Structural studies of the integral membrane component of the twin-arginine transport system, TatC

Sarah E. Rollauer

Department of Biochemistry

Sir William Dunn School of Pathology

Brasenose College



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

Trinity term 2013

Structural studies of the integral membrane component of the twin-arginine transport system, TatC

Sarah Rollauer

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the Univeristy of Oxford, Trinity term 2013

ABSTRACT

The twin-arginine protein transport (Tat) system is one of two general protein transport systems in the cytoplasmic membranes of bacteria, and is conserved in the thylakoid membranes of plants. The Tat system transports fully folded proteins of varying sizes across the membrane. This feat is achieved by the use of just three conserved integral membrane proteins, TatA, TatB and TatC. Passenger proteins are designated for transport by the system by the use of an N-terminal extension, termed a signal peptide which contains an invariant twin arginine motif. The TatC component of the transport system is responsible for recognising the signal peptide, as well as binding to TatB and the TatA components. TatC therefore emerges as the functional and organisational 'core' of the Tat system. There is a wealth of functional data relating to the TatC protein, but no high resolution structural information on TatC was available to interpret this data. In order to gain an understanding of the Tat system at the molecular level, a structure was required of TatC.

Preliminary expression screening using green fluorescent protein had previously identified the TatC protein from the thermophilic bacterium *Aquifex aeolicus* as being amenable for structural work. This study purified the *A. aeolicus* TatC construct in a variety of detergents, used native mass spectrometry and light scattering techniques to assay the sample homogeneity and set up extensive crystallisation trials. Following optimisation of a selenomethionine minimal media growth protocol, crystals grown from selenomethionine-substituted TatC purified in lauryl maltose neopentyl glycol diffracted to 3.5Å resolution, and allowed the structure of TatC to be solved by single wavelength anomalous dispersion.

Analysis of the novel structure of TatC, in combination with an isothermal titration calorimetric binding assay, gave information on how TatC binds to signal peptides. Further structural analysis combined with additional experimental data from collaborators allowed a model to be proposed for how TatC binds to the two additional integral membrane components of the system, TatB and TatA, as part of the transport mechanism. Subsequent work was undertaken in this study to attempt to gain a co-crystal structure of TatC with the signal peptide. Construct and detergent screening was carried out in order to aim towards high resolution structural characterisation of the TatBC complex.

Acknowledgements

I would like to thank, first and foremost, my supervisor, Susan Lea, for all her advice and support over the past four years of my DPhil. I am so glad of the opportunity to have worked in her lab, and appreciated the opportunity to work freely on this difficult project, without any objections to the copious amounts of money I spent on various detergents! I would also like to thank my second supervisor, Ben Berks, for his never ending enthusiasm for all things relating to TatC.

Thanks also go to Tracy Palmer (University of Dundee) and everyone in her lab for their contributions to my project. I also would like to thank the past members of the Berks lab who laid the ground work for the project, particularly Dr. Mike Tarry. I am grateful for the native mass spectrometry, by Julien Marcoux, and MD simulations by Phill Stansfeld.

I want to thank everyone in the Lea lab for their advice and help with experiments during the course of both my Part II, and my DPhil. Thanks to Dr. Janet Deane for teaching me the ropes when I first arrived in the lab during my undergraduate degree, and to Dr. Steven Johnson for all his help with MALLS, Dr. Patrizia Abrusci for helping me to get my head around any molecular biology issues which arose. Thanks to Dr. Pietro Roversi for accompanying me on the endless weekend synchrotron trips until TatC was solved. A massive thanks extends to everyone else in the lab (past and present) who have made day to day life so enjoyable (apart from the ping-pong…). I am sure that beer and cider at 4pm on a Friday afternoon will not be a regular feature of anywhere else I go on to work!

To the members of Bravo alpha november tango echo romeo, I thank you all for being such a great group of friends to me over the course of our DPhils. Our many nights out, nights in

and excursions made life away from the lab so much fun, providing a great distraction and many good memories.

Thanks to my family, for your love, support and the much welcomed day trips out of Oxford (even if we would always get lost on all our walks). And to Eachan, who has been there for me during the highs, and (many) lows of the TatC project. Thanks for keeping me motivated and supporting me, you always seemed to know it was going to work when I most definitely didn't.

And last but not least, TatC, I have to thank you for finally playing ball…

Publications resulting from this thesis

Rollauer, S. E., M. J. Tarry, J. E. Graham, M. Jaaskelainen, F. Jager, S. Johnson, M. Krehenbrink, S.-M. Liu, M. J. Lukey, J. Marcoux, M. A. McDowell, F. Rodriguez, P. Roversi, P. J. Stansfeld, C. V. Robinson, M. S. P. Sansom, T. Palmer, M. Hogbom, B. C. Berks & S. M. Lea, (2012) Structure of the TatC core of the twin-arginine protein transport system. *Nature* **492**: 210-214.

Younis, R., L. E. H. Bingle, **S. Rollauer**, D. Munera, S. J. Busby, S. Johnson, J. E. Deane, S. M. Lea, G. Frankel & M. J. Pallen, (2010) SepL Resembles an Aberrant Effector in Binding to a Class 1 Type III Secretion Chaperone and Carrying an N-Terminal Secretion Signal. *J. Bacteriol.* **192**: 6093-6098.

(Both articles are appended at the end of this thesis in **Appendix 8.2** (Younis *et al*) and **Appendix 8.6** (Rollauer *et al*))

List of abbreviations

ABC	Dimethylamine borane complex
APol	Amphipol
ASU	Asymmetric unit
ATP	Adenosine triphosphate
BN-PAGE	Blue native polyacrylamide gel electrophoresis
CC	Correlation coefficient
CCP4	Collaborative computation project 4
CD	Circular dichronism
CGMD	Course grain molecular dynamics
CLP	Cubic lipid phase
CMC	Critical micelle concentration
CT	C-terminus
CV	Column volume
DDM	n-dodecyl β-D maltopyranoside
DM	n-decyl β-D maltopyranoside
DMNG	Decyl maltose neopentyl glycol
DPPC	dipalmitoylphosphatidylcholine
EGC	Electronic grade coating
EM	Electron microscopy
ESRF	European synchrotron facility
FL	Full length
F-SEC	Fluorescence size exclusion chromatography
GFP	Green fluorescence protein
GPCR	G-protein coupled receptor
HT	High throughput
IMV	Inner membrane vescicle
IPTG	Isopropyl β-D-1-thiogalatopyranoside
ITC	Isothermal titration calorimetry
LDAO	Dodecyl-N,N-Dimethylamine-oxide
LLG	Log likelihood gradient

LMNG	Lauryl maltose neopentyl glycol
MALLS	Multi angle laser light scattering
MBCD	β-methyl cyclodextran
MCS	Multiple cloning site
MD	Molecular dynamics
ML	Maximum likelihood
MWC	Molecular weight cut off
NCS	Non crystallographic symmetry
NM	n-nonyl β-D maltopyranoside
NMR	Nuclear magnetic resonance
NT	N-terminus
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Poly ethylene glycol
PMF	Proton motive force
RMDF	Root mean square fluctuation
SAD	Single wavelength anomalous dispersion
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Sec	General secretory pathway
SEC	Size exclusion chromatography
SHC	Sirohydrochlorin
SRF	Self rotation function
SRP	Signal recognition particle
Tat	Twin arginine transport
TB	Terrific broth
TEV	Tobacco etch virus
TM	Transmembrane
UDAO	undecyl-N,N-Dimethylamine-oxide
UDM	n-undecyl β-D maltopyranoside
WT	Wild-type

Table of contents

1 Introduction 1
1.1 Protein transport across bacterial membranes 1
Transport of fully folded proteins? 4
Tat system substrates 6
Protein targeting to the Tat system 8
1.2 Integral membrane components of the Tat system 11
TatA and the TatA complex 12
TatB 14
TatC 15
The TatBC complex 15
A model for the transport system 17
1.3 Structural studies on protein transport systems 19
The structure of the Sec translocon, the SEC YEG channel 20
Structures of the Sec translocon in complex with partner proteins 24
Structural information on the oligomeric state of the Sec translocon 26
Structural insight into signal peptide binding in the Sec system 27
1.4 Structural Characterisation of the Tat system integral membrane components 28
TatA 31
The TatBC receptor complex 35
1.5 Identification of a TatC construct for structural studies 37
1.6 Thesis aims 41

2 *A. aeolicus* TatC detergent screening and crystallisation trials 43
2.1 Purification in dodecyl-maltopyranoside and optimisation of the initial crystal hits 49
Subsequent crystallisation trials in DDM yield only low resolution crystals 56
2.2 Multi angle laser light scattering (MALLS) and native mass spectroscopy reveals that TatC in DDM is a highly dynamic, oligomeric complex 60
2.3 Purifications in detergents that result in heterogeneous TatC species were unsuccessful in crystallisation trials 62
2.4 Purification of TatC in cymal-4 produces a homogenous sample 66
Crystallisation trials of TatC in cymal-4 68
Other techniques attempted to improve the crystals grown from in cymal-4 73
Dehydration 73
Increasing the size of the hydrophilic regions of the sample 75
Additive screening and detergent removal 77
Use of amphipols 78
2.5 Methylation of TatC in cymal-4 gives a new, higher resolution crystal form 79
Data collection and processing 81

Analysis of the asymmetric unit (ASU) 85
Difficulties in reproducing the 4Å cymal-4 crystal form… 90
2.6 Phasing attempts with the 4Å cymal-4 crystal form…. 92
Heavy atom soaks 93
Molecular replacement. 96
2.7 Summary 96

3 Determining the structure of *A. aeolicus* TatC 98
3.1 Purification of *A. aeolicus* TatC in LMNG 99
3.2 Crystallisation trials with *A. aeolicus* TatC in LMNG 103
3.3 Optimisation of the TatC LMNG H32 crystal form 107
3.4 Selenomethionine preparation of TatC in LMNG 109
Growth in Molecular Dimensions minimal media 111
Growth in B834 cells 114
Inhibition of the L56 methionine biosynthesis pathway 115
3.5 Crystallisation trials of SeMet TatC 119
3.6 Structure solution of TatC 123
Finding selenium sites…. 126
Density modification procedures to resolve the phase and hand ambiguities 131
Refinement of heavy atom parameters against more datasets 133
3.7 Model building and refinement 137
3.8 Summary 144

4 The structure of TatC and insights into signal peptide binding 145
4.1 The overall structure of TatC. 146
Molecular dynamic studies of TatC 150
4.2 Functionally important regions of the TatC structure. 154
4.3 Signal peptide binding by TatC 160
Structural insights into the interaction 165
A potential route for signal peptide insertion into the membrane? 175
4.4 Summary 180

5 Attempts to crystallise TatC and the signal peptide in complex 182
5.1 Peptide soaks 183
5.2 Co-crystallisation of *A. aeolicus* TatC with the signal peptide 184
5.3 *A. aeolicus*-like signal peptide sequences 186
5.4 Signal peptide fusions 187
Purification of the N-terminal fusion constructs 192
Purification of the C-terminal fusion constructs 197
5.5 Summary 197

6 Interactions of TatC with the other membrane components of the Tat system.....198
6.1 Self interactions between TatC monomers.....199
6.2 Interactions with TatA.....205
6.3 Interactions with TatB.....207
A model for the position of the rest of the TatB protein?.....212

7 Towards structural characterisation of the TatBC complex 216
7.1 Biophysical assessment of the *E. coli* TatBC complex.....217
7.2 Initial characterisation of the *A. aeolicus* TatBC complex.....221
7.3 Molecular biology to create constructs of A. aeolicus TatB suitable for co-expression with *A. aeolicus* TatC-GFP_{his} for structural studies.....224
7.4 Co-expression and co-purification screening of the Strep-tag and 3C-Strep-tag *A. aeolicus* TatB constructs with *A. aeolicus* TatC-GFP_{his}.....231
FL, 1-94 and 1-64 length constructs.....237
TM constructs.....239
7.5 Detergent screening of the four most promising *A. aeolicus* $TatB_{Strep}$ and $TatB_{3C\text{-}Strep}$ constructs.....241
7.6 Large scale purifications of *A. aeolicus* TatBStrepC from the co-expression vectors to assess the complex homogeneity by MALLS.....246
7.7 Summary and future perspectives.....252
7.8 Concluding remarks: A hypothetical model for the Tat translocase……..... 253
Main unanswered questions.....255

8 Appendix..... 257
8.1 Structural Characterisation of CbiX, the terminal enzyme of siroheme biosynthesis from *Paracoccus pantotrophus*.....257
Siroheme synthesis.....257
Expression, purification and structure determination of CbiX.....259
The structure of *P. pantotrophus* CbiX.....263
Model for substrate binding and ferrochelatase activity.....266
Co-ordination to the metalled tetrapyrrole.....268
Binding the metal in the metal only bound state.....271
8.2 SepL resembles an Aberrant Effector in Binding to a Class 1 type III Secretion Chaperone and Carrying an N-terminal Secretion Signal.....273
8.3 Construct directory.....275
8.4 Standard laboratory procedures.....278
TEV protease purification.....280
Alteration to TatC purification protocol for high CMC detergents.....281
Western blotting.....281
Agarose gel electrophoresis.....283
8.5 Molecular biology.....284

Mutagenesis of *A. aeolicus* TatC in pWaldo 285
Signal peptide binding site mutations 286
Signal peptide fusion RR to KK mutations 286
Creating *A. aeolicus* TatB constructs for co-expression with *A. aeolicus* TatC in pWaldo 287
8.6 The structure of the TatC core of the twin-arginine protein transport system 293

9 References 295

Index of Tables and Figures

1. Introduction

Figure 1.1.1 Protein transport into the bacterial periplasm 3
Figure 1.1.2 The sequence of the Tat signal peptide from the *E. coli* SufI protein..... 9
Figure 1.2.1 The overall topology of the three essential integral membrane components of the *E. coli* Tat system 11
Figure 1.2.2 A mechanism for the Tat transport cycle in *E. coli* and thylakoid membranes..... 18
Figure 1.3.1 Crystal structure of the SecYEG channel from *Methanococcus jannaschii* ... 22
Figure 1.4.1 Structural studies on the Tat system integral membrane components 34
Figure 1.5.1 A cartoon of the GFP fusion constructs 39
Figure 1.5.2 The pWaldo vector map and multiple cloning site 41

2. *Aquifex aeolicus* TatC detergent screening and crystallisation trials

Figure 2.1 The structures of two detergents used in this study 44
Table 2.1 Detergents used in initial F-SEC studies as promising candidates for *A. aeolicus* TatC purifications..... 48
Figure 2.1.1 The purification of TatC in DDM 51
Table 2.2 DDM optimisation screen 1..... 55
Figure 2.1.2 Crystals obtained in the DDM optimisation screen 1 55
Table 2.3 Broad range crystallisation screens used with the *A. aeolicus* TatC construct ... 57
Figure 2.1.3 New crystallisation conditions found in the broad screen for TatC in DDM . 57
Figure 2.1.4 Crystals of methylated TatC in DDM 59
Figure 2.2.1 Assessment of the oligomeric state of TatC in DDM by multi angle laser light scattering (MALLS) 60
Figure 2.2.2 Native mass spectroscopy of TatC in DDM 61
Figure 2.3.1 An overlay of the size exclusion chromatograms in the final step of the TatC purifications for each of the detergents tested..... 63
Figure 2.3.2 Multi angle laser light scattering (MALLS) analysis of TatC in NM..... 65
Figure 2.4.1 The size exclusion chromatogram of *A. aeolicus* TatC in cymal-4 66
Figure 2.4.2 Multi angle laser light scattering analysis of TatC in cymal-4 67
Figure 2.4.3 Crystals grown from broad screens of TatC purified in cymal-4..... 69
Table 2.4 Cymal-4 crystallisation optimisation screen 1 72
Table 2.5 Cymal-4 crystallisation optimisation screen 2 72
Figure 2.4.4 Crystals of TatC in cymal-4 grown from a purification using a 50kDa concentrator device 73
Figure 2.4.5 Multi angle leaser light scattering analysis of the TatC-GFP fusion constructs purified in cymal-4 76

Figure 2.5.1 Crystals of methylated TatC in cymal-4 identified in broad screen crystallisation trials 80
Table 2.6 The data quality statistics used to analyse the processed diffraction data 83
Table 2.7 The data processing statistics from two datasets from TatC cyaml-4 crystals.....84
Figure 2.5.2 Data processing of the ID29 TatC dataset 86
Table 2.8 An estimate of the solvent content of the two TatC datasets 87
Figure 2.5.3 The self rotation function (SRF) calculated for the two TatC datasets.....89
Figure 2.5.4 Crystals of TatC in cymal-4 grown in the optimisation trays 91
Table 2.9 A summary of the best diffraction data collected from heavy atom soaks attempted with the cymal-4 crystal form 95
Figure 2.6.1 The anomalous correlations against resolution for the highest resolution dataset of the K_2PtCl_4 soaked TatC crystals 95

3. Determining the structure of *Aquifex aeolicus* TatC

Figure 3.1 The structure of lauryl maltose neopentyl glycol (LMNG) 99
Figure 3.1.1 Purification of TatC in LMNG 100
Figure 3.1.2 Assessment of the homogeneity of TatC in LMNG 101
Figure 3.2.1 Crystals of TatC in LMNG grown in the MemGold broad screen 104
Table 3.1 Data processing statistics for the crystal grown from TatC in LMNG 105
Table 3.2 An estimate of the solvent content of the ASU 105
Figure 3.2.1 The self rotation function for the H32 crystal form of TatC in LMNG.....106
Table 3.3 LMNG optimisation screen 1 108
Table 3.4 LMNG optimisation screen 2 108
Figure 3.3.1 Crystals of TatC grown in the optimisation screens 108
Figure 3.4.1 Expression trials of *A. aeolicus* TatC-GFP in various minimal media conditions 112
Table 3.5 Expression conditions tested for TatC-GFP in the B834 (DE3) strain 115
Table 3.6 Minimal media recipe for the metabolic inhibition of methionine biosynthesis 117
Figure 3.4.2 The native mass spectra of TatC in LMNG 118
Figure 3.5.1 Crystals of the SeMet sample of TatC 120
Figure 3.5.2 X-ray energy scan carried out on the SeMet crystals at the selenium edge...121
Table 3.7 Maximum resolutions and overall anomalous correlations for seven SeMet datasets 123
Figure 3.6.1 Crystals of SeMet TatC in LMNG.....124
Table 3.8 The data processing statistics for two datasets from one of the SeMet TatC crystals 126
Figure 3.6.2 Data processing graphs for Dataset 1.....127
Figure 3.6.3 The w=0 Harker section of the anomalous difference Patterson for dataset 1 128
Figure 3.6.4 The result of the solvent flattening at 78% solvent.....133

Table 3.9 Data collection strategies used for collecting the datasets from the SeMet crystals 134
Table 3.10 Data processing statistics for the five datasets collected from additional SeMet TatC crystals 135
Table 3.11 The phasing power for each dataset as calculated by SHARP 136
Table 3.12 Positions of the Selenium atoms after refinement against the first two datasets (Initial) and after refinement against all seven datasets (Final) 136
Figure 3.7.1 The 73% solvent flattened electron density map. 138
Figure 3.7.2 Electron density maps for the 3.5 Å resolution structure of TatC 142
Table 3.13 Refinement statistics for the final model of TatC 144

4. The structure of TatC and insights into signal peptide binding

Figure 4.1.1 The structure of *A. aeolicus* TatC 147
Figure 4.1.2 An overlay of *A. aeolicus* TatC structures 150
Figure 4.1.3 Membrane bilayer deformation by TatC in course grain molecular dynamic simulations. 151
Figure 4.1.4 Thermal motions of TatC and stability in a bilayer 152
Figure 4.2.1 Sequence alignment of the *A. aeolicus* and *E. coli* TatC sequences 155
Table 4.1 Mutations that inactive *E. coli* TatC 157
Figure 4.2.2 Functional and conserved residues of TatC shown on the structure 158
Figure 4.3.1 Interaction of *A. aeolicus* in LMNG with *E. coli* signal peptides. 164
Figure 4.3.2 Identification of the signal peptide binding site on TatC. 166
Table 4.2 *E. coli* TatC residues implicated in signal peptide binding by three types of experiment. 168
Table 4.3 Primers used to introduce alanine mutations of the residues identified with a functional role in binding signal peptides 170
Figure 4.3.3 Isothermal titration calorimetry of *A. aeolicus* TatC variants. 171
Figure 4.3.4 Characterisation of the residues involved in the signal peptide interaction with TatC 172
Figure 4.3.5 A potential pathway for the h region of the signal peptide into the membrane, promoted by TatC 177

5. Attempts to crystallise TatC and the signal peptide in complex

Figure 5.4.1 A summary of the TatC-signal peptide fusion constructs created in this study 188
Table 5.1 Primers used to create the signal peptide fusion constructs 190
Figure 5.4.2 Expression trials of the TatC-GFP signal peptide fusion constructs 191
Figure 5.4.3. Purification in LMNG of the NT TatC signal peptide fusions 193
Figure 5.4.4 Isothermal titration calorimetry of the short signal peptide with the NT9 signal peptide fusion constructs of TatC 194

Figure 5.4.5 Multi angle laser light scattering (MALLS) of the TatC NT9 signal peptide fusion constructs. 195

6. Interactions of TatC with the other membrane components of the Tat system

Table 6.1 *E. coli* residues implicated in interactions with other Tat components.....200
Figure 6.1 Sites of interaction with the other Tat components. 200
Figure 6.1.1 Crystallographic dimer of *A. aeolicus* TatC 202
Figure 6.1.2 Evolutionary co-variance analysis of the *A. aeolicus* TatC structure. 203
Figure 6.2.1 TatA docking to the *A. aeolicus* TatC crystal structure. 206
Figure 6.3.1 A model for the interaction site between TatC and TatB 209
Figure 6.3.2 A model for the amphipathic helix of TatB interaction with TatC and the signal peptide. 213

7. Towards structural characterisation of the TatBC complex

Figure 7.1.1 Biophysical characterisation of the *E. coli* TatBC complex. 218
Figure 7.2.1 The *A. aeolicus tat* operon. 222
Figure 7.2.2. Purification of the genomic *A. aeolicus* TatBC construct in pWaldo. 223
Table 7.1 Characteristics of the pWaldo vector, and the two compatible vectors chosen for use for TatB co-expression, CDFDuet-1 and pACYCDuet-1.....225
Figure 7.3.1 Design of *A. aeolicus* TatB constructs for co-expression with *A. aeolicus* TatC-GFP_{his} in pWaldo. 226
Figure 7.3.2. Agarose gel electrophoresis of the PCR products for four different length *A. aeolicus* TatB fragments. 228
Table 7.2 Summary of the Strep-tag and 3C-Strep tag constructs created.....230
Figure 7.3.3 An overview of the *A. aeolicus* TatB construct and detergent screening process.....232
Table 7.3 Expression conditions for the five different length constructs of TatB. 233
Table 7.4 The High Throughput nickel magnetic bead purification protocol.....234
Figure 7.4.1. Level of *A. aeolicus* TatC-GFP present in the nickel bead eluate 235
Figure 7.4.2. Dot blot of the HT nickel magnetic bead elutate from co-expression and co-purification of TatC with the 20 constructs of TatB. 236
Figure 7.4.3. Western blot analysis of a SDS-PAGE gel of the FL, 1-94 and 1-64 fragments from the co-expression screening. 238
Figure 7.4.4. Western blot analysis of the TM fragments from the co-expression screening. 240
Table 7.5 The four best *A. aeolicus* TatB constructs and corresponding expression conditions, chosen for detergent screening with co-expressed *A. aeolicus* TatC-GFP....242
Figure 7.5.1. Assessment of the level of protein purified in each detergent. 243
Figure 7.5.2 Assessment of the homogeneity of each of the four TatB constructs by F-SEC. 245

Figure 7.6.1 Large scale preparations of TatC with three TatB constructs in LMNG. 248
Figure 7.6.2 MALLS analysis of each of the three TatB constructs copurified with TatC in LMNG 250
Table 7.6 Molecular masses for the *A. aeolicus* TatBC constructs 250
Figure 7.8.1 A hypothetical molecular model for the initial steps in Tat translocation ... 254

8. Appendix

Figure 8.1.1 The terminal step in the siroheme biosynthesis pathway. 258
Figure 8.1.2 Biophysical characterization of CbiX from *P. pantotrophus* 261
Table 8.1 The data processing statistics for the native and Pt derivativised CbiX datasets 262
Table 8.2 Refinement statistics for the final model of CbiX 264
Figure 8.1.3 Structural comparison of *P. pantotrophus* CbiX with *B. subtilis* HemH. 265
Table 8.3 Residues shown to be involved in metal co-ordination. 267
Fig 8.1.4. Investigating the active site of *P. panatrophus* CbiX. 269
Table 8.3.1 *A. aeolicus* TatC-GFP_{His} in the Xho1 and BamH1 sites of the pWaldo vector 275
Table 8.3.2 Highlighted in green, the positions of the single Ala point mutations introduced into the *A. aeolicus* TatC-GFP_{His} construct 275
Table 8.3.3 The TatC-signal peptide fusion constructs of *A. aeolicus* TatC-GFP_{His} created in pWaldo. 276
Table 8.3.4 The sequence of the 5 different lengths of A. aeolicus TatB fragements created in this study 277
Table 8.3.5 The sequence of the Strep-tag and 3C-strep CT modifications 277
Table 8.4.1 Competent cell transformation protocol 278
Table 8.4.2 Phosphate buffered saline (PBS) recipe used in this study 278
Table 8.4.3 The composition of the SDS-PAGE reducing loading dye and the gel (resolving and stacking). 279
Table 8.4.4 Immunodetection protocol for electro-transferred proteins. 282
Table 8.5.1 The Quikchange mutagenesis PCR protocol 284
Table 8.5.2 Primers used to introduce three further alanine mutations of the residues identified with a functional role in binding signal peptides 285
Table 8.5.3 Primers to create the KK mutant versions of the five signal peptide fusions. 286
Table 8.5.4 Primers used to amplify TatB fragments from the genomic TatBC pWaldo vector for insertion into MCS1 of CDFDuet-1 and MCS1 of pACYC 287
Table 8.5.5 The touchdown PCR protocol for amplification of the TatB fragments from the genomic *A. aeolicus* TatBC pWaldo vector 288
Table 8.5.6 The ‘gentle’ transformation protocol. 289
Table 8.5.7 Primers used to insert an 8 amino acid Streptavidin (Strep-tag) at the CT of each TatB construct in CFFDuet-1 and pACYCDuet-1 291
Table 8.5.8 Primers for the generation of a 3C-cleavage site prior to the Streptavidin-tag 292

Chapter 1: Introduction

1.1 Protein transport across bacterial membranes

Protein transport across bacterial membranes is a fundamental process which allows bacteria to survive in, and adapt to, their surroundings. The number of proteins requiring transport varies, but in the Gram negative bacterium, *Escherichia coli*, it has been estimated at 25-30% of the proteins produced are transported from their point of synthesis in the cytoplasm, across, or into, the cytoplasmic membrane in order to perform their function (Palmer *et al.*, 2005). Transport across this membrane can result in their localisation to the periplasmic space, an important sub-compartment of the bacterial cell where many functionally essential processes and reactions occur. For instance, the periplasm is the location of electron transport proteins which work in conjunction with proteins embedded in the inner membrane to provide energy for the cell, it is the location of certain nutrient binding proteins, and of receptors involved in chemotaxis (Klein, 2005).

Proteins may also be secreted beyond the bacterial outer membrane to the extracellular milieu (Sandkvist, 2001), or even injected into eukaryotic cells in the case of some pathogenic bacterial protein toxins (Mota & Cornelis, 2005). This can occur directly, by forming a channel from the bacterial cytoplasm across both inner and outer membranes to these locations, or it can occur in a step wise manner, whereby proteins that have already been localised to the periplasm can be transported onwards, across the outer membrane (Tseng *et al.*, 2009). Bacteria have therefore evolved a number of protein transport machines, embedded within their membranes, in order to transport a wide array of substrate proteins to a variety of final destinations (Chatzi *et al.*, 2013). Each machine has

evolved to fulfil a specific task, and differs in the destination of the protein transported, the timing of its use, and the types of proteins it transports.

The bacterial transport machines may adopt very different overall architectures (Koster *et al.*, 2000), but they are all presented with the common predicament of maintaining the integrity of the membrane. This must remain sealed to other proteins, ions and small molecules, in order to sustain, for example, the crucial energy transducing ion gradients across the bacterial cytoplasmic membrane (Papanikou *et al.*, 2007).

A simple mechanism to ensure the integrity of the membrane involves the creation of a very small pore through the membrane, through which the passage of substrates can be highly regulated. Fully folded proteins, which can have very large diameters are incompatible for transport by such a system. However, in their unfolded state, proteins have a much smaller and more uniform cross section, allowing them to pass through such a pore. This logic is exploited by the ubiquitous Secretory (Sec) protein transport system (**Figure 1.1.1 A**), which transports proteins in their unfolded state (Driessen, 2001) across the bacterial cytoplasmic membrane, and across the eukaryotic endoplasmic reticulum (ER) membrane. It also acts to insert most membrane proteins into the membrane (Scotti *et al.*, 2000). The system uses energy from nucleotide hydrolysis to 'thread' unfolded proteins through a tightly regulated, and small, channel in the membrane (Van den Berg *et al.*, 2004) termed the translocon.

'Threading' of unfolded proteins is the mechanism used for the vast majority of proteins transported, and it is exploited for instance by some of the bacterial secretion systems, which have been categorised into eight broad types. One such example is the Type III Secretion System (T3SS), a complex nano-machine composed of over 20 different protein components, present in some Gram-negative bacteria (Galan & Wolf-Watz, 2006). The

T3SS can be divided into two sub categories. Firstly the flagellar system, which transports the components of the flagellar hook and filament. The assembled flagellum mediates bacterial motion and swimming (Erhardt *et al.*, 2010). The second category is the translocation-associated system, which is involved in the direct injection of toxic 'effector' proteins from the bacterial cytoplasm into a eukaryotic cell cytoplasm (Aizawa, 2001). Transport of the unfolded effector protein occurs through a narrow, needle-like structure (Schraidt & Marlovits, 2011) which protrudes from a basal body structure, which spans both the inner and outer membranes. Located below the base of the basal body, in the cytoplasm, is an ATPase which may provide the energy necessary to unfold the effector proteins before they are threaded through the needle (Minamino *et al.*, 2008). Effectors are delivered to the system bound to specific chaperone proteins, which act to prevent their

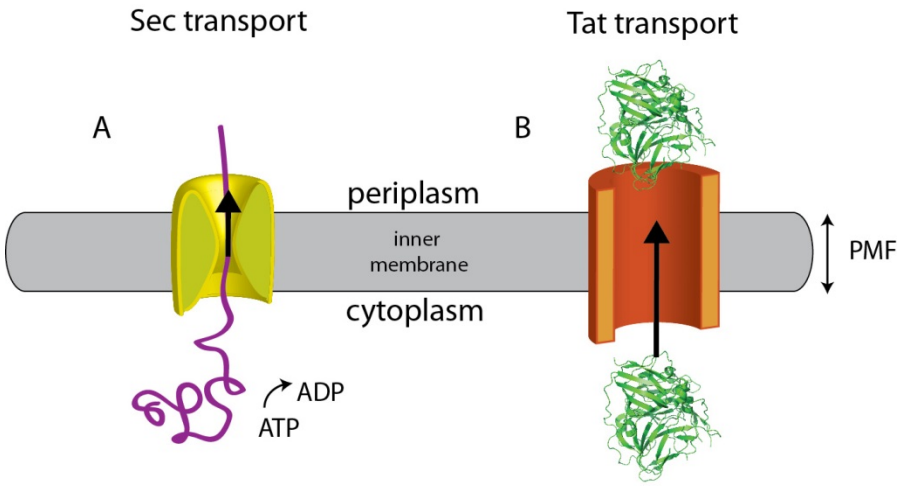


Figure 1.1.1 Protein transport into the bacterial periplasm.

Transport occurs by two major routes, the ubiquitous Secretory (Sec) system (**A**) which transports proteins in the unfolded state by passage through a narrow pore (Van den Berg *et al.*, 2004), powered by ATP usage by chaperones that contact the pore protein (Economou & Wickner, 1994). (**B**) The twin arginine transport (Tat) system transports fully folded proteins across the membrane, and is powered solely by the PMF (Mould & Robinson, 1991) (Cline *et al.*, 1992).

premature secretion, and may retain the protein in a semi-unfolded state prior to export (Page & Parsot, 2002). As an aside to the main focus of this thesis, an effector protein-chaperone complex from the enteropathogenic *E. coli* T3SS has been studied. The published results (Younis *et al.*, 2010) are given in **Appendix 8.2.**

Transport of fully folded proteins?

Remarkably, given the necessity to retain the membrane seal, a protein transport system exists in the cytoplasmic membranes of bacteria which transports fully folded proteins across the cytoplasmic membrane (Berks, 1996). This transport system is referred to as the 'Twin Arginine Transport', or Tat, system (**Figure 1.1.1 B**) and along with the bacterial Sec system, forms the general route for protein export into the periplasm. The Tat system was initially characterised in plant thylakoids, where it carries out transport of proteins from the stroma across the thylakoid membrane into the thylakoid lumen (Mould & Robinson, 1991).

The transport of folded proteins has been demonstrated by various experiments. For example, both *in vivo* and *in vitro* the Tat targeted alkaline phosphatase PhoA is only transported to the periplasm (DeLisa *et al.*, 2003) or into inner membrane vesicles (IMV) (Panahandeh *et al.*, 2008) if oxidative disulphide bridge formation is able to occur in the cell cytoplasm or IMV reaction setup. The requirement for protein folding before Tat transport was also shown by Tat targeting the cytochrome *c* protein, which was only Tat transported following the incorporation of the heme cofactor and folding (Sanders *et al.*, 2001). However, it is known that in certain cases, the Tat system can transport specifically

engineered single unfolded strands, indicating that protein folding is not a strict mechanistic requirement for Tat transport (Richter *et al.*, 2007).

Unlike the Sec system, the Tat transport system is faced with the formidable task of transporting proteins which can have very large cross sections across the membrane, whilst retaining the seal and integrity of the membrane. Furthermore, the cross section of the proteins the system transports varies significantly, from as little as 10 Å (Richter *et al.*, 2007) to over 70 Å (Berks *et al.*, 2000), suggesting that it does not use a uniformly sized channel. To date, there is only one other protein transport system which seems capable of achieving such a difficult task, the seemingly unrelated peroxisomal import pathway (Schliebs *et al.*, 2010).

Not only do the Sec and Tat systems differ in the folded state of the proteins they transport, but the power sources utilised by the Sec system and Tat systems are also very different. The Sec system initially requires the hydrolysis of ATP directly to help thread the passenger protein through the pore. The subsequent transport steps can then be driven by the proton motive force (PMF) (Driessen, 1992). The Tat system relies solely on the PMF (Mould & Robinson, 1991, Cline *et al.*, 1992), although the molecular mechanism of energy transduction is not understood. The combination of the difference in the folded state of the substrate, and the different energy utilisation suggests that transport by the Tat system will require a very different mechanism to the Sec system, and implies the Tat system arose independently. The need to export proteins post-folding, such as those proteins which assemble with redox-sensitive co-factors, presumably acted to drive the elution of the Tat system in bacteria and the evolutionary related thylakoids (Celedon & Cline, 2013).

Tat system substrates

The first Tat system substrates characterised were proteins which have a non-covalently bound cofactor for their vital role in the many metabolic processes that occur in the periplasm (Berks, 1996), and these tend to be either metal-sulphur clusters or nucleotide containing co-factors. Enzymes of energy metabolism, whether intra-cellular or extra-cellular and in both prokaryotes and eukaryotes, extensively make use of exquisitely designed metal containing co-factors in order to carry out complex redox chemistry, which would prove impossible using protein chemistry alone (Crane *et al.*, 1997). The pathways and the enzymes involved in the manufacture of cofactors and prosthetic groups, are therefore of critical importance. One such enzyme, CbiX from the denitrifying bacterium *Paracoccus pantotrophus,* acts as a chelatase to insert iron into the metal-free tetrapyrrole sirohydrochlorin to produced siroheme (Bali *et al*, unpublished). This co-factor is necessary to perform the redox reactions of the assimilatory sulphite and nitrite reductases, and also as an intermediate to another vital prosthetic group, heme d_1 (Tripathy *et al.*, 2010). The CbiX enzyme has therefore been studied and structurally characterised as an aside to this thesis, the results given in **Appendix 8.1.**

In the case of transport of these cofactor containing proteins, the Tat system allows for the protein to fold around its cofactor before transport (Halbig *et al.*, 1999), avoiding the need for a specific export pathway for the cofactor if the protein had instead been transported unfolded by the Sec pathway. However, not all co-factor containing proteins use the Tat system. Those with co-factors that are only found in periplasmic proteins tend to be Sec transported. It has been postulated that this may be because there is no pre-existing dedicated insertion machinery for these co-factors in the cytoplasm (Palmer & Berks, 2012) whereas co-factors that are also found in cytoplasmic proteins can take advantage of

the pre-existing cytoplasmic machinery for their insertion prior to Tat transport (Berks, 1996).

The need to attain a fully folded state is also a requirement for the transport of hetero-oligomeric complexes, whereby only one protein in the oligomer has a Tat transport signal. If the protein were to be transported in an unfolded conformation, this would result in only one member of the complex being located correctly. As one protein carries a non-targeted protein into the periplasm, this is known as a 'hitchhiker' mechanism (Rodrigue *et al.*, 1999).

Some periplasmic proteins that bind metal ions face the problem of acquiring their correct metal ions. The periplasm contains competing metal ions, and here incorrect ions could bind to the protein instead. The Tat system allows such metallo-proteins to acquire their metal ion in the cytoplasm where metal insertion can be finely controlled and regulated (Tottey *et al.*, 2008).

Protein folding may also dictate why certain substrates use the Tat pathway. Compared to *E. coli*, where Tat transport only accounts for around 6% of all secreted proteins (Lee *et al.*, 2006b), it is known that archaea living in high salt environments transport a much higher proportion of their proteins by the Tat system (Rose *et al.*, 2002). This maybe due to the inability of some proteins to fold *de novo* in the high salt concentrations found in the periplasm (Pohlschröder *et al.*, 2004). Proteins exhibiting fast folding kinetics in the cytoplasm under certain conditions are also Tat transported, as these proteins are clearly incompatible with the need to maintain the unfolded state for Sec transport. This shows bacteria are able to adapt and use different transport machines depending on their environment.

In addition to these water soluble substrate proteins, the range of Tat pathway substrates is increased by its ability to transport membrane proteins. In *E.coli* there are five such examples, each membrane protein containing a single transmembrane helix (TM) located at the C-terminus (CT) of the protein (Hatzixanthis *et al.*, 2003). In bacteria and plants, an important example includes the Rieske iron-sulphur protein. This protein forms an essential component of the cytochrome bc_1 and b_6f complexes, present within proteins of the respiratory and photosynthetic energy transducing pathways (Bachmann *et al.*, 2006). It is interesting to note that the Rieske protein from actinobacteria, which is larger and contains three TM domains is inserted into the membrane by the combined actions of both the Sec system and the Tat systems (Keller *et al.*, 2012) .

As the examples of the types of proteins requiring Tat transport are so varied, the substrates encompass proteins with significantly varying functions, from energy metabolism (Berks, 1996) to cell division (Stanley *et al.*, 2001) and cell motility (Ding & Christie, 2003). Many of these processes are key to bacterial virulence, and thus the presence of a functional Tat pathway is increasingly being shown to be necessary for virulence in pathogenic bacteria (De Buck *et al.*, 2008).

Protein targeting to the Tat system

Given the that the Tat pathway is able to transport such a large repertoire of structurally and functionally distinct proteins, it is of crucial importance to the bacterium that only those proteins destined for movement across the membrane are transported, and that proteins are transported by the appropriate transport pathway. The bacterium achieves this

specificity by the use of specialised targeting signals on the proteins to be transported (Hegde & Bernstein, 2006).

The signal is in the form of an N-terminal extension on the protein to be transported, referred to as a 'signal peptide'. The presence of a Tat signal peptide alone is necessary and sufficient for recognition and transport by the Tat system: *in vitro* the peptide is able to competitively inhibit Tat transport of other substrates (Alami *et al.*, 2002) and the addition of a Tat signal peptide in place of the Sec peptide results in the Tat transport of the passenger protein (Blaudeck *et al.*, 2003).

Certain features of the signal sequence are conserved between both the Tat and Sec system (von Heijne, 1985). The peptide is composed of three regions, which are highlighted on the *E. coli* Tat substrate SufI sequence in **Figure 1.1.2**. There is a positively charged n-region, a hydrophobic h-region, and a polar c-region. Within the c-region, a conserved sequence of small amino acids at -1 and -3, typically AXA, dictates the position for the cleavage of the signal peptide by the signal peptidase. This peptidase is a periplasmically located protease which can cleave the signal peptide from the passenger protein following the completion of protein transport (Lüke *et al.*, 2009).

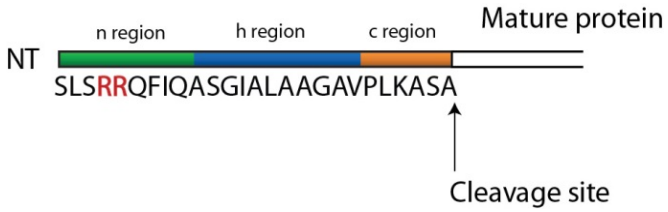


Figure 1.1.2 The sequence of the Tat signal peptide from the *E. coli* SufI protein.

The general regions are shown: the positively charged n-region, the hydrophobic h-region and the polar c-region. The n-region contains the consensus sequence, containing the twin arginine pair (red). The cleavage site is located in the c-region, at the junction with the mature protein.

For Tat substrates, the n-region contains a highly conserved consensus sequence; Ser-Arg-Arg-X-Phe-Leu-Lys (X= any polar amino acid) (Berks, 1996). Within this sequence, the presence of the invariant twin arginine motif gives rise to the pathways name. The requirement for this motif is demonstrated by the fact that the conservative substitution of both the twin arginines to twin lysines (KK) *in vivo* will, in general, abolish transport of a protein by the Tat system (Stanley *et al.*, 2000). However, this requirement is not as strict as first thought, as experimental evidence has shown that some substitutions introduced in Tat substrates can be partially tolerated, which suggested the first arginine of the twin arginine pair being more crucial than the second (DeLisa *et al.*, 2002, Ize *et al.*, 2002). In contrast, sequence analysis of Tat substrates shows there is a single naturally occurring exception to the 'RR' motif in known signal peptides, in which the first arginine is replaced by a lysine (Hinsley *et al.*, 2001).

The contribution from the other residues in the consensus motif, which are present in over 50% of substrate sequences, seem to act to increase the efficiency of Tat transport, but their exact effect can vary between proteins (Mendel *et al.*, 2008) and the sensitivity of the reporter used to assess Tat transport. Furthermore, transport depends on the exact nature of the replacement residue, certain amino acids can be tolerated at some positions, when others can not (Lausberg *et al.*, 2012).

Despite their similar overall architectures (Heijne, 1990), the signal peptides used by the Sec system are generally shorter, typically18-26 amino acids compared to the average 37 amino acid Tat peptide (Bendtsen *et al.*, 2005). The c-region of Tat peptides are enriched for basic residues, compared to the Sec signals, and have been demonstrated to act as Sec 'avoidance' signals (Bogsch *et al.*, 1997). Sec peptides also have a more hydrophobic (h-) region. Interestingly, the specificity of a Tat signal peptide can be altered to that of a Sec peptide by subtly altering the hydrophobicity of either the n- or h-regions (Cristobal *et al.*,

1999). A key challenge of the Tat system is that it must efficiently discriminate between these two types of similar signal peptide, to ensure only the desired proteins with the specific Tat signal peptide are transported.

1.2 Integral membrane components of the Tat system

In Gram-negative bacteria the Tat system performs the difficult task of selectively transporting only those proteins harbouring the Tat specific signal peptide, in a fully folded conformation (DeLisa *et al.*, 2003), by the use of just three conserved integral membrane proteins refered to as TatA, TatB and TatC (**Figure 1.2.1**) (Bogsch *et al.*, 1998). The homologous of these proteins in plant thylakoids are referred to as Tha4, Hcf106 and cpTatC.

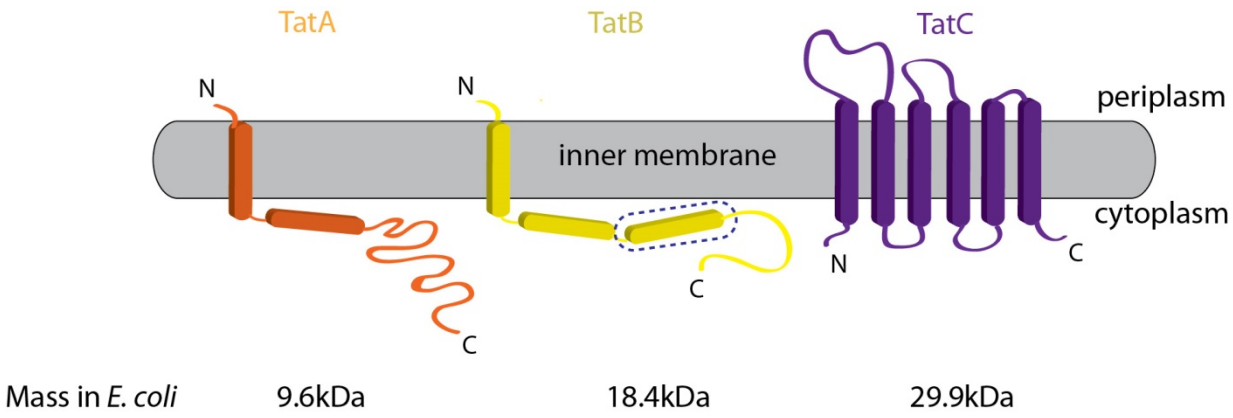


Figure 1.2.1 The overall topology of the three essential integral membrane components of the *E. coli* Tat system.

Helices are shown as cylinders, and loops/unstructured regions shown as lines. The topologies are confirmed by biochemical experiments, and the secondary structure shown based on the NMR structure of TatA (Rodriguez *et al.*, 2013), and based on sequence predictions and biochemical experiments, for TatB (Koch *et al.*, 2012) and TatC (Behrendt *et al.*, 2004) (Punginelli *et al.*, 2007). The additional amphipathic helix in TatB, compared to TatA, is highlighted in a dashed box.

In *E. coli* a fourth Tat protein, TatE, is encoded by a second *tat* locus and is predicted to be very similar to the TatA protein, with 53% sequence identity (Sargent *et al.*, 1998). It has been hypothesised that *tatE* arose as a cryptic gene duplication of the *tatA* gene (Yen *et al.*, 2002). *tatE* has been shown to be able to complement the phenotype of a Δ*tatAtatE* mutant strain indicating a similar function (Sargent *et al.*, 1999). However, it is generally thought to be less crucial than TatA since many more substates are mislocalised in a *tatA* mutant strain than in a *tatE* mutant strain (Sargent *et al.*, 1998), and *tatE* is expressed at much lower levels than TatA (Jack *et al.*, 2001). Overexpression of just *tatABC* in *E. coli* was shown to be sufficient to give export of an *E. coli* substrate, it is assumed that these three proteins form the minimal translocase (Sargent *et al.*, 2001).

A defining feature of the three Tat membrane proteins is their ability to interact and form complexes within the membrane. These complexes will now be discussed.

TatA and the TatA complex

TatA in *E. coli* is just under 10kDa in mass and sequence analysis predicted a single, short transmembrane (TM) helix, followed by a single amphipahtic helix and an unstructured CT (Settles *et al.*, 1997). The position of the termini of this protein has been debated, but recent evidence implies the N-terminus (NT) of the protein is located in the periplasm, and the CT remains strictly cytosolic and does not alter topology during transport (Koch *et al.*, 2012). The overall structure was confirmed by a solution NMR structure of the *Bacillus subtilis* TatA protein (Hu *et al.*, 2010) and also of the *E. coli* TatA protein (Rodriguez *et al.*, 2013).

TatA in resting membranes exists as small homo-oligomeric complexes which are known to increase in size during Tat transport. In the resting state, the small homo-oligomeric

complexes may be composed of tetrameric units, as inferred from the fluorescence microscopy of Yellow Fluorescent Protein (YFP) labelled TatA protein in *E. coli* (Leake *et al.*, 2008). During transport, these complexes can increase in size to an average of 25 TatA molecules per complex. Higher oligomerisation states have been inferred in thylakoids, where the TatA component can be crosslinked as small (tetrameric units) in resting membranes, and then into larger oligomers during transport (Dabney-Smith *et al.*, 2006). Formation of the larger oligomers requires both the TM and the amphipathic helix of TatA (Dabney-Smith & Cline, 2009).

A requirement for a high stoichiometry of TatA in the translocase is implied by an assessment of the level of transcription of each *tat* gene by β-galactosidase translational fusions which showed TatA is produced in *E. coli* at a 20 fold higher level than TatB and at a 50-fold higher level than TatC (Jack *et al.*, 2001). This is also seen at the protein level; inspection of the stoichiometric amounts of each of the three proteins in native membranes in a recent study in the chloroplast Tat system reported a 20 fold excess of the TatA protein compared to the TatC protein (Celedon & Cline, 2012) and where titrating TatA into the transport system, suggested 26 molecules of TatA on average are needed per translocation event.

The presence of oligomers of TatA, along with other evidence (Mori & Cline, 2001), implies TatA is the protein translocating element of the system. The mechanism by which TatA functions is debated, with two main theories. Firstly, it may polymerise around the protein to be transported to make a custom sized transport pore. Electron microscopy studies, which revealed pore-like particles of varying size (Gohlke *et al.*, 2005), coupled with the requirement for bound substrate to initiate TatA oligomerisation (Dabney-Smith *et al.*, 2006) provide evidence for a substrate-specific pore. The nature of the interior of the pore is unknown, and recent reports suggest a flipping mechanism, where by the

amphiphatic helices of TatA flip into the membrane to line the transport channel (Walther *et al.*, 2013). This mechanism would reqire the CT tails to remain inside at all times, to fit with the known TatA topolgy. Alternatively, given the short length of the TM helix, TatA oligomers may function to weaken the membrane such that a protein located close to the surface of the bilayer is able to pass through (Rodriguez *et al.*, 2013).

TatB

TatB in *E. coli* is 18.4kDa in mass and shares 25% identity with TatA, such that it is likely the two proteins evolved from a common ancestor (Sargent *et al.*, 1999) . In some Tat systems, such as that present in the Gram-positive bacterium *B. subtilis*, only TatA and TatC are present. The TatA, termed $TatA_d$, seems to act in a dual 'TatA-TatB' role (Jongbloed *et al.*, 2004). Consistent with this, substitutions at the NT of TatA can suppress a TatB deficiency in *E. coli* (Blaudeck *et al.*, 2005). However, in wild-type (WT) *E.coli*, both TatA and TatB are required for function, implying they fulfill non-identical roles (Sargent *et al.*, 1999).

Although no structure is avaliable for the TatB protein, given its sequence similarity to TatA, it is likely to be similar to TatA. The sequence from *E. coli* is predicted to have a single, short transmembrane helix, followed by two amphipathic helices, and an unstructured, cytoplasmically located CT tail, which has been shown to be dispensible for function (Lee *et al.*, 2002). The major difference compared to the TatA protein in *E. coli* is the insertion of a second amphipathic helix in the structure (Maldonado *et al.*, 2011b) shown in a dashed box in **Figure 1.2.1.**

TatC

At 29.9kDa in *E.coli*, the TatC component is the largest and most highly conserved of all the Tat components (Yen *et al.*, 2002). The conservation extends to the cpTatC protein from thylakoids. The cpTatC protein contains an additonal leader sequence compared to the bacterial protein, approximately 40 amino acids in length (Ma & Cline, 2013) to allow cpTatC import into the chloroplast. The conservation between the rest of the TatC sequence from *Pisum sativum* and *E. coli* is around 35% identity, and there are many identical residues across multiple sequence alignements of bacterial and plant sequences (Allen *et al.*, 2002). TatC from *E. coli* was predicted to have six TM helices, with both its NT and CT inside the cytoplasm (Bogsch *et al.*, 1998). These predications have been confirmed by the use of reporter fusion proteins inserted into predicted loop regions of the protein (Behrendt *et al.*, 2004, Jeong *et al.*, 2004) and by cysteine accessibility studies with membrane permeable and impermeable thiol reactive regents (Punginelli *et al.*, 2007). However no further structural details were known.

The TatBC complex

In native membranes, TatC is known to form a tight complex with TatB (Bolhuis *et al.*, 2001). Within this complex, TatB is estimated to be present in stoichiometric amounts to TatC. A 1:1 stoichiometry was implied by the construction of a fusion protein of TatB-TatC which remained functional for Tat transport. Size exclusion chromatography (SEC) of the purified TatBC complex expressed at native levels (McDevitt *et al.*, 2006), blue native polyacrylamide gel electrophoresis (BN-PAGE) of isolated *E. coli* complexes, low resolution EM studies (Tarry *et al.*, 2009b) and BN-PAGE of the TatBC complex in plant

thylakoids (Celedon & Cline, 2012), implies that this complex is a large oligomer, containing 6-8 copies of both TatB and TatC.

The TatBC complex in *E. coli* is composed of distinct and stable TatC and TatB sub-complexes (Orriss *et al.*, 2007). The TatC sub-complex is proposed to itself be composed of a core building block of a local dimer of TatC (Buchanan *et al.*, 2002, Maldonado *et al.*, 2011a). Cysteine scanning studies suggest the TatB sub-complex may be composed of tetramer units (Lee *et al.*, 2006a).

The TatBC complex functions in the crucial step of recognizing and binding the twin arginine signal peptides on the passenger proteins designated for Tat transport (Cline & Mori, 2001). A model has been proposed whereby signal peptide recognition is sequential. Site-specific crosslinking has confirmed TatC is involved in the initial recognition of the twin arginine motif where TatB contacts the entire signal peptide (Alami *et al.*, 2003) as well as the mature folded passenger protein (Maurer *et al.*, 2010). The contact between TatB and the signal peptide is strictly dependent on the presence of TatC (Alami *et al.*, 2003), making TatC the primary receptor for the signal peptide.

Site specific crosslinking (Zoufaly *et al.*, 2012, Ma & Cline, 2013), in combination with an array of mutational data (Holzapfel *et al.*, 2007) and the sites of suppressor mutations allowing transport of sequence defective signal peptides (Lausberg *et al.*, 2012, Kreutzenbeck *et al.*, 2007, Strauch & Georgiou, 2007) map the region of TatC involved in the interaction with the signal peptide to the extreme NT and predicted first cytoplasmic loop. These studies give an indication of where the binding site may lie, but the lack of structural data on TatC makes this information hard to interpret, and does not give details on the interaction at the molecular level, or how the binding site is specific for RR versus KK peptides.

Evidence for a hairpin-like insertion of the signal peptide into the membrane has been implied in thylakoids in which passenger proteins with the NT of the signal peptide covalently fused to TatC can still undergo transport (Fincher *et al.*, 1998). This observation has been supported by experiments in *E. coli* derived inner membrane vesicles which showed that TatC mediated transport of the signal peptide CT across the membrane independently of TatB or the PMF (Fröbel *et al.*, 2012a). This lead to the suggestion that TatC does not just specifically recognise the RR signal peptide, but can also catalyse signal peptide insertion into the membrane.

A model for the transport mechanism

In order to bring about the process of Tat transport the TatA and TatBC complexes interact (Mori & Cline, 2002). The knowledge of the composition of the resting complexes comes from affinity purifications of tagged proteins from the *E. coli* bacterial system reconstituted into inner membrane vesicles (Sargent *et al.*, 2001) and various biochemical crosslinking data in the plant thylakoid system (Mori & Cline, 2002). The order in which these complexes may interact has generated a potential model of the transport mechanism. An overview of the mechanism, as it is currently understood, is given in **Figure 1.2.2.** In this model the Tat system builds an intact translocon only when needed. This would aid in the task of maintaining the membrane seal when the system is not transporting.

Firstly, signal peptides have been shown to be able to peripherally bind to membranes independently of the Tat system components (Bageshwar *et al.*, 2009). Membrane association would allow them to undertake a random, two-dimensional search of the membrane surface (**Figure 1.2.2 A**) until they encounter a TatBC receptor complex. The TatBC complex can then specifically bind (**Figure 1.2.2 B**) (Gérard & Cline, 2006), and

then insert the signal peptide (**Figure 1.2.2 C**) into the membrane mediated primarily by TatC (Fröbel *et al.*, 2012a), as discussed previously.

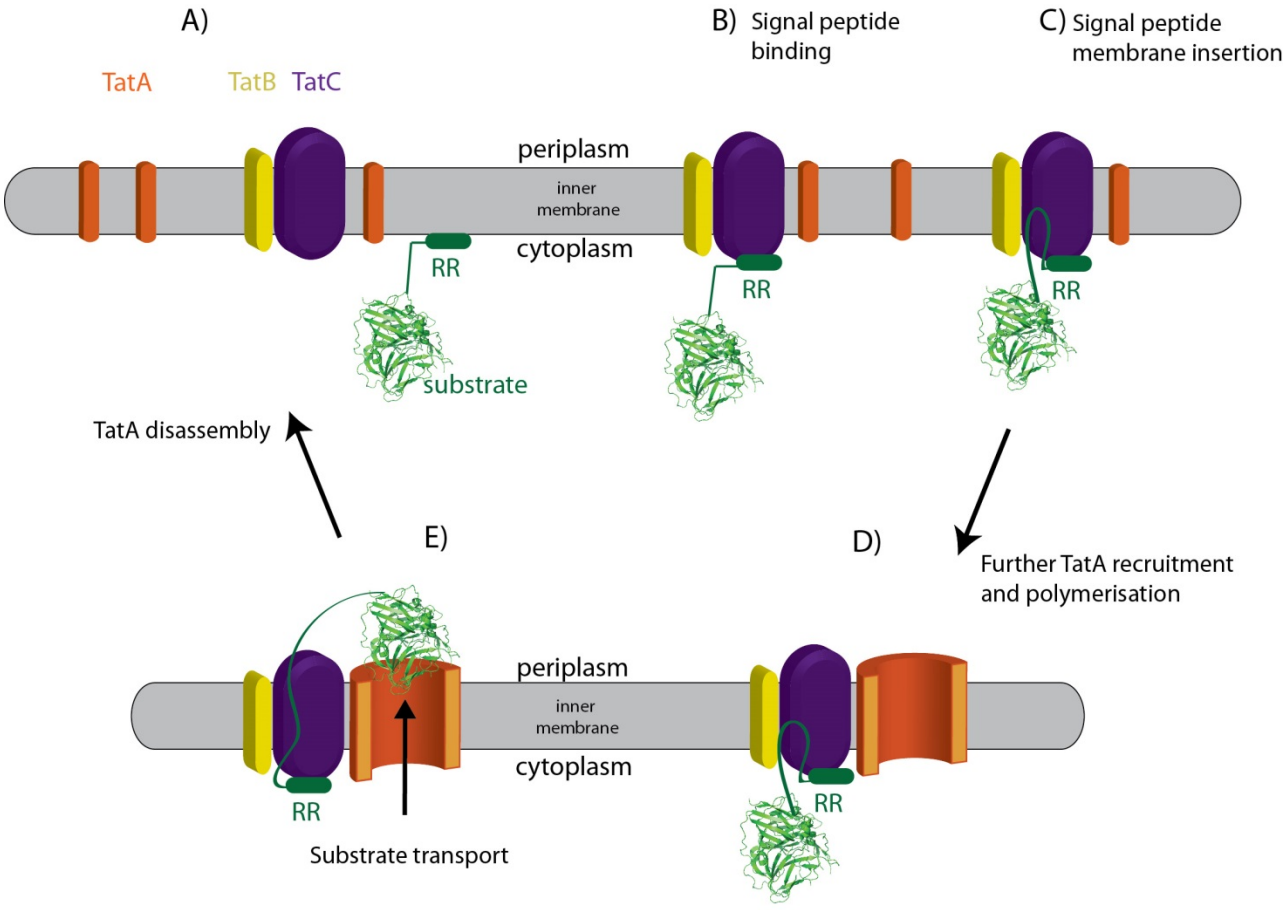


Figure 1.2.2 A mechanism for the Tat transport cycle in *E. coli* and thylakoid membranes.

TatB (yellow) and TatC (purple) form a tight, 1:1 complex (Bolhuis *et al.*, 2001). TatA (orange) exists as discrete, smaller complexes, although evidence exists that monomeric TatA is able to contact and bind to the TatBC complex (Fröbel *et al.*, 2011) **(A)**. The signal peptide of the folded protein (green) may bind to the membrane and undergo a random 2D search. It can then bind a TatBC complex, **(B)** with the interaction mediated primarily by TatC. TatC can then insert the signal peptide into the membrane, forming a hairpin **(C)** (Fröbel *et al.*, 2012a). TatA is then recruited in a PMF dependent manner to the TatBC complex (Mori & Cline, 2002) **(D)** where it can oligomerise and allow passage of the substrate across the membrane **(E)** (Dabney-Smith *et al.*, 2006).

Upon signal peptide binding to the TatBC complex the small, potentially tetrameric, TatA complexes in the membrane are recruited, in a PMF-dependent manner (Mori & Cline, 2002), to the receptor complex, where they undergo oligomerisation (**Figure 1.2.2 D**) (Dabney-Smith *et al.*, 2006). It is possible to purify the TatBC complex with a small amount of TatA present even in the non-substrate bound state (Bolhuis *et al.*, 2001), and to cross link TatBC to TatA before the substrate is bound (Fröbel *et al.*, 2011), indicating that there maybe a binding site on the TatBC receptor for some permanent, monomeric TatA. TatA is then able to contact the signal sequence, in a PMF dependent manner. This may form an advanced, PMF dependent stage of interaction of the signal peptide with the Tat translocase as previously noted in plant thylakoids (Gérard & Cline, 2007).

Signal peptide binding may then induce a conformational change in the receptor complex to recruit more TatA to bind to the TatBC complex(Dabney-Smith *et al.*, 2006). The TatA oligomer in the TatABC complex then mediates protein transport across the membrane (**Figure 1.2.2 E**) (Mori & Cline, 2002). The TatC component emerges as the 'core' of the transport system which can act as a framework to bind signal peptides and the two other components of the Tat system. Upon completion of transport, the TatA complex can disassemble (Cline & McCaffery, 2007). The signal peptide is then processed by the signal peptidase (Lüke *et al.*, 2009), although the exact timing of this step in relation to the completion of transport is still not clear.

1.3 Structural studies on protein transport systems

Each of the Tat translocase proteins have been the target of extensive mutagenesis and biochemical studies aimed at understanding the nature of their interactions with the signal

peptide, and with the other membrane components of the system (Fröbel *et al.*, 2012b) . However, in order to gain detailed information on the mechanism of action of the Tat system at the molecular level, and rationalise the data currently available, structures of the individual Tat components, and of the complexes they form, are required.

Structures of proteins can suggest how they function, and mechanisms tested by the design of experiments based on their structure and available biochemical data from complementary biochemical data can be rationalised. These points are well exemplified by the Sec protein transport system. The ubiquitous Sec system, introduced in **section 1.1**, transports unfolded proteins across membranes (Manting & Driessen, 2000). The molecular mechanism of Sec transport is the best understood of any transport system due to a decade's worth of structural studies.

The Sec system is the major route for the transport of proteins across the bacterial cytoplasmic membranes. It involves the use of a protein conducting channel in the membrane, the translocon (Brundage *et al.*, 1990). The Sec system can transport proteins which have already been fully translated and have been kept in an unfolded state by the use of chaperones or, it can transport proteins as they are being translated, by engaging with the proteins as they emerge from a ribsome. In both these types of transport the signal peptide on the transported substrate is recognised by targeting factors that target the protein to the translocon (Valent *et al.*, 1997).

The structure of the Sec translocon, the SecYEG channel

The membrane embedded Sec translocon in eukaryotes is composed of three integral membrane proteins, α, β and γ subunits, that are the minimal membrane components necessary for Sec transport (Brundage *et al.*, 1990). These are highly conserved and are

referred to as SecY, E and G in prokaryotes and archaea. A complex of these proteins was predicted to form a protein conducting channel, initially demonstrated by photocrosslinking experiments in which this subunit was able to contact and surround the polypeptide chain during its translocation (Mothes, 1994) and electrophysiology studies which showed the presence of an aqueous interior to the channel (Simon & Blobel, 1991).

Initially, a medium (8Å) resolution electron microscopy structure of the Sec translocon became available, which gave some information on the oligomeric state of the channel (Breyton *et al.*, 2002), but important details of the transport mechanism were revealed upon the solution of a 3.2 Å crystal structure of the Sec translocon published in 2004 (Van den Berg *et al.*, 2004). The structure provided an explanation for how this system can transport unfolded proteins and still allow the bilayer seal to remain intact.

The crystal structure of the Sec translocon from the archaea *Methanococcus jannaschii* revealed a membrane embedded channel (**Fig 1.3.1)** with an hourglass shape when viewed from the bilayer. A narrow constriction in the centre forms a ring of 'pore' lining residues. This constriction is located above an additional short helix that appeared to be blocking the end of the channel and was referred to as the 'plug helix'. The architecture of the α subunit was shown to be arranged as two sections, the first section composed of TM helices 1-6 and the second section composed of helices 7-10, forming a 'clam shell' arrangement. The plug helix is helix 2b, and thus loops out from the first section. The γ subunit stretches across the back face of the channel, seemingly stabilising and linking the two clam shell halves. The β subunit is located peripherally, and does not form part of the core of the channel explaining why this protein is dispensable for function.

Further X-ray crystal structures have been solved from a variety of bacteria (Zimmer *et al.*, 2008) (Tsukazaki *et al.*, 2008) all showing a similar overall architecture implying that the structure and therefore transport mechanism are highly conserved.

The structure suggested that translocation of passenger proteins occurs, by threading them through the narrow pore in the centre of the α subunit. It also showed how the channel is able to maintain the permeability barrier to small molecules. The structure suggested that the membrane seal is made by the channel itself, using the ring of residues in the constriction site to form a gasket around the translocating protein, as it travels through the channel in an extended conformation. The plug, when the pore is not translocating, can remain in place and help stabilise the closed channel. The structure has rationalised many biochemical and electrophysiological studies on how the seal is maintained (Saparov *et al.*, 2007), and raised questions about florescence quenching experiments that suggest the channel may not be a barrier to small molecules (Hamman *et al.*, 1997).

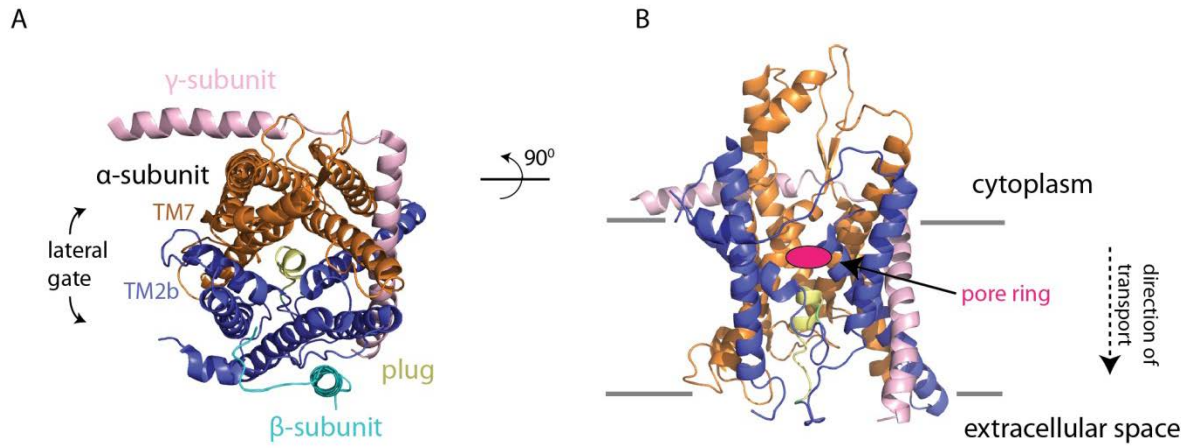


Figure 1.3.1 Crystal structure of the SecYEG channel from *Methanococcus jannaschii*.

(A). The channel viewed from the cytoplasmic face with the γ subunit coloured pink, the β subunit coloured cyan and the central α subunit shown in orange (TM 1-6) and blue (TM 7-12). The plug helix (2b) is shown in yellow. **(B)** Side view, with the β subunit removed in order to more clearly show the central constriction. The position of the side chain residues which point into the constriction and constitute the 'pore ring' is shown by the purple oval.

Mapping the mutations which allow for the transport of passenger proteins with defective signal peptides (the *prl* mutations) (Derman, 1993) in combination with the evident ‘clam shell’ shape, suggested a mechanism involving partial intercalation of the signal sequence into the membrane. This implied the lateral gate between TM2b and TM7 can be opened, to allow the signal peptide to bind in this membrane exposed pocket to open the pore and rationalised the photo crosslinks seen between the signal peptide and the membrane lipid (Plath *et al.*, 1998).

Molecular Dynamics (MD) simulations are being increasingly used as a complementary technique to study membrane proteins. These *in silico* experiments can predict structural changes in membrane proteins and crucially they can help to understand how they behave in a membrane environment (Lindahl & Sansom, 2008) something that currently cannot be achieved by high resolution structural techniques or easily achieved by other experimental method. In the case of the Sec translocon channel simulations were carried out to confirm the feasibility of lateral gate opening and plug displacement, as was suggested by the structure. Furthermore, they were able to observe how membrane lipids interact with a channel with an open lateral gate (Gumbart & Schulten, 2007).

The SecYEG structure exemplifies how much a single crystal structure can rationalise existing biochemical and mutagenesis data, and provide details on the transport mechanism. As mentioned, there is an abundance of these types of data for the Tat system proteins, and so gives further incentive to structurally characterise these proteins to understand the Tat mechanism in greater detail.

Structures of Sec translocon in complex with partner proteins

Unlike the Tat system, in which the three membrane proteins and the presence of a PMF are sufficient to transport passenger proteins, the Sec membrane channel is simply a passive pore. In order to bring about transport, it needs to associate and transiently bind to soluble partner proteins which localise the substrate to the channel and provide the energy to thread it through. Therefore, despite the many insights into the mechanism that were provided by the structure of the Sec translocon alone, elucidating in detail each step of the transport process has involved solving structures of the Sec channel whilst bound to its channel partners.

In bacteria, the translocon can associate with the ATPase SecA to carry out post translational protein transport. SecA is an ATPase which binds to pre-proteins and, along with the SecB chaperone, helps to retain them in the unfolded state. SecA provides the energy to thread the proteins through the translocon (Economou & Wickner, 1994). The determination of the crystal structure of SecYEG bound to SecA shows how SecA associates with the channel (Zimmer *et al.*, 2008), and how it may be opened by the signal sequence. The structure revealed that SecA docking onto SecYEG causes the lateral gate of the Sec translocon to partially open and displace the plug thus capturing a snapshot of the first step in the protein translocation mechanism. This structure implied a position of the translocating protein which was consistent with systematic disulphide crosslinking (Bauer & Rapoport, 2009). However, a structure of SecYEG:SecA in complex with a passenger protein chain is still missing, and would confirm the details of the route the protein takes through both proteins.

The mechanism by which SecA may act to convert the energy from ATP hydrolysis into movement of the protein through the Sec translocon channel was revealed through crystal

structures of the ATPase from many different species, with different nucleotides bound (Hunt *et al.*, 2002). These structures show large scale differences in the conformation of two of the domains in SecA, which led to the postulation of a mechanism of 'pushing' the protein through the translocon channel by a 'two finger helix'. The structures have therefore generated theories on the molecular basis for energy utilisation by the Sec transport system, suggesting structures may aid our understanding of how the PMF drives Tat transport, which is currently not understood. A structure of SecA in complex with a synthetic peptide led to insight into how SecA is able to transport such a large range of protein substrates that can differ in sequence significantly (Zimmer & Rapoport, 2009) by use of a β-strand augmentation mechanism (Remaut & Waksman, 2006).

The ribosome is the partner protein used for cotranslational protein transport through the Sec channel (Matlack *et al.*, 1998) The understanding of the mechanism of cotranslational translocation has been aided by cryo electron microscopy (EM) studies of the Sec channel in complex with a ribosome, and the fitting of the crystal structures of the constituent components into the electron density. The position of the ribsome on the translocon implied the direct passage of the precursor from the exit tunnel of the ribosome into the Sec channel.

There has been a point of conflict between structures of Sec translocon-ribsome complexes obtained from detergent solubilised channels (Becker *et al.*, 2009), and those within a nano-disc environment (Frauenfeld *et al.*, 2011) surprisingly the former saw no conformational change of the SecY channel whereas the latter saw the opening of the lateral gate. Unfortunately, in the nano-disc reconstruction, the nascent chain could not be fully traced through the channel and appeared to have been already released in to the membrane, and thus gave no real biological insight into the exact positioning of the nascent chain during protein transport.

Structural information on the oligomeric state of Sec translocon

The issue of whether the Sec translocon is required to form higher oligomers for function, and whether this oligomeric state changes during transport, is debated (Osborne & Rapoport, 2007). Although the crystal structure of the Sec translocon revealed that just a single copy of the complex can form a channel through the centre of the α component (Van den Berg *et al.*, 2004), verified by *in vivo* crosslinking experiments (Osborne & Rapoport, 2007, Park & Rapoport, 2012) oligomeric Sec translocon channels have been seen in intact membranes by freeze fracture EM (Hanein *et al.*, 1996), and their presence detected by other biochemical studies (Deville *et al.*, 2011, Mori *et al.*, 2003).

The medium (8Å) resolution 3D electron microscopy structures of the purified Sec translocon calculated from 2D crystals indicated that a dimer of SecYEG molecules were present (Breyton *et al.*, 2002), potentially in a 'back to back' arrangement, consistent with crosslinking data (Deville *et al.*, 2011). EM of 2D crystals of the SecYEG bound to a pre-protein mimic (Hizlan *et al.*, 2012) indicated that just one copy of a dimer may be functional for transport. However, this structure was of too low resolution to understand at the molecular level the reason for the apparent asymmetry in SecYEG channel function. To date, an unequivocal, high resolution X-ray crystal structure of a SecYEG oligomer has proven difficult to obtain, the channel is monomeric in the structure of the channel alone and in complex with either SecA or the ribsome. This maybe, in part, due to the solubilisation of the Sec translocon channel from the membrane by detergents, a necessary step in order to study these proteins at high resolution by X-ray crystallography, causing these fragile complexes to break apart into their monomeric components (Bessonneau *et al.*, 2002). By contrast, the 2D crystals which have given the dimer reconstructions (Hizlan *et al.*, 2012, Breyton *et al.*, 2002) mimic the native membrane environment, and so

minimise the problems associated with detergent solubilisation, which may suggest that this is the natural, functional state.

How molecules of the Sec translocon are arranged with respect to one another in biological membranes during protein transport is therefore one of the major outstanding issues for the Sec field. It provides an example of how a biological relevant high resolution structure of a unstable membrane protein oligomer has proved very elusive, a problem that will be of relevance to the oligomer forming proteins of the Tat system.

Structural insight into signal peptide binding in the Sec system

The signal peptide recognition particle (SRP) is the soluble protein involved in the crucial step of recognising and binding to the NT signal peptide present on proteins to be co-translationally Sec translocated. SRP binds the signal peptide as it emerges from the ribosome, and then targets the pre-protein-ribosome complex to the Sec translocon channel (Doudna & Batey, 2004). The receptor role of SRP is analogous to that of the TatBC complex which is the only component of the Tat system which specifically recognises the twin arginine-containing signal peptides present on proteins to be Tat transported.

SRP contains a C-terminal, methionine rich ‘M-domain’ which is the required domain for signal peptide binding (Lütcke H., 1992). The specific details on how SRP performs its recognition function were provided firstly by an apo-crystal structure of the M-domain. The structure revealed the M-domain has a groove lined by flexible hydrophobic residues which could allow SRP the structural plasticity it needs to bind signal sequences of varying length and sequence (Keenan *et al.*, 1998). Subsequently, the crystal structure of SRP in complex with a bound signal sequence was solved, which showed the h-region of the signal peptide bound as an α helix in the groove and a conformational change in the

structure of the SRP M-domain, indicating an induced fit mechanism during the recognition (Janda *et al.*, 2010). These structures have paved the way to understanding the recognition interaction in more detail. However, more structures of SRP with different peptides bound will still be needed to delineate the specificities of the interaction, and truly understand the determinants of binding to different h-regions, with different sequences. In contrast, given the TatC interaction with signal peptide is much more sequence specific, due to the highly conserved consensus motif, a co-structure of TatC and the signal peptide will shed much light on to the residues involved in binding all Tat signal peptides.

The signal peptide-Sec translocon interaction has been visualised in cyro-EM reconstructions of the ribosome-translocon complexes (Frauenfeld *et al.*, 2011) and a reconstruction by EM of 2D crystals of a Sec channel bound to a truncated pre-protein (Hizlan *et al.*, 2012). Both structures seemed to confirm the biochemical experiments, and the structural evidence, which suggested the lateral gate between helices 2 and 7 can open, to allow the signal peptide to intercalate between these helices, where it is partially exposed to the membrane. However, this was visualised at low resolution, and so the exact atomic details are still required.

1.4 Structural characterisation of the Tat system integral membrane components

The three major components of the Tat system are integral membrane proteins. This class of protein is notoriously difficult to study structurally, as evident from the lack of high resolution structures in the protein data bank (PDB). There are around 350 entries for novel membrane protein structures compared to almost 90,000 total deposited structures (Sciara & Mancia, 2012). The reasons for the difficulties in the structural characterisation of membrane proteins broadly fall into two major categories, obtaining a useable level of

expression of the membrane protein of interest and the stability of the purified membrane protein.

To structurally study proteins by either X-ray crystallography or by nuclear magnetic resonance (NMR) requires large (milligram) amounts of pure protein. For X-ray crystallography, material is required to set up hundreds of crystallisation trials in order to screen as many conditions as possible to find one in which the protein can crystallise. Milligrams of pure protein material is often very difficult to achieve. Membrane proteins tend to be naturally present at low levels in biological membranes. The rare examples in which these proteins are naturally abundant, such as in the case of bacteriorhodopsin in which it is the only protein present in the purple membranes of *Haliobacter salinaria* (Lozier *et al.*, 1975), can be exploited to easily obtain a high yield of protein for structural study (Henderson & Unwin, 1975). However, most targets must be over-expressed. Membrane proteins are particularly hard to recombinantly over-express in bacteria, the most common expression host for soluble proteins. Often over-expression of the membrane protein can be toxic to the cells, or can saturate the membrane embedded Sec channel which is used to co-translationally insert them into the membrane (Wagner *et al.*, 2007). These issues and the recent developments to overcome the expression problem will be discussed in further detail in **section 1.5**.

Having expressed the membrane protein of interest in a host organism, the protein must be extracted from the membrane in order to be purified. This membrane solubilisation step is achieved by detergents which also act to maintain the membrane protein in solution (more detail on detergents and their mechanism of action will be discussed in **Chapter 2**). In doing so the membrane protein is removed from the environment of the membrane, where specific combinations of lipids can help to stabilise the protein. The role of lipids in both membrane protein folding and as a topology determinant are being increasing understood

(Bogdanov *et al.*, 2002). The loss of the lipid-protein interactions can contribute to the membrane protein being unstable in detergents (Guan *et al.*, 2006), particularly those membrane proteins from eukaryotic organisms. The unstable protein may undergo aggregation and so will be unsuitable for structural study (Matar-Merheb *et al.*, 2011). Or, given that certain lipids are required for the function of some membrane proteins (Dowhan, 1997), the lack of specific lipids may render the membrane protein to no longer be functional in the detergent solubilised state. Membrane proteins from bacteria which have adapted to survive in extreme environments, such as thermophilic bacterial proteins, tend to be more stable upon membrane extraction.

The rationale that thermophilic bacterial membrane proteins have evolved to become more thermostable and hence tend to be more stable in detergents, has lead to the approach of introducing thermo-stabilising mutations into eukaryotic membrane proteins. The more stable, mutant proteins can be more robust in detergents and more amenable to crystallisation (Warne *et al.*, 2008). This technique has been most powerfully demonstrated in the G protein-coupled receptors (GPCR) field (Lebon *et al.*, 2011, Doré *et al.*, 2011).

Bovine rhodopsin is a GPCR which is extremely abundant in the eye and can be easily kept in a single active conformation. These two properties, coupled with its extreme natural stability in detergents, allowed it to form crystals diffracting to high resolution and represented the first member of the eukaryotic GPCR family to be solved by X-ray crystallography (Palczewski *et al.*, 2000). However, rhodopsin was the exception as most other GPCRs are easily switched between two conformational states and are very unstable in detergents. In order to overcome the instability, a high throughput ligand binding assay in conjunction with heating has lead to the identification of thermostable generated by systematic alanine scanning of each residue in the GPCR sequence (Serrano-Vega *et al.*, 2008). In many cases, four or more mutations need to be combined to obtain a receptor that

can be stable at over 20^{o}C higher temperatures than the native protein. These thermostable proteins seem to be held in one conformational state, and are more stable in certain detergents, which can aid in their crystallisation and structure determination. The first example of the success of this technique for high resolution structure determination was for the thermostabilised turkey β_1 adrenergic receptor, solved in complex with a high affinity antagonist to 2.7Å resolution (Warne *et al.*, 2008). There are more examples engineering membrane proteins to attempt to improve their ability to crystallise, these will be discussed in detail in **Chapter 2.**

The membrane proteins of the Tat system

Alongside the more general problems discussed above related to the structural characterisation of membrane proteins, the membrane protein components of the Tat components also undergo self-oligomersation (discussed in **section 1.2**), which may lead to heterogeneous samples unsuitable for high resolution X-ray or NMR studies. The oligomerised membrane proteins also form complexes. The transient existence of some of these (**Figure 1.2.2**) will make the structural characterisation of these complexes very difficult. Thus, currently, there is very little structural information on the Tat system proteins.

TatA

As discussed in **section 1.2**, the TatA component is thought to exist in small, tetrameric units within the resting membrane (Leake *et al.*, 2008) which can homo-oligomerise to form larger structures upon recruitment to a substrate-bound TatBC complex (Dabney-Smith *et al.*, 2006) with an average of 25 subunits of TatA (Leake *et al.*, 2008). Extracting TatA from membranes with detergents results in TatA homo-oligomers of varying sizes, as

assessed by BN-PAGE. These complexes range from 100kDa to over 500kDa and can be seen to run as a mixture of bands (Oates *et al.*, 2005). Although *in vivo* oligomerisation requires the TatBC complex (Dabney-Smith *et al.*, 2006) in a PMF dependent manner (Cline & Mori, 2001), detergent extraction seems to be able to drive the protein to oligomerise. This resulting mixture of oligomeric states is normally incompatible with structure determination by X-ray crystallography or NMR, but is amenable to single particle electron microscopy (EM), in which different sized particles can be grouped into different classes, and independently averaged within each class.

3-Dimensional (3D) reconstructions of TatA complexes by random conical tilt of single particle electron microscopy images revealed particles of different sizes (Gohlke *et al.*, 2005). The reconstruction (**Figure 1.3.1 A(i)**) show capped 'doughnut'-like structures which are proposed to be transport 'pores' closed with a lid. These structures have internal pore diameters that vary from 30-70 Å consistent with the range of diameters of Tat substrates suggesting a possible mechanism for the ability of TatA oligomers to bring about transport of passenger proteins that vary greatly in their size. Despite giving further evidence to a pore-like mode of transport, the molecular details of the TatA protein oligomerisation were not able to be provided by this study. Site specific spin-label electron paramagnetic resonance (EPR) studies show that TatA TM helices are arranged as a single walled ring in detergent, which would suggest the walls of the channel-like structures seen by EM are made up of TatA TM regions packing together (White *et al.*, 2010). Further steps towards a molecular understanding of the TatA structure were made when the solution NMR structure of an individual $TatA_d$ molecule from *B. subtilis* was obtained from a monomer of the protein purified in the detergent DPC (Hu *et al.*, 2010). This revealed an L-shaped molecule with a 13 amino acid TM helix that is roughly 25 Å in length and so shorter than the typical width of a bilayer (40-50 Å). The TM forms a 90^{o}

angle with the amphiphatic helix. The remaining CT residues, from 49-70, were unstructured. This structure confirmed the secondary structural elements predicted from the sequence analysis, but did not shed much light onto the mechanism of TatA action, which is known to involve complexes larger than a monomer.

Attempts have been made to understand how this monomer may build up into larger structures. A solution NMR model of a nonomer of *E. coli* TatA molecules has more recently been proposed (Rodriguez *et al.*, 2013)(**Figure 1.4.1 A (iii)).** This model was achieved by using an *E. coli* TatA variant which appeared to maintain its oligomeric conformation when monomerised and exploiting the ability to finely control the balance of this monomer to oligomer state by varying the protein:detergent ratio. This allowed NMR assignments of this structure, which would have been unobtainable by traditional NMR methods. The TatA monomer was very similar to that of the *B. subtilis* $TatA_d$ monomer. Despite the oligomer model having a similar overall diameter to those seen in the EM studies (~100 Å), the TatA construct used in the NMR had the 40 non-essential CT residues truncated, so these extra residues are not accounted for. Furthermore, its overall shape, with the amphipathic helices pointing outwards, seems incompatible with the overall shape seen in the EM studies.

MD simulations of the structure within a dipalmitoylphosphatidylcholine (DPPC) bilayer allowed for a model of the TatA mechanism of action to be proposed. As discussed previously for the Sec translocon, the power of these simulations comes not just from the details they give on the membrane protein internal motions, but also on how the protein behaves and interacts with its natural bilayer environment. In the case of TatA, the short transmembrane helices of the TatA oligomer model were able to distort the membrane, and a water filled channel was seen to form partially across the bilayer, in the centre of the

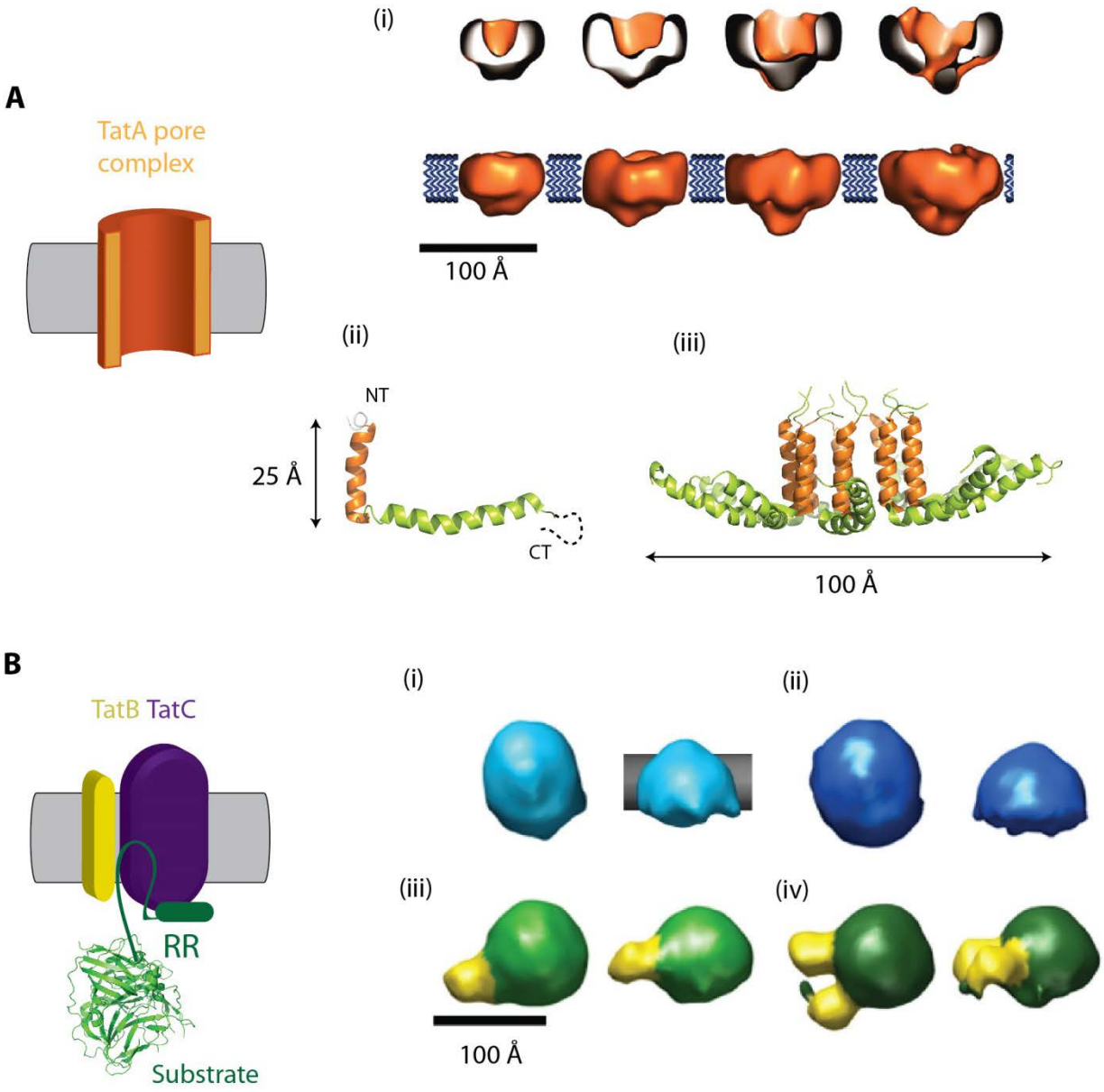


Figure 1.4.1 Structural studies on the Tat system integral membrane components.

(A) The TatA protein is believed to form the protein conducting element of the Tat system.

(i) 3D EM reconstructions of *E. coli* TatA extracted and purified from native membranes in the detergent $C_{12}E_9$. The top row shows a slice through representative size classes of the protein, and the bottom row shows the inferred position of complexes in a bilayer (Gohlke *et al.*, 2005). **(ii)** The solution NMR structure of $TatA_d$ from *B. subtilis* purified in the detergent dodecyl-phosphocholine (DPC) where the TM helix is shown in orange, and the amphipathic helix in yellow (Hu *et al.*, 2010). **(iii)** The model of a TatA oligomer (nine subunits) from *E. coli* using solution state NMR techniques (Rodriguez *et al.*, 2013) and coloured as in **(A) (ii).**

(B). The TatBC complex is known to bind the signal peptide on the protein to be transported. A 3D reconstruction by single particle negative stain electron microscopy of the *E. coli* TatBC complex shows small **(i)** and large **(ii)** complexes. TatBC (green) purified with the SufI precursor (yellow) showed either one **(iii)** or two **(iv)** molecules of SufI bound (Tarry *et al.*, 2009b). The structures are shown in two images, rotated by 90°.

structure. This experiment suggests that this arrangement of TatA molecules in a bilayer could perturb the membrane and thus allow passage across the membrane for a substrate protein anchored nearby to TatBC. These experiments may begin to help understand the mechanism of TatA, by providing a model of interactions to test. Furthermore, it may provide an approach to obtain structural information for other weakly associated membrane proteins.

TatBC receptor complex

So far the TatB and TatC proteins have only been structurally characterised by low resolution EM studies of the TatBC complex (Tarry *et al.*, 2009b). The *E. coli* TatBC complex was recombinantly expressed and purified and samples analysed by single particle EM. A random conical tilt approach allowed for the 3D reconstruction of the particles. These maps revealed particles with a thickness slightly larger than that of a membrane, that were roughly symmetrical when viewed from above the predicted membrane. Two slightly different sized classes were seen **(Figure 1.4.1.B (i and ii)**).

The TatBC complex was also co-purified with the *E. coli* SufI protein, an archetypical Tat substrate, the structure of which has been solved to high resolution by X-ray crystallography (Tarry *et al.*, 2009a). Imaging these complexes revealed two additional distinct classes; those that appeared to have one region of extra density compared to the apo structure, and some with two such regions. These extra regions of density fit the crystal structure of SufI, implying that substrate binds to the periphery of the complex. The overall size of the TatBC complex decreased on association with SufI, implying that the binding of substrate seems to introduce a change in the shape of the overall complex. The authors

interpreted the fact that the complex decreases in size, yet is not more densely packed, as being due to the loss of TatBC subunits from the complex.

These low resolution images fail to give any useable information on the structures of the TatB and TatC proteins, or how they may be are arranged. Additionally, no molecular details were provided on the nature of the crucial targeting interaction between the signal peptide and the receptor complex. The extensive functional data in the literature on these two proteins (reviewed in (Fröbel *et al.*, 2012b)) were therefore still not able to be rationalised. Thus, high resolution structural detail is needed on the TatBC complex.

It has been shown that within the TatBC complex, TatC and TatB can each form structurally autonomous, stable, and defined units (Orriss *et al.*, 2007). Thus, in an analogous manner to the high resolution structure of the stable, monomeric form of the Sec translocon (Van den Berg *et al.*, 2004), we considered that breaking the TatBC complex into its constituent, stable sub-components and structurally tackling these alone may prove more tractable than tackling both components together.

Of the two sub-components, TatC is thought to form the ‘core’ of the TatBC complex (Orriss *et al.*, 2007). TatC is the most highly conserved of all the Tat components (Yen *et al.*, 2002), it forms the primary signal peptide binding site within the TatBC complex (Alami *et al.*, 2003) and it is likely that the binding site for a single signal peptide is located within a single molecule of TatC (Tarry *et al.*, 2009b, Celedon & Cline, 2012). TatC can also insert the signal peptide into the membrane independently of TatB or the PMF (Fröbel *et al.*, 2012a). Furthermore, TatC acts as the platform within the Tat translocase, onto which the other two membrane components of the system can bind (Mori & Cline, 2002). These factors make TatC an attractive, and important, structural target.

TatC in *E. coli* has also been the subject of an array of studies aimed at understanding residues important for function, by genetic screening (Kneuper *et al.*, 2012) or targeted substitutions(Buchanan *et al.*, 2002, Holzapfel *et al.*, 2007), which localise across the sequence. Therefore, a TatC structure would provide a basis to rationalise these data and give insight into the transport mechanism. The structure of TatC could be combined with the various data on how TatC interacts with the signal peptide, and TatB, and TatA, (Zoufaly *et al.*, 2012) to allow for a model of the Tat translocase, which would otherwise be very difficult to study structurally, to be constructed by a 'bottom up' approach.

1.5 Identification of a TatC construct for structural studies

The technique of X-ray crystallography was selected to obtain a high resolution structure of the TatC component since at around 29kDa in mass the protein is quite large in size for study by traditional solution state NMR techniques (Yu, 1999) even before any additional mass contribution from detergents is taken into account.

As introduced in the previous section, the first major bottleneck to structural characterisation of membrane proteins is achieving useable level of expression. The difficulties in over-expression have led to the development of new bacterial expression strains which are better able to produce membrane proteins. C41 (DE3) and C43 (DE3) strains are BL21 derivatives selected on the basis that they were resistant to death upon induction of a previously toxic membrane protein. The C41 and C43 cells were later genetically characterised, and were shown to have mutations in the *lacUV5* promoter which dictates the expression of the T7 RNA polymerase used to drive expression of the recombinant membrane protein (Miroux & Walker, 1996). This observation has led to the engineering of the 'Lemo' strains, in which the activity of the T7 RNA polymerase can be

more precisely controlled by its inhibitor, T7 lysozyme (Wagner *et al.*, 2008). Subsequent strains have been developed, including the Lemo56 strain (Tarry *et al.*, 2012), showing improved expression of more membrane proteins compared to that of the original Lemo21 strain. (unpublished data, Biomembrane institute, University of Stockholm).

Alongside the development of *E. coli* strains for expression, new expression hosts for membrane proteins are emerging. The bacterium *Lactococcus lactis* has been successfully used for a range of prokaryotic and eukaryotic membrane proteins (Kunji *et al.*, 2005), and some targets were shown to be more resistant to degradation in this bacterium compared to *E. coli*. Yeast systems using strains such as *Saccharomyces cerevisiae* and *Pichia pastoris* are being increasingly used for membrane protein expression. The use of the *Saccharomyces cerevisiae* system is being aided by the attempts to understand the key players involved in membrane protein synthesis pathways in this organism (Bonander *et al.*, 2005), and exploiting the knowledge of such pathways to increase the yield of expression.

Even with the advances in the development of more optimal expression strains, not all membrane proteins will express to useable levels. To increase the chance of finding a construct that is amenable to over-expression, sequences of the membrane protein from different homologs can be tested. In the case of TatC, a protein that is present in most bacteria and archaea (Bogsch *et al.*, 1998), there was a wealth of sequences from which to select. Furthermore, as the TatC sequence is well conserved (Yen *et al.*, 2002), a structure of one such sequence would allow for the interpretation of the functional data carried out in *E. coli*, and the conclusions derived from one TatC structure can be applied to other bacterial TatC proteins.

As seen for the Sec system, membrane and soluble proteins from archaea and thermophilic bacteria may be able to crystallise more efficiently. As discussed in the previous section, the improvement in crystallisation may be due to their additional stability (Wiener, 2004), a known determinant for obtaining well ordered, and well diffracting crystals (Sonoda *et al.*, 2011). To aid in the structural characterisation of TatC, over 30 TatC proteins from thermophilic bacteria were chosen for expression screening. For each of these constructs the various membrane protein-optimised strains, along with a range of expression conditions would ideally be tested. In order to efficiently test expression parameters in a high throughput manner a GFP-fusion system had been developed (Drew *et al.*, 2001). The sensitivity of the detection of the levels of GFP by fluorescence, as little as 10µg of GFP per litre of culture (Drew *et al.*, 2006), allows for protein expression trials to be conducted on a very small scale.

The GFP-fusion system (**Figure 1.5.1**) is appropriate for integral membrane proteins which have their CT inside the bacterial cell, and is thus applicable for TatC, given its confirmed

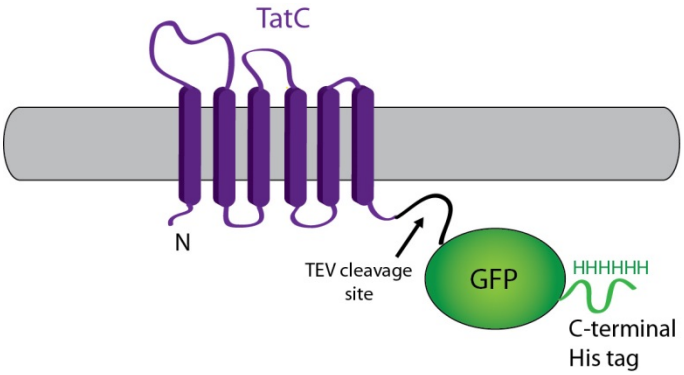


Figure 1.5.1 A cartoon of the GFP fusion constructs.

GFP fusion constructs were created from over 30 different thermophilic TatC sequences for high throughput construct screening. The GFP moiety is located at the CT of the protein, separated from the TatC protein by a TEV protease cleavage site for use in down stream purifications. There is a His tag at the CT of the GFP protein to allow affinity purification by nickel affinity.

topology (Punginelli *et al.*, 2007). The bacteria express the membrane protein, and only if this folds properly and is correctly inserted into the membrane will the GFP moiety also be expressed, and will give fluorescence. If the protein is mis-folded and targeted to inclusion bodies the GFP will no longer fold and thus will give no signal. The GFP fluorescence of a cell culture will therefore give a direct read out of the level of correct expression of the construct. The construct used also contains a His-tag at the CT of the GFP, to allow for nickel affinity purification and a TEV protease site between the GFP and TatC to allow for removal of the GFP_{his} domain.

At the start of this project, GFP fusions of over 30 thermophilic TatC sequences were created by insertion of the protein encoding sequence between the XhoI and BamHI sites of the GFP-fusion vector pWaldo (**Figure 1.5.2**). The molecular biology to create the constructs was performed by Dr. James Graham and Donna Bryan (Department of Biochemistry, University of Oxford). Dr. James Graham, along with Dr. Mike Tarry and Mari Jääskeläinen (Biomembrane Unit, University of Stockholm) screened these constructs using GFP fluorescence to identify those amenable to over-expression.

The expression screening identified the *Aquifex aeolicus* TatC protein (sequence in **Appendix 8.3, Table 8.3.1**) as giving good overexpression in the Lemo56 (DE3) strain under certain expression conditions. Preliminary detergent screening (Dr. M Tarry, Biomembrane Unit, University of Stockholm) indicated this protein was suitable for structural studies. Having initially identified this construct, further work would need to be carried out before high resolution crystals would be obtained.

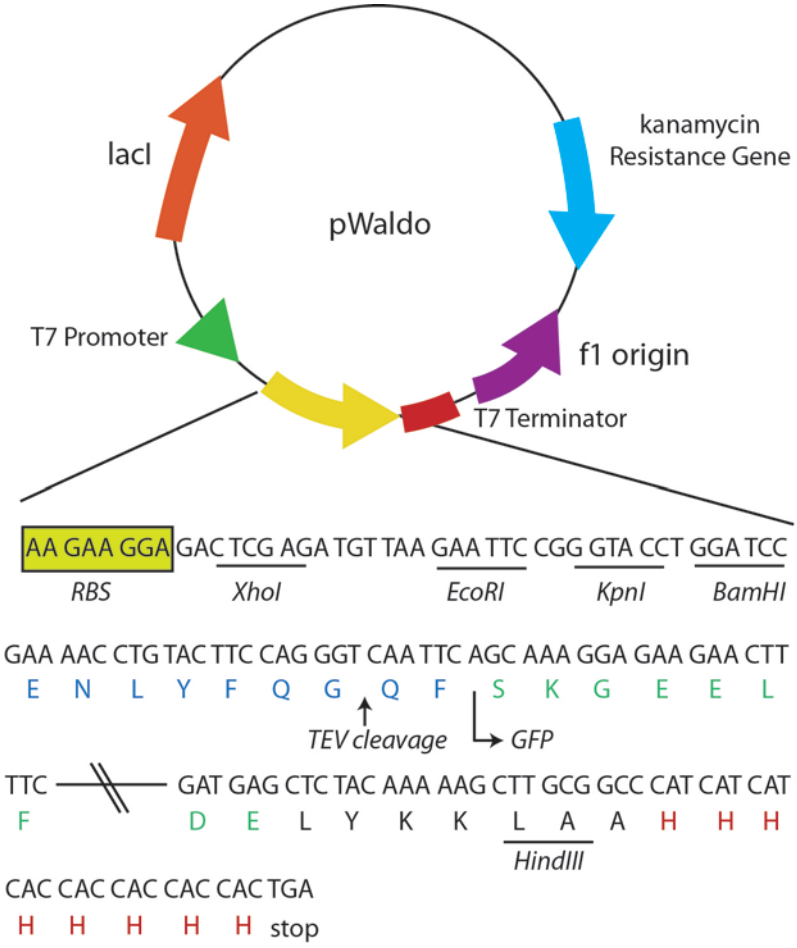


Figure 1.5.2 The pWaldo vector map and multiple cloning site.

1.6 Thesis Aims

The main aim of this dissertation is to build on the preliminary work with the *A. aeolicus* TatC construct to solve the structure of TatC, and allow for the interpretation of the abundance of literature in the field to understand how this protein may carry out its vital roles within the Tat system; specific signal peptide recognition and binding the other Tat components. To these aims, the following approach was taken:

- Carry out detergent screening using an initial list of promising detergents identified by Dr. M Tarry (Biomembrane Unit, University of Stockholm), to identify the best detergents for use in crystallisation - **Chapters 2 and 3**

- Intensive crystallisation trials of the most promising samples with a view to gaining high resolution diffraction data - **Chapters 2 and 3**
- Establish a selenomethionine expression protocol to give 100% incorporation of selenomethionine into TatC, to allow the diffraction data to be phased - **Chapter 3**
- To analyse the structure of TatC in the light of published functional data relating to signal peptide binding, and establish a functional assay to enable the dissection of signal peptide recognition by TatC – **Chapter 4**
- Attempt to gain a co-crystal structure of TatC and the signal peptide in complex, with a view to understanding the molecular details of this interaction - **Chapter 5**
- To analyse the structure of TatC in the light of published functional data relating to the interactions of TatC with the two other membrane components of the Tat system, with a view to produce a model for how these interact with TatC - **Chapter 6**
- Create TatB expression constructs and establish a high-throughput screening protocol to identify the optimal ones with a view to the structural characterisation of the TatBC complex at high resolution - **Chapter 7**

Chapter 2: *Aquifex aeolicus* TatC detergent screening and crystallisation trials

A TatC-GFP_{his} construct using the TatC protein from the thermophilic bacterium *A. aeolicus* had been previously identified, by Dr. M Tarry and Mari Jääskeläinen, as giving good over expression in *E. coli* L56 (DE3) cells (Biomembrane institute, Stockholm). This expression screening over-came the first bottle neck to membrane protein structure determination; gaining a good level of construct expression. The next step and second major bottleneck in membrane protein structural biology is to obtain a purified sample of the membrane protein in a stable, and mono-disperse state. This state is most often achieved by use of detergents. The stability of the protein and the success of obtaining crystals can vary depending on the detergent used during the purification (Tate, 2010).

Detergents can be used to solubilise the bacterial membranes and release the membrane protein of interest from the membrane for purification. Detergent molecules can vary significantly in their architecture, but are always amphiphatic and generally consist of a polar headgroup and a hydrophobic tail. An example of two detergents used in this study are given in **Figure 2.1**. The headgroup can vary greatly between different detergents, from maltose groups (shown in the figure) to ionic groups, such as in sodium dodecyl sulphate (SDS). The hydrophobic chain usually consists of a carbon acyl chain, which can differ in both length, bonding, and modifications, such as the ring present on the chain of the cymal class of detergent (**Figure 2.1 B**). The differences in structure alter how detergents behave in aqueous solutions. A key property detergents all share is that they are able to exist as monomers in solution but also as self ‘aggregates’ containing multiple molecules, known as micelles (Garavito & Ferguson-Miller, 2001). In the aqueous buffers used for protein

purifications, the hydrophobic effect drives detergents to arrange themselves in micelles so as to shield their hydrophobic regions from the aqueous solution and expose their hydrophilic groups. The capacity to form a micelle increases as the concentration of the detergent increases. Beyond a certain concentration, referred to as the critical micelle concentration (CMC), micelles are able to form. Increasing the detergent concentration above the CMC the concentration of micelles increases and the concentration of monomers remains constant. The CMC is dependent on the structure of the detergent, for example longer chain detergents from the same family tend to have a lower CMC. The CMC may also dependent on factors such as the ionic strength of the solution and temperature (Helenius & Simons, 1975). For use in membrane protein purifications, detergents must be used at a concentration above their CMC so that the detergent has the ability to form micelles around the membrane protein. For solubilisation, the concentration used may be many times above the CMC.

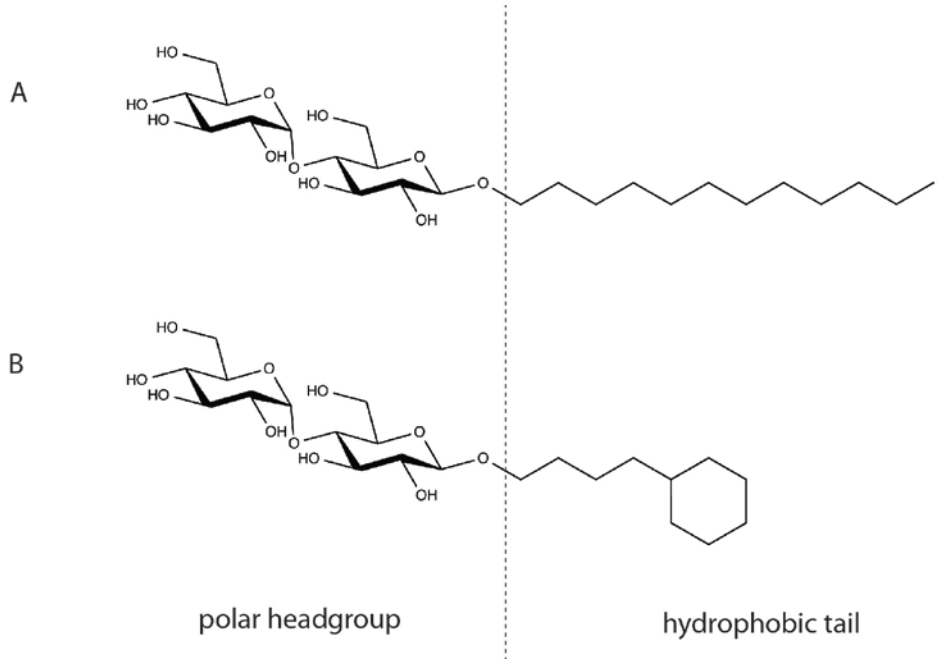


Figure 2.1 The structures of two detergents used in this study.

(**A**) n-dodecyl β-D maltopyranoside (**B**) cymal-4

Detergents solubilise membrane proteins from the membrane by partitioning into and then disrupting the membrane, which releases the membrane proteins. The hydrophobic transmembrane regions of the membrane protein become surrounded by the hydrophobic region of the detergent, encapsulating them into a membrane protein-detergent micelle. Shielding the hydrophobic regions of the protein can help to prevent aggregation. However, the efficiency with which a given detergent is able to solubilise a membrane protein out of the membranes is protein specific, detergents that efficiently extract one protein may not extract another. The reasons for these differences are not fully understood. (Arnold & Linke, 2001). Furthermore, even if a membrane protein is efficiently solubilised, it may still undergo aggregation. Aggregation tends occur more often for a membrane protein purified in so-called ‘harsh’ detergents, a term describing those detergents with shorter acyl chains which cover the hydrophobic transmembrane regions less than ‘milder’ detergents with longer acyl chains. Detergents with smaller, highly charged headgroups are also generally more denaturing, and so classified as harsh.

If there is no aggregation of the membrane protein in a short chain detergent, it is though that a protein purified in these detergents may be more amenable to crystallisation. Statistics show that, on average, α-helical membrane proteins crystallised in detergents with 12 carbon atoms in the acyl chain give lower resolution structures than those in shorter chain detergents (Sonoda *et al.*, 2010). Given the detergent ‘belt’ surrounding the hydrophobic centre of the membrane protein, crystals contacts between the central, transmembrane segments are less likely to occur and the majority of crystal contacts in crystals grown in the standard crystallisation methods are between the hydrophilic regions of the protein. These regions include the loops and any extra-membranous domains. However, the detergent micelle can cover not just the hydrophobic transmembrane regions of the membrane protein, but it can also occlude the hydrophilic loop regions hampering

their ability to form these crystal contacts. There is some evidence from comparisons of the sizes of micelles around the mitochondrial ATP/ADP carrier that shorter chain detergents form a micelle of smaller radius around the membrane protein (Kunji *et al.*, 2008), and this then leaves more of the proteins hydrophilic loop regions exposed out of the micelle. In general, this could be advantageous for crystallisation, as it allows more of the protein to potentially come into contact with itself and form crystal contacts, during the process of forming a lattice (Bill *et al.*, 2011). If most of the protein is buried within a micelle, as is predicted for the longer chain detergents, it may hamper the formation of the crucial crystal packing contacts. In the case where a short chain detergent has an extremely denaturing effect on a membrane protein, thermostabilising mutations of the membrane protein have been used to prevent the proteins aggregation. For example, the β_1-adrenergic receptor structure was solved after thermostabilisation allowed for the protein to be stably exchanged into the short chain detergent octylthioglycoside (OTG) for crystallisation, a detergent which proved extremely denaturing for the native protein (Warne *et al.*, 2008). In addition to allowing the use of the OTG detergent, the thermostabilising mutations may keep the receptor in a single conformation, which may have aided the crystallisation.

Membrane proteins, like their soluble counterparts, often oligomerise as part of their function. Detergents may solubilise a variety of oligomerisation states, and thus give a highly polydisperse sample (Privé, 2007). The extent of both protein aggregation in a certain detergent and some insight into the monodispersity of a solubilised GFP-tagged sample can be initially estimated in a high throughout manner by fluorescence size exclusion chromatography (F-SEC) (Drew *et al.*, 2006). This technique makes use of the GFP tag on the protein of interest (**Figure 1.4.1**) that was initially used in the HT screening to identify the *A. aeolicus* construct as giving very good expression. Very small amounts of detergent solubilised crude membranes can be injected onto a size exclusion column

with an in-line fluorescence detector attached and the GFP fluorescence monitored (Kawate & Gouaux, 2006). Inspection of the fluorescence trace allows quick analysis of the state of the solubilised protein. The amount of protein running in the void of the size exclusion column to gives an estimate of the extent of aggregation and the shape of the peak gives an indication of the monodispersity of the sample. F-SEC can be repeated for protein solubilised in a variety of detergents to provide insight into those detergents which best stabilise a particular membrane protein of interest.

F-SEC gives a crude measure of membrane protein stability. There are a growing number of assays that can quantitatively membrane protein stability. Examples include the use of a radiolabelled ligand of a membrane protein in binding assays. If the solubilised membrane protein binds with the same affinity as the native protein within a membrane it gives evidence that the protein is stable. The assays can be repeated at increasing temperatures, to give a measure of the thermostability of the protein (Serrano-Vega *et al.*, 2008). These assays require very little, unpurified protein and are quick to perform. In the same manner as F-SEC, the assay can be repeated in different detergents to determine which detergents give the most stable sample. A fluorescence based unfolding assay has been developed which allows for a membrane proteins stability in different detergents to be measured (Alexandrov *et al.*, 2008, Sonoda *et al.*, 2011).

F-SEC was carried out previously by Dr. M Tarry (University of Stockholm) for the *A. aeolicus* TatC-GFP$_{His}$ construct. This experiment resulted in an initial list of promising detergents shown in **Table 2.1** which showed both efficient solubilisation of TatC-GFP and good monodispersion of the protein sample. Two more recently available detergents were used as part of this study are also included in **Table 2.1**.

Out of the detergents identified in the screening, dodecyl-maltopyranoside (DDM) is the

Detergent		CMC (mM) in 0.2M NaCl	% in purification buffers (w/v)
1	n-Dodecyl β-D maltopyranoside (DDM)	0.12	0.03
2	n-Undecyl β-D maltopyranoside (UDM)	0.59	0.08
3	n-Decyl β-D maltopyranoside (DM)	1.6	0.2
4	n-Nonyl β-D maltopyranoside (NM)	6	0.5
5	Cymal-4	7.3	0.8
6	Undecyl-N,N-Dimethylamine-oxide (UDAO)	3.21	0.15
7	Dodecyl-N,N-Dimethylamine-N-oxide (LDAO)	0.14	0.1
8	Lauryl maltose neopentyl glycol (LMNG)	0.01	0.02
9	Decyl maltose neopentyl glycol (DMNG)	0.036	0.02

Table 2.1 Detergents identified in initial F-SEC studies as promising candidates for *A. aeolicus* TatC purifications. Rows 1-7 were identified by Dr. M Tarry (University of Stockholm) and 8-9 identified in this study. Critical micelle concentrations (CMCs) are given along with the concentrations used in purification buffers.

most widely used detergent for membrane protein crystallography (Newstead *et al.*, 2008). DDM is a mild detergent with 12 carbon atoms in the acyl chain (**Figure 2.1 A**) which has been shown to retain the activity of most of the membrane proteins that it is used to solubilise and purify (Privé, 2007). This detergent formed the basis of the preliminary work on the *A. aeolicus* TatC construct. An initial TatC purification carried out by Dr. M Tarry in dodecyl-maltopyranoside (DDM) (Anatrace) gave pure protein which was subjected to initial crystallisation trials. This yielded some preliminary crystal hits, of very small, fine hair-like crystals. The most promising of these were in the MemSys/Memstart broad screen (Molecular dimensions) condition B8 which contained 0.1M magnesium chloride, 0.1M sodium acetate pH 4.6, 30% v/v PEG 400. These diffracted to very low resolution (~20Å), which indicated they were protein crystals. These crystal hits would, however, need to be optimised in order to collect the higher resolution data (<4Å) that would enable the structure to be solved.

2.1 Purification in dodecyl-maltopyranoside and optimisation of the initial crystal hits

TatC was purified in DDM in order to try and optimise the crystal hits. An initial protocol for the growth and purification was optimised as part of this study, and outlined below. 50µl aliquots of L56 cells were transformed with 1µl of TatC-GFP_{His} pWALDO (kanamycin resistance) according to the protocol in **Appendix 8.4 Table 8.4.1**, and an overnight culture grown in LB from one colony of the transformation reaction. A 1:1000 dilution of this culture was used to inoculate 1L Terrific Broth (TB) (Fisher), supplemented with 50µg/ml kanamycin and 8ml of 50% (v/v) sterile glycerol. 1L of cells were grown in a 3L flask at 37^{o}C with shaking at 180 rpm for 9 hours until the culture reached an optical density at 600nm (OD_{600}) of 4.0-5.0. The cells were induced with 0.1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16 hours at 24^{o}C.

6L of cells were harvested at 4,500 rpm in a JLA 8.100 rotor, and frozen at -20^{o}C for storage. Prior to use, cells were thawed at 4^{o}C overnight. Cells were re-suspended in 130ml PBS (recipe in **Appendix 8.4 Table 8.4.2**) supplemented with a protease inhibitor tablet (Roche) and a few granules of both DNAse and lysozyme (Sigma). Cells were re-suspended by mixing in a beaker with a magnetic stirrer on a stirrer plate for 40mins, and then homogenised three times in a hand held homogeniser before passage five times through an Emulsifex C5 (Avestin) cell disrupter at 4^{o}C.

The lysate was clarified by centrifugation at 15,000 rpm in a JA 25.50 rotor (Beckman). The soluble fraction was taken and subject to ultra centrifugation at 50,000 rpm in a Type 70 Ti rotor (Beckman) to pellet the membrane fraction. The supernatant was discarded, and the membranes re-suspended in 66ml PBS with a hand held homogeniser. The 66ml of membranes were then solubilised for two hours at 4^{o}C in 66ml of 2x solubilisation buffer (PBS supplemented with 40mM imidazole (Sigma) and 2% (w/v) DDM added as a solid).

To remove any un-solubilised proteins the sample was ultra-centrifuged at 50,000 rpm in a Type 70 Ti rotor (Beckman). The soluble fraction was then loaded on to a 5ml Ni-NTA column (Qiagen) pre-equilibrated in PBS at a flow rate of 1ml/min and the flow-through loaded again to allow as much his-tagged protein to bind as possible. The column was then washed with 5 column volumes (CV) of PBS supplemented with 25mM imidazole and 0.03% (w/v) DDM, then with two column volumes of the same buffer but containing 40mM imidazole. The bound protein was then eluted from the column with the same buffer supplemented with 250mM imidazole. Green fractions containing TatC-$GFP_{his,}$ as confirmed by SDS-PAGE (**Figure 2.1.1 A**, nickel column fractions) (see **Appendix 8.4 Table 8.4.3** for SDS-PAGE protocol) were pooled, and incubated overnight with 2mg of his-tagged TEV protease (for preparation protocol, see **Appendix 8.4.1**) along with 5mM sodium citrate pH 6.3, 0.3mM oxidised glutathione and 3mM reduced glutathione (added from stock solutions of 1M sodium citrate, pH 6.3, 50mM oxidised and 150mM reduced glutathione). TEV protease is sensitive to high concentrations of imidazole, and so the reaction was carried out in 10kDa molecular weight cut-off (MWC) snakeskin dialysis tubing (Fisher) and dialysed against 20mM Sodium Hepes pH 7.5, 150mM NaCl and 0.03% w/v DDM.

The cleavage reaction was centrifuged at 4,000 rpm in a bench top centrifuge to remove precipitated TEV protease. Complete cleavage was confirmed by reducing SDS-PAGE (**Figure 2.1.1 A)** and the reaction mix loaded on to a pre-equilibrated 5ml Ni-NTA column for reverse purification. The his-tagged TEV and GFP bound to this column, leaving pure TatC in the flow-through. This was again confirmed by SDS-PAGE (**Figure 2.1.1 A**). The flow-though was concentrated to 5ml using 100kDaMWC centrifugal concentrator devices (Millipore), and subjected to size exclusion chromatography (SEC) by injection on to an S200 16/60 size exclusion column (GE healthcare) pre-equilibrated in 20mM Hepes pH

7.5, 150mM NaCl and 0.03% w/v DDM. The elution trace is shown in **Figure 2.1.1 B**. Fractions across the peak were pooled and concentrated to ~6mg/ml using the 100kDaMWC centrifugal devices. The yield of TatC in DDM was 1.5mg from 6L and the final sample of TatC was free of contaminating proteins (**Figure 2.1.1 C**) and so ready for crystallisation trials.

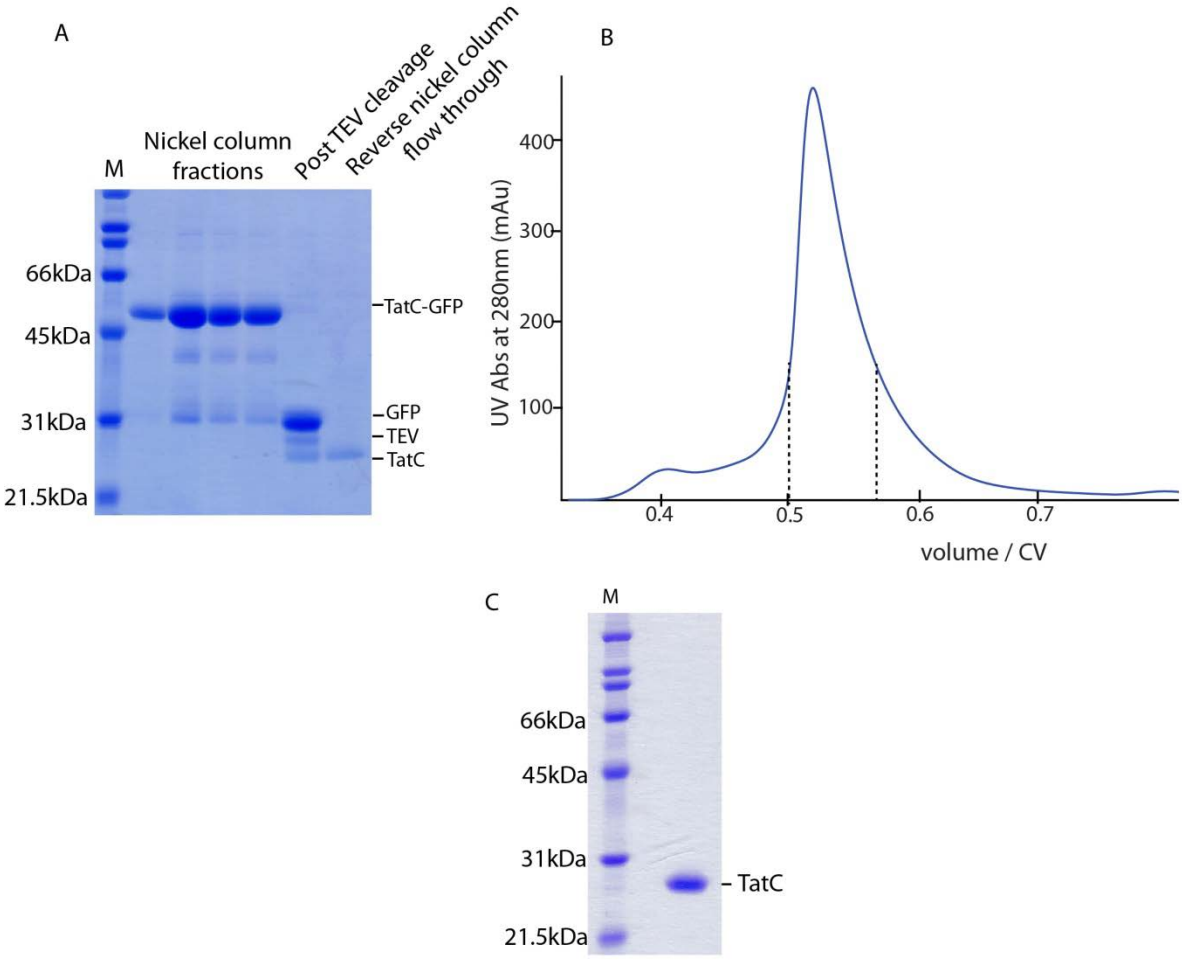


Figure 2.1.1 The purification of TatC in DDM.

(A) 12% (w/v) polyacrylamide SDS-PAGE showing the initial stages in the purification of TatC from L56 (DE3) cells. **(B)** The S200 16/60 size exclusion profile of TatC in DDM post reverse Nickel column. The dashed lines indicate the region of the peak pooled for concentration **(C)** 12% SDS-PAGE showing the final, purified TatC sample. M: Molecular weight standards, Broad range (Biorad).

Crystallisation of a protein occurs when a protein undergoes precipitation from a saturated solution and forms an ordered crystalline lattice. The extent to which the protein precipitates is dependent on factors within that condition, such as the concentration of the salts, PEG (which often acts as the precipitant within the condition) and the pH of the buffer, as well as the concentration of the protein (McPherson, 2001). Once a condition which gives crystallisation has been identified, these parameters can be slightly altered, to attempt to grow crystals which may diffract better. In order to try to reproduce and further optimise the initial crystallisation hit identified by Dr M. Tarry, a 96 condition optimisation tray was made by hand which varied the concentration of PEG, salt and the pH of the buffer (DDM Optimisation screen 1, **Table 2.2**).

For purified membrane proteins, the approaches to produce 3D crystals can be divided into two major categories, ‘in detergent’ and ‘in cubic lipid phase’ (CLP). Currently, most membrane protein structures solved to date are from the ‘in detergent’ method. This method uses detergent purified protein, as discussed above, and sets up crystallisation trials with the membrane protein-detergent mixture as if it were a soluble protein. However, there are many drawbacks with using this method. As discussed above, the reliance (in general) on the hydrophilic loop regions to make the majority of the crystal contacts may be occluded by the micelle and will hamper crystal formation. In addition, the native membrane can exert lateral pressure on membrane proteins and reduce their conformational flexibility. When a membrane protein is in detergent the constraint is reduced and the protein may be more likely to exist as a heterogeneous mixture of conformational states not amenable for crystallisation (Fethiere, 2007).

The second crystallisation method, CLP, is becoming an increasingly popular and accounts for around 100 membrane protein structures in the PDB (~10% of total membrane protein structures) from an increasing variety of membrane protein types, including two new

GPCR structures (Rosenbaum *et al.*, 2011, Rasmussen *et al.*, 2011). The CLP method exploits the ability of certain lipids, such as monoolein, to form a 3D lipid array which contains aqueous channels running through the structure. This phase is known as the cubic phase (Caffrey, 2009). This phase is dynamic and can undergo phase transitions during the process of crystal growth. This crystallisation method firstly reconstitutes membrane proteins into the 3D cubic phase. The reconstitution removes the membrane protein from the potentially unstable environment of the detergent micelle, and places it in a more native membrane-like environment. This environment may help to increase the protein stability, and reduce conformational flexibility. Following membrane protein reconstitution, the cubic phase is exposed to conditions that promote crystallisation. It is speculated that the cubic phase can aid the growth of membrane protein crystals by allowing the lateral diffusion of membrane protein molecules through the 3D array of lipids to a nucleating crystal site (Caffrey & Cherezov, 2009). The contacts between membrane protein molecules in a crystal grown in CLP form between the transmembrane regions of neighbouring molecules, and additional contacts between the hydrophilic domains allow stacks to form on to of one another to create a 3D crystal (Caffrey *et al.*, 2012). These types of crystals are referred to as Type I crystals, and all example of CLP membrane protein crystals to date are of this type. CLP crystallisation trials were attempted for *A. aeolicus* TatC purified in lauryl maltose neopentyl glycol (purifications will be discussed in **Chapter 3**), but CLP was not successful in yielding crystals. All crystallisation in this thesis will from now on refer to the ‘in detergent’ method.

In order to carry out the ‘in detergent’ crystallisation method with purified TatC in DDM, the sitting wells of a MRC crystallisation plate (Molecular Dimensions) were coated in 1.5µl of electronic grade coating (EGC) 1700 solution (Griener 3M) in order to help form the small drops from the detergent-protein solution. A Hydra 96 liquid handling robot

(Alpha biotech) was used to transfer 75µl of the mother liquor solution from the 96 condition screen across to the reservoir well of the 96 MRC crystallisation plate. Protein and mother liquor were then pipetted into the drops next to the reservoir with the Oryx Nano crystallisation robot (Douglas instruments), before the tray was sealed with StarSeal Advanced polyolefin film (Starlab). This is referred to as the 'sitting drop vapour diffusion' method. The presence of protein and buffer in the drop increases the water concentration in the drop and water will evaporate from the drop to the mother liquor reservoir. The evaporation increases the protein and precipitant concentration in the drop, and protein crystal nucleation can occur (Asherie, 2004).

DDM micelles have a reported molecular weight of around 66kDa (Strop & Brunger, 2005). To try and facilitate their removal during the protein concentration process, centrifugal concentrators with a filter size cut off of 100kDaMWC were used. This size cut off may allow the free micelles to pass through the filter but will retain the protein-detergent complex. In order to test whether this is the case, a refractive index detector coupled to a size exclusion column can be used to observe a peak corresponding to free micelles within the sample. Alternatively, a colorimetric detergent assay could be used to check for how much excess detergent is present after concentration compared to the protein:detergent complex alone (Urbani & Warne, 2005). To try to ensure the removal of any detergent micelles and monomers, an equimolar ratio of β-methyl cyclodextran (MBCD) (Sigma) to the detergent concentration was added to the purified protein immediately prior to setting up the crystallisation drops. MBCD has been hypothesised to help improve membrane protein crystallisation by sequestering detergent monomers present in the final sample, and by removing loosely bound detergent from the protein-detergent complex (Li *et al.,* 2008). MBCB was added from a concentrated (60mM) stock of the MBCD, to ensure the final protein concentration was minimally affected.

			%(v/v) PEG 400						% (v/v) PEG 400					
			28	29	30	32	33	35	28	29	30	32	33	35
			1	2	3	4	5	6	7	8	9	10	11	12
[$MgCl_2$] (M)	A	0.08	0.1M sodium acetate pH 4.6						0.1M sodium acetate pH 4.8					
	B	0.09												
	C	0.1												
	D	0.12												
	E	0.08	0.1M sodium acetate pH 5.0						0.1M sodium acetate pH 5.2					
	F	0.09												
	G	0.1												
	H	0.12												

Table 2.2 DDM Optimisation screen 1. The screen was based around 0.1M magnesium chloride, 0.1M sodium acetate pH 4.6, 30% (v/v) PEG 400. The region of the screen in which crystals grew is outlined in the red box.

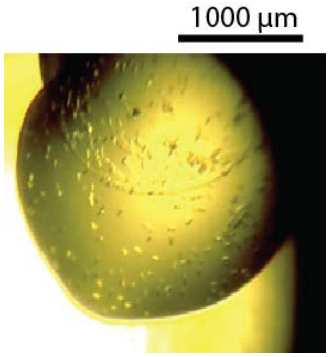


Figure 2.1.2 Crystals obtained in the DDM optimisation screen 1.

Crystals were grown in the region of the optimisation tray highlighted in red in **Table 2**

The protein concentration used to set up the trays, along with the ratio of the protein:mother liquor (25:75, 50:50, 75:25) and the drop size (200nl or 400nl) was varied. Once sealed, the trays were incubated at 21^{o}C.

The optimisation screen yielded crystals under the conditions shown in **Table 2.2**. An example of the type of crystal obtained in this condition is shown in **Figure 2.1.2**. In order to test the diffraction of these crystals, a selection were harvested with Cryoloops

(Hampton research) using the mother liquor solution as a cryo-protectant, and flash frozen in liquid nitrogen. These crystals were exposed to X-rays at the Diamond synchrotron under a cryo stream in order to minimise the radiation damage (Garman & Mitchell, 1996). However, although these crystals were shown to be protein and they did diffract, they did so to very low resolution (>20Å). This could have either been due to their very small size or disorder. No larger crystals were obtained, and so no further optimisation of this condition was attempted.

Subsequent crystallisation trials in DDM yield only low resolution crystals

In order to try to gain better crystals that may give higher resolution data, TatC in DDM was subjected to a further array of broad screen crystallisation trials. TatC in DDM was concentrated to a variety of different concentrations (3, 9 and 12 mg/ml) and high throughput broad screen crystallisation trials set up using the commercially available membrane protein specific crystal screens. These screens each contain 96 different conditions, and thus by using as many of them as possible in parallel should give a large sampling of the potential crystallisation space. A table of the screens used is given in **Table 2.3**. Another parameter that was varied was the temperature with the trays incubated at 21, 12 or 4^{o}C.

These trials yielded crystals in quite different conditions from the original hit, and appeared to exhibit different morphologies, indicating that the protein is now crystallising in new crystal forms. The best ones are shown in **Figure 2.1.3**. However, despite the much larger size of most of these crystals relative to the initial hits (**Figure 2.1.3**), these new crystals all still diffracted to very low (>20Å) resolution. None of the crystal forms could be indexed.

Crystallisation screen	Design Rationale
MemSys (Molecular Dimensions)	A systematic screen of a wide range of pH values. Provides an initial basis for assaying membrane protein solubility (Iwata, 2003).
MemStart (Molecular Dimensions)	Spans conditions successful for 33 high resolution membrane protein structures. Includes salts and PEGs previously found to be successful in the laboratory of Prof. Iwata (Iwata, 2003).
MemGold (Molecular Dimensions)	Set of 96 conditions covering the most recent α- helical and β-barrel protein crystallisation conditions from the PDB (Newstead *et al.*, 2008).
PGA (Molecular Dimensions)	Uses poly-γ-glutamic acid (PGA) polymers as the precipitant. Shown to be suitable for membrane proteins (Hu *et al.*, 2008).

Table 2.3 Broad range crystallisation screens used with the *A. aeolicus* TatC construct

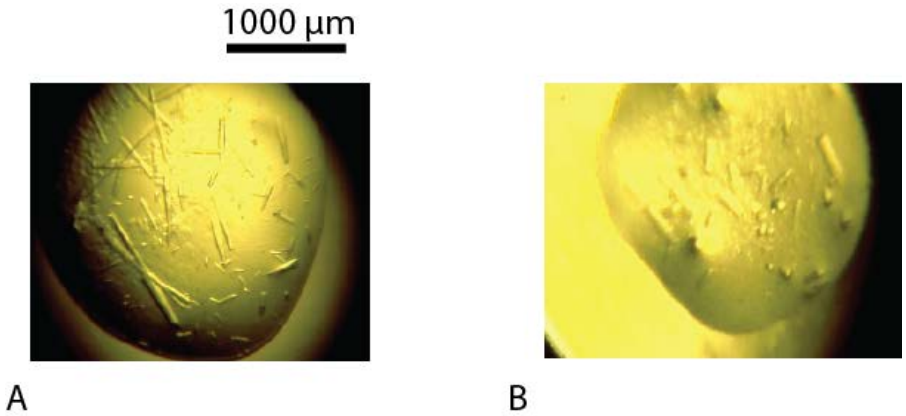


Figure 2.1.3 New crystallisation conditions found in the broad screen for TatC in DDM.

(A) 0.07M NaCl, 0.05M sodium citrate pH 4.5, 22% (v/v) PEG 400 (**B)** 0.02M $MgCl_2$/0.1M KCl, 0.025M sodium citrate pH 4.0, 33% (v/v) PEG 400

As a final attempt to gain better crystals in DDM, purified TatC was subjected to reductive methylation. This technique has been shown to aid crystallisation of proteins that had previously proved difficult to crystallise (Walter *et al.*, 2006). The protocol involves specific methylation of exposed lysine residues on the protein, therefore potentially changing the proteins solubility properties, meaning precipitation may occur under new conditions. In the case of TatC, which has 14 lysine residues, this may alter the way in which the protein can pack together in a crystal lattice and thus may allow a more ordered lattice to form.

The methylation reaction was carried out on the TatC protein after the reverse nickel column and before the final size exclusion step. The protein was diluted to <1mg/ml, and 20µl of freshly prepared 1M dimethylamine borane complex (ABC) and 40µl 1M formaldehyde per ml of protein solution was added. The sample was incubated at 4^{o}C in the dark for 2 hours with mixing on a rotating wheel at 40 rpm. This step was then repeated. The reaction was completed by the final addition of 10µl ABC to each ml of solution, and left mixing overnight at 4^{o}C in the dark. The next day tris buffer at pH 7.5 was added to a final concentration of 20mM to quench the reaction. The reaction was then spun at 13,000 rpm on the bench top centrifuge to remove any precipitated protein, concentrated to 5ml, and injected on the S200 16/60 size exclusion column. The fractions were then treated as for the normal DDM sample.

To test the solubility of the methylated TatC sample compared to the native protein 1µl of the protein was mixed with 1µl of a standard crystallisation condition from the Stura footprint screen (Molecular dimensions). The condition contains 0.2M imidazole malate, pH 7.0 and 20% w/v PEG 4000. This condition represents an ‘average’ of the Stura screen and is useful when determining an appropriate protein concentration at which to begin crystallisation trials (the ‘Stura test’). A protein that is ‘optimally’ concentrated will start to

show signs of precipitation after roughly 1min. For the native protein, this was at a concentration around 6mg/ml (i.e. the initial crystallisation concentration trialled) but for the methylated protein the concentration was around 4.3mg/ml. This indicated that the methylation had altered the solubility of the protein, and the methylated protein was subjected to the broad screen crystallisation trials described as for the native protein.

The crystal trials again gave new crystal forms. An example is given in **Figure 2.1.4**. However, even after optimising these crystals as before by varying PEG, salt, pH, protein concentration and temperature, the diffraction did not improve beyond 20Å.

The lack of improvement in the diffraction quality of the DDM-purified TatC crystals suggested that TatC in DDM was a heterogeneous species. This could have accounted for its inability to pack into a regular lattice. Although the SEC profile for the purification shows a single peak, the TatC species present may not have been a single, well defined species.

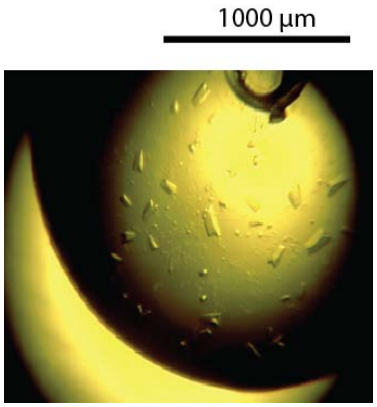


Figure 2.1.4 Crystals of methylated TatC in DDM.

Crystals were grown in 0.05M NaCl, 0.5M sodium phosphate pH 6.2 , 16% (w/v) PEG 4000

2.2 Multi angle laser light scattering (MALLS) and native mass spectroscopy reveals that TatC in DDM is a highly dynamic, oligomeric complex

There is evidence that the *E. coli* TatC protein is present within an *E. coli* TatBC complex as an oligomer (Orriss *et al.*, 2007, Buchanan *et al.*, 2002), with a minimal unit composed of a dimer (Maldonado *et al.*, 2011a). Although the oligomerisation state of *A. aeolicus* TatC was not known, it was possible it could also form higher order species within membranes. It is therefore possible that different detergents may solubilise a certain oligomerisation state of TatC, or possibly a mixture of these.

To try and dissect the nature of the DDM TatC species Steven Johnson at the Sir William Dunn School of Pathology (University of Oxford) carried out Multi Angle Laser Light

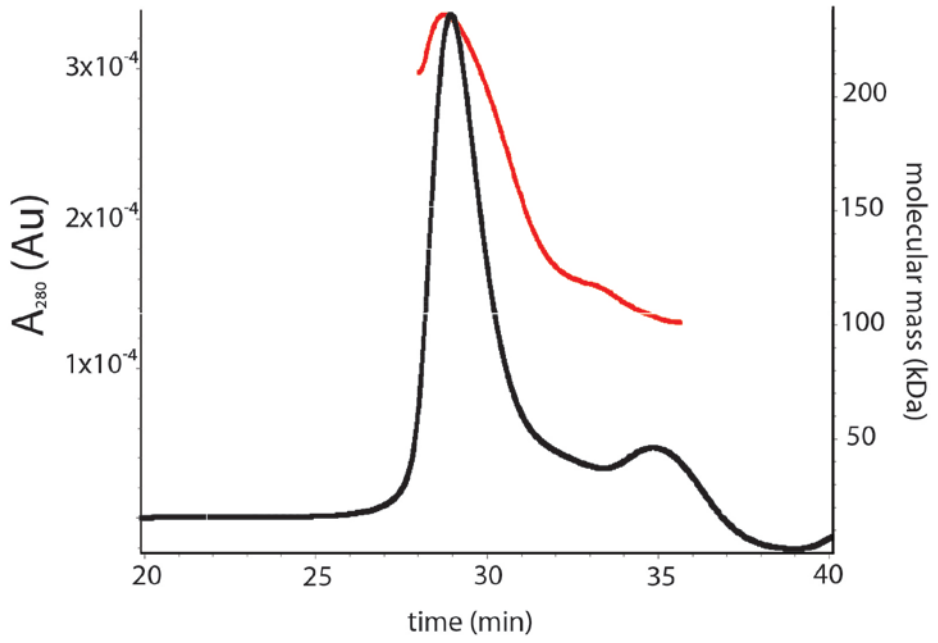


Figure 2.2.1 Assessment of the oligomeric state of TatC in DDM by Multi Angle Laser Light Scattering (MALLS).

The UV absorbance trace of the TatC sample in DDM is shown in black, and the calculated molecular mass is given in red. The left axis gives the UV absorbance, and the right axis the molecular mass, at any given point in the elution profile calculated from light scattering of the sample. Experiment carried out on a 300 µM sample of TatC in SEC buffer with 0.03% DDM.

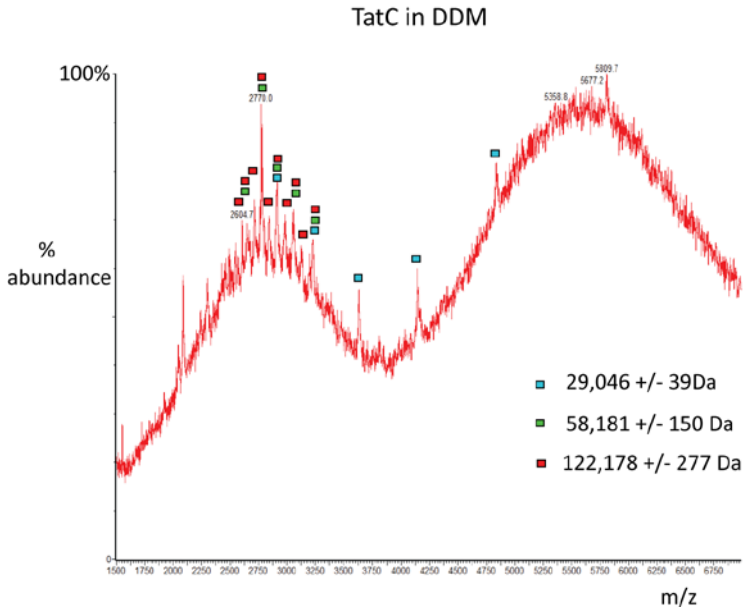


Figure 2.2.2 Native mass spectroscopy of TatC in DDM.

The ionisation states identified for three different oligomeric TatC species are labelled in blue (monomer) green (dimer) and red (tetramer)

Scattering (MALLS) experiments with TatC in DDM (**Figure 2.2.1).** This analysis used a multi angle light photometer (Dawn Heleos, Wyatt) connected to a high pressure liquid chromatography (HPLC) system with an S200 10/300 GL size exclusion column attached.

The calculated mass showed a profile characteristic of an extremely heterogeneous species with the calculated molecular mass of the protein and detergent complex varying across the elution peak from 100kDa to over 220kDa at the higher end. This indicated that TatC was present as an oligomer, but that this complex is potentially not stable, and was dynamic. This resulted in the sample being a mixture of different species.

The oligomeric nature of TatC in DDM was also demonstrated by native mass spectrometry measurements, carried out by Dr Julien Marcoux (Dept of Chemistry, University of Oxford). The spectra (**Figure 2.2.2**) showed a mixture of protein-alone masses consistent with a TatC monomer (29kDa), dimer (58kDa) and tetramer (122kDa).

This fits with the MALLS experiments, which indicated the sample contained TatC in an oligomeric state. The oligomeric heterogeneity shown by MALLS may explain why the DDM sample of TatC gave crystals that were poorly ordered and so diffracted to low resolution.

2.3 Purifications in detergents that result in heterogeneous TatC species were unsuccessful in crystallisation trials

Given that the sample of TatC purified in DDM was highly heterogeneous, it was decided to purify TatC in other detergents and test the homogeneity of the resulting TatC sample. It was hypothesised that a more homogenous sample may increase the chance of giving higher resolution crystals. The initial F-SEC experiments characterised TatC in different detergents and gave a list of detergents which seemed promising, (**Table 1**). In order to test the homogeneity of TatC in each detergent, large scale (6L) preparations of TatC in each of the detergents in **Table 1** was carried out, using the same protocol as in **section 2.1** The same detergent was used to solubilise the membranes and in the downstream purification buffers. Each detergent was used at 1% (w/v) to solubilise the membranes, and then at a concentration dependent upon its CMC and recommended by the manufacturers (Anatrace) (**Table 2.1).** For slight alterations in the preparation protocol for those detergents which had very high CMCs see **Appendix 6.4.2**.

The elution position of a protein from a SEC column depends on the molecular mass of the protein and its shape (Khan, 2004). In the case of a purified membrane protein, its size will not only be determined by its oligomerisation state, but also by the amount of detergent bound. The chromatogram in **Figure 2.3.1** shows a comparison of the SEC profiles for the tested detergents relative to the DDM preparation. The traces showed how much the TatC

elution profile varied depending on the detergent used in the purification. For example, there were two elution peaks present in the DMNG purification (blue trace). SDS-PAGE confirmed that both these peaks contained pure TatC. MALLS analysis showed each of these peaks to correspond to a different, dynamic population of TatC oligomers (results not shown). The profiles of the UDM (cyan) and DM (brown) purifications showed a very similarly shaped elution peak to the DDM sample, but shifted towards a smaller mass. This

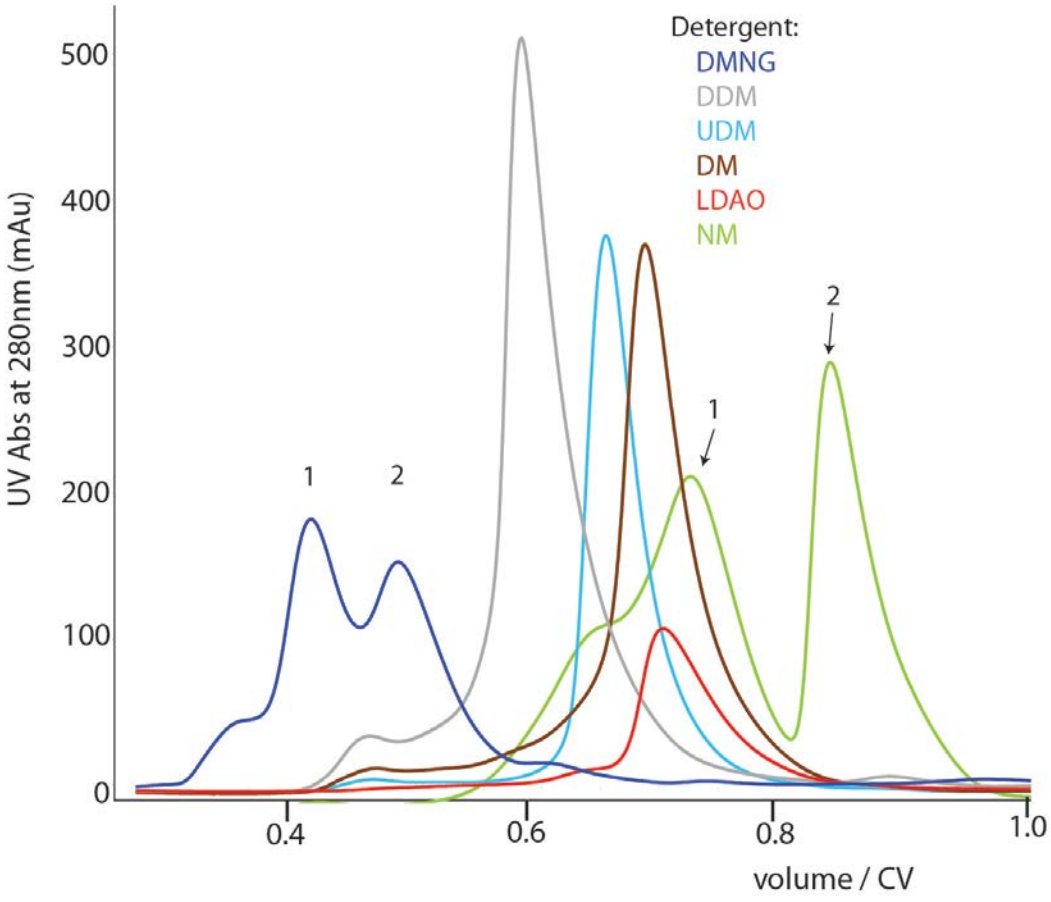


Figure 2.3.1 An overlay of the size exclusion chromatograms in the final step of the TatC purifications for each of the detergents tested.

These traces were all run on the S200 16/60 column after an injection of 5ml of the TatC sample from the reverse nickel affinity step in the SEC buffer containing the appropriate detergent.

is not unsurprising given these detergents are structurally related to DDM, except one or two carbon chain shorter in the acyl chain. Shorter acyl chains causes the size of the micelles the detergent forms to be smaller (Oliver *et al.*, 2013). Extensive broad screen crystallisation trials in UDM and DM detergents, following the same approach as for DDM described in **section 2.1.2,** gave no crystals in any conditions, indicting the TatC species in these detergents is potentially even more heterogeneous than that in DDM.

The purification in NM (green), a detergent that is one carbon chain shorter than DM showed a very different elution profile, however. Firstly, there were two elution peaks, as for DMNG. However, this time the second peak was expected to be free GFP, due to a difference in preparation method that meant the reverse nickel column could not be run until after this size exclusion step (**Appendix 6.4.2**). SDS-PAGE confirmed this was the case (result not shown). The first peak, which contained the TatC, looked to have a shoulder, which SDS-PAGE showed was due to a small amount of uncut TatC-GFP fusion. Once these GFP species were removed by carrying out the reverse nickel column step, the final TatC sample was analysed by MALLS to see if it was similar to DDM.

The MALLS trace for TatC in NM (**Figure 2.3.2**) now contained a micelle peak, as the protein was concentrated with a 30kDa MWC concentrator device which is not large enough to allow the NM micelles to pass through, and peaks for the protein present in the sample. The calculated mass rose rapidly across the protein peaks, from a relatively small mass around 50kDa to above 150kDa at the top end of the peak. The mass seemed to briefly ‘plateau’ at certain points, with masses that could be indicative of dimers or trimers of TatC. However, this is difficult to confirm as the calculated mass trace is considerably sloped, which implies that the sample was very dynamic and interchanging.

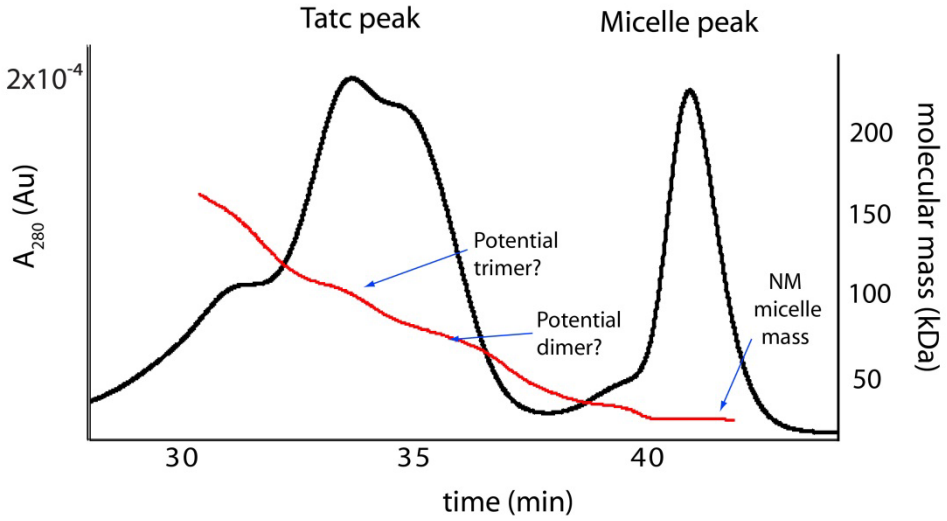


Figure 2.3.2 Multi angle laser light scattering (MALLS) analysis of TatC in NM.

The UV trace is shown in black, and the calculated mass of the total protein-detergent complex across the elution peak given in red, for a sample of TatC at 300µM in SEC buffer with 0.5% (w/v) NM.

The LDAO trace (red) showed a single peak. However, this preparation was never subject to crystallisation trials or MALLS experiments. Any attempt to concentrate it led to the protein precipitating out of solution. It also precipitated at 4^{o}C, overnight. The other homologous detergent identified in the initial screen was UDAO which has a one carbon atom shorter acyl chain than LDAO. Purification in this detergent proved unsuccessful as the protein precipitated on the nickel column. This was also true of purification in the detergent sodium deoxycholic acid which was also tested.

Overall, the preparations in the different detergents showed that the state in which TatC is purified from the membranes is highly dependent on the detergent. All of the detergents discussed so far produced poor protein homogeneity and either poorly diffracting crystals (DDM) or lack of any crystallisation success at all (detergents in **Figure 2.3.1**).

2.4 Purification of TatC in cymal-4 produces a homogenous sample

The purification of TatC in the detergent cymal-4 showed a very different elution profile to those given above. The SEC trace (**Figure 2.4.1**) showed a single peak, which was very symmetrical. This indicated that the TatC species present in cymal-4 may be homogenous. To test this MALLS was run with purified samples of TatC (**Figure 2.4.2**).

For the initial purification of TatC in cymal-4, until the mass of TatC in detergent was determined, 30kDa MWC concentrator devices were used to ensure that no TatC was lost during the concentration steps. This is why the MALLS trace for this sample, as for NM, also had a separate micelle peak, as the micelles are larger than the cut off MW of the concentrator. This then enabled the mass of the micelles to be accurately determined as 30kDa.

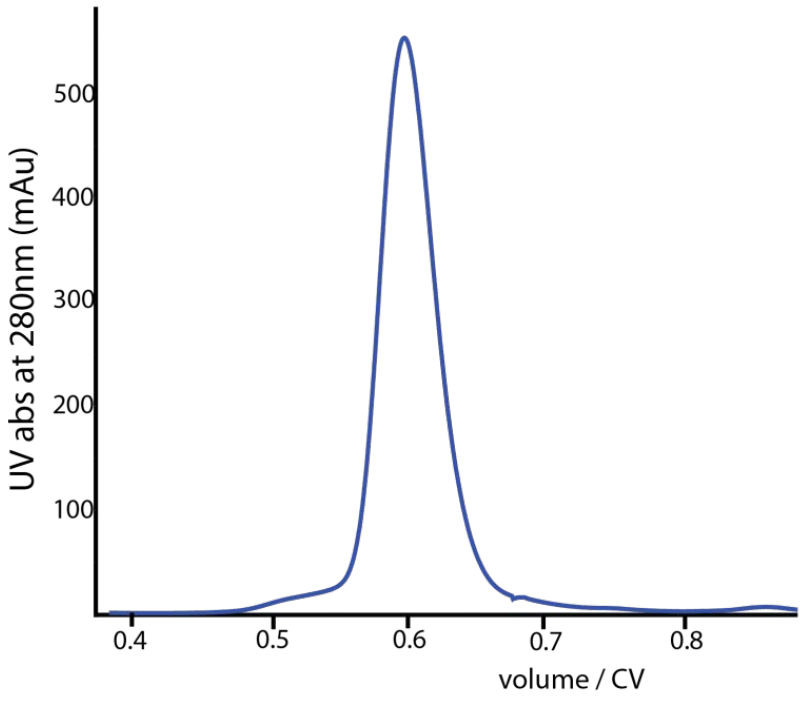


Figure 2.4.1 The size exclusion chromatogram of *A. aeolicus* TatC in cyaml-4 .

1ml of TatC was injected onto an S200 10/300 size exclusion column pre equilibrated in 20mM hepes pH 7.5, 150mM NaCl and 0.8% (w/v) cymal-4

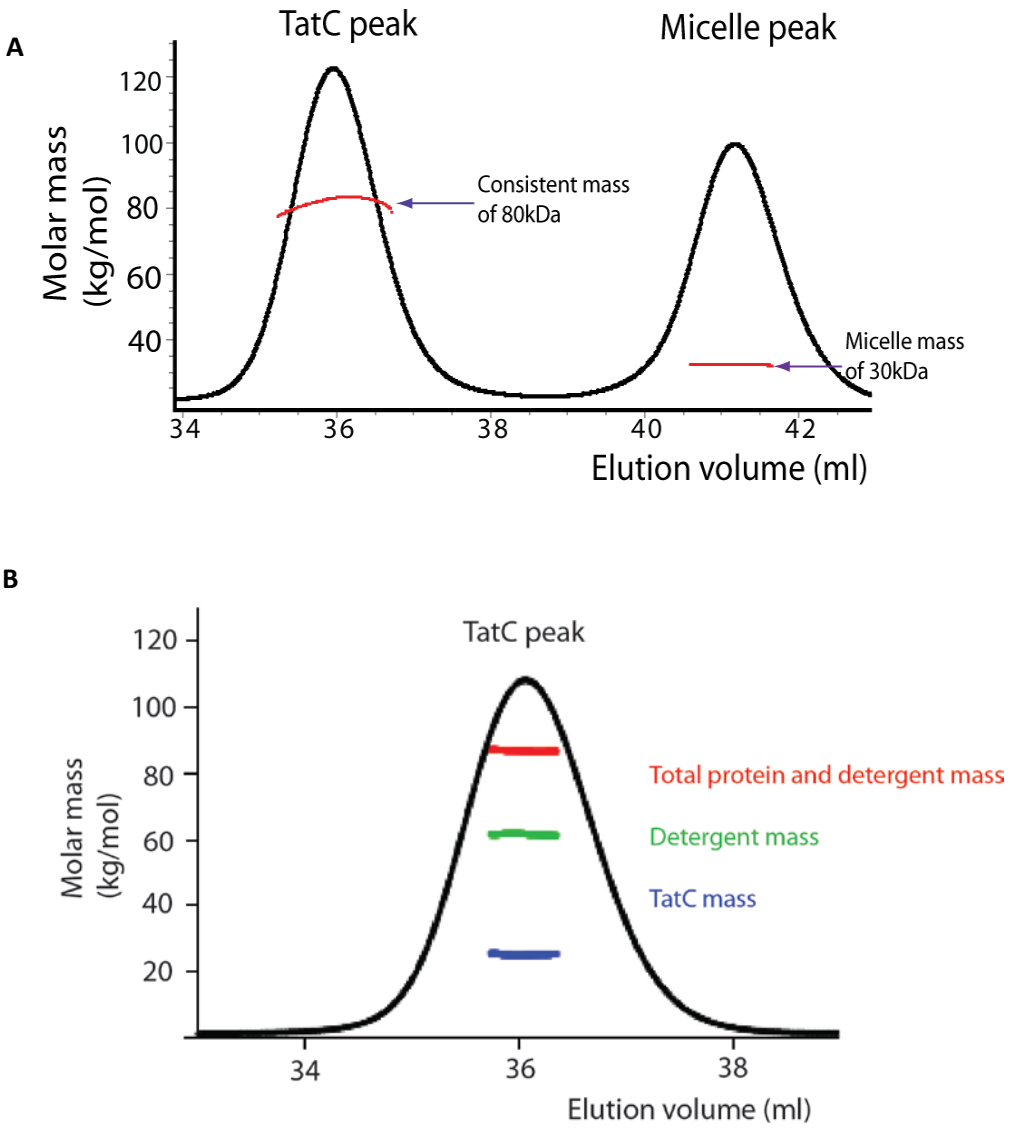


Figure 2.4.2 Multi angle laser light scattering analysis of TatC in cymal-4.

(A) The peak containing TatC in cymal-4 is the first peak, and the second peak is due to the cymal-4 micelles. The calculated masses for each of these species is given in red. **(B)** Protein conjugate analysis of the TatC peak.

Inspection of the calculated molecular mass for the TatC in cymal-4 showed that in contrast to DDM and the other detergents tested in **section 2.3**, the total protein-detergent mass was very consistent all the way across the peak. This indicated that there was a single

species of TatC present in cymal-4. Furthermore, protein conjugate analysis (which is able to dissect apart the contributions to the calculated mass from the detergent or protein components) gave the same, homogenous mass profile for the TatC component alone (**Figure 2.4.2 B**). The calculated molecular mass in this detergent of 25kDa was consistent with a monomer of TatC. MALLS experiments run at different concentrations of TatC in cymal-4 showed exactly the same profile and calculated mass and showed no concentration dependent oligomerisation. The MALLS therefore indicated that cymal-4 may be good to use for crystallography, as there was a single and stable protein species present in the sample.

Crystallisation trials of TatC in cymal-4

TatC in cymal-4 was concentrated to around 6mg/ml and the 'Stura solubility test' (discussed in **section 2.1**) was used to show this concentration formed a good starting point for crystallisation. As discussed previously, high throughput crystallisation trials were carried out with this sample using commercial screens, at varying temperatures, drop ratios, and protein concentrations. These initial crystal trials were successful in finding conditions in which the protein would crystallise. These conditions were different from those used to grow crystals of TatC in DDM, and again these crystals have a different morphology. The best initial crystals are shown in **Figure 2.4.3.**

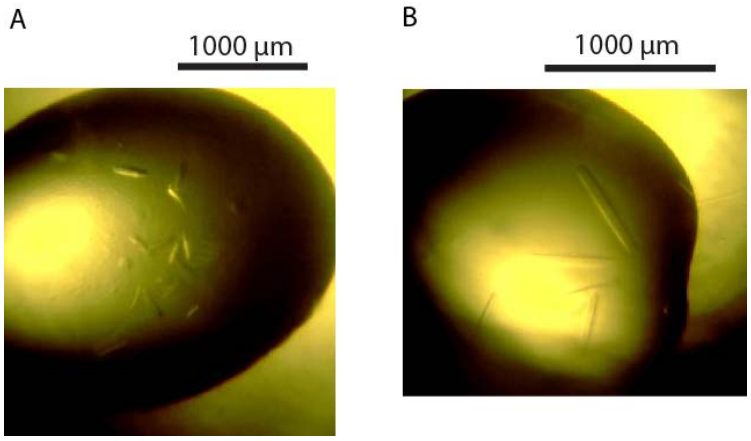


Figure 2.4.3 Crystals grown from broad screens of TatC purified in cymal-4.

Both crystal types were grown in 0.1M KCl, 0.02M Tris pH 7.0, 20% (w/v) PEG 4000 (**A**) 25:75 protein:mother liquor (**B**) 40:60 protein: mother liquor. In both cases the protein concentration was ~5.5mg/ml

The diffraction from these crystals was tested at the synchrotron, as described for the DDM crystals. Diffraction was significantly improved compared to the DDM crystals, with maximum diffraction in the best direction reaching 10Å, an improvement on the 20Å resolution diffraction for crystals in DDM. An advantage to using cymal-4 over DDM for crystallisation is that cymal-4 is a very short chain detergent, with just a four carbon atom stretch capped with a cyclohexane ring in the acyl chain as opposed to a stretch of 12 carbon atoms in the acyl chain of DDM (**Figure 2.1**). As discussed at the start of this chapter, a membrane protein purified in a shorter chain detergent will have more of its hydrophilic loop regions exposed, as so may be better able to form crystal contacts to other protein molecules. The longer chain detergents, such as DDM, form a much larger micelle around the membrane protein and so can occlude the crucial hydrophilic regions and so hamper the crystal lattice formation, perhaps contributing to the disorder of the crystals of TatC grown in DDM.

Above the CMC, detergent micelles are in a constant equilibrium with detergent monomers. If the detergent concentration drops below the CMC, the micelles will break up

and so expose the hydrophobic parts of the membrane protein leading to protein aggregation. For this reason, the concentration of the detergent must be kept above the CMC, which means that there are often free micelles present alongside the membrane protein. It is known that excess detergent micelles can hamper the formation of a well ordered protein crystal lattice (Privé, 2007). Following the MALLS experiment, which confirmed the mass of the cymal-4 micelles as 30kDa and the mass of the TatC in detergent as over 80kDa, a more appropriate cut off concentrator device was selected to use in the final stages of concentration. The use of 50kDa MWC concentrators still allowed the protein to be concentrated, but now the micelles passed through the membrane. The depletion of the micelles was confirmed by re-running the MALLS using samples of cymal-4 buffer concentrated in this way (data not shown).

As an additional step to avoid excess detergent concentration prior to setting up crystallisation trays, the protein from the reverse nickel column was concentrated to 1ml and injected on a S200 10/30 SEC column instead of the S200 16/60 SEC column used previously. This resulted in the protein eluting in a much smaller volume (2ml rather than 10ml) and it was therefore almost 5-fold more concentrated. As the column effectively separates the protein from the free micelles these fractions contain the same amount of buffer detergent as fractions from the S200 16/60 size exclusion column but a higher concentration of protein. As the concentration required for crystallography is now much less, this acts to minimise the amount of free detergent that could potentially also become concentrated, by reducing the degree of concentration required.

Following these attempts to reduce the amount of excess detergent in the sample the protein was again subjected to the broad screen crystallisation trials, discussed previously, to see if the sample, now devoid of micelles, may crystallise in new conditions. This sample yielded crystals from different conditions of the broad screens, which demonstrated

that removal of excess detergent had an effect on the crystallisation properties of the TatC sample. The best crystals are shown in **Figure 2.4.4 A**, grown in 0.05M NaCl, 0.02M glycine pH 10.0 and 33% (w/v) PEG 1000. To test the diffraction, the crystals were harvested in a cryo-protectant solution of mother liquor and 20% (v/v) glycerol, flash frozen in liquid nitrogen and exposed to X-rays at the Diamond synchrotron. The maximum diffraction reached approximately 8Å. These crystals therefore diffracted to slightly better resolution than the initial cymal-4 crystals and this could be as a result of the detergent removal. This crystal condition was subjected to extensive optimisation using optimisation screens made up as in **Table 2.4** (PEG, pH and salt screen) and **Table 2.5** (PEG versus protein concentration screen) with different concentrations of protein and various temperatures.

Upon setting up trays using the optimisation screens and varying the protein concentration, drop ratios and temperatures, crystals were found to grow in a number of drops, indicated by the red boxes in **Table 2.4** and **Table 2.5.** The best crystals were found in the 'PEG only' screen, which had a constant pH and constant NaCl concentration. There was a clear trend whereby increasing the concentration of PEG led to a decrease in the size and quality of the crystals.

The best crystals were obtained in 30% PEG 1000 at 21^{o}C (**Figure 2.4.4.B**) and these diffracted to 7-8Å. These screens therefore showed the crystals were reproducible, but still did not diffract to significantly higher resolution than those from the initial broad screen. Data at this resolution would not be useful in solving the structure of TatC.

			%(v/v) PEG 1000						% (v/v) PEG 1000					
			30	31	32	33	34	35	30	31	32	33	34	35
			1	2	3	4	5	6	7	8	9	10	11	12
[NaCl] (M)	A	0.03	0.02M glycine pH 9.6						0.02M glycine pH 9.8					
	B	0.04												
	C	0.05												
	D	0.06												
	E	0.03	0.02M glycine pH 10.0						0.02M glycine pH 10.2					
	F	0.04												
	G	0.05												
	H	0.06												

Table 2.4 Cymal-4 crystallisation optimisation screen 1. Based on 0.05M NaCl, 0.02M glycine pH 10.0 and 33% (w/v) PEG 1000

			% PEG 1000											
			29.5%											35%
			29.5	30	30.5	31	31.5	32	32.5	33	33.5	34	34.5	35
			1	2	3	4	5	6	7	8	9	10	11	12
% [protein] relative to that in the initial hit	110	A												
	107	B												
	105	C												
	100	D												
	95	E												
	92	F												
	90	G												
	89	H												



Table 2.5 Cymal-4 crystallisation optimisation screen 2. Based on 0.05M NaCl, 0.02M glycine pH 10.0 and 33% (w/v) PEG 1000, where 100% [protein] relative to initial hit was exact concentration of protein in the initial hit. The red square marks the region in which crystals grew, and more specifically 'B' marks the position that the crystal in **Figure 2.4.4 B** grew.

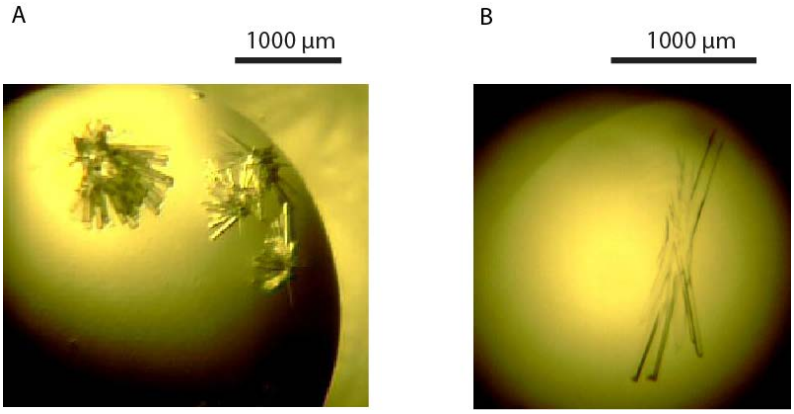


Figure 2.4.4 Crystals of TatC in cymal-4 grown from a purification using a 50kDa concentrator device.

(A) The crystals grown in the initial broad screen condition 0.05M NaCl, 0.02M glycine pH 10.0, 33% w/v PEG 1000. (**B)** Larger crystals grown upon optimisation. These were grown in the presence of 30% PEG 1000, the remaining components were unaltered. These were grown in position 'B' in **Table 2.5.**

Other techniques attempted to improve the crystals grown from TatC in cymal-4

To try and push the resolution of the above crystal form higher, a commonly used technique for post-crystallisation improvement was tried, dehydration. Furthermore in an attempt to find a new, better diffracting, crystal form in this detergent, other, pre-crystallisation techniques were used; fusion to a soluble protein and additive screening.

Dehydration

Crystals are often over 50% solvent, and large solvent channels can exist throughout the crystalline lattice. There is a general trend for crystals having a larger solvent content to be less well ordered, and thus diffract to lower resolution (Frey, 1994). Dehydration methods have, therefore, been developed to help remove solvent from protein crystals. This approach has been shown to be very successful in a small number of cases, for example by increasing diffraction from 6Å to 2.3 Å for a bacterial enzyme (Manjasetty *et al.*, 2001)

and has been recently used on crystals of the respiratory complex I (Efremov & Sazanov, 2011), a large integral membrane protein complex with 55 transmembrane (TM) helices. There are several methods available to carry out crystal dehydration ranging from capillary tubes with small holes connected to a higher salt solution, to exposing the crystal to mother liquor solutions containing progressively more of the precipitant solution (eg the PEG) (Heras & Martin, 2005).

A different dehydration approach was used for the crystals grown from the screens in **Table 2.4** and **Table 2.5.** They were taken to Diamond and exposed to X-rays at room temperature whilst being subjected to online dehydration, using a specially installed dehumidifier, HC1b (Sanchez-Weatherby *et al.*, 2009). This has several advantages to the experiments mentioned above as it allows for much greater control over the dehydration process, and for monitoring diffraction at each step in the process. It has been shown that the controlled dehydration of some crystals can lead to the partial collapse of a semi-disordered protein lattice which can then re-orientate to form a more ordered lattice (Russi *et al.*, 2011) . This may explain how the technique can help give such a drastic improvement in resolution for some crystals.

The experiment was conducted on around 15 different TatC crystals from the optimisation trays (**Tables 2.4 and Table 2.5)** and showed that the solvent content of the crystals was reduced to roughly 70% of the starting humidity over a 10-20 minute time course, but the diffraction remained at a constant 7Å. Beyond this point diffraction was lost never recovered.

Increasing the size of the hydrophilic regions of the sample

The concept that the detergent micelle can occlude not only the hydrophobic, transmembrane sections of the membrane protein, but also cover the hydrophilic loop regions which are often involved in crystal contact formation, has lead to approaches to increase the size of the hydrophilic regions on the membrane protein to aid crystallisation. Given that *A. aeolicus* TatC is predicted to have only very short loop regions, this method may have been advantageous to promote ordered crystal contact formation. One method to achieve a greater exposed hydrophilic area for the formation of crystal contacts is to create fusions of the membrane protein of interest to a soluble protein. This method has proven successful for solving the structure of the β_2 adrenergic G-protein couple receptor (GPCR), in which T4 lysozyme was fused to the protein in place of the third loop between transmembrane helices (Rosenbaum *et al.*, 2007). In this case, the loop that was replaced was also expected to be very flexible, and so addition of lysozyme, a protein which is known for its tendency to crystallise to high resolution with ease, was also hoped to help reduce this disorder and overcome an entropic barriers to crystallisation. Indeed this method proved better at producing high resolution crystals than crystallisation with an antibody fragment made against the protein (2.4 Å resolution compared to 3.4Å).

The fusion approach requires engineering of the constructs to add on the soluble protein. In the case of the *A. aeolicus* TatC-GFP construct, there is already a soluble protein which has a tendency to crystallise fused to the CT of TatC, the GFP fusion which aided in the initial screening stages of the project. This could behave in a similar fashion to the lysozyme fusions, so purification of the full length construct in cymal-4 was carried out. The purification was done in the same way as discussed previously, except there was no TEV cleavage step and instead the protein was concentrated and injected onto the SEC column straight after the initial Nickel affinity column stage. To check that this sample in cymal-4

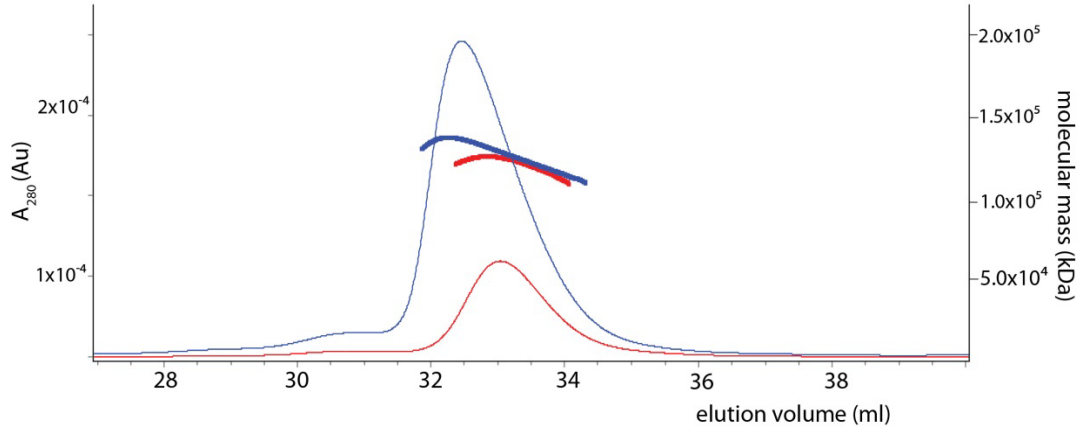


Figure 2.4.5 Multi angle laser light scattering analysis of theTatC-GFP fusion constructs purified in cymal-4.

The blue traces are from a 300 µM sample, and the red traces are from a 75 µM sample run in SEC buffer with 0.8% (w/v) cymal-4. The thin lines are the UV trace, and the corresponding calculated masses across the peak are given by the thick lines across the elution peaks.

was similarly homogenous to TatC alone in cymal-4, the MALLS experiment was repeated for the fusion construct in cymal-4 at two concentrations (**Figure 2.4.5**).

The calculated mass across the peak was not quite as flat and consistent as that for TatC. The mass of TatC-GFP varied with protein concentration, the more concentrated sample (blue) gave a mass that was much more sloped, and tending to higher MW than the lower (red) concentration. Protein conjugate analysis showed that the calculated protein mass became slightly larger as the concentration was increased, and the mass profile was sloped. This indicated that the GFP was potentially starting to cause some oligomerisation, or aggregation of the protein sample. However, as the mass heterogeneity did not appear as drastic as in the DDM or NM samples for example, the fusion was still subjected to the standard broad screen crystallisation trials discussed in the previous sections. However, these trays gave no crystal hits. This could have been due to the slight heterogeneity seen in the MALLS experiments, or due to the linker incorporating the TEV cleavage site

sequence, between TatC and the GFP domain, being 12 amino acids and potentially flexible.

Another emerging technique which also acts to increase the hydrophilic area of the membrane protein is the use of antibody fragments as crystallisation chaperones. Antibodies from mice (Rasmussen *et al.*, 2007) and more recently special antibodies from camelids, 'nanobodies', have been used successful to crystallise membrane proteins. Nanobodies are derived from antibodies from llamas which naturally do not contain the light chain. This means the variable region is encoded by a single gene, and thus is made from a single protein domain (Muyldermans *et al.*, 2001). Antibodies can be raised in the llama and the genes for the variable region isolated and selected for binding the protein of interest. These can bind and stabilise a certain conformation of the protein which may help in the crystallisation of membrane proteins, such as the GPCRs which have a variety of conformational states as part of their function. The ability of a nanobody-GPCR complex to crystallise and have its structure determined to high resolution has been shown for a T4-lysozyme fusion of the β_2 adrenergic receptor (Steyaert & Kobilka, 2011) solved in an active agonist-bound state. The nanobody chaperone technique was not attempted as part of this study to solve the structure of TatC, but was tried unsuccessfully to aid high resolution crystallisation of *A. aeolicus* TatC by a competitor (Personal communication, Ass't Prof. Bil Clemons).

Additive screening and detergent removal

Some membrane protein crystals have shown an increase in quality when grown in a mixture of detergents, which proved to be key to the success of gaining high resolution crystals which were never obtainable from a single detergent alone (Koronakis *et al.*,

2000). To try and most efficiently screen whether a combination of the cymal-4 detergent used in purification with another detergent can give higher resolution crystals an additive screen, MemAdvantage (Molecular Dimensions), was used. This screen is in a high throughput format and provides 96 different additives to add to a 96 well crystallisation tray. The screen includes a mixture of additional detergents and additives that have been successful in the structure determination of other membrane proteins in the PDB. The additive screen was used in the optimisation screens in **Table 2.4** and **Table 2.5**. However, there was no improvement in the diffraction quality of the crystals grown.

As cymal-4 is a detergent with a high CMC (7mM in 0.2M NaCl), it is used in all the purification buffers at high concentration (0.8% w/v). It was hypothesised that the large amount of detergent in the buffer may contribute to the disorder within the crystals. As discussed earlier, the removal of the excess free micelles did lead to an improvement in the diffraction resolution of crystals. Thus, attempts were made to remove more detergent from the protein, either by the use of different cyclodextrans (γ as well as β) which may be more effective at the removal of this detergent with a 'bulky' cyclohexane group (Signorell *et al.*, 2007), or by simply reducing the amount of detergent in the purification buffer to the lowest tolerated level (before the protein precipitates) of 0.6% (w/v). Neither of these techniques significantly improved the crystals grown.

Use of amphipols

To combat the problem of working with detergents which need to be kept in the purification buffer at a certain minimal concentration and the presence of excess micelles, compounds called amphipathic polymers (amphipols, or APols) have been developed (Tribet *et al.*, 1996). APols have a much higher affinity for the hydrophobic regions of

membrane proteins such that only a small amount needs to be present in the buffer, as it will bind with high affinity, essentially irreversibly to the membrane protein and will not dissociate even at very low APol concentrations. This slow rate of dissociation is because APols are larger than detergents, and so associate with the membrane protein at multiple contact points. APols cannot solubilise membrane proteins, so detergents are used for the initial extraction step and are then exchanged for APols. However, the potential benefits resulting from the complete removal of free detergent micelles from the buffer using APols for membrane protein X-ray crystallography have yet to be demonstrated. Crystals of a model membrane protein, the mitochondrial cytochrome bc_1, have been grown in the presence of APol and detergent, but only diffracted to >20Å (Popot *et al.*, 2011).

Despite the lack of proven success with gaining high resolution crystals, exchange of *A. aeolicus* TatC from cymal-4 into APol A8-35 (Anatrace), a polyacrylate APol, was attempted. Upon optimisation of the detergent exchange protocol, a relatively monodispersed sample of TatC in A8-35 was achieved (as assessed by MALLS, data not shown) and was functional (as assessed by a calorimetric binding assay (discussed in **Chapter 4**), data not shown). However, at these experiments were conducted at the point of successful phasing from another crystal form (**Chapter 3**) the APol-TatC sample was not subject to crystallisation trials.

2.5 Methylation of TatC in cymal-4 gives a new, higher resolution crystal form

Amongst the attempts discussed above to gain crystals with a much improved diffraction quality from the TatC cymal-4 preparation, the methylation approach employed for the DDM species was also attempted in the hope this would alter the surface properties of TatC enough to improve the crystal packing. The same protocol was used as in **section 2.1**,

and native mass spectrometry by Dr. Julien Marcoux, (Dept of Chemistry) was able to confirm the expected mass increase for the methylation of all 14 lysine residues, and the methylation of the NT (420 Da mass increase). However, this time there was no obvious change in the solubility of the species, as determined by the crude 'Stura solubility test' (as in **section 2.1**) Crystallisation trials were therefore initially set up a using a protein concentration of 6mg/ml, and subsequently at a range of concentrations. After a week, crystals were observed in new conditions for the 6mg/ml sample, compared to those for the native protein, and again as for TatC in DDM the methylation did seem to have had an affect on the surface properties or solubility of the protein, altering the conditions in which it will crystallise.

Crystals were grown in MemGold condition G12, 0.001M zinc sulphate, 0.05M hepes pH 7.8, 28% (w/v) PEG 600. This drop is shown in **Figure 2.5.1**, and there are three crystals highlighted. Each of these three crystals were harvested in a cryo solution of mother liquor supplemented with 30% (v/v) PEG 400, flash frozen, and exposed to X-rays at the European Synchrotron and Radiation Facility (ESRF) in Grenoble. The crystals circled with a red line diffracted poorly (roughly 7Å) as seen before, but the crystal circled with a blue line displayed a different morphology to the other two and gave diffraction up to 4.0Å resolution, which is much better than any of the other TatC crystals ever shot before. Upon

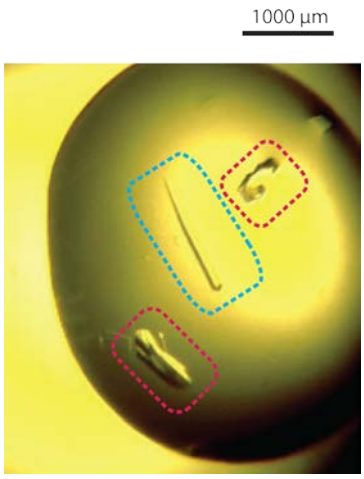


Figure 2.5.1 Crystals of methylated TatC in cymal-4 identified in broad screen crystallization trials.

Crystals were grown in 0.001M zinc sulphate, 0.05M hepes pH 7.8, 28% w/v PEG 600. The crystals highlighted by red lines gave poor diffraction, and the crystal in blue gave diffraction to 4.0 Å.

optimisation of this condition, by varying the concentration of PEG, salt and the pH of the buffer, a subsequent crystal was grown (**Figure 2.5.4** highlighted in blue) that gave a second high resolution dataset, collected at Diamond, which diffracted to 4.16 Å resolution.

Data collection and processing

Images from both the 4.0Å crystal and the 4.16Å crystal were collected at the ESRF or Diamond under cryo conditions, and using the rotation method, by mounting the crystal onto a rotating goniometer. Two images were initially collected, 90^{o} apart to allow for indexing. Indexing by Mosflm (Leslie, 1992), which identifies spot positions on each image and attempts to identify a unit cell with dimensions and symmetry consistent with this pattern, suggested the data were orthorhombic with a lattice symmetry of P222, and an estimated mosaicisity of 0.4^{o}. The knowledge of the symmetry of the crystal defined which reflections are equivalent. A strategy was then formulated for a complete dataset collection, so as to give an intensity measurement for each unique reflection, over as small a rotation range as possible (to minimise radiation damage). The EDNA programme (Incardona *et al.*, 2009) was used to calculate this strategy, which advised for data collection of the ESRF dataset to begin with a starting angle of 58^{o}, with an oscillation per image of 0.15^{o} and exposure for 0.2 seconds, over a total rotation range of 135^{o} .

Data processing analyses the complete set of the diffraction images to give a list of the Miller indices for each unique reflection (h, k, l) and the intensity of each reflection I_{hkl} along with an estimate of the error of that intensity, σ_{hkl}. Upon scaling and merging this .mtz file allows for determination of the space group, by analysis of the statistics for

merging symmetry related reflections. The final step allows for the structure factor magnitudes, |F|, to be calculated from the intensity measurements.

Images were processed by collaborative computation project 4 (CCP4) programmes as part of the 'Xia2' automated processing pipeline (Winter, 2010). The main stages in the data reduction to the final, merged mtz file involve the following stages;

1. Indexing the data to identify the unit cell and symmetry, and labelling each spot with the appropriate Miller index.
2. Integration to give an intensity value for each reflection and an estimate of its error.
3. Scaling to bring the reflections observed across multiple images on the same scale.
4. Merging, whereby symmetry related reflections and those reflections which have been recorded redundantly, are averaged together to give a single intensity and error measurement for that particular unique *hkl* reflection.
5. Truncation, in which the merged intensities are converted to structure factor amplitudes.

The indexing and integration were performed using XDS (Kabsch, 2010). This programme is optimised for collection with Pilatus detectors in which data has been collected by fine slicing and thus reflections are partially recorded and distributed across a number of images. XDS fits a 3-dimensional profile to the spots and uses this to estimate the pixel area underneath each reflection. It takes into account the background by creating a mask around each spot and subtracting this from the spot intensities.

During the data collection process, there are many factors which can affect the diffraction from the crystal, such as orientation within the beam, radiation damage, B factors, and small fluctuations in the intensity of the beam. The data are therefore scaled to bring the intensities from reflections which are related (redundantly measured reflections, either

identical measurements of the reflection or measurements of those reflections that are symmetry related) but have been measured separately, onto the same scale. For this data set SCALA was used to carry out the scaling and merging (Evans, 2006). This programme creates a scaling profile based on redundant measurements of the same reflection, and uses this to scale the frames. This outputs statistics for the quality of the data, which can help in determining the true symmetry of the data. Data processed in too high symmetry will show high merging statistics from merging two values which are not symmetry related, and thus the intensities are not the same.

Data processing statistic	**Interpretation**
$< I/\sigma(I) > = \frac{1}{N}\sum_{h}^{N} \frac{\lvert I_{(h)} \rvert}{\sigma\,(I_{(h)})}$	A measure of the average signal to noise for a reflection h when merged over N observations within a resolution bin.
$R_{\text{merge}} = \frac{\sum_h \sum_{i=1}^{N} \lvert I_{(h)} - \bar{I}_{(h)} \rvert}{\sum_h \sum_{i=1}^{N} I_{(h)i}}$	This reports the agreement between multiple measurements of the same reflection within a dataset. This allows for a measure of the internal consistency of the data after scaling and merging, however it will increase with an increased number of measurements of the reflection (i.e. increased multiplicity).
$R_{meas} = \frac{\sum_h \frac{1}{N-1}^{\frac{1}{2}} \sum_{i=1}^{N} \lvert I_{i(h)} - \bar{I}_{(h)} \rvert}{\sum_h \sum_{i=1}^{N} I_{(ih)}}$	This measurement gives the multiplicity independent merging R-value, and thus is not artificially inflated upon collecting more redundant datasets (which contain more measurements of the same reflection, and thus are more accurate in determining intensity).
$CC_{1/2}$	The linear correlation coefficient of the intensity measurements of a reflection from two random halves of the dataset.

Table 2.6 The data quality statistics used to analyse the processed diffraction data

The merged intensities outputted from SCALA are fed into Truncate (French & Wilson, 1978), which allows the structure factor magnitudes, |F|, to be estimated from each intensity measurement. It will also generate a set of free R flags on 5% of the data (Brunger, 1992). Certain key statistics are calculated in order to help give an assessment of the quality of the diffraction data collected (**Table 2.6)**. In order to see if the improvement in diffraction has enabled a usable data set to be collected, these statistics need to be examined. The statistics for the processing of the ESRF and Diamond cymal-4 TatC datasets are given in **Table 2.7.**

Data collection	**Dataset 1 (ESRF)**	**Dataset 2 (Diamond)**
Space group	P 2_12_12	P 2_12_12
Cell dimensions		
a,*b*,*c* (Å)	77, 125, 59	82.7, 114.4, 57.2
$\alpha\ \beta\ \gamma$ (^{o})	90, 90 ,90	90, 90, 90
Wavelength	0.98	0.976
Resolution (Å)	125.54-4.00 (4.10-4.00)	82.7-4.16 (4.27-4.16)
R_{merge}	0.094 (0.532)	0.052 (0.363)
R_{meas}	0.099 (0.555)	0.059 (0.404)
$I/\sigma I$	15.2 (4.9)	14.2 (3.8)
$CC_{1/2}$	0.98 (0.95)	0.99 (0.93)
Completeness (%)	99.9 (100.0)	88.2 (87.5)
Redundancy	12.2 (12.2)	4.9 (5.1)
Total observations	55640 (3932)	19887 (1517)
Total unique	4556 (323)	4056 (298)

Table 2.7 The data processing statistics from two datasets from TatC cymal-4 crystals. The first dataset was collected from a crystal at the ESRF and the second from a separate, reproduced crystal at Diamond. The values for the inner shell are given in brackets.

Any symmetry operation that includes translational elements leads to the systematic absences of certain reflections in the diffraction pattern. The absence arises due to the additional destructive phase difference between identical molecules that have been translated along an integer fraction of the unit cell axis. If the translation is along a screw axis, this will lead to the destruction of certain reflections along the corresponding reciprocal axis. In the case of a 2_1 screw axis along the b axis, the screw axis allows reflections to be possible only for 0k0 when k=2n, and every second reflection (where κ= odd number) must be absent (i.e. the intensity is zero). The $P2_12_12$ space group of the TatC datasets was identified by inspection of the pattern of systematic absences. The 2_1 screw axes along the h and k axis gave the characteristic pattern of intensity absences for every odd numbered reflection (**Figure 2.5.2 A** and **B**). However the pattern of absences on the k axes contained a single outlier to this pattern for the ESRF dataset on the k axis, highlighted in pink on the figure, which would not be consistent with the 2_1 screw axis. The lack of pattern in the absences along l (**Figure 2.5.2 C)** implied a lack of screw axis in this direction. R_{merge} against batch for all the images was very consistent during the dataset collection and thus there was no reason to exclude any batches during the data processing. The R_{merge} against resolution for the overall dataset is shown in **Figure 2.5.2 D**, which shows that the R_{merge} is around 50% at the 4Å limit.

Analysis of the asymmetric unit (ASU)

At this stage of the data processing, there is no phase information to go with this dataset. However, there are certain pieces of phase independent information that can be extracted and used to estimate the unit cell contents.

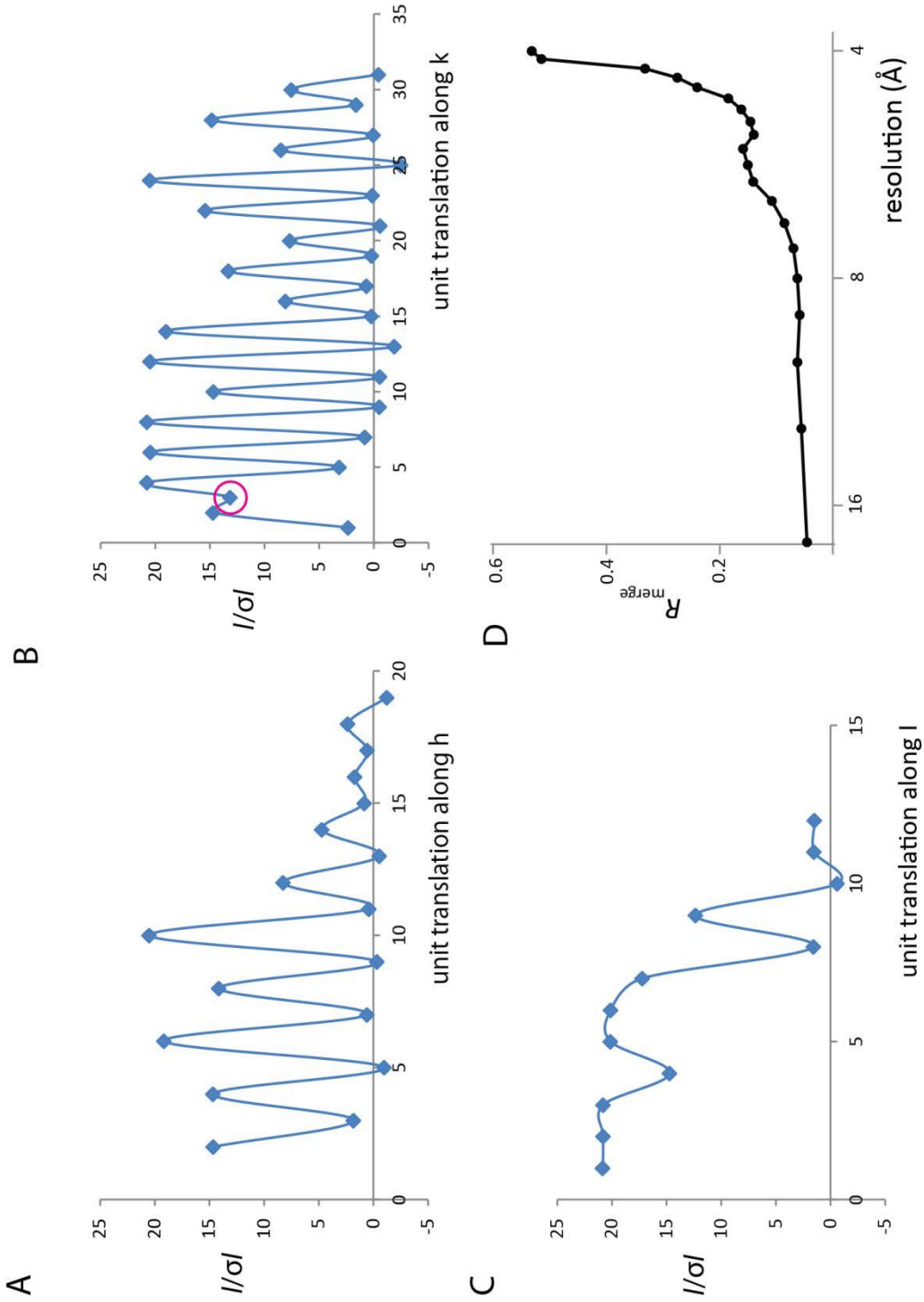


Figure 2.5.2 Data processing of the ID29 TatC dataset.

The systematic absences in the $I/\sigma I$ values observed along the 2 crystal axis, (**A**) h and (**B**) k indicates the presence of a screw axis along both of these axes. The single outlying reflection on the k axis is marked with a pink circle. The lack of pattern in the absences in (**C**) implies there is no screw axis along l. (**D**) The R_{merge} against resolution.

Proteins often crystallise with more than one copy of the molecule in the asymmetric unit (Bricogne, 1974). The number of molecules in the asymmetric unit is useful to know as multiple copies within the asymmetric unit can assist in non-crystallographic symmetry (NCS) techniques to help in structure solution and phase improvement by NCS averaging (Zhang *et al.*, 2001) and knowledge of the solvent content is crucial for density modification techniques (Wang, 1985).

Based on the analysis of solved crystal structures, there is a correspondence between the volume of the asymmetric unit and the corresponding molecular weight of the protein. The ratio of these two values is the Matthews coefficient, V_M; this bears a relationship to the solvent content of the crystal (Matthews, 1968). The solvent content of the ASU can be calculated using the CCP4 'MATTHEWS' programme, which requires the unit cell dimensions and the number of residues in the protein sequence. Using a calculated set of probability distributions for the Matthews coefficient given the resolution limit of the data (Kantardjieff & Rupp, 2003), it can give a set of Matthews probabilities for the presence of a certain number of copies given the diffraction limit of the dataset (i.e. P(4.00) in the case of the 4Å dataset) or given all structures in the PDB, P(total). For the dataset (ID29) these statistics are given in **Table 2.8.**

Number of copies/ASU	% solvent		Matthews coefficient, V_M ($Å^3Da^{-1}$)		P (4.00) ID29 P(4.16) I03		P (total)	
	1	2	1	2	1	2	1	2
1	73.1	75.2	4.58	5.00	0.03	0.01	0.02	0.01
2	46.2	50.1	2.29	2.50	0.97	0.98	0.98	0.98
3	19.4	26.3	1.53	1.67	0.00	0.01	0.00	0.01

Table 2.8 An estimate of the solvent content of the two cymal-4 TatC datasets. Calculated by the CCP4 Matthews programme. Dataset 1 from the initial crystal grown in the broad screen and the data collected at the ESRF, and dataset 2 from the crystal grown in the optimisation trays and data collected at Diamond.

For both datasets, inspection of the P(4.00)/P(4.16) and P(total) indicated that the most likely content is two copies in the asymmetric unit. However, the extremely low probability of one copy with a solvent content of 73/75%, can not be ruled out as high solvent content could relatively poor diffraction.

The self rotation function (SRF) can give more insight into the ASU contents, by giving the relative orientational relationships between molecules within the ASU (Rupp, 2009). The SRF reveals the presence of all rotational and screw axes within the unit cell, the orientation of these axes and the degree of rotation required to bring a pair of symmetry related copies into coincidence. The SRF can also reveal the presence of NCS, the presence of this type of symmetry gives additional peaks in the SRF to those peaks arising from the crystallographic symmetry. The SRF involves the calculation of the Patterson function for the data around the origin, and then the rotation of this function. Peaks in the SRF arise when, at a certain rotational angle, the Patterson function overlays exactly with itself. The angle at which the coincidence between the Patterson maps occurs corresponds to the rotational symmetry of the crystal. The SRF can be calculated by rotation in space defined by spherical polar coordinates. These coordinates specify three angles (Φ,Ψ,κ) allowing separation of the amount of rotation (given as κ) required to bring the symmetry related copies in coincidence, from the orientation of the rotational axes (Φ,Ψ). There will always be a significant peak at the origin, and also at the rotation positions which correspond to the crystal symmetry.

The SRF function was calculated using the CCP4 programme POLARRFN. The output is given as a stereographic projection, plotted in sections of κ, where the longitude and latitude of the sections mark the Φ and Ψ angles. In the presence of different types of crystal symmetry there will be different peaks on different κ rotation sections of the SRF. Peaks on the κ=180 section will reveal the presence of all 2-fold rotational or screw

symmetry axes, whereas peaks on the κ=120 section will reveal all 3-fold symmetry axes. The positions of the peaks on the sections mark the locations where the rotational symmetry axes enter and exit the sphere.

The SRF of the TatC datasets show the crystals have 222 point group symmetry. This type of symmetry has three 2-fold axes, and so was revealed by inspection of the κ=180^{o} section (**Figure 2.5.3**). The peaks arising due to the three 2-fold axes are highlighted in red, and labelled according to the specific axis they have arisen from. For both the ESRF and Diamond datasets there were no other significant peaks in either the κ=180^{o} section or any other rotation section, apart from those expected for the three 2-fold axes. However, NCS could be present within these crystals if the axis is parallel to the crystallographic 2-folds and so the peaks arising from this axis may be obscured by the crystallographic symmetry peaks.

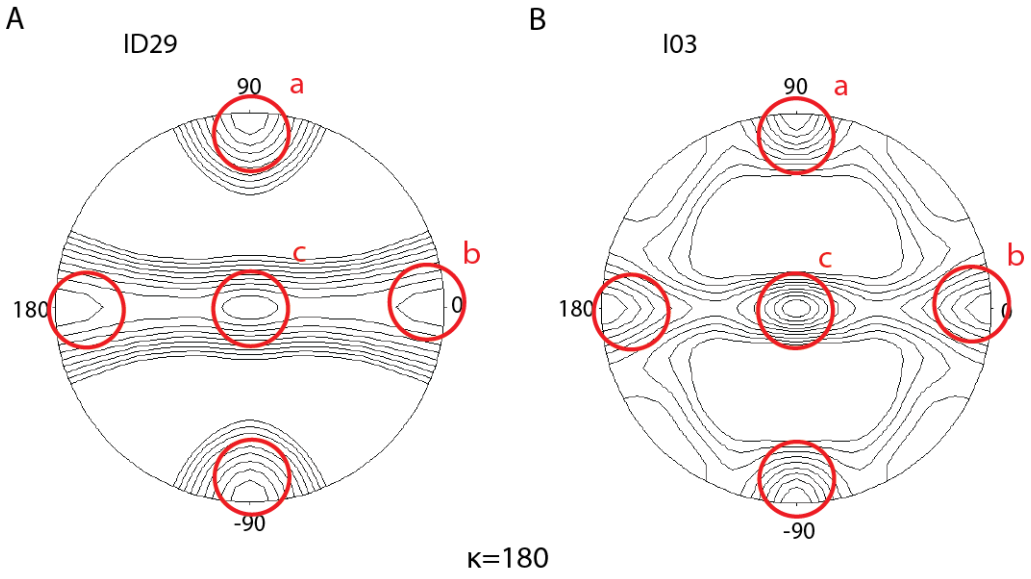


Figure 2.5.3 The Self Rotation Function (SRF) calculated for the two TatC datasets

Calculated using data from 15 to 4.1Å for the ESRF (**A**) (ID29) and the Diamond datasets (**B**) (I03), showing he κ=180^{o} section. The peaks corresponding to the three two-fold symmetry axes are highlighted on the ID29 and I03 dataset, along with the corresponding crystal axis they come from.

Despite the agreement between the ESRF and Diamond datasets on the space group and content of the ASU, the two datasets differed in their cell dimensions. The *a* axis of the Diamond cell was slightly longer at 82.7Å compared to the ESRF cell length of 77Å, the *b* axis much shorter at 114.4Å compared to 125Å, the *c* axes were more similar at 59Å compared to 57.2Å. The differences in the cell dimensions of these two high resolution datasets indicated the protein molecules in the lattice may have been arranged differently in these two crystals. This indicates that the contacts mediating the crystal lattice are may not have been very strong, and may be slightly 'slippery', which would have enabled them to pack into a slightly different orientation.

Difficulties in reproducing the 4 Å crystal form

The more ordered packing arrangement of TatC was clearly very difficult to achieve, as overall, these two datasets were by far the best collected for this crystal form. This was despite the fact that large numbers of crystals grew in many of the optimisation trays (**Figure 2.5.4**) and in total approximately 85 crystals of this type were shot at various X-ray beam lines. Many different types of morphology were observed, from 'plate-like' crystals seen in **A** and **C,** to the more 'rod-like' crystals seen in **B, D, E** and **F**. It was not surprising that the crystals observed from optimisation of this condition look different, given the initial broad screen hit contained crystal forms varying in their morphology and diffraction quality (**Figure 2.5.1**).

When both the plate and rod crystal were tested for their diffraction at various beam lines at Diamond, the 'plate-like' crystals gave poor (~10Å) diffraction, and when iMosflm was able to index these images, it gave the symmetry as hexagonal. The hexagonal form was therefore a very different crystal form to the 4Å orthorhombic form. As these plate-like

crystals looked very different to the rod type, this could have been an indicator of which crystals will be the ones that diffract to higher resolution, and thus the rods could have been targeted for heavy atom soaks when attempting to phase the data.

However, despite the fact that all the 'rod-like' crystals in **Figure 2.5.4 B**, **D, E** and **F** indexed as orthorhombic, with point group 222, they also exhibited variation in their diffraction quality, most datasets collected were only in the 8-10 Å region, and a handful collected with maximal resolution in the 4.5-7Å range. For example, those crystals from drops **D** and **E** gave maximal diffraction to around 8Å, whereas those from **B** faired better, giving a dataset to 4.6 Å.

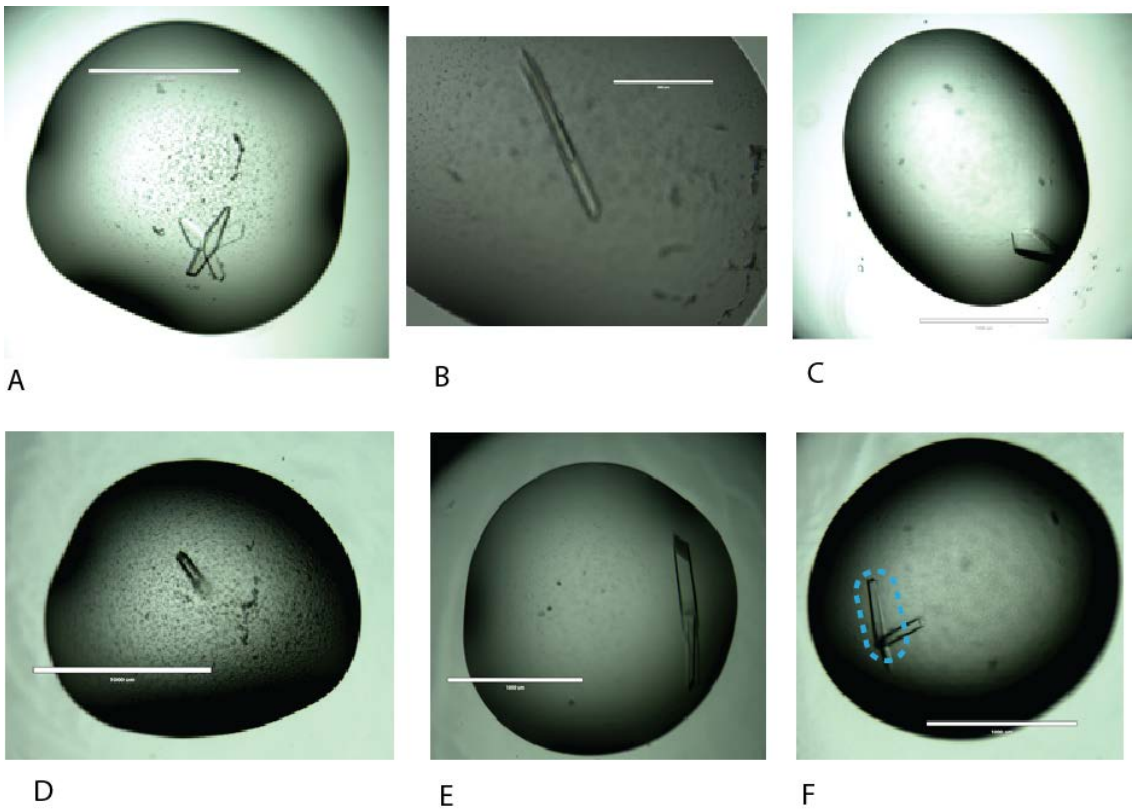


Figure 2.5.4 Crystals of TatC in cymal-4 grown in the optimisation trays.

The scale bar in all the images is set to 1000µM, with the exception of (**B**), where it is set to 400µm. The crystal highlighted in blue dashes in (**F**) gave the second high resolution dataset (Diamond dataset).

The variation in crystal quality was again observed within the same drop. The crystal which gave the second high (4.16 Å) resolution Diamond dataset, shown highlighted in blue in **Figure 2.5.4 F,** adjoined a second crystal which only gave a data set to 5.3Å resolution. This lower resolution crystal seemed to suffer in the beam as it was no longer giving any diffraction by the end of the data collection, whereas the other, higher resolution, crystal from the drop did not suffer from any obvious radiation damage effects.

This second high resolution dataset was collected only after shooting a very large number of poor crystals. The plate-like crystals were ruled out after they were found to be a different crystal form which always diffracted to low resolution, but it was still random as to which of the rod type crystals would give a 4Å dataset. There were no obvious morphological features which were able to distinguish a 'good' rod from a poor one. Repeat trays were set up using the exact condition that this crystal was grown in with the same batch of protein, and yet would only succeed in producing poor crystals. This high resolution crystal form was therefore tremendously difficult to reproduce.

Phasing attempts with the 4.0Å cymal-4 TatC crystal form

At this point the cymal-4 crystals were still the best TatC crystals available. Attempts were therefore made to phase the data. As there were no homologous structures solved for TatC experimental information was required in order to phase any datasets. This can be achieved by isomorphous replacement or anomalous dispersion methods, alone or in combination. Isomorphous replacement involves the binding of heavy atoms to regular points within the crystal lattice, either by co-crystallisation or by soaking pre-grown crystals. A dataset of this derivative (F_{HP}) is used along side the native dataset (F_P) is used to locate the positions of the heavy atoms within the unit cell by inspection of the isomorhous difference

Patterson maps. Once located, the known heavy atom positions allow for the phases of the heavy atoms to be determined. This information allows calculation of phase probability distributions of the protein structure factor phases (Blow & Rossmann, 1961). This method relies on having very little difference between the unit cells between the native and the derivative datasets otherwise the reflections are no longer equivalent.

Anomalous dispersion relies on the presence within the dataset of an anomalously scattering element, and then in a similar manner to isomorphous replacement relies on the solution of this substructure of marker atoms to allow generation of initial protein phase estimation (Blow & Rossmann, 1961). This method has become the most popular way to solve new structures, thanks to the ease of a method that allows the incorporation of the anomalous scatterer selenium into the protein itself, by substitution of methionine for selenomethionine. A more detailed description of the anomalous scattering method will be given in **Chapter 3**, along with the details on producing selenomethionine substituted TatC (SeMet TatC). The anomalous dispersion method, for reasons discussed in **Chapter 3**, was the preferred route to try and gain phase information. However, a SeMet preparation of TatC in cyaml-4 proved even more difficult to give crystals that gave 4Å data, the only crystals that were obtained did not diffract beyond 20Å.

Heavy atom soaks

As SeMet TatC did not give high resolution data, heavy atom soaks were carried out in an attempt to gain some isomorphous or anomolous data from the crystals grown of native TatC in cymal-4. Singular, rod type crystals were selected for this approach as this was the morphology of the two crystals which diffracted to high resolution.

A variety of soak compounds were selected to provide a spectrum of metal ions with different binding preferences which would be soluble at high concentrations (mM range) at the pH of the mother liquor (pH 7.8). Mercury compounds have been shown to be the most successful in phasing novel membrane protein crystals, followed by the platinum derivative K_2PtCl_4. The success of the mercury compounds maybe due to the large amounts of cysteine residues normally found in membrane protein loops and disulphide bonds (Morth *et al.*, 2006). However, the *A. aeolicus* TatC construct contains no native cysteine residues, which could make derivatisation more difficult.

Soaks were performed using a 10mM stock of the heavy atom made in mother liquor. Crystals were transferred from the drop to the solution for varying amounts of time between 5 and 15 minutes. During this period, there was no indication that the soaks were having any detrimental effects on the crystals i.e. there was no obvious colour change or cracks appearing, which could indicate the crystal lattice was being disrupted. The soaked crystals were then transferred to an appropriate cryo protectant solution, before being flash frozen in liquid nitrogen. Crystals were taken to I03 at diamond, where fluorescence scans were carried out on the mounted crystal, immediately prior to data collection. This allowed for crystals which showed an absorbance edge characteristic of the soaked heavy atom to be exposed to a beam of X-rays at a wavelength above the atoms absorption edge. Datasets could be collected from a few of the soaked crystals with the best for three of the compounds given in **Table 2.9.**

Looking at the overall anomalous correlations (CC_{anom}) within the datasets, which is an measure of how much anomalous signal there is present, indicated anomolous differences in the Friedel pairs were evident. However, upon inspection of the correlations against resolution, the CC_{anom} was not as strong as expected at low resolution for Pt and dropped below 0.2 at very low resolution, for example at 7.5 Å for the Pt soak data (**Figure 2.5.5**).

The magnitude of the CC_{anom} in the data proved too low as attempts to find anomalously scattering sites with heavy atom search programmes, such as Hyss (Grosse-Kunstleve & Adams, 2003) and SHARP (de La Fortelle & Bricogne, 1997), had no success.

The overall resolution limit was again quite varied, with only one crystal diffracting to better than 5Å. This could have been due to some disruption to the lattice packing upon soaking, but was most likely due to the intrinsic disorder already present in the crystal, as the native crystals were so varied in their quality. This would pose a massive challenge in any attempt to systematically assess which soaks would be best to use, as derivatives that

Compound soaked into crystals	**Best, complete data set resolution and space group**	**Wavelength (Å)**	**Overall Anomalous Correlation within dataset**
K_2PtCl_4	4.6 Å (P 2_12_12)	0.92	0.017
$HgCl_2$	5.5 Å (P 2_12_12)	1.00	0.21
$GdCl_3.xH_2O$	9.1 Å ($C222_1$)	1.71	0.614

Table 2.9 A summary of the best diffraction data collected from heavy atom soaks attempted with the cymal-4 crystal form.

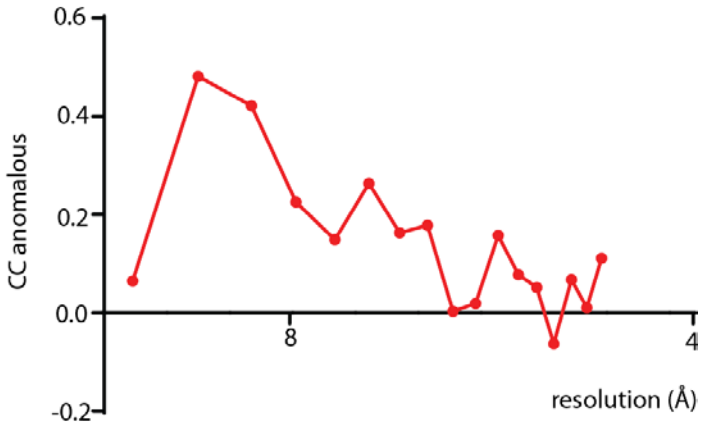


Figure 2.5.5 The anomalous correlations against resolution for the highest resolution dataset for the K_2PtCl_4 soaked TatC crystals.

may have potentially been useful may happen to be soaked into a poor crystal, and thus no data to a useable resolution would be extracted.

Molecular replacement

An interesting point to note is that following structure solution of *A. aeolicus* TatC from a different crystal form (described in **Chapter 3**) extensive attempts to phase the two cymal-4 high resolution datasets (**Table 2.7**) by molecular replacement were unsuccessful. This could indicate a large scale conformational change of the structure in this crystal form, but given the structural analysis presented in **Chapter 4.1,** such a large scale movement is unlikely. These crystals could potentially have been ordered detergent crystals with a unit cell compatible with an object the size of TatC. This theory could be tested by SDS-PAGE of the cymal-4 crystals to check whether they contain TatC protein.

2.6 Summary

The TatC construct from *A. aeolicus* proved to be very amenable to crystallisation. In certain detergents crystals could be readily obtained in different mother liquor conditions. The initial low resolution of the diffraction was significantly improved once thorough detergent screening was able to identify that purification in cymal-4 gave a homogenous sample of TatC. The best two datasets from crystals grown in this detergent, cymal-4, contained data to 4Å. Although this was only of medium resolution, it was a large step forward. However, these higher resolution crystals proved extremely tricky to reproduce efficiently enough to allow phasing by heavy atom soaks. Furthermore, preparations of SeMet TatC (**Section 3.2**), which would allow for anomalous phasing, proved even more

difficult to obtain good crystals from and no datasets were collected. Having exhausted a large range of techniques to use with the cymal4-TatC species to improve both the reproducibility and resolution, a new crystal form needed to be found in order to solve the structure of TatC.

Chapter 3: Determining the structure of *Aquifex aeolicus* TatC

TatC in cymal-4 appeared homogenous by Multi Angle Laser Light Scattering (MALLS) and gave crystals that diffracted to reasonable resolution. The crystals were very difficult to reproduce, and attempts to grow selenomethionine substituted crystals only produced extremely poor crystals. The extremely variable diffraction quality of the crystals generated in cymal-4 made attempts at phasing with heavy atoms very difficult. As described in **Chapter 2**, an exhaustive range of efforts were employed in an attempt to improve this crystal form, and to find a different, better crystal form than TatC in cymal-4. The lack of success led to attempts to find another detergent which gave an equally homogenous sample of TatC for crystallisation trials.

The initial list of promising detergents (**Chapter 2, Table 2.1**) identified by F-SEC gave no other species that satisfied this criteria. However, during the course of this study, a new family of detergents were developed by Anatrace, the neopentyl glycol class of detergents. This class of detergents is based around a quaternary carbon atom with two lipophilic tails, and the two hydrophilic head groups based on maltose sugars, as in the DDM type detergents. Neopentyl glycols have shown improved protein stability and more successful crystallisation for a subset of tested membrane proteins (Chae *et al.*, 2010). The two members of this family, lauryl maltose neopentyl glycol (LMNG) and decyl maltose neopentyl glycol (DMNG) have the same headgroup, but differ in the length of the two hydrophobic chain carbon chain tails (ten and eight atoms respectively). A feature of these detergents are their the extremely low CMC values (**Chapter 1, Table 1.1**), which are up to ten fold lower than most popular currently available detergents.

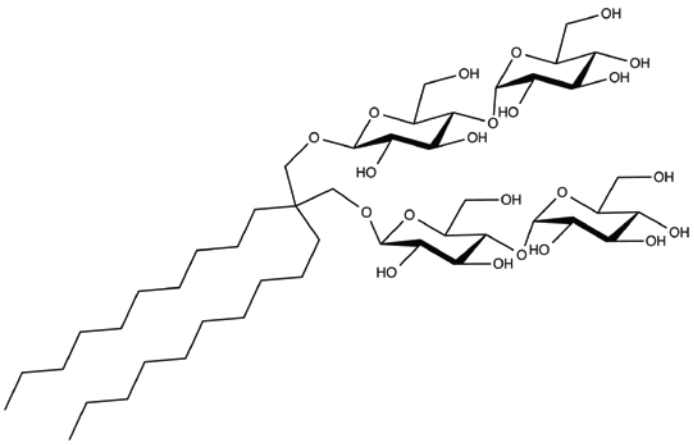


Figure 3.1 The structure of lauryl maltose neopentyl glycol (LMNG).

As discussed in **Chapter 2**, the DMNG detergent was shown to be able to solubilise TatC, but upon purification resulted in a mixture of two species of TatC oligomer (**Figure 2.3.1)**. TatC purified in DMNG was not subjected to crystallisation trials. The other detergent in this class, LMNG, was also found to solubilise TatC and so a purification in this detergent was carried out.

3.1 Purification of *A. aeolicus* TatC in LMNG

The purification was carried out using the same protocol as described in **section 2.1**, except the detergent was used at 1% (w/v) in the solubilisation buffer, and 0.02% (w/v) in all subsequent buffers. The results of the purification are shown in **Figure 3.1.1**. Initially, SDS-PAGE gel showed trace amounts of GFP in the reverse nickel column flow through. The protocol was amended, and the nickel column was eluted, washed in 3CV 1M imidazole (Sigma) pH 7.0, then 10 CV of dialysis buffer, before the flow through (1) was run back over the column, and the flow through (2) collected. SDS-PAGE confirmed this was free of GFP, and so it was concentrated to 5ml and injected on the S200 16/60 column

(GE healthcare) as previously described (**section 2.1**). The elution trace is shown in **Figure 3.1.1 A.**

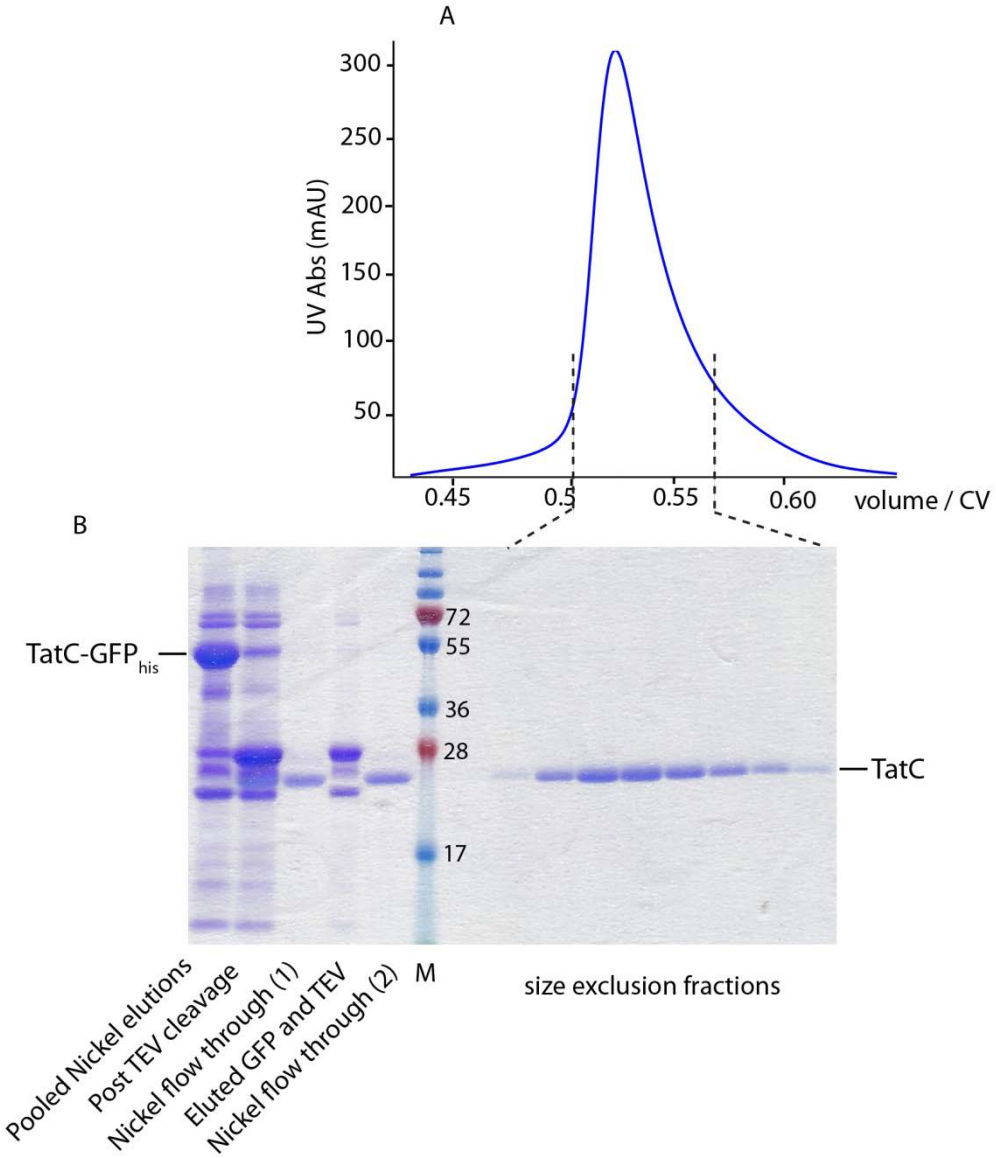


Figure 3.1.1 Purification of TatC in LMNG.

(A) Size exclusion of a sample of TatC post-reverse Nickel purification ran on an S200 16/60 column (GE healthcare) in 150mM NaCl, 20mM hepes pH 7.5 and 0.02% LMNG **(B)**, 15% (w/v) SDS-PAGE showed the majority of the TatC-GFP$_{His}$ eluted from the Ni column was cleaved by the TEV protease. The flow-through from the nickel column was subjected to the reverse nickel column step again, after elution of the bound GFP, TEV and uncut fusion from the column. This flow through (2) was concentrated and applied to the S200 16/60 column, SDS-PAGE confirms the SEC peak contains pure TatC. **M** Pageruler Plus Prestained Protein Ladder (Fermentas), molecular weights given in kDa..

MALLS was used in order to investigate the oligomeric state of TatC in LMNG (**Figure 3.1.2 A)**. The UV elution profile from the S200 10/300 (GE healthcare) column used for the MALLS experiment was very symmetrical and the mass calculated for the total protein:detergent complex at each point across the peak was extremely consistent.

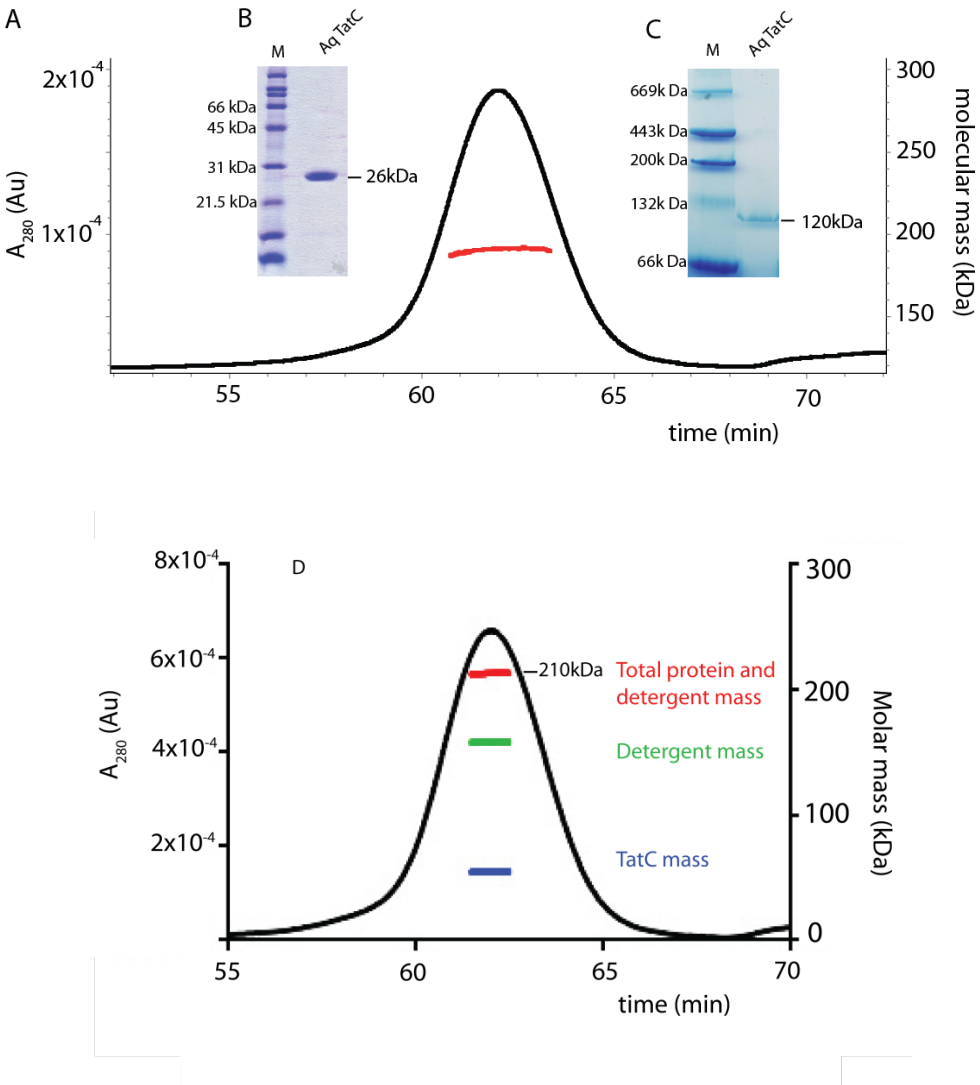


Figure 3.1.2 Assessment of the homogeneity of TatC in LMNG.

(A) Multi angle laser light scattering (MALLS) of 50µM TatC in LMNG. The black line gives the UV profile, and the red line gives the calculated molecular mass at each point across the peak for the TatC-LMNG complex. **(B)** 15% (w/v) SDS-PAGE confirmed the purity of the final TatC sample. **(C)** BN-PAGE of the sample showed a single band **(D)** Protein conjugate analysis of a 150 µM sample of TatC in LMNG. The apparent mass of TatC given by each technique is labelled.

Subsequent protein conjugate analysis (Steven Johnson, Sir William Dunn School of Pathology) showed that the protein component alone had a flat mass profile (**Figure 3.1.2 D**) but with a lower molecular weight. The mass profile therefore indicated that there was only one species of TatC present in the LMNG. Furthermore, experiments conducted at other concentrations of sample showed that the calculated mass was independent of concentration, and thus the species was stable and did not undergo any concentration dependent oligomerisation. The conjugate analysis gave the total protein mass as 53kDa. This was consistent with a dimer of TatC in this detergent.

The MALLS also showed that using the 100kDa MW cut off concentrator devices (Millipore) for this sample resulted in an undetectable (as measured by the refractive index) level of LMNG micelles being present in the final sample, which would be advantageous for any subsequent crystallisation trials.

Figure 3.1.2 B shows the final SDS-PAGE gel of TatC in LMNG, which confirmed the purity of the final sample, and **Figure 3.1.2 C** is the equivalent sample run on a blue native polyacrylamide gel (BN-PAGE) (Dr M. Krehenbrink, Department of Biochemistry, Oxford). BN-PAGE also showed a single band, which confirmed the single species seen by MALLS. The difference in the total mass of TatC seen by each of these techniques (marked on **Figure 3.1.2**) relates to how the different techniques work. SDS-PAGE is a denaturing technique, the TatC sample was boiled and then mixed with SDS to assist its partial unfolding. This treatment will tend to breakup any higher order complexes the protein may form. The SDS molecules will bind to the protein to give it a net negative charge and so allow it to move through the gel by electrophoresis, where its migration is dependent on its size and mass to charge ratio. In general, unfolded proteins bind the SDS in proportion to their molecular weight, and so their migration through the gel will be proportional to their overall size. However, the additional hydrophobic stretches present in

membrane proteins can bind SDS disproportionately, and so cause membrane proteins to run at a slightly altered mass (Rath *et al.*, 2009) and would explain why TatC runs slightly differently than expected (around 26kDa compared to the true molecular weight of 28kDa). MALLS is an 'in solution' technique run on the native protein in the purification buffer and can directly determine the mass of the overall protein:detergent micelle, as well as the mass of the protein component alone within this complex (by protein conjugate analysis). As the MALLS experiment is run in the standard purification buffer, it allows protein complexes and oligomeric states to remain intact. As the mass is measured by the amount of light scattered by the species present, the technique does not require the sample to be perturbed, for example by binding SDS or binding a dye, and so the mass that is observed is not altered is any way. This is in contrast to BN-PAGE, in which a coomassie G-250 dye binds to the protein:detergent complex to provide the necessary negative charge for electrophoresis. This dye can dramatically alter the amount of detergent/lipid bound in the protein:detergent/lipid complex (Crichton *et al.*, 2013) and so altering the apparent mass of the complex when run on the gel. This would explain the difference between the mass seen from the MALLS and the BN-PAGE; the coomassie dye can remove and replace some of the detergent bound with in the protein:detergent complex, whereas MALLS gives a measure of the unperturbed complex.

3.2 Crystallisation trials with *A. aeolicus* TatC in LMNG

TatC purified in LMNG was subject to broad screen crystallisation trials using the high throughput screens used previously (**Table 2.3.**). The 'Stura test' (**section 2.2)** was used to ascertain a starting concentration to use for these screens, which was around 8mg/ml. The

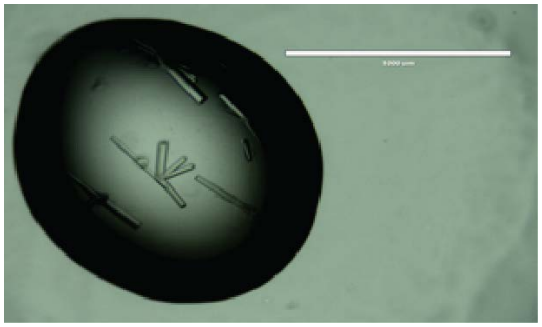


Figure 3.2.1 Crystals of TatC in LMNG grown in the MemGold broad screen.

Crystallisation condition was 0.2M ammonium sulphate, 0.02M NaCl, 0.02M sodium acetate pH4.0 and 33% w/v PEG 200, and protein concentration of 7mg/ml. Scale bar size is 1000µm

screens were set up in the same manner as described in **section 2.2**, at two mother liquor: protein drop ratios (50:50, 25:75) in a volume of 400nl, and at various temperatures. After a few weeks, crystals had grown in the broad screen MemGold (molecular dimensions) F11 condition, 50:50 drop ratio (**Figure 3.2.1)**. In order to test the diffraction from these crystals, a selection from the drop were harvested in the mother liquor, to act as the cryo protectant, and exposed to X-rays at the Diamond beam line I03. These crystals were found to diffract to reasonably high resolution with spots reaching to around 4.5Å in the best directions. An initial data set was therefore collected and processed as described for the cymal-4 datasets in **section 2.5,** using the automated data processing package, Xia2 (Winter, 2010). The data processing statistics for the best dataset collected from crystals in this drop are given in **Table 3.1.** These crystals were a different crystal form from the cymal-4 crystal form. They belonged to the rhombohedral space group H32, with a unit cell of 124.8, 124.8 and 214.5 Å along a, b and c respectively.

Information on the asymmetric unit (ASU) of the unit cell was again investigated at this stage. The solvent content analysis by MATTHEWS, as described in **section 2.5** was carried out on this dataset, and the result given in **Table 3.2.** This analysis gave very similar statistics to the Cymal-4 orthorhombic dataset, with the most likely scenario of two copies of TatC within the ASU.

Data collection	**Initial LMNG crystal hit**
Space group	H32
Cell dimensions *a*, *b*, *c* (Å) *α*, *β*, *ɤ* (°)	 124.8, 124.8, 214.5 90, 90 , 120
Wavelength Resolution (Å)	1.072 96.52-4.53 (4.65-4.53)
R_{merge} R_{meas} $I/\sigma I$ $CC_{1/2}$	0.051 (0.618) 0.055 (0.653) 20.2 (4.4) 0.94 (0.88)
Completeness (%) Redundancy Total observations Total unique	99.9 (100) 9.5 (9.9) 37427 (2864) 3928 (288)

Table 3.1 Data processing statistics for the crystal grown from TatC LMNG.

Number of copies in the ASU	**% solvent**	**Matthews coefficient, V_M (Å^{3}Da^{-1})**	**P (4.53)**	**P (total)**
1	78.8	5.82	0.01	0.01
2	57.7	2.91	0.73	0.59
3	36.6	1.94	0.26	0.40
4	15.6	1.46	0.01	0.00

Table 3.2 An estimate of the solvent content of the ASU. Calculated by the CCP4 programme MATTHEWS.

Calculation of the SRF by POLARRFN within the CCP4 suite of programmes (Winn *et al.*, 2011), as in **section 2.5**, revealed a peak on the κ=120° section indicative of the three-fold axis (**Figure 3.2.2 A**), and six peaks on the κ=180° axis derived from the crystallographic two-fold axis (shown in red, **Figure 3.2.1 B**). This dataset also gave an extra peak on this section of the SRF, shown in blue in **Figure 3.2.1 B.** This peak does not arise from the crystallographic axes. This indicated the presence of a potential extra symmetry axis within the ASU, due to the presence of two TatC molecules, related by non-crystallographic symmetry. However, this peak was not always present in the datasets collected subsequently (**section 3.4**), which implied this peak may be noise within the data. With one copy of TatC within the asymmetric unit, the solvent content of these crystals would be approximately 78%.

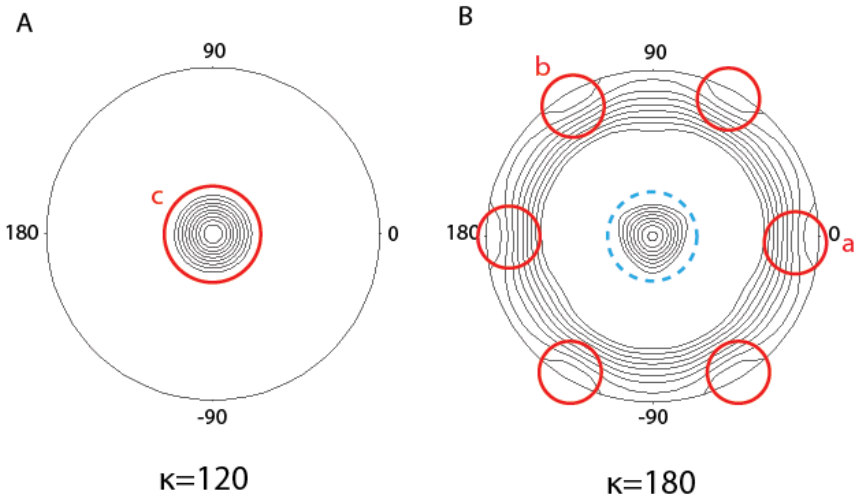


Figure 3.2.1 The self rotation function for the H32 crystal form of TatC

(**A**) κ=120° and the (**B**) κ=180° sections calculated with data from 15 to 4.53 Å resolution. Peaks in red correspond to the crystallographic axis, and the peak in blue is a non-crystallographic axis.

3.3 Optimisation of the TatC LMNG H32 crystal form

To optimise the crystals gained in the broad screen condition, the initial strategy involved altering the mother liquor conditions rather than altering the TatC LMNG sample. The following optimisation trays were set up in which the parameters of the mother liquor condition, such as pH, salt concentration and % PEG were varied (**Table 3.3** and **Table 3.4**).

Within two weeks of setting up the optimisation trays crystals were obtained in many drops of the trays. These crystals exhibited a similar morphology to those of the initial hit: long rods with a triangular cross section (**Figure 3.3.1).** Furthermore, unlike the highly variable crystals obtained in the cymal-4 optimisation trays, all the crystals grown appeared to show the same morphology, indicating just one crystal form was growing within these conditions. These crystals were therefore harvested and their diffraction tested in order to assess firstly if they were the same crystal form as the initial hit, and secondly if the resolution of the diffraction was reproducible.

Two crystals taken from each of the drops above provided a good example of the quality of the data collected from these optimisation trays. They each gave a full dataset to 3.97 Å and 4.00 Å respectively, and the space group indexed as H32 for both, with unit cell dimensions of 123.0, 123.0, 215.1 Å and 123.1, 123.1, 217.0 Å respectively.

These crystals gave an improvement in the maximal resolution of the initial diffraction from 4.53 to 4Å, so they were now of a similar quality to the two best datasets obtained from crystals grown in cymal-4. However, crucially, all the crystals shot from these trays were of same space group, and all had very similar unit cell dimensions. This was in sharp

			%v/v PEG 200						% v/v PEG 200					
			31	32	32.5	33	34	35	31	32	32.5	33	34	35
			1	2	3	4	5	6	7	8	9	10	11	12
[NaCl/ $(NH_4)_2SO_4$] (M)	A	0.01 / 0.2 0.015 /0.2 0.02 /0.2 0.03 /0.2	0.02M sodium acetate pH 3.84						0.02M sodium acetate pH 3.9					
	B													
	C													
	D													
	E	0.01 /0.2 0.015 /0.2 0.02/ 0.2 0.03/ 0.2	0.02M sodium acetate pH 4.0						0.02M sodium acetate pH 4.1					
	F													
	G													
	H													

Table 3.3 LMNG crystallisation optimisation screen 1. Based on 0.2M $(NH_4)_2SO_4$ /0.02M NaCl, 0.02M Sodium acetate pH 4.0, 33% v/v PEG 200

			% of initial molecular dimensions condition in water											
			100 %											93%
			100	99.5	99	98.5	98	97.5	97	96.5	96	95	94	93
			1	2	3	4	5	6	7	8	9	10	11	12
% [protein] relative to that in the initial hit	110	A												
	107	B												
	105	C												
	100	D												
	95	E												
	92	F												
	90	G												
	89	H												

Table 3.4 LMNG crystallisation optimisation screen 2. Based on 0.2M $(NH_4)_2SO_4$ /0.02M NaCl, 0.02M Sodium acetate pH 4.0, 33% v/v PEG 200

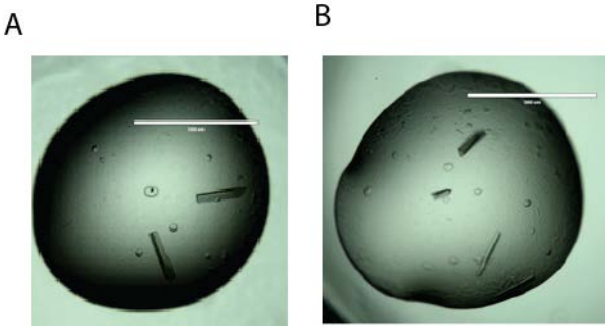


Figure 3.3.1 Crystals of TatC grown in the optimisation screens.

The positions in the screens that the crystals were obtained in are marked on **Tables 3.3** and **3.4**. Scale bar set to 1000µM.

contrast to that seen in the better of the cymal-4 crystals. The reproducibility of these crystals in this detergent was also seen from different batches of protein purified. These observations confirmed that this crystal form was suitable to use in phasing attempts.

3.4 Selenomethionine preparation of TatC in LMNG

As discussed in **section 2.5**, phasing of a structure with that has no suitable homology models to allow for molecular replacement can be achieved by two main methods, used alone or in combination; isomorphous replacement and anomalous dispersion techniques. In recent years, the most widely used technique to solve novel protein structure is anomalous dispersion, and the most popular anomalous scattering atom used is selenium (Walden, 2010). This is due to the approach pioneered by Henderickson *et al* in which recombinantly expressed proteins are grown in media supplemented with selenomethionine (Henderickson, 1990) which can then be stably incorporated into a protein in place of methionine. Crystals grown from this batch of protein therefore contains the anomalous element, with no need to soak in extra compounds. A significant advantage of this technique is that all the information required to phase the data is (theoretically) present in one crystal, so there is no issue with non-isomorphism, as in isomorphous replacement methods. This method also negates the time consuming trial and error involved in finding appropriate heavy atom soaks (Sun *et al.*, 2002). Anomalous dispersion techniques, using the selenomethionine approach, was therefore the first choice in attempts to phase the data from the LMNG crystals.

In order to obtain sufficient anomalous signal from selenomethionine containing protein crystals ideally every methionine in the protein will be substituted with selenomethionine. Protein preparations in which there is a mixture of selenomethionine and methionine in the

protein of interest can lead to difficulties in phasing, as the sites of the anomalously scattering elements needs to be located. If not all the sites are fully occupied, and to varying extents, this can lead to the failure to determine the anomalous substructure (Bushnell *et al.*, 2001). The issue in obtaining full selenomethionine incorporation is that bacterial cells are capable of synthesising their own methionine, and thus can incorporate this into the over-expressed proteins. To avoid this bacterial cells are grown in minimal medium supplemented with selenomethionine. This is designed to encourage the cells to take up the selenomethionine, and incorporate this into all their proteins. This results in high levels of selenomethionine incorporation, in most bacterial, *E. coli* expression strains (Van Duyne *et al.*, 1993).

As well as obtaining fully selenomethionine substituted protein, expression conditions must give a sufficient yield of TatC in order to obtain enough protein to carry out crystal optimisation trials. A reduction in yield compared to the native levels could be tolerated to some extent, as having identified a crystallisation condition more focused optimisation screens are set up, which utilises less protein than screening all the broad screen trays with various protein concentrations, temperatures etc. However, obtaining a good yield can often be difficult in minimal media. A growth condition therefore needed to be found in which the requirement for both full substitution and a workable yield, were met.

A commonly used and widely available medium for selenomethionine substitution is available from Molecular Dimensions. This is based on M9 minimal media, but is supplemented with all the amino acids (with the exception of methionine), vitamins and glucose. These supplements can help improve the yields of protein relative to a standard minimal medium . There is also a glucose-free nutrient mix which can be supplemented with alternative carbon sources such as glycerol (Molecular Dimensions). These minimal

media recipes formed the basis of the initial attempts to obtain fully selenomethionine substituted TatC protein.

Growth in Molecular Dimensions minimal media

Expression trials were carried out in the Molecular Dimensions minimal media in order to find a condition in which the protein would express. In order to do this, the GFP tag on the TatC construct was utilised to assess the levels of expression. This was achieved by harvesting 1ml of the pre induction culture and 1ml of the post induction culture, the supernatant discarded and the pellet re-suspended in 100µl PBS. The re-suspended cells were then transferred to the well of a black, 96 well clear bottomed plate (Greiner). The level of GFP fluorescence was then read using a HT fluorescence plate reader (SpectraMaxM5, Molecular devices), at an excitation wavelength of 485nm and emission wavelength of 538nm. Control fluorescence readings were also carried out with TatC-GFP expressed with the standard protocol in TB, to assess what fluorescence reading correspond to a good level of expression. The pre-induction reading was subtracted from the post-induction readings to give an overall change in fluorescence after overnight induction (**Figure 3.3.1)**. This illustrates that a good level of expression of the construct corresponded to a change in fluorescence of around 20,000 absorbance units.

For the analysis of selenomethionine labelling media, L56 cells were transformed as normal with the pWaldo TatC-GFP vector (**2.1**) and an overnight culture in 100ml LB set up from one colony. This culture was harvested, washed twice in PBS, and a 1:100 dilution used to inoculate a culture of molecular dimensions (MD) minimal medium supplemented with 40mg/l L-selenomethionine, (Acros Organics). The cells were grown at 37^{o}C for 11 hours until they reached an OD_{600} of 3.0. This was slightly lower than the OD_{600} that the

cells usually reach when grown in TB. The cells were then induced with IPTG and left overnight at 24^{o}C as normal.

The fluorescence readings for this growth are plotted (**Figure 3.4.1, MD1**) , it can be seen that there was very low expression under these conditions. The cells in this experiement grew mores slowly than cultures in TB and reached an OD_{600} of 3.0 compared to 5.0. To gain a higher cell density prior to induction, the growth was repeated, except the cultures

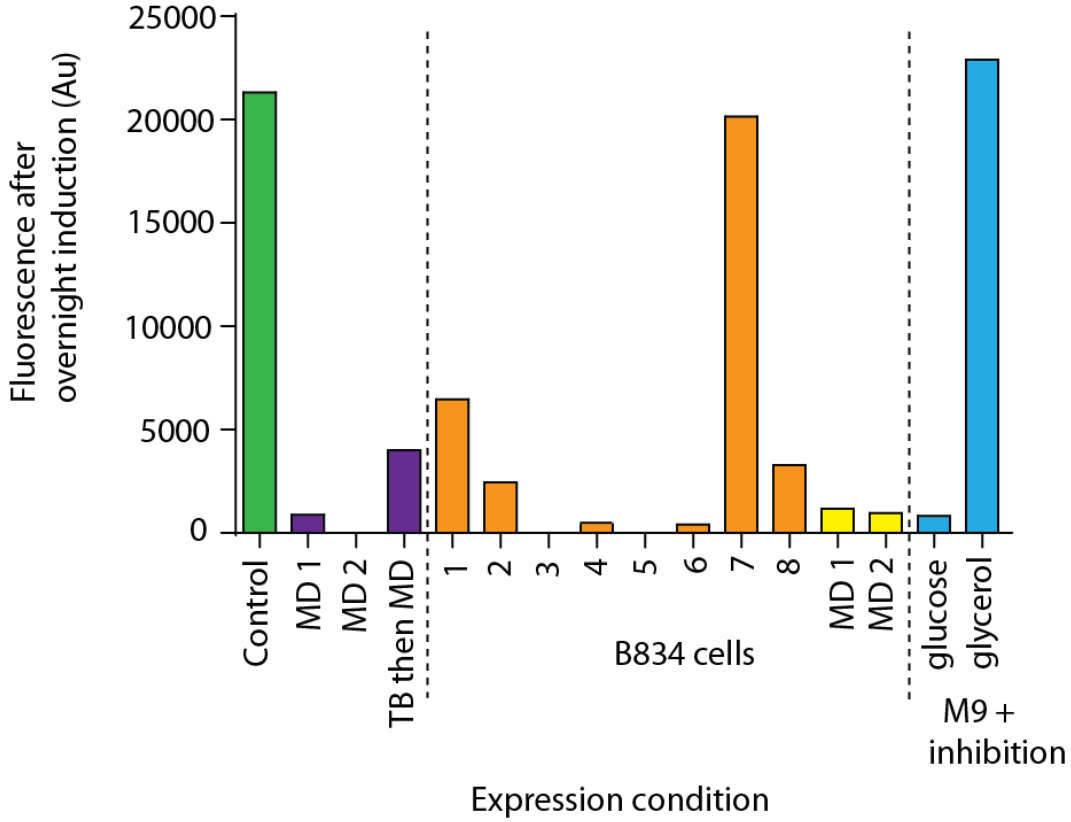


Figure 3.4.1 Expression trials of *A. aeolicus* TatC-GFP in various minimal media conditions.

Expression levels were monitored by the change in GFP fluorescence of the cells after overnight induction. **MD1 and MD2-** growth of L56 cells in Molecular Dimensions (MD)selenomethionine minimal medium. **TB then MD-** growth in TB until OD_{600} reached for induction, then growth in the MD medium. **B834 cells: Conditions 1-8:** in rich medium as **Table 3.5. MD1 and MD2** the expression parameters identified in Condition 7 repeated in the MD selenomethionine minimal media. **M9 plus inhibition;** in L56 cells, using either **glucose** or **glycerol** as the carbon source.

were left to grow overnight after being set up in the afternoon. This allowed them to grow for much longer (~16 hours) and reach a higher OD_{600} (4.0). However, after overnight induction the fluorescence reading (**Figure 3.4.1, MD2**) indicated that again there was no expression of the protein in this media.

The low levels of expression could have been attributed to the cells being under stress growing in the minimal media. One technique that has been used previously is to grow cells in rich media until they are almost at the correct OD_{600} for induction, then harvesting them from the medium, washing the cells three times in PBS before using them to inoculate a culture of the minimal medium. This method was tried for TatC. The cells were grown to OD_{600} 4.0 before being harvested, washed and used to inoculate (without overall dilution) a culture of the MD minimal medium. This approach was tried using either glycerol or glucose as the carbon source. The fluorescence readings showed expression, but at roughly $1/8^{th}$ the level of the normal protein expression (**Figure 3.4.1, TB then MD**).

In order to see if scaling up this cell growth condition would lead to sufficient protein for crystal trials, 8L of cells were grown in this condition and the TatC purified as normal. The yield of protein was extremely low, 0.15mg from 8L compared to 1.2mg from 6L TB, which was in proportion to that expected from the fluorescence. However, this quantity of protein could still have been useful for setting up some crystal reproduction screens, if the protein was fully labelled with selenomethionine. In order to asses whether this was the case, native mass spectrometry was carried out on the selenomethionine substituted TatC protein (Dr Julien Marcoux, Department of Chemistry, University of Oxford). The mass difference of selenomethionine compared to methionine is 47Da. Thus with four methionines present in the *A. aeolicus* TatC sequence, SeMet TatC would be expected to show a 188Da mass increase compared to native TatC. The native mass spectrometry was able to obtain a very accurate mass of TatC, which was necessary to observe this relatively

small mass increase. The experiment was also able to assess if there was a single TatC species present, or whether there was a mixture of protein species with differing levels of SeMet incorporation.For this sample, there were two species present TatC with all Met and no SeMet incorporated, and TatC with all SeMet and no Met (Data not shown). Furthermore, these were present in an approximate 50:50 mixture. This sample would have proven very difficult, if not impossible to used for phasing the data, and no efforts were made in attempting to crystallise protein made from this expression condition.

Growth in B834 cells

L56 cells are still able to incorporate endogenously synthesized methionine into the TatC protein, despite the presence of externally added selenomethionine. One way of avoiding this problem is to use the *E. coli* expression strain B834 (DE3), which is a methionine auxotroph (Invitrogen). This cell line is knocked out for methionine biosynthesis, and thus has been specially designed for selenomethionine expression. This avoids any risk of the cells introducing their own methionine into the protein and so the cells have to use the supplied selenomethionine (Walden, 2010). However, a potential issue with using this cell line is that it has not been optimised for membrane protein expression, and so yields maybe low.

Initial expression trials using B834 cells were carried out in rich (TB) medium in eight different conditions (**Table 3.5)**. This would show if it is possible to express TatC in rich medium, before attempting growth in the minimal media. If an expression condition could be identified that gives growth in TB, this would form the basis of expression trials in the minimal media.

Condition number	[IPTG] mM	Induction temperature (^{o}C)	OD_{600} at induction
1	0.1	24	4.0-5.0
2	0.4	24	4.0-5.0
3	1	24	4.0-5.0
4	0.1	30	4.0-5.0
5	0.4	30	4.0-5.0
6	1	30	4.0-5.0
7	0.1	24	0.6
8	0.4	24	0.6

Table 3.5 Expression conditions tested for TatC-GFP in the B834 (DE3) strain. These trials were conducted in rich medium (TB and 0.4% (v/v) glycerol). The cells were all induced overnight, and the fluorescence readings taken as described previously.

A high level of expression was only seen in one condition, number 7 (**Figure 3.4.1, B834 cells 1-8)**. This condition (induction at OD_{600} of 0.6 with 0.1mM IPTG, at 24^{o}C overnight) gave expression that was comparable to the L56 cells in rich medium. This condition was used as a start point for expression trials in the minimal medium (Molecular Dimensions). Cells were either inoculated into the minimal medium straight from the overnight culture (**MD1** in **Figure 3.4.1**) or were first grown in TB before being extensively washed and then inoculated into the minimal media (**MD2** in **Figure 3.4.1**). However, in both these cases there was no detectable expression of TatC.

Inhibition of the L56 methionine biosynthesis pathway

The attempts to gain near 100% incorporation of SeMet into TatC by expression in minimal media with a methionine auxotroph strain, B834, gave very low expression. This

was not entirely unexpected, given that the L56 strain was initially chosen for TatC expression as the strain has been optimised for membrane protein expression. Thus, expression trials in the L56 cells were revisited as these had at least given some protein expression in the minimal medium. However, a way of inhibiting the cells own methionine synthesis pathways needed to be found, in order to stop normal methionine production and subsequent incorporation into the TatC protein. This was attempted by feedback inhibition of the methionine biosynthesis pathway (Doublié, 2007). This feedback method involved growth in standard M9 minimal medium (**Table 3.6**), at 37^{o}C until the cells reach the OD_{600} required for induction. At this point, a mixture of certain amino acids were added to the medium to inhibit the cellular methionine biosynthesis pathway. Isoleucine, lysine and threonine act to inhibit an aspartokinase, the first enzyme of the methionine biosynthesis pathway. Leucine and phenylalanine act in synergy with lysine. The cells were grown for another 30mins at 37^{o}C, before induction with the normal concentration of IPTG.

This protocol was tested with either glucose or glycerol, at different concentrations, as carbon sources. **Figure 3.4.1, 'M9 plus inhibition'** showed a representative reading with glucose, indicating there was no detectable expression in this M9 media. However, in the presence of glycerol there was a very high level of TatC-GFP expression.

The high level of TatC expression did not necessarily mean all the protein expressed by the cells was fully substituted with selenomethionine. To test this, 8L cultures of this particular condition were set up and the TatC protein purified in LMNG as described previously. Samples were subjected to native mass spectrometry, the spectra are shown in **Figure 3.4.2**. This gave the mass of methionine-containing (native) TatC as 29166Da, and SeMet TatC as 29361Da. The mass addition of 195Da corresponds very well to the 188Da mass difference expected if all four methionines are selenomethionine substituted. Notably, the mass spectrometry identified only 100% SeMet incorporated TatC protein. The series of

Component	**Reason for use**	**Final concentration in 1L media**
10x Stock of M9 salts 60g Na_2HPO_4 30g KH_2O_4 10g NH_4Cl 5g NaCl	Provides a source of nitrogen for amino acid synthesis, a source of phosphate and a pH buffer	1x
$MgS0_4$	Source of Mg and sulphate	2µM
$CaCl_2$	Source of calcium	0.1µM
Thiamine	Vitamin B_1	$5x10^{-5}$ % w/v
Glycerol	Source of carbon and energy	0.5% w/v
Kanamycin	Maintains selection for the pWaldo TatC-GFP plasmid	50µg/ml
Components added 30 minutes prior to induction		
Amino acid mix Threonine Lysine Phenylalanine Leucine Isoleucine Valine	Allows for the inhibition of the cells own methionine synthesis by metabolic feedback inhibition.	100mg/L: Threonine Lysine Phenylalanine 50mg/L: Leucine Isoleucine Valine
Selenomethionine	For incorporation into proteins instead of methionine	60mg/L

Table 3.6 Minimal medium recipe for the metabolic inhibition of methionine biosynthesis. The components in the last two columns were added 30 minutes prior to induction, in order to inhibit the methionine metabolic pathway in the bacteria.

additional peaks seen for the native and the SeMet sample probably correspond to two non-covalent lipid adducts bound to TatC. These adducts could correspond to cardiolipin which is of a similar mass to the 1400Da adducts observed.

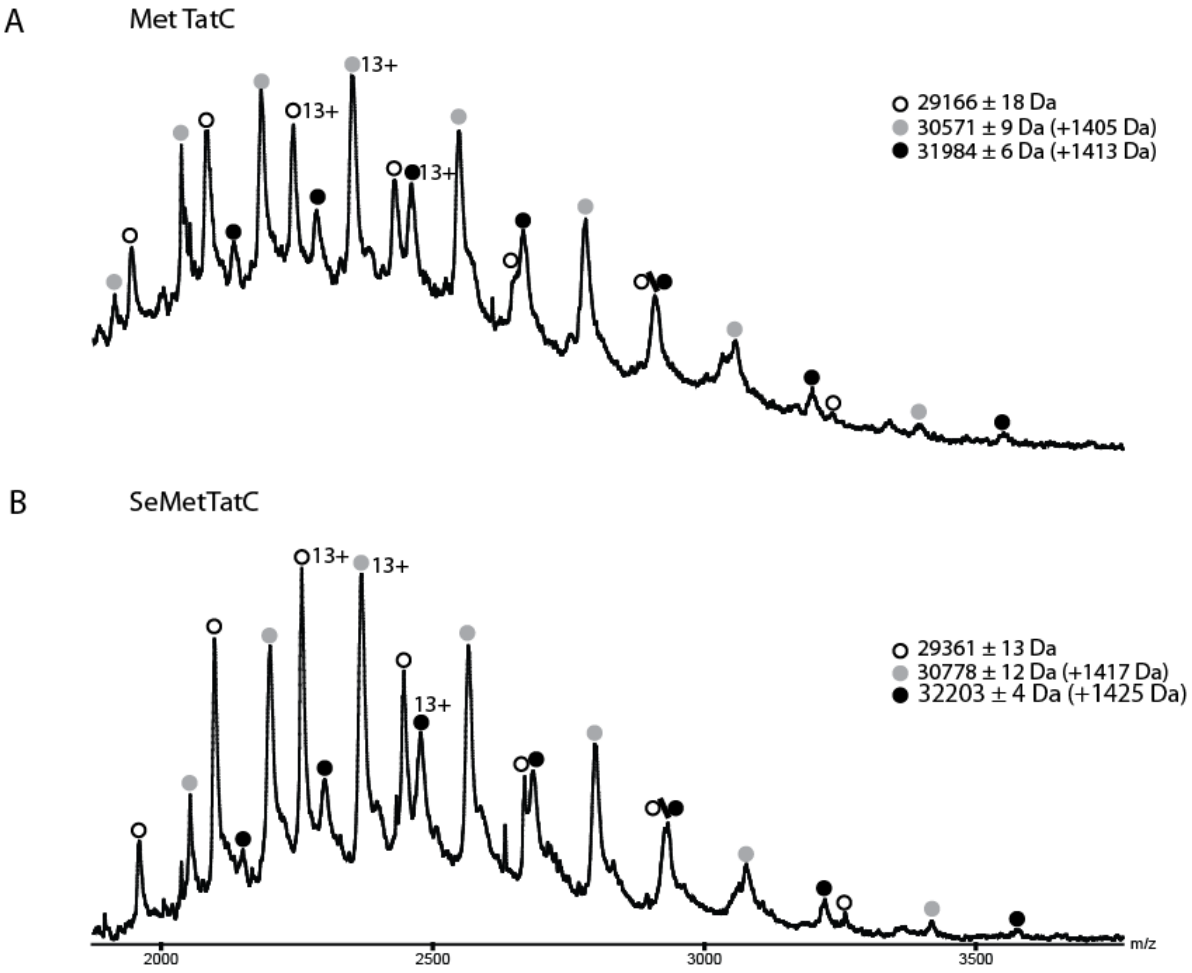


Figure 3.4.2 The native mass spectra of TatC.

Spectra of (**A**) the native TatC protein and the (**B**) the SeMet labeled protein. Both proteins show lipid binding (cardiolipin ~1400Da). The spectrum also showed this species was the only species present.

3.5 Crystallisation trials of SeMet TatC

The growth condition gave fully substituted selenomethionine TatC, and with a very good yield (3mg from 8L culture). In order to try to grow crystals from this condition the protein was purified in LMNG as described previously, except that the final size exclusion buffer was supplemented with 5mM β-mercaptoethanol (Sigma) to prevent the oxidation of the selenium atom from -2 to +2. The absorption of these oxidation states is different, and so the anomalous signal will be diminished if a mixture of the reduced and oxidised form of the element is present in the crystal. Furthermore, the reduced form of selenium is less sensitive to radiation damage (Holton, 2007).

The SeMet TatC protein was concentrated to a similar concentration to the native protein and optimisation screens around the native TatC crystallisation in **Table 3.3** and **Table 3.4** were set up. Alongside these screens, more high throughput (HT) broad crystallisations screens were set up (**Table 2.3**) in case the selenomethionine substitution had changed the surface properties of the protein enough to prevent formation of the same crystals as the native protein.

Fortunately, the SeMet version of the protein formed crystals in the condition in **Figure 3.2.1** with similar efficiency as the native protein. Crystals had a similar triangular 'rod' morphology and again grew within two weeks. The best crystals grow in the optimisation screen in **Table 4** and are shown in **Figure 3.5.1 A** and **B.**

The optimised crystals were harvested, and flash frozen in mother liquor in order to be exposed to X-rays at the Diamond I02 beam line. This beam line is tuneable meaning the wavelength, and thus the energy of the X-ray beam, can be altered. By incrementally altering the energy of the X-rays exposed to the selenium-containing crystals it is possible

to measure the fluorescent emission of X-rays from the crystal at right angles to the incident beam. This experiment is termed a fluorescent scan.

A fluorescence scan was carried out across the selenium K-edge (**Figure 3.5.2 A)** which showed the increase in detected fluorescence at the characteristic X-ray energy corresponding to the absorption by selenium of X-ray photons. These X-rays are of the appropriate energy to induce a transfer of an electron from an inner, K-electron shell of the selenium atom to one of the unoccupied orbitals of higher energy, or remove it completely from the atom. Subsequently, there is a corresponding fluorescent emission of a photon as an electron from a higher shell drops back to fill the hole in the inner electron shell. It is this fluorescence which is measured in the scan. The useful outcome of X-ray absorption by selenium is that the total coherent X-ray scattering from the crystal is reduced, and also shifted in phase. The atom is now said to scatter 'anomalously'. The result of the anomalous scattering by selenium within the crystals is the centrosymmetry of reciprocal space is broken; both the intensity and phase of Friedel mates F_h and F_{-h} are no longer the

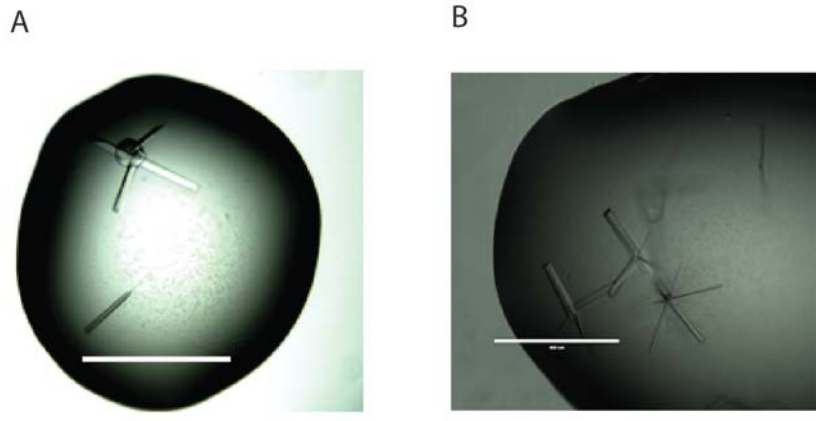


Figure 3.5.1 Crystals of the SeMet sample of TatC.

Crystals grown in the optimizations screens in **Table 3.4.** SeMet TatC purified in LMNG at a final concentration of 8mg/ml. Scale bar in (**A**) set to 1000µm, in (**B**) to 400µm.

complex conjugate of one another. The intensity of each member of the Friedel pair must be measured in the diffraction experiment in order to determine their structure factor amplitudes. Anomalous difference Patterson maps, calculated from the corresponding

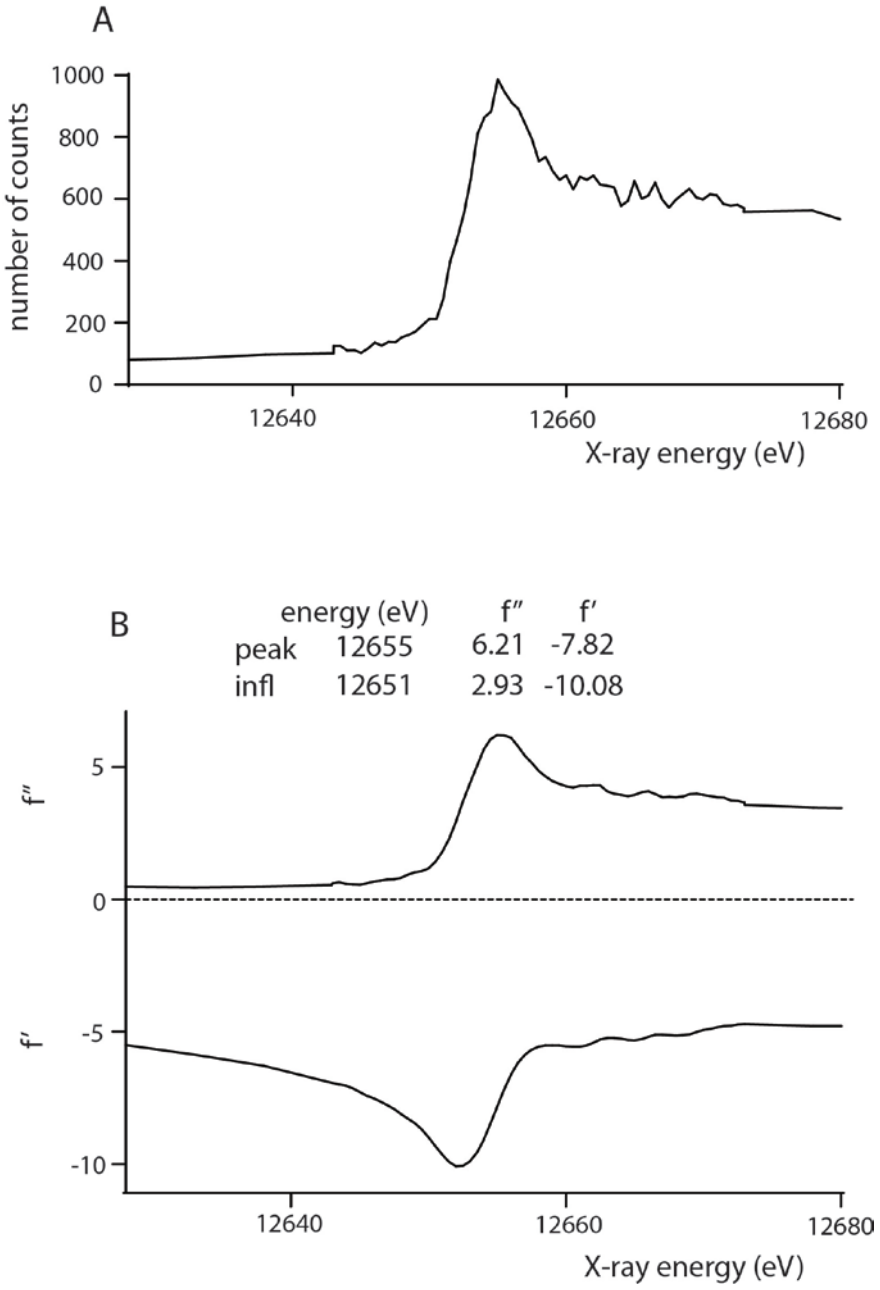


Figure 3.5.2 X-ray energy scan carried out on the SeMet crystals at the selenium edge.

(A) Fluorescence scan showing x-ray energy versus the number of photons counted emitted from the crystals. **(B)** The variation in f' and f" values calculated from the fluorescence scan by the program CHOOCH. The energy values at the peak and inflection points, along with the number of anomalous electrons, are quoted.

amplitude differences between $|F_h|$ and $|F_{-h}|$, can allow the positions of the anomalously scattering selenium atoms to be derived. Knowledge of the selenium atom substructure can subsequently allow an initial phase distribution for each reflection to be calculated (Rupp, 2009).

As well as the mass data from the native mass spectrometry experiment, the fluorescence scan demonstrated the presence of selenium within the crystals. The scan then allowed for the calculation of the real, f' and imaginary, f" components of the anomalous scattering factors (**Figure 3.5.2 B**) by use of the programme CHOOCH (Evans & Pettifer, 2001). For data collection, the wavelength of the X-ray beam was altered to the 'peak' wavelength at which there was maximal anomalous differences, f". For this particular beam line, the peak in the fluorescence scan corresponded to an energy of 12655eV, with a corresponding X-ray wavelength of 0.98 Å.

Datasets were collected from the SeMet crystals grown in the optimisation trays (a selection of crystals are shown in **Figure 3.5.1**). The crystals all diffracted allowing collection of complete datasets in the H32 space group. Data were processed using the Xia2 automated processing suite (Winter, 2010) run in anomalous mode to ensure that Friedel pairs were not merged. The maximal resolution of these crystals was slightly poorer than those from the native protein optimisation trays. A summary of the resolutions and anomalous correlation within each dataset are given in **Table 3.7**.

These datasets underwent extensive reprocessing in order to try and maximise their resolution, reduce the noise and boost the anomalous correlations. The datasets were then used in autoSHARP (Vonrhein *et al.*, 2007) in a phasing attempt. Unfortunately, this failed to find any selenium sites in the data. However, these experiments did confirm the presence of selenium within the protein crystals, and proved that the well diffracting

crystal form is reproducible for SeMet TatC. In order to find the selenium sites within the data and thus solve the structure, the resolution of the diffraction from the crystals would need to be slightly higher.

Crystal	Resolution (Å)	Overall anomalous correlation
1	4.11	0.369
2	4.7	0.772
3	4.67	0.584
4	4.7	0.262
5	4.64	0.407
6	4.44	0.181

Table 3.7 The maximum resolution and overall anomalous correlations for seven SeMet datasets. Collected from the SeMet version of the H32 LMNG crystal form. Resolution determined by R_{merge}, $I/\sigma I$, and completeness.

3.6 Structure solution of TatC

Bigger crystals of the crystal form above grew in the same optimisation trays as those described in **section 3.5**, with slightly altered protein concentrations, after a subsequent two weeks of growth compared to those in **section 3.5.** The new crystals (**Figure 3.6.1**) had the similar triangular rod morphology as those harvested previously, but were roughly twice the size. This difference in size was probably due to a combination of the longer growth time and the difference in the concentration of protein within the drops. These

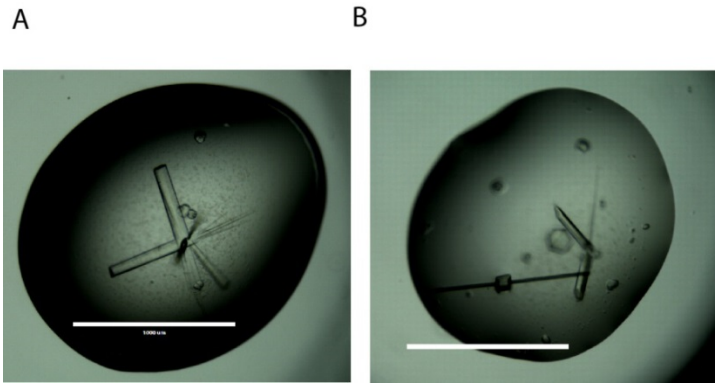


Figure 3.6.1 Crystals of SeMet TatC in LMNG.

Crystals grown in the same optimisation trays as those in Figure 3.3.3 at a protein concentration of 8.65mg/ml and a drop ratio of 40:60 protein:mother liquor. Scale bar set to 1000µm.

crystals were harvested in mother liquor as before and flash frozen for data collection at I04-1 beam line at Diamond.

The I04-1 beam line, unlike I02, is not tuneable, and the X-ray wavelength is fixed. For this data collection the beam was set at 0.92 Å with a corresponding energy of 13530eV. As this is above the selenium K-edge the X-ray photon will still have appropriate energy to interact with the selenium K-shell electrons and scatter anomalously. Datasets collected at this wavelength should therefore still contain anomalous signal to allow for phasing.

Two datasets were collected from one of the crystals in **Figure 3.4.1A**. The first dataset was collected using a strategy designed for optimal anomalous completeness, as advised by EDNA (Incardona *et al.*, 2009) and summarised in **Table 3.9**. In order to achieve anomalous completeness, not only does the unique wedge of data covering the asymmetric reciprocal unit need to be collected, but also the centrosymmetrically opposed wedge or Bijvoet pairs from symmetry related wedges must be collected to ensure an intensity measurement for each member of the required Friedel pair is measured.

The first dataset underwent automated processing by Xia2 (Winter, 2010), using XDS and SCALA, and gave a resolution of 3.62 Å. This was significantly better than any TatC LMNG crystal, SeMet or native, ever shot previously. The data processing statistics (**Table 3.8**) also show good anomalous correlations in the data (0.578 overall and 0.019 in the inner shell). Inspection of a selection of the data processing graphs for this dataset (**Figure 3.6.2**) showed that the strong anomalous correlations were present to moderate resolution. The anomalous completeness of this dataset was not quite 100% (**Figure 3.6.2 C**) despite 90^{o} worth of data being collected which should be sufficient given the symmetry of the space group. Upon inspection of the R_{merge} for this dataset against batch number (**Figure 3.6.2 A)**, it became apparent that the data from the crystal was very consistent during the whole length of the data collection, and the R_{merge} for batches at the end of the run was still very low (<0.1). This implied that the crystal was not suffering from radiation damage from its exposure to the beam and that more data of comparable quality from this region of this crystal could potentially have been collected to boost the anomalous completeness of the data obtained.

Further data collection was carried out by furhter exposing the same region of the crystal using the data collection parameters given by EDNA, except the number of images were doubled (so 180^{o} of data were collected), to ensure maximal anomalous completeness and to increase the redundancy of the data collected. The main problem with data for phasing by anomalous scattering is the signal to noise level of the data. The measured intensity differences between the Friedel pairs is very small compared to the intensity measurement itself. Given that the signal to noise of the measure intensity data for each reflection is a function of the square root of the number of observations, increasing the redundancy of the data increases the signal to noise of the intensity measurements for each reflection. This can greatly improve the accuracy and precision of the anomalous differences between

Data Processing	**Dataset 1**	**Dataset 2**
Space group	H32	H32
Cell dimensions α, β, ɤ (Å) *α, b, c*	 124.2, 124.2, 216.0 90, 90, 120	 124.2, 124.2, 215.6 90, 90, 120
Wavelength Resolution (Å)	0.92 76.13-3.62 (3.71-3.62)	0.92 48.26-3.79 (3.87-3.79)
R_{merge} $I/\sigma I$ $CC_{1/2}$	0.032 (0.610) 22.8 (2.2) 0.97 (0.75)	0.044 (0.772) 27.3 (3.6) 0.99 (0.91)
Completeness (%) Redundancy Anomalous correlations Anomalous completeness (%) Anomalous multiplicity	96.2 (97.9) 5.1 (5.4) 0.578 (0.019) 81.8 (82.1) 2.9 (3.0)	99.6 (100) 9.8 (10.5) 0.553 (0.036) 99.5 (100) 5.2 (5.3)
Total observations Total unique	37124 (2862) 7236 (529)	64284 (6570) 5095 (483)

Table 3.8 The data processing statistics for two datasets from one of the SeMet TatC crystals. Crystals shown in **Figure 3.4.1 A.**

Friedel pairs (Rose, 2001). Anomalous completeness was achieved, as shown in the processing statistics for this dataset (**Table 3.8 Dataset 2**). The automated processing for this dataset by Xia2 (Winter, 2010) was carried out as for Dataset 1, expect the scaling programme 'Aimless' was used for scaling instead of SCALA (Evans, 2006). Furthermore, the resolution was still reasonably high for this dataset, 3.8 Å, and the anomalous correlations were comparable to those in Dataset 1.

Finding selenium sites

The two processed dataset .mtz files were merged together using the programme 'CAD' (CCP4). This merged mtz file was then searched for potential selenium sites, by use of the

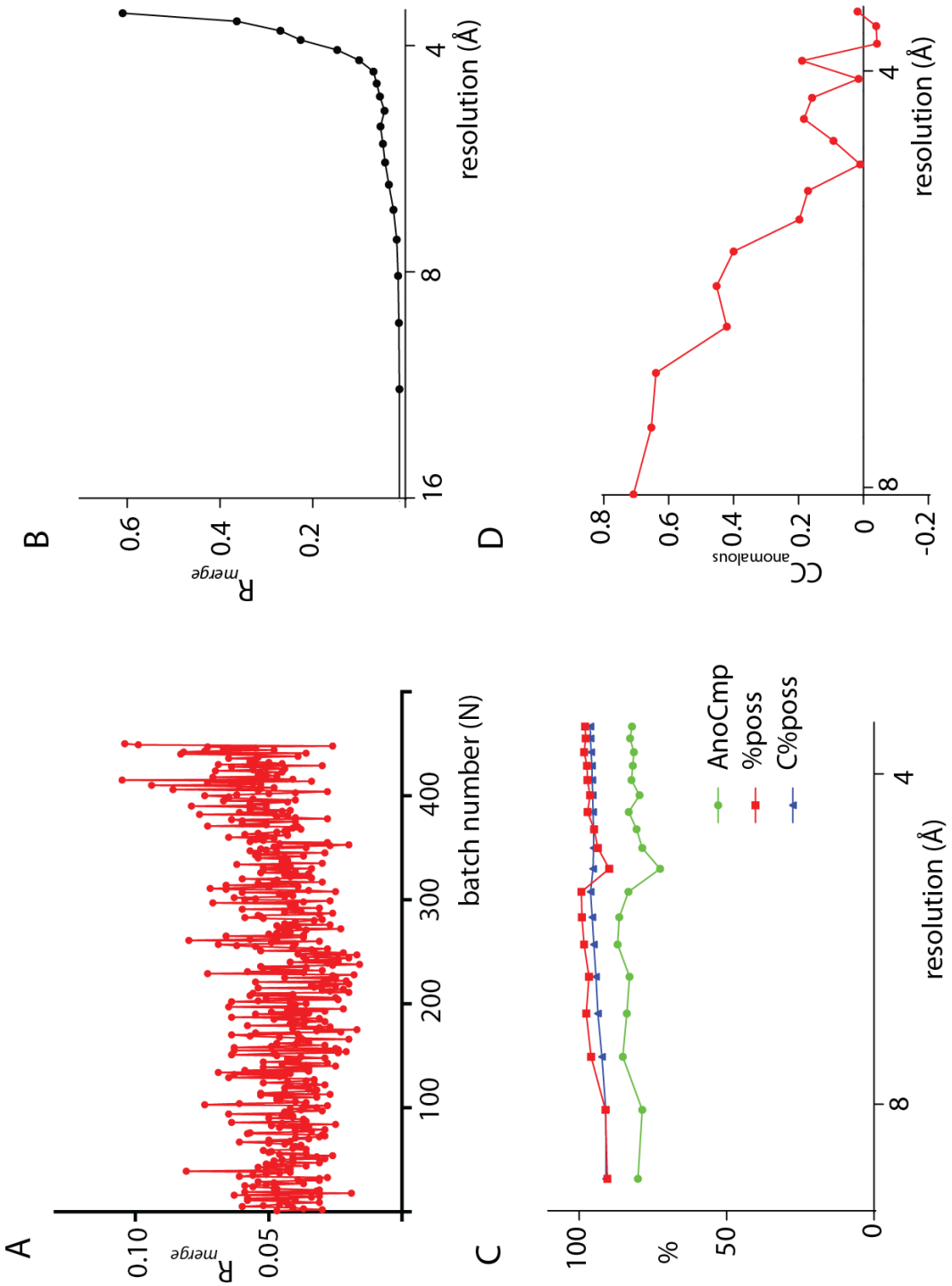


Figure 3.6.2 Data processing graphs for Dataset 1

Dataset collected from SeMet TatC crystals at the I04-1 beam line, and processed using Xia2. **(A)** R_{merge} verses image (batch) number **(B)** R_{merge} verses resolution, **(C)** The completeness of the data, with the % anomalous completeness given in green **(D)** The anomalous correlation plotted as a function of the resolution of the data.

programme Hyss (Grosse-Kunstleve & Adams, 2003), from the Phenix suite of crystallographic programmes. Hyss (Hybrid Substructure Search) uses both direct search methods and Patterson based search methods to identify potential sites of anomalous signal that account for the differences in the structure factor amplitudes between the Friedel pairs in the dataset. Patterson methods involved the calculation of Patterson maps from the fourier transform of the square of the structure factor amplitudes (or the square of the difference in the amplitudes between Friedel pairs, in the case of anomalous Patterson maps), and they contain interatomic distance information. Self vectors are the interatomic vectors between an atom and the symmetry related copy of itself. Harker sections are the special planes onto which self vectors between an atom and its symmetry related copy are found. When plotted for the anomalous difference Patterson maps, a Harker section will provide information on the positions of the symmetry related selenium atoms in the unit cell. An example of the w=0 Harker section of the anomalous Patterson for dataset 1 (calculated by autoSHARP (Vonrhein *et al.*, 2007)) demonstrates peaks arising from the presence of the anomalous scatterer (**Figure 3.6.3)**.

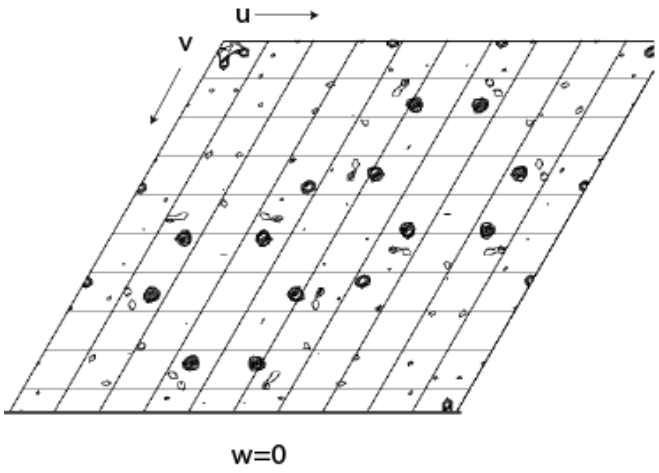


Figure 3.6.3 The w=0 Harker section of the anomalous difference Patterson map of dataset 1.

Calculated using 25-5Å range of data**.** The significant peaks arise from the presence of the anomalously scattering selenium atoms.

HySS was tasked with finding four selenium sites, as it was assumed from the lack of consistency seen in the SRF non-crystallographic peaks there was one copy of TatC within the unit cell, and TatC has four selenomethionines present. The programme was able to locate three potential sites. The positions of these three sites were then refined using the programme SHARP (Statistical Heavy Atom Phasing and Refinement (de La Fortelle & Bricogne, 1997)). This is a heavy atom refinement programme which uses maximum likelihood (ML) methods and allows factors such as the co-ordinate positions, B-factors and occupancies of the sites to be refined, as well as the f' and f" components of the anomalous scattering to be altered. For these datasets, the f' and f" values had to be estimated by the programme CROSSEC (CCP4) as the fixed wavelength beam line did not allow these to be experimentally determined by fluorescence scan. CROSSEC estimated, f'=-2.0 electrons and f"=3.4 electrons at a wavelength of 0.92 Å.

An initial Wilson plot of the data gave an estimate of the overall B factors of the data as 130Å^2, and so the B factors for these three initial selenium sites were then fixed at this value. The occupancies of these sites were then refined, and the first two sites identified remained occupied (occupancy =1.0) but the occupancy of the third site was severely diminished. This site was therefore removed from the heavy atom substructure. Subsequent B factor refinement of the remaining two sites was carried out, upon convergence coefficients for a map of the Log-Likelihood Gradient (LLG) were calculated.

The SHARP LLG maps (de La Fortelle & Bricogne, 1997) were now inspected. These ‘residual maps’ showed significant positive density at two separate positions, indicating the presence of two extra anomalous scattering atom sites, which were added to the substructure. All four of these potential sites underwent a final round of refinement in SHARP in which all the parameters, except the site occupancy, were refined until convergence. After convergence SHARP calculates a phase probability distribution for

each reflection and the centroid phase (Blow & Crick, 1959). A calculated figure of merit (FOM or *m*) provides a level of confidence for the calculated centroid phase for each reflection, and is used to weigh down poorly phased reflections. The centroid protein electron density maps are calculated from the observed structure factor amplitudes of each reflection, decreased by the FOM, and the centroid phase for the reflection. The overall FOM for phasing with these datasets converged to 0.29 (acentric reflections) and 0.098 (centric reflections).

As the anomalous data had only been collected at one wavelength (Single wavelength anomalous dispersion, SAD) the phase ambiguity, present as the probability phase distribution for each reflection, needs to be resolved. This can be achieved by collecting data from the same crystal at a different wavelength, the multi wavelength anomalous dispersion method (MAD) which allows for dispersive differences to resolve the phase ambiguity (Hendrickson & Ogata, 1997). As a fixed wavelength beam line had been used for the data collection discussed here, this approach was not possible.

SAD can also be combined with the isomorphous replacement method, by collecting dataset from crystals with the addition of a heavy atom (SIRAS). However, as discussed previously, (**section 3.3**) heavy atom derivitisation can lead to problems with non-isomorphism and finding a derivative that binds within the crystal can be very time consuming (Sun *et al.,* 2002). Alternatively, the development of powerful solvent flattening programmes have allowed for density modification to resolve this ambiguity within the SAD data (Wang, 1985). The initial centroid electron density map can be thought of as a superposition of the map of the molecule and a structure calculated with the wrong phases. In principal, only one of these will have electron density that is consistent with prior prejudices about electron density distributions. The centroid electron density map is therefore subjected to methods that exploit what is known about the physical

properties of protein structure and the properties of the solvent. These methods are termed solvent flattening, or density modification, and they attempt to improve the quality of the map.

Density modification procedures to resolve the phase and hand ambiguities

The solvent flattening programme used within SHARP is Solomon (Abrahams & Leslie, 1996). This makes use of a 'solvent mask' applied around regions defined as solvent, and it carries out 'solvent flipping'. This method inverts the grid point values in the solvent region, rather than setting them to a predefined, low value (as in solvent flattening). Phases and modified structure factor amplitudes are then calculated from the flattened map. These calculated values are then recombined with the experimental phase probability distributions, in such a manner as to minimise the bias involved in combining the new and experimental phases, and used to produce a new map. The solvent flattening process is then repeated on the new map in an iterative manner.

Solvent flipping by Solomon was run on the centroid maps calculated in SHARP, using an estimated solvent content of 78% based on the MATTHEWS solvent content predication for this crystal form, given one copy of the protein in the asymmetric unit (**Table 3.2**). Solomon outputs statistics for each round of the solvent flattening procedure, and the statistics can be compared for the starting map and for the final, solvent flattened map.

The I(contrast) statistic gives the ratio of the standard deviation of the electron density of the solvent before the flattening procedure compared to the standard deviation of the electron density of the protein. The solvent electron density is expected to be low and for the protein it should be high. The I(contrast) ratio is expected to drop during successful rounds of solvent flattening. The I(SigmaA) statistic gives the overall correlation on $|E|^2$

calculated in SIGMAA. This is the correlation between the square normalised structure factor amplitudes of the observed data and the modified map. This correlation is expected to rise during the solvent flattening procedure.

The substructure solution has been carried out in Patterson space, which is inherently centrosymmetric, thus the substructure solution is independent of the final hand. Both this hand and the inverse hand of the substructure will satisfy the anomalous signal. For the H32 space group the inverse hand is obtained by inversion of the substructure enantiomoph, i.e. changing all the fractional coordinate vectors from x to $-x$. However, the final structure will only be correct in one of these hands. The correct set of phases should yield an electron density map that has features consistent with protein density. As 'protein-like density' is essentially what is tested for in solvent flattening programmes, the solvent flattening can be carried out in both hands to see which hand yields the better statistics. The corresponding protein centroid electron density map is therefore calculated in both hands, and both maps subject to solvent flattening, Only the correct enantiomorph of the substructure is expected to give a map that improves during cycles of solvent flattening (Rupp, 2009). Solvent flattening can therefore be used to resolve both the inherent 2-fold phase ambiguity, but also the hand ambiguity of the substructure solution.

For solvent flattening in the original hand, the I(contrast) started at 0.56 and reduced to 0.31, and the I(SigmaA) increased from 0.19 to 0.44. For the inverse hand, the I(contrast) had a larger decrease, from 0.55 to 0.27 and the I(SigmaA) had a greater increase, from 0.21 to 0.51, which implied the inverse hand of the substructure was the correct hand. However, these statistics were not drastically different, and might not have been reliable alone in determining the correct hand. To investigate if the inverted hand gave a plausible protein-like electron density map, the solvent flattened maps were inspected in COOT

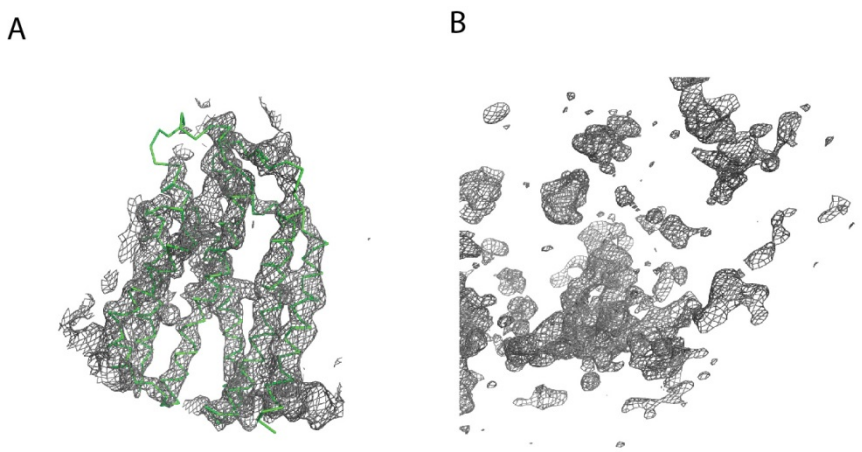


Figure 3.6.4 The result of the solvent flattening at 78% solvent.

Solvent flattening run in the (**A**) inverse hand and (**B**) the original hand contoured at 1.5σ**.** Electron density maps calculated using the solvent flattened phases as given by Solomon, and the observed amplitudes. The cartoon of the final structure is shown in green within the density in (**A**).

(Emsley & Cowtan, 2004). The maps confirmed the protein-like electron density (**Figure 3.6.4 A**) for the inverted hand, compared to the clear noise map produced for the initial hand (**Figure 3.6.4 B**).

Refinement of heavy atom parameters against more datasets

More data were obtained to help improve the phases obtained in the calculation of the initial experimental electron density map to help in the subsequent task of model building. This was achieved by collecting subsequent datasets from more TatC crystals from the optimisation conditions at the I04-1 beam line **(Table 3.9)** The data collection strategy was again altered in some circumstances to increase the number of images and thus the degrees of data collected. This was to ensure full anomalous completeness and to obtain higher redundancy in the datasets, to increase the accuracy of the intensity measurements.

Dataset	Φ start (^{o})	Δ Φ (^{o})	Exposure time per image / % Transmission	Total number of images collected	^{o} data collected
1	150	0.2	0.5 sec/ 100%	450	90
2	150	0.2	0.5sec/ 100%	900	180
3	111	0.2	0.3 sec/ 100%	850	170
4	0	0.2	0.2 sec/ 100%	1800	360
5	99	0.2	0.5sec/ 100%	1205	241
6	159	0.2	1.5sec/ 100%	300	60
7	219	0.2	1.5sec/ 100%	300	60

Table 3.9 Data collection strategies used for collecting the datasets from the SeMet crystals. The data collection strategies used were advised by EDNA (Incardona *et al.*, 2009).

The images were again processed by the automatic processing pipeline Xia2 (Winter, 2010) run in anomalous mode to keep the Friedel pairs unmerged. The processing statistics are given in **Table 3.10** for these five datasets. These five datasets, along with the first two datasets collected (**Table 3.8**) were input into autoSHARP (Vonrhein *et al.*, 2007) for scaling and merging, and outputted as a final, merged .mtz. Analysis of the *R*-factor on |F| for these datasets showed they all scaled together very well.

autoSHARP also runs its own analysis on each dataset. The programme will calculate, for the 25-5Å resolution range and the entire resolution range of the dataset, the SRF, and will plot Harker sections of the anomalous difference Patterson maps (example in **Figure 3.6.3**). The four selenium sites found previously in HySS were now subjected to further refinement in SHARP against this merged .mtz file. SHARP refines the parameters as before, but now included the scaling and non-isomorphism parameters for each dataset. Following the refinement in SHARP, statistics are produced for each dataset which

Data processing statistic	**Dataset 3**	**Dataset 4**	**Dataset 5**
Space group	H32	H32	H32
Cell dimensions			
a, *b*, *c* (Å)	124.2, 124.2, 215.6	124.3, 124.3, 215.3	123.5, 123.5 216.4
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Wavelength	0.92	0.92	0.92
Resolution	76.15-3.89 (3.99-3.89)	96.27-3.99 (4.09-3.99)	95.90 – 3.50(3.59-3.50)
R_{merge}	0.054 (0.523)	0.07 (0.769)	0.033 (0.949)
$I/\sigma I$	21.1 (3.5)	30.8 (4.9)	13.9 (1.4)
$CC_{1/2}$	0.98 (0.93)	0.99 (0.92)	0.93 (0.46)
Anomalous correlations	0.453 (0.533)	0.655 (0.027)	0.026 (0.048)
Anomalous completeness	95.0 (78.7)	99.8 (100)	87.7 (87.6)
Anomalous multiplicity	3.7 (3.6)	10.6 (11.2)	1.8 (1.8)
Completeness	98.9 (92.5)	99.7 (100)	96.8 (96.6)
Redundancy	6.9 (6.4)	19.9 (21.3)	3.3 (3.4)

Data processing statistic	**Dataset 6**	**Dataset 7**
Space group	H32	H32
Cell dimensions		
a, *b*, *c* (Å)	124.0, 124.0, 216.1	124.1, 124.1, 216.4
α, β, γ (°)	90, 90, 120	90, 90, 120
Wavelength	0.92	0.92
Resolution	52.09-3.50 (3.59-3.50)	76.26-3.73 (3.88-3.73)
R_{merge}	0.034 (1.166)	0.043 (0.701)
$I/\sigma I$	13.5 (1.7)	11.8 (2.0)
$CC_{1/2}$	0.94 (0.43)	0.93 (0.56)
Anomalous correlations	0.302 (0.045)	0.568 (0.028)
Anomalous completeness	86.4 (90.1)	87.8 (79.9)
Anomalous multiplicity	0.302 (0.045)	1.7 (1.6)
Completeness	97.2 (98.1)	96.7 (91.0)
Redundancy	3.3 (3.4)	3.3 (3.1)

Table 3.10 Data processing statistics for the five datasets collected from additional SeMet TatC crystals.

describe their overall phasing power (**Table 3.11**). This statistic is a measure of how effective a particular dataset is in determining the phase angles based on the ML statistics the programme implements. The phasing power drops off with resolution, and should be above 1 if the data are to contribute to the phasing. The resolution cut off at which this drops below 1 for each dataset is given in the table.

Dataset	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Overall Phasing power	0.883	0.43	0.632	0.694	0.680	0.610	0.797
The resolution at which the phasing power drops below 1 (Å)	5.4	6.22	5.4	5.4	5.4	5.8	5.4

Table 3.11 The phasing power for each dataset as calculated by SHARP.

The final co-ordinate positions of each of the selenium atoms after this round of refinement is given in **Table 3.12**, along with the initial starting positions (after the refinement against the first two datasets).

	Atomic coordinate					
	x		**y**		**z**	
	Initial (after 2 datasets)	Final	Initial (after 2 datasets)	Final	Initial (after 2 datasets)	Final
Se 1	0.2617	0.2622	0.1946	0.1957	-0.3811	-0.3879
Se 2	0.3870	0.3879	0.2781	0.2794	-0.3962	-0.3959
Se 3	0.1359	0.1354	0.1642	0.1160	-0.3709	-0.3709
Se 4	0.2972	0.2964	0.2188	0.2196	-0.4701	-0.4694

Table 3.12 Positions of the Selenium atoms after refinement against the first two datasets (Initial) and after refinement against all seven datasets (Final). Co-ordinates are quoted in the correct, inverted hand.

Having optimised the various substructure parameters to give the best possible initial phase estimates, the next step involved phase improvement by optimisation of the solvent flattening. Optimisation was achieved by varying the overall total solvent content of the asymmetric unit, and running the Solomon solvent flattening programme at these different overall unit cell solvent contents. This gave an optimal solvent content of 73%, with an I(contrast) that decreased from 0.54 to 0.28, and an I(SigmaA) that increased from 0.25 to 0.52.

As the solvent flattening statistics alone were not definitive in determining the hand from the substructure generated from the first two datasets, the solvent flattening was now re-run at this optimised solvent content, with the positions of the selenium sites inverted (back to the original hand). The run gave an I(contrast) that decreased from 0.61 to 0.38 and an I(SigmaA) that increased from 0.11 to 0.31, and these parameters are clearly worse than the inverted hand statistics above. The solvent flattening statistics now more confidently demonstrate the inverted hand was correct.

3.7 Model building and refinement

The model building programme COOT (Emsley & Cowtan, 2004) was used to view the optimally (73%) solvent-flattened map calculated from the structure factor amplitudes, FBshalsol, and the phases, PHIBshalsol, calculated from Solomon (**Figure 3.7.1**). The density, although only medium resolution, clearly showed regions of α helical secondary structure, consistent with the prediction that TatC has six transmembrane helices (Bogsch *et al.*, 1998). These helices were well connected. The loop regions between helices were very well defined, and although the exact positions of residues within the helices was not

obvious, the overall connectivity of the structure made crude tracing of the backbone chain relatively easy.

Building the secondary structure of the model was achieved in COOT (Emsley & Cowtan, 2004) by the use of the 'place helix here' automated helix builder tool, which will place the backbone of an α helix into a defined region of density. The connecting loops were then built manually. The positions of the selenium atoms allowed for the directionality of the chain to be determined, as the methionine residues in the sequence are non-uniformly distributed, and clustered towards the CT. The SeMet positions also gave initial anchor points to help assign the sequence to the regions of placed secondary structure.

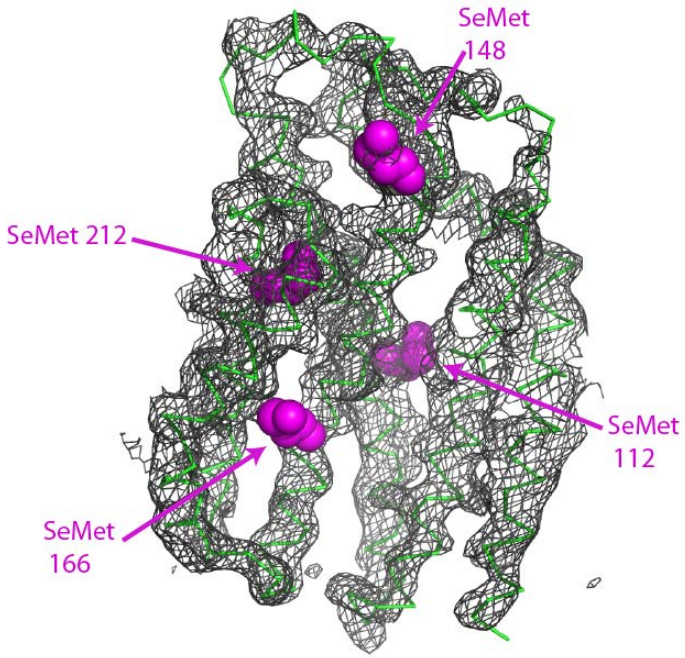


Figure 3.7.1 The 73% solvent flattened electron density map.

Calculated using the solvent flattened phases as given by Solomon, and the observed amplitudes, contoured at 1.5σ. The density is shown (grey). The positions of the selenomethionine side chains (pink spheres) and the final TatC model positioned in the density (green cartoon) are shown.

The model building process is aided by the iterative cycles between model building, which involves placing the chain into the real space electron density and reciprocal space refinement, in which the fit between the observed and calculated structure factor amplitudes are optimised. The reciprocal space refinement step minimises the many small and correlated local stereochemical and conformational errors left over from the real space model building.

The reciprocal space refinement programme used was autoBUSTER (Bricogne, 2011). This programme refines the atom positions, overall B-factors, bulk solvent corrections and anisotropy corrections against the experimental data whilst restraining the bond lengths, angles and atomic parameters expected from protein chemistry. These chemical parameters allow the inclusion of additional restraints during the refinement procedure, to improve the ratio of observed data to parameters to be refined. The residual between the observed experimental structure factors and the experimental structure factor amplitudes is the target function that the programme aims to minimise.

As the quality of the model begins to improve so too will the model-derived calculated phases (α_{calc}) compared to the initial experimentally derived solvent flattened phases. More detail will be observed in the electron density map resulting from use of the calculated phases in the structure factor equation. At the end of each stage of refinement a difference electron density map can be produced using the calculated phases, and amplitudes from the difference between the observed structure factor amplitudes $|F_{obs}|$, and the calculated structure factor amplitudes, $|F_{calc}|$ calculated from the model. A difference map in which $|F_{obs}|$ and $|F_{calc}|$ have equal weighting $|F_{obs}|$-$|F_{calc}|$, is known as the ‘residual’ map. This map emphasises errors in the current model and allows the identification of features which are either present in the current model and not in the data, or regions in which there should be something in the model when there currently is not. This map can help guide the model

building to move regions of structure/side chains to better fit the data. A map calculated from $2|F_{obs}|-|F_{calc}|$ allows for greater emphasis to be placed on the observed data to avoid bias from the model derived phases. This map is easier to interpret than the $|F_{obs}|-|F_{calc}|$ map, as it looks like protein density. Maps computed from both sets of coefficients were used during the model building process. The improvement in the quality of the map is shown in **Figure 3.7.1 A** and **B** in the region around the residue SeMet 148.

At each stage of model building the agreement of the model with the experimental data has to be verified. This is assessed by the *R*-factor statistic given below, in which the overall fit between the experimental data and the model is quantified by comparing the scaled structure factor amplitudes from the experimentally observed data ($|F|_{obs}$) and those structure factor amplitudes calculated from the current model ($|F|_{calc}$).

$$\boldsymbol{R} = \frac{\sum_{\boldsymbol{h}} ||\boldsymbol{F}|_{\boldsymbol{obs}} - |\boldsymbol{F}|_{\boldsymbol{cal}}|}{\sum_{\boldsymbol{h}} |\boldsymbol{F}_{\boldsymbol{obs}}|}$$

In order avoid over fitting the data and model bias a cross-validation statistic, R_{free} is used. This statistic is calculated as for the previous *R* factor except that it compares the calculated structure factors to those observed structure factors in 5% of the data which have not been used for structure solution. This randomly selected 5% of data is 'flagged' during the data processing steps by the programme TRUNCATE. An absolute difference of less than 5% between the R and the R_{free} for the final model is generally acceptable. A greater difference can indicate the data are being over-fit. If the model is improving to fit the data better, both R values are expected to fall during the model building and refinement process. The dataset used in refinement was dataset 5 in **Table 3.10**, which contained data to 3.5 Å resolution.

The initial model had high *R* factors (R/R_{free} 0.5094/0.4621) and many rounds of manual building using the difference maps described above were needed in order to determine the correct register of sequence within the density. The model was initially improved to give *R* factors of 0.37/0.34 (R/R_{free}). The biggest difficulty was getting the register of the sequence in the first two helices correct, as there are no methionines present in the structure until half way through helix three. The lack of anchor points in this region and the quite 'tubular' shape of the density in the helix was problematic. However, after the rounds of manual building, the rest of the model was now of a sufficient quality that the model could be fed into the automated model building programme, BUCCANEER (Cowtan, 2006) as a starting model with no external experimental phase information. This produced a model with trimmed side-chains that was a significantly better fit to the data, after adding back the side-chains another round of autoBUSTER refinement gave R/R_{free} of 0.31/0.29.

At this point, the phases had improved sufficiently that a non-protein feature of the map was able to emerge. At the top of helix one, at the interface between two molecules of TatC on the crystallographic axis, density emerges for a molecule of the LMNG detergent used in the purification. The initial density for the detergent is shown in **Figure 3.7.2 C,** contoured at 2.5σ prior to incorporation of the detergent molecule. The tetrahedral carbon atom in the centre makes the shape of the detergent extremely unusual which fitted the shape of the density very well this indicated that the density was detergent rather than another ligand such as PEG present in the crystallisation condition.

It is not unheard of to see detergents molecules or lipids that have been used during the membrane protein purification ordered within the crystal packing. A molecule of the detergent LDAO was seen between crystallographic dimers in a structure of rhodopsin (Li *et al.*, 2004), wrapped around a potentially flexible hinge region, stabilising its conformation aiding in the crystallisation of the protein. The naturally occurring lipid,

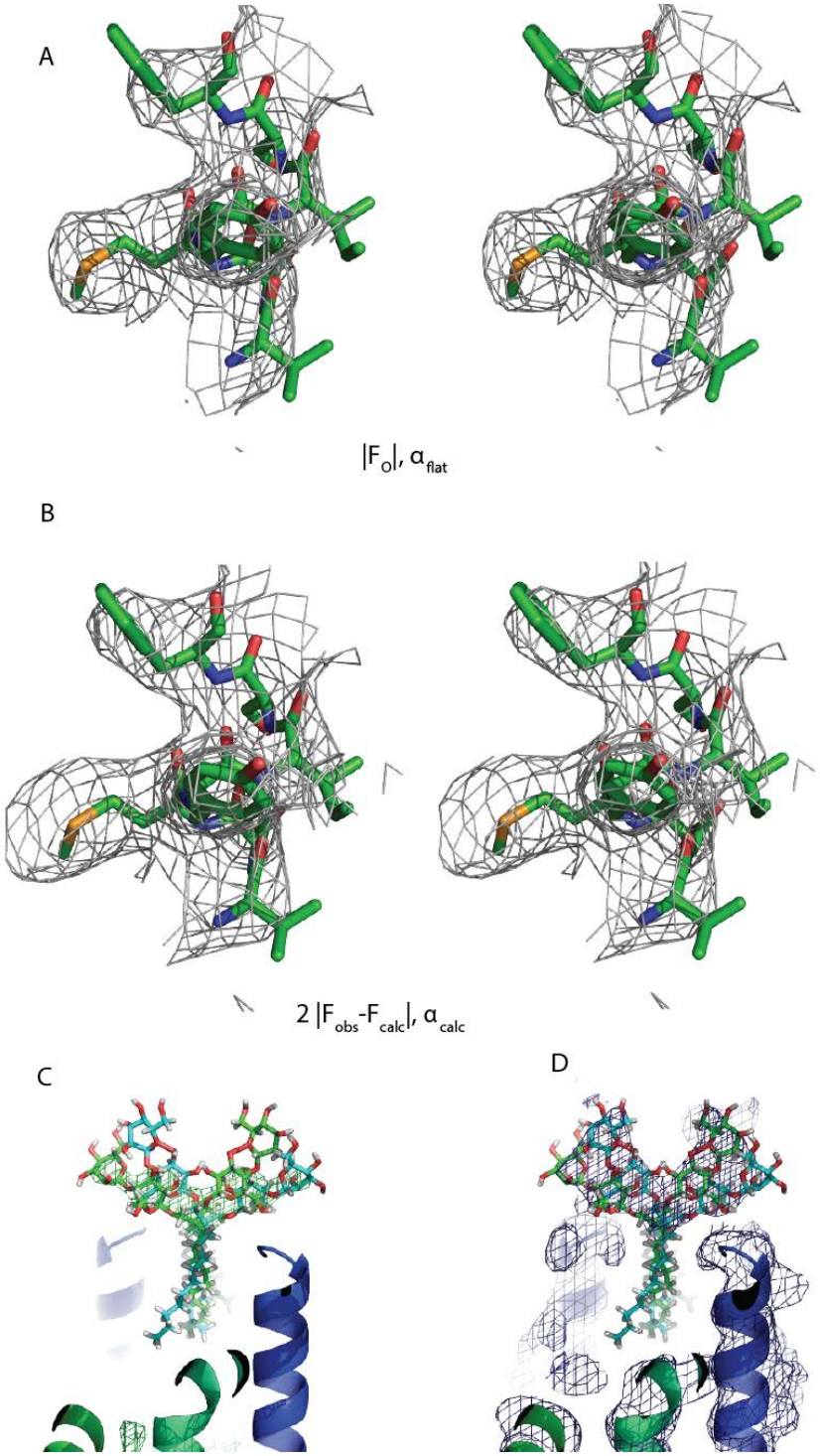


Figure 3.7.2 Electron density maps for the 3.5 Å resolution structure of TatC.

(A) and (B) electron density for the region around SeMet residue 148 shown as stereographic pairs. (**A)** for the initial solvent flattened map from Solomon (**B)** the final, refined map. (**C)** The initial density for the detergent molecule on the 2 fold crystallographic axis contoured at 2.5σ in the $|\mathbf{F_{obs}}-\mathbf{F_{calc}}|$, α_{calc} map **(D)** The density in the final 2 $|\mathbf{F_{obs}}-\mathbf{F_{calc}}|$, α_{calc} map contoured at 1.0σ.

cardiolipin, carried over from the membrane solubilisation stage has been observed in membrane protein crystal structures (McAuley *et al.*, 1999, Palsdottir *et al.*, 2003) giving insight into the specificities of lipid-membrane protein interactions.

As there have been no previously determined crystal structures that contain LMNG as a ligand a model of the detergent needed to be constructed. The programme Chemdraw (PerkinElmer) was used to draw the molecule and then a dictionary of stereochemical restraints for use in refinement was generated using Grade (Global Phasing Ltd). Following the introduction of the LMNG molecule, with an occupancy set to 0.5 in refinement due to its position on the 2-fold crystallographic axis, further rounds of refinement produced a final model of TatC with an R/R_{free} of 0.252/0.288. The final density for the LMNG molecule can be seen in **Figure 3.7.2 D,** where the density of the maltose head groups is much stronger than for the hydrophobic carbon tails, which are potentially more flexible. The final model comprised a continuous stretch of residues from Glu 5 to Leu 229, as some residues at the extreme NT and CT of TatC were disordered in the crystal, and no density for these was visible during the process of refinement.

Validation of the model involves the assessment of the local stereochemistry of the structure, which is not given by the global R factor statistic. MolProbity (Chen *et al.*, 2010) is a web-based server that can asses the Ramachandran plot for each amino acid in the structure, as well as look at the quality of the rotomers for each amino acid side chain. It can also add hydrogen atoms to the model, and use this to check for atomic clashes in the structure. The final refinement statistics are given in **Table 3.13**. The model has one Ramachandran outlier, and the rest are in favoured or allowed regions. The structure has a 'Molprobity score' of 2.8, this score takes into account validation statistics for all the structures in the PDB of comparable resolution. It places the structure in the 93rd percentile for structures in the 3.5 $\pm$ 0.25 Å resolution range, where the 100th percentile is the best.

Refinement statistics	Refinement against dataset 5
Resolution (Å)	22-3.5 (3.9-3.5)
Number of reflections	7900
R_{work}/R_{free}	25.2% / 28.8% (24.0% / 24.8%)
Number of atoms	
Protein	1804
Ligand (LMNG)	151
Water	0
B factors (Å^2)*	
Protein	159
Ligand (LMNG)	162
R.m.s deviations	
Bond lengths (Å)	0.006
Bond angles (^{o})	1.2

Table 3.13 Refinement statistics for the final model of TatC.
***** The Wilson B-factors are 133Å^2

3.7 Summary

TatC in LMNG proved to be a highly homogenous sample which enabled high quality, reproducible crystals to be grown. The optimisation of growth of the L56 (DE3) expression cells in selenomethionine minimal media enabled a high yield of fully SeMet substituted TatC protein to be purified. Crystallisation of this sample in the condition identified for the native protein gave better crystals, with the best crystals diffracting to 3.5Å resolution. These data enabled the structure of TatC to be solved by the single wavelength anomalous dispersion method. Difficult model building followed, but eventually a final structure of TatC was built which showed good agreement to the experimental data (R/R_{free} 25.2% / 28.8%) and displayed good stereochemistry, as shown by the MolProbity score. The completed structure could now be analysed.

Chapter 4: The structure of TatC and insights into signal peptide binding

Following screening of a number of constructs which identified the TatC sequence from *A. aeolicus* as amenable to over expression, intensive detergent homogeneity screening in this study (**Chapter 2**) allowed high quality crystals of *A. aeolicus* TatC to be grown and the structure solved by Selenium-SAD using selenomethionine derivatised protein crystals (**Chapter 3**).

No detailed structural information was previously available for any member of the TatC family of proteins, from either the bacterial or chloroplast branches of the Tat system. Various biochemical studies had confirmed that the topology of the protein, with both the N and C-termini located within the bacterial cytoplasm (or in the stroma of the chloroplast) (Jeong *et al.*, 2004). Sequence analysis predicted TatC would have six transmembrane helices (Bogsch *et al.*, 1998) which was demonstrated experimentally by cysteine accessibility experiments (Punginelli *et al.*, 2007). However, attempts to further investigate how these helices are orientated with respect to one another, by disulphide mapping, were inconclusive. The helical nature of the *B. subtilis* $TatC_d$ protein, homolgous to *E. coli* TatC, has been demonstrated by circular dichronism (CD) experiments (Nolandt *et al.*, 2009) which also indicated that, on avaerage, the helices showed significant tilt in the membrane, but no further details on the structure are were obtained by this method.

The work presented in this study allowed the structure of TatC to be visualised for the first time at high resolution, and many interesting features have become apparent.

4.1 The overall structure of TatC

TatC crystallised with one copy of the protein in the asymmetric unit, along with one ordered LMNG detergent molecule located on a crystallographic 2-fold (**Figure 4.1.1A)**. The position of the lipid bilayer is indicated based on molecular dynamic simulations of TatC within a membrane environment (Dr Phillip Stansfeld, Dept of Biochemistry, University of Oxford)

The structure (**Figure 4.1.1A)** confirms the presence of the predicted six transmembrane helices. Using the structure to perform a PDBeFold (EMBL-EBI) search of the PDB, which searches for similar structures deposited in the PDB, brings back no significant hits Therefore the overall arrangement and organisation of these helices shows a novel architectural arrangement not seen to date in any other solved membrane protein structures. Interestingly, the structure is unrelated to that of the ubiquitous SecYEG translocon (Van den Berg *et al.*, 2004) which was discussed in the Introduction, even though this also binds signal peptides of a similar overall architecture (Berks, 1996), present on passenger proteins to be transported by the Sec system. The distinct mechanisms of protein transport brought about by these two systems leads to very different structures.

Figure 4.1.1 B(i) and **(ii)** show how the overall arrangement of helices in the bilayer forms a ‘curved’ shape when viewed from either the cytoplasmic or periplasmic face. The surface view in **Figure 4.1.1 C**, which is shaded to show the depth of the structure, highlights how this curvature of the helices creates a large ‘cleft’ in the centre of the protein which extends upwards and becomes narrower as it approaches the periplasm. This shape seems to resemble a ‘cupped hand’ shape, with the transmembrane helices forming a curved wall and the periplasmic cap curved over the top. The gap between helices 5 and 6 gives a prominent ‘groove’ from the periplasm to the cytoplasm.

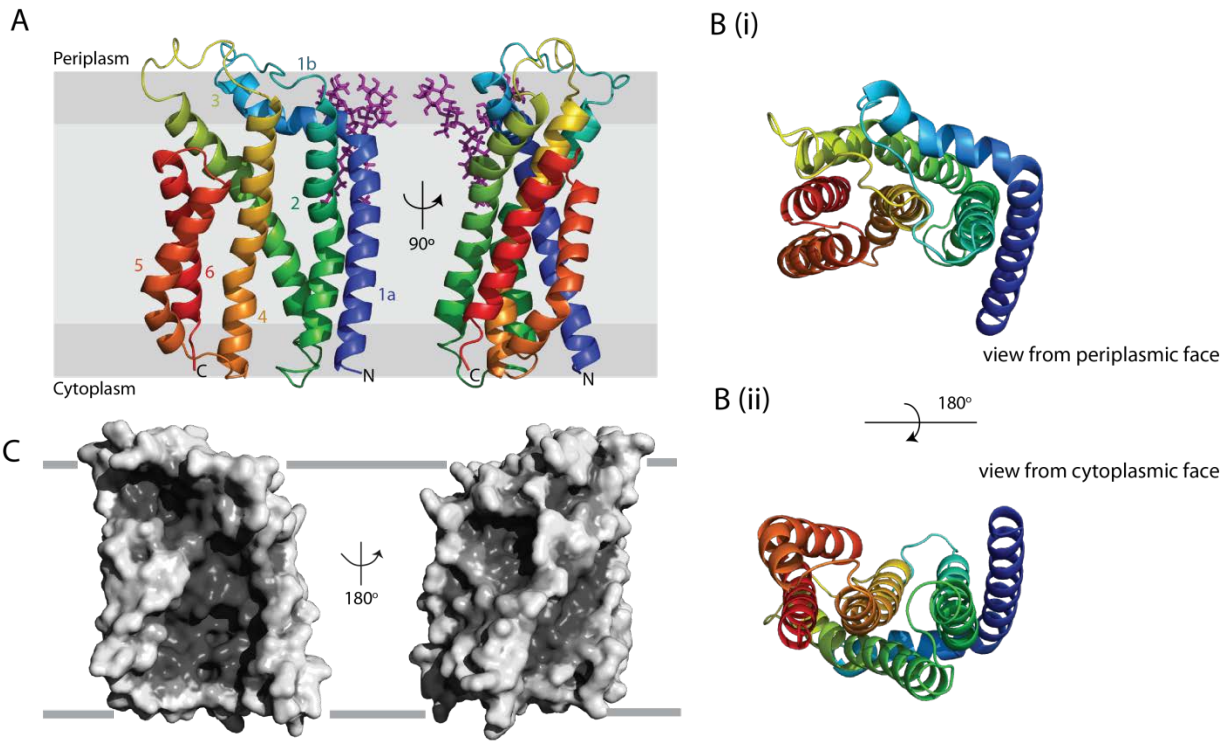


Figure 4.1.1 The structure of *A. aeolicus* TatC.

(A) Cartoon representation of TatC, coloured from blue at the NT to red at the CT. The position of the bilayer inferred from MD simulations. The semi-ordered molecule of LMNG is shown in purple sticks. Cartoon representation in (**B)** as in (**A**), but shown from the **(i)** the periplasmic face and **(ii)** the cytoplasmic face. (**C)** Surface representation of TatC, shaded to show depth. The left hand view is the same as the left hand view in (**A**).

This surface view also shows how elongated the structure is, rather than being a compact collection of helices arranged in a cluster, the trans-membrane helices are spread out in a more linear fashion. This gives the protein a surface area of around 16,500 $Å^2$ which is relatively large given the size of the protein.

The helices within TatC contain quite severe inter-helical kinks and are also very tilted with respect to one another and to the membrane normal, as can be seen very clearly in the

side view in **Figure 4.1.1 A.** Helix 1, for example, is tilted almost 40^{o} with respect to the membrane normal, and is tilted almost 60^{o} away from the cytoplasmic end of helix 3.

The final model is missing 11 residues at the CT and four residues at the NT. Helix 1 is relatively straight as it spans the membrane, but as it reaches the periplasmic face of the bilayer, it bends around residue 30 causing the rest of the helix to form an interfacial helix. This interfacial helix has many lysine and aspartate residues orientated towards the periplasm. The interfacial helix then loops back to form helix 2. Helix 2 is kinked at a point in the centre of the membrane around Pro 78. On the last turn of helix 2 is residue Phe 87, pointing towards helix 1, followed by the first cytoplasmic loop of around 12 residues. Helix 3 curves as it cuts through the membrane, and kinks in the centre of the membrane which causes it to span across the back 'face' of the protein, past helices 4, 5 and 6. At the end of helix 3, the protein forms a second periplasmic loop, which packs closely and tightly with the first periplasmic interfacial helix and loop, forming extensive interactions.

Helix 4 follows on from this loop, and is relatively straight as it spans the membrane, except for a small kink towards the cytoplasmic face. This leads into the second, shorter cytoplasmic loop, before helix 5 starts. Helix 5 and helix 6 are both very short helices compared to the rest of the TatC helices, being composed of around 22 residues as opposed to the 28-30 residues of helices 1-4. Helix 5 has a very severe kink around residue Phe187 near the cytoplasmic face. The final periplasmic loop between helices 5 and 6 is only four residues and helix 6 comes straight down through the membrane.

Interestingly residue Glu 165, on helix 4, is in the centre of the 'cupped' hand of TatC. This places an exposed negatively charged residue in the middle of the bilyaer. This lone, charged residue is surrounded by non-polar, hydrophobic residues; as expected for the

membrane embedded part of the protein. The unusual location of this resuidue is highlighted by the electrostatic surface representation of TatC, shown in **Figure 4.3.2** in which the surface is coloured according to the surface charge. The red (negative) patch of Glu 165 clearly stands out amongst the white (non-charged) residues. This residue is highly consered as a glutamate or a glutamine, suggesting a conserved functional role within the Tat transport system.

Following publication of the TatC structure presented in this study (Rollauer *et al.*, 2012) (**Appendix 8.6**) two additional, lower resolution structures of the TatC protein from *A. aeolicus* have very recently become available (Ramasamy *et al.*, 2013). These crystals were grown in different conditions to that crystallised here. Firstly a construct of *A. aeolicus* TatC fused at its CT to T4-lysozyme was solved in DDM at 6.8 Å resolution (PDB code 4HTT). Secondly, an *A. aeolicus* TatC construct engineered to have two surface entropy reducing mutations was purified and crystallised in DHPC (di-heptanoyl phosphatidylcholine) and gave only a single dataset to 4.0 Å resolution (PDB code 4HTS). These structures were solved by molecular replacement using the PDB co-ordinates of the TatC structure presented in this study (PDB code 4B4A).

These new structures are very similar to the structure presented here, an overlay of which is shown in **Figure 4.1.2**. These have an RMSD when compared to the LMNG structure of 0.85 Å over 224 residues (for the 4HTS structure) and 1.0Å over 222 residues (for the 4HTT structure), RMSD values which lie within the experimental error. The structures are extremely well aligned; the only slight difference, noted by Ramasamy *et al*, is in the loop region between helix 5 and helix 6, where helix 5 of structure 4HTS structure (TatC in DHP) has an extra turn at its CT, shown by the red arrow in **Figure 4.1.2**.

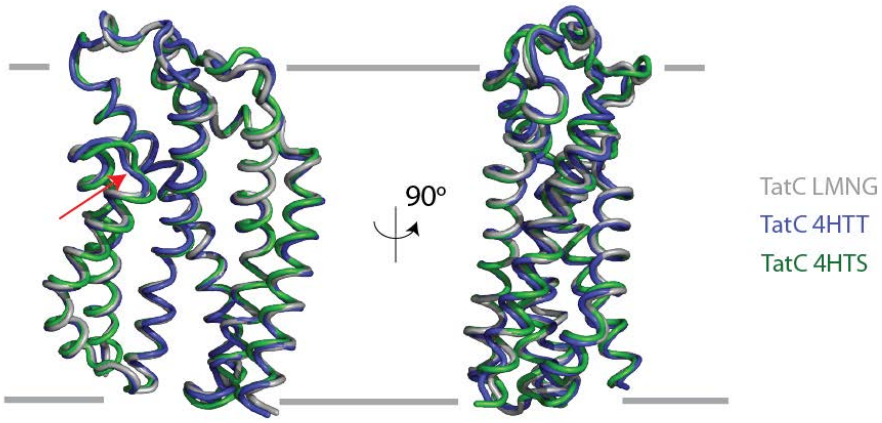


Figure 4.1.2 An overlay of *A. aeolicus* TatC structures.

The structure presented in this study is shown in grey (TatC LMNG) with the two structures solved subsequently by Ramasamy *et al* are shown in green (TatC in DHPC (4HTS)) and blue (TatC-T4 lysozyme in DDM (4HTT)). The red arrow indicates the region of slight structural deviation.

Molecular dynamics studies of TatC

In order to explore how TatC may behave in a membrane environment, molecular dynamic (MD) simulations were carried out by Dr Phillip Stansfeld (Department of Biochemistry, University of Oxford). TatC was placed in a bilayer composed of dipalmitoylphosphatidylcholine (DPPC) molecules, and coarse grain molecular dynamics (CGMD) simulations run for 1μs (Scott *et al.*, 2008). **Figure 4.1.3 A** shows that during the course of the simulation TatC becomes almost completely buried within the bilayer. This is not surprising given that the height of TatC (~53 Å at the longest point) is on a similar scale to the depth of a bilayer (~50Å). Very little of the cytoplasmic loops are exposed. More of the structure is exposed on the periplasmic side of the membrane, with the loops able to protrude outwards to some extent. The interfacial helix at the top of helix 1 is mostly within the bilayer.

These simulations also suggested how TatC may affect the membrane. **Figure 4.1.3 B** shows the membrane thickness, normalised over the MD trajectory. This simulation shows

that TatC introduces a significant ‘thinning’ of the membrane around the short helices 5 and 6. The membrane is drawn towards the kink in helix 5. There are three charged or polar residues in this region, Arg188, Lys189 and Tyr190, which ‘pull’ the polar head groups of the membrane upwards. Similarly, at the periplasmic end of helix 5 and 6 there is deformation of the membrane downwards towards the top of the helix. Some deformation is also seen at the NT of helix 1, and again there is a run of polar and charged residues up to position 13 on this helix which could be responsible for this.

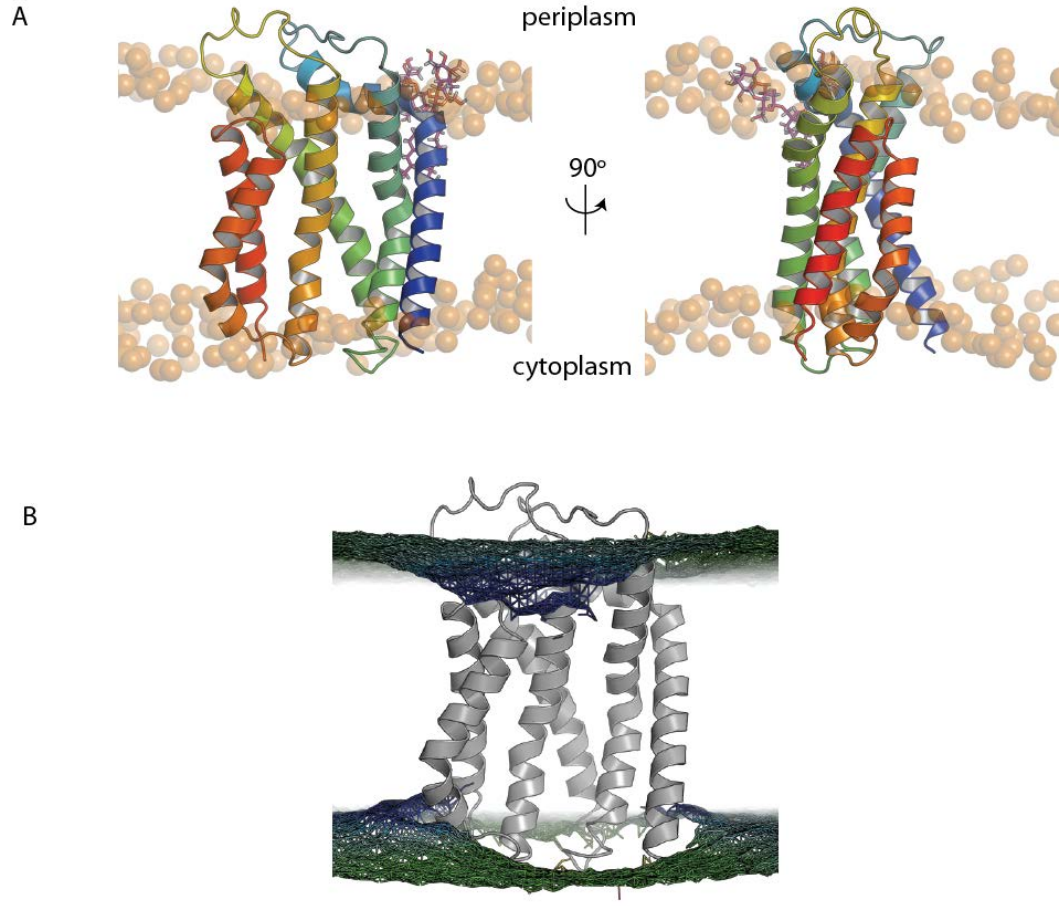


Figure 4.1.3 Membrane bilayer deformation by TatC in course grain molecular dynamic simulations.

(A) front and side view of TatC in a DPPC bilayer after 1μs of CGMD simulations. The lipid head groups are shown as orange transparent spheres. TatC is almost completely buried in, and induces deformation of, this membrane. (**B)** Membrane deformation is illustrated by the thickness of the bilayer, where the positions of the lipid head groups are shown as a mesh, coloured from blue (thin) to red (thick) and TatC as a grey cartoon. All simulations were run, and this figure produced by Dr Phillip Stansfeld (University of Oxford).

Membrane deformation by TatC is intriguing and may be functional. TatC could initiate membrane weakening around helices 5 and 6 to aid membrane perturbation by the TatA component at a later stage of the transport process and bring about precursor protein transport by either of the debated mechanisms of action, by further deforming the membrane (Brüser & Sanders, 2003), or by building a passage way across the membrane for the precursor to travel (Gohlke *et al.*, 2005).

Atomistic molecular dynamic simulations were also run over a 100ns time period. These enabled the internal motions of TatC to be assessed, to see how stable the structure is in a bilayer.

These simulations showed that TatC within a bilayer is remarkably stable. The Cα root mean square flucation (RMSF) reaches a plateau of 3 Å RMSF within 50ns. This is really a very small change for a membrane protein and is much lower than is seen in other membrane protein structures. Most have an RMSF of around 4Å, some never plateau and

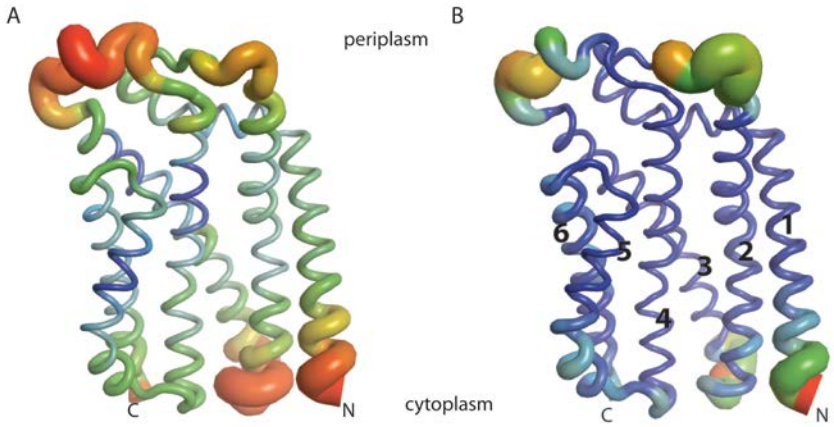


Figure 4.1.4 Thermal motions of TatC and stability in a bilayer.

Putty representations of TatC, where the thickness and colour of the tubes represents (**A)** the crystallographic B-factors (**B)** the RMS fluctuations of C_α during 100ns of atomistic MD studies, from thin/blue (small) to thick/red (large).

can still be increasing even after 100ns of simulation (Dr P. Stansfeld, personal communication). The small motions that do occur can be represented by plotting the relative Cα RMS flucations over the time of the 100ns simulations onto the structure (**Figure 4.1.4 B**).

The simulations shows that motion of TatC is restricted to distinct sections of the periplasmic loop regions, to the very NT of helix 1, and to the cytoplasmic loop region at the end of helix 3. There is almost no motion of the transmembrane region of TatC, the transmembrane helices are very stable and stay anchored within the membrane. The only motion of a helix within the membrane, which is not observed in all simulations, is that of the cytoplasmic end of helix 5, which 'relaxes' to form a shallower kink. This movement was noted by Ramamsay *et al*. which led them to claim this helix might have some inherent flexiblity. More experiments will need to be carried out to assess if motion can occur here and if this subtle straightening could be of functional importance. The subtle motions are also echoed in the atomic B-factors found in the crystal structure, these can also be plotted onto the structure in a similar manner as for the RMS flucations (**Figure 4.1.4 B)**. The crystallisation B-factors of the structure correspond closely to the regions that fluctate most during the simulation.

Given the extremely similar stuctures observed for TatC crystallised in three different detergents (**Figure 4.1.2**) indicates that, despite the difference in the detergents used in purification, and the very different crystallisation conditions (pH, salt concentrations, precipitants), the structure seen here maybe a favoured conformation for TatC to adopt. Therefore, if TatC does undergo any drastic, large scale conformational re-modelling as part of the Tat transport mechanism it can not be predicted based on the current structures alone. In order to assess if TatC does undergo more large scale conformational changes, crystal structures of TatC in different conformations would be needed. The alternating

conformations seen in crystal structures of small molecule transporters, for example the LacY family of metabolite transporters or ABC cassette transporters (Forrest *et al.*, 2011), have been successful in providing evidence for how these families of transporters function.

A further point of interest in the MD simulations concerns the negatively charged residue Glu 165 (*E. coli* Glu 170) which is located in centre of the membrane. In the MD studies this residue is seen to be hydrated and can attract a lipid headgroup, further distorting the membrane. This suggests that the position of this residue within the membrane is highly unfavorable, and may either play a role in membrane distortion or may be involved in a crucial interaction with a partner protein within the cell. This will be discussed in **Chapter 5**.

Given that the Tat system is powered by a proton electrochemical gradient (or proton motif force, PMF) (Mould & Robinson, 1991) it was speculated that this charged residue could be involved in the mechanim by which the PMF is utilised. Although the MD simulations showed Glu 165 to be involved in a short hydrogen-bond (H-bond) network with Ser 107 and Trp 85, this network is isolated from the membrane surface and thus is unlikely to participate in the movement of protons across the membrane. This was the only observation of an internal H-bond network within the TatC structure, and no other surface-linked channels that could potentially accomodate a flow of ions are present. TatC also bears no resemblance to an ion antiporter. Thus, how the PMF is utilised by the Tat system is not obvious from the TatC structure alone.

4.2 Functionally important regions of the TatC structure

The distinct and intriguing structural features of TatC described above, indicate that TatC is highly structurally adapted for its role in the unique Tat transport system. As seen for the

crystal structure of the SecY channel (Van den Berg *et al.*, 2004), the TatC structure could therefore potentially give some insight into understanding the mechanism of action of TatC within the transport system. To help elucidate the importance of the observed structral features, data from years of previous work in the Tat system can be drawn upon.

The TatC protein has been widely studied in *E. coli*, and several studies aimed at investigating the functional importance of residues within the protein have been carried out (reviewed in (Fröbel *et al.*, 2012b). Due to the high level of sequence conservation between

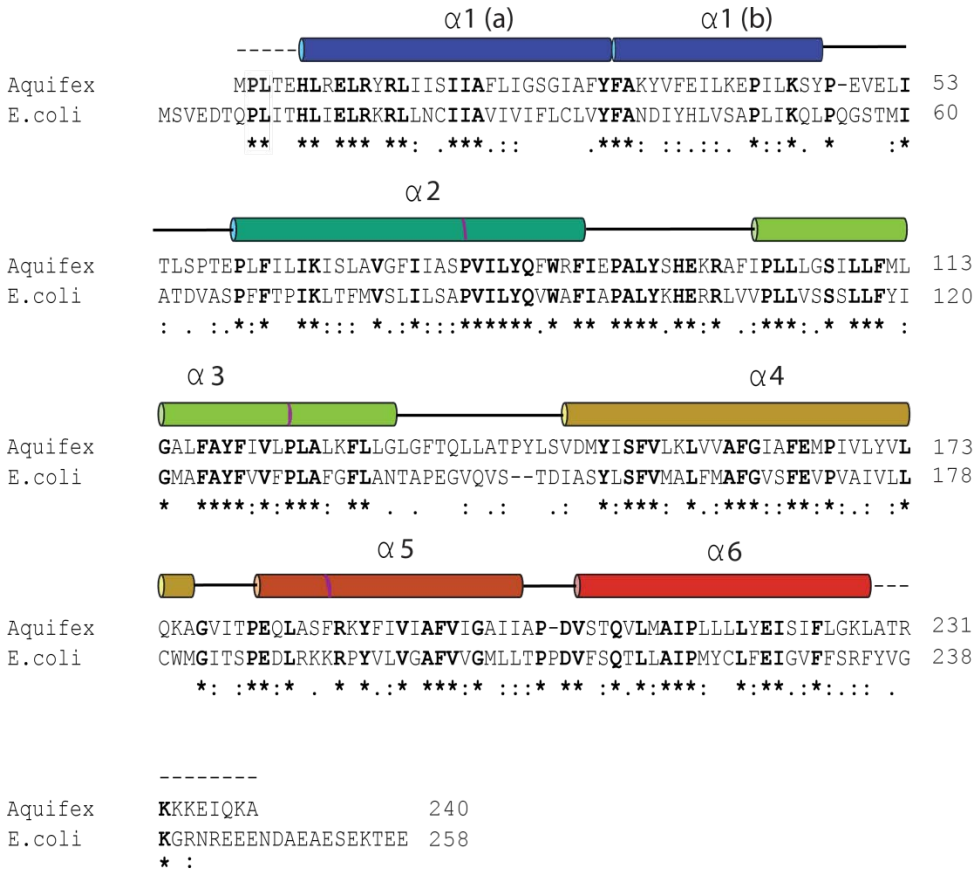


Figure 4.2.1 Sequence alignment of the *A. aeolicus* and *E. coli* TatC sequences.

The positions of the helices are coloured in **Figure 4.1.1A** and the kinks within each helix labeled with a purple line. Disordered residues in the crystal are shown with a dash. Alignment was performed by ClustalW2 (Larkin et al., 2007) and identical residues are shown with a star, highly conserved and conserved residues shown with colons and dots respectively.

the *E. coli* and the *A. aeolicus* proteins (40% identical and 80% similar as from the scoring algorithm from ClustalW2 (Larkin *et al.*, 2007)) (sequence alignement in **Figure 4.2.1)** the functional data present in the literature can be intepreted in light of the *A. aeolicus* TatC structure.

Many studies in the literature have mutated residues in the *E. coli* TatC protein and investigated the effect on Tat system function, however these have reported conflicting phenotypes for the same TatC substitutions. This is most likely due to the difference in sensitiviy of the transport assays used, and the different levels of expression of the mutant proteins, whereby high over expression of the mutant Tat component can compensate for a decrease in function. Therefore, the most systematic, unbiased studies with unambiguous functional results were used to give a shortlist of putative *E. coli* residues which seem to be important for *E. coli* TatC function.

Firstly, residues identified in a random mutagensis screen which blocked Tat transport (Kneuper *et al.*, 2012) were listed (shown as green in **Table 4.1**). The second category of residues, shown in orange in **Table 4.1**, are from studies in which conserved TatC residues were selectively mutated to alanine and Tat transport was inhibited (Buchanan *et al.*, 2002) (Holzapfel *et al.*, 2007). Residues which have been implicated in TatC function by both types of study are shown in yellow. These residues, coloured according to **Table 4.1**, were plotted as CPK-spheres onto the TatC structure (**Figure 4.2.2 A**). This analysis showed that the functionally important residues cluster into three main regions of the TatC structure, the ‘cytoplasmic face’ at the NT end of helix 1 and in the first cytoplasmic loop between helix 2 and 3 (marked as region (i)); the membrane-exposed termini of helices 5 and 6 (marked as region (ii)); and the periplasmic loop region at the top of the structure.

***E. coli* residue**	***A. aeolicus* residue**	**Genetically selected inactivating substitutions in *E. coli* TatC**	**Transport affected by Ala substitution in *E. coli* TatC**
His12	His6		His12
Arg17	Arg11		Arg17
Ser46	Lys40	S46F	
Pro48	Pro42	P42L, P48S	Pro48
Met59	Leu52	M59K	
Ile60	Ile53	I60N	
Asp63	Ser56	D63V	
Val64	Pro57	V64E	
Ser66	Glu59	S66P	
Phe68	Leu61	F68S	
Thr70	Ile63	T70R	
Leu74	Ile67	L74P	
Phe94	Phe87	F94S	Phe94
Pro97	Pro90		Pro97
Leu99	Leu92	L99P, L99Q	Leu99
Tyr100	Tyr93		Tyr100
Glu103	Glu96	E103G, E103K	Glu103
Tyr126	Tyr119		Tyr126
Pro131	Pro124	Pro131L	
Leu137	Leu130	L137H	
Val145	Leu138	V145E	
Ser148	Thr141	S148P	
Asp150	Ser145	D150G, D150V	
Ile151	Val146	I151T	
Glu170	Glu163		Glu170
Met205	Ala200	M205K, M205R	
Asp211	Asp203		Asp211
Gln215	Gln209	Q215R	
Leu225	Leu219	L225P	

Table 4.1 Mutations that inactive *E. coli* TatC. The corresponding residue in *A. aeolicus* is given based on the sequence alignment in **Figure 4.2.1**. The colouring given is that used in **Figure 4.4.2 (A)**. Genetically selected substitutions from (Kneuper *et al.*, 2012), alanine substitutions from (Holzapfel *et al.*, 2007) and (Buchanan *et al.*, 2002), which include only class I and class II mutations, as defined by Buchanan *et al*. Only mutations from stably membrane integrated TatC protein have been included.

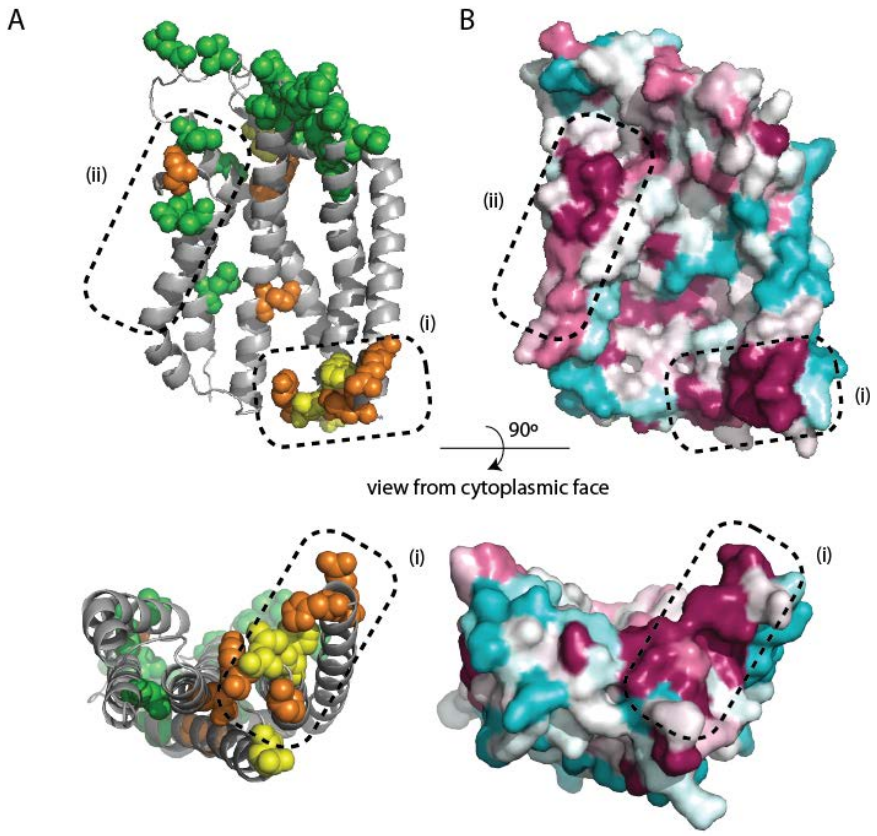


Figure 4.2.2 Functional and conserved residues of TatC shown on the structure.

(A) TatC shown in grey, orientated as in **Figure 4.1.1A**, with the inactivating residues shown in **Table 4.1** plotted onto the structure with their side chains shown as spheres. **(B)** sequence conservation at the surface of TatC displayed in the Consurf colour scheme (Ashkenazy *et al.*, 2010), from maroon (highly conserved) to blue (highly variable). The dotted regions in (**A)** and (**B)** labeled (i) and (ii) highlight the functional and conserved regions of the structure.

The residues in the periplasmic loop regions are very tightly packed, and form a periplasmic cap over the rest of the structure. This arrangement may be directly involved in TatC function, or could have a very significant role in maintaining the overall architecture of the helices in the membrane, by stabilising their orientation. A structural role for this region may explain why so many of the residues in the periplasmic cap are sensitive to mutation, as any change in the structure of the cap region could have detrimental structural effects on the rest of TatC, resulting in the proteins inability to function as its fold is no

longer correct. The positions of regions (i) and (ii), however, have no such obvious structural explanation. These regions could be directly involved in TatC function.

To further understand the contributions of these regions to TatC function, a sequence conservation analysis was undertaken, using the programme 'Consurf' (Ashkenazy *et al.*, 2010). The structure was submitted to the online server, which then creates a sequence alignment of many thousands of TatC sequences. The server then plots onto the structure the sequence alignment in the form of a colour code, in which regions of the sequence that display high conservation across all the sequences are plotted maroon, and those that are not well conserved, and highly variable, are plotted blue. This analysis is shown on the surface representation of TatC in **Figure 4.2.2. B**.

The Consurf analysis identifies two regions of highly conserved sequence, which correspond very well to the two proposed functionally important regions identified in **Figure 4.2.2.A**. This gives an indication that these regions may have an important role in how TatC functions, and thus residues and structural elements in these regions have been conserved. Only a few residues in the periplasmic cap region are conserved, and many are not conserved, suggesting this region may not participate directly in the function of TatC.

TatC is known to bind to signal peptides of Tat substrate proteins (Alami *et al.*, 2003). TatC is also known to form a stable, 1:1 complex with TatB in the membrane and can also bind to the TatA component (Bolhuis *et al.*, 2001). It was therefore speculated that these highly conserved, functionally important regions may be involved in one or more of these functions. The interaction with the signal peptide will be discussed in this Chapter, while the structural insight into TatC interactions with the other Tat integral membrane components, TatA and TatB, will be discussed in **Chapter 5.**

4.3 Signal peptide binding by TatC

Substrate protein binding by the TatBC complex has been demonstrated by various techniques; the use of antibodies against the thylakoid TatB and TatC proteins that blocked substrate binding to thylakoids (Cline & Mori, 2001), the co-purification of substrate protein with TatB and TatC when detergent extracted from membranes (Tarry *et al.*, 2009b), the identification of suppressor mutations within TatBC that allow for the transport of previously defective signal peptides (Kreutzenbeck *et al.*, 2007) and crosslinking between the signal peptide and the TatBC complex by chemical and site directed methods (Alami *et al.*, 2003, Gérard & Cline, 2006). A subset of these studies also demonstrated binding by TatBC occurs independently of the PMF.

Crosslinking studies have demonstrated that the TatC component forms the primary recognition site for the recognition of the RR motif. The crosslinks between the substrate and TatC are still present even without the presence of TatB component (Alami *et al.*, 2003) indicating TatC is able to bind the RR region of the signal peptide alone, with no requirement for TatB. Furthermore, despite the apparent multivalency of both the *E. coli* and thylakoid TatBC complexes, this may not be a strict requirement for function of the Tat pathway. Various lines of evidence suggest it is likely the binding site for a signal peptide is present within a single molecule of TatC. Engineered dimer fusions of *E. coli* TatC in which one molecule included a known inactivating mutation (F94A, implicated in signal peptide binding by TatC) showed that the system was still active for transport (Maldonado *et al.*, 2011a) indicating that only one of the TatC units within the dimer is needed to bind a precursor. Consistent with this, it was also observed that multiple precursors can bind to the *E. coli* receptor complex, by single particle EM of co-purified *E. coli* $SufI_{His}$ and TatBC (Tarry *et al.*, 2009b) indicating there is more than one receptor binding site within the TatBC complex. Additionally, experiments with plant thylakoid

membranes using immunoblotting and scintillation counting, coupled with titration of labeled precursors to calculate saturation binding isotherms, concluded that each cpTatC protein present in the thylakoid is competent to bind a single precursor protein (Celedon & Cline, 2012).

In this study, Isothermal Titration Calorimetry (ITC) was carried out to determine the binding affinity for the signal peptide interaction with *A. aeolicus* TatC *in vitro*, and obtain direct evidence for the stoichiometry of binding. Synthetic peptides have been shown *in vitro* to competitively inhibit Tat transport (Alami *et al.*, 2002) and thus alone are functional for binding. Thus, in order to test if purified *A. aeolicus* TatC in LMNG is able to bind signal peptides, binding was directly assayed by using a synthetic peptide consisting of the 26 amino acids of the signal peptide sequence from the *E. coli* protein SufI (an archetypal Tat transported passenger protein). This peptide encompasses all three regions of tat signal peptides, the positively charged n-region the hydrophobic h-region and the polar c-region (**Figure 4.3.1 (A)** and **(B)**).

ITC is a heat based binding assay, in which the temperature change during the course of a reaction is used to determine the K_D for that interaction, as well as the enthalpy change and the binding stoichiometry according to the following equation:

$$\Delta G = -RT\ln K_a \ = \Delta H - T\Delta S$$

The ITC calorimeter consists of two cells, a reference and a reaction cell. Temperature differences between these two cells, filled with protein and water respectively, is used to monitor the heat change in the reaction chamber as the ligand is titrated into the protein cell. This is read out as the power needed to be applied (or removed) to the experimental cell to keep the temperature the same as that in the reference cell.

The ITC experiments were performed using a iTC_{200} isothermal titration calorimeter (MicroCal), and the synthetic peptides were produced by Generon. TatC was loaded into the cell at roughly 100µM, the exact concentration determined by UV spectroscopy at 280nm in 6M guanadinium hydrochloride, using an extinction coefficient of 24,870 $M^{-1}cm^{-1}$. This is the 'ideal' extinction coefficient as determined by the server 'Protparam' (Expasy), and as such may not be completely accurate. In order to determine this more accurately, amino acid analysis would need to be run on purified protein samples at various UV absorbances to determine an exact extinction coefficient. The peptide was dissolved in the TatC SEC purification buffer at a concentration of 1mM and used as the titrant. The titrations were carried out at 25^{o}C, using a reference power of 3cal s^{-1} and a stirring speed of 1000 rpm. Isotherms were concatenated together, where necessary, using the CONCAT software, and then analysed by SEDPHAT (Houtman *et al.*, 2007), using a hetero-association model (A + B ↔ AB) using ΔH and K_D as fitting parameters. Due to the inaccuracies in weighing the peptide, the concentration of this was also allowed to float.

The analysis of the isotherm of TatC binding to the full length peptide (**Figure 4.3.1 A)**, allows both K_D and stoichiometry to be determined by performing non-linear regression using fitting functions within the CONCAT software. The analysis gives a 1:1 binding stoichiometry between TatC and the signal peptide. This calculated ratio coincides with the data suggesting that the binding site for the signal peptide resides within one TatC molecule. The analysis gives a moderate K_D of 3µM for the interaction.

Previously, attempts to measure the interaction *in vitro* with purified, tagged *E. coli* TatABC complexes by monitoring effects from the intrinsic tryptophan fluorescence levels demonstrated binding, but gave no measure of the strength or stoichiometry of this (de Leeuw *et al.*, 2002). A recent florescence resonance energy transfer (FRET) experiment using *E. coli* derived inner membrane vesicles, and fluorescence donors/acceptors on

substrates and the TatC/TatB components, was able to measure the K_D as ~20nM (Whitaker *et al.*, 2012). The difference to the K_d calculated here could be the use of the *E. coli* substrate in the ITC assays, which is not necessarily optimised for *A. aeolicus* TatC binding. Indirect measurements of the interaction between transported proteins and the Tat translocase in plant thylakoids, by the use of SDS-PAGE in combination with immunoblotting or fluorography measured a much tighter K_D of 1nM (Celedon & Cline, 2012). However, this experiment was done using an engineered substrate that has non-physiological, tighter binding affinity, which could be part of the reason for the difference to the K_D calculated here by ITC. Additionally, as both the FRET and fluorography experiments were carried out with either inner membrane vesicles or with isolated thylakoids, where presumably TatC in both experiments is part of the TatBC receptor complex embedded in a membrane environment, this may alter the affinity compared to TatC in a detergent solution.

The result of repeating the ITC measurements with the same peptide, but this time in which the twin arginine motif has been mutated to twin lysine, a change which both *in vivo* (Stanley *et al.*, 2000) and *in vitro* (Alami *et al.*, 2002) has been shown to inhibit the Tat transport of previously competent proteins, is shown in **Figure 4.3.1 B**. The TatC protein was completely unable to bind to this peptide, which demonstrated the interaction observed with the 'RR' peptide is highly specific to the conserved twin-arginine motif.

It has been shown that TatC, within the TatBC complex, binds closely to the RR motif within the n-region of the peptide, whereas TatB contacts the region more downstream of this (Alami *et al.*, 2003). Thus TatC maybe involved in specifically recognising the n-region and the conserved RR motif within the peptide (Alami *et al.*, 2003) (Gérard & Cline, 2006). Therefore, a short, nine amino acid, version of this SufI peptide, comprised

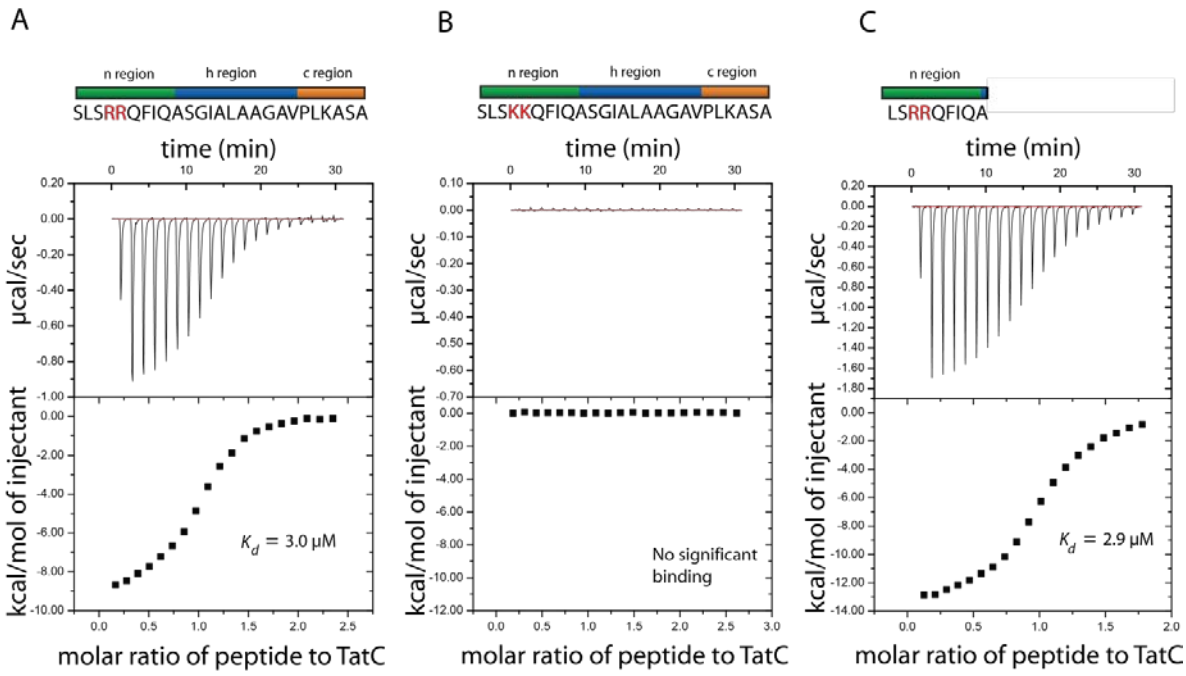


Figure 4.3.1 Interaction of *A. aeolicus* in LMNG with *E. coli* signal peptides.

ITC was used to assess the interaction between TatC and various synthetic peptides **(A)** the 26 amino acid *E. coli* SufI signal peptide **(B)** a functionally inactive KK mutant of the peptide, **(C)** a short, nine amino acid version of the *E. coli* SufI peptide, encompassing the n region and the RR motif.

mainly of the n-region around the RR motif, was tested for binding to TatC in LMNG. This peptide was amidated at the CT during synthesis (Generon) to avoid the negative charge of the C-terminal carboxylic acid potentially interfering with the chemistry of the peptide. The ITC was carried out as for the two previous experiments.

The results for the interaction with this peptide is shown in **Figure 4.3.1 C**. This showed that this peptide interacted with the same affinity as the longer peptide. This implied that all the information needed for TatC to recognise the signal peptide is present within this n-region, and the remaining h- and c-regions are not specifically interacting with TatC, at least by this *in vitro* assay.

Structural insights into the interaction

Having shown that TatC is functional for signal peptide binding, the structure was examined to see if there was any information on how TatC might be specifically recognising the signal peptide.

Given the signal peptide contains an electrostatically positive invariant twin arginine motif it may be recognised by a negative region on the surface of TatC. **Figure 4.3.2 A** shows the surface electrostatic representation of the structure, calculated using the pymol plugin APBS (Baker *et al.*, 2001), with regions of negative charge shown in red and regions of positive charge in blue. As expected, most of the region of the protein exposed to the membrane is not very charged, with the intriguing exception of the negative glutamate 165 residue, which sits right in the centre of the curved cavity, as discussed in **section 4.1**. The potential role of this residue will be discussed in further detail in **Chapter 6**.

Inspecting the rest of the structure shows that there is a positive region around the NT end of helix 1, as was noted in the MD simulations which could contribute to how TatC distorts the membrane in this region. There is also a patch of charge at the position where the periplasmic cap region is expected to contact the lipid head groups. If the structure is viewed from the cytoplasmic face, there is a region of negative charge that coincides well with the highy conserved and functionally important region (i) in **Figure 4.2.2.** This could suggest that this negative patch on the structure is important for function, and that this function could potentially be linked to the recognition of signal peptides with the electrostatically positive twin arginine motif.

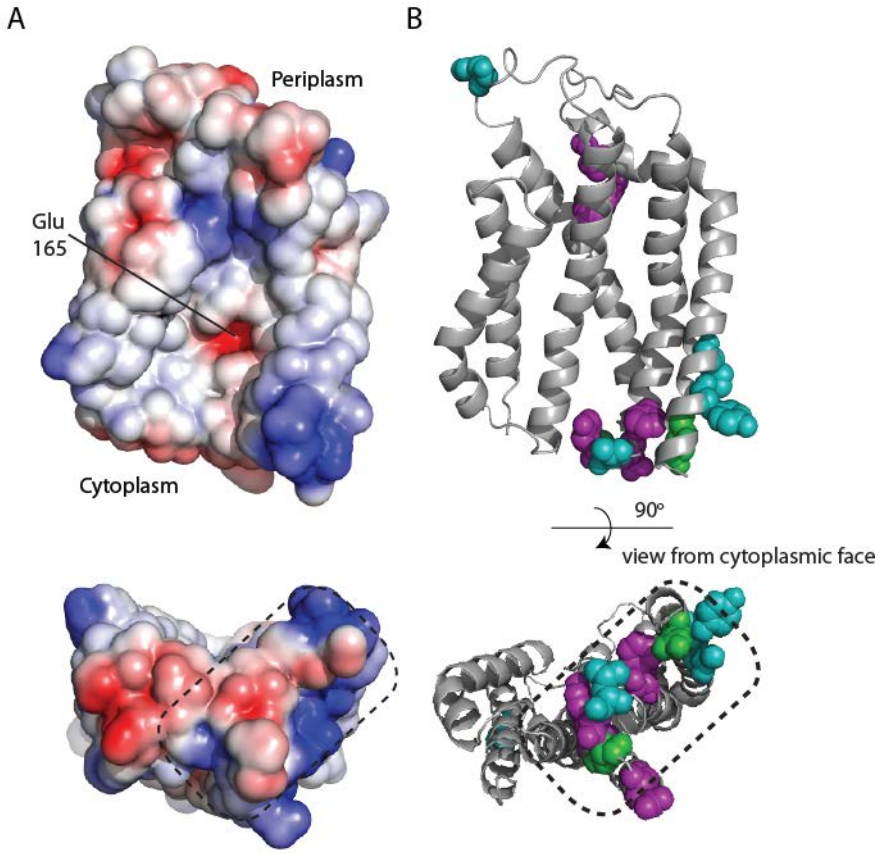


Figure 4.3.2 Identification of the signal peptide binding site on TatC.

(A) The electrostatic surface representation of TatC, showing exposed patches of negative charge (red) positive charge (blue), and non-polar regions in white. **(B)** Residues implicated in signal peptide binding identified by inhibition of crosslinking of the signal peptide to TatC upon TatC mutation (purple) (Holzapfel *et al.*, 2007) photo crosslinking from positions in TatC to the signal peptide (green) (Zoufaly *et al.*, 2012) and genetic studies (blue) (Kreutzenbeck *et al.*, 2007, Lausberg *et al.*, 2012, Strauch & Georgiou, 2007). Colour code given in **Table 4.2.**

Previous studies have shown that the signal peptide maybe able to contact the surface of the membrane independently of the Tat translocase, to form a lipid bound, functionally important intermediate (Bageshwar *et al.*, 2009). Once peripherally associated they would be able to undergo a two-dimensional search of the membrane surface, until they are specifically recognised by TatC. This face of the protein, which is barely exposed out of the membrane (as in the MD simulations in **section 4.2**) presents itself as the initial site that peptides will encounter on TatC. This negatively charged patch will then be well

positioned to engage and bind to signal peptides; which maybe now present in an α helical conformation (San Miguel *et al.*, 2003). However, given the twin lysine version of the peptide, which would be expected to have similar electrostatics to the twin arginine version, does not bind to TatC (**Figure 4.3.1 B**) the interaction must also be mediated by the shape of the peptide, as well as charge.

Again, previous studies have been undertaken in which various *E. coli* residues have been implicated in being involved in this interaction. Genetic studies allowed for the investigation of which residues, upon mutation to alanine, can block previously observed crosslinks between TatC and the signal peptide (Holzapfel *et al.*, 2007). Site specific photo crosslinking has also been carried out, in which a light activated cross linker was placed at various positions in TatC (Zoufaly *et al.*, 2012). Genetic screens also identified residues that upon mutation can suppress defective peptides that harbour sequence alterations that made them previously incompetent for transport (Kreutzenbeck *et al.*, 2007, Lausberg *et al.*, 2012, Strauch & Georgiou, 2007).

The residues identified in these studies are given in **Table 4.2**. The identified residues can then be plotted as spheres onto the structure of TatC, which is shown in **Figure 4.3.2 B**. The residues identified in more than one study are coloured according to that shown in the **Table 4.2**. Mapping the residues onto the structure shows that most now localise to patch (i) in **Figure 4.2.2**, indicating this may be a binding site. Of the three residues plotted that are located elsewhere, Phe135 is located on periplasmic loop 2, Pro 42 and Tyr 119 are facing each other on helices 1 and 3. Pro 42 causes the kink in the interfacial helix of helix 1. This is a key region of structure, and along with the Tyr 119 it is possible this interaction helps to maintain the architecture of these helices within the membrane. Mutation to alanine may cause a partial collapse of this region of the TatC structure, which may effect signal peptide binding to the cytoplasmic end of TM helix 1 and explain the inhibition of

***E. coli* residue**	***A. aeolicus* residue**	**Signal peptide crosslinking to *E. coli* TatC blocked when substituted to Ala**	**Position of photocrosslink from *E. coli* TatC to substrate**	**Suppressor of signal peptide defects in *E. coli* TatC**
Glu15	Glu9		Glu15	
Lys18	Tyr12			K18E, K18M
Asn22	Ile16			N221
Pro48	Pro42	Pro48		
Phe94	Phe87	Phe94		F94S
Leu99	Leu92			L99P
Tyr100	Tyr93	Tyr100		
Lys101	Ser94		Lys101	
Glu103	Glu96	Glu103		
Tyr126	Tyr119	Tyr126		
Pro142	Phe135			P142S

Table 4.2 *E. coli* TatC residues implicated in signal peptide binding by three types of experiment. The equivalent *A. aeolicus* residue is given, along with the colour code used in **Figure 4.3.2 B.** Purple residues (Holzapfel *et al.*, 2007), green residues (Zoufaly *et al.*, 2012) and blue residues (Kreutzenbeck *et al.*, 2007, Lausberg *et al.*, 2012, Strauch & Georgiou, 2007).

the previously obeserved crosslinks between the signal peptide and TatC.

A binding site on patch (i) in **Figure 4.2.2** is also consistent with the predicted positions of the residues identified very recently in the plant TatC homologue, cpTatC, as compentant to cysteine crosslink to the RR region of the signal peptide (Ma & Cline, 2013). In order to test whether the residues in this region have a direct effect on signal peptide binding, and allow their effect to be more accurately quantified, the ITC assay was repeated using the short SufI peptide (as in **Figure 4.2.1 C)** and various alanine substitutions of the TatC protein.

Initially, five residues that are located in patch (i) and were previously identified as important for signal peptide binding (**Table 4.2**) were mutated to alanine (*A. aeolicus* E9A, F87A, L92A, Y93A and E96A) along with a control mutation, D205A. This residue has been implicated in being important for TatC function (Buchanan *et al.*, 2002), but is

located in the second highly conserved patch, (ii) on the structure. This would therefore not be expected to be involved in signal peptide binding, and should exhibit a WT-like $K_{D.}$

The primers in **Table 4.3** were used to introduce the mutation to the pWaldo TatC-GFP vector, using the Quikchange mutagenesis XL kit (Agilent), and following the manufacturers protocol (See **Appendix 8.5**). The proteins were then expressed and purified in LMNG as described in **section 3.1**, and the ITC carried out with the short SufI peptide as described previously. The isotherms (**Figure 4.3.3)** showed that the control mutation, D205A, had no effect on the signal peptide binding, and this still binds with the same affinity as WT (3μM). However, the other residues displayed reduced binding affinities.

The severity in the reduction of the K_D of the signal peptide interaction roughly fell into two categories, substitutions that led to a ten-fold reduction in affinity (E9A and L92A) and substitutions that gave a greater than 50-fold reduction (over 200μM K_d, E96A, Y93A and F87A) compared to the WT K_D of 3μM. The F87A protein exhibited such a low level of binding (**Figure 4.3.3 A)** that the exact value for the K_D could not be determined as the experiment was not appropriately designed to measure such a weak $K_{D.}$ A higher concentration of peptide would need to be used in order to determine the K_D for this interaction.

The calculated K_D's for the signal peptide interaction with the substituted TatC proteins can be plotted on the structure (**Figure 4.3.4**). The residues which reduced the K_D to ~50 μM are shown in pink, and those that reduced the K_D to over 200 μM are shown in red. This 'heat map' indicates that residues from patch (i) are involved in signal peptide recognition, and shows the residues exposed on the cytoplasmic loop and the glutamate at the end of helix 1 may form part of a linear patch that can recognise the signal peptide.

E. coli residue	*A. aeolicus* residue	Forward primer (5' to 3')	Reverse Primer (5' to 3')
Glu 15	**E9A**	TGC CTC TAA CGG AAC AT CTG AGA GCC CTT AGG TAT CGT TTA ATA ATA TC	GA TAT TAT TAA ACG ATA CCT AAG GGC TCT CAG ATG TTC CGT TAG AGG CA
Phe 94	**F87A**	CTT TAC CAG TTC TGG CGT GCC ATA GAG CCA GCC CTT TAC	GTA AAG GGC TGG CTC TAT GGC ACG CCA GAA CTG GTA AAG
Leu 99	**L92A**	CTGGCGTTTTATAGA GCCAGCCGCTTACTC CCATGAAAAAAGAG CTT	AAGCTCTTTTTTCATGGG AGTAAGCGGCTGGCTCT ATAAAACGCCAG
Tyr 100	**Y93A**	GGCGTTTTATAGAGC CAGCCCTTGCCTCCC ATGAAAAAAGAGCTT TTAT	ATAAAAGCTCTTTTTTCA TGGGAGGCAAGGGCTGG CTCTATAAAACGCC
Glu 96	**E96A**	CCA GCC CTT TAC TCC CAT GCA AAA AGA GCT TTT ATC CCG	CGGGATAAAAGCTCTTT TTGCATGGGAGTAAAGG GCTGG
Asp 211	**D205A**	GGT GCG ATA ATC GCT CCG GCT GTT TCC ACA CAG GTT TTG	CAAAACCTGTGTGGAAA CAGCCGGAGCGATTATC GCACC

Table 4.3 Primers used to introduce alanine mutations of the residues identified with a functional role in binding signal peptides. The template for the mutagenesis was the WT TatC-GFP pWaldo vector
.

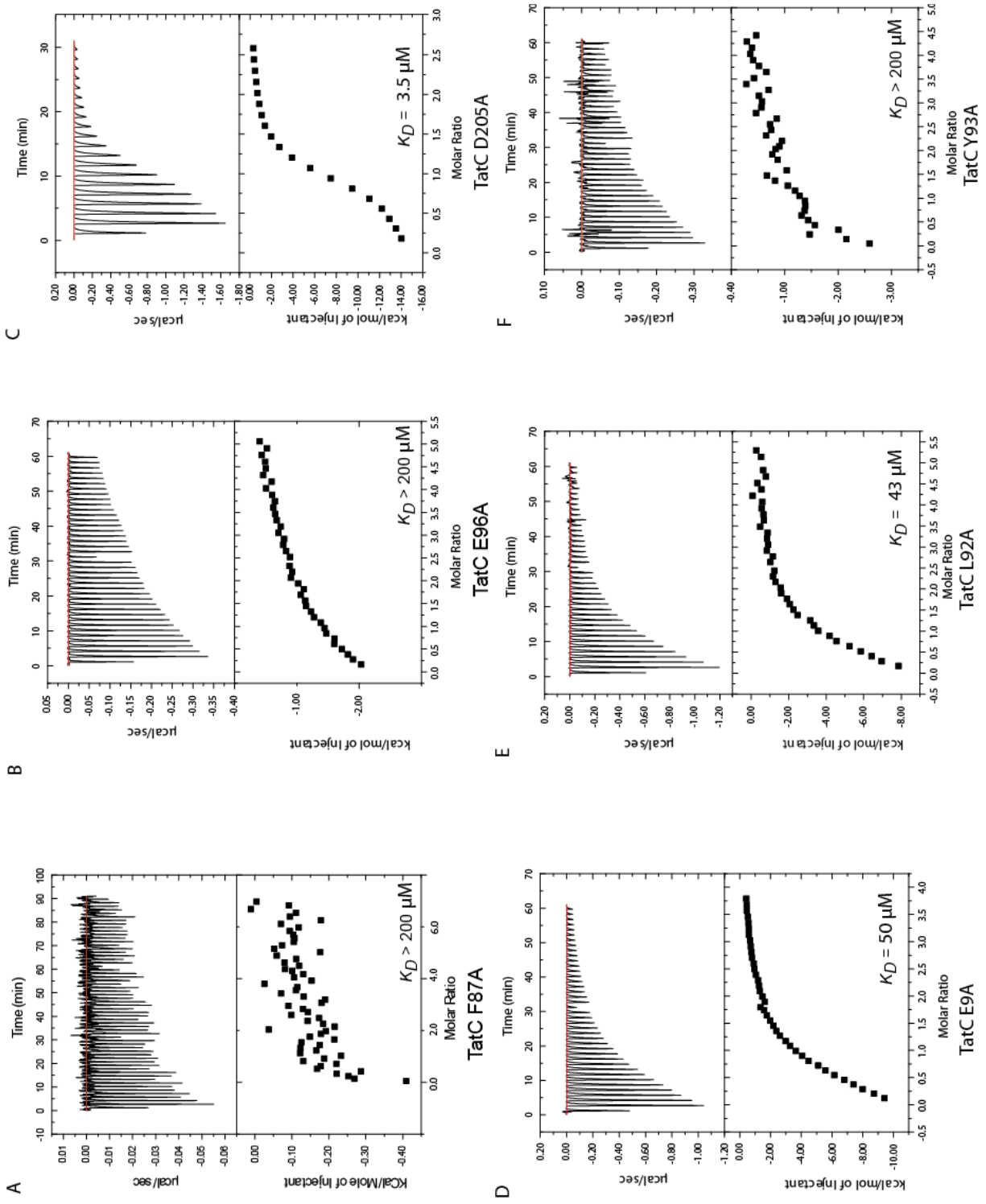


Figure 4.3.3 Isothermal titration calorimetry of *A. aeolicus* TatC variants.

(A-F) Variants of TatC purified in LMNG and titrated with the SufI short signal peptide. All mutations in patch (i) **(A-B, D-F)** show a reduction in the binding affinity, K_d, whereas D205A **(C)**, located in patch (ii), shows a WT K_d.

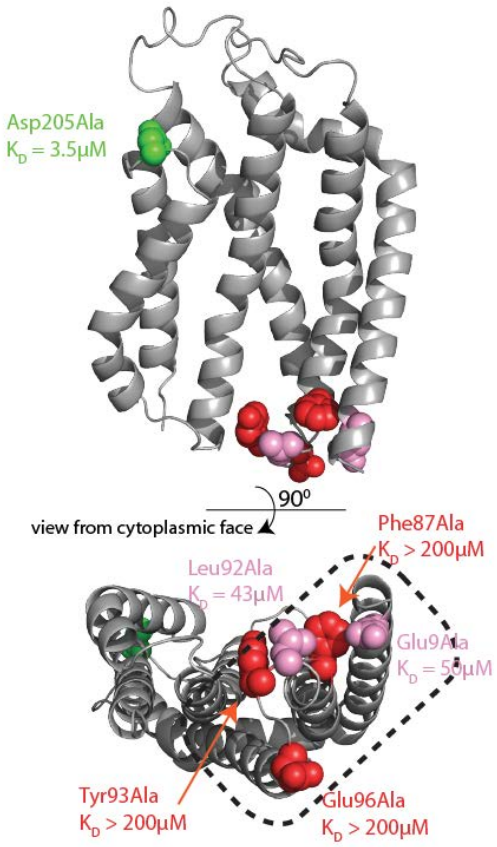


Figure 4.3.4 Characterisation of the residues involved in the signal peptide interaction with TatC.

Effects of the indicated amino acid substitutions on the affinity of *A. aeolicus* TatC for the short SufI signal peptide. Substitutions that affect binding are coloured in pink and red, and those with no effect are shown in green, as from the ITC experiments (**Figure 4.3.3)**. Residue side chains are represented as spheres.

To confirm whether this region of the cytoplamsically exposed face of TatC is alone involved in binding, further subsitutions along the entire face would need to be made, and their binding affinity tested.

The greatest effect on signal peptide binding K_D out of all the substitutions tested was the F87A substitution. The F87 residue is removed from the membrane surface slightly, on helix 3 rather than on the loop where the other three nearby residues reside (E96A, L92A and Y93A). The reduction in binding seen with the F87A substitution could be due to a

direct contact with the signal peptide. The F87 aromatic group could potentially π bond to the Phe residue at the +2 position from the RR pair, consistent with data that the Phe residue on the peptide is important for tight binding of the peptide to TatC (Gérard & Cline, 2007). The F87 side chain is orientated so it points down towards the loop, and so could be able to contact the peptide. Alternatively, given the key location of F87 at the end of the helix 2, before the loop on which the other three important residues tested by ITC reside, the reduction in signal peptide binding could be due to a structural effect. Mutation of F87 could alter the conformation of this entire loop, so that the three important residues are now no longer held in the correct orientation to bind to the peptide. A structural effect could explain why this mutation is picked up in so many studies in *E. coli* that address signal peptide binding, as altering this is equivalent to altering all the three other residues on this loop.

The ITC experiments showed that the two glutamate residues (E9 and E96) affected binding and so suggests that they could be involved in binding to the positively charged groups on the twin arginine residues. Crystal structures of other proteins involved in recognition of conserved arginine residues, such as the yeast mitochondrial processing peptidase (MPP), revealed the arginine binding site consists of a conserved glutamate residue that forms a salt bridge to an arginine side chain (Taylor *et al.*, 2001). It is interesting to note that during the molecular dynamic studies of TatC in a membrane (**section 4.2**), one of the regions of flexibility in the otherwise very static structure was the end of transmembrane helix 1, where Glu 9 is located, and also the cytoplasmic loop region around Glu 96. The simulations could imply that inherent flexibility in this region of the structure would allow the two glutamate residues to be able to move in closer to one another upon each binding an arginine residue within the RR motif. This small scale

conformational change could be a signal to initiate the subsequent steps (TatA recruitment, followed by passage across the membrane) of Tat transport (Mori & Cline, 2002).

Primers to introduce alanine substitutions for the final three residues identified as being important for the interaction of TatC with the signal peptide (**Table 4.2)** and which also reside in patch (i), L12A, L16A and S94A were designed and the constructs made (**Appendix 8.5.1**). These proteins were purified and the ITC carried out by Michael Roberts (Part II student) as in **section 4.3**. These ITC experiments showed no alteration in the K_d of the TatC-signal peptide interaction for these three TatC substitutions compared to WT (Michael Roberts-Part II thesis). These results indicate that the region further up the TM 1, at position 12 and 16, does not directly contribute to the binding of the signal peptide. Inspection of the structure shows these residues are both further up helix 1, and so removed from patch (i), but their side chains also point away from the binding site and outwards into the lipid membrane. An explanation for the observed suppression of signal peptide defects by the equivalent *E. coli* residues (Kreutzenbeck *et al.*, 2007, Lausberg *et al.*, 2012) could be due to indirect structural effects such as altering the conformation of the functionally important E9 residue further down this helix.

The S94A substitution had no effect on signal peptide binding. Despite the S94 residues location on the cytoplasmic loop in patch (i), immediately adjacent to the L92 and Y93 residues and two residues removed from the E96 residue, it is not involved in this surface exposed binding site for the consensus sequence. The S94 residue is not conserved (it is Lys101 in *E. coli*) and has not been implicated as being functionally important in other studies. Its ability to photo crosslink to signal peptides may arise due to its location on this loop and proximity to the signal peptide, once the signal peptide has been held in place on TatC by the other residues in this region.

A potential route for signal peptide insertion into the membrane?

Evidence of a loop, or hairpin, mechanism of insertion of the signal peptide into the membrane (Fincher *et al.*, 1998) was very recently shown to be due to the action of TatC, independent of either the other two transmembrane proteins of the translocase or the Proton Motive Force (PMF) (Fröbel *et al.*, 2012a). In a serious of elegant experiments using *E. coli* derived inner membrane vesicles (IMV) containing TatABC or TatAC, and *in vitro* synthesised precursor, Frobel *et al*. were able to show cleavage of signal peptides at the peptidase site in the h-region occurs in the absence of TatB and precursor translocation and is promoted by the presence of TatC. The ability of the periplasmically located signal peptidase to cleave the signal peptide indicates the h-region has been transferred across the membrane. The cleavage was independent of added protonophore, demonstrating the insertion was not dependent on the PMF.

TatC therefore not only has the ability to specifically recognise the twin arginine motif of the signal peptide, but is able to insert these into the membrane. Periplasmic location of the h-region of the signal peptide would explain the previously observed site of suppressor mutations that allow transport of defective peptide located on the periplasmic portion of the TM helix of TatB (Kreutzenbeck *et al.*, 2007, Lausberg *et al.*, 2012), as well as the site of the suppressor mutation in TatC on the periplasmic loop 4 (**Figure 4.3.2 B**) (Strauch & Georgiou, 2007). The suggested 'hairpin' arrangement of the signal peptide in the membrane would still allow the signal peptide to remain captured at its n-region by the RR motif bound in the cytoplasmically exposed binding patch (i) of TatC identified in this study, and so rationalise data which demonstrated a signal peptide covalently linked to TatC is still competent for passenger transport (Gérard & Cline, 2006).

Signal peptide insertion into the membrane is also consistent with data in which the signal peptide is seen to bind to the thylakoid translocase in two modes; a loosely associated state, in which the precursor is easily washed away with salt buffers from thylakoid membranes and the signal peptide remains accessible to added proteases, and a tightly bound state, in which the precursor is resistant to salt washes (Gérard & Cline, 2007). The loosely bound state could represent the signal peptide bound at its n-region to the surface exposed patch (i) on TatC (**Figure 4.3.4**) potentially mediated by electrostatic interactions between the two conserved and functionally important glutamate residues (**Figure 4.3.3**). This interaction would be consistent with the susceptibility to salt removal in the thylakoid washes. The tightly bound state may represent the signal peptide once it has been subsequently inserted into the membrane by TatC. The conversion between these two states in the thylakoid membrane was found to be dependent on the PMF.

As this study identified no difference in the binding affinity between the long and short peptide (**Figure 4.3.1**), it is not clear whether the insertion of the h-region is into a specific binding pocket within TatC, or whether TatC can provide a more passive ‘passageway’ for the peptide. Given the concave nature of the membrane exposed ‘cleft’ shape that TatC makes in the membrane, near to the exposed signal peptide binding site (**Figure 4.1.1A**), which connects to the periplasmic side of the membrane, it is tempting to envisage this ready made ‘channel’ across the membrane could provide this passageway for the h-region of the signal peptide. Modelling experiments performed by Dr Phillip Stansfeld were able to sensibly fit a full length signal peptide from patch (i) through the channel on the TatC structure to the periplasmic groove.

The first indication as to the path of the signal peptide though the membrane were given in recent, unpublished work using florescence dye quenching (S.Musser *et al*, unpublished).

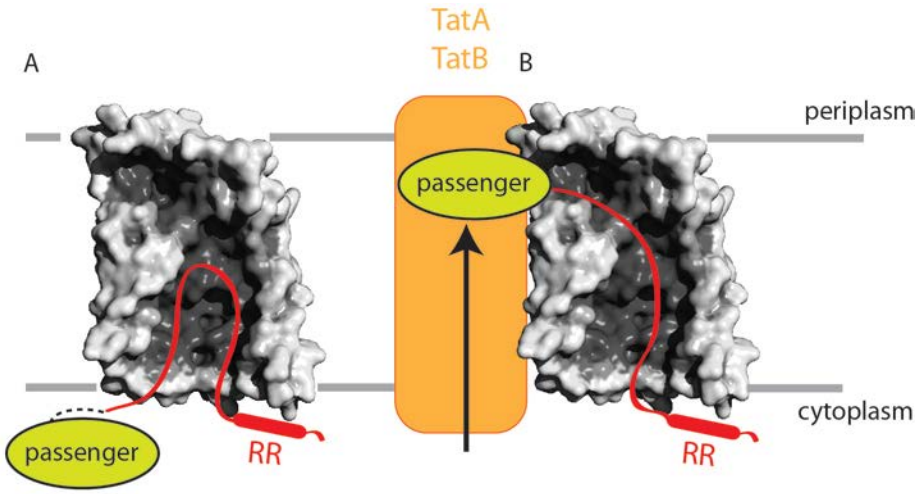


Figure 4.3.5 A potential pathway for the h region of the signal peptide into the membrane, promoted by TatC.

(A) The h-region of the signal peptide (red line) is thought to insert into the membrane, firstly by a hairpin mechanism. This could be envisaged by use of the 'cavity' within TatC (shown as a shaded surface), with the n-region and its RR motif (red tube) bound at the surface exposed binding site (patch (i)) located proximal to this. (**B)** Unpublished evidence (Musser *et al*.) that the signal peptide can alter its conformation when bound to TatC, during the later stages of translocation, so the peptide is fully transported across the membrane (Gérard & Cline, 2007, Fröbel *et al.*, 2012a), potentially through the rest of the cavity leading to the periplasmic space. This would need to be coupled to TatA recruitment, to allow the passenger protein to cross the membrane.

A fluorescence quencher was placed on the end of different length phospholipids and a fluorescence probe was present at different positions on the signal peptides. This allowed determination of the positions of certain amino acids on the signal peptide within the membrane before and after engaging with the TatBC receptor complex and gave further evidence for this 'hairpin insertion'. They were able to rationalise their results on the basis of the TatC structure; the observed central cleft in TatC could form the passageway for the signal peptide to cross the membrane. If the surface exposed patch identified above binds the RR proximal n-region, this would allow the h- and c-region of the peptide to loop into the membrane, whilst the consensus sequence remains 'locked down' on the surface exposed binding site (**Figure 4.3.5**).

In addition, the quenching experiments showed the hairpin was able to change conformation, which led Musser *et al*. to predict a 'hinge' region in the signal peptide, capable of alternating the conformation from a tight hairpin (**Figure 4.3.5 A)** to one in which the peptide more fully crosses the membrane later in the Tat transport cycle, and travels all the way through the TatC cleft (**Figure 4.3.5 B)**. This change would require TatA recruitment to allow the passenger protein attached to the end of the signal peptide to also cross the membrane. The signal peptide would now be placed near the periplasmic edge of the membrane, where it would be accessible to the periplasmically located signal peptidase and be cleaved (Fröbel *et al.*, 2012a). The alteration in signal peptide conformation was seen to be independent of the PMF. In the absence of TatA or TatB, and therefore the absence of passenger protein transport, Fröbel *et al* had to engineer an unstructured region into the passenger protein to allow the peptide to adopt the deep hairpin conformation, and so for the signal peptide cleavage to occur.

The major dispute with the data discussed above is exactly what role the PMF plays in the membrane insertion of the signal peptide. The alteration between loose and tight binding states in thylakoids was dependent on the PMF (Gérard & Cline, 2007), whereas the model proposed by Musser *et al*. and the signal peptide insertion seen by Frobel *et al*. were both PMF independent. Whether the 'advanced stage' of signal peptide binding seen by Gérard *et al*. constitutes just the hairpin insertion, or represents a subsequent transport step in which TatBC localised TatA is able to bind the signal peptide (Fröbel *et al.*, 2011) and potentially bury it in a pocket between all three Tat components, is not clear. Further experimental information will be required to assess the role the PMF plays in the molecular details of the conformation of the signal peptide in the membrane.

Docking experiments carried out as part of the recent paper describing the two subsequent *A. aeolicus* TatC structures (Ramasamy *et al.*, 2013) proposed a slightly different model

from the cartoon in **Figure 4.3.5**. They modelled the first 15 amino acid sequence of the *A. aeolicus* version of the SufI peptide, in an α helical conformation as has been shown for the *E. coli* version in a membrane (San Miguel *et al.*, 2003) and observed the n-region binding straight into the membrane exposed cleft of TatC; with no clear interactions between the signal peptide and the TatC residues on the surface exposed binding site, region (i). They propose that conserved a lysine residue of the signal peptide, at position 8 in the *A. aeolicus* SufI peptide is therefore positioned to bind to the negatively charged residue in the centre of the membrane, Glu165. However, the lysine in the signal peptide sequence they used for docking is not conserved in other SufI signal peptide sequences, and was not present in the peptide used in the ITC experiments performed in this study. In addition, the docked model did not provide an appropriate explanation for the surface exposed residues implicated previously in binding to the n-region, and that have been shown in this study to have a direct effect in binding the peptide. To test if the residues within this cavity are involved in forming the specific, initial recognition site, along with those on the surface exposed binding site proposed in this study, an alanine substitution of the glutamate 165, and substitutions of other residues in this cavity, will be tested.

Docking versions of the *E. coli* short signal peptide was also attempted as part of this study by Dr Phillip Stansfeld (Department of Biochemistry, University of Oxford). The peptide was seen to localise to the surface exposed face, in patch (i), and the conserved Glu 9 and Glu 96 were able bind to an arginine of the RR pair. However, running the simulations with the KK version of the same peptide, under the same conditions, gave the same result (not shown). This mutant peptide should not bind to TatC (**Figure 4.3.1 B**), thus the interactions driving the peptide in the simulations to bind to the TatC structure could not be interpreted as meaningful.

It is interesting to note that in an analogous fashion to the Sec translocon, TatC is also able to insert the signal peptide into the membrane. However, the evidence discussed here suggests the two proteins do so in very different ways; the Sec channel allows its signal peptide to intercalate between two of its helices, which partially move apart (Gumbart & Schulten, 2007) where as TatC may make use a 'passage way' across the membrane, formed from the unique arrangement of its helices.

4.4 Summary

The structure of TatC revealed a novel membrane protein architecture, and many unusual and distinctive features. The behaviour of TatC within the membrane was investigated using molecular dynamic simulations, which showed the structure is very stable within a membrane environment. The wealth of information in the literature regarding functional importance of residues within the *E. coli* protein was exploited to identify two functionally important patches on the protein, which correlated very well with the regions of the structure identified as being highly conserved.

One of these patches, located on the cytoplasmic face of the structure, contained residues previously implicated in signal peptide binding by biochemical studies. These residues were tested for their importance in mediating the signal peptide interaction by use of an ITC assay with a short region of the SufI signal peptide and mutant versions of *A. aeolicus* TatC purified in LMNG. This allowed the exposed binding site to be probed, and the residues within it directly assayed for their contribution to binding. Results from other groups, both published and unpublished, indicate the signal peptide is able to undergo deeper insertion into the membrane, potentially by insertion into the lipid exposed 'cavity' within the centre of the structure of TatC, near this surface exposed binding site. However,

despite these studies, a molecular level understanding of how the twin arginine motif is specifically recognised is still elusive. To more thoroughly understand this, a structure of TatC and the signal peptide in complex is required.

Chapter 5: Attempts to crystallise TatC and the signal peptide in complex

The structure of *A. aeolicus* TatC described in **Chapter 4** has provided some clues into the potential nature of the interaction of TatC with the signal peptide, and has given an important platform for the design of future experiments that can test specific hypotheses, as has been shown by glimpses at other groups unpublished work. However, the understanding of the crucial interaction between TatC and the signal peptide would be aided greatly by obtaining a structure of TatC in complex with the signal peptide; this would provide atomic detail on exactly which residues contact the twin arginine pair, and what contacts the other residues of the motif make with the residues in patch (i) on TatC.

As discussed in the introduction, the structures of the signal peptide targeting factor, SRP, from the Sec protein transport system both alone (Keenan *et al.*, 1998) and in complex with the Sec signal peptide (Janda *et al.*, 2010) have begun to unravel the determinants of how this signal peptide is recognised by the soluble SRP factor. However, given that SRP is faced with the more difficult task of binding Sec signal peptides that vary in length and sequence more than Tat signal peptides, many structures are likely to be needed to fully understand the requirements for binding. As the Tat signal peptide contains the almost invariant and necessary RR motif, a single co-crystal structure of it in complex with TatC will represent exactly how the peptide is recognised by TatC in all cases. To this end, a number of approaches have been attempted in this study: peptide soaks, co-crystallisation, trying different signal peptides, and fusion of the signal peptide to the termini of TatC.

5.1 Peptide soaks

In order to try to obtain structures of proteins in complex with their ligand, an approach that has been successful in many cases (Hassell *et al.*, 2007) involves the use of pre-grown crystals and the addition of a highly concentrated solution of the ligand in the mother liquor crystallisation condition. This approach seemed feasible for the *A. aeolicus* TatC H32 crystals. The nine amino acid 'short' SufI peptide is relatively small (~1kDa), and the solvent channels within the H32 crystal form of TatC are very large, with the entire crystal having a solvent content of ~73% (**3.3**). Furthermore, the potential signal peptide binding site on TatC (patch (i) in **Figure 4.2.2**) in the H32 crystals is completely exposed to the solvent channels. The patch is not occluded by the other copies of the TatC molecule in the crystal, and is not involved in forming the crystal contacts. This would allow the peptide to access the binding site on all the TatC molecules within the crystal.

Pre-grown crystals of TatC in LMNG, either native or SeMet crystals from the optimisation trays in **Chapter 3,** were soaked in various concentrated solutions of the short peptide. Solutions of the peptide at either 20mM, 10mM or 5mM were used, made up in the mother liquor crystallisation condition. However, the peptide was not fully soluble in this condition and some precipitation was seen following a 13,000 rpm centrifugation of the 20mM stock. The actual concentration of this saturated stock will therefore have been less than 20mM. Crystals were soaked by addition of the crystals to 1µl of the stock, and crystals were fished after various amounts of time, 10, 15, 18, 20 and 25 minutes.

The crystals were then harvested with cryo loops (Hampton research) and flash frozen in a cryo protectant solution of mother liquor, free of peptide. Crystals were exposed to X-rays at the ID29 beam line at the ESRF. Nine datasets were collected, varying in resolution from 3.53Å to 4.20Å. These datasets were used in autoBUSTER (Bricogne, 2011) to refine

the *A. aeolicus* TatC model. The resulting electron density difference maps, α_{calc} $|F_o|-|F_c|$, were inspected in COOT (Emsley & Cowtan, 2004) to see if there was density that was not accounted for by the model. No obvious peptide-like density appeared in these maps.

However, as discussed in **section 4.1**, the small scale motions seen in the atomistic MD studies in the vicinity of the proposed signal peptide binding site (patch (i)) on the cytoplasmic face of TatC could indicate that this region is flexible. TatC may be able to alter conformation subtly in this region upon signal peptide binding. The alteration in the protein structure could be enough to alter the crystal contacts between TatC molecules within this crystal form, and break up the crystal lattice. As helix 1 is involved in contacting helix1 from the neighbouring TatC molecule, a subtle motion of these two helices could cause this crystal contact to break apart.

5.2 Co-crystallisation of *A. aeolicus* TatC with the signal peptide

If conformational change does occur to any great extent, a new crystal form of TatC, with different contacts could form. In order to screen for new conditions, the protein and peptide mixture would need to be subject to crystallisation trials. In order to conduct these trials, SeMet TatC in LMNG was expressed and purified (as in **section 3.3**), as this protein species gave better crystals in the optimisation screens than the native TatC. The protein was then mixed with a concentrated stock of the short peptide, made up in the SEC LMNG buffer.

The ITC experiments in **section 4.3** showed that a three times molar ratio of signal peptide to TatC allowed for the saturation of all the TatC protein with peptide. This ratio was the minimal molar excess that needed to be added to the purified protein before pipetting crystallisation trays, to ensure full occupancy of all the TatC molecules within a crystal

with peptide. Different ratios of the peptide to TatC were tested. A ratio slightly above the minimal ratio of three molar signal peptide:TatC was tested of 5:1, along with a higher ratio of 10:1. These mixtures were made by adding a small volume of a 20mM stock of the peptide in buffer to the final TatC sample, at a concentration of approximately 8mg/ml. These mixtures were incubated overnight at 4^{o}C to equilibrate. For crystallisation trials, the same optimisation trays were used as for TatC in LMNG without the signal peptide, as in **section 3.3**. These screens were in case the H32 crystal form is still able to be produced in the presence of the peptide, and the soaks were unsuccessful for reasons of solubility of the peptide in the mother liquor stock. However, if there is a small alteration in the structure of TatC upon binding the signal peptide, the H32 crystal form may no longer be able to form. A new crystal form with different packing interactions may be able to grow in a different mother liquor condition. In order to test this, the TatC-signal peptide mix was subject to the MemGold broad crystallisation screen (Molecular Dimensions), as described in **section 2.3**. Trays were incubated at 21^{o}C.

No crystals in any new conditions grew in the broad screens, but the protein crystallised in many of the optimisation screens from the original condition in **section 3.3**. These crystals were harvested, as described previously, and exposed to X-rays at the I04-1 beam line at Diamond. The crystals were of the same form as for the non-peptide crystals, H32, and the same unit cell as previously seen. Datasets from 11 different crystals were collected, with maximal resolution of the diffraction of the datasets varying from 3.8 to 4.4 Å. The datasets were used, as with the soaked datasets, to run rounds of refinement against the final TatC model. Again, inspection of the difference maps gave no peaks of peptide-like density.

The interaction measured between the peptide and the protein has a quite low affinity, 3μM. Interactions of this strength are often very difficult to co-crystallise (Yamada *et al.*,

2005), and even co-crystal structures of interactions that are much tighter, such as in the nanomolar affinity range can still remain elusive (Wallrapp *et al.*, 2008). Furthermore, the conditions of the mother liquor, the low pH (0.02M sodium acetate, pH 4.0) and the PEG concentration, may further reduce the affinity of the protein for the peptide and no longer allow binding, making a co-crystal structure in this condition, with this peptide, even harder to achieve.

5.3 *A. aeolicus*-like signal peptide sequences

Increasing the strength of the interaction between the peptide and TatC was explored, in the hope a stronger binding peptide could stay bound to TatC within the LMNG H32 crystallisation condition. A peptide with a sequence that is more similar to the consensus sequence of Tat signal peptide from *A. aeolicus* Tat substrate proteins, rather than the *E. coli* SufI sequence, may bind to *A. aeolicus* TatC more tightly. In order to test this, a new peptide was designed, based on sequence alignments for known and predicted *A. aeolicus* signal peptides. This encompassed only seven amino acids of the highly conserved consensus sequence, SRRDFLK. Unlike the SufI peptide, this was not amidated on the CT.

Repeating the ITC with the new peptide showed no binding of the peptide to TatC. This could potentially be as the sequence is shorter than that in the *E. coli* peptide, or that the lack of amidation at the CT may now have caused electrostatic interference between the peptides, causing it to self associate. A second peptide was therefore designed, which was slightly longer and has no negative charge, with the sequence SRRGFLKLS and was amidated on the CT. However, ITC showed this peptide was still unable to bind to TatC. Future work will entail designing an extensive range of peptide sequences based on the *A.*

aeolicus consensus sequence, and including exact, longer sequences from *A. aeolicus* Tat transported proteins, in order to find one that can bind.

5.4 Signal peptide fusions

If a different sequence that binds with a stronger K_d cannot be found, then a way to help improve the efficiency of binding of the SufI short peptide to TatC was explored. In the Sec transport field, the only available co-crystal structure of SRP in complex with the Sec signal peptide was recently achieved by the use of fusion constructs. The signal peptide was fused to the SRP protein via a flexible linker, allowing the co-crystal structure to be solved to 3.5Å resolution (Janda *et al.*, 2010). Other examples exist of peptide fusions to the termini of their interaction partner proteins that have been able to give co-crystal structures (Pellegrini *et al.*, 2002, Strushkevich *et al.*, 2011). It was speculated that this approach could also be feasible for TatC, by fusing the signal peptide to the protein this could hold the peptide closer to the binding site. This could help to maintain the contact between the two during the crystallisation process, so if the peptide does dissociate it is held close enough to rebind into the binding site very efficiently. The fusion would effectively enhance the affinity of the signal peptide for TatC.

The signal peptide is thought to bind near the NT region of the protein, at the start of helix 1 and in the first cytoplasmic loop (patch (i)). However, it is not known in exactly which orientation the peptide binds to TatC. In order to try and obtain a construct for crystallography that has the signal peptide fused to TatC in a manner that allows for it to undergo productive binding within the binding site, different constructs of TatC-signal peptide fusions were created.

Figure 5.4.1 shows a cartoon summary of the constructs created. The nine amino acid peptide was fused to either the NT or the CT, and then a subsequent construct created in which there was more of the peptide present (in the case of the NT fusion) or a sequence of alanine residues (in the case of the CT fusion) to act as a linker between the protein and the peptide. The longer constructs were created in case the fused peptide is too far away from the binding site (in the case of the CT fusions) or too constricted (in the case of the NT fusions) to be able to bend into the correct orientation for binding.

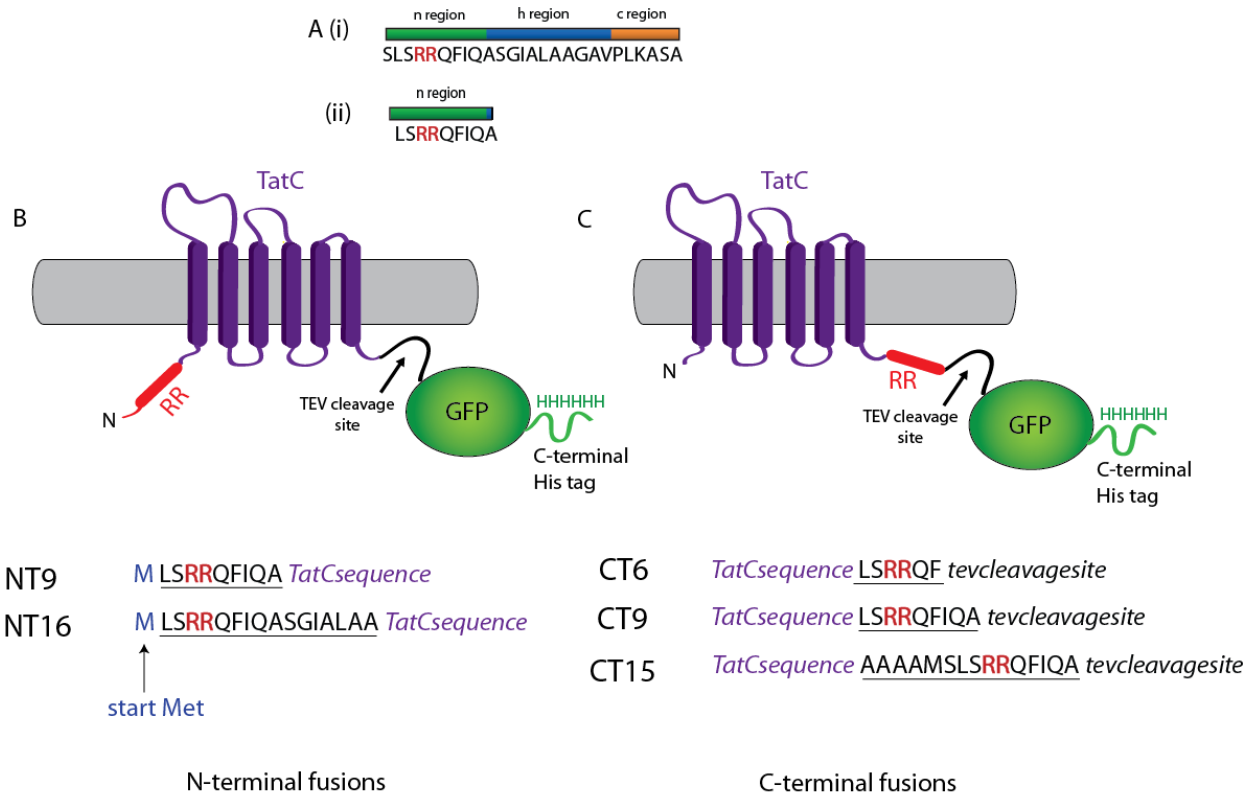


Figure 5.4.1 A summary of the TatC-signal peptide fusion constructs created in this study.

The sequence of the long, **(A i)** and the short, **(A ii)** versions of the SufI signal peptide. (**B**) fusions to the NT of *A. aeolicus* TatC-GFP in the pWaldo vector, of either 9 or 16 amino acids of the SufI peptide. (**C**) fusions to the CT of TatC-GFP, of either 6, 9 or 15 amino acids of the SufI peptide. The 15 amino acid construct included poly alanine as a linker.

The signal peptide sequence was fused immediately after the start methionine residue in the NT fusion constructs, and it was fused at the CT before the TEV cleavage site, in order for the peptide to be retained on the construct after the cleavage step. The fusions were created by the Quikchange mutagenesis method (Agilent) which allowed amino acids to be inserted into the TatC-GFP_{His} pWaldo construct.

The primers in **Table 5.1** were used to create the fusions by the stepwise addition of amino acids to the TatC-GFP construct. In the case of the NT fusion constructs, nine amino acids were added firstly, and then another seven amino acids of the peptide were added to this. In the case of the CT fusions, however, difficulties in the Quikchange mutagenesis reaction meant that initially only the first six amino acids were able to be added to the construct (CT6), followed by the additional three to create the nine amino acid construct (CT9). To the CT9 construct an additional six amino acids were added (the first two of the long SufI) and a string of alanines to act as the linker. Details of the procedure used are given in **Appendix 8.5**, and the full resulting TatC sequence given in **Appendix 8.3, Tables 8.3.4** and **8.3.5**.

Expression trials were carried out for the five TatC-GFP-signal peptide fusion constructs in L56 cells in order to test if these constructs can be expressed. The level of expression of the protein was assayed by the GFP fluorescence of the cell pellet from the culture, as described for the selenomethionine minimal media expression trials in **section 3.4**. The TB medium and expression cells (L56 (DE3)), were kept constant but the induction parameters of temperature (24^{o}C and 18^{o}C) and the concentration of IPTG (0.1mM or 0.4mM) were varied.

Construct	Forward primer (5' to 3')	Reverse Primer (5' to 3')	Template	Resulting sequence of the construct
NT9	T CAG GAA AAA CTC CCG GAA ATG CTC TCT AGG AGG CAA TTT ATA CAA GCA CCT CTA ACG GAA CAT CTG AGA G	C TCT CAG ATG TTC CGT TAG AGG TGC TTG TAT AAA TTG CCT CCT AGA GAG CAT TTC CGG GAG TTT TTC CTG A	WT TatC-GFP pWaldo	MLSRRQFIQA..(TatC in pWaldo)..
NT16	C TCT AGG AGG CAA TTT ATA CAA GCA TCT GGA ATA GCA CTC GCA GCA CCT CTA ACG GAA CAT CTG AGA GAG C	G CTC TCT CAG ATG TTC CGT TAG AGG TGC TGC GAG TGC TAT TCC AGA TGC TTG TAT AAA TTG CCT CCT AGA G	NT9	MLSRRQFIQASGIALAA..(TatC in pWaldo)…
CT6	GG AAG AAG AAG GAA ATT CAA AAG GCA CTC TCT AGG AGG CAA TTT CCT GGA TCC GAA AAC CTG TAC TTC CA	TG GAA GTA CAG GTT TTC GGA TCC AGG AAA TTG CCT CCT AGA GAG TGC CTT TTG AAT TTC CTT CTT CTT CC	WT TatC-GFP pWaldo	..(TatC in pWaldo).. LSRRQFPGSENLYFQG
CT9	AAG GCA CTC TCT AGG AGG CAA TTT ATA CAA GCA CCT GGA TCC GAA AAC CTG TAC TTC	GAA GTA CAG GTT TTC GGA TCC AGG TGC TTG TAT AAA TTG CCT CCT AGA GAG TGC CTT	CT6	..(TatC in pWaldo).. LSRRQFIQAPGSENLYFQG
CT15	GG AAG AAG AAG GAA ATT CAA AAG GCA GCT GCT GCT GCT ATG TCT CTC TCT AGG AGG CAA TTT ATA CAA GC	GC TTG TAT AAA TTG CCT CCT AGA GAG AGA CAT AGC AGC AGC AGC TGC CTT TTG AAT TTC CTT CTT CTT CC	CT9	..(TatC in pWaldo).. AAAAMSLSRRQFIQAPGSENLYFQG

Table 5.1 Primers used to create the signal peptide fusion constructs. The primers, template used and the sequence of the resulting fusion (sequence of the fused signal peptide, and any additional linker residues, highlighted in purple and the TEV site highlighted in blue). The rest of the TatC-GFP sequence is omitted for clarity.

The fluorescence readings from the expression trials (**Figure 5.4.2**) showed that none of the fusion constructs expressed as well as the WT. For each construct, the expression condition that gave the highest levels of GFP fluorescence was used to grow 6L of culture, and each protein purified in LMNG as described in **section 3.3**.

In order to test if these constructs were able to productively bind the signal peptide fused to them, the ITC assay was used, in which the titration was repeated with the short SufI peptide. It would be expected that if the fused peptide is able to bind in the binding site, then the titrated peptide will be able to do so less efficiently. This could manifest itself as a reduction in the K_d of the observed interaction for the fusion construct. However, any reduction in binding to the externally added signal peptide may also be due to steric effects

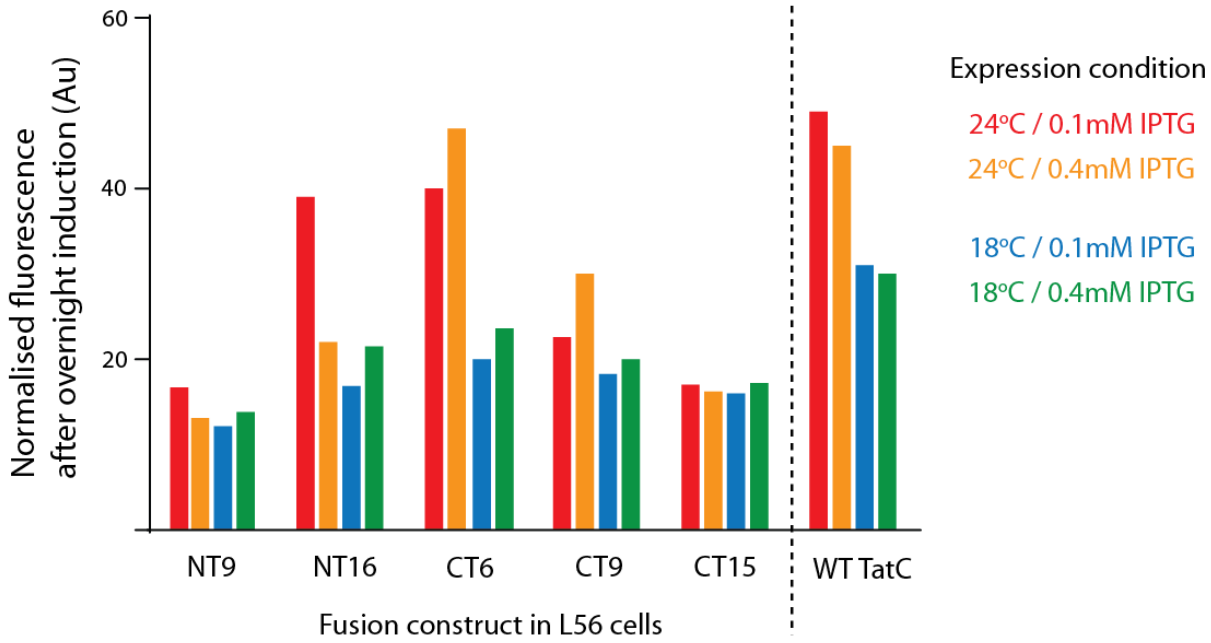


Figure 5.4.2 Expression trials of the TatC-GFP signal peptide fusion constructs.

Expression was assayed by the change in fluorescence after induction under four different conditions. All constructs were expressed in L56 (DE3) cells and TB medium. WT TatC expression levels are shown for comparison, where the standard expression condition for the WT as 24°C, 0.1mM IPTG (red bar).

from the extra amino acids at the cytoplasmic side of the protein, especially for those constructs with fusions to the NT. In order to try and ascertain if the reduction of the titrated peptide binding is just due to the presence of the extra amino acids, constructs were created in which the consensus RR motif was mutated to KK. Based on previous observations (Alami *et al.*, 2002, Stanley *et al.*, 2000) and on the ITC experiment in **Figure 4.3.1C**, the KK version of the peptide should not bind to TatC. ITC with the KK fusion construct would give a measure of the steric hindrance effect on titrated signal peptide binding that a string of amino acids of the corresponding size and shape of the RR peptide. If a reduction in the K_d for the RR peptide is seen beyond that for the KK, this could indicate that the RR peptide is able to specifically bind to TatC.

The primers in **Appendix 8.5, Table 8.5.3** were used to create KK substitutions from RR signal peptide fusion constructs, for all five of the fusions. The KK constructs were expressed under the conditions identified in **Figure 5.4.2** for the corresponding RR construct.

Purification of the N-terminal fusion constructs

The fusion constructs were purified in LMNG as described for the WT TatC protein in **section 3.3.** For the NT9 construct, the SEC profile for the RR and KK constructs is shown overlaid with the WT in **Figure 5.4.3 A(i)**. Firstly, the difference the area under the UV peaks showed the reduction in yield obtained for the NT9RR construct; around a third of the WT yield. The yield of the NT9KK construct was almost at WT levels. Secondly, the shape of the peak for the RR construct was very different from that of both the WT and the NT9KK. The peak was no longer the characteristic asymmetric shape seen for the WT on

the S200 16/60 column (**Figure 3.1.1**). Instead, it was much broader, and more symmetrical.

SDS-PAGE analysis of the NT9RR construct (**Figure 5.4.3 A(ii)**) showed the characteristic band for TatC at the position expected. The same band on SDS-PAGE was seen for the NT9KK construct. The purified protein samples were subject to native mass

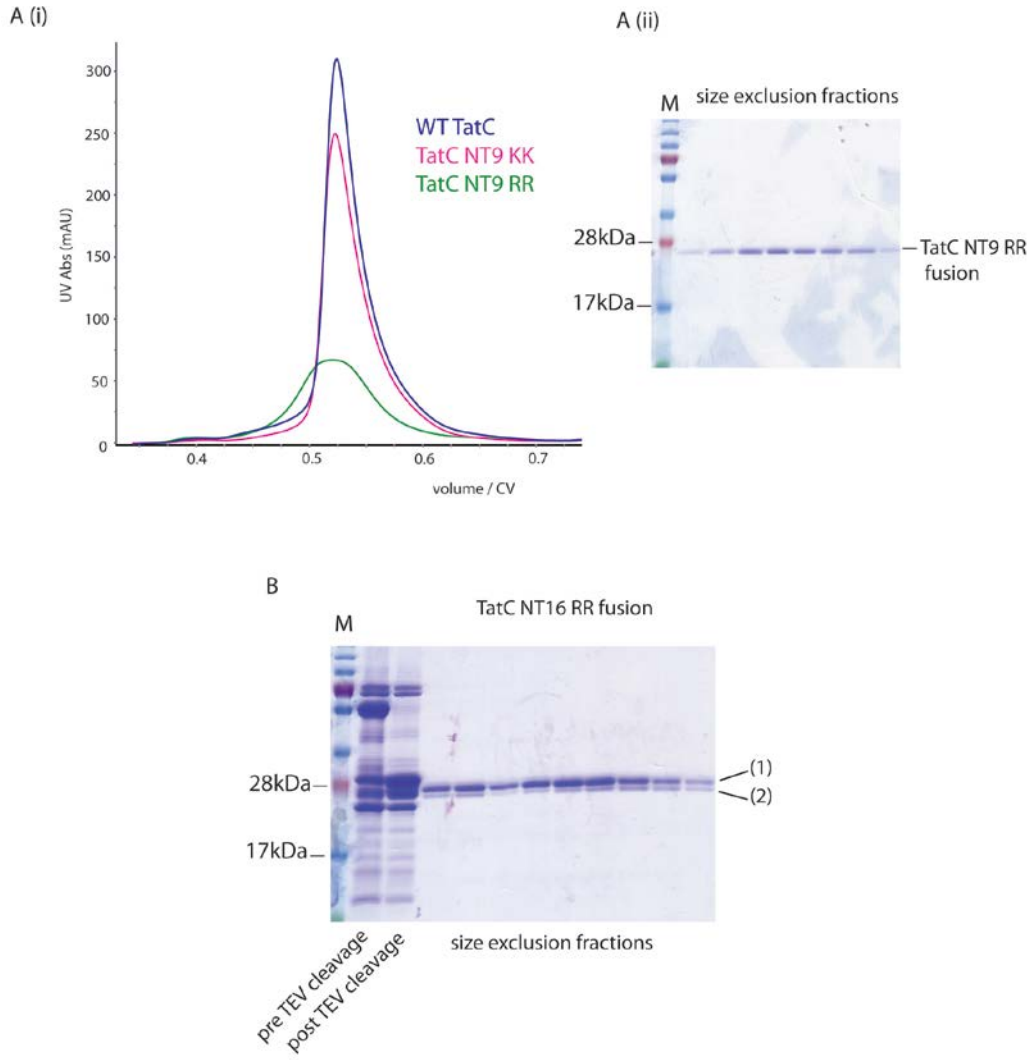


Figure 5.4.3. Purification in LMNG of the NT TatC signal peptide fusions.

(A) (i) S200 16/60 SEC traces for the NT9 RR (green), NT9KK (pink) and WT TatC (blue) overlaid. **(A) (ii)** SDS-PAGE of the SEC fractions for the NT9RR fusion. **(B)** SDS-PAGE of the SEC fractions for the NT16RR shows two bands for the TatC fusion protein. **M**:Markers, Pageruler Plus Prestained Protein Ladder (Fermentas), both SDS-PAGE at 15% (w/v).

spectroscopy (Dr Julien Marcoux, Dept of Chemistry, University of Oxford) which confirmed the mass difference expected for the fusion.

ITC was then carried out for these two constructs (as described in **section 4.3**) to test if binding of the titrated SufI peptide was inhibited to a greater extent in the RR construct than the KK construct (**Figure 5.4.4 A).** The calculated K_d for the interaction between the signal peptide and the NT9RR TatC fusion was around 70µM, a greater than 10-fold reduction compared to the K_d with the WT of 3µM. However, repeating the experiment with the NT9KK construct also gave a reduction in binding, (**Figure 5.4.4 B**), and gave a K_d comparable to that for NT9RR.

As both the NT9RR and NT9KK TatC fusion proteins showed a similar reduction in the binding affinity for the short peptide, it indicated inhibition in binding was not specific to

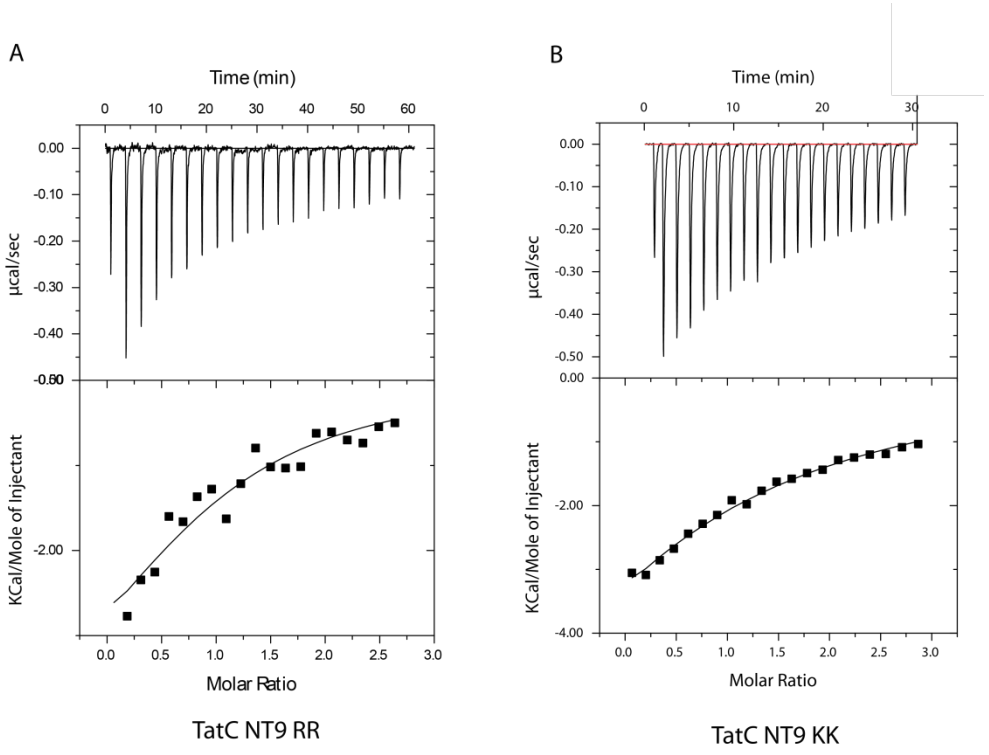


Figure 5.4.4 Isothermal titration calorimetry of the short signal peptide with the NT9 signal peptide fusion constructs of TatC.

Isotherms for the RR signal peptide fusion **(A)** and the KK signal peptide fusion **(B).**

the NT9RR peptide binding within the pocket on TatC. This implies the inhibition may be due to the steric effect resulting from the fusion of nine additional amino acids near to the TatC binding site. This ITC experiment did not give a clear answer as to whether the NT9 fusion peptide was actually able to bind in the binding site on TatC.

In order to investigate the difference in the shape of the SEC elution peak for the RR compared to the KK fusion (**Figure 5.4.3 A(i)**), multi angle laser light scattering (MALLS) was carried out (Dr Steven Johnson, Dunn School of Pathology).

The MALLS traces showed the NT9KK construct (**Figure 5.4.5 A**) had a mass that is slightly higher than WT, in line with the mass increase expected for the ~1kDa that has been fused onto the TatC protein. The calculated mass for this construct was extremely consistent all the way across the peak, which showed there was a single species present, exactly as was seen for WT TatC in LMNG (**Figure 3.1.2 A)**.

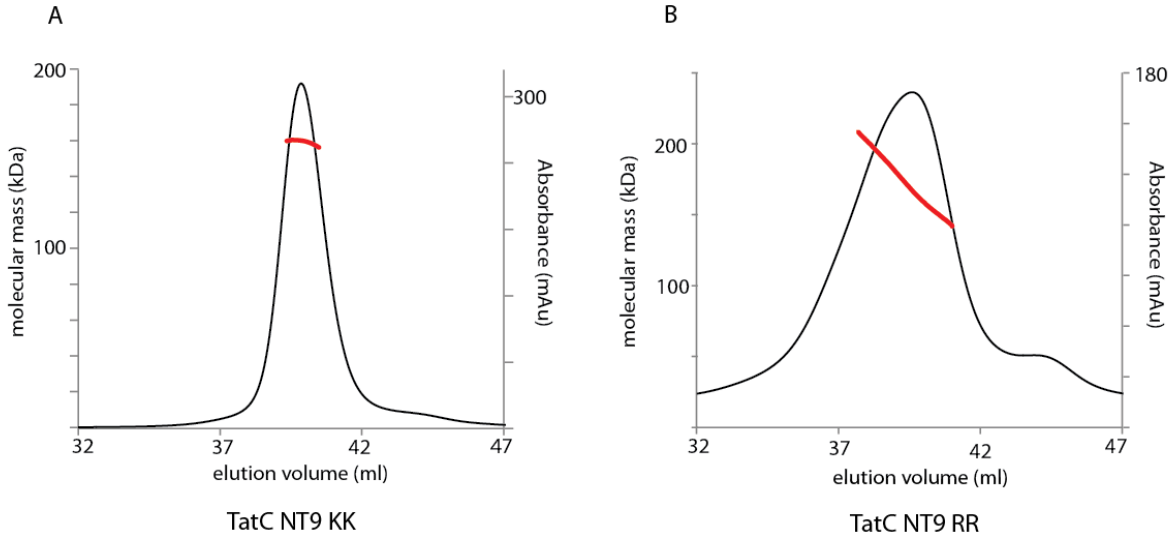


Figure 5.4.5 Multi angle laser light scattering (MALLS) of the TatC NT9 signal peptide fusion constructs.

The UV profile is given in black, and the calculated mass across each point of the peak is given in red for 70µM of the KK variant **(A)** and the RR variant **(B)** of the NT9 TatC fusion protein.

The MALLS trace for the NT9RR was very different (**Figure 5.4.5 B**). This showed that the mass (for the protein in LMNG) at the start of the peak was around 140kDa, slightly below that for the TatCNT9KK, but then the mass profile drastically rose across the peak, and at the leading edge the mass was over 200kDa. The profile was indicative of protein oligomerisation. The observed oligomerisation could be due to the signal peptide fused to one molecule of TatC being able to bind to another molecule of TatC which is nearby in solution (**Figure 5.1.1**). However, what was formed by the 'signal peptide crosslinks' was not a single, distinct species. The sample maybe a mixture of some cross linked TatC molecules and some bound to its own fused peptide. This heterogeneous mixture of species was not a good candidate to use in crystallography.

The other NT fusion construct, NT16, behaved similarly to the NT9 in the SEC elution and the ITC. Furthermore, SDS-PAGE of this longer construct, (**Figure 5.4.4 B**) showed that unlike the NT9 constructs, the fusion protein was cleaved. The SEC fractions contained two protein bands, band (1) was most likely the full length fusion protein, and band (2) could be TatC in which the fusion peptide had been clipped off, and so runs at a slightly smaller molecular weight. Cleavage was also seen for both the NT16RR and the NT16KK (gel not shown), so could be caused by the excess amino acids in the peptide having greater flexibility and so more accessible to cytoplasmic proteases which clip regions of disordered proteins during the purification. Additionally both the RR and the KK NT16 fusion proteins were not very stable in solution as both precipitated at 4^{o}C. For these reasons, these fusion constructs were not amenable for crystallography and so have not been subjected to crystallisation trials.

Purification of the C-Terminal fusion constructs

The CT6 and CT9 fusion constructs behaved in a similar fashion to the NT9 and NT16 fusions. The CT15RR construct under went such extensive oligomerisation that the protein aggregated and was not amenable to purification.

In summary, the experiments in this section indicated that the approach of fusing the signal peptide to the either termini of TatC will most likely not be successful in giving a co-crystal structure. ITC was not able to show the fusions were specifically able to bind within the binding site. MALLS suggested the TatC present within these samples was oligomerising. The oligomerisation may have been mediated by the fused signal peptide not always binding to its associated TatC molecule, and binding to a neighbouring TatC molecule and thus forming a heterogeneous sample. In the case of CT15RR, the oligomerisation caused the protein solubility to alter such that it was not able to be purified. The longer constructs, CT15KK, and both the NT16 RR and KK, also underwent clipping, giving a mixture of two differently sized TatC species seen by SDS-PAGE.

5.5 Summary

Chapter 4 identified a cytoplasmically exposed patch (i) on the structure of TatC that had been previously implicated in signal peptide binding, and ITC with variants of TatC in this region began to enable the important of the residues in this site to be assessed. However, in order gain molecular detail on the interaction, attempts were made in this chapter to obtain a co-crystal structure of TatC with the signal peptide. Various approaches proved unsuccessful, soaking crystals, co-crystallisation, finding a stronger binding peptide and fusion of the peptide to the termini of TatC. Attempts to gain this structure are still on-going.

Chapter 6: Interactions of TatC with the other membrane components of the Tat system

In addition to the TatC component, a fully functional Tat system in both Gram negative bacteria and in plant thylakoids is composed of two other integral membrane components, TatA (Sargent *et al.*, 1998) and TatB (Settles *et al.*, 1997). Affinity purifications of these components from *E. coli* membranes have shown that these three integral membrane components exist as two distinct complexes within the membrane (de Leeuw *et al.*, 2002), which may transiently associate as part of the transport mechanism (Mori & Cline, 2002).

TatC can stably bind, in a 1:1 stoichiometry to TatB (Bolhuis *et al.*, 2001) to form the TatBC receptor complex, in which TatC binds the signal peptide near the conserved twin arginine motif (Gérard & Cline, 2006) whereas TatB binds adjacent to the region of the signal peptide further downstream of this (Alami *et al.*, 2003). A small amount of TatA can be associated with this complex, even in the absence of signal peptide (Bolhuis *et al.*, 2001) (Fröbel *et al.*, 2011). Upon signal peptide binding, TatC is able to specifically recognise and insert this peptide into the membrane (Fröbel *et al.*, 2012a). Subsequently, TatC is able to provide a platform onto which many copies of TatA can be recruited and polymerise, to bring about protein transport across the membrane (Dabney-Smith *et al.*, 2006).

TatC therefore emerges as the central and organisational ‘core’ of the Tat system. Understanding the molecular details of the interactions with the other Tat membrane components is crucial to try and gain some understanding of how the coordinated mechanism of transport occurs.

Chapter 4 discussed the overall structure of TatC, and what can be learnt about the interaction of TatC with the signal peptide. Analysis of the structure showed two patches on the surface that were both important for function and highly conserved (**Figure 4.2.2**). The cytoplasmically exposed patch (i) was shown to be involved in the specific recognition of the signal peptide. It was speculated that the remaining patch (ii), on the groove between TM5 and TM6, therefore may be involved in one of the crucial interactions between TatC and either the TatA or TatB component. In order to investigate which regions of TatC may bind the other two Tat membrane components, information from the *E. coli* system on contacts between TatC and the TatA and TatB components has been obtained by various techniques. These technquies include photoaffinity crosslinking studies (Zoufaly *et al.*, 2012), cysteine crosslinking, genetic interactions (Kneuper *et al.*, 2012) and mutational studies which lead to destabilization of the TatBC complex (Buchanan *et al.*, 2002). These residues are listed in **Table 6.1** and plotted as spheres onto the structure of TatC (**Figure 6.1**).

The data from the interaction studies map to different regions of the TatC structure; this information can, in part, be used to try and understand how TatC interacts with each of the other components, which will now be discussed in turn.

6.1 Self interactions between TatC monomers

Within the *E. coli* TatBC complex, it is thought that the TatC and TatB components can exist as distinct sub-complexes (Orriss *et al.*, 2007). Within the TatC sub-complex, the TatC component it is thought to form a local dimer. TatC dimeric fusions that still retain function (Maldonado *et al.*, 2011a), and the observation that two distinct inactivating

***E. coli* residue**	***A. aeolicus* residue**	**Contact data in *E. coli***
Val3	Not in our construct	TatB contact by photoaffinity labelling
Ile10	Thr4 (disordered)	TatB contact by photoaffinity labelling
Ile14	Arg8	TatB contact by photoaffinity labelling
Asp63	Ser56	TatC contact by photoaffinity labelling
Asp150	Ser145	TatA contact by photoaffinity labelling TatB contact by photoaffinity labelling
Met205	Ala200	Cysteine crosslinks to TatB L9C Genetic interaction between TatC M205R variant and TatB F2L and/or F6L variants
Asp211	Asp205	Destabilises interactions with TatB TatA contact by photoaffinity labelling

Table 6.1 *E. coli* residues implicated in interactions with other Tat components. The corresponding *A. aeolicus* residue is given, and the colouring scheme used as in **Figure 5.1**. Photoaffinity labelling from (Zoufaly *et al.*, 2012), cysteine crosslinks and the Met 205 interactions from (Kneuper *et al.*, 2012), Asp211 destabilising effect from (Buchanan *et al.*, 2002).

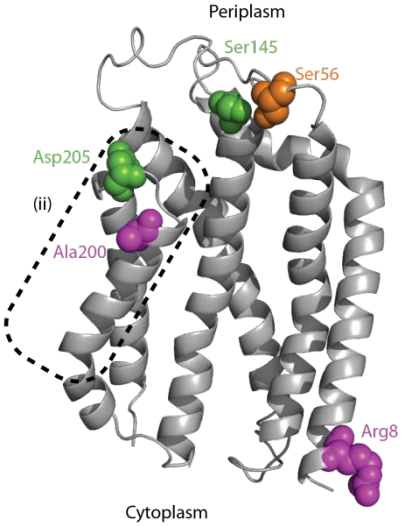


Figure 6.1 Sites of interaction with the other Tat components.

The positions in *E. coli* that have been reported to interact with both TatA and TatB (green), TatB (magenta) and TatC (orange), as given in **Table 6.1**, are mapped onto the structure of *A. aeolicus* TatC. The highlighted region corresponds to patch (ii) in **Figure 4.2.2**.

mutations in TatC are suppressed when expressed together (Buchanan *et al.*, 2002) give evidence for this.

Although the stoichiometry of the *A. aeolicus* TatC component in a native membrane is unknown, the MALLS analysis of *A. aeolicus* TatC purified in LMNG revealed this species was a dimer in this detergent solution (**Figure 3.1.2 A**). It was therefore speculated that this dimer may be of relevance, and that one of the packing interactions seen in the crystal may represent the interaction of the dimer seen in solution. Inspection of the crystal packing revealed only one potentially biologically plausible packing interaction, in which both copies have the correct topology (**Figure 6.1.1 A**) i.e. the N- and C-termini of both molecules are on the same side of the membrane (Behrendt *et al.*, 2004). Furthermore, it could easily be imagined how this dimer could position itself within a membrane, as the level of the membrane on the first monomer is level with where the membrane is predicted to be on the second copy. The PDB of the dimer was also subject to molecular dynamics simulations (Phillip Stansfeld, Department of Biochemistry, University of Oxford) in a DPPC bilayer, as described for the monomer (section 4.1). The dimer was able to equilibrate into the bilayer, and was comparably stable as the monomer simulation.

Inspection of the interface of this dimer shows it is formed between TM1 and TM2, and it buries a relatively large surface area; almost 760 $Å^2$. However, there is a lack of clear, specific interactions between the monomers in this interface. ePISA analysis (EMBL-EBI), which looks at the nature of the interactions between multimeric protein complexes, judged this interface as not biologically significant. Furthermore, conservation analysis of the dimer interaction surface by the program Consurf (Ashkenazy *et al.*, 2010) showed the residues on this surface are not conserved (Figure 6.1.1). The interface is coloured overwhelmingly in blue, meaning these residues are hyper-variable across many sequences. This argues that no specific interactions are maintained in this interface, which would be

expected if the interface were involved in forming the local TatC dimer across different bacterial species. In order to investigate the relevance of the crystallographic dimer further, evolutionary co-variance analysis of the TatC structure was performed by Phillip Stansfeld (Department of Biochemistry, University of Oxford). This analysis was performed on 2414 non-identical TatC sequences using the program PSICOV (Jones *et al.*, 2012), which allows for the prediction of sequence-based contacts identified within multiple sequence alignments of correlated mutations. This approach is therefore purely sequence based and does not require structures from more TatC sequences.

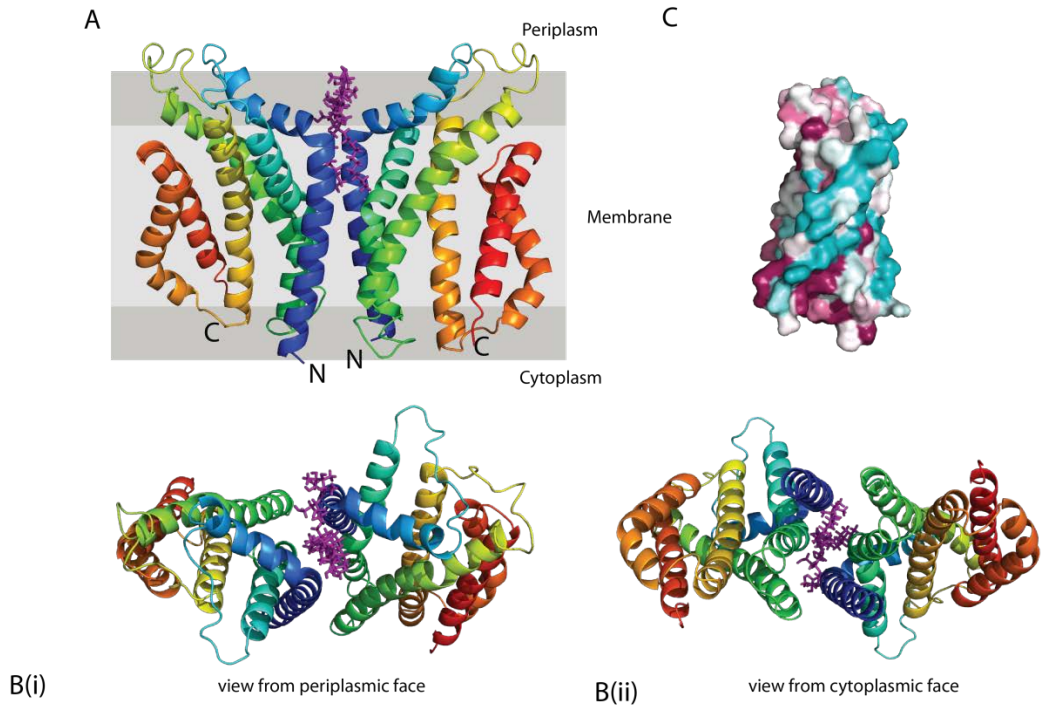


Figure 6.1.1 Crystallographic dimer of *A. aeolicus* TatC.

(A) and **(B)** TatC is shown in a cartoon representation coloured from blue at the NT to red at the CT. The semi-ordered detergent molecule is shown in purple sticks. The position of the bilayer is drawn based on molecular dynamics (MD) studies of TatC in a DPPC bilayer (Phillip Stansfeld, Department of Biochemistry, University of Oxford). **(C)** Consurf (Ashkenazy *et al.*, 2010) analysis of the sequence conservation of the surface of TatC plotted from maroon (highly conserved) to blue (highly variable) onto the dimer interface surface composed of TM helix 1 and 2.

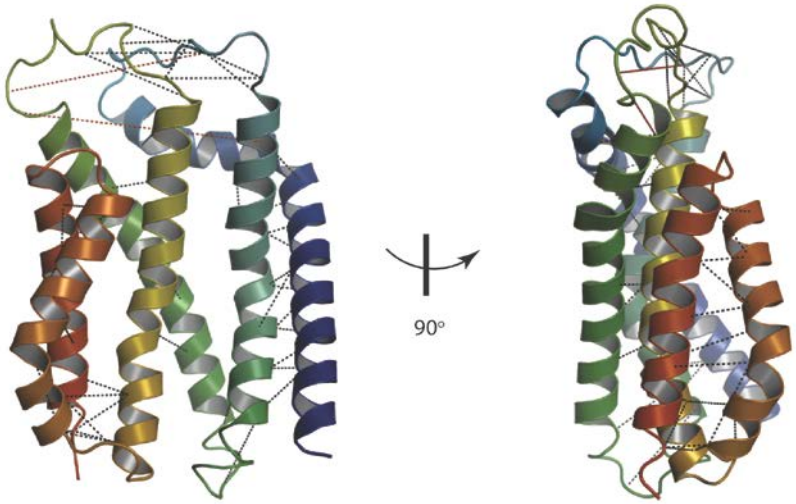


Figure 6.1.2 Evolutionary co-variance analysis of the *A. aeolicus* TatC structure.

Prediction by PSICOV (Jones *et al.*, 2012) with 2414 non-identical TatC sequences. Significant correlated positions within plausible physical contact (judged by a Cα-CαÅ) are plotted with black dashed lines. Well correlated positions at a similar depth in the membrane that would be plausible candidates for inter-subunit contacts are shown with red dashed lines. Analysis and figure produced by Dr. P Stansfeld (Department of Biochemistry, University of Oxford).

Correlated positions with a prediction confidence of greater than 0.6 and on the opposite side of the membrane were plotted onto the *A. aeolicus* structure (**Figure 6.1.2**). Those correlated positions that are within contact distance (deemed to be a Cα-Cα distance of less than 15Å) are shown with black lines. There are relatively few contacts predicted for the central 'core' region of TatC, but many for the maintenance of the TM1/TM2 and TM5/TM6 architecture. The PISCOV gives more evidence against the relevance of the crystallographic dimer, as all the TM1-TM2 correlations are better explained by intramolecular contact, rather than intermolecular contacts across the crystallographic dimer interface.

Interestingly, the ordered LMNG detergent molecule in the crystal (shown as purple sticks in Figure 6.1.1 A), sits between these two copies of TatC, and could help to stabilise this interface when TatC is purified in the LMNG detergent. The presence of this detergent molecule explain why this dimer forms in this detergent, rather than via specific, conserved

interactions between TatC monomers. Furthermore, the known TatC-TatC photoaffinity crosslink in *E. coli*, (**Figure 6.1**) localises to the periplasmic loops of TatC and is not consistent with this crystallographic dimer, providing strong evidence against the relevance of this dimer.

The other major packing interaction seen in this crystal involves TatC packing against an upside down copy of itself, which is an implausible biological dimer given the known topology in the membrane. The two subsequent *A. aeolicus* TatC structures (Ramasamy *et al.*, 2013) also gave no further indication on the nature of the TatC dimer, as the crystal packing in both these crystal forms did not give a biologically relevant dimer. Monomers were packed up side down against one another, or tilted with respect to one another such that they could not sit in a membrane. Currently, these TatC structures are therefore unable to provide evidence on how the local TatC dimer may form. Further crosslinking data may be required to enable an interface to be mapped onto the structure. However, if the contacts between TatC molecules are weak, or limited in number, further crosslinking experiments may prove difficult to obtain the data needed to understand the oligomer structure.

As an aside, it is interesting to note that the H32 TatC crystal contains two major inter-transmembrane crystal contacts (as well as a contact between the periplasm loops of three TatC molecules). **Chapter 2** contained a discussion on the rationale for why the ‘in detergent’ crystals tend to consist of crystal contacts between the hydrophilic regions of the membrane protein, given the large detergent belt around the center of the protein. This reasoning also explained why, on average, short chain detergents tend to give higher resolution crystals, as more of the loops are exposed to allow crystal contacts to form between loops. *A. aeolicus* TatC purified in LMNG, which with 12 carbon atoms is considered a ‘long’ chain detergent, and crystallised using the ‘in detergent’ method from major contacts between the trans-membrane regions of the protein, has therefore behaved

somewhat unexpectedly. The large detergent belt surrounding TatC did not hamper it packing against itself to form high reasonable high resolution crystals, and a single molecule of LMNG seen in the crystal may have helped to stabilise the packing interface.

6.2 Interactions with TatA

TatC can bind a small amount of TatA in resting membranes (Fröbel *et al.*, 2011), but will recruit large amounts upon the formation of a TatBC-substrate complex, whereby TatA can then homo-oligomerise and acts to bring about transport of the protein across the membrane (Dabney-Smith *et al.*, 2006). The interaction between TatC and TatA is therefore much more transient, and could involve many TatA monomers binding onto a single TatC molecule.

The presence of a photo crosslink between these two proteins on the periplasmic loop of TatC, (**Figure 6.1**) suggests the unstructured, periplasmically located NT tail of the TatA TM helix can reach this point. There is a lack of experimental data on how the TM TatA protein may interact with the membrane region of TatC; the photoaffinity contact data between TatC D211 and TatA has been questioned by recent, unpublished results from Tracy Palmers Laboratory, Division of Molecular Microbiology, University of Dundee (these results will be discussed in more detail in **section 6.3**). These recent results could suggest the photoaffinity crosslink to TatA in this region is an artefact. Alternatively, it could suggest that TatA binds adjacent to the TatB binding site, or TatA and TatB compete for the same binding site.

In the absence of more experimental data, a homology model of the TM region of *A. aeolicus* TatA based on the NMR structure of the *Bacillus subtilis* $TatA_d$ monomer (Hu *et*

al., 2010) was subjected to in silico docking and MD simulations (Dr Phil Stansfeld, Department of Biochemistry, University of Oxford). **Figure 6.2.1** shows the results of 3 such experiments, using (**A**) AutoDock Vina (Trott & Olson, 2010) , (**B**) ZDOCK (Pierce *et al.*, 2011) and (**C**) Course Grain MD simulations (Monticelli *et al.*, 2008).

All three of these experiments suggest that TatA binds to the concave face of TatC. This is of interest, as it would provide a role for the single, negatively charged residue within the membrane, Glu 165, on the concave face of TatC, shown as the red patch on the electrostatic surface representation in **Figure 4.3.2.** As discussed in **section 4.1**, the MD simulations suggest that the position of this residue in the bilayer is unfavorable, which implies that this residue has a possible interaction partner within the cell; this maybe provided by the TM region of TatA. It is interesting to note that the docking experiments carried out by (Ramasamy *et al.*, 2013) docked a signal peptide into this groove. This may

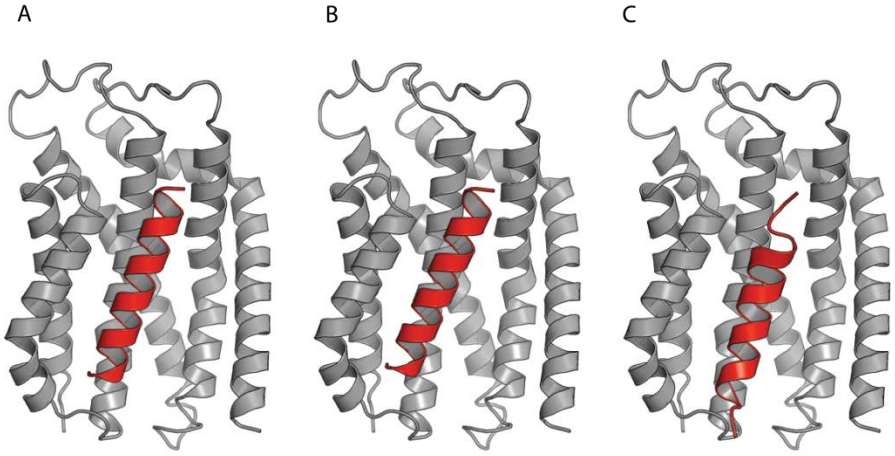


Figure 6.2.1 TatA docking to the *A. aeolicus* TatC crystal structure.

(A) homology model of the *A. aeolicus* TatA TM helix was produced based on the *B. subtilis* $TatA_d$ NMR structure (red) and docked onto the *A. aeolicus* TatC structure with 3 methods **(A).** Autodock Vina and **(B)** ZDOCK, showing the top ranked result and **(C)** Course Grain Molecular Dynamic (CGMD) simulation, in which this result is given in 2 of out 5 simulations. In the two other simulations it is seen to bind near TM5/6, the binding site proposed for TatB (see **scetion 5.3**). In one simulation TatA does not bind TatC. Experiments performed and figure produced by Dr Phillip Stansfeld (Department of Biochemistry, University of Oxford).

have been driven by the fact that they modeled the signal peptide as a helix, suggesting that this face of TatC is structurally well suited to binding helical regions of proteins.

Inspection of TatA sequences from bacteria shows the presence of a conserved, and functionally essential, polar glutamine residue on the TM helix (Greene *et al.*, 2007). This could provide a H-bonding interaction site for the Glu165 residue. In the plant Tat system, the charge on both of these residues are reversed, to a glutamine in the equivalent position on the TatC protein and a glutamate on the TM of the TatA equivalent protein (Dabney-Smith *et al.*, 2006), hinting that a relationship between the two residues could be of crucial importance. Furthermore, it has been shown in thylakoids that the TatA Glu 8 residue is necessary to recruit TatA to the TatBC complex, indicating this residue is required for the initial interaction (Dabney-Smith *et al.*, 2003). The relatively large and concave nature of this face of TatC could therefore 'nucleate' polymerisation of TatA in a circular manner, around a protein to be transported.

Subsequent to these experiments, recent, unpublished data (Cline *et al*, University of Florida) from the plant homologues of these proteins has now identified an inter-membrane cross link between a residue adjacent to the equivalent Glu 165 residue in the TatC, and the TatA TM helix, which would give further weight to the model proposed above. This would give the central, membrane exposed cavity within TatC two roles, for the passage of signal peptides across the membrane (**Figure 4.3.5**) and for TatA binding.

6.3 Interactions with TatB

As discussed previously, TatB binds to TatC in a 1:1 stoichiometric complex (Bolhuis *et al.*, 2001). The interaction with TatC is known to be primarily mediated by the functionally

essential transmembrane helix of TatB (De Leeuw *et al.*, 2001) as removal of the C-terminal cytoplasmically exposed regions still enables TatB to bind to TatC (Maldonado *et al.*, 2011b). The TM helix of TatB is predicted to be short, with an N-out, C-in topology (Koch *et al.*, 2012), similar to the short TM helix of the homologous TatA protein structurally characterised in a detergent micelle by solution state NMR (Hu *et al.*, 2010, Rodriguez *et al.*, 2013). Mapping the interactions observed between TatC and TatB onto the TatC structure, (**Figure 6.1)** shows they localise to the periplasmic end of the second functionally important and highly conserved patch of the structure, identified as patch (ii) in **Figure 6.1**. This lies on the 'groove' between TM5 and TM6, highlighted in dashes. The corresponding contact from TatC M205C residue map to the periplasmic end of its transmembrane helix of TatB (Kneuper *et al.*, 2012). This indicates patch (ii) on TatC maybe involved in packing against the TM region of TatB.

The evolutionary co-variance analysis (**Figure 6.1.2**) showed that there is a high degree of correlated positions between TM5 and TM6, indicating that the architecture of these helices and thus the surface exposed 'groove' between them, is well conserved. This further suggests these helices form an interaction surface for another membrane protein component. In order to investigate this further, limited cysteine disulphide scanning mutagenesis was carried out between the *E. coli* TatC and the *E. coli* TatB proteins in these areas by Franziska Jäger (Division of Molecular Microbiology, University of Dundee) (**Figure 6.3.1 B)**. These experiments identified a list of residues on the *E. coli* TatB and *E. coli* TatC involved in interaction with the L10 residue on *E. coli* TatB and the M205 residue on the *E. coli* TatC.

Some insight into a model for the TatC-TatB TM helix interaction was provided by the crystal packing of TatC. It was noted that the packing interactions in the *A. aeolicus* TatC crystal, previously discussed as dimer interfaces of TatC in **section 6.1**, either involved

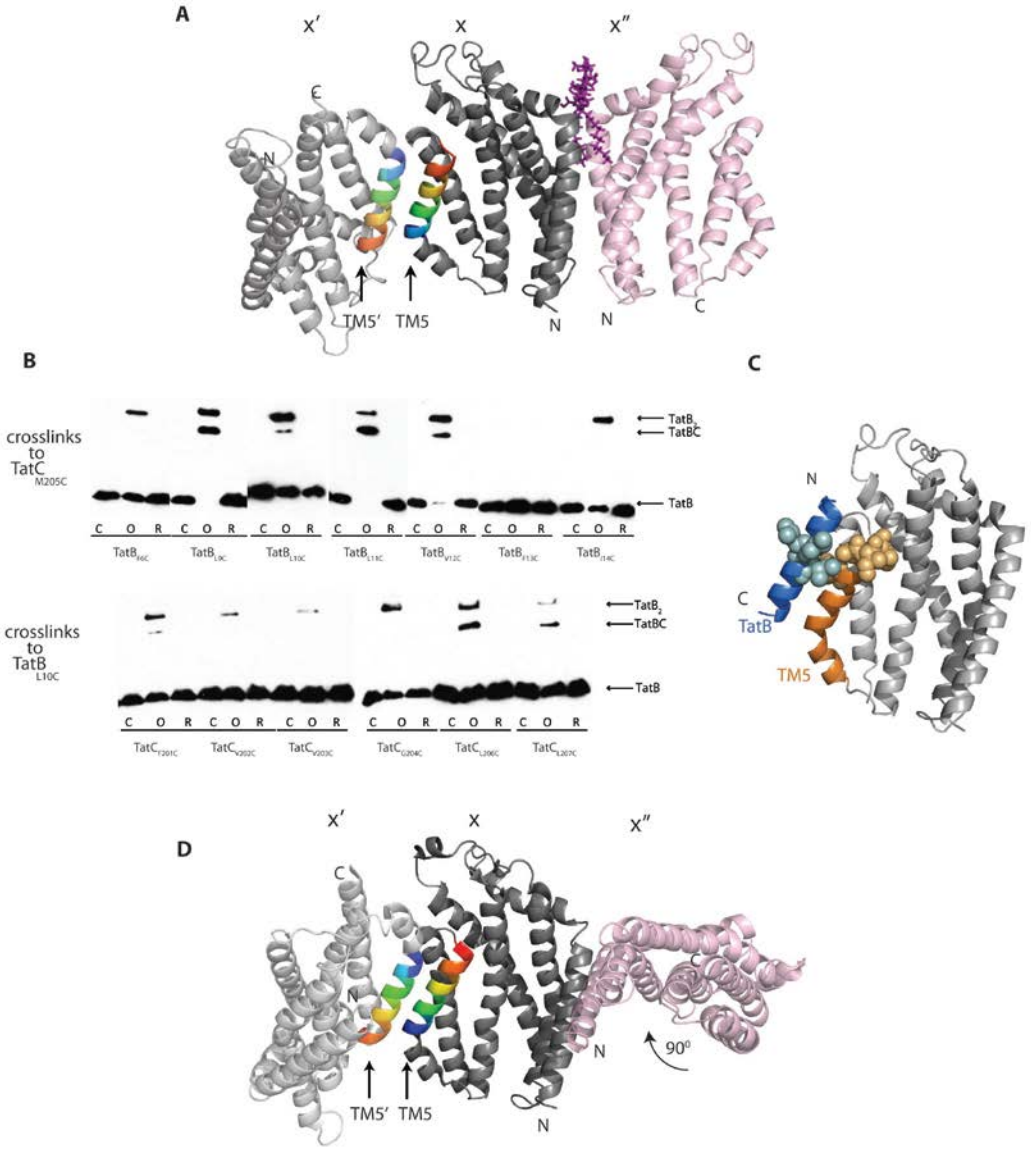


Figure 6.3.1 A model for the interaction site between TatC and TatB.

(A). The crystal contacts seen in the *A. aeolicus* TatC crystals grown in this study. The bilayer compatible dimer (shown in **Figure 5.1.1**) between molecule X and X",stabilised by the detergent molecule, LMNG (purple sticks), whereas the contact between X and X' involves a packing interaction between antiparallel TM5 helices, shown coloured from blue (NT) to red (CT). **(B)** Disulphide crosslinking between *E. coli* TatB and TatC variants detected by immunoblotting with TatB antibodies. Lanes are untreated (C), oxidized with Cu(II) phenanthroline (O) or oxidized then reduced with DTT (R). These experiments were carried out by Franziska Jäger (Division of Molecular Microbiology, University of Dundee). **(C)** The complex between TatC and the TM helix of TatB modeled on the *A. aeolicus* TatC X-X' crystal contact. Positions in *E. coli* TatC that disulphide bond to *E. coli* TatB L10C variant (yellow), or positions in *E. coli* TatB that form a disulphide bond to a *E. coli* TatC M205C variant (blue) are mapped onto the model. **(D)** The crystal contacts between X and X' are conserved in both crystal forms of the two subsequent *A. aeolicus* TatC crystal structures (Ramasamy *et al.*, 2013), shown here in the DHPC crystal form (PBD code 4HTS), whereas the interactions between X and X" are different, and rotated by 90^{o}.

molecule X (dark grey) and X" (pink) mediated via the bound LMNG molecule (**Figure 6.3.1A)**, or involved X (dark grey) packing against an upside down copy, X' (light grey). The latter interaction, mediated by the short section of TM5 packing against an upside down version of itself, is not plausible as a model for how the biological TatC dimers may form, but it does show how the proposed TatB contact site is able to pack against a short TM helix, of the correct topology to the TM helix of TatB (NT towards the periplasm, CT in the cytoplasm) (Koch *et al.*, 2012).

The X-X' packing interaction could therefore explain how the TatB helix is able to pack against TatC at this highly conserved and functionally important 'groove' between TM5 and TM6 of TatC (**Figure 4.2.2 B**). In combination with the cysteine crosslinks observed in this region (**Figure 6.3.1**) a model was proposed for the TatB helix packed against the TatC structure. This model is shown in **Figure 6.3.1 C**, in which the position of the blue TatB helix is taken from the position of the short region of the inverted TM5 helix of molecule X'. The residues shown as yellow spheres are the positions in the *E. coli* TatC that form crosslinks to the *E. coli* TatB L10C variant, and the residues shown as blue spheres are the positions in *E. coli* TatB that crosslink to the *E. coli* TatC M205C variant (from **Figure 6.3.1 B**).

The model proposed for TatB packing onto TatC is also compatible with recent nickel affinity purifications carried out in Tracy Palmers Laboratory (Division of Molecular Microbiology, University of Dundee). An alanine mutant of the *E. coli* TatC at position D211 (*A. aeolicus* D205) failed to pull down *E. coli* TatB from $C_{12}E_9$ solubilised membranes (unpublished). This experiment further suggests that the packing of TatB may occur at this highly conserved site on TatC.

Interestingly the packing interaction between TatC molecules X and X' mediated by the TM5 helices, is conserved in both of the two subsequent crystal forms of *A. aeolicus* TatC solved (Ramasamy *et al.,* 2013). An example of the *A. aeolicus* TatC in DPPC crystal packing is shown in **Figure 6.3.1 D**, where the X and X' molecules are seen to pack together by their TM5 helices in the same way as in **Figure 6.3.1 A**. The conservation of this crystal contact between all three available crystal forms gives further evidence that this interaction is highly favorable, and thus may represent a biologically meaningful way to pack helices at this point on a TatC monomer.

In reference to the discussion on how TatC may interact with other TatC monomers to form a local dimer (**section 6.1**), it is interesting to note that in contrast to this conserved crystal contact, very different contacts were observed in all three crystal structures between molecules X (dark grey) and molecule X" (pink), as seen in **Figure 6.3.1 A** and **D**. In the LMNG TatC crystal, TatC was seen to pack against itself with an interface formed between helices 1 and 2, in which both molecules were parallel to one another, thus giving the potential to form a membrane compatible dimer, as discussed in **section 5.1**. However, in the other two crystal forms, this was not the case. In the DPPC crystal form, molecule X" was seen to be in a similar location to X" in the LMNG crystal, except it was now rotated by 90^{o}, so it no longer formed a bilayer compatible dimer. The rotation at this interface is more exaggerated in the *A. aeolicus* DDM crystal form (not shown), in which molecule X" had rotated a full 180^{o}, so that it was inverted and thus packed upside down against molecule X at this interface.

The lack of conservation in the crystal packing at the interface created by TM1 and TM2 also provides further evidence against the TatC dimer seen in **Figure 6.1.1**, being biologically meaningful. The fact that the TM1-TM2 interface can pack against itself at such different angles argues that this interface is not important for forming a stable and

biologicaly relevant dimer, as it can form very different interactions with itself in a crystal, which are not maintained between detergents and the crystal forms. The interactions between these faces may therefore be very weak.

A model for the position of the rest of the TatB protein?

The model proposed in **Figure 6.3.1 C** provides an explanation for how just the TM part of TatB may interact with TatC. However, additional contacts have been identified between TatB and TatC that map to the cytoplasmic NT of TatC (**Figure 6.1),** near the proposed signal peptide binding site, patch (i), on TatC. Evidence also exists for an interaction between TatB and signal peptides. The association of KK peptides with Tat translocase, that fail to cross link to TatC (Panahandeh *et al.*, 2008) has been attributed to the interaction between the signal peptide and TatB. Crosslinks between the two form, but only in the presence of TatC (Alami *et al.*, 2003). These crosslinks formed between TatB and the folded domain of the precursor (Maurer *et al.*, 2010) or across the entire region of the signal peptide, including the h-region downstream of the RR motif on the signal peptide (Alami *et al.*, 2003). These experiements indicate that the interaction occurs sequentially, so that TatC binds firstly to the RR region, then TatB binds further downstream of this (Gérard & Cline, 2006) and to the precursor.

Consistent with the crosslinks observed interactions between TatB and the signal peptide, and the photo-affinity contact observed between TatB and TatC on the cytoplasmic NT of TatC (Zoufaly *et al.*, 2012), is an arrangement in which the amphipathic helix of TatB stretches across the cytoplasmic face of TatC, from the TM5/TM6 end. This would then allow it to reach the signal peptide binding site at the TM1 end of TatC, and contact the signal peptide (**Figure 6.3.2**).

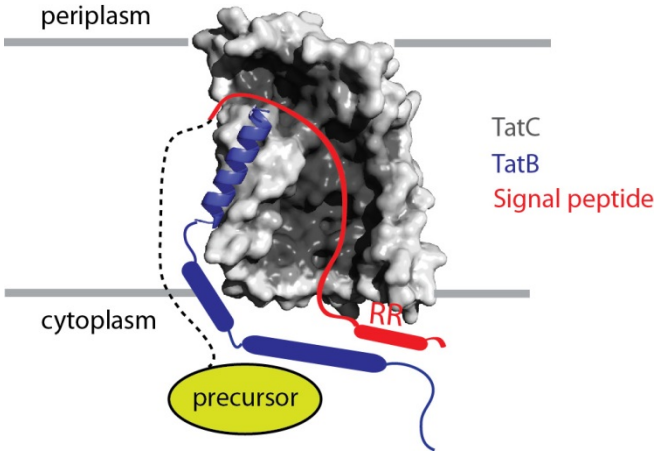


Figure 6.3.2 A model for the amphipathic helix of TatB interaction with TatC and the signal peptide.

TatC is shown as a surface, shaded in grey to show the depth of the structure. The signal peptide is shown in red, and its position modeled based on data discussed in **Figure 4.3.5**. TatB is shown in blue, with the TM helix position taken from the model in **Figure 6.3.1 C**. The signal peptide, if bound in this way, could contact the NT end of the TatB TM helix, and explain the suppressor mutations at this site. The position of the amphipathic helix is modeled so it can reach the signal peptide binding site at the NT of TatC. This model would then explain the cross links to TatC in this region, but also to the mature domain of the precursor.

Positions on TatB that have been shown to photo crosslink to the bound precursor are located at the periplasmic NT, on the TM helix. (Maurer *et al.*, 2010). Consistent with this, the sites of suppressor mutations in TatB, which can allow for the transport of signal peptides with sequence alterations that made them previously incompetent for transport, localise to the periplasmic end of the TatB TM helix (Kreutzenbeck *et al.*, 2007). Furthermore, the NT end of the TatB TM helix is the site of synergistic suppressor mutations of sequence variant defective signal peptides when used in combination with mutations on patch (i) of TatC (Lausberg *et al.*, 2012). This gives evidence for a tight cooperation between TatB and TatC during the binding of signal peptides.

The periplasmic location of the suppressor mutations could suggest the TM helix of TatB is involved in mediating long range conformational effects on binding, which are somehow transmitted to the cytoplasmic location of the signal peptide binding site on patch (i). Or, given the mounting evidence discussed in **4.3.5**, that the signal peptide undergoes an advanced stage of binding (Gérard & Cline, 2007) (Fröbel *et al.*, 2012a), the suppressor mutations are consistent with the h-region of the signal peptide being translocated across the membrane into a deeply membrane embedded binding pocket. This idea builds on the model proposed in **section 4.3.5** for how the signal peptide maybe inserted into the membrane by TatC, and subsequently transverse the membrane. If the NT of the TatB TM helix sits in this groove on TatC, it would place it by the hypothesized 'exit site' on TatC for the signal peptide h-region (S. Musser *et al*, unpublished).

Cleavage of signal peptides in the absence of TatB, and thus in the absence of transport, is still able to occur (Fröbel *et al.*, 2012a). The presence of TatB acts to inhibit the premature signal peptide cleavage by the periplasmically located signal peptidase. Thus, speculating even further on the model shown in **Figure 6.3.2**, the NT end of the TM of TatB may act to help regulate the position of the signal peptide and stop it from travelling further though TatC and becoming exposed to the periplasm. The top of the helix may act to 'secure' the h-region of the signal peptide into the binding pocket between TatB and TatC, and stop it drifting further through the membrane. This could explain why the TM of TatB, and TM5/TM6 of TatC are short, it might be a necessity that the helices do not fully span the side of TatC, in order for the signal peptide to be kept from fully passing through the membrane until transport has been achieved.

The model in **Figure 6.3.2** would allow for the rest of the signal peptide to be bound in a deep pocket mediated by both TatB and TatC (Fröbel *et al.*, 2012a), and for TatB to make its observed contacts to the h-region, as well as to the other regions of the signal peptide

(Alami *et al.*, 2003). It could be easily imagined how the amphipathic helix running across the face of TatC would then also allow for TatB to make the observed extensive contacts to the mature domain of the precursor protein (Maurer *et al.*, 2010), and thus the model would explain much of the crosslinking and genetic data from *E. coli* published to date on the TatBC-signal peptide interaction, and the tight cooperation between these two proteins. However this proposed model (**Figure 6.3.2**) gives no indication on how the TatBC complexes interact with other TatBC complexes during the formation (at least in *E. coli* and thylakoids) of the multivalent receptor complex, in which it is though there may be six to eight copies of each protein (Tarry *et al.*, 2009b, Celedon & Cline, 2012).

Currently, there is conflicting evidence on the exact structural arrangement of the TatBC complex. Overexpression and purification of the *E. coli* TatB and TatC components indicated that TatC may form the structural core, onto which TatB is able to assemble (Orriss *et al.*, 2007). However disulphide cross linking results proposed an alternative model, in which TatB may form a homo-oligomeric ring, potentially a tetramer, mediated by the TM and amphipathic helices of TatB, with TatC arranged at the periphery of a TatB complex (Lee *et al.*, 2006a). High resolution structural information on the TatBC complex is needed to try and resolve these issues.

Chapter 7: Towards structural characterisation of the TatBC complex

Previous data available on the sites of interaction between TatC and the two other membrane components, TatA and TatB were interpreted in **Chapter 6**. These contact data, along with additional experiments from Tracy Palmer's Laboratory (University of Dundee) and Phillip Stansfeld, (Department of Biochemistry, University of Oxford) enabled the potential site of interaction with the TatA TM helix to be localised to the concave face of TatC. The interaction with the TatB TM helix was localised to the 'groove' between TM5 and TM6, in the second of the two functionally important and highly conserved patches on the TatC structure identified in **Chapter 4**.

In order to gain a more detailed understanding on the interaction of TatC with TatA and TatB, further structural information on TatC in complex with these two proteins is required. Despite the evidence that a single molecule of TatA may be permanently associated with the TatBC complex in resting membranes (Fröbel *et al.*, 2011), the ability of TatA to homo-oligomerise to varying degrees when extracted with detergent (Gohlke *et al.*, 2005) would make a TatC-TatA complex a challenging high resolution structural target. However, TatBC is thought to form a permanent, tight complex in the membrane and so has more potential to be extracted from the membrane by detergents in a homogenous manner, and so more amenable to crystallographic studies. High resolution structural information would allow the model proposed in **Figures 6.3.1** and **6.3.2,** for the interaction of TatB with TatC to be tested, and may provide further information on how the multivalent TatBC receptor complex is built up from its constituent components.

As discussed in the introduction, previous attempts to understand the structure of the TatBC receptor complex involved cryo-electron microscopy (EM) of purified *E. coli*

TatBC complexes. The *E. coli* TatBC complexes used in the EM study were purified in digitionin and ran as a large complex on Blue-Native PAGE (BN-PAGE), approximately 440kDa in size (Tarry *et al.,* 2009b). This mass is consistent with evidence in *E. coli* that the TatBC complex is multimeric (Bolhuis *et al.,* 2001) and could contain up to eight copies of each component. The cryo-EM study revealed a large sphere, with an appropriate width of a bilayer (Tarry *et al.,* 2009b). The low resolution of these reconstructions gave no information on the interaction between TatC and TatB, or on how the components interact to build up the larger oligomeric complex

Given the extremely low resolution of the 3D reconstructions, attempts to dock the *A. aeolicus* crystal structure into this complex proved difficult, and did not give a consistent or sensible answer. The size and shape of the *A. aeolicus* TatC protein seems inconsistent with the object observed in the EM. This is reminiscent of the situation, discussed in the introduction, of the issues surrounding the Sec translocon channel and its oligomerisation state. The high resolution crystal structures contain a single copy of the SecYEG channel, but this is in conflict with some biochemical evidence which implies the channel acts as a dimer in a membrane. In order to now try and understand the stoichiometry of the purified *E. coli* TatBC complex, and asses its potential for use in higher resolution structural studies, the complex was purified and subject to MALLS analysis.

7.1 Biophysical assessment the *E. coli* TatBC complex

The *E. coli* TatBC complex can be recombinantly expressed and purified from *E. coli* DADE (*ΔtatABCDΔtatE*) cells. Dr. M Krenhenbrink and Dr. F Rodriguez (Department of Biochemistry, University of Oxford) purified the *E. coli* TatBC complex, using a His tag at the CT of TatC, as described in (Orriss *et al.,* 2007), but in the LMNG detergent. The

complex eluted from the size exclusion column as a single peak (not shown) and the SDS-PAGE analysis (**Figure 7.1.1 A**) showed the presence of both *E. coli* TatC and *E. coli* TatB. The TatB protein is seen to run at around 40kDa, which is much higher than expected given its molecular weight of 18.4kDa. The difference in the position of the TatB protein from that expected for its size may be to due to a number of factors. As discussed previously, the extent of the migration of a protein in SDS-PAGE is due to its size and the amount of the SDS detergent that it binds. For most proteins, the amount bound is linearly proportional to its size when denatured. However, it has been shown that many membrane

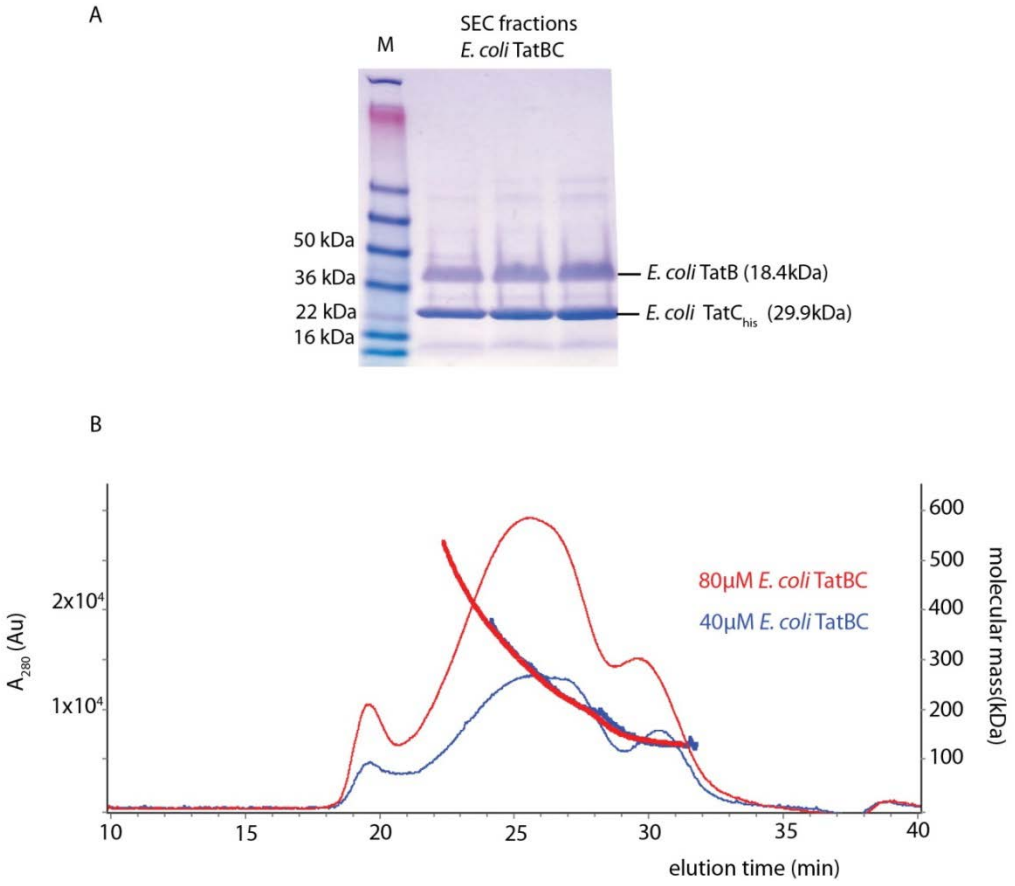


Figure 7.1.1 Biophysical characterisation of the *E. coli* TatBC complex.

(A) SDS-PAGE shows the SEC fractions pooled and concentrated for further studies, run on a Bio-rad TGX precast gel, 4-20%. The *E. coli* TatC band is shown to run as expected, but the *E. coli* TatB band runs much higher than its molecular weight (18.4kDa). **(B)** MALLS analysis for two concentrations of the *E. coli* TatBC complex, 80µM (red) and 40µM (blue). The UV trace is given as the thin line, and the corresponding calculated molecular mass across the peak is given as the thick line. This profile shows the extremely heterogeneous nature of the purified *E. coli* TatBC complex.

proteins, due to the additional hydrophobic stretches of sequence they contain, may not bind the same amount of SDS molecules as a soluble protein of the same size would bind (Rath *et al.*, 2009). The *E.coli* TatBC samples are not boiled prior to running the gel, and so the proteins may not be fully denatured. Both these factors would result in the protein potentially not binding an amount of SDS that is linearly proportionally to its mass. Futhermore, not denaturing the protein may enable any higher order oligomers present in the sample to remain intact when run on the gel. These factors combined can cause the band to run at a position which is not consistent with the monomer mass of TatB. The large apparent mass may therefore represent an oligomer, or a monomer, both with unusual SDS binding properties. On the basis of the SDS-PAGE alone, it is not possible to tell these possibilities apart. The TatB band also stained very differently to the TatC band. The gel was stained using coomassie (R-250) dye, which works by forming strong, non-covalent forces with the protein in the gel, by a combination of van der Waals and electrostatic interactions. The differential staining seen in **Figure 7.1.1 (A)** could potentially be due to the very high proportion of neagtively charged, acidic residues in the *E. coli* TatB sequence, residues which are known to interact less efficiently with the negatively charged coomassie dye. These may cause TatB to bind fewer dye molecules overall than would be expected.

In order to measure the stoichiometry of the purified *E. coli* $TatBC_{his}$ complex sample purified in LMNG, two concentrations of this sample, 40µM and 80µM, were subject to MALLS analysis, by Dr. Steven Johnson (Sir William Dunn School of Pathology) as described in **2.3**. The result is shown in **Figure 7.1.1 B**, where the UV trace for each concentration (red, 80µM and blue, 40µM) is given as the thin line, and the calculated molecular mass at each point across the peak is given as the thick line across the corresponding peak.

The UV trace showed there is more than a single peak. The difference from the preparative column maybe due to the greater resolving power of the S200 10/300 (GE healthcare) analytical column, enabling the sample to be dissected into its constituent pieces more efficiently. This indicated a mixture of species were present, which was further evident upon the calculation of the molecular mass for these samples. The calculated mass profile, shown in **Figure 7.1.1 B** for each sample is the total protein:detergent complex mass. The mass profiles showed the lack of homogeneity within the samples. The mass was just over 100kDa at the lower end of the peak, and rose to over 500kDa at the upper end of the peak. This gave a massive change in mass across the peak. Furthermore, the mass did not plateau at the upper end of the peak, it continued to rise rapidly, which was indicative of excessive oligomerisation or aggregation.

Protein conjugate analysis of these sample, which allowed the mass from the protein component alone to be calculated, gave a similar mass profile for the protein alone as the profile of the protein:detergent complex. The change in mass across the UV peak was therefore due to different species of protein present in the sample. As the calculated mass of the protein continued to vary all across the peak implying the complex did not form a stable, single distinct species, it proved impossible to try and rationalise what species of TatBC might have been present.

Upon running a SDS-PAGE of the fractions collected from the MALLS analytical column (not shown), the fractions towards the end of the peak seemed to contain just *E. coli* TatC. As the His-tag is on the CT of TatC, and TatB is being pulled down with TatC, there could be an excess of the TatC component. If the complex is dynamic and some TatB could dissociate during the purification, the affinity step may pull out non-stoichiometric amounts of TatBC. This may add to the observed heterogeneity of the mixture of complexes, as well as the proteins tendency to oligomerise.

Clearly this sample is not suitable for high resolution structural studies. A sample of TatBC would need to be found that was single, and homogenous. As the TatC protein from *A. aeolicus* was able to be purified in such a way, it was speculated that the TatBC complex from this organism may be able to give a homogenous sample for crystallography.

7.2 Initial characterisation of the *A. aeolicus* TatBC complex

In order to ascertain whether the *A. aeolicus* TatB and *A. aeolicus* TatC proteins might be suitable for structural studies, an initial experiment was firstly carried out to ascertain if *A. aeolicus* TatB and TatC can be co-purified, and to give an idea as to the nature of the SEC profile on an analytical column. This was achieved by co-expressing and co-purifying TatB and TatC-GFP from the only construct available at this point, which was created from the genomic operon of the *tatB* and *tatC* genes.

A. aeolicus has an unusual arrangement of the *tatB* and *tatC* genes on its chromosome. The reading frame of the *tatC* gene overlaps with the reading frame of the *tatB* gene, such that there is an +1 frameshifted internal ribosome binding site within the *tatB* gene, 93 residues into the 111 residue TatB sequence, so the TatC protein is out of frame with TatB (**Figure 7.2.1 A**).

This genomic piece of DNA had been previously cloned (Dr Sai-Man Lui, Department of Biochemistry, University of Oxford) into the XhoI and BamHI sites of the pWaldo vector, to give a GFP_{His} fusion protein separated by a TEV cleavage site, as for the *A. aeolicus* TatC construct used in this study (**Figure 7.2.1 B**). This therefore has the T7 promoter upstream of the TatB sequence, but no additional promoter or ribosome binding site for the

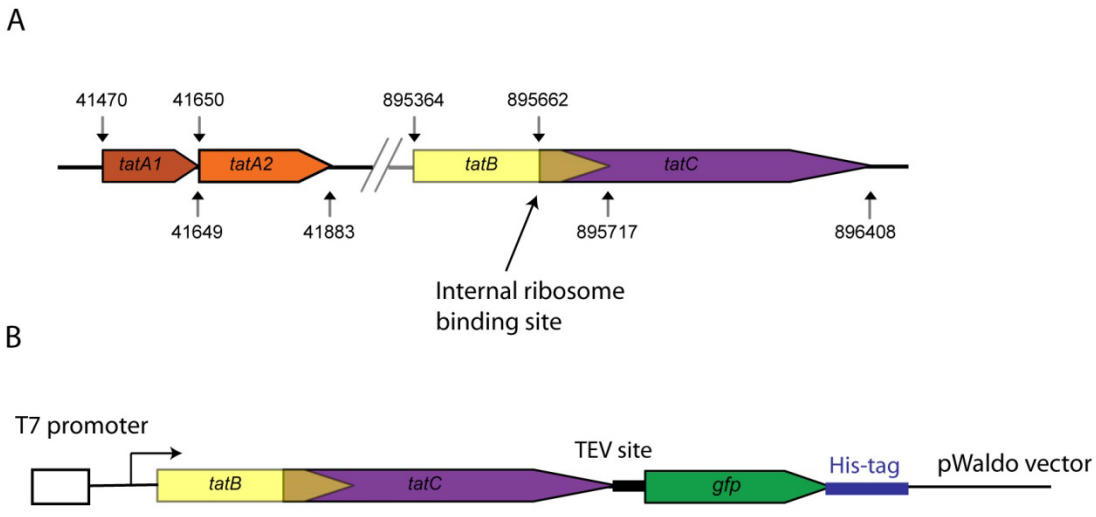


Figure 7.2.1 The *A. aeolicus tat* operon.

(A). The *tatC* gene overlaps with the *tatb* gene with a +1 frameshift such that there is an internal ribosome binding site for *tatC* initiation. The numbering refers to co-ordinates used by TIGR (www.tigr.com). Figure adapted from Dr Sai-Man Lui, Department of Biochemistry, University of Oxford. **(B).** The genomic *A. aeolicus* TatBC construct in pWaldo, with a GFP_{His} fusion to the CT of TatC separated by a TEV cleavage site.

TatC sequence, other than that encoded within in the *tatB* gene. This arrangement may not be particularly efficient at initiating the translation of TatC when recombinantly expressed in *E. coli*, so the yield from this construct would therefore expected to be low, but should suffice for this initial scouting step.

The expression and purification of this construct was attempted in LMNG, as described in **section 3.1**, except the final size exclusion step was run by injection of 200µl of concentrated protein from the reverse Nickel affinity step onto an S200 10/300 analytical column (GE healthcare) pre-equilibrated in LMNG SEC buffer. The UV elution trace is shown in **Figure 7.2.2 A**. This showed that the peak was single, and symmetrical. SDS-PAGE analysis of the SEC fractions, and the final, concentrated sample **Figure 7.2.2 B**, showed a band at the characteristic position for *A. aeolicus* TatC (below the 28kDa

marker). There was also the presence of a second band, running above the 15kDa marker. As *A. aeolicus* TatB is predicted to be 12.7 kDa, this band could have corresponded to *A. aeolicus* TatB, as the *E. coli* TatB ran higher than expected for its mass on an SDS-PAGE gel (**Figure 7.1.1 A**). Furthermore, the band did not stain very well, again similarly to that seen for the *E. coli* TatB protein.

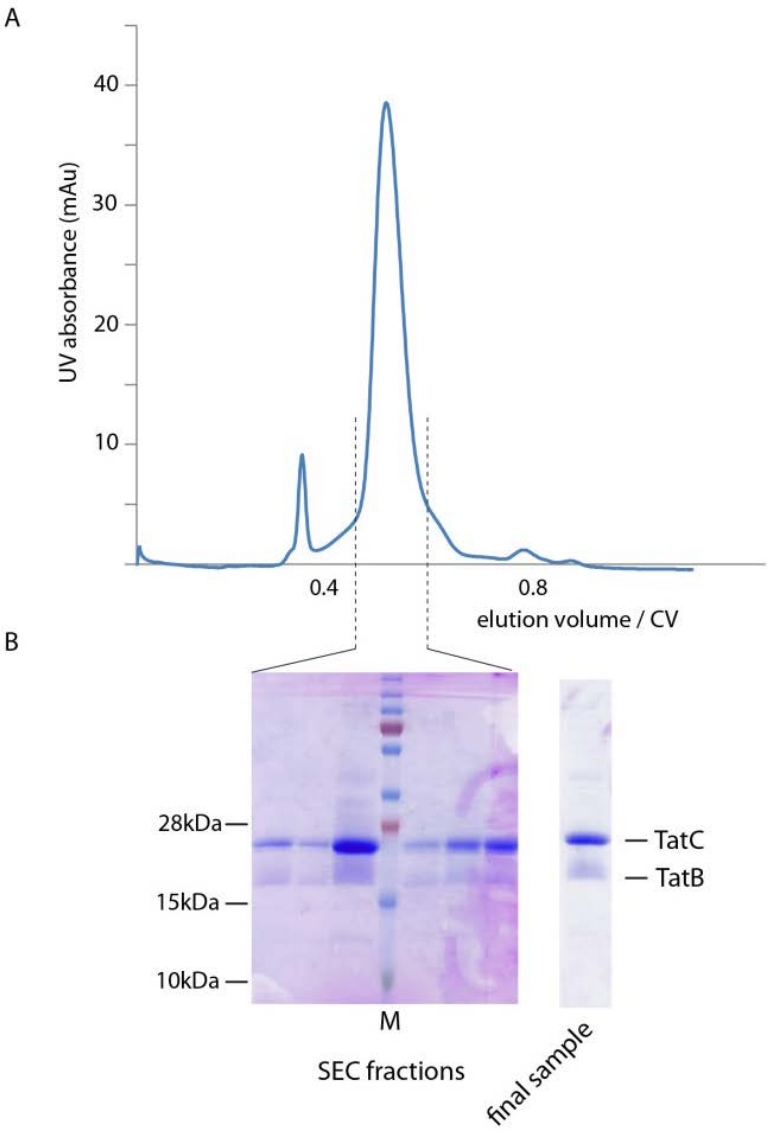


Figure 7.2.2. Purification of the genomic *A. aeolicus* TatBC construct in pWaldo.

(A) 200µl of protein from the reverse Nickel affinity step was injected onto a S200 10/300 SEC column (GE healthcare). **(B)** Fractions from across the peak were analysed by 15% SDS-PAGE and the final, concentrated sample shown in the last lane.

A lack of available antibodies against the *A. aeolicus* TatB protein meant that western blotting could not be carried out to confirm the identity of the band observed. Therefore, the band was cut out of the gel, trypsin digested and the resulting fragments subject to mass spectrometry, which allowed then to be identified by proteomics analysis (Proteomics Facility, Sir William Dunn School of Pathology, University of Oxford). The top hit from this database was for the keratin protein, this was dismissed as an obvious laboratory contaminant. The next hit was the sequence of the *A. aeolicus* TatB protein. Some other contaminating *E. coli* proteins were also present, which were carried over from the purification.

The yield purified from 6L of culture of the genomic construct was low, around 0.2mg, and so the MALLS analysis could not be run on this sample. However, this experiment did confirm that *A. aeolicus* TatB and TatC are amenable to over-expression and co-purification, and the species produced looks reasonably homogenous, as assessed by the nature of the SEC peak. It was decided to therefore attempt to structurally characterise the TatB and TatC proteins from *A. aeolicus*.

7.3 Molecular biology to create constructs of *A. aeolicus* TatB suitable for co-expression with *A. aeolicus* TatC-GFP$_{his}$ for structural studies

The construct of the genomic *A. aeolicus* TatBC-GFP$_{his}$ used above was unable to give ayield suitable for structural studies. In order to improve this, the proteins would need to be cloned into separate multiple cloning sites (MCS), each with their own T7 promoter and ribosome binding site, or into separate expression vectors suitable for recombinant co-expression in *E. coli* in order to increase the level of their expression.

The *A. aeolicus* TatC-GFP fusion construct in pWaldo (**Figure 7.3.1 A)** expressed highly and gave pure, functional protein. It was therefore decided to leave this construct in this vector and design constructs of *A. aeolicus* TatB suitable to be co-expressed alongside the pWaldo TatC construct. The pWaldo vector has only one multiple cloning site (**Figure 1.5.2**), so the *A. aeolicus* TatB protein could not be cloned into a second multiple cloning site on this vector. This meant TatB would have to be cloned into a compatible vector. The vectors selected were the pACYCDuet-1(Novagen) and the CDFDuet-1(Novagen), both with a different antibiotic resistance and origin of replication compared to pWaldo, **Table 7.1.** These two vectors are different in their copy number, which allows for a vector with a lower copy number (pACYCDuet-1) and a comparable copy number (CDFDuet-1) to pWaldo to be tested.

There are no available structures of the TatB protein from any organism, although it is predicted to be similar to the homolgous TatA protein, but with a substantial insertion in the amphipathic helix, before the CT (Lee *et al.*, 2002).To gain insight into the boundaries of the expected secondary structure of the *A. aeolicus* TatB protein, the sequence of the *A. aeolicus* TatB protein was subjected to multiple sequences alignments to other annotated *tatB* sequences, and secondary structure analysis servers (JPred (University of Dundee), TMHMM (CBS, Denmark), TMpred (EMBnet-CH)). This analysis, depicted in **Figure**

	pWaldo	**CDFDuet-1**	**pACYCDuet-1**
Antibiotic resistance	Kanamycin	Spectinomycin	Chloramphenicol
Origin of replication	f1	CloDF13-derived	P15A
Copy number	40	20-40	~10

Table 7.1 Characteristics of the pWaldo vector, and the two compatible vectors chosen for use for TatB co-expression, CDFDuet-1 and pACYCDuet-1.

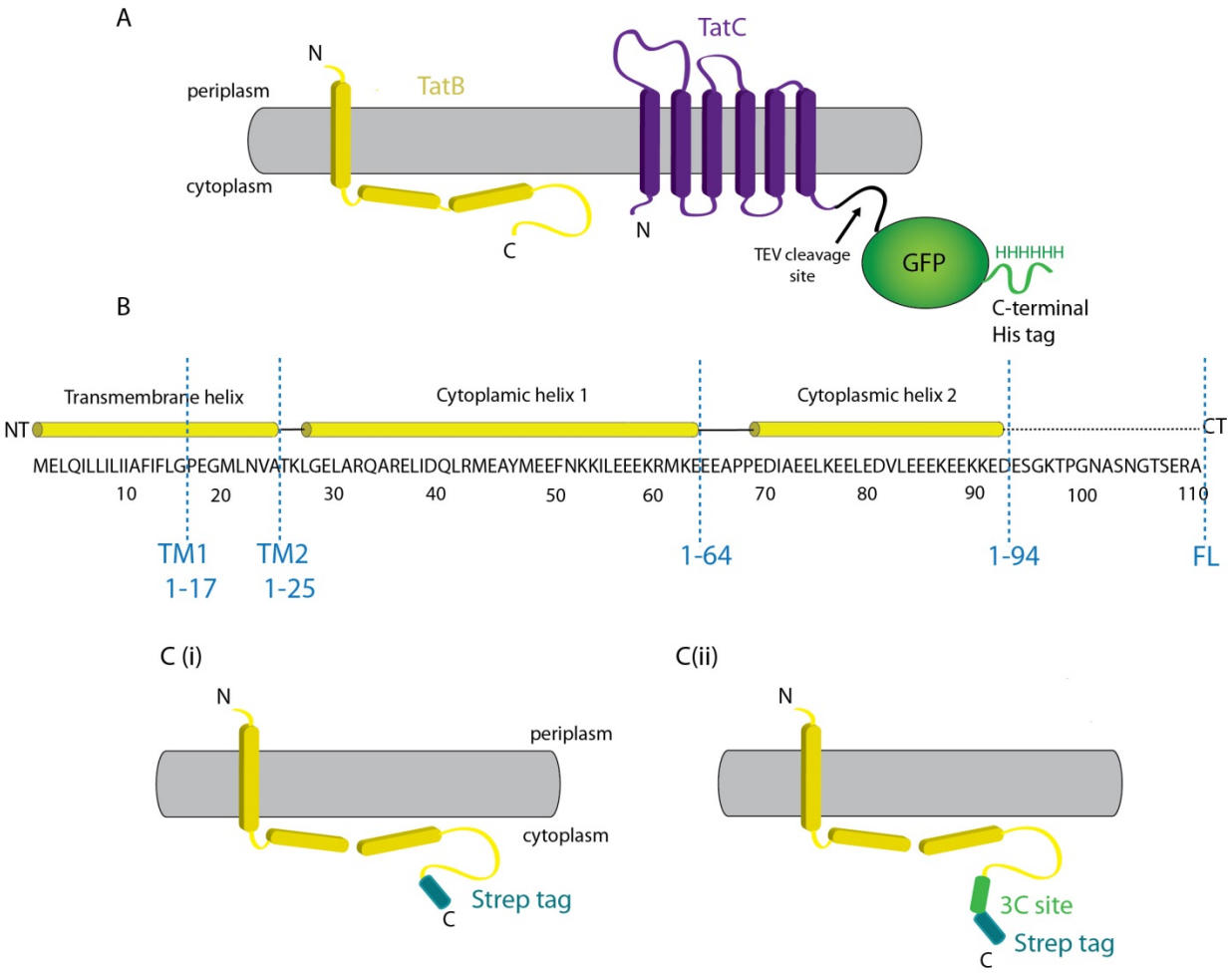


Figure 7.3.1 Design of *A. aeolicus* TatB constructs for co-expression with *A. aeolicus* TatC-GFP$_{his}$ in pWaldo.

(A) A schematic view of predicted secondary structure of TatB, with helices shown as tubes and flexible/loop regions as thin lines, and TatC in pWaldo shown as the NT fusion to GFP$_{his}$, separated by a TEV cleavage site. **(B)** The sequence of *A. aeolicus* TatB, with the predicted secondary structural elements shown above, where α helices are shown as tubes, loops as black lines and the region of disorder at the CT shown as a dashed line. The five different length constructs are shown marked onto the sequence with blue dashed lines. Versions of the constructs were created with either a CT Streptavidin tag, **(C i)** or a 3C cleavage site prior to the CT Streptavidin tag, **(C ii)** to allow for a dual purification system using the His tag on TatC and a Streptavidin tag on TatB.

7.3.1 B, showed that as for the *E. coli* TatB, *A. aeolicus* TatB is also predicted to have a single TM helix, and an amphipathic helix, potentially split into two by a short loop region, and an unstructured region at the extreme CT.

The consensus of the prediction servers was used to design constructs of *A. aeolicus* TatB for co-expression with *A. aeolicus* TatC, with two main aims in mind. The TatB protein in complex with TatC wants to be both structurally investigated, but also the interaction of the TatBC complex with the signal peptide also would be assessed by ITC, to see if TatB has any affect on this interaction, such as increasing the affinity for the peptide. These two aims potentially require different constructs of TatB; functional (ITC) binding assays would ideally require the full-length (FL) protein, whereas having less of the unstructured regions might assist in structural studies. All the constructs would need to contain the TM helix of TatB, as this has been shown to be involved in maintaining the interaction with TatC (De Leeuw *et al.*, 2001).

For these reasons, firstly a full length (FL) construct was designed. Secondly, a construct missing the predicted unstructured C-terminal 17 amino acids, which are also thought to be non-essential for *E. coli* TatB function (Lee *et al.*, 2002) containing residues 1-94 was designed. A third construct encompassing the TM helix and just the first cytoplasmic helix was designed (residues 1-64) in order to have less of the cytoplasmic helix that may be connected to the rest of the protein via a flexible loop; this may help in crystallisation. Finally, two constructs were designed to contain just the predicted TM helix of TatB. A short (residues 1-17) construct and a longer version (residues 1-25) of this helix were designed, due to the different predictions for the exact length of this helix. The five lengths of the *A. aeolicus* TatB fragments designed are shown marked onto **Figure 7.3.1 B** with blue dashes.

The polymerase chain reactions (PCR) was used to generate four different length DNA fragments from the *A. aeolicus tatB* gene (**Figure 7.3.2**). The primers designed to produce each of the five different length *A. aeolicus* TatB constructs in the MCS1 of both the CDFDuet-1 and pACYCDuet-1 vectors, along with full details of the molecular biology

carried out, is given in **Appendix 8.5.2.** The correct insertion of each fragment into both vectors was confirmed by sequencing with the ACYCDuetUP1 primer (Source Bioscience).

The molecular biology produced untagged versions of the TatB constructs in these vectors. However, in order to avoid purifying a mixture of TatC and TatBC complexes, due to the presence of the affinity tag on TatC-GFP_{his}, which may have added to the heterogeneity seen in the *E. coli* TatBC sample, it was decided to create tagged constructs of *A. aeolicus* TatB, with an additional affinity tag to the His tag, to allow for a dual affinity purification method. This will try to ensure that, despite potential differences in the levels of expression of these two proteins, from different copy number plasmids, that a 1:1 complex of the two will be purified.

The tag chosen for this role was the Streptavidin tag (referred to from now on as the Strep-tag), as this tag can be easily used alongside the his tag; the eluate from one column can be loaded straight onto the next column. This tag is also small, as just an eight amino acid

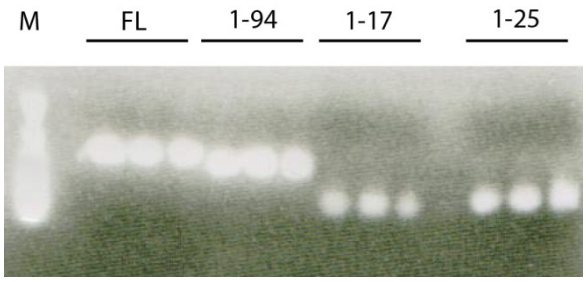


Figure 7.3.2. Agarose gel electrophoresis of the PCR products for four different length *A. aeolicus* TatB fragments.

For each PCR, three repeats are shown. The 1-64 construct is not shown, and details of how this construct was created is given in **Appendix 8.5.2**, along with the full details of the PCR conditions and subsequent molecular biology to create the final constructs. Markers, 100bp ladder, Promega.

sequence is required to bind to *Strep*-tactin resin, meaning that the tag should not interfere with the protein function. Given the topology and the predicted secondary structure of the TatB protein, it was decided to have the Strep-tag at the C-terminus, shown in **Figure 7.3.1 C(i)**. Commercial vectors are available that have Strep-Tag sequences already encoded in them, however none are available which are suitable for co-expression with pWaldo and have the Strep-tag encoded at the correct terminus of the construct. So, the 8 amino acid sequence of the Strep-tag (WSHPQFEK) was added onto the constructs in the pACYCDuet-1 and CDFDuet-1 vectors, using the primers in **Appendix 8.5.2 Table 5** and the Quikchange (Agilent) protocol.

It was also decided that the option of removing this tag maybe useful during the purification protocol, so a cleavage site was engineered into the constructs, before the Strep-Tag (shown in **Figure 7.3.1 (ii)**. As a TEV site protease site was already used in the TatC-GFP_{his} construct to cleave this His-tag was used, a 3C cleavage site was engineered. A different cleavage site was used, as this would allow removal of the Strep tag at a later stage in the purification, after running the reverse nickel and Strep affinity column. As with the TEV cleavage site, the eight amino acid 3C cleavage site is also short with the sequence LEVLFQGP. This sequence was added to the Strep-tag constructs by another round of Quikchange insertion (Agilent) using the primers in **Appendix 8.5.2 Table 3.** The confirmation of the correct insertion of the Tag and cleavage site into each construct was achieved by sequencing with the ACYCDuet-1UP primer (Source Bioscience).

The molecular biology described above resulted in 20 constructs of *A. aeolicus* TatB compatible for co-expression and dual affinity tag purification with *A. aeolicus* TatC-GFP_{his} in pWaldo; five different lengths of TatB each in two vectors (pACYCDuet-1 and CDFDuet-1) and each either Strep-tagged or 3Csite-Strep-tagged. The constructs are summerised in **Table 7.2**

	Length of *A. aeolicus* TatB fragment (residue numbers)	**Tag/cleavage site**	**Vector**	**Expected Molecular weight of TatB protein (kDa)**
1	FL	Strep-tag	CDFDuet-1	13.7
2			pACYCDuet-1	
3		3C-Strep-tag	CDFDuet-1	14.7
4			pACYCDuet-1	
5	1-94	Strep-tag	CDFDuet-1	12.1
6			pACYCDuet-1	
7		3C-Strep-tag	CDFDuet-1	13.0
8			pACYCDuet-1	
9	1-64	Strep-tag	CDFDuet-1	8.6
10			pACYCDuet-1	
11		3C-Strep-tag	CDFDuet-1	9.5
12			pACYCDuet-1	
13	1-17 (TM short)	Strep-tag	CDFDuet-1	3.0
14			pACYCDuet-1	
15		3C-Strep-tag	CDFDuet-1	3.9
16			pACYCDuet-1	
17	1-25 (TM long)	Strep-tag	CDFDuet-1	3.8
18			pACYCDuet-1	
19		3C-Strep-tag	CDFDuet-1	4.7
20			pACYCDuet-1	

Table 7.2 Summary of the Strep-tag and 3C-Strep tag constructs created. Each length of *A. aeolicus* TatB fragment designed (FL, 1-94, 1-64, 1-17 and 1-25) was cloned into two vectors (CDFDuet-1 and pACYCDuet-1), with either a Strep-tag, or with a 3C site prior to the Strep-tag (3C-Strep). The molecular weight of each construct is given in kDa.

Ideally, at least one construct from each of the five different lengths of TatB (FL, 1-94, 1-17, 1-25 and 1-64) either Strep-tagged and with or without a 3C cleavage site, from either vector, would be identified which gave expression, and amenable to co-purification in an

homogenous manner, in order to give protein of each length, to carry out ITC and crystallography.

Screening was carried out to identify which would be the best construct to use for each of the five lengths of TatB fragments desired. An overview of the logic behind the screening process is given in **Figure 7.3.3.** The boxes in blue depict the first steps involved in finding which construct (from those in **Table 7.2**) and expression condition can produce TatB protein for each of the five lengths of TatB fragments. These steps will be discussed in section **5.4.4.** Having identified one such construct for each length of TatB fragment, detergent screening (green box in **Figure 7.3.3**) aims to identify the best detergents to use in the purification of these constructs with *A. aeolicus* TatC-GFP. This step will be discussed in **section 7.5.**

7.4 Co-expression and co-purification screening of the Strep-tag and 3C-Strep-Tag *A. aeolicus* TatB constructs with *A. aeolicus* TatC-GFP$_{his}$

Co-expression trials were carried out in a High-Throughput (HT) manner. 50µl aliquots of L56 (DE3) cells were transformed (**Appendix 8.4 Table 1)** with both the *A. aeolicus* TatC-GFP$_{his}$ construct, and each of the 20 constructs of TatB. Overnight cultures were set up in 1ml of LB and 50µg/ml kanamycin, supplemented with 34µg/ml chloramphenicol (pACYCDuet-1) or 50µ/ml spectinomycin (CDFDuet-1) at 37^{o}C. 100µl of this culture was used to inoculate 5ml of TB, supplemented with 0.4% w/v sterile glycerol and the appropriate antibiotics at the concentrations given above. Cultures were grown in 10ml 24 well deep-well HT plates (Promega) sealed with gas permeable seals (Starlab) at 37^{o}C, 180 rpm until they reached an OD_{600} of 4.0-5.0. They were then induced with either 0.1mM IPTG or 0.4mM IPTG (final concentration), (**Table 7.3**) and left overnight at 24^{o}C.

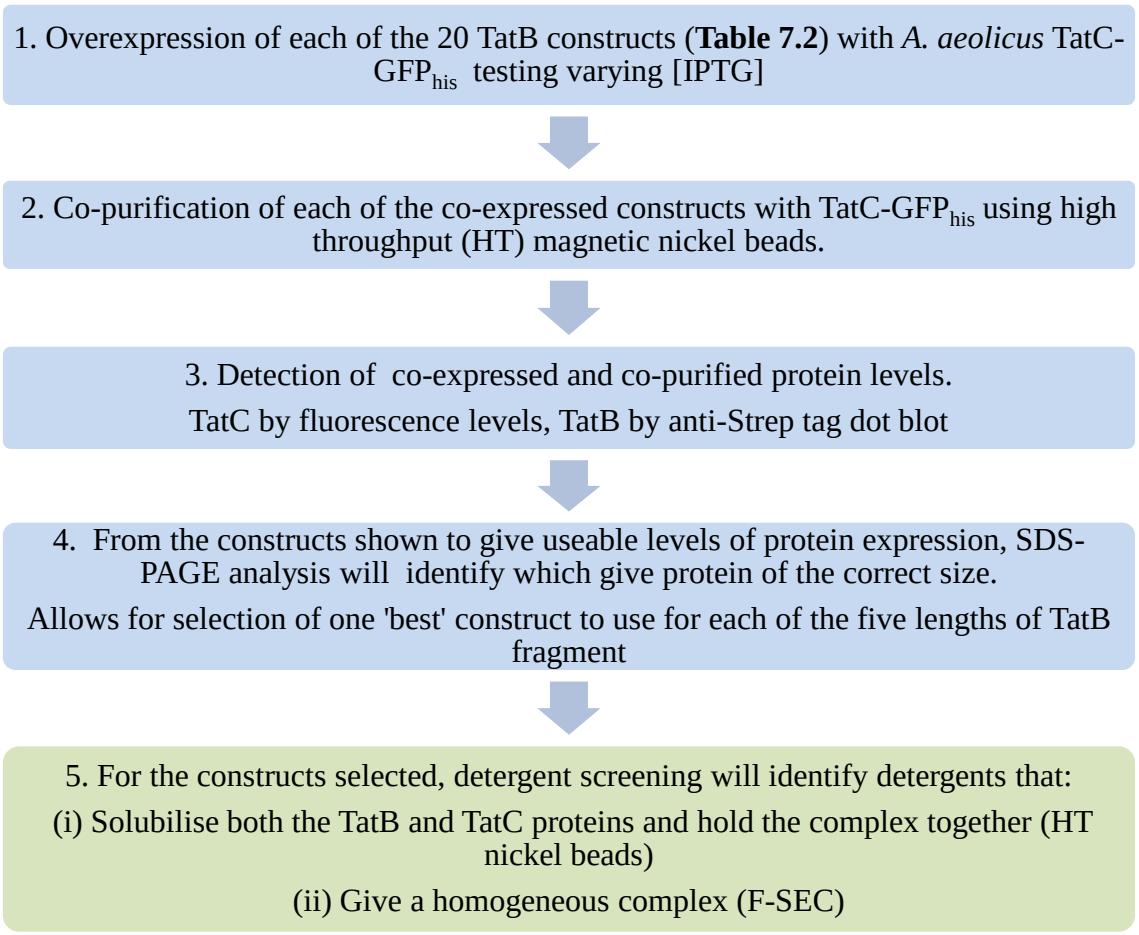


Figure 7.3.3 An overview of the *A. aeolicus* TatB construct and detergent screening process.

The process firstly will identify a single, best construct for each of the five length of *A. aeolicus* TatB fragments (blue boxes) and then the best detergent to use for each construct when co-purified with *A. aeolicus* TatC (green box).

The setup in **Table 7.3** allows scouting of eight 'conditions', A-H, for each length construct of TatB, in which the fragment was expressed from a different vector, with either a different Strep-tag/3C-Strep tag, at a different concentration of IPTG. The first step in the screening process aimed to identify which of the eight conditions were the best to express each TatB fragment. This was achieved by using the screening steps described in the blue boxes in **Figure 7.3.3**.

Tag and vector		Condition	**TatB insert**					
			FL	**1-94**	**1-17**	**1-25**	**1-64**	**TatC only**
			Induction condition					
Strep tag	CDFDuet-1	A	0.1mM IPTG					
		B	0.4 mM IPTG					
	pACYCDuet-1	C	0.1mM IPTG					
		D	0.4mM IPTG					
3C-Strep tag	CDFDuet-1	E	0.1mM IPTG					
		F	0.4 mM IPTG					
	pACYCDuet-1	G	0.1mM IPTG					
		H	0.4 mM IPTG					

Table 7.3 Expression conditions for the five different length constructs of TatB. Each construct was in pACYCDuet-1 and CDFDuet-1 with either a Strep-tag, or a 3Csite-Strep-tag (A-H) co-expressed with TatC-GFP**.** Induction was achieved with different concentrations of IPTG, at 24^{o}C overnight

At the end of the growth, 1.5ml of each culture was transferred to a 96 well, 1.6ml deepwell block (Nunc) and the cell pellet harvested by centrifugation at 4,000 rpm for 30mins at 4^{o}C. These pellets were frozen at -80^{o}C.

In order to assess the level of co-expression with TatC-GFP in these cell pellets, a High Throughput (HT) nickel affinity co-purification with TatC-GFP$_{his}$ in DDM of each of the TatBC constructs in each condition was carried out with nickel magnetic beads (Qiagen) (**Table 7.4**). DDM was chosen for this experiment as this detergent is known to be extremely mild (Privé, 2007) and can retain the integrity of some membrane protein complexes. The amount of protein present in the elutate was assayed. The level of TatC-GFP present was assessed by reading the GFP fluorescence, and the level of TatB was assessed by an anti-Strep Dot Blot of the elutions.

1. The cell pellets were lysed in 105µl of lysis buffer (100mM NaH_2PO_4 pH7.5, 600mM NaCl, 20mM Imidazole, 2mg/ml lysozyme (Sigma) and 8kU Benzonase (Sigma) and protease inhibitor (Roche)) by mixing at room temperature (RT) for 30 minutes at 180rpm. The re-suspension was aided by pipetting the pellet with a multi channel pipette (Ranin). The pellets were then subject to three freeze-thaw cycles between -80^{o}C and RT.

2. 105µl of 2.5% w/v DDM (Anatrace) was then added to each cell suspension, and mixing at RT continued for another 45mins, and aided by mixing with the multi channel pipette (Ranin). Once solubilisation was complete, the HT block was spun for 30mins at 4,000 rpm at 4^{o}C.

3. 15µl of magnetic Nickel-NTA affinity beads (Qiagen) were then pipetted into the wells of a HT, 96 well ELISA U-bottomed microplate (Greiner bio-one). 210µl of the soluble supernatant of the crude lysates was added to each of these wells, and were incubated for 1hr at RT at 180 rpm to affinity bind the tagged TatC-GFP_{His}

4. The plate was then transferred onto a magnet, to allow the removal of the unbound protein fraction, whilst retaining the magnetic beads on the plate. The beads were then washed twice with 50mM NaH_2PO_4 pH7.5, 300mM NaCl, 10mM imidazole with 0.05% w/v DDM (Anatrace) by addition of 200µl of buffer and incubation with shaking at RT for 10 minutes.

5. To elute bound protein from the beads, 50µl of the buffer above, supplemented with 250mM imidazole, was added to the beads, incubated and mixed for 10 minutes at RT. The supernantant was used to determine the amount of protein purified.

Table 7.4 The High Throughput nickel magnetic bead purification protocol. This was used to co-purify TatC-GFP_{his} and TatB from the constructs and expression conditions given in **Table 7.3.**

The eluted sample from the beads was then transferred to a black, 96 well, clear bottomed plate (Grenier bio-one). The level of GFP fluorescence was then assayed for using a HT fluorescence plate reader (SpectraMaxM5, Molecular devices), at an excitation wavelength of 485nm and emission at 538nm.

The levels of TatC-GFP fluorescence when co-expression and co-purification with each construct is given in **Figure 7.4.1**. This shows the varying levels of TatC-GFP that has been purified from the co-expression cultures, along with the TatC-GFP alone as a control. There is a reasonable level of purification of TatC from most vector/Tag/IPTG conditions when co-expressed and purified with each of the five length of constructs of TatB. This confirmed that TatC has expressed, and was able to be purified in this HT manner.

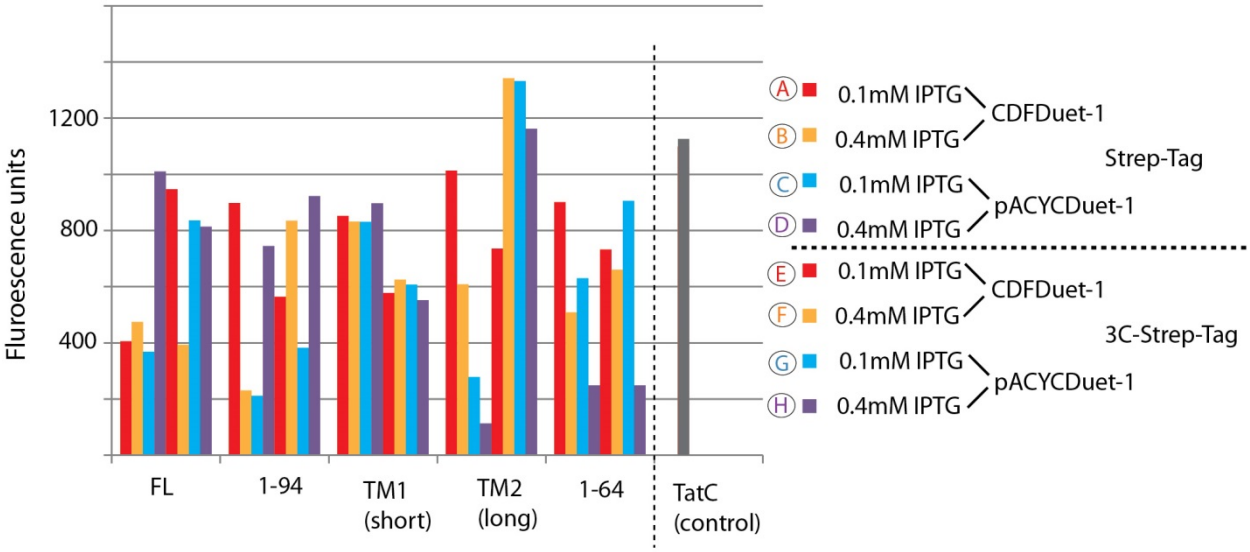


Figure 7.4.1. Level of *A. aeolicus* TatC-GFP present in the nickel bead eluate.

Following co-expression and co-purification with the 20 constructs of *A. aeolicus* TatB, each expressed in two conditions (**Table 7.3**), the level of TatC was assessed by reading the GFP fluorescence in a HT fluorescence plate reader (SpectraMax M5, Molecular devices) at excitation wavelength of 485nm and emission at 538nm.The first 4 rows of each length TatB are for the Strep constructs, and the second 4 rows for the 3C-Strep constructs. The TatC-GFP pWaldo construct alone (grey) is shown as a positive control.

In order to assess the levels of Strep-tag TatB and 3C-Strep-tag TatB that had co-expressed and co-purified with the TatC-GFP, 2µl of the HT magnetic nickel bead (Qiagen) elutate were subjected to an anti-Strep dot blot. This was achieved using the protocol in **Appendix 8.4.3**, with a 1:1000 dilution of the primary anti-Strep antibody (Qiagen) and a 1:1500 dilution of the secondary antibody (stabilised peroxidase conjugated goat anti-mouse, 10µg/ml, Thermo Scientific). The result of the dot blot is shown in **Figure 7.4.2**

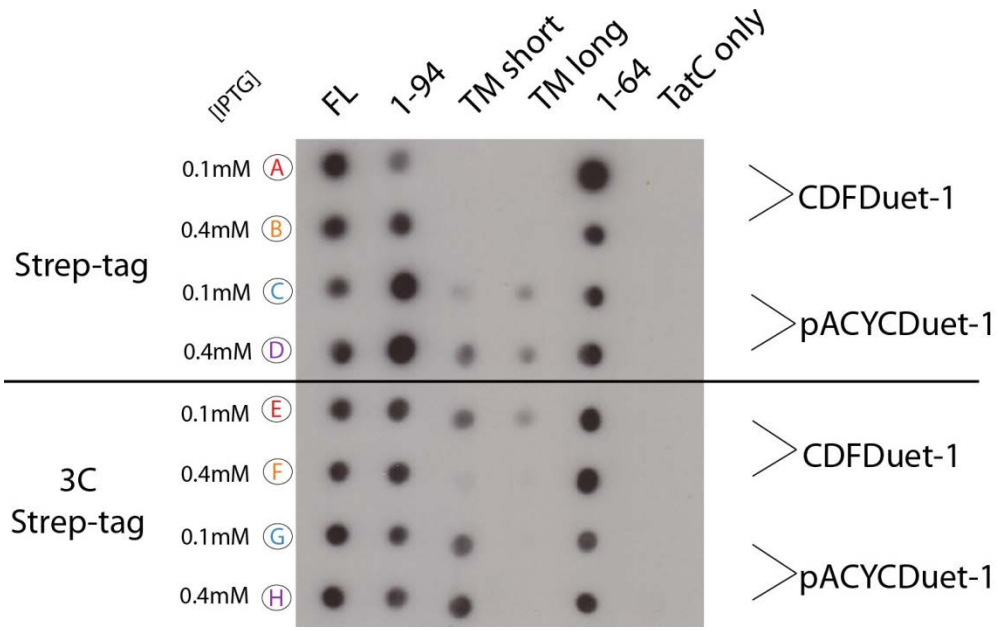


Figure 7.4.2. Dot blot of the HT nickel magnetic bead elutate from co-expression and co-purification of TatC with the 20 constructs of TatB.

The blot has a grid set up, as in **Table 7.3**, where each of the five length constructs are given across the top of the blot, (FL, 1-94, TM short (1-17), TM long (1-25), 1-64) followed by the TatC-GFP alone as a negative control. The Strep-tag constructs, in each of the two vectors (CDFDuet-1 and pACYCDuet-1) and at each of the two IPTG concentrations are given at the top half of the blot, and the 3C-Strep-tag constructs in the bottom half of the blot. 2µl of the elution was spotted onto the nitrocellulose membrane (GE healthcare) and subjected to incubation with the anti-Strep antibody (Qiagen) followed by incubation with the anti-mouse antibody (Thermo Scientific). The TatC alone elution serves as a negative control.

The dot blot showed that there was signal for all the FL constructs, Strep-tagged and 3C-Strep-tagged in both vectors, for both concentrations of IPTG tested. This was also the case for the 1-94 length constructs, and the 1-64 length constructs, showing that all these constructs had co-expressed and co-purified with TatC-GFP. However, there was a very reduced signal for both of the TM helix constructs, where the short (1-17) construct appeared to give more signal than the long (1-25) construct. In order to investigate the nature of the species giving signal on the dot blot, to ensure the constructs are of the correct expected size, the elutates were analysed by SDS-PAGE.

Fl, 1-94 and 1-64 length constructs

5µl of each of the FL, 1-94 and 1-64 construct elutions were run on a 15% SDS-PAGE gel. These gels were then subject to western blotting with the anti-Strep antibody (Qiagen) at a 1:750 dilution, and the secondary antibody (anti mouse, Thermo Scientific) at a 1:1,500 dilution using the protocol given in **Appendix 8.5.3.**

Figure 7.4.3 shows the anti-Strep western blot of the elutions. **A** shows the Strep-tag constructs, where A, B, C and D represent the vectors and IPTG concentration in **Table 7.3,** and **Figure 7.4.3 B** shows the 3C-Strep-tag versions of the constructs where again the letters refer to those in **Table 7.3.**

- Strep-Tag constructs: Inspection of the position of the protein bands showed condition A gave a single band for each of the constructs at the expected positions, marked with a red asterisk. There was some ‘smearing’ of the band, but this could be similar to the diffuse way TatB runs on an SDS-PAGE gel, as seen for the FL *E.*

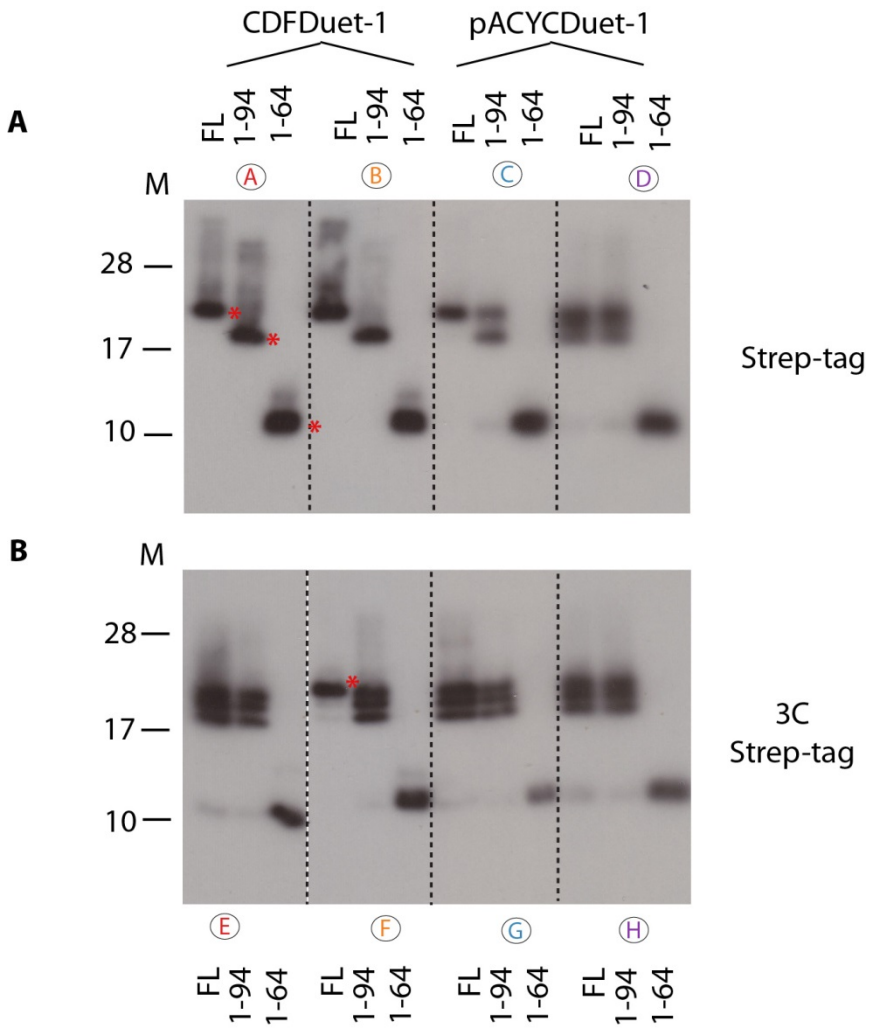


Figure 7.4.3. Western blot analysis of a SDS-PAGE gel of the FL, 1-94 and 1-64 fragments from the co-expression screening.

15% (w/v) SDS-PAGE of 15µl of the eluate from the HT co-expression and co-purification with TatC-GFP using magnetic nickel beads (Qiagen). **(A)** The Strep-tagged constructs and **(B)** the 3C-Strep tag constructs. Constructs in the CDFDuet-1 vector are shown in the first half of the gels, and the 3C-Strep constructs in the second half of the gels. Lane blocks are labeled according to the condtions in **Table 5.** For the constructs chosen to use for further studies, the red asterisks mark the bands of TatB protein at the correct size. Marker molecular masses shown in kDa.

coli and *A. aeolicus* TatB proteins (**Figures 7.1.1A** and **7.2.2**). Bands of the same size and diffusion were seen for each of the constructs in the B condition, where the constructs were in the same vector (CDFDuet-1) but induced at a different induction concentration of IPTG. The 1-64 in the pACYC vector (conditions C and

D) behaved similarly, but the FL and the 1-94 construct appeared to now give additional bands, indicative of degradation, or SDS-resistant aggregation of the samples. Thus, the Strep-tagged/CDFDuet-1 vector combination expressed with 0.1mM IPTG for the FL, 1-94 and 1-64 constructs, i.e. condition A, were chosen to pursue further.

- 3C-Strep-tag constructs: the 1-64 constructs gave a single band of the expected size for all conditions, however these were less intense than for the Strep-tag constructs, conditions A and B. For this construct, it was therefore decided to concentrate on the CDFDuet-1 Strep-tag construct, with induction at 0.1mM IPTG (condition A). The FL and 1-94 constructs again gave multiple bands of unexpected sizes. This is true for all of the constructs, apart from the FL in condition F, marked in **Figure 7.4.3 B** with a red asterisk. This construct, FL 3C-Strep-tag in CDFDuet-1 at 0.4mM IPTG induction, was therefore chosen pursue further.

The co-expression screening identified at least one construct and expression condition for each of the FL, 1-94 and 1-64 length of TatB sequences to take forward into further purification trials.

TM constructs

In order to assess if the small amount of signal seen on the dot blot for the two TM helix constructs were indeed the TatB protein, these were also subjected to western blotting. To enable the low molecular weight proteins to be visualized, a 20% Tris-Tricine gel (Schagger, 2006) (**Figure 7.4.4)** was run and blotted as described previously.

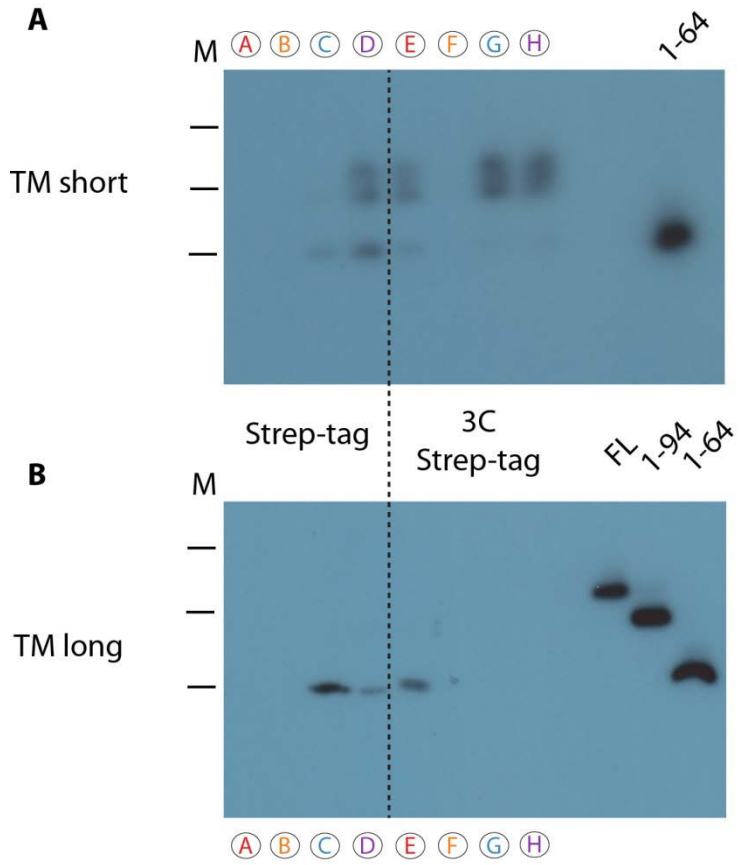


Figure 7.4.4. Western blot analysis of the TM fragments from the co-expression screening.

20% (w/v) Tris-tricine gel of the **(A)** TM short (1-17) constructs and the **(B)** TM long (1-25) TatB constructs eluted from the magnetic nickel beads (Qiagen). The first half of both gels shows the Strep-tag constructs, and the second half shows the 3C-Strep tag constructs. Control eluates from the FL, 1-94 and 1-64 samples are shown on both gels and labeled accordingly. Marker positions are shown for the Fermentas pre-stained Page ruler plus.

Figure 7.4.4 A shows the Western Blot of elution's from the co-expression and co-purification of the short TM helix (1-17), where the first half of the gel is loaded with the Strep-tag constructs, and the second half of the gel with the 3C-Strep-tag constructs. Each of the different vector and induction conditions used (A-H) are as given in **Table 7.3.** The same layout is used in **Figure 7.4.4 B**, but the eluate is from the longer, TM (1-25) construct. The control constructs are from condition A.

The blots showed that the TM short construct gave a mixture of bands at the incorrect molecular weight expected for the six of the constructs, and thus the construct may have undergone aggregation. The TM long (1-25) however, gave single bands in three of the conditions, C, D and E. These bands ran smaller than the 1-64 construct, as would be expected given their molecular weight of 3.8kDa (Strep-tag) and 4.7 kDa (3C-Strep-tag). However, these bands were of very weak intensity compared to the three bands from the control constructs, the FL, 1-94 and 1-64 which have been chosen to further pursue. This indicated that these TM constructs may not be useful for structural studies, and so no TM helix constructs were taken any further in the screening process.

7.5 Detergent screening of the four most promising *A. aeolicus* $TatB_{Strep}$ and $TatB_{3C\text{-}Strep}$ constructs

This initial experiment identified for each of the FL, 1-94 and 1-64 length fragments, a vector and expression condition that gave tagged *A. aeolicus* TatB protein of the correct size, and which did not appear to be undergoing degradation or aggregation when co-expressed with TatC-GFP. These are listed in **Table 7.5**.

From now on, 'FL_{strep}, FL_{strep}+3C, $1\text{-}94_{strep}$ and $1\text{-}64_{strep}$' will exclusively refer to the constucts in the vectors and from the expression conditions in **Table 7.5**, where 'FL_{strep}+3C' refers to the construct with the 3C site prior to the CT Strep-tag.

In order to determine which detergent is best to use in the purification of these complexes of *A. aeolicus* TatBC, detergent screening was carried out. This aimed to identify which detergents are able to co-purifiy the complex, as although a particular detergent may be able to solubilise TatC from the membrane, it may not be able to also solubilse the TatB component. Furthermore, even if the detergent is able to successfully solubilise both of the

	Length of construct	Tag/3C site	Vector	Induction condition [IPTG]
1	FL	Strep	CDFDuet-1	0.1mM
2		3C-Strep	CDFDuet-1	0.4mM
3	1-94	Strep	CDFDuet-1	0.1mM
4	1-64	Strep	CDFDuet-1	0.1mM

Table 7.5 The four best *A. aeolicus* TatB constructs and corresponding expression conditions, chosen for detergent screening with co-expressed *A. aeolicus* TatC-GFP.

proteins from the membrane, it may not be able to effectively maintain the complex between the two proteins, and allow them to stay together. As the intial HT co-purification attempt described above was carried out in DDM, this detergent is clearly able to do this. Once detergents have been identified that are able to solubilse and maintain the intergirty of the complex, the homogeneity of the complex in this detergent needs to be screened. This can be done using the F-SEC system following the GFP fluorescence from the TatC-GFP fusion.

The detergents used for these experiments were selected from the initial list of detergents that were shown to be able to solubilse TatC (**Chapter 1, Table 1**) to give eight starting detergents to use (shown in **Figure 7.5.1 A**). In order to screen these detergents by F-SEC, L56 cells were transformed with both the TatC-GFP in pWaldo and the TatB constructs given in **Table 7.5**. 100ml cultures were grown in TB in a 250ml flask, as described for the HT expression growths. 25ml of these cultures were harvested, and the pellets re-suspended in 840µl of the lysis buffer used previoulsy. The pellets were left to resuspend for 30 minutes at RT by mixing on a rotating wheel at 40 rpm. The cell pellet was split into eight, by transfering 8x150µl of each pellet to a 96 well, 1.6ml deepwell block (Nunc). Each pellet was solubilsed in one of 8 detergents at 1.25% w/v final concentration. The HT magentic bead protocol was then repeated, as in **Table 7.5**, except each wash and elution

buffer was supplemented with the solubilisation buffer at a w/v dependent on its CMC, and listed in **Chapter 1, Table 1.**

50µl of the elutions from this affinity step were assayed for GFP fluroescence as described in the previous section, and are shown plotted in **Figure 7.5.1 A**. This confirmed the presence of TatC-GFP when co-expressed with all 4 of the TatB constructs. The level of TatC purified in each detergent differed quite considerably, with DMNG (Anatrace) able to purify the greatest amount of TatC-GFP.

In order to now assess whether the TatB constructs were also co-purified with TatC in each of these detergents, 2µl of the elution from each detergent were subject to the anti-Strep dot blot as descibed previously. The result of this is shown in **Figure 7.5.1 B**. There was

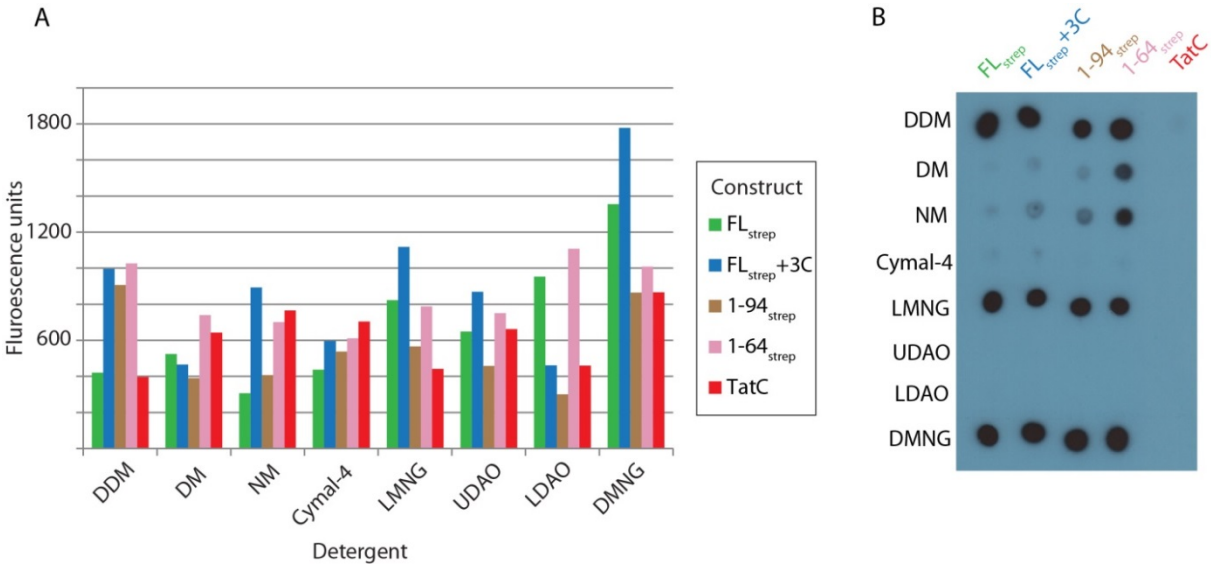


Figure 7.5.1. Assessment of the level of protein purified in each detergent.

The levels of **(A)** TatC-GFP given by fluorescence reading, as described previously and **(B)** TatB estimated by an anti-Strep dot blot, as described previously, for each of the four constructs of TatB (Table 5) co-expressed and co-purified with TatC in eight different detergents using HT magnetic nickel beads (Qiagen).

clearly very strong signal for all of the constructs co-purified in DDM, LMNG and DMNG. Interestingly, even detergents related to DDM, such as NM, failed to allow the co-purification of complexes from the FL and 1-95 fragments, whereas the 1-64 was able to stick to TatC in the NM and DM detergenents, but at a lower level than in the other two detergents.

This experiement indicated that the three detergents, DDM, LMNG and DMNG allowed the *A. aeolicus* TatBC complex to hold together during the purification. The next stage in the detergent screening processes involved assessesment of the homogenity of the complexes in these three detergents, in order to see if any appear better to use than others. This was achieved by monitoring the fluorescence from the TatC-GFP protein, by an in-line fluorescence detector connected to a size exclusion column (F-SEC). The shape of the fluorescence traces would give an indication how reasonably mono-dispersed and singlar the species present within these complexes for each detergent (Drew *et al.*, 2005).

A S200 5/150 analytical column (volume of 3ml) was connected to an in-line Prominence Fluorescence detector (Shimadzu), reading at a excitation wavelength of 395nm and emission at 509nm. The column was allowed to equilibrate in 5 column volumes of SEC buffer (20mM hepes pH 7.5, 150mM NaCl supplemented with 0.02% w/v LMNG, 0.02% w/v DMNG or 0.03% w/v DDM (all detergents were from Anatrace). A 100µl loop was connected, and 30µl of each of the elutions was injected in turn onto the column in the appropriate detergent, and the fluorescence monitored. The F-SEC traces for each co-expression construct, including the TatC-GFP alone control for comparison, are shown in **Figure 7.5.2.**

Traces in all three detergents for every construct, including the TatC-GFP alone control, showed two peaks. The second peak arose from the presence of free GFP, which can be

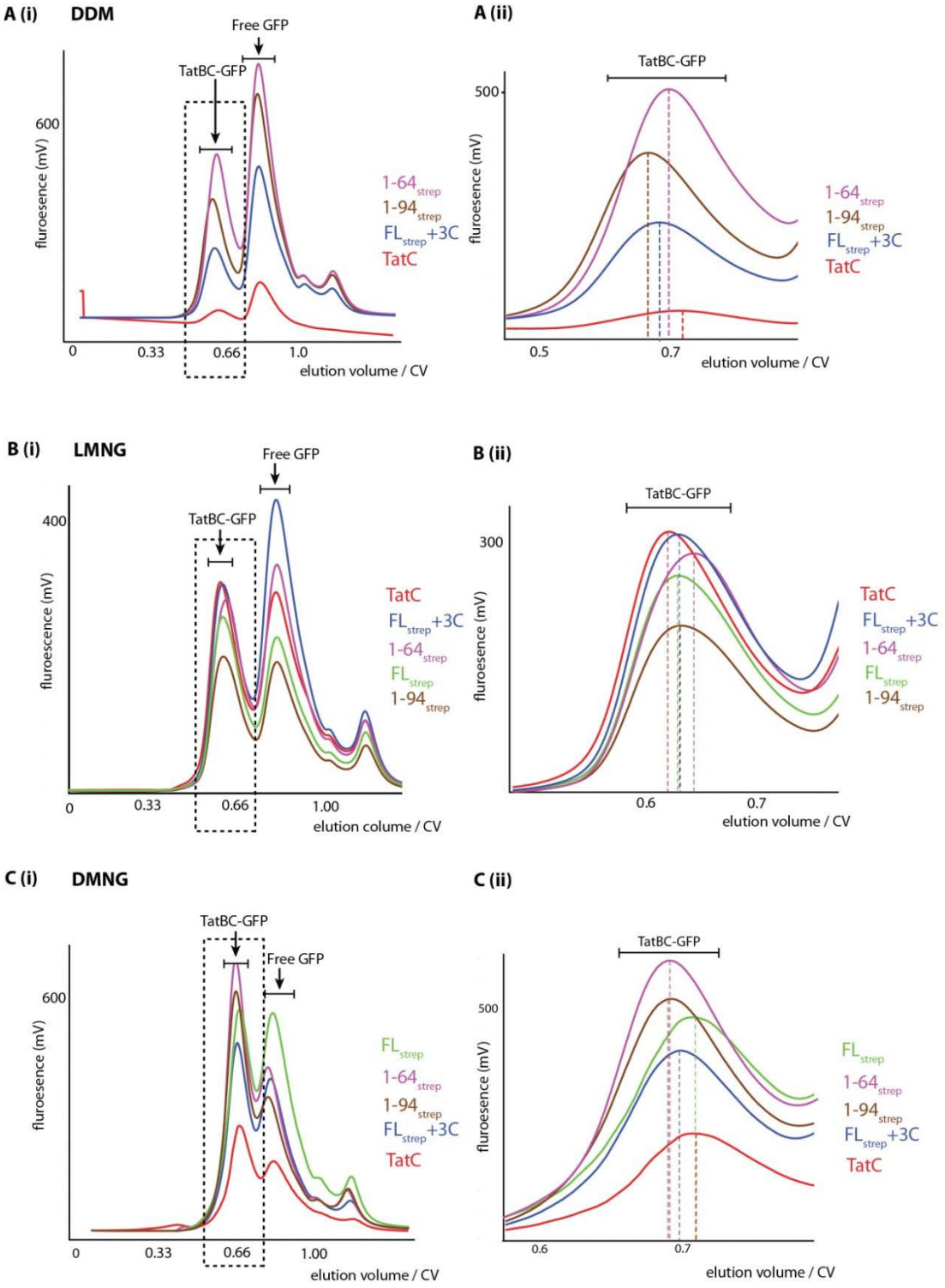


Figure 7.5.2 Assessment of the homogeneity of each of the four TatB constructs by F-SEC.

The TatB constructs from **Table 7.5** were co-expressed and co-purified with TatC-GFP and 30µl of the HT nickel magnetic bead (Qiagen) elutions were injected onto an S200 5/150 column (GE healthcare) pre-equilibrated in the appropriate detergent buffer and the fluorescence monitored by an inline Prominence fluorescence detector (Shimadzu) at excitation of 395nm and emission at 509nm. Traces shown in **(A)** DDM **(B)** LMNG and **(C)** DMNG**.** The fluorescence traces from one column volume are shown in **(i)**, and the peak of interest (containing the TatBC-GFP) shown zoomed in **(ii)** with the elution volumes marked as dashed lines.

purified in this HT experiment, as there is no additional membrane purification step (as for the large scale purifications, **section 3.1**). This allowed any soluble GFP, which may have been cleaved from the TatC-GFP during expression or purification, to also be purified. This peak always ran much smaller than the fusion, and thus was resolved from the TatC-GFP peak. Its presence therefore did not affect the assessment of the shape of the TatC containing peak. The traces showed that for each detergent, there was a single peak for the TatBC-GFP complex, which was of a very similar shape to the TatC-GFP alone peak. The elution position of the TatBC-GFP complex peak had subtly shifted position compared to the TatC-GFP alone for all the constructs, in each of the three detergents (**Figure 7.5.1 (ii))** which demonstrated the alteration in mass of the complex now that the TatB protein was present.

As all three of these detergents gave reasonably symmetrical and homogenous looking peaks, they all therefore seemed promising to use in large scale purifications of TatC. LMNG was selected as the initial detergent to carry out large scale purifications as this detergent had previously shown to be successful in purifiying TatC alone to homogeneity.

7.6 Large scale purifications of *A. aeolicus* $TatB_{strep}C$ from the co-expression vectors to assess the complex homogeneity by MALLS

Having firstly identified suitable over-expression constructs of TatB for three different lengths of the protein that allowed for co-expression and co-purification with TatC, F-SEC subsequently identified three promising detergents were identified that give homogenous complexes.

In order to know the true homogeneity of these three complexes and to try to understand the oligomerisation state of the complexes, Multi Angle Laser Light Scattering (MALLS) was carried out. This would be able to show if these *A. aeolicus* TatBC complexes were more homogenous than the sample of the *E. coli* version (**Figure 7.1.1 B**). In order to run MALLS experiments, large scale purifications were carried out from 6L cultures, to ensure enough protein material would be obtained. For these studies, it was decided to just use the FL_{Strep} construct identified, rather than both this and the FL_{Strep} +3C, as they seemed to behave similarly in the detergent screening. The 3C-containing construct can always be later used if the Strep-tag needs to undergo cleavage.

The co-expression of the three remaining constructs was carried out in 6L of TB, as discussed in this chapter for the small scale studies. The co-purification was then carried out in LMNG, in a similar manner as discussed for TatC-GFP in **section 3.1**, with a few differences in some steps. The TEV cleavage buffer was altered to 50mM NaH_2PO_4 pH 8.0, 300mM NaCl with 0.02% w/v LMNG. The reverse Nickel column was equilibrated in this dialysis buffer, and the column run as before. The flow through from the reverse Nickel column step was collected, and concentrated to 10ml. This was then applied to a 1ml *Strep*-Tactin (Qiagen) affinity column, at a slow flow rate, >0.3ml/ml, to bind the Strep-tag at the CT of the TatB constructs. The flow-through from this was loaded onto the column for a second time, in an attempt to bind as much $TatB_{strep}C$ as possible.

The column was then washed in 10 CV of the dialysis buffer, and $TatB_{strep}C$ eluted in dialysis buffer supplemented with 2.5mM desthiobiotin (Sigma). The protein containing fractions, as assessed by UV reading at A_{280} and SDS-PAGE, were concentrated to 5ml in a 100kDaMW cut off concentrator device (Millipore) and injected onto an S200 16/60 SEC column (GE healthcare) equilibrated in 20mM hepes pH7.5, 150mM NaCl and 0.02% w/v LMNG.

The UV profile of the SEC traces for each of the three Strep-tag TatB constructs is given in **Figure 7.6.1 A**. SDS-PAGE analysis of the SEC fractions showed a band expected for TatC, and one band that corresponded to each of the TatB constructs, **Figure 7.6.1 B**. The

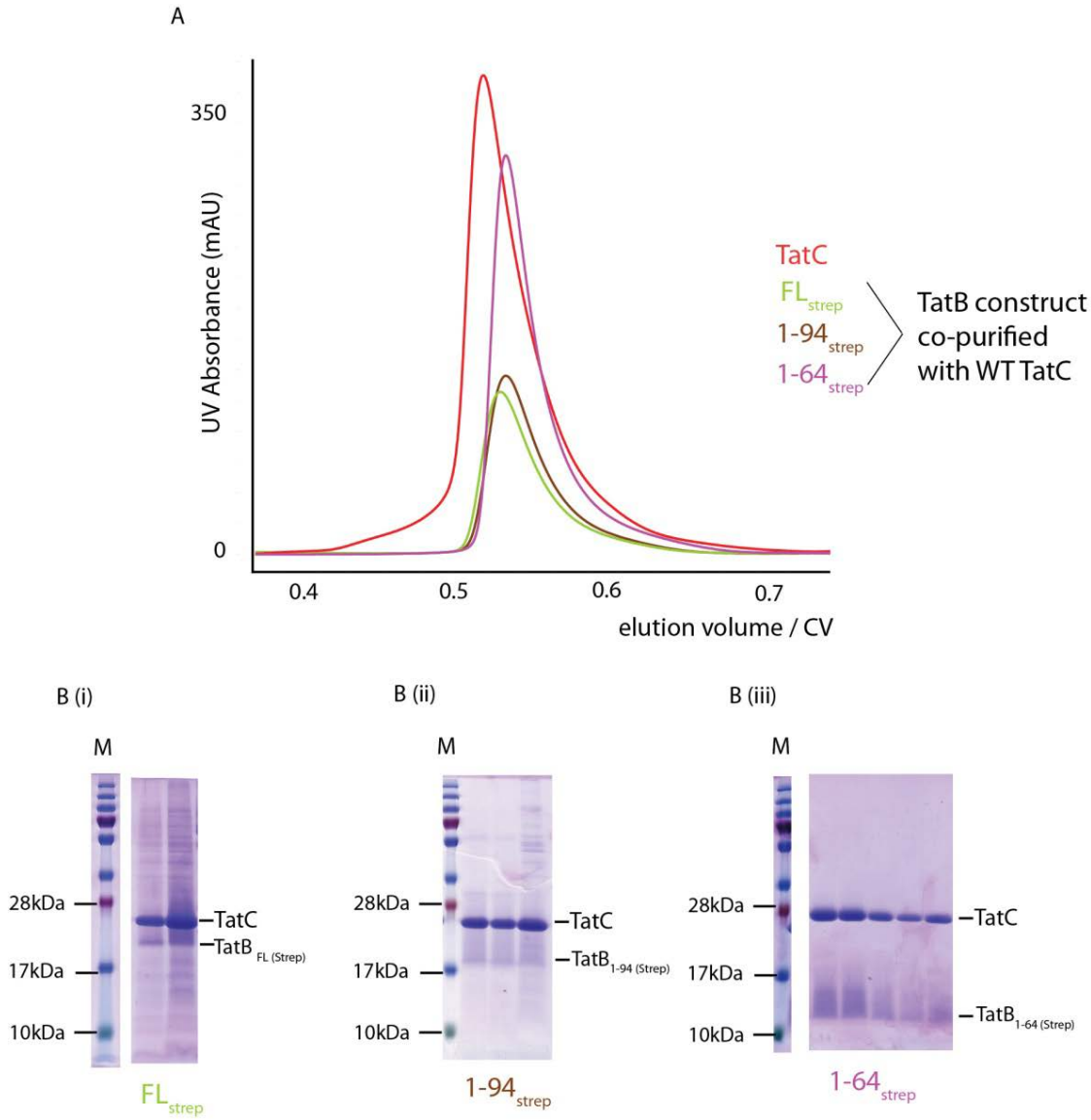


Figure 7.6.1 Large scale preparations of TatC with three TatB constructs in LMNG.

(A) The UV SEC traces from injection of 5ml of the *Strep*-Tacti column (Qiagen) elutate on the S200 16/60 column (GE healthcare) shown overlaid with the TatC alone purified in LMNG. **(B).** SDS-PAGE analysis of the SEC fractions from **(i)** TatC-$TatB_{FLstrep}$ **(ii)** TatC-$_{TatB1-94Strep}$ and **(iii)** TatC-$TatB_{1-64Strep}$

presence of *A. aeolicus* TatB in these bands were confirmed by western blotting with the anti-Strep antibody (not shown). Native mass spectrometry also confirmed the presence of TatC, and the correct mass of the TatB protein expected in each of the purified samples (Dr Julien Marcoux, Department of Chemistry, University of Oxford).

The SEC traces all had a very similar shape to that of the WT TatC, and showed a slight shift in the elution position. The peak heights were lower for the 1-94 and FL constructs, which demonstrated the reduced yield of these longer constructs. However, there was enough protein to allow MALLS experiments to be run at two concentrations for each construct, and the traces are shown in **Figure 7.6.2**. The overall shapes of the calculated mass profiles for each construct were generally quite flat and consistent across the peak. There was only a small difference in the mass for the two concentrations for each construct. There was, however, some hint of a slight increase in mass for the highest concentration of the $TatB_{1\text{-}64}TatC$.

Protein conjugate analysis on these samples allowed the mass of the protein alone to be calculated. The mass profiles for the protein alone had the exact same shape distribution as that shown in the figure for the protein:detergent mass profile. Given the nice shape of the mass profiles for these constructs, it indicated that these samples may be homogenous enough to crystallise. The value of the calculated molecular masses for each TatBC complex at the two concentrations tested is given in **Table 7.6**, along with the expected mass for a 1:1 complex of the *A. aeolicus* TatB construct with TatC.

The conjugate analysis shows that the measured masses for the FL and the 1-94 constructs is roughly equivalent to that expected for a 1:1 complex. What is interesting is that these masses show that the TatC dimer (which has a mass of 56kDa) seen in this detergent by MALLS for TatC alone (**Figure 3.1.2 A**) no longer appears to be present. The presence of

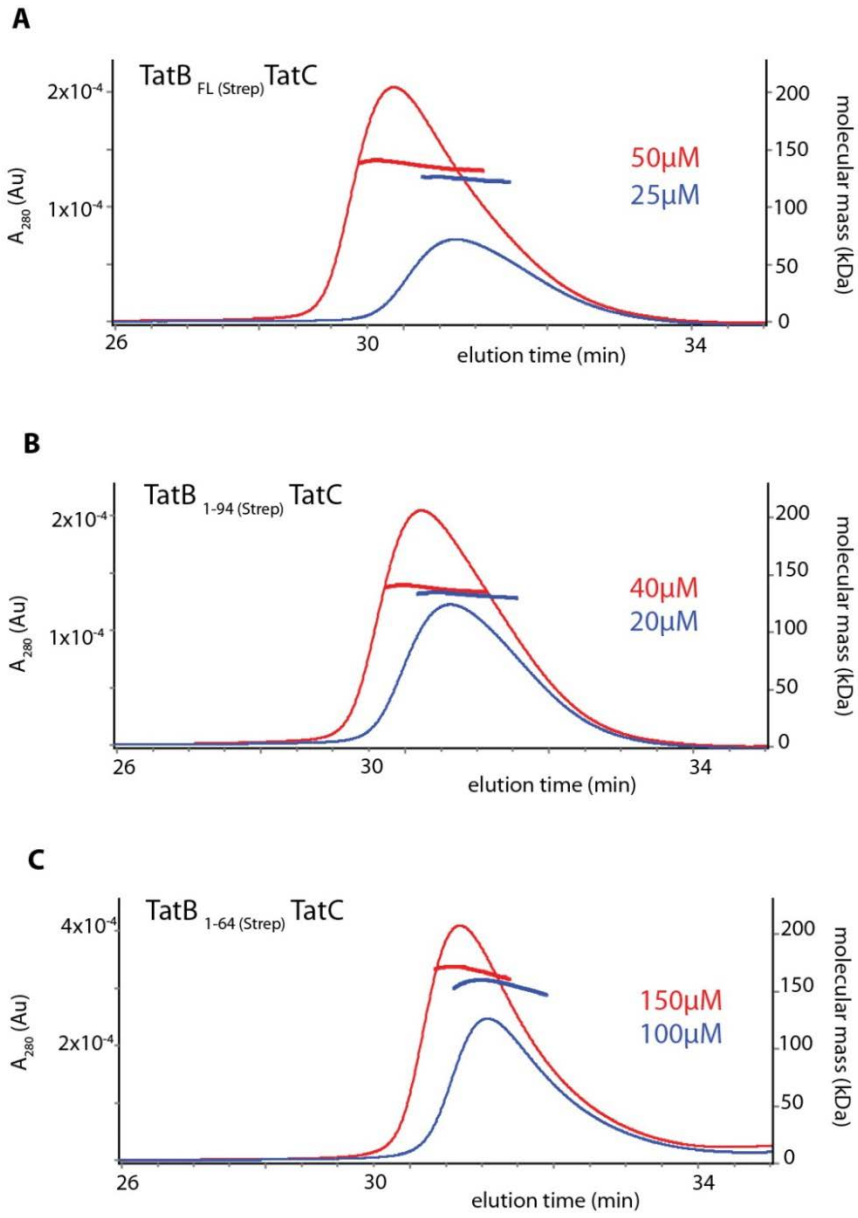


Figure 7.6.2 MALLS analysis of each of the three TatB constructs copurified with TatC in LMNG.

(A) the FL_{Strep} TatB construct **(B)** the $1\text{-}94_{Strep}$ construct and **(C)** the $1\text{-}64_{Strep}$ construct purified with TatC in LMNG.

Two concentrations are shown for each construct. The UV trace is given as the thin lines, for the higher (red) and the lower (blue) concentrations. The corresponding calculated molecular mass across the peak for the total protein:detergent complex is given as the thick lines.

Construct	Mass of TatB construct	Expected mass for a 1:1 complex (kDa)	Measured molecular mass (kDa)	Concentration in MALLS experiment (µM)
TatB $_{FL(Strep)}$	13.8 kDa	41.9	36.4	25
			45.6	50
TatB $_{1\text{-}94(Strep)}$	12.2 kDa	40.3	45.1	20
			47.0	40
TatB $_{1\text{-}64(Strep)}$	8.6 kDa	36.7	62.0	150
			56.1	100

Table 7.6 Molecular masses for the *A. aeolicus* TatBC constructs. The expected mass for a 1:1 complex with *A. aeolicus* TatC(28.1kDa) are given, along with the calculated molecular masses for the complex from the protein conjugate analysis of the MALLS experiments (two different concentrations per construct)

the TatB component must have therefore some how altered the interactions between TatC and itself to cause TatC to no longer dimerise.

Interestingly, these detergent solubilized complexes indicate that larger oligomerisation is not occurring, unlike the sample of the *E. coli* TatBC complex. As there is no evidence from studies in *A. aeolicus* membranes as to the oligomeric state of their TatBC complex, this may mean that the *A. aeolicus* proteins may not form a large TatBC complex consisting of eight copies of each protein as in *E. coli* (McDevitt *et al.*, 2006), but instead may exist in the membrane as smaller receptor complexes, containing just one TatB and just one TatC. Alternatively, the lack of oligomerisation may be an artifact of detergent solubilisation.

If these samples truly consist of single TatBC units, a crystal structure from these samples still may not answer the question on how the larger, multivalent TatBC complex is built up, but would give molecular details on the 1:1 complex between the TatB and TatC proteins. Given the tight cooperation during the function of these two proteins (discussed in **section 6.3**) and the evidence that multivalency of the TatBC complex is unlikely to be strict requirement for its function (discussed in **section 4.3**), a TatBC structure will be a great step forward in understanding how the two proteins function together in binding signal peptides.

The masses for the shorter, 1-64, construct are harder to rationalise. The increase in mass seen for this construct could be due to more TatB binding to the TatC, so the ratio of the 2 proteins is no longer 1:1. This could be due to $TatB_{1-64}$ being shorter than the FL and 1-94, and thus some of the sequence of the additional amphipathic helix compared to TatA (de Leeuw *et al.*, 2002) has thus been removed, causing it to now be more ‘TatA’ like; which is known to be able to associate in greater stoichiometric amounts to TatC than the TatB

component. Peptide analysis will be carried out on these samples to assess the relative stoichiometry of the proteins present.

7.7 Summary and future perspectives for the structural characterisation of the TatBC complex

Cloning five different length fragments of *A. aeolicus* TatB into compatible vectors to allow for co-expression, and co-purification by a dual affinity tag method, was followed by intensive construct screening, and the subsequent detergent screening of the best constructs for 3 lengths of TatB, FL, 1-94 and 1-64. This led to the large scale purification of the complexes in LMNG, which were analysed by MALLS.

The MALLS traces of the *A. aeolicus* TatBC FL and 1-94 complexes looked very encouraging, especially compared to the heterogeneity seen in the MALLS trace of the *E. coli* TatBC complex (**Figure 7.1.1 B**). This indicated the *A. aeolicus* TatBC complexes purified in this study are homogenous enough for crystallization trials, and thus the constructs will undergo intensive crystallisation trials. Future work will also involve using these constructs and this purification system to asses the nature of the complexes formed in the remaining two detergents (DDM and DMNG) by MALLS, and additional promising ones will also be subject to extensive crystallisation trials.

ITC with the short and long SufI signal peptide will also be carried out, to assess for any difference in the affinity for these peptides now with TatB present. If this does lead to an increase in the affinity for the signal peptide, attempts will be made to co-crystallise this entire ternary complex.

7.8 Concluding remarks: A hypothetical model for the Tat translocase

Drawing together the current information in the Tat field, along with the structure of TatC and the other data presented in this thesis, allows for the 'cartoon' model (**Figure 1.4**) discussed in the introduction, to be updated. **Figure 7.8.1 A** shows the structure of TatC, in surface representation, along with the TM helix of TatB modeled into the TM5/TM6 'groove', based on the data presented from the cysteine disulphide crosslinking experiments and the crystal packing interactions (**Figure 6.2.1**), and the amphipathic helices speculatively placed running across the cytoplasmic face of TatC, satisfying photo affinity contact data (Zoufaly *et al.*, 2012).

The signal peptide may firstly associate with the membrane (Bageshwar *et al.*, 2009), undergo a two-dimensional search of the surface until it encounters the 'surface exposed binding' site on TatC, at the end of TM1 and on the first cytoplasmic loop, where the RR region around the RR motif is bound, and the twin arginines themselves potentially contacted directly by the two conserved glutamate residues (**Figure 4.3.4**).

Upon the consensus motif binding at this region of the TatC structure, TatC can then insert the peptide into the membrane (Fröbel *et al.*, 2012b), perhaps using the membrane exposed cavity on the concaved surface of TatC (Musser, unpublished) shown in **Figure 7.8.1 B**. The peptide may undergo deeper insertion into the membrane, potentially in a PMF dependent manner (Gérard & Cline, 2007), into a pocket between TatB and TatC (**Figure 7.8.1 C**) (Fröbel *et al.*, 2012a), allowing the NT of TatB to contact the h-region of the peptide (Alami *et al.*, 2003) and explaining the involvement of the periplasmic end of TatB in signal peptide binding (Lausberg *et al.*, 2012).

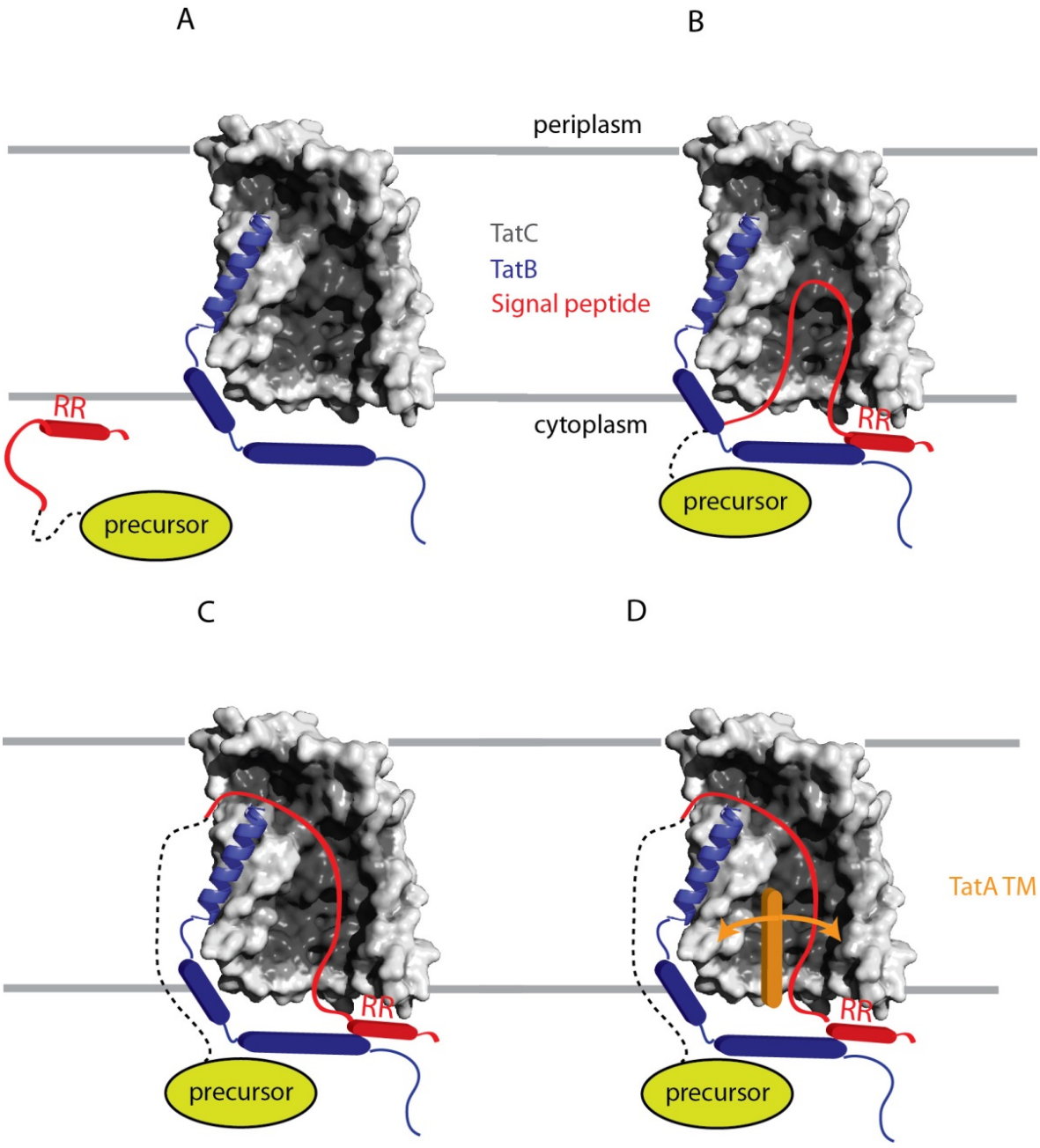


Figure 7.8.1 A hypothetical molecular model for the initial steps in Tat translocation.

(A) A model for the TatBC complex, based on the structure presented in this study and unpublished data on the path of the signal peptide. **(B)**TatC can bind signal peptides at their RR containing n-region at the surface exposed binding site identified in **Figure 4.3.4**, and can insert the signal peptide into the membrane (Fröbel *et al.*, 2012a), which may enter a deeper binding site, **(C)**, also mediated by TatB (Gérard & Cline, 2007). (**D**) Finally the TM region of TatA can be subsequently recruited to the TatBC complex (Dabney-Smith *et al.*, 2006) to bring about transport. This may occur at the concave face of TatC (**Figure 5.2.1**).

Finally, the TM region of TatA is then recruited to TatC, perhaps to its concave face (**Figure 7.8.1 D**) based on the docking experiments in **Figure 6.2.1**, and unpublished crosslinking results (Cline *et al,* University of Florida). This recruitment may be mediated by the negatively charged Glu165 residue embedded on this face of TatC, in the center of the bilayer (**Figure 4.3.2 A**). The structure of the *A. aeolicus* TatC protein has therefore allowed for much data within the Tat field to be rationalized, and for a potential model of the crucial initial step of signal peptide recognition to be proposed. The structure can pave the way for the design of experiments to test these hypotheses.

Main unanswered questions

However, questions still remain to be answered, such as the exact molecular details of the TatC-signal peptide interaction, and on how TatC binds to TatB. These questions will be addressed by building on the preliminary experiments discussed in this study (**Chapters 5** and **7**).

Other key questions include how the Tat system is powered by the PMF. The TatC structure revealed no obvious clues as to how this is utilized; there were no obvious hydrogen bond networks through the structure, or evidence for TatC acting as an ion antiporter, as has been suggested (Alder & Theg, 2003). The exact role played by the PMF is also still a matter of debate. It seems clear the PMF is needed for TatA recruitment to the TatBC receptor complex (Mori & Cline, 2002), but how it affects the stages of signal peptide binding to the TatBC receptor complex, is unclear (Gérard & Cline, 2007) (Fröbel *et al.*, 2012a). Given the structures of the SecA protein, either empty or with nucleotide bound (Hunt *et al.*, 2002), revealed evidence for how the SecA protein utilities ATP hydrolysis to change its conformation and drive the precursor though the Sec translocon

channel with its two-finger helix, further structures of TatC with the other subunits or signal peptides may help in elucidating the mechanism by which the PMF is utilized.

Finally, the structure of the multivalent TatBC complex is still unknown. The TatC structure gave good evidence for how TatB may associate in the 1:1 complex (**Figure 6.3.2**), but not how this is built up further to create the oligomeric complex. Further structures will therefore be required to rationalise how the receptor complex is arranged. This issue is reminiscent of the Sec translocon oligomer debate; like TatBC, it appears that mechanistically one copy of the Sec translocon is sufficient to transport precursors (Van den Berg *et al.*, 2004), but the known existence of dimers in the membrane has led to many attempts to structurally characterize and understand the oligomerisation (Breyton *et al.*, 2002, Mitra *et al.*, 2005), but a satisfactory answer is still outstanding. Thus, the organisation of the oligomeric TatBC complex, and its mechanistic function, may remain elusive for a while longer.

Appendix 8.1: Structural characterisation of CbiX, the terminal enzyme of Siroheme Biosynthesis, from *Paracoccus pantotrophus*

Siroheme synthesis

Co-factors carry out many essential roles in a wide range of biological processes. Of these, the modified tetrapyrrole class of co-factors contain chlorophyll, heme, cobalamin (vitamin B_{12}), heme d_1, the coenzyme F_{430}, and siroheme; thus are involved in many vital processes. For example, siroheme is a known prosthetic group for assimilatory sulphite and nitrite reductases, which carry out the six electron reduction of sulphite to sulphide and nitrile to ammonia respectively, a process essential for all life on earth. Siroheme is also used as an intermediate in the production of heme d_1 in denitrifying bacteria, giving it dual function as a co-factor in its own right, and also in the synthesis of another key co-factor (Tripathy *et al.*, 2010).

Encompassing just three synthetic steps, the biosynthesis pathway of siroheme is the simplest of all the known tetrapyrrole derived co-factors. The first two steps of the pathway involve the modification of the tetrapyrrole ring to make sirohydrochlorin (SHC), whereas the final step involves the insertion of iron into SHC to make siroheme, catalysed by a class II ferrochelatase (**Figure 6.1.1**) (Warren *et al.*, 1994).

Despite the limited number of steps in siroheme synthesis, there are no reports of how the two closely related denitrifying bacteria *Paracoccus denitrifcans* and *Paracoccus pantotrophus*, make siroheme *de novo*. These organisms have the dual need for siroheme as a co-factor in the nitrite reductases and as an intermediate for heme d_1 synthesis, thus it is of crucial importance to understand the pathway by which they are able to make

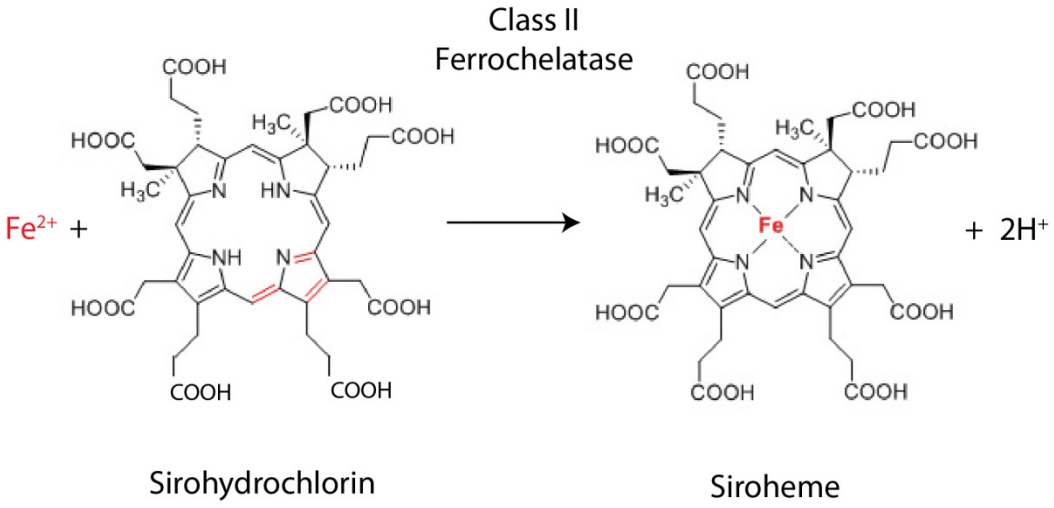


Figure 8.1.1 The terminal step in the siroheme biosynthesis pathway.

The modified tetrapyrrole, sirohydrochlorin, has an Fe^{2+} ion inserted in to the center of the ring to produce siroheme; a reaction catalyzed by a Class II ferrochelatase enzyme.

siroheme. There is a characterised candidate enzyme for the first of the two steps which modify the tetrapyrrole, but the identities of the enzymes involved in both the second step of the modification, and the class II chelatase that carries out the final step of metal chelation into the tetrapyrrole (**Figure 8.1.1**), are not known.

The class II chelatases can specifically insert a divalent metal ion into a range of modified tetrapyrroles; for example cobaltochelatases can insert Co^{2+} (CbiK, $CbiX^{L}$ or $CBiX^{S}$), and ferrochelatases Fe^{2+} (SirB), into SHC during cobalamin and siroheme biosynthesis respectively (Leech, 2002). Other ferrochelatases, such as HemH, act under anaerobic conditions to convert protoporphyrin IX into heme (Hansson & Hederstedt, 1994). A defining feature of this class of enzyme is overall structures of enzymes are highly similar, irrespective of both the metal ion they insert, and the tetrapyrrole substrate used, suggesting evolution from a common ancestor (Fodje *et al.*, 2001).

Some insight into the identity of the enzyme responsible for the final, metal chelation step in *Paracoccus denitrifcans* and *Paracoccus pantotrophus* siroheme synthesis was provided by inspection of the genome of *P. denitrificans*, in which there is a gene annotated as CbiX. In other bacteria, the CbiX enzyme acts as a cobaltochelatase to sirohydrochlorin (SHC), in the process of making vitamin B_{12} anaeobically (Leech, 2002). As *P. denitrificans* is known to make this vitamin aerobically by a separate pathway, it was speculated that this gene does not act as a cobaltochelatse in this way. Furthermore, as this gene is situated downstream of the operon of genes involved in the synthesis of heme d_1 via a siroheme intermediate, it was hypothesised that CbiX may play the role of the ferrochelastse in the terminal step of siroheme production (Shilpa Bali, Department of Biochemistry, University of Oxford). To these ends, CbiX from the related bacterium *P. pantotrophus* was characterised both functionally and structurally.

Functional work (carried out by Dr. Shilpa Bali, Department of Biochemistry, University of Oxford) involving the complementation of an *E. coli CysG* mutant strain (a multifunctional enzyme with the ability to act as a ferrochelatase (Stroupe *et al.*, 2003)) showed that CbiX does indeed act as a ferrochelatase, rather than a cobaltochelatase. Subsequently, the structure of CbiX from *P. pantotrophus* was determined to assess its membership of the structurally related Type II metal chelatase family of enzymes. Complementation studies with point mutants of CbiX were subsequently carried out based on residues implicated from sequence alignments and the structure to allow the role that each may play in the reaction mechanism to be dissected.

Expression, purification, crystallisation and structure determination of *P. pantotrophus* CbiX

The gene for CbiX from *P. pantotrophus* was cloned into pET14b with a NT his-tag (Dr. Shilpa Bali). Over expression of CbiX and its variants was performed in the *E. coli* strain

BL21 codonplus (RIPL) (Stratagene). Cells were grown at 37 °C in 500 ml volumes of Luria Bertani (LB) broth in 2L flasks from overnight starter cultures to an OD_{600} of 0.8−1.0 and transferred to 18 °C before induction with 0.1 mM IPTG. After further incubation for 16 hours the cells from 2 l culture were harvested and re-suspended in 25 ml of Buffer A (50 mM Tris-HCl, 200 mM NaCl, pH 7.5), containing a trace amount of DNaseI (Sigma) and EDTA-free protease inhibitor tablet (Roche). The total cell lysate were obtained by sonication on ice for 5 min and removing the insoluble material by centrifuging at 15,000g for 40 min. The total cell lysate was applied to 5 ml of Chelating Sepharose fast flow column (5 mL) (GE, Healthcare), previously charged with $NiCl_2$ and equilibrated with Buffer A. The column was first washed with 3 column volumes of Buffer A followed by washes with 10 column volumes of 50 mM Tris-HCl, 200 mM NaCl, 25mM Imidazole (pH 7.5), and the protein was eluted with 50 mM Tris-HCl (pH 8.0), 400 mM imidazole. Protein containing fractions were concentrated to 5ml (30kDa MW cut off concentrators, Millipore) and injected onto a s75 16/60 size exclusion column (GE healthcare) pre equilibrated in 50mM Tris pH 7.5, 200mM NaCl and 0.5mM DTT. The fractions from the column were pooled, concentrated as before and the purity of the final was checked by running 15% SDS-PAGE gels **(Figure 8.1.2 A).** Multi angle laser light scattering (MALLS) analysis confirmed the protein was homogenous and monomeric (Steven Johnson, Sir William Dunn School of Pathology) (**Figure 8.1.2 B**).

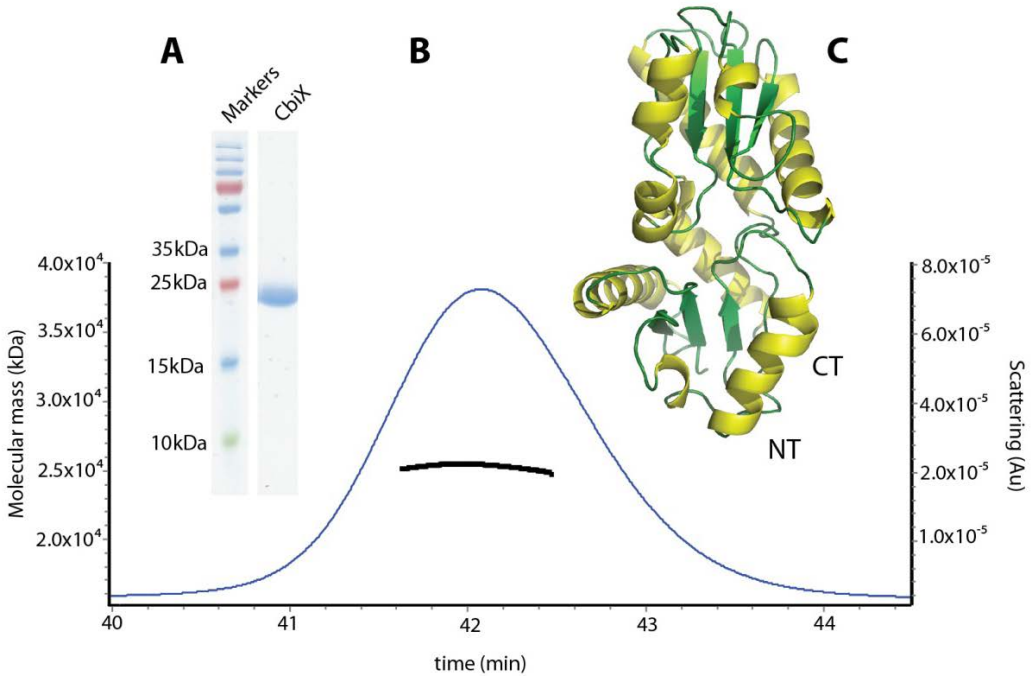


Figure 8.1.2 Biophysical characterization of CbiX from *P. pantotrophus*.

(A) 15% (w/v) SDS-PAGE of purified CbiX. **(B)** MALLS analysis of 100µM CbiX, where the blue line gives the UV trace, and the black line the calculated molecular mass at each point across the peak. (**C**) The crystal structure of CbiX, with loop regions in light green, β strands in dark green and helices in yellow.

Purified CbiX was concentrated to 22 mg ml^{-1} using 10 kDa MW cut off centrifugal concentrator devices (Millipore). Crystallisation was carried out using the sitting drop vapour diffusion method at 21°C. Crystals were obtained in 0.2 M imidazole malate pH 5.5, 24% PEG 600 at a ratio of 50:50 ratio of protein:mother liquor and 400 nl drop volume. These were flash frozen in liquid nitrogen in a solution of mother liquor and 10% ethylene glycol.

Native crystal data were collected at the ESRF, beamline ID29 (λ = 0.87Å). The best crystals diffracted to 1.9 Å resolution and belonged to the space group $P6_122$, with a

MATTHEWS (Matthews, 1968) estimate of 1 copy/ASU and an estimated solvent content of 54.1%. For phasing, crystals were soaked in a solution of mother liquor plus 5% v/v of a saturated K_2PtCl_4 stock solution for 5 minutes. Crystals were then harvested and flash frozen in the same solution as for the native crystals. Data were collected at the Diamond light source, beamline I04-1 (λ=0.92Å). Data were processed using Xia2 pipeline (Winter, 2010), and the data processing statistics for both these data sets are given in **Table 8.1.1**.

For the derivative data, autoSHARP (Vonrhein *et al.*, 2007) automated heavy atom finder was used to identify the first Platinum (Pt) site. The hand was determined at this point by

Data Processing	**Native**	**Pt derivative**
Space group	$P6_122$	$P6_122$
Cell dimensions		
α, β, ɤ (Å)	130.45, 130.45, 56.85	129.99, 129.99, 56.34
α, b, c	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Wavelength (Å)	0.87	0.92
Resolution (Å)	62.2-1.90 (1.95-1.90)	56.3-2.68 (2.75-2.68)
R_{merge}	0.078 (0.751)	0.094 (0.962)
$I/\sigma I$	24.9 (3.7)	30.1 (6.5)
Completeness (%)	100.0 (100.0)	99.8 (99.2)
Redundancy	14.2 (14.6)	37.0 (34.2)
Anomalous correlations	-	0.899 (0.011)
Anomalous completeness (%)	-	100.0 (99.8)
Anomalous multiplicity	-	20.5 (18.3)
Total observations	327577	306279
Total unique	23026	8272

Table 8.1.1 The data processing statistics for the native and Pt derivatised CbiX datasets. The numbers in brackets are for the highest resolution shell.

SOLOMON (Abrahams & Leslie, 1996) solvent flattening, showing the original hand was the correct one (CC for observed verses solvent flattened E's was 17.5 v 11.7%). Inspection of SHARP LLG maps (de La Fortelle & Bricogne, 1997) then identified two further Platinum sites, which were added to the substructure. This gave final phasing statistics of an overall phasing power of 2.0, which was greater than 1.0 at 3.3 Å resolution, and an overall FOM of 0.37 (acentric) and 0.11 (centric).

The solvent content was then optimised to 52.8%, and Bucanneer (Cowtan, 2006) used to build into the solvent flattened map, followed by rounds of manual refinement in COOT (Emsley & Cowtan, 2004) and autoBUSTER (Bricogne, 2011). The final model is a continuous stretch from residues 3-228. There is one protomer of CbiX in the asymmetric unit, along with one ordered malic acid and one ethylene glycol molecule. The final R/R_{free} values are 19.4%/23.7% with refinement statistics given in **Table 8.1.2.** 99.55% of the residues lie in the favoured region of the Ramachandran plot, and there are no outliers. The structure has a MolProbity score of 0.98, making the structure in the 100th percentile for all structures in the resolution range 1.90Å ($\pm$ 0.25Å).

The structure of *P. pantotrophus* CbiX

The crystal structure reveals the protein has a bi-lobed architecture, in which the two domains are similar to one another and are pseudo-symmetrically arranged. The two domains contain a core comprising 3 β-strands, which is enclosed by 3 α-helices and at least one other small, single turn α-helix (**Figure 8.1.2 C**). The two domains split the structure into NT and CT halves, with the exception of the final CT helix crossing back in towards the NT to complete this domain. The two domains are separated by a shallow cleft, or groove, and the domains are slightly kinked with respect to one another. The observed

Refinement statistics	**Refinement against the native dataset**
Resolution (Å)	65.2-1.9
Number of reflections	22942
R_{work}/R_{free}	19.4/23.7
Number of atoms	
Protein	1674
Water	239
B factors	
Protein	31.8
R.m.s deviations	
Bond lengths (Å)	0.01
Bond angles (^{o})	0.98

Table 8.1.2 Refinement statistics for the final model of CbiX.

α,β Rossman fold is characteristic of the structures available of the Type II chelatase family of metal-chelatase enzymes (Fodje *et al.*, 2001).

Indeed, a DALI search of the PDB using the CbiX coordinates brings up numerous significant hits from chelatases, the top models being from structures of the *Bacillus subtilis* ferrochelatase, HemH, which is the terminal enzyme of the 'classical' heme biosynthesis pathway (Al-Karadaghi *et al.*, 1997) where it acts to insert Fe^{2+} into protporphyrin IX. Despite a low sequence identity of 12.9% (Clustal X), these models overlay with an average RMSD of 2.85 Å over 204 residues.

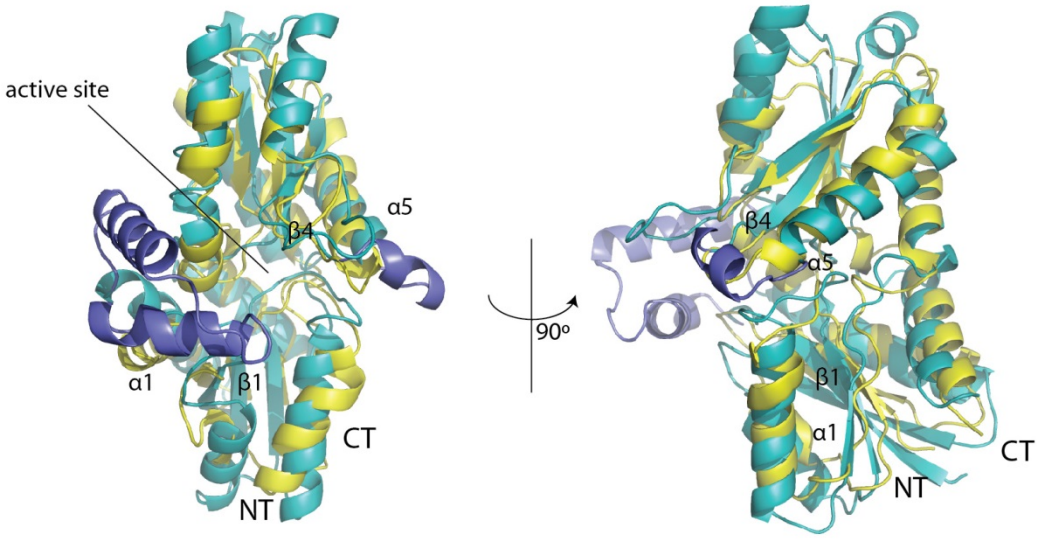


Figure 8.1.3 Structural comparison of *P. pantotrophus* CbiX with *B. subtilis* HemH.

CbiX in yellow and HemH in cyan. The additional helices present on the HemH structure (Al-Karadaghi *et al*., 1997) are coloured in purple. The active site, in the center of the molecule, is labeled.

An overlay of the structure of *P. pantotrophus* CbiX and *B. subtilis* HemH (**Figure 8.1.3**) reveals the highly conserved secondary structural elements in both the NT and CT domains, with CbiX having α-helices and β-sheets of similar length to HemH in approximately similar locations. The major differences between these two structures are in the exact orientation of these elements, CbiX contain helices in slightly skewed orientations compared to those in HemH, and thus they do not overlay completely perfectly. The active site of HemH has been mapped to the cleft in between the two domains, and CbiX maintains a similarly shaped cleft in its structure. The overlay does, however, show that CbiX represents a 'trimmed down' version of HemH, as CbiX lacks two insertions present in HemH. The enzyme, HemH contains two additional helices in the NT, between β1 and α1 of CbiX, and also an extra helix in the CT, between β4 and α5 of CbiX. These extra helices of HemH protrude outwards from the main body of the structure,

and are coloured in slate in **Figure 8.1.3**. This leaves the active site of CbiX much less 'buried' within the structure than that of HemH

Model for substrate binding and ferrochelatase activity

Ferro- and cobaltochelatases have been crystallised with a variety of substrate and product analogues, (both metalled and non-metalled) and with metal ions alone bound in the active site (Al-Karadaghi *et al.*, 1997, Romão *et al.*, 2011, Hansson *et al.*, 2006). These structures have provided a wealth of information regarding the mechanism by which the Type II chelatases function, identifying key residues involved in each of the steps of the enzyme mechanism.

The mechanism for metal ion chelation is thought to proceed via the binding of the metal ion in the active site, and subsequent binding of the tetrapyrrole substrate. The substrate binds in such a manner to cause distortion of the porphryin ring (Lecerof *et al.*, 2000), which then allows the perfectly placed ion to ligate into the center of the ring (Al-Karadaghi *et al.*, 2006). The residues involved in co-ordinating to the metal in both the metal only and metalled-tetrapyrrole state are therefore of crucial importance to the enzyme mechanism.

The attempts to co-crystallise *P. pantotrophus* CbiX with SHC (which is extremely oxygen sensitive and thus must be handled in a glove box) were unsuccessful. Therefore, in an attempt to try and identify the key residues in this enzyme, the structure was compared to those of other chelatases bound to various ligands. The crystal structures of interest to this study are from the ferrochelatse from *B. subtilis* (HemH), which has been crystallised with both a metalled tetrapyrrole (Lecerof *et al.*, 2000) and with a metal ion only (Hansson *et al.*, 2006). The cobalt chelatase from *Salmonella enterica* (Se-CbiK) has been crystallised

with a metalled tetrapyrrole, and *Desulfovibrio vulgaris* CbiK has been crystallised with the metal only (Romão *et al.*, 2011). A list of the residues both structurally and functionally implicated in either co-ordination to the metal in the terrapyrole, or in the metal only bound state, are shown in **Table 8.3**. These structures formed the basis for sequence and structural comparison to CbiX, to identify its potentially important functional residues. These are also given in **Table 8.3**.

Table 8.1.3 shows both the ferrochelatase HemH and the colbalt chelatase, CbiK, use two residues to co-ordinate to the metal, in both the metal only and the metalled-tetrapyrrole state. One of these is common in both states, the His 183 residue in HemH and the His 145 residue of CbiK; giving this residue a dual functional role.

Co-ordinates to metal in the:	**HemH (Ferro chelatase)**	**CbiK (Cobalto chelatase)**	**Structurally and sequence equivalent residue in CbiX**	**Functionally important in CbiX ?**
Tetrapyrrole ring	Tyr 13	Phe 10	His 10	No
			Gln 11	No
Tetrapyrrole ring and metal only state	His 183	His 145	His 127	Yes
Metal only state	Glu 264	His 216	Asp 190	No
			His 187	Yes

Table 8.1.3 Residues shown to be involved in metal co-ordination. Residues demonstrated for the ferrochelatase *Bacillus subtilis* HemH (Lecerof *et al.*, 2000) (Hansson *et al.*, 2006) and the cobalt chelatase CbiK (*Salmonella enterica* and *Desulfovibrio vulgaris* (Romão et al., 2011)). The structurally and sequence equivalent residues in *P. pantotrophus* CbiX are given, along with their functional importance assessed by the complementation studies.

Co-ordination to the metalled tetrapyrrole

In order to understand how a metalled tetrapyrrole may bind to CbiX, a tetrapyrrole substrate was modeled into the active site of *P. pantotrophus* CbiX by using the PDB co-ordinates for the HemH ferrochelatase from *B. subtilis* in complex with a metalled product analogue, N-MeMP-Mg^{2+} (Lecerof *et al.*, 2000) (**Figure 8.1.4 A**). The model demonstrates where the substrate sits in *P. pantotrophus* CbiX, in the active site cleft created by the gap between the two domains.

In order to compare the details of the active site residues and architecture present in *P. pantotrophus* CbiX to how the HemH ferrochelatase binds to the metal in the tetrapyrrole N-MeMP-Mg^{2+}, an overlay of the two structures in the active site region is shown in **Figure 8.1.4 B.** An overlay was also done between *P. pantotrophus* CbiX and the typical cobaltochelatase CbiK from *S. enterica* (Se-CbiK) bound to metalled SHC (Romão *et al.*, 2011) in this same region of the active site (**Figure 8.1.4 C**).

The two HemH residues, highlighted and labeled in cyan in **Figure 8.1.4 B,** that have been shown to be functionally involved in binding to the metal in the tetrapyrrole are His 183 and Tyr 13. These two equivalent residues (shown in **Figure 8.1.4 C)** in Se-CbiK are His 145 and Phe 10. Residues present in the structure of *P. pantotrophus* CbiX at the equivalent structural positions to these are His 127 at the first position, and His 10 and Glu 12 at the second position; these are labeled in yellow. Thus, these residues of CbiX may putatively be involved in binding the metal in the tetrapyrrole.

To test these residues for functionality, mutations were assayed for their ability to complement the *E. coli CysG* knockout strain (Dr. S. Bali, Department of Biochemistry, University of Oxford). This data showed that the His 127 residue is essential for the ferrochelatase activity. This was not surprising, given His 127 of CbiX overlays exactly

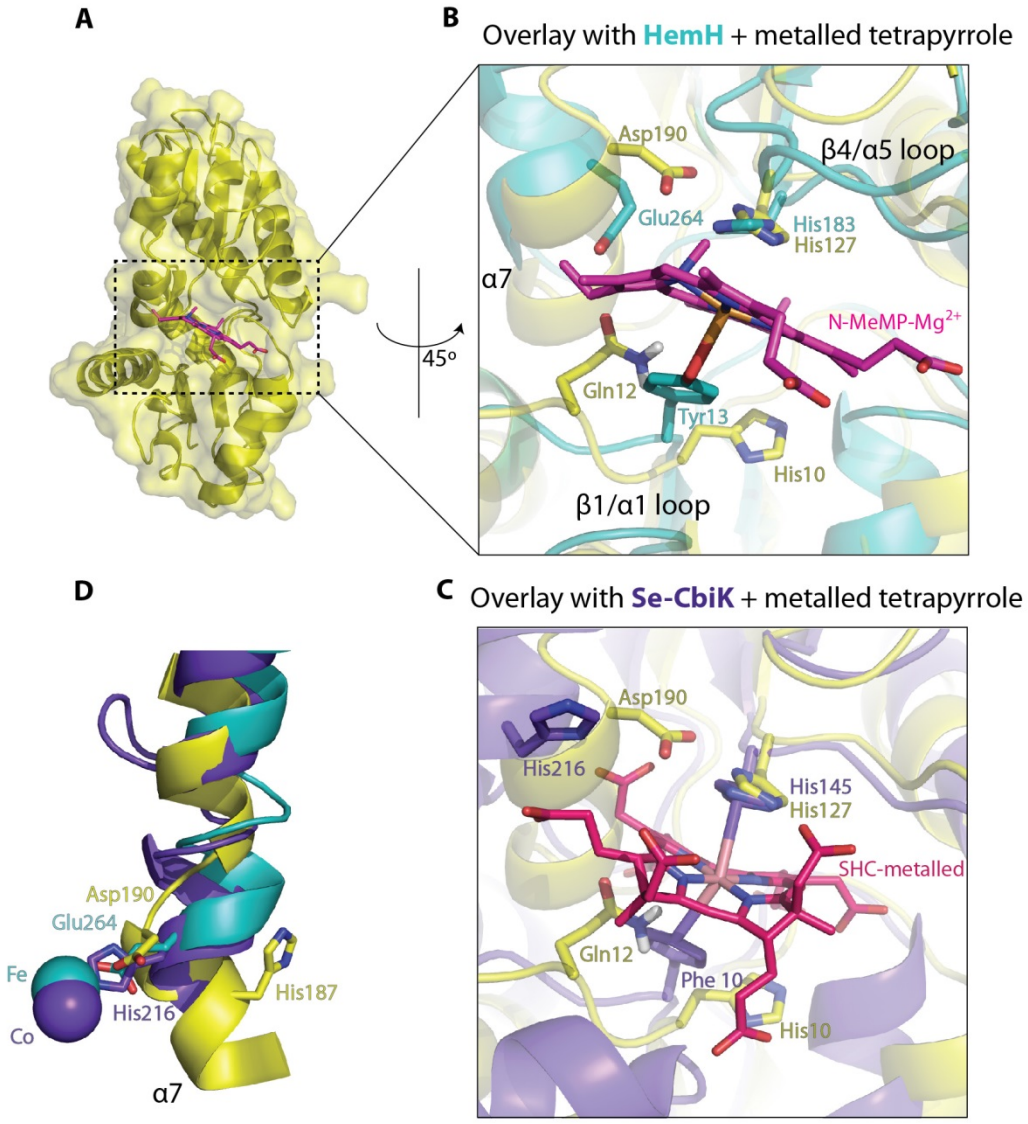


Fig 8.1.4. Investigating the active site of *P. panatrophus* CbiX.

(A) A ribbon and surface representation of *P. panatrophus* CbiX (yellow), showing a product analogue in the active site (pink), based on the co-crystal structure from *B. subtilis* HemH. (**B)** A zoomed in view of the active site, *P. panatrophus* CbiX shown in yellow and *B. subtilis* HemH in cyan. The product analogue, N-MeMP-Mg^{2+} is shown in pink. Residues important in HemH function, and structurally equivalent CbiX residues are shown as pink sticks. (**C)** The same view as in (**B)** but overlaid with CbiK (Salmonella) and SHC. (**D)** Helix 7, which contains a second residue important for metal ion binding in the substrate-free structures of *B. subtilis* HemH-Fe (cyan) and the *D. vulgaris* CbiKp-Co (purple), are shown co-crystallised with their metal ions (spheres), overlaid with *P. panatrophus* CbiX (yellow).

with that of His 183 from HemH (from *B. subtilis*) and with His 145 of CbiK (from *S. enterica*) in the binding pocket; thus the highly conserved architecture of this residue in CbiX and the two ferro- and cobaltochelatases implies it is of great functional significance. Furthermore, this residue has a dual role, as in the metal only bound structures of HemH and CbiK, this residue is also shown to bind the free metal. This implies His 127 of CbiX may act in the similar way.

The second ligand coordinating the metal in the tetrapyrrole in the ferro- (Tyr 13) and cobaltochelatase (Phe 10) is shown in **Figure 8.1.4 B** and **Figure 8.1.4 C**. Both these residues are present on a loop in the NT domain of the structure, in between the first β-strand and the first α-helix (β1/α1 loop), and points up into the active site cleft.

CbiX from *P. pantotrophus* does not have a Tyr or a Phe residue at this position; instead it has a histidine (His 10) and a glutamine (Gln 12) present on the equivalent loop, and they both point upwards towards the active site. We speculated that one of these two residues could be acting in a similar role to the Tyr13 in HemH, and so these were both mutated and tested for their enzyme activity. Unfortunately, mutants in both of these residues were able to completment the CysG knockout in *E. coli*, suggesting that they both are not essential for the chelatase function. This also indicates that neither of these residues are fulfilling the role of Tyr 13 in *P. pantotrophus* CbiX, therefore, this protein must instead have another residue capable of binding the metal in the product and abstracting hydrogen(s) for the chelatase function. There are two other conserved glutamines at positions 19 and 20 in the primary amino acid sequence of CbiX and these might be involved in the similar role, however, these were not yet tested for their *in vivo* function.

Binding the metal in the metal only bound state

Metal binding is likely to precede the binding of the porphyrin ring, and can be assessed by the inspection of crystal structures of HemH (Hansson *et al.*, 2006) in complex with Fe^{2+} and CbiK in complex with Co^{2+} ion only (Romão *et al.*, 2011).

As described earlier, these structures revealed one of the two residues involved in binding the free metal in both the HemH and the CbiK is also involved in binding to the metal in the metalled tetrapyrrole, His183 (HemH) and His154 (CbiK).

The identity of the second residue in HemH was shown to be Glu264, **Figure 8.1.4 D** demonstrates this ligand coordinating to the metal ion (this shows the end of helix α7 from **B** and **C** twisted round). For the cobalt chelatase CbiK, the His216 residue coordinates to the Co^{2+}. In the *P. pantotrophus* CbiX structure there is an aspartate residues, Asp 190, which is present very near to this position in structure (**Figure 8.1.4 B** and **C**), which points towards the potential position where the metal would be coordinated in the HemH and CbiK structures. Therefore, a variant of this Asp residue was tested to check for its *in vivo* function in *P. pantotrophus* CbiX. However, the complementation assay showed that Asp190 is not essential for function. This therefore implies that Asp190 residue, despite its similar chemical characteristics and location in the active site in our structure, does not play a part in this crucial, metal binding role.

Based on the multiple sequence alignments, it was speculated that the nearby residue His 187 or Glu 186 of CbiX may play an equivalent role in its chelatase activity, and the *in vivo* analysis confirmed that at least His 187 is indeed crucial. The residue His 187 may therefore play a role in metal binding. This makes CbiX from *P. pantotrophus* unusual in having a His residue perform this role, rather than the Glu or Asp residue(s) in other ferro-chelatases.

In contrast, the related cobaltochelatases, commonly use His in this role. In the co-crystal structure of CbiK from *D. vulgaris* with Co^{2+} bound at the active site (purple, **Figure 8.1.4 D**) which shows His 216 residue is capable of coordinating to the metal ion (Romão *et al.*, 2011).

Overlaying this helix of the CbiX structure with CbiK-Co^{2+} and HemH-Zn^{2+}, reveals CbiX residue His 187 is around the other side of the helix, and actually points away from the active site (**Figure 8.1.4 D**). In order to explain the functional data, it is likely that this helix must undergo a reasonably large scale re-arrangement, in order to allow this His 187 to come round into the active site and thus bind the metal. One explanation might be that in our CbiX structure, as in the available structures from both ferro- and cobalt- chelatases, this helix is discontinuous with a small kink just before these residues. This may allow the helix to tilt round upon metal binding, and thus allow His187 residue to point into the active site and coordinate to the metal. Or it is likely that this structure represents the inactive, resting state of the enzyme.

In conclusion, the CbiX protein from *P. pantotrophus,* previously annotated as a cobaltochelatase, has been characterized as a ferrochelatase. The solution of the crystal structure revealed that despite its low sequence conservation to the class II chelatases, it indeed has the characteristic structure and active site seen in this class. By using the structures of other members of this class in complex with metals and metalled tetrapyrroles, the mechanistically important residues have been identified, and their functional importance confirmed by *in vivo* assays. These experiments has revealed how the denitrifying bacteria are able to carry out the terminal step in siroheme synthesis, a co-factor required not only in their nitrite reductases enzymes in order to assimilate nitrogen, but also as an intermediate for heme d_1 synthesis.

Appendix 8.2: SepL Resembles an Aberrant Effector in Binding to a Class 1 Type III Secretion Chaperone and Carrying an N-Terminal Secretion Signal

For this paper, molecular biology was carried out to create constructs of SepL, SepD and CesL suitable for co-purification studies. Combinations of the proteins were purified, and MALLS used to assess the stoichiometry of the complexes formed.

Appendix 8.3: Construct directory

All constructs listed, with the exception of that in **Table 8.3.1**, were created as part of this study.

```
        10         20         30         40         50         60
MPLTEHLREL RYRLIISIIA FLIGSGIAFY FAKYVFEILK EPILKSYPEV ELITLSPTEP
        70         80         90        100        110        120
LFILIKISLA VGFIIASPVI LYQFWRFIEP ALYSHEKRAF IPLLLGSILL FMLGALFAYF
       130        140        150        160        170        180
IVLPLALKFL LGLGFTQLLA TPYLSVDMYI SFVLKLVVAF GIAFEMPIVL YVLQKAGVIT
       190        200        210        220        230        240
PEQLASFRKY FIVIAFVIGA IIAPDVSTQV LMAIPLLLLY EISIFLGKLA TRKKKEIQKA
       250        260        270        280        290        300
PGSENLYFQG QFSKGEELFT GVVPILVELD GDVNGHKFSV SGEGEGDATY GKLTLKFICT
       310        320        330        340        350        360
TGKLPVPWPT LVTTLTYGVQ CFSRYPDHMK RHDFFKSAMP EGYVQERTIS FKDDGNYKTR
       370        380        390        400        410        420
AEVKFEGDTL VNRIELKGID FKEDGNILGH KLEYNYNSHN VYITADKQKN GIKANFKIRH
       430        440        450        460        470        480
NIEDGSVQLA DHYQQNTPIG DGPVLLPDNH YLSTQSALSK DPNEKRDHMV LLEFVTAAGI
       490        500
THGMDELYKK LAAHHHHHHH H
```

Table 8.3.1 *A. aeolicus* TatC-GFP$_{His}$ in the Xho1 and BamH1 sites of the pWaldo vector. (Source, Dr. J. Graham and Donna Byrne previously of the Department of Biochemistry, University of Oxford). The TatC start Met is highlighted in red, the TEV cleavage site, at the CT of TatC, is highlighted in blue, and the His tag (at the CT of the GFP) is highlighted in yellow. The TatC sequence is in black text, and the GFP in green.

```
        10         20         30         40         50         60
MPLTEHLREL RYRLIISIIA FLIGSGIAFY FAKYVFEILK EPILKSYPEV ELITLSPTEP
        70         80         90        100        110        120
LFILIKISLA VGFIIASPVI LYQFWRFIEP ALYSHEKRAF IPLLLGSILL FMLGALFAYF
       130        140        150        160        170        180
IVLPLALKFL LGLGFTQLLA TPYLSVDMYI SFVLKLVVAF GIAFEMPIVL YVLQKAGVIT
       190        200        210        220        230        240
PEQLASFRKY FIVIAFVIGA IIAPDVSTQV LMAIPLLLLY EISIFLGKLA TRKKKEIQKA

PGSENLYFQ
```

Table 8.3.2 Highlighted in green, the positions of the single Ala point mutations introduced into the *A. aeolicus* TatC-GFP$_{His}$ construct. Mutants produced using the primers described in **Chapter 4.3.**

Fusion construct	Resulting *A. aeolicus* TatC sequence
NT9	MLSRRQFIQAPLTEHLRELRYRLIISIIAFLIGSG IAFYFAKYVFEILKEPILKSYPEVELITLSPTEPLFILIKISLAVGFIIASPVILYQFWR FIEPALYSHEKRAFIPLLLGSILLFMLGALFAYFIVLPLALKFLLGLGFTQLLATPYLSV DMYISFVLKLVVAFGIAFEMPIVLYVLQKAGVITPEQLASFRKYFIVIAFVIGAIIAPDV STQVLMAIPLLLLYEISIFLGKLATRKKKEIQKAPGSENLYFQGQ
NT16	MLSRRQFIQASGIALAAPLTEHLRELRYRLIISIIAFLIGSG IAFYFAKYVFEILKEPILKSYPEVELITLSPTEPLFILIKISLAVGFIIASPVILYQFWR FIEPALYSHEKRAFIPLLLGSILLFMLGALFAYFIVLPLALKFLLGLGFTQLLATPYLSV DMYISFVLKLVVAFGIAFEMPIVLYVLQKAGVITPEQLASFRKYFIVIAFVIGAIIAPDV STQVLMAIPLLLLYEISIFLGKLATRKKKEIQKAPGSENLYFQGQ
CT6	MPLTEHLRELRYRLIISIIAFLIGSGIAFYFAKYVFEILKEPILKSYPEVELITLSPTEP LFILIKISLAVGFIIASPVILYQFWRFIEPALYSHEKRAFIPLLLGSILLFMLGALFAYF IVLPLALKFLLGLGFTQLLATPYLSVDMYISFVLKLVVAFGIAFEMPIVLYVLQKAGVIT PEQLASFRKYFIVIAFVIGAIIAPDVSTQVLMAIPLLLLYEISIFLGKLATRKKKEIQKA LSRRQFPGSENLYFQGQFS
CT9	MPLTEHLRELRYRLIISIIAFLIGSGIAFYFAKYVFEILKEPILKSYPEVELITLSPTEP LFILIKISLAVGFIIASPVILYQFWRFIEPALYSHEKRAFIPLLLGSILLFMLGALFAYF IVLPLALKFLLGLGFTQLLATPYLSVDMYISFVLKLVVAFGIAFEMPIVLYVLQKAGVIT PEQLASFRKYFIVIAFVIGAIIAPDVSTQVLMAIPLLLLYEISIFLGKLATRKKKEIQKA LSRRQFIQAPGSENLYFQGQFS
CT15	MPLTEHLRELRYRLIISIIAFLIGSGIAFYFAKYVFEILKEPILKSYPEVELITLSPTEP LFILIKISLAVGFIIASPVILYQFWRFIEPALYSHEKRAFIPLLLGSILLFMLGALFAYF IVLPLALKFLLGLGFTQLLATPYLSVDMYISFVLKLVVAFGIAFEMPIVLYVLQKAGVIT PEQLASFRKYFIVIAFVIGAIIAPDVSTQVLMAIPLLLLYEISIFLGKLATRKKKEIQKA AAAAMSLSRRQFIQAPGSENLYFQGQFS

Table 8.3.3 The TatC-signal peptide fusion constructs of *A. aeolicus* TatC-GFP$_{His}$ created in pWaldo. The TatC sequence is shown in black text. The start Met is highlighted in red, the sequence of the fused signal peptide highlighted in purple, and the TEV cleavage site in blue. The sequence of the GFP at the CT side of the TEV site is omitted for clarity.

Construct length	*A. aeolicus* TatB sequence
FL	10 20 30 40 50 60 MELQILLILI IAFIFLGPEG MLNVATKLGE LARQARELID QLRMEAYMEE FNKKILEEEK 70 80 90 100 110 RMKEEEAPPE DIAEELKEEL EDVLEEEKEE KKEDESGKTP GNASNGTSER A
1-94	10 20 30 40 50 60 MELQILLILI IAFIFLGPEG MLNVATKLGE LARQARELID QLRMEAYMEE FNKKILEEEK 70 80 90 RMKEEEAPPE DIAEELKEEL EDVLEEEKEE KKED
1-64	10 20 30 40 50 60 MELQILLILI IAFIFLGPEG MLNVATKLGE LARQARELID QLRMEAYMEE FNKKILEEEK RMKE
1-25	10 20 MELQILLILI IAFIFLGPEG MLNVA
1-17	10 MELQILLILI IAFIFLG

Table 8.3.4 The sequence of the 5 different lengths of *A. aeolicus* TatB fragments created in this study. Primers, and cloning of these fragments in the two vectors (pACYCDuet-1 and CDFDuet-1) are given in **Appendix 8.5.2.**

CT modification	Sequence
Strep Tag	WSHPQFEK
3C-Strep Tag	LEVLFQGP WSHPQFEK

Table 8.3.5 The sequence of the Strep-tag and the 3C-Strep-tag CT modifications to each of the 5 length constructs of *A. aeolicus* TatB, in each of the two vectors (pACYCDuet-1 and CDFDuet-1).Primer and mutagenesis protocol to create these modifications are given in **Appendix 8.5.2 .**

8.4 Appendix: Standard Laboratory procedures

All competent cells were transformed using the 'heat shock' protocol, given in **Table 1.**

Step	Procedure
1	Thaw 50µl of cells on ice, and add 1µl of plasmid DNA
2	Incubate on ice for 30 mins
3	Incubate at 42^{o}C for 2 mins, before incubating on ice for 2 mins
4	Add 250µl of 2xTY broth, and incubate at 37^{o}C with shaking
5	Plate 100µl of cells onto appropriate antibiotic plate, and incubated at 37^{o}C overnight.

Table 8.4.1 Competent cell transformation protocol

In TatC purifications, the PBS buffer was made according to the recipe given in **Table 2.**

Component	Concentration
NaCl	140mM
KCl	2.7mM
Na_2HP0_4	10mM
KH_2PO_4	1.8mM

Table 8.4.2 Phosphate buffered saline (PBS) recipe used in this study

SDS-PAGE gels were made up as the recipe in **Table 8.4.3**. Samples to run on the gel were prepared by mixing with loading dye (Table 1) at a final concentration of x1. These were boiled for 5 mins at 95^{o}C, before being centrifuged at 13,000 rpm for 2 mins. Samples were loaded, and the SDS-PAGE gel run at 200V in a Mini-PROTEAN Tera

Electrophoresis System (Biorad) in 25mM Tris, 190mM glycine and 1% (w/v) SDS buffer. The gel was then stained with Coomassie stain solution (0.12% (w/v) Coomaisse Brilliant Blue R-250, 40% (v/v) methanol and 10% (v/v) glacial acetic acid) until protein bands became visible, and then de-stained overnight in a solution of 30% (v/v) methanol and 10% (v/v) glacial acetic acid.

Solution	**Component**	**Final Concentration**
1 x loading dye	Tris-HCl (pH 6.8) Glycerol SDS Bromophenol Blue DTT	0.05M 10% (v/v) 2% (w/v) 0.1% (w/v) 10mM
SDS-PAGE (12 or 15 %)	<u>Resolving gel</u> acrylamide/methylene bisacrylamide solution (30% w/v) Tris-HCl pH 8.8 SDS Ammonium Persulphate (APS) Tetramethylethylenediamine (TEMED) <u>Stacking gel</u> acrylamide/methylene bisacrylamide solution (30% w/v) Tris-HCl pH 6.8 SDS APS TEMED	 12 or 15% (w/v) 0.4M 0.1% (w/v) 0.1% (w/v) 0.01% (v/v) 5.1% (w/v) 0.13M 0.1% (w/v) 0.1% (w/v) 0.01% (v/v)

Table 8.4.3 The composition of the SDS-PAGE reducing loading dye and the gel (resolving and stacking).

8.4.1 TEV protease purification

50µl of Rosetta-2 competent cells were transformed (**Table 8.4.1**) with 1µl of TEV_{His} in pET28, and plated on 34µg/ml chloramphenicol and 50g/ml kanamycin LB agar plates, and incubated overnight at 37^{o}C. An overnight culture from a single colony in 100ml of LB was set up at 37oC. A 1:100 dilution was used to inoculate 6L of TB (Fisher) supplemented with 0.4% (v/v) glycerol and 34µg/ml chloramphenicol and 50g/ml kanamycin. Growth at 37oC was carried out until an OD_{600} of 0.6 was reached, at which point the cells were cooled to 20oC and induced with 1mM IPTG overnight.

Cells were harvested at 4,500rpm in a JLA 8.100 rotor, and resuspended in lysis buffer, 50 mM NaPi (pH 7.0), 300 mM NaCl supplemented with DNase, lysozyme (Sigma) and a protease inhibitor tablet (Roche). Cells were lysed by passage through an Emusiflex C5 (Avestin) cell disruptor at 4^{o}C. Cell debris was cleared by centrifugation at 20,000 rpm with a JLA 25.50 rotor for 30 minutes. Cleared lysate was applied to a 5ml Ni-NTA column (Qiagen) pre equilibrated in lysis buffer. This was washed in 5 column volumes (CV) of lysis buffer, followed by 5CV of 20mM imidazole 50 mM NaPi (pH 7.0), 300 mM NaCl. Bound protein was eluted in 350mM imidazole 50 mM NaPi (pH 7.0), 300 mM NaCl, and the protein containing fractions (as shown by a nanodrop reading at 280nm) were pooled to 10ml. These were injected onto a S75 26/60 size exclusion column pre equilibrated in 25 mM NaPi (pH 7.0), 200 mM NaCl, 2mM EDTA and 3mM DTT. Fractions containing the TEV protease, as assessed by SDS-PAGE, were pooled to 1mg/ml, and protein flash frozen in 1ml aliquots in 0.1% (v/v) glycerol.

8.4.2 Alteration to TatC purification protocol for high CMC detergents

For those two detergents with high CMC's; a slight alteration to the protocol described for purification in DDM was implemented. A high working concentration in the purification buffers of 0.8% w/v (Cymal-4) and 0.5% w/v (NM) would require a lot of expensive detergent to be used in the 1L of dialysis buffer for the TEV protease cleavage step. Thus, to remove the imidazole from the TatC-GFP_{His} elutions, to allow the TEV protease to function, and to allow the reaction to be run over a second Ni-NTA column, a more buffer-efficient method was used. This involved injecting 5ml of the pooled elutions onto a S200 16/60 pre washed in SEC buffer. As soon as the TatC-GFP fractions eluted, the column was switched into sodium azide (0.02 % (w/v)). This enabled over $1/5^{th}$ less buffer to be used to remove the imidazole. The TEV protease was then added into the pooled fractions at 4^{o}C, and the subsequent purification continued as in **2.1.**

8.4.3 Western blotting

SDS-PAGE was run, as described previously, but upon completion the gel was blotted, in order to transfer across the protein on the gel to a membrane. This was achieved by placing two extra thick blot papers (Biorad) into transfer buffer (0.05M Tris, 0.04M glycine, 0.04% (w/v) SDS, 20% (v/v) methanol), and onto the lower plate of the TransBlot (Biorad) transfer cell. The membrane (HyBond-P PVDF, GE healthcare), pre-activated for a few minutes in methanol, was then submerged in the transfer buffer, before being placed on top of the filter paper. The gel was removed from the glass plates, placed in transfer buffer for a few seconds, before being placed on the membrane. A final set of two filter paper stacks in transfer buffer were placed on top of the gel. The top plate was placed onto the stack, and the Transblot cell run at 22V for 60 minutes.

The membrane was then subject to immunodetection, according to the protocol in **Table 8.4.4.**

1. Membrane incubation in blocking buffer (PBS, 0.1% (v/v) tween-20 and 2% (w/v) skimmed milk powder) for 1 hr at room temperature, with mixing.
2. Membrane incubation in primary antibody overnight at 4^{o}C.
3. Membrane washed in blocking buffer 3x15 mins at 4^{o}C
4. Membrane incubated in secondary antibody for 1 hour at 4^{o}C.
5. Membrane washed in blocking buffer 3x15 mins at 4^{o}C.
6. Membrane washed once in PBS for 5 mins
7. Membrane incubated in Amersham ECL western blotting analysis system (GE healthcare) for 1 min, and excess removed.
8. Exposed to Amersham Hyperfilm ECL (GE healthcare) the dark room, before developing with the Xograph developing machine.

Table 8.4.4 Immunodetection protocol for electro-transferred proteins.

For Dot-blots, 2µl of protein spotted onto a nitrocellulose HyBond-P membrane (GE healthcare). The membrane was allowed to dry, and then subjected to immunodetection as in **Table 8.4.4**, with a few exceptions; all steps were done at room temperature, and step 2 was only carried out for 1 hour.

The dilutions, in blocking buffer, used for a western blot were 1:750 of the primary antibody (QIA*express Strep*-Tag Antibody, Qiagen) and a 1:1,500 dilution of the secondary antibody (HRP conjugated goat anti-mouse, Thermo Scientific). For a dot blot, the same antibodies were used but the primary dilution was used at 1:1,000.

Agarose gel electrophoresis

DNA samples were combined with 6 x Blue/Orange loading dye (Promega) and loaded onto a 1% (w/v) agarose gel made in TAE buffer (40mM Tris-acetate pH 8.3, 2mM disodium EDTA) and supplemented with 1µg/ml ethidium bromide. The gel was run at 70V in a Mini-Sub Cell GT Cell (Bio-rad) in TAE buffer supplemented with 0.5µg/ml ethidium bromide. Bands were visualised in a UV light box (Gel Doc 2000, Bio-rad).

Appendix 8.5: Molecular biology

8.5.1 Mutagenesis of *A. aeolicus* TatC in pWaldo

Signal peptide binding site mutations

In order to create mutations in the sequence of *A. aeolicus* TatC in pWaldo, the Quikchange II XL method (Agilent Technologies) was used. The supplied Pfu DNA polymerase was used to amplify the vector using the specified primers, according to **Table 8.5.1** using the component recipe specified in the manufactures instructions, and using 10ng of the template plasmid.

Number of cycles	Stage	Temperature (^oC)	Time
1	Initial denaturation	95	60 sec
18	Denaturation	95	50 sec
	Annealing	T_a	50 sec
	Extension	68	$T_{extension}$
1	Final extension	68	10 mins

Table 8.5.1 The Quikchange mutagenesis (Agilent technologies) PCR protocol, where the $T_{annealing}$ (T_a) and $T_{extension}$ were varied depending on the construct.

The primers used for three additional mutations to test for signal peptide binding (**4.3**) with purified TatC are given in **Table 8.5.2.** The protocol in Table 1 was used, with a Ta of 60^oC and a $T_{extension}$ of 7 mins.

Following amplification, the template DNA was digested with Dpn1 at 37^{o}C for 1 hour. Following the addition of β-mercatopethanol (as per manufacturers instructions), the amplified DNA was then transformed in the XL-Gold cells by the 'gentle' transformation proceedure (**Table 8.5.6**) and plated onto 50μg/ml kanamycin LB-agar plates. Colonies were used to inoculate 5ml of an LB overnight culture. Cells were harvested, and DNA prepared (Qiagen mini preparation kit), the correct sequences confirmed by sequencing with the T7 forward primer (Source Bioscience).

***E. coli* residue**	***A. aeolicus* residue**	**Forward primer (5' to 3')**	**Reverse Primer (5' to 3')**
Lys 18	**Tyr 12**	ACG GAA CAT CTG AGA GAG CTT AGG GCA CGT TTA ATA ATA TCT ATA ATT GCC	GGC AAT TAT AGA TAT TAT TAA ACG TGC CCT AAG CTC TCT CAG ATG TTC CGT
Asn 22	**Ile 16**	AGA GAG CTT AGG TAT CGT TTA ATA GCG TCT ATA ATT GCC TTC CTT ATA GGC	GCC TAT AAG GAA GGC AAT TAT AGA CGC TAT TAA ACG ATA CCT AAG CTC TCT
Lys 101	**Ser 94**	TTT ATA GAG CCA GCC CTT TAC GCC CAT GAA AAA AGA GCT TTT ATC	GAT AAA AGC TCT TTT TTC ATG GGC GTA AAG GGC TGG CTC TAT AAA

Table 8.5.2 Primers used to introduce three further alanine mutations of the residues identified with a functional role in binding signal peptides. The template for the mutagenesis was the WT TatC-GFP pWaldo vector.

Signal peptide fusion RR to KK mutations

In order to create the 'control' signal peptide fusions (**5.4**), the RR motif on the five fusion constructs were mutated to KK. The primers used on each of the five constructs are given in **Table 8.5.3**, using the PCR protocol in **Table 8.5.1** with a Ta of 60^{o}C and a $T_{extension}$ of 7 mins.

Construct	**Forward primer (5' to 3')**	**Reverse Primer (5' to 3')**	**Fusion Template**	**Resulting sequence of the construct**
NT9 KK	AAA CTC CCG GAA ATG CTC TCT AAG AAG CAA TTT ATA CAA GCA CCT CTA	TAG AGG TGC TTG TAT AAA TTG CTT CTT AGA GAG CAT TTC CGG GAG TTT	NT9	MLS**KK**QFIQA..(Tat C in pWaldo)..
NT16 KK	AAA CTC CCG GAA ATG CTC TCT AAG AAG CAA TTT ATA CAA GCA TCT GGA	TCC AGA TGC TTG TAT AAA TTG CTT CTT AGA GAG CAT TTC CGG GAG TTT	NT16	MLS**KK**QFIQASGIA LAA..(TatC in pWaldo)…
CT6 KK	GAA ATT CAA AAG GCA CTC TCT AAG AAG CAA TTT CCT GGA TCC GAA AAC	GTT TTC GGA TCC AGG AAA TTG CTT CTT AGA GAG TGC CTT TTG AAT TTC	CT6	..(TatC in pWaldo).. LS**KK**QFPGSENLYF QG
CT9 KK	G GAA ATT CAA AAG GCA CTC TCT AAG AAG CAA TTT ATA CAA GCA CCT GGA T	A TCC AGG TGC TTG TAT AAA TTG CTT CTT AGA GAG TGC CTT TTG AAT TTC C	CT9	..(TatC in pWaldo).. LS**KK**QFIQAPGSEN LYFQG
CT15 KK	GCT GCT GCT ATG TCT CTC TCT AAG AAG CAA TTT ATA CAA GCA CCT GGA	TCC AGG TGC TTG TAT AAA TTG CTT CTT AGA GAG AGA CAT AGC AGC AGC	CT15	..(TatC in pWaldo).. AAAAMSLS**KK**QFI QAPGSENLYFQG

Table 8.5.3 Primers to create the KK mutant versions of the 5 signal peptide fusions, created from the primers in Chapter 4, Table 4. The KK residues are highlighted in red.

8.5.2 Creating *A. aeolicus* TatB constructs for coexpression with *A. aeolicus* TatC in pWaldo

The primers in **Table 8.5.4** were used to amplify fragment of *A. aeolicus* TatB from genomic TatBC in pWaldo according to the touchdown PCR protocol given in **Table 8.5.5.**

TatB construct size	Forward primer (5' to 3')	Reverse primer (5' to 3')	Template
Full length (FL)	GAG ATA TAC CAT GGA ACT TCA GAT ACT GCT CAT TCT TAT AAT AGC TTT TA	CGA ATT CGG ATC CCT AAG CTC TCT CAG ATG TTC CGT TAG AGG CAT T	Genomic TatBC pWaldo
1-94		CGA ATT CGG ATC CCT AAT CTT CCT TCT TCT CCT CCT TTT CTT CTT CTA A	Genomic TatBC pWaldo
1-25		CGA ATT CGG ATC CCT AGG CAA CGT TGA GCA TAC CTT CGG GTC CGA GAA AT	Genomic TatBC pWaldo
1-17		CGA ATT CGG ATC CCT ATC CGA GAA ATA TAA AAG CTA TTA TAA GAA TGA GC	Genomic TatBC pWaldo
1-64	C CTT GAA GAA GAA AAA AGA ATG AAA GAG TAG GGA TCC GAA TTC GAG CTC GGC GCG C	GCG CGC CGA GCT CGA ATT CGG ATC CCT ACT CTT TCA TTC TTT TTT CTT CTT CAA GG	$TatB_{1-94}$ in CDFDuet/pACYC

Table 8.5.4 Primers used to amplify TatB fragments from the genomic TatBC pWaldo vector for insertion into MCS1 of CDFDuet-1 and MCS1 of pACYC. The forward primer contains an Nco1 restriction site, and the reverse a BamH1 restriction site. The 1-64 construct in both vectors was created after insertion of the 1-94 sequence into both vectors, by a Quikchange deletion reaction.

Number of Cycles	Stage	Temperature (^{o}C)	Time
1	Initial denaturation	94	5 mins
10	Denaturation	94	20 sec
	Annealing	55 50	2 sec
	Elongation	72	25 sec
20	Denaturation	94	20 sec
	Annealing	60	2 sec
	Elongation	72	25 sec
1	Final extension	72	10 mins

Table 8.5.5 The touchdown PCR protocol for amplification of the TatB fragments from the genomic *A. aeolicus* TatBC pWaldo vector. The annealing temperature in the 2nd stage was incrementally decreased from 55 to 50^{o}C in 0.5^{o}C steps.

These primers contained an NcoI restriction site on the forward primer, and a BamHI restriction site on the reverse primer, to enable the ligation into the multiple cloning site 1 (MCS-1) of the pACYCDuet-1 (Novagen) and the CDFDuet-1 (Novagen).

Subsequent to the specific PCR amplification of the fragments, they were subject to a PCR purification (Qiagen). These inserts, along with the purified empty vectors (Qiagen mini preparation kit) were subject to a two step double digest, in which NcoI was used firstly, followed by BamHI (Tango, Thermo Scientific) for 15 hours at 37^{o}C, with inactivation at 80^{o}C for 20 minutes. The digestion was carried out on 3ng of DNA with 40U of enzyme, with the buffer as specified in the manufactures instructions.

The inserts were treated with ExoI and DpnI (Fast digest, NEB) and the vectors with Alkaline Phosphatase (Antarctic, NEB) at 37^{o}C for 1hr, with inactivation at 80^{o}C for 20mins. The vectors and inserts were then treated to a second PCR purification (Qiagen).

Ligation was achieved using varying ratios of vector:insert (1:1 to 1:2) at 16^{o}C for 16 hours with T4 DNA ligase (Roche).

DNA was transformed using a 'gentle protocol' into DH5α cells (made in house), according to **Table 8.5.6** and plated on the appropriate antibiotic resistant plates (34 µg/ml chloramphenicol, pACYCDuet-1, and 10µg/ml spectinomycin, CDFDuet-1).

Step	**Procedure**
1	80µl cells thawed on ice, and transferred to an eppendorf tube
2	4µl of ligation mix added and incubated for 30 mins
3	Incubated at 42^{o}C for 45 seconds, and placed on ice for 2 mins
4	450µl of 2xTY broth added, and cells transferred to a 15ml falcon tube for incubation at 37^{o}C for 1 hr
5	Cells harvested in eppendorf tube by centrifugation at 10,000rpm for 10mins at 20^{o}C.
6	Pellet re-suspended in 100µ of 2xTY and plated onto antibiotic plates for incubation at 37^{o}C overnight.

Table 8.5.6 The 'gentle' transformation protocol.

Colonies were picked, and used to inoculate 5ml of an overnight culture. Plasmid DNA was isolated by a mini-preparation kit (Qiagen) and the inserts confirmed by sequence g with the DuetUP-1 primer (Source bioscience).

The 1-64 construct was made subsequently to the FL, 1-94, 1-25 and 1-17 constructs, using the primes in **Table 8.5.4** to delete 90 base pairs from the 1-94 construct by the Quikchange mutagenesis method (Table 1) using a T_a of 53^{o}C and a $T_{extension}$ of 7 mins.

In order to add the 8 amino acid Strep II tag (Strep-tag) onto these constructs, the primers in **Table 8.5.7** were used, according to the protocol in **Table 8.5.1**, with a T_a of 53^{o}C and a $T_{extension}$ of 7 mins.

Again, Strep-tag fusions and the 3C-Strep-tag fusions of the 1-64 construct were made subsequently to those of the FL, 1-94, 1-25 and 1-17 constructs, using the primes in Table 4 and 5 to delete 90 base pairs from the 1-94 Strep-tag and 1-94 3C-Strep-tag constructs by the Quikchange mutagenesis method (Table 1) using a T_a of 53^{o}C and a $T_{extension}$ of 7 mins.

These vectors were then used as the templates for the addition of the 8 amino acid 3C cleavage site, at the NT side of the Strep tag, using the primers in **Table 8.5.8** and the PCR protocol as for the Strep tag fusions.

Streptavidin tag at CT	Forward primer (5' to 3')	Reverse primer (5' to 3')	Template
FL	AAC GGA ACA TCT GAG AGA GCT TGG AGC CAC CCG CAG TTC GAA AAA TAG GGA TCC GAA TTC GAG CTC	GAG CTC GAA TTC GGA TCC CTA TTT TTC GAA CTG CGG GTG GCT CCA AGC TCT CTC AGA TGT TCC GTT	Un-tagged TatB FL in pCDFDuet-1/ pACYCDuet-1
1-94	GAA AAG GAG GAG AAG AAG GAA GAT TGG AGC CAC CCG CAG TTC GAA AAA TAG GGA TCC GAA TTC GAG CTC GGC	GCC GAG CTC GAA TTC GGA TCC CTA TTT TTC GAA CTG CGG GTG GCT CCA ATC TTC CTT CTT CTC CTC CTT TTC	Un-tagged $TatB_{1-94}$ in pCDFDuet-1/ pACYCDuet-1
1-25	GAA GGT ATG CTC AAC GTT GCC TGG AGC CAC CCG CAG TTC GAA AAA TAG GGA TCC GAA TTC GAG CTC	GAG CTC GAA TTC GGA TCC CTA TTT TTC GAA CTG CGG GTG GCT CCA GGC AAC GTT GAG CAT ACC TTC	Un-tagged $TatB_{1-25}$ in pCDFDuet-1/ pACYCDuet-1
1-17	CTT ATA ATA GCT TTT ATA TTT CTC GGA TGG AGC CAC CCG CAG TTC GAA AAA TAG GGA TCC GAA TTC GAG CTC GGC GCG	CGC GCC GAG CTC GAA TTC GGA TCC CTA TTT TTC GAA CTG CGG GTG GCT CCA TCC GAG AAA TAT AAA AGC TAT TAT AAG	Un-tagged $TatB_{1-17}$ in pCDFDuet-1/ pACYCDuet-1
1-64	C CTT GAA GAA GAA AAA AGA ATG AAA GAG TGG AGC CAC CCG CAG TTC GAA AAA TAG G	C CTA TTT TTC GAA CTG CGG GTG GCT CCA CTC TTT CAT TCT TTT TTC TTC TTC AAG G	Strep-tagged $TatB_{1-94}$ in pCDFDuet-1/ pACYCDuet-1

Table 8.5.7 Primers used to insert an 8 amino acid Streptavidin (Strep-tag) at the CT of each TatB construct in CFFDuet-1 and pACYCDuet-1. Insertion was achieved by a Quikchange mutagenesis reaction. For the 1-64 construct, this was again created by a Quikchange deletion reaction from the 1-94 Strep tag construct in both vectors.

3C site prior to Strep-tag at CT	Forward primer (5' to 3')	Reverse primer (3' to 5')	Template
FL	AAC GGA ACA TCT GAG AGA GCT CTG GAA GTT CTG TTC CAG GGG CCC TGG AGC CAC CCG CAG TTC GAA	TTC GAA CTG CGG GTG GCT CCA GGG CCC CTG GAA CAG AAC TTC CAG AGC TCT CTC AGA TGT TCC GTT	Strep-tagged TatB FL in pCDFDuet-1/ pACYCDuet-1
1-94	GAA AAG GAG GAG AAG AAG GAA GAT CTG GAA GTT CTG TTC CAG GGG CCC TGG AGC CAC CCG CAG TTC GAA AAA	TTT TTC GAA CTG CGG GTG GCT CCA GGG CCC CTG GAA CAG AAC TTC CAG ATC TTC CTT CTT CTC CTC CTT TTC	Strep-tagged $TatB_{1-94}$ in pCDFDuet-1/ pACYCDuet-1
1-25	GAA GGT ATG CTC AAC GTT GCC CTG GAA GTT CTG TTC CAG GGG CCC TGG AGC CAC CCG CAG TTC GAA	TTC GAA CTG CGG GTG GCT CCA GGG CCC CTG GAA CAG AAC TTC CAG GGC AAC GTT GAG CAT ACC TTC	Strep-tagged $TatB_{1-25}$ in pCDFDuet-1/ pACYCDuet-1
1-17	CTT ATA ATA GCT TTT ATA TTT CTC GGA CTG GAA GTT CTG TTC CAG GGG CCC TGG AGC CAC CCG CAG TTC GAA AAA TAG	CTA TTT TTC GAA CTG CGG GTG GCT CCA GGG CCC CTG GAA CAG AAC TTC CAG TCC GAG AAA TAT AAA AGC TAT TAT AAG	Strep-tagged $TatB_{1-17}$ in pCDFDuet-1/ pACYCDuet-1
1-64	C CTT GAA GAA GAA AAA AGA ATG AAA GAG CTG GAA GTT CTG TTC CAG GGG CCC TGG A	TCC AGG GCC CCT GGA ACA GAA CTT CCA GCT CTT TCA TTC TTT TTT CTT CTT CAA GG	3C-Strep-tagged $TatB_{1-94}$ in pCDFDuet-1/ pACYCDuet-1

Table 8.5.8 Primers for the generation of a 3C-cleavage site prior to the Streptavidin-tag. Insertion of the 8 amino acid 3C cleavage site, N-terminal to the Streptavidin-tag was achieved by a Quikchange reaction. For the 1-64 construct, this was again created by a Quikchange deletion reaction from the 1-94 3C-Strep-tag construct in both vectors.

Appendix 8.6 The structure of the TatC core of the twin-arginine protein transport system

9.0 References

Abrahams, J. P. & A. G. W. Leslie, (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallographica Section D* **52**: 30-42.

Aizawa, S.-I., (2001) Bacterial flagella and type III secretion systems. *FEMS Microbiology Letters* **202**: 157-164.

Al-Karadaghi, S., R. Franco, M. Hansson, J. A. Shelnutt, G. Isaya & G. C. Ferreira, (2006) Chelatases: distort to select? *Trends in Biochemical Sciences* **31**: 135-142.

Al-Karadaghi, S., M. Hansson, S. Nikonov, B. Jönsson & L. Hederstedt, (1997) Crystal structure of ferrochelatase: the terminal enzyme in heme biosynthesis. *Structure (London, England : 1993)* **5**: 1501-1510.

Alami, M., I. Luke, S. Deitermann, G. Eisner, H. Koch, J. Brunner & M. Muller, (2003) Differential interactions between a twin-arginine signal peptide and its translocase in Escherichia coli. *Mol Cell* **12**: 937 - 946.

Alami, M., D. Trescher, L.-F. Wu & M. Müller, (2002) Separate Analysis of Twin-arginine Translocation (Tat)-specific Membrane Binding and Translocation in Escherichia coli. *Journal of Biological Chemistry* **277**: 20499-20503.

Alder, N. N. & S. M. Theg, (2003) Energetics of Protein Transport across Biological Membranes: A Study of the Thylakoid ΔpH-Dependent/cpTat Pathway. *Cell* **112**: 231-242.

Alexandrov, A. I., M. Mileni, E. Y. T. Chien, M. A. Hanson & R. C. Stevens, (2008) Microscale Fluorescent Thermal Stability Assay for Membrane Proteins. *Structure* **16**: 351-359.

Allen, S. C. H., C. M. L. Barrett, N. Ray & C. Robinson, (2002) Essential Cytoplasmic Domains in the Escherichia coli TatC Protein. *Journal of Biological Chemistry* **277**: 10362-10366.

Arnold, T. & D. Linke, (2001) The Use of Detergents to Purify Membrane Proteins. In: Current Protocols in Protein Science. John Wiley & Sons, Inc., pp.

Asherie, N., (2004) Protein crystallization and phase diagrams. *Methods* **34**: 266-272.

Ashkenazy, H., E. Erez, E. Martz, T. Pupko & N. Ben-Tal, (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research* **38**: W529-W533.

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

Bageshwar, U. K., N. Whitaker, F.-C. Liang & S. M. Musser, (2009) Interconvertibility of lipid- and translocon-bound forms of the bacterial Tat precursor pre-SufI. *Molecular Microbiology* **74**: 209-226.

Baker, N. A., D. Sept, S. Joseph, M. J. Holst & J. A. McCammon, (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences* **98**: 10037-10041.

Bauer, B. W. & T. A. Rapoport, (2009) Mapping polypeptide interactions of the SecA ATPase during translocation. *Proceedings of the National Academy of Sciences* **106**: 20800-20805.

Becker, T., S. Bhushan, A. Jarasch, J.-P. Armache, S. Funes, F. Jossinet, J. Gumbart, T. Mielke, O. Berninghausen, K. Schulten, E. Westhof, R. Gilmore, E. C. Mandon & R. Beckmann, (2009) Structure of Monomeric Yeast and Mammalian Sec61 Complexes Interacting with the Translating Ribosome. *Science* **326**: 1369-1373.

Behrendt, J., K. Standar, U. Lindenstrauß & T. Brüser, (2004) Topological studies on the twin-arginine translocase component TatC. *FEMS Microbiology Letters* **234**: 303-308.

Bendtsen, J., H. Nielsen, D. Widdick, T. Palmer & S. Brunak, (2005) Prediction of twin-arginine signal peptides. *BMC Bioinformatics* **6**: 167.

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

Berks, B. C., (1996) A common export pathway for proteins binding complex redox cofactors? *Molecular Microbiology* **22**: 393-404.

Bessonneau, P., V. Besson, I. Collinson & F. Duong, (2002) The SecYEG preprotein translocation channel is a conformationally dynamic and dimeric structure. *EMBO J* **21**: 995-1003.

Bill, R. M., P. J. F. Henderson, S. Iwata, E. R. S. Kunji, H. Michel, R. Neutze, S. Newstead, B. Poolman, C. G. Tate & H. Vogel, (2011) Overcoming barriers to membrane protein structure determination. *Nat Biotech* **29**: 335-340.

Blaudeck, N., P. Kreutzenbeck, R. Freudl & G. A. Sprenger, (2003) Genetic Analysis of Pathway Specificity during Posttranslational Protein Translocation across the *Escherichia coli* Plasma Membrane. *Journal of bacteriology* **185**: 2811-2819.

Blaudeck, N., P. Kreutzenbeck, M. Müller, G. A. Sprenger & R. Freudl, (2005) Isolation and Characterization of Bifunctional *Escherichia coli* TatA Mutant Proteins That Allow Efficient Tat-dependent Protein Translocation in the Absence of TatB. *Journal of Biological Chemistry* **280**: 3426-3432.

Blow, D. M. & F. H. C. Crick, (1959) The treatment of errors in the isomorphous replacement method. *Acta Crystallographica* **12**: 794-802.

Blow, D. M. & M. G. Rossmann, (1961) The single isomorphous replacement method. *Acta Crystallographica* **14**: 1195-1202.

Bogdanov, M., P. N. Heacock & W. Dowhan, (2002) A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition. *EMBO J* **21**: 2107-2116.

Bogsch, E., S. Brink & C. Robinson, (1997) Pathway specificity for a ΔpH-dependent precursor thylakoid lumen protein is governed by a 'Sec-avoidance' motif in the transfer peptide and a 'Sec-incompatible' mature protein. *EMBO J* **16**: 3851-3859.

Bogsch, E. G., F. Sargent, N. R. Stanley, B. C. Berks, C. Robinson & T. Palmer, (1998) An Essential Component of a Novel Bacterial Protein Export System with Homologues in Plastids and Mitochondria. *Journal of Biological Chemistry* **273**: 18003-18006.

Bolhuis, A., J. Mathers, J. Thomas, C. Barrett & C. Robinson, (2001) TatB and TatC form a functional and structural unit of the twin-arginine translocase from *Escherichia coli*. *J Biol Chem* **276**: 20213 - 20219.

Bonander, N., K. Hedfalk, C. Larsson, P. Mostad, C. Chang, L. Gustafsson & R. M. Bill, (2005) Design of improved membrane protein production experiments: Quantitation of the host response. *Protein Science* **14**: 1729-1740.

Breyton, C., W. Haase, T. A. Rapoport, W. Kuhlbrandt & I. Collinson, (2002) Three-dimensional structure of the bacterial protein-translocation complex SecYEG. *Nature* **418**: 662-665.

Bricogne, G., (1974) Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Crystallographica Section A* **30**: 395-405.

Bricogne, G. e. a., (2011) BUSTER 2.11.2. *(Global Phasing Ltd., 2011)*.

Brundage, L., J. P. Hendrick, E. Schiebel, A. J. M. Driessen & W. Wickner, (1990) The purified *E. coli* integral membrane protein SecY E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* **62**: 649-657.

Brunger, A. T., (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**: 472-475.

Brüser, T. & C. Sanders, (2003) An alternative model of the twin arginine translocation system. *Microbiological Research* **158**: 7-17.

Buchanan, G., E. de Leeuw, N. Stanley, M. Wexler, B. Berks, F. Sargent & T. Palmer, (2002) Functional complexity of the twin-arginine translocase TatC component revealed by site-directed mutagenesis. *Mol Microbiol* **43**: 1457 - 1470.

Bushnell, D. A., P. Cramer & R. D. Kornberg, (2001) Selenomethionine Incorporation in Saccharomyces cerevisiae RNA Polymerase II. *Structure* **9**: R11-R14.

Caffrey, M., (2009) Crystallizing Membrane Proteins for Structure Determination: Use of Lipidic Mesophases. *Annual Review of Biophysics* **38**: 29-51.

Caffrey, M. & V. Cherezov, (2009) Crystallizing membrane proteins using lipidic mesophases. *Nat. Protocols* **4**: 706-731.

Caffrey, M., D. Li & A. Dukkipati, (2012) Membrane Protein Structure Determination Using Crystallography and Lipidic Mesophases: Recent Advances and Successes. *Biochemistry* **51**: 6266-6288.

Celedon, J. M. & K. Cline, (2012) Stoichiometry for binding and transport by the twin arginine translocation system. *The Journal of Cell Biology* **197**: 523-534.

Celedon, J. M. & K. Cline, (2013) Intra-plastid protein trafficking: How plant cells adapted prokaryotic mechanisms to the eukaryotic condition. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1833**: 341-351.

Chae, P. S., S. G. F. Rasmussen, R. R. Rana, K. Gotfryd, R. Chandra, M. A. Goren, A. C. Kruse, S. Nurva, C. J. Loland, Y. Pierre, D. Drew, J.-L. Popot, D. Picot, B. G. Fox, L. Guan, U. Gether, B. Byrne, B. Kobilka & S. H. Gellman, (2010) Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nat Meth* **7**: 1003-1008.

Chatzi, K. E., M. F. Sardis, S. Karamanou & A. Economou, (2013) Breaking on through to the other side: protein export through the bacterial Sec system. *Biochemical Journal* **449**: 25-37.

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

Cline, K., W. F. Ettinger & S. M. Theg, (1992) Protein-specific energy requirements for protein transport across or into thylakoid membranes. Two lumenal proteins are transported in the absence of ATP. *Journal of Biological Chemistry* **267**: 2688-2696.

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

Cline, K. & H. Mori, (2001) Thylakoid ΔpH-dependent precursor proteins bind to a cpTatC-Hcf106 complex before Tha4-dependent transport. *The Journal of Cell Biology* **154**: 719-730.

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

Crane, B. R., L. M. Siegel & E. D. Getzoff, (1997) Structures of the Siroheme- and Fe_4S_4-Containing Active Center of Sulfite Reductase in Different States of Oxidation: Heme Activation via Reduction-Gated Exogenous Ligand Exchange. *Biochemistry* **36**: 12101-12119.

Crichton, P. G., M. Harding, J. J. Ruprecht, Y. Lee & E. R. S. Kunji, (2013) Lipid, detergent and Coomassie dye affect the migration of small membrane proteins in

blue native gels: mitochondrial carriers migrate as monomers not dimers. *Journal of Biological Chemistry*.

Cristobal, S., J.-W. de Gier, H. Nielsen & G. von Heijne, (1999) Competition between Sec- and TAT-dependent protein translocation in Escherichia coli. *EMBO J* **18**: 2982-2990.

Dabney-Smith, C. & K. Cline, (2009) Clustering of C-Terminal Stromal Domains of Tha4 Homo-oligomers during Translocation by the Tat Protein Transport System. *Molecular Biology of the Cell* **20**: 2060-2069.

Dabney-Smith, C., H. Mori & K. Cline, (2003) Requirement of a Tha4-conserved Transmembrane Glutamate in Thylakoid Tat Translocase Assembly Revealed by Biochemical Complementation. *Journal of Biological Chemistry* **278**: 43027-43033.

Dabney-Smith, C., H. Mori & K. Cline, (2006) Oligomers of Tha4 Organize at the Thylakoid Tat Translocase during Protein Transport. *Journal of Biological Chemistry* **281**: 5476-5483.

De Buck, E., E. Lammertyn & J. Anné, (2008) The importance of the twin-arginine translocation pathway for bacterial virulence. *Trends in Microbiology* **16**: 442-453.

de La Fortelle, E. & G. Bricogne, (1997) [27] Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In: Methods in Enzymology. Charles W. Carter, Jr. (ed). Academic Press, pp. 472-494.

de Leeuw, E., T. Granjon, I. Porcelli, M. Alami, S. B. Carr, M. Müller, F. Sargent, T. Palmer & B. C. Berks, (2002) Oligomeric Properties and Signal Peptide Binding by Escherichia coli Tat Protein Transport Complexes. *Journal of Molecular Biology* **322**: 1135-1146.

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

DeLisa, M. P., P. Samuelson, T. Palmer & G. Georgiou, (2002) Genetic Analysis of the Twin Arginine Translocator Secretion Pathway in Bacteria. *Journal of Biological Chemistry* **277**: 29825-29831.

DeLisa, M. P., D. Tullman & G. Georgiou, (2003) Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proceedings of the National Academy of Sciences* **100**: 6115-6120.

Derman, A. I., Puziss, J. W., Bassford, P. J. & Beckwith, J., (1993) A signal sequence is not required for protein export in prlA mutants of *Escherichia coli*. *EMBO J* **12**: 879-888.

Deville, K., V. A. M. Gold, A. Robson, S. Whitehouse, R. B. Sessions, S. A. Baldwin, S. E. Radford & I. Collinson, (2011) The Oligomeric State and Arrangement of the Active Bacterial Translocon. *Journal of Biological Chemistry* **286**: 4659-4669.

Ding, Z. & P. J. Christie, (2003) Agrobacterium tumefaciens Twin-Arginine-Dependent Translocation Is Important for Virulence, Flagellation, and Chemotaxis but Not Type IV Secretion. *Journal of bacteriology* **185**: 760-771.

Doré, Andrew S., N. Robertson, James C. Errey, I. Ng, K. Hollenstein, B. Tehan, E. Hurrell, K. Bennett, M. Congreve, F. Magnani, Christopher G. Tate, M. Weir & Fiona H. Marshall, (2011) Structure of the Adenosine A_{2A} Receptor in Complex with ZM241385 and the Xanthines XAC and Caffeine. *Structure* **19**: 1283-1293.

Doublié, S., (2007) Production of Selenomethionyl Proteins in Prokaryotic and Eukaryotic Expression Systems. In: Macromolecular Crystallography Protocols. J. Walker & S. Doublié (eds). Humana Press, pp. 91-108.

Doudna, J. A. & R. T. Batey, (2004) Structural insights into the signal recognition particle. *Annual Review of Biochemistry* **73**: 539-557.

Dowhan, W., (1997) Molecular basis for membrane phospholipid diversity:Why Are There So Many Lipids? *Annual Review of Biochemistry* **66**: 199-232.

Drew, D., M. Lerch, E. Kunji, D.-J. Slotboom & J.-W. de Gier, (2006) Optimization of membrane protein overexpression and purification using GFP fusions. *Nat Meth* **3**: 303-313.

Drew, D., D.-J. Slotboom, G. Friso, T. Reda, P. Genevaux, M. Rapp, N. M. Meindl-Beinker, W. Lambert, M. Lerch, D. O. Daley, K.-J. Van Wijk, J. Hirst, E. Kunji & J.-W. De Gier, (2005) A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Science* **14**: 2011-2017.

Drew, D. E., G. von Heijne, P. Nordlund & J.-W. L. de Gier, (2001) Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. *FEBS Letters* **507**: 220-224.

Driessen, A. J. M., (1992) Bacterial protein translocation: kinetic and thermodynamic role of ATP and the protonmotive force. *Trends in Biochemical Sciences* **17**: 219-223.

Driessen, A. J. M., (2001) SecB, a molecular chaperone with two faces. *Trends in Microbiology* **9**: 193-196.

Economou, A. & W. Wickner, (1994) SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. *Cell* **78**: 835-843.

Efremov, R. G. & L. A. Sazanov, (2011) Structure of the membrane domain of respiratory complex I. *Nature* **476**: 414-420.

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

Erhardt, M., K. Namba & K. T. Hughes, (2010) Bacterial Nanomachines: The Flagellum and Type III Injectisome. *Cold Spring Harbor Perspectives in Biology* **2**.

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

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

Fethiere, J., (2007) Three-dimensional crystallisation of membrane proteins. *Methods in Molecular Biology* **363**.

Fincher, V., M. McCaffery & K. Cline, (1998) Evidence for a loop mechanism of protein transport by the thylakoid ΔpH pathway. *FEBS Letters* **423**: 66-70.

Fodje, M. N., A. Hansson, M. Hansson, J. G. Olsen, S. Gough, R. D. Willows & S. Al-Karadaghi, (2001) Interplay between an AAA module and an integrin I domain may regulate the function of magnesium chelatase. *Journal of Molecular Biology* **311**: 111-122.

Frauenfeld, J., J. Gumbart, E. O. v. d. Sluis, S. Funes, M. Gartmann, B. Beatrix, T. Mielke, O. Berninghausen, T. Becker, K. Schulten & R. Beckmann, (2011) Cryo-EM structure of the ribosome–SecYE complex in the membrane environment. *Nat Struct Mol Biol* **18**: 614-621.

French, S. & K. Wilson, (1978) On the treatment of negative intensity observations. *Acta Crystallographica Section A* **34**: 517-525.

Frey, M., (1994) Water structure associated with proteins and its role in crystallization. *Acta Crystallographica Section D* **50**: 663-666.

Fröbel, J., P. Rose, F. Lausberg, A.-S. Blümmel, R. Freudl & M. Müller, (2012a) Transmembrane insertion of twin-arginine signal peptides is driven by TatC and regulated by TatB. *Nat Commun* **3**: 1311.

Fröbel, J., P. Rose & M. Müller, (2011) Early Contacts between Substrate Proteins and TatA Translocase Component in Twin-arginine Translocation. *Journal of Biological Chemistry* **286**: 43679-43689.

Fröbel, J., P. Rose & M. Müller, (2012b) Twin-arginine-dependent translocation of folded proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 1029-1046.

Galan, J. E. & H. Wolf-Watz, (2006) Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **444**: 567-573.

Garavito, R. M. & S. Ferguson-Miller, (2001) Detergents as Tools in Membrane Biochemistry. *Journal of Biological Chemistry* **276**: 32403-32406.

Garman, E. F. & E. P. Mitchell, (1996) Glycerol concentrations required for cryoprotection of 50 typical protein crystallization solutions. *Journal of Applied Crystallography* **29**: 584-587.

Gérard, F. & K. Cline, (2006) Efficient Twin Arginine Translocation (Tat) Pathway Transport of a Precursor Protein Covalently Anchored to Its Initial cpTatC Binding Site. *Journal of Biological Chemistry* **281**: 6130-6135.

Gérard, F. & K. Cline, (2007) The Thylakoid Proton Gradient Promotes an Advanced Stage of Signal Peptide Binding Deep within the Tat Pathway Receptor Complex. *Journal of Biological Chemistry* **282**: 5263-5272.

Gohlke, U., L. Pullan, C. McDevitt, I. Porcelli, E. de Leeuw, T. Palmer, H. Saibil & B. Berks, (2005) The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *Proc Natl Acad Sci USA* **102**: 10482 - 10486.

Greene, N. P., I. Porcelli, G. Buchanan, M. G. Hicks, S. M. Schermann, T. Palmer & B. C. Berks, (2007) Cysteine Scanning Mutagenesis and Disulfide Mapping Studies of the TatA Component of the Bacterial Twin Arginine Translocase. *Journal of Biological Chemistry* **282**: 23937-23945.

Grosse-Kunstleve, R. W. & P. D. Adams, (2003) Substructure search procedures for macromolecular structures. *Acta Crystallographica Section D* **59**: 1966-1973.

Guan, L., I. N. Smirnova, G. Verner, S. Nagamori & H. R. Kaback, (2006) Manipulating phospholipids for crystallization of a membrane transport protein. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 1723-1726.

Gumbart, J. & K. Schulten, (2007) Structural Determinants of Lateral Gate Opening in the Protein Translocon†. *Biochemistry* **46**: 11147-11157.

Halbig, D., T. Wiegert, N. Blaudeck, R. Freudl & G. A. Sprenger, (1999) The efficient export of NADP-containing glucose-fructose oxidoreductase to the periplasm of Zymomonas mobilis depends both on an intact twin-arginine motif in the signal peptide and on the generation of a structural export signal induced by cofactor binding. *Eur J Biochem* **263**: 543-551.

Hamman, B. D., J.-C. Chen, E. E. Johnson & A. E. Johnson, (1997) The Aqueous Pore through the Translocon Has a Diameter of 40-60 Å during Cotranslational Protein Translocation at the ER Membrane. *Cell* **89**: 535-544.

Hanein, D., K. E. S. Matlack, B. Jungnickel, K. Plath, K.-U. Kalies, K. R. Miller, T. A. Rapoport & C. W. Akey, (1996) Oligomeric Rings of the Sec61p Complex Induced by Ligands Required for Protein Translocation. *Cell* **87**: 721-732.

Hansson, M. & L. Hederstedt, (1994) Purification and characterisation of a water-soluble ferrochelatase from *Bacillus subtilis*. *European Journal of Biochemistry* **220**: 201-208.

Hansson, M. D., T. Karlberg, M. A. Rahardja, S. Al-Karadaghi & M. Hansson, (2006) Amino Acid Residues His183 and Glu264 in *Bacillus subtilis* Ferrochelatase Direct and Facilitate the Insertion of Metal Ion into Protoporphyrin IX. *Biochemistry* **46**: 87-94.

Hassell, A. M., G. An, R. K. Bledsoe, J. M. Bynum, H. L. Carter, III, S.-J. J. Deng, R. T. Gampe, T. E. Grisard, K. P. Madauss, R. T. Nolte, W. J. Rocque, L. Wang, K. L. Weaver, S. P. Williams, G. B. Wisely, R. Xu & L. M. Shewchuk, (2007) Crystallization of protein-ligand complexes. *Acta Crystallographica Section D* **63**: 72-79.

Hatzixanthis, K., T. Palmer & F. Sargent, (2003) A subset of bacterial inner membrane proteins integrated by the twin-arginine translocase. *Molecular Microbiology* **49**: 1377-1390.

Hegde, R. S. & H. D. Bernstein, (2006) The surprising complexity of signal sequences. *Trends in Biochemical Sciences* **31**: 563-571.

Heijne, G., (1990) The signal peptide. *J. Membrain Biol.* **115**: 195-201.

Helenius, A. & K. Simons, (1975) Solubilization of membranes by detergents. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* **415**: 29-79.

Henderickson, W. A., JR Horton, DM LeMaster, (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J* **9**: 1665-1675.

Henderson, R. & P. N. T. Unwin, (1975) Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**: 28-32.

Hendrickson, W. A. & C. M. Ogata, (1997) Phase determination from multiwavelength anomalous diffraction measurements. In: Methods in Enzymology. Charles W. Carter, Jr. (ed). Academic Press, pp. 494-523.

Heras, B. & J. L. Martin, (2005) Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallographica Section D* **61**: 1173-1180.

Hinsley, A. P., N. R. Stanley, T. Palmer & B. C. Berks, (2001) A naturally occurring bacterial Tat signal peptide lacking one of the 'invariant' arginine residues of the consensus targeting motif. *FEBS Letters* **497**: 45-49.

Hizlan, D., A. Robson, S. Whitehouse, Vicki A. Gold, J. Vonck, D. Mills, W. Kühlbrandt & I. Collinson, (2012) Structure of the SecY Complex Unlocked by a Preprotein Mimic. *Cell Reports* **1**: 21-28.

Holton, J., (2007) XANES measurements of the rate of radiation damage to selenomethionine side chains. *Journal of Synchrotron Radiation* **14**: 51-72.

Holzapfel, E., G. Eisner, M. Alami, C. M. L. Barrett, G. Buchanan, I. Lake, J.-M. Betton, C. Robinson, T. Palmer, M. Moser & M. Muller, (2007) The Entire N-Terminal Half of TatC is Involved in Twin-Arginine Precursor Binding. *Biochemistry* **46**: 2892-2898.

Houtman, J. C. D., P. H. Brown, B. Bowden, H. Yamaguchi, E. Appella, L. E. Samelson & P. Schuck, (2007) Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: Application to adaptor protein complexes in cell signaling. *Protein Science* **16**: 30-42.

Hu, T.-C., J. Korczynska, D. K. Smith & A. M. Brzozowski, (2008) High-molecular-weight polymers for protein crystallization: poly-ϒ-glutamic acid-based precipitants. *Acta Crystallographica Section D* **64**: 957-963.

Hu, Y., E. Zhao, H. Li, B. Xia & C. Jin, (2010) Solution NMR Structure of the TatA Component of the Twin-Arginine Protein Transport System from Gram-Positive Bacterium *Bacillus subtilis*. *Journal of the American Chemical Society* **132**: 15942-15944.

Hunt, J. F., S. Weinkauf, L. Henry, J. J. Fak, P. McNicholas, D. B. Oliver & J. Deisenhofer, (2002) Nucleotide Control of Interdomain Interactions in the Conformational Reaction Cycle of SecA. *Science* **297**: 2018-2026.

Incardona, M.-F., G. P. Bourenkov, K. Levik, R. A. Pieritz, A. N. Popov & O. Svensson, (2009) EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *Journal of Synchrotron Radiation* **16**: 872-879.

Iwata, S., (2003) Methods and Results in Crystallization of Membrane Proteins. . *UL Biotechnology Series, 4. Ed* **ISBN: 0-9636817-9-6**.

Ize, B., F. Gérard & L.-F. Wu, (2002) In vivo assessment of the Tat signal peptide specificity in Escherichia coli. *Archives of Microbiology* **178**: 548-553.

Jack, R., F. Sargent, B. Berks, G. Sawers & T. Palmer, (2001) Constitutive expression of *Escherichia coli* tat genes indicates an important role for the twin-arginine translocase during aerobic and anaerobic growth. *J Bacteriol* **183**: 1801 - 1804.

Janda, C. Y., J. Li, C. Oubridge, H. Hernandez, C. V. Robinson & K. Nagai, (2010) Recognition of a signal peptide by the signal recognition particle. *Nature* **465**: 507-510.

Jeong, K. J., Y. Kawarasaki, J. Gam, B. R. Harvey, B. L. Iverson & G. Georgiou, (2004) A Periplasmic Fluorescent Reporter Protein and its Application in High-throughput Membrane Protein Topology Analysis. *Journal of Molecular Biology* **341**: 901-909.

Jones, D. T., D. W. A. Buchan, D. Cozzetto & M. Pontil, (2012) PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics* **28**: 184-190.

Jongbloed, J. D. H., U. Grieger, H. Antelmann, M. Hecker, R. Nijland, S. Bron & J. M. Van Dijl, (2004) Two minimal Tat translocases in Bacillus. *Molecular Microbiology* **54**: 1319-1325.

Kabsch, W., (2010) XDS. *Acta Crystallographica Section D* **66**: 125-132.

Kantardjieff, K. A. & B. Rupp, (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Science* **12**: 1865-1871.

Kawate, T. & E. Gouaux, (2006) Fluorescence-Detection Size-Exclusion Chromatography for Precrystallization Screening of Integral Membrane Proteins. *Structure* **14**: 673-681.

Keenan, R. J., D. M. Freymann, P. Walter & R. M. Stroud, (1998) Crystal Structure of the Signal Sequence Binding Subunit of the Signal Recognition Particle. *Cell* **94**: 181-191.

Keller, R., J. de Keyzer, A. J. M. Driessen & T. Palmer, (2012) Co-operation between different targeting pathways during integration of a membrane protein. *The Journal of Cell Biology* **199**: 303-315.

Khan, N. A., (2004) Protein Purification Protocols, 2nd edn, edited by PAUL CUTLER (Methods in Molecular Biology, Volume 244, Series Editor J. M. WALKER). Humana Press, Totowa, NJ, 2003, 484 pp. ISBN: 1-58829-067-0. *Biomedical Chromatography* **18**: 408-408.

Klein, D. W. P., Lansing M; Harley, John, (2005) Microbiology. *Boston: McGraw-Hill Higher Education*.

Kneuper, H., B. Maldonado, F. Jäger, M. Krehenbrink, G. Buchanan, R. Keller, M. Müller, B. C. Berks & T. Palmer, (2012) Molecular dissection of TatC defines critical

regions essential for protein transport and a TatB–TatC contact site. *Molecular Microbiology* **85**: 945-961.

Koch, S., M. J. Fritsch, G. Buchanan & T. Palmer, (2012) Escherichia coli TatA and TatB Proteins Have N-out, C-in Topology in Intact Cells. *Journal of Biological Chemistry* **287**: 14420-14431.

Koronakis, V., A. Sharff, E. Koronakis, B. Luisi & C. Hughes, (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**: 914-919.

Koster, M., W. Bitter & J. Tommassen, (2000) Protein secretion mechanisms in Gram-negative bacteria. *International Journal of Medical Microbiology* **290**: 325-331.

Kreutzenbeck, P., C. Kröger, F. Lausberg, N. Blaudeck, G. A. Sprenger & R. Freudl, (2007) Escherichia coli Twin Arginine (Tat) Mutant Translocases Possessing Relaxed Signal Peptide Recognition Specificities. *Journal of Biological Chemistry* **282**: 7903-7911.

Kunji, E. R. S., K. W. Chan, D. J. Slotboom, S. Floyd, R. O'Connor & M. Monné, (2005) Eukaryotic membrane protein overproduction in *Lactococcus lactis*. *Current Opinion in Biotechnology* **16**: 546-551.

Kunji, E. R. S., M. Harding, P. J. G. Butler & P. Akamine, (2008) Determination of the molecular mass and dimensions of membrane proteins by size exclusion chromatography. *Methods* **46**: 62-72.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson & D. G. Higgins, (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Lausberg, F., S. Fleckenstein, P. Kreutzenbeck, J. Fröbel, P. Rose, M. Müller & R. Freudl, (2012) Genetic Evidence for a Tight Cooperation of TatB and TatC during Productive Recognition of Twin-Arginine (Tat) Signal Peptides in *Escherichia coli*. *PLoS ONE* **7**: e39867.

Leake, M. C., N. P. Greene, R. M. Godun, T. Granjon, G. Buchanan, S. Chen, R. M. Berry, T. Palmer & B. C. Berks, (2008) Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by in vivo single-molecule imaging. *Proceedings of the National Academy of Sciences* **105**: 15376-15381.

Lebon, G., T. Warne, P. C. Edwards, K. Bennett, C. J. Langmead, A. G. W. Leslie & C. G. Tate, (2011) Agonist-bound adenosine A_{2A} receptor structures reveal common features of GPCR activation. *Nature* **474**: 521-525.

Lecerof, D., M. Fodje, A. Hansson, M. Hansson & S. Al-Karadaghi, (2000) Structural and mechanistic basis of porphyrin metallation by ferrochelatase. *Journal of Molecular Biology* **297**: 221-232.

Lee, P., G. Buchanan, N. Stanley, B. Berks & T. Palmer, (2002) Truncation analysis of TatA and TatB defines the minimal functional units required for protein translocation. *J Bacteriol* **184**: 5871 - 5879.

Lee, P. A., G. L. Orriss, G. Buchanan, N. P. Greene, P. J. Bond, C. Punginelli, R. L. Jack, M. S. P. Sansom, B. C. Berks & T. Palmer, (2006a) Cysteine-scanning Mutagenesis and Disulfide Mapping Studies of the Conserved Domain of the Twin-arginine Translocase TatB Component. *Journal of Biological Chemistry* **281**: 34072-34085.

Lee, P. A., D. Tullman-Ercek & G. Georgiou, (2006b) The Bacterial Twin-Arginine Translocation Pathway. *Annual Review of Microbiology* **60**: 373-395.

Leech, H. K., E. Raux-Deery, P. Heathcote & M.J.Warren, (2002) Production of cobalamin and siroheme in Bacillus megarterium: an investigation into the role of the

bracnhpoint chelatases sirohydrochlorin (SirB) and sirohydrochlorin cobalt cheltases (CbiX). *Biochemistry Society Transactions* **30**: 610-613.

Leslie, A. G. W., (1992) MOSFLM! . *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallography*.

Li, J., P. C. Edwards, M. Burghammer, C. Villa & G. F. X. Schertler, (2004) Structure of Bovine Rhodopsin in a Trigonal Crystal Form. *Journal of Molecular Biology* **343**: 1409-1438.

Li, L., S. Nachtergaele, A. M. Seddon, V. Tereshko, N. Ponomarenko & R. F. Ismagilov, (2008) Simple Host−Guest Chemistry To Modulate the Process of Concentration and Crystallization of Membrane Proteins by Detergent Capture in a Microfluidic Device. *Journal of the American Chemical Society* **130**: 14324-14328.

Lindahl, E. & M. S. P. Sansom, (2008) Membrane proteins: molecular dynamics simulations. *Current Opinion in Structural Biology* **18**: 425-431.

Lozier, R. H., R. A. Bogomolni & W. Stoeckenius, (1975) Bacteriorhodopsin: a light-driven proton pump in Halobacterium Halobium. *Biophysical journal* **15**: 955-962.

Lüke, I., J. Handford, T. Palmer & F. Sargent, (2009) Proteolytic processing of Escherichia coli twin-arginine signal peptides by LepB. *Archives of Microbiology* **191**: 919-925.

Lütcke H., S. H., K. Römisch, A. J. Ashford, and B. Dobberstein, (1992) The methionine-rich domain of the 54 kDa subunit of signal recognition particle is sufficient for the interaction with signal sequences. *EMBO J* **11**: 1543-1551.

Ma, X. & K. Cline, (2013) Mapping the Signal Peptide Binding and Oligomer Contact Sites of the Core Subunit of the Pea Twin Arginine Protein Translocase. *The Plant Cell Online* **25**: 999-1015.

Maldonado, B., G. Buchanan, M. Müller, B. C. Berks & T. Palmer, (2011a) Genetic Evidence for a TatC Dimer at the Core of the *Escherichia coli* Twin Arginine (Tat) Protein Translocase. *Journal of Molecular Microbiology and Biotechnology* **20**: 168-175.

Maldonado, B., H. Kneuper, G. Buchanan, K. Hatzixanthis, F. Sargent, B. C. Berks & T. Palmer, (2011b) Characterisation of the membrane-extrinsic domain of the TatB component of the twin arginine protein translocase. *FEBS Letters* **585**: 478-484.

Manjasetty, B. A., N. Croteau, J. Powlowski & A. Vrielink, (2001) Crystallization and preliminary X-ray analysis of dmpFG-encoded 4-hydroxy-2-ketovalerate aldolase-aldehyde dehydrogenase (acylating) from *Pseudomonas* sp. strain CF600. *Acta Crystallographica Section D* **57**: 582-585.

Manting, E. H. & A. J. M. Driessen, (2000) Escherichia coli translocase: the unravelling of a molecular machine. *Molecular Microbiology* **37**: 226-238.

Matar-Merheb, R., M. Rhimi, A. Leydier, F. Huché, C. Galián, E. Desuzinges-Mandon, D. Ficheux, D. Flot, N. Aghajari, R. Kahn, A. Di Pietro, J.-M. Jault, A. W. Coleman & P. Falson, (2011) Structuring Detergents for Extracting and Stabilizing Functional Membrane Proteins. *PLoS ONE* **6**: e18036.

Matlack, K. E. S., W. Mothes & T. A. Rapoport, (1998) Protein Translocation: Tunnel Vision. *Cell* **92**: 381-390.

Matthews, B., (1968) Solvent content of protein crystals. *J Mol Biol* **Apr 28;33(2):491-7**.

Maurer, C., S. Panahandeh, A.-C. Jungkamp, M. Moser & M. Muller, (2010) TatB Functions as an Oligomeric Binding Site for Folded Tat Precursor Proteins. *Mol. Biol. Cell* **21**: 4151-4161.

McAuley, K. E., P. K. Fyfe, J. P. Ridge, N. W. Isaacs, R. J. Cogdell & M. R. Jones, (1999) Structural details of an interaction between cardiolipin and an integral membrane protein. *Proceedings of the National Academy of Sciences* **96**: 14706-14711.

McDevitt, C. A., G. Buchanan, F. Sargent, T. Palmer & B. C. Berks, (2006) Subunit composition and in vivo substrate-binding characteristics of *Escherichia coli* Tat protein complexes expressed at native levels. *FEBS Journal* **273**: 5656-5668.

McPherson, A., (2001) Crystallization of Proteins and Protein–Ligand Complexes. In: eLS. John Wiley & Sons, Ltd, pp.

Mendel, S., A. McCarthy, J. P. Barnett, R. T. Eijlander, A. Nenninger, O. P. Kuipers & C. Robinson, (2008) The *Escherichia coli* TatABC System and a *Bacillus subtilis* TatAC-type System Recognise Three Distinct Targeting Determinants in Twin-arginine Signal Peptides. *Journal of Molecular Biology* **375**: 661-672.

Minamino, T., K. Imada & K. Namba, (2008) Molecular motors of the bacterial flagella. *Current Opinion in Structural Biology* **18**: 693-701.

Miroux, B. & J. E. Walker, (1996) Over-production of Proteins in *Escherichia coli*: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *Journal of Molecular Biology* **260**: 289-298.

Mitra, K., C. Schaffitzel, T. Shaikh, F. Tama, S. Jenni, C. L. Brooks, N. Ban & J. Frank, (2005) Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature* **438**: 318-324.

Monticelli, L., S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman & S.-J. Marrink, (2008) The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of Chemical Theory and Computation* **4**: 819-834.

Mori, H. & K. Cline, (2001) Post-translational protein translocation into thylakoids by the Sec and ΔpH-dependent pathways. *Biochim Biophys Acta* **1541**: 80 - 90.

Mori, H. & K. Cline, (2002) A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid ΔpH/Tat translocase. *The Journal of Cell Biology* **157**: 205-210.

Mori, H., T. Tsukazaki, R. Masui, S. Kuramitsu, S. Yokoyama, A. E. Johnson, Y. Kimura, Y. Akiyama & K. Ito, (2003) Fluorescence Resonance Energy Transfer Analysis of Protein Translocase: SecYE from *Thermophilus* HB8 forms a constitutive oligomer in membranes. *Journal of Biological Chemistry* **278**: 14257-14264.

Morth, J. P., T. L.-M. Sorensen & P. Nissen, (2006) Membrane's Eleven: heavy-atom derivatives of membrane-protein crystals. *Acta Crystallographica Section D* **62**: 877-882.

Mota, L. J. & G. R. Cornelis, (2005) The bacterial injection kit: Type III secretion systems. *Annals of Medicine* **37**: 234-249.

Mothes, W., S. Prehn, and T. A. Rapoport, (1994) Systematic probing of the environment of a translocating secretory protein during translocation through the ER membrane. *EMBO J. 1994 September 1; 13(17): 3973–3982.* .

Mould, R. M. & C. Robinson, (1991) A proton gradient is required for the transport of two lumenal oxygen-evolving proteins across the thylakoid membrane. *Journal of Biological Chemistry* **266**: 12189-12193.

Muyldermans, S., C. Cambillau & L. Wyns, (2001) Recognition of antigens by single-domain antibody fragments: the superfluous luxury of paired domains. *Trends in Biochemical Sciences* **26**: 230-235.

Newstead, S., S. Ferrandon & S. Iwata, (2008) Rationalizing α-helical membrane protein crystallization. *Protein Science* **17**: 466-472.

Nolandt, O. V., T. H. Walther, S. Roth, J. Bürck & A. S. Ulrich, (2009) Structure analysis of the membrane protein TatCd from the Tat system of *B. subtilis* by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1788**: 2238-2244.

Oates, J., C. M. L. Barrett, J. P. Barnett, K. G. Byrne, A. Bolhuis & C. Robinson, (2005) The Escherichia coli Twin-arginine Translocation Apparatus Incorporates a Distinct Form of TatABC Complex, Spectrum of Modular TatA Complexes and Minor TatAB Complex. *Journal of Molecular Biology* **346**: 295-305.

Oliver, R. C., J. Lipfert, D. A. Fox, R. H. Lo, S. Doniach & L. Columbus, (2013) Dependence of Micelle Size and Shape on Detergent Alkyl Chain Length and Head Group. *PLoS ONE* **8**: e62488.

Orriss, G. L., M. J. Tarry, B. Ize, F. Sargent, S. M. Lea, T. Palmer & B. C. Berks, (2007) TatBC, TatB, and TatC form structurally autonomous units within the twin arginine protein transport system of Escherichia coli. *FEBS Letters* **581**: 4091-4097.

Osborne, A. R. & T. A. Rapoport, (2007) Protein Translocation Is Mediated by Oligomers of the SecY Complex with One SecY Copy Forming the Channel. *Cell* **129**: 97-110.

Page, A.-L. & C. Parsot, (2002) Chaperones of the type III secretion pathway: jacks of all trades. *Molecular Microbiology* **46**: 1-11.

Palczewski, K., T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. L. Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto & M. Miyano, (2000) Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science* **289**: 739-745.

Palmer, T. & B. C. Berks, (2012) The twin-arginine translocation (Tat) protein export pathway. *Nat Rev Micro* **10**: 483-496.

Palmer, T., F. Sargent & B. C. Berks, (2005) Export of complex cofactor-containing proteins by the bacterial Tat pathway. *Trends in Microbiology* **13**: 175-180.

Palsdottir, H., C. G. Lojero, B. L. Trumpower & C. Hunte, (2003) Structure of the Yeast Cytochrome bc1 Complex with a Hydroxyquinone Anion Qo Site Inhibitor Bound. *Journal of Biological Chemistry* **278**: 31303-31311.

Panahandeh, S., C. Maurer, M. Moser, M. P. DeLisa & M. Müller, (2008) Following the Path of a Twin-arginine Precursor along the TatABC Translocase of *Escherichia coli*. *Journal of Biological Chemistry* **283**: 33267-33275.

Papanikou, E., S. Karamanou & A. Economou, (2007) Bacterial protein secretion through the translocase nanomachine. *Nat Rev Micro* **5**: 839-851.

Park, E. & T. A. Rapoport, (2012) Bacterial protein translocation requires only one copy of the SecY complex in vivo. *The Journal of Cell Biology* **198**: 881-893.

Pellegrini, L., D. S. Yu, T. Lo, S. Anand, M. Lee, T. L. Blundell & A. R. Venkitaraman, (2002) Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature* **420**: 287-293.

Pierce, B. G., Y. Hourai & Z. Weng, (2011) Accelerating Protein Docking in ZDOCK Using an Advanced 3D Convolution Library. *PLoS ONE* **6**: e24657.

Plath, K., W. Mothes, B. M. Wilkinson, C. J. Stirling & T. A. Rapoport, (1998) Signal Sequence Recognition in Posttranslational Protein Transport across the Yeast ER Membrane. *Cell* **94**: 795-807.

Pohlschröder, M., K. Dilks, N. J. Hand & R. Wesley Rose, (2004) Translocation of proteins across archaeal cytoplasmic membranes. *FEMS Microbiology Reviews* **28**: 3-24.

Popot, J.-L., T. Althoff, D. Bagnard, J.-L. Banères, P. Bazzacco, E. Billon-Denis, L. J. Catoire, P. Champeil, D. Charvolin, M. J. Cocco, G. Crémel, T. Dahmane, L. M. de la Maza, C. Ebel, F. Gabel, F. Giusti, Y. Gohon, E. Goormaghtigh, E. Guittet, J. H. Kleinschmidt, W. Kühlbrandt, C. Le Bon, K. L. Martinez, M. Picard, B. Pucci, J. N. Sachs, C. Tribet, C. van Heijenoort, F. Wien, F. Zito & M. Zoonens, (2011) Amphipols From A to Z*. *Annual Review of Biophysics* **40**: 379-408.

Privé, G. G., (2007) Detergents for the stabilization and crystallization of membrane proteins. *Methods* **41**: 388-397.

Punginelli, C., B. Maldonado, S. Grahl, R. Jack, M. Alami, J. Schroder, B. C. Berks & T. Palmer, (2007) Cysteine scanning mutagenesis and topological mapping of the *Escherichia coli* twin-arginine translocase TatC Component. *Journal of bacteriology* **189**: 5482-5494.

Ramasamy, S., R. Abrol, Christian J. M. Suloway & William M. Clemons Jr, (2013) The Glove-like Structure of the Conserved Membrane Protein TatC Provides Insight into Signal Sequence Recognition in Twin-Arginine Translocation. *Structure* **21**: 777-788.

Rasmussen, S. G. F., H.-J. Choi, J. J. Fung, E. Pardon, P. Casarosa, P. S. Chae, B. T. DeVree, D. M. Rosenbaum, F. S. Thian, T. S. Kobilka, A. Schnapp, I. Konetzki, R. K. Sunahara, S. H. Gellman, A. Pautsch, J. Steyaert, W. I. Weis & B. K. Kobilka, (2011) Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* **469**: 175-180.

Rasmussen, S. G. F., H.-J. Choi, D. M. Rosenbaum, T. S. Kobilka, F. S. Thian, P. C. Edwards, M. Burghammer, V. R. P. Ratnala, R. Sanishvili, R. F. Fischetti, G. F. X. Schertler, W. I. Weis & B. K. Kobilka, (2007) Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature* **450**: 383-387.

Rath, A., M. Glibowicka, V. G. Nadeau, G. Chen & C. M. Deber, (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences* **106**: 1760-1765.

Remaut, H. & G. Waksman, (2006) Protein–protein interaction through β-strand addition. *Trends in Biochemical Sciences* **31**: 436-444.

Richter, S., U. Lindenstrauss, C. Lücke, R. Bayliss & T. Brüser, (2007) Functional Tat Transport of Unstructured, Small, Hydrophilic Proteins. *Journal of Biological Chemistry* **282**: 33257-33264.

Rodrigue, A., A. Chanal, K. Beck, M. Müller & L.-F. Wu, (1999) Co-translocation of a Periplasmic Enzyme Complex by a Hitchhiker Mechanism through the Bacterial Tat Pathway. *Journal of Biological Chemistry* **274**: 13223-13228.

Rodriguez, F., S. L. Rouse, C. E. Tait, J. Harmer, A. De Riso, C. R. Timmel, M. S. P. Sansom, B. C. Berks & J. R. Schnell, (2013) Structural model for the protein-translocating element of the twin-arginine transport system. *Proceedings of the National Academy of Sciences*.

Rollauer, S. E., M. J. Tarry, J. E. Graham, M. Jaaskelainen, F. Jager, S. Johnson, M. Krehenbrink, S.-M. Liu, M. J. Lukey, J. Marcoux, M. A. McDowell, F. Rodriguez, P. Roversi, P. J. Stansfeld, C. V. Robinson, M. S. P. Sansom, T. Palmer, M. Hogbom, B. C. Berks & S. M. Lea, (2012) Structure of the TatC core of the twin-arginine protein transport system. *Nature* **492**: 210-214.

Romão, C. V., D. Ladakis, S. A. L. Lobo, M. A. Carrondo, A. A. Brindley, E. Deery, P. M. Matias, R. W. Pickersgill, L. M. Saraiva & M. J. Warren, (2011) Evolution in a family of chelatases facilitated by the introduction of active site asymmetry and protein oligomerization. *Proceedings of the National Academy of Sciences* **108**: 97-102.

Rose, J. P., Chia-KuieWu, Zhi-Jie Liu, Gary Newton, Bi-Cheng Wang, (2001) Using single wavlength anomolous scattering data for in-house protein structure determination. *The Rigaku Journal* **18**: 2.

Rose, R. W., T. Brüser, J. C. Kissinger & M. Pohlschröder, (2002) Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Molecular Microbiology* **45**: 943-950.

Rosenbaum, D. M., V. Cherezov, M. A. Hanson, S. G. F. Rasmussen, F. S. Thian, T. S. Kobilka, H.-J. Choi, X.-J. Yao, W. I. Weis, R. C. Stevens & B. K. Kobilka, (2007) GPCR Engineering Yields High-Resolution Structural Insights into β2-Adrenergic Receptor Function. *Science* **318**: 1266-1273.

Rosenbaum, D. M., C. Zhang, J. A. Lyons, R. Holl, D. Aragao, D. H. Arlow, S. G. F. Rasmussen, H.-J. Choi, B. T. DeVree, R. K. Sunahara, P. S. Chae, S. H. Gellman, R. O. Dror, D. E. Shaw, W. I. Weis, M. Caffrey, P. Gmeiner & B. K. Kobilka, (2011) Structure and function of an irreversible agonist-β_2 adrenoceptor complex. *Nature* **469**: 236-240.

Rupp, B., (2009) Biomolecular crystallography. Garland Science.

Russi, S., D. H. Juers, J. Sanchez-Weatherby, E. Pellegrini, E. Mossou, V. T. Forsyth, J. Huet, A. Gobbo, F. Felisaz, R. Moya, S. M. McSweeney, S. Cusack, F. Cipriani & M. W. Bowler, (2011) Inducing phase changes in crystals of macromolecules: Status and perspectives for controlled crystal dehydration. *Journal of Structural Biology* **175**: 236-243.

San Miguel, M., R. Marrington, P. M. Rodger, A. Rodger & C. Robinson, (2003) An Escherichia coli twin-arginine signal peptide switches between helical and unstructured conformations depending on the hydrophobicity of the environment. *European Journal of Biochemistry* **270**: 3345-3352.

Sanchez-Weatherby, J., M. W. Bowler, J. Huet, A. Gobbo, F. Felisaz, B. Lavault, R. Moya, J. Kadlec, R. B. G. Ravelli & F. Cipriani, (2009) Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallographica Section D* **65**: 1237-1246.

Sanders, C., N. Wethkamp & H. Lill, (2001) Transport of cytochrome c derivatives by the bacterial Tat protein translocation system. *Molecular Microbiology* **41**: 241-246.

Sandkvist, M., (2001) Type II Secretion and Pathogenesis. *Infection and Immunity* **69**: 3523-3535.

Saparov, S. M., K. Erlandson, K. Cannon, J. Schaletzky, S. Schulman, T. A. Rapoport & P. Pohl, (2007) Determining the Conductance of the SecY Protein Translocation Channel for Small Molecules. *Molecular Cell* **26**: 501-509.

Sargent, F., E. Bogsch, N. Stanley, M. Wexler, C. Robinson, B. Berks & T. Palmer, (1998) Overlapping functions of components of a bacterial Sec-independent protein export pathway. *EMBO J* **17**: 3640 - 3650.

Sargent, F., U. Gohlke, E. de Leeuw, N. Stanley, T. Palmer, H. Saibil & B. Berks, (2001) Purified components of the *Escherichia coli* Tat protein transport system form a double-layered ring structure. *Eur J Biochem* **268**: 3361 - 3367.

Sargent, F., N. Stanley, B. Berks & T. Palmer, (1999) Sec-independent protein translocation in Escherichia coli: a distinct and pivotal role for the TatB protein. *J Biol Chem* **274**: 36073 - 36083.

Schagger, H., (2006) Tricine-SDS-PAGE. *Nat. Protocols* **1**: 16-22.

Schliebs, W., W. Girzalsky & R. Erdmann, (2010) Peroxisomal protein import and ERAD: variations on a common theme. *Nat Rev Mol Cell Biol* **11**: 885-890.

Schraidt, O. & T. C. Marlovits, (2011) Three-Dimensional Model of Salmonella's Needle Complex at Subnanometer Resolution. *Science* **331**: 1192-1195.

Sciara, G. & F. Mancia, (2012) Highlights from recently determined structures of membrane proteins: a focus on channels and transporters. *Current Opinion in Structural Biology* **22**: 476-481.

Scott, K. A., P. J. Bond, A. Ivetac, A. P. Chetwynd, S. Khalid & M. S. P. Sansom, (2008) Coarse-Grained MD Simulations of Membrane Protein-Bilayer Self-Assembly. *Structure* **16**: 621-630.

Scotti, P. A., M. L. Urbanus, J. Brunner, J.-W. L. de Gier, G. von Heijne, C. van der Does, A. J. M. Driessen, B. Oudega & J. Luirink, (2000) YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J* **19**: 542-549.

Serrano-Vega, M. J., F. Magnani, Y. Shibata & C. G. Tate, (2008) Conformational thermostabilization of the β1-adrenergic receptor in a detergent-resistant form. *Proceedings of the National Academy of Sciences* **105**: 877-882.

Settles, A. M., A. Yonetani, A. Baron, D. R. Bush, K. Cline & R. Martienssen, (1997) Sec-Independent Protein Translocation by the Maize Hcf106 Protein. *Science* **278**: 1467-1470.

Signorell, G. A., T. C. Kaufmann, W. Kukulski, A. Engel & H.-W. Rémigy, (2007) Controlled 2D crystallization of membrane proteins using methyl-β-cyclodextrin. *Journal of Structural Biology* **157**: 321-328.

Simon, S. M. & G. Blobel, (1991) A protein-conducting channel in the endoplasmic reticulum. *Cell* **65**: 371-380.

Sonoda, Y., A. Cameron, S. Newstead, H. Omote, Y. Moriyama, M. Kasahara, S. Iwata & D. Drew, (2010) Tricks of the trade used to accelerate high-resolution structure determination of membrane proteins. *FEBS Letters* **584**: 2539-2547.

Sonoda, Y., S. Newstead, N.-J. Hu, Y. Alguel, E. Nji, K. Beis, S. Yashiro, C. Lee, J. Leung, A. D. Cameron, B. Byrne, S. Iwata & D. Drew, (2011) Benchmarking Membrane Protein Detergent Stability for Improving Throughput of High-Resolution X-ray Structures. *Structure* **19**: 17-25.

Stanley, N., T. Palmer & B. Berks, (2000) The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in Escherichia coli. *J Biol Chem* **257**: 11591 - 11596.

Steyaert, J. & B. K. Kobilka, (2011) Nanobody stabilization of G protein-coupled receptor conformational states. *Current Opinion in Structural Biology* **21**: 567-572.

Strauch, E.-M. & G. Georgiou, (2007) *Escherichia coli* tatC Mutations that Suppress Defective Twin-Arginine Transporter Signal Peptides. *Journal of Molecular Biology* **374**: 283-291.

Strop, P. & A. T. Brunger, (2005) Refractive index-based determination of detergent concentration and its application to the study of membrane proteins. *Protein Science* **14**: 2207-2211.

Stroupe, M. E., H. K. Leech, D. S. Daniels, M. J. Warren & E. D. Getzoff, (2003) CysG structure reveals tetrapyrrole-binding features and novel regulation of siroheme biosynthesis. *Nat Struct Mol Biol* **10**: 1064-1073.

Strushkevich, N., F. MacKenzie, T. Cherkesova, I. Grabovec, S. Usanov & H.-W. Park, (2011) Structural basis for pregnenolone biosynthesis by the mitochondrial monooxygenase system. *Proceedings of the National Academy of Sciences* **108**: 10139-10143.

Sun, P. D., S. Radaev & M. Kattah, (2002) Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases. *Acta Crystallographica Section D* **58**: 1092-1098.

Tarry, M., S. J. R. Arends, P. Roversi, E. Piette, F. Sargent, B. C. Berks, D. S. Weiss & S. M. Lea, (2009a) The *Escherichia coli* Cell Division Protein and Model Tat Substrate SufI (FtsP) Localizes to the Septal Ring and Has a Multicopper Oxidase-Like Structure. *Journal of Molecular Biology* **386**: 504-519.

Tarry, M., K. Skaar, G. v. Heijne, R. R. Draheim & M. Högbom, (2012) Production of human tetraspanin proteins in *Escherichia coli*. *Protein Expression and Purification* **82**: 373-379.

Tarry, M. J., E. Schafer, S. Chen, G. Buchanan, N. P. Greene, S. M. Lea, T. Palmer, H. R. Saibil & B. C. Berks, (2009b) Structural analysis of substrate binding by the TatBC component of the twin-arginine protein transport system. *Proceedings of the National Academy of Sciences* **106**: 13284-13289.

Tate, C., (2010) Practical Considerations of Membrane Protein Instability during Purification and Crystallisation. In: Heterologous Expression of Membrane Proteins. I. Mus-Veteau (ed). Humana Press, pp. 187-203.

Taylor, A. B., B. S. Smith, S. Kitada, K. Kojima, H. Miyaura, Z. Otwinowski, A. Ito & J. Deisenhofer, (2001) Crystal Structures of Mitochondrial Processing Peptidase Reveal the Mode for Specific Cleavage of Import Signal Sequences. *Structure* **9**: 615-625.

Tottey, S., K. J. Waldron, S. J. Firbank, B. Reale, C. Bessant, K. Sato, T. R. Cheek, J. Gray, M. J. Banfield, C. Dennison & N. J. Robinson, (2008) Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature* **455**: 1138-1142.

Tribet, C., R. Audebert & J.-L. Popot, (1996) Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. *Proceedings of the National Academy of Sciences* **93**: 15047-15050.

Tripathy, B. C., I. Sherameti & R. Oelmüller, (2010) Siroheme: An essential component for life on earth. *Plant Signaling & Behavior* **5**: 14-20.

Trott, O. & A. J. Olson, (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**: 455-461.

Tseng, T.-T., B. Tyler & J. Setubal, (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiology* **9**: S2.

Tsukazaki, T., H. Mori, S. Fukai, R. Ishitani, T. Mori, N. Dohmae, A. Perederina, Y. Sugita, D. G. Vassylyev, K. Ito & O. Nureki, (2008) Conformational transition of Sec machinery inferred from bacterial SecYE structures. *Nature* **455**: 988-991.

Urbani, A. & T. Warne, (2005) A colorimetric determination for glycosidic and bile salt-based detergents: applications in membrane protein research. *Analytical Biochemistry* **336**: 117-124.

Valent, Q. A., J.-W. L. De Gier, G. v. Heijne, D. A. Kendall, C. M. Ten Hagen-Jongman, B. Oudega & J. Luirink, (1997) Nascent membrane and presecretory proteins synthesized in *Escherichia coli* associate with signal recognition particle and trigger factor. *Molecular Microbiology* **25**: 53-64.

Van den Berg, B., W. M. Clemons, I. Collinson, Y. Modis, E. Hartmann, S. C. Harrison & T. A. Rapoport, (2004) X-ray structure of a protein-conducting channel. *Nature* **427**: 36-44.

Van Duyne, G. D., R. F. Standaert, P. A. Karplus, S. L. Schreiber & J. Clardy, (1993) Atomic Structures of the Human Immunophilin FKBP-12 Complexes with FK506 and Rapamycin. *Journal of Molecular Biology* **229**: 105-124.

von Heijne, G., (1985) Signal sequences: The limits of variation. *Journal of Molecular Biology* **184**: 99-105.

Vonrhein, C., E. Blanc, P. Roversi & G. Bricogne, (2007) Automated Structure Solution With autoSHARP. In: Macromolecular Crystallography Protocols. S. Doublié (ed). Humana Press, pp. 215-230.

Wagner, S., L. Baars, A. J. Ytterberg, A. Klussmeier, C. S. Wagner, O. Nord, P.-Å. Nygren, K. J. van Wijk & J.-W. de Gier, (2007) Consequences of Membrane Protein Overexpression in *Escherichia coli*. *Molecular & Cellular Proteomics* **6**: 1527-1550.

Wagner, S., M. M. Klepsch, S. Schlegel, A. Appel, R. Draheim, M. Tarry, M. Högbom, K. J. van Wijk, D. J. Slotboom, J. O. Persson & J.-W. de Gier, (2008) Tuning *Escherichia coli* for membrane protein overexpression. *Proceedings of the National Academy of Sciences* **105**: 14371-14376.

Walden, H., (2010) Selenium incorporation using recombinant techniques. *Acta Crystallographica Section D* **66**: 352-357.

Wallrapp, F., D. Masone & V. Guallar, (2008) Electron Transfer in the P_{450}cam/PDX Complex. The QM/MM e-Pathway. *The Journal of Physical Chemistry A* **112**: 12989-12994.

Walter, T. S., C. Meier, R. Assenberg, K.-F. Au, J. Ren, A. Verma, Joanne E. Nettleship, R. J. Owens, David I. Stuart & J. M. Grimes, (2006) Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. *Structure* **14**: 1617-1622.

Walther, Torsten H., C. Gottselig, Stephan L. Grage, M. Wolf, Attilio V. Vargiu, Marco J. Klein, S. Vollmer, S. Prock, M. Hartmann, S. Afonin, E. Stockwald, H. Heinzmann, Olga V. Nolandt, W. Wenzel, P. Ruggerone & Anne S. Ulrich, (2013) Folding and Self-Assembly of the TatA Translocation Pore Based on a Charge Zipper Mechanism. *Cell* **152**: 316-326.

Wang, B., (1985) Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymology* **115:90-112**.

Warne, T., M. J. Serrano-Vega, J. G. Baker, R. Moukhametzianov, P. C. Edwards, R. Henderson, A. G. W. Leslie, C. G. Tate & G. F. X. Schertler, (2008) Structure of a β_1-adrenergic G-protein-coupled receptor. *Nature* **454**: 486-491.

Warren, M. J., E. L. Bolt, C. A. Roessner, A. I. Scott, J. B. Spencer & S. C. Woodcock, (1994) Gene dissection demonstrates that the Escherichia coli cysG gene encodes a multifunctional protein. *Biochem. J.* **302**: 837-844.

Whitaker, N., U. K. Bageshwar & S. M. Musser, (2012) Kinetics of Precursor Interactions with the Bacterial Tat Translocase Detected by Real-time FRET. *Journal of Biological Chemistry* **287**: 11252-11260.

White, G. F., S. M. Schermann, J. Bradley, A. Roberts, N. P. Greene, B. C. Berks & A. J. Thomson, (2010) Subunit Organization in the TatA Complex of the Twin Arginine Protein Translocase: A site-directed EPR spin labelling study. *Journal of Biological Chemistry* **285**: 2294-2301.

Wiener, M. C., (2004) A pedestrian guide to membrane protein crystallization. *Methods* **34**: 364-372.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin & K. S. Wilson, (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D* **67**: 235-242.

Winter, G., (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**: 186-190.

Yamada, Y., M. Inoue, T. Shiba, M. Kawasaki, R. Kato, K. Nakayama & S. Wakatsuki, (2005) Structure determination of GGA-GAE and [gamma]1-ear in complex with peptides: crystallization of low-affinity complexes in membrane traffic. *Acta Crystallographica Section D* **61**: 731-736.

Yen, M.-R., Y.-H. Tseng, E. Nguyen, L.-F. Wu & M. Saier, (2002) Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Archives of Microbiology* **177**: 441-450.

Younis, R., L. E. H. Bingle, S. Rollauer, D. Munera, S. J. Busby, S. Johnson, J. E. Deane, S. M. Lea, G. Frankel & M. J. Pallen, (2010) SepL Resembles an Aberrant Effector

in Binding to a Class 1 Type III Secretion Chaperone and Carrying an N-Terminal Secretion Signal. *J. Bacteriol.* **192**: 6093-6098.

Yu, H., (1999) Extending the size limit of protein nuclear magnetic resonance. *Proceedings of the National Academy of Sciences* **96**: 332-334.

Zhang, K. Y. J., K. D. Cowtan & P. Main, (2001) Phase improvement by iterative density modification. In: International Tables for Crystallography Volume F: Crystallography of biological macromolecules. M. G. Rossmann & E. Arnold (eds). Springer Netherlands, pp. 311-324.

Zimmer, J., Y. Nam & T. A. Rapoport, (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455**: 936-943.

Zimmer, J. & T. A. Rapoport, (2009) Conformational Flexibility and Peptide Interaction of the Translocation ATPase SecA. *Journal of Molecular Biology* **394**: 606-612.

Zoufaly, S., J. Fröbel, P. Rose, T. Flecken, C. Maurer, M. Moser & M. Müller, (2012) Mapping Precursor-binding Site on TatC Subunit of Twin Arginine-specific Protein Translocase by Site-specific Photo Cross-linking. *Journal of Biological Chemistry* **287**: 13430-13441.