

Investigation of the DNA uptake machinery in naturally competent bacteria

Augustinas Silale

Department of Biochemistry

and

University College

University of Oxford

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

Trinity Term 2019

40,000 words

Investigation of the DNA uptake machinery in naturally competent bacteria

Augustinas Silale

Department of Biochemistry and University College, University of Oxford

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

Trinity Term 2019

Abstract

Natural transformation is a mode of horizontal gene transfer in bacteria which involves uptake of naked DNA from the environment. The acquired DNA can be incorporated into the bacterial genome and confer useful traits, such as antibiotic resistance and increased virulence. The mechanistic details of how specialised transport proteins allow passage of DNA through the cell envelope remain elusive.

The membrane protein ComEC forms a channel for DNA transport into the cytoplasm but the molecular details of its structure and function are unknown. An extensive expression screen identified ComEC from *Moorella thermoacetica* as a candidate for structural and biochemical analysis. Low resolution structural data suggest that the two extramembranous domains of ComEC are located on the same side of the membrane and that they interact.

Biophysical and biochemical studies of the C-terminal soluble domain of ComEC showed that it has manganese(II)-dependent nuclease activity. It was shown that the nuclease activity of the C-terminal domain is required for efficient transformation in *B. subtilis*. It is likely that the C-terminal domain degrades the non-transforming strand during DNA uptake.

Attempts at isolating DNA uptake complexes for biochemical and structural analysis from competent *Bacillus subtilis* cells revealed that the competence machinery components are present at very low levels. No native DNA uptake complexes could be isolated.

ComFA is a cytoplasmic membrane-associated ATPase which is thought to power DNA uptake through the ComEC channel. A thermophilic homologue of ComFA was expressed, purified and shown to bind both single- and double-stranded DNA. ComFA was also shown to have ATPase activity. A unique N-terminal extension domain was identified in the thermophilic ComFA, but its function remains unknown.

Together these data contribute to our understanding of DNA uptake across the cytoplasmic membrane.

Acknowledgements

I would like to thank my supervisor Ben Berks for giving me the opportunity to carry out my DPhil research in his lab, for the guidance in pushing this difficult project forward, and for the help with writing this thesis. I would also like to thank my second supervisor Susan Lea for helpful discussions and for access to electron microscopy facilities.

Thanks to the various people who helped me with experiments or provided their services. Justin Deme and Susan Lea for EM data acquisition and analysis and advice on crystallisation trials; Errin Johnson for training in EM sample preparation and operating the T12 microscope; Sabrina Liberatori for the mass spectrometry experiments; Steve Johnson for the SEC-MALS experiments; David Staunton for the DSC experiments and for training in biophysical methods; Phill Stansfeld for bioinformatics insights.

Thank you to my colleagues in the Department of Biochemistry who made my time here very enjoyable. In particular, thanks to Felicity Alcock for the guidance when I started out my DPhil, for the helpful discussions and, most importantly, for the friendship; to my course mates Natalie Barber and Phil Bräuer for the fun times and for looking out for me; to Despoina Mavridou for the mentorship and encouragement; to Sam Hickman for the general support and help with applying for postdoc positions.

A special thanks to the Oxford University Company of Archers who made sure I did not lose my mind by constantly working. Also thanks for the opportunities to learn skills—not just archery—that I would not have learned otherwise.

Many thanks to the Wellcome Trust for generous funding.

Finally, I am grateful to my parents for the unwavering support throughout my education.

Table of Contents

Chapter 1 : Introduction	1
1.1. Horizontal gene transfer in bacteria.....	2
Conjugation	3
Transduction.....	3
Natural transformation	4
1.2. The clinical significance of natural transformation.....	5
1.3. The DNA uptake machinery of naturally competent bacteria	7
The current model for DNA transport	7
DNA binding by the competence pilus.....	9
DNA uptake preferences.....	10
DNA transport by the competence pilus.....	10
DNA at the outer surface of the cytoplasmic membrane	11
DNA transport through the cytoplasmic membrane	12
The fate of the DNA in the cytoplasm.....	12
Competence regulation.....	13
1.4. The role of ComEC in natural transformation.....	13
ComEC is most likely a DNA transporter.....	13
ComEC topology and domain arrangement.....	14
ComEC biogenesis	16
ComEC oligomeric state.....	16
ComEC involvement in other cellular processes	17
1.4. Unresolved issues in DNA transport across the inner membrane.....	18
The structure and function of ComEC	18
Conversion of dsDNA to ssDNA.....	20

The energy source for DNA transport	21
1.5. Project aims.....	22
Chapter 2 : Materials and methods	23
2.1. Culturing of bacteria.....	24
2.2. Molecular biology.....	30
Polymerase chain reactions	30
Genomic DNA sources.....	30
Molecular cloning methods	30
Restriction cloning	31
TA cloning.....	40
KLD reaction cloning	40
In-Fusion cloning	40
Gene synthesis.....	41
2.3. Construction of B. subtilis chromosomal mutant strains.....	41
Construction of plasmids for chromosomal integration.....	41
Transformation and mutant selection.....	43
2.4. B. subtilis transformation assays	43
2.5. Protein sample analysis.....	45
SDS-PAGE.....	45
Western blotting.....	46
Protein concentration determination	47
Mass spectrometry	48
2.6. Protein expression and purification	48
Expression screens	48
ComEC homologue expression screen.....	48

MthComEC truncation construct expression screen	49
Detergent screen	49
Purification of MthComEC	51
Purification of BsuComEC-TS	53
Purification of BsuComFC-TS	55
Purification of MthCTD and MthCTD-DADA	55
Purification of MthComFA	56
Purification of MthComFC	57
2.7. Electron microscopy	58
Grid preparation	59
Data collection	59
Single particle analysis.....	60
2.8. Biophysical methods	61
Thermal shift assays	61
Differential scanning calorimetry.....	62
Circular dichroism spectroscopy	63
2.9. Biochemical assays	64
Electrophoretic mobility shift assays.....	64
Phosphomonoesterase and phosphodiesterase assays	66
Nuclease assays	67
2.10. Crystallisation trials.....	67
<i>Chapter 3 : Production and electron microscopy studies of recombinant ComEC... 69</i>	
3.1. Introduction	70
3.2. Selection of ComEC homologues	71
3.3. Testing ComEC homologue expression in E. coli	75

3.4. Detergent screening	78
3.5. Developing a purification protocol for ComEC.....	78
Initial purification of MthComEC and KoxComEC.....	80
MthComEC co-purifies with the expression host Sec machinery.....	82
Removal of Sec contaminants.....	85
3.6. Mass spectrometry of MthComEC.....	86
3.7. N-terminus of MthComEC is cleaved	87
3.8. MthComEC SEC-MALLS analysis.....	89
3.9. MthComEC limited proteolysis	90
3.7. MthComEC N-terminal deletion constructs	92
3.9. Negative stain electron microscopy 3D reconstruction of MthComEC	94
Chemical crosslinking	96
MthComEC N-terminal truncation construct.....	98
3.10. Cryo-electron microscopy studies of MthComEC.....	99
Chemical crosslinking and removal of excess detergent	100
Amphipol trapping	101
3.10. Summary	105
<i>Chapter 4 : Expression and purification of ComEC from Bacillus subtilis</i>	<i>107</i>
4.1. Introduction	108
Competence regulation in Bacillus subtilis	108
Artificial induction of competence.....	110
Protease- and sporulation-deficient Bacillus subtilis	110
4.2. Construction of a supercompetent and protease-deficient Bacillus subtilis strain	111
4.3. Chromosomal tagging of competence proteins.....	114

4.4. Pull downs of ComEC-TwinStrep from supercompetent cells	117
4.5. Purification of ComFC-TwinStrep	121
4.7. Summary	123
Chapter 5 : Characterisation of the C-terminal domain of ComEC.....	126
5.1. Introduction	127
Degradation of the non-transforming DNA strand	127
The metallo- β -lactamase family of hydrolases	128
5.2. The metallo- β -lactamase domain of ComEC.....	130
5.3. Expression and purification of the C-terminal domain	133
MthCTD is not toxic to E. coli	135
5.4. Biophysical characterisation of MthCTD.....	136
Circular dichroism spectroscopy	136
Thermal shift assay.....	138
Differential scanning calorimetry.....	141
5.5. Attempts at MthCTD crystallisation	143
5.6. MthCTD activity assays.....	144
Assessment of DNA binding.....	144
Phosphoesterase assays	144
Nuclease assay.	149
5.7. Transformability of ComEC metal binding mutants in Bacillus subtilis	151
5.8. Attempts at producing MthNTD	153
5.9. Summary	153
Chapter 6 : Production and characterisation of ComFA and ComFC	155

6.1. Introduction	156
6.2. Bioinformatics analysis of MthComFA.....	158
6.3. Bioinformatics analysis of MthComFC.....	162
6.4. Production of recombinant MthComFA	163
6.5. MthComFA SEC-MALLS analysis	166
6.6. MthComFA DNA binding assay.....	167
6.7. MthComFA metal binding assay.....	169
6.8. MthComFA ATPase assay	171
6.9. MthComFA limited proteolysis.....	172
6.10. MthComFA crystallisation attempts.....	174
6.11. Production of recombinant MthComFC.....	174
6.12. Co-expression of MthComFA and MthComFC.....	175
6.13. Co-expression of MthComFA and MthComEC	176
6.13. Summary	178
Chapter 7 : Conclusions and future directions	181
7.1. Towards the structure of ComEC.....	182
7.2. Characterisation of DNA uptake complexes	183
7.3. The function of the C-terminal domain of ComEC.	184
Bibliography.....	185

List of Figures

Figure 1.1. Major modes of horizontal gene transfer.....	2
Figure 1.2. Current model for DNA uptake in <i>Bacillus subtilis</i>	8
Figure 1.3. Summary of current knowledge about ComEC and its soluble domains.	15
Figure 1.4. Topology predictions for ComEC from different organisms.	19
Figure 3.1. Expression vector and construct schematics	75
Figure 3.2. ComEC expression test.....	77
Figure 3.3. Detergent screen for MthComEC solubilization.....	79
Figure 3.4. ComEC purification workflow.....	80
Figure 3.5. Initial attempts at KoxComEC and MthComEC purification.....	81
Figure 3.6. Purification of MthComEC-sfGFP-His ₈	82
Figure 3.7. Purification of MthComEC-Strep.	84
Figure 3.8. MthComEC-sfGFP-TwinStrep purification with MgATP wash.....	86
Figure 3.9. Peptide coverage of MthComEC.	87
Figure 3.10. The N-terminus of MthComEC is cleaved	88
Figure 3.11. MthComEC SEC-MALLS analysis..	89
Figure 3.12. MthComEC limited proteolysis.....	91
Figure 3.13. Expression and purification of N-terminal MthComEC truncations.....	93
Figure 3.14. Negative stain EM of ComEC.....	95
Figure 3.15. GraFix procedure and negative stain EM of fixed MthComEC.....	97
Figure 3.16. Negative stain EM of truncated MthComEC.	98
Figure 3.17. Cryo-EM studies of MthComEC in LMNG.....	100
Figure 3.18. EM studies of MthComEC treated with the GraDeR technique.....	102

Figure 3.19. Cryo-EM studies of MthComEC in amphipols.	104
Figure 4.1. Regulation of the master competence transcription factor ComK in <i>Bacillus subtilis</i>	109
Figure 4.2. Supercompetent <i>Bacillus subtilis</i>	113
Figure 4.3. Chromosomal tagging of <i>B. subtilis</i> competence proteins..	116
Figure 4.4. Initial ComEC-TwinStrep pulldowns in DDM and LMNG.	119
Figure 4.5. Optimised pulldown of ComEC-TwinStrep.	120
Figure 4.6. Purification of ComFC-TwinStrep.	122
Figure 5.1. The active site of a binuclear MBL endonuclease..	130
Figure 5.2. Alignment of ComEC CTD sequences.....	132
Figure 5.3. MthCTD expression constructs.....	134
Figure 5.4. Purification of MthCTD and MthCTD-DADA.....	135
Figure 5.5. MthCTD secondary structure analysis using circular dichroism spectroscopy.....	138
Figure 5.6. MthCTD metal binding assay.....	140
Figure 5.7. Differential scanning calorimetry of MthCTD.	142
Figure 5.8. EMSA attempts with MthCTD	144
Figure 5.9. DNA, bpNPP and pNPP hydrolysis..	146
Figure 5.10. Phosphoesterase assays.....	147
Figure 5.11. MthCTD nuclease assay.	150
Figure 5.12. Transformation assay of a <i>B. subtilis</i> ComEC metal ion binding mutant	152
Figure 6.1. Organisation of the <i>comF</i> operon in different organisms.	159
Figure 6.2. ComFA bioinformatics analysis.	161

Figure 6.3. Bioinformatics analysis of ComFC.....	163
Figure 6.4. Purification of MthComFA.....	165
Figure 6.5. SEC-MALLS analysis of MthComFA.....	167
Figure 6.6. MthComFA DNA binding assay.....	168
Figure 6.7. Purification of MthComFA in the presence of Zn ²⁺	169
Figure 6.8. MthComFA metal binding analysis.....	171
Figure 6.9. MthComFA ATPase assay.....	172
Figure 6.10. The domain structure of MthComFA analysed by limited proteolysis.	173
Figure 6.11. MthComFC NiNTA purification.....	175
Figure 6.12. MthComFA and MthComFC co-expression.....	176
Figure 6.13. MthComEC and MthComFA co-expression.....	178

List of Tables

Table 1.1. World Health Organisation priority pathogens list for new antibiotics..	6
Table 2.1. Composition of the growth media used in this study.....	24
Table 2.2. Antibiotics used in this study.....	25
Table 2.3. <i>E. coli</i> strains used in this study.....	26
Table 2.4. <i>B. subtilis</i> strains used in this study.	28
Table 2.5. Plasmids constructed in earlier studies.	32
Table 2.6. Plasmids constructed during this project.....	33
Table 2.7. Primers used in this study.	37
Table 2.8. Buffers used for SDS-PAGE.	45
Table 2.9. List of antibodies.....	47
Table 2.10. Detergents used for screening..	50
Table 2.11. Recipes for EMSAs.....	66
Table 3.1. List of organisms from which ComEC homologues were cloned.	73

List of Abbreviations

APol	Amphipol
ATP	Adenosine triphosphate
bpNPP	Bis(para-nitrophenyl) phosphate
C₁₂E₉	Polyoxyethylene-9 dodecyl ether
CD	Circular dichroism
CFU	Colony forming unit
CTD	C-terminal domain
CTF	Contrast transfer function
Cymal-6-NG	2,2-bis(3'-cyclohexylbutyl) propane-1,3-bis- β -D-maltopyranoside
DDM	n-Dodecyl- β -D-Maltopyranoside
DM	n-Decyl- β -D-Maltopyranoside
DMNG	Decyl maltose neopentyl glycol
DSC	Dynamic scanning calorimetry
dsDNA	Double-stranded DNA
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
GDN	Glyco-diosgenin
gDNA	Genomic DNA
GFP	Green fluorescent protein
GraDeR	Gradient Detergent Removal
GraFix	Gradient Fixation

HRP	Horseradish peroxidase
IPTG	β -d-1-thiogalactopyranoside
KLD	Kinase-ligase-DpnI
LB	Lysogeny broth
LDAO	Lauryldimethylamine-N-oxide
LMNG	Lauryl maltose neopentyl glycol
MALLS	Multi-angle laser light scattering
MBL	Metallo-beta-lactamase
MCS	Multiple cloning site
NTD	N-terminal domain
OB	Oligonucleotide binding
OG	n-Octyl- β -D-Glucopyranoside
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PMF	Proton motive force
PMSF	Phenylmethylsulfonyl fluoride
pNPP	Para-nitrophenyl phosphate
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
sfGFP	Superfolder green fluorescent protein
ssDNA	Single-stranded DNA
T4P	Type IV pilus
T4SS	Type IV secretion system

TB	Terrific broth
TBS	Tris-buffered saline
tDNA	Transforming DNA
TEV	Tobacco etch virus
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside

Chapter 1: Introduction

1.1. Horizontal gene transfer in bacteria

Bacteria are the most ubiquitous and adaptable organisms on the planet. They can be found in almost any environmental niche—from boiling-hot hydrothermal vents to our gastrointestinal tracts to glaciers. The combination of short generation times, relatively high mutation rates, and possession of mechanisms for acquiring novel DNA sequences enables bacteria to rapidly adapt to changes in their environment (Wiedenbeck & Cohan, 2011). The mechanisms for acquisition of DNA in a non-parent-daughter cell relationship are collectively called horizontal gene transfer. The main three modes of horizontal gene transfer are conjugation, transduction and natural transformation (Figure 1.1).

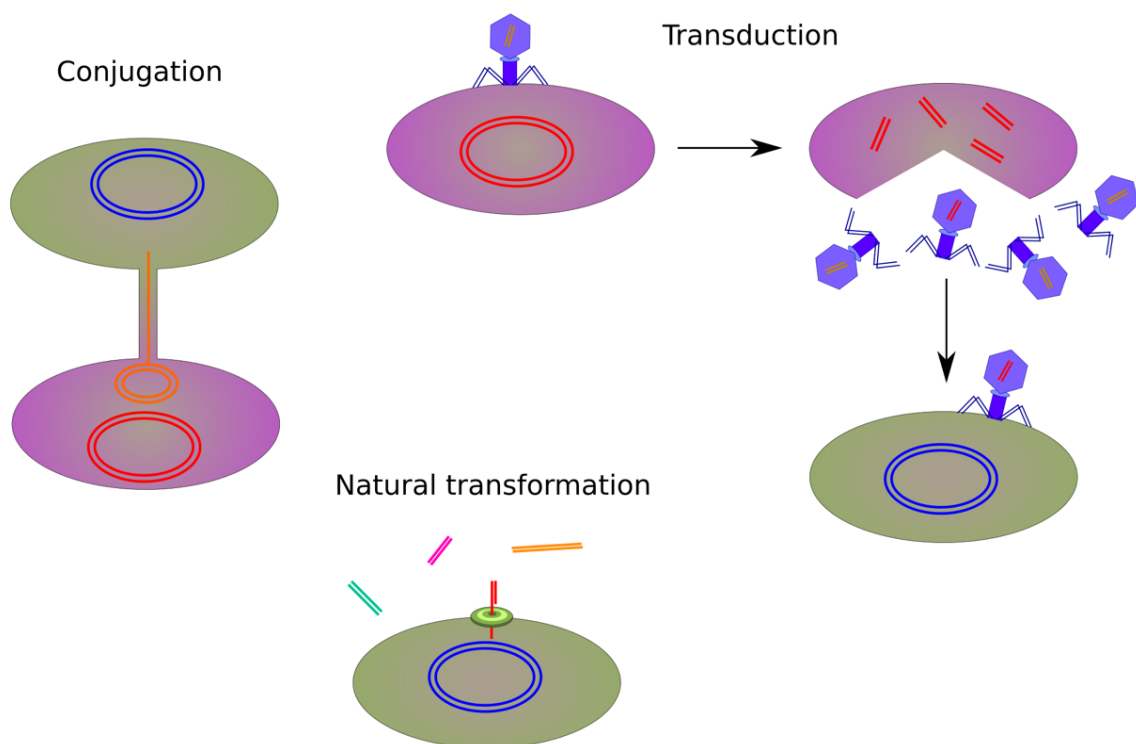


Figure 1.1. Major modes of horizontal gene transfer. Conjugation involves formation of a physical bridge between two bacteria, through which the donor plasmid (orange) is passed to the recipient cell (blue DNA). Transduction is transfer of bacterial DNA via bacteriophages. When a bacterial cell is infected, the virus (purple) hacks the cellular machinery to replicate itself. When new virion particles are packaged, sometimes pieces of the host cell DNA (red) are incorporated as well. Bacteriophages can then transfer this DNA to other cells they infect. Natural transformation is the process of taking up naked DNA from the environment via specialised machinery in the cell envelope (green).

Conjugation

Conjugative plasmids, such as the F plasmid, encode the machinery required to transfer copies of themselves from their host to other bacteria (Waksman, 2019). The DNA is prepared for transfer by the 'relaxosome', which nicks the DNA, initiates rolling circle replication and guides the resulting single stranded DNA (ssDNA) to the export machinery. A type IV secretion system (T4SS) encoded by the mobile plasmid forms a conjugation pilus between the donor and the recipient cell. The conjugation pilus is a hollow tube composed of a helical assembly of protein-phospholipid units. The ssDNA is transferred through the lumen of this tube (Costa *et al*, 2016). Cytoplasmic ATPases associated with the T4SS of the donor cell drive directional DNA transport into the recipient cell (Waksman, 2019). Conjugative plasmids may confer useful traits, such as antibiotic resistance genes.

Transduction

Bacteriophages can be agents of horizontal gene transfer. When a tailed double stranded DNA (dsDNA) phage binds to a suitable receptor on the cell surface, structural rearrangements in the virion result in the tail punching through the cell membrane. The viral genome is ejected through the lumen of the tail (Wang *et al*, 2019). In many tailed phages the DNA inside the capsid is so tightly packed by a packaging motor during phage assembly that the pressure built up inside the capsid pushes out the entire genome through the tail and into the host (Molineux & Panja, 2013). Other phages eject their genome partially and the host transcription machinery pulls in the rest of the genome from the capsid via the tail. Phages subvert the host biosynthetic machinery to produce more copies of viral DNA. This newly synthesised viral DNA is packaged into virions by an

Chapter 1

ATPase packaging nanomotor, the terminase (Molineux & Panja, 2013). DNA encoding parts of the host genome can be incorporated into virions (Chiang *et al*, 2019). Generalised transduction occurs when the terminase cuts newly synthesised DNA at specific sites for packaging into virions. Sometimes sites similar to those used for cutting the viral genome are recognised on the host chromosome by the terminase and host DNA is erroneously incorporated into a virion. In specialised transduction, phages insert themselves into the host genome at specific sites by homologous recombination and can become dormant (form lysogens). Certain stimuli, such as cell stress, reactivate the phage which excises itself from the host genome and during this process can also pick up an adjacent piece of host DNA. All newly synthesised virions will carry this piece of DNA in their genome. Transducing particles formed by either mode of transduction can spread the bacterial DNA that they carry by infecting other cells which in turn can incorporate the foreign DNA into their genome by homologous recombination

Natural transformation

Natural transformation is the process of taking up naked DNA from the environment. It does not require cell-to-cell contact or intermediary agents such as phages but does involve specialised cellular machinery for DNA uptake. Bacteria that express this machinery are said to be in a state of 'competence' for DNA uptake – hence, the terms 'natural transformation' and 'natural competence' will be used interchangeably. Although only about 80 bacterial species have been experimentally shown to be naturally competent, homologues of core competence genes are found widely across the bacterial phyla in both Gram-negative and Gram-positive bacteria (Johnston *et al*, 2014; Pimentel & Zhang, 2018).

It has been proposed that bacteria break down the imported DNA into nucleotides, which are costly to synthesise *de novo*, and, therefore, the DNA is used as a source of nutrients (Redfield, 1993). This is unlikely to be the main purpose of natural transformation because the multiprotein transport machinery is also costly to produce. In some species DNA uptake is sequence-specific to limit the incoming DNA to that from closely related strains. This sequence similarity enables homologous recombination to repair a damaged genome or to excise costly mobile genetic elements (Croucher *et al*, 2016; Blokesch, 2017). Some organisms co-upregulate competence with predation mechanisms. For example, *Vibrio cholerae* co-expresses the type VI secretion system, which acts as a molecular harpoon that delivers lethal toxins into adjacent bacteria (Borgeaud *et al*, 2015). Competent *Streptococcus pneumoniae* co-expresses fratricide toxins and their cognate immunity proteins, resulting in killing of both non-competent kin cells and closely related cells (Claverys & Håvarstein, 2007; Johnsborg *et al*, 2008). Lysis of neighbouring cells increases the DNA pool available for uptake. Killing of 'foreign' cells in particular provides access to novel DNA sequences. The aforementioned properties of naturally competent bacteria, together with the knowledge that imported DNA can be incorporated into the host genome, firmly ground natural transformation as a major mechanism of horizontal gene transfer.

1.2. The clinical significance of natural transformation

The discovery of natural transformation by Frederick Griffith in the 1920s concurrently proved its clinical relevance. A non-pathogenic pneumococcal strain lacking a polysaccharide capsule was transformed into a pathogenic strain by co-inoculation of mice with heat-killed pathogenic bacteria (Griffith, 1928). This illustrates that natural

transformation is important for the spread of virulence traits and can also facilitate the spread of antibiotic resistance genes (Chi *et al*, 2007; Domingues *et al*, 2012). Indeed, the World Health Organisation priority pathogen list for new antibiotics includes several naturally competent bacteria (Table 1.1).

Table 1.1. World Health Organisation priority pathogens list for new antibiotics. Species shown to be naturally competent are in red. Adapted from the WHO report “Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics” (2017).

Priority 1: Critical
<i>Acinetobacter baumannii</i> <i>Pseudomonas aeruginosa</i> <i>Enterobacteriaceae</i> (inc. <i>Klebsiella pneumoniae</i> , <i>Escherichia coli</i> , <i>Enterobacter</i> spp., <i>Serratia</i> spp., <i>Proteus</i> spp., <i>Providencia</i> spp, <i>Morganella</i> spp.)
Priority 2: High
<i>Enterococcus faecium</i> <i>Staphylococcus aureus</i> <i>Helicobacter pylori</i> <i>Campylobacter</i> spp. <i>Salmonella</i> spp. <i>Neisseria gonorrhoeae</i>
Priority 3: Medium
<i>Streptococcus pneumoniae</i> <i>Haemophilus influenzae</i> <i>Shigella</i> spp.

1.3. The DNA uptake machinery of naturally competent bacteria

In spite of how widespread natural transformation is among bacteria, how long ago it was discovered and the important role it plays in the antibiotic resistance crisis, the molecular mechanisms of DNA uptake remain elusive. The cell envelope poses significant hurdles for DNA import. In Gram-positive bacteria, the DNA has to cross a peptidoglycan layer of around 33 nm thickness (Matias & Beveridge, 2005), which corresponds to around 10 helical turns of B-form DNA, or over 100 base pairs. After that, the DNA must cross the cytoplasmic membrane. Although the peptidoglycan layer is much thinner in Gram negative bacteria, an outer membrane must be crossed in addition to the inner membrane. DNA is a long, polyanionic species and so its diffusion through peptidoglycan is sterically hindered and its movement across membranes thermodynamically highly unfavourable. Consequently, specialised machinery is required to allow passage of DNA across the cell envelope.

The current model for DNA transport

The current understanding of DNA uptake in natural transformation is derived from studies of numerous organisms (Dubnau & Blokesch, 2019). *Bacillus subtilis* is one of the best studied models for natural transformation. Most core competence genes have been first identified in *B. subtilis* (Albano *et al*, 1987; Hahn *et al*, 1987). This overview will draw on studies from the various naturally competent organisms but focus on the *B. subtilis* system and the *B. subtilis* competence gene nomenclature will be used throughout this work. A summary of what is currently known about the system is shown in Figure 1.2.

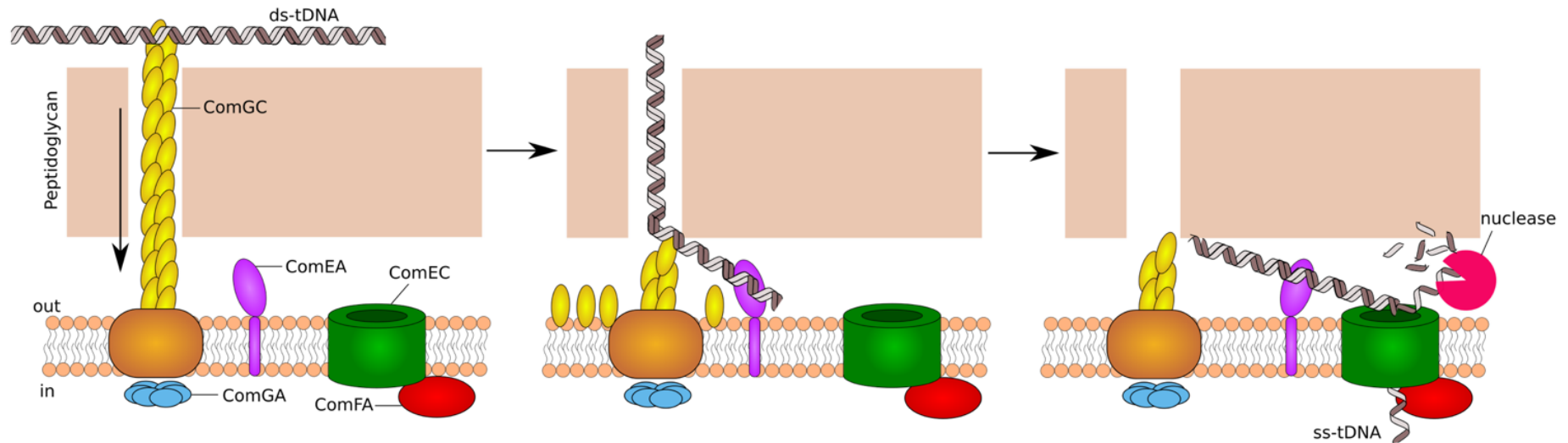


Figure 1.2. Current model for DNA uptake in *Bacillus subtilis*. Double-stranded transforming DNA (tDNA) is engaged by the competence pilus, composed of the major pilin, ComGC, and minor pilins (not shown). Pilus extension is powered by the ComGA assembly ATPase. Retraction of the pilus via depolymerisation of the basal end pulls the bound tDNA through the thick peptidoglycan layer. tDNA is bound by the membrane-anchored dsDNA binding protein ComEA. Binding to ComEA may trap the DNA to facilitate its transport by a Brownian ratchet mechanism. One strand of the tDNA is degraded by an unknown nuclease, and the other strand is transported across the cell membrane into the cytoplasm via the ComEC channel. This step may be powered by the ComFA ATPase.

DNA binding by the competence pilus

The first step in DNA uptake is binding of DNA to the cell surface and transport through the peptidoglycan layer (or the outer membrane in Gram-negative organisms) towards the cytoplasmic membrane. This is performed by competence pili, which resemble type IV pili (T4P) and the related type II secretion system components (Dubnau & Blokesch, 2019). The only known exception is *Helicobacter pylori*, which uses a T4SS to import DNA (Hofreuter *et al*, 2001). The *comGABCDEFGG* operon in *B. subtilis* encodes the pilus components required for DNA binding and transport through the peptidoglycan (Chung & Dubnau, 1998; Chen *et al*, 2006; Briley *et al*, 2011). ComGC is the major pilin (Chen *et al*, 2006). ComGD, ComGE and ComGG are minor pilins that are also required for pilus assembly, though their specific functions are unknown (Chung & Dubnau, 1998). The pilin genes encode prepilin presursors which must be processed by the prepilin peptidase ComC to produce mature pilins capable of pilus assembly (Chung & Dubnau, 1995, 1998). ComGA and ComGB resemble accessory proteins that are required for T4P assembly in *Pseudomonas aeruginosa* (Nunn *et al*, 1990). ComGB is a conserved transmembrane protein. ComGA is the assembly ATPase that powers extraction of pilins from the membrane so that they can assemble into a pilus structure (Chen *et al*, 2006; Briley *et al*, 2011). Protruding competence pilus structures up to 2-3 μm in length have been observed in *Streptococcus pneumoniae* (Laurenceau *et al*, 2013), but not *B. subtilis*. It is possible that in *B. subtilis* the *comG* locus encodes a 'pseudopilus' that is only just long enough to span the peptidoglycan layer (Chen *et al*, 2006). There are no structures available of competence pili, but a recent cryo-electron tomography study of a twitching motility pilus in *Myxococcus xanthus* sheds light on how the machinery that forms T4P is organised (Chang *et al*, 2016). Proteinaceous multimeric rings in the inner membrane, the

peptidoglycan layer and the outer membrane form a path for pilus secretion through the cell envelope. Cytoplasmic ATPases that power pilus assembly and retraction can dock to the base of the structure.

DNA uptake preferences

The Gram-negative bacteria *Haemophilus influenzae* and *Neisseria* spp. only take up DNA that has specific short DNA uptake sequences (Danner *et al*, 1980; Goodman & Scoocca, 1988). In *Neisseria* spp. binding specificity is conferred by the ComP minor pilin (Berry *et al*, 2013; Cehovin *et al*, 2013; Berry *et al*, 2016). These bacteria discriminate the DNA sequences they take up in order to protect themselves from deleterious sequences, such as mobile genetic elements. DNA binding by T4P is not sequence-specific in most naturally competent bacteria, including *B. subtilis*.

It is assumed that the competence pilus usually binds and transports dsDNA. There is evidence *S. pneumoniae*, *Pseudomonas stutzeri* and *N. gonorrhoeae* can also be transformed by ssDNA (Barany & Boeke, 1983; Meier *et al*, 2002; Hepp *et al*, 2016). However, transformation with ssDNA is 30-60-fold less efficient in *S. pneumoniae* and *P. stutzeri*, and in *N. gonorrhoeae* a double-stranded specific DNA uptake sequence must be present within the ssDNA.

DNA transport by the competence pilus

Once the DNA is bound by the competence pilus, it must be brought to the proximity of the cytoplasmic membrane (Figure 1.2). A recent *V. cholerae* microscopy study showed that fluorescently labelled DNA binds to the tip of the pilus and the pilus retracts, pulling the DNA towards the cell (Ellison *et al*, 2018). In Gram-negative bacteria, the assembly ATPase PilB is thought to be exchanged for a retraction ATPase, PilT, which powers

disassembly of the pilus by extracting pilins back into the membrane resulting in pilus retraction. However, no such retraction ATPase has been identified in Gram-positive bacteria. It might be the case that ComGA is bifunctional and can power pilus retraction as well as assembly. Interestingly, recent studies have shown T4P retraction can still occur in Gram-negative bacteria with deleted retraction ATPases, though at greatly reduced rate and force (Ellison *et al*, 2018; Zöllner *et al*, 2019). It is possible that in Gram-positive bacteria DNA is pulled in by the likely thermodynamically favourable disassembly of the pilus subunits back into the cytoplasmic membrane.

DNA at the outer surface of the cytoplasmic membrane

Following uptake across the cell wall, in *B. subtilis* the incoming DNA is trapped in the proximity of the cytoplasmic membrane by the membrane-anchored dsDNA binding protein ComEA (Inamine & Dubnau, 1995; Provvedi & Dubnau, 1999) (Figure 1.2). Fluorescently labelled ComEA forms foci in the presence of DNA under competence conditions (Ellison *et al*, 2018). This step is critical, as ComEA deletion mutants are non-transformable (Inamine & Dubnau, 1995; Provvedi & Dubnau, 1999; Seitz & Blokesch, 2013). In Gram-negative cells, ComEA provides a Brownian ratchet mechanism by preventing DNA from slipping back out of the periplasm through the outer membrane T4P secretin pore (Hepp & Maier, 2016). Furthermore, in Gram-negative bacteria T4P-mediated DNA uptake and transport through the inner membrane is temporally uncoupled and DNA-ComEA complexes can persist in the periplasm (Gangel *et al*, 2014). It is unclear whether this is the case in *B. subtilis*.

DNA transport through the cytoplasmic membrane

At some stage, one DNA strand (the 'non-transforming strand') is degraded (Mejean & Claverys, 1993) and the transforming strand passes through the membrane into the cytoplasm, most likely via ComEC (Figure 1.2) (Draskovic & Dubnau, 2005; Dubnau & Blokesch, 2019). It is unclear how the transport across the membrane is initiated or energised. In Gram-positive bacteria the movement of DNA through ComEC may be powered by the cytoplasmic ATPase ComFA (Londoño-Vallejo & Dubnau, 1994b, 1994a; Chilton *et al*, 2017; Diallo *et al*, 2017). No such ATPase has been confirmed in Gram-negative bacteria.

The fate of the DNA in the cytoplasm

Cytoplasmic ssDNA binding proteins associated with the DNA transport machinery, SsbB and DprA, bind to the incoming DNA and promote loading of the RecA ATPase that catalyses hybridisation of the transforming strand with homologous regions in the host genome (Kidane & Graumann, 2005; Kramer *et al*, 2007; Yadav *et al*, 2014). The transforming strand can then be incorporated into the genome by homologous recombination. DNA that does not have homology to the host genome presumably gets degraded to release nucleotides to be used as nutrients. It is likely that ssDNA, rather than dsDNA, is imported across the cytoplasmic membrane because it is a more efficient substrate for homologous recombination (Dubnau & Blokesch, 2019). Additionally, the cytoplasmic ssDNA binding proteins may aid in DNA transport by providing a Brownian ratchet mechanism that prevents the DNA strand from slipping back out through the ComEC transport channel.

Competence regulation

In contrast to the other major mechanisms of horizontal gene transfer, natural transformation is an active decision undertaken by the bacterium. Whereas during conjugation or transduction the bacterium has no choice whether to accept the incoming DNA or not, the bacterium chooses when to express the DNA uptake machinery required for natural transformation. Although the core competence proteins are similar in all bacteria, conditions for competence induction vary considerably between different species (Johnston *et al*, 2014). *S. pneumoniae* becomes competent in early exponential phase. Expression of the alternative sigma factor ComX is induced by the competence-stimulating peptide of streptococci (Peterson *et al*, 2004). ComX then induces global changes in gene expression, including upregulation of DNA uptake proteins. *B. subtilis* competence is upregulated in stationary phase during nutrient starvation. Environmental signals in *B. subtilis* converge to activate a master transcription factor, ComK, which upregulates expression of proteins that form the DNA uptake machinery, among over 100 other proteins (Turgay *et al*, 1998; Berka *et al*, 2002). It is likely that the competence machinery has remained similar to that of a common ancestor, while regulation mechanisms have diverged in order to fit the lifestyle of individual species.

1.4. The role of ComEC in natural transformation

ComEC is most likely a DNA transporter

ComEC is present in all naturally competent bacteria (Johnston *et al*, 2014). Bioinformatics predictions suggest that it is a polytopic membrane protein with 9-13 transmembrane segments and two large hydrophilic loops. Fluorescence microscopy indicates that in *B. subtilis* it is found mostly at the cell poles and co-localises with the

competence proteins ComGA and ComFA (Kaufenstein *et al*, 2011). The stability of ComFA is dependent on ComEC expression (Kramer *et al*, 2007). Cells lacking ComEC can still bind DNA, and the DNA becomes resistant to DNase I because the initial steps of DNA uptake mediated by the competence pilus are not affected in *comEC* mutants (Inamine & Dubnau, 1995; Stingl *et al*, 2010; Seitz & Blokesch, 2013). Fluorescently labelled DNA can only be internalised into the cytoplasm in the presence of ComEC (Corbinais *et al*, 2016). These data strongly suggest that ComEC is the inner membrane transporter that allows passage of the transforming DNA strand into the cytoplasm.

ComEC topology and domain arrangement

There are very limited biochemical data available on ComEC. A summary of the current knowledge is shown in Figure 1.3A. The *B. subtilis* homologue (BsuComEC) has been the most extensively studied. Fusions of *lacZ* and *phoA* were used to investigate the topology of BsuComEC (Draskovic & Dubnau, 2005). The study found 7-8 transmembrane regions, which are fewer than predicted by bioinformatics methods, and two large extramembranous loops that are likely located on the outside of the cell. The central 5-10 transmembrane helices form a conserved 'competence' domain that may form the channel for DNA transport and contains a two-histidine motif of unknown function (LaBrie *et al*, 2019). The N-terminal transmembrane helices and the N-terminal loop are annotated in Pfam as 'domain of unknown function 4131' (DUF4131) and the C-terminal domain (CTD) as a metallo- β -lactamase (MBL) domain (El-Gebali *et al*, 2019). A bioinformatics study used homology modelling to investigate the BsuComEC loops and also found that the CTD has an MBL fold and discovered that the soluble part of DUF4131 (from now on referred to as the NTD) takes on an oligonucleotide binding (OB) fold which

could bind ssDNA (Baker *et al*, 2016). Indeed, most ComEC homologue soluble loops are predicted to fold into OB and MBL domains by homology modelling in Phyre2 (Figure 1.3B).

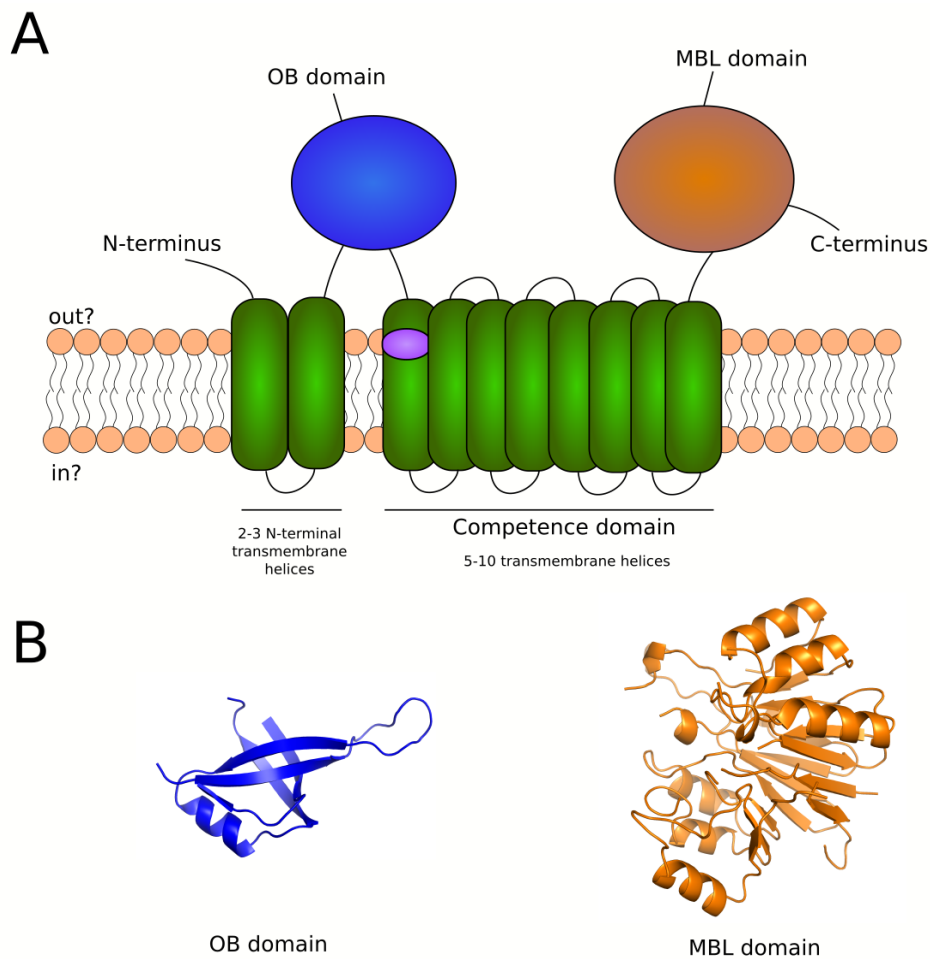


Figure 1.3. Summary of current knowledge about ComEC and its soluble domains. (A) Schematic representation of ComEC. It should be noted that the number of transmembrane helices has not been conclusively established for BsuComEC, and likely differs for different homologues. The location of the conserved two-histidine motif (H ϕ XX ϕ SG ϕ H, where ϕ is a hydrophobic residue) is indicated in purple. (B) Residues 99-167 corresponding to part of the NTD of *Moorella thermoacetica* ComEC were modelled with 92.7% confidence based on a fragment of the crystal structure of *Staphylococcus aureus* SsbA (PDB 5XGT). Residues 545-802 corresponding to the CTD of *M. thermoacetica* ComEC were modelled with 100% confidence based on part of the crystal structure of *S. pneumoniae* CbpE (PDB 2BIB). Phyre2 was used to make the homology models (Kelley *et al*, 2015). OB – oligonucleotide-binding; MBL – metallo- β -lactamase.

ComEC biogenesis

ComEC does not have a signal peptide. Two or three extreme-N-terminal transmembrane helices likely direct it to the Sec translocon for membrane insertion (Tsirigotaki *et al*, 2016) (Figure 1.3). Interestingly, BsuComEC requires the competence-induced oxidoreductase pair BdbC and BdbD for the formation of a critical intramolecular disulphide bond within the NTD (Draskovic & Dubnau, 2005). However, the cysteine residues involved are not conserved in all homologues so this is likely species-specific.

ComEC oligomeric state

ComEC contains 600-800 amino acid residues. Considering that the width of a dsDNA helix is around 20 Å, transport of ssDNA would require a channel that is up to 10 Å in diameter, though likely smaller. It is unclear if a monomer of ComEC would be of sufficient size for ssDNA transport. Conjugation pili also transport ssDNA. The cryo-electron microscopy structure of the pED208 pilus revealed an internal lumen with a diameter of 28 Å (Costa *et al*, 2016). However, it is known that during conjugation ssDNA is transported as a nucleoprotein complex (Ilangoan *et al*, 2015). Therefore, the conjugation pilus lumen is larger than would be required for ssDNA alone. The Sec translocon forms a hydrophilic pore that allows passage of an unfolded polypeptide chain (Tsirigotaki *et al*, 2016), which may be considered as an approximation of ssDNA. The translocon consists of the SecYEG heterotrimer (Berg *et al*, 2004). SecY has ten transmembrane helices and forms the translocation channel pore. In the resting state, the pore has an hourglass shape with a 5 Å constriction formed by hydrophobic residues. The conformation of SecY is stabilised by SecE which binds externally to the transmembrane helices of SecY. The pore is gated by a loop of SecG on the cytoplasmic

side and by a helical plug on the periplasmic side. These features allow formation of a pore large enough to conduct a polypeptide chain while maintaining the membrane permeability barrier. Thus, it is possible that a ComEC monomer contains enough transmembrane helices to form a stable aqueous pore for DNA transport.

There are weak experimental data suggesting that ComEC may form dimers and potentially higher order oligomers. When isolated membranes containing BsuComEC were incubated at 37°C, formation of higher molecular weight species was observed in anti-ComEC Western blots (Draskovic & Dubnau, 2005). Oligomerisation was solely mediated by four cysteine residues within the C-terminal part of the competence domain. These cysteine residues are not conserved and it remains unclear whether the oligomerisation is a species-specific observation and whether it is functionally important.

ComEC involvement in other cellular processes

Gene transfer agents are domesticated phage-like elements that package only host DNA, are incapable of replication and contribute to horizontal gene transfer in some bacteria. ComEC, together with cytoplasmic competence proteins, has been shown to be essential for gene transfer agent-mediated DNA acquisition in *Rhodobacter capsulatus* and *Bartonella* (Brimacombe *et al*, 2015; Québatte *et al*, 2017). The authors of the two studies suggest that DNA released into the periplasm by gene transfer agents is transported across the cytoplasmic membrane by ComEC. This observation is consistent with the role of ComEC in natural transformation.

There is at least one reported case of competence genes being involved in processes unrelated to DNA uptake. Rabinovich *et al* (2012) found that competence proteins are

upregulated in *Listeria monocytogenes* during intracellular growth in macrophages. This is achieved by controlled phage excision from the *comK* locus, which controls expression of late competence genes. The *comG* operon and *comEC* were required for optimal intracellular growth, indicating that the competence pilus and the cytoplasmic membrane DNA transporter perform yet unknown functions.

1.4. Unresolved issues in DNA transport across the inner membrane

The structure and function of ComEC

As there is no structure of ComEC, many details regarding its topology and function remain unclear. The topological model derived from reporter protein fusions (Draskovic & Dubnau, 2005) suggests that the final transmembrane segment exists within the predicted MBL fold, which seems improbable. Furthermore, topology predictions for different homologues suggest inconsistent assignment to compartments of the two soluble domains due to the different numbers of transmembrane segments predicted to separate them, even though the whole intervening sequence is conserved (Figure 1.4). An unpublished report claims that the NTD and CTD interact when co-expressed in *E. coli* (Dubnau & Blokesch, 2019). This implies that the two domains should be on the same side of the membrane in all organisms. Some ComEC homologues lack one or both extramembranous domains (Pimentel & Zhang, 2018), which means that they are either encoded elsewhere in the genome or that other proteins are capable of fulfilling their function. While it would make sense for both the NTD and the CTD to be on the extracellular side of the membrane where they can interact and engage the incoming DNA, this has not been conclusively shown and the function of the two domains remains

Chapter 1

unclear. Solving the topology problem and elucidating the biochemical functions of the two extramembranous domains is critical to understanding how ComEC works.

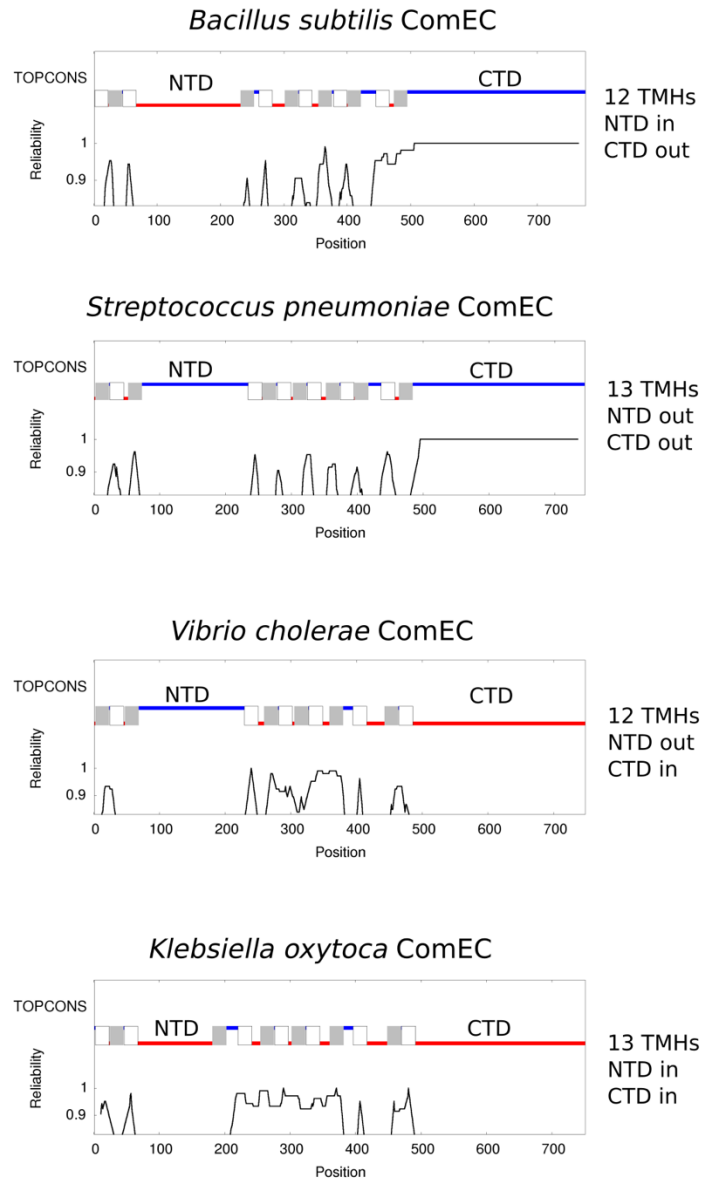


Figure 1.4. Topology predictions for ComEC from different organisms. The vertical axis shows the reliability score for each amino acid position indicated on the horizontal axis. A summary of the prediction for each ComEC sequence is shown on the right. Predicted outside-to-inside transmembrane segments are shown as white rectangles and inside-to-outside segments by grey rectangles. Blue – outside, red – inside. TMH – transmembrane helix. Topology predictions were done in TOPCONS (Tsirigos *et al*, 2015).

A recombinant ComEC production system would facilitate complete characterisation of ComEC. However, attempts at producing ComEC from various species in *E. coli* expression systems have been reported to be unsuccessful (Inamine & Dubnau, 1995; Yeh *et al*, 2003; Diallo *et al*, 2017). This may be because ComEC is stabilised by other competence proteins. Alternatively, as it has been estimated that competent *B. subtilis* has only around 200 ‘dimers’ of ComEC per cell (unpublished; reported by Hahn *et al*, 2005), ComEC may have structural features that impede efficient biosynthesis at high levels of expression. Purification of native ComEC from competent cells is also challenging due to low levels of expression which makes it difficult to isolate enough material for biochemical analysis. Therefore, the complexes that transport DNA through the cytoplasmic membrane have not been characterised by pulldown experiments and are not well understood.

Conversion of dsDNA to ssDNA.

It is known that only ssDNA reaches the cytoplasm during natural transformation (Mejean & Claverys, 1993; Dubnau & Blokesch, 2019). The membrane-bound extracytoplasmic nuclease EndA is required for transformation in *S. pneumoniae*. EndA has been shown to degrade the non-transforming strand with 5′-to-3′ polarity and the transforming strand enters the cytoplasm with 3′-to-5′ polarity (Puyet *et al*, 1990; Mejean & Claverys, 1993; Berge *et al*, 2002; Bergé *et al*, 2013). No equivalent nuclease involved in non-transforming strand degradation has been identified in other organisms. NucA in *B. subtilis* fragments the incoming DNA into smaller pieces but does not degrade the non-transforming strand (Provvedi *et al*, 2001). It has been shown that the MBL proteins are hydrolases, and many are known to act as nucleases (Dominski, 2007). Therefore, it has

been proposed that the CTD of ComEC may perform the conversion of double-to-single-stranded DNA (Baker *et al*, 2016; Dubnau & Blokesch, 2019), though this has never been shown experimentally.

The energy source for DNA transport

It is unclear how DNA transport via ComEC is energised. DNA is a long, highly negatively charged polymer that must be transported against a negative membrane potential. In *B. subtilis* it has been shown that DNA is pulled in by a powerful molecular motor that transports DNA at a rate of 80 bp s⁻¹ and that is capable of working against external forces of up to 50 pN (Maier *et al*, 2004). It has been shown that the proton motive force (PMF) is required for DNA uptake in bulk transformation experiments (van Nieuwenhoven *et al*, 1982). A study measuring individual DNA uptake events using optical tweezers found that the addition of protonophores severely impeded the rate of DNA transport, and that this effect could not be attributed to reduced cellular ATP levels (Maier *et al*, 2004). These studies, however, do not elucidate whether pilus retraction or transport via ComEC is the PMF requiring step. Other data indicate that ATP is required for DNA uptake across the inner membrane. Gram-positive bacteria have a cytoplasmic membrane-associated ATPase, ComFA, which has the signature motifs of a DEAD-box helicase/translocase and is required for efficient transformation in *S. pneumoniae* and *B. subtilis* (Londoño-Vallejo & Dubnau, 1994a, 1994b; Diallo *et al*, 2017; Chilton *et al*, 2017). The ATPase activity of ComFA is stimulated by ssDNA, suggesting that ComFA reels in DNA through the ComEC pore using its ATP-dependent translocase activity (Diallo *et al*, 2017). There are two issues with this hypothesis. Firstly, transformation can still occur in *B. subtilis* *comFA* mutants, albeit at low efficiency. Secondly, no clear functional analogue of

ComFA has been found in Gram-negative bacteria. The only ComFA homologue in Gram-negative bacteria is the DEAD-box helicase primosomal assembly protein, PriA (Matthey & Blokesch, 2016). The homology between ComFA and PriA mostly extends to the conserved DEAD-box helicase motifs, and PriA has never been shown to directly interact with the transformation machinery. Therefore, PriA is unlikely to be a functional analogue of ComFA. It may be the case that another ATPase fulfils this role in Gram-negative bacteria, or that ComFA in Gram-positive bacteria is involved in homologous recombination rather than DNA uptake.

1.5. Project aims

The general aim of this thesis is to gain mechanistic detail into how ComEC functions and to address some of the aforementioned gaps in the understanding of the DNA uptake process in natural transformation. Subsequent chapters will address the following specific aims:

- Developing a method for production of recombinant ComEC to investigate the structure, oligomeric state and function of the protein.
- Purification of putative competence complexes from *B. subtilis* to better understand how the competence machinery is organised *in vivo*.
- Production and biochemical studies of the ComEC CTD to investigate the function of this MBL fold in DNA transport.
- Production of recombinant ComFA and ComFC to investigate the machinery that potentially powers DNA transport through ComEC.

Chapter 2: Materials and methods

2.1. Culturing of bacteria

Media for growing *E. coli* and *B. subtilis* are detailed in Table 2.1. Agar plates were prepared by adding 20 g/l agar to the appropriate medium. Media and agar plates were supplemented with antibiotics, where appropriate, at the concentrations specified in Table 2.2. The list of *E. coli* strains used in this study can be found in Table 2.3, and the list of *B. subtilis* strains used in this study in Table 2.4.

Table 2.1. Composition of the growth media used in this study.

Medium	Components	Concentration
Lysogeny broth (LB)	Tryptone	10 g/l
	Yeast extract	5 g/l
	Sodium Chloride	10 g/l
Terrific broth (TB)	Tryptone	12 g/l
	Yeast extract	24 g/l
	Glycerol	0.4%
	Potassium phosphate, monobasic	23.13 g/l
	Potassium phosphate, dibasic	123.41 g/l
SPC	Magnesium sulphate	1 mM
	Glucose	0.5% (w/v)
	Yeast Extract	0.2% (w/v)
	Casamino acids	0.025% (w/v)
	Ammonium sulphate	0.2 g/l
	Potassium phosphate, dibasic	1.83 g/l
	Potassium phosphate, monobasic	0.6 g/l
	Trisodium citrate dihydrate	0.1 g/l

Medium	Components	Concentration
SP11	Magnesium sulphate	3.5 mM
	Calcium chloride	1 mM
	Glucose	0.5% (w/v)
	Yeast Extract	0.1% (w/v)
	Casamino acids	0.01% (w/v)
	Ammonium sulphate	0.2 g/l
	Potassium phosphate, dibasic	1.83 g/l
	Potassium phosphate, monobasic	0.6 g/l
	Trisodium citrate dihydrate	0.1 g/

Table 2.2. Antibiotics used in this study.

Antibiotic	Concentration used with <i>E. coli</i>	Concentration used with <i>B. subtilis</i>
Ampicillin (Amp)	100 µg/ml	-
Chloramphenicol (Cam)	25 µg/ml	5 µg/ml
Erythromycin (Erm)	-	1 or 5 µg/ml
Kanamycin (Kan)	50 µg/ml	-
Lincomycin	-	25 µg/ml
Spectinomycin (Spec)	100 µg/ml	100 µg/ml

Table 2.3. *E. coli* strains used in this study.

Strain	Parental strain	Genotype	Description	Resistance	Source
MC1061	K-12	<i>F⁻ Δ(ara-leu)7697 [araD139]_{B/r} Δ(codB-lacI)3 galK16 galE15 λ⁻ e14⁻ mcrA0 relA1 rpsL150(Str^R) spoT1 mcrB1 hsdR2</i>	Used for cloning and storage of plasmids for <i>B. subtilis</i> transformation.	Streptomycin	(Casadaban & Cohen, 1980)
TOP10	MC1061	<i>F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ⁻</i>	Used for general cloning and storage of plasmids in the form of glycerol stocks.	Streptomycin	Invitrogen
BL21 (DE3) pLysS	B	<i>F⁻ ompT lon hsdSB (rB⁻ mB⁻) gal dcm (DE3) pLysS</i>	Carries the DE3 lysogen, which confers expression of T7 RNA polymerase (T7 RNAP). The pLysS plasmid confers low level T7 lysozyme expression, which reduces basal expression levels from the T7 promoter by inhibiting T7 RNAP. Deficient in OmpT and Lon proteases. Used for protein expression.	Cam	Invitrogen
BL21 Star (DE3)	BL21 (DE3)	<i>F⁻ ompT lon hsdSB (rB⁻ mB⁻) gal dcm rne131 (DE3)</i>	Lacks a functional RNase E, resulting in longer transcript half-life. Used for protein expression.	-	Invitrogen

Chapter 2

Rosetta 2 (DE3)	BL21 (DE3)	<i>F⁻ ompT hsdSB(rB- mB-) gal dcm</i> (DE3) pRARE2	Carries the pRARE2 plasmid that confers expression of 7 rare tRNAs. Used for protein expression.	Cam	Novagen
C41 (DE3)	BL21 (DE3)	<i>F⁻ ompT lon hsdSB (rB- mB-) gal dcm</i> (DE3) <i>lacUV5*</i>	Mutations in the <i>lacUV5</i> promoter of T7RNAP result in lower T7RNAP expression levels. Useful for toxic and membrane protein overexpression.	-	(Miroux & Walker, 1996)
C43 (DE3)	BL21 (DE3)	<i>F⁻ ompT lon hsdSB (rB- mB-) gal dcm</i> (DE3) <i>lacUV5**</i>	Mutations in the <i>lacUV5</i> promoter of T7RNAP result in lower T7RNAP expression levels . Useful for toxic and membrane protein overexpression.	-	(Miroux & Walker, 1996)
L56 (DE3) (Mt56 (DE3))	BL21 (DE3)	<i>F⁻ ompT lon hsdSB (rB- mB-) gal dcm</i> (DE3) <i>t7rnapi_{mt56}</i>	Substitution in T7 RNAP decreasing T7 promoter affinity. Useful for toxic and membrane protein overexpression.	-	(Baumgarten <i>et al</i> , 2017)

Table 2.4. *B. subtilis* strains used in this study.

Strain	Parental strain	Genotype	Description	Resistance	Source
168+	168	<i>trpC2⁺</i>	Wild type strain with restored tryptophan prototrophy.	-	R Daniel (Newcastle)
1A976 (SCK6)	1A751	<i>his nprR2 nprE18 ΔaprA3 ΔeglS102 ΔbglT bglSRV lacA::P_{xyIA}-comK (Erm^R)</i>	Supercompetent strain containing a xylose-inducible <i>comK</i> expression cassette integrated into the <i>lacA</i> locus.	Erm	(Zhang & Zhang, 2011)
KO7-S	PY79	<i>PY79 ΔnprE ΔaprE Δepr Δmpr ΔnprB Δvpr Δbpr ΔsigF</i>	Non-sporulating, deficient in all seven secreted proteases.	-	Bacillus Genetic Stock Centre
KO7-S-K	KO7-S	<i>KO7-S lacA::P_{xyIA}comK (Erm^R)</i>	KO7-S strain with xylose-inducible ComK supercompetence cassette in the <i>lacA</i> locus.	Erm	This study
AS1	KO7-S	<i>KO7-S amyE::P_{comGA}lacZ-gfp (Spec^R) lacA::P_{xyIA}comK (Erm^R)</i>	Supercompetent fluorescent reporter strain produced by transformation with plasmid pPG35. Contains a <i>lacZ-gfp</i> fusion under control of the competence-inducible <i>comGA</i> promoter. Used for fluorescence microscopy.	Spec, Erm	This study
AS2	KO7-S	<i>KO7-S amyE::P_{comGA}lacZ-gfp (Cam^R) lacA::P_{xyIA}comK (Erm^R)</i>	Supercompetent fluorescent reporter strain produced by transformation with plasmid pPG40. Contains a <i>lacZ-gfp</i> fusion under control of the competence-inducible <i>comGA</i> promoter. Genomic DNA used for transformation efficiency assays.	Cam, Erm	This study

Chapter 2

AS3	KO7-S	KO7-S <i>comEC-TS</i>	Expresses ComEC with a C-terminal TwinStrep-tag. Constructed using the pMAD-ComEC-TS plasmid.	-	This study
AS3K	AS3	KO7-S <i>comEC-TS</i> <i>lacA::P_{xyIA}comK (Erm^R)</i>	Expresses ComEC with a C-terminal TwinStrep-tag. Carries the xylose-inducible ComK supercompetence cassette.	Erm	This study
AS4	KO7-S	KO7-S <i>comFA-His</i>	Expresses ComFA with an N-terminal His ₈ -tag. Constructed using the pMAD-ComFA-His plasmid.	-	This study
AS4K	AS4	KO7-S <i>comFA-His</i> <i>lacA::P_{xyIA}comK (Erm^R)</i>	Expresses ComFA with an N-terminal His ₈ -tag. Carries the xylose-inducible ComK supercompetence cassette.	Erm	This study
AS5	KO7-S	KO7-S <i>comFC-TS</i>	Expresses ComFC with a C-terminal TwinStrep-tag. Constructed using the pMAD_ComFC-TS plasmid.	-	This study
AS5K	AS4	KO7-S <i>comFC-TS</i> <i>lacA::P_{xyIA}comK (Erm^R)</i>	Expresses ComFC with a C-terminal TwinStrep-tag. Carries the xylose-inducible ComK supercompetence cassette.	Erm	This study
AS6	168+	168+ <i>comEC-DADA</i>	168+ strain with D573A and D575A substitutions in the ComEC amino acid sequence. Constructed using the pMAD-ComEC-DADA plasmid.	-	This study

2.2. Molecular biology

Polymerase chain reactions

All polymerase chain reactions (PCRs) were carried out using the Q5 High-Fidelity DNA polymerase 2X Master Mix (NEB), or OneTaq DNA polymerase 2X Master Mix with Standard Buffer (NEB), or KOD DNA polymerase (Merck) or PfuUltra II Fusion HS DNA polymerase (Agilent) according to the manufacturers' instructions. Primer annealing temperatures were determined using the manufacturers' recommendations. The primers used in this study are given in Table 2.7.

Genomic DNA sources

B. subtilis genomic DNA (gDNA) for the construction of chromosomally tagged *B. subtilis* mutants and transformation efficiency assays was prepared from overnight cultures using the PureLink Genomic DNA Mini Kit (ThermoFisher Scientific).

gDNA used for the amplification of the *comEC* homologues was obtained from the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ), with the exception of the *Klebsiella oxytoca* genomic DNA, which was extracted from cells using the PureLink Genomic DNA Mini kit.

Molecular cloning methods

Table 2.5 lists plasmids made by others that were used in this work. The plasmids made during this project are listed in Table 2.6 together with the cloning strategies used to make them. The methods used to construct the plasmids were as follows.

Restriction cloning

Inserts were amplified by PCR with primers incorporating appropriate restriction sites at the 5' and 3' end of the target gene. Plasmids or gDNA containing the target sequence were used as templates for PCR. PCR products were gel-extracted from 1% agarose gels (Invitrogen UltraPure agarose dissolved in TAE buffer, 40 mM Tris-acetate pH 8.2, 1 mM EDTA) after staining with SYBR Safe stain (Invitrogen). The QIAquick Gel Extraction Kit (QIAGEN) was used for gel extraction. Vectors and purified inserts were digested with restriction enzymes from NEB for 1-16 h at 37°C. Digested vectors were treated with calf intestinal phosphatase (NEB) for 1 h at 37°C. Digested inserts and vectors were purified from 1% agarose gels. Inserts were ligated into vectors at an insert-to-vector molar ratio of 3:1 with T4 DNA ligase (NEB) for 1 h at room temperature or 4°C overnight. Chemically competent *E. coli* TOP10 or MC1061 cells were transformed with ligation reactions and plated on LB agar plates with appropriate antibiotics. Clones were put into liquid culture, grown overnight, and plasmids were extracted using the QIAprep Spin Miniprep Kit (QIAGEN). The plasmids were screened for correct inserts by restriction digestion and confirmed by Sanger sequencing at SourceBioscience.

Table 2.5. Plasmids constructed in earlier studies.

Plasmid	Description	Resistance	Source
pWaldo-sfGFP	Based on the pWaldo-GFPd vector (Drew <i>et al</i> , 2006). The GFP-coding sequence was swapped for a superfolder GFP (sfGFP)-coding sequence (Pédélecq <i>et al</i> , 2006) using BamHI and HindII restriction sites. Inserts cloned using NdeI and BamHI/KpnI restriction sites are in-frame with a TEV protease cleavage site followed by the sfGFP sequence and a His ₈ -tag.	Kan	David Drew (Stockholm)
pGEM-T Easy	Plasmid for TA cloning.	Amp	Promega
pCDFDuet-1	Vector for protein overexpression of two genes. A T7 promoter directs expression from two multiple cloning sites (MCS). Fuses an N-terminal His ₆ -tag to the coding sequence in MCS1, and a C-terminal S-tag to the coding sequence in MCS2.	Spec	Novagen
pMAD	Vector with a temperature-sensitive replication origin. Used for markerless allelic replacement in <i>B. subtilis</i> .	Amp (<i>E. coli</i>)/Erm (<i>B. subtilis</i>)	(Arnaud <i>et al</i> , 2004)
pQE80-L	Vector for protein overexpression in <i>E. coli</i> . T5 promoter-driven expression. Fuses an N-terminal His ₆ -tag to the coding sequence.	Amp	QIAGEN
pPG35	Plasmid carrying a <i>lacZ-gfp</i> fusion under the control of a <i>comGA</i> promoter. Integrates into the <i>amyE</i> locus of <i>B. subtilis</i> .	Amp (<i>E. coli</i>)/Spec (<i>B. subtilis</i>)	(Gamba <i>et al</i> , 2015)
pPG40	Plasmid carrying a <i>lacZ-gfp</i> fusion under the control of a <i>comGA</i> promoter. Integrates into the <i>amyE</i> locus of <i>B. subtilis</i> .	Amp (<i>E. coli</i>)/Cam (<i>B. subtilis</i>)	(Gamba <i>et al</i> , 2015)
pHT01	This plasmid was used in nuclease assays.	Amp (<i>E. coli</i>)/ Cam (<i>B. subtilis</i>)	MoBiTec

Chapter 2

Table 2.6. Plasmids constructed during this project.

Plasmid	Backbone	Insert	Cloning method	Resistance
pWsf-BhaComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Bacillus halodurans</i> DSM-18197 (Bha) gDNA with primers AS1 and AS2.	Kan
pWsf-ChyComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Carboxydotherrhus hydrogenofomans</i> DSM 6008 (Chy) gDNA with primers AS3 and AS4.	Kan
pWsf-DthComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Desulfurobacterium thermolithotrophum</i> DSM 11699 (Dth) gDNA with primers AS5 and AS6.	Kan
pWsf-GthComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Geobacillus thermodenitrificans</i> DSM 465 (Gth) gDNA with primers AS7 and AS8.	Kan
pWsf-HthComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Hydrogenobacter thermophilus</i> DSM 6534 (Hth) gDNA with primers AS9 and AS10.	Kan
pWsf-KoxComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Klebsiella oxytoca</i> DSM 7342 (Kox) gDNA with primers AS11 and AS12.	Kan
pWsf-KpnComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Klebsiella pneumoniae</i> subsp. <i>rhinoscleromatis</i> DSM 16231 (Kpn) gDNA with primers AS13 and AS14.	Kan
pWsf-MthComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Moorella thermoacetica</i> DSM 521 (Mth) gDNA with primers AS15 and AS16.	Kan

Chapter 2

pWsf-TbrComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Thermoanaerobacter brockii</i> subsp. <i>finnii</i> DSM 3389 (Tbr) gDNA with primers AS17 and AS18.	Kan
pWsf-TigComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Thermus igniterrae</i> DSM 12459 (Tig) gDNA with primers AS21 and AS22.	Kan
pWsf-TisComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Thermus islandicus</i> DSM 21543 (Tis) gDNA with primers AS23 and AS24.	Kan
pWsf-TscComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Thermus scotoductus</i> DSM 8553 (Tsc) gDNA with primers AS25 and AS26.	Kan
pWsf-TthComEC	pWaldo-sfGFP	<i>comEC</i>	Restriction cloning. Insert amplified from <i>Thermus thermophilus</i> DSM 7039 (Tth) gDNA with primers AS27 and AS28	Kan
pWsf-TthComEC-CO	pWaldo-sfGFP	<i>comEC</i>	Plasmid carrying codon-optimised TthComEC coding sequence was cut at engineered 5' NdeI and 3' BamHI sites to liberate the TthComEC coding sequence which was then ligated in to pWaldo-sfGFP.	Kan
pWsf-VchComEC-CO	pWaldo-sfGFP	<i>comEC</i>	Plasmid carrying codon-optimised <i>Vibrio cholerae</i> (Vch) ComEC coding sequence was cut at engineered 5' NdeI and 3' BamHI sites to liberate the VchComEC coding sequence which was then ligated in to pWaldo-sfGFP.	Kan
pWsf-MthComEC-TS	pWsf-MthComEC	<i>comEC</i>	C-terminal His ₈ -tag exchanged for a TwinStrep-tag (WSHPQFEKGGGSGGGSGGSAWSHPQFEK) using the KLD reaction and primers AS29 and AS30.	Kan

Chapter 2

pWsf-MthComEC- Δ 83-156	pWsf-MthComEC-TS	<i>comeEC-ΔNT</i>	Nucleotides corresponding to amino acid residues 83-156 of MthComEC were deleted using the KLD reaction and primers AS67 and AS68.	Kan
pWsf-MthComEC- Δ 2-259	pWsf-MthComEC-TS	<i>comeEC-ΔNT</i>	Nucleotides corresponding to amino acid residues 2-259 of MthComEC were deleted using the KLD reaction and primers AS69 and AS70.	Kan
pWsf-MthComEC- Δ -2-156	pWsf-MthComEC-TS	<i>comeEC-ΔNT</i>	Nucleotides corresponding to amino acid residues 2-156 of MthComEC were deleted using the KLD reaction and primers AS71 and AS72.	Kan
pCDFDuet-MthCTD	pCDFDuet-1	<i>comeEC-CTD</i>	The coding sequence for the CTD of MthComEC (residues 541-809) was inserted into cloning site 1 of the vector in-frame with the His ₆ -tag using In-Fusion cloning. The insert was amplified with primers AS31 and AS32, and the vector with primers AS33 and AS34.	Spec
pCDFDuet-MthCTD-DADA	pCDFDuet-MthCTD	<i>comeEC-CTD-DADA</i>	The coding sequence for the CTD of MthComEC was mutagenized to encode the D611A and D613A substitutions using KLD reaction and primers AS35 and AS36.	Spec
pCDFDuet-TEV-MthCTD	pCDFDuet-MthCTD	<i>comeEC-CTD</i>	A TEV cleavage site inserted after the His ₆ -tag at the start of multiple cloning site 1 using the KLD reaction and primers AS37 and AS38.	Spec
pCDFDuet-TEV-MthCTD-DADA	pCDFDuet-MthCTD-DADA	<i>comeEC-CTD-DADA</i>	A TEV cleavage site inserted after the His ₆ -tag at the start of multiple cloning site 1 using the KLD reaction and primers AS37 and AS38.	Spec

Chapter 2

pQE80-L-MthComFA	pQE80-L	<i>comFA</i>	The nucleotide sequence corresponding to MthComFA residues 15-600 was amplified from <i>M. thermoacetica</i> DSM 521 gDNA using primers AS73 and AS74 and ligated into pQE80-L by restriction cloning.	Amp
pQE80-L-MthComFC	pQE80-L	<i>comFC</i>	MthComFC coding sequence was amplified from <i>M. thermoacetica</i> DSM 521 gDNA using primers AS75 and AS76 and ligated into pQE80-L by restriction cloning.	Amp
pMAD-ComEC-TS	pMAD	<i>comEC-TS</i>	Cloning strategy described in section 2.3.	Amp (<i>E. coli</i>)/ Erm (<i>B. subtilis</i>)
pMAD-ComFA-His	pMAD	<i>comFA-His</i>	Cloning strategy described in section 2.3.	Amp (<i>E. coli</i>)/ Erm (<i>B. subtilis</i>)
pMAD_ComFC-TS	pMAD	<i>comFC-TS</i>	Cloning strategy described in section 2.3.	Amp (<i>E. coli</i>)/ Erm (<i>B. subtilis</i>)
pMAD-ComEC-DADA	pMAD	<i>comEC-DADA</i>	Cloning strategy described in section 2.3.	Amp (<i>E. coli</i>)/ Erm (<i>B. subtilis</i>)

Table 2.7. Primers used in this study.

Name	Sequence 5'-to-3' (restriction sites underlined)	Restriction site
AS1	ATATATCATATGGCTTCCCGAATCATTTTC	NdeI
AS2	ATATAGGATCCATCTGTCGTCATCACCC	BamHI
AS3	CGATATCATATGGAAATTTATCTTGCTTTCATTTTTTAC	NdeI
AS4	ATATAGGATCCATAATTTTCGACAATAAGCCC	BamHI
AS5	ATATATCATATGGAAAAACATACTCTCTATATC	NdeI
AS6	ATATAGGATCCTTTTCTTACCTCTTAGGTTTTG	BamHI
AS7	CGATATCATATGAAAGGGCAAGCCGTTTAC	NdeI
AS8	ATATAGGATCCTGGTTTCATCACTTGAAAGGTTCC	BamHI
AS9	ATATATCATATGGGGGAGACAGAACGG	NdeI
AS10	ATATAGGATCCTAACAAGAGTGGCAGTAAAAAAGC	BamHI
AS11	CTATATCATATGCGTTTACCACAACCTTGCG	NdeI
AS12	ATATAGGTACCGCCGCCTTCCCGCCAAAC	KpnI
AS13	ATATATCATATGCTGCCGCTTCTCTGG	NdeI
AS14	ATATAGGATCCCCGTTATCGGCGGG	BamHI
AS15	CTATATCATATGCGTGATCGACCGCTC	NdeI
AS16	ATATAGGATCCAGGGAGAACAGTCTTACCTG	BamHI
AS17	CGATATCATATGGAATTTCCATTTTATATCTTACTG	NdeI
AS18	ATATAGGATCCCTCCTTAAGAACAGTCCTTATATATG	BamHI
AS19	ATATATCATATGGGTCTGGGCCTGGGG	NdeI
AS20	ATATAGGTACCGCCCAGGCGTAGCCCAGGTC	KpnI
AS21	ATATATCATATGTCCGGCCCTGAGGC	NdeI
AS22	ATATAGGATCCCAGGCGTAGCCGAAGAC	BamHI
AS23	ATATATCATATGAGACCCGCCCTCCAG	NdeI

Chapter 2

AS24	ATATAGGATCCCCACGCGTAGCCGAAGGAG	BamHI
AS25	ATATATCATATGTCCGGCCCTAGGGCATTG	NdeI
AS26	ATATAGGATCCCCAGGCGTAGCCAAAGAGG	BamHI
AS27	CTATATCATATGAGGCCTTCTTGGACCTG	NdeI
AS28	ATATAGGATCCCCAGGCGTAGCCGAAGAG	BamHI
AS29	ACCGCCAGAACCTCCACCTTTTCAAAGTGC GGATGGCTCCATTGTAGA GCTCATCCATGCC	-
AS30	GGATCGGGAGGTTTCAGCGTGGAGCCACCCGCAGTTCGAGAAAACTC AAGCTTGCGGCCC	-
AS31	ATCACCATCATCACCACAGCGGGCAGCCGGGAGAAC	-
AS32	GAGCTCGAATTCGGATCCTGTTAAGGGAGAACAGTCTTTACCTGC	-
AS33	CTAGTTCTCCCGCTGCCCGCTGTGGTGATGATGGTGATG	-
AS34	GTAAAGACTGTTCTCCCTTAACAGGATCCGAATTCGAGC	-
AS35	TGCTCATCTGGGCGGCCTCATG	-
AS36	GTAGCCGGGTGGGTGCTGACGAC	-
AS37	TACTTCCAGGGGCAGCCGGGAGAACTA	-
AS38	CAGGTTTTCGCTGTGGTGATGATGGTGATG	-
AS39	ATAGTCGACACGATCTCCAATAAACG	Sall
AS40	ATAGGATCCGTAACCGGAGGCATTTG	BamHI
AS41	GTTCGTCTCTGTTATATCTGATGT	-
AS42	TAAAAAAGACTGCCGAGAAATCAG	-
AS43	CGGCAGTCTTTTTTATTTCTCGAACTGCGGGTGG	-
AS44	ATAACAGAGACGAACTGGAGCCATCCGCAGTTTG	-
AS45	CTGCAGACGCGTCGAACGATCTCCAATAAACGTGCAGAGC	-
AS46	GGGCGATATCGGATCGTAACCGGAGGCATTTGCTGC	-

Chapter 2

AS47	AGGAGGTGTTGCACTTAAATAAAC	-
AS48	AAGTGAAAAAACATAATGAGCTCC	-
AS49	GATGGTGATGTGCCGACATATAGCACGCCTCCTTTCG	-
AS50	ATCACCATCACCCTCCTCGAATGTGCCAGTTGAAAAAACAG	-
AS51	TGCCATGGTACCCGGAGGAGGTGTTGCACTTAAATAAACG	-
AS52	GGGCGATATCGGATCAAGTGAAAAACATAATGAGCTCCC	-
AS53	ATCATAATTTTTTTGATATGGATTAAC	-
AS54	ATTCTTGCAGAAGCAAATTAAAG	-
AS55	GCTTCTGATCAAGGTAAAAGATGAC	-
AS56	TAAATGATTCTGTTTTATGCCGA	-
AS57	AAACAGAATCATTTATTTCTCGAACTGCGGGTGGC	-
AS58	ACCTTGATCAGAAGCTCGGCATGGAGCCATCCG	-
AS59	CTGCAGACGCGTCGAATCATAATTTTTTTGATATGGATT	-
AS60	GGGCGATATCGGATCATTCTTGCAGAAGCAAAT	-
AS61	CGTCTCTGTTATATCTGATGTATC	-
AS62	CAGTTTTCTATCATCAGCGTAC	-
AS63	GCGCAGCGTGCGTCAGAATTAAAG	-
AS64	AAGCGCATATCGGAGAGGCGGAG	-
AS65	CTGCAGACGCGTCGACGTCTCTGTTATATCTGATGTATCA	-
AS66	GGGCGATATCGGATCCAGTTTTCTATCATCAGCGTACCGA	-
AS67	CCAGGTACTAATAAAAAAACCCATGGC	-
AS68	GGCGACGTTCTCCGCGTC	-
AS69	CATATGTATATCTCCTTCTTAAAGTTAAAC	-
AS70	AACCTGGGGGTTTTCCAT	-
AS71	CATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGG	-

AS72	GGCGACGTTCTCCGCGTC	-
AS73	ATAGCATGCCAAGGTCCCCTCTACC	SphI
AS74	ATAGTCGACTTACCGCGTCAAGTAAC	Sall
AS75	TATTGGATCCATGCCGGCGTTTATC	BamHI
AS76	AATAAAGCTTTCATGGTGAAATATTTGCAATTG	HindIII

TA cloning

Difficult to clone inserts were initially cloned into pGEM-T Easy (Promega) (Table 2.6). PCR products were gel-purified and incubated with Taq DNA polymerase and 0.2 mM dATP for 30 min at 70°C to attach A-overhangs onto the termini of the PCR products. The PCR products were then ligated into the linearised pGEM-T Easy vector with T-overhangs by incubation with T4 DNA ligase (Promega) for 16 h at 4°C. The ligation reactions were transformed into *E. coli* and screened as above.

KLD reaction cloning

Where specified in Table 2.6, point mutations, insertions and deletions were performed using the Kinase-Ligase-DpnI (KLD) enzyme mix from the Q5 Site-Directed Mutagenesis Kit (NEB). The kinase phosphorylates the ends of PCR products to promote efficient ligation by the ligase. DpnI digests methylated template DNA. All primers for PCR reactions involving the KLD reaction were designed using the NEBaseChanger (<http://nebasechanger.neb.com/>) online tool. Products of the KLD reaction were transformed into *E. coli* and screened as above.

In-Fusion cloning

Some constructs that could not be obtained using the aforementioned cloning methods were successfully constructed using the In-Fusion HD Cloning Kit (Takara Bio)

according to the manufacturer's instructions and using their online primer design tool. This method relies on complementary ssDNA overhangs resected by an enzyme mix at the termini of the vector and the insert. The overhangs anneal at sites of complementarity and the resulting circular construct is ligated by the DNA repair machinery of *E. coli* after transformation. Plasmids were isolated from transformants and screened as above.

Gene synthesis

The *comEC* genes of *M. thermoacetica*, *T. thermophilus* and *V. cholerae* were codon-optimised for expression in *E. coli* and synthesised by GeneArt (Invitrogen). The synthetic DNA sequences encoding the ComEC homologues were cut out from vectors supplied by GeneArt using NdeI and BamHI restriction sites, and the *comEC* genes were ligated into pWaldo-sfGFP.

2.3. Construction of B. subtilis chromosomal mutant strains

Construction of plasmids for chromosomal integration

The pMAD plasmid was used to make specific alterations to the *B. subtilis* chromosomal sequence without leaving resistance genes or other unwanted alterations (Arnaud *et al*, 2004). The pMAD plasmid features a thermosensitive origin of replication that is functional in *B. subtilis* at 30°C. It contains an erythromycin resistance gene and also allows blue/white colony selection for the presence of the plasmid on 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal) plates.

For making the pMAD-ComEC-TS plasmid, a 1,900 bp fragment incorporating the *comEC* gene was amplified from *B. subtilis* subsp. *subtilis* 168 genomic DNA using primers AS39 and AS40 and ligated into pGEM-T Easy using TA cloning. A TwinStrep tag was

inserted at the C-terminus of the *comEC* coding sequence using In-Fusion cloning: the TwinStrep tag was amplified with primers AS43 and AS44 from pWsf-MthComEC-TS, and the vector was amplified with primers AS41 and AS42. The fragment containing the tagged coding sequence was amplified using primers AS45 and AS46 and ligated using the In-Fusion reaction into pMAD digested with BamHI and Sall restriction enzymes. For the pMAD-ComFA-His plasmid, a 1,604 bp fragment was amplified using primers AS47 and AS48 and ligated into pGEM-T Easy as above. A His₈-tag was inserted after the start codon of the *comFA* coding sequence using the KLD reaction and primers AS49 and AS50. The tagged fragment was amplified using primers AS51 and AS52 and ligated into pMAD as above. For the pMAD-ComFC-TS plasmid, a 1,232 bp fragment was amplified using primers AS53 and AS54 and ligated into pGEM-T Easy as above. A TwinStrep tag with a SA linker was inserted at the C-terminus of the *comFC* coding sequence using the In-Fusion reaction: the tag was amplified with primers AS57 and AS58 from pWsf-MthComEC-TS, the vector was amplified with primers AS55 and AS56. The tagged fragment was amplified using primers AS59 and AS60 and ligated into pMAD as above. For the pMAD-ComEC-DADA plasmid, a 1,200 bp fragment was amplified as above using primers AS61 and AS62 and ligated into pGEM-T Easy. Nucleotides coding for ComEC residues D573 and D575 were mutated to alanine-coding nucleotides using KLD cloning and primers AS63 and AS64. The fragment containing the mutated codons was amplified using primers AS65 and AS66 and cloned into pMAD as above. All plasmid insert sequences were confirmed by sequencing.

Transformation and mutant selection

B. subtilis chromosomal mutant strains AS3, AS4, AS5 and AS6 (Table 2.4) were constructed using the aforementioned pMAD plasmids and a protocol adapted from that used by the N Stanley-Wall laboratory (Dundee). *B. subtilis* cells were grown to competence as described in section 2.4. 3 µg of the appropriate pMAD plasmid were added to 500 µl competent cells in glass test tubes and incubated for 1 h at 30°C, 180 rpm. Cells were plated on LB agar plates with MLS (25 µg/ml lincomycin and 1 µg/ml Erm) and incubated at 30°C for 2 days. Two colonies were picked per transformation and used to inoculate 5 ml LB-MLS, which were then cultured overnight at 37°C with 180 rpm shaking to promote integration of the temperature-sensitive plasmid. The following morning the cultures were serially diluted 10^{-4} to 10^{-6} and plated on LB-MLS agar plates with 100 µg/ml X-gal, followed by overnight incubation at 37°C. The following day 3-4 blue colonies were picked and used to inoculate 5 ml LB cultures, which were then cultured at 30°C overnight without shaking to promote excision of the pMAD plasmid from the chromosome. The following morning the cultures were incubated for a further 4 h at 30°C with 180 rpm shaking. The cultures were serially diluted 10^{-6} to 10^{-8} and plated on LB-X-gal agar plates which were incubated at 37°C overnight. The following day white colonies were replica-patched on LB-X-gal and LB-MLS agar plates followed by overnight incubation at 37°C. Patched colonies that were white and MLS-sensitive were screened by PCR and sequencing to confirm alterations to the chromosome.

2.4. B. subtilis transformation assays

B. subtilis transformation efficiency assays were based on the transformation protocol of Bennallack *et al* (2014). Overnight LB cultures were used to inoculate 15 ml of pre-

Chapter 2

warmed SPC medium at a starting OD₆₀₀ of 0.3-0.4. Cultures were grown at 37°C, 180 rpm shaking in 250 ml flasks. Regular cell density measurements were taken to follow cultures to stationary phase (3-4 h). The cultures were then diluted by adding an equivalent volume of pre-warmed SPII medium, and grown for another 1.5 h until competence. Genomic DNA from *B. subtilis* AS2 cells was prepared using the PureLink Genomic DNA Mini Kit. 1 µg of DNA was added to 500 µl competent cells in sterile glass test tubes. Cells were incubated with DNA for 30 min at 37°C, shaking at 180 rpm. 300 µl of LB were then added, and the cells were incubated for a further 30 min. Control transformations with no added DNA were carried out in parallel. The number of transformants was estimated by plating 10⁰-10⁻² serial dilutions of the transformation reaction on LB-Cam agar plates and counting the number of transformants after overnight incubation at 37°C. The total number of viable cells in the transformation reactions (colony forming units, CFU) was estimated by plating 10⁻⁶ and 10⁻⁷ dilutions on LB agar plates without antibiotics and counting the colonies after overnight incubation at 37°C. The transformation efficiency is defined as the transformants-to-CFU ratio.

Transformation efficiency under supercompetence conditions was determined following the protocol by Zhang & Zhang (2011) for preparing and transforming competent cells. An overnight LB culture was diluted to OD₆₀₀=1.0 in LB with 1% xylose and grown for 2 h at 37°C, shaking at 180 rpm. 1 µg of gDNA was added to 500 µl supercompetent cells which were then incubated for 1.5 h at 37°C, 180 rpm. Cells were plated and transformation efficiencies calculated as above.

2.5. Protein sample analysis

SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was routinely used for qualitative estimation of protein sample purity and abundance. Table 2.8 shows the composition of buffers used for SDS-PAGE. Protein samples were denatured by adding 6x SDS-PAGE sample buffer prior to electrophoresis. Mini-PROTEAN TGX 4-20% precast gels (Bio-Rad) were used in Mini-PROTEAN Tetra vertical electrophoresis cells (Bio-Rad) with 1x SDS-PAGE running buffer. Gels were run at constant current (35 mA) for 40 min at room temperature. After electrophoresis gels were either used for Western blotting or stained. Coomassie blue stain was routinely used for staining gels, followed by destaining. Silver staining was performed using the Pierce Silver Stain Kit (Thermo Scientific) according to the manufacturer's instructions.

Table 2.8. Buffers used for SDS-PAGE.

Buffer	Components	Concentration
1x SDS-PAGE running buffer	Tris	25 mM
	Glycine	192 mM
	SDS	0.1% w/v
6x SDS-PAGE sample buffer	Tris-HCl	375 mM
	SDS	10% w/v
	Glycerol	50% v/v
	Bromophenol blue	0.1% w/v
Coomassie blue stain	Coomassie Brilliant Blue R250	0.1% w/v
	Methanol	50% v/v
	Glacial acetic acid	10% v/v

Destaining solution	Methanol	15% v/v
	Glacial acetic acid	10% w/v

Western blotting

Western blotting was routinely used to identify tagged proteins in complex samples or to confirm the tagged protein identity in pure samples. After electrophoresis, proteins from the SDS-PAGE gel were transferred onto nitrocellulose membranes either using the semi-dry Trans-Blot Turbo Transfer System (Bio-Rad) or wet transfer in 20 mM Tris, 192 mM glycine and 20% methanol using the TE 22 Mini Tank Transfer Unit (GE Healthcare). After transfer the membranes were stained with Ponceau S staining solution (0.1 % (w/v) Ponceau S, 5% (v/v) acetic acid) to inspect for transfer artefacts, followed by destaining in water. Membranes were blocked with 2% (w/v) milk powder solution in TBST (20 mM Tris-HCl pH 7.6, 137 mM NaCl, 0.02% Tween-20) for 20 min to 1 h at room temperature. Membranes were then probed with primary antibodies in 2% (w/v) milk powder solution in TBST at the dilutions indicated in Table 2.9. This incubation was for 1 h at room temperature or at 4°C overnight. The membranes were then washed three times (5 min each) with TBST, incubated with secondary antibody coupled to horseradish peroxidase (HRP) in TBST for 45 min and washed three times (5 min each) with TBST again. Alternatively, for the detection of Strep-tagged proteins membranes were blocked with 3% (w/v) bovine serum albumin in TBST and incubated with a 1:100,000 dilution of StrepTactin-HRP (Bio-Rad) for 1 h at room temperature and washed as above. Western blots were developed by exposing Amersham Hyperfilm ECL (GE) to membranes sprayed with Immobilon Forte HRP substrate (Merck). The film was developed in an X-ray film processor (Optimax).

Table 2.9. List of antibodies.

Antibody	Dilution	Source
Primary		
Monoclonal anti-GFP 3H9 (mouse)	1:1,000	ChromoTek
Monoclonal anti-Strep-tag II (mouse)	1:2,000	QIAGEN
Monoclonal anti-polyhistidine H1029 (mouse)	1:1,000	Sigma-Aldrich
Secondary		
Polyclonal anti-mouse antibody-peroxidase conjugate (goat)	1:10,000	Sigma-Aldrich

Protein concentration determination

For complex samples, such as lysates and membrane fractions, the DC Protein Assay kit (Bio-Rad) was with BSA as the calibration standard to estimate the total protein concentration.

For pure protein samples, UV absorbance measurements were used to determine the concentration. A Nanodrop 2000c spectrophotometer (Thermo Scientific) was used to measure sample absorbance at 280 nm. Protein concentration was determined using the Lambert-Beer law, $A = \varepsilon cl$, where A is absorbance, c is molar concentration, ε is the molar extinction coefficient at 280 nm, and l is the optical path length in cm. The following extinction coefficients at 280 nm were used as calculated by the ProtParam tool (<https://web.expasy.org/protparam/>): 134,190 M⁻¹ cm⁻¹ (MthComEC), 23,950 M⁻¹ cm⁻¹ (MthCTD and MthCTD-DADA), 100,685 M⁻¹ cm⁻¹ (MthComFA).

Mass spectrometry

All mass spectrometry sample processing and peptide identification was done by Sabrina Liberatori at the Advanced Proteomics Facility, University of Oxford. Samples were submitted for mass spectrometry analysis in the form of gel bands. The bands were subjected to overnight in-gel trypsin digestion. The tryptic peptides were subjected to liquid chromatography-mass spectrometry analysis.

2.6. Protein expression and purification

All cultures were started from cells freshly transformed with plasmid DNA. Starter and expression cultures contained standard concentrations of the appropriate antibiotics.

Expression screens

ComEC homologue expression screen

E. coli expression strains transformed with the pWaldo-sfGFP vectors carrying the *comEC* homologues were used to inoculate 300 µl TB starter cultures in a 2.2 ml 96-well plate with a gas-permeable seal. The starter cultures were incubated overnight at 37°C with 225 rpm shaking. The expression test was carried out in 24-well format plates. 4 ml TB with appropriate antibiotics was inoculated with 50 µl of starter culture. The plate was then covered with a gas-permeable seal and incubated at 37°C with 225 rpm shaking for either 4 or 8 hours. At this stage cells were induced with 0.1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) and grown at 18°C with 180 rpm shaking for 16-20 hours. Cells were harvested by centrifugation at 3,200g for 20 min at 4°C. Pellets were resuspended in 1 ml cold phosphate-buffered saline (PBS; 10 mM phosphate pH 7.4, 2.7 mM KCl, 137 mM NaCl) and kept on ice. Aliquots of the resuspended pellets were dispensed into 96-well plates. OD₆₀₀ and sfGFP fluorescence (485 nm excitation/509 nm

emission) measurements were taken using a CLARIOstar Plus plate reader (BMG LABTECH). Cells were diluted to give a 1 ml sample of $OD_{600}=1.0$, pelleted by centrifugation and solubilised in SDS-PAGE sample buffer. Pellet samples were analysed by SDS-PAGE and Western blotting.

MthComEC truncation construct expression screen

MthComEC NTD truncation constructs (pWsf-MthComEC- Δ 83-156, pWsf-MthComEC- Δ 2-259 and pWsf-MthComEC- Δ 2-156) were screened for expression as in the previous section but modified to enable a comparison of the truncation construct expression levels to that of the full-length construct (pWsf-MthComEC-TS in Rosetta 2 (DE3) cells) present on the same screening plate. When the cells carrying the full-length construct reached an $OD_{600}=0.6$, expression of all the construct/strain combinations on the plate was induced with 0.1 mM IPTG. Induction, harvesting, washing and OD_{600} /fluorescence readings were as previously described. The samples were then lysed by sonication, and centrifuged to separate the inclusion bodies from the soluble fraction. The lysate fractions were subsequently analysed by SDS-PAGE and Western blotting.

Detergent screen

Membranes were isolated from a 500 ml *E. coli* Rosetta 2 (DE3) cell culture expressing MthComEC-sfGFP-His₈ as described below. Membranes were resuspended in PBS buffer, split into 10 tubes, and detergent stock solutions and PBS buffer were added to a final volume of 500 μ l and a final detergent concentration shown in Table 2.10. After 1 h incubation on a spinning wheel at 4°C, insoluble material was cleared by centrifugation at 30,000g for 1 h at 4°C. Supernatants were retained, and 50 μ l of each sample was subjected to size exclusion chromatography on a Superose 6 10/30 column (GE

Healthcare) connected to an AKTA chromatography system with an in-line fluorescence detector (Hitachi) set to measure GFP fluorescence. The mobile phase was PBS buffer with 0.03% DDM.

Table 2.10. Detergents used for screening. All detergents were purchased from Anatrace, except digitonin which was purchased from Calbiochem.

Detergent	Abbreviation	Critical micellar concentration (CMC) (w/v) ^a	Final concentration in solubilisation assay (w/v)
n-Decyl- β -D-Maltopyranoside	DM	0.087%	1%
n-Dodecyl- β -D-Maltopyranoside	DDM	0.0087%	1%
Decyl maltose neopentyl glycol	DMNG	0.0034%	1%
Lauryl maltose neopentyl glycol	LMNG	0.001%	1%
Lauryldimethylamine-N-oxide	LDAO	0.023%	1%
2,2-bis(3'-cyclohexylbutyl) propane-1,3-bis- β -D-maltopyranoside	Cymal-6-NG	Unknown	1%
Polyoxyethylene-9 dodecyl ether	C ₁₂ E ₉	0.003%	1%
n-Octyl- β -D-Glucopyranoside	OG	0.53%	2%
Digitonin	-	0.02-0.03%	1.5%
Glyco-diosgenin	GDN	0.0021%	1.5%

^a As specified by the manufacturer.

Purification of MthComEC

The following protocol is for expression and purification of MthComEC-TS. Variations to the protocol are described at the end of the section and in Chapter 2. 100 ml starter TB cultures with Kan and Cam were inoculated with *E. coli* Rosetta2 (DE3) cells transformed with the pWsf-MthComEC-TS plasmid and grown overnight at 37°, 180 rpm. The next day starter cultures were diluted 1:80 in 500 ml fresh TB medium with Kan and Cam per 2 l baffled culture flask. The cultures were grown at 37°C with 180 rpm shaking until OD₆₀₀=0.6-0.8. The cultures were then induced with 0.1 mM IPTG, the temperature was lowered to 18°C and cells were grown for another 18-22 hours. Cells were harvested by centrifugation at 6,000g for 30 min at 4°C. Cell pellets were flash-frozen in liquid nitrogen and stored at -80°C until use.

Cell pellets were thawed in a room temperature water bath and resuspended in 4°C lysis buffer (40 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA). For each gram wet weight of cell pellet 5 ml of lysis buffer was used. The solution was supplemented with a few crystals of lysozyme and DNase I and incubated at room temperature for 30 min with stirring. The cells were then homogenised using a Dounce homogeniser and lysed by passing the solution twice through a French pressure cell press at 20 kpsi. The lysate was clarified by centrifugation at 30,000g for 30 min at 4°C. The supernatant was subjected to ultracentrifugation at 200,000-235,000g for 1.5 h at 4°C. The membrane-containing pellets were rinsed and resuspended in binding buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA). The total protein concentration of the membrane fraction was determined using the DC Protein Assay kit. The membranes were then either used immediately, stored at 4°C overnight, or flash-frozen in liquid nitrogen and stored at -80°C.

Chapter 2

Membranes were solubilised in binding buffer at a 1:2 membrane protein-to-detergent mass ratio by addition of LMNG to a final concentration of 1% (w/v). After incubation for 1 h at 4°C with gentle stirring, insoluble material was pelleted by ultracentrifugation at 200,000-235,000g for 30 min at 4°C. The supernatant was passed through two 1 ml StrepTrap HP columns (GE Healthcare) connected in series that had been pre-equilibrated in binding buffer with 0.02% LMNG (w/v). The columns were washed with 20 column volumes (CV) of binding buffer with 0.02% LMNG. To remove the bound Sec machinery components, a further 15 CV wash with binding buffer, 0.02% LMNG, 5 mM MgCl₂ and 5 mM ATP was required. After another 20 CV wash with binding buffer containing 0.02% LMNG, MthComEC-sfGFP-TS was eluted with 6 CV binding buffer with 0.02% LMNG and 2.5 mM desthiobiotin (IBA). Fractions that were fluorescent under blue light were pooled. TEV protease was added to the pooled fractions at a protease-to-protein mass ratio of 1:10, followed by dialysis for 16 h at 4°C in binding buffer containing 0.02% LMNG, 3 mM reduced glutathione and 0.3 mM oxidised glutathione. If precipitation occurred after overnight dialysis, the sample was filtered through a 0.22 µm syringe filter. To remove the cleaved sfGFP-TS fragment released from MthComEC and any uncleaved fusion protein, the sample was passed through two 1 ml StrepTrap HP columns connected in series. The flowthrough was collected and concentrated in a spin concentrator with a 100 kDa cut-off membrane (Amicon Ultra, Merck) to less than 0.5 ml. The sample was then loaded on a Superdex 200 10/300 GL column (GE Healthcare) equilibrated in 20 mM Tris pH 8.0, 150 mM NaCl, 0.02% LMNG. Size exclusion fraction purity was analysed by SDS-PAGE and Coomassie staining. Fractions containing pure MthComEC were further concentrated if required. Samples were either used immediately, stored short-term at 4°C or flash-frozen in liquid nitrogen and stored at -80°C.

Chapter 2

The MthComEC Δ 2-156 truncation mutant was purified as above, except the MgATP wash step was omitted, 1% cyclofos-6 was used for solubilisation, and 0.25% cyclofos-6 was included instead of LMNG in subsequent purification steps.

The His₈-tagged variants of MthComEC and KoxComEC (produced using the pWsf-MthComEC and pWsf-KoxComEC plasmids) were purified as above, but with the following modifications. EDTA was omitted from buffers after membrane isolation. After clearing the solubilised membranes by ultracentrifugation, samples were supplemented with 20 mM imidazole and recirculated on a 1 ml HisTrap FF column (GE Healthcare) at 4°C for 4-16 h. The column was washed with 20 CV binding buffer with 0.03% LMNG and 20 mM imidazole, and then for another 20 CV with buffer in which the imidazole concentration had been increased to 30 mM. Proteins were eluted with binding buffer containing 250 mM imidazole and 0.03% LMNG. The MthComEC construct was then either cleaved with TEV protease and processed as above or directly subjected to size exclusion chromatography as above. The KoxComEC sample was not cleaved with TEV protease.

The MthComEC construct with a C-terminal StrepTag II was expressed and processed as above, except the MgATP wash, TEV protease cleavage and the reverse affinity purification steps were omitted.

Purification of BsuComEC-TS

The following protocol for BsuComEC-TS purification from *B. subtilis* is the final, optimised protocol. Variations to the protocol are described in Chapter 4. *B. subtilis* strain AS3K was streaked out on LB-Erm (1 µg/ml) agar plates and grown overnight. A single colony was picked and used to inoculate a 50 ml overnight TB-Erm starter culture. The starter culture was diluted in TB-Erm to OD₆₀₀~0.1 and incubated at 37°C with 180 rpm

Chapter 2

shaking until $OD_{600} \sim 1$ (2-3 h). The expression cultures were grown in 2 L baffled culture flasks and no more than 500 ml TB per flask was used. Competence gene expression was then induced by adding 1% (w/v) D-xylose (FluoroChem) to the TB cultures. Cells were then incubated for 3-3.5 h at 37°C with 180 rpm shaking. Cells were harvested by centrifugation and stored as for MthComEC.

Cell pellets were thawed and resuspended in *Bacillus* lysis buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA) supplemented with 1 mM PMSF, DNase I and lysozyme. The resuspended cells were incubated at room temperature for 30 min whilst stirring. Cells were then homogenised with a Dounce homogeniser and lysed with two passes through a French press cell at 20 kpsi. The lysate was then sonicated in a 130 W Vibra-Cell sonicator at 60% amplitude for 1 min at 10 s on/off intervals to reduce the viscosity of the sample. The lysate was clarified by centrifugation and membranes were isolated as for MthComEC. Membranes were solubilised in *Bacillus* lysis buffer with 1% LMNG (1:2 protein-to-detergent ratio by mass) and 50 U avidin (IBA BioLock reagent) for 1 h at 4°C. Insoluble material was removed by ultracentrifugation, and the soluble fraction was loaded on a 1 ml StrepTactin XT column (IBA). The column was washed as described for MthComEC, including the MgATP wash. BsuComEC was eluted with StrepTactin XT elution buffer (IBA) supplemented with 0.02% LMNG. The eluate was concentrated and subjected to size exclusion chromatography on a Superose 6 3.2/300 Increase column in 20 mM Tris pH 8.0, 150 mM NaCl, 0.02% LMNG. The size exclusion fractions were analysed by SDS-PAGE, Western blotting and silver staining.

Purification of BsuComFC-TS

BsuComFC-TS was expressed in *B. subtilis* strain AS5K in the same conditions as BsuComEC-TS. The purification protocol was also the same as that for BsuComEC-TS (unless otherwise specified), except the sample was purified in GDN (1.5% (w/v) for solubilisation and 0.02% in purification buffers) and the MgATP wash was omitted.

Purification of MthCTD and MthCTD-DADA

Starter TB cultures were inoculated with BL21 (DE3) pLysS cells transformed with the pCDFDuet-MthCTD, pCDFDuet-TEV-MthCTD, pCDFDuet-MthCTD-DADA or pCDFDuet-TEV-MthCTD-DADA plasmid and grown overnight at 37°C. 50 ml starter cultures were diluted 1:80 in TB with Cam and Spec and grown at 37°C with 180 rpm shaking. 500 ml of culture were used per 2 l baffled culture flask. When cultures reached OD₆₀₀=0.6-0.8 expression was induced by adding 0.5 mM IPTG. The cells were grown for a further 3-4 hours at 30°C. Cells were harvested by centrifugation and stored as for MthComEC. Cell pellets were thawed and resuspended in buffer A (50 mM Tris-HCl pH 8.0, 300 mM NaCl) supplemented with lysozyme and 1 mM PMSF. After a 30 min incubation at room temperature with stirring, the cells were homogenised and lysed by sonication in a 130 W Vibra-Cell sonicator at 60% amplitude for a total of 10 min on ice. The lysate was clarified as for MthComEC. The supernatant was filtered through a 0.45 µm syringe filter, supplemented with imidazole to a final concentration of 20 mM and incubated with NiNTA agarose beads (QIAGEN) at 4°C for 4-18 hours with gentle stirring. Beads with bound protein were collected on a gravity flow column, washed with 30 CV buffer A with 20 mM imidazole, 30 CV buffer A with 30 mM imidazole and eluted with 5 CV buffer A with 125 mM imidazole. The eluate was immediately diluted 2-fold with imidazole-free buffer A and stored at 4°C. The diluted eluate was concentrated in a spin concentrator

Chapter 2

with a 10 kDa cut-off membrane (Amicon Ultra, Merck) to less than 5 ml. The sample was then subjected to size exclusion chromatography on a Superdex 200 10/300 GL, Superdex 75 10/300 or HiLoad Superdex 75 10/60 prep-grade column (GE) on an ÄKTA FPLC. The mobile phase was 20 mM Tris-HCl pH 8.0, 150 mM NaCl. Fractions containing pure protein were concentrated if required and stored at 4°C.

After elution from the NiNTA resin, constructs containing the TEV protease cleavage site were incubated with TEV protease (1:100 mass ratio of protease to target protein) and dialysed against buffer A for 16 h at 4°C. The samples were then passed through NiNTA beads to remove the cleaved His₆-tag and any uncleaved protein, and the eluate was concentrated and subjected to size exclusion chromatography as above.

Purification of MthComFA

E. coli Rosetta 2 (DE3) cells transformed with the pQE80-L-MthComFA plasmid were grown overnight in TB with Amp and Cam. The 50 ml overnight culture was diluted 1:100 in TB with Amp and Cam and grown at 37°C, 200 rpm until OD₆₀₀~1.0. Expression was then induced by adding 0.5 mM IPTG. The temperature was lowered to 18°C and cells were grown for 16 h. Cell pellets were harvested and stored as above. Pellets were resuspended in ComFA lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, 2 mM β-mercaptoethanol), using 5 ml buffer for every gram wet weight of cell pellet. The cell suspension was supplemented with 1 mM PMSF, DNase I and lysozyme and incubated at room temperature for 30 min with stirring. Cell lysis by sonication and lysate clarification was done as for MthCTD. The clarified lysate was loaded on a 1 ml or 5 ml HisTrap FF column (GE Healthcare), washed with 20 CV buffer with 20 mM imidazole, then with 20 CV buffer with 30 mM imidazole, and then eluted with a 30-200 mM imidazole gradient.

NiNTA eluate fractions containing MthComFA were pooled and flash diluted ~10-fold in cation exchange buffer (20 mM Tricine-NaOH pH 8.5, 2 mM β -mercaptoethanol). The diluted protein was loaded on a 4.7 ml HiScreen SP HP column (GE Healthcare) pre-equilibrated in cation exchange buffer. The column was washed with cation exchange buffer with successively increasing NaCl concentrations (5 CV each at 20 mM, 100 mM and 150 mM) before a gradient elution step (150-500 mM NaCl). The fractions containing full-length MthComFA were concentrated in a spin concentrator with a 30 kDa cut-off membrane (Amicon Ultra, Merck) and loaded either on a Superdex 200 10/300 GL or HiLoad Superdex 75 10/60 prep-grade size exclusion column equilibrated in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP). Fractions containing pure MthComFA were used immediately or concentrated, flash-frozen in liquid nitrogen and stored at -80°C.

Purification of MthComFC

MthComFC was expressed in *E. coli* Rosetta 2 (DE3) cells from the pQE80-L-MthComFC plasmid. The expression and purification protocol was identical to that of MthComFA until NiNTA purification. The clarified lysate was recirculated on a 1 ml HisTrap HP column for 45 min at room temperature, washed with 20 CV buffer with 20 mM imidazole, 20 CV buffer with 30 mM imidazole, and bound proteins were eluted with a 30 mM-500 mM imidazole gradient. Fractions containing MthComFC were pooled and precipitate was removed by filtration through a 0.22 μ m filter. The sample was concentrated and subjected to size exclusion chromatography on a Superdex 200 10/300 GL column equilibrated in 20 mM Tris-HCl pH 8.0, 150 mM NaCl.

2.7. Electron microscopy

Individual protein molecules are too small to be observed by light microscopy due to the diffraction limit of visible light. However, shorter wavelength radiation, such as electrons, can be used to image proteins. Electron microscopy (EM) is a widely used technique for protein structural analysis. The two main modes of EM used for studying protein samples are negative staining and cryo-electron microscopy (cryo-EM) (De Carlo & Harris, 2011; Cheng *et al*, 2015). In negative stain EM, the protein particles are embedded in a heavy atom stain (such as uranyl acetate) that scatters electrons well, resulting in a high-contrast image. However, only the overall shape (at ~ 20 Å resolution) of the protein can be observed by negative stain EM because it is the imprint of the particles in the stain that is imaged, not the particles themselves. In cryo-EM, protein molecules are suspended in vitreous ice and are imaged directly. The images contain the high resolution information about the protein particles. But the signal from the protein particles is very weak, often not much above the noise level. This is because proteins are composed of low atomic number elements (H, C, N, O, S) which do not scatter electrons well, unlike heavy metal atoms. Image processing methods are used to average the signal from individual particles in the same orientation to increase the signal-to-noise ratio and retrieve the high resolution information. The relative 3D orientations of individual particles can be determined from 2D EM images by computational methods and 3D structural models of proteins can be reconstructed. Recent EM hardware and software improvements have allowed routine 3D structure determination of proteins at resolutions of less than 4 Å (Bai *et al*, 2015).

Grid preparation

For negative stain EM, copper grids coated with carbon (300 mesh) were glow discharged at 15 mA for 20 seconds in a Leica EM ACE200 glow discharge unit. Five microliters of protein purified as above were applied to the carbon side of the grid and incubated for 2.5 minutes at room temperature. The protein concentration was 0.5 µg/ml. The grid was then blotted with Whatman paper and washed twice by successively dipping the grid into two water drops and blotting after each wash. The washed grids were dipped into a drop of 2% aqueous uranyl acetate solution, blotted, and floated on a second uranyl acetate solution drop for 30 seconds, followed by blotting for the final time. The grids were then left to dry in the dark for 10 minutes before placing them into a storage box. The grids were imaged no earlier than the day after making them to allow them to dry fully.

For cryo-EM, Quantifoil holey carbon grids were glow discharged as above and mounted in a FEI Vitrobot Mark IV device. Protein samples at 0.5-2 mg/ml concentration (as indicated in Chapter 3) were applied to the mounted grids and incubated for 20 s. The grids were then blotted and vitrified by plunge-freezing into liquid ethane. The grids were transferred to storage boxes and stored in liquid nitrogen until imaged. Justin Deme, Susan Lea and Adam Costin (Sir William Dunn School of Pathology, University of Oxford UK) either made or assisted with making the grids for cryo-EM.

Data collection

Negatively stained grids were screened on a Tecnai T12 transmission electron microscope (TEM) operating at 120 kV before data collection to qualitatively determine protein sample dispersity, particle density and quality of staining. The grids were imaged

at 68,000x magnification. Negative stain datasets were automatically collected using the EPU software (FEI) on a Talos F200C TEM (FEI) operating at 200 kV. Magnification was set to 120,000x, corresponding to 0.89 Å/pixel. Defocus was varied from -3.5 to -0.5 µm. Justin Deme (Sir William Dunn School of Pathology, University of Oxford UK) either collected or assisted with collection of negative stain data on the Talos F200C TEM.

Cryo-EM data were collected on a Talos Arctica TEM operating at 200 kV with a Falcon 3 Direct Electron Detector. Movies were collected at a sampling of 0.78 Å/pixel. Defocus was varied from -2.5 to -0.5 µm. Justin Deme and Susan Lea (Sir William Dunn School of Pathology, University of Oxford UK) collected the cryo-EM data.

Single particle analysis

For negative stain data, the contrast transfer function (CTF) was estimated using CTFFIND4 (Rohou & Grigorieff, 2015). Particles were semi-automatically picked in EMAN2 using swarm mode (Tang *et al*, 2007). Particle coordinates were exported, and further data processing was done in SIMPLE (Reboul *et al*, 2018). Particles were extracted and classified into 2D classes using the prime2D programme. Poor quality 2D classes consisting of non-protein particles or broken particles were discarded. Selected 2D class averages were used for reconstructing a 3D model using ini3D_from_cavgs.

For cryo-EM data, movies were motion-corrected with Simple-unblur (Reboul *et al*, 2018) and the CTF was estimated using CTFFIND4. Particles were picked and classified into 2D class averages either as above or using on-the-fly data analysis in SIMPLE.

2.8. Biophysical methods

Thermal shift assays

Metal binding by proteins was screened using a thermal shift assay (Nettleship *et al*, 2008). The assay uses the fluorescent dye SYPRO Orange (Invitrogen) to monitor temperature-dependent protein unfolding. SYPRO Orange fluorescence is quenched in aqueous solution but is high in a hydrophobic environment. At low temperatures the target protein is folded and its hydrophobic core is buried within its structure. As the temperature increases, the target protein starts to unfold, exposing hydrophobic residues to which SYPRO Orange can bind, resulting in an increased fluorescence signal. The thermal denaturation curve (fluorescence *versus* temperature) of a typical single domain protein is sigmoidal, with the lowest fluorescence value corresponding to the folded state and the maximal fluorescence value corresponding to the unfolded state. The inflection point of the sigmoidal curve is defined as the melting temperature of the protein, T_m . Ligands, such as cofactors and substrates, tend to stabilise proteins. This stabilisation is reflected by an increase in the T_m in the presence of a ligand.

Thermal shift assays for metal binding to MthComFA, MthCTD and MthCTD-DADA were carried out in 96-well format in a Stratagene MX3005P real-time PCR machine. 25 μ l reactions contained 12 μ M MthComEC-CTD (WT or DADA) or 2.25 μ M MthComFA, 1 mM of the metal salt under investigation, 1:5,000 diluted SYPRO Orange dye, 18 mM Tris-HCl pH 8.0 and 135 mM NaCl. The sample was heated from 25°C to 94°C in 1°C increments, allowing 30 s for equilibration at each set temperature. Fluorescence readings were taken for each temperature increment at the end of the equilibration period. Data were

analysed in TSA-CRAFT software, which performs fitting of the data to the Boltzmann equation to extract the T_m (Lee *et al*, 2019):

$$f(T) = f_{min} + \frac{f_{max} - f_{min}}{1 + e^{\frac{T_m - T}{a}}}$$

where $f(T)$ is fluorescence at a particular temperature (T), f_{min} is fluorescence before thermal melting, f_{max} is fluorescence after thermal melting, and a is the enthalpy of the reaction. All data fits had a coefficient of determination (R^2) of >0.99. The thermal shift induced by ligands is defined as the T_m value with ligand present subtracted from the T_m value in the absence of ligand.

Differential scanning calorimetry

Differential scanning calorimetry (DSC) directly measures the change in heat capacity during protein unfolding (Johnson, 2013). A sample cell and a reference cell are incrementally heated at the same time. DSC instruments measure the power required to heat each cell. As the temperature increases and the protein in the sample cell begins to unfold, more power is required to keep the sample cell at the same temperature as the reference cell. This is because protein unfolding involves breakage of non-covalent interactions within the protein structure which is an endothermic process. The signal reaches a maximum when half of the protein is unfolded, and decreases as the protein nears complete denaturation. The thermogram (heat capacity vs temperature) of a protein can be used to determine the T_m of the protein in the specific buffer conditions of the experiment. Protein stabilisation by ligands can be observed in the thermogram and is reflected in an increase in the T_m .

DSC measurements were carried out with MthCTD and MthCTD-DADA. The DSC experiments were performed by David Staunton (Department of Biochemistry, University of Oxford) on a Malvern VP Capillary DSC instrument. The sample cell contained 18.6 μM protein in 20 mM Tris-HCl, 150 mM NaCl and, where appropriate, 1 mM MnCl_2 . The temperature was raised from 10°C to 110 °C at a rate of 200°C/h. Runs without protein were used for baseline subtraction. Data were analysed using the instrument manufacturer's software. The baseline was subtracted from the raw data and the data were normalised to protein concentration. The data were fitted to a two-state unfolding model to extract the T_m .

Circular dichroism spectroscopy

When circularly polarised light passes through a chiral molecule the resulting radiation may become elliptically polarised due to differential absorption of left-handed and right-handed polarised light by chiral molecules (Kelly *et al*, 2005). This difference in absorbance of left-handed and right-handed polarised light is defined as circular dichroism (CD). Proteins are chiral molecules. CD spectroscopy can be used to study protein secondary structure (Greenfield, 2006). The interaction of circularly polarised light in the 180-240 nm UV region with protein is principally due to the amide bond chromophore. The amide bonds of the polypeptide backbone are arranged in different arrays in α -helices and β -sheets. Thus, different secondary structure elements give rise to characteristic signals in the CD spectrum of proteins (Greenfield, 2006). α -helical elements give negative signal at 208 nm and 222 nm and positive signal at 193 nm. β -sheets give negative signal at 218 nm and positive signal at 195 nm. Disordered regions show weak CD signal above 210 nm and a strong negative signal at 195 nm. Contributions

of the different secondary structure elements are convoluted in protein CD spectra. The different contributions can be deconvoluted by comparing the CD spectrum of the target protein to the CD spectrum of a protein of known structure (Louis-Jeune *et al*, 2012).

CD spectra of MthCTD and MthCTD-DADA were collected on a Jasco J-815 Spectropolarimeter. The proteins were dialysed into 20 mM sodium phosphate pH 7.4 buffer, diluted to 0.1 mg/ml and pipetted into a 1 mm pathlength cuvette. A buffer reference scan was performed before data collection. Data were collected at 20°C in the 190-280 nm range at 0.5 nm intervals. Nine scans were averaged for each sample. The data were converted from CD units (mdeg) to mean residue ellipticity ($\text{deg cm}^2 \text{dmol}^{-1}$) by normalising the data in CD units to pathlength and the mean residue molecular weight. The converted data in the 190-240 nm range were submitted to the K2D3 server (Louis-Jeune *et al*, 2012) for secondary structure deconvolution.

2.9. Biochemical assays

Electrophoretic mobility shift assays

Protein binding to DNA was assayed in electrophoretic mobility shift assays (EMSAs). The assay is based on the differential migration of free and protein-bound oligonucleotides (Hellman & Fried, 2007). Free DNA oligonucleotides are small and negatively charged and, thus, exhibit fast migration rates in gels. The electrophoretic mobility of a protein-DNA complex depends on its size and net charge. Protein-DNA complexes with a net negative charge will migrate slower than free DNA because the complex is larger, resulting in disappearance of free DNA signal and appearance of bands exhibiting slower migration. Protein-DNA complexes with a net neutral or positive charge will not enter the gel and so DNA binding is inferred from disappearance of the free DNA

signal. Protein-DNA binding constants can be estimated by titrating the amount of protein against a constant concentration of DNA.

The ssDNA oligonucleotide used for EMSAs with MthComFA and MthCTD was a 5'-Cy5-labelled (ACTG)₈ concatemer. Cy5 is a far-red fluorescent dye with excitation/emission maxima at 649/666 nm. The dsDNA oligonucleotide was prepared by annealing an unlabelled complementary oligonucleotide to the labelled ssDNA oligonucleotide. The two oligonucleotides were mixed together at equimolar ratios (50 μ M each) in 20 mM Tris-HCl pH 8.0, 1 mM EDTA buffer, heated to 95°C and cooled to 25°C over the course of 1 h in a thermocycler. The annealed oligonucleotides and the ssDNA oligonucleotide were subjected to electrophoresis as described below for quality control purposes. Formation of a single species in the annealed sample that migrated slower than ssDNA indicated dsDNA formation. The oligonucleotides were a gift from Tim Nott (Department of Biochemistry, University of Oxford).

DNA and protein (MthCTD or MthComFA) were mixed in 20 mM Tris-HCl pH 8.0, 37.5 mM NaCl and 10% glycerol. The DNA concentration was constant at 50 nM for both ssDNA and dsDNA. The DNA was incubated with increasing concentrations of protein, as indicated in figure legends. DNA and protein mixtures were incubated on ice for 30 min before electrophoresis. Non-denaturing gels and buffers were used to preserve protein-DNA interactions. Gel and buffer recipes can be found in table 2.11. Gels were cast using a Mini-PROTEAN Tetra Handcast system (Bio-Rad) at 0.75 mm thickness. Gels were pre-run in 1x non-denaturing PAGE buffer for 30 min at 80 V at 4°C. Samples were then loaded and subjected to electrophoresis for 1 h at 80 V at 4°C. Gels were imaged in a Fuji FLA-

3000 laser scanner equipped with a 633 nm helium-neon excitation laser and a 675 nm light receiving filter.

Table 2.11. Recipes for EMSAs.

Reagent	Components	Concentrations
6% polyacrylamide non-denaturing gels	Tris	25 mM
	Glycine	192 mM
	ProtoGel 30% solution (37.5:1 acrylamide to bisacrylamide ratio) (Geneflow)	20% (v/v)
	Ammonium persulphate	0.6% (w/v)
	Tetramethylethylenediamine	0.12% (v/v)
1x non-denaturing PAGE buffer	Tris	25 mM
	Glycine	192 mM

Phosphomonoesterase and phosphodiesterase assays

Non-nucleotide phosphoester substrates were used to assay the phosphoesterase activity of MthCTD and MthCTD-DADA. *p*-Nitrophenyl phosphate (pNPP) or bis-(*p*-nitrophenyl) phosphate (bpNPP) (both from Sigma) was used, respectively, as the substrate for phosphomonoesterase and phosphodiesterase assays, as reported previously (Rodriguez *et al*, 2014; Chen *et al*, 2004). Reactions were performed in 40 mM glycine-NaOH pH 9.5, 60 mM NaCl and 4 mM MnCl₂, unless otherwise specified. The final enzyme concentration was 1 µM and the final substrate concentration was 0.8 mM. The buffer and enzyme components were pre-mixed and 100 µl aliquots were dispensed into the wells of a clear flat-bottom 96-well plate (200 µl capacity per well). 25 µl of substrate were dispensed into the reaction wells using the injector system of a CLARIOstar Plus plate

reader (BMG LABTECH) and mixed by shaking at 225 rpm for 10 s. The reactions were carried out at 30°C. The rate of *p*-nitrophenol release was monitored by measuring absorbance at 405 nm for 10 minutes. A linear regression fit was performed in MARS Data Analysis software (BMG LABTECH) to extract the slopes from the absorbance over time curves. The amount of *p*-nitrophenol released was calculated using $\epsilon_{405}(\textit{p-nitrophenol})=18,000 \text{ M}^{-1} \text{ cm}^{-1}$ and the Lambert-Beer equation.

Nuclease assays

The nuclease activity of MthCTD and MthCTD-DADA was assayed by incubating the proteins with DNA substrate in different buffer conditions followed by agarose gel electrophoresis to inspect the samples for DNA degradation. 50 μM protein was incubated with 1.5 μg double-stranded supercoiled plasmid DNA (pHT01) at 37°C. The final buffer composition was 60 mM Tris-HCl pH 8.0, 75 mM NaCl with either 5 mM MnCl_2 or 5 mM MgCl_2 , or without any metal salt. Samples were taken from the reactions at different time points and immediately quenched by adding 6x purple gel loading dye (NEB) with 50 mM EDTA and chilled on ice. DNase I was used as a positive control. The samples were analysed by agarose gel electrophoresis as described in section 2.2.

2.10. Crystallisation trials

Crystallisation trials were set up with MthCTD and MthComFA. Fresh purified protein preparations or thawed aliquots of flash-frozen purified protein were used to set up the trials. Before setting up crystallisation trials protein samples were centrifugated for 15 min at 10,000*g* to remove any aggregates. The formulated crystallisation screens named in Chapters 5 and 6 were obtained from Molecular Dimensions. The Stura Footprint screen was used to determine the protein concentration at which to set up the screens.

Chapter 2

The sitting drop vapour diffusion method was used to screen crystallisation conditions. A Mosquito LCP crystallization robot (Ttpabtech) was used to set up all screens. 100 μl of each precipitant solution from a commercial screen were aliquoted into reservoir wells of a 96-well crystallisation screening plate. Crystallisation drops were set up at 1:1 and 1:2 reservoir solution-to-protein ratios (v/v) and sealed immediately with a clear film. The drop size was 0.4 μl or 0.6 μl for the 1:1 and 1:2 ratio drops, respectively. Crystallisation plates were incubated at 20°C. The plates were visually inspected under a stereomicroscope for precipitate and crystal formation. The plates were inspected immediately after setting them up, then every day for 5-7 days, then every few weeks.

Chapter 3: Production and electron microscopy studies of recombinant ComEC

3.1. Introduction

ComEC is central to understanding how during natural transformation DNA is transported through membranes against a negative membrane potential while maintaining the integrity of the membrane permeability barrier (Chapter 1). To understand the mechanism of transport by ComEC, recombinant protein for structural and biochemical studies is required. Three previous studies reported that their authors were unable to express ComEC in *E. coli* (Yeh *et al*, 2003; Diallo *et al*, 2017; Inamine & Dubnau, 1995). However, these studies tried a very limited number of ComEC homologues that might simply not fold well in *E. coli*.

An extensive expression screen testing many ComEC homologues in different *E. coli* expression strains would cover substantial “expression space” and, hopefully, yield a homologue/strain combination suitable for recombinant protein production. The screening process can be accelerated by fusing a green fluorescent protein (GFP) tag to the target protein and using fluorescence as a readout for expression level (Drew *et al*, 2006).

Once a suitable expression condition is identified, a purification protocol can be developed. A challenge specific to membrane protein biochemistry is choosing a suitable detergent for protein purification. The detergent must be able to simulate the native lipid environment and stabilise the folded state of the membrane protein. Fluorescence-size exclusion chromatography (FSEC) is a useful method for screening the properties of GFP-tagged proteins in different detergents (Kawate & Gouaux, 2006). GFP fluorescence allows highly sensitive detection of the tagged protein in an impure sample, and size exclusion chromatography is used to assess sample size and dispersity.

Once recombinant protein is available, structural studies can go ahead. The majority of protein structure models deposited to date in the Protein Data Bank (PDB, <https://www.rcsb.org/>) have been obtained using X-ray crystallography. However, crystallisation screens require milligram quantities of pure protein, which is unachievable for many, especially membrane, proteins. Electron microscopy (EM) is an attractive structural method for proteins that express poorly. In recent years, software and hardware improvements allowed sub-4 Å resolution single particle reconstructions of proteins smaller than 200 kDa (Bai *et al*, 2015). Because ComEC is a novel type of transport protein, its structure is of immense interest. The structure of ComEC will be key to understanding the process of DNA transport across the bacterial cell membrane and how the protein interacts with other components of the competence machinery. Even a low resolution structure would shed light on the topology of the protein, which is still not clear from bioinformatics predictions and experimental data (Draskovic & Dubnau, 2005; Baker *et al*, 2016) (Chapter 1).

The following sections address in detail the selection of ComEC homologues, expression testing, optimisation of protein purification, and electron microscopy studies of ComEC during this project.

3.2. Selection of ComEC homologues

ComEC homologues to be expressed in the heterologous expression host, *E. coli*, were chosen to include examples exhibiting at least one of three characteristics hypothesised to aid protein expression.

First of all, the biosynthetic machinery and membrane lipid composition of *E. coli* may be more compatible with proteins from organisms that are closely related to *E. coli*. The

Chapter 3

phylogenetic tree of life by Hug *et al* (2016) was used to qualitatively evaluate how different bacterial species that have ComEC homologues are related to *E. coli* at the phylum or class level.

Secondly, homologues from thermophilic bacteria were included in the expression screen because they are usually intrinsically more stable due to their evolutionary adaption to function at high temperatures. Therefore, thermophile ComEC homologues are likely more amenable to structural studies (Szilágyi & Závodszky, 2000).

Thirdly, smaller proteins may express better because they have fewer opportunities to misfold. ComEC annotation in Pfam (El-Gebali *et al*, 2019) (PF03772) suggests that the two most prominent ComEC domain architectures contain either three or two domains (Chapter 1). It was hypothesised that ComEC homologues lacking the C-terminal domain may express better because they are smaller and there will be one less domain that the biosynthetic machinery needs to target to and fold in the periplasm.

Most ComEC protein coding sequences were cloned by amplifying the nucleotide sequence from genomic DNA obtained from the German Collection of Microorganisms and Cell Cultures (<https://www.dsmz.de/catalogues/catalogue-microorganisms.html>). The *T. thermophilus* and *V. cholerae* sequences were codon-optimised for expression in *E. coli* and made synthetically (GeneArt). The final list of homologues that were successfully cloned and their properties are shown in Table 3.1.

Chapter 3

Table 3.1. List of organisms from which ComEC homologues were cloned.

Organism	Phylum (class)	Cell envelope	Optimal growth temperature	Contains CTD?	No. of residues	Shown to be naturally competent? ^a	DSMZ Strain Number ^b
<i>Bacillus halodurans</i>	Firmicutes (Bacilli)	Gram-positive	30°C	Yes	776	No	DSM 18197
<i>Carboxydotherrnus hydrogenofomans</i>	Firmicutes (Clostridia)	Gram-positive	65°C	Yes	725	No	DSM 6008
<i>Desulfurobacterium thermolithotrophum</i>	Aquificae (Aquificae)	Gram-negative	70°C	No	483	No	DSM 11699
<i>Geobacillus thermodenitrificans</i>	Firmicutes (Bacilli)	Gram-positive	60°C	Yes	748	No	DSM 465
<i>Hydrogenobacter thermophilus</i>	Aquificae (Aquificae)	Gram-negative	70°C	No	421	No	DSM 6534
<i>Klebsiella oxytoca</i>	Proteobacteria (Gammaproteobacteria)	Gram-negative	28°C	Yes	751	No	DSM 7342
<i>Klebsiella pneumoniae</i> subsp. <i>rhinoscleromatis</i>	Proteobacteria (Gammaproteobacteria)	Gram-negative	37°C	Yes	741	No	DSM 16231
<i>Moorella thermoacetica</i>	Firmicutes (Clostridia)	Gram-positive	55°C	Yes	809	No	DSM 521
<i>Thermoanaerobacter brockii</i> subsp. <i>finnii</i>	Firmicutes (Clostridia)	Gram-positive	65°C	Yes	757	No	DSM 3389
<i>Thermus igniterrae</i>	Deinococcus-Thermus (Deinococci)	Gram-negative	70°C	Yes	686	No	DSM 12459
<i>Thermus islandicus</i>	Deinococcus-Thermus (Deinococci)	Gram-negative	65°C	Yes	675	No	DSM 21543

Chapter 3

<i>Thermus scotoductus</i>	Deinococcus-Thermus (Deinococci)	Gram- negative	65°C	Yes	687	No	DSM 8553
<i>Thermus thermophilus</i>	Deinococcus-Thermus (Deinococci)	Gram- negative	70°C	Yes	677	Yes	DSM 7039
<i>Vibrio cholerae</i>	Proteobacteria (Gammaproteobacteria)	Gram- negative	28°C	Yes	748	Yes	DSM 106276

^a As reviewed by (Johnston *et al*, 2014)

^b Detailed information, including alternative strain names, can be found in the DSMZ microorganism catalogue (<https://www.dsmz.de/catalogues/catalogue-microorganisms.html>).

3.3. Testing ComEC homologue expression in *E. coli*

The coding sequences of the selected ComEC homologues were either amplified from genomic DNA or made synthetically and ligated in to the pET28a(+)-based expression vector pWaldo (Waldo *et al*, 1999). The pWaldo vector features IPTG-inducible expression of the inserted gene and fuses at the C-terminus of the expressed protein a tobacco etch virus (TEV) protease cleavage site followed by superfolder green fluorescent protein (sfGFP) and then an octa-histidine (His₈)-tag (Waldo *et al*, 1999; Drew *et al*, 2006) (Figure 3.1). The superfolder variant of GFP was used because the C-terminus of ComEC is likely located in the periplasm where commonly used GFP variants do not fold (Pédelacq *et al*, 2006; Dinh & Bernhardt, 2011).

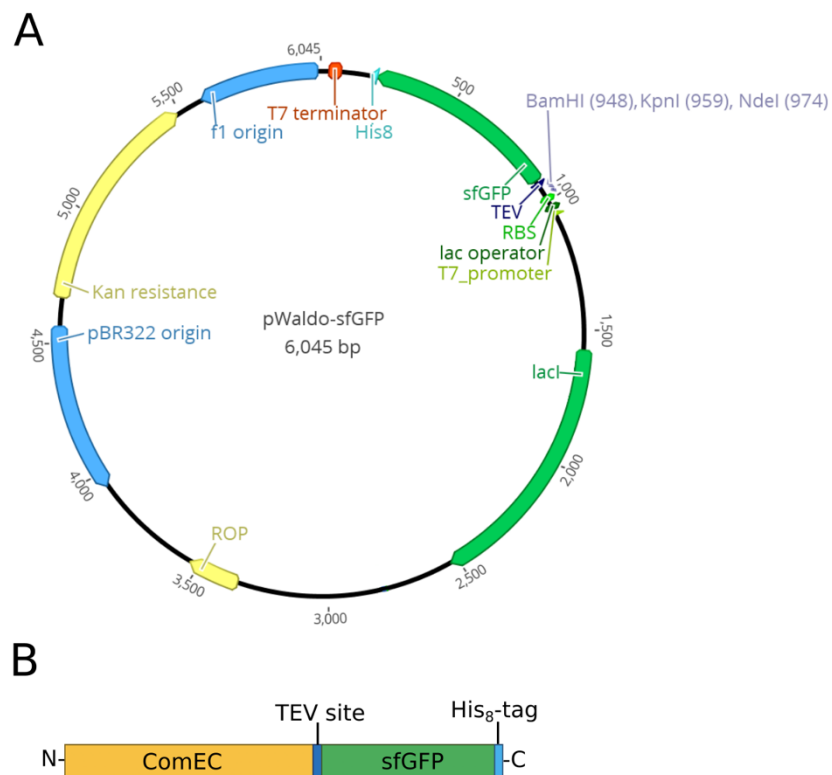


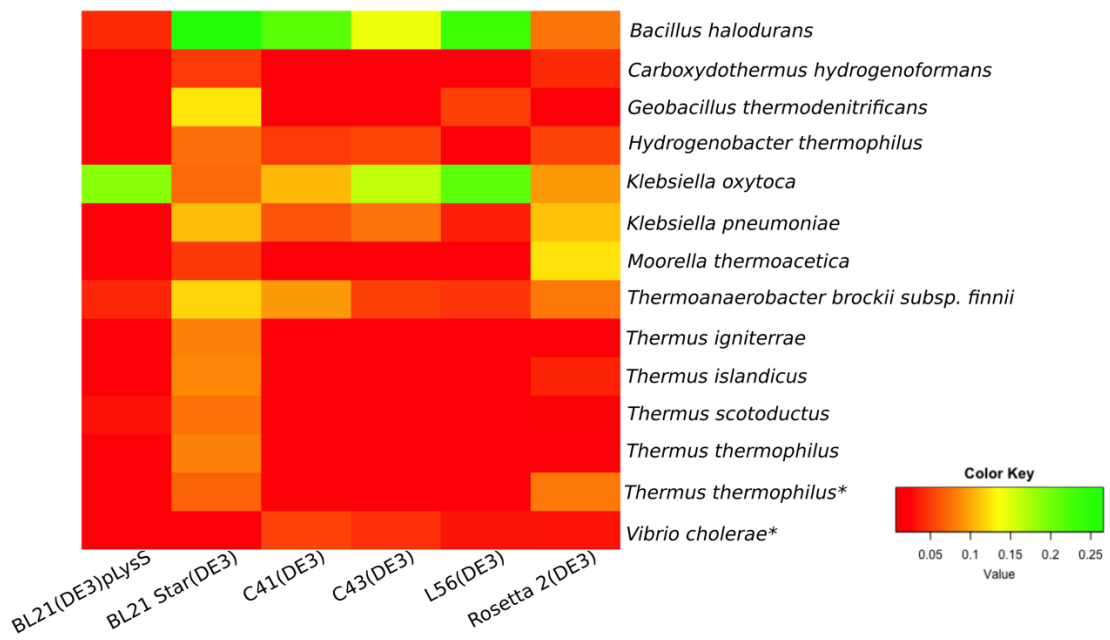
Figure 3.1. Expression vector and construct schematics. (A) Map of pWaldo-sfGFP plasmid with annotated features. ComEC nucleotide sequences were cloned in using NdeI/BamHI or NdeI/KpnI restriction sites. (B) ComEC-sfGFP fusion construct schematic.

Chapter 3

Expression plasmids harbouring the ComEC homologue coding sequences were tested in six *E. coli* expression strains with different characteristics detailed in Chapter 2. After overnight induction, expression was monitored by measuring the ratio of sfGFP fluorescence to OD₆₀₀ (Figure 3.2A). These ratios were then normalised to the fluorescence/OD₆₀₀ reading from *E. coli* L56 (DE3) cells expressing a TatC membrane protein GFP fusion at levels suitable for crystallization (Rollauer *et al*, 2012). Most homologue/strain combinations showed no or very poor expression. Promising homologue/strain combinations were analysed by Western blotting (Figure 3.2B). The ideal homologue/strain combination would have a high level of sfGFP fluorescence and show a single band corresponding to full-length protein on a Western blot probed with an anti-GFP antibody. However, in practice the latter criterion is rarely satisfied because flexible linkers between proteins and tags are susceptible to proteolysis, resulting in significant amounts of free sfGFP (Figure 3.2B). Therefore, the ratio of full-length protein to degradation products (bands lower than the expected full-length molecular weight) seen in Western blots was taken in to account when selecting which conditions to follow up.

Several homologue/strain combinations were identified that merited further characterisation: *Klebsiella oxytoca* ComEC in strain C43 (DE3), *Bacillus halodurans* ComEC in strain BL21 Star (DE3), *Moorella thermoacetica* ComEC in strain Rosetta 2 (DE3), and *Desulfurobacterium thermolithotrophum* ComEC in strain BL21 Star (DE3) (not shown in Figure 3.2). Unfortunately, upon scale-up of the cultures some combinations no longer produced full-length protein (*B. halodurans*, *D. thermolithotrophum*).

A



B

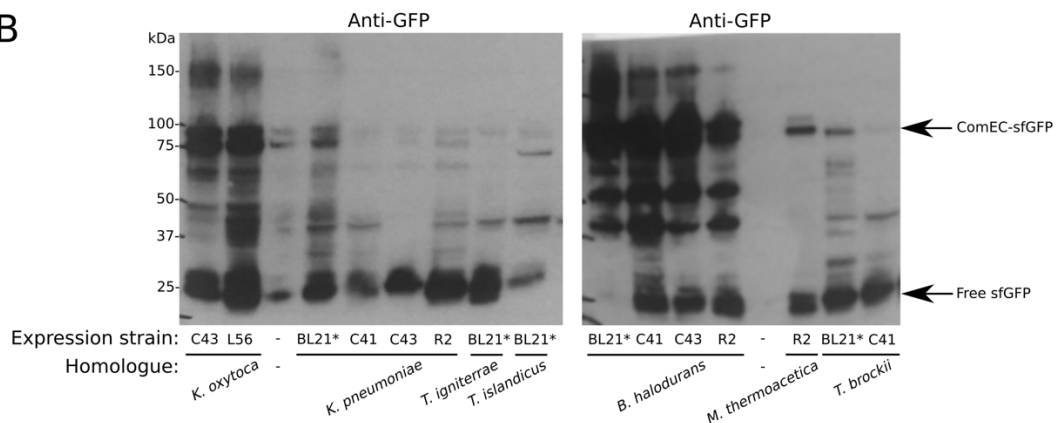


Figure 3.2. ComEC expression test. (A) Ratio of sfGFP fluorescence to OD₆₀₀ for different homologue/strain combinations from a representative expression test. For each combination, overnight cultures were diluted and grown in TB medium with appropriate antibiotics for 4 h at 37°C with 225 rpm shaking, induced with 0.1 mM IPTG and grown overnight at 18°C with 225 rpm shaking. Cells were harvested by centrifugation and resuspended in PBS. Fluorescence and OD₆₀₀ readings were taken in 96-well format in a CLARIOstar Plus plate reader. The data were normalised to the fluorescence/OD₆₀₀ reading from *E. coli* L56 (DE3) cells expressing GFP-tagged TatC from *Aquifex aeolicus* (Rollauer *et al*, 2012). The source organisms of the ComEC homologues are presented on the vertical axis. The *E. coli* expression strains are shown on the horizontal axis. Homologues marked with an asterisk were codon-optimised for expression in *E. coli*. (B) Representative anti-GFP blots of whole cells solubilized in SDS sample buffer. The homologue/strain combinations correspond to those in (A). The same number of cells was solubilized and loaded in each lane.

Subsequent sections will focus mostly on ComEC from *M. thermoacetica* (from now on termed MthComEC) expressed in *E. coli* Rosetta 2 (DE3) cells. This combination showed a moderate expression level and only two prominent bands on the Western blot—one for full-length protein and one for free sfGFP (Figure 3.2B).

3.4. Detergent screening

FSEC was used to identify a suitable detergent for solubilisation of the sfGFP-tagged MthComEC protein from *E. coli* membranes (Kawate & Gouaux, 2006). Crude detergent-solubilized membranes were analysed. Ten detergents with a wide range of chemical properties were screened (see Chapter 2). MthComEC solubilised in C₁₂E₉, GDN, digitonin, Cymal-6-NG or LMNG gave a symmetrical gel filtration peak indicating that the sample was monodisperse (Figure 3.3A). LMNG and C₁₂E₉ were taken forward for purification trials, but only LMNG (Figure 3.3B) was picked as the final working detergent because it had a higher solubilisation efficiency (50-80% versus 30-50% for C₁₂E₉) and the protein tended to precipitate less during purification in LMNG.

3.5. Developing a purification protocol for ComEC

The workflow used for purification is shown in Figure 3.4. Isolated *E. coli* membranes were solubilised in detergent and subjected to an affinity purification step. The sample was then treated with TEV protease to cleave the site between ComEC and the sfGFP-His₈ tag. The subsequent reverse affinity purification step removed the released tag, any remaining uncleaved tagged protein and contaminants that interact with the affinity resin, while the released product was collected in the flow-through. The flow-through was then concentrated and subjected to size exclusion chromatography as a polishing step as well as acting as quality control of the dispersity of the final product.

Chapter 3

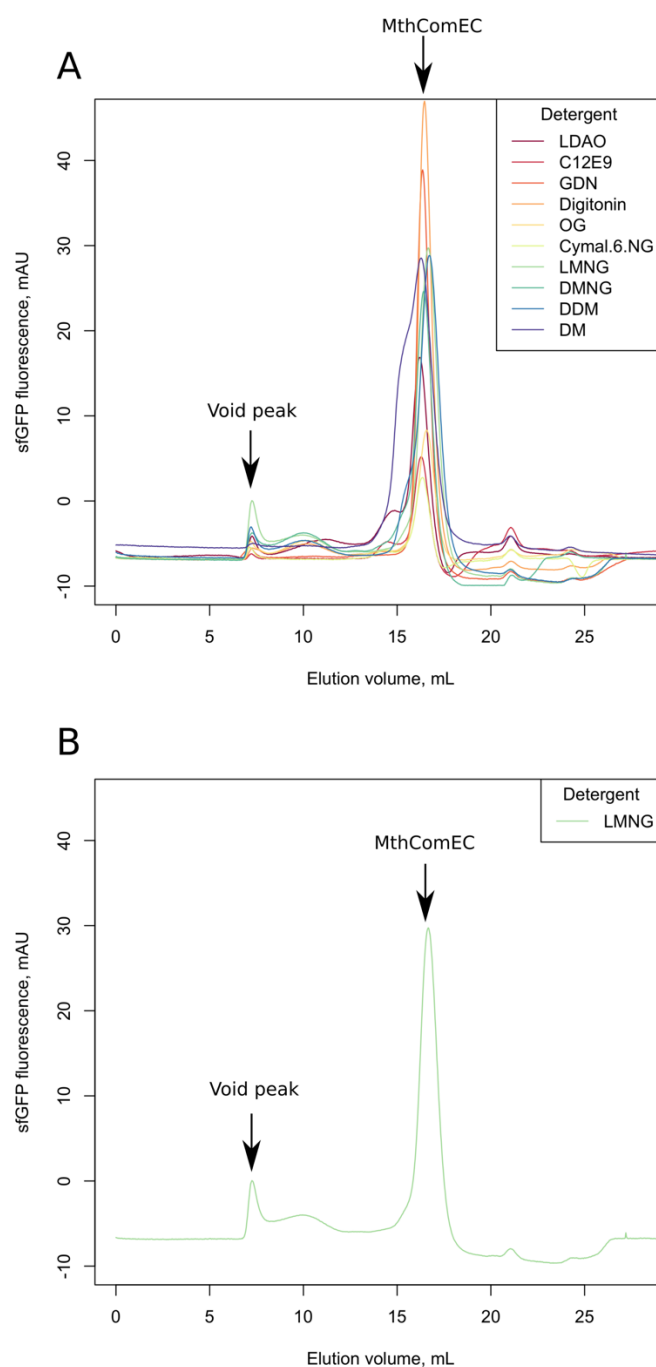


Figure 3.3. Detergent screen for MthComEC solubilization. (A) Superose 6 10/30 column FSEC traces from MthComEC-sfGFP-containing membranes solubilized in ten different detergents. The membranes were isolated from a culture of *E. coli* Rosetta 2 (DE3) cells expressing MthComEC-sfGFP, split into ten equal aliquots, solubilised in the different detergents as described in Chapter 2. 50 μ l samples were injected onto the size exclusion column using an autosampler. (B) LMNG FSEC trace shown for clarity.

Chapter 3

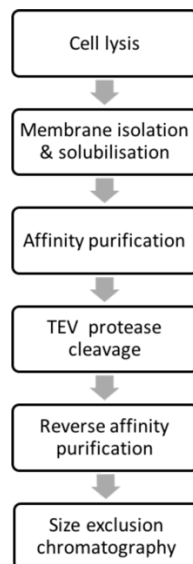


Figure 3.4. ComEC purification workflow. Membranes were solubilised in either LMNG or DDM, as explained in the text. After solubilisation, the insoluble material was removed by ultracentrifugation. The affinity steps were performed either with NiNTA or StrepTactin columns.

Initial purification of MthComEC and KoxComEC

Initially, MthComEC and *K. oxytoca* ComEC (KoxComEC) were purified using this workflow but skipping the TEV cleavage and reverse affinity steps. KoxComEC samples showed degradation in Western blots and formed aggregates, as shown by size exclusion chromatography, when purified in DDM and LMNG (Figure 3.5A and B). This homologue was not pursued further. The initial attempts to purify MthComEC in LMNG produced preparations that showed degradation and aggregation as well, but full-length protein that did not elute in the column void peak was present (Figure 3.5C). The purification protocol, therefore, had to be optimised to isolate the full-length protein.

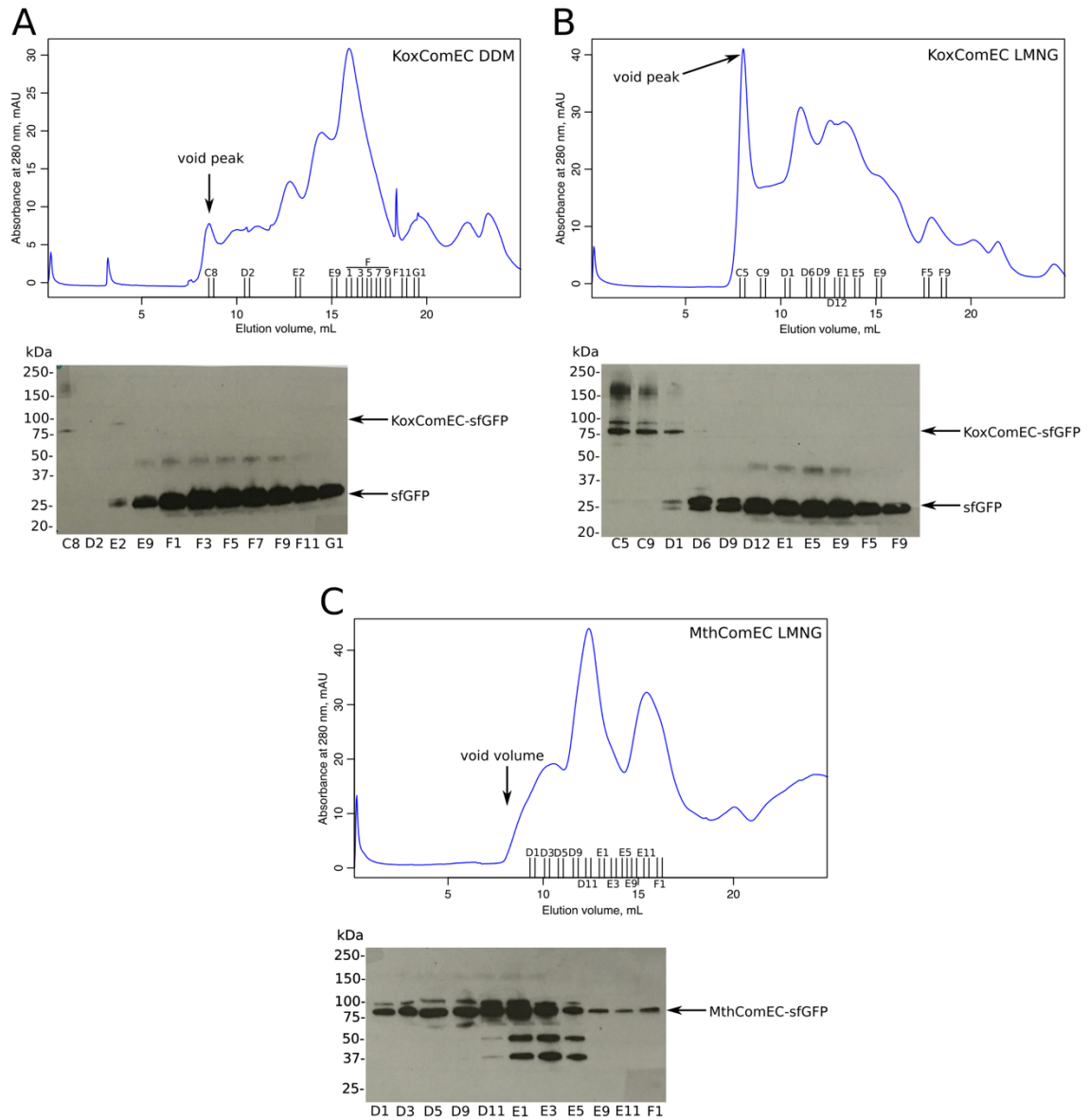


Figure 3.5. Initial attempts at KoxComEC and MthComEC purification. Superdex 200 10/300 GL size exclusion chromatography traces of KoxComEC purified in DDM (A), KoxComEC purified in LMNG (B) and MthComEC purified in LMNG (C). Anti-GFP Western blots of indicated fractions from the size exclusion runs are shown below each trace. Membranes with KoxComEC and MthComEC containing ~250 mg and ~170 mg total protein, respectively, were used for solubilisation. Solubilised samples were purified using a NiNTA affinity step followed by size exclusion chromatography.

MthComEC co-purifies with the expression host Sec machinery

The His₈-tag at the C-terminus of the sfGFP fusion was used initially for the affinity purification step in the MthComEC purification. The protein eluted with several contaminants (Figure 3.6A). The contaminants did not bind to the NiNTA resin in the reverse affinity purification step and could not be separated from MthComEC using size exclusion chromatography (Figure 3.6C and D). The contaminants were identified by peptide fingerprinting mass spectrometry as expression host proteins SecA, SecY, SecE and SecG. These proteins form the Sec translocon which is involved in membrane protein biogenesis (Tsirigotaki *et al*, 2016). With the exception of SecA, the presence and amount of the contaminant proteins differed substantially between purifications.

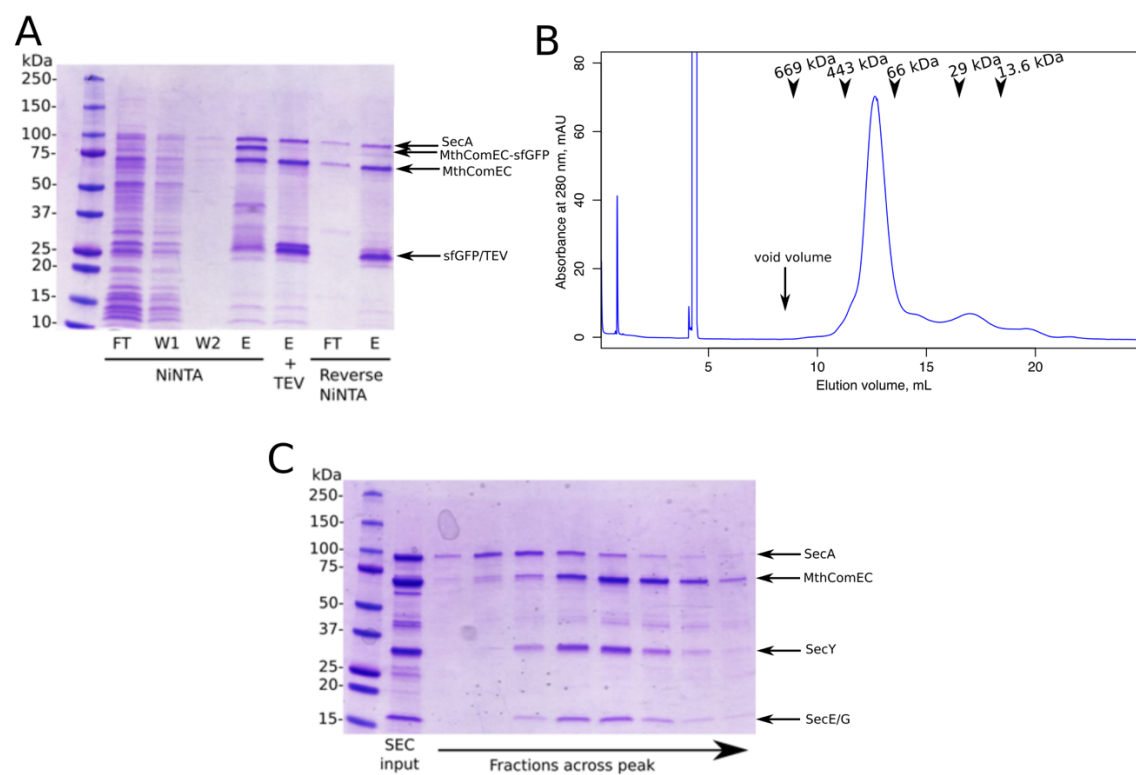


Figure 3.6. Purification of MthComEC-sfGFP-His₈. (A) SDS-PAGE gel of steps from the NiNTA purification, TEV protease cleavage and reverse NiNTA purification. FT, flow-through; W, wash; E, eluate. (B) Superdex 200 10/300 GL size exclusion chromatography trace of the concentrated flow-through from the reverse affinity purification. Markers above the trace indicate the elution volumes of standard soluble proteins of known molecular weight. (C) SDS-PAGE gel of fractions from (D).

In the attempt to obtain a pure ComEC preparation, many steps of the purification protocol were varied, starting with the cell density at which the cells were induced for ComEC expression and the duration of the expression period. Washing the crude membrane fraction at higher salt concentrations or with pH 11.0 buffer was trialled. Different membrane protein-to-detergent ratios were tested for membrane solubilisation. Conditions for optimizing ComEC-sfGFP binding to the affinity resin in both the initial and reverse affinity steps were also extensively tested by varying the pH, temperature, salt concentration, detergent concentration during binding, batch binding *versus* column binding, binding time, and buffering reagent. An ion exchange chromatography step after the reverse affinity step was also trialled but did not significantly improve the purity of the preparation and resulted in extremely low yields (data now shown). The His₈-tag on sfGFP was exchanged for a TwinStrep-tag to allow affinity purification using StrepTactin resin. However, this construct also co-purified with the Sec machinery indicating that ComEC interacts directly with the Sec translocon components and that their co-purification is not the result of non-specific interactions with the affinity matrix.

The general secretory (Sec) protein export pathway is used by most proteins that either integrate into or translocate through the *E. coli* cytoplasmic membrane (Tsirigotaki *et al*, 2016) and so it can be assumed that MthComEC uses the Sec pathway for membrane insertion. SecYEG comprise the transmembrane pore-forming translocon complex whilst the cytoplasmic ATPase SecA is involved in energising the transport of soluble protein domains through the SecYEG channel. Attaching a sfGFP fusion to the extracytoplasmic C-terminal metallo- β -lactamase-like domain of ComEC substantially increases the amount

Chapter 3

of protein that needs to be threaded through the SecYEG pore into the periplasm, potentially jamming the Sec machinery. Therefore, it was hypothesised that removing the sfGFP domain would prevent jamming. A MthComEC construct was cloned in which the TEV cleavage site, the sfGFP domain, and the His₈-tag were replaced with a C-terminal StrepTag II. Using this expression construct MthComEC was completely pure after affinity purification using a StrepTactin resin (Figure 3.7A). However, the yield was extremely low and size exclusion chromatography showed significant aggregation (Figure 3.7B and C). This suggests that MthComEC is stabilised by the sfGFP tag during expression – possibly by association with the jammed Sec machinery.

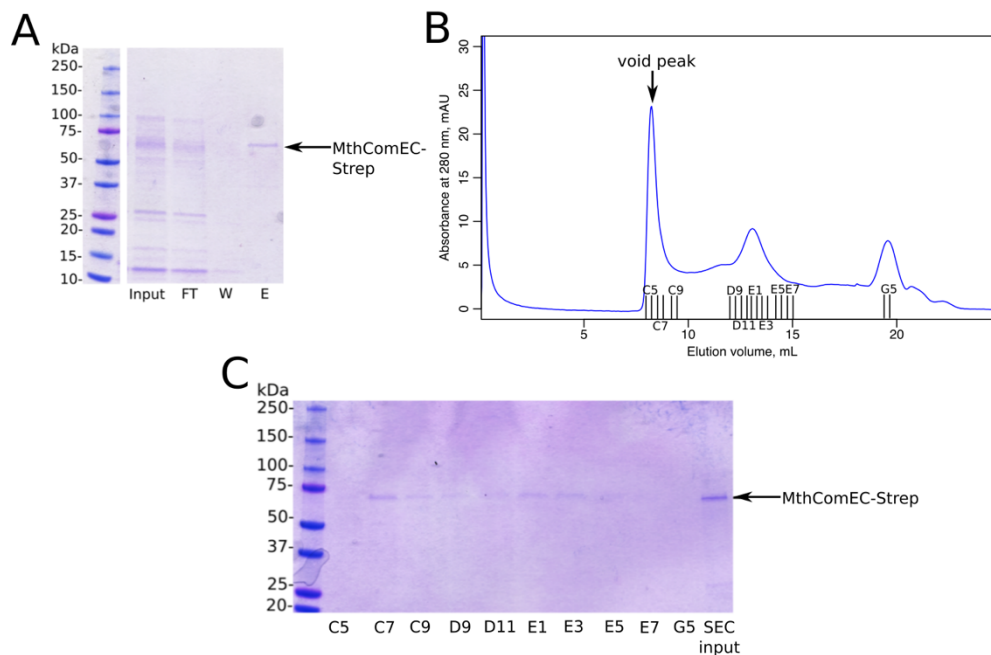


Figure 3.7. Purification of MthComEC-Strep. (A) SDS-PAGE gel of samples from different stages of StrepTactin affinity purification. FT, flow-through, W, wash, E, eluate. (B) Superdex 200 10/300 GL size exclusion trace of the concentrated eluate from (A). (C) SDS-PAGE gel of fractions from (B).

Removal of Sec contaminants.

It was next hypothesised that association of SecA with MthComEC may represent a substrate-bound stage of the SecA peptide translocation cycle. If this is the case, then provision of ATP and Mg^{2+} might allow SecA to complete the cycle and dissociate from MthComEC. To test this proposal, the sfGFP-TwinStrep-tagged construct was solubilised in LMNG and allowed to bind to a StrepTactin column. The column was washed with binding buffer, and then with buffer containing 5 mM ATP and 5 mM MgCl_2 . Most of SecA and SecY was found in the wash flow-through (Figure 3.8A). The peak of MthComEC in the subsequent size exclusion step eluted 0.4 ml later than material purified without the MgATP wash, and now comprised highly pure protein (Figure 3.8B and C). The identity of MthComEC was confirmed by mass spectrometry and the results are detailed in section 3.6. The approximate molecular weight of the pure species, as estimated from a standard curve of soluble proteins, was 160 kDa. This optimised purification has been reproduced more than ten times during the course of the current work.

Nevertheless, the overall MthComEC yield remained very low at 30-50 μg of pure protein per litre of cell culture. In an attempt to alleviate this problem the expression conditions were optimised extensively by varying the cell density at induction, inducer concentration, time and temperature of induction, but without improvement in yield. Next, a synthetic *M. thermoacetica* ComEC gene that is codon-optimised for expression in *E. coli* was tested. However, purified protein yields were no better than when the protein was expressed using the original coding sequence. An alternative rhamnose-inducible expression system was also tested but gave even worse expression than the IPTG-inducible system. The detergent-to-protein ratio used to extract the protein and the

binding time to affinity resin was optimised. However, even after all these improvements the protein yield did not exceed around 50 µg pure MthComEC per litre of cell culture.

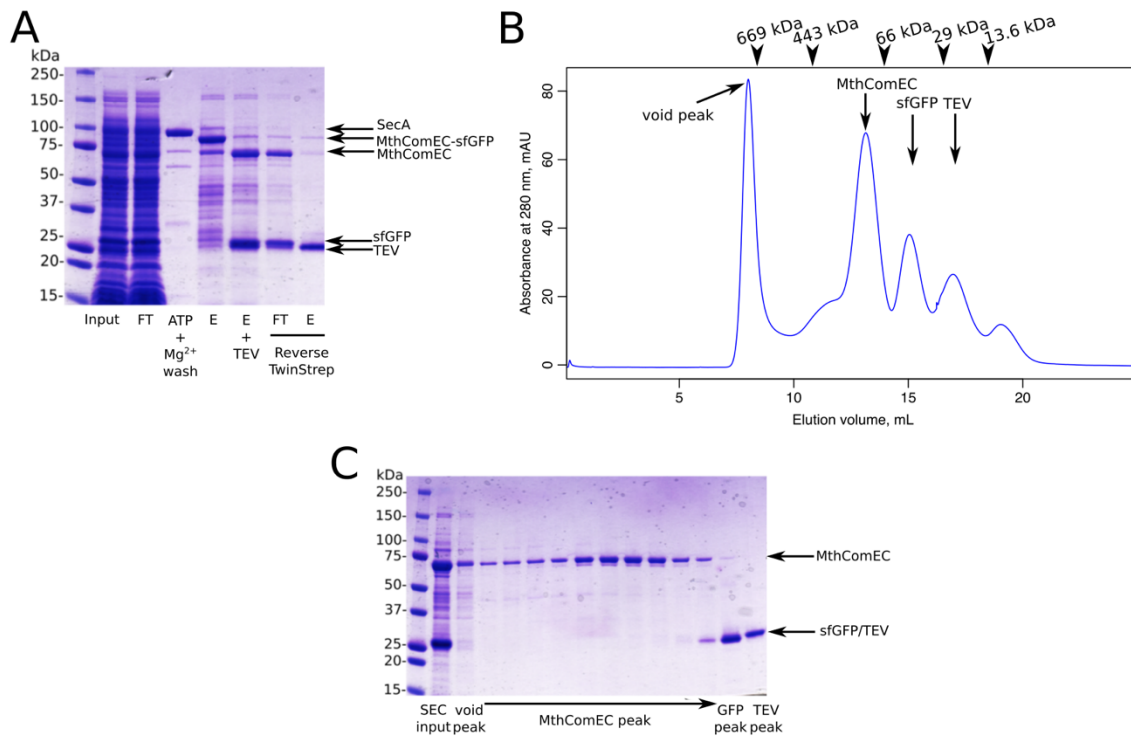


Figure 3.8. MthComEC-sfGFP-TwinStrep purification with MgATP wash. (A) Coomassie gel of samples from different steps of StrepTactin affinity purification, TEV protease cleavage, and reverse affinity purification. FT, flow-through; E, eluate. Most of SecA dissociates from MthComEC and is found in the ATP+Mg²⁺ wash fraction. (B) Superdex 200 10/300 GL size exclusion trace of the reverse affinity purification flow-through. (C) SDS-PAGE gel of peak fractions from (B). Note that the MthComEC peak here elutes about 0.4 ml later than the MthComEC/Sec peak in Figure 3.6C. Markers above the trace indicate the elution volumes of standard soluble proteins of known molecular weight.

3.6. Mass spectrometry of MthComEC

The identity of purified MthComEC was confirmed by mass spectrometry of trypsin-digested SDS-PAGE gel bands (Figure 3.9). Peptides corresponding to most of the N- and C-terminal soluble domains were identified (residues 83-262 and 539-809, respectively). No peptides were identified for the extreme N-terminal transmembrane helices (residues

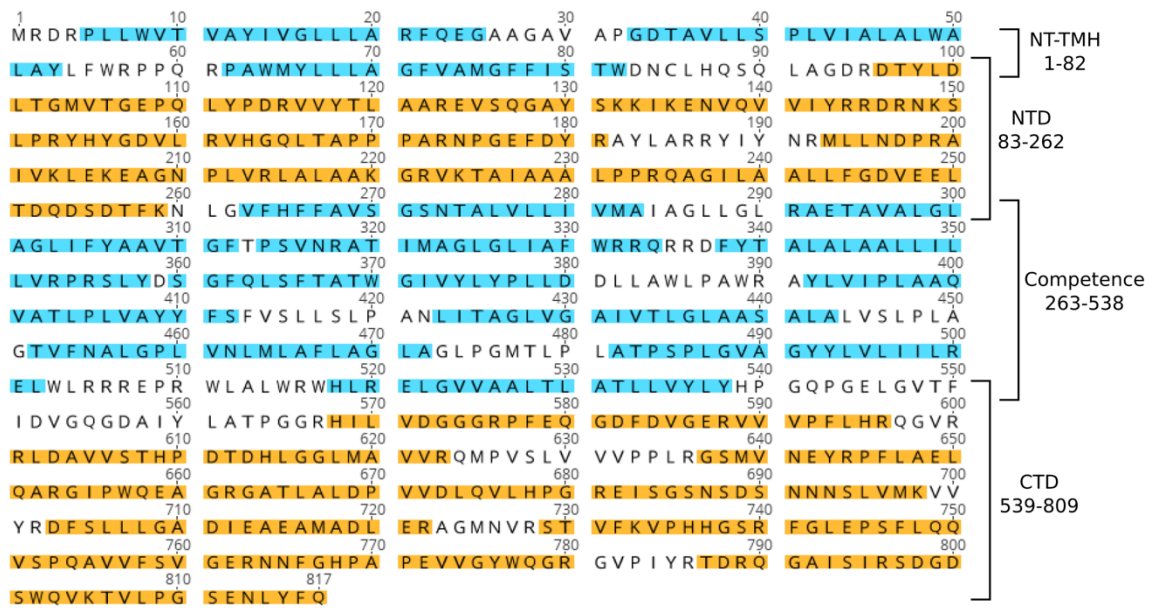


Figure 3.9. Peptide coverage of MthComEC. A SDS-PAGE gel band containing MthComEC was cut out, digested with trypsin at 37°C overnight and submitted for mass spectrometry analysis. Regions of MthComEC that were identified in the mass spectrometer are highlighted in orange. The predicted transmembrane helices of MthComEC are highlighted in blue.

1-82) and for the competence domain (residues 263-538). This is probably because the transmembrane segment-rich regions do not generate many tryptic peptides of an appropriate size for mass spectrometry experiments and because the hydrophobic peptides from the transmembrane regions might not be detected in the mass spectrometer. As peptides for the NTD and CTD were identified from a single gel band, it can be assumed that the intervening competence domain is also present.

3.7. *N-terminus of MthComEC is cleaved*

It was observed that MthComEC migrates faster than expected in SDS-PAGE gels, with an apparent molecular weight of around 70 kDa rather than the expected 89 kDa (as seen in gels in figures 3.6-3.8). Although membrane proteins often migrate abnormally on SDS-PAGE gels, this observation warranted further exploration. The mass spectrometry and SDS-PAGE data could account for the presence of soluble domains and the competence

domain, but not for the presence of the N-terminal transmembrane helices. The MthComEC-sfGFP-His₈ construct with an additional StrepTag II at the N-terminus was made. The N-terminal StrepTag II was not detected in a Western blot probed with StrepTactin-HRP (Figure 3.10A), but the C-terminal sfGFP tag was detected with an anti-GFP antibody (Figure 3.10B). This showed that the construct with N- and C-terminal tags was expressed but the N-terminal StrepTag II was not present, suggesting that the N-terminus of MthComEC is processed in *E. coli*. It is possible that the host signal peptidase I interprets the alanine-rich amino acid sequence between the first and second transmembrane helices as a signal cleavage site (Figure 3.9)(Auclair *et al*, 2012). MthComEC purified as in Figure 3.8 was sent for N-terminal sequencing by Edman degradation but no sequence was obtained, possibly because the N-terminus is modified or because the N-terminus starts with two consecutive proline residues. Thus, the exact N-terminal amino acid residue of purified MthComEC remains unknown.

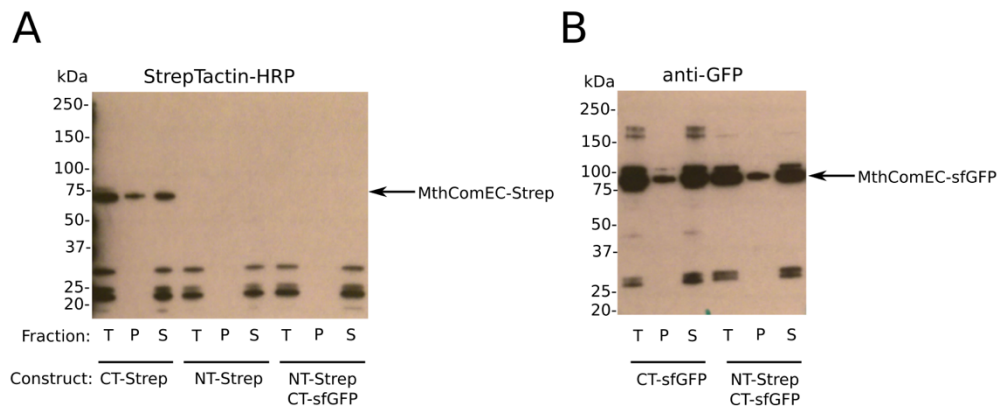


Figure 3.10. The N-terminus of MthComEC is cleaved. Cells expressing the following constructs were lysed: MthComEC with a C-terminal StrepTag II, MthComEC with an N-terminal StrepTag II, MthComEC-sfGFP-His₈ and MthComEC-sfGFP-His₈ with an additional N-terminal StrepTag II. The lysate (T) was subjected to low speed centrifugation to separate inclusion bodies (P) and the soluble fraction (S). Expression conditions were the same as for MthComEC-sfGFP-TS (Chapter 2). The fractions were analysed in Western blots probed with StrepTactin HRP (A) and anti-GFP (B). The CT-Strep construct and CT-sfGFP constructs are positive controls for probing.

3.8. *MthComEC* SEC-MALLS analysis

Size exclusion chromatography with multiangle laser light scattering (SEC-MALLS) also indicated that the calculated protein mass of *MthComEC* was smaller than expected (around 70 kDa at the centre of the peak vs 89.3 kDa expected) (Figure 3.11). SEC-MALLS revealed variability in protein mass across the absorbance peak, indicating that the sample is heterogeneous and, therefore, unlikely to crystallize. This was unexpected because *MthComEC* is very pure as seen in SDS-PAGE gels. The mass of *MthComEC* is consistent with that of a monomer. The N-terminal region not detected in mass spectrometry analysis corresponds to a molecular weight of 10.5 kDa. If the entire region was missing in the pure *MthComEC* sample, the molecular weight would be 78.8 kDa, which is close to the average protein mass indicated by SEC-MALLS.

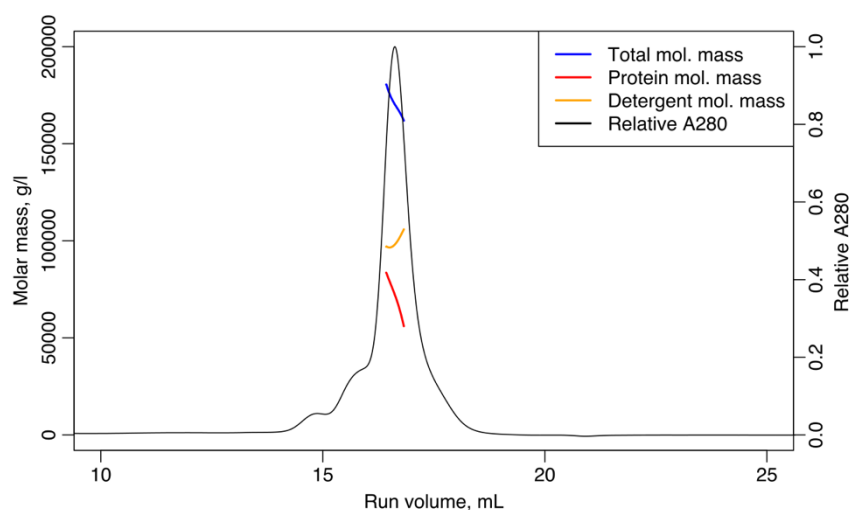


Figure 3.11. *MthComEC* SEC-MALLS analysis. The SEC-MALLS analysis was carried out by Steven Johnson at the Sir William Dunn School of Pathology (University of Oxford). A chromatography system with a Superdex 200 10/300 GL size exclusion column connected to a multiangle light photometer (Dawn Heleos, Wyatt) was used. Light scattering data were used to calculate the total molar mass (blue) at any point in the elution profile. Protein-detergent conjugate analysis was performed to deconvolute the mass contributions of detergent (orange) and protein (red). Purified *MthComEC* at 1.0 mg/ml was analysed. The mobile phase was TBS with 0.02% LMNG.

3.9. *MthComEC limited proteolysis*

MthComEC was assayed by limited proteolysis to identify any regions that might be misfolded and, therefore, susceptible to proteolysis. Digestion with trypsin showed that purified MthComEC is readily cleaved into two fragments of around 16 kDa (EC_16) and 60 kDa (EC_60) (Figure 3.12A). Gel bands containing the uncleaved MthComEC, EC_60 and EC_16 were cut out, subjected to in-gel trypsin digestion and the tryptic peptides were identified by mass spectrometry (Figure 3.12B). Peptides from EC_16 covered residues 96-144, which form part of the NTD. Peptides from EC_60 covered residues 154-234 and 568-817, corresponding to the rest of the NTD and the CTD, respectively. No peptides corresponding to the extreme N-terminal transmembrane helices and the competence domain were identified, due to reasons explained in section 3.6. The results suggest that EC_60 contains the NTD from residue 154, the competence domain and the CTD. The NTD region contained in EC_16 (residues 83-144) has a molecular weight of 7 kDa, or less than half that of the EC_16 fragment as estimated from SDS-PAGE analysis. Therefore, the EC_16 fragment and, by extension, MthComEC must contain at least some part of the N-terminal transmembrane helical region.

The region between residues 144 and 154 was susceptible to protease cleavage. Phyre2 modelling predicts that NTD residues 99-167 form an OB fold, but the fold for the rest of the NTD is unknown (Chapter 1, Figure 1.3). It is unclear whether the 144-154 region is accessible to protease cleavage because it is at an interface between two folds or because the NTD is misfolded.

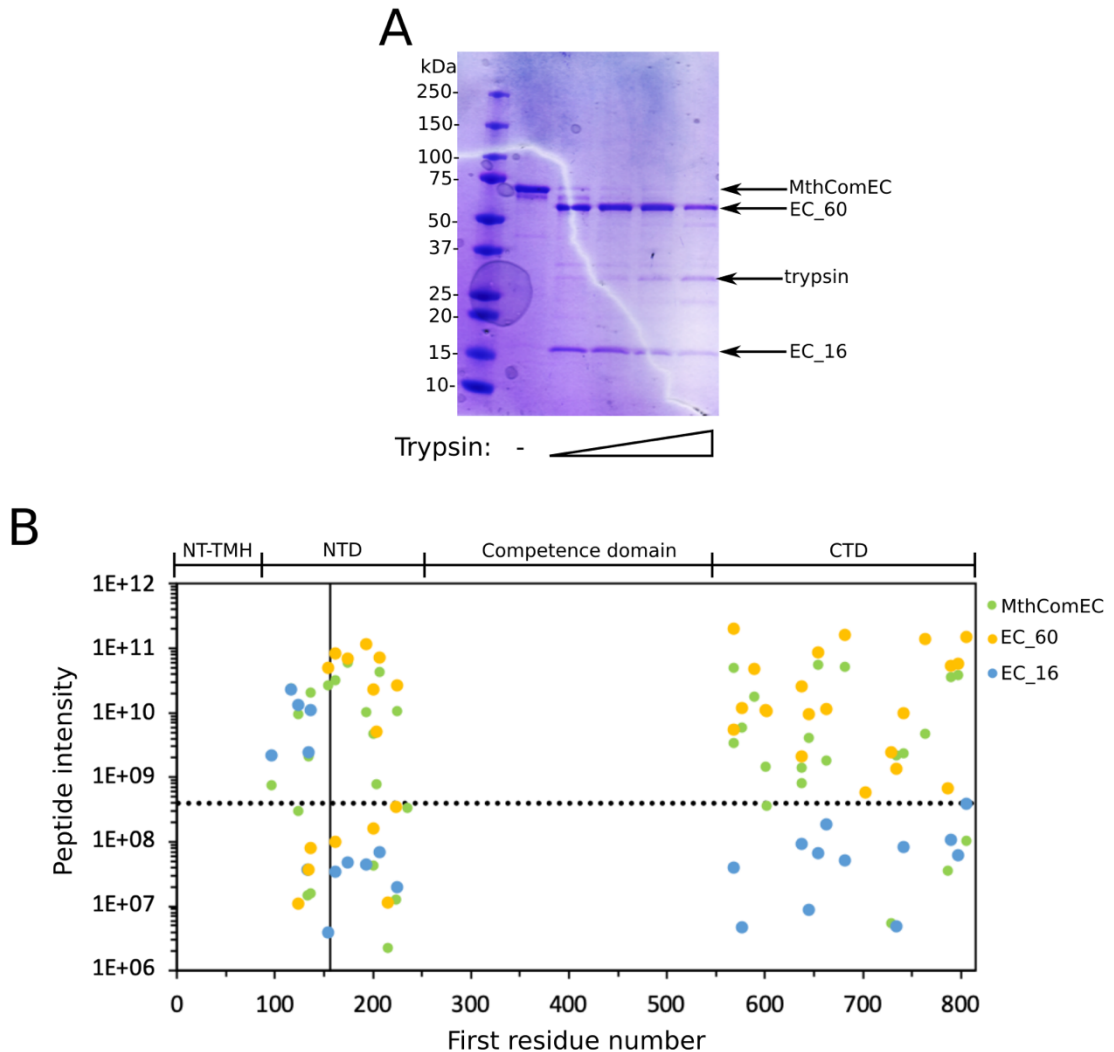


Figure 3.12. MthComEC limited proteolysis. (A) SDS-PAGE analysis of MthComEC protease digestion. Trypsin was added to 1.4 μ g of purified MthComEC at protease-to-protein mass ratios of 1:80, 1:40, 1:20 and 1:10 in TBS buffer with LMNG. The reactions were incubated for 30 min at room temperature and then quenched with SDS-PAGE sample buffer containing 10 mM EDTA and 3 mM PMSF, followed by heating at 90°C for 3 min and SDS-PAGE. MthComEC was readily cleaved by trypsin into ~60 kDa (EC_60) and ~16 kDa (EC_16) fragments. The experiment was repeated at the highest trypsin concentration, the sample was analysed by SDS-PAGE and the MthComEC, EC_60 and EC_16 bands were cut out and submitted for peptide fingerprinting mass spectrometry analysis. (B) Peptide intensity plotted against residue number for each gel band. The cut-off line for significance (dashed line) was chosen as the highest intensity of a CTD peptide identified in the EC_16 sample. The solid vertical line indicates residue number 154. Notably, no peptides were observed for the N-terminal transmembrane helices (NT-TMH) and the competence domain.

3.7. *MthComEC N-terminal deletion constructs*

The proteolysis and mass spectrometry data could indicate that the protease-susceptible N-terminal region is misfolded. Therefore, N-terminal truncation constructs were designed to test whether they express better than the full-length protein (Figure 3.13A). The constructs correspond to deletion of part of the N-terminal loop (construct $\Delta 83-156$), deletion of the N-terminal transmembrane helices and part of the N-terminal loop (construct $\Delta 2-156$), deletion of the whole N-terminal part of the protein up to the competence domain (construct $\Delta 2-259$). Around half of the constructs showed a similar level of expression to full-length MthComEC in an expression test (Figure 3.13B). Construct $\Delta 2-156$ expressed in BL21 Star (DE3) cells showed an approximately 13-fold increase in expression levels compared to wild type MthComEC. Most of the protein with the $\Delta 2-156$ truncation produced was present in the lysate supernatant rather than the low speed pellet containing inclusion bodies (Figure 3.13C). Isolated membranes were green in colour but conventional detergents such as LMNG and DDM were not able to solubilise this construct efficiently. Only the relatively harsh detergents Anzergent-14 and Cyclofos-6 showed efficient solubilization of the truncated MthComEC, suggesting that the construct might be misfolded (Figure 3.13D). When purified in Cyclofos-6, there was significant precipitation of the $\Delta 2-156$ construct during normal handling steps. The final product, however, had a good size exclusion chromatography profile (Figure 3.13E) and seemed larger (between 443 kDa and 669 kDa) than the full-length protein, which suggested that it might be an oligomer. Although this sample looked promising, negative stain electron microscopy revealed that it was very heterogeneous (see section 3.9), and characterisation of this construct was not pursued further.

Chapter 3

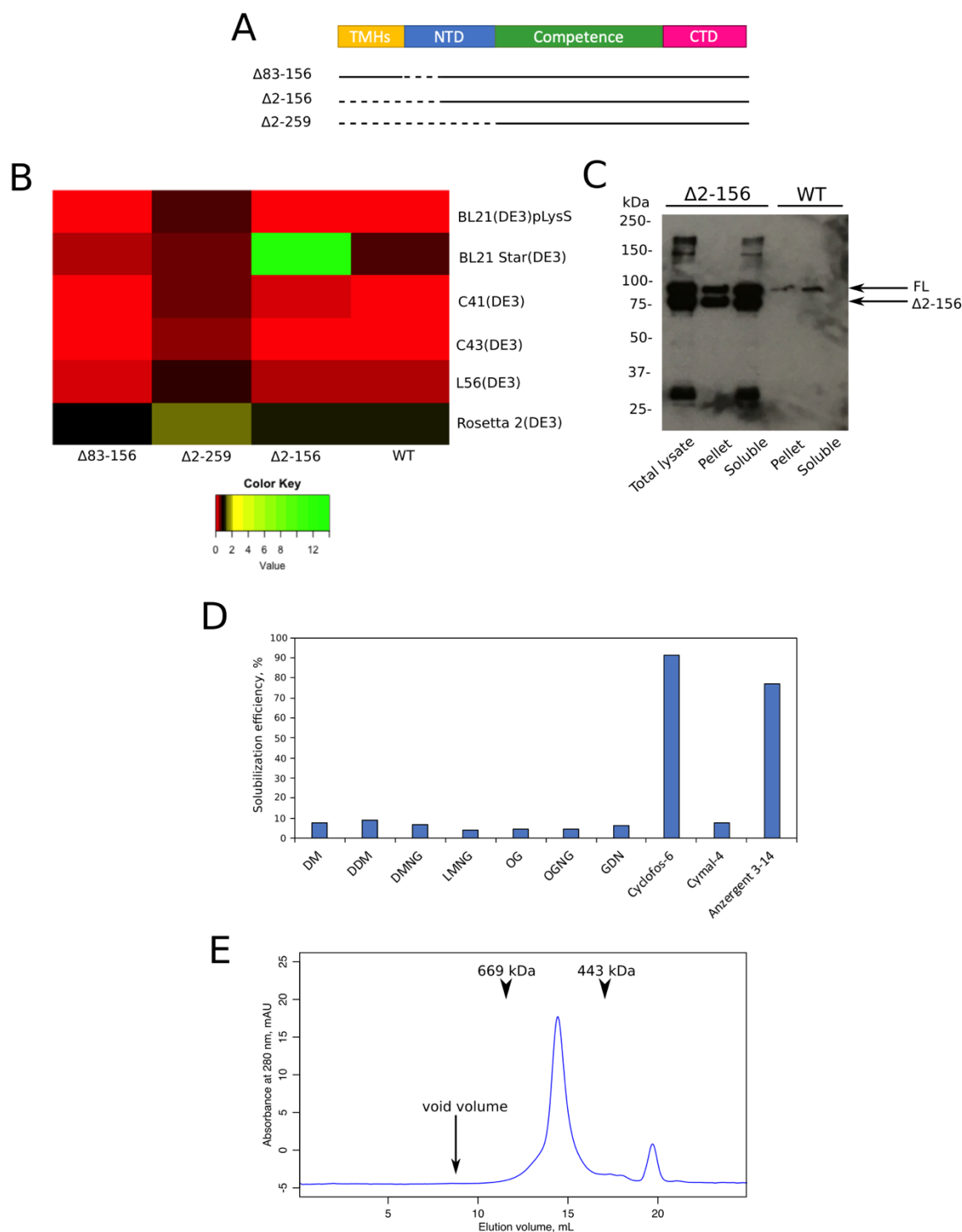


Figure 3.13. Expression and purification of N-terminal MthComEC truncations. (A) Schematic representation of the MthComEC N-terminal deletion constructs. The deleted regions are shown as dashed lines. (B) Ratio of sfGFP fluorescence to OD₆₀₀ for each N-terminal deletion construct/expressor strain combination. The expression test was done as detailed in Figure 3.2 and Chapter 2. The readings were normalised to that of wild-type MthComEC expressed in *E. coli* Rosetta 2 (DE3) cells. The colour key values represent fold-changes in expression. (C) Anti-GFP blot of lysate fractions from Δ2-156-expressing BL21 Star (DE2) cells and wild-type MthComEC-expressing Rosetta 2 (DE3) lysate fractions. Cells cultures expressing the constructs were adjusted to OD₆₀₀=1 and lysed by sonication (total lysate). The lysate was subjected to a low-speed

centrifugation step to separate inclusion bodies (Pellet) from the soluble fraction. Notably, the $\Delta 2$ -156 construct runs as a double band because of folded/unfolded states of sfGFP. (D) Solubilization efficiency test of MthComEC $\Delta 2$ -156 by various detergents. *E. coli* membranes expressing MthComEC $\Delta 2$ -156 were solubilised in the indicated detergents for 1 h at 4°C and sfGFP fluorescence readings were taken. Insoluble material was cleared by ultracentrifugation and fluorescence readings were taken again. The ratio of fluorescence after and before centrifugation is the solubilization efficiency. (E) Superose 6 10/300 size exclusion chromatography trace of MthComEC $\Delta 2$ -156 in Cyclofos-6.

3.9. Negative stain electron microscopy 3D reconstruction of MthComEC

MthComEC is the first ComEC protein successfully produced in a heterologous expression system. Crystallisation of purified MthComEC was not attempted because of the low yield of the purification. Instead, the sample was characterised by EM. Negative stain EM is a good method for characterising protein dispersity and overall shape before proceeding to more costly and time consuming cryo-EM data collection. A major concern with the suitability of MthComEC for EM was that it may be too small to be clearly seen in electron micrographs, since the detergent/protein complex is only around 160 kDa as determined by SEC-MALLS.

Figure 3.14A shows a representative electron micrograph of uranyl-acetate stained MthComEC with a reasonably homogeneous population of particles. The 2D class averages and the 3D model (Figure 3.14B and C, respectively) show that MthComEC has a round base, which possibly corresponds to the LMNG micelle-coated transmembrane region, and two protruding densities that are contacting each other, which might correspond to the soluble periplasmic domains.

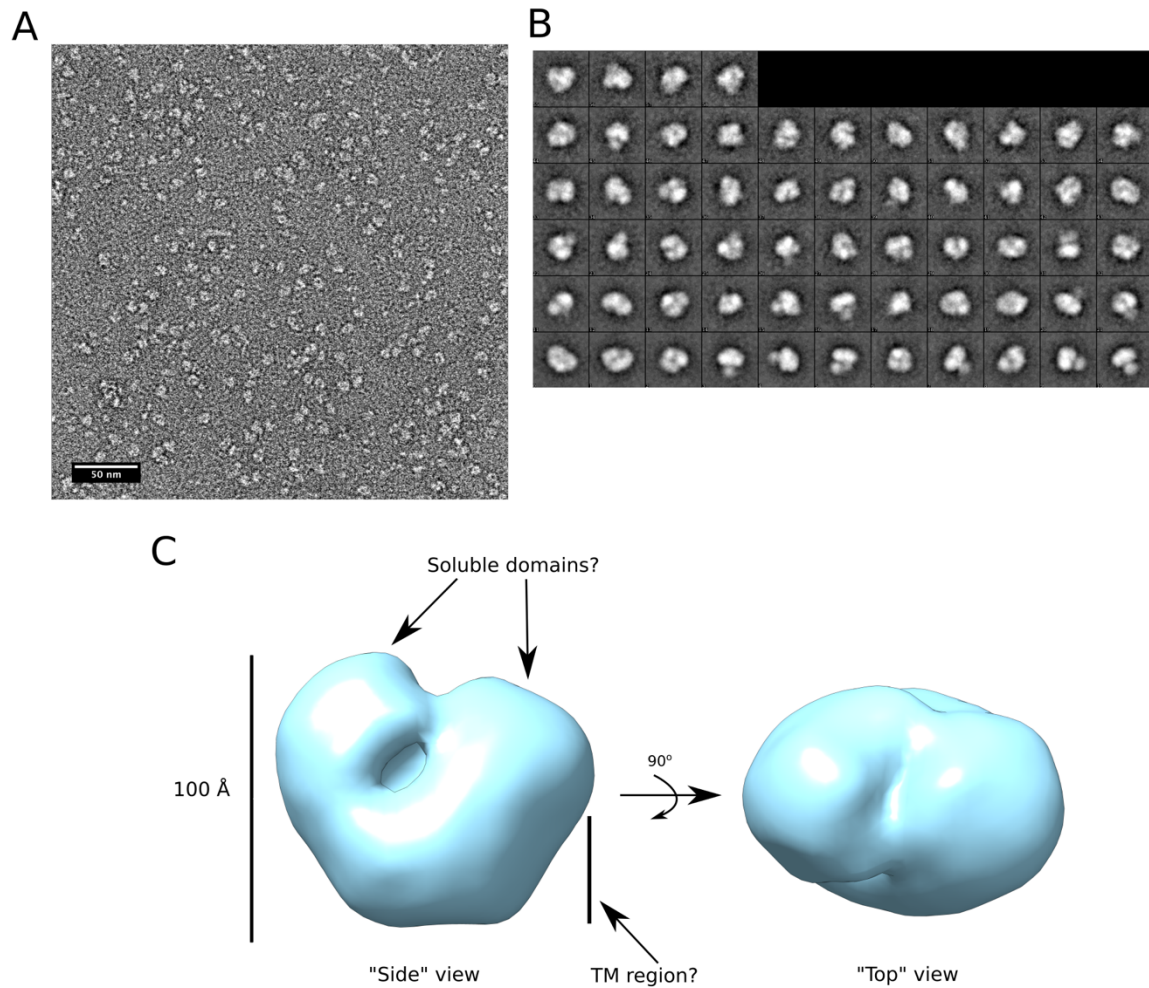


Figure 3.14. Negative stain EM of ComEC. (A) Representative micrograph of MthComEC stained with 2% uranyl acetate. Protein was applied at 0.5 $\mu\text{g/ml}$. Scale bar – 50 nm. A dataset of 54 micrographs was collected and 7,138 particles were picked. (B) Selected 2D class averages, which were used to reconstruct the 3D model (C). The “top” view is rotated 90° outwards from the page relative to the “side” view.

Chemical crosslinking

The GraFix (Gradient Fixation) EM sample preparation technique was used in an attempt to chemically fix the two soluble domains and reduce any possible interdomain flexibility (Kastner *et al*, 2007). The sample was applied to the top of a double glutaraldehyde and glycerol gradient and subjected to sedimentation in an ultracentrifuge (Figure 3.15A). The gradient was then fractionated and fractions with high absorbance at 280 nm were collected (Figure 3.15B). Tris-HCl pH 8.0 was added to the fractions to quench glutaraldehyde, which reacts with the primary amine group of Tris to form a secondary amine linkage. The pooled fractions were then dialysed in glycerol-free buffer, the sample was concentrated, subjected to size exclusion chromatography, and concentrated again (Figure 3.15C).

Negative stain EM grids of the final sample were imaged and analysed by single particle analysis as before (Figure 3.15D, E and F). The 2D classes (Figure 3.15E) exhibited more detail than the sample without fixation (Figure 3.14B). It is unclear whether this is due to the sample characteristics or because many more particles were used for reconstructing the fixed sample model (24,519 particles vs 7,138 particles for the non-fixed sample). A 3D model reconstructed from the fixed sample data showed similar overall features to the non-fixed sample (Figure 3.14C), namely a base with two protruding densities (Figure 3.15F). The 3D models from both samples suggest that the two extramembranous domains are on the same side of the membrane. This arrangement is inconsistent with a previous study that mapped the topology of *Bacillus subtilis* ComEC using truncation fusions to reporter enzymes (Draskovic & Dubnau, 2005). The resolution limit due to negative staining does not allow a reliable assignment of specific domains to the EM map densities.

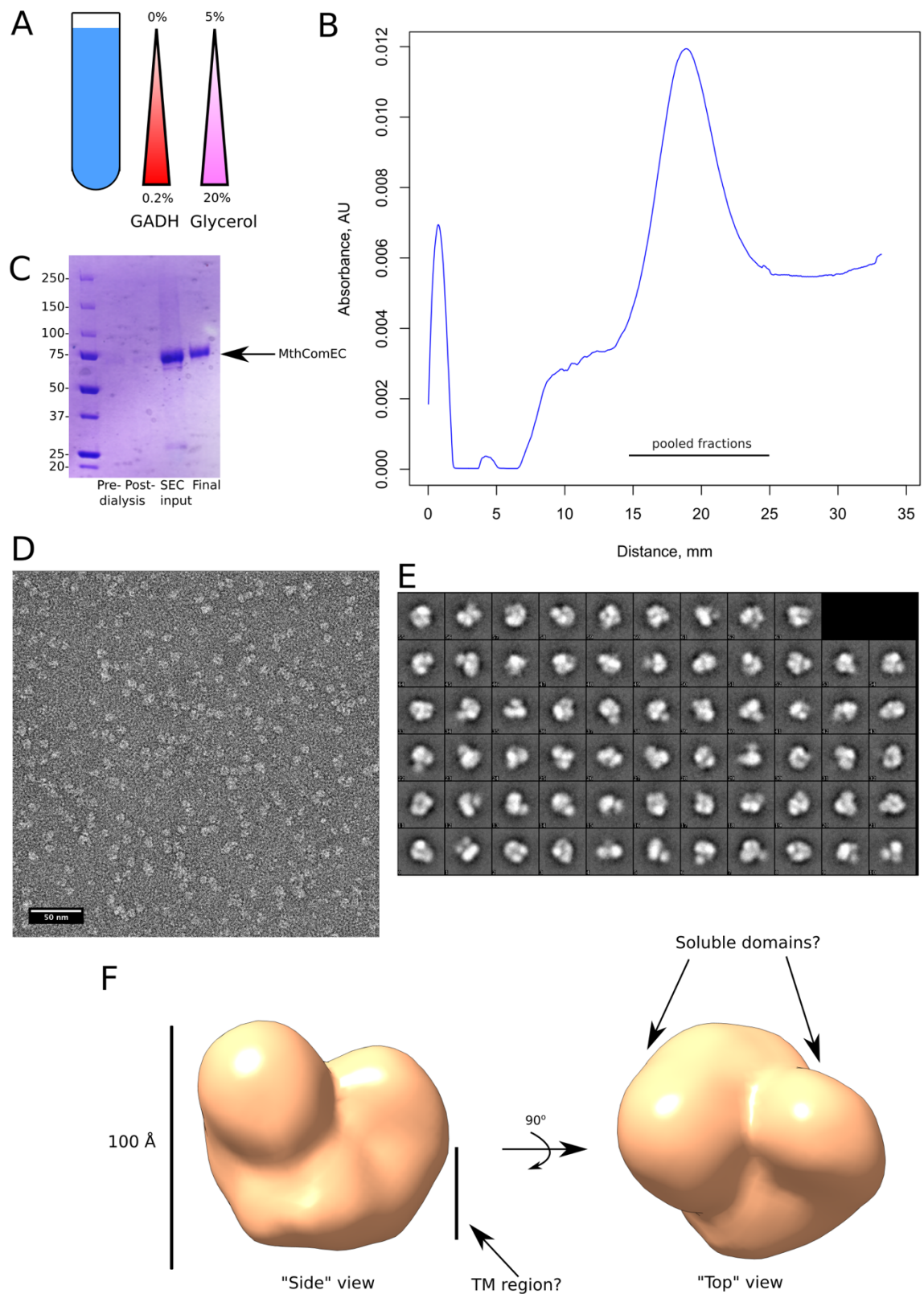


Figure 3.15. GraFix procedure and negative stain EM of fixed MthComEC. (A) Setup of the GraFix experiment (Kastner *et al*, 2007). GADH, glutaraldehyde. (B) Absorbance at 280 nm trace during gradient fractionation from the top (0 mm) of the ultracentrifuge tube. (C) SDS-PAGE gel of post-GraFix sample handling. The fractions indicated in (B) were pooled (Pre-dialysis) and dialysed against crosslinker and glycerol-free buffer (Post-dialysis). The sample was then concentrated

(SEC Input) and subjected to size exclusion chromatography on a Superdex 200 10/300 GL column, and concentrated again to produce the final sample used for EM. The sample was applied to carbon support grids at 0.5 $\mu\text{g}/\text{ml}$ and stained with 2% uranyl acetate. A dataset of 134 micrographs was collected and a representative micrograph is down in (D). (E) Selected 2D class averages from 24,519 particles picked from the dataset. (F) 3D model reconstructed from the 2D class averages.

MthComEC N-terminal truncation construct

EM studies of small proteins are still challenging and so bigger proteins and complexes present a better chance of solving a high resolution structure. The MthComEC $\Delta 2$ -156 construct appeared to be much bigger than the wild type MthComEC (around 500 kDa *versus* 160 kDa) (Figure 3.13E). If this increase in size corresponds to a physiologically relevant higher order oligomer of MthComEC, the truncated construct might be a better EM target than the full-length protein. Although size exclusion indicated monodispersity (Figure 3.13E), particles on the grid were heterogeneous and lacked clear shape (Figure 3.16), in contrast to the wild type MthComEC purified in LMNG as seen in Figures 3.14A and 3.15D. Thus, the $\Delta 2$ -156 truncation construct unsuitable for structure determination.

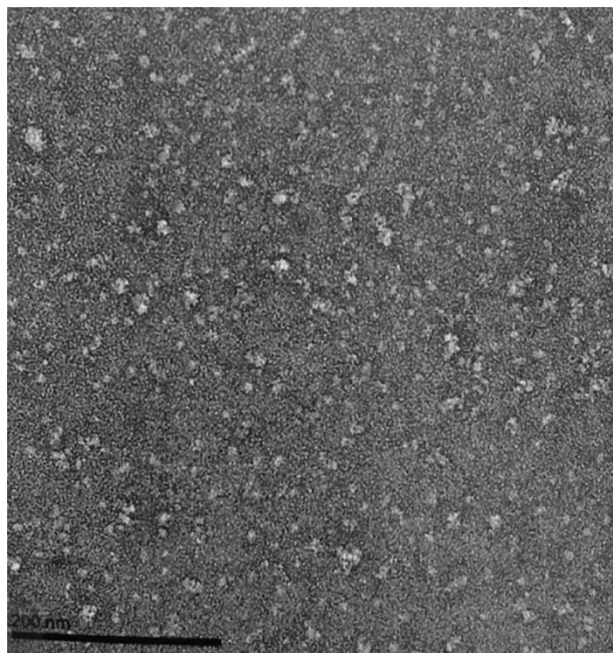


Figure 3.16. Negative stain EM of truncated MthComEC. Representative uranyl-acetate stained EM grid of MthComEC $\Delta 2$ -156 purified in Cyclofos-6. The sample was applied at 0.55 $\mu\text{g}/\text{ml}$.

3.10. Cryo-electron microscopy studies of *MthComEC*

MthComEC is much smaller than proteins routinely studied by cryo-EM and single particle analysis, which are usually at least 200 kDa in size. However, the surprisingly detailed negative stain reconstructions for such a small particle, and the low amounts of sample required prompted attempts at cryo-EM characterisation of *MthComEC*.

Purified *MthComEC* in LMNG was applied to holey carbon grids and vitrified in liquid ethane using standard procedures (Chapter 2). A representative micrograph is shown in Figure 3.17A. In comparison to negative stain, the signal-to-noise ratio in the cryo-EM micrographs was very poor and the low contrast made it hard to see and pick particles. Furthermore, the particle density was not very good. Well-separated particles were rather sparse, and some aggregates were also seen (Figure 3.17A, circles). The grids did not look much different when the concentration of applied protein was increased twofold. It was observed that if the sample was concentrated too much, strings of particles were formed (Figure 3.17B). This can be explained by considering that the detergent micelles get concentrated together with the protein (more so if using a molecular weight cut-off filter of less than 100 kDa) and aggregate. This has previously been observed in our laboratory with other protein samples solubilized in LMNG.

A dataset of around 30,000 particles was collected from grids that look like the one in Figure 3.17A. Some representative 2D class averages are shown in Figure 3.17C. The classes showed a tri-lobed shape. However, compared to the negative stain class averages (e.g. Figure 3.15B), the three densities were much more equal in size with no thicker ‘base’ density. It is likely that this similarity of the three densities is resolution-limiting because it makes correct alignment of the particles difficult. In addition, the flexibility in the

positions of the extramembranous domains might also contribute to poor alignment. Finally, the poor contrast (background *versus* particle signal) made particle picking very difficult and reduced the quality of the data.

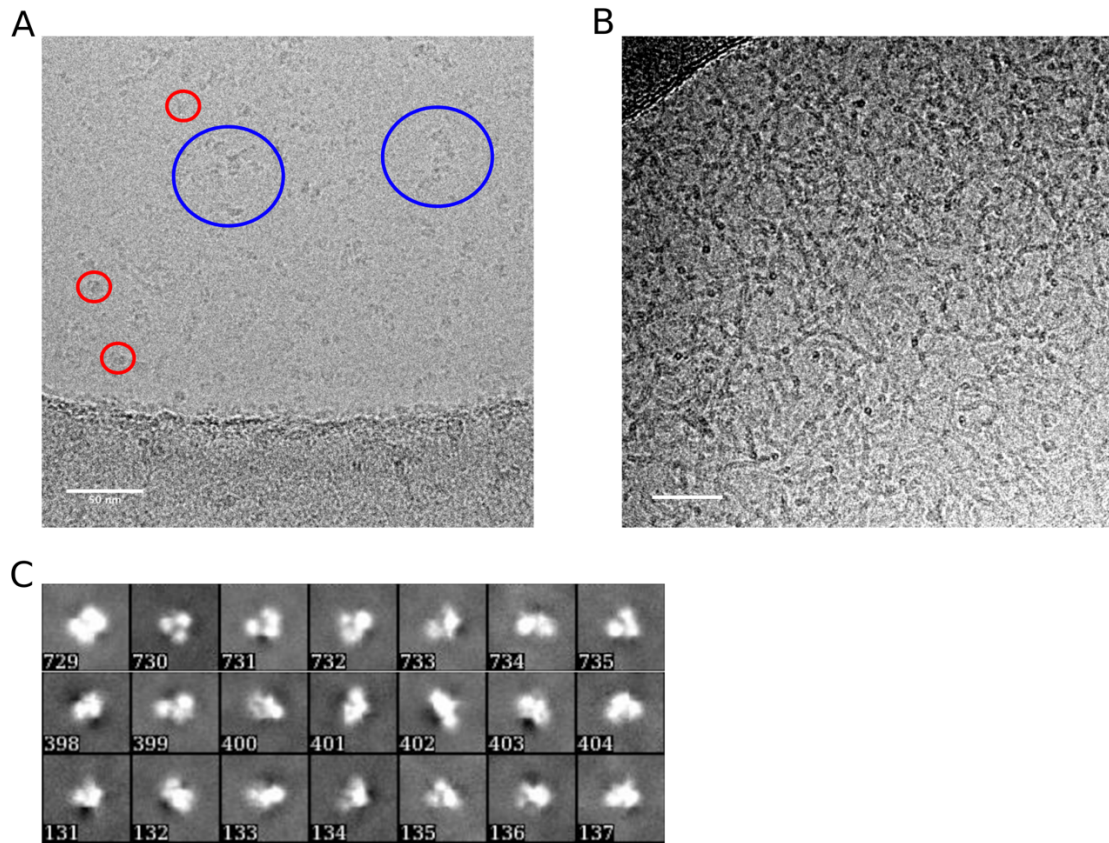


Figure 3.17. Cryo-EM studies of MthComEC in LMNG. (A) Representative micrograph of vitrified MthComEC in LMNG. Red circles show three single MthComEC particles. Blue circles show clumps of particles. Protein was applied at a concentration of 0.5 mg/ml. (B) A micrograph of a MthComEC sample which was very dilute after size exclusion chromatography and was concentrated more than 10-fold before making grids. The sample was applied at 0.11 mg/ml. Scale bars – 50 nm. (C) Representative 2D class averages from a SIMPLE 2D stream.

Chemical crosslinking and removal of excess detergent

GraFix was combined with the GraDeR method (gradient-based detergent removal) to fix the flexible domains of MthComEC and to remove excess detergent, which contributes to the background signal in cryo-EM (Kastner *et al*, 2007; Hauer *et al*, 2015) (Figure 3.18A). In this approach, in addition to a gradient of cross-linker and glycerol, an

inverse gradient of LMNG was applied. As the protein moved down the glycerol and cross-linker gradients, the concentration of LMNG decreased. The gradient was then fractionated, the sample was dialysed against buffer lacking LMNG, glycerol and cross-linker, and then concentrated, subjected to size exclusion chromatography and concentrated again (Figure 3.18B). Figure 3.18C shows that signs of aggregation could be seen on negative stain grids. The 2D class averages also looked less detailed than before (Figure 3.18D). Furthermore, the GraDeR/GraFix-treated particles looked very clumpy (Figure 3.18E). The 2D class averages also looked ‘fuzzy’ and lacked detail (Figure 3.18F). Due to these difficulties, the GraDeR approach was abandoned.

Amphipol trapping

Amphipols (APols) are amphipathic polymers that have a strongly hydrophilic backbone with hydrophobic moieties (Tribet *et al*, 1996; Popot *et al*, 2011; Zoonens *et al*, 2014). The hydrophobic chains of the APol bind to the transmembrane region of the membrane protein while the hydrophilic backbone keeps the complex in solution. Because the dissociation rate of APol from membrane protein is very low, no additional APol is required in the buffer. This is unlike detergent-solubilised samples where the concentration of detergent has to be maintained above the critical micellar concentration (CMC) so that the detergent forms micelles. Free detergent monomers and empty micelles can contribute to background in cryo-EM micrographs. It has been shown that ‘trapping’ membrane proteins in APols after detergent extraction can lead to improved cryo-EM data (Mio & Sato, 2018).

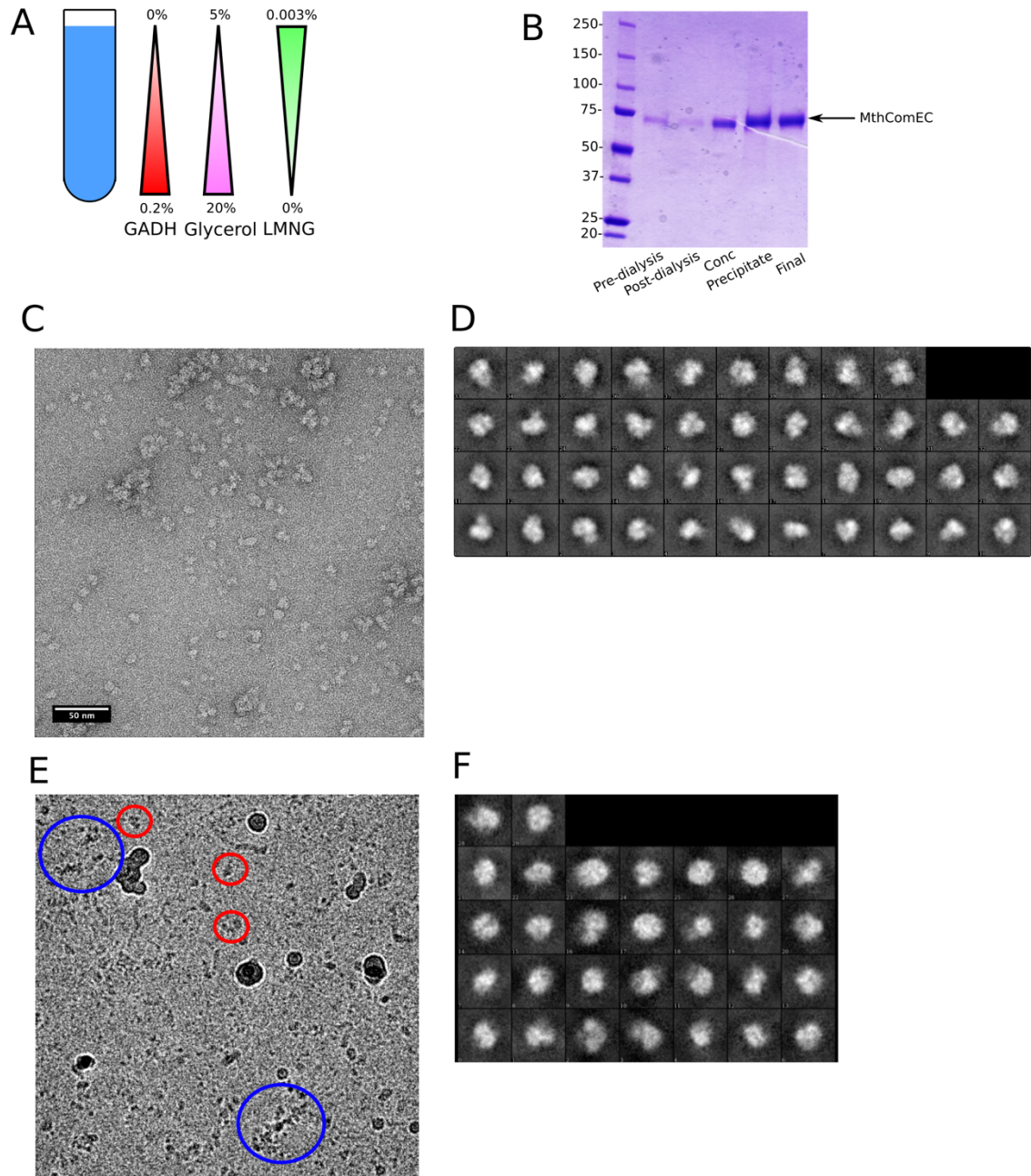


Figure 3.18. EM studies of MthComEC treated with the GraDeR technique. (A) Setup of the GraDeR/GraFix experiment (Hauer *et al*, 2015; Kastner *et al*, 2007). GADH, glutaraldehyde. (B) SDS-PAGE gel of post-GraDeR/GraFix sample handling fractions. After gradient fractionation, pooled fractions (Pre-dialysis) were dialysed against glycerol-, crosslinker- and LMNG-free buffer (Post-dialysis). The sample was concentrated (Conc) but precipitate formed (Precipitate). The final sample was subjected to size exclusion chromatography on a Superose 6 5/150 column, and concentrated again before making grids (Final). (C) Negative stain electron micrograph of MthComEC treated with GraDeR/GraFix and (D) the 2D class averages generated from this data. The sample was applied at 0.5 $\mu\text{g/ml}$ and stained with 2% uranyl acetate. (E) Cryo-electron micrograph of MthComEC after GraDeR/GraFix treatment and (F) the 2D class averages generated from this data. The sample was applied at 0.56 mg/ml. A dataset of 211 micrographs was collected and around 12,000 particles were used in 2D classification. In (E), red circles – individual particles, blue circles – aggregates.

APol trapping of MthComEC before cryo-EM was attempted in order to reduce the image background and improve particle stability. MthComEC was purified in LMNG and incubated with APol A8-35 (Anatrace) and bio-beads. The bio-beads absorb detergent from the buffer allowing the APol to bind to the transmembrane region of MthComEC. The resulting MthComEC/APol/LMNG complexes were analysed by size exclusion chromatography (Figure 3.19A, coloured traces). Increasing the ratio of APol-to-protein shifted the complex to earlier elution volumes, indicating an increase in size. When MthComEC was incubated with bio-beads alone and applied to the size exclusion column equilibrated with buffer without any LMNG, it still formed a protein-detergent complex of the same size as MthComEC in LMNG-containing buffer (Figure 3.19A, solid and dashed black traces). Thus, as reported previously for other proteins, LMNG seems to be tightly associated with MthComEC (Hauer *et al*, 2015). Alternative detergent removal methods, such as dialysis and dilution, were also unable to remove all of the bound LMNG as assessed by size exclusion chromatography without LMNG in the mobile phase. Therefore, it is likely that the increase in molecular weight observed in Figure 3.19A represents increasing amounts of A8-35 binding to a stable MthComEC-LMNG complex, potentially adding more disordered density to the transmembrane region.

When imaged in cryo-grids, the APol-bound MthComEC sample did not show an obvious decrease in background signal or improved particle distribution. Individual particles were still hard to see and there was a lot of clumping (Figure 3.19B). The 2D class averages (Figure 3.19C) still showed a tri-lobed density with approximately equal lobes. They did not look significantly different to those from MthComEC in LMNG (Figure 3.17C). No signs of secondary structure were visible within the 2D class average density.

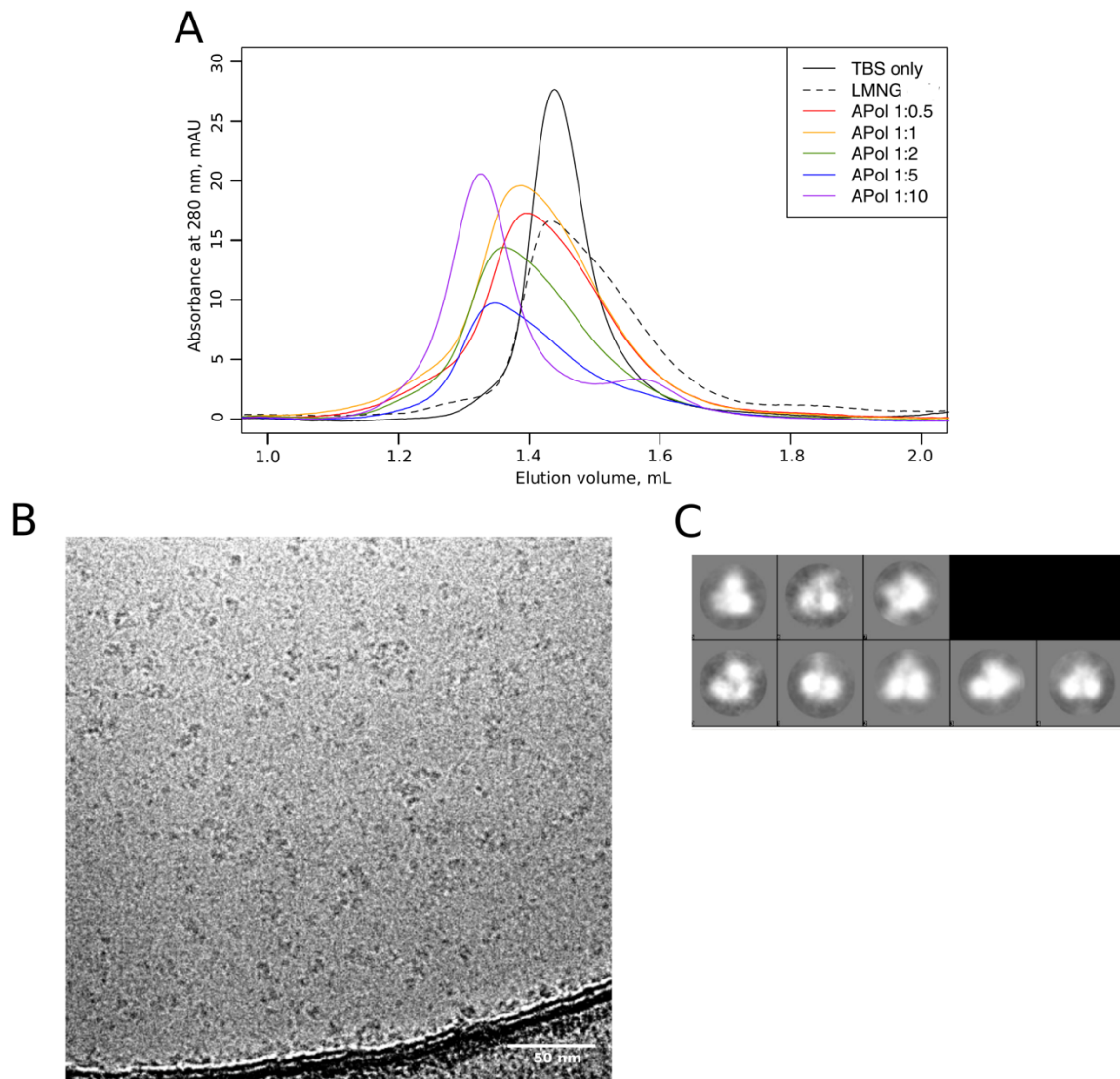


Figure 3.19. Cryo-EM studies of MthComEC in amphipols. (A) Superdex 200 3.2/300 column size exclusion chromatography traces of MthComEC mixed with amphipols. Pure protein was mixed with amphipol A8-35 in the indicated mass ratios of protein-to-amphipol, incubated on ice for 45 min, and then incubated with biobeads on a roller at 4°C overnight. The MthComEC/amphipol mixtures were run in detergent-free TBS buffer. TBS, Tris-buffered saline. APol, amphipol 8-35 (Anatrace). (B) Representative cryo-EM micrograph of MthComEC trapped in amphipols using a protein-to-amphipol ratio of 1:10. Sample was applied to grids at 1 mg/ml. (C) 2D class averages from particles picked from grids of MthComEC in amphipols.

It was concluded that MthComEC is an unlikely candidate for high-resolution structure determination by cryo-EM. Conformational heterogeneity due to domain flexibility and the small size of the particles likely limit the accuracy of particle alignment.

ComEC potentially interacts with other components of the competence machinery. If such complexes can be purified, their larger size would make them better targets for structural analysis by cryo-EM than pure ComEC. The following chapter describes attempts to purify such extended ComEC complexes from *B. subtilis*.

3.10. Summary

In the pursuit of a better understanding of DNA transport across the bacterial cytoplasmic membrane during natural transformation, a heterologous expression system and purification protocol for the ComEC DNA transport channel was developed. Expression tests and initial purification attempts identified ComEC from *Moorella thermoacetica* as a promising target. After considerable optimization, an expression and purification protocol was established that includes a step to wash away expression host contaminants. The resulting purified MthComEC was not an ideal structural target. Crystallisation was not possible due to very low yields and the protein was too small for high resolution cryo-EM studies. A complex of ComEC with other components of the competence machinery would present a better target for cryo-EM. This possibility is pursued in the following chapter (Chapter 4).

Previous biochemical work indicated that ComEC may form homodimers (Draskovic & Dubnau, 2005). Therefore, purification of a monomer was surprising. This could indicate that ComEC is capable of transporting DNA as a monomer. However, it remains possible that ComEC oligomerises around DNA during transport, or that the lack of oligomerisation is a heterologous overexpression or purification artefact.

Negative stain EM allowed the first glimpse at the overall shape of the cytoplasmic membrane DNA transporter. The observed low-resolution EM density map has important

Chapter 3

implications for the topology of the extramembranous domains of ComEC. Although the resolution is too low to confidently assign domains to the density, the shape of the particle suggests that the N- and C-terminal domains are located on the same side of the membrane. Functional characterisation of the C-terminal domain of MthComEC is described below in Chapter 5.

Chapter 4: Expression and purification of ComEC from *Bacillus subtilis*

4.1. Introduction

Production of recombinant ComEC from *M. thermoacetica* was described in Chapter 3. The size of recombinant MthComEC (160 kDa protein/detergent complex) was too small for high-resolution cryo-EM and single particle analysis. Recent successes in the literature report that purification of proteins from native sources for structural work can lead to co-purification of partner proteins (Glenwright *et al*, 2017; Lauber *et al*, 2018). It was hypothesised that pulldowns of ComEC or other competence proteins from a naturally competent bacterium may lead to isolation of large DNA uptake complexes that may be of suitable size and yield for structural biology and biochemical characterisation.

Competence regulation in Bacillus subtilis

Natural transformation of *B. subtilis* was first reported by Spizizen (1958). Since then studies of competence regulation have shown that in *B. subtilis* competence is a developmental stage that is tightly regulated by cell density and nutrient availability via a complex network of regulators (Dubnau, 1991; Hamoen *et al*, 2003). A simplified schematic for competence regulation in *B. subtilis* is shown in Figure 4.1, the key player in which is the master competence transcription factor, ComK. In non-competent cells ComK is bound by MecA and targeted for proteolytic degradation by ClpCP (Turgay *et al*, 1998). Several signalling pathways converge on the regulation of ComK levels, quorum sensing being one of them (Hamoen *et al*, 2003). When cultures experience stationary phase stress, the ComX signalling peptide is expressed, modified by the isoprenyltransferase ComQ and exported into the medium. Modified ComX is recognised by the membrane-bound sensor histidine kinase ComP, which phosphorylates the response regulator ComA that in turn induces expression of the ComS peptide. ComS binds to the MecA/ClpCP complex and inhibits proteolysis of ComK (Turgay *et al*, 1998)

(Figure 4.1). ComK binds to AT-rich ComK-boxes upstream of over 100 genes, including the *comK* gene itself (Berka *et al*, 2002; Ogura *et al*, 2002). Thus, inhibition of MecA/ClpCP by ComS leads to accumulation of ComK due to its own self-induction. However, this occurs only in 5-10% of cells in a stationary phase culture. Only those cells become competent which already express a sufficient level of ComK when ComS is expressed. If a cell does not express ComK at all, it cannot be induced by ComS. It has been shown that in *B. subtilis* competence induction in the population is a stochastic process and depends on variations in the basal transcription level of ComK, in addition to other transcriptional regulators that repress ComK expression (Maamar & Dubnau, 2005; Smits *et al*, 2005; Leisner *et al*, 2008). Cells that accumulate ComK undergo global changes in gene expression, including production of late competence proteins that form the DNA uptake machinery (Chapter 1). Upon entering the competent state, *B. subtilis* remains competent for approximately 2 h after which the rate of ComK degradation exceeds the rate of synthesis and cells exit the competent state (Turgay *et al*, 1998).

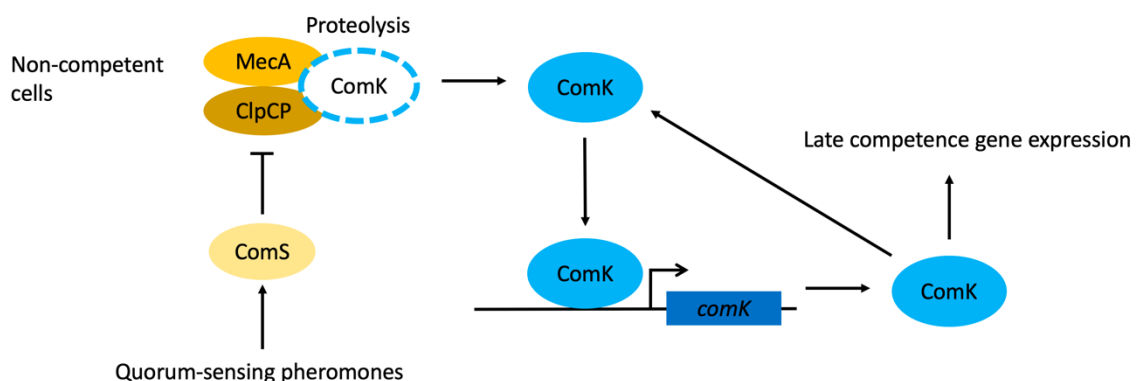


Figure 4.1. Regulation of the master competence transcription factor ComK in *Bacillus subtilis*.

Artificial induction of competence

Only a small proportion of *B. subtilis* cells become competent and express the DNA uptake machinery. This impedes the efficiency of transformation of bulk cultures and poses a challenge to isolation of native DNA uptake complexes for biochemical and structural characterisation of native DNA uptake complexes due to low yields. Several approaches have been used to maximise the proportion of *B. subtilis* cells that become competent, all of them involving overexpression of ComK with or without ComS (Hahn *et al*, 1996; Liu *et al*, 1996; Duitman *et al*, 2007; Nijland *et al*, 2010; Zhang & Zhang, 2011; Rahmer *et al*, 2015). In strains carrying ectopic *comK/comS* genes under the control of an inducible promoter, virtually all cells in the population become competent upon adding inducer. In addition, being able to regulate competence development with an inducer removes the need for specific environmental cues such as stationary phase stress which are normally required for competence development. This allows induction of competence in cells grown in rich media. Overexpression of the *B. subtilis* master transcription factor ComK, rather than specific late competence genes directly responsible for DNA uptake, hopefully, retains the native stoichiometries of the DNA uptake complex components. Such overexpression systems may enable growth of sufficient biomass expressing the DNA uptake machinery so that enough material can be isolated to study DNA uptake complexes *in vitro*.

Protease- and sporulation-deficient Bacillus subtilis

A potential problem when purifying proteins from *B. subtilis* is proteolytic degradation. *B. subtilis* has at least seven secreted proteases with various mechanisms of action (Simonen & Palva, 1993). Although proteolytic degradation is usually an issue when expressing heterologous proteins (Li *et al*, 2004), it is possible that exposing cytoplasmic

components of the DNA uptake machinery to the extracellular proteases during cell lysis could lead to degradation of the cytoplasmic components. Therefore, a protease- and sporulation-deficient *B. subtilis* strain, KO7-S, was obtained from the Bacillus Genetic Stock Centre (unpublished) to be used in pull-down experiments. KO7-S is deficient in all seven secreted proteases. Additionally, KO7-S also lacks the sporulation transcription factor σ^F (Decatur & Losick, 1996) and is consequently unable to form spores. Spores are resistant to some disinfectants and can contaminate laboratory equipment when using wild type *B. subtilis* strains.

The subsequent sections of this chapter report on generating supercompetent *B. subtilis* KO7-S strains expressing chromosomally tagged competence proteins and attempts at purifying them.

4.2. Construction of a supercompetent and protease-deficient Bacillus subtilis strain

The supercompetence system used in this study was designed by Zhang & Zhang (2011). A supercompetence cassette, composed of ComK under the control of a xylose inducible promoter, a xylose repressor gene and an erythromycin resistance gene, is flanked by regions homologous to the 5' and 3' sequences of the *lacA* locus of *B. subtilis* (Figure 4.2A). A strain can be made supercompetent by transforming it with genomic DNA from another supercompetent strain using conventional transformation methods.

The KO7-S strain was transformed with genomic DNA from strain 1A976, which carries the supercompetence cassette, to generate strain KO7-S-K. The supercompetent KO7-S-K cells were then transformed with plasmid pPG35 (Gamba *et al*, 2015) to generate strain AS1. The pPG35 plasmid integrates into the *amyE* locus and contains a LacZ-GFP fusion

under the control of the ComK-inducible *comGA* promoter. LacZ-GFP expression in the AS1 strain was assessed by fluorescence microscopy before and after induction with 1% xylose (Figure 4.2B). Prior to induction some cells showed background GFP fluorescence. This may be due to stochastic low-level *comK* expression, which induces low levels of LacZ-GFP production. After induction with xylose all cells in the field of view were fluorescent and the signal was higher than in the non-induced cells (Figure 4.2B). This experiment showed that all KO7-S-K cells in the culture were competent when induced with xylose.

Transformation efficiency assays were carried out to compare the different *B. subtilis* strains. Competent *B. subtilis* 168+ and KO7-S cells were prepared using the conventional protocol that involves nutrient starvation (Chapter 2). When transformed with gDNA carrying a chloramphenicol resistance gene, the 168+ and KO7-S strains did not show statistically significant differences in transformation efficiency (Figure 4.2C). This experiment showed that the KO7-S strain does not have any inherent DNA uptake defects that could be related to deletion of its protease genes and sporulation sigma factor. KO7-S and KO7-S-K cells were prepared in supercompetence conditions (growth in LB supplemented with 1% xylose) and incubated with gDNA carrying a chloramphenicol resistance gene. As expected, no transformants were detected for KO7-S because this strain does not have the supercompetence cassette and cannot become competent when grown in LB (Figure 4.2D). The supercompetent KO7-S-K strain showed the same transformation efficiency as KO7-S grown under normal competence conditions (Figure 4.2C and D). These data show that possession of the supercompetence cassette relieves the requirement of starvation for competence development. However, the transformation efficiency of the supercompetent KO7-S-K strain grown with xylose

induction was not greater than of KO7-S grown under normal competence conditions. This is surprising, as there should be more competent cells in a supercompetent culture of KO7-S-K and, thus, the transformation efficiency should be higher. It is possible that not enough gDNA was used in the assay (1 μ g).

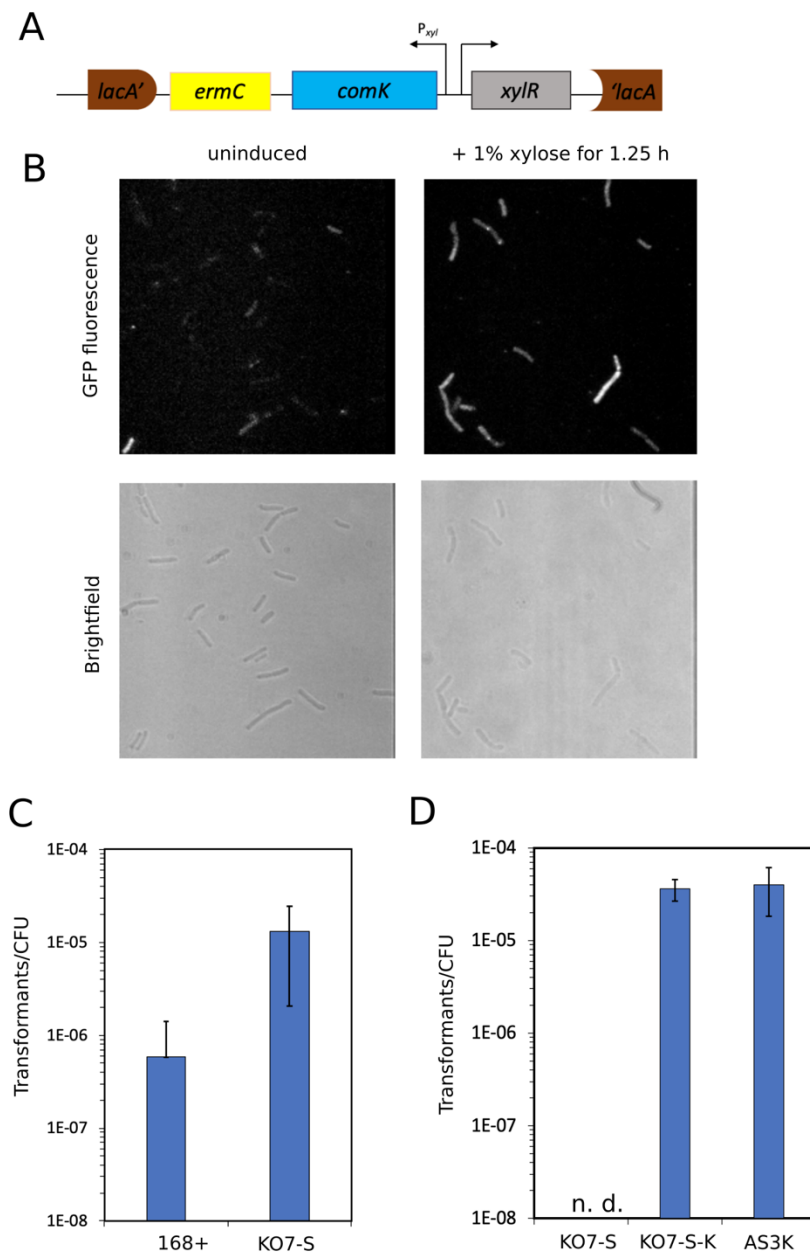


Figure 4.2. Supercompetent *Bacillus subtilis*. (A) The xylose-inducible supercompetence cassette designed by Zhang & Zhang (2011). (B) Fluorescence and brightfield microscopy of strain AS1 without and with xylose induction. Strain AS1 was grown overnight in LB medium at 37°C with 180 rpm shaking. The culture was diluted to $OD_{600}=0.5$ in LB medium with or without 1% xylose and further grown for 1.25 h. Cells were collected by centrifugation, resuspended in PBS buffer and spotted on agarose pads. The pads were imaged in a Nanoimager microscope (Oxford

Nanoimaging) in brightfield mode and in fluorescence mode with a 532 nm solid state excitation laser at 100x magnification. (C) Transformation efficiencies of the wild type 168+ *B. subtilis* cells and the protease- and sporulation-deficient KO7-S cells. Competent cells were prepared using the conventional nutrient starvation method (Chapter 2). The average of three independent experiments is shown. The two conditions were not significantly different. (D) Transformation efficiencies of the KO7-S, KO7-S-K and AS3K strains grown under supercompetence conditions (Chapter 2). N. d., no transformants detected. The average of two independent experiments is shown. Error bars represent the standard deviation. CFU, colony forming units.

4.3. Chromosomal tagging of competence proteins

Previous microscopy studies showed that competence proteins co-localise to discrete foci within cells even without added external DNA. (Hahn *et al*, 2005; Kramer *et al*, 2007; Kaufenstein *et al*, 2011). Co-localisation of DNA uptake proteins in the absence of transforming DNA suggests that the DNA uptake machinery is assembled and primed for accepting DNA upon competence induction. Therefore, it might be possible to isolate these complexes using biochemical methods. In order to purify the putative DNA uptake complexes from competent *B. subtilis* cells, at least one of their components had to be tagged with an affinity tag. The pMAD vector (Arnaud *et al*, 2004) was used to scarlessly insert affinity tags in the chromosome of the KO7-S strain. The tags were positioned so that there would be minimal disruption to the structure of the *comEABC* and *comFABC* operons (Figure 4.3A). The cytoplasmic membrane DNA channel ComEC and the poorly characterised cytoplasmic membrane-associated protein ComFC were tagged with a C-terminal TwinStrep tag. The cytoplasmic ATPase ComFA was tagged with an N-terminal His₈-tag. These strains were then transformed with gDNA from strain 1A976, generating the supercompetent strains AS3K (ComEC-TwinStrep), AS4K (ComFA-His) and AS5K (ComFC-TwinStrep).

The transformation efficiency of strain AS3K was compared to that of KO7-S and KO7-S-K under supercompetence conditions (Figure 4.2D). The transformation efficiency of

AS3K was the same as that of KO7-S-K showing that the tag does not impair ComEC function. The transformation efficiencies of AS4K and AS5K were not quantified. However, the tagged strains must be functional because they could be transformed with 1A976 gDNA to make them supercompetent.

In order to confirm that the tagged proteins were expressed, strains KO7-S-K, AS3K, AS4K and AS5K were grown in TB medium with 1% xylose to induce competence. After a 2 h induction period at 37°C, the cells were harvested and fractionated into total lysate, low-speed pellet, soluble lysate, high-speed pellet (membranes) and high-speed supernatant fractions. The fractions from different strains were analysed by Western blotting (Figure 4.3B, C, D). In strain AS3K fractions, ComEC-TwinStrep was detected only in the membrane fraction (Figure 4.3B), likely because the expression levels of ComEC-TwinStrep even under supercompetence conditions are too low to allow detection in the lysate fractions. ComFA-His could not be detected in any fraction of AS4K cells (Figure 4.3C). The presence of the tag on the chromosome was confirmed by sequencing. It is unlikely that ComFA-His is not expressed in this strain because the strain is competent. It is possible that the expression levels are too low to allow detection by the anti-His antibody, or that the tag is cleaved in the cytoplasm of *B. subtilis*. The ComFA-His construct was not investigated further. ComFC-TwinStrep was detected in all fractions of AS5K (Figure 4.3D). ComFC-TwinStrep presence in the low-speed pellet fraction suggests that either some of the protein is aggregated under supercompetence conditions or that the cell lysis step was not efficient. ComFC-TwinStrep could be detected in both the membrane fraction and the membrane supernatant fraction, suggesting that some ComFC-TwinStrep is membrane-associated. This was not unexpected because ComFC has

been shown to co-localise at polar foci with the pilus assembly ATPase ComGA (Kaufenstein *et al*, 2011).

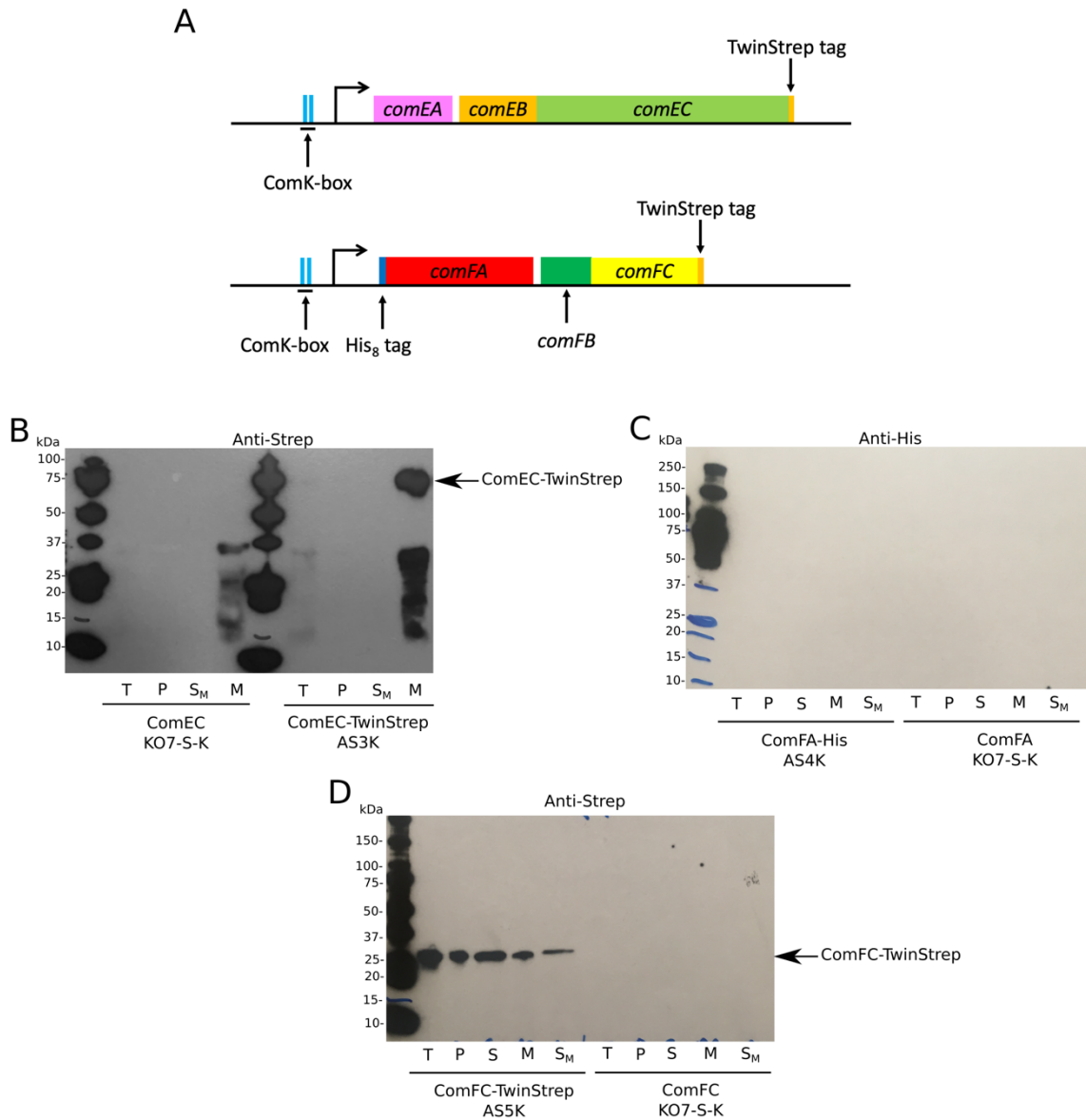


Figure 4.3. Chromosomal tagging of *B. subtilis* competence proteins. (A) The organisation of the *comEABC* and *comFABC* operons and the location of the affinity tags that were inserted in to the chromosome of *B. subtilis* KO7-S in this study. (C-D) Western blots probed with the indicated antibodies of cellular fractions of the tagged and the untagged *B. subtilis* strains grown in TB with 1% xylose. T, total lysate; P, low-speed pellet; S_M , supernatant from the ultracentrifugation step; M, pellet from the ultracentrifugation step containing the membrane fraction.

4.4. Pull downs of ComEC-TwinStrep from supercompetent cells

Purification of ComEC-TwinStrep from supercompetent AS3K cells was attempted in order to isolate a DNA uptake complex that is large enough for high resolution cryo-EM studies. Suitable detergents that could solubilise ComEC efficiently had to be identified first. Membranes of AS3K cells grown in TB medium with 1% xylose were isolated and solubilised in 1% DDM, 1% LMNG or 1.5% GDN. Insoluble material was removed by ultracentrifugation, and the samples were analysed by Western blotting (Figure 4.4A). DDM and LMNG solubilised all of ComEC, but hardly any ComEC was detected in the GDN sample after ultracentrifugation.

A 4 litre TB-xylose culture of AS3K cells was harvested 3 hours after induction of supercompetence. The membrane fraction was isolated, split in half and solubilised in either DDM or LMNG. After StrepTrap purification, the samples were concentrated and injected onto a Superdex 200 6 3.2/300 analytical column (Figure 4.4B and C). Almost no signal was observed for the DDM sample. The LMNG sample trace contained broad overlapping peaks. A Western blot of the size exclusion fractions showed that ComEC-TwinStrep in LMNG eluted in a broad peak, but the DDM sample was lost (Figure 4.4D). The size exclusion fractions of the LMNG-solubilised ComEC-TwinStrep sample were analysed by SDS-PAGE and silver staining (Figure 4.4E). The silver stained gel showed that the fractions from the size exclusion column had many proteins present. The bands indicated in Figure 4.4E were cut out and submitted to peptide fingerprinting mass spectrometry analysis for identification. The biotin carboxyl carrier protein likely bound to the StrepTrap column via the interaction of biotin with the streptavidin in the column resin. GroL and pyruvate carboxylase have never been implicated in competence, and their involvement in DNA uptake is unlikely. The function of the YceH protein is unknown

but it is associated with general and cell envelope stress in *B. subtilis* (Petersohn *et al*, 2001; Eiamphungporn & Helmann, 2008). The only two proteins known to be involved in competence that were identified were ComEC and RecA, and their respective band intensities in the silver stained gel seemed to correlate from fraction F10 to G3, suggesting that ComEC and RecA form a complex. The purification protocol had to be optimised in order to remove the contaminant proteins so that the interaction between ComEC and RecA could be investigated further.

To prevent the binding of biotin carboxyl carrier protein, 8 U of avidin per litre of cell culture were included in the lysate. Avidin blocks the biotin groups of biotinylated *B. subtilis* proteins and prevents them from binding to the affinity column. A StrepTactin XT affinity column was also used, which is more selective for the TwinStrep tag than the StrepTrap column and, thus, should bind less endogenous biotinylated protein. GroL is a folding chaperone with ATPase activity. In order to remove GroL from the sample, the StrepTactin XT column was washed with MgATP-containing buffer. ComEC-TwinStrep was purified from a 6 litre TB-xylose culture of AS3K cells in LMNG with the aforementioned modifications. The size exclusion trace was again very broad (Figure 4.5B), indicating sample heterogeneity. However, the silver stained gel showed that the fractions contained mostly pure ComEC-TwinStrep (Figure 4.5C). The only minor contaminant was 30S ribosomal protein RpsB. No RecA was observed in the pulldown. RecA is a cytoplasmic competence protein involved in integration of transforming DNA by homologous recombination (Yadav *et al*, 2014). It has ATPase activity and was likely washed away in the MgATP wash. When the MgATP wash was omitted, RecA was again observed in the sample, but so was GroL and other unidentified contaminants (not shown).

Chapter 4

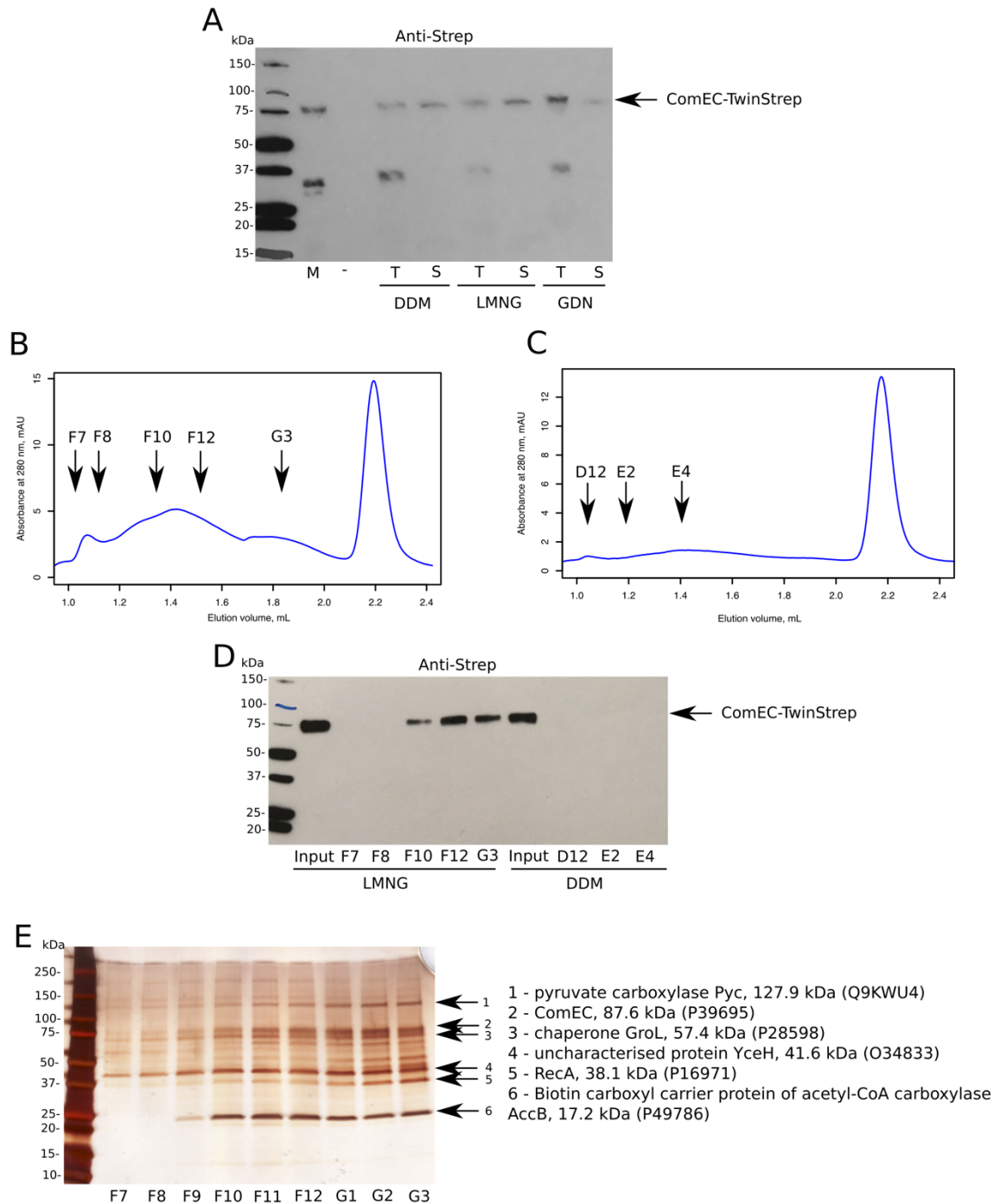


Figure 4.4. Initial ComEC-TwinStrep pulldowns in DDM and LMNG. (A) ComEC-TwinStrep detergent solubilisation efficiency test. An anti-Strep Western blot of AS3K membranes from a 500 ml TB-xylose culture solubilised in 1% DDM, 1% LMNG and 1.5% GDN (membrane protein-to-detergent mass ratio of 1:2). M, membrane fraction; T, total solubilised membranes; S, solubilised membranes after pelleting insoluble material by ultracentrifugation. Size exclusion chromatography traces of affinity-purified ComEC-TwinStrep solubilised in LMNG (B) and DDM (C) on a Superdex 200 3.2/300 column. The void volume for this column is approximately 1.0 ml. (D) Anti-Strep Western blot of size exclusion fractions from (B) and (C). (E) Silver-stained SDS-PAGE gel of fractions across the size exclusion profile from (B). The identities of the bands indicated by the numbered arrows were confirmed by mass spectrometry and are listed on the right, together with their molecular weights and UniProt accession numbers.

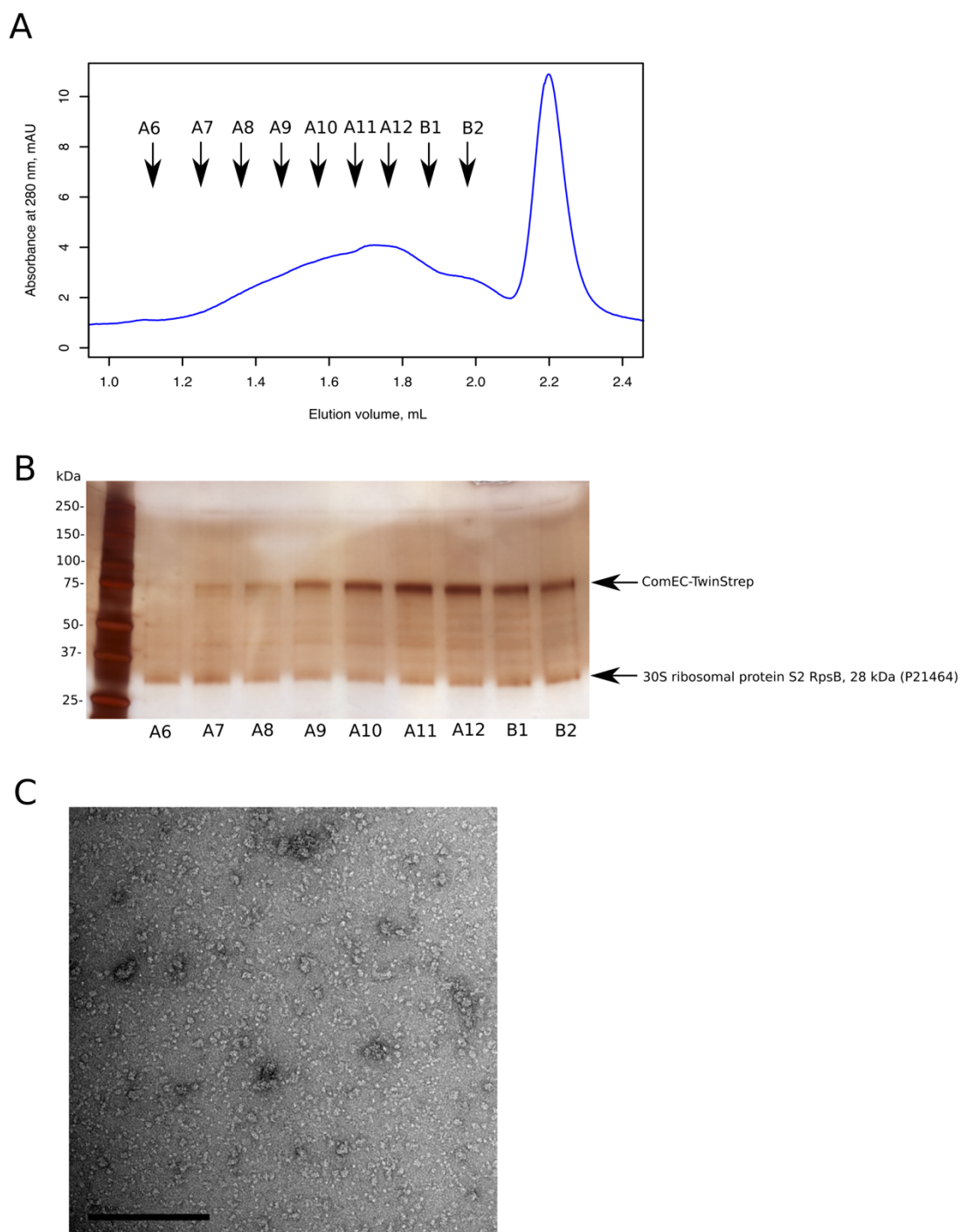


Figure 4.5. Optimised pulldown of ComEC-TwinStrep. (A) Size exclusion chromatography profile on a Superose 6 3.2/300 Increase column of affinity-purified ComEC-TwinStrep from *B. subtilis* using the optimised purification protocol. The void volume of the column is at 1.0 ml. The purification was done in LMNG. (B) Silver-stained SDS-PAGE gel of the size exclusion fractions (C) Representative electron micrograph of a negatively stained sample from fraction A11.

The pure ComEC-TwinStrep fractions from size exclusion chromatography were visualised in negative stain EM (Figure 4.5C). Although the silver stain gel indicated that the protein is mostly pure (Figure 4.5B), the particles looked very heterogeneous. This sample was not investigated further.

The peak of the broad size exclusion chromatography profile of ComEC-TwinStrep was approximately at 1.75 ml. The MthComEC monomer elutes at 1.8 ml on a Superose 6 3.2/300 column (not shown). Therefore, the purification results suggest that a significant portion of ComEC-TwinStrep is also purified as a monomer from *B. subtilis*. This is in disagreement with a previous study that suggested that ComEC forms dimers (Draskovic & Dubnau, 2005). However, the signal in our size exclusion data was very small and the profile was very broad, so the exact oligomeric state or states cannot be determined from the data in Figure 4.5A.

4.5. Purification of ComFC-TwinStrep

ComFC-TwinStrep was expressed in the initial supercompetence induction test in strain AS5K (Figure 4.3D). ComFC-TwinStrep was found in the soluble fraction and in the membrane fraction. Therefore, two purification strategies were attempted.

First, ComFC-TwinStrep was treated as a 'soluble' protein and purified without detergent present in buffers. A 2 litre TB-xylose culture was grown for 3 hours, harvested, lysed, the lysate was clarified and the supernatant was loaded on a 1 ml StrepTrap HP column (Chapter 2). After washing the column, bound material was eluted, concentrated and subjected to size exclusion chromatography (Figure 4.6A). The main peak in the size exclusion profile eluted just after the void peak (fraction 5). Fractions across this peak were analysed by SDS-PAGE (Figure 4.6B). The gel revealed a faint 27 kDa species and a

prominent 60 kDa species. The identity of the 27 kDa species was confirmed to be ComFC-TwinStrep by Western blotting (not shown), and the 60 kDa species was confirmed as GroL by fingerprinting mass spectrometry analysis. As for ComEC, the yield of ComFC was very low. Co-purification with the folding chaperone GroL indicated that ComFC may be misfolding.

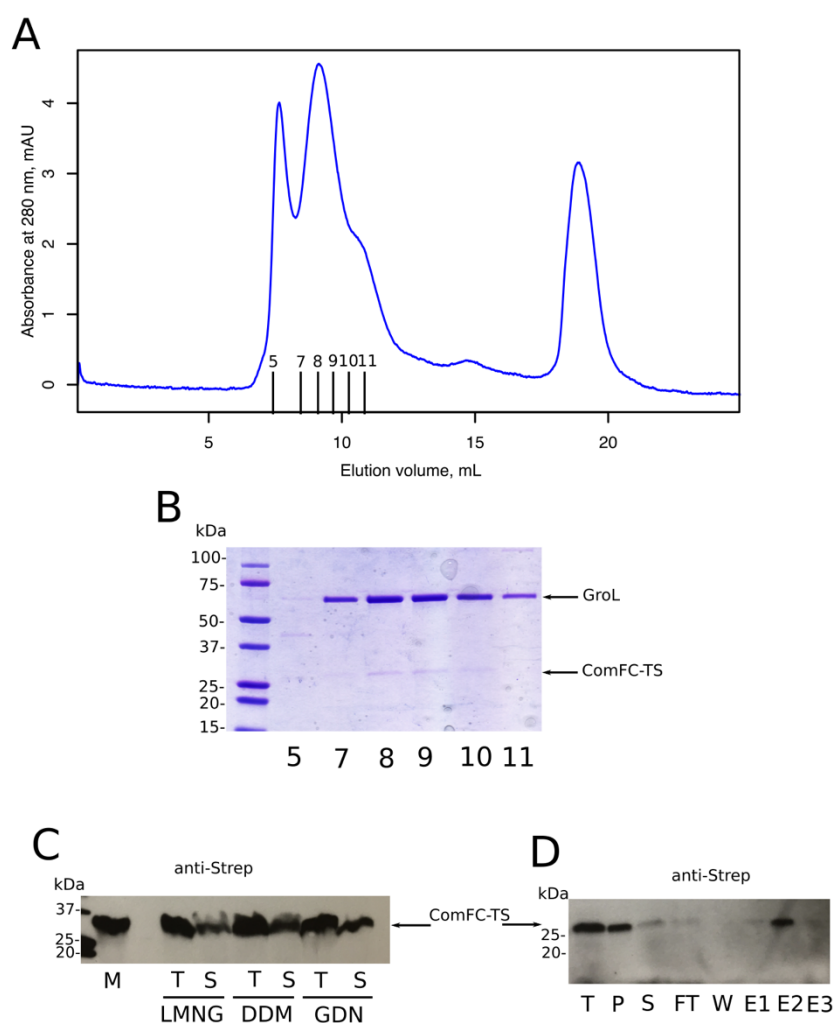


Figure 4.6. Purification of ComFC-TwinStrep. (A) Superdex 200 10/300 size exclusion chromatography trace of ComFC-TwinStrep purified as a ‘soluble’ protein from a 2 litre culture. (B) SDS-PAGE analysis of fractions from (A). (C) Anti-Strep Western blot of ComFC-TwinStrep detergent solubilisation test. The isolated membrane fraction (M) was solubilised in 1% LMNG, 1% DDM, and 1.5% GDN at a membrane protein-to-detergent mass ratio of 1:2 for 1 h at 4°C (T). Insoluble material was pelleted by ultracentrifugation and the soluble material was collected (S). (D) Anti-Strep Western blot of ComFC-TwinStrep purified in GDN from a 4 litre culture. T – membranes solubilised in 1.5% GDN; P – pellet from ultracentrifugation step after solubilisation and S – the supernatant; FT – StrepTrap column flow-through; W – column wash; E – eluate.

The second purification strategy was to purify ComFC-TwinStrep as one would a membrane protein. A 4 litre TB-xylose culture was processed as described in Chapter 2 in order to isolate the membrane fraction. A detergent solubilisation test was performed with LMNG, DDM and GDN (Figure 4.6C). All three detergents solubilised ComFC-TwinStrep efficiently. LMNG and DDM had been trialled with ComEC-TwinStrep (Figure 4.4) and did not result in isolation of competence complexes. Therefore, GDN was trialled with ComFC-TwinStrep. The rest of the membrane fraction from the 4 litre culture was solubilised in GDN, the insoluble material was pelleted by ultracentrifugation and the supernatant was loaded on a StrepTrap column. The column was washed and bound protein was eluted in three fractions (Figure 4.6D). Due to unknown reasons, the solubilisation efficiency was not as good as in the detergent test (Figure 4.6C), and very little protein eluted from the column. No signal was observed in an SDS-PAGE gel stained with Coomassie stain (not shown). ComFC-TwinStrep could only be detected by Western blotting (Figure 4.6D). Because the yields were vanishingly small this construct was not pursued further.

4.7. Summary

In this chapter purification of chromosomally tagged ComEC and ComFC competence proteins from *B. subtilis* was trialled in an attempt to isolate native DNA uptake complexes for structural and biochemical analysis. Although an artificial competence induction system (Zhang & Zhang, 2011) was used to maximise expression of the DNA uptake machinery, the yields of ComEC and ComFC were very low. In addition, ComEC co-purified with the stress-associated protein YceH and both ComEC and ComFC co-purified with the folding chaperone GroL. It is unlikely that the co-purified proteins are involved in

competence. It is more likely that *B. subtilis* cells are stressed under supercompetence conditions. It can be concluded that purification of the native DNA uptake machinery from *B. subtilis* for *in vitro* analysis is not a viable approach.

As discussed in Chapter 1, many competence proteins have been shown to co-localise at polar foci in competent *B. subtilis* cells even in the absence of transforming DNA. This suggests that they form a complex. But no such complex was isolated in this study. It is possible that the competence proteins are only loosely associated in the absence of DNA. Once DNA transport begins, the whole complex may be held together by the long DNA molecule that is being transported. Another explanation could be that the DNA uptake complexes are sensitive to detergents, and even mild detergents such as LMNG destabilise the complexes. ComEC and ComFC may also be only peripherally associated with the rest of the DNA uptake machinery and dissociate from it during purification.

The only potential competence complex that was observed is between ComEC and RecA. ComEC co-purified with many cellular proteins, most of which are not competence-related. This makes it difficult to establish whether the ComEC-RecA interaction is genuine or if RecA interacts with one of the cellular contaminants that happen to co-purify with ComEC.

ComEC homologues express poorly in *E. coli* (Chapter 3). The data presented in this chapter suggest that ComEC cannot be expressed to high levels in the native organism *B. subtilis*. These observations strongly suggest that there must be some structural features of ComEC which prevent its efficient biosynthesis. If ComEC transports DNA as a monomer or dimer, only one or two copies of ComEC are required for a single transformation event.

Therefore, there is no evolutionary pressure to achieve high expression levels which would require losing the structural features of ComEC that impede its biosynthesis.

Chapter 5: Characterisation of the C-terminal domain of ComEC

5.1. Introduction

Degradation of the non-transforming DNA strand

It is well-established that during dsDNA uptake in naturally competent bacteria only one DNA strand is transported into the cytoplasm and the other strand is degraded. This has been conclusively shown in *S. pneumoniae* (Mejean & Claverys, 1993) where degradation of the non-transforming strand is performed by the membrane-bound extracellular endonuclease EndA (Puyet *et al*, 1990; Berge *et al*, 2002). Berge *et al* showed that EndA can only degrade DNA if the competence pilus assembly ATPase ComGA is active. The proteins that transport DNA through the cytoplasmic membrane, ComEC and ComFA, are not required for DNA degradation. These data show that EndA degrades the non-transforming strand in the proximity of the membrane after the competence pilus pulls the dsDNA in through the peptidoglycan. The rate of degradation of the non-transforming strand and the rate of uptake of the transforming strand is similar at around 100 bp s⁻¹ in *S. pneumoniae* (Mejean & Claverys, 1993). Degradation occurs in the 5'-to-3' direction while uptake occurs in the 3'-to-5' direction. Similar kinetics and opposite polarities suggest that the two processes occur concomitantly. *endA* null mutants are not transformable (Berge *et al*, 2002). These observations suggest that the degradation of the non-transforming strand by EndA is required for transport of the transforming strand into the cytoplasm in *S. pneumoniae*.

B. subtilis does not have a preferred DNA uptake direction (Vagner *et al*, 1990) and it has no EndA homologues. The membrane-associated extracellular endonuclease NucA fragments the transforming dsDNA in *B. subtilis* thereby increasing the number of free DNA ends available for initiation of DNA transport through the inner membrane (Provvedi

et al, 2001). *nucA* null mutants have a decreased rate of DNA uptake, but NucA is not essential for transformation. NucA does not degrade the non-transforming strand unlike EndA in *S. pneumoniae*. Nevertheless, one strand of DNA is still degraded in *B. subtilis* (Vagner *et al*, 1990), although the nuclease responsible remains unknown. In contrast to *S. pneumoniae*, it has been reported that degradation of the non-transforming strand in *B. subtilis* requires the cytoplasmic membrane DNA transport channel ComEC and the cytoplasmic ATPase ComFA (Inamine & Dubnau, 1995; Provvedi *et al*, 2001). The role of ComFA in DNA degradation is unclear. By contrast, the C-terminal domain (CTD) of ComEC, which has a metallo- β -lactamase (MBL) fold and is likely extracellular, has been proposed to act as the nuclease responsible for converting dsDNA to ssDNA (Inamine & Dubnau, 1995; Baker *et al*, 2016; Dubnau & Blokesch, 2019). However, the proposed nuclease activity of ComEC has never been experimentally demonstrated.

The metallo- β -lactamase family of hydrolases

MBL family members are found in both prokaryotes and eukaryotes and vary in their biological functions and substrate specificities (Daiyasu *et al*, 2001; Dominski, 2007). All members of the family have a common $\alpha\beta/\beta\alpha$ structure, which consists of two core β -sheets and a number of external α -helices. The $\alpha\beta/\beta\alpha$ fold was first identified in the crystal structure of the *Bacillus cereus* zinc-dependent MBL (Carfi *et al*, 1995), from which the name of the family is derived. MBLs share five conserved sequence motifs (I-V) that normally bind two catalytic metal ions through aspartate and histidine residues (Dominski, 2007; Palzkill, 2013). The motif II sequence H-X-H-X-D-H is the signature motif of the family. Most characterised MBLs bind zinc ions. However, iron-specific and metal-promiscuous MBLs have also been reported (Schilling *et al*, 2003; Carrondo *et al*, 2000).

Whilst canonical MBLs hydrolyse the amide bond of the β -lactam ring in β -lactam class antibiotics, other members of the family include nucleases involved in DNA repair and RNA maturation (Daiyasu *et al*, 2001; Dominski, 2007). Some of the MBL nucleases, such as Artemis (DNA repair) and CPSF-73 (mRNA processing), lack motif V and instead contain the β -CASP subdomain which defines a separate subfamily (Callebaut *et al*, 2002).

Ribonuclease Z (RNase Z) is a well-studied MBL nuclease. It has endonuclease activity and cleaves the 3' end of tRNAs during tRNA maturation. Figure 5.1A shows the binuclear Zn^{2+} catalytic centre of RNase Z from *B. subtilis* (de la Sierra-Gallay *et al*, 2005). RNase Z contains all five MBL motifs (Dominski, 2007). One Zn^{2+} is coordinated by D67, H68 (both motif II) and H269 (motif V), and the second Zn^{2+} by H63, H65 (both motif II) and H140 (motif III). D211 (motif IV) bridges both metal centres and D37 (motif I) positions H68. A water molecule between the two Zn^{2+} sites is also observed in the crystal structure.

The catalytic activity of MBLs arises from activation of a water molecule bound to the two metal ions. Active site residues not involved in metal binding deprotonate the water molecule and the two metal centres stabilise the resulting hydroxide (Figure 5.1B) (de la Sierra-Gallay *et al*, 2005; Llarrull *et al*, 2007; Pettinati *et al*, 2016). Additional interactions of the substrate with the metal centres as well as with residues not shown in Figure 5.1 are also important for stabilising developing charges on the substrate during hydrolysis.

In subsequent sections of this chapter the amino acid sequence features of the ComEC CTD are compared to that of RNase Z, and purification and functional characterisation of the CTD from *M. thermoacetica* ComEC (MthCTD) is presented.

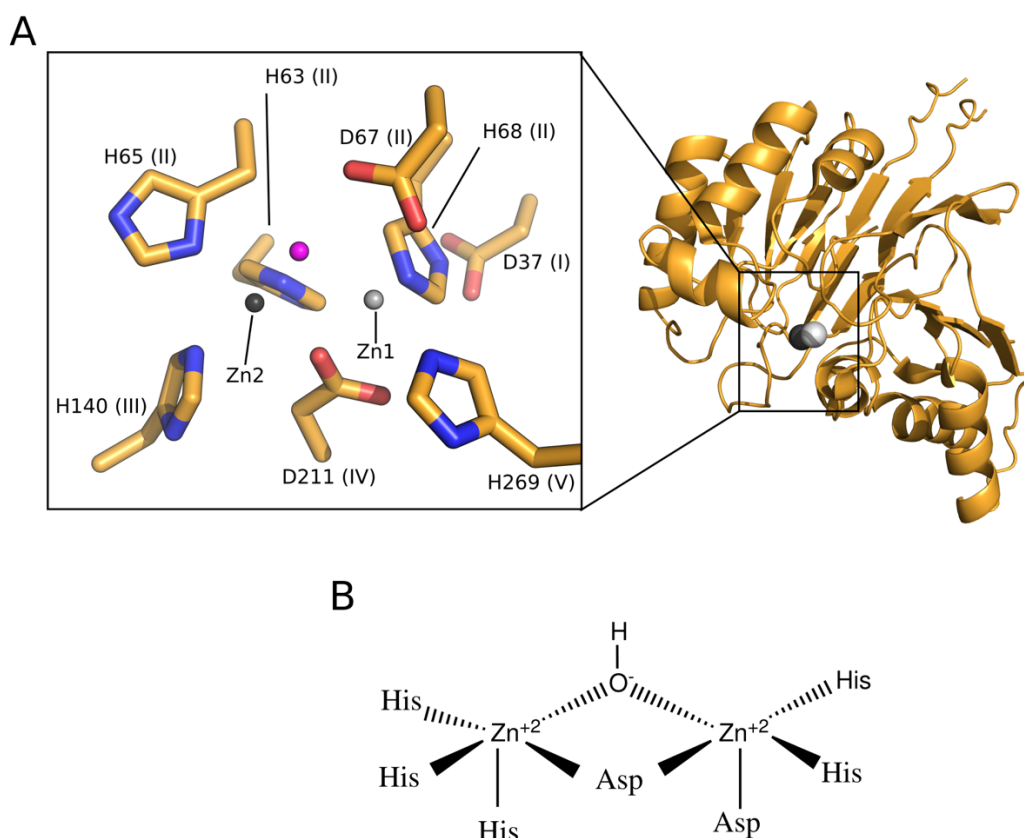


Figure 5.1. The active site of a binuclear MBL endonuclease. (A) RNase Z binds two Zn^{2+} ions (light and dark grey) via the conserved residues in motifs II-V (motif number in parentheses next to each residue label). D37 of motif I stabilises the metal-binding H68 of motif II. A water molecule (pink) is observed between the two Zn^{2+} ions in the crystal structure (PDB 1Y44)(de la Sierra-Gallay *et al*, 2005). (B) A plausible generic basis for MBL catalysis could be deprotonation of a water molecule by the active site residues (not shown) and stabilisation of the resulting hydroxide by the metal centres. The hydroxide can then perform a nucleophilic attack on the substrate.

5.2. The metallo- β -lactamase domain of ComEC

The ComEC CTD is predicted to have an MBL fold based on sequence similarity to other members of the MBL superfamily (El-Gebali *et al*, 2019) and on this basis a homology model of *B. subtilis* ComEC CTD (BsuCTD) has been reported (Baker *et al*, 2016). The structural similarity of the CTD to MBL domains was confirmed by submitting a number of ComEC CTD homologue sequences for modelling on the Phyre2 server (Kelley *et al*, 2015). This programme predicts MBL folds for all the tested sequences with >99% confidence (Figure 1.3B).

ComEC CTD sequences from 12 phylogenetically diverse organisms were aligned (Figure 5.2). Four key observations can be made from the alignment. First, MBL motifs I, II, IV and V are conserved, but motif III is not present. Motif III directly interacts with one of the metal ions through a conserved histidine residue (H140 in RNase Z)(Figure 5.1A). The intervening sequence between motifs II and IV does not contain any conserved histidine or aspartate residues that could fulfil the role of motif III. This could mean that a residue previously unobserved in MBL metal binding centres is performing the role of motif III, or that one of the metal binding sites is disrupted in the ComEC CTD.

Second, motif II of most ComEC CTDs has an aspartate present in position 3 instead of the canonical histidine (H65 in RNase Z). No characterised MBLs have this feature. The structure of the MBL rubredoxin:oxygen oxidoreductase shows that although this enzyme has a glutamate residue in position 3 it can still bind to an iron centre (Carrondo *et al*, 2000). It is difficult to predict what effect, if any, the histidine-to-aspartate substitution could have on metal binding and the catalytic activity of ComEC CTD.

Third, motifs II and IV of *S. pneumoniae* ComEC CTD (SpnCTD) are different from those in the other sequences. In addition to an aspartate at position 3, motif II of this homologue has an asparagine in position 1 instead of the canonical histidine, and a glutamate in position 5 instead of the canonical aspartate. In motif IV of SpnCTD, the aspartate bridging the two metal ions (D211 in RNase Z) is substituted with an asparagine. Such extensive deviation from the canonical MBL motifs suggests that the SpnCTD may be metal binding-deficient and catalytically inactive. If the function of the ComEC CTD is indeed degradation of the non-transforming DNA strand, an inactive CTD would explain the requirement for the EndA nuclease in *S. pneumoniae* (Mejean & Claverys, 1993; Berge *et al*, 2002). Many

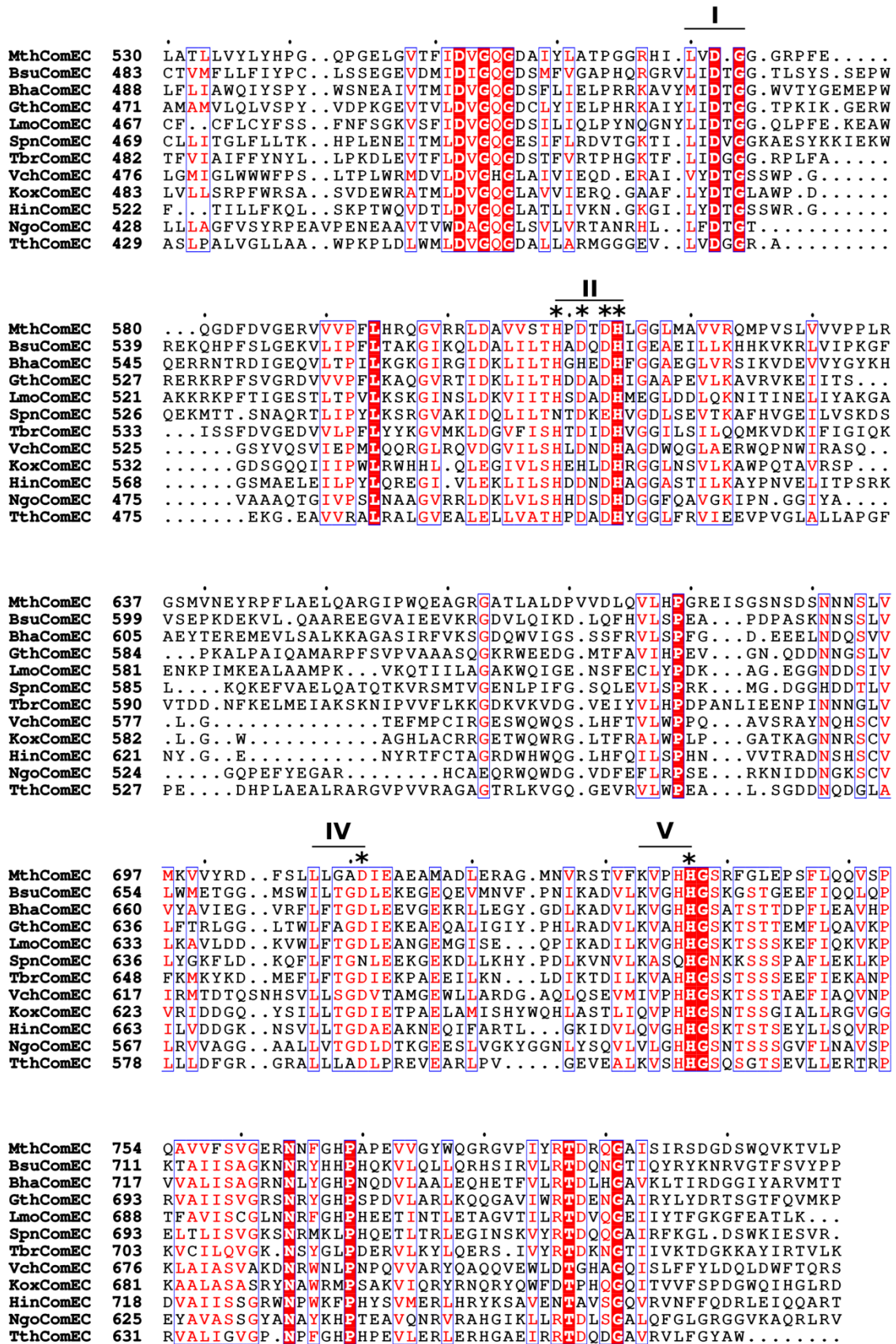


Figure 5.2. Alignment of ComEC CTD sequences. The conserved MBL motifs I, II, IV and V are annotated above the alignment. Motif III is absent in the ComEC CTD proteins (Daiyasu *et al*, 2001). The predicted metal-binding residues are marked with asterisks. Mth – *M. thermoacetica*,

Bsu – *B. subtilis*, Bha – *B. halodurans*, Gth – *Geobacillus thermodenitrificans*, Lmo – *Listeria monocytogenes*, Spn – *S. pneumoniae*, Tbr – *Thermoanaerobacter brockii* subsp. *finnii*, Vch – *V. cholerae*, Kox – *K. oxytoca*, Hin – *H. influenzae*, Ngo – *N. gonorrhoeae*, Tth – *Thermus thermophilus*.

ComEC homologues do not have a CTD (Pimentel & Zhang, 2018), so it is unclear why *S. pneumoniae* would retain the inactive domain. It is also unclear how the ComEC homologues lacking the CTD would replace its nuclease activity.

Finally, the alignment of the CTDs reveals many conserved residues that are not part of the five MBL motifs (Figure 5.2, highlighted in red). It is unclear what the roles of these residues are in CTD function.

5.3. Expression and purification of the C-terminal domain

To enable *in vitro* studies of the ComEC CTD, the CTD of *M. thermoacetica* ComEC (MthCTD) was heterologously produced in the *E. coli* cytoplasm. The coding sequence for amino acids 541-809 of MthComEC was cloned into the first multiple cloning site of the pCDFDuet-1 vector in-frame with an N-terminal MGSSHHHHHS tag (Figure 5.3A). *E. coli* BL21 (DE3) pLysS cells transformed with this construct expressed an abundance of MthCTD (Figure 5.4A and B). After NiNTA affinity purification, both high and low molecular weight contaminants were present. After size exclusion chromatography, pure MthCTD was obtained. The identity of the protein was confirmed by peptide fingerprinting mass spectrometry. The yield was approximately 6 mg of pure protein per litre of cell culture.

The two aspartate residues in motif II of MthCTD (positions 611 and 613 in the full-length protein) are predicted to bind to different metal ions in the catalytic centre. Substitution of both of these residues with alanines was used to generate the MthCTD-

DADA variant (Figure 5.3B). MthCTD-DADA was expressed and purified under the same conditions as MthCTD (Figure 5.4C and D). The yield of pure MthCTD-DADA was comparable to that of MthCTD. Both MthCTD and MthCTD-DADA samples were monodisperse, as shown by sharp symmetrical peaks in size exclusion chromatography traces (Figure 5.4A and C). The elution volume of both wild type and mutant MthCTD was identical and between the elution volumes of 43 kDa and 13.6 kDa reference proteins, indicating that the MthCTD samples were monomeric (30 kDa expected monomer size).

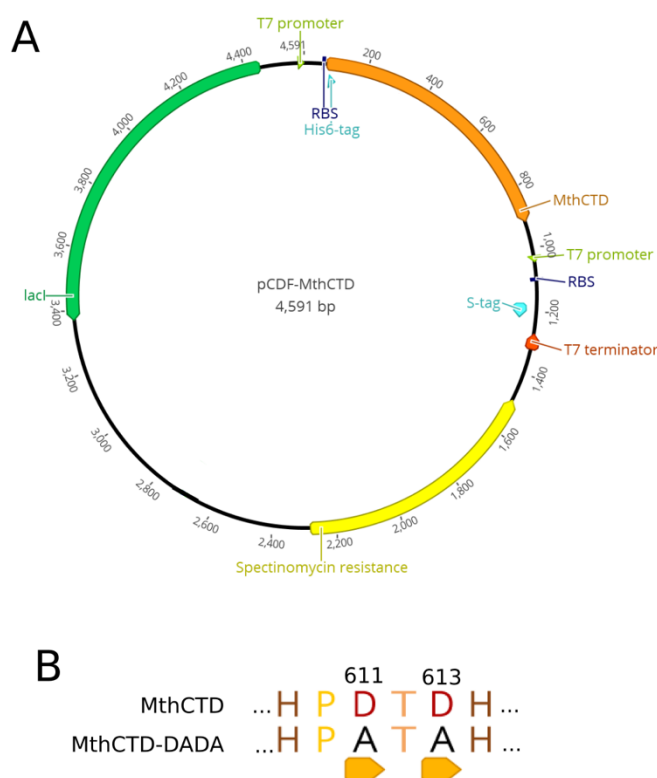


Figure 5.3. MthCTD expression constructs. (A) Plasmid map of pCDF-MthCTD. (B) The MBL motif II in MthCTD and MthCTD-DADA.

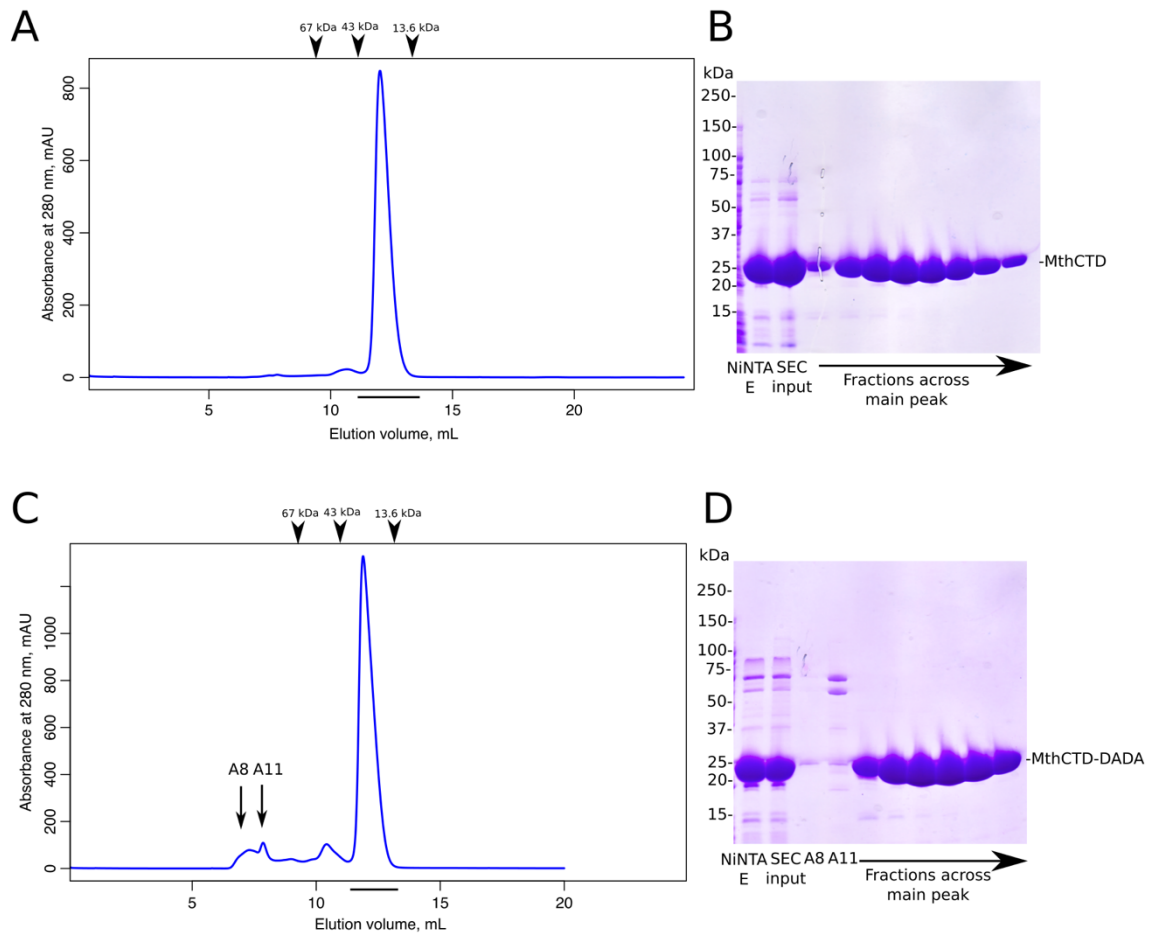


Figure 5.4. Purification of MthCTD and MthCTD-DADA. Superdex 75 10/300 GL size exclusion chromatography traces and SDS-PAGE gels of NiNTA eluates (E) and size exclusion fractions for (A-B) MthCTD and (C-D) MthCTD-DADA.

MthCTD is not toxic to E. coli

Nuclease toxins are prevalent in bacterial competition systems. For example, many bacteriocins and type VI secretion system effectors are nucleases (Papadakos *et al*, 2012; Durand *et al*, 2014). Unless target cells express the appropriate immunity proteins, nuclease toxins delivered to the cytoplasm cause cell death. Overexpression of a nuclease in the cytoplasm should also lead to death of the expression host due to degradation of genomic DNA. Therefore, if the ComEC CTD is a nuclease, the high level of expression of MthCTD was surprising. The BsuCTD has been proposed to have Zn^{2+} -dependent nuclease

activity (Baker *et al*, 2016). The intracellular concentration of Zn^{2+} in *E. coli* can reach up to 100 μM , but most of it is likely bound by various host enzymes and almost none is freely available (Outten & O'Halloran, 2001). A possible explanation for lack of toxicity is that MthCTD is not metallated properly in *E. coli* because there is not enough of the appropriate metal ion available in the cytoplasm. MthCTD would also not degrade host genomic DNA if it is not an endonuclease, if it does not fold correctly in the cytoplasm, or if the conditions in the *E. coli* cytoplasm prevent its activity.

5.4. Biophysical characterisation of MthCTD

Circular dichroism spectroscopy

During translation in naturally competent organisms the CTD of ComEC is threaded through the Sec translocon and folds in the extracellular/periplasmic space. Therefore, expression in the *E. coli* cytoplasm could result in MthCTD misfolding. Introducing substitutions into the amino acid sequence of a protein can also lead to changes in structure as metal binding sites may have a structural function in addition to a catalytic function. Consequently, the secondary structure content of MthCTD and the MthCTD-DADA mutant was investigated by circular dichroism (CD) spectroscopy to assess whether the two proteins are folded with similar secondary structure content. The CD spectra of both MthCTD and MthCTD-DADA show a strong positive signal at 195 nm and a broad negative band in the 208-222 nm region (Figure 5.5A). These signals are distinct from those of disordered species, which have negative signal below 200 nm and weak signal above 210 nm (Greenfield, 2006). Therefore, it can be concluded that purified MthCTD and MthCTD-DADA are folded. MthCTD has slightly higher signal at 195 nm and lower

signal at 208 nm, which indicates that it has a higher α -helical content than that of MthCTD-DADA.

The CD spectra of MthCTD and MthCTD-DADA were submitted to the K2D3 server (Louis-Jeune *et al*, 2012) for secondary structure deconvolution. K2D3 compares the input CD spectrum to a non-redundant set of reference CD spectra calculated from crystal structures in order to estimate the secondary structure content of the input spectrum. As expected from the shape of the CD spectra, MthCTD had a higher α -helical content than MthCTD-DADA by 5% (Figure 5.5B). The β -sheet content was similar for both samples. MthCTD and MthCTD-DADA had a lower estimated β -sheet content than the *S. pneumoniae* phosphorylcholine esterase CbpE (Figure 5.5B). CbpE was the top hit from Phyre2 homology modelling of MthCTD (Figure 1.3B). It is possible that the structural differences arise due to lack of a metal bound in MthCTD as the spectra were collected without any metal ions in the buffer. Whether the presence of metals affects the secondary structure of MthCTD has not yet been tested.

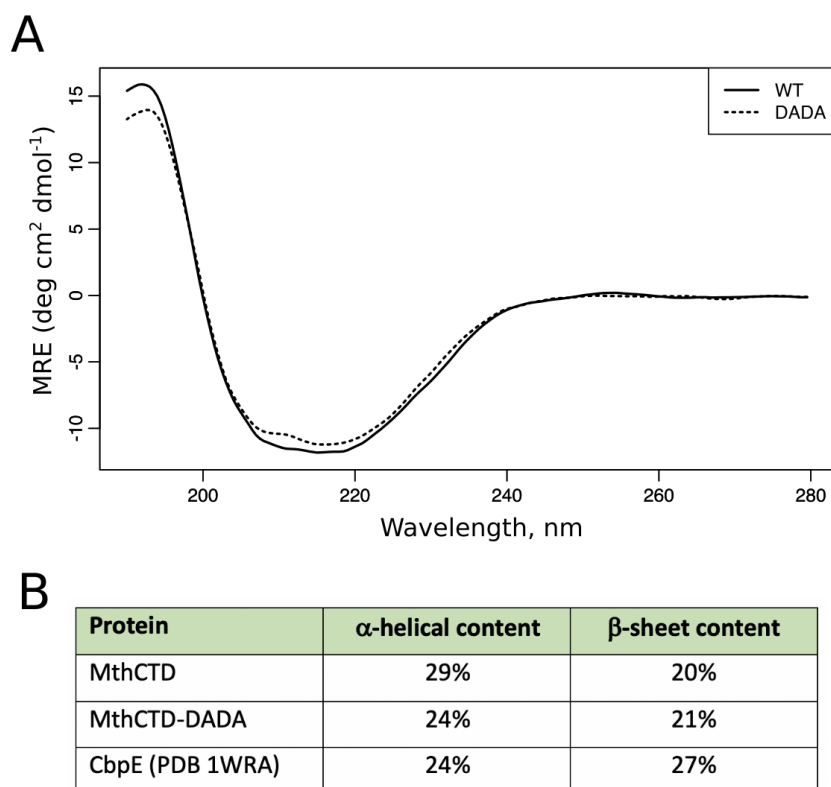


Figure 5.5. MthCTD secondary structure analysis using circular dichroism spectroscopy. (A) CD spectra of MthCTD (solid line) and MthCTD-DADA (dashed line) in 20 mM sodium phosphate pH 7.4 buffer. Spectra were collected in the 190-280 nm wavelength range at 20°C. The mean residue ellipticity is plotted against wavelength. (B) Deconvolution of MthCTD and MthCTD-DADA CD spectra. Secondary structure content was estimated in K2D3 (Louis-Jeune *et al*, 2012) from CD data in the 190-240 nm region. For comparison, the secondary structure content of a metallo- β -lactamase-like protein, CbpE, is shown as determined from the crystal structure (PDB 1WRA). Experimental details are described in Chapter 2.

Thermal shift assay

Thermal shift assays were carried out to determine the metal binding preferences of MthCTD. Initially, the assays were carried out with a His₆-tagged version of the protein (Figure 5.6A). Only Mn²⁺ gave a statistically significant increase in the T_m (13°C) for the MthCTD sample. The MthCTD-DADA variant was not stabilised by Mn²⁺, indicating that D611 and D613 are metal binding residues in MthCTD. Notably, although BsuCTD has been proposed to bind Zn²⁺ (Baker *et al*, 2016) neither MthCTD nor MthCTD-DADA was stabilised by Zn²⁺ (Figure 5.6A).

The MthCTD and Mth-DADA constructs were modified to include a TEV cleavage site between the N-terminal His₆-tag and the MthCTD sequence in order to prevent potential false positive results due to the His₆-tag binding to the metals. The tag-less constructs (MthCTD-TEV and MthCTD-DADA-TEV) were assayed with Mn²⁺ and Zn²⁺ (Figure 5.6B). Interestingly, the tag-less MthCTD-TEV construct was more thermostable than the tagged MthCTD construct (63.2°C vs 55.4°C) and the MthCTD-DADA-TEV construct (59.5°C) in the absence of added metals. Only Mn²⁺ increased the T_m of MthCTD-TEV substantially (by 4.75°C). No change in the T_m of MthCTD-DADA-TEV was observed in the presence of Mn²⁺. A mild destabilisation (by 2.4°C) of MthCTD-TEV was observed in the presence of Zn²⁺. These results suggest that, out of the metals tested, MthCTD binds only Mn²⁺.

It should be noted that not all metals could be assayed using the thermal shift assay. Co²⁺, Ni²⁺ and Cu²⁺ were also tested, but the fluorescence signal was very weak in these conditions. This could be due to interactions of the metal ion with the fluorescent SYPRO Orange dye used in the assay.

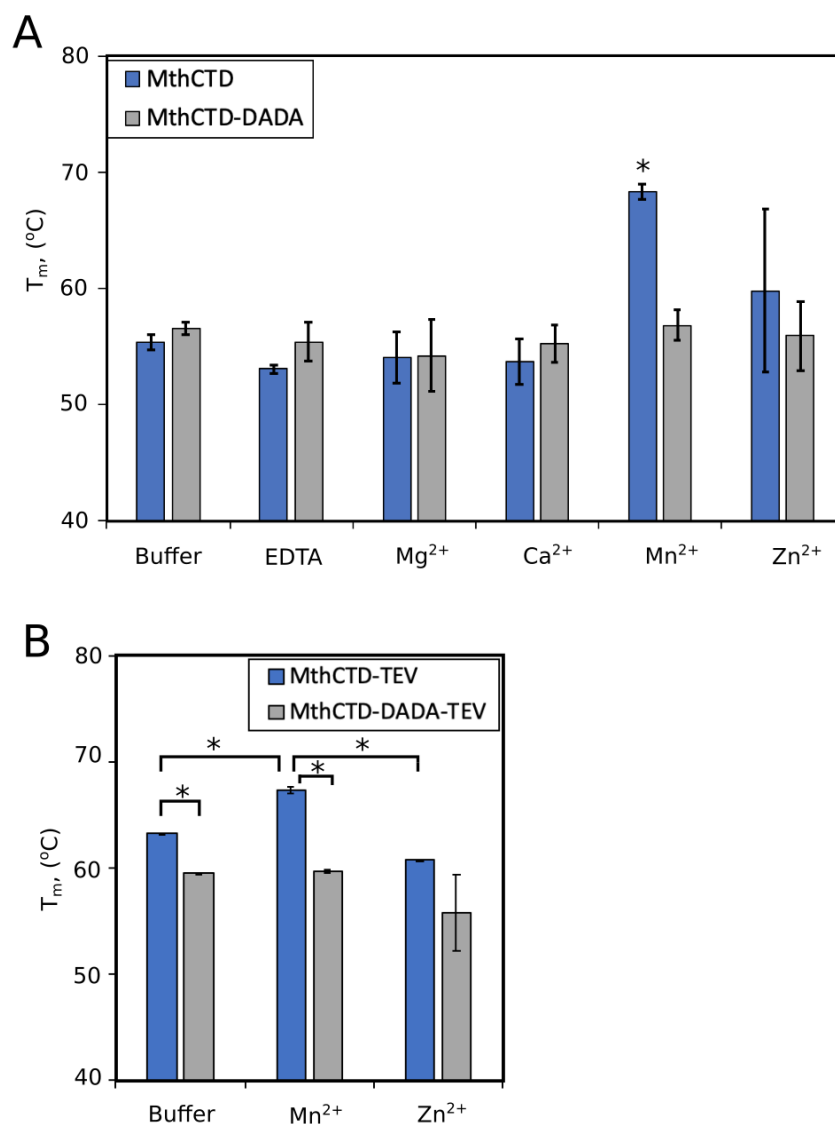


Figure 5.6. MthCTD metal binding assay. The melting temperatures, T_m , are plotted in the presence of different additives. The assay was done in 18 mM Tris-HCl pH 8.0, 135 mM NaCl, 1:5,000 dilution of SYPRO Orange dye. The concentration of EDTA and metal salt chlorides (with the exception of zinc where the sulphate salt was used) was 1 mM. The protein concentration was 12 μ M. (A) His-tagged MthCTD and MthCTD-DADA thermal stability in the presence of different metals. The asterisk denotes statistically significant differences ($p < 0.05$, Student's t -test) compared to the MthCTD sample without additives ("Buffer"). $n = 2$, except for buffer and Zn²⁺ conditions where $n = 3$. The replicates were done with protein from different preparations. (B) Thermal stability of MthCTD and MthCTD-DADA samples where the His-tag has been cleaved off by TEV protease. Asterisks denote statistically significant differences ($p < 0.05$, Student's t -test). $n = 3$. The replicates were done with protein from one preparation. Error bars represent the standard deviation.

Differential scanning calorimetry

Differential scanning calorimetry was used to confirm the thermal shift assay results. For the MthCTD (His-tagged) sample, a thermal shift of 14.3°C was observed in the presence of Mn^{2+} (Figure 5.7A), which is in agreement with the thermal shift assay data that showed a thermal shift of 13°C. The MthCTD-DADA (His-tagged) sample had two melting points in the absence of Mn^{2+} at 44°C and 58.7 °C (Figure 5.7B and C). The higher T_m corresponds well to that of MthCTD without Mn^{2+} (Figure 5.7C). The additional lower T_m of MthCTD-DADA without Mn^{2+} , which was not observed in the thermal shift assay, could be the result of MthCTD-DADA existing in two different solution states which differ in their melting temperatures. Alternatively, one part of the protein might unfold first, resulting in the first peak, and the rest of the protein unfolds at the higher temperature. Addition of Mn^{2+} to the MthCTD-DADA sample resulted in the disappearance of the lower melting point and did not significantly shift the MthCTD-DADA peak to higher temperatures. This indicates that MthCTD-DADA is still interacting with Mn^{2+} , although not in a manner that thermostabilises the protein.

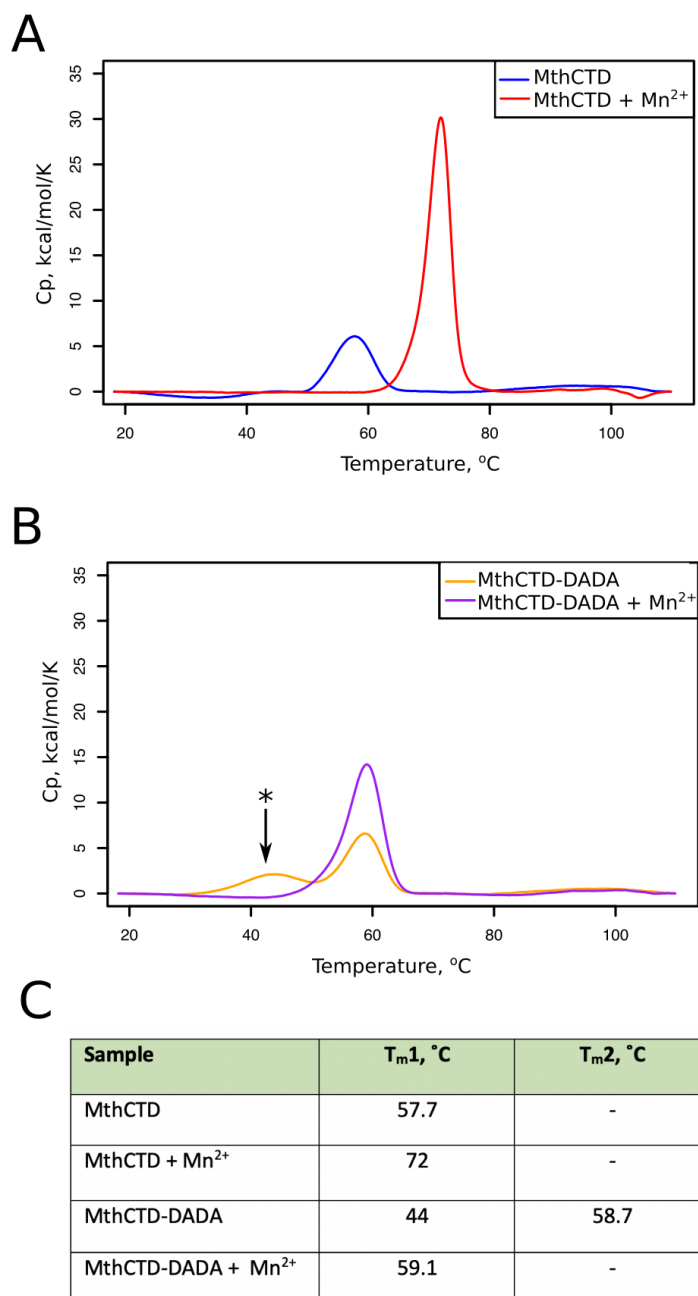


Figure 5.7. Differential scanning calorimetry of MthCTD. DSC measurements were performed by David Staunton (Department of Biochemistry, University of Oxford) on a Malvern VP Capillary DSC instrument. Experimental details can be found in Chapter 2. Thermograms of MthCTD (A) and MthCTD-DADA (B) in TBS buffer with and without 1 mM MnCl₂ are shown. Cp – partial protein heat capacity in kcal/mol/K. Peaks in the thermograms correspond to protein unfolding transitions. The asterisk notes the T_{m1} peak in the MthCTD-DADA sample without Mn²⁺. (C) Summary of T_m values extracted from the thermograms.

5.5. Attempts at MthCTD crystallisation

A structure of the MthCTD would allow a better understanding of how it binds metals and performs its function. Crystallisation of MthCTD (His-tagged) was attempted. After size exclusion chromatography pure MthCTD was concentrated in a spin concentrator with a 10 kDa cut-off membrane to 10 mg/ml. Stura Footprint screens (Molecular Dimensions) (Stura *et al*, 1992) were set up at 5 mg/ml and 10 mg/ml protein concentrations in order to estimate the appropriate concentration at which to set up crystallisation screens. The best concentration for setting up crystallisation trials would give a 50% to 50% ratio of clear drops to drops with precipitate. The 5 mg/ml footprint screen resulted in a clear drop to precipitate ratio of 60% to 40%, while the 10 mg/ml screen had a ratio of 40% to 60%. Therefore, crystallisation trials were set up at the intermediate concentration of 8 mg/ml. The following commercial crystallisation screens were used to set up crystallisation trials: Morpheus, Structure 1+2, JCSG+, PACT premier, Clear Strategy 1, Clear Strategy 2, MIDAS+ and PGA-LM. No hits were obtained. Crystallisation trials were then set up in the presence of 0.1 mM MnCl_2 , which should stabilise the protein as indicated by thermal stability assays. A number of conditions gave crystals, most of them in the PACT premier screen. All of the conditions with crystals had phosphate as the buffering reagent. When 'mock' screens with TBS and 0.1 mM MnCl_2 were set up, crystals with the same morphology appeared in the same conditions, showing that the crystals were of manganese phosphate.

5.6. MthCTD activity assays

Assessment of DNA binding

Electrophoretic mobility shift assays (EMSAs) were attempted in order to determine whether MthCTD (His-tagged) binds ssDNA or dsDNA and how tightly. However, MthCTD did not bind DNA in EMSA assays (Figure 5.8). It is possible that the binding conditions were not appropriate for MthCTD-DNA complex formation. There was no Mn^{2+} in the binding buffer, which could be required at the CTD active site to provide positive charges to interact with the DNA backbone. Alternatively, the binding affinity of MthCTD to DNA may be too low to be observed in the tested protein concentration range. MthCTD DNA binding activity warrants further investigation using biophysical methods, such as isothermal titration calorimetry.

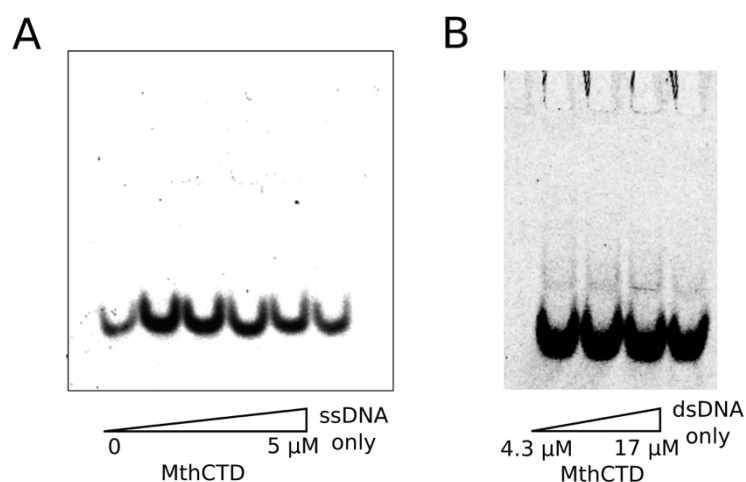


Figure 5.8. EMSA attempts with MthCTD. EMSA gels of MthCTD with (A) ssDNA and (B) dsDNA. DNA concentration was 50 nM. The DNA was 5'-Cy5-labelled. Protein and DNA samples were incubated on ice for 30 min before electrophoresis in a 6% non-denaturing gel.

Phosphoesterase assays

DNA hydrolysis involves cleavage of the phosphate backbone by a nucleophile. As discussed at the beginning of this chapter, MBLs likely hydrolyse DNA by activating a water

molecule to perform a nucleophilic attack on the backbone phosphodiester linkage (Figure 5.9A).

Bis-(*p*-nitrophenyl) phosphate (bpNPP) is a non-nucleotide phosphodiester analogue substrate (Figure 5.9B). bpNPP hydrolysis results in formation *p*-nitrophenol, which absorbs at 405 nm, and of *p*-nitrophenyl phosphate (pNPP). Hydrolysis of pNPP yields free phosphate and *p*-nitrophenol (Figure 5.9C). Nucleases are phosphodiesterases and so should cleave bpNPP but not pNPP. pNPP is cleaved by phosphomonoesterases, such as alkaline phosphatase.

The phosphoesterase activity of MthCTD (cleaved His-tag) was assayed with bpNPP and pNPP across a range of pH values and in the presence of Mn^{2+} (Figure 5.10A). As anticipated, no hydrolysis of pNPP was observed at any pH value because MthCTD is expected to be a phosphodiesterase. Phosphodiesterase activity appeared at pH 7.5 and increased with the increasing pH, reaching a maximum at pH 9.5. It should be noted that assays were also done at pH 10.0 but precipitate formation interfered with the A_{405} reading. The precipitate was brown in colour and only appeared in conditions with Mn^{2+} , suggesting that the precipitate is MnO_2 . Precipitation was also eventually visible at pH 9.5 but not during the course of the assay. High activity at basic pH is consistent with the mechanism of action of an MBL nuclease (Figures 5.1, 5.9). At basic pH, residues that deprotonate the metal-bound water molecule can transfer the acquired proton to the bulk solvent faster due to the abundance of hydroxide ions. Faster 'regeneration' of these residues means that they can deprotonate another water molecule sooner, resulting in higher rates of catalysis.

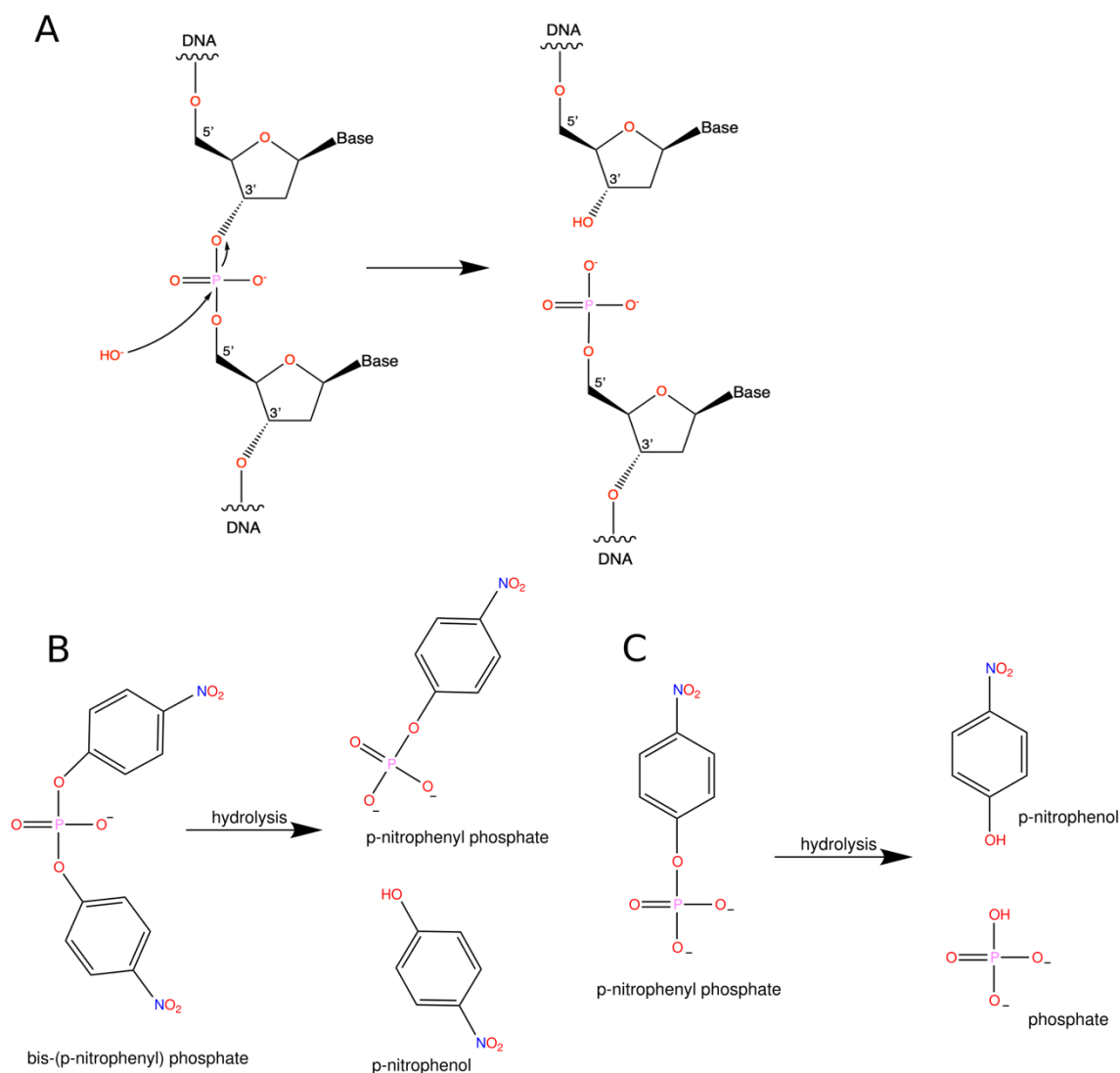


Figure 5.9. DNA, bpNPP and pNPP hydrolysis. (A) Enzyme-catalysed hydrolysis of the phosphodiester bond on the DNA backbone occurs via a bimolecular nucleophilic substitution (S_N2) mechanism (Yang, 2011). The nucleophile in the case of MBLs is a hydroxide. Whether the phosphate group ends up on the 5' or 3' nucleotide depends on the nuclease. (B) bpNPP is a non-nucleotide analogue phosphodiesterase substrate. It is hydrolysed into pNPP and *p*-nitrophenol, presumably via the same reaction mechanism as DNA. (C) pNPP is a non-nucleotide analogue substrate of phosphomonoesterases. It is hydrolysed into *p*-nitrophenol and phosphate.

Metal ions were screened for activation of the phosphodiesterase activity of MthCTD (cleaved His-tag) (Figure 5.10B). Metal ions that could not be tested in the thermal shift assay were included in this screen. However, only in the presence of Mn^{2+} did the MthCTD show statistically significant phosphodiesterase activity. This activity was dependent on the metal binding aspartates in motif II of MthCTD since the MthCTD-DADA variant was

inactive under the same conditions. These data strongly suggest that MthCTD is a Mn^{2+} -dependent nuclease.

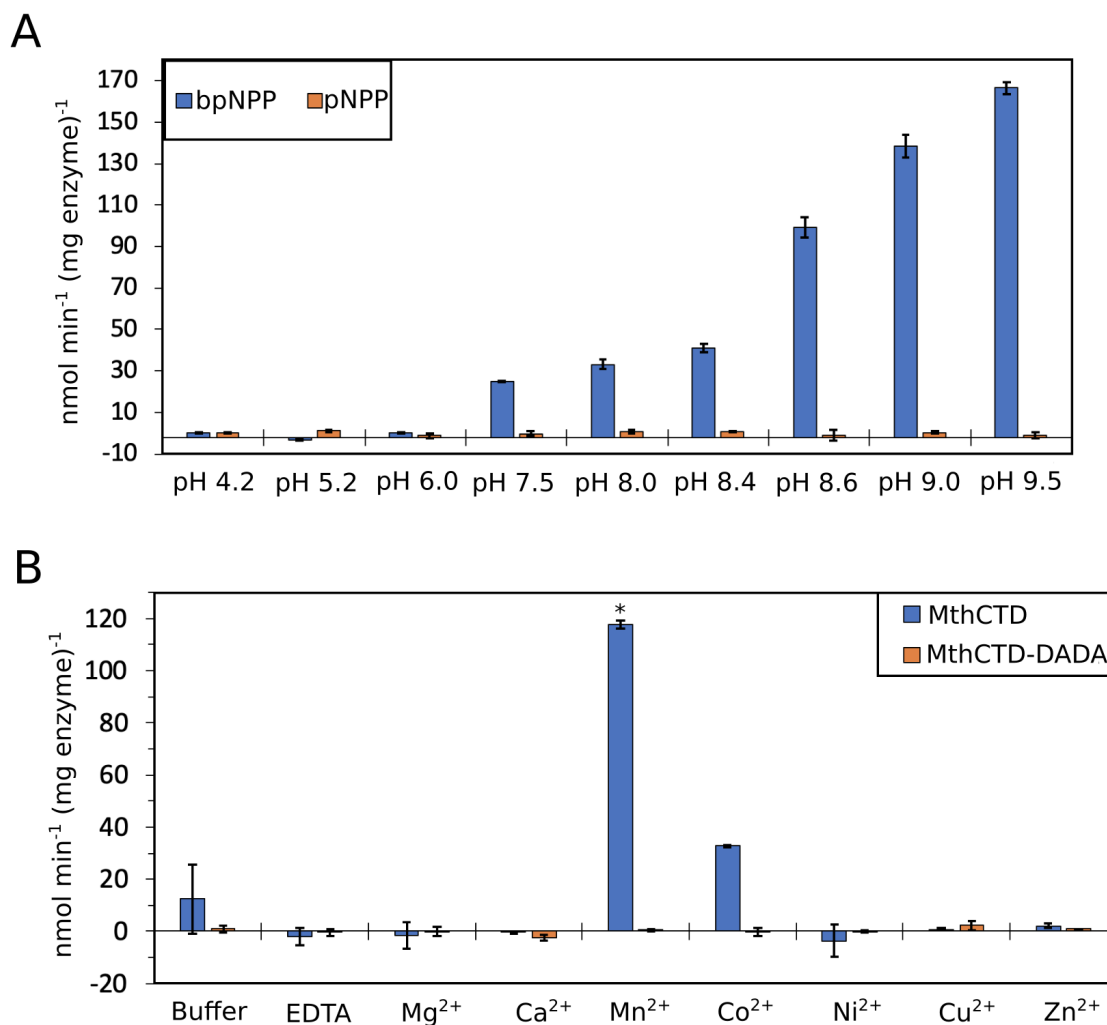


Figure 5.10. Phosphoesterase assays. Phosphoesterase activity was tested in different conditions. Production of *p*-nitrophenol was monitored by measuring A_{405} over time. Full experimental details are in Chapter 2. (A) MthCTD phosphoesterase assays at different pH conditions and in the presence of Mn^{2+} were carried out with bpNPP and pNPP. (B) MthCTD and MthCTD-DADA screen for metal-dependent phosphodiesterase activity using bpNPP as substrate at pH 9.5. Background rate of substrate hydrolysis was subtracted in all experiments. The error bars show the standard deviation, $n = 3$, protein from one preparation was used in all experiments. The asterisk denotes the condition in which the rate of *p*-nitrophenol release was significantly different to that in the buffer control ($p < 0.05$, Student's *t*-test).

If the ComEC CTD is indeed a nuclease, then it likely degrades the non-transforming strand during DNA uptake in organisms that do not have a homologue of EndA, which performs the nuclease function in *S. pneumoniae* (Mejean & Claverys, 1993; Berge *et al*, 2002). *B. subtilis* takes up DNA during natural transformation at a rate of 80 bp s⁻¹ (Maier *et al*, 2004). Assuming that (a) the rate of degradation of the non-transforming strand is similar to the rate of DNA uptake in *B. subtilis* as is the case in *S. pneumoniae* (Mejean & Claverys, 1993), (b) that every single phosphodiester bond of the non-transforming strand is cleaved, and (c) that a single CTD is responsible for degradation of the whole non-transforming strand, the rate of hydrolysis by the CTD should be ~80 phosphodiester bonds per second. The rate of bpNPP hydrolysis by MthCTD in the presence of Mn²⁺ was 117 nmol hydrolysed bpNPP per minute per milligram of MthCTD (Figure 5.10B), which corresponds to ~0.06 bpNPP molecules hydrolysed per second per molecule of MthCTD. Thus, the rate of bpNPP hydrolysis by MthCTD is three orders of magnitude smaller than would be required for degradation of the non-transforming strand in *B. subtilis*. This could be due to several reasons. First of all, the chemical nature of the *p*-nitrophenyl groups of bpNPP hardly resembles the deoxyribose sugar moiety of DNA (Figure 5.9). It is possible that bpNPP binding to the active site of MthCTD is suboptimal compared to DNA binding, which would result in lower catalysis rates. Secondly, it is possible that the 30°C temperature at which the assay was carried out was not optimal for MthCTD activity. *M. thermoacetica* is a thermophilic organism that grows optimally at 55°C and so MthCTD may require elevated temperatures for full activity. Thirdly, many ComEC monomers could associate during DNA transport and the multiple CTDs could form a more efficient DNA degradation machinery than a monomer. However, the evidence for ComEC

oligomerisation is weak (Chapter 1). Finally, the CTD may require the NTD and the competence domain for optimal activity.

Nuclease assay.

The phosphodiesterase substrate bpNPP is only an analogue of the phosphodiester bond of the DNA backbone. Therefore, a nuclease assay with DNA was performed in order to test whether MthCTD is a *bona fide* nuclease. MthCTD and MthCTD-DADA were purified without DNase I in the lysis buffer to prevent contaminating nuclease activity. Purified MthCTD and MthCTD-DADA were incubated with a dsDNA substrate in the absence of metals or in the presence of Mn^{2+} or Mg^{2+} (Figure 5.11A). The DNA used for the assay was circular supercoiled double-stranded plasmid DNA (pHT01, 8 kbp). Degradation of DNA over time could be observed by the appearance of a smear of degradation products that were smaller than the intact plasmid, as also seen with the DNase I positive control (Figure 5.11A). DNase I can be activated by Ca^{2+} , Mn^{2+} or Mg^{2+} ions. The higher activity of DNase I in the presence of Mn^{2+} than in the presence of Mg^{2+} is consistent with previously reported data (Pan *et al*, 1998). MthCTD only had nuclease activity in the presence of Mn^{2+} , but not with Mg^{2+} or without metal present. MthCTD-DADA did not show activity in any of the conditions. This assay conclusively shows that MthCTD is a Mn^{2+} -dependent DNA nuclease. Degradation of circular plasmid DNA also implies that MthCTD is an endonuclease.

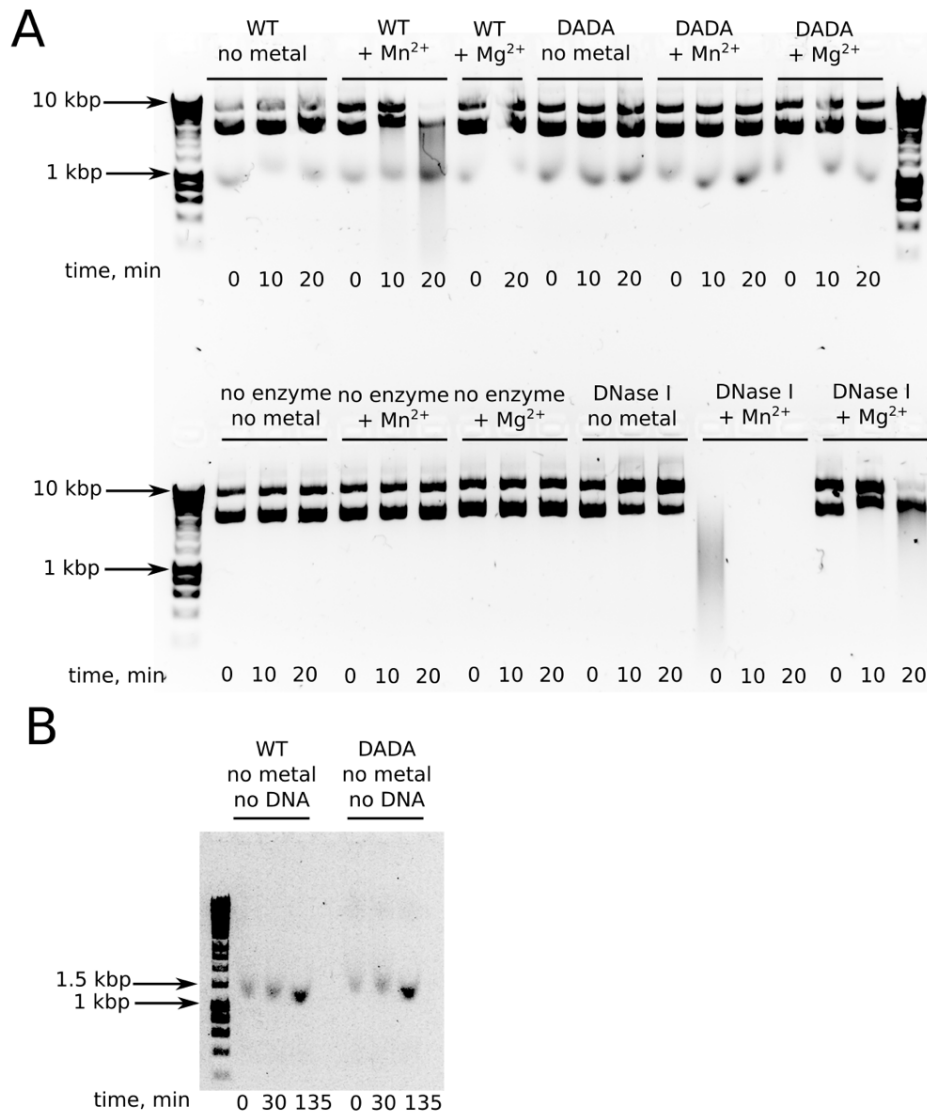


Figure 5.11. MthCTD nuclease assay. 1% (w/v) agarose gels stained with SYBR Safe DNA stain. The nuclease assay was carried out with His-tagged MthCTD and MthCTD-DADA. (A) *Top* Supercoiled double-stranded plasmid DNA was incubated at 37°C with 50 μM MthCTD or MthCTD-DADA with no metal present, with 5 mM MnCl₂ or with 5 mM MgCl₂. *Bottom* Reactions without enzyme were included as negative controls. Reactions with 2 U DNase I were included as positive controls. Aliquots from each reaction were taken immediately after adding enzyme (0 min) and after 10 min and 20 min. Aliquots were quenched by adding 6x SDS-containing purple loading dye (NEB) and 50 mM EDTA. (B) It was observed that both MthCTD and MthCTD-DADA produce a band above the 1 kbp marker even in the absence of DNA. It is likely because the proteins interact with the SYBR Safe DNA dye.

Although it is difficult to quantify the results of the nuclease assay, a qualitative assessment of the gel in Figure 5.11A suggests that MthCTD is much less catalytically active than DNase I in the presence of Mn²⁺. As mentioned previously, the thermophilic

nature of MthCTD could result in low activity due to a suboptimal temperature at which the assay was carried out (37°C). The assay was attempted at 42°C as well, but the MthCTD sample showed signs of precipitation. It is also possible that MthCTD requires the NTD and the competence domain of ComEC, or even other competence proteins for optimal activity.

It was observed that whenever MthCTD and MthCTD-DADA were included in agarose gel wells, a 1-1.5 kbp band appeared on the gel (Figure 5.11B). The initial hypothesis was that the proteins co-purify with DNA fragments from the lysate of *E. coli* and that these fragments are accessible to the SYBR Safe DNA dye. However, the A_{260}/A_{280} ratio of the purified protein samples was between 0.5 and 0.6 (preparation-to-preparation variability). If nucleic acid contaminants were present, the A_{260}/A_{280} ratio would be >1 (Zhang & Heyer, 2011). It is likely that the MthCTD and MthCTD-DADA proteins directly bind the SYBR Safe dye.

5.7. Transformability of ComEC metal binding mutants in *Bacillus subtilis*

The experiments above show that MthCTD has a nuclease function that is metal-dependent. However, the importance of this activity in the context of DNA transport was unclear and had to be tested *in vivo*. Both BsuCTD and MthCTD have the non-canonical aspartate in position 3 of MBL motif II (Figure 5.12A). The *comEC* gene of *B. subtilis* 168+ was mutated to produce strain AS6 with the DADA substitutions in motif II of the BsuCTD that are equivalent to those of MthCTD-DADA. Transformation efficiency assays with genomic DNA carrying a chloramphenicol resistance gene were carried out to test the function of the mutant *comEC* gene (Figure 5.12B). The BsuCTD-DADA mutant could still be transformed, but had a transformation efficiency that was 750-fold lower than that of

wild type *B. subtilis*. It is unclear if the *B. subtilis* ComEC-DADA mutant is correctly folded, stable and expressed at wild type levels in competent cells. However, if the observed defect in transformation is not due to biosynthesis-related issues, the results can mean one of two things. First, the DADA substitutions did not completely kill the nuclease activity of BsuCTD and the observed transformation efficiency is due to residual activity of the CTD, or the activity is substituted inefficiently by an unknown nuclease. Second, the ComEC nuclease function is not absolutely required for DNA uptake but rather increases the efficiency of DNA transport.

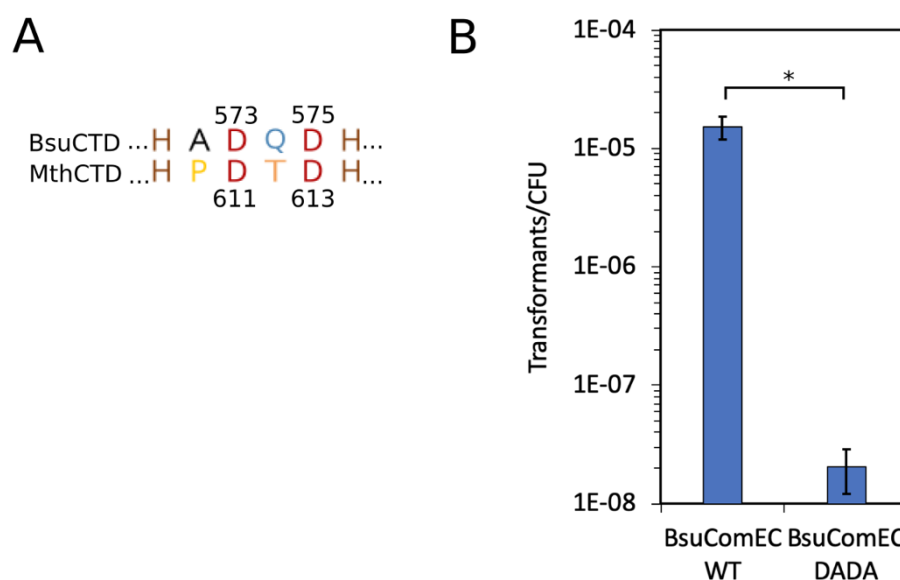


Figure 5.12. Transformation assay of a *B. subtilis* ComEC metal ion binding mutant. (A) BsuComEC-CTD and MthCTD MBL motif (II). The codons for aspartate residues 573 and 575 of BsuComEC-CTD were mutated to alanine codons on the chromosome of *B. subtilis* 168+. (B) Transformation efficiency of wild-type *B. subtilis* 168+ and the *B. subtilis* ComEC metal ion binding mutant (DADA). Cells were transformed with genomic DNA carrying a chloramphenicol resistance gene. The number of colonies that grew on selective media was normalised to the number of colonies that grew on plates without chloramphenicol selection. The error bars show the standard deviation. The assay was repeated with independent cultures; n = 2 for BsuCTD WT and n = 3 for BsuComEC DADA, * - p = 0.003 < 0.05 (Student's t-test).

5.8. Attempts at producing MthNTD

The ComEC NTD is predicted to have an oligonucleotide binding fold (Figure 1.3)(Baker *et al*, 2016). The function of the NTD of ComEC and how it interacts with the C-terminal nuclease domain to transport DNA is unclear. Similarly to the MthCTD, efforts were made to express and characterise MthNTD (not shown). When expressed by itself, MthNTD goes to inclusion bodies and very little to none is produced in soluble form in *E. coli*. It has been recently stated in the literature that the NTD and CTD interact when co-expressed in *E. coli* (Dubnau & Blokesch, 2019). In an attempt to stabilise MthNTD, co-expression from the pCDFDuet-1 vector of MthCTD and MthNTD was attempted. However, the presence of MthCTD did not rescue MthNTD. Therefore, the ComEC NTD could not be characterised.

5.9. Summary

During DNA transport through the cytoplasmic membrane one strand of DNA is degraded. In *S. pneumoniae* this degradation is performed by the EndA endonuclease but in other organisms the identity of the nuclease is unknown. The CTD of ComEC has been proposed to perform the nuclease function. Although there are no structural ComEC data available, bioinformatics analyses provide compelling evidence that the ComEC CTD has a MBL fold, but with some alterations to its canonical motifs. Therefore, the CTD likely represents a unique adaptation of the MBL fold to DNA degradation and/or transport.

MthCTD and the motif II mutant MthCTD-DADA were successfully overexpressed in *E. coli* and purified. The protein yield was high (6 mg per litre of cell culture) and the proteins were pure, stable and monodisperse. Unfortunately, crystals of MthCTD could not be obtained. Thermal shift assays and dynamic scanning calorimetry identified Mn^{2+} as the

only metal ion that could thermostabilise MthCTD. Mn^{2+} was also required for MthCTD phosphodiesterase and nuclease activities. It was shown conclusively for the first time that the CTD of ComEC is a metal-dependent endonuclease.

MthCTD phosphodiesterase and nuclease activity rates in *in vitro* assays were modest. This may be because the MthCTD originates from a thermophilic organism and the temperatures at which the assays were performed prevented the enzyme from exhibiting its full activity. Low MthCTD activity at 30°C explains why it is expressed so well in the cytoplasm of *E. coli* and does not kill the expression host.

The metal binding motif II DADA mutation was tested *in vivo* in *B. subtilis*. The BsuCTD-DADA variant had a 750-fold reduction in transformation efficiency. Assuming that the observed effect is not due to altered expression levels of the variant, the result implies that the nuclease activity of ComEC CTD does play an important role in DNA uptake during natural transformation. The CTD may perform the role of *S. pneumoniae* EndA in naturally competent organisms that lack EndA.

Chapter 6: Production and characterisation of ComFA and ComFC

6.1. Introduction

A very important and unresolved issue regarding DNA uptake across the cytoplasmic membrane during natural transformation is the energy source used to drive transport. As discussed in Chapter 1, both the PMF and ATP have been implicated as sources of energy (van Nieuwenhoven *et al*, 1982; Maier *et al*, 2004; Kramer *et al*, 2007; Diallo *et al*, 2017; Chilton *et al*, 2017). The strongest evidence for the requirement of ATP comes from studies of the cytoplasmic ComFA proteins that are required for DNA uptake in Gram-positive bacteria such as *B. subtilis* and *S. pneumoniae*. ComFA proteins are members of the DEAD-box helicase/translocase family (Londono-Vallejo & Dubnau, 1993; Chilton *et al*, 2017). DEAD-box helicases are superfamily 2 (SF2) helicases that hydrolyse ATP to unwind short, 10-12 bp RNA duplexes (Linder & Jankowsky, 2011). DEAD-box helicases contain a two consecutive recombinase A-like domains that have a number of conserved motifs important for ATP binding and hydrolysis and for RNA binding. It has been shown that mutations in the DEAD-box helicase motifs in *B. subtilis* ComFA (BsuComFA) have severe effects on transformation efficiency (Londoño-Vallejo & Dubnau, 1994b; Takeno *et al*, 2011; Chilton *et al*, 2017). *S. pneumoniae* ComFA (SpnComFA) has ssDNA-stimulated ATPase activity *in vitro* and *comFA* null mutants do not yield detectable transformants (Diallo *et al*, 2017). ComFA is cytoplasmic but localised to the cell membrane (Londoño-Vallejo & Dubnau, 1994a). It is known to colocalise with the cytoplasmic membrane DNA transporter ComEC. It also forms complexes with ComFC, an essential but poorly characterised cytoplasmic competence protein, and DprA, a cytoplasmic ssDNA binding competence protein that is involved in promoting homologous recombination of the imported DNA (Hahn *et al*, 2005; Kramer *et al*, 2007; Kaufenstein *et al*, 2011; Diallo *et al*,

2017). Together these observations support the idea that ComFA might be an ATPase that powers DNA import into the cytoplasm.

The main argument against a DNA uptake ATPase function for ComFA is that there have been no clear competence-associated ComFA homologues identified in Gram-negative bacteria even though it would be anticipated that the DNA uptake mechanism would be the same in all bacteria. The closest homologue of *B. subtilis* ComFA in Gram-negative bacteria is the PriA protein which is a DEAD-box DNA helicase/translocase that is known to be involved in primosome assembly (Jezewska & Bujalowski, 2000). Attempts at deleting the *priA* gene in *V. cholerae* were unsuccessful (Matthey & Blokesch, 2016), which is consistent with its essential function in restarting stalled replication forks. Insertional inactivation of *priA* in *N. gonorrhoeae* had pleiotropic effects including a lack of transformability (Kline & Seifert, 2005). However, it is unclear whether the transformation defect was due to impaired DNA uptake or general defects in DNA replication. Therefore, the involvement of PriA in DNA uptake by Gram-negative bacteria remains unclear.

ComFA is encoded in the same operon as ComFC in Gram-positive bacteria. ComFC is conserved in Gram-negative organisms, where it is referred to as ComF (Matthey & Blokesch, 2016). ComF is annotated as a phosphoribosyltransferase-like protein in Pfam based on sequence similarity (El-Gebali *et al*, 2019). It has been shown that deletion of *comFC* in *B. subtilis* causes a 300-fold reduction in transformation efficiency (Sysoeva *et al*, 2015). Similarly, no transformants are detected when *comF/comFC* is deleted in *V. cholerae*, *H. influenzae* and *S. pneumoniae* (Seitz & Blokesch, 2013; Larson & Goodgal, 1992; Diallo *et al*, 2017). Notably, both *comF* and *comEC* deletion mutants in *V. cholerae*

accumulate incoming DNA in the periplasm. This suggests that both proteins are involved in DNA uptake across the inner membrane and that they might be interacting (Seitz & Blokesch, 2013). However, the exact role of ComFC/ComF in natural transformation is unknown. A recent study demonstrated that SpnComFA and *S. pneumoniae* ComFC (SpnComFC) can form both homooligomers and a ComFA/ComFC complex. (Diallo *et al*, 2017). This was shown by bacterial two-hybrid assays and purification of the SpnComFA/SpnComFC complex expressed in *E. coli*. The SpnComFA/ComFC complex is around 500 kDa in size with an apparent 1:1 stoichiometry of the two components. The complex forms apparently six-fold symmetric ring-shaped objects of around 15 nm in diameter as seen in negative stain EM 2D class averages. However, the authors did not present a reconstructed EM density or report attempts at obtaining a high resolution structure in cryo-EM. The SpnComFA/SpnComFC complex presents the first macromolecular complex involved in DNA uptake across the cytoplasmic membrane to be isolated.

In Chapter 3 the successful recombinant expression of *M. thermoacetica* ComEC was reported. In order to investigate the interactions of ComEC with potential partner proteins we attempted to produce ComFA and ComFC from the same bacterial source.

6.2. Bioinformatics analysis of *MthComFA*

The *comF* operon in *B. subtilis* consists of the three genes *comFA*, *comFB* and *comFC* (Figure 6.1). The *comFB* gene is not conserved, its deletion does not have any effect on transformation in *B. subtilis*, and it has been proposed to perform a yet unknown regulatory function (Sysoeva *et al*, 2015). In *S. pneumoniae* and many other Gram-positive organisms, the *comF* operon has only two genes, *comFA* and *comFC*. BLAST analysis of the

M. thermoacetica DSM 521 genome also revealed a two-gene operon organisation as expected for a Gram-positive organism (Figure 6.1). In Gram-negative bacteria, ComF probably constitutes the only gene in the operon (Matthey & Blokesch, 2016).

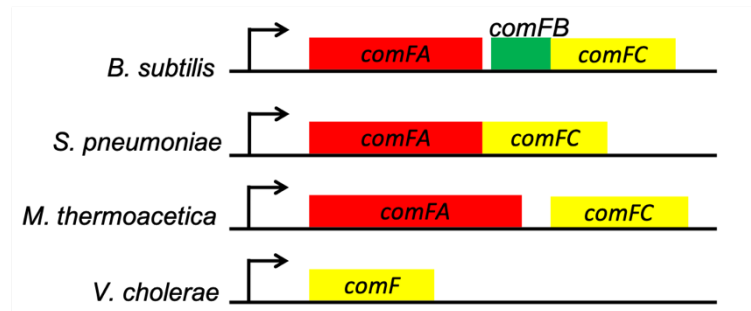


Figure 6.1. Organisation of the *comF* operon in different organisms. Homologues of the DEAD-box helicase/translocase ComFA are shown in red. Homologues of the phosphoribosyltransferase-like ComFC are shown in yellow. *B. subtilis* ComFB is shown in green.

Amino acid sequence alignment of ComFA from *B. subtilis* (BsuComFA), *S. pneumoniae* and *M. thermoacetica* showed conservation of eight canonical DEAD-box helicase/translocase motifs (Figure 6.2A) (Tanner & Linder, 2001; Chilton *et al*, 2017). Motifs I and II, also known as the Walker A and B motifs, are found in many different ATPases and are associated with ATP binding and hydrolysis. Motif I binds the ATP phosphates and coordinates a Mg^{2+} ion. Motif II also binds the Mg^{2+} ion, and the glutamate of motif II coordinates the catalytic water molecule that hydrolyses the γ -phosphate of ATP. Motifs Q, III and VI also contribute to ATP binding. Motifs Ia and Ib interact with the phosphate backbone of the RNA. The ATP binding motifs are located in a cleft between the two helicase domains, and the RNA binds on the other side of this cleft (Figure 6.2B). It has been previously shown that the aforementioned motifs are

important for the function of ComFA in natural transformation of *B. subtilis* (Londoño-Vallejo & Dubnau, 1994b; Takeno *et al*, 2011; Chilton *et al*, 2017). In addition, the (CxxC)₂ zinc finger of BsuComFA has been shown to be important for BsuComFA stability and efficient transformation (Chilton *et al*, 2017). MthComFA and SpnComFA also contain the (CxxC)₂ zinc finger motif (Figure 6.2A). NMR spectra of the first 72 residues of SpnComFA, which include the zinc finger, indicate that the region is disordered yet it is still able to bind zinc (Diallo *et al*, 2017). Therefore, it is likely that MthComFA also binds zinc. MthComFA has an N-terminal extension of around 120 amino acids not present in the other homologues (Figure 6.2A). An N-terminal extension relative to other ComFA proteins is a unique feature of ComFA from *M. thermoacetica* and closely related organisms. This extension has no obvious homology to proteins of known fold. Because there are no structures of ComFA available, the MthComFA sequence was submitted to Phyre2 for homology modelling (Kelley *et al*, 2015). Only residues 244-590 were modelled based on the crystal structure of the *E. coli* transcription-repair coupling factor (PDB 2EYQ), which has a DEAD-box helicase domain (Figure 6.2B and C). The unmodeled N-terminal extension and the putatively disordered stretch containing the zinc finger are interesting regions warranting further investigation.

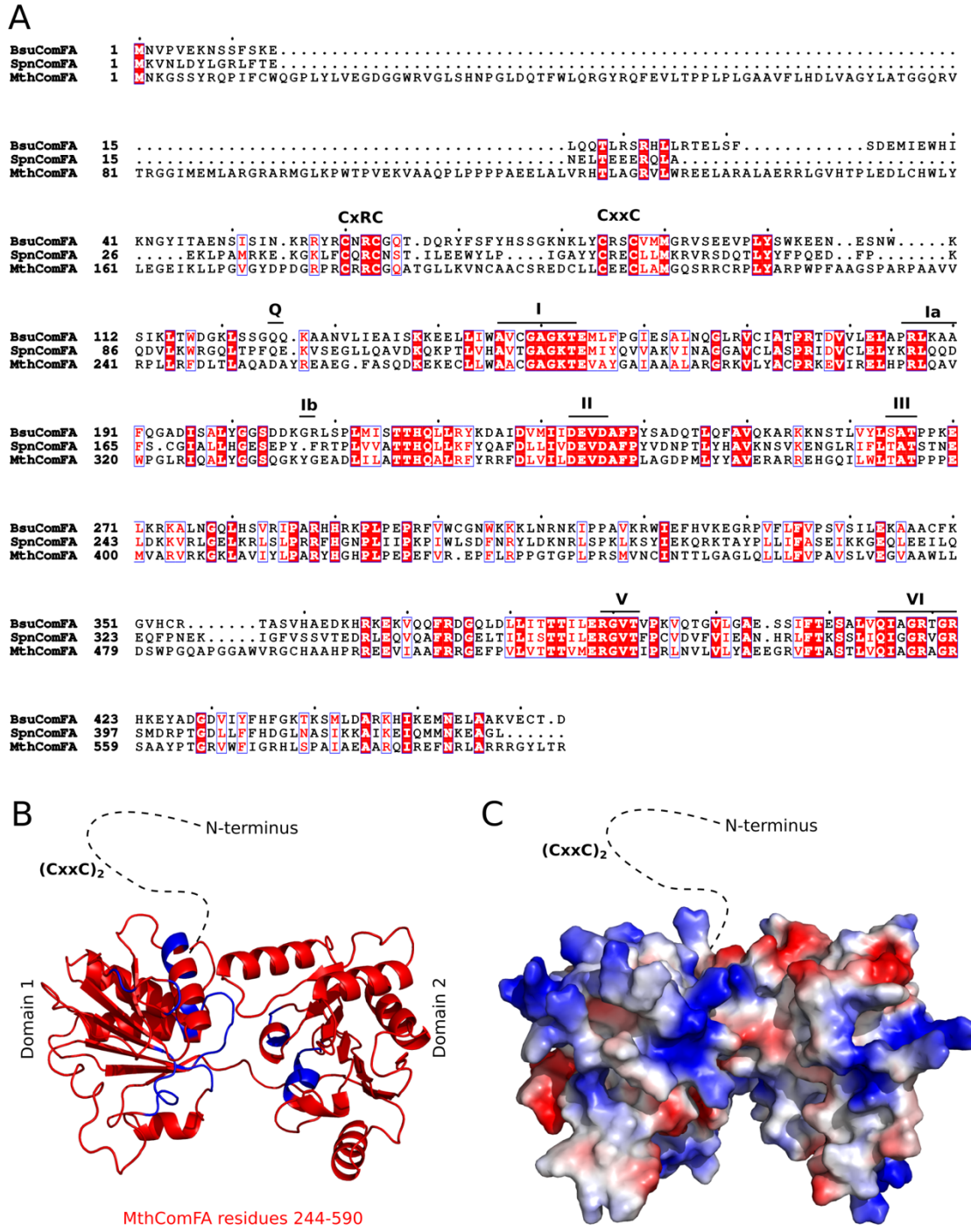


Figure 6.2. ComFA bioinformatics analysis. (A) Sequence alignment of *B. subtilis*, *S. pneumoniae* and *M. thermoacetica* ComFA amino acid. The CxxC motifs and the DEAD-box helicase/translocase motifs are annotated above the alignment (Linder & Jankowsky, 2011; Chilton *et al*, 2017). Sequences were aligned with T-Coffee and visualised in ESPrpt (Notredame *et al*, 2000; Robert & Gouet, 2014). Cartoon (B) and surface electrostatics (C) representations of the MthComFA model from Phyre2 (Kelly *et al*, 2015). Only residues 244-590 were modelled. In the cartoon representation, the conserved DEAD-box helicase/translocase motifs annotated in (A) are in blue. In the surface electrostatics representation blue represents a positive charge and red represents a negative charge.

6.3. Bioinformatics analysis of MthComFC

The ComFC sequences from *B. subtilis*, *S. pneumoniae*, *M. thermoacetica* and *V. cholerae* (VchComF) were aligned (Figure 6.3A). VchComF aligned well with the ComFC sequences from Gram-positive organisms, but it does have an N-terminal extension of around 60 amino acids that is not present in the other sequences (Figure 6.3A). All ComFC sequences contain four conserved CxxC zinc finger-like motifs (Sysoeva *et al*, 2015). While it was shown that the CxxC motifs of BsuComFA and SpnComFA bind zinc (Chilton *et al*, 2017; Diallo *et al*, 2017), it is unknown if ComFC also binds zinc as it has never been studied in detail. The MthComFC sequence was submitted to the Phyre2 server (Kelly *et al*, 2015) for homology modelling. The region containing the CxxC motifs has an unknown fold and was not modelled in Phyre2. The crystal structure of a *T. thermophilus* adenine phosphoribosyltransferase (PDB 5ZGO) was used as a template for building the homology model of residues 69-211 of MthComFC (Figure 6.3B and C). Phosphoribosyltransferases catalyse phosphoryl transfer reactions on 5-phosphoribosyl- α 1-pyrophosphate (PRPP) and are involved in nucleotide synthesis and salvage pathways or act as PRPP binding regulators (Musick & Nyhan, 1981; Sinha & Smith, 2001). They have a canonical PRPP binding motif which is conserved in the ComFC sequence alignment (Figure 6.3A). It is unclear why a phosphoribosyltransferase is required for DNA transport across the inner membrane or how it interacts with the rest of the transport machinery (Seitz & Blokesch, 2013; Sysoeva *et al*, 2015; Diallo *et al*, 2017). No *in vitro* biochemical characterisation or residue mutational analysis has been carried out on ComFC. Therefore, ComFC and its conserved motifs present a very interesting target for study.

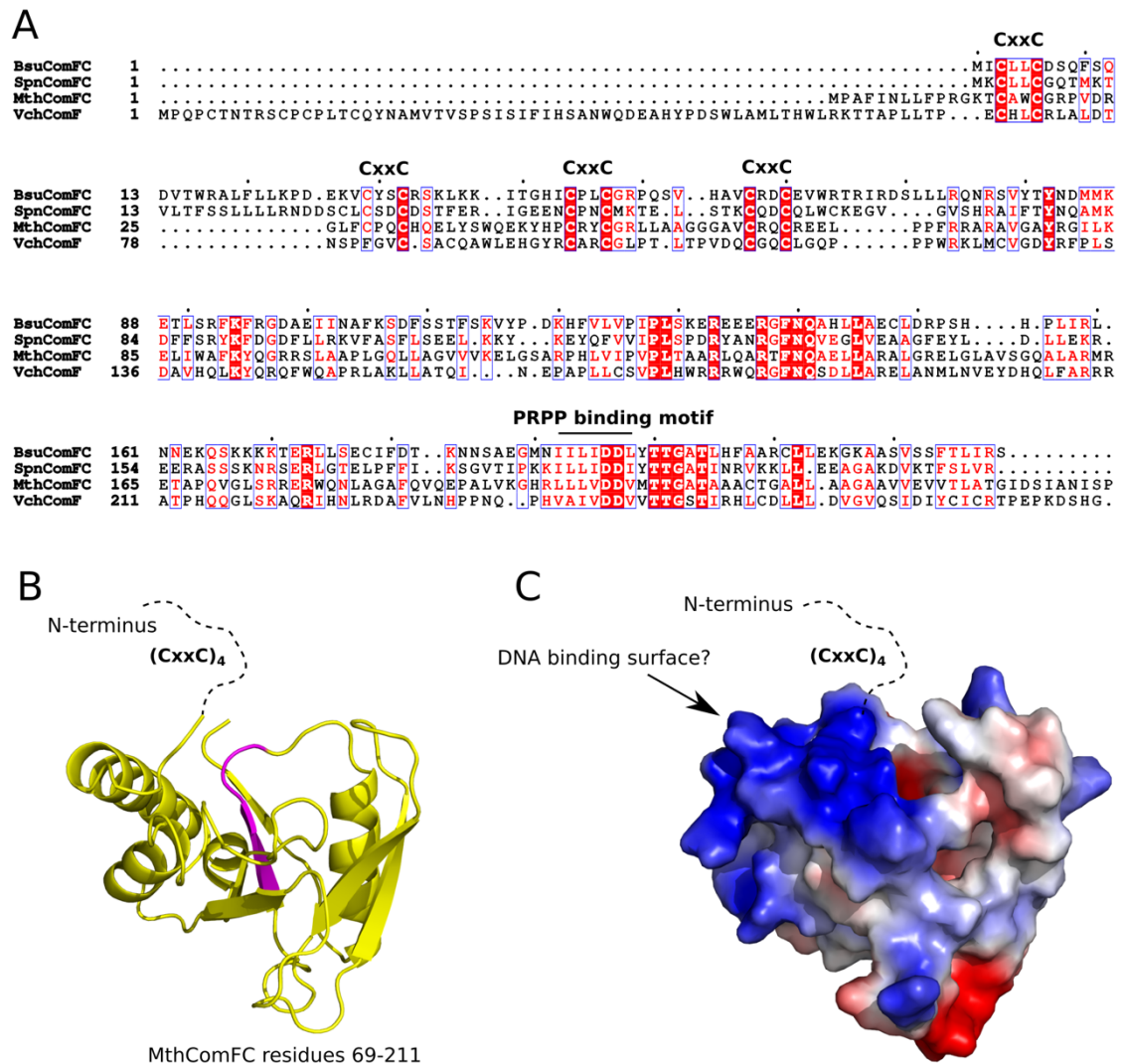


Figure 6.3. Bioinformatics analysis of ComFC. (A) Sequence alignment of *B. subtilis*, *S. pneumoniae*, *M. thermoacetica* and *V. cholerae* ComFC/ComF sequences. The CxxC motifs and the PRPP binding motif are annotated above the alignment (Sinha & Smith, 2001; Sysoeva *et al*, 2015). Sequences were aligned with Clustal Omega and visualised in ESPript (Sievers & Higgins, 2014; Robert & Gouet, 2014). (B) Cartoon and (C) surface electrostatics representations of the MthComFC model from Phyre2. Only residues 69-211 were modelled. The PRPP binding motif is depicted in magenta in (B). Blue represents a positive charge and red represents a negative charge in (C).

6.4. Production of recombinant MthComFA

The nucleotide sequence corresponding to amino acids 15-600 of MthComFA was cloned into a pQE80-L vector in-frame with the N-terminal His-tag-coding sequence MRGSHHHHHHGSAC. The resultant pQE80-L-MthComFA plasmid was transformed into *E.*

coli Rosetta 2 (DE3) cells, and the MthComFA protein was expressed and purified as described in Chapter 2. During NiNTA elution, the 66.4 kDa MthComFA protein co-eluted with a 50 kDa species (Figure 6.4A). Western blotting for the N-terminal His-tag of MthComFA showed that the 50 kDa species is a C-terminal truncation of MthComFA (Figure 6.4B). No other species that could correspond to the cleaved C-terminal part of MthComFA were observed in the SDS-PAGE gel of the NiNTA elution fractions (Figure 6.4A). The 50 kDa species should lack approximately 150 C-terminal residues to account for the 16.4 kDa size difference compared to the full-length protein. The 66.4 kDa and 50 kDa species remained associated after size exclusion chromatography (Figure 6.4C and D). The inability to separate the two fragments of different sizes on a size exclusion column indicated that full-length MthComFA forms a complex with the truncated species. It was hypothesised that full-length MthComFA and the 50 kDa C-terminal truncation may have different surface charge properties that could allow separation of the two species by ion exchange chromatography. Full-length MthComFA contains 55 acidic residues (aspartates and glutamates) and 77 basic residues (lysines and arginines), giving a net charge of +22, assuming all side chains are ionised. A 50 kDa truncation that lacks 150 C-terminal residues of the full-length MthComFA would contain 45 acidic residues and 58 basic residues, giving a net charge of +13. Therefore, full-length MthComFA should interact more strongly than the 50 kDa truncation with the negatively charged resin of a cation exchange column. During cation exchange chromatography, 50 kDa truncation eluted at around 150 mM NaCl, while the full-length MthComFA eluted at around 250 mM NaCl (Figure 6.4E-F). The cation exchange fractions containing full-length protein were collected, concentrated and subjected to size exclusion chromatography. This resulted in pure full-length MthComFA (Figure 6.4G and H).

Chapter 6

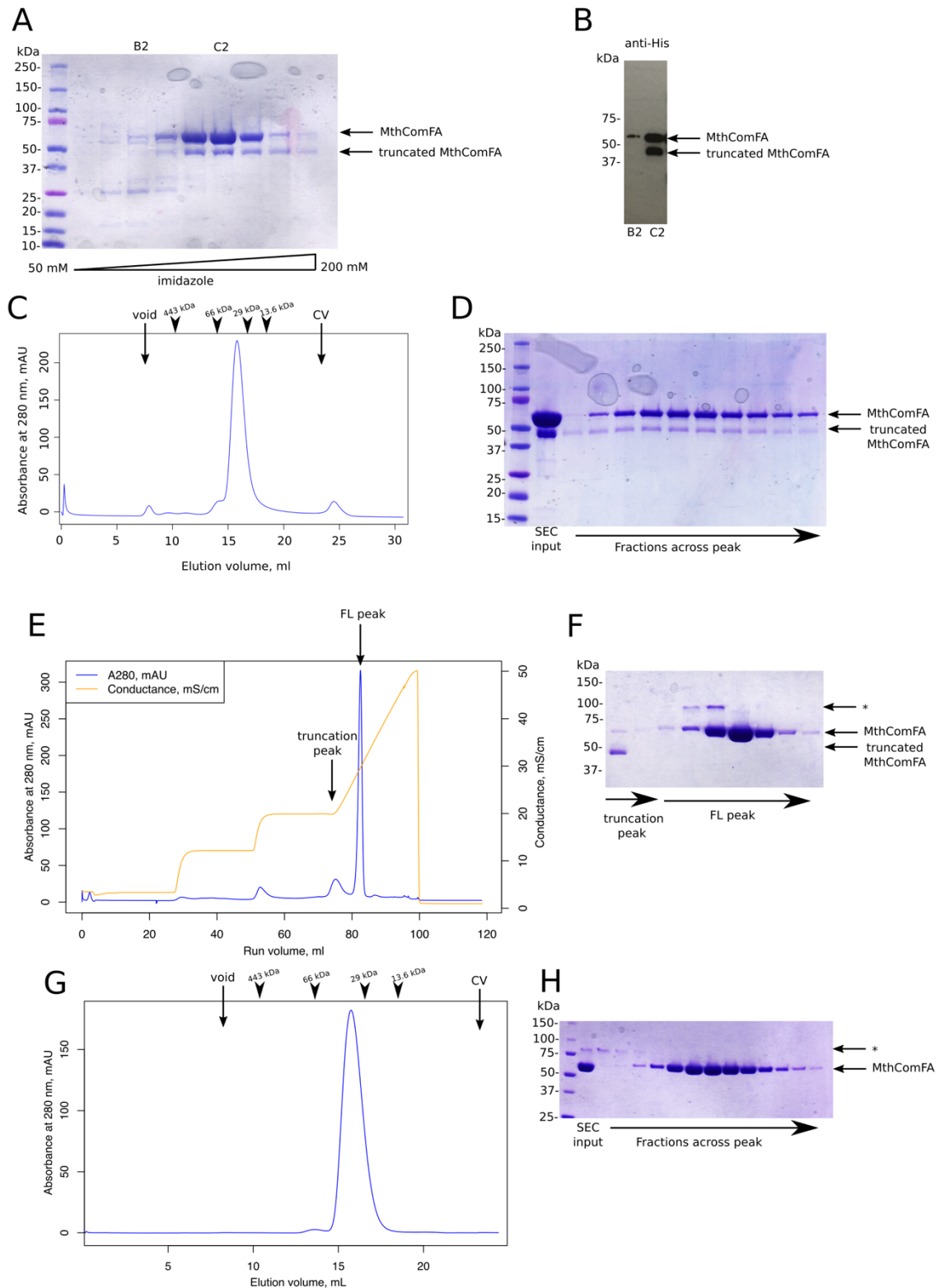


Figure 6.4. Purification of MthComFA. (A) SDS-PAGE gel of an imidazole gradient elution from NiNTA resin. (B) Anti-His Western blot of fractions B2 and C2 from the NiNTA elution. (C) Superdex 200 10/300 GL size exclusion chromatography trace of concentrated fractions from (A). CV, column volume. (D) SDS-PAGE gel of fractions across the main peak in (C). (E) Cation exchange chromatography of a NiNTA elution. The sample was flash-diluted in low ionic strength buffer before loading on a HiScreen SP HP 4.7 ml column. The column was washed with successive increases of NaCl (20 mM, 100 mM and 150 mM) before a gradient elution step (150-500 mM

NaCl), as reflected in the conductance trace. FL, full-length MthComFA. (F) SDS-PAGE gel of material eluted from the cation exchange column. (G) Superdex 200 10/300 GL size exclusion chromatography trace of fractions from the cation exchange FL peak. (H) SDS-PAGE gel of fractions across the main peak in (G). N. B. the material used in (E-H) came from a different purification than that used for (A-D). The contaminant indicated by an asterisk was occasionally present after NiNTA purification.

6.5. *MthComFA SEC-MALLS analysis*

Purified MthComFA eluted from the size exclusion column much later than expected for a monomeric 66.4 kDa protein. The apparent molecular weight of purified MthComFA was ~35 kDa as estimated from the size exclusion elution volumes of reference globular proteins (Figure 6.4G). The reason for the inconsistency could be that the target protein is interacting strongly with the column resin, resulting in an increased retention time (Arakawa *et al*, 2010). Size exclusion chromatography-multi-angle laser light scattering (SEC-MALLS) analysis also showed a delayed elution volume of MthComFA (peak at ~17 ml *versus* ~14 ml expected for a 66 kDa protein on a Superdex 200 10/300 column)(Figure 6.5). In addition, a tailing peak was observed in the SEC-MALLS A_{280} absorbance trace. Peak tailing was not observed in preparative size exclusion chromatography experiments, possibly because the flow rate in the SEC-MALLS experiment was lower. The delayed elution volume and peak tailing are consistent with the protein interacting with the Superdex 200 resin used in the size exclusion column. The light scattering data from the SEC-MALLS experiment showed that the MthComFA sample is a mixture of monomer and dimer. The estimated molar mass at the front of the absorbance peak roughly corresponds to a dimeric species (122 kDa) and gradually drops to a mass consistent with a monomer (66.4 kDa). It was previously reported that SpnComFA was also purified as a mixture of oligomeric states (Diallo *et al*, 2017). However, in this case the two states could be resolved by size exclusion chromatography. The mixed monomer-dimer population in

the MthComFA SEC-MALLS analysis could be due to dimers being broken by strong interactions with the column resin. Therefore, it remains unclear whether the protein is fully dimerised in solution.

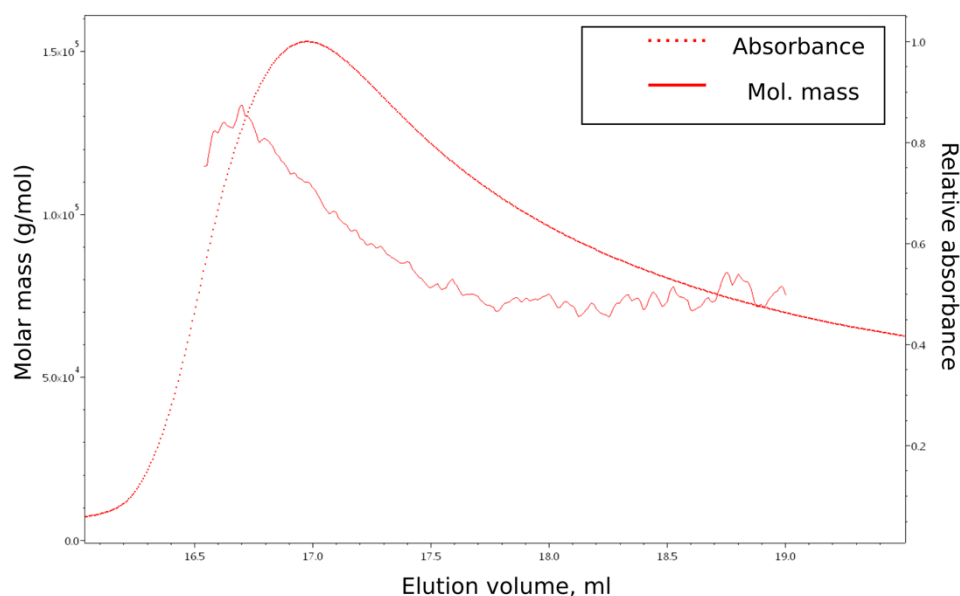


Figure 6.5. SEC-MALLS analysis of MthComFA. The SEC-MALLS analysis was carried out by Steven Johnson at the Sir William Dunn School of Pathology (University of Oxford). A chromatography system with a Superdex 200 10/300 GL size exclusion column connected to a multiangle light photometer (Dawn Heleos, Wyatt) was used. Light scattering data were used to calculate the total molar mass (solid red trace) at any point in the elution profile (red dots). The protein concentration in the injected sample was 0.9 mg/ml. The experiment was done in 20 mM Tris-HCl pH 8.0, 150 mM NaCl buffer.

6.6. MthComFA DNA binding assay

MthComFA was assayed for DNA binding activity. Initially, microscale thermophoresis with 5'-Cy5-labelled oligonucleotides was attempted but no binding was observed and there were signs of protein precipitation in the thermophoresis capillaries (not shown). Electrophoretic mobility shift assays (EMSAs) were then performed. Binding to both ssDNA and dsDNA was observed, with half-maximal binding at around 350 nM MthComFA for both species (Figure 6.6). This is an unexpected result because ComFA is thought to

interact only with ssDNA as it emerges in the cytoplasm from the membrane pore formed by ComEC. SpnComFA binding to ssDNA with a similar affinity has been previously shown (Diallo *et al*, 2017) but analysis of binding to dsDNA was not reported.

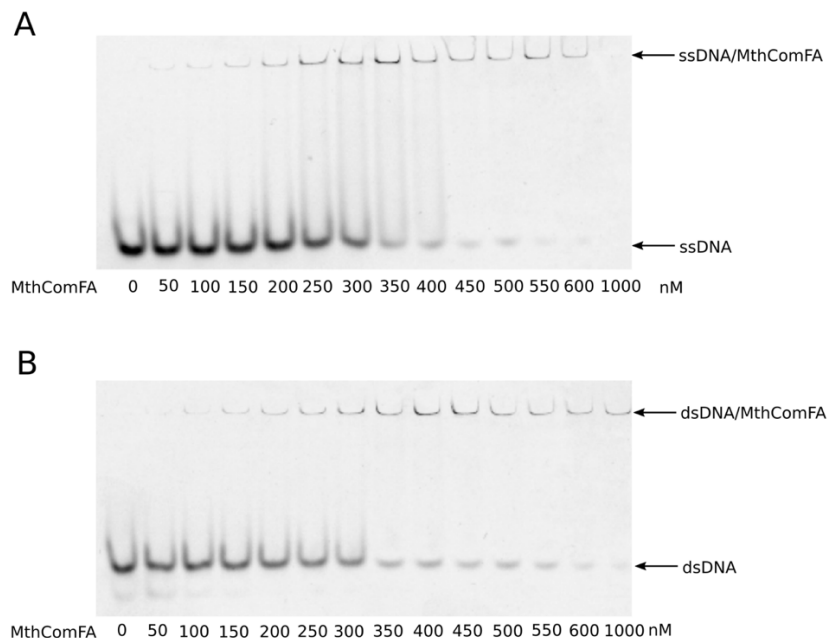


Figure 6.6. MthComFA DNA binding assay. ssDNA (A) and dsDNA (B) binding to MthComFA. 50 nM DNA (ACTG)₈ oligonucleotides labelled with a 5'-Cy5 dye were incubated with protein for 30 min on ice and free and bound DNA was separated on 6% native PAGE gels. The gels were imaged in a Fuji FLA-3000 laser scanner. Representative gels are shown.

Although MthComFA clearly binds both ssDNA and dsDNA in an EMSA, DNA did not co-purify during isolation of the protein as judged by an A_{260}/A_{280} ratio of 0.5-0.6 (nucleic acid contamination would yield a ratio of >1 (Zhang & Heyer, 2011)), and failure of the protein to stain with the SYBR Safe DNA stain after agarose gel electrophoresis. It is possible that the salt in the lysis buffer, extensive washing of the affinity column or the cation exchange step removed any bound DNA.

6.7. *MthComFA* metal binding assay

It was previously shown that BsuComFA and SpnComFA bind Zn^{2+} (Chilton *et al*, 2017; Diallo *et al*, 2017). A preparation of MthComFA in the presence of ZnCl_2 at all steps of purification was therefore attempted to ascertain whether the properties of the protein would change in the presence of this metal ion. In contrast to the behaviour of the protein purified without Zn^{2+} , the Zn^{2+} -treated sample eluted in two peaks on the size exclusion column, with the second peak eluting after the column volume, presumably due to strong interactions with the column resin (Figure 6.7). The protein was unstable in the presence of Zn^{2+} and precipitated within minutes of coming off the size exclusion column.

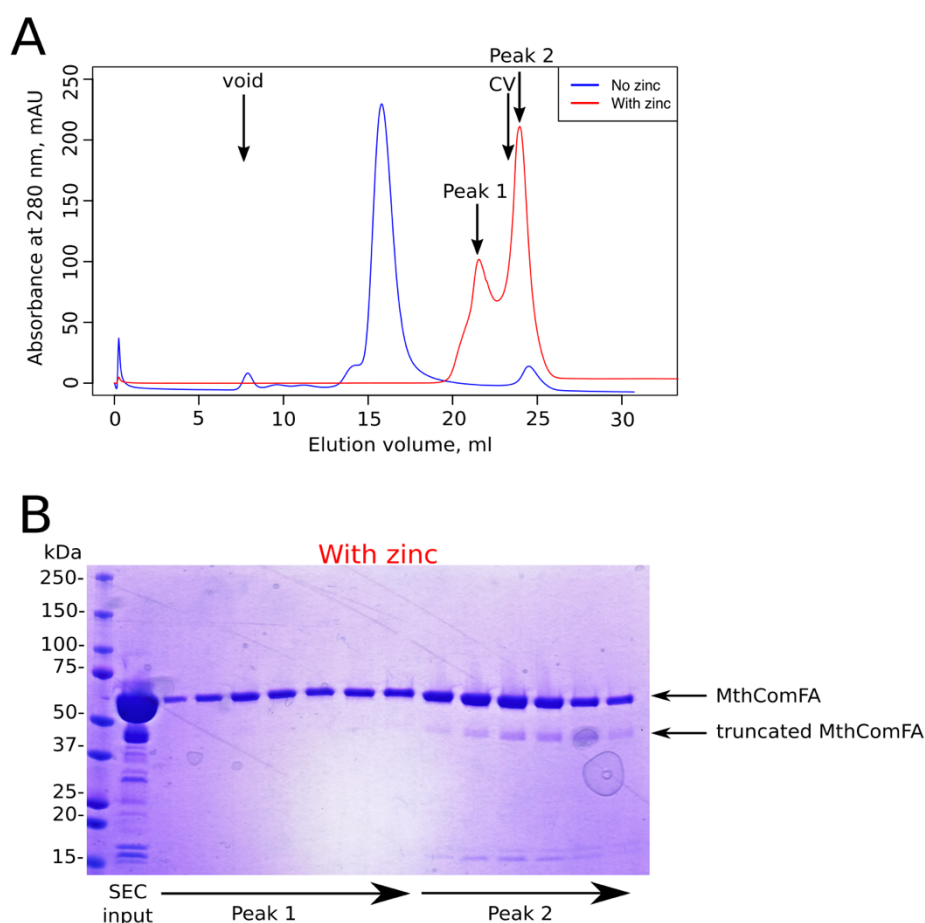


Figure 6.7. Purification of MthComFA in the presence of Zn^{2+} . (A) Size exclusion chromatography traces with and without 0.1 mM Zn^{2+} on a Superdex 200 10/300 GL column in 20 mM Tris-HCl, 150 mM NaCl buffer. The trace without Zn^{2+} is the same as that depicted in Figure 6.5C. CV, column

volume. (B) Coomassie-stained SDS-PAGE gel of fractions from the size exclusion run in the presence of Zn^{2+} .

A thermal shift assay was performed to determine the metal ion binding preference of MthComFA (Figure 6.8A). The T_m of MthComFA without additives was 70.4°C , as expected for a thermophilic protein. The only metal ion that produced a thermal shift was Zn^{2+} . However, the shift was negative (-8.3°C), indicating that the protein structure is less stable in the presence of Zn^{2+} . This is consistent with the observation that MthComFA precipitates when purified in Zn^{2+} -containing buffers. Figure 6.8B shows that in the presence of Zn^{2+} MthComFA begins to unfold much earlier than without Zn^{2+} ($\sim 40^\circ\text{C}$ vs $\sim 65^\circ\text{C}$). However, full unfolding is reached at approximately the same temperature ($\sim 75^\circ\text{C}$). Normally, the transition from the folded to unfolded state occurs over a narrow temperature range, as seen for the condition without Zn^{2+} . The transition over a wide temperature range for the Zn^{2+} -bound protein indicates that several unfolding events with different melting temperatures could be convoluted in the fluorescence data. Zn^{2+} may destabilise one portion or domain of MthComFA while the rest of the protein is not affected.

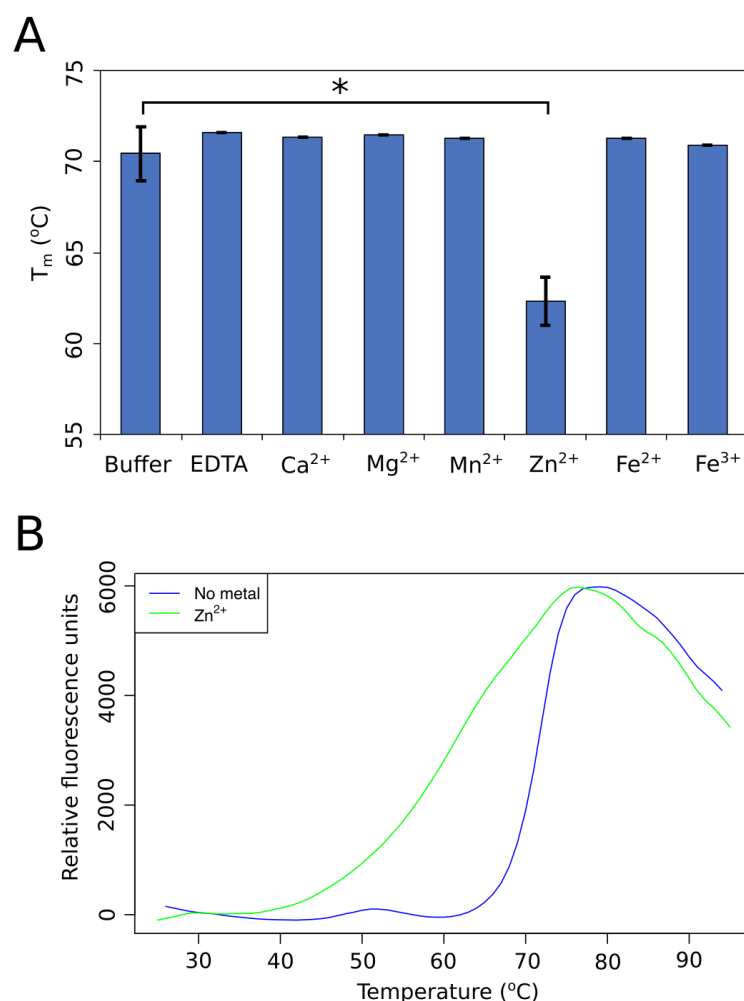


Figure 6.8. MthComFA metal binding analysis. (A) Melting temperature (T_m) of MthComFA with different additives. EDTA was added to a final concentration of 10 mM. Metal ions were added to a final concentration of 1 mM. Metal ions were added as the chloride salt or, for Zn^{2+} and Fe^{2+} , the sulphate salt. MthComFA concentration was 2.25 μM . $n = 3$ (different protein purifications) for buffer and Zn^{2+} conditions, $n = 1$ for the rest. The asterisk denotes statistical significance ($p < 0.05$, Student's t-test). (B) Representative MthComFA melting curves in the absence of added metals and in the presence of Zn^{2+} . The fluorescence values were normalised to the same scale.

6.8. MthComFA ATPase assay

MthComFA exhibited ATPase activity (Figure 6.9) as previously shown for SpnComFA (Diallo *et al*, 2017). Without added DNA MthComFA hydrolysed approximately 100 molecules of ATP per second per molecule of MthComFA (ATP/s/MthComFA) when supplied with 1.5 mM ATP (Figure 6.9). Diallo *et al* reported that SpnComFA hydrolysed

200-300 ATP/s/SpnComFA when no DNA was supplied. The rates of ATP hydrolysis by MthComFA and SpnComFA are roughly comparable, although the lower rate of hydrolysis by MthComFA may be because it is a thermophilic enzyme and the reaction was carried out at 30°C. The rate of ATP hydrolysis by SpnComFA was stimulated almost 100-fold in the presence of ssDNA but not in the presence of dsDNA (Diallo *et al*, 2017). The effects of ssDNA and dsDNA on the activity of MthComFA have not yet been tested.

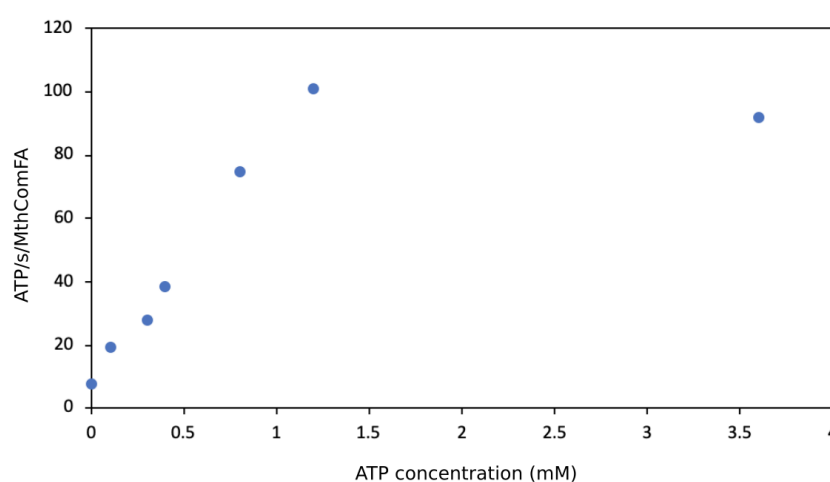


Figure 6.9. MthComFA ATPase assay. A nicotinamide adenine dinucleotide (NADH)-coupled microplate ATP hydrolysis assay (Gilbert & Mackey, 2000) was performed with 0.3 μ M MthComFA and increasing concentrations of ATP. The rate of hydrolysis (molecules of ATP hydrolysed per second per molecule of MthComFA) are plotted for different ATP concentrations. The assay was done at 30°C in buffer containing 20 mM Tris-HCl pH 8.0, 0.5 mM $MgCl_2$, 1.5 mM KCl, 0.2 mM NADH, 0.5 mM phosphoenolpyruvate, 4 U pyruvate kinase, and 12 U lactate dehydrogenase. Loss of A_{340} (due to loss of NADH) over time was measured, and the slopes of the absorbance *versus* time curves were used to determine the rate of ATP hydrolysis. The data points shown are from a test run, $n = 1$.

6.9. MthComFA limited proteolysis

The domain structure of MthComFA was investigated by limited proteolysis. When digested with trypsin, the 66.4 kDa MthComFA protein readily cleaves into two stable fragments of around 45 kDa and 16 kDa (Figure 6.10A), suggesting that there is a flexible, protease-accessible loop between these two fragments. Mass spectrometry analysis indicates that the 16 kDa band corresponds to the N-terminal extension unique to

MthComFA, and the 45 kDa band corresponds to the zinc finger and the core DEAD-box helicase fold (Figure 6.10B). Although it has been suggested that the zinc finger containing region might be disordered (Diallo *et al*, 2017), presence of peptides corresponding to that region in the 45 kDa fragment suggests that the region is resistant to trypsin digestion and likely structured. The 16 kDa and 45 kDa fragments remain strongly associated after cleavage, as shown by size exclusion chromatography (Figure 6.10C and D). This suggests that the N-terminal extension of MthComFA could be important for the function or stability of the helicase domain.

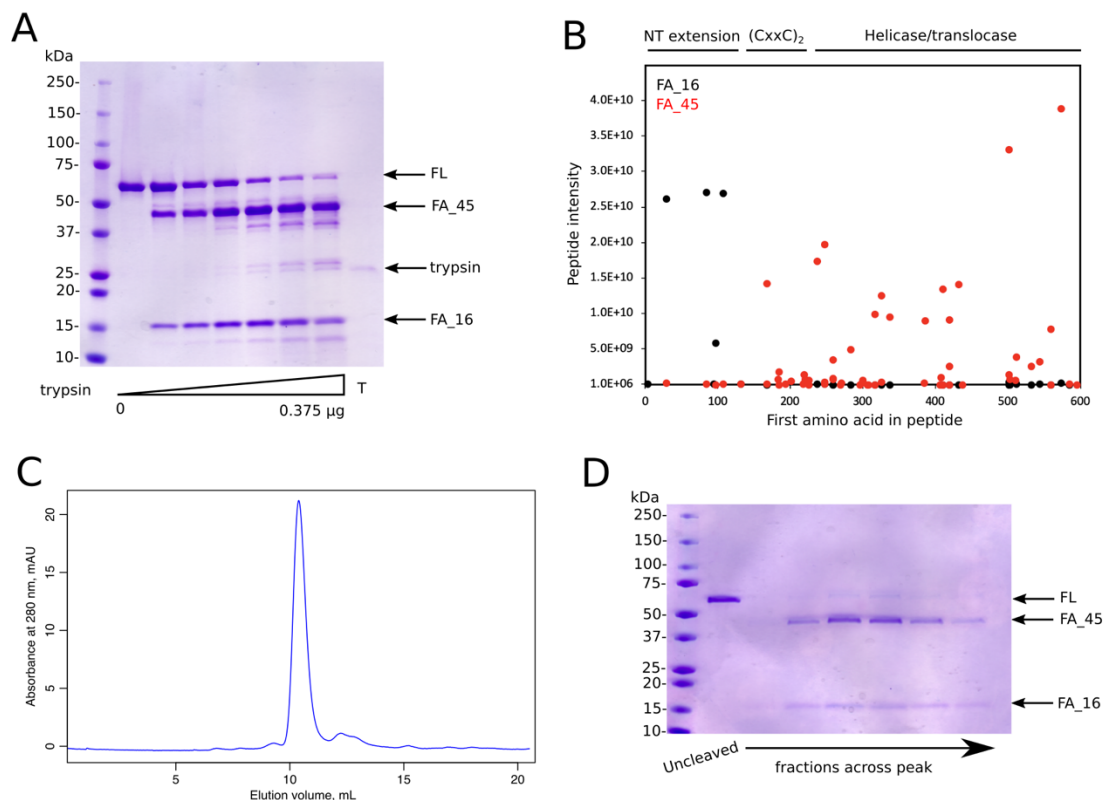


Figure 6.10. The domain structure of MthComFA analysed by limited proteolysis. (A) SDS-PAGE gel of MthComFA digested with different amounts of trypsin. Trypsin was added to 3.75 μ g of MthComFA at trypsin-to-protein mass ratios of 1:320, 1:160, 1:80, 1:40, 1:20, 1:10. The samples were incubated for 30 min at room temperature before quenching with 1 mM PMSF and 50 mM EDTA. FL, full-length MthComFA; FA_45, 45 kDa proteolytic fragment; FA_16, 16 kDa proteolytic fragment; T, trypsin only lane. (B) Peptide intensity plotted against the first amino acid residue of the peptide (numbered according to the full-length protein) from mass spectrometry analysis of the FA_16 (black) and FA_45 (red) bands. (C) Superdex 75 10/300 GL size exclusion chromatography run of trypsin-digested MthComFA. Trypsin was quenched by adding 1 mM PMSF

and 10-fold excess of soybean trypsin inhibitor (Sigma) before size exclusion chromatography. The fractions across the main peak were analysed by SDS-PAGE (D).

6.10. *MthComFA* crystallisation attempts

Crystallization of MthComFA was attempted even though the protein was not very stable. The Stura Footprint Screen (Molecular Dimensions) was used to estimate the concentration at which the crystallization trials should be set up (Stura *et al*, 1992). The ratio of clear drops to drops containing precipitate of 1:1 in the Stura screen indicated that trials should be set up at a concentration of around 3 mg/ml. Sitting drop vapour diffusion crystallization trials were then set up with fresh batches of protein at 2.9-3.2 mg/ml using the commercial screens PACT, PGA, Structure 1+2, MIDAS+ and JCSG+ (Molecular Dimensions). Drops were set up at 1:1 and 1:2 ratios of protein sample to reservoir solution. In addition, screens with protein samples doped with 2 mM AMPPNP, a non-hydrolysable ATP analogue, were also set up as the nucleotide could stabilise the protein structure. Because the limited proteolysis assay indicated that there may be a flexible loop between two domains of MthComFA (Figure 6.10), screens were set up with a 1:100 mass ratio of trypsin to MthComFA. Unfortunately, no hits were obtained in any of the aforementioned conditions over the course of several months.

6.11. *Production of recombinant MthComFC*

MthComFC was cloned into the pQE80-L vector and expressed in *E. coli* Rosetta 2 (DE3) cells. Pure protein eluted from the NiNTA column in a single peak (Figure 6.11). Around 300 mM imidazole was required to elute the protein. This is unusual for His₆-tagged proteins, which normally elute at 100-150 mM imidazole concentrations. It is possible that the protein was forming either oligomers or aggregates which could

enhance binding to the column resin via avidity effects. Considering that the protein completely precipitated immediately after NiNTA elution, aggregation is the more likely explanation for the strong binding to the column. After removing the precipitated protein, the sample was concentrated and subjected to size exclusion chromatography, but no eluted protein was observed (not shown). Even if the eluate from the initial NiNTA affinity column was collected into tubes containing imidazole-free buffer precipitation still occurred. Because the MthComFC protein was unstable, no further studies were carried out on this construct.

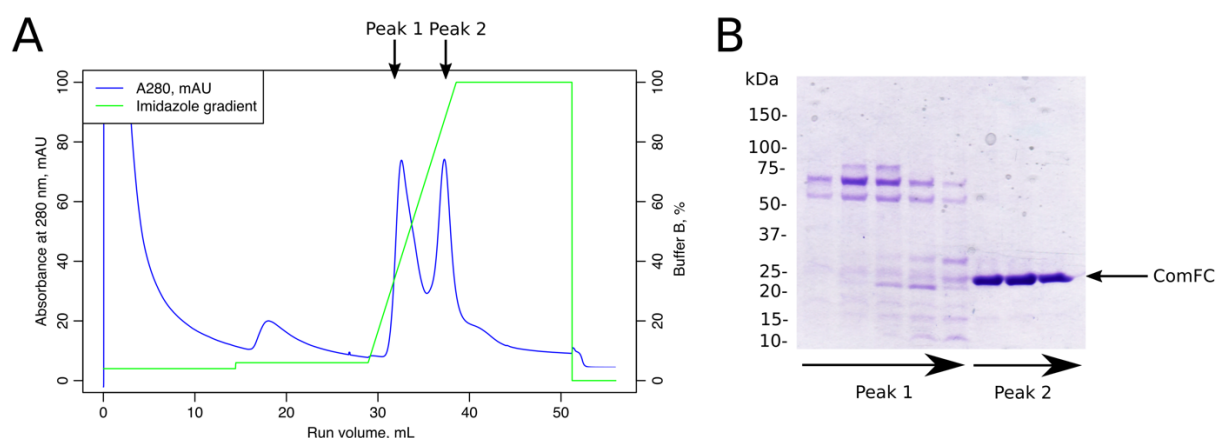


Figure 6.11. MthComFC NiNTA purification. (A) Imidazole gradient elution (30 mM-500 mM) from a NiNTA column. (B) SDS-PAGE analysis of fractions across peaks 1 and 2. The identity of ComFC was confirmed by Western blotting with an anti-His antibody (not shown).

6.12. Co-expression of MthComFA and MthComFC

MthComFA and MthComFC were co-expressed in an attempt to obtain a complex suitable for structural studies by EM. Both proteins were expressed from a pCDFDuet-1 plasmid in *E. coli* Rosetta 2 (DE3) cells. The MthComFA construct had an N-terminal MGSSHHHHHHS tag, and the MthComFC construct had an N-terminal TwinStrep tag. Cells expressing both proteins were lysed, and the soluble fraction was loaded on a 1 ml

StrepTrap HP column. The column was washed and bound proteins were eluted with desthiobiotin-containing buffer. Samples from different purification stages were analysed by Western blotting (Figure 6.12). The StrepTactin-HRP blot showed that most of MthComFC was insoluble but some was also present in the elution fraction. The anti-His blot, however, indicated that all of MthComFA was in the pellet fraction. Attempts at optimising the expression conditions were unsuccessful – three different strains of *E. coli* (BL21 (DE3) pLysS, BL21 Star (DE3), Rosetta 2 (DE3)) were tested, expressing at either 30°C for 3 h or at 18°C for 16 h, but no soluble MthComFA was detected (not shown). It is unclear why the MthComFA construct expresses well by itself but not in the presence of MthComFC.

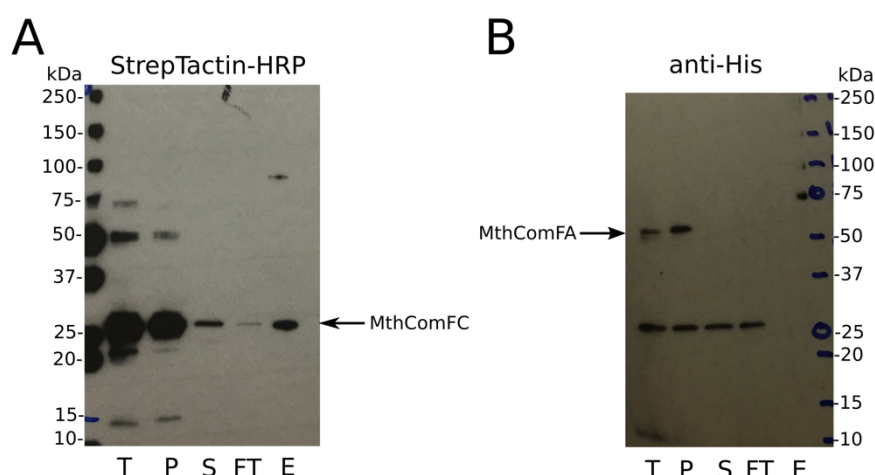


Figure 6.12. MthComFA and MthComFC co-expression. StrepTactin-HRP (A) and anti-His (B) blots of different StrepTrap purification stages. T, total lysate; P, insoluble fraction; S, soluble fraction; FT, column flow-through; E, elution. The MthComFA and MthComFC bands are indicated by arrows; the other bands correspond to either cross-reacting cellular proteins or degradation products.

6.13. Co-expression of MthComFA and MthComEC

To test if MthComFA and MthComEC form a complex, the two proteins were co-expressed. Although the replication origins of the pQE80-L-MthComFA and pWsf-

MthComEC-TS plasmids are incompatible, both proteins were expressed when the plasmids were co-transformed into *E. coli* Rosetta 2 (DE3) cells. The cells were lysed, the lysate was clarified in a low-speed centrifugation step, and the lysate supernatant was subjected to ultracentrifugation. The ultracentrifugation pellet contained the membrane fraction. MthComEC-sfGFP could be detected in the lysate supernatant and membrane fractions (Figure 6.13A). Interestingly, although MthComFA can be purified without detergents (section 6.3) (Chilton *et al*, 2017; Diallo *et al*, 2017), a significant portion of the protein was also found in the membrane fraction (Figure 6.13B). This is consistent with other studies which reported that *B. subtilis* ComFA is membrane-associated. (Londoño-Vallejo & Dubnau, 1994a, 1994b; Kramer *et al*, 2007). It is tempting to speculate that treating MthComFA as a membrane protein and having detergents in the purification buffers might make it more stable and amenable to structural analysis.

The membrane fraction of the cells co-expressing MthComEC-sfGFP and MthComFA was solubilised with LMNG. After clearing the insoluble material, the soluble fraction was split in two and subjected to NiNTA or StrepTrap affinity column purification (Figure 6.13C and D). Both MthComEC-sfGFP and MthComFA were present in the solubilised membrane fraction. However, MthComEC-sfGFP did not co-purify with MthComFA and vice versa. MthComEC-sfGFP was found only in the StrepTrap eluate and MthComFA was found only in the NiNTA eluate. This experiment showed that the proteins do not form a stable complex in a heterologous expression system, at least in the presence of detergent and in the absence of actively transported DNA.

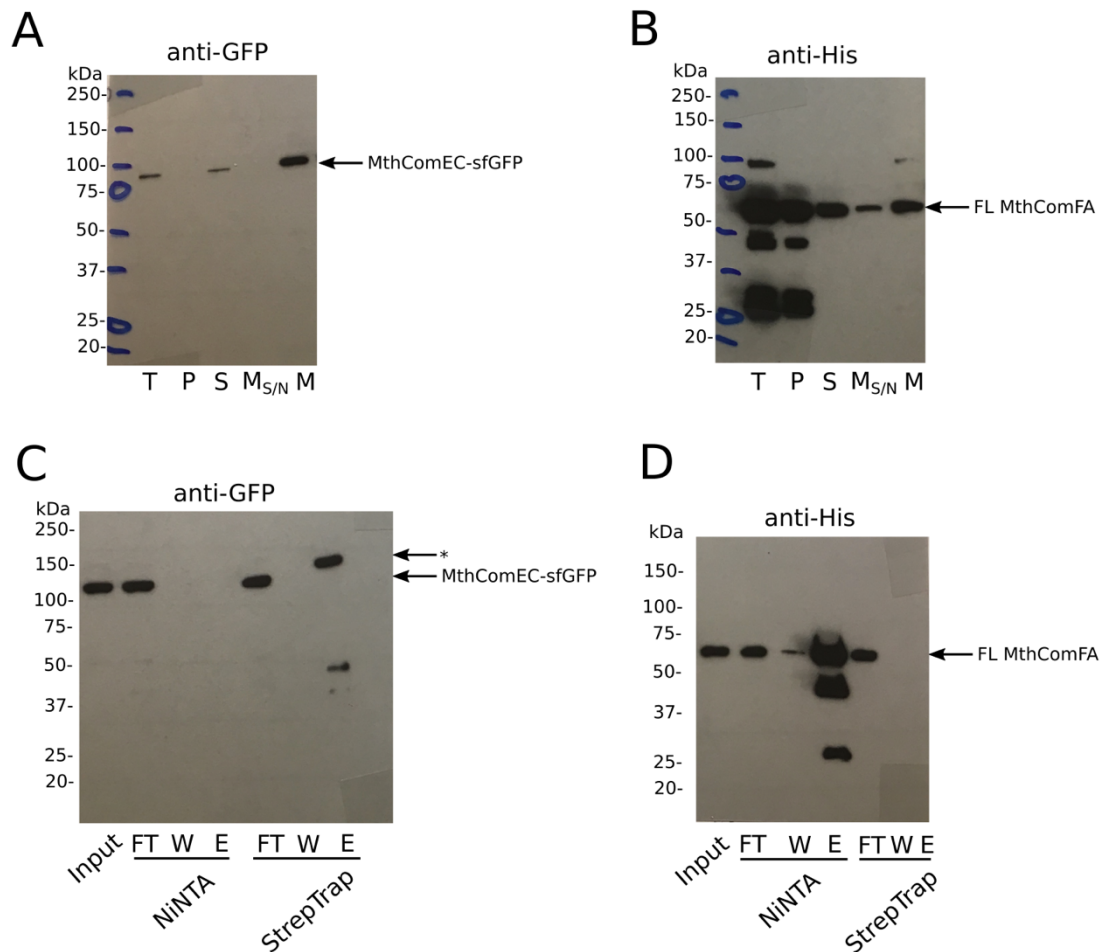


Figure 6.13. MthComEC and MthComFA co-expression. MthComEC-sfGFP-TwinStrep and MthComFA-His were co-expressed in *E. coli*. Membranes were isolated, solubilised with LMNG and purified either using a NiNTA or a StrepTactin column. Anti-GFP (A, C) and anti-His (B, D) Western blots of samples from different purification stages are shown. T, total lysate; P, low speed pellet; S, low speed soluble fraction; M_{S/N}, supernatant from high speed spin; M, pellet from high speed spin containing the membranes; Input, soluble fraction of LMNG-solubilised membranes; FT, flow-through; W, wash; E, eluate. Notably, the membrane fraction was 2.5x more concentrated than the T, P, S and M_{S/N} fractions. The band marked with an asterisk is MthComEC-sfGFP but appears to run higher than expected because the SDS-PAGE gel was broken and not pieced together correctly during transfer.

6.13. Summary

Bioinformatics analysis of MthComFA revealed that it has the same zinc-binding and DEAD-box helicase/translocase motifs as previously investigated homologues. In addition, MthComFA has a unique N-terminal extension. MthComFC sequence analysis revealed that its (CxxC)₄ motifs and PRPP binding motif are conserved in both Gram-positive and Gram-negative species.

MthComFA was expressed and purified from *E. coli*, likely in dimeric form. The protein was destabilised by zinc, bound ssDNA and dsDNA and exhibited ATPase activity. The N-terminal extension of MthComFA was resistant to trypsin digestion in limited proteolysis experiments and remained tightly associated with the rest of the protein after cleavage. The putatively disordered zinc-binding region was also resistant to proteolysis, suggesting that it is structured. Crystallisation attempts were unsuccessful, likely due to sample instability.

DNA binding by SpnComFA was studied previously, and SpnComFA was not reported to bind dsDNA. Here we have shown that MthComFA binds both ssDNA and dsDNA. It is unclear what is the significance of this. It is possible that the ability to bind both ssDNA and dsDNA reflects the binding mode of MthComFA, for example, if MthComFA binds DNA just via the backbone and is insensitive to its geometry. It would be interesting to test if and how ATP binding affects the affinity of MthComFA for ssDNA and dsDNA.

The function of the N-terminal extension of MthComFA is unknown. However, tight association between the extension and the core DEAD-box helicase domains suggests that the extension is somehow important for the function of MthComFA.

MthComFC could be expressed but immediately precipitated after purification. MthComFA and MthComFC did not co-express. MthComFA and MthComEC co-expressed but did not interact under the conditions tested. It is possible that the correct conditions for optimal expression induction and protein folding were not identified. Different expression plasmid and construct combinations may yield better results.

Although standard purification protocols for isolating SpnComFA/ComFC and BsuComFA were successful in previous studies, treating these proteins as membrane-bound species may give better results. These proteins have been shown to be membrane associated, and it is likely that the lack of a local hydrophobic environment results in structural instability.

Chapter 7: Conclusions and future directions

7.1. Towards the structure of ComEC

The aim of this project was to gain a better understanding of DNA uptake across the cytoplasmic membrane during natural transformation. ComEC is the central protein involved in this process. ComEC forms the aqueous pore that allows passage of DNA into the cytoplasm. No structural and very limited biochemical data are available on ComEC. The precise topology and functional oligomeric state of ComEC is unknown. The roles of its extramembranous domains in DNA transport are unclear. Many of these questions could be answered by obtaining a structure of ComEC.

In chapter 3 we performed an extensive expression screen of ComEC homologues in *E. coli* hoping that a homologue suitable for structural and biochemical analysis would be identified. ComEC from the thermophilic Gram-positive bacterium *Moorella thermoacetica* was the only homologue that we could purify. After substantial efforts to develop a purification protocol, pure ComEC was obtained. Biochemical characterisation by limited proteolysis identified a trypsin susceptible region in the N-terminal extramembranous domain of ComEC. SEC-MALLS analysis revealed that heterologously produced *M. thermoacetica* ComEC is a monomer. Negative stain EM analysis revealed that its two soluble extramembranous domains are located on the same side of the membrane and likely interact. High resolution structural analysis of ComEC was hindered by its small size and low expression level.

ComEC expression levels were too low to attempt crystallisation. It is possible that by screening even more ComEC homologues from diverse organisms eventually one that is expressed at a level suitable for crystallisation will be identified. An alternative approach,

such as cell-free expression (Schwarz *et al*, 2007) could be taken to try to improve ComEC production.

Technological solutions in cryo-EM could allow determination of high resolution structures of proteins as small as ComEC. Recently, a novel laser phase plate for TEM has been reported (Schwartz *et al*, 2019). Once fully developed, the laser phase plate will improve the contrast of cryo-EM micrographs and allow high resolution cryo-EM studies of small proteins.

7.2. Characterisation of DNA uptake complexes

In chapter 4 we attempted to obtain a DNA uptake complex from *B. subtilis* that would be suitable for structural studies by cryo-EM. ComEC, ComFA and ComFC were tagged with affinity tags on the *B. subtilis* chromosome. A supercompetence induction system was used to increase the expression level of the DNA transport machinery. No DNA uptake machinery complexes were isolated. The tagged ComEC and ComFC proteins co-purified with stress associated *B. subtilis* proteins. The purified samples were very heterogeneous, as assessed by size exclusion chromatography and negative stain electron microscopy.

Although isolation of DNA uptake complexes has proven challenging, alternative approaches can be used to study the composition of DNA uptake complexes. Different components of the DNA transport machinery could be fused to the APEX peroxidase (Ganapathy *et al*, 2018). In the presence of hydrogen peroxide and biotin-phenol, APEX generates phenoxyl radicals that covalently bind to nearby proteins. The resulting biotinylated proteins can be purified and analysed by mass spectrometry. In this way all, low-resolution structural data about the composition of DNA uptake complexes can be obtained.

7.3. The function of the C-terminal domain of ComEC.

The CTD of ComEC has been proposed to degrade the non-transforming DNA strand during DNA transport in bacteria that do not have the dedicated endonuclease EndA. In chapter 5 we experimentally demonstrated the nuclease activity of the CTD of *M. thermoacetica* ComEC. The CTD was expressed and purified from *E. coli*, and thermal stability assays were used to determine that it binds Mn^{2+} . Activity assays with a phosphodiesterase analogue substrate and with plasmid DNA showed conclusively that the CTD is a metal-dependent nuclease. A metal binding-deficient variant of the CTD was shown to lack nuclease activity. When the equivalent metal binding mutations were introduced to the *comEC* gene in *B. subtilis* 168+, the mutant strain showed a 750-fold decrease in transformation efficiency.

Our combined *in vitro* and *in vivo* approach to showing that the CTD nuclease function is important in natural transformation can be applied to studying other MBL motifs of the CTD. An interesting experiment to do would be to exchange motif II of MthCTD for that of SpnCTD, which should be catalytically inactive. Furthermore, it would be interesting to see if a nuclease-defective *comEC* mutant in *B. subtilis* could be complemented in *trans* with an active recombinantly expressed CTD.

Bibliography

- Albano M, Hahn J & Dubnau D (1987) Expression of competence genes in *Bacillus subtilis*. *J. Bacteriol.* **169**: 3110–3117
- Arakawa T, Ejima D, Li T & Philo JS (2010) The critical role of mobile phase composition in size exclusion chromatography of protein pharmaceuticals. *J. Pharm. Sci.* **99**: 1674–1692
- Arnaud M, Chastanet A & Débarbouillé M (2004) New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. *Appl. Environ. Microbiol.* **70**: 6887–91
- Auclair SM, Bhanu MK & Kendall DA (2012) Signal peptidase I: cleaving the way to mature proteins. *Protein Sci.* **21**: 13–25
- Bai X chen, McMullan G & Scheres SH. (2015) How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40**: 49–57
- Baker JA, Simkovic F, Taylor HMC & Rigden DJ (2016) Potential DNA binding and nuclease functions of ComEC domains characterized *in silico*. *Proteins Struct. Funct. Bioinforma.* **84**: 1431–1442
- Barany F & Boeke JD (1983) Genetic transformation of *Streptococcus pneumoniae* by DNA cloned into the single-stranded bacteriophage f1. *J. Bacteriol.* **153**: 200–10

- Baumgarten T, Schlegel S, Wagner S, Löw M, Eriksson J, Bonde I, Herrgård MJ, Heipieper HJ, Nørholm MHH, Slotboom DJ & de Gier J-W (2017) Isolation and characterization of the *E. coli* membrane protein production strain Mutant56(DE3). *Sci. Rep.* **7**: 45089
- Bennallack PR, Burt SR, Heder MJ, Robison RA & Griffiths JS (2014) Characterization of a novel plasmid-borne thiopeptide gene cluster in *Staphylococcus epidermidis* strain 115. *J. Bacteriol.* **196**: 4344–50
- Berg B van den, Clemons WM, Collinson I, Modis Y, Hartmann E, Harrison SC & Rapoport TA (2004) X-ray structure of a protein-conducting channel. *Nature* **427**: 36–44
- Berge M, Moscoso M, Prudhomme M, Martin B & Claverys J-P (2002) Uptake of transforming DNA in Gram-positive bacteria: a view from *Streptococcus pneumoniae*. *Mol. Microbiol.* **45**: 411–421
- Bergé MJ, Kamgoué A, Martin B, Polard P, Campo N & Claverys J-P (2013) Midcell Recruitment of the DNA Uptake and Virulence Nuclease, EndA, for Pneumococcal Transformation. *PLoS Pathog.* **9**: e1003596
- Berka RM, Hahn J, Albano M, Draskovic I, Persuh M, Cui X, Sloma A, Widner W & Dubnau D (2002) Microarray analysis of the *Bacillus subtilis* K-state: Genome-wide expression changes dependent on ComK. *Mol. Microbiol.* **43**: 1331–1345
- Berry J-L, Xu Y, Ward PN, Lea SM, Matthews SJ & Pelicic V (2016) A Comparative Structure/Function Analysis of Two Type IV Pilin DNA Receptors Defines a Novel Mode of DNA Binding. *Structure* **24**: 926–934

- Berry JL, Cehovin A, McDowell MA, Lea SM & Pelicic V (2013) Functional Analysis of the Interdependence between DNA Uptake Sequence and Its Cognate ComP Receptor during Natural Transformation in *Neisseria* Species. *PLoS Genet.* **9**: e1004014
- Blokesch M (2017) In and out—contribution of natural transformation to the shuffling of large genomic regions. *Curr. Opin. Microbiol.* **38**: 22–29
- Borgeaud S, Metzger LC, Scrignari T & Blokesch M (2015) The type VI secretion system of *Vibrio cholerae* fosters horizontal gene transfer. *Science (80-.).* **347**: 63–67
- Briley K, Dorsey-Oresto A, Prepiak P, Dias MJ, Mann JM, Dubnau D & Dubnau D (2011) The secretion ATPase ComGA is required for the binding and transport of transforming DNA. *Mol. Microbiol.* **81**: 818–30
- Brimacombe CA, Ding H, Johnson JA & Beatty JT (2015) Homologues of Genetic Transformation DNA Import Genes Are Required for *Rhodobacter capsulatus* Gene Transfer Agent Recipient Capability Regulated by the Response Regulator CtrA. *J. Bacteriol.* **197**: 2653–63
- Callebaut I, Moshous D, Mornon J-P & de Villartay J-P (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res.* **30**: 3592–3601
- Carfi A, Pares S, Duée E, Galleni M, Duez C, Frère JM & Dideberg O (1995) The 3-D structure of a zinc metallo-beta-lactamase from *Bacillus cereus* reveals a new type of protein fold. *EMBO J.* **14**: 4914–4921

- De Carlo S & Harris JR (2011) Negative staining and cryo-negative staining of macromolecules and viruses for TEM. *Micron* **42**: 117–31
- Carrondo MA, Frazão C, Silva G, Gomes CM, Matias P, Coelho R, Sieker L, Macedo S, Liu MY, Oliveira S, Teixeira M, Xavier A V., Rodrigues-Pousada C & Le Gall J (2000) Structure of a dioxygen reduction enzyme from *Desulfovibrio gigas*. *Nat. Struct. Biol.* **7**: 1041–1045
- Casadaban MJ & Cohen SN (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J. Mol. Biol.* **138**: 179–207
- Cehovin A, Simpson PJ, McDowell MA, Brown DR, Noschese R, Pallett M, Brady J, Baldwin GS, Lea SM, Matthews SJ & Pelicic V (2013) Specific DNA recognition mediated by a type IV pilin. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 3065–3070
- Chang Y-W, Rettberg LA, Treuner-Lange A, Iwasa J, Søgaaard-Andersen L & Jensen GJ (2016) Architecture of the type IVa pilus machine. *Science* **351**: aad2001
- Chen I, Provvedi R & Dubnau D (2006) A macromolecular complex formed by a pilin-like protein in competent *Bacillus subtilis*. *J. Biol. Chem.* **281**: 21720–21727
- Chen S, Yakunin AF, Kuznetsova E, Busso D, Pufan R, Proudfoot M, Kim R & Kim S-H (2004) Structural and functional characterization of a novel phosphodiesterase from *Methanococcus jannaschii*. *J. Biol. Chem.* **279**: 31854–62
- Cheng Y, Grigorieff N, Penczek PA & Walz T (2015) A primer to single-particle cryo-electron microscopy. *Cell* **161**: 438–449

- Chi F, Nolte O, Bergmann C, Ip M & Hakenbeck R (2007) Crossing the barrier: Evolution and spread of a major class of mosaic *pbp2x* in *Streptococcus pneumoniae*, *S. mitis* and *S. oralis*. *Int. J. Med. Microbiol.* **297**: 503–512
- Chiang YN, Penadés JR & Chen J (2019) Genetic transduction by phages and chromosomal islands: The new and noncanonical. *PLOS Pathog.* **15**: e1007878
- Chilton SS, Falbel TG, Hromada S & Burton BM (2017) A conserved metal binding motif in the *Bacillus subtilis* competence protein ComFA enhances transformation. *J. Bacteriol.* **199**: e00272-17
- Chung YS & Dubnau D (1995) ComC is required for the processing and translocation of ComGC, a pilin-like competence protein of *Bacillus subtilis*. *Mol. Microbiol.* **15**: 543–551
- Chung YS & Dubnau D (1998) All seven *comG* open reading frames are required for DNA binding during transformation of competent *Bacillus subtilis*. *J. Bacteriol.* **180**: 41–5
- Claverys J-P & Håvarstein LS (2007) Cannibalism and fratricide: mechanisms and *raisons d'être*. *Nat. Rev. Microbiol.* **5**: 219–229
- Corbinais C, Mathieu A, Kortulewski T, Radicella JP & Marsin S (2016) Following transforming DNA in *Helicobacter pylori* from uptake to expression. *Mol. Microbiol.* **101**: 1039–1053
- Costa TRD, Ilangoan A, Ukleja M, Redzej A, Santini JM, Smith TK, Egelman EH & Waksman G (2016) Structure of the Bacterial Sex F Pilus Reveals an Assembly of a Stoichiometric Protein-Phospholipid Complex. *Cell* **166**: 1436-1444.e10

- Croucher NJ, Mostowy R, Wymant C, Turner P, Bentley SD & Fraser C (2016) Horizontal DNA Transfer Mechanisms of Bacteria as Weapons of Intragenomic Conflict. *PLoS Biol.* **14**: e1002394
- Daiyasu H, Osaka K, Ishino Y & Toh H (2001) Expansion of the zinc metallo-hydrolase family of the β -lactamase fold. *FEBS Lett.* **503**: 1–6
- Danner DB, Deich RA, Sisco KL & Smith HO (1980) An eleven-base-pair sequence determines the specificity of DNA uptake in *Haemophilus transformation*. *Gene* **11**: 311–318
- Decatur A & Losick R (1996) Identification of additional genes under the control of the transcription factor sigma F of *Bacillus subtilis*. *J. Bacteriol.* **178**: 5039–41
- Diallo A, Foster HR, Gromek KA, Perry TN, Dujeancourt A, Krasteva P V., Gubellini F, Falbel TG, Burton BM & Fronzes R (2017) Bacterial transformation: ComFA is a DNA-dependent ATPase that forms complexes with ComFC and DprA. *Mol. Microbiol.* **105**: 741–754
- Dinh T & Bernhardt TG (2011) Using superfolder green fluorescent protein for periplasmic protein localization studies. *J. Bacteriol.* **193**: 4984–7
- Domingues S, Harms K, Fricke WF, Johnsen PJ, da Silva GJ & Nielsen KM (2012) Natural Transformation Facilitates Transfer of Transposons, Integrations and Gene Cassettes between Bacterial Species. *PLoS Pathog.* **8**: e1002837

- Dominski Z (2007) Nucleases of the metallo- β -lactamase family and their role in DNA and RNA metabolism. *Crit. Rev. Biochem. Mol. Biol.* **42**: 67–93
- Draskovic I & Dubnau D (2005) Biogenesis of a putative channel protein, ComEC, required for DNA uptake: Membrane topology, oligomerization and formation of disulphide bonds. *Mol. Microbiol.* **55**: 881–896
- Drew D, Lerch M, Kunji E, Slotboom D-J & de Gier J-W (2006) Optimization of membrane protein overexpression and purification using GFP fusions. *Nat. Methods* **3**: 303–313
- Dubnau D (1991) The regulation of genetic competence in *Bacillus subtilis*. *Mol. Microbiol.* **5**: 11–18
- Dubnau D & Blokesch M (2019) Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annu. Rev. Genet.* **53**: annurev-genet-112618-043641
- Duitman EH, Wyczawski D, Boven LG, Venema G, Kuipers OP & Hamoen LW (2007) Novel methods for genetic transformation of natural *Bacillus subtilis* isolates used to study the regulation of the mycosubtilin and surfactin synthetases. *Appl. Environ. Microbiol.* **73**: 3490–6
- Durand E, Cambillau C, Cascales E & Journet L (2014) VgrG, Tae, Tle, and beyond: the versatile arsenal of Type VI secretion effectors. *Trends Microbiol.* **22**: 498–507
- Eiamphungporn W & Helmann JD (2008) The *Bacillus subtilis* σ^M regulon and its contribution to cell envelope stress responses. *Mol. Microbiol.* **67**: 830–848

- El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE & Finn RD (2019) The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**: D427–D432
- Ellison CK, Dalia TN, Vidal Ceballos A, Wang JC-Y, Biais N, Brun Y V. & Dalia AB (2018) Retraction of DNA-bound type IV competence pili initiates DNA uptake during natural transformation in *Vibrio cholerae*. *Nat. Microbiol.* **3**: 773–780
- Gamba P, Jonker MJ & Hamoen LW (2015) A Novel Feedback Loop That Controls Bimodal Expression of Genetic Competence. *PLOS Genet.* **11**: e1005047
- Ganapathy US, Bai L, Wei L, Eckart KA, Lett CM, Previti ML, Carrico IS & Seeliger JC (2018) Compartment-Specific Labeling of Bacterial Periplasmic Proteins by Peroxidase-Mediated Biotinylation. *ACS Infect. Dis.* **4**: 918–925
- Gangel H, Hepp C, Müller S, Oldewurtel ER, Aas FE, Koomey M & Maier B (2014) Concerted Spatio-Temporal Dynamics of Imported DNA and ComE DNA Uptake Protein during Gonococcal Transformation. *PLoS Pathog.* **10**: e1004043
- Gilbert SP & Mackey AT (2000) Kinetics: A Tool to Study Molecular Motors. *Methods* **22**: 337–354
- Glenwright AJ, Pothula KR, Bhamidimarri SP, Chorev DS, Baslé A, Firbank SJ, Zheng H, Robinson C V., Winterhalter M, Kleinekathöfer U, Bolam DN & van den Berg B (2017) Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**: 407–411

- Goodman SD & Scoocca JJ (1988) Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. *Proc. Natl. Acad. Sci. U. S. A.* **85**: 6982–6986
- Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**: 2876–90
- Griffith F (1928) The Significance of Pneumococcal Types. *J. Hyg. (Lond).* **27**: 113–59
- Hahn J, Albano M & Dubnau D (1987) Isolation and characterization of *Tn917lac*-generated competence mutants of *Bacillus subtilis*. *J. Bacteriol.* **169**: 3104–3109
- Hahn J, Luttinger A & Dubnau D (1996) Regulatory inputs for the synthesis of ComK, the competence transcription factor of *Bacillus subtilis*. *Mol. Microbiol.* **21**: 763–775
- Hahn J, Maier B, Haijema BJ, Sheetz M & Dubnau D (2005) Transformation proteins and DNA uptake localize to the cell poles in *Bacillus subtilis*. *Cell* **122**: 59–71
- Hamoen LW, Venema G & Kuipers OP (2003) Controlling competence in *Bacillus subtilis*: shared use of regulators. *Microbiology* **149**: 9–17
- Hauer F, Gerle C, Fischer N, Oshima A, Shinzawa-Itoh K, Shimada S, Yokoyama K, Fujiyoshi Y, Stark H, Scheres SHW & al. et (2015) GraDeR: Membrane Protein Complex Preparation for Single-Particle Cryo-EM. *Structure* **23**: 1769–1775
- Hellman LM & Fried MG (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2**: 1849–61

- Hepp C, Gangel H, Henseler K, Günther N & Maier B (2016) Single-Stranded DNA Uptake during Gonococcal Transformation. *J. Bacteriol.* **198**: 2515–23
- Hepp C & Maier B (2016) Kinetics of DNA uptake during transformation provide evidence for a translocation ratchet mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **113**: 12467–12472
- Hofreuter D, Odenbreit S & Haas R (2001) Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Mol. Microbiol.* **41**: 379–391
- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hermsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC & Banfield JF (2016) A new view of the tree of life. *Nat. Microbiol.* **1**: 16048
- Ilangovan A, Connery S & Waksman G (2015) Structural biology of the Gram-negative bacterial conjugation systems. *Trends Microbiol.* **23**: 301–10
- Inamine GS & Dubnau D (1995) ComEA, a *Bacillus subtilis* integral membrane protein required for genetic transformation, is needed for both DNA binding and transport. *J. Bacteriol.* **177**: 3045–3051
- Jezewska MJ & Bujalowski W (2000) Interactions of *Escherichia coli* Replicative Helicase PriA Protein with Single-Stranded DNA [†]. *Biochemistry* **39**: 10454–10467

- Johnsborg O, Eldholm V, Bjørnstad ML & Håvarstein LS (2008) A predatory mechanism dramatically increases the efficiency of lateral gene transfer in *Streptococcus pneumoniae* and related commensal species. *Mol. Microbiol.* **69**: 245–253
- Johnson CM (2013) Differential scanning calorimetry as a tool for protein folding and stability. *Arch. Biochem. Biophys.* **531**: 100–109
- Johnston C, Martin B, Fichant G, Polard P & Claverys J-P (2014) Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat. Publ. Gr.* **12**:
- Kastner B, Fischer N, Golas MM, Sander B, Dube P, Boehringer D, Hartmuth K, Deckert J, Hauer F, Wolf E, Uchtenhagen H, Urlaub H, Herzog F, Peters JM, Poerschke D, Lührmann R & Stark H (2007) GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**: 53–55
- Kaufenstein M, Van Der Laan M & Graumann PL (2011) The three-layered DNA uptake machinery at the cell pole in competent *Bacillus subtilis* cells is a stable complex. *J. Bacteriol.* **193**: 1633–1642
- Kawate T & Gouaux E (2006) Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**: 673–81
- Kelley LA, Mezulis S, Yates CM, Wass MN & Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**: 845–858
- Kelly LA, Mezulis S, Yates C, Wass M & Sternberg M (2015) The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat. Protoc.* **10**: 845–858

- Kelly SM, Jess TJ & Price NC (2005) How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* **1751**: 119–139
- Kidane D & Graumann PL (2005) Intracellular protein and DNA dynamics in competent *Bacillus subtilis* cells. *Cell* **122**: 73–84
- Kline KA & Seifert HS (2005) Mutation of the *priA* gene of *Neisseria gonorrhoeae* affects DNA transformation and DNA repair. *J. Bacteriol.* **187**: 5347–55
- Kramer N, Hahn J & Dubnau D (2007) Multiple interactions among the competence proteins of *Bacillus subtilis*. *Mol. Microbiol.* **65**: 454–464
- de la Sierra-Gallay IL, Pellegrini O & Condon C (2005) Structural basis for substrate binding, cleavage and allostery in the tRNA maturase RNase Z. *Nature* **433**: 657–661
- LaBrie SD, Dimond ZE, Harrison KS, Baid S, Wickstrum J, Suchland RJ & Hefty PS (2019) Transposon Mutagenesis in *Chlamydia trachomatis* Identifies CT339 as a ComEC Homolog Important for DNA Uptake and Lateral Gene Transfer. *MBio* **10**: e01343-19
- Larson TG & Goodgal SH (1992) Donor DNA processing is blocked by a mutation in the *com101A* locus of *Haemophilus influenzae*. *J. Bacteriol.* **174**: 3392–3394
- Lauber F, Deme JC, Lea SM & Berks BC (2018) Type 9 secretion system structures reveal a new protein transport mechanism. *Nature* **564**: 77–82

- Laurenceau R, Péhau-Arnaudet G, Baconnais S, Gault J, Malosse C, Dujeancourt A, Campo N, Chamot-Rooke J, Le Cam E, Claverys J-P & Fronzes R (2013) A Type IV Pilus Mediates DNA Binding during Natural Transformation in *Streptococcus pneumoniae*. *PLoS Pathog.* **9**: e1003473
- Lee P-H, Huang XX, Teh BT & Ng L-M (2019) TSA-CRAFT: A Free Software for Automatic and Robust Thermal Shift Assay Data Analysis. *SLAS Discov. Adv. Life Sci. R&D* **24**: 606–612
- Leisner M, Stingl K, Frey E & Maier B (2008) Stochastic switching to competence. *Curr. Opin. Microbiol.* **11**: 553–559
- Li W, Zhou X & Lu P (2004) Bottlenecks in the expression and secretion of heterologous proteins in *Bacillus subtilis*. *Res. Microbiol.* **155**: 605–610
- Linder P & Jankowsky E (2011) From unwinding to clamping — the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* **12**: 505–516
- Liu L, Nakano MM, Lee OH & Zuber P (1996) Plasmid-amplified *comS* enhances genetic competence and suppresses *sinR* in *Bacillus subtilis*. *J. Bacteriol.* **178**: 5144–52
- Llarrull LI, Fabiane SM, Kowalski JM, Bennett B, Sutton BJ & Vila AJ (2007) Asp-120 locates Zn² for optimal metallo-beta-lactamase activity. *J. Biol. Chem.* **282**: 18276–85
- Londono-Vallejo JA & Dubnau D (1993) *comF*, a *Bacillus subtilis* late competence locus, encodes a protein similar to ATP-dependent RNA/DNA helicases

- Londoño-Vallejo JA & Dubnau D (1994a) Membrane association and role in DNA uptake of the *Bacillus subtilis* PriA anaigoue ComF1. *Mol. Microbiol.* **13**: 197–205
- Londoño-Vallejo JA & Dubnau D (1994b) Mutation of the putative nucleotide binding site of the *Bacillus subtilis* membrane protein ComFA abolishes the uptake of DNA during transformation. *J. Bacteriol.* **176**: 4642–4645
- Louis-Jeune C, Andrade-Navarro MA & Perez-Iratxeta C (2012) Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins Struct. Funct. Bioinforma.* **80**: 374–381
- Maamar H & Dubnau D (2005) Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. *Mol. Microbiol.* **56**: 615–624
- Maier B, Chen I, Dubnau D & Sheetz MP (2004) DNA transport into *Bacillus subtilis* requires proton motive force to generate large molecular forces. *Nat. Struct. Mol. Biol.* **11**: 643–649
- Matias VRF & Beveridge TJ (2005) Cryo-electron microscopy reveals native polymeric cell wall structure in *Bacillus subtilis* 168 and the existence of a periplasmic space. *Mol. Microbiol.* **56**: 240–251
- Matthey N & Blokesch M (2016) The DNA-Uptake Process of Naturally Competent *Vibrio cholerae*. *Trends Microbiol.* **24**: 98–110
- Meier P, Berndt C, Weger N & Wackernagel W (2002) Natural transformation of *Pseudomonas stutzeri* by single-stranded DNA requires type IV pili, competence state and *comA*. *FEMS Microbiol. Lett.* **207**: 75–80

- Mejean V & Claverys JP (1993) DNA processing during entry in transformation of *Streptococcus pneumoniae*. *J. Biol. Chem.* **268**: 5594–5599
- Mio K & Sato C (2018) Lipid environment of membrane proteins in cryo-EM based structural analysis. *Biophys. Rev.* **10**: 307–316
- Miroux B & Walker JE (1996) Over-production of Proteins in *Escherichia coli*: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *J. Mol. Biol.* **260**: 289–298
- Molineux IJ & Panja D (2013) Popping the cork: mechanisms of phage genome ejection. *Nat. Rev. Microbiol.* **11**: 194–204
- Musick WDL & Nyhan WL (1981) Structural Features of the Phosphoribosyl-Transferases and Their Relationship to the Human Deficiency Disorders of Purine and Pyrimidine Metabolism. *Crit. Rev. Biochem.* **11**: 1–34
- Nettleship JE, Brown J, Groves MR & Geerloff A (2008) Methods for Protein Characterization by Mass Spectrometry, Thermal Shift (ThermoFluor) Assay, and Multiangle or Static Light Scattering. In pp 299–318. Humana Press
- van Nieuwenhoven MH, Hellingwerf KJ, Venema G & Konings WN (1982) Role of proton motive force in genetic transformation of *Bacillus subtilis*. *J. Bacteriol.* **151**: 771–6
- Nijland R, Burgess JG, Errington J & Veening J-W (2010) Transformation of environmental *Bacillus subtilis* isolates by transiently inducing genetic competence. *PLoS One* **5**: e9724

- Notredame C, Higgins DG & Heringa J (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**: 205–217
- Nunn D, Bergman S & Lory S (1990) Products of three accessory genes, *pilB*, *pilC*, and *pilD*, are required for biogenesis of *Pseudomonas aeruginosa* pili. *J. Bacteriol.* **172**: 2911–9
- Ogura M, Yamaguchi H, Kobayashi K, Ogasawara N, Fujita Y & Tanaka T (2002) Whole-genome analysis of genes regulated by the *Bacillus subtilis* competence transcription factor ComK. *J. Bacteriol.* **184**: 2344–51
- Outten CE & O’Halloran T V (2001) Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. *Science* **292**: 2488–92
- Palzkill T (2013) Metallo- β -lactamase structure and function. *Ann. N. Y. Acad. Sci.* **1277**: 91–104
- Pan CQ, Uumer JS, Herzka A & Lazarus RA (1998) Mutational analysis of human DNase I at the DNA binding interface: Implications for DNA recognition, catalysis, and metal ion dependence. *Protein Sci.* **7**: 628–636
- Papadakos G, Wojdyla JA & Kleanthous C (2012) Nuclease colicins and their immunity proteins. *Q. Rev. Biophys.* **45**: 57–103
- Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC & Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**: 79–88

- Petersohn A, Brigulla M, Haas S, Hoheisel JD, Volker U & Hecker M (2001) Global Analysis of the General Stress Response of *Bacillus subtilis*. *J. Bacteriol.* **183**: 5617–5631
- Peterson SN, Sung CK, Cline R, Desai B V., Snesrud EC, Luo P, Walling J, Li H, Mintz M, Tsegaye G, Burr PC, Do Y, Ahn S, Gilbert J, Fleischmann RD & Morrison DA (2004) Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Mol. Microbiol.* **51**: 1051–1070
- Pettinati I, Brem J, Lee SY, McHugh PJ & Schofield CJ (2016) The Chemical Biology of Human Metallo- β -Lactamase Fold Proteins. *Trends Biochem. Sci.* **41**: 338–355
- Pimentel ZT & Zhang Y (2018) Evolution of the Natural Transformation Protein, ComEC, in Bacteria. *Front. Microbiol.* **9**: 2980
- Popot J-L, Althoff T, Bagnard D, Banères J-L, Bazzacco P, Billon-Denis E, Catoire LJ, Champeil P, Charvolin D, Cocco MJ, Crémel G, Dahmane T, de la Maza LM, Ebel C, Gabel F, Giusti F, Gohon Y, Goormaghtigh E, Guittet E, Kleinschmidt JH, et al (2011) Amphipols from A to Z. *Annu. Rev. Biophys.* **40**: 379–408
- Provvedi R, Chen I & Dubnau D (2001) NucA is required for DNA cleavage during transformation of *Bacillus subtilis*. *Mol. Microbiol.* **40**: 634–644
- Provvedi R & Dubnau D (1999) ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. *Mol. Microbiol.* **31**: 271–280
- Puyet A, Greenberg B & Lacks SA (1990) Genetic and structural characterization of *endA*. *J. Mol. Biol.* **213**: 727–738

- Québatte M, Christen M, Harms A, Körner J, Christen B & Dehio C (2017) Gene Transfer Agent Promotes Evolvability within the Fittest Subpopulation of a Bacterial Pathogen. *Cell Syst.* **4**: 611-621.e6
- Rabinovich L, Sigal N, Borovok I, Nir-Paz R & Herskovits AA (2012) Prophage excision activates listeria competence genes that promote phagosomal escape and virulence. *Cell* **150**: 792–802
- Rahmer R, Morabbi Heravi K & Altenbuchner J (2015) Construction of a Super-Competent *Bacillus subtilis* 168 Using the *P_{mtlA}-comKS* Inducible Cassette. *Front. Microbiol.* **6**: 1431
- Reboul CF, Eager M, Elmlund D & Elmlund H (2018) Single-particle cryo-EM-Improved *ab initio* 3D reconstruction with SIMPLE/PRIME. *Protein Sci.* **27**: 51–61
- Redfield RJ (1993) Genes for breakfast: The have-your-cake and-eat-It-too of bacterial transformation. *J. Hered.* **84**: 400–404
- Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**: W320–W324
- Rodriguez F, Lillington J, Johnson S, Timmel CR, Lea SM & Berks BC (2014) Crystal Structure of the *Bacillus subtilis* Phosphodiesterase PhoD Reveals an Iron and Calcium-containing Active Site. *J. Biol. Chem.* **289**: 30889–30899
- Rohou A & Grigorieff N (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**: 216–221

- Rollauer SE, Tarry MJ, Graham JE, Jääskeläinen M, Jäger F, Johnson S, Krehenbrink M, Liu S-M, Lukey MJ, Marcoux J, McDowell MA, Rodriguez F, Roversi P, Stansfeld PJ, Robinson C V., Sansom MSP, Palmer T, Högbom M, Berks BC & Lea SM (2012) Structure of the TatC core of the twin-arginine protein transport system. *Nature* **492**: 210–214
- Schilling O, Wenzel N, Naylor M, Vogel A, Crowder M, Makaroff C & Meyer-Klaucke W (2003) Flexible metal binding of the metallo- β -lactamase domain: Glyoxalase II incorporates iron, manganese, and zinc in vivo. *Biochemistry* **42**: 11777–11786
- Schwartz O, Axelrod JJ, Campbell SL, Turnbaugh C, Glaeser RM & Müller H (2019) Laser phase plate for transmission electron microscopy. *Nat. Methods* **16**: 1016–1020
- Schwarz D, Junge F, Durst F, Frölich N, Schneider B, Reckel S, Sobhanifar S, Dötsch V & Bernhard F (2007) Preparative scale expression of membrane proteins in *Escherichia coli*-based continuous exchange cell-free systems. *Nat. Protoc.* **2**: 2945–2957
- Seitz P & Blokesch M (2013) DNA-uptake machinery of naturally competent *Vibrio cholerae*. *Proc. Natl. Acad. Sci.* **110**: 17987–17992
- Sievers F & Higgins DG (2014) Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences. In *Methods in molecular biology (Clifton, N.J.)* pp 105–116.
- Simonen M & Palva I (1993) Protein secretion in *Bacillus species*. *Microbiol. Rev.* **57**: 109–37

- Sinha SC & Smith JL (2001) The PRT protein family. *Curr. Opin. Struct. Biol.* **11**: 733–739
- Smits WK, Eschevins CC, Susanna KA, Bron S, Kuipers OP & Hamoen LW (2005) Stripping *Bacillus*: ComK auto-stimulation is responsible for the bistable response in competence development. *Mol. Microbiol.* **56**: 604–614
- Spizizen J (1958) Transformation of Biochemically Deficient Strains of *Bacillus subtilis* by Deoxyribonucleate. *Proc. Natl. Acad. Sci. U. S. A.* **44**: 1072–8
- Stingl K, Müller S, Scheidgen-Kleyboldt G, Clausen M & Maier B (2010) Composite system mediates two-step DNA uptake into *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 1184–9
- Stura EA, Nemerow GR & Wilson IA (1992) Strategies in the crystallization of glycoproteins and protein complexes. *J. Cryst. Growth* **122**: 273–285
- Sysoeva TA, Bane LB, Xiao DY, Bose B, Chilton SS, Gaudet R & Burton BM (2015) Structural characterization of the late competence protein ComFB from *Bacillus subtilis*. *Biosci. Rep.* **35**: e00183
- Szilágyi A & Závodszky P (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8**: 493–504
- Takeno M, Taguchi H & Akamatsu T (2011) Role of ComFA in controlling the DNA uptake rate during transformation of competent *Bacillus subtilis*. *J. Biosci. Bioeng.* **111**: 618–623

- Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I & Ludtke SJ (2007) EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**: 38–46
- Tanner NK & Linder P (2001) DExD/H box RNA helicases: From generic motors to specific dissociation functions. *Mol. Cell* **8**: 251–262
- Tribet C, Audebert R & Popot J-L (1996) Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. *Proc. Natl. Acad. Sci.* **93**: 15047–15050
- Tsirigos KD, Peters C, Shu N, Kuhl L, Elofsson A, Käll L & Elofsson A (2015) The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**: W401–W407
- Tsirigotaki A, De Geyter J, Šoštaric' N, Economou A & Karamanou S (2016) Protein export through the bacterial Sec pathway. *Nat. Rev. Microbiol.* **15**: 21–36
- Turgay K, Hahn J, Burghoorn J & Dubnau D (1998) Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *EMBO J.* **17**: 6730–8
- Vagner V, Claverys J -P, Ehrlich SD & Méjean V (1990) Direction of DNA entry in competent cells of *Bacillus subtilis*. *Mol. Microbiol.* **4**: 1785–1788
- Waksman G (2019) From conjugation to T4S systems in Gram-negative bacteria: a mechanistic biology perspective. *EMBO Rep.* **20**:
- Waldo GS, Standish BM, Berendzen J & Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17**: 691–695

- Wang C, Tu J, Liu J & Molineux IJ (2019) Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. *Nat. Microbiol.* **4**: 1049–1056
- Wiedenbeck J & Cohan FM (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* **35**: 957–976
- Yadav T, Carrasco B, Serrano E & Alonso JC (2014) Roles of *Bacillus subtilis* DprA and SsbA in RecA-mediated Genetic Recombination. *J. Biol. Chem.* **289**: 27640–27652
- Yang W (2011) Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **44**: 1–93
- Yeh YC, Lin TL, Chang KC & Wang JT (2003) Characterization of a ComE3 homologue essential for DNA transformation in *Helicobacter pylori*. *Infect. Immun.* **71**: 5427–5431
- Zhang X-P & Heyer W-D (2011) Quality control of purified proteins involved in homologous recombination. *Methods Mol. Biol.* **745**: 329–43
- Zhang X-Z & Zhang Y-HP (2011) Simple, fast and high-efficiency transformation system for directed evolution of cellulase in *Bacillus subtilis*. *Microb. Biotechnol.* **4**: 98–105
- Zöllner R, Cronenberg T & Maier B (2019) Motor properties of PilT-independent type 4 pilus retraction in gonococci. *J. Bacteriol.* **201**: e00778-18

Zoonens M, Zito F, Martinez KL & Popot J-L (2014) Amphipols: A General Introduction and Some Protocols. In *Membrane Proteins Production for Structural Analysis* pp 173–203. New York, NY: Springer New York