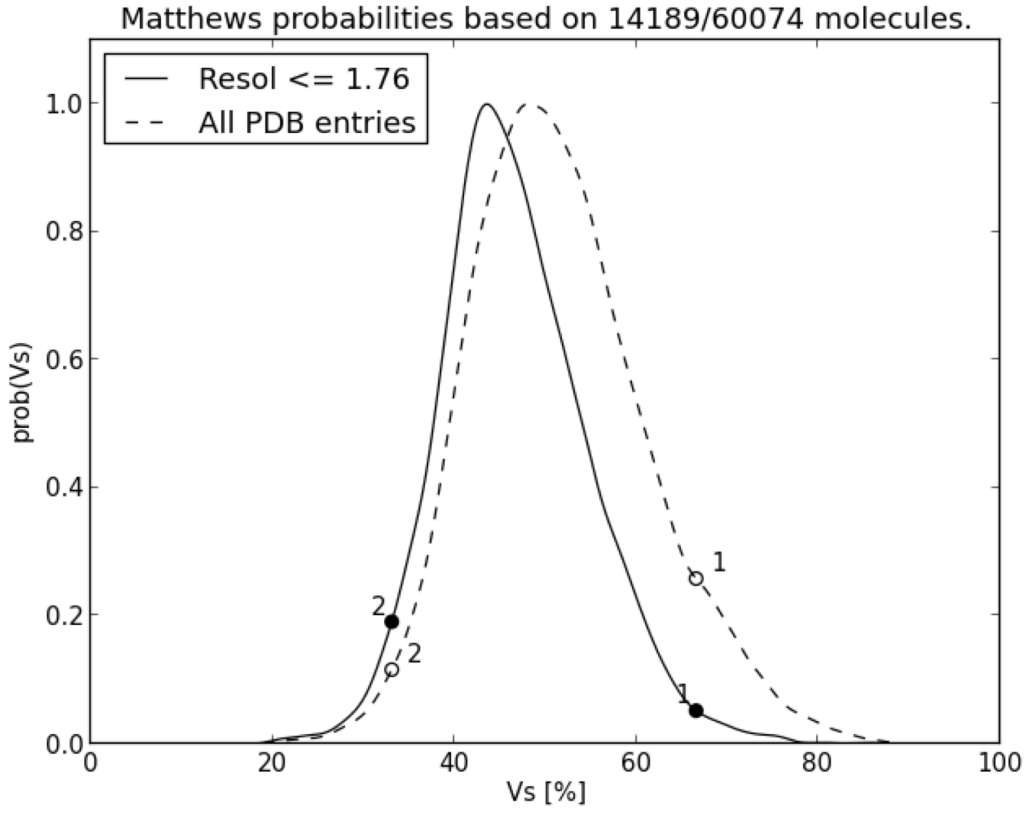


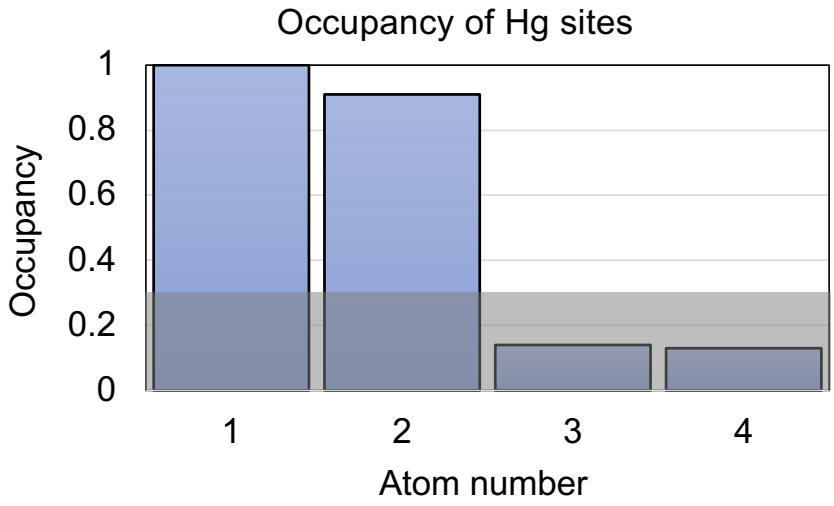
FIGURE A.13: **Ah-TssA Nt1-Nt2.his$_6$ Matthews calculation. a)** Table of the possible number of molecules within the asymmetric unit with the corresponding percentage solvent content. This indicates either one or two molecules in the asymmetric unit, however with unfavourable % solvent values **b)** A graphical representation of the data presented in a. Figure created using http://www.ruppweb.org/mattprob/default.html (Matthews, 1968)

A.7 Structure determination and initial model building Ah-TssA his_6.Nt2

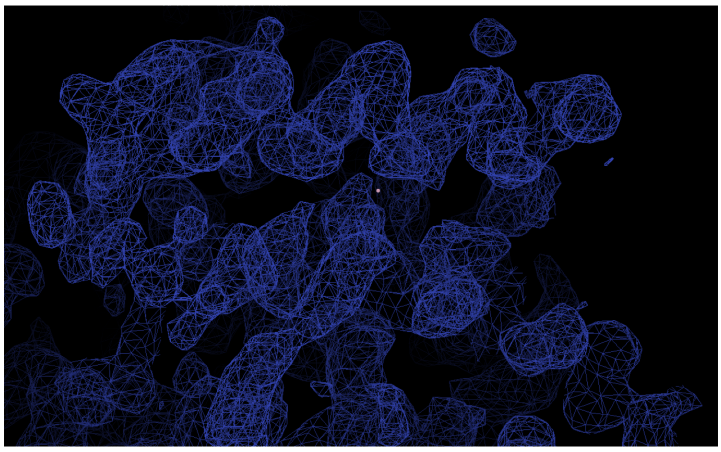
The Ah-TssA his_6.Nt2 dataset was collected from an Ah-TssA his_6.Nt2 mercury derivative crystal exposed to X-rays with a wavelength of 1.0088 Å, the resultant dataset had an anomalous mid-slope of 1.061 (3.2, table 6). Attempts to calculate initial phases used the CRANK2 software pipeline (Skubák and Pannu, 2013), with the optional SHELX software package (Sheldrick, 2008). CRANK2 is an experimental phasing pipeline which integrates the substructure determination, density modification, iterative model building and refinement into a single manageable package. Initial analysis with SHELXC identified anomalous signal > 0.8 d"/sig for the full extent of the data. Subsequently, SHELXD identified four mercury sites, and inspection of them indicated a high occupancy for only two of the four sites (figure A.14a). This is consistent with a single mercury atom for each of the two Ah-TssA his_6.Nt2 molecules within the asymmetric unit. These sites were passed to SHELXE which calculated the initial phases and carried out subsequent rounds of density modification and model building. Both hands were tested, and the original hand was selected. Observation of the electron density map obtained from both hands confirmed that the original hand produced the best map as can be seen in figure A.14b whereby continuous electron density corresponding to α-helices can easily be seen.

SHELXE also generated an initial model of the protein from the initial electron density map (figure A.15 (upper)). The initial map and model, along with the protein sequence of Ah-TssA his_6.Nt2 were used by Buccaneer to carry out five cycles of iterative automated model building and refinement (Cowtan, 2006), generating a nearly complete model for Ah-TssA Nt2 (figure A.15 (lower)). The model generated from Buccaneer had a final R_{work} and R_{free} of 0.27 and 0.33, respectively. Initial observation of the model suggested two molecules of Ah-TssA Nt2 present in the asymmetric unit, in agreement with the Matthews calculation. Manual observation of the resultant model did not identify any large errors in the model building process that conflicted with the electron density map, as a consequence this model was subsequently used as the search model in molecular replacement of the Ah-TssA Nt1-Nt2.his_6 data.

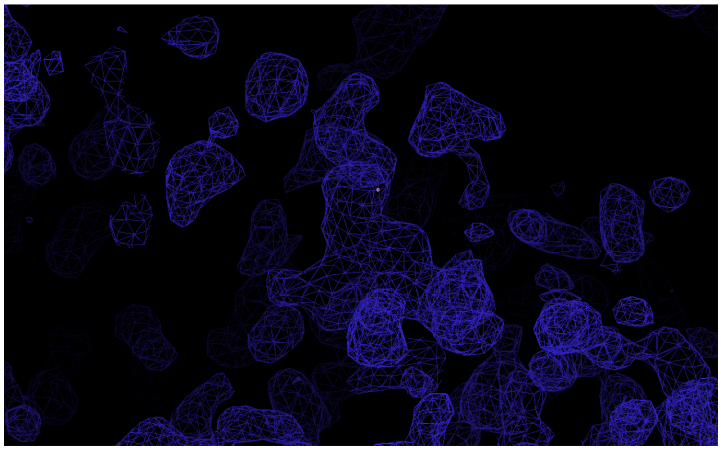
a)



b)

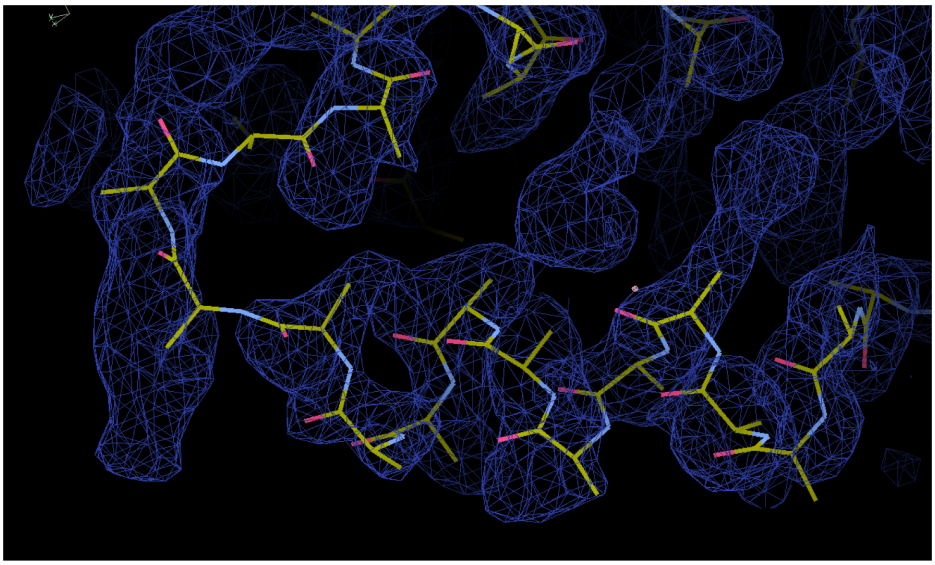


Original hand

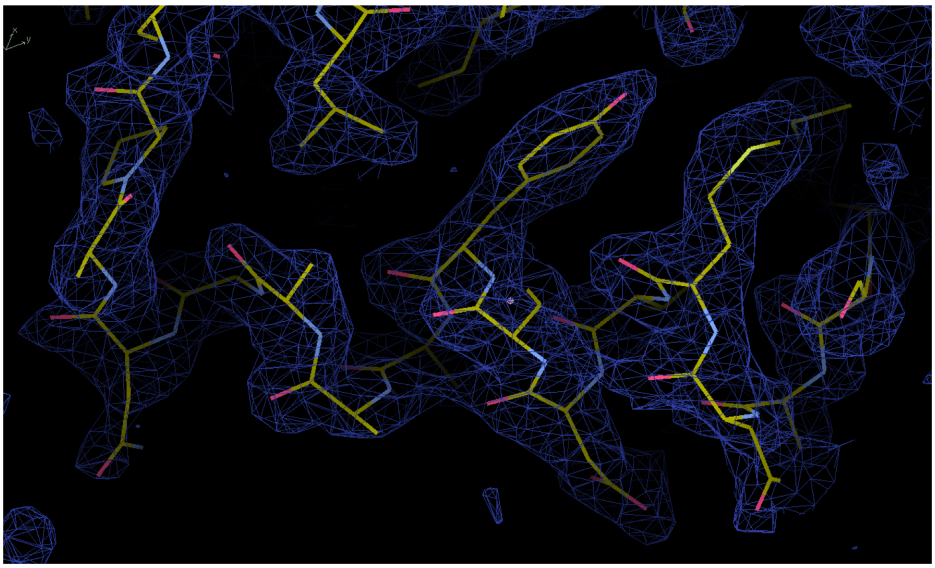


Inverted hand

FIGURE A.14: **Initial phase determination of Ah-TssA his$_6$.Nt2. a)** Graph representing the occupancy for Mercury atoms found by SHELXD. Values lower than 0.3 (grey area) are not included in subsequent phase determination calculations. **b)** 2Fo-Fc maps for both hands generated from the initial phase determination of Ah-TssA his$_6$.Nt2 by SHELXE. The original hand (upper) shows continuous density with distinct secondary structure features.



Initial map



Final Map

FIGURE A.15: **Initial vs. final electron density map for Ah-TssA Nt2. a)** Initial map and model generated by SHELXE. **b)** Final map and model generated by Buccaneer.

A.8 Structure determination, model building and validation of Ah-TssA Nt1-Nt2.his$_6$

As crystals of Ah-TssA Nt1-Nt2.his$_6$ were presumed to contain copies of the Ah-TssA Nt2 domain, the model built for the Ah-TssA Nt2 domain was used as a search model in molecular replacement (PhaserMR (McCoy et al., 2007)) to provide initial phases for the Ah-TssA Nt1-Nt2.his$_6$ data. Calculation using the Matthews coefficient estimated two molecules of Ah-TssA Nt1-Nt2.his$_6$ in the asymmetric unit; therefore two molecules of Ah-TssA Nt2 were used as search models to generate the initial phases for the Ah-TssA2 Nt1-Nt2.his$_6$ data.

Four solutions, each of two molecules in the asymmetric unit, were found with the top refined TFZ of 15.7, an overall LLG of 1879, and no packing clashes, indicating a successful solution had been found. Initial inspection of the resultant 2Fo-Fc/Fo-Fc electron density map indicated the presence of unaccounted electron density, however, upon refinement of the map it could be identified that this density could be occupied by additional molecules of Ah-TssA Nt2 (figure A.16a) rather then density for the missing Nt1 domain. As a consequence, phaserMR was repeated searching for four molecules of Ah-TssA Nt2, which generated improved statistics (refined TFZ of 38.3, an overall LLG of 7187, and no packing clashes). Analysis of the resultant electron density map indicated a complete lattice consisting of four molecules of Ah-TssA Nt2 with no space for additional residues (figure A.16b) that would constitute the Ah-TssA Nt1 domain. Therefore it was concluded that only the Ah-TssA Nt2 domain had crystallised, possibly a result of adventitious protease clipping during crystal growth resulting in the loss of the Nt1 domain and its associated linker. This could also explain why optimisation of the crystallisation conditions did not yield further crystals and why all three crystallisations result in the apparent production of so few crystals. However, rather serendipitously, this did produce a dataset with greater resolution (1.8 Å) than that for Ah-TssA his$_6$.Nt2 crystals (2.3 Å) alone. Therefore, hereinafter the structure of the Nt2 domain derived from attempts to crystallise the Nt1-Nt2 protein will be referred to as Ah-TssA Nt2 model. The dataset at 1.8 Å was used to build the final Ah-TssA Nt2 model (figure A.16c). The solution to PhaserMR was subjected to multiple rounds of manual model building and refinement, including the addition of solvent where appropriate until the final model was built and the map could no longer be interpreted which resulted in a final R_{work} and R_{free} of 0.27 and 0.30, respectively. The final refinement statistics can

be found in 3.2, table 6. The final model from Ah-TssA Nt2 crystals contains four chains of Ah-TssA Nt2 (Chains A-D) and 431 waters. Structure validation was carried out using the Molprobity validation server (Chen et al., 2010). For the given resolution of the data, this placed the structure in the 100 th percentile when compared to other structures in the database (see 3.2, table 6 for details). In addition, analysis of the Ramachandran angles indicated that 98.8 % were favourable, with no outliers. The model was subsequently deposited under the PDB ID: 6G7B.

A.9 Structural analysis of Ah-TssA Nt2 domain

The model for Ah-TssA Nt2 crystals was shown to contain four chains of Ah-TssA Nt2 within the asymmetric unit (Chains A-D) (figure A.18a), where each chain is comprised of 146 residues (D231-E376). The Nt2 domain of Ah-TssA consists of seven α-helices with helices 1-6 forming an antiparallel helical bundle. $\alpha 7$ is perpendicular to this bundle, where the protein would transition into the linker to the CTD (figure A.18b). Analysis of the structure reveals that Ah-TssA Nt2 appears to be present as dimers within the asymmetric unit (figure A.18c). This is consistent with the molecular weight estimation by gel filtration, where Ah-TssA Nt2 migrates with an apparent molecular weight equivalent to a dimer (3.1). Furthermore, during this study the structure of the *E. coli* TssA (Ec-TssA) Nt2 was determined (Zoued et al., 2016), allowing for a direct structural comparison between both TssA Nt2 domains. In general, both structures are very similar (figure A.18d), such that a structural alignment of the Ec-TssA Nt2 and Ah-TssA generates an RMSD difference of 2.3 Å. This is unsurprising as both are TssA2 representatives, and share 25 % sequence identity as discussed in 3.2.

A.9.1 Biological assembly of Ah-TssA Nt2 domain

The structure of Ah-TssA Nt2 domain was analysed using the ePISA (Krissinel and Henrick, 2007) server to identify if the interfaces generated by the four protein chains had any biological significance. This identified three unique interfaces between the protein molecules (A<>B, C<>D), (A<>D, B<>C) and (A<>A, B<>B, C<>C, D<>D). Details of these interfaces are shown in figure A.19. This showed that the interface relating chain A<>B and C<>D (figure A.19a) have the largest buried surface area ($\sim$ 650 Å^2), suggesting that this is the

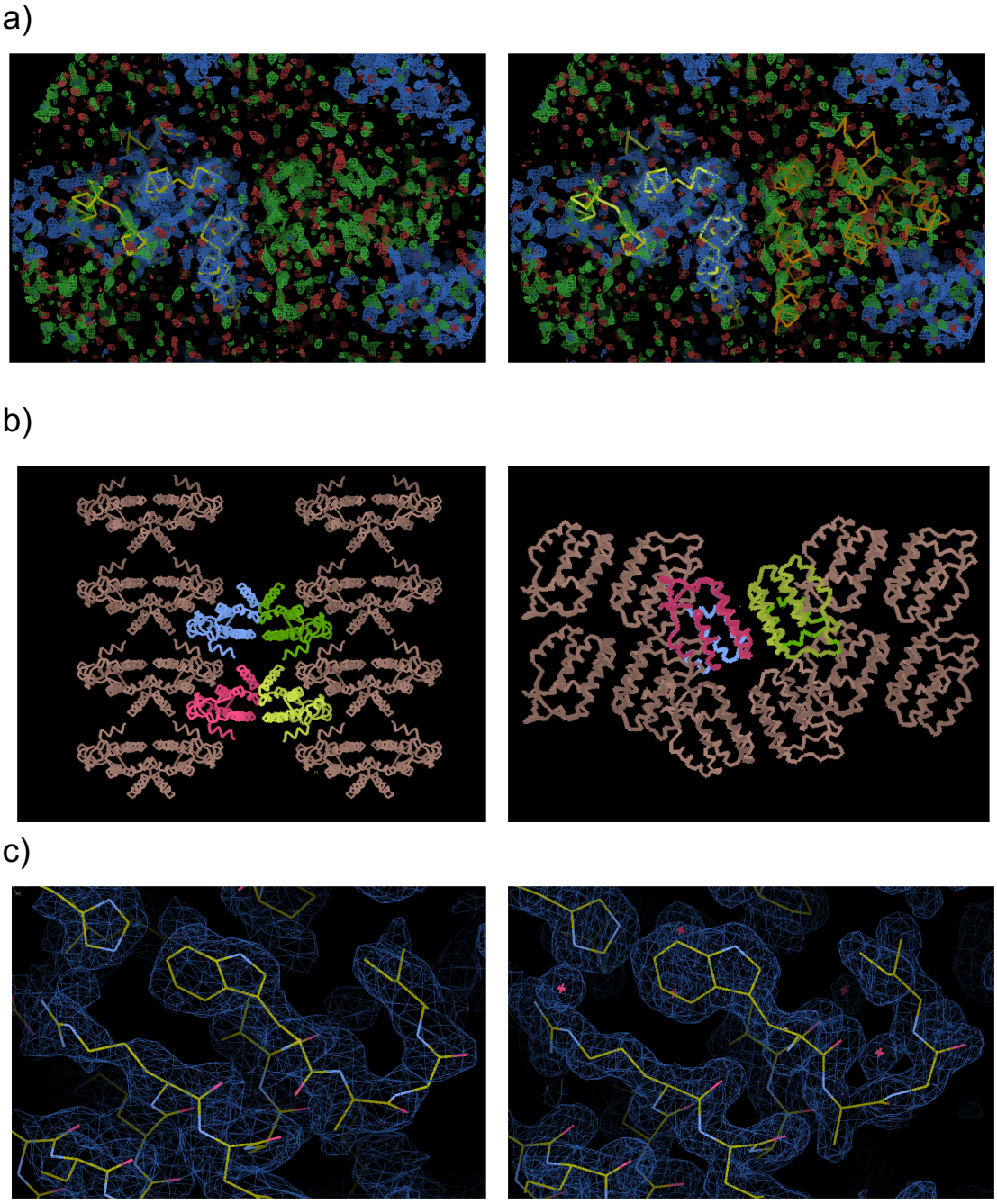


FIGURE A.16: **Ah-TssA Nt1-Nt2.his$_6$ structure determination. a)** (left) Initial map generated after phaserMR, the placed Ah-TssA Nt2 domain is shown in yellow. There is additional density unaccounted for and space in the lattice for additional residues. (right) Superimposed dimer of Ah-TssA Nt2 (orange) indicating that the density can be accounted for by an additional Ah-TssA Nt2 domain. Fo-Fc and 2Fo-Fc maps at contour levels of 2.5 and 2.9 RMSD, respectively. **b)** Packing of the Ah-TssA Nt2 molecule in the lattice. (left) View down the a-axis. (right) View down the c-axis. Both views indicate that there is no space in the lattice for an addition (Ah-TssA Nt1) domain. **c)** Representative electron density map for the lower resolution Ah-TssA Nt2 dataset (left) and higher resolution Ah-TssA Nt1-Nt2 dataset (right), both 2Fo-Fc maps at a contour level of 1.0 RMSD.

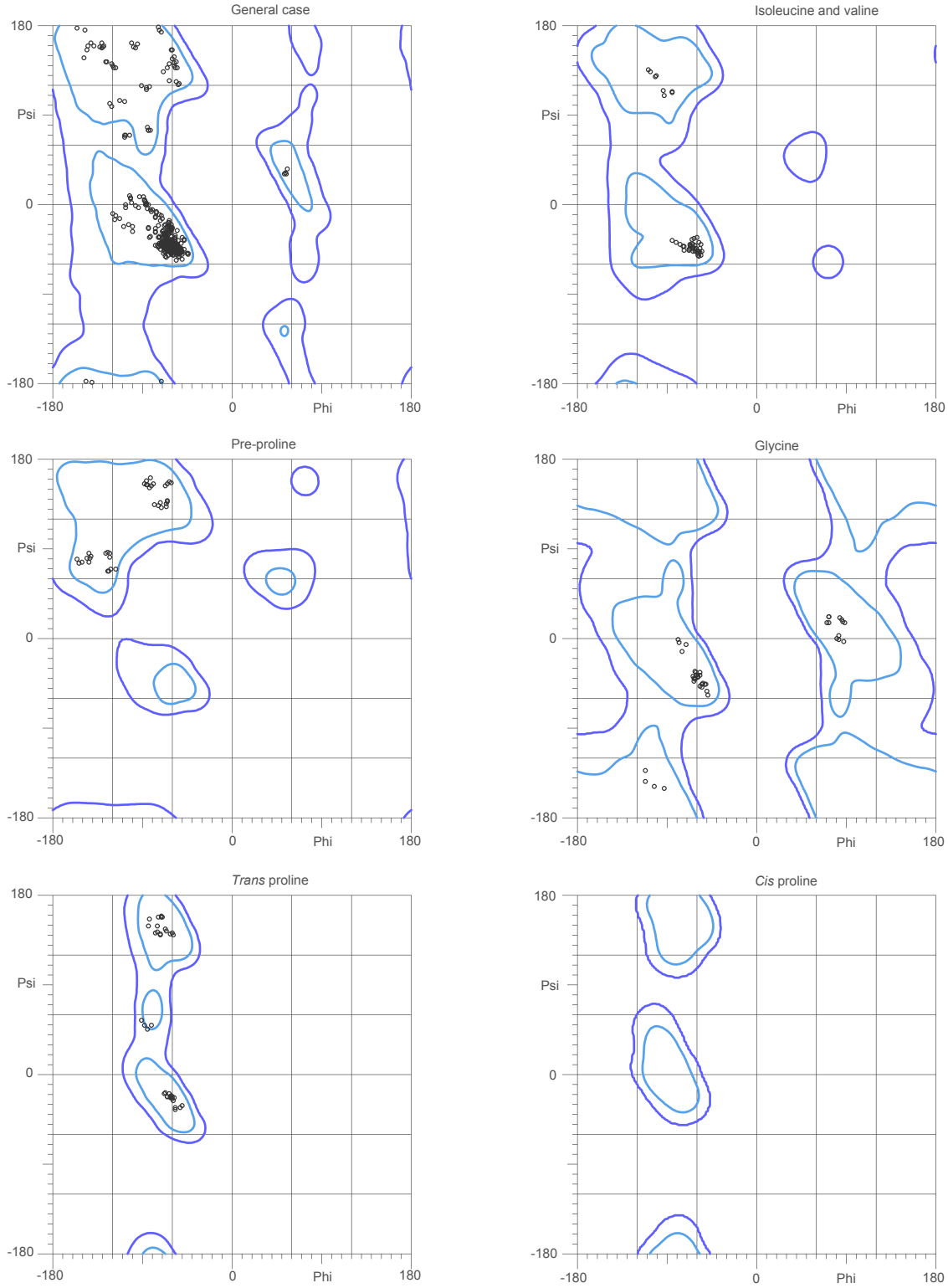


FIGURE A.17: **Ah-TssA Nt2 model Ramachandran analysis.** Analysis of the phi and psi torsion angles for different residue classes. These include the general case, isoleucine and valine, pre-proline, glycine, trans proline and cis proline. Light blue contour levels denote regions for favoured Ramachandran angles, dark blue allowed angles, and outside these are outliers. This map indicates there are no Ramachandran outliers within the Ah-TssA Nt2 model. Figure created using Chen et al., 2010.

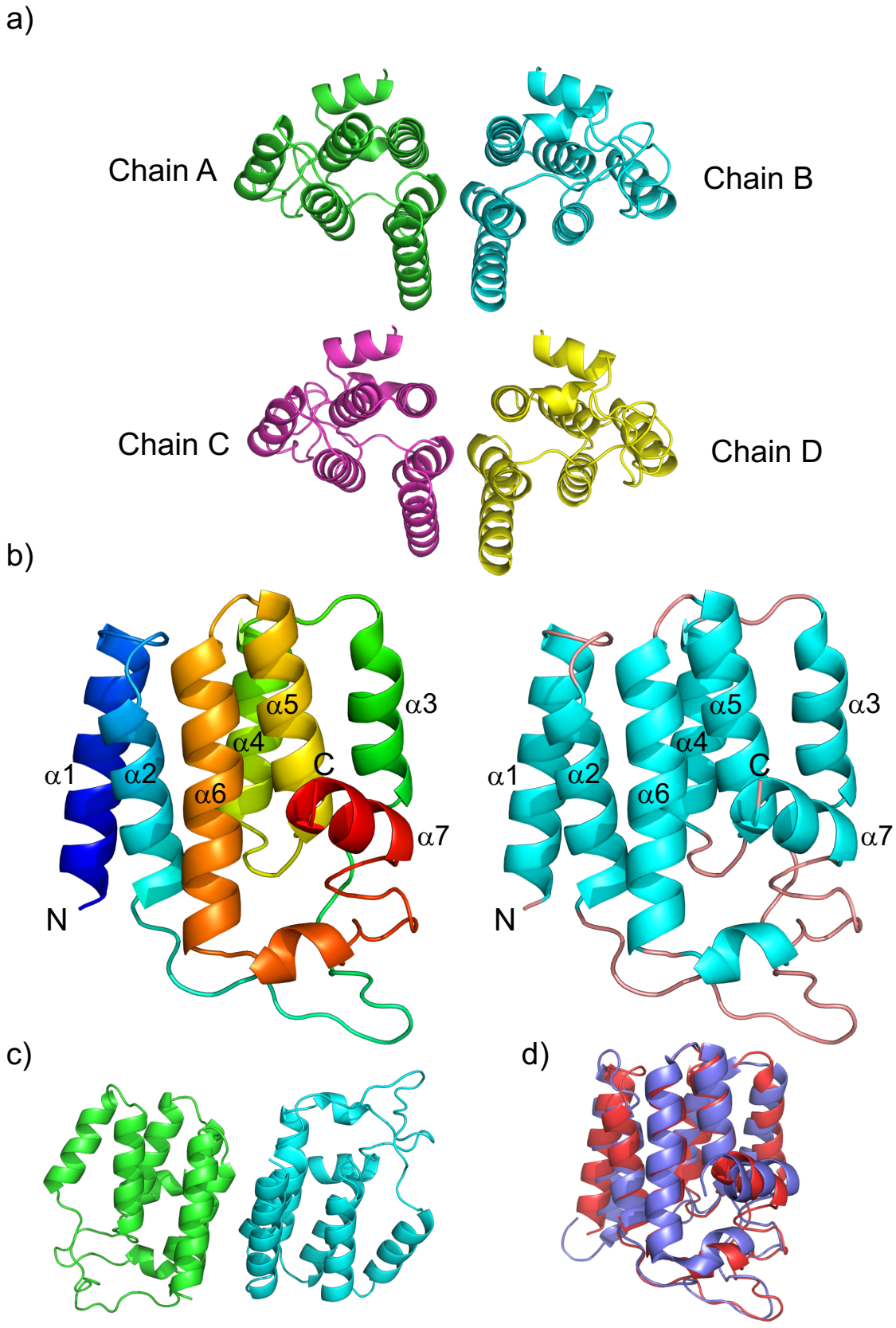


FIGURE A.18: **Structural analysis of the Ah-TssA Nt2 domain. a)** The content of the asymmetric unit from TssA Nt1-Nt2.his$_6$ crystals. Four molecules of Ah-TssA Nt2 appear to be arranged as two dimers (A<>B, C<>D). **b)** (left) Chainbow representation of Ah-TssA Nt2. (right) Helical representation of Ah-TssA Nt2 indicating the structure is comprised of seven α-helices. **c)** An alternative view of the Ah-TssA Nt2 proposed dimer. **d)** Structural alignment of the Ah-TssA Nt2 domain (red) and Ec-TssA Nt2 domain (blue). The alignment indicates a strong structural relationship between the domains.

biological dimer displayed by Ah-TssA Nt2 as initially expected. Consequently, the asymmetric unit of Ah-TssA Nt1-Nt2.his$_6$ crystals are comprised of two Ah-TssA Nt2 dimers as initially described.

A.10 Structure determination, model building and validation of Ah-TssA Nt2-CTD

While an initial Matthews calculation predicts between 10 - 18 copies of Ah-TssA his$_6$.Nt2-CTD within the asymmetric unit, with the most probable being 14, SEC-MALLS analysis suggests the complex is comprised of 10 molecules (0.3 MDa ($\sim$ 10 x 30kDa)) (3.1). The initial strategy for solving the Ah-TssA his$_6$.Nt2-CTD structure was to use the Ah-TssA Nt2 domain model as the search model. The identification that the Ah-TssA Nt2 domain is present as a dimer in solution (3.1), and is present as a dimer within the Ah-TssA Nt2 crystal, suggested that Ah-TssA Nt2 will also be present as a dimer within the Ah-TssA2^A his$_6$.Nt2-CTD crystal structure. Therefore, molecular replacement was carried out using dimers of Ah-TssA Nt2 as the search model. PhaserMR was able to successfully place five dimers of Ah-TssA Nt2 within the asymmetric unit generating a single solution with a refined TFZ of 23.52, an overall LLG of 2614.9, and two packing clashes. Initial inspection of the resultant electron density map indicated that molecular replacement had been successful. However, there was no obvious symmetry generated by the Nt2 dimers (figure A.20a). In addition, there were large gaps in the lattice suggesting space for addition residues (figure A.20b). Furthermore, additional features could be observed in the Fo-Fc map (figure A.20b (right) + c). This un-modelled density appeared to contain features that could be attributed to a star-like structure with five-fold symmetry, consistent with observations of the Ah-TssA CTD domain oligomer when analysed via negative stain EM (3.2, figure 6a). The final refinement of this solution from PhaserMR had a R_{work} and R_{free} of 0.41 and 0.45, respectively. To aid model building, Parrot (Cowtan, 2010) was utilised to take advantage of the local non-crystallographic symmetry (NCS) generated from the 10 Ah-TssA Nt2 molecules within the asymmetric unit with 90 unique symmetry operators between the Ah-TssA Nt2 domains. After four cycles the figure of merit had improved from 0.42 to 0.76 and the subsequent electron density map had sufficient density to identify secondary structure features consistent with α-helices within the un-modelled region. This enabled the manual addition of

a)

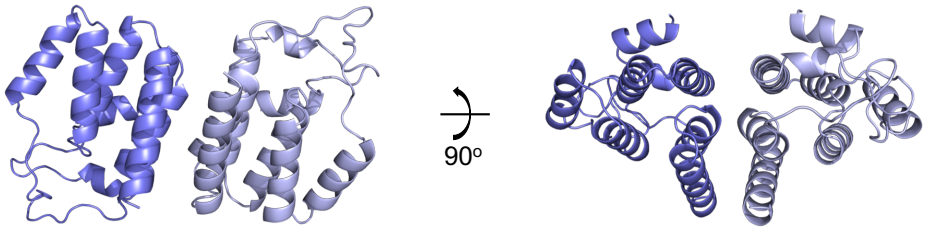
Nt2 Chains	Average interface (A^2)	Average Δ G (kcal/mol)	Average Δ G P-Value	No. of H-Bonds	No. of salt bridges
A<>B, C<>D	646.9	-9.3	0.205	3	3

H-Bonds

##	Structure 1	Dist. [Å]	Structure 2
1	B:ARG 248[NH1]	3.16	A:ASP 231[OD1]
2	B:HIS 261[NE2]	2.64	A:GLN 249[OE1]
3	B:GLN 249[OE1]	2.65	A:HIS 261[NE2]

Salt bridges

##	Structure 1	Dist. [Å]	Structure 2
1	B:ARG 248[NH1]	3.16	A:ASP 231[OD1]
2	B:ARG 248[NH2]	3.26	A:ASP 231[OD1]
3	B:ARG 248[NH2]	3.87	A:ASP 231[OD2]



b)

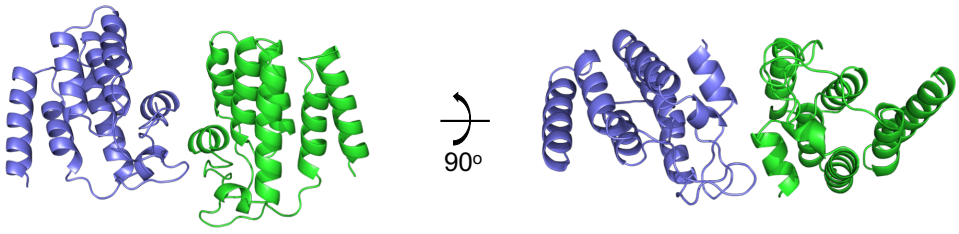
Nt2 Chains	Average interface (A^2)	Average Δ G (kcal/mol)	Average Δ G P-Value	No. of H-Bonds	No. of salt bridges
B<>C, D<>A	346.9	0.8	0.798	5	6

H-Bonds

##	Structure 1	Dist. [Å]	Structure 2
1	B:ARG 293[NH1]	2.86	C:GLU 369[OE2]
2	B:ARG 372[NH2]	2.78	C:GLY 276[O]
3	B:ARG 372[NH2]	3.59	C:GLU 356[O]
4	B:ASP 290[OD1]	2.93	C:ARG 372[NH2]
5	B:ASP 290[OD2]	2.88	C:ARG 372[NH1]

Salt bridges

##	Structure 1	Dist. [Å]	Structure 2
1	B:ARG 293[NH1]	2.86	C:GLU 369[OE2]
2	B:ARG 293[NH1]	3.93	C:GLU 369[OE1]
3	B:ASP 290[OD1]	3.31	C:ARG 372[NH1]
4	B:ASP 290[OD1]	2.93	C:ARG 372[NH2]
5	B:ASP 290[OD2]	2.88	C:ARG 372[NH1]
6	B:ASP 290[OD2]	3.48	C:ARG 372[NH2]



c)

Nt2 Chains	Average interface (A^2)	Average Δ G (kcal/mol)	Average Δ G P-Value	No. of H-Bonds	No. of salt bridges
A<>A, B<>B, C<>C, D<>D	157.6	-1.9	0.375	1	0

H-Bonds

##	Structure 1	Dist. [Å]	Structure 2
1	B:THR 268[OG1]	3.67	B:ASP 300[OD2]

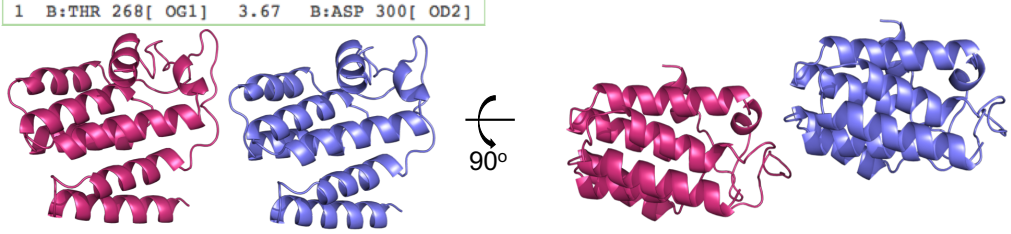


FIGURE A.19: **ePISA analysis of the Ah-TssA Nt2 structure.** For each interface the chains involved, average interface size, average delta G, average delta G P-value, number of hydrogen bonds (including what they are), number of salt bridges (including what they are) and the protein assembly that forms the interface are shown. **a)** Interface analysis for chains A<>B and C<>D. **b)** Interface analysis for chains A<>D and B<>C. **c)** Interface analysis for chains A<>A, B<>B, C<>C and D<>D.

poly-ala α-helices to the map (figure A.20d), generating a molecule which appeared to be arranged in D_5 symmetry. After successive rounds of manual model building, refinement and NCS averaging the final model for the Ah-TssA Nt2-CTD crystal was completed with a final R_{work} and R_{free} of 0.26 and 0.33, respectively. The final model consisted of 10 molecules of Nt2-CTD (chains A-J) which largely cover the full-length of the construct, apart from residues in the interdomain linker. However, there was a lack of density for the interdomain linker; therefore the assignment of Nt2 domains to CTD domains was challenging (see A.11 for details). Analysis of the Ramachandran angles in the final model indicated 90.5 % favoured angles and 1.35 % outliers (figure A.21). Final model validation statistics can be found in 3.2, table 7.

A.11 Structural analysis of Ah-TssA2^{A} Nt2-CTD

The final Ah-TssA Nt2-CTD model was shown to contain 10 subunits of the Ah-Nt2 domain organised as five dimers (figure A.22a (left)). In addition, 10 monomers of the Ah-TssA2^{A} CTD, were identified. Each Ah-TssA CTD monomer consists of a helical bundle consisting of five antiparallel α-helices. The 10 Ah-TssA CTD monomers are arranged as an oligomer which displays D_5 symmetry. Furthermore, the structure of the Ah-TssA CTD domain shows a striking similarity to that of the Ec-TssA CTD domain, but with 41 residues missing at the C-terminus as a consequence of sequence differences. Full details of the assembly of the Ah-TssA CTD and comparisons with the Ec-TssA CTD domain can be found in 3.2.

A.11.1 Defining the Ah-TssA Nt2-CTD asymmetric unit

Upon completion of the Ah-TssA Nt2-CTD model, it was not obvious which Ah-TssA Nt2 domains belonged to their corresponding Ah-TssA CTD due to a lack of density for the interdomain linker between the Ah-TssA Nt2 and Ah-TssA CTD. However, it is clear that the Ah-TssA Nt2 domains are maintained as dimers within the crystal lattice, consistent with the structure of the Ah-TssA Nt2 domain in isolation. This suggests that the assembly of Ah-TssA Nt2-CTD oligomer does not disturb the Ah-TssA Nt2 dimer. As shown in figure A.22a (left) the initial placement of the Ah-TssA Nt2 domains do not correspond with the arrangement of the Ah-TssA CTD in the pentameric ring, such that it was difficult to assign

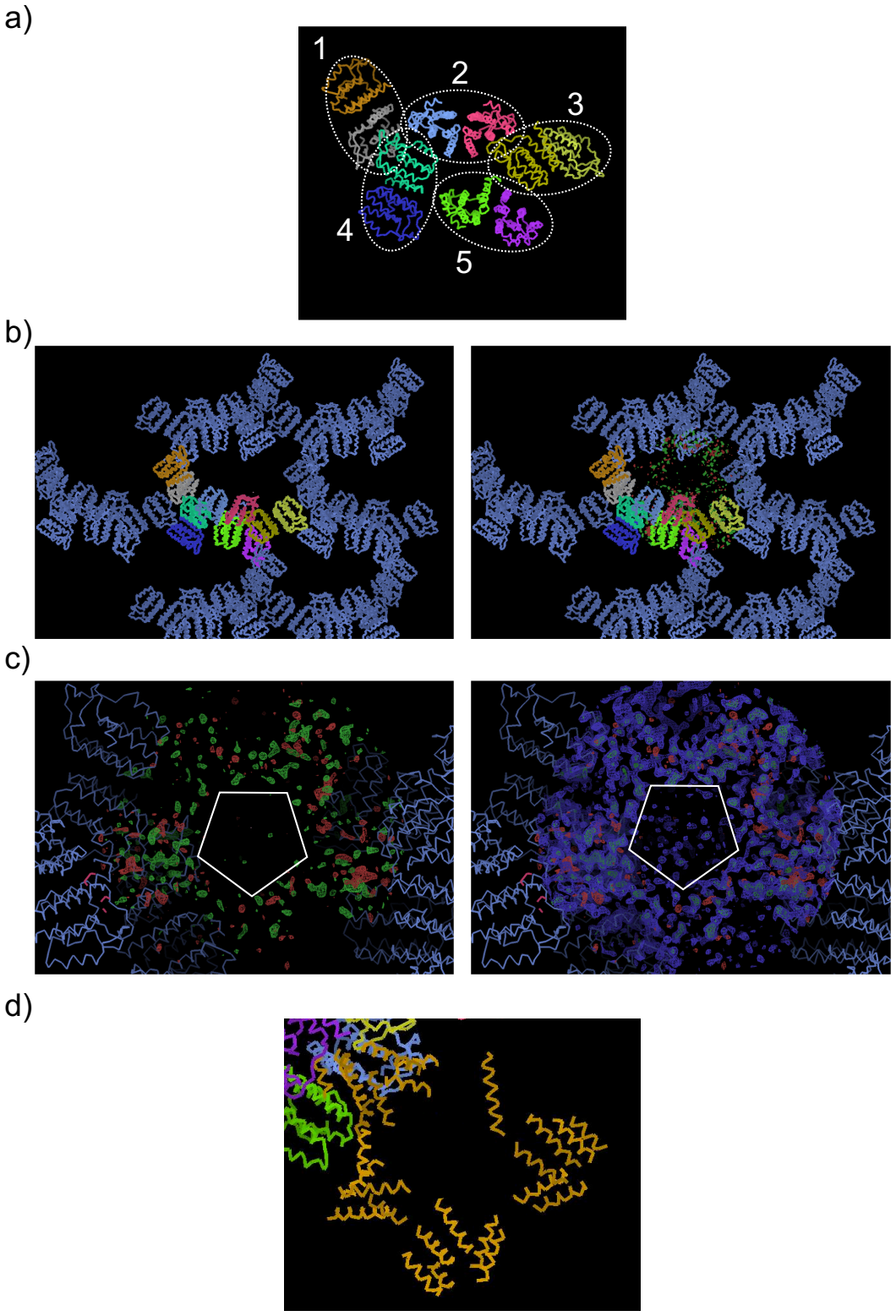


FIGURE A.20: **Structure determination of Ah-TssA his$_6$.Nt2-CTD. a)** Initial placement of the five Ah-TssA Nt2 dimer within the Ah-TssA his$_6$.Nt2-CTD asymmetric unit. **b)** View of the symmetry-related Ah-TssA Nt2 molecules within the Ah-TssA his$_6$.Nt2-CTD crystal lattice. (right) Fo-Fc map indicating the presence of additional residues not currently modelled. **c)** (left) A closer view of the Fo-Fc map around the un-modelled density. The map appears to display five-fold symmetry with a distinct hole in the centre. (right) The same view however with map output from parrot displayed. **d)** preliminary placement of poly-ala α-helices, they are starting to form a molecule with D_5 symmetry. The Fo-Fc and map output from parrot have an RMSD of 3.0 and 1.3 respectively.

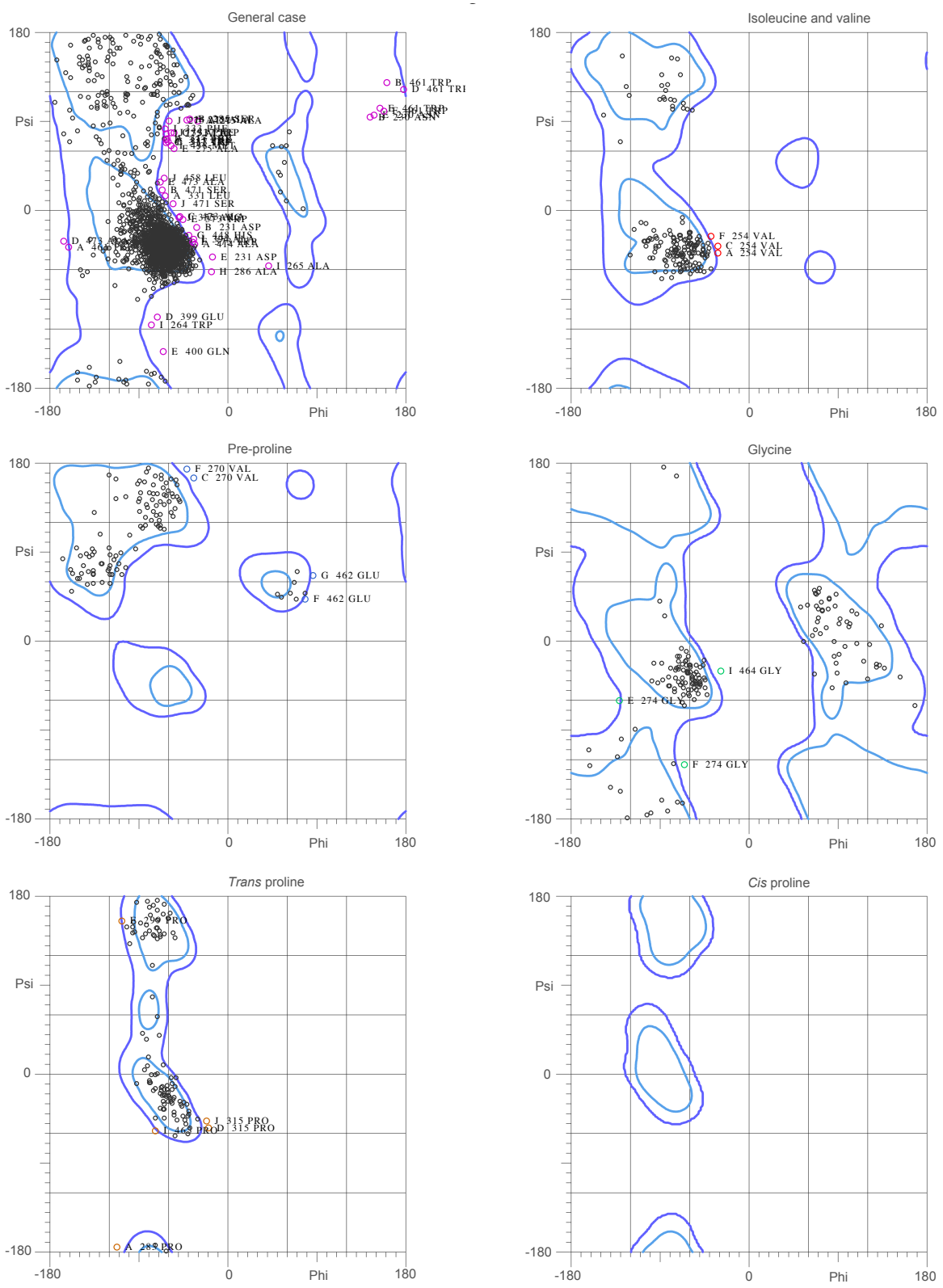


FIGURE A.21: **TssA2^A Nt2-CTD Ramachandran plot.** Analysis of the phi and psi torsion angles for different residue classes. These include the general case, isoleucine and valine, pre-proline, glycine, trans proline and cis proline. Light blue contour levels denote regions for favoured Ramachandran angles, dark blue allowed angles, and outside these are outliers. Figure created using Chen et al., 2010.

the Ah-TssA Nt2 to a corresponding Ah-TssA CTD. The relative position of the different Nt2 domains in the asymmetric unit were compared to the CTDs to identity the most appropriate arrangement of the Ah-TssA Nt2 domains around the Ah-TssA CTD oligomer.

This involved inspection of the symmetry-related molecules in the crystal lattice (figure A.22b) to reveal the distribution of the Ah-TssA Nt2 dimers surrounding the Ah-TssA CTD oligomers. Due to the absence of between 7 and 14 residues within the Ah-TssA Nt2-CTD interdomain linkers in the final Ah-TssA Nt2-CTD model, an immediate relationship between the Ah-TssA Nt2 domains and Ah-TssA CTDs could not be determined. Therefore, the maximal distance these missing residues would allow, assuming they adopted an unstructured loop, would be between $\sim$ 25-50 Å depending on the chain, details of this are given in table A.1, and this can be used to match the domains. This produces a steric limitation to which Ah-TssA Nt2 domain can be attached to which Ah-TssA CTD. Systematic elimination of the symmetry-related Ah-TssA Nt2 molecules surrounding the Ah-TssA CTD N-termini identified the position for the five dimers surrounding the Ah-TssA CTD oligomer, an example of this is given in figure A.22c. This allowed for the full assignment of the Ah-TssA Nt2 domains to their corresponding Ah-TssA CTD, generating the newly defined biological asymmetric unit (figure A.22a (right)), (3.2 figure 7a). As shown in paper 3.2, figure 7a this produces an assembly whereby the Ah-TssA Nt2 domains can sample a broad range of space about the plane of the CTD oligomer due to the inherent flexibility of the interdomain linker. This highlights the mobility the Ah-TssA Nt2 domain has with regards to the position of the CTD ring.

TABLE A.1: Missing Ah-TssA Nt2-CTD interdomain residues

Nt2-CTD chain	**No. of missing residues**	**Maximal distance of a Nt2 domain**
A	13 (375-388)	45.5 Å (3.5 Å x 13)
B	14 (374-388)	49 Å (3.5 Å x 14)
C	7 (376-383)	24.5 Å (3.5 Å x 7)
D	13 (375-388)	45.5 Å (3.5 Å x 13)
E	14 (374-388)	49 Å (3.5 Å x 14)
F	14 (374-388)	49 Å (3.5 Å x 14)
G	14 (374-388)	49 Å (3.5 Å x 14)
H	12 (376-388)	42 Å (3.5 Å x 12)
I	14 (374-388)	49 Å (3.5 Å x 14)
J	14 (374-388)	49 Å (3.5 Å x 14)

On further observation of the crystal lattice, it is clear that the mobility of the Ah-TssA Nt2 domains was critical for the formation of a complete lattice within the Ah-TssA his$_6$.Nt2-CTD crystal. The crystal packing down the a and c axis (figure A.22b) shows that crystal contacts between Ah-TssA Nt2-CTD complexes are only maintained via contacts made by Ah-TssA Nt2 domains. In addition, the Ah-TssA Nt2 domain-mediated contacts do not follow the strict five-fold symmetry of the Ah-TssA CTD ring. Therefore the mobility of the Ah-TssA Nt2 domains breaks the five-fold symmetry of the whole assembly allowing it to pack. This would explain the lack of obvious symmetry observed in the self-rotation function discussed in paper 3.1.

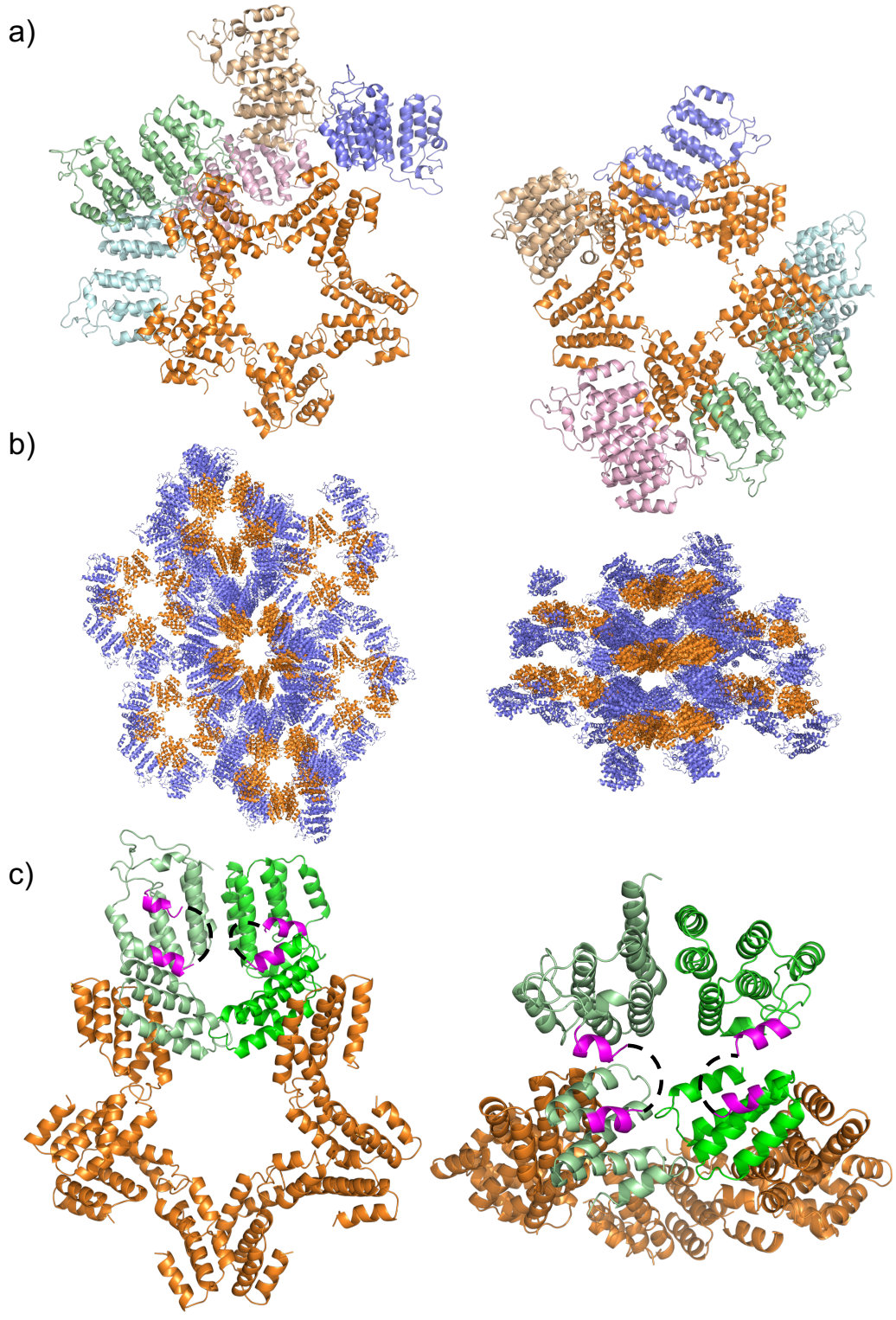


FIGURE A.22: **Ah-TssA Nt2-CTD structure assembly. a)** (left) Initial model of the Ah-TssA his$_6$.Nt2-CTD asymmetric content. The five Ah-TssA Nt2 dimers (green, blue, pink, brown and slate) in their initial placement with respect to the Ah-TssA CTD oligomer (orange). (right) The redefined asymmetric unit after the Ah-TssA Nt2 domains had been assigned to their respective CTD domain. **b)** View of the Ah-TssA his$_6$.Nt2-CTD crystal packing. Nt2 domains are coloured in slate, while the Ah-TssA CTD oligomers are shown in orange. (left) View parallel with the A axis of the unit cell. (right) View parallel with the C axis of the unit cell. **c)** Pairing of the Nt2 domains with their corresponding CTD domains. Paired Ah-TssA Nt2 and Ah-TssA CTDs shown in the same shade of green. The C-terminus and N-terminus of the paired Ah-TssA Nt2 and Ah-TssA CTDs respectively are shown in magenta. Black dashed lines indicate the missing interdomain linker.

Appendix B

Extended methodology - TssK

B.1 Construct design

DNA constructs of *tssK* generated from *B. cenocepacia* (Bc) H111 and *B. pseudomallei* (Bp) D286 for crystallisation studies were utilised throughout this project. This section provides some extended detail on their acquisition or how they were produced as part of this body of work.

B.1.1 Bc-TssK constructs provided by Mark Thomas

A Bc-TssK construct was provided by Mark Thomas, Department of Infection and Immunity, Sheffield Medical School. This was a full-length construct (codons 1-448) generated from *tssK* (I35_RS01705) from *B. cenocepacia* H111. Full details of the cloning can be found in 3.3, table 1.

B.1.2 Generation of the Bc-TssK$_{SN}$, Bc-TssK$_{H}$ and Bp-TssK construct

Two domain constructs of Bc-TssK (Bc-TssK$_{SN}$ and Bc-TssK$_{H}$) were generated as during the structure determination of the full-length Bc-TssK, a second crystal form was discovered which corresponded to a clipped variant of Bc-Tssk which had undergone adventitious proteolysis during crystal growth. This crystal form was identified to contain only the shoulder and neck domains of Bc-TssK. This guided the design of new constructs suitable for crystallisation. Further details of this observation can be found in 3.4. The first construct, Bc-TssK$_{SN}$, encoding residues 1-317 of *tssK* (I35_RS01705) from *B. cenocepacia* H111, while the second,

Bc-TssK$_H$, consists of codons 318-448 (figure B.1. In parallel, a DNA construct of the full-length *tssK* (T6SS gene cluster 1, BPSL3110) from *B. pseudomallei* D286 Bp-TssK (residues 1-448), was produced (figure B.3).

Primers for the constructs were generated using the NEB primer design tool suitable for use with the NEB Q5 polymerase kit. The primers for Bc-TssK$_{SN}$, Bc-TssK$_H$ and Bp-TssK construct used are shown in 3.4, table 4. The PCR reaction mixture was made up as described in table 2.6 using 2 μL at 100 ng/μL pACYduet-1/Bc-TssK (1-448) plasmid or *B. pseudomallei* D286 gDNA as the DNA template. The PCR reaction was carried out in the thermocycler as in table 2.7, and the reaction was run in duplicate to improve the chances of success. The amplified DNA products of the reaction were subsequently analysed on a 1 % agarose gel (figure B.2).

The Bc-TssK$_{SN}$, Bc-TssK$_H$ and Bp-TssK PCR products were then purified as described in 2.1.7. The purified PCR products, in addition to 30 μL of 100 ng/μL purified pET21a vector, were then digested separately, using restriction enzymes *Xho*1 and *Nde*1 according to 2.2.3. This allowed for the incorporation of a C-terminal 6xhis affinity tag present in the pET21a vector (figure 2.1) to the Bc-TssK$_{SN}$ and Bc-TssK$_H$ protein products. Following restriction digest, 35 μL at $\sim$ 40 ng/μL, 2x 35 μL at $\sim$ 40 ng/μL, 2x 35 μL at $\sim$ 35 ng/μL and 35 μL at $\sim$ 50 ng/μL were recovered for the Bc-TssK$_{SN}$, Bc-TssK$_H$, Bp-TssK and pET21a restriction products, respectively. Ligation reactions were set up to generate the final constructs as in 2.2.4.

The successful incorporation of the Bc-TssK$_{SN}$, Bc-TssK$_H$ and Bp-TssK gene fragments into their respective pET21a vector was initially determined through colony PCR, set up as described in 2.1.10. The successful colonies were then propagated, their plasmids liberated through miniprep, and sent for sequencing. No errors were found for the two Bc-TssK clones, but sequencing of the Bp-TssK clone showed that it had been truncated after residue 328 as a result of an Xho1 site in the *tssK* BPSL3110 gene that had not been identified (figure B.3). Since this corresponded to a construct representing the complete Bp-TssK shoulder and neck domains (Bp-TssK$_{SN}$) the construct was used to express this particular domain combination.

FIGURE B.1: **Sequence details of the Bc-TssK domain constructs.** The constructs generated were used for protein expression for subsequent crystallisation trials. The protein sequence for each construct is indicated in different colours, according to the included key. The full-length Bc-TssK construct has been omitted as it spans the full gene.

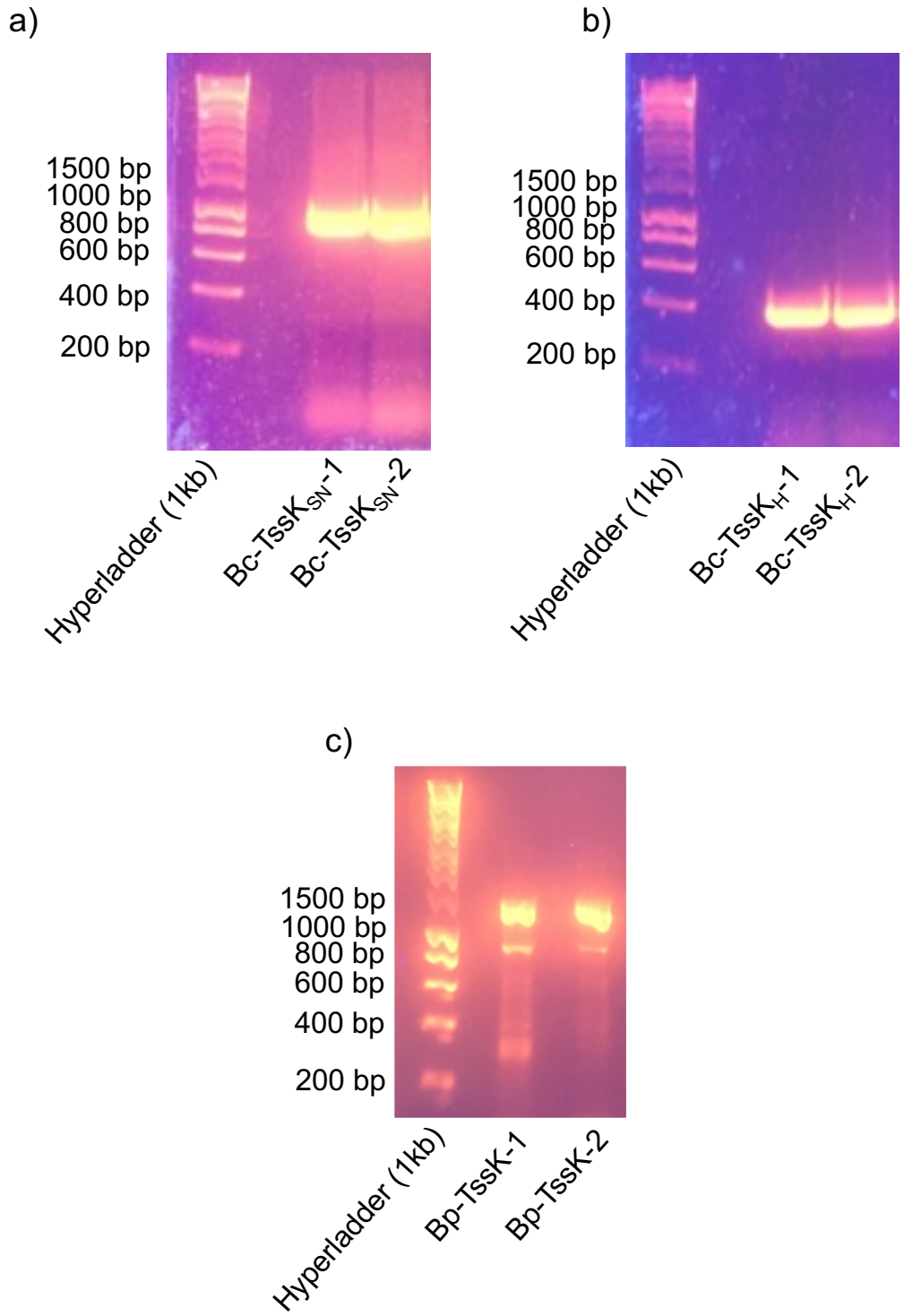


FIGURE B.2: **Agarose gel of the PCR reactions for Bc-TssK domain and Bp-TssK constructs.** Each sample lane consists of 4 μL DNA product, mixed with 1 μL loading buffer. **a)** Amplified DNA from *B. cenocepacia tssK* I35_RS01705 (codons 1-317), marked lanes of Bc-TssK$_{SN}$ indicates successful amplification at $\sim$ 950 bp, the expected size for the gene fragment. **b)** Amplified DNA from *B. cenocepacia tssK* I35_RS01705 (codons 318-448), marked lanes of Bc-TssK$_{H}$ indicate successful amplification at $\sim$ 400 bp, the expected size for the gene fragment. **c)** Amplified DNA from *B. pseudomallei tssK* BPSL3110 (codons 1-448), marked lanes of Bp-TssK indicates successful amplification at $\sim$ 1200 bp, the expected size for the gene.

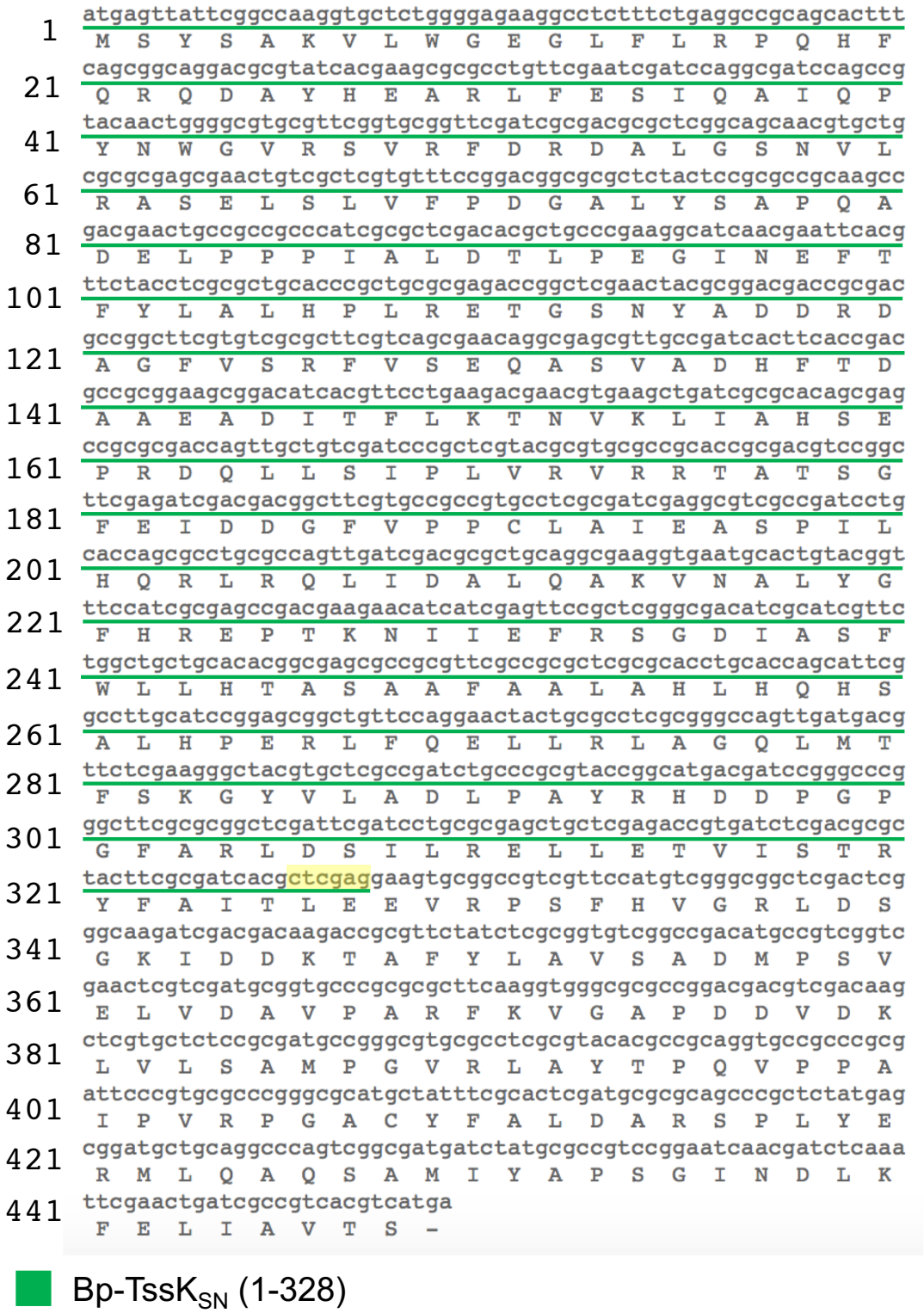


FIGURE B.3: **BPSL3110 (Bp-TssK) DNA and translated protein sequence.** The DNA sequence of BPSL3110 (Bp-TssK) indicates there is an Xho1 site within the gene (yellow) which was missed during construct design. However, this generated a shoulder and neck domain construct (codons 1-328) comparable to Bc-TssK$_{SN}$. The sequence of the Bp-TssK$_{SN}$ construct is indicated in green.

B.2 Protein production of Bc-TssK and Bp-TssK domain constructs

While the Bc-TssK provided by Mark Thomas had an established expression protocol (details shown in paper 3.3), expression of the Bc-TssK$_{SN}$.his$_6$, Bc-TssK$_H$.his$_6$ and Bp-TssK$_{SN}$.his$_6$ domain constructs were unknown, and therefore initial expression trials were carried out.

B.2.1 Test expression of Bc-TssK$_{SN}$.his$_6$, Bc-TssK$_H$.his$_6$ and Bp-TssK$_{SN}$.his$_6$

To test if the Bc-TssK$_{SN}$.his$_6$, Bc-TssK$_H$.his$_6$ and Bp-TssK$_{SN}$.his$_6$ produced soluble protein, and in sufficient quantity when expressed in *E. coli*, small-scale test expressions were set up as described in 2.3.1 using BHI media supplemented with ampicillin (100 μg/μL). The result of the test expressions were successful in producing soluble protein at the expected molecular weight for the Bc-TssK$_{SN}$.his$_6$, Bc-TssK$_H$.his$_6$ and Bp-TssK$_{SN}$.his$_6$ constructs (35.6, 14.5 and 36.7 kDa, respectively). The associated SDS-PAGE gel of the test expression can be seen in figure B.4.

B.2.2 Large-scale protein production of Bc-TssK and Bp-TssK constructs for crystallisation

To produce sufficient protein for downstream structural studies large-scale protein expression of the Bc-TssK$_{SN}$.his$_6$, TssK$_H$.his$_6$ and Bp-TssK$_{SN}$.his$_6$, was carried out. To achieve this 1 L of BHI media supplemented with 100 μg/μL ampicillin was used as the expression media. Large-scale over expression was carried out as in 2.3.2 with 30 °C used as the incubation temperature after induction with 1 mM IPTG. Due to the heavy over expression of the constructs, incubation of the cultures after induction was left for 2-3 hours. The resultant over expression profiles of the Bc-TssK$_{SN}$.his$_6$, Bc-TssK$_H$.his$_6$ and Bp-TssK$_{SN}$.his$_6$ domain constructs were similar to the test expressions and so are not shown.

B.3 Protein purification of the Bc-TssK and Bp-TssK constructs

This section will detail the purification process for the TssK constructs used in crystallisation trials as part of this project.

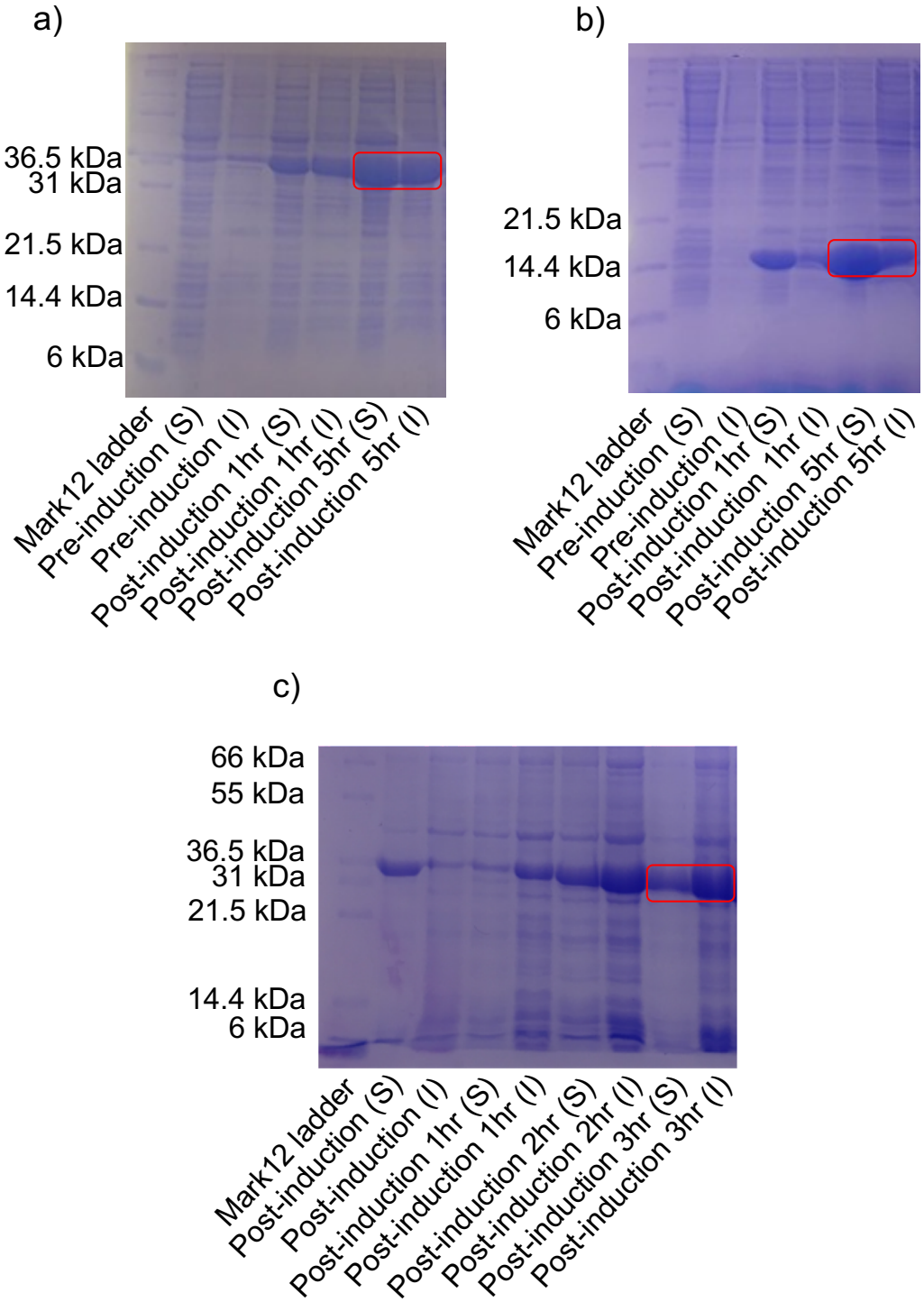


FIGURE B.4: **Test expressions of the Bc-TssK$_{SN}$.his$_6$, Bc-TssK$_H$.his$_6$ and Bp-TssK$_{SNH}$.his$_6$ constructs.** 6/12 % SDS-PAGE gels of the pre-induction and post-induction samples. **a)** Test expression of the Bc-TssK$_{SN}$.his$_6$ indicates the production of suitable quantities of soluble protein at the expected molecular weight (35.6 kDa). **b)** Test expression of the Bc-TssK$_H$.his$_6$ indicates the production of suitable quantities of soluble protein at the expected molecular weight (14.5 kDa). **c)** Test expression of the Bp-TssK$_{SN}$.his$_6$ indicates the production of suitable quantities of soluble protein at the expected molecular weight (36.7 kDa). Bands of interest are indicated in red boxes.

B.3.1 Protein purification of Bc-TssK$_{SN}$.his$_6$

From 1 l of BHI media, 3 g of cell paste was recovered from the over expression of Bc-TssK$_{SN}$.his$_6$. The cells were re-suspended and subsequently lysed as described in section 2.3.4, which, after centrifugation and separating from the cell debris, yielded 16 mL CFE at 5.3 mg/mL. The sample was applied to a Ni-NTA column (figure B.5a) as described in 2.4.1. Fractions 16-18, which equated to volumes around the peak found at 53 mL, were taken and combined to form the final sample. Analysis of the purification can be seen by the SDS-PAGE in figure B.5b. The final sample was concentrated to 30 mg/mL in 300 μL of 50 mM Tris (pH 8.0), before being diluted down to 10 mg/mL for subsequent crystallisation trials.

B.3.2 Protein purification of Bc-TssK$_{H}$.his$_6$

From 1 l of BHI media, 3 g of cell paste was recovered from the over expression of Bc-TssK$_{H}$.his$_6$. The cells were re-suspended and subsequently lysed as described in section 2.3.4, which, after centrifugation and separating from the cell debris, yielded 17 mL CFE at 4.9 mg/mL. The sample was applied to a Ni-NTA column (figure B.6a) as described in 2.4.1. Fractions 13-16, which equated to volumes around the peak found at 47 mL, were taken and combined to form the final sample. Analysis of the purification can be seen by the SDS-PAGE in figure B.6b. The final sample was concentrated to 28 mg/mL in 280 μL of 50 mM Tris (pH 8.0), before being diluted down to 10 mg/mL for subsequent crystallisation trials.

B.3.3 Protein purification of Bp-TssK$_{SN}$.his$_6$

From 1 l of BHI media, 4.5 g of *E. coli* cell paste was recovered. This pellet was subsequently re-suspended and lysed as described in 2.3.4 before being centrifuged to separate the CFE from the cell debris. This produced 19 mL at 8 mg/mL Bp-TssK$_{SN}$.his$_6$ expressing CFE. The sample was initially applied to a DEAE FF 5mL (GE healthcare) anion exchange column (figure B.7a) and eluted as described in 2.4.2. Fractions 16-20 inclusive, surrounding the peak at $\sim$ 63 mL, were pooled to generate a 12.5 mL sample, and 4 M Ammonium Sulphate was then added to a final concentration of 1.6 M, to precipitate Bp-TssK$_{SN}$.his$_6$ as described in 2.4.3. The resultant sample was concentrated to 1.3 mL before being applied to a gel filtration

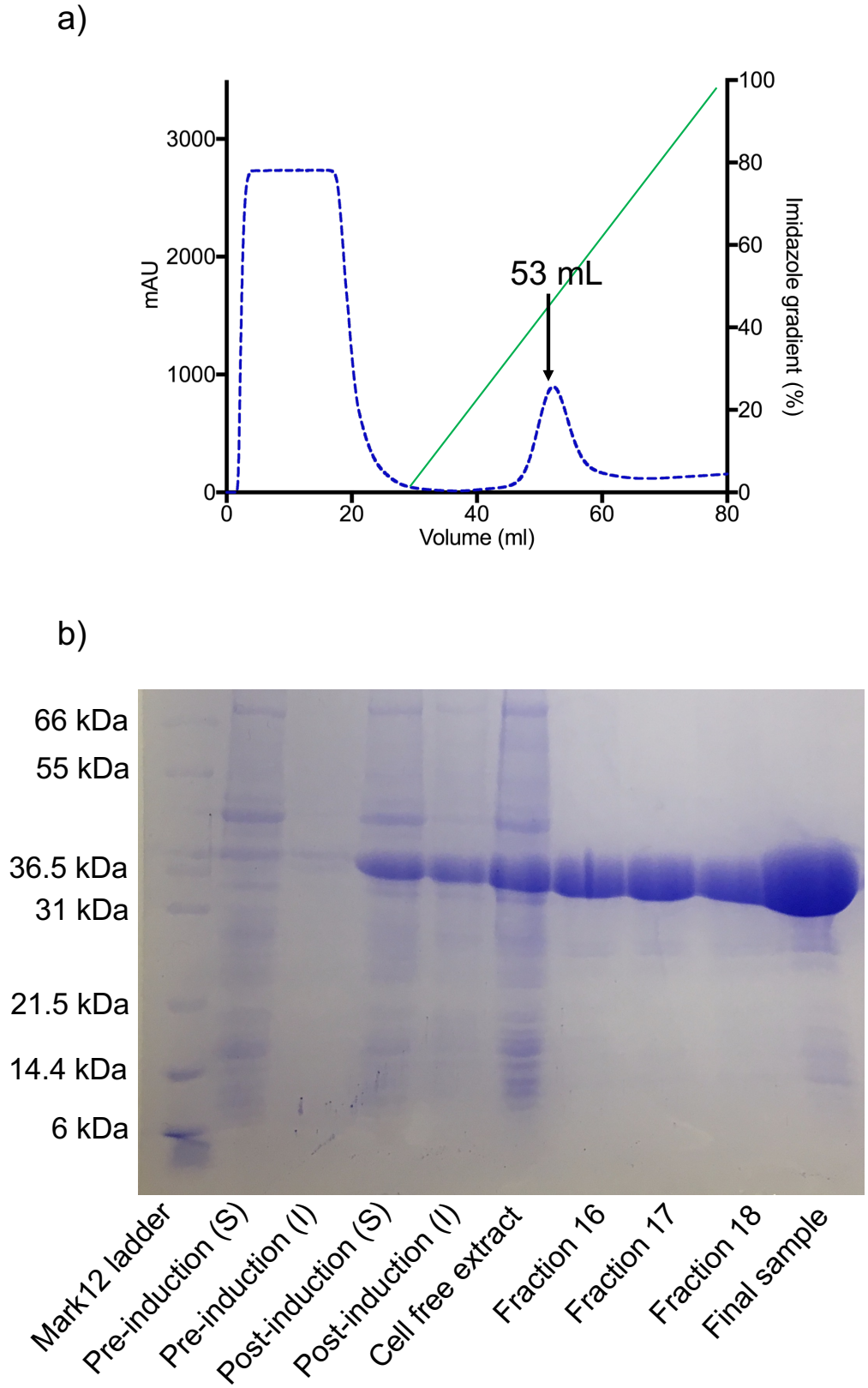


FIGURE B.5: **Protein purification of Protein purification of Bc-TssK$_{SN}$.his$_6$. a)** Ni-NTA chromatography purification trace of Bc-TssK$_{SN}$.his$_6$. The green line indicates the percentage of imidazole being applied to the column. The trace indicates a peak at $\sim$ 53 mL indicated by the arrow, which is the elution peak corresponding to Bc-TssK$_{SN}$.his$_6$. **b)** 4/12% gradient SDS-PAGE gel of the purification of Bc-TssK$_{SN}$.his$_6$. The SDS-PAGE gel shows successful purification indicating that the final sample of Bc-TssK$_{SN}$.his$_6$ are at the expected molecular weight ($\sim$ 35.6 kDa), and is > 95 % pure.

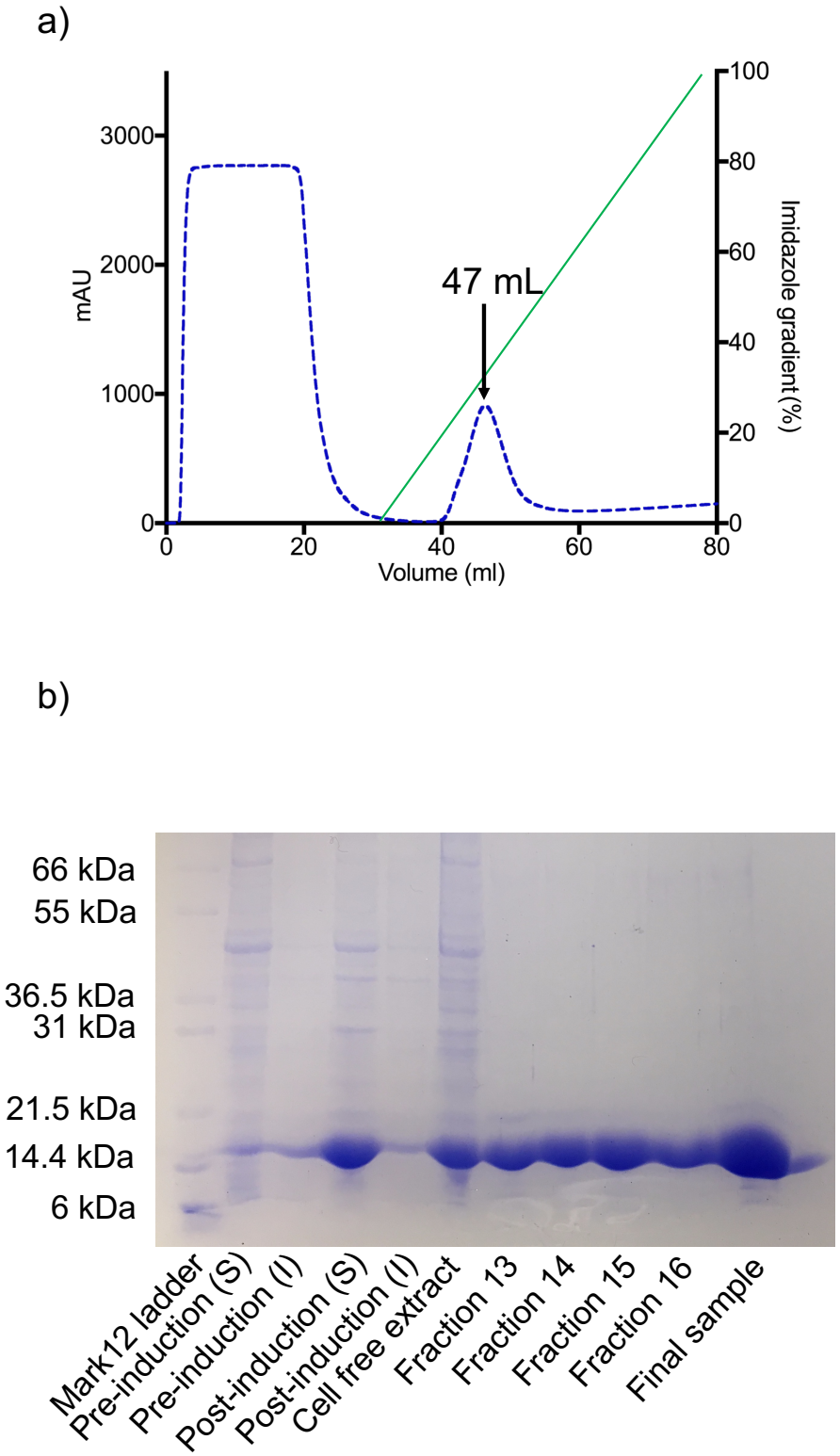


FIGURE B.6: **Protein purification of Bc-TssK$_H$.his$_6$. a)** Ni-NTA chromatography purification trace of Bc-TssK$_H$.his$_6$. The green line indicates the percentage of imidazole being applied to the column. The trace indicates a peak at $\sim$ 47 mL indicated by the arrow, which is the elution peak corresponding to Bc-TssK$_H$.his$_6$. **b)** 4/12% gradient SDS-PAGE gel of the purification of Bc-TssK$_H$.his$_6$. The SDS-PAGE gel shows successful purification indicating that the final sample of Bc-TssK$_H$.his$_6$ are at the expected molecular weight ($\sim$ 14.5 kDa), and is > 95 % pure.

column (figure B.7a) as described in 2.4.4. Fractions 10-12 inclusive, surrounding the peak seen at ~62 mL, were taken and concentrated to form the final sample. SDS-PAGE analysis of the purification can be seen in figure B.7b. The final sample of native Bp-TssK$_{SN}$.his$_6$ was 37.5 mg/mL in a volume of 250 μL 50 mM Tris (pH 8.0). This was diluted down to 10 mg/mL for use in crystallisation trials.

B.4 Crystallisation trials of Bc-TssK and Bp-TssK domains.

As part of this project purified protein samples of Bc-TssK, SeMet-Bc-TssK, Bc-TssK$_{SN}$.his$_6$, TssK$_H$.his$_6$ and Bp-TssK$_{SN}$.his$_6$ were concentrated as described in 2.5.1 and subjected to crystallisation trials. The details of the crystallisation for Bc-TssK and SeMet-Bc-TssK are described in 3.4. Details of the crystallisation of Bc-TssK$_{SN}$.his$_6$, TssK$_H$.his$_6$ and Bp-TssK$_{SN}$.his$_6$ are described below.

B.4.1 Crystallisation of Bc-TssK$_{SN}$.his$_6$

Crystallisation trials of Bc-TssK$_{SN}$.his$_6$ were carried out as in 2.5.2, using protein at 10 mg/mL. The plates were kept at 17 °C, and after four days of incubation, crystals had appeared in multiple conditions, with a variety of crystal forms. The crystallisation conditions and crystal morphologies are listed in table B.1 and an example is shown in B.8a. All protein crystals were of suitable size for subsequent data collection and cryo-protected according to 2.5.6.

B.4.2 Crystallisation of Bc-TssK$_H$.his$_6$

Initial crystallisation trials of Bc-TssK$_H$.his$_6$ were carried out as in 2.5.2 at 10 mg/mL. After a month of incubation crystals formed in a single condition (0.1 M MES pH 6.0 and 4.0 M NaCl (pH clear B9)). The crystals initially grew as thin rods, which after optimisation, grew as a combination of thick and thin rods emanating from single nucleation sites (figure B.8b). The crystals were large enough for subsequent data collection.

B.4.3 Crystallisation of Bp-TssK$_{SN}$.his$_6$

Initial crystallisation trials of Bp-TssK$_{SN}$.his$_6$ were carried out as in 2.5.2 at 10 mg/mL, after a month of incubation this resulted in a single condition that generated crystals. The crystals had grown in (Proplex G6) 0.1 M Na acetate pH 5.0, and 2.0 M Na formate with a needle-like

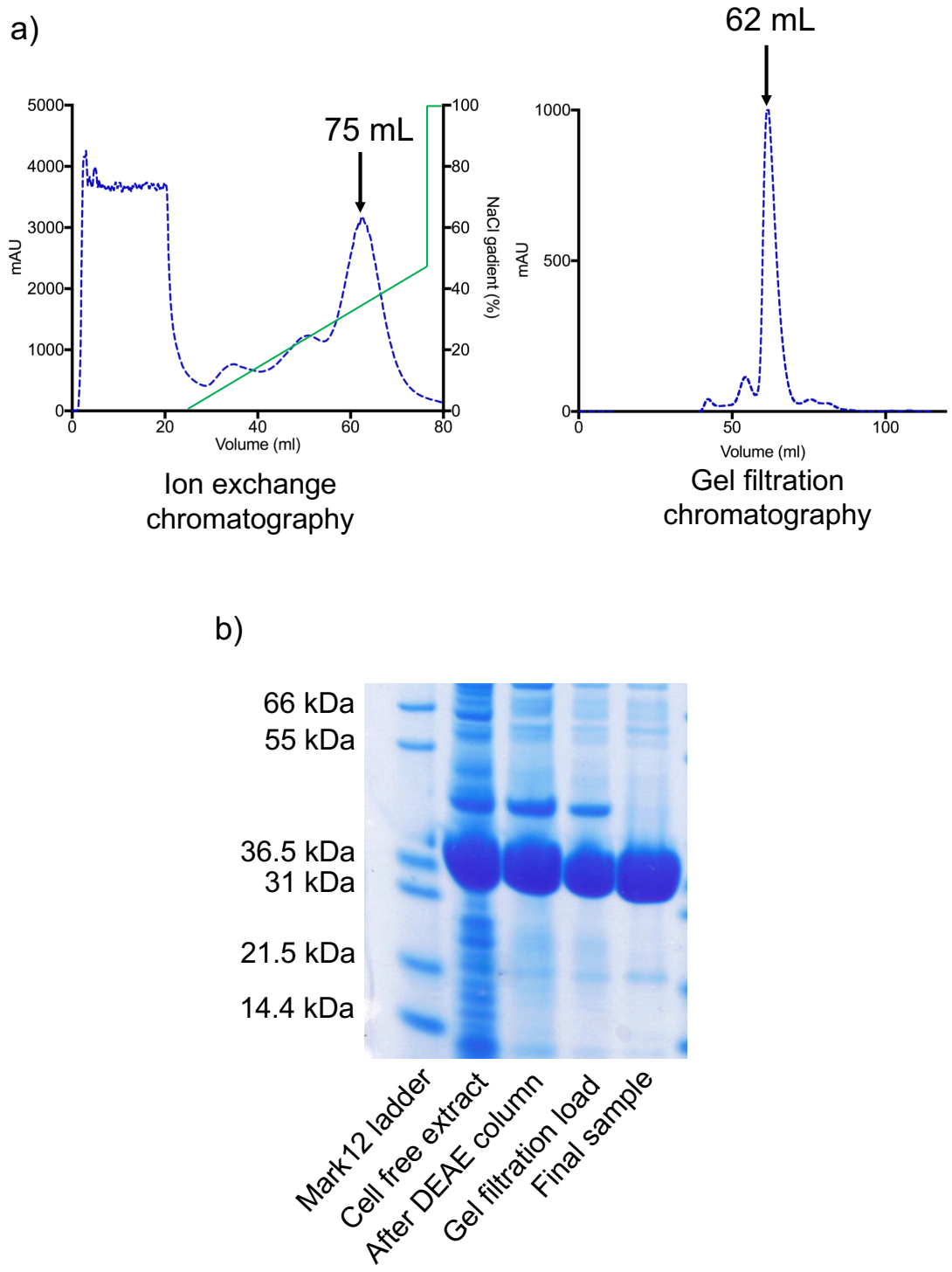


FIGURE B.7: **Protein purification of Bp-TssK$_{SN}$.his$_6$. a)** Ion exchange purification trace of Bc-TssK, where the chromatogram indicates a peak at 63 mL, which corresponds to the point of elution for Bc-TssK.**b)** Analytical gel filtration purification trace, where the chromatogram indicates a peak at $\sim$ 62 mL, which corresponds with an apparent molecular weight of $\sim$ 140 kDa. **b)** The associated SDS-PAGE gel shows successful purification indicating that the final sample of Bp-TssK$_{SN}$.his$_6$ are at the expected molecular weight ($\sim$ 36.7 kDa), and is > 95 % pure.

TABLE B.1: Crystallisation conditions of Bc-TssK$_{SN}$

Well condition	**Crystal morphology**
PACT: (E8,G9,G11,H8,H9,H11,H12)	Cubic
JCSG: (B8,E8,F7)	Cubic
Morpheus: (B3,C1,C5,G3)	Cubic
$AmSO_4$ E6	Cubic
Proplex: (A6,A12,B4,C6,D3,E3,F10,G7,G8)	Cubic
Ph Clear: (E3,E4,E5,E6)	Cubic
PACT: E10	Triangular
Proplex: E9	Triangular
Ph Clear: B12	Triangular
Morpheus: G5	Bi pyrimids
$AmSO_4$: (E4,E5)	Bi pyrimids
PACT: (E11,E12)	Rods
$AmSO_4$: E3	Rods
Proplex: F9	Rods

morphology emanating from a single nucleation site (figure B.8c). While thin, the crystals were large enough for subsequent data collection.

B.5 X-ray data collection of Bc-TssK and Bp-TssK constructs

B.5.1 Data collection and analysis of Bc-TssK

Diffraction quality of the Bc-TssK protein crystals with and without selenomethionine incorporation were initially tested as described in 2.6. As discussed in 3.3, a preliminary native and derivative dataset for crystals of Bc-TssK were collected and are presented in 3.3, table 3. Subsequently, higher-resolution datasets of the selenomethionine derivative and native crystals were collected. Three datasets, covering 0-200 ^{o} (2000 images), were collected from the same selenomethionine derivative crystal from three locations. Initial processing of the datasets, independently, identified significant radiation damage after 120 ^{o} (image 1200), as determined by a sharp inflection in the R_{merge} value per batch. Therefore these batches were removed, and the remaining data were merged to improve the anomalous signal. Details of the final native and selenomethionine derivative datasets, along with the corresponding processing statistics are described in 3.4, table 1.

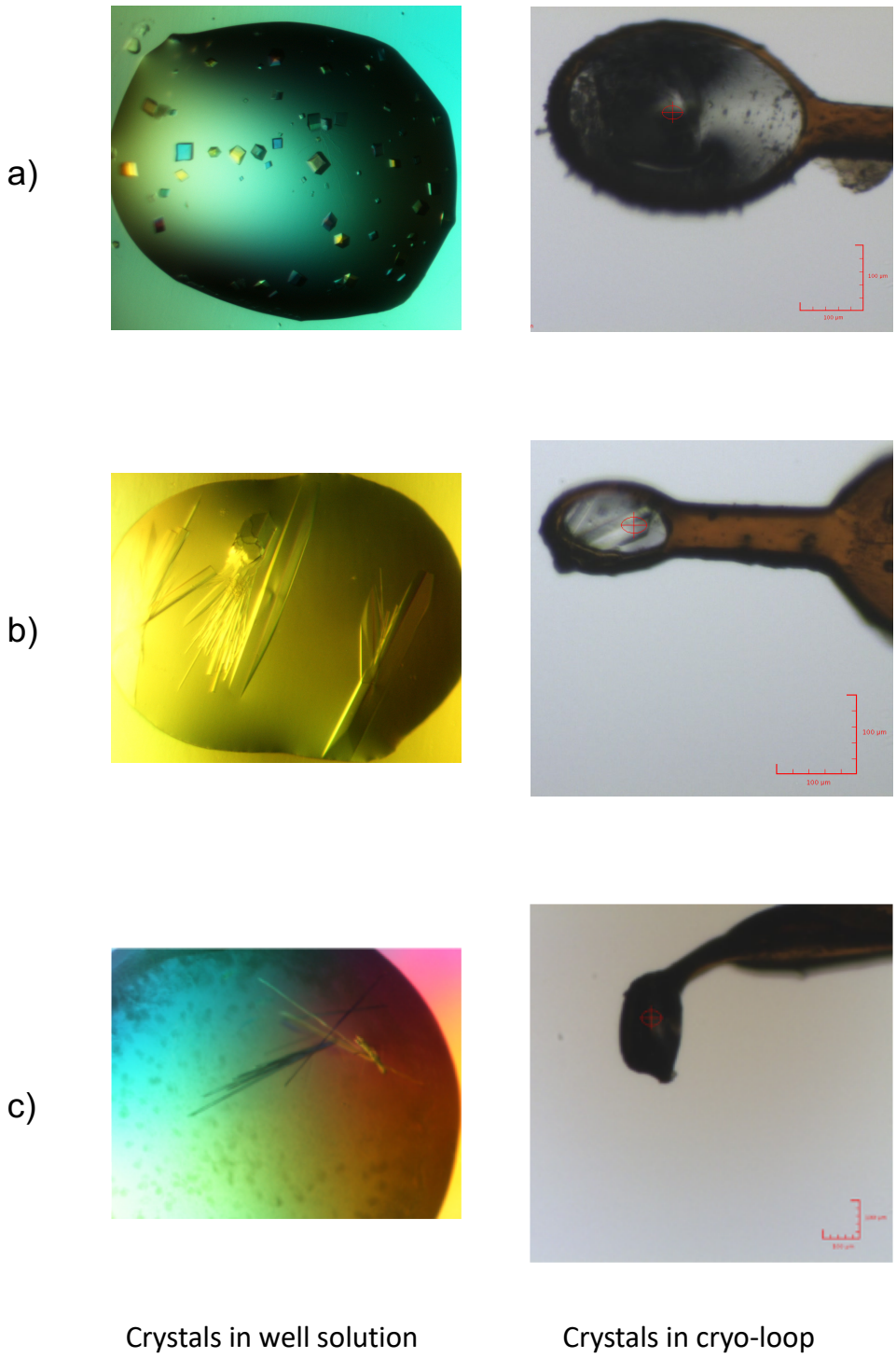


FIGURE B.8: **Crystallisation of Bc-TssK$_{SN}$.his$_6$, Bc-TssK$_H$.his$_6$ and Bp-TssK$_{SN}$.his$_6$.** (Left) Images of the crystals taken while they are still growing, (Right) images of the crystals once they had been mounted onto the beamline. **a)** Crystals of Bc-TssK$_{SN}$.his$_6$ grown in 0.2M Magnesium chloride hexahydrate, 0.1M Tris pH 7.0, 10 % w/v PEG 8000. The crystals have a morphology of cubes. **b)** Crystals of Bc-TssK$_H$.his$_6$ grown in 0.1 M MES pH 6.0 and 4.0 M NaCl. The crystals appear to have grown as rods. **c)** Crystals of Bp-TssK$_{SN}$.his$_6$ were grown in 0.1 M Na acetate pH 5.0, and 2.0 M Na formate. The crystals appear to have grown as needles.

The cell parameters, gave an estimate for two or three molecules within the asymmetric unit, as determined through the Matthews calculation (figure B.9). Analysis of a self-rotation function on the Bc-TssK native data, indicate the presence of a non-crystallographic two-fold axis (3.3, figure 3), suggesting that there are two molecules of Bc-TssK within the asymmetric unit.

B.5.2 Data collection and analysis of Bc-TssK$_H$.his$_6$

Diffraction quality of the Bc-TssK$_H$.his$_6$ protein crystals were initially tested as described in 2.6. Initial tests indicated weak diffraction with reflections only observed to $\sim$ 6.5 Å. Nevertheless, a native dataset was collected on DLS MX beamline I04 using an X-ray beam of wavelength 0.9795 Å, a start phi of 0 º, 0 º delta, 0.1 º oscillation, an exposure time of 0.05 s exposure and transmission of 100 %. 3600 images (360 º) were collected and processed using the Xia2 automatic pipelines (Winter, 2010) (see 2.6.3 for details). Full details of the processing statistics for the data set can be seen in B.2. An initial estimation using the Matthews calculation suggests the most probable number of Bc-TssK$_H$.his$_6$ molecules within the asymmetric unit to be 14, but with considerable uncertainty (B.10).

TABLE B.2: Data collection and processing statistics of Bc-TssK$_H$.

Data collection	**Native**
Wavelength (Å)	0.9795
Resolution range (Å)	157.1-6.47 (6.58-6.47)
Space group	$P6_2$ 2 2
Unit cell (a,b,c) (Å)	314.3, 314.3, 70.5
(α,β,γ) (º)	90.0, 90.0, 120.0
Total reflections	160419 (12205)
Unique reflections	4436 (312)
Completeness (%)	100.0 (100.0)
Multiplicity	36.2 (39.1)
Mean I/σ	7.8 (3.7)
Wilson B factor (Å^2)	378
R_{merge}	0.394 (1.360)
R_{pim} (I)	0.066 (0.218)
R_{meas}	0.400 (1.378)
CC 1/2	0.971 (0.210)

Data for the higher resolution shell shown in parentheses.

a)

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0072	0.0298	6.13	79.92	49900.00
2	0.3693	0.4887	3.06	59.85	99800.00
3	0.6208	0.4796	2.04	39.77	149700.00
4	0.0027	0.0019	1.53	19.69	199600.00

b)

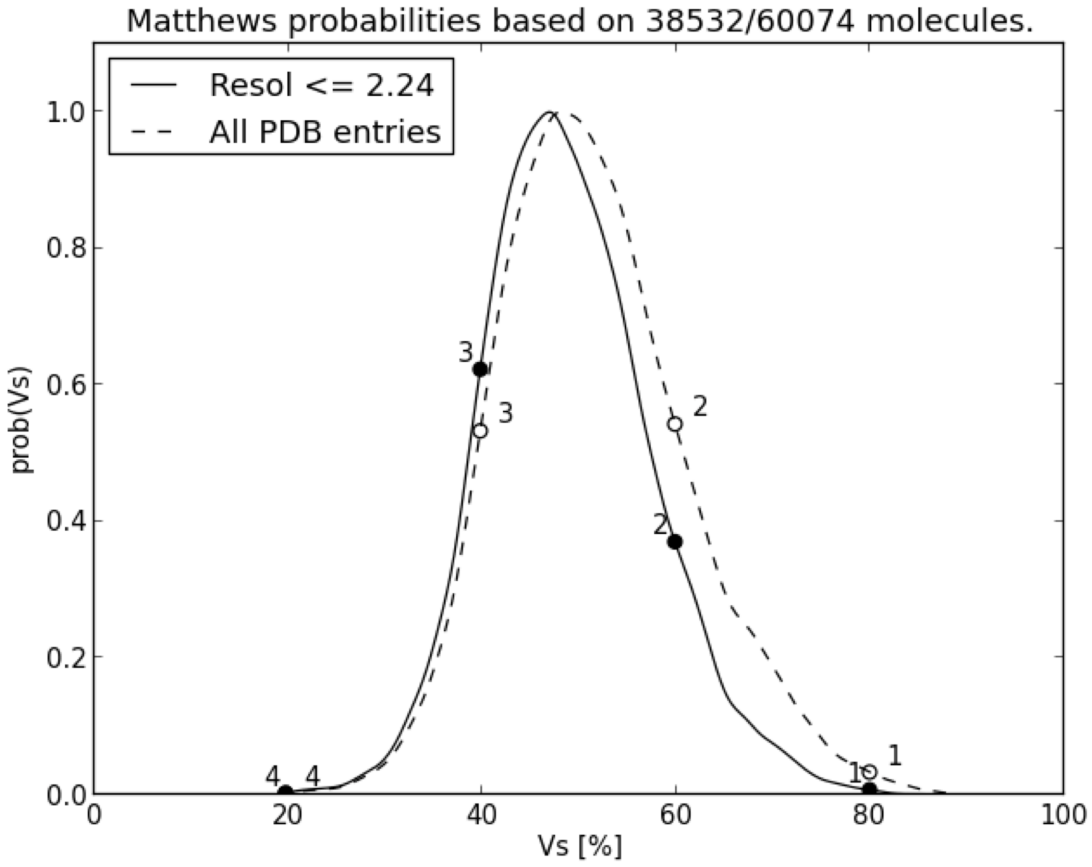


FIGURE B.9: **Bc-TssK Matthews calculation. a)** Table of the possible number of molecules within the asymmetric unit with their respective probability and corresponding percentage solvent content. This would suggest either two or three molecules in the asymmetric units, with similar probability. **b)** A graphical representation of the data presented in a. Figure created using http://www.ruppweb.org/mattprob/default.html (Matthews, 1968)

a)

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0000	0.0000	32.30	96.19	15548.00
2	0.0000	0.0000	16.15	92.38	31096.00
3	0.0002	0.0002	10.77	88.57	46644.00
4	0.0016	0.0016	8.07	84.77	62192.00
5	0.0043	0.0043	6.46	80.96	77740.00
6	0.0085	0.0085	5.38	77.15	93288.00
7	0.0183	0.0183	4.61	73.34	108836.00
8	0.0322	0.0322	4.04	69.53	124384.00
9	0.0447	0.0447	3.59	65.72	139932.00
10	0.0719	0.0719	3.23	61.92	155480.00
11	0.1031	0.1031	2.94	58.11	171028.00
12	0.1400	0.1400	2.69	54.30	186576.00
13	0.1594	0.1594	2.48	50.49	202124.00
14	0.1595	0.1595	2.31	46.68	217672.00
15	0.1305	0.1305	2.15	42.87	233220.00
16	0.0761	0.0761	2.02	39.06	248768.00
17	0.0319	0.0319	1.90	35.26	264316.00
18	0.0119	0.0119	1.79	31.45	279864.00
19	0.0041	0.0041	1.70	27.64	295412.00
20	0.0013	0.0013	1.61	23.83	310960.00
21	0.0005	0.0005	1.54	20.02	326508.00
22	0.0000	0.0000	1.47	16.21	342056.00
23	0.0000	0.0000	1.40	12.40	357604.00

b)

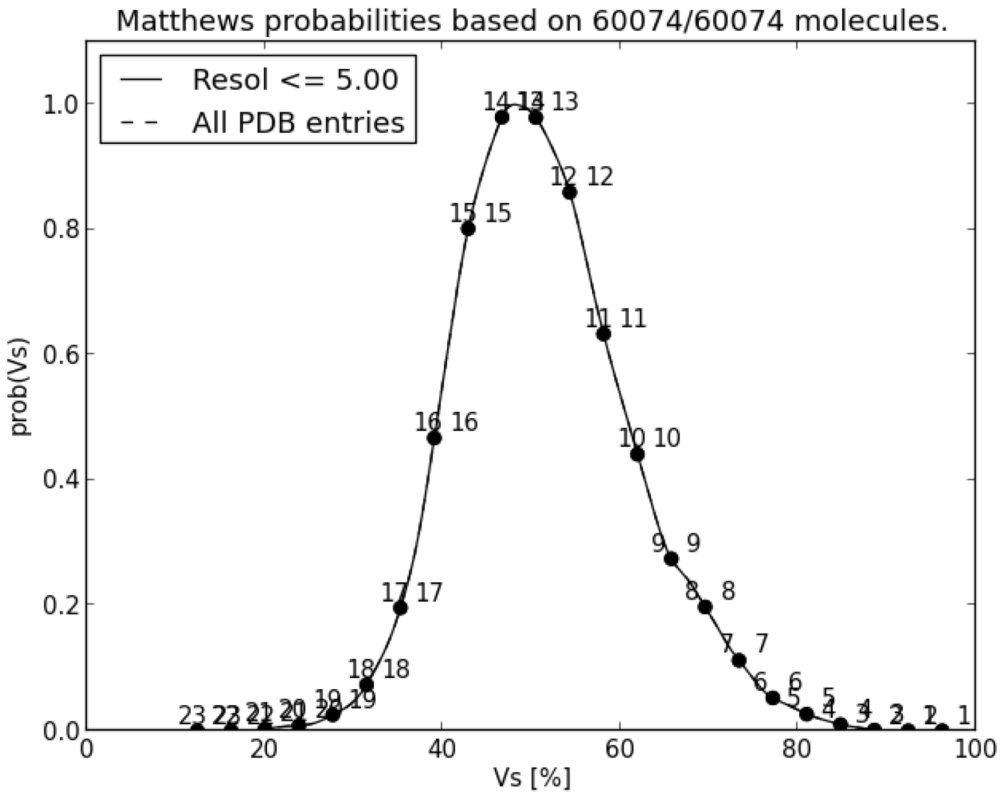


FIGURE B.10: **Bc-TssK$_H$.his$_6$ Matthews calculation. a)** Table of the possible number of molecules within the asymmetric unit with their respective probability and corresponding percentage solvent content. This would suggest about 14 molecules in the asymmetric unit as indicated by the red arrow. **b)** A graphical representation of the data presented in a. Figure created using http://www.ruppweb.org/mattprob/default.html (Matthews, 1968)

B.5.3 Data collection and analysis of Bc-TssK$_{SN}$.his$_6$

Diffraction quality of the Bc-TssK$_{SN}$.his$_6$ protein crystals were initially tested as described in 2.6. Out of the multiple conditions that produced diffraction quality crystals, the best data were collected from crystals grown in 0.2 M Magnesium chloride hexahydrate, 0.1 M Tris pH 7.0, 10 % w/v PEG 8000 (JCSG B8). A native data set was collected on DLS MX beamline I04-1 using an X-ray beam of wavelength 0.9282 Å, a start phi of 0 º, 0 º delta, 0.1 º oscillation, an exposure time of 0.05 s exposure and transmission of 100 %. 2000 images were collected and processed using the Xia2 automatic pipelines (Winter, 2010), see 2.6.3 for details. Full details of the processing statistics of the dataset can be found in 3.4, table 1. Initial estimation predicts a single molecule of Bc-TssK$_{SN}$.his$_6$ within the asymmetric unit as determined through the Matthews calculation (figure B.11).

B.5.4 Data collection and analysis of Bp-TssK$_{SN}$.his$_6$

Diffraction quality of the Bp-TssK$_{SN}$.his$_6$ protein crystal was initially tested as described in 2.6. A subsequent native data set was collected on DLS MX beamline I04 using an X-ray beam of wavelength 0.9795 Å, a start phi of 0 º, 0 º delta, 0.1 º oscillation, an exposure time of 0.05 s exposure and transmission of 100 %. 2000 images were collected and processed using the Xia2 automatic pipelines (Winter, 2010), see 2.6.3 for details. Full processing statistics of the dataset can be found in 3.4, table 1. Initial estimation predicts a single molecule of Bp-TssK$_{SN}$.his$_6$ within the asymmetric unit as determined through the Matthews calculation (figure B.12).

B.6 Structure determination, model building and validation of Bc-TssK

B.6.1 Structure determination, and initial model building of Bc-TssK

The Bc-TssK selenomethionine dataset, as shown in 3.4, table 1, has an anomalous mid-slope of 1.142. Attempts to calculate initial phases utilised the SHELX and CRANK2 software pipelines (Sheldrick, 2008; Skubák and Pannu, 2013). Initial analysis with SHELXC identified anomalous signal > 0.8 d''/sig for the full extent of the data. Subsequently, SHELXD

a)

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	1.0000	1.0000	2.84	56.68	37143.00
2	0.0000	0.0000	1.42	13.36	74286.00

b)

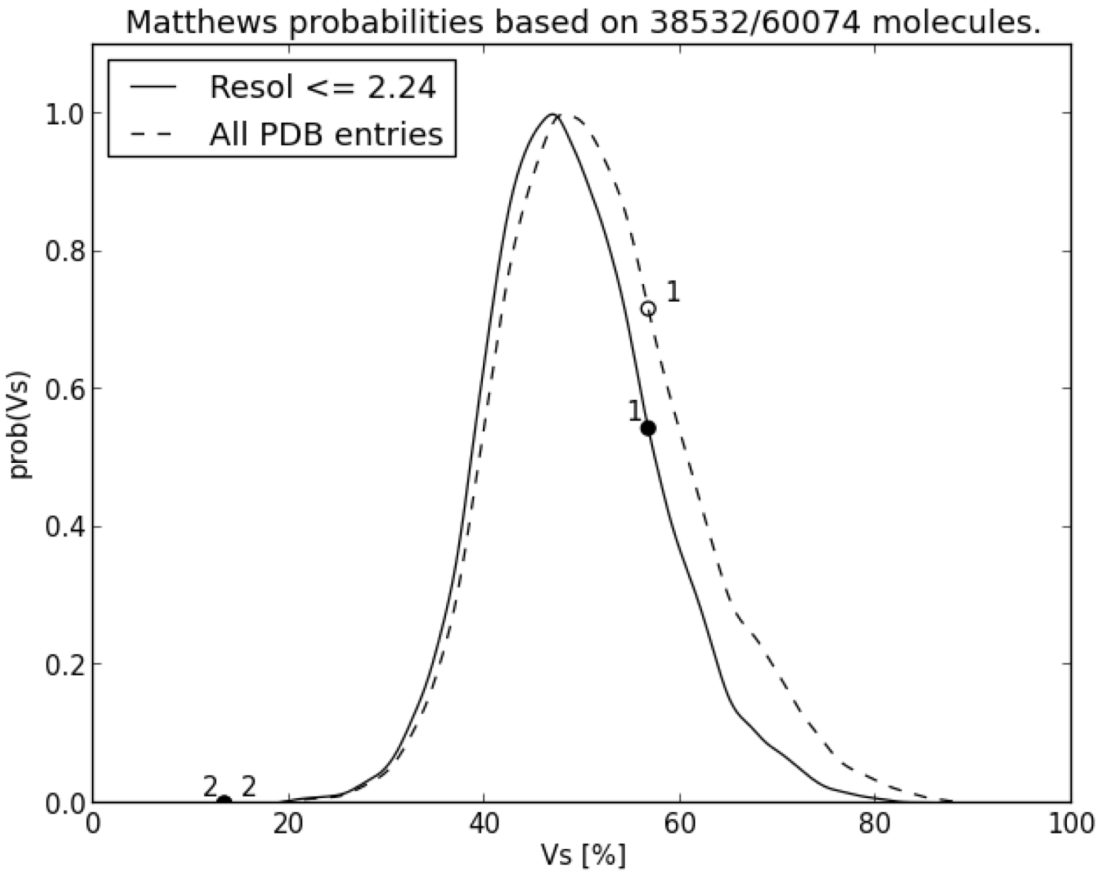


FIGURE B.11: **Bc-TssK$_{SN}$.his$_6$ Matthews calculation. a)** Table of the possible number of molecules within the asymmetric unit with their respective probability and corresponding percentage solvent content. This would Indicate a single molecule of Bc-TssK$_{SN}$.his$_6$ within the asymmetric unit. **b)** A graphical representation of the data presented in a. Figure created using http://www.ruppweb.org/mattprob/default.html (Matthews, 1968)

a)

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	1.0000	1.0000	2.74	55.11	37533.00
2	0.0000	0.0000	1.37	10.21	75066.00

b)

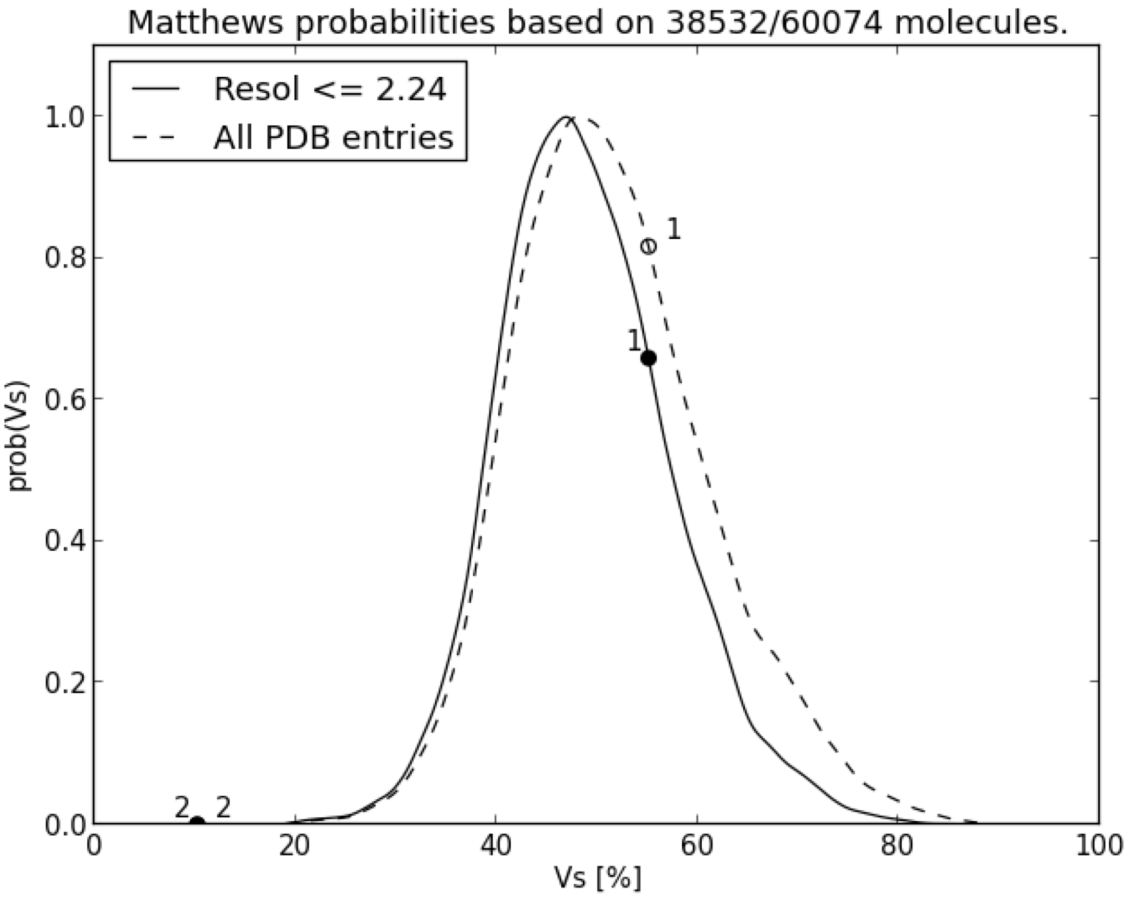


FIGURE B.12: **Bp-TssK$_{SN}$.his$_6$ Matthews calculation. a)** Table of the possible number of molecules within the asymmetric unit with their respective probability and corresponding percentage solvent content. This would suggest a single molecule of Bp-TssK$_{SN}$.his$_6$ **b)** A graphical representation of the data presented in a. Figure created using http://www.ruppweb.org/mattprob/default.html (Matthews, 1968)

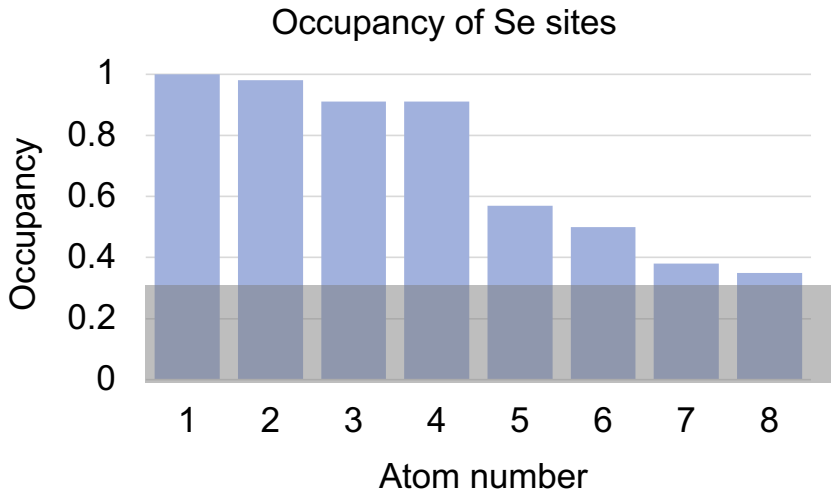
identified 22 possible Selenium sites, which was reduced to only eight by CRANK2 to remove atoms with an occupancy < 0.3 (figure B.13a). Inspection of the sites indicates a high occupancy for four of the eight sites far less than the 16 sites expected within the asymmetric unit. These sites were passed to Multicomb and SOLOMON within the CRANK2 software package which carried out phase combination and density modification, respectively. Observation of the map from the selected hand identified continuous electron density corresponding to α-helices confirming that the correct hand has been chosen (figure B.13b). An initial model was then built utilising Parrot, Buccaneer and REFMAC5. The initial model had an R_{work} and R_{free} of 0.39 and 0.44, respectively (figure B.14) and contained two molecules of TssK in the asymmetric unit. However, the model comprised only the shoulder and neck domains, and there was a distinct lack of electron density for either head domain (figure B.15b,c), even though there is sufficient space in the crystal lattice (figure B.15a).

The Bc-TssK model was subsequently refined against a higher resolution (2.2 Å) native set (3.4, table 1) and underwent multiple rounds of manual model building and refinement, with a final R_{work} and R_{free} of 0.26 and 0.29, respectively. In parallel, the structure determination of Bc-$TssK_{SN}$ using data to 1.6 Å (3.4, table 1) (see B.7) allowed this high-resolution model to be superimposed onto both chains of Bc-TssK (Chain A, 0.50 Å RMSD over 289 equivalenced residues, Chain B, 0.44 Å RMSD over 294 equivalenced residues) The refinement using the higher-resolution model had an R_{work} and R_{free} statistics of 0.25 and 0.28, respectively.

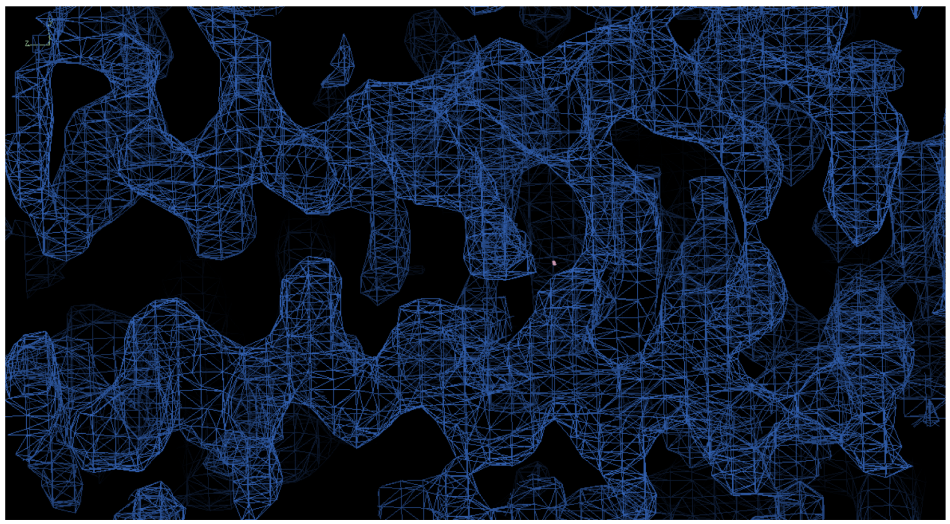
B.6.2 Model building of the Bc-TssK head domain

At this stage, density could be observed for only one of the head domains (chain A) (B.15b,c). Model building within this region enabled the placement of two α-helices and seven β-strands B.16a. However, the electron density was of insufficient quality to allow the placing of all the residues belonging to this region. Subsequently, the structure of the *E. coli* TssK (Ec-TssK) ortholog was published and included a high-resolution model of the head domain (PDB ID: 5M2Y). As discussed in 3.4, Ec-TssK has 33 % sequence identity to Bc-TssK (35 % in the head domain (321-448)), therefore a structural superposition of the Ec-TssK onto the secondary structure identified for the Bc-TssK head domain was used as a guide to identify the chain trace of the domain (2.3 Å RMSD over 31 equivalenced residues) (figure B.16b).

a)

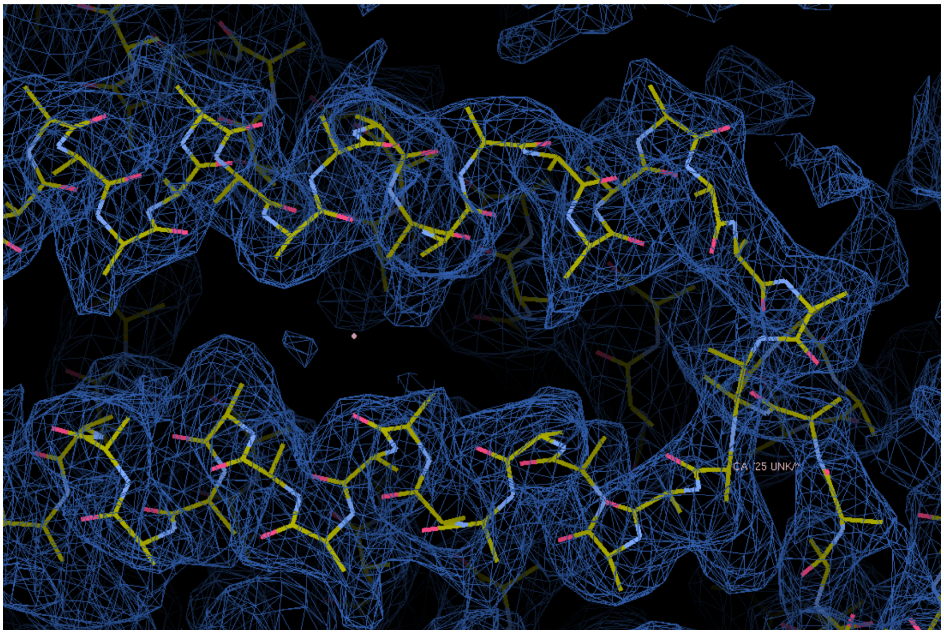


b)



Selected hand

FIGURE B.13: **Experimental phasing of Bc-TssK. a)** Table of the number of Selenium atoms found in the asymmetric unit and the occupancy for each atom. **b)** Initial electron density map of the selected hand generated from CRANK2.



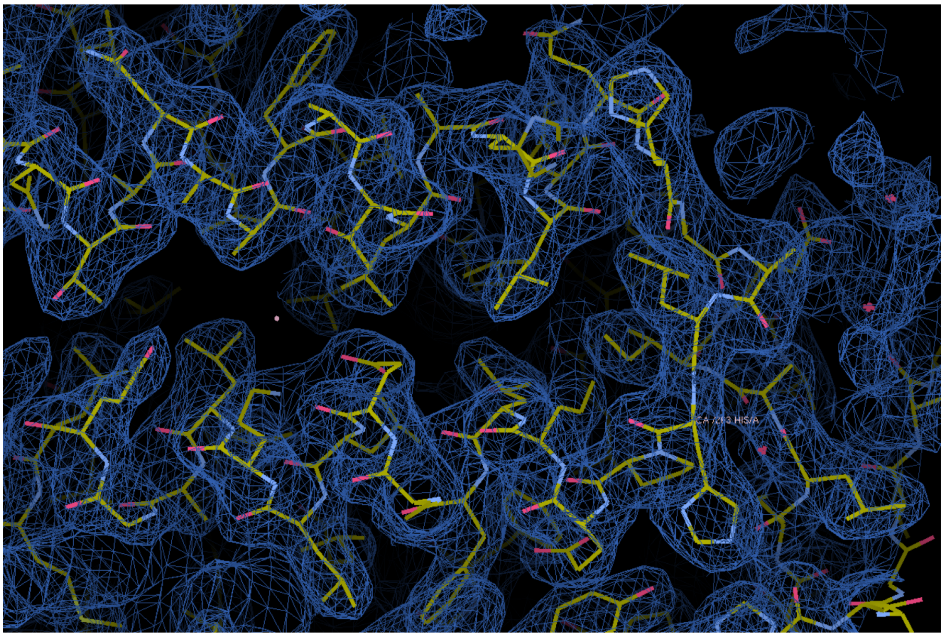


FIGURE B.14: **Initial and final electron density map and model of Bc-TssK. a)** Electron density map and model after initial model building. **b)** Final electron density map and model once building had been complete.

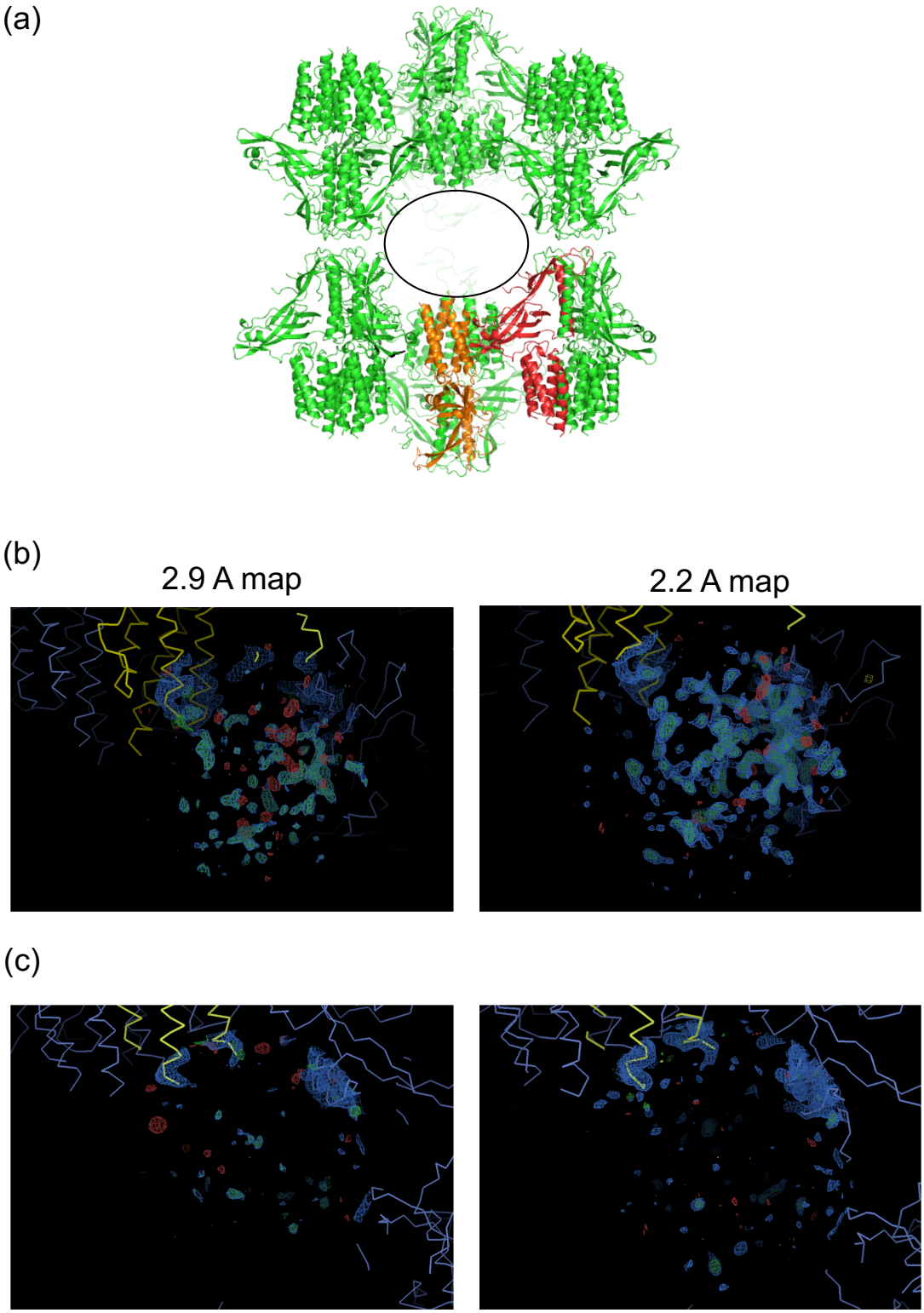


FIGURE B.15: **Packing of Bc-TssK Matthews calculation within the crystal lattice. a)** View of the Bc-TssK crystal lattice down the A axis indicating space remaining for the missing head domains. The two Bc-TssK monomers are coloured red (chain A) and orange (chain B). The black circle indiates the possible location for the head domain. **b)** 2Fo-Fc and Fo-Fc maps generated around the location for a potential head domain in Chain A. (Left) 2.9 Å map, (Right) 2.2 Å map. **c)** 2Fo-Fc and Fo-Fc maps generated around the location for a potential head domain in Chain B. (Left) 2.9 Å map, (Right) 2.2 Å map. All maps are contoured as 2Fo-Fc = 1.0, Fo-Fc = 3.0 RMSD.

Analysis of the quality of the electron density for the head domain of chain A shows that it is highly variable. Whereby, this domain makes no contacts with the shoulder and neck domains with which it is associated with and solely creates crystal contacts with an adjacent shoulder domain in the lattice. It is noticeable the density associated with residues near the crystal contact are distinctly better than those on the opposite face (figure B.16c). Therefore, the lack of equivalent adventitious crystal contacts and the inherent mobility of the head domain could indicate why this domain is absent in Chain B. As mentioned in 3.4, α8 of the Bc-TssK has been modelled as an α-helix, whereas the equivalent residue range in Ec-TssK is an extended loop. As can be seen in figure B.16d, there are density features consistent with side chains for this helix, suggesting these residues adopt a different fold. The final model of the head domain consists of 99 poly-alanine residues modelled as eight β-stands and three α-helices (figure B.16e).

B.6.3 Final model build and model validation of Bc-TssK.

The final Bc-TssK model, now with the head domain for chain A, underwent addition of solvent where appropriate (figure B.14) resulting in an R_{work} and R_{free} of 0.23 and 0.27, respectively. The final refinement statistics can be found in 3.4, table 1. The final model from Bc-TssK crystals contains two chains of Bc-TssK (Chains A and B), 108 waters and three ethylene glycol molecules. Chain A contains residues 3-447 which comprises the shoulder, neck and head domains with loops 188-121, 219-235 and 318-320 absent. Whereas, chain B contains residues 3-316, which form the shoulder and neck domain, but no head domain and loops 188-121, 219-235 missing (figure B.20a).

Structure validation was carried out using the Molprobity validation server (Chen et al., 2010). For the given resolution of the data, this placed the structure in the 99 th percentile when compared to other structures in the database, see 3.4, table 1 for details. In addition, analysis of the Ramachandran angles indicated that 94.4 % were favourable, with 0.7 % outliers B.17. The model shall be deposited prior to the submission of 3.4.

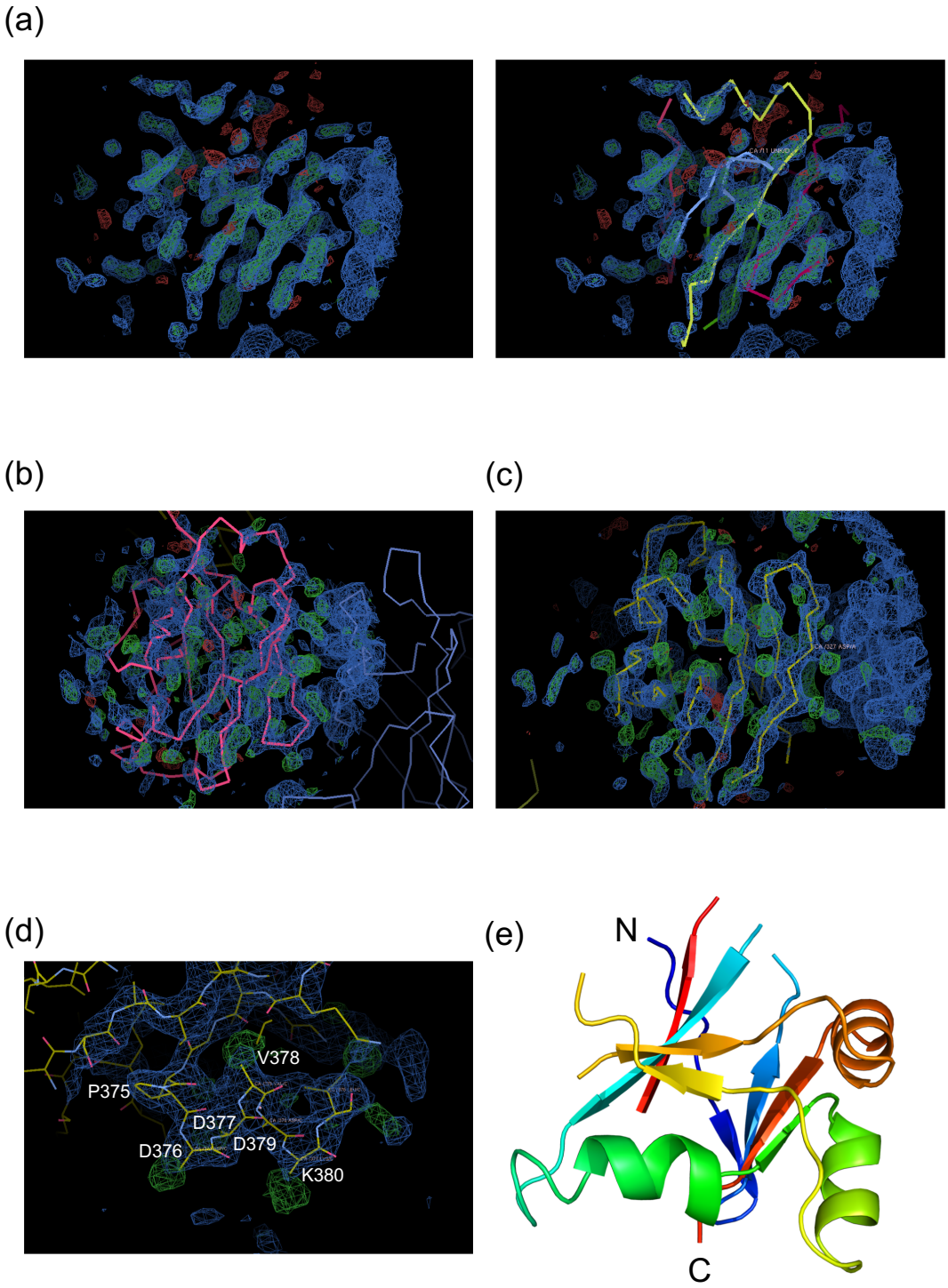


FIGURE B.16: **Structure determination of the Bc-TssK head domain.** **a)** (Left) Electron density around the region of the asymmetric unit believed to be the head domain of Chain A. (Right) Initial model building of the head domain within the unaccounted for electron density. **b)** Resultant electron density map after refinement with the Ec-TssK head domain structure superimposed on the Bc-TssK head domain initial structure. **c)** Final electron density map after the Bc-TssK head domain had been built. **d)** Electron density surrounding $\alpha 8$, the side chains have been labelled in white. **e)** Final model of the Bc-TssK head domain coloured in chainbow. All electron density maps have an Fo-Fc = 3.0 and 2Fo-Fc = 1.0.

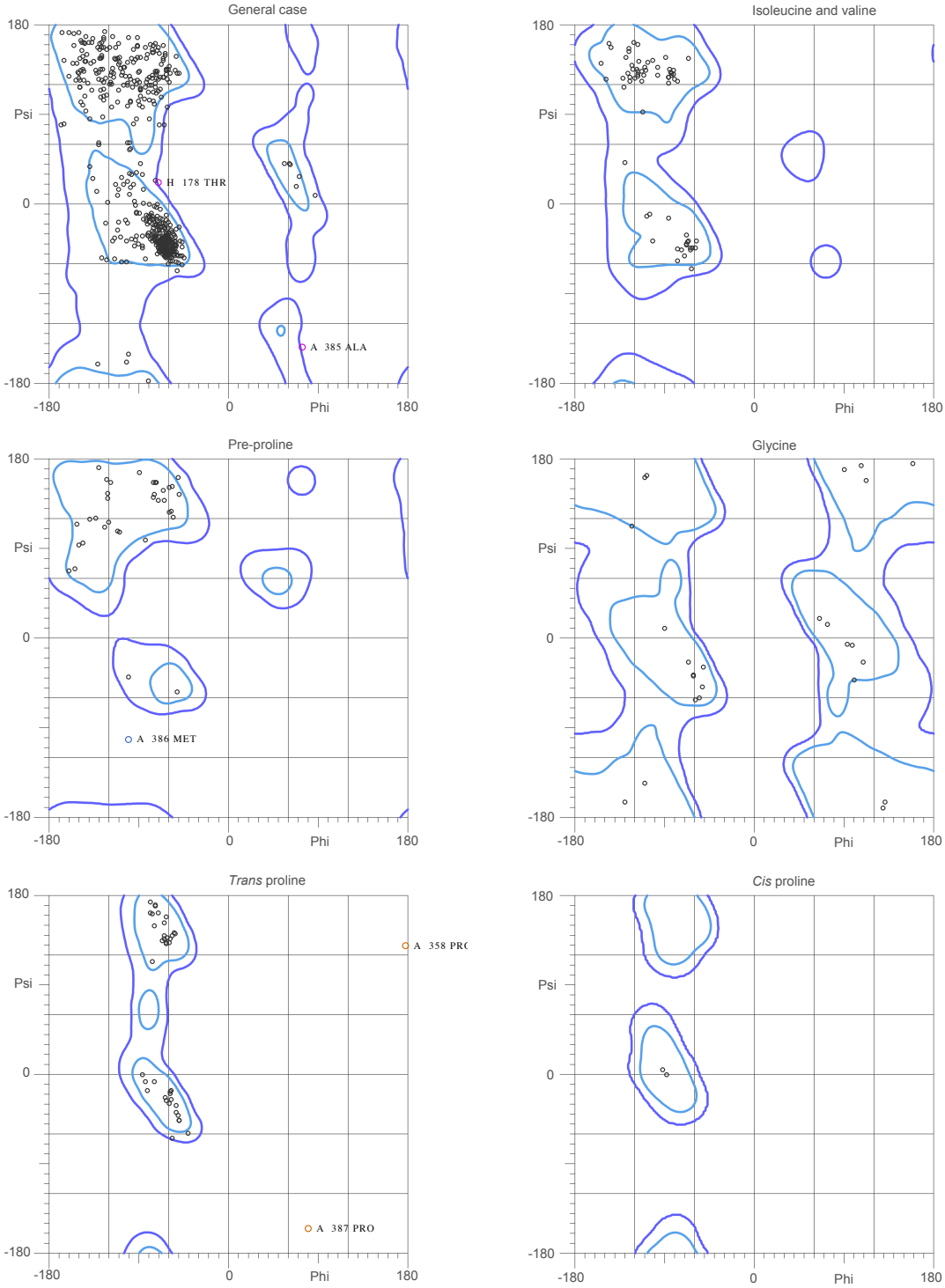


FIGURE B.17: **Bc-TssK structure Ramachandran plot.** Analysis of the phi and psi torsion angles for different residue classes. These include the general case, isoleucine and valine, pre-proline, glycine, trans proline and cis proline. Light blue contour levels denote regions for favoured Ramachandran angles, dark blue allowed angles, and outside these are outliers. Figure created using Chen et al., 2010.

B.7 Structure determination, model building and validation of Bc-TssK$_{SN}$

The initial model built for the shoulder and neck domains of Bc-TssK, using data to 2.2 Å, was used as a search model in molecular replacement (PhaserMR (McCoy et al., 2007)) to provide initial phases for the Bc-TssK$_{SN}$.his$_6$ data to 1.6 Å. Calculation using the Matthews coefficient estimated a single molecule of Ah-TssA Nt1-Nt2.his$_6$ in the asymmetric unit; therefore a single copy of chain A was used as a search model. A unique solution was found with a refined TFZ of 50.2, an overall LLG of 6977, and no packing clashes, indicating a successful solution had been found. The model was subjected to multiple rounds of manual model building and refinement, including the addition of solvent where appropriate, resulting in an R_{work} and R_{free} of 0.20 and 0.22, respectively. The final refinement statistics can be found in 3.4, table 1. The final model from Bc-TssK$_{SN}$.his$_6$ crystals contains a single chain of Bc-TssK$_{SN}$ (Chains A) (residues 4-316), 196 waters and four ethylene glycol molecules. Structure validation was carried out using the Molprobity validation server (Chen et al., 2010) placing the structure in the 100 th percentile (see 3.4, table 1 for details). Analysis of the Ramachandran angles indicated that 97.2 % were favourable, with no outliers B.18. The model will be deposited prior to the submission of 3.4.

B.8 Structure determination, model building and validation of Bp-TssK$_{SN}$

The Bc-TssK$_{SN}$ model was used as a search model in molecular replacement (PhaserMR (McCoy et al., 2007)) to provide initial phases for the Bp-TssK$_{SN}$.his$_6$ data to 2.2 Å. Calculation using the Matthews coefficient suggested that the crystals contain a single molecule of Bp-TssK$_{SN}$.his$_6$ in the asymmetric unit. A single solution was found with a refined TFZ of 46.4, an overall LLG of 2197, and no packing clashes, indicating a successful solution had been found. The model underwent multiple rounds of manual model building and refinement, including the addition of solvent where appropriate, resulting in an R_{work} and R_{free} of 0.22 and 0.25, respectively. The final refinement statistics can be found in 3.4, table 1. The final model from Bp-TssK$_{SN}$.his$_6$ crystals contains a single chain of Bp-TssK$_{SN}$ (Chains A) (residues 3-315), 62 waters and two ethylene glycol molecules. Structure validation was

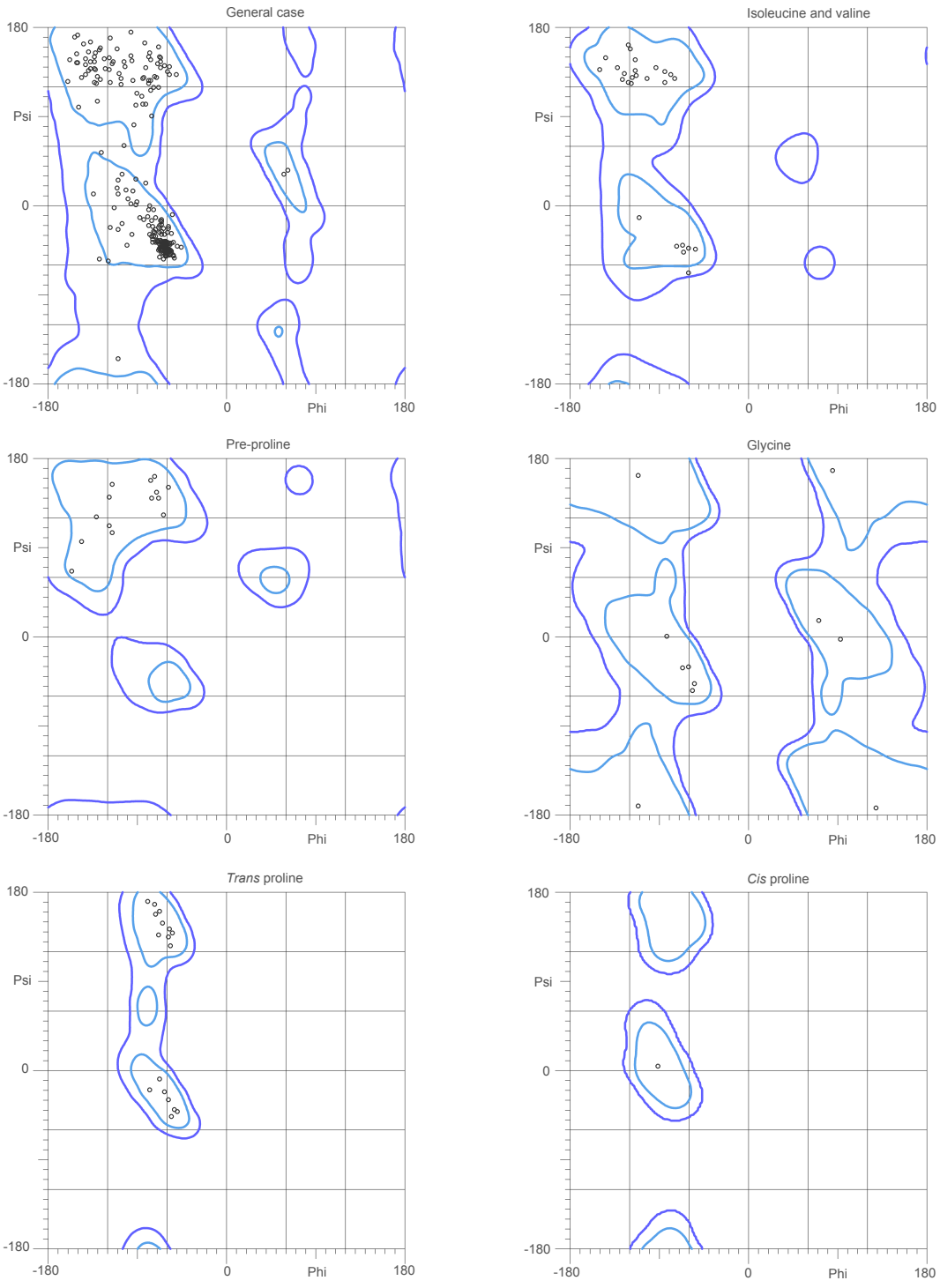


FIGURE B.18: **Bc-TssK$_{SN}$ structure Ramachandran plot.** Analysis of the phi and psi torsion angles for different residue classes. These include the general case, isoleucine and valine, pre-proline, glycine, trans proline and cis proline. Light blue contour levels denote regions for favoured Ramachandran angles, dark blue allowed angles, and outside these are outliers. Figure created using Chen et al., 2010.

carried out using the Molprobity validation server (Chen et al., 2010) placing the structure in the 98 th percentile (see 3.4, table 1 for details). Analysis of the Ramachandran angles indicated that 95.4 % were favourable, with no outliers B.19. The model will be deposited prior to the submission of 3.4.

B.9 Analysis of Bc-TssK trimerisation in the shoulder domain

As described in 3.4, TssK assembles to generate a trimer (B.20c). The assembly of this molecule features interactions mediated through both the shoulder and neck domain of each subunit. A more detailed analysis of trimerisation carried out in the shoulder domain, reveals a highly intertwined interface (B.20c) involving intimate interactions between all three subunits. Specifically, if each subunit within the Bc-TssK is labelled X, Y and Z, respectively, three distinct interactions are involved in generating the TssK assembly. The first involves the generation of a hydrophobic core through residues contributed from α1 of each of the subunits. The other areas of interaction arise from the interdigitation of the two distinct faces of the extended loop between β5 and 6 of subunit X, with the adjacent subunit, Y. One of them involves residues on the inner face of the β5/6 loop (X-N137 and X-F138) generating contacts with the N-terminal end of Y-α1 (Y-Q18) and the N-terminal residues of subunit Z (Z-K6 and Z-L8). The other interaction is made by the outer face of the β5/6 loop, interacting with the adjacent loop between β4 and 5 of subunit Y (Y-T113 and Y-N114). Both interfaces involve contributions of both the main chain and side chain and are predominantly hydrophilic.

B.10 Bc-TssK is structurally related to RBPs

As has been discussed in 3.4, Bc-TssK and Bp-TssK, like Ec-TssK, show a distinct structural resemblance to phage receptor binding proteins (RBPs) (figure B.21). A comparison of Bc-TssK with RBPs from *lactococcal* phage p2, 1358 and TP901-1 indicates a commonality in their quaternary structure, such that they all generate biological trimers. In addition, they all contain a shoulder, neck and head domain albeit that they are variable in size and the architecture of them varies significantly depending on the species. This is most evident when comparing the neck domains of each molecule, such that in the case of Bc-TssK, the

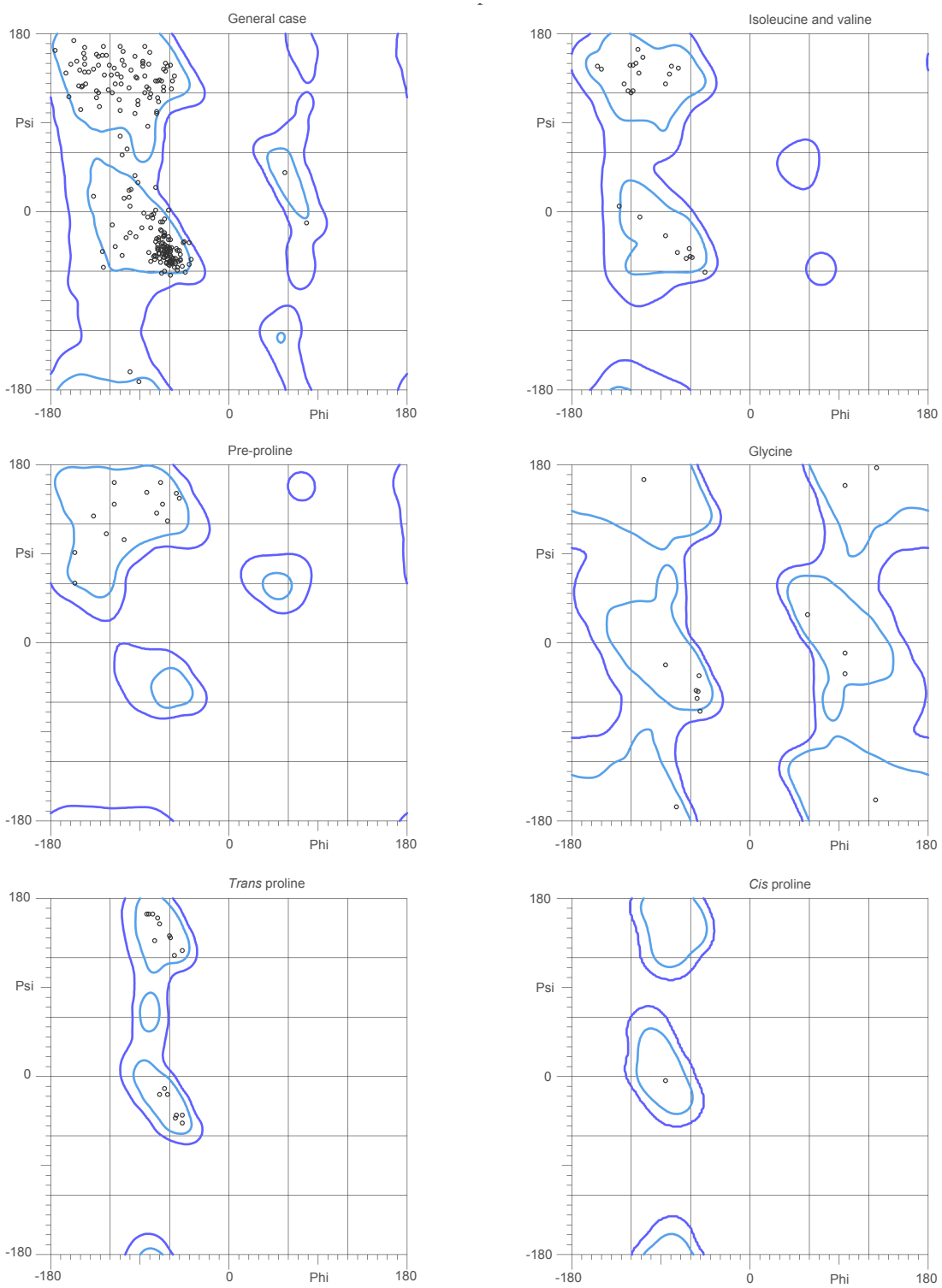


FIGURE B.19: **Bp-TssK$_{SN}$ structure Ramachandran plot.** Analysis of the phi and psi torsion angles for different residue classes. These include the general case, isoleucine and valine, pre-proline, glycine, trans proline and cis proline. Light blue contour levels denote regions for favoured Ramachandran angles, dark blue allowed angles, and outside these are outliers. Figure created using Chen et al., 2010.

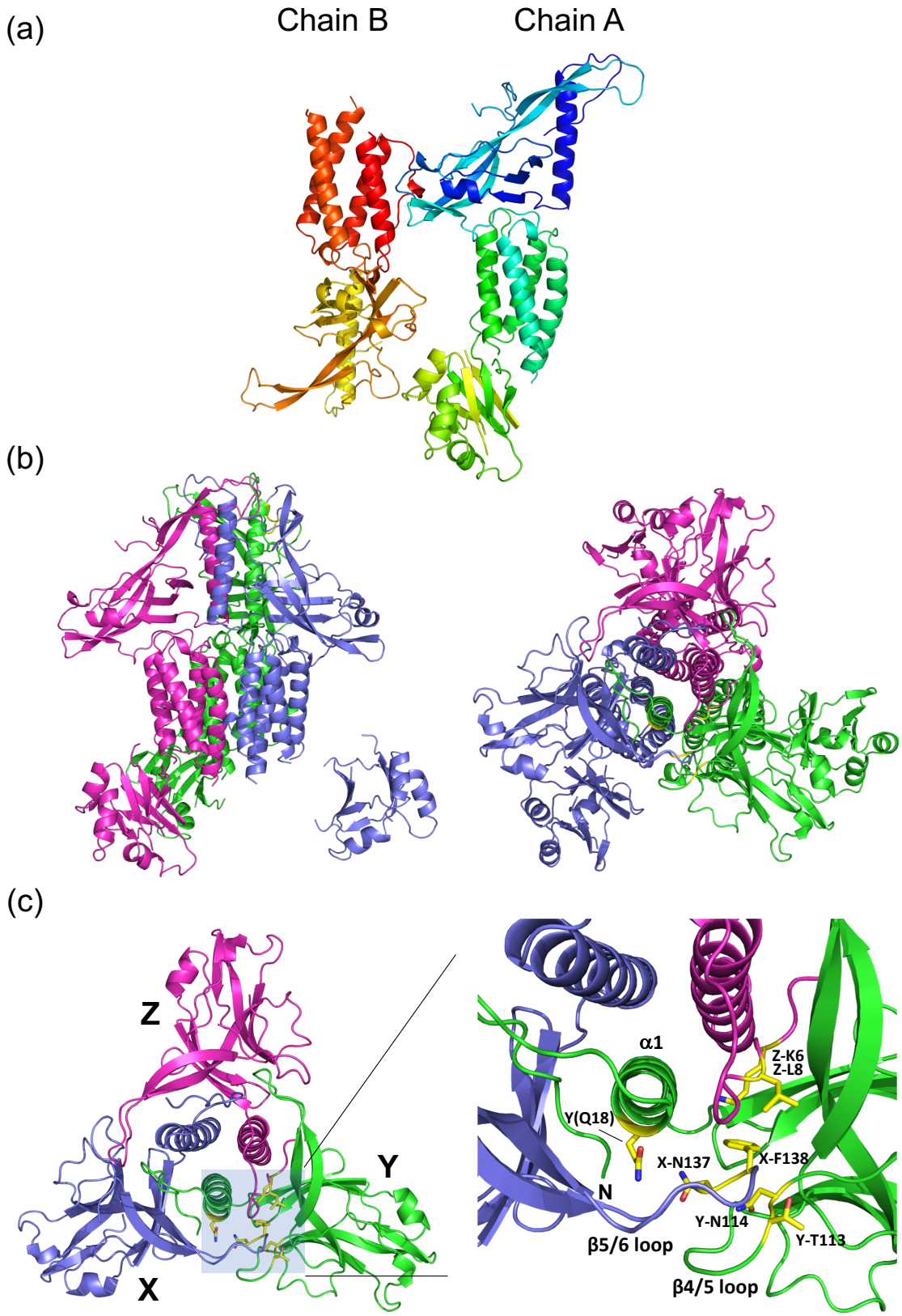


FIGURE B.20: **Structural analysis of the Bc-TssK Trimer. a)** Composition of the Bc-TssK asymmetric unit in the final model. **b)** (Left) Side view of the Bc-TssK trimer. (Right) View down the three-fold axis of the Bc-TssK trimer. **c)** (Left) View of the Bc-TssK trimer down the three-fold axis with only the shoulder domain visible. Each subunit in the assembly is labelled. (Right) View of the trimerisation interface generated in the shoulder domain. key residues are shown in yellow and labelled.

neck domain consists of α-helical bundles, whereas in P2 and TP901-1 their respective neck domains consist of a beta prism, while 1358 appears to have a distinctly truncated neck domain. What is clear is the conservation of the shoulder domain, particularly the internal trimerisation interface involving the N-terminal residues and α1 found at the core, indicating a possible conserved role for this domain.

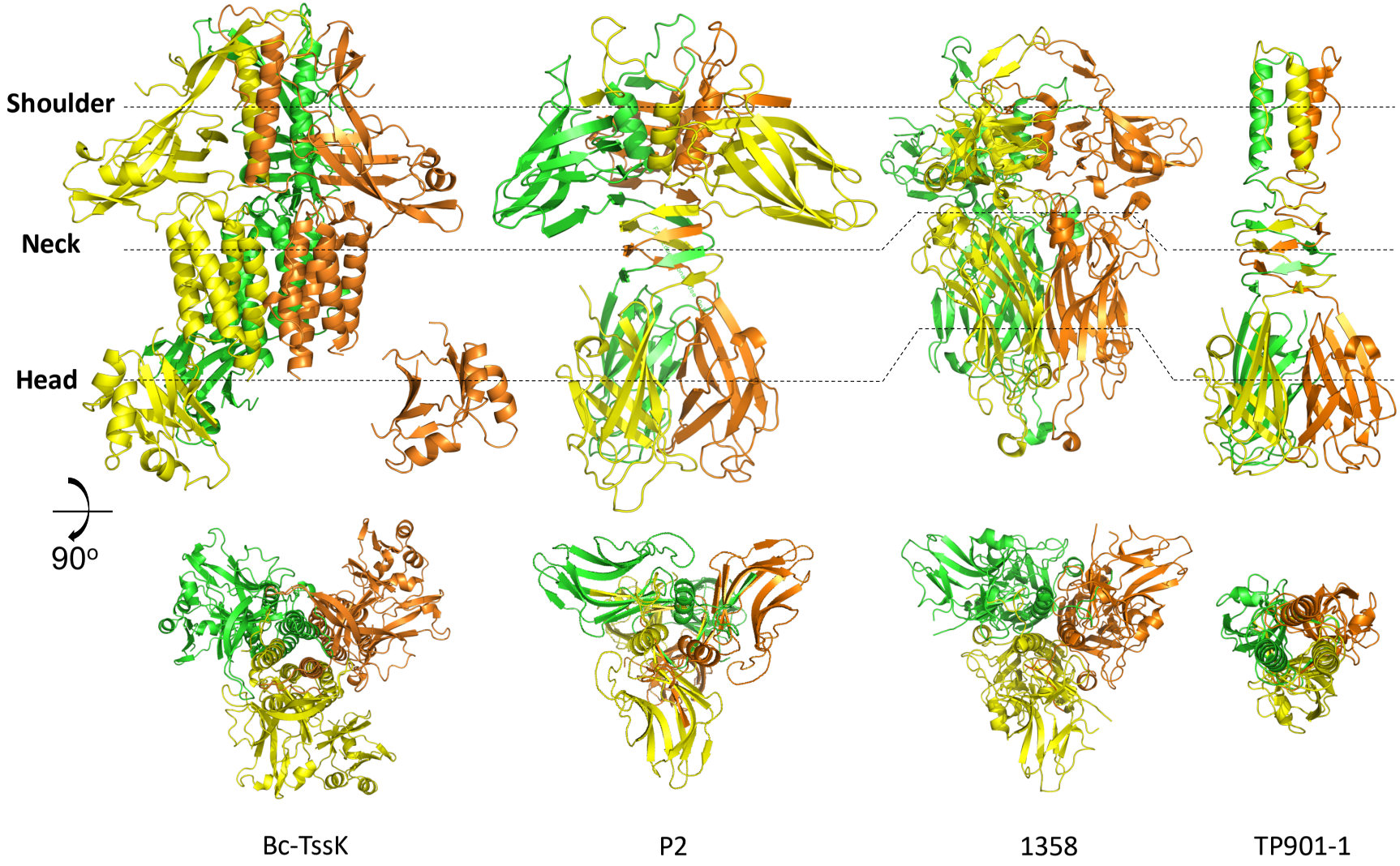


FIGURE B.21: **Structural comparison of Bc-Tssk with RBPs.** Side-by-side view of Bc-TssK with *lactococcal* phage P2, 1358 and TP901-1 RBPs. Dotted lines pass through an approximate centre for each domain. Figure generated using PBDs, P2-4RGG, 1358-4I99 and TP901-1-3U68

Bibliography

Abendroth, Jan et al. (2004). "The structure of the cytoplasmic domain of EpsL, an inner membrane component of the type II secretion system of *Vibrio cholerae*: An unusual member of the actin-like ATPase superfamily". In: *Journal of Molecular Biology* 344, pp. 619–633.

Abendroth, Jan et al. (2005). "The X-ray structure of the type II secretion system complex formed by the N-terminal domain of EpsE and the cytoplasmic domain of EpsL of *Vibrio cholerae*". In: *Journal of Molecular Biology* 348, pp. 845–855.

Abendroth, Jan et al. (2009). "The three-dimensional structure of the cytoplasmic domains of EpsF from the type 2 secretion system of *Vibrio cholerae*". In: *Journal of Structural Biology* 166, pp. 303–315.

Abrahams, J P and A G W Leslie (1996). "Methods used in the structure determination of bovine mitochondrial F1 ATPase". In: *Acta Crystallographica Section D Biological Crystallography* 52, pp. 30–42.

Alvarez-Martinez, Cristina E and Peter J Christie (2009). "Biological diversity of prokaryotic type IV secretion systems." In: *Microbiology and molecular biology reviews : MMBR* 73, pp. 775–808.

Aschtgen, Marie Stéphanie et al. (2008). "SciN is an outer membrane lipoprotein required for type VI secretion in enteroaggregative *Escherichia coli*". In: *Journal of Bacteriology* 190, pp. 7523–7531.

Aschtgen, Marie Stéphanie et al. (2010). "The SciZ protein anchors the enteroaggregative *Escherichia coli* Type VI secretion system to the cell wall". In: *Molecular Microbiology* 75, pp. 886–899.

Aschtgen, Marie Stéphanie et al. (2012). "The C-tail anchored TssL subunit, an essential protein of the enteroaggregative *Escherichia coli* Sci-1 Type VI secretion system, is inserted by YidC". In: *MicrobiologyOpen* 1, pp. 71–82.

Aubert, Daniel F, Ronald S Flannagan, and Miguel A Valvano (2008). "A novel sensor kinase-response regulator hybrid controls biofilm formation and type VI secretion system activity in *Burkholderia cenocepacia*". In: *Infection and Immunity* 76, pp. 1979–1991.

Aubert, Daniel F., Sherry Hu, and Miguel A. Valvano (2015). "Quantification of type VI secretion system activity in macrophages infected with *Burkholderia cenocepacia*". In: *Microbiology (United Kingdom)* 161, pp. 2161–2173.

Aubert, Daniel F. et al. (2016). "A Burkholderia Type VI Effector Deamidates Rho GTPases to Activate the Pyrin Inflammasome and Trigger Inflammation". In: *Cell Host & Microbe* 19, pp. 664–674.

Balakrishnan, Lekshmy, Colin Hughes, and Vassilis Koronakis (2001). "Substrate-triggered recruitment of the TolC channel-tunnel during type I export of hemolysin by *Escherichia coli*". In: *Journal of Molecular Biology* 313, pp. 501–510.

Ballister, Edward R et al. (2008). "In vitro self-assembly of tailorable nanotubes from a simple protein building block." In: *Proceedings of the National Academy of Sciences of the United States of America* 105, pp. 3733–3738.

Barret, Matthieu et al. (2011). "Genomic analysis of the type VI secretion systems in *Pseudomonas* spp.: Novel clusters and putative effectors uncovered". In: *Microbiology* 157, pp. 1726–1739.

Basler, M. et al. (2012). "Type VI secretion requires a dynamic contractile phage tail-like structure". In: *Nature* 483, pp. 182–186.

Battye, T. Geoff G. et al. (2011). "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM". In: 67, pp. 271–281.

Benz, Juliane et al. (2012). "Structural Insights into the Effector – Immunity System Tse1/Tsi1 from *Pseudomonas aeruginosa*". In: *PLoS ONE* 7, pp. 2112–2117.

Bladergroen, M. R., K. Badelt, and H. P. Spaink (2003). "Infection-Blocking Genes of a Symbiotic *Rhizobium leguminosarum* Strain That Are Involved in Temperature-Dependent Protein Secretion". In: *Molecular Plant-Microbe Interactions* 16, pp. 53–64.

Bönemann, Gabriele et al. (2009). "Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion". In: *EMBO Journal* 28, pp. 315–325.

Boyer, Frédéric et al. (2009). "Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources?" In: *BMC genomics* 10, p. 104.

Brooks, Teresa M et al. (2013). "Lytic activity of the *Vibrio cholerae* type VI secretion toxin VgrG-3 is inhibited by the antitoxin TsaB". In: *Journal of Biological Chemistry* 288, pp. 7618–7625.

Brunet, Yannick R. et al. (2014). "Type VI secretion and bacteriophage tail tubes share a common assembly pathway". In: *EMBO Reports* 15, pp. 315–321.

Brunet, Yannick R. et al. (2015). "The Type VI Secretion TssEFGK-VgrG Phage-Like Baseplate Is Recruited to the TssJLM Membrane Complex via Multiple Contacts and Serves As Assembly Platform for Tail Tube/Sheath Polymerization". In: *PLoS Genetics* 11, e1005545.

Campos, Manuel et al. (2010). "Detailed structural and assembly model of the type II secretion pilus from sparse data". In: *Proceedings of the National Academy of Sciences* 107, pp. 13081–13086.

Cascales, Eric and Peter J Christie (2004). "*Agrobacterium* VirB10, an ATP energy sensor required for type IV secretion." In: *Proceedings of the National Academy of Sciences of the United States of America* 101, pp. 17228–33.

Chandran, Vidya et al. (2009). "Structure of the outer membrane complex of a type IV secretion system". In: *Nature* 462, pp. 1011–1015.

Chen, Vincent B. et al. (2010). "*MolProbity*: all-atom structure validation for macromolecular crystallography". In: *Acta Crystallographica Section D* 66, pp. 12–21.

Chen, Yahua et al. (2011). "Regulation of Type VI Secretion System during *Burkholderia pseudomallei* Infection". In: *Infection and Immunity* 79, pp. 3064–3073.

Cherrak, Yassine et al. (2018). "Biogenesis and structure of a type VI secretion baseplate". In: *Nature Microbiology* 3, pp. 1404–1416.

Chopra, Ashok K and Clifford W Houston (1999). *Enterotoxins in Aeromonas-associated gastroenteritis*.

Chow, Janet and Sarkis K Mazmanian (2010). "A Pathobiont of the Microbiota Balances Host Colonization and Intestinal Inflammation". In: *Cell Host & Microbe* 7, pp. 265–276.

Christie, Peter J, Neal Whitaker, and Christian González-Rivera (2014). "Revised Mechanism and Structure of the Bacterial Type IV Secretion Systems". In: *Biochim Biophys Acta* 184312.

Cianfanelli, Francesca R., Laura Monlezun, and Sarah J. Coulthurst (2016). "Aim, Load, Fire: The Type VI Secretion System, a Bacterial Nanoweapon". In: *Trends in Microbiology* 24, pp. 51–62.

Cisneros, David A, Gerard Pehau-Arnaudet, and Olivera Francetic (2012). "Heterologous assembly of type IV pili by a type II secretion system reveals the role of minor pilins in assembly initiation". In: *Molecular Microbiology* 86, pp. 805–818.

Cornelis, Guy (2006). "The type III secretion injectisome." In: *Nature reviews Microbiology* 4, pp. 811–825.

Costa, Tiago R. D. et al. (2015). "Secretion systems in Gram-negative bacteria: structural and mechanistic insights". In: *Nature Reviews Microbiology* 13, pp. 343–359.

Cowtan, Kevin (2006). "The Buccaneer software for automated model building." In: *Acta Crystallographica Section D: Biological Crystallography* 62, pp. 1002–1011.

— (2010). "Recent developments in classical density modification." In: *Acta crystallographica. Section D, Biological crystallography* 66, pp. 470–478.

De Maayer, Pieter et al. (2011). "Comparative genomics of the type VI secretion systems of *Pantoea* and *Erwinia* species reveals the presence of putative effector islands that may be translocated by the VgrG and Hcp proteins". In: *BMC Genomics* 12, p. 576.

Delepelaire, P. (2004). "Type I secretion in gram-negative bacteria". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1694, pp. 149–161.

Dix, S. R. et al. (2018a). "TssA from *Aeromonas hydrophila*: Expression, purification and crystallographic studies." In: *Acta Crystallogr. Sect. F.* 74, pp. 578–582.

Dix, Samuel R. et al. (2018b). "Structural insights into the function of type VI secretion system TssA subunits". In: *Nature communications* 9.1, p. 4765.

Dong, T. G. et al. (2013). "Identification of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholerae*". In: *Proceedings of the National Academy of Sciences* 110, pp. 2623–2628.

Duda, Robert L et al. (1985). "Mass distribution of a probable tail-length-determining protein in bacteriophage T4." In: *Proceedings of the National Academy of Sciences* 82, pp. 5550–5554.

Durand, Eric et al. (2012). "Structural characterization and oligomerization of the TssL protein, a component shared by bacterial type VI and type IVb secretion systems". In: *Journal of Biological Chemistry* 287, pp. 14157–14168.

Durand, Eric et al. (2015). "Biogenesis and structure of a type VI secretion membrane core complex". In: *Nature* 523, pp. 555–560.

Emsley, P. et al. (2010). "Features and development of *Coot*". In: *Acta Crystallographica Section D* 66, pp. 486–501.

English, Grant et al. (2014). "Biochemical analysis of TssK, a core component of the bacterial Type VI secretion system, reveals distinct oligomeric states of TssK and identifies a TssK-TssFG subcomplex." In: *The Biochemical journal* 461, pp. 291–304.

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

Evans, Philip (2006). "Scaling and assessment of data quality". In: *Acta Crystallographica Section D: Biological Crystallography* 62, pp. 72–82.

Evans, Philip R (2011). "An introduction to data reduction: Space-group determination, scaling and intensity statistics". In: *Acta Crystallographica Section D: Biological Crystallography* 67, pp. 282–292.

Evans, Philip R and Garib N Murshudov (2013). "How good are my data and what is the resolution?" In: *Acta Crystallographica Section D: Biological Crystallography* 69, pp. 1204–1214.

Farenc, C. et al. (2014). "Molecular Insights on the Recognition of a *Lactococcus lactis* Cell Wall Pellicle by the Phage 1358 Receptor Binding Protein". In: *Journal of Virology* 88, pp. 7005–7015.

Felisberto-Rodrigues, Catarina et al. (2011). "Towards a structural comprehension of bacterial type vi secretion systems: Characterization of the TssJ-TssM complex of an *Escherichia col*i pathovar". In: *PLoS Pathogens* 7, e1002386.

Flannagan, Ronald S. et al. (2012). "*Burkholderia cenocepacia* disrupts host cell actin cytoskeleton by inactivating Rac and Cdc42". In: *Cellular Microbiology* 14, pp. 239–254.

Galan, J E and Hans Wolf-watz (2006). "Protein delivery into eukaryotic cells by type III secretion machines." In: *Nature* 444, pp. 567–73.

Gasteiger, Elisabeth et al. (2005). "Protein Identification and Analysis Tools on the ExPASy Server". In: *The Proteomics Protocols Handbook*, pp. 571–607.

Ge, Peng et al. (2015). "Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states." In: *Nature structural & molecular biology* 22, pp. 377–82.

Haden, Russell L. (1923). "Elective localization in the eye of bacteria from infected teeth". In: *Archives of Internal Medicine* 32, pp. 828–849.

Hanahan, Douglas (1983). "Studies on transformation of *Escherichia coli* with plasmids". In: *Journal of Molecular Biology* 166, pp. 557–580.

Heymann, J Bernard et al. (2013). "Three-dimensional structure of the toxin-delivery particle antifeeding prophage of *serratia entomophila*". In: *Journal of Biological Chemistry* 288, pp. 25276–25284.

Ho, BrianT., TaoG. Dong, and JohnJ. Mekalanos (2014). "A View to a Kill: The Bacterial Type VI Secretion System". In: *Cell Host & Microbe* 15, pp. 9–21.

Hood, Rachel D et al. (2010). "A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria." In: *Cell host & microbe* 7, pp. 25–37.

Hu, Bo et al. (2015). "Visualization of the type III secretion sorting platform of *Shigella flexneri*". In: *Proceedings of the National Academy of Sciences* 112, pp. 1047–1052.

Janda, J M and S L Abbott (1998). "Evolving concepts regarding the genus *Aeromonas*: an expanding Panorama of species, disease presentations, and unanswered questions." In: *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 27, pp. 332–344.

Junker, Mirco, Richard N. Besingi, and Patricia L. Clark (2009). "Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion". In: *Molecular Microbiology* 71, pp. 1323–1332.

Kabsch, Wolfgang (2010). "Xds". In: *Acta Crystallographica Section D: Biological Crystallography* 66, pp. 125–132.

Kanonenberg, Kerstin, Christian K W Schwarz, and Lutz Schmitt (2013). "Type I secretion systems - a story of appendices". In: *Research in Microbiology* 164, pp. 596–604.

Kapitein, Nicole et al. (2013). "ClpV recycles VipA/VipB tubules and prevents non-productive tubule formation to ensure efficient type VI protein secretion". In: *Molecular Microbiology* 87, pp. 1013–1028.

Karplus, P Andrew and Kay Diederichs (2015). "Assessing and maximizing data quality in macromolecular crystallography". In: *Current Opinion in Structural Biology* 34, pp. 60–68.

Kimbrough, Tyler G and Samuel I Miller (2000). "Contribution of *Salmonella typhimurium* type III secretion components to needle complex formation". In: *Proceedings of the National Academy of Sciences* 97, pp. 11008–11013.

Koronakis, Vassilis, Jeyanthy Eswaran, and Colin Hughes (2004). "Structure and function of TolC: The Bacterial Exit Duct for Proteins and Drugs". In: *Annu. Rev. Biochem.* 73, pp. 467–489.

Korotkov, Konstantin V., Maria Sandkvist, and Wim G. J. Hol (2012). "The type II secretion system: biogenesis, molecular architecture and mechanism". In: *Nature Reviews Microbiology* 10, pp. 336–351.

Korotkov, Konstantin V et al. (2011). "Structural and Functional Studies on the Interaction of GspC and GspD in the Type II Secretion System". In: *PLoS Pathog* 7, e1002228.

Koskiniemi, S. et al. (2013). "Rhs proteins from diverse bacteria mediate intercellular competition". In: *Proceedings of the National Academy of Sciences* 110, pp. 7032–7037.

Krissinel, Evgeny (2012). "Enhanced fold recognition using efficient short fragment clustering." In: *Journal of molecular biochemistry* 1, pp. 76–85.

Krissinel, Evgeny and Kim Henrick (2007). "Inference of Macromolecular Assemblies from Crystalline State". In: *Journal of Molecular Biology* 372, pp. 774–797.

Kubori, T et al. (1998). "Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system." In: *Science* 280, pp. 602–5.

Kubori, Tomoko et al. (2000). "Molecular characterization and assembly of the needle complex of the *Salmonella typhimurium* type III protein secretion system". In: *Proceedings of the National Academy of Sciences* 97, pp. 10225–10230.

Kudryashev, Mikhail et al. (2015). "Structure of the Type VI Secretion System Contractile Sheath". In: *Cell* 160.5, pp. 952–962.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4". In: *Nature* 227, pp. 680–685.

Lara-Tejero, María et al. (2011). "A sorting platform determines the order of protein secretion in bacterial type III systems". In: *Science* 331, pp. 1188–1191.

Leal, Ricardo M. F. et al. (2011). "Experimental procedure for the characterization of radiation damage in macromolecular crystals". In: *Journal of Synchrotron Radiation* 18, pp. 381–386.

Lee, Song-Hua et al. (2007). "*Burkholderia pseudomallei* animal and human isolates from Malaysia exhibit different phenotypic characteristics". In: *Diagnostic Microbiology and Infectious Disease* 58, pp. 263–270.

Leiman, Petr G and Mikhail M Shneider (2012a). "Contractile tail machines of bacteriophages." In: *Advances in experimental medicine and biology* 726, pp. 93–114.

Leiman, Petr G. and Mikhail M. Shneider (2012b). "Contractile tail machines of bacteriophages". In: *Advances in Experimental Medicine and Biology* 726, pp. 93–114.

Leiman, Petr G et al. (2009). "Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin." In: *Proceedings of the National Academy of Sciences of the United States of America* 106, pp. 4154–9.

Leiman, Petr G et al. (2010). "Morphogenesis of the T4 tail and tail fibers". In: *Virology Journal* 7, p. 355.

Leo, Jack C, Iwan Grin, and Dirk Linke (2012). "Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367, pp. 1088–1101.

Leyton, Denisse L, Amanda E Rossiter, and Ian R Henderson (2012). "From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis." In: *Nature reviews. Microbiology* 10, pp. 213–225.

Liang, Xiaoye et al. (2015). "Identification of divergent type VI secretion effectors using a conserved chaperone domain". In: *Proceedings of the National Academy of Sciences* 112, pp. 9106–9111.

Lin, Jer-Sheng, Lay-Sun Ma, and Erh-Min Lai (2013). "Systematic Dissection of the *Agrobacterium* Type VI Secretion System Reveals Machinery and Secreted Components for Subcomplex Formation". In: *PLoS ONE* 8, e67647.

Loquet, Antoine et al. (2012). "Atomic model of the type III secretion system needle". In: *Nature* 486, pp. 276–279.

Lossi, Nadine S et al. (2011). "Structure-function analysis of HsiF, a gp25-like component of the type VI secretion system, in *Pseudomonas aeruginosa*." In: *Microbiology (Reading, England)* 157, pp. 3292–305.

Low, Harry H et al. (2014). "Structure of a type IV secretion system". In: *Nature* 508, pp. 550–553.

Lycklama a Nijeholt, Jelger A. and Arnold J.M. Driessen (2012). "The bacterial Sec-translocase: Structure and mechanism". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367, pp. 1016–1028.

Lyczak, Jeffrey B, Carolyn L Cannon, and Gerald B Pier (2002). "Lung Infections Associated with Cystic Fibrosis Lung Infections Associated with Cystic Fibrosis". In: *Clinical microbiology reviews* 15, pp. 194–222.

Ma, Amy T et al. (2009). "Translocation of a *Vibrio cholerae* Type VI Secretion Effector Requires Bacterial Endocytosis by Host Cells". In: *Cell Host & Microbe* 5, pp. 234–243.

Ma, Lay Sun, Jer Sheng Lin, and Erh Min Lai (2009). "An IcmF family protein, ImpLM, is an integral inner membrane protein interacting with ImpKL, and its Walker a motif is required for type VI secretion system-mediated Hcp secretion in *Agrobacterium tumefaciens*". In: *Journal of Bacteriology* 191, pp. 4316–4329.

MacIntyre, Dana L. et al. (2010). "The *Vibrio cholerae* type VI secretion system displays antimicrobial properties". In: *Proceedings of the National Academy of Sciences* 107, pp. 19520 –19524.

Marlovits, Thomas C et al. (2004). "Structural insights into the assembly of the type III secretion needle complex". In: *Science* 306, pp. 1040–1042.

Marlovits, Thomas C et al. (2006). "Assembly of the inner rod determines needle length in the type III secretion injectisome". In: *Nature* 441, pp. 637–640.

Matias, Valério R. F. et al. (2003). "Cryo-Transmission Electron Microscopy of Frozen-Hydrated Sections of *Escherichia coli* and *Pseudomonas aeruginosa*". In: *Journal of Bacteriology* 185, pp. 6112–6118.

Matthews, B.W. (1968). "Solvent content of protein crystals". In: *Journal of Molecular Biology* 33, pp. 491–497.

McCoy, Airlie J. et al. (2007). "Phaser crystallographic software". In: *Journal of Applied Crystallography* 40, pp. 658–674.

Miller, JH (1972). "Experiments in Molecular Genetics". In: *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY*.

Mougous, Joseph D et al. (2006). "A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus". In: *Science* 312, pp. 1526–1530.

Murshudov, G. N., A. A. Vagin, and E. J. Dodson (1997). "Refinement of Macromolecular Structures by the Maximum-Likelihood Method". In: *Acta Crystallographica Section D* 53, pp. 240–255.

Murshudov, Garib N. et al. (2011). "REFMAC5 for the refinement of macromolecular crystal structures". In: *Acta Crystallographica Section D: Biological Crystallography* 67, pp. 355–367.

Nazarov, Sergey et al. (2018). "Cryo-EM reconstruction of Type VI secretion system baseplate and sheath distal end". In: *The EMBO Journal* 37, e201797103.

Nguyen, Van Son et al. (2017). "Type VI secretion TssK baseplate protein exhibits structural similarity with phage receptor-binding proteins and evolved to bind the membrane complex". In: *Nature Microbiology* 2, pp. 1–9.

Nivaskumar, Mangayarkarasi and Olivera Francetic (2014). "Type II secretion system: A magic beanstalk or a protein escalator". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1843, pp. 1568–1577.

Nivaskumar, Mangayarkarasi et al. (2014). "Distinct docking and stabilization steps of the pseudopilus conformational transition path suggest rotational assembly of type IV pilus-like fibers". In: *Structure* 6, pp. 685–696.

Palmer, Tracy and Ben C Berks (2012). "The twin-arginine translocation (Tat) protein export pathway". In: *Nature Reviews Microbiology* 10, pp. 483–496.

Pei, Xue-Yuan et al. (2011). "Structures of sequential open states in a symmetrical opening transition of the TolC exit duct." In: *Proceedings of the National Academy of Sciences of the United States of America* 108, pp. 2112–2117.

Piddock, Laura J V (2006). "Multidrug - resistance efflux pumps — not just for resistance". In: *Nature Reviews Microbiology* 4, pp. 629–636.

Pietrosiuk, Aleksandra et al. (2011). "Molecular basis for the unique role of the AAA+ chaperone ClpV in type VI protein secretion." In: *The Journal of biological chemistry* 286, pp. 30010–21.

Planamente, Sara et al. (2016). "TssA forms a gp6-like ring attached to the type VI secretion sheath". In: *The EMBO Journal* 35, pp. 1613–1627.

Poole, Stephen J et al. (2011). "Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (cdi) and rearrangement hotspot (rhs) systems". In: *PLoS Genetics* 7.

Pukatzki, S. et al. (2006). "Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system". In: *Proceedings of the National Academy of Sciences* 103, pp. 1528–1533.

Pukatzki, S. et al. (2007). "Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin". In: *Proceedings of the National Academy of Sciences* 104, pp. 15508–15513.

Pukatzki, Stefan, Steven B. McAuley, and Sarah T. Miyata (2009). "The type VI secretion system: translocation of effectors and effector-domains". In: *Current Opinion in Microbiology* 12, pp. 11–17.

Py, Béatrice, Laurent Loiseau, and Frédéric Barras (2001). "An inner membrane platform in the type II secretion machinery of Gram-negative bacteria". In: *EMBO reports* 2, pp. 244–248.

Radics, Julia, Lisa Königsmaier, and Thomas C Marlovits (2014). "Structure of a pathogenic type 3 secretion system in action". In: *Nature Structural and Molecular Biology* 21, pp. 82–87.

Read, R. J. and A. J. Schierbeek (1988). "A phased translation function". In: *Journal of Applied Crystallography* 21, pp. 490–495.

Rhodes, Gale (2006). *Crystallography Made Crystal Clear*. 3rd. Elsevier Inc.

Robb, Craig S., Francis E. Nano, and Alisdair B. Boraston (2012). "The structure of the conserved type six secretion protein TssL (DotU) from Francisella novicida". In: *Journal of Molecular Biology* 419, pp. 277–283.

Robert, Xavier and Patrice Gouet (2014). "Deciphering key features in protein structures with the new ENDscript server". In: *Nucleic Acids Research* 42, pp. 320–324.

Romling, U et al. (1994). "Epidemiology of chronic *Pseudomonas aeruginosa* infections in cystic fibrosis." In: *The Journal of infectious diseases* 170, pp. 1616–1621.

Rosa, Leonardo T. et al. (2017). "Structural basis for high-affinity adipate binding to AdpC (RPA4515), an orphan periplasmic-binding protein from the tripartite tricarboxylate transporter (TTT) family in *Rhodopseudomonas palustris*". In: *FEBS Journal* 284, pp. 4262–4277.

— (2019). "A New Mechanism for High-Affinity Uptake of C4-Dicarboxylates in Bacteria Revealed by the Structure of Rhodopseudomonas palustris MatC (RPA3494), a Periplasmic Binding Protein of the Tripartite Tricarboxylate Transporter (TTT) Family". In: *Journal of Molecular Biology* 431, pp. 351–367.

Rosales-Reyes, Roberto et al. (2012). "*Burkholderia cenocepacia* type VI secretion system mediates escape of type II secreted proteins into the cytoplasm of infected macrophages". In: *PLoS ONE* 7, pp. 1–14.

Rosenow, EC (1919). "Studies on elective localization". In: *J. Dent. Research* 1, pp. 205–249.

Rupp, B (2009). *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. 1st. New York: Garland Science.

Russell, Alistair B., S. Brook Peterson, and Joseph D. Mougous (2014). "Type VI secretion system effectors: poisons with a purpose". In: *Nature Reviews Microbiology* 12, pp. 137–148.

Russell, Alistair B et al. (2011). "Type VI secretion delivers bacteriolytic effectors to target cells". In: *Nature* 475, pp. 343–349.

Russell, Alistair B et al. (2012). "A widespread bacterial type VI secretion effector superfamily identified using a heuristic approach." In: *Cell host & microbe* 11, pp. 538–49.

Russell, Alistair B et al. (2013). "Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors". In: *Nature* 496, pp. 508–512.

Salomon, Dor et al. (2014). "Marker for type VI secretion system effectors." In: *Proceedings of the National Academy of Sciences of the United States of America* 111, pp. 9271–6.

Schlieker, Christian et al. (2005). "ClpV, a unique Hsp100/Clp member of pathogenic proteobacteria". In: *Biological Chemistry* 386, pp. 1115–27.

Schraidt, Oliver and Thomas C Marlovits (2011). "Three-dimensional model of Salmonella's needle complex at subnanometer resolution". In: *Science* 331, pp. 1192–1195.

Schwarz, Sandra et al. (2010). "*Burkholderia* Type VI Secretion Systems Have Distinct Roles in Eukaryotic and Bacterial Cell Interactions". In: *PLoS Pathog* 6, e1001068.

Sciara, G. et al. (2010). "Structure of *lactococcal* phage p2 baseplate and its mechanism of activation". In: *Proceedings of the National Academy of Sciences* 107, pp. 6852–6857.

Shalom, Gil, Jonathan G. Shaw, and Mark S. Thomas (2007). "In vivo expression technology identifies a type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages". In: *Microbiology* 153, pp. 2689–2699.

Sheldrick, George M. (2008). "A short history of SHELX". In: *Acta Crystallographica Section A: Foundations of Crystallography* 64, pp. 112–122.

Shikuma, Nicholas J et al. (2014). "Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures". In: *Science* 343, pp. 529–533.

Shneider, Mikhail M et al. (2013). "PAAR-repeat proteins sharpen and diversify the type VI secretion system spike." In: *Nature* 500, pp. 350–353.

Silverman, Julie M et al. (2013). "Haemolysin Coregulated Protein Is an Exported Receptor and Chaperone of Type VI Secretion Substrates". In: *Molecular Cell*, pp. 584–593.

Skubák, Pavol and Navraj S Pannu (2013). "Automatic protein structure solution from weak X-ray data". In: *Nature Communications* 4, p. 2777.

Skubák, Pavol, Willem-Jan Waterreus, and Navraj S Pannu (2010). "Multivariate phase combination improves automated crystallographic model building". In: *Acta Crystallographica Section D* 66, pp. 783–788.

Spinelli, Silvia et al. (2006a). "*Lactococcal* bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses". In: *Nature Structural and Molecular Biology* 13, pp. 85–89.

Spinelli, Silvia et al. (2006b). "Modular structure of the receptor binding proteins of Lactococcus lactis phages: The RBP structure of the temperate phage TP901-1". In: *Journal of Biological Chemistry* 281, pp. 14256–14262.

Srinivasa Rao, P. S. et al. (2004). "Use of proteomics to identify novel virulence determinants that are required for *Edwardsiella tarda* pathogenesis". In: *Molecular Microbiology* 53, pp. 573–586.

Suarez, Giovanni et al. (2008). "Molecular Characterization of a Functional Type VI Secretion System from a Clinical Isolate of *Aeromonas hydrophila*". In: *Microb Pathog* 44, pp. 344–361.

Suarez, Giovanni et al. (2010). "Role of Hcp, a type 6 secretion system effector, of *Aeromonas hydrophila* in modulating activation of host immune cells". In: *Microbiology* 156, pp. 3678–3688.

Tamura, Koichiro et al. (2013). "MEGA6: Molecular evolutionary genetics analysis version 6.0". In: *Molecular Biology and Evolution* 30, pp. 2725–2729.

Taylor, Nicholas M.I. et al. (2016). "Structure of the T4 baseplate and its function in triggering sheath contraction". In: *Nature* 533, pp. 346–352.

Trokter, Martina et al. (2014). "Recent advances in the structural and molecular biology of type IV secretion systems". In: *Current Opinion in Structural Biology* 27, pp. 16–23.

Wang, Jing et al. (2017). "Cryo-EM structure of the extended type VI secretion system sheath-tube complex". In: *Nature Microbiology* 2, pp. 1507–1512.

Wenren, Larissa M et al. (2013). "Two Independent Pathways for Self-Recognition in *Proteus mirabilis* Are Linked by Type VI-Dependent Export". In: *mBio* 4, e00374–13.

Whitmore, A (1912). "On the bacteriology of an infective disease occurring in Rangoon". In: *British Medical Journal* 13, pp. 1306–1308.

Winn, Martyn D. et al. (2011). "Overview of the *CCP*4 suite and current developments". In: *Acta Crystallographica Section D* 67, pp. 235–242.

Winter, G. (2010). "*xia2*: an expert system for macromolecular crystallography data reduction". In: *Journal of Applied Crystallography* 43, pp. 186–190.

Winter, Graeme and Katherine E. McAuley (2011). "Automated data collection for macromolecular crystallography". In: *Methods* 55, pp. 81–93.

Worrall, Liam J, Emilie Lameignere, and Natalie Cj Strynadka (2011). "Structural overview of the bacterial injectisome". In: *Current Opinion in Microbiology* 14, pp. 3–8.

Yamamoto, T (1967). "Presence of rhapidosomes in various species of bacteria and their morphological characteristics." In: *Journal of bacteriology* 94, pp. 1746–56.

Yang, G et al. (2006). "Photorhabdus virulence cassettes confer injectable insecticidal activity against the wax moth". In: *Journal of Bacteriology* 188, pp. 2254–2261.

Zhang, Heng et al. (2013). "Structure of the type vi effector-immunity complex (Tae4-Tai4) provides novel insights into the inhibition mechanism of the effector by its immunity protein". In: *Journal of Biological Chemistry* 288, pp. 5928–5939.

Zheng, Jun, Brian Ho, and John J Mekalanos (2011). "Genetic Analysis of Anti-Amoebae and Anti-Bacterial Activities of the Type VI Secretion System in *Vibrio cholerae*". In: *PLoS ONE* 6.

Zheng, Jun and Ka Yin Leung (2007). "Dissection of a type VI secretion system in *Edwardsiella tarda*". In: *Molecular Microbiology* 66, pp. 1192–1206.

Zhong, Dalian et al. (2012). "The *Salmonella* type III secretion system inner rod protein PrgJ is partially folded". In: *Journal of Biological Chemistry* 287, pp. 25303–25311.

Zoued, Abdelrahim et al. (2013). "TssK is a trimeric cytoplasmic protein interacting with components of both phage-like and membrane anchoring complexes of the type VI secretion system". In: *Journal of Biological Chemistry* 288, pp. 27031–27041.

Zoued, Abdelrahim et al. (2016). "Priming and polymerization of a bacterial contractile tail structure". In: *Nature* 531, pp. 59–63.

Zoued, Abdelrahim et al. (2017). "TssA: The cap protein of the Type VI secretion system tail". In: *BioEssays* 39, pp. 1–9.