Structural and Computational Analysis of the 
Escherichia coli Chaperone Protein DmsD

by

Charles M. Stevens
B.Sc., University of Calgary, 2006

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in the
Department of Molecular Biology & Biochemistry
Faculty of Science

© Charles M. Stevens 2012
SIMON FRASER UNIVERSITY
Fall 2012

All rights reserved.
However, in accordance with the Copyright Act of Canada, this work may
be reproduced, without authorization, under the conditions for
“Fair Dealing.” Therefore, limited reproduction of this work for the
purposes of private study, research, criticism, review and news reporting
is likely to be in accordance with the law, particularly if cited appropriately.
Approval

Name: Charles M. Stevens  
Degree: Doctor of Philosophy  
Molecular Biology & Biochemistry  
Title of Thesis: Structural and Computational Analysis of the Escherichia coli Chaperone Protein DmsD  
Examining Committee:  
Chair: David Baillie, Professor  
Mark Paetzel  
Senior Supervisor  
Associate Professor  
Christopher T. Beh  
Supervisor  
Associate Professor  
Edgar Young  
Supervisor  
Associate Professor  
Michel Leroux  
Internal Examiner  
Professor  
Molecular Biology & Biochemistry  
Michael Murphy  
External Examiner  
Professor, Microbiology & Immunology  
University of British Columbia  
Date Defended/Approved: September 19th 2012
Partial Copyright License

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the “Institutional Repository” link of the SFU Library website (www.lib.sfu.ca) at http://summit/sfu.ca and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author’s written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2011
Abstract

In Gram-negative bacteria, the secretion of proteins that contain redox cofactors is accomplished using the twin arginine translocase (TAT) system, so named because the cofactor containing secretory proteins contain an N-terminal leader peptide with a twin arginine motif. The redox enzyme maturation proteins (REMPs) are molecular chaperones that prevent TAT substrate translocation until the preprotein is folded and its cofactor is incorporated. REMPs then direct the substrate to the TAT translocase. The work presented here explores, from a structural biology perspective, a model REMP: Escherichia coli DmsD.

DmsD was crystallized and the structure determined and refined to 2.0 Å resolution. This is the first structure of E. coli DmsD, and contains clear electron density for all 204 amino acid residues in the protein molecule. This was complemented by NMR analysis that characterized the local backbone dynamics of the protein. The dynamic properties of DmsD were also explored by molecular dynamics simulation. These analyses have identified three flexible regions of DmsD, two of which contribute to the putative leader peptide binding site. The third flexible region is located in a patch of residues that were implicated in GTP binding in the homologue TorD.

A method for the purification of active TAT leader peptides was devised and used in the generation of a selectively labeled sample for NMR analysis. Chemical shift perturbation analyses are consistent with a hydrophobic groove on the surface of DmsD interacting with the DmsA leader peptide. DmsA is the cofactor containing catalytic subunit of DMSO reductase and the specific substrate of DmsD. The leader peptide binding groove on DmsD overlaps with the previously identified “hot pocket” which is predicted to interact with the twin-arginine motif of the leader peptide, and encompasses regions of the x-ray structure that form crystal contacts with crystallization reagents. Finally, a new model for TAT leader peptide binding is proposed which combines all of the available data.

Keywords: DmsD; Chaperone, Twin-Arginine-Translocase (TAT); Redox Enzyme Maturation Protein (REMP); X-Ray Crystallography; Nuclear Magnetic Resonance (NMR).
In Memory of my Parents
Acknowledgements

I would like to acknowledge the efforts of my senior supervisor, Dr. Mark Paetzel, and my supervisory committee Dr. Edgar Young and Dr. Chris Beh for their guidance. Their expertise and insight have been invaluable throughout the course of this project.

I would also like to express my gratitude to our lab manager, Deidre de Jong-Wong for her help and support in making sure that I was always organized. Her help when vetting ideas and for making sure that I had remained on track and productive, even when progress was slow was essential to the work described in this thesis.

The past and present members of the Paetzel lab have also been extremely helpful, whether it was discussing ideas or proofreading papers. I would like to acknowledge their role in this work. Dr. Kelly Kim, Dr. Jae-Yong Lee, Dr. David Oliver, Dr. Anat Feldman, Suraaj Aulakh, Alison Li, Cindy Li, Dan Chiang, Sung-Eun Nam, Linda Zhang, Apollos Kim, Ivy Cheung, Chuanyun Luo, Yuliya Shapova, Minfei Fu, and Michael Ungerer all deserve my gratitude for everything that they have done to help me over the past few years.

I would like to thank the professors and graduate students who had given me my start in research, Dr. Raymond J. Turner, Dr. Michael Surette, Dr. Catherine Chan and Dylan Silver. Finally I’d like to thank Sarah and Nadya for reading long papers that don’t make sense to them and for putting up with my schedule and keeping me sane while under pressure and working long hours.
Table of Contents

Approval ...................................................................................................................... ii
Partial Copyright License ........................................................................................ iii
Abstract ..................................................................................................................... iv
Dedication .................................................................................................................. v
Acknowledgements ................................................................................................... vi
Table of Contents ...................................................................................................... vii
List of Tables ............................................................................................................. x
List of Figures ........................................................................................................... xi
Glossary ..................................................................................................................... xii

1. Introduction & background .................................................................................. 1
  1.1. Why study the TAT system? .......................................................................... 1
  1.2. Type II protein secretion .............................................................................. 1
  1.3. The twin arginine translocase ..................................................................... 2
  1.4. A model TAT substrate protein: DMSO reductase ....................................... 4
  1.5. The TAT leader peptide .............................................................................. 6
  1.6. Redox enzyme maturation proteins .............................................................. 8
  1.7. The structure of all-α (TorD-like) REMPs .................................................... 11
  1.8. The structure of α/β REMPs ..................................................................... 14
  1.9. The REMP-leader peptide interaction ......................................................... 17
  1.10. Proofreading and substrate assembly ......................................................... 19
  1.11. GTPase / guanine nucleotide binding ......................................................... 20
  1.12. Oligomerization ....................................................................................... 22
  1.13. Outstanding issues in the literature ............................................................. 24
  1.14. Objectives ................................................................................................ 26

2. X-ray crystallographic analysis of E. coli DmsD .................................................. 28
  2.1. Publication & contributions ......................................................................... 28
  2.2. Introduction ................................................................................................ 28
  2.3. Materials & methods ................................................................................... 29
     2.3.1. Cloning ................................................................................................. 29
     2.3.2. Expression and purification of E. coli DmsD .......................................... 30
     2.3.3. Crystallization ..................................................................................... 31
     2.3.4. Data collection ................................................................................... 31
     2.3.5. Structure determination and refinement .............................................. 34
     2.3.6. Structural analysis ............................................................................... 34
     2.3.7. Protein data bank accession codes ...................................................... 35
     2.3.8. Figure preparation ............................................................................... 35
  2.4. Results and discussion ................................................................................... 35
     2.4.1. P3_21 crystal Form ............................................................................... 35
        2.4.1.1. Purification & crystallization ............................................................ 35
        2.4.1.2. Structure solution ......................................................................... 36
        2.4.1.3. Protein fold .................................................................................. 36
        2.4.1.4. Structural comparison to other REMPs ......................................... 40
        2.4.1.5. The putative TAT leader peptide binding pocket on E. coli DmsD .......................................................................................................................... 42
3. Purification of the DmsA leader peptide by co-expression with the DmsD chaperone ................................................. 52

3.1. Publication & contributions ............................................. 52
3.2. Introduction .................................................................. 52
3.3. Materials and methods .................................................. 54
  3.3.1. Cloning DmsA_L ..................................................... 54
  3.3.2. Cloning DmsD ....................................................... 54
  3.3.3. Cloning the DmsA_L-DmsD fusion ............................ 55
  3.3.4. Expression ............................................................ 55
  3.3.5. Purification of DmsD, the DmsA_L-DmsD fusion or the DmsA_e/DmsD complex .................................................... 56
  3.3.6. Purification of DmsA_L ............................................ 57
  3.3.7. Amino-terminal sequencing of the DmsA_L peptide ...... 57
  3.3.8. Formation of DmsA_e/DmsD complex from purified peptide and protein ...................................................... 57
  3.3.9. Analytical size-exclusion chromatography in-line with multi-angle light scattering analysis (SEC-MALS) .................... 58
3.4. Results & discussion ..................................................... 58
  3.4.1. Expression and purification ...................................... 58
  3.4.2. Reassembly of the DmsA_L/DmsD complex ................ 61
  3.4.3. The DmsA_L-DmsD fusion protein ............................ 63
3.5. Concluding remarks .................................................... 66

4. NMR backbone amide assignment, backbone relaxation and chemical shift perturbation analysis of DmsD ................................................... 67

4.1. Publication & contributions ............................................. 67
4.2. Introduction .................................................................. 67
4.3. Methods and experiments ............................................. 68
  4.3.1. Cloning, expression and purification .......................... 68
  4.3.2. NMR data collection .............................................. 69
4.4. Results & discussion ..................................................... 71
  4.4.1. Assignments .......................................................... 71
  4.4.2. Secondary structure ............................................... 72
  4.4.3. Dynamics ............................................................. 72
  4.4.4. Chemical shift perturbation ..................................... 74
  4.4.5. Structure of the bound leader peptide ........................ 75

5. Molecular dynamics simulation ........................................ 76
5.1. Publication & contributions ............................................. 76
5.2. Introduction .................................................................. 76
5.3. Materials & methods .................................................... 77
  5.3.1. Molecular dynamics simulation of DmsD .................... 77
5.3.2. Docking & molecular dynamics of the DmsA₅/DmsD complex .................. 78
5.3.3. Molecular model for the DmsA₅-DmsD fusion protein .......................... 79
5.3.4. Generation of a new model for DmsD peptide binding ...................... 80
5.4. Results & discussion .............................................................................. 81
  5.4.1. Molecular dynamics of the DmsD protein ......................................... 81
  5.4.2. Docking & molecular dynamics simulation of the DmsA₅/DmsD complex ......................................................................................... 84
  5.4.3. Molecular model for the DmsA₅-DmsD fusion protein ...................... 86
  5.4.4. A new model for DmsD/DmsA₅ peptide binding ............................... 90

6. Summary and conclusion ........................................................................... 92
  6.1. The structure & dynamics of the DmsD protein ...................................... 92
  6.2. The DmsA₅/DmsD interaction .............................................................. 94
  6.3. Future directions .................................................................................... 96

References ..................................................................................................... 99

Appendices .................................................................................................... 112
Appendix A. Cloning details ......................................................................... 113
Appendix B. Statistics in crystallography ....................................................... 114
Appendix C. Size-exclusion chromatography standard curves ...................... 116
Appendix D. Alignment of protein sequences for E.coli REMP s .................. 119
List of Tables

Table 1.1. Statistics for α-type REMP X-ray crystal structures................................. 12
Table 1.2. Statistics for α/β REMP X-ray crystal structures ...................................... 12
Table 1.3. Statistics for REMP NMR structures........................................................... 13
Table 2.1. Crystallographic statistics ........................................................................... 33
Table 3.1. Molecular mass of protein constructs: calculated and observed by SEC-MALS............................................................................................................... 65
Table 5.1. Proposed interactions between DmsD and the truncated DmsA_L............... 85
Table 5.2. Contacts between the full length DmsA leader and DmsD ......................... 89
List of Figures

Figure 1.1. A schematic diagram of the twin arginine translocase.............................................. 3
Figure 1.2. Cofactors of the TAT substrates.................................................................................. 6
Figure 1.3. Sequence alignment of the E. coli leader peptides......................................................... 7
Figure 1.4. A schematic representation of the TAT and sec leader peptides................................. 7
Figure 1.5. Phylogram of the REMP family proteins in E. coli divided into all α-helical and α/β types. ......................................................................................................................... 10
Figure 1.6. The α-helical REMPs visualized as cartoon ribbons..................................................... 11
Figure 1.7. Structures of the α/β REMP proteins............................................................................ 16
Figure 1.8. The putative leader peptide binding surface of DmsD.................................................. 18
Figure 1.9. Cartoon rendering and schematic of the modeled binding pocket on a single lobe of the S. massilia TorD structure (PDB: 1n1c).................................................................. 22
Figure 1.10. Schematic representation of the monomeric, dimeric and trimeric forms of the TorD-like REMPs. ......................................................................................................................... 24
Figure 2.2. Purification and crystallization of DmsD expressed in pTDms67 ................................. 36
Figure 2.2. Sample electron density contoured at 2.0 σ for the 3efp structure and 3u41 structure. ................................................................................................................................. 37
Figure 2.3. The fold of the DmsD protein....................................................................................... 38
Figure 2.4. Sequence alignment of E. coli DmsD with homologous proteins................................. 39
Figure 2.6. Electron Density for the hinge region of E. coli DmsD................................................ 41
Figure 2.7. The putative leader peptide binding pocket on E. coli DmsD................................. 43
Figure 2.8. Purification and crystallization of DmsD expressed in pET-15b ................................. 44
Figure 2.9. General structural features of the E. coli DmsD structure in the P212121 space group ............................................................................................................................... 45
Figure 2.10. Comparison of flexibility in each of the DmsD structures...................................... 47
Figure 2.11. The presence of ligands observed on the surface of the crystalized E. coli DmsD molecule.............................................................................................................................. 49
Figure 3.1. Schematic of the DmsA leader peptide .......................................................... 53
Figure 3.2. Co-expressed DmsA_L and DmsD ................................................................. 60
Figure 3.3. Denatured purification of the complex ......................................................... 61
Figure 3.4. Reconstitution of the DmsA_L/DmsD complex from purified components. .................................................................................................................. 62
Figure 3.6. Purification of the DmsA_L-DmsD fusion protein ......................................... 63
Figure 3.7. Leader peptide moieties on the DmsA_L-DmsD fusion protein dimer are not accessible to free DmsD .............................................................................. 64
Figure 3.8. Schematics of the possible interactions between the protomers in the fusion protein dimer. ........................................................................................................... 65
Figure 4.1. Assigned \textsuperscript{15}N-HSQC TROSY spectrum of DmsD (pH 6.5, 25°C). .......... 71
Figure 4.2. Secondary structure prediction and NOE relaxation of the DmsD protein .............................................................................................................................. 73
Figure 4.3. Chemical shift perturbation analysis of DmsD in the presence and absence of the DmsA_L peptide ...................................................................................... 75
Figure 5.1. Root mean squared fluctuation (RMSF) plotted against residue number for the DmsD protein simulation ................................................................. 82
Figure 5.2. DmsD after 58 ns of simulation ...................................................................... 82
Figure 5.3. Docking and molecular dynamics simulation experiments to predict the DmsD/DmsA leader peptide interactions ................................................................. 83
Figure 5.4. The DmsD protein and the full length DmsA leader peptide after 50 ns of simulation .............................................................................................................. 86
Figure 5.5. A proposed model for the DmsA_L-DmsD fusion protein dimer .................... 87
Figure 5.6. Proposed model for DmsA_L/DmsD interaction ........................................... 90
Glossary

(v/v) A volume / volume ratio. 10 mL in a 1 L solution is a 1% (v/v) solution

Å Ångstrom (10⁻¹⁰ m)

B-factor Also called ‘temperature-factor’. Describes the variation in the position of an atom in an X-ray crystal structure. Higher B-factor values are indicative of higher disorder or mobility.

Chaotrope A chemical that interrupts the secondary structure of protein, causing denaturation

Chemical shift Resonant frequency of a nucleus relative to a standard. It is reported in ppm.

\[ \delta = \frac{V - V_0}{V_0} \]

Where \( V_0 \) is the resonant frequency of that standard.

Completeness The number of reflections from a data set divided by the number expected at a given resolution

Constraints In molecular dynamics, fixed bond lengths or bond angles.

Crystal A regular repeat of molecules in three dimensions

Crystallographic refinement An iterative process of improving the agreement between the molecular model and the crystallographic data.

CV Column Volume

Cytoplasm The area encased by the plasma membrane of a cell

de novo Anew, from the beginning, with no prior data.

Didermic Any bacterium that has both an inner and outer membrane separated by a periplasmic space

DmsA The catalytic subunit of DMSO reductase, it uses a molybdopterin cofactor to reduce DMSO to DMS.

DmsD The REMP chaperone that is specific for DMSO reductase, encoded by the \( dmsD \) gene

Docking The process of fitting a molecule to a partner molecule that maximizes favourable interactions while minimizing unfavourable interactions

Electron density map A 3 dimensional representation of the electrons present in a crystal structure, averaged across the entire crystal. Mathematically, electron density is the Fourier transform of the structure factors and is described by the following equation:

\[ \rho(x, y, z) = \frac{1}{V} \sum_{hkl} F_{hkl} \exp[-2\pi i(hx + ky + zl)] \]

such that \( \rho(xyz) \) is the electron density function
<table>
<thead>
<tr>
<th>Elution volume</th>
<th>The volume of mobile phase between the start of the separation and the exit of a solute from a size exclusion chromatography column.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Force-field (molecular dynamics)</td>
<td>Forces derived from a set of parameters and potential functions that are applied to atoms that are being simulated. These parameters are derived from empirical observation and are used in the potential functions to calculate the forces that each atom exerts on each other atom.</td>
</tr>
<tr>
<td>Holoenzyme</td>
<td>A complex that consists of an enzyme and cofactor.</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear Single Quantum Correlation, a 2 dimensional NMR experiment that results in spectrum with the chemical shift of $^1$H on one axis and the non-hydrogen nucleus (generally $^{15}$N or $^{13}$C for protein NMR) on the other axis.</td>
</tr>
<tr>
<td>in vitro</td>
<td>Latin for “within glass”. Describes experiments that isolate specific parts of an organism and examines them in a synthetic context.</td>
</tr>
<tr>
<td>in vivo</td>
<td>Latin for “within the living”. Describes experiments that take place in a living organism.</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-thiogalactopyranoside.</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilodalton, a non-SI unit for molecular mass equal to 1000 Daltons.</td>
</tr>
<tr>
<td>MALS</td>
<td>Multi Angle Light Scattering.</td>
</tr>
<tr>
<td>Matthews coefficient</td>
<td>The specific volume of a crystal ($V_m$); $V_m = \frac{V}{nM}$ where V is the volume of the unit cell, n is the number of asymmetric units, and M is the molecular mass of the asymmetric unit contents.</td>
</tr>
<tr>
<td>Molecular replacement</td>
<td>A method for deriving initial phases for diffraction data using a known homologous structure.</td>
</tr>
<tr>
<td>Molybdozyme</td>
<td>An enzyme with a molybdenum based cofactor.</td>
</tr>
<tr>
<td>Molybodpterin guanine dinucleotide</td>
<td>A cofactor used by some reductase enzymes to facilitate the transfer of electrons to an electron acceptor such as DMSO.</td>
</tr>
<tr>
<td>Molybtopterin</td>
<td>A class of cofactors that bind to molybdenum used in redox reactions. They are also known as pyranopterin-dithiolates.</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance, the re-emission of magnetism from nuclei in a magnetic field.</td>
</tr>
<tr>
<td>NOESY</td>
<td>Nuclear Overhauser Effect Spectroscopy. An NMR experiment that correlates resonances from atoms that are within 5 Å of each other, though are not necessarily connected by a bond.</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide Gel Electrophoresis.</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction, a molecular biology technique to amplify selected DNA sequence by several orders of magnitude.</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank.</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene Glycol.</td>
</tr>
<tr>
<td>Periodic boundary condition</td>
<td>A computational method for simulating a large system by modeling a small repeating section. The section must be large enough that any part of the section cannot be impacted by its equivalent in any adjacent section.</td>
</tr>
<tr>
<td>Periplasm</td>
<td>The area between the inner and outer membrane of a Gram negative bacterium. This region contains the peptidoglycan component of the cell wall.</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polyvinylidene fluoride.</td>
</tr>
<tr>
<td>REMP</td>
<td>Redox enzyme maturation protein, a member of a family of chaperone proteins, which are each specific for a single TAT substrate. These proteins are responsible for the correct folding and complex formation of substrate proteins, cofactor insertion, and preventing translocation until the assembly is complete. They may also be involved in membrane targeting.</td>
</tr>
<tr>
<td>Resonance assignment</td>
<td>The identification of the source atom of a resonance in an NMR spectrum.</td>
</tr>
<tr>
<td>Restraints</td>
<td>A set of limitations on an atom, bond or molecule that prevent motion beyond a set point through the imposition of a penalty for deviating from an ideal value.</td>
</tr>
<tr>
<td>R-factor &amp; R_free</td>
<td>Model quality indicators that report the agreement between a molecular model and X-ray diffraction data. R_free is calculated as R, but uses a 5% subset of randomly selected reflections that are set aside and not used in the refinement of the structural model.</td>
</tr>
</tbody>
</table>

\[
R = \frac{\sum h |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum h |F_{\text{obs}}|}
\]

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rmerge</td>
<td>An X-ray data quality indicator that reports the agreement between multiple measurements of the same reflections.</td>
</tr>
</tbody>
</table>

\[
R_{\text{merge}} = \frac{\sum_h \sum_i |l_i - \langle l_i \rangle|}{\sum_h \sum_i l_i}
\]
RMSD
The root mean squared deviation between corresponding atoms in superposed protein structures.

\[
RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \delta_i^2}
\]

Where \(\delta\) is the distance between \(N\) pairs of equivalent atoms.

RMSF
Root Mean Squared Fluctuation: The RMSD of an atom from its average position calculated for each frame of a simulation and added together, divided by the total number of frames.

\[
RMSF = \frac{\sum RMSD}{j}
\]

Where \(j\) is the total number of frames being considered.

RR
Twin Arginine, typically used in reference to the twin arginine consensus motif S/T-R-R-x-F-L-K, wherein X may be any amino acid.

SDS
Sodium Dodecyl Sulphate.

Space group
Denotes the symmetry that is present in a unit cell.

TAT proofreading
A mechanism by which the REMP prevents membrane targeting of a substrate protein until it has been properly folded and assembled.

TAT quality control
A mechanism by which the TAT translocon rejects improperly folded or assembled proteins.

TBS
Tris Buffered Saline: 20 mM Tris-HCl pH = 8.0, 100 mM NaCl.

TOCSY
Total Correlation Spectroscopy.

Tris
Tris(hydroxymethyl)aminomethane, a buffering agent with a \(pK_a\) of 8.06 at 25 °C.

TROSY
Transverse Relaxation Optimized Spectroscopy.

Unit cell
The basic repeating unit of a crystal.

Void volume
The volume of a chromatography column that is occupied by the mobile phase.

X-ray
A form of electromagnetic radiation that has a wavelength from 0.01 to 10 nm.

Zimm fit method
A plot of \(K^*c/R(\Theta)\) vs. \(\sin^2(\Theta/2)\). This plot fits a polynomial in \(\sin^2(\Theta/2)\) to the data, thereby obtaining \(M_w\) and \(<r_g^2>\) from the intercept and slope at a zero angle.
1. **Introduction & background**

1.1. **Why study the TAT system?**

The TAT system exports fully folded proteins and protein complexes. Those that do not fold correctly are not secreted to the periplasm, and are subsequently degraded by the cell by an intrinsic quality control mechanism (Matos et al., 2008; Lindenstrauss et al., 2010). The TAT has been harnessed to purify large quantities of functional recombinant protein that are laborious to produce at chemical amounts (Fisher et al., 2008). A greater understanding of the TAT pathway interactions, including the roles of accessory chaperones can enhance the production of these proteins, and enable the large-scale production of commercially important and therapeutic proteins that are otherwise challenging to synthesise (Brüser, 2007).

The TAT is used by several pathogenic bacteria to secrete virulence factors and beta-lactamase enzymes that are responsible for antibiotic resistance (McDonough et al., 2005). Disruption of TAT translocation can interfere with colonization in *Salmonella* species and *Pseudomonas aeruginosa* (Ochsner et al., 2002; Reynolds et al., 2011). These functions, coupled with the accessibility of the periplasmic face of the TAT to small molecules and the absence of the TAT translocation system in mammals, distinguishes the TAT as a novel target for antibacterial drugs.

1.2. **Type II protein secretion**

There are 6 major classifications for protein secretion in Gram negative bacteria. Type II secretion refers to the export of the substrate protein into the periplasmic space through either the general secretory system (sec) or the TAT, optionally followed by a second export stage facilitated by secretin proteins that form a large pore, permitting the substrates to cross the outer membrane (Johnson et al., 2006).
The majority of secreted proteins from bacteria use the sec system, this translocase is comprised of channel subunits SecYEG and are homologous to the eukaryotic Sec61 protein channel (Hartl et al., 1990; Meyer et al., 1999). Substrates of the sec translocon contain an amino-terminal leader peptide that targets the protein for secretion (Blobel and Dobberstein, 1975; von Heijne, 1985). During translation, the nascent polypeptide is maintained in an unfolded state by the chaperone protein SecB. The polypeptide is then threaded through the pore, powered by the ATPase activity of SecA. Upon completion of translocation, the leader peptide is removed by leader peptidase I (LepB) and protein folding takes place in the periplasm (Manting et al., 2000; Yahr and Wickner, 2000).

The sec leader peptide is typically 15 - 25 amino acid residues in length and can be divided into a polar N-terminal-region, a hydrophobic H-region and a polar C-region. If the leader peptide is to be removed in mature protein the C-terminal end of the leader peptide will contain a leader peptidase recognition sequence, alanine - x - alanine, wherein x may be any amino acid (Perlman and Halvorson, 1983). The sec translocation system is reviewed in greater detail by Papanikou et al. (2007).

1.3. The twin arginine translocase

The twin arginine translocase was discovered by observing that several exported proteins contained a characteristic motif in their N-terminal leader peptides. A consensus sequence for this motif was then assembled through multiple sequence alignment: (S/T)-R-R-x-F-L-K wherein x can be any amino acid residue (Berks, 1996). In this seminal work, Berks also demonstrated that the majority of the proteins that contained this sequence were periplasmic proteins that incorporate redox co-factors. It was then discovered that these proteins are exported by a sec-independent mechanism (Weiner et al., 1998).

The TAT in gram negative bacteria consists of four proteins, the genes for tatABC are encoded by one operon; the fourth, tatE gene, is encoded separately and is derived from a duplication of TatA (Sargent et al., 1998; Weiner et al., 1998). The TAT transmembrane complex is made up of three proteins, TatA, TatB and TatC (Figure 1.1)
The channel is comprised of an oligomer that contains a variable number of either the 9.6 kDa TatA or the 7.0 kDa TatE proteins which have an overlapping function (Sargent et al., 1998; Gohlke et al., 2005). The 18.4 kDa TatB and 28.9 kDa TatC, which are found exclusively in a 1:1 stoichiometric ratio, serve a regulatory function and can interact with the TAT leader peptide and the transport chaperone proteins associated with TAT substrates (Bolhuis et al., 2001; Papish et al., 2003; Maurer et al., 2010).

When a substrate protein interacts with the TatBC complex a number of TatA monomers will be recruited and assembled into the variable diameter pore, permitting the substrate to traverse the membrane. The translocation process is powered, in both plant thylakoid membranes and in bacteria, by the proton motive force (Mould et al., 1991; Alder and Theg, 2003; Bageshwar and Musser, 2007; Gérard and Cline, 2007).

![Figure 1.1. A schematic diagram of the twin arginine translocase.](image)

TatA/E are shown in green, TatB is blue and TatC is red.

There has been some structural analysis of the channel proteins including an assembly of the TatBC complex, both alone and bound to the substrate protein SufI, by single particle electron microscopy and random conical tilt reconstruction (Tarry et al., 2009). This demonstrated the conformational change in the TatBC complex upon substrate binding and provided some insight into the mechanism of the channel function. In addition to this structure, a conical tilt reconstruction of the TatA polymer has revealed that the channel is flexible in its diameter with a variable number of TatA monomers.
incorporated into the complex (Gohlke et al., 2005). This was recently supplemented by an NMR solution structure of the *Bacillus subtilis* TatA protein (Hu et al., 2010). Solid state NMR analysis has provided an image of how the TatA monomer behaves in the membrane and gives clues to the details of polymer assembly (Walther et al., 2010).

An analogous TAT system exists in chloroplasts (cpTAT) with many similar features, reviewed by Cline and Dabney-smith (2008). The TAT proteins Tha4, Hcf106 and cpTatC are homologous to the bacterial TatAB & C proteins respectively. The cpTAT is powered by the proton motive force. An influx of protons is observed across the thylakoid membrane during translocation, leading to the inference of the ΔpH gradient as the energy source for the cpTAT.

A twin arginine translcoase has also been identified in Gram positive bacteria, which consists of two variants of each of the TatA and TatC proteins which form two distinct translocons, TatAdCd and TatAyDy (Ridder et al., 2009). The NMR solution structure of TatAd has been solved and is largely similar to that of the Gram positive TatA protein (Hu et al., 2010), however very little is known regarding its function (Barnett et al., 2008).

### 1.4. A model TAT substrate protein: DMSO reductase

The substrate proteins for the TAT system are often soluble periplasmic reductases that contain redox sensitive cofactors (Weiner et al., 1998). The TAT, however, is also capable of inserting proteins into the membrane (Hatzixanthis et al., 2003). The substrate protein for DmsD is the DMSO reductase protein complex, a heterotrimer comprised of DmsA, DmsB and DmsC, along with a molybdopterin (MPT) guanine dinucleotide (MGD) cofactor (Rothery et al., 1995). DmsA is the 87.4 kDa molybdenum containing catalytic subunit, DmsB a 23.1 kDa protein that contains a [Fe-S] cluster, and DmsC, an integral membrane protein that anchors the complex to the membrane (Bilous et al., 1988). The DmsABC complex was thought to be located attached to the cytosolic face of the plasma membrane using fusion reporter proteins (Sambasivarao et al., 1990; Rothery and Weiner, 1993), though more recent analysis has found that there was an anomaly with the reporter protein integration and DMSO
reductase has been found in periplasmic isolations (Sambasivarao et al., 2001; Stanley et al., 2002).

The assembly of the DmsABC complex begins with DmsA being translated and bound to DmsD; DmsB and DmsC then bind to the DmsA-DmsD complex. The DmsD chaperone protein is hypothesised to coordinate this assembly process (Li et al., 2010). Cofactor loading then takes place, with the [Fe-S] being integrated into the DmsB subunit. The Mo-MGD cofactor is likely inserted into DmsA through an interaction with the enzyme responsible for the last stage of the Mo-MGD biosynthetic pathway, MobA. This interaction was observed in studies with the homologous TorA molybdozyme precursor (Genest et al., 2008). Not all TAT substrates are molybdozymes, and therefore interactions may occur with other cofactor biosynthetic enzymes, or in some cases where a redox cofactor is not required this stage would be omitted entirely. Once the assembly is complete, the enzyme complex is targeted to the TatBC complex in the membrane.

Once translocated to the periplasm, the DMSO reductase enzyme complex catalyzes the electron transfer reaction in which DMSO is reduced to dimethylsulfide (DMS). This enzyme uses a molybdopterin guanine dinucleotide redox cofactor (Figure 1.2) in which the molybdenum atom is converted between the redox states Mo\textsuperscript{VI} and Mo\textsuperscript{IV}, transporting 2 electrons to DMSO at each cycle and producing a water molecule as a by-product.

Structural information is available for the DMSO reductase enzyme from \textit{Rhodobacter capsulatus} (PDB: 4DMR) (McAlpine et al., 1998). The primary sequence of the \textit{R. capsulatus} DMSO reductase bears a higher similarity to the \textit{E. coli} Trimethylamine N-oxide (TMAO) reductase at 48 \%, than to the \textit{E. coli} DMSO reductase at 26 \% (Shaw et al., 1996) however, it is likely that the active sites of the \textit{E. coli} and \textit{R. capsulatus} DMSO reductase enzymes are similar (George et al., 2007).
Cofactors of the TAT substrates.

Molybdopterin and molybdopterin guanine dinucleotide are the primary redox cofactors found within the TAT substrates. The presence of one of these cofactors catalyzes the reduction of terminal electron acceptors such as DMSO or TMAO.

1.5. The TAT leader peptide

Though the functions of the TAT substrates are diverse, each substrate protein must contain a TAT leader peptide. The TAT leader peptide shares many features with the canonical sec leader peptide, each bears a polar N-terminal region, a central hydrophobic region and a C-terminal region with the leader peptidase I cleavage site (Tullman-Ercek et al., 2007). The most prominent difference between the sec and TAT leader peptides is the presence of the conserved twin arginine motif (Figure 1.3), this is found at the interface of the N-terminal region and the hydrophobic region of the peptide (Figure 1.4) (Berks, 1996). In addition to this, TAT leader peptides tend to be longer than their sec counterparts, have a less hydrophobic H-region and will typically contain a positive charge in the C-terminal region of the leader peptide (Figure 1.4). This charged
region has been dubbed the 'sec avoidance motif' (Cristóbal et al., 1999; Tullman-Ercek et al., 2007). These features have been used to discriminate sec leader peptides from TAT leader peptides with confidence using bioinformatic analysis (Bagos et al., 2010).

![Figure 1.3. Sequence alignment of the E. coli leader peptides.](image)
The conserved regions (yellow) correspond directly to the RR motif, the two arginine residues are absolutely conserved (red). These sequences were obtained from a list maintained by Dr. Tracy Palmer (Palmer, 2012).

![Figure 1.4. A schematic representation of the TAT and sec leader peptides.](image)
The TAT leader peptide is longer, contains the RR motif (S-R-R-x-F-L-K), has a less hydrophobic H-region and a positively charged C-region.
The twin arginine motif functions to bind to protein chaperones and prevent degradation of the substrate, target the folded protein to the membrane, and bind to the translocon (Sambasivarao et al., 2000). Not all of these functions are conserved across all of the leader peptides. In the case of TtrD from A. fulgidus, the conserved arginine residues are required for translocation but not for interaction with the chaperone protein (Coulthurst et al., 2012). While the RR motif is the most salient feature of these leader peptides, the functional region of the leader peptide is not limited to that region. Disruption of the hydrophobic region has been shown to abolish the function of the TorA leader peptide, or change the specificity to that of a different REMP by exchanging the hydrophobic region of the leader peptide for that of another substrate (Buchanan et al., 2008; Shanmugham et al., 2012).

In solution, the twin arginine leader peptide has been shown to be unstructured through NMR and molecular dynamics studies (San Miguel et al., 2003). Additionally, the TorA leader peptide appears to be degraded in the cytosol when it is expressed in strains that lack the corresponding chaperone protein. In the case of TAT substrates such as SufI, that do not bind to these chaperones, the leader peptide is resistant to proteolytic degradation (Genest et al., 2006). The leader peptide is expected to have an α-helical conformation when it is bound to a chaperone protein, and this has been demonstrated in the NMR solution structure of E. coli NapD chaperone bound to residues 1-24 of the 35 residue NapA leader peptide (PDB 2PQ4). A similar helical conformation has been predicted for the TorA leader peptide based on the NMR structural analysis of the NarG peptide in solution (Buchanan et al., 2008; Zakian et al., 2010).

### 1.6. Redox enzyme maturation proteins

The chaperone proteins that bind to a nascent TAT leader peptide of a given substrate are collectively referred to as Redox Enzyme Maturation Proteins (REMPs) (Turner et al., 2004). The two most thoroughly studied examples of this class of proteins are TorD and DmsD. The TorD protein was discovered as a part of the TorCAD operon that codes for the Trimethylamine-N-oxide (TMAO) reductase protein, though it was not
part of the final holoenzyme complex (Pommier et al., 1998). The next of these redox enzyme chaperone proteins to be characterized was the DmsD protein encoded by the *ynfl* gene, a chaperone specific for the dimethylsulfoxide (DMSO) reductase molybdozyme (Oresnik et al., 2001).

Biochemical analysis of the members of the REMP family has revealed several roles for REMPs. The first is binding and protection of leader peptides. Each REMP binds exclusively to a single pre-protein substrate, for example, DmsD binds to preDmsA, TorD binds to preTorA (Ray et al., 2003; Turner et al., 2004; Chan et al., 2009). The other functions of these proteins are to prevent translocation until the substrate protein is properly folded, and to maintain the substrate protein in a form that is competent to receive a cofactor (Ray et al., 2003; Hatzixanthis et al., 2005; Genest et al., 2006, 2009).

The REMPs were sub-classified into families according to sequence similarity and structural content (Figure 1.5). These include the TorD-like, all α-helical, group containing TorD, DmsD, NarW/J and YcdY among others, and the α/β group that have a mixture of α-helical and β-sheet secondary structure. This group consists of the ferredoxin-like NapD, the unique HybE and FdhE, and the thioredoxin-like HyaE families (Figure 1.7). Structural studies of the REMP family has yielded several X-ray and NMR structures (Table 1.1, 1.2 & 1.3).
Figure 1.5. Phylogram of the REMP family proteins in E. coli divided into all α-helical and α/β types.

This phylogram is based on a sequence alignment of all of the known REMPs in E. coli. This alignment and separate alignments for each type may be found in appendix D. The phylogram was generated using the program FigTree (Rambaut, 2007).
1.7. The structure of all-α (TorD-like) REMPs

The TorD protein was the first REMP with a structure available (PDB: 1n1c) (Figure 1.6 A) and exhibits a domain swapping interaction between two chains, which forms a barbell-like shape with two lobes connected by an unstructured linker region from each of the monomers (Tranier et al., 2003). This structure revealed a novel all α-helical fold contained within each lobe, which is comprised of 11 α-helices arranged in a bundle. The dimerization occurs such that the 6 N-terminal α-helices of one monomer interact with the 4 C-terminal α-helices of the other monomer. A hinge at residues 125-135 between parallel α-helices 7 & 8 connects the two subdomains and is present in all of the structures of TorD homologs, but the sequence of the hinge is not well conserved.

A. TorD from S. massilia PDB: 1n1c B. DmsD from S. typhimurium PDB: 1s9u C. A. fulgidus NarJ, AF0173. PDB: 2o9x D. A. fulgidus AF01060 TtrD, PDB: 2idg E. E. coli DmsD PDB: 3efp F. E. coli DmsD PDB: 3cw0.

Figure 1.6. The α-helical REMPs visualized as cartoon ribbons.
### Table 1.1. Statistics for α-type REMP X-ray crystal structures

<table>
<thead>
<tr>
<th>Organism &amp; Protein</th>
<th>Identity to E. coli DmsD</th>
<th>Length (AA)</th>
<th>PDB</th>
<th>Resolution (Å)</th>
<th>R/Rfree</th>
<th>Space Group</th>
<th>Coordinate error (Å)*</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. massilia</em> TorD</td>
<td>19.61 %</td>
<td>209</td>
<td>1n1c</td>
<td>2.40</td>
<td>0.224 / 0.255</td>
<td>P2;2;1;2;</td>
<td>0.31</td>
</tr>
<tr>
<td><em>S. typhimurium</em> DmsD</td>
<td>76.96 %</td>
<td>204</td>
<td>1s9u</td>
<td>1.38</td>
<td>0.163 / 0.185</td>
<td>C2</td>
<td>0.05</td>
</tr>
<tr>
<td><em>E. coli</em> DmsD</td>
<td>100%</td>
<td>204</td>
<td>3efp</td>
<td>2.01</td>
<td>0.180 / 0.219</td>
<td>P3;21</td>
<td>0.13</td>
</tr>
<tr>
<td><em>E. coli</em> DmsD</td>
<td>100%</td>
<td>204</td>
<td>3u41</td>
<td>2.50</td>
<td>0.228 / 0.263</td>
<td>P2;2;1;2;</td>
<td>0.32</td>
</tr>
<tr>
<td><em>E. coli</em> DmsD</td>
<td>100%</td>
<td>204</td>
<td>3cw0</td>
<td>2.40</td>
<td>0.197 / 0.240</td>
<td>P4;2;1;2;</td>
<td>0.40</td>
</tr>
<tr>
<td><em>A. fulgidus</em> NarJ</td>
<td>15.69 %</td>
<td>159</td>
<td>2o9x</td>
<td>3.40</td>
<td>0.239 / 0.285</td>
<td>P6;22</td>
<td>1.25</td>
</tr>
<tr>
<td><em>A. fulgidus</em> TtrD (AF0160)</td>
<td>12.75 %</td>
<td>174</td>
<td>2idg</td>
<td>2.69</td>
<td>0.247 / 0.297</td>
<td>P2;2;1;2;</td>
<td>3.82</td>
</tr>
<tr>
<td><em>A. fulgidus</em> TtrD</td>
<td>12.75%</td>
<td>176</td>
<td>2yjm</td>
<td>1.84</td>
<td>0.216 / 0.245</td>
<td>C2</td>
<td>0.20</td>
</tr>
<tr>
<td><em>A. fulgidus</em> TtrD</td>
<td>12.75%</td>
<td>176</td>
<td>2y6y</td>
<td>2.20</td>
<td>0.218 / 0.238</td>
<td>P3;21</td>
<td>0.16</td>
</tr>
<tr>
<td><em>A. fulgidus</em> TtrD</td>
<td>12.75%</td>
<td>176</td>
<td>2xol</td>
<td>1.35</td>
<td>0.152 / 0.178</td>
<td>P2;1</td>
<td>0.05</td>
</tr>
</tbody>
</table>

* Based on R value

### Table 1.2. Statistics for α/β REMP X-ray crystal structures

<table>
<thead>
<tr>
<th>Organism &amp; Protein</th>
<th>Length (AA)</th>
<th>PDB</th>
<th>Resolution (Å)</th>
<th>R/Rfree</th>
<th>Space Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>R. palustris</em> HupG</td>
<td>133</td>
<td>2qsi</td>
<td>1.80</td>
<td>0.155 / 0.201</td>
<td>P1</td>
</tr>
<tr>
<td><em>S. typhimurium</em> HyaE</td>
<td>134</td>
<td>2es7</td>
<td>2.80</td>
<td>0.285 / 0.338</td>
<td>P2;1</td>
</tr>
<tr>
<td><em>S. flexneri</em> HyaE</td>
<td>132</td>
<td>2qgv</td>
<td>2.70</td>
<td>0.253 / 0.292</td>
<td>P2;1</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> FdhE</td>
<td>309</td>
<td>2fiy</td>
<td>2.10</td>
<td>0.199 / 0.243</td>
<td>P222;1</td>
</tr>
</tbody>
</table>
Table 1.3. Statistics for REMP NMR structures

<table>
<thead>
<tr>
<th>Organism &amp; Protein</th>
<th>Type</th>
<th>Length (AA)</th>
<th>PDB</th>
<th>Total No. of Restraints*</th>
<th>Backbone RMSD*</th>
<th>Heavy Atom RMSD*</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli NapD</td>
<td>α/β</td>
<td>87</td>
<td>2jsx</td>
<td>1340</td>
<td>8.63 +/- 2.06</td>
<td>9.71 +/- 2.18</td>
</tr>
<tr>
<td>E. coli NapD &amp; NapA&lt;sub&gt;L&lt;/sub&gt;</td>
<td>α/β</td>
<td>87</td>
<td>2pq4</td>
<td>Unknown</td>
<td>2.56 +/- 0.53</td>
<td>2.80 +/- 0.49</td>
</tr>
<tr>
<td>E. coli HyaE</td>
<td>α/β</td>
<td>132</td>
<td>2hfd</td>
<td>1523</td>
<td>2.5 +/- 0.70</td>
<td>2.92 +/- 0.59</td>
</tr>
<tr>
<td>S. typhimurium HyaE</td>
<td>α/β</td>
<td>134</td>
<td>2jzt</td>
<td>1541</td>
<td>5.94 +/- 1.13</td>
<td>6.68 +/- 1.17</td>
</tr>
<tr>
<td>S. typhimurium HyaE</td>
<td>α/β</td>
<td>134</td>
<td>2gzp</td>
<td>1191</td>
<td>3.78 +/- 0.71</td>
<td>4.11 +/- 0.65</td>
</tr>
<tr>
<td>E. coli HybE</td>
<td>α/β</td>
<td>162</td>
<td>2kc5</td>
<td>4257</td>
<td>10.82 +/- 3.00</td>
<td>10.94 +/- 2.95</td>
</tr>
</tbody>
</table>

* Data from NRG-CING (Doreleijers et al., 2012)

There are now 3 solved structures of DmsD in the PDB (Figure 1.6 B, E, F), the first from Salmonella typhimurium (PDB: 1s9u) (Qiu et al., 2008) and two from E. coli in different crystal forms (PDB: 3efp; 3cw0) (Ramasamy and Clemons, 2009). These three structures superpose quite well, with a C-alpha RMSD of 0.472 Å, though density is not observed for the hinge region of the S. typhimurium DmsD structure. As was mentioned before, the monomeric REMP exemplified by DmsD is architecturally very similar to a single lobe of the dimeric TorD barbell, consisting of 11 α-helices. The largest differences between these two structures are the longer C-terminal helix in TorD, and the angle of helix 7 of TorD, which is found immediately before the hinge region that is nearly perpendicular to that of the equivalent DmsD helix. Finally a single turn helix is observed following the hinge region in the TorD structure, which is absent from DmsD. It is tempting to speculate as to whether these structural differences in the regions close to the hinge region are related to the dimerization of the REMP molecule, or due to the sequence variation between DmsD and TorD. Structural data for multiple forms of a single REMP would be of great help in drawing those conclusions.
The only other structures of a TorD like REMP are from the Archaean *Archaeoglobus fulgidus*. The proteins AF0173 and AF0160, now known as TtrD (Figure 1.6 C, D), have structural similarity to the TorD family. The AF0173 protein (PDB: 2o9x), originally thought to be a DmsD (Kirillova et al., 2007), has been reclassified as a NarJ (Sargent, 2007). This AF0173 protein is truncated to 153 residues but is structurally similar to the first 153 residues of DmsD and TorD proteins. From the same organism, Af0160 also known as TtrD (PDB: 2idg, 2yjm, 2yby, 2xo1) (Coulthurst et al., 2012) is a member of the TorD-like family as classified by SCOP (Murzin et al., 1995). It has a similar overall architecture to other TorD-like REMPs and binds to the leader peptide of TtrA, a tetraethionate reductase.

### 1.8. The structure of α/β REMPs

Each of these α/β REMPs shares some structural and functional characteristics with the TorD-like REMPs. They each bind to the twin arginine leader peptide of their substrate proteins, and there is evidence for a proofreading function and homodimerization. This is in spite of a very different overall structure for these proteins, the α/β REMPs do not have a single common fold, though several of them exhibit either ferredoxin-like folds for NapD chaperones (Maillard et al., 2007), or thioredoxin-like folds for HyaE and HupG (Parish et al., 2008)

The NMR solution structure of NapD was solved alone (A) and in complex with the NapA leader peptide (Figure 1.7 B) and revealed a ferredoxin-like fold (Maillard et al., 2007). The structure of the complex showed the helical conformation of the NapA leader peptide when bound by the REMP. The NapD structure consists of a single 4 stranded β sheet with two helices packed against a single face of the sheet. The structure has been complemented with some biochemical analysis, demonstrating that the NapA_L and NapD interaction is favourable from an enthalpic and entropic point of view (Maillard et al., 2007).

HyaE has a thioredoxin-like fold, consisting of a twisted, 5-stranded beta-sheet contained within 5 α-helices (Parish et al., 2008). There have been 3 structures solved of HyaE, an x-ray structure *E. coli* (Figure 1.7 D), and both x-ray and NMR solution
structures from *S. typhimurium* (Figure 1.7 E F). The three HyaE structures superpose with RMSD values less than 2.0 Å² when the alpha carbons are superposed. There is little biochemical data to explain any structure-function relationships for the HyaE proteins, though they have been shown to bind to the HyaA enzyme precursor by a bacterial two hybrid experiment (Dubini and Sargent, 2003). This fold is known to be shared by one other REMP, *Rhodopseudomonas palustris* HupG (PDB: 2qsi) (Figure 1.7 G). The structure of HupG exhibits a domain swapping interaction in which a helix and strand from one monomer contribute to the structure of the second monomer.

*Pseudomonas aeruginosa* FdhE (PDB: 2fiy) (Figure 1.7 C) and *E. coli* HybE (PDB: 2kc5) (Figure 1.7 H) (Shao et al., 2009) have folds that are unique in the PDB. FdhE, like the other REMPs, binds to and plays a role in the biogenesis of the redox enzyme formate dehydrogenase, but it is not part of the completed holoenzyme. Little is known about how these functions are carried out by the FdhE, or to what extent the functional roles are conserved. Based on biochemical studies on the FdhE in the *E. coli* homolog, there is evidence that leader peptide recognition is not the primary means by which FdhE proteins act on their substrate, the FdnGHI complex (Schlindwein et al., 1990; Stewart et al., 1991; Lüke et al., 2008).

HybE is structurally distinct from the other α/β REMPs (Shao et al., 2009). It bears a three-layer sandwich fold, in which there are two α-helices that encase a core made up of 8 beta strands. Structural analysis has led to the proposition of a HybO leader peptide-binding cleft that is comprised of conserved residues, though this is yet to be confirmed through biochemical analysis (Shao et al., 2009). The diverse family of REMPs is an intriguing case of convergent evolution as the α/β REMPs are evolutionarily distinct from the TorD-like REMPs yet share very similar functions.
Figure 1.7. Structures of the α/β REMP proteins.

A. the NapD protein from *E. coli* alone (PDB: 2jsx) and B. in complex with the NapA leader peptide shown in white (PDB: 2pq4). C. FdhE from *P. Aeruginosa* (PDB: 2fiy). D. HyaE from *E. coli* (PDB: 2hfd), and E. *S. typhimurium* HyaE solved by X-ray crystallography (PDB: 2es7) and F. NMR (PDB: 2jzt; 2gzp). G. *R. palustris* HupG (PDB: 2qsi) as a domain swapped dimer, with one monomer shown in black, while the other is shown in white. H. *E. coli* HybE (PDB: 2kc5).
1.9. The REMP-leader peptide interaction

The best characterized role of the REMP is binding to the leader peptide of TAT substrates. Upon their discovery, it was noted that TorD binds specifically to pre-TorA, while DmsD binds with high affinity to pre-DmsA and with lower affinity to pre-TorA (Pommier et al., 1998; Oresnik et al., 2001; Ray et al., 2003). The affinity of each of these proteins for their proper substrates, and for the substrates of other REMPs, have been quantified and compared. While there was some promiscuity discovered in *in vitro* studies wherein DmsD was able to bind to the leader peptide of pre-TorA, DmsD and TorD, REMP proteins are not interchangeable *in vivo* (Genest et al., 2005; Chan et al., 2009). This is consistent with the discovery that the regions of the sequence of TorD that define specificity are not well conserved among REMP family members (Genest et al., 2009).

Using isothermal calorimetry and far-western blotting, several experiments have characterized the DmsA/DmsD binding interaction. DmsD was shown to bind to the DmsA leader peptide when it was genetically fused to the N-terminus of glutathione-S-transferase (GST), but not when it was fused downstream of GST, suggesting that the leader peptide must be N-terminal to the substrate protein in order to be bound by DmsD (Winstone et al., 2006); however the dimerization of the GST fusion may have interfered with this result. The measured Kₐ for the interaction between DmsD and the DmsA leader peptide was 0.2 µM, a value comparable to that reported for the TorD and TorA leader peptide interaction at 1 µM (Hatzixanthis et al., 2005; Winstone et al., 2006).

REMP-substrate interaction is not solely dependent on the presence of an unaltered twin arginine motif. The hydrophobic region of the leader peptide is also required for interaction with the REMP, as was discovered using truncated leader peptide constructs and protein interaction assays with TorD, DmsD, FdhE, NarJ, NarW, NapD and their substrates (Hatzixanthis et al., 2005; Chan et al., 2009). Further exploration revealed that the hydrophobic region confers specificity for a given REMP and can be exchanged (Shanmugham et al., 2012).

When it is bound to a REMP, the leader peptide is protected from degradation by cytosolic proteases (Genest et al., 2006). This work explored the state of the pre-TorA
protein in the presence and absence of TorD, noting that the N-terminal 41 residues were removed from pre-TorA in strains deficient in the TorD gene, while the leader peptide was unaltered when TorD was present (Genest et al., 2006).

**Figure 1.8. The putative leader peptide binding surface of DmsD.**

A. Conserved motifs and residues shown to be important for leader peptide binding are mapped to the structure of *E. coli* DmsD (pdb: 3efp). These residues are as follows: 18, 22, 64, 68, 72, 75, 76, 86, 87, 93, 95, 123, 124, 126, 127, 147, 151, 154, 172, 175 with the following mutations having the largest effect: 72, 75, 76, 86, 124, 126, 127, 147, 151, 172 while mutating residues 87 and 95 increased the affinity of DmsD for the leader peptide.

B. The same protein after docking and molecular dynamics simulation. The leader peptide is shown in dark green and the residues within 5 Å of the leader peptide highlighted in green.

C. These regions of DmsD are highlighted on a schematic of the DmsD protein.

With the rapidly accumulating wealth of structural information pertaining to the REMP family, protein interaction data can be visualized in terms of the structural features that are responsible for a given function. In studies involving DmsD, leader peptide binding is performed by residues that were identified by bioinformatic analysis as conserved residues, followed by mutation of these residues and binding assays (Chan et al., 2008). The mutated residues that were found to be important for binding include 18,
22, 64, 68, 72, 75, 76, 86, 87, 93, 95, 123, 124, 126, 127, 147, 151, 154, 172, 175, with residues 72, 75, 76, 86, 124, 126, 127, 147, 151, 172 contributing most to the binding interaction. The mutation of two residues, 87 and 95, increased the affinity of the DmsD protein for the DmsA leader peptide. Together, many of these residues form a “hot pocket” on the surface of the DmsD molecule in the region of the molecule between helices 5 & 6 and helices 7 & 8 (Chan et al., 2008). Two conserved $123E(P/x)(x/P)DH$ and $72WxxLF$ motifs that are conserved across the entire REMP family were also implicated in binding (Figure 1.8) (Turner et al., 2004).

In the case of the TorD / TorA interaction, the REMP has been shown to interact with the core of the targeted substrate protein. Residues 79 and 83 of TorD are primarily responsible for this interaction, congruent with observations regarding REMP-substrate specificity, these residues are not conserved between members of the REMP family (Genest et al., 2008). This behaviour is not unique to TorD; NarJ was also found to interact with the core of the NarG protein and the NarG leader peptide (Vergnes et al., 2006). These findings have spurred a system-wide analysis for the REMP proteins in which it was discovered that HybE binds to the mature HybO protein, while FdhE binds to the full FdoG pre-protein, but not the leader peptide or the mature protein alone (Chan et al., 2010). A study involving the analysis of A. fulgidus AF1060, now renamed TtrD, has demonstrated that the portion of the TtrA leader peptide that is required for binding to the TtrD chaperone includes residues 7-17, which begins immediately after the RR motif (Coulthurst et al., 2012).

1.10. Proofreading and substrate assembly

The proofreading function of the REMPs can be summarized as a mechanism by which the substrate is held in the cytosol by the chaperone until it is correctly folded, with each of the subunits and cofactors in place. The TAT substrates that require REMPs to fold properly are not efficiently translocated when REMPs are unavailable, while proteins such as GFP, that do not require chaperones are readily translocated when fused to a TAT leader peptide (Ray et al., 2003). This is distinct from TAT quality control, a process by which misfolded substrate proteins are rejected by the TAT pore (Jack et al., 2004).
The mechanism through which the folded state of the substrate protein is sensed by the REMP remains unknown, but is an interesting topic for further study.

For many TAT substrates, the incorporation of a redox sensitive cofactor such as Ni-Fe or molybdopterin guanine dinucleotide (MGD) is essential for the function of the mature enzyme (Rothery et al., 1995). The REMP has a role in this phase of substrate development beyond preventing translocation until assembly is complete. TorD has been shown to interact with the molybdopterin (MoPt) cofactor biosynthetic enzyme MobA, as well as with a precursor to MGD known as Mo-MoPt. It has been hypothesized that the REMP binding interaction with the substrate maintains the substrate in a form competent to have the cofactor loaded (Genest et al., 2008).

Once the substrate is properly folded and the cofactors inserted, the complex is targeted to the translocon. An interaction between DmsD and the membrane bound TatBC protein complex has been observed (Papish et al., 2003), indicating that the REMP has a role in this targeting event. The REMP is not always required for the leader peptide to interact with the channel, as fusion proteins with the DmsA leader peptide can be translocated by the TAT in the absence of DmsD. However, the REMP is required when the substrate protein is incapable of folding correctly without a chaperone (Ray et al., 2003).

### 1.11. GTPase / guanine nucleotide binding

The structure of TorD from *Shewanella massilia* had a DTT molecule, a part of the crystallization buffer, bound that was used to generate a hypothetical GTP binding site (Tranier et al., 2003; Hatzixanthis et al., 2005). Fluorescence quenching was used to determine the binding affinity of TorD for each of the guanosine nucleotides GTP, GDP and GMP, which were found to have $K_d$ values of 370, 385 and 338 µM respectively. This suggested that it was the guanosine moiety that was involved in the binding. Further, the affinity of TorD bound to the leader peptide for guanosine nucleotides was slightly higher with a $K_d$ of 200 µM. Mutagenesis and binding studies led to the identification of a putative GTP binding pocket in *E. coli* TorD comprised of residues 11, 16, 33, 41 and 42 (Hatzixanthis et al., 2005) (Figure 1.9).
It was recently noted that the dimeric form of TorD has a magnesium dependant GTPase activity, despite the absence of any classical nucleotide binding motifs (Guymer et al., 2010). The residues responsible for the TorD GTP binding and GTPase activity of TorD do not appear to form a larger pocket that spans across the two lobes, but rather form two separate binding sites that participate in cooperative binding wherein the binding of one GTP molecule enhances the molecule's affinity for another through subtle structural changes that propagate through the TorD dimer (Guymer et al., 2010).

Mutagenesis experiments have revealed a critical role for residue D68, as when it is mutated to a tryptophan, the specificity is altered such that the GTPase activity is converted to an ATPase activity, and TorD is no longer able to perform the proofreading function. This residue is part of a “G-4” guanosine specificity motif with the sequence TVRD. The motif lies in the loop that is situated between helices 4 and 5, which was found to be important for guanine nucleotide binding (Guymer et al., 2010).

The proposed location of this G-4 specificity loop, near the leader peptide binding groove, is consistent with the observation that having the leader peptide bound enhances the affinity of TorD for GTP. (Hatzixanthis et al., 2005; Guymer et al., 2010). Guymer et al. (2010) also suggest that GTPase activity may be a function of the trimeric form of TorD that may have been present in their samples. The multimer formation of REMP proteins will be addressed in greater detail below.
Figure 1.9. Cartoon rendering and schematic of the modeled binding pocket on a single lobe of the \textit{S. massilia} TorD structure (PDB: 1n1c).

A pocket colored blue in this figure is formed in \textit{S. massilia} TorD by residues (with \textit{E. coli} equivalents in parentheses) L12 (C11), L16 (W15), L33 (I32), F41 (W40), and W42 (F41) 31. Also highlighted in red in this figure are residues 65-68 of \textit{E. coli} TorD, which comprise the G-4 specificity loop and D68, which confers the specificity of the GTPase activity.

### 1.12. Oligomerization

An interesting property of TorD-like REMP s is their ability to form homo-oligomers. This has been observed in both DmsD and TorD, they each exist in both monomeric and dimeric form. While higher order oligomers has been observed in DmsD, TorD oligomerization appears to be limited to a trimeric form (Tranier et al., 2002; Sarfo et al., 2004). The structure of the dimeric form of TorD isolated from \textit{S. massilia} has been solved, and a domain swapped conformation is observed, that has a barbell shape wherein each lobe of the barbell has a structure similar to the monomeric structures of DmsD and NarJ (Tranier et al., 2003; Kirillova et al., 2007; Qiu et al., 2008; Ramasamy and Clemons, 2009). The dimeric form of TorD was observed to have a higher affinity for the TorA leader peptide than the more abundant monomer (Tranier et al., 2002). All of the folded forms of DmsD bind to the DmsA leader peptide with equal affinity and it has
been established that leader peptide binding does not cause the dimerization event in DmsD or in TorD (Winstone et al., 2006).

A schematic of each of the monomeric, dimeric and trimeric forms of a TorD-like REMP has been modeled based on the monomeric and dimeric REMP structures. In these models, the hinge region crosses onto the adjacent lobe so that each individual chain of the multimer contributes to two lobes of the quaternary structure. In this model, the interactions between the N-terminal and C-terminal regions are maintained as they appear in the monomeric form and the domain swapped dimer (Figure 1.10).

While the majority of the protein observed in an over-expressed purification of a given TorD-like REMP is in the monomeric form, there is a fraction that appears as a dimer, and as a higher-order oligomer. It is possible, in some cases, to interconvert between these forms. When the pH of an S. massilia TorD preparation is lowered to the 3.0 range and then returned to physiological pH, a population of monomeric TorD is converted to its dimeric form (Tranier et al., 2002). This process does not disrupt the α-helical character of the protein, but a hydrophobic surface is exposed. When a similar process is attempted with DmsD, at pH 5.0, a distinctive “ladder” of bands is observed when visualized by native PAGE that are not representative of higher-order oligomers, but were rather described as folding intermediates whose travel through the gel matrix has been impaired (Sarfo et al., 2004). The absence of multimeric species was confirmed with size exclusion chromatography and light scattering data that demonstrate that there is exclusively one oligomeric form present in the sample. Restoring the pH to a physiological value does not cause DmsD to form oligomers, but rather aggregates that may be restored to monomers through chaotropic denaturation and refolding (Sarfo et al., 2004).
Figure 1.10. Schematic representation of the monomeric, dimeric and trimeric forms of the TorD-like REMPs.

In each case individual lobes (circles) are made up of protomers (light and dark shaded regions of the same protein), which are connected by the hinge region. The Monomeric form is shown in panel A, dimeric form in panel B, and trimeric form in panel C.

1.13. Outstanding issues in the literature

At the time that this work was started, in 2006, all of the biochemical analysis and protein characterization work had been performed using data from the DmsD or TorD proteins from *E. coli*. The only available TorD-like REMP structures were *S. typhimurium* DmsD (1s9u) (Qiu et al., 2008) *S. massilia* TorD (Tranier et al., 2003) and AF0173 (Kirillova et al., 2007), now known to be a NarJ (Sargent, 2007). Additionally, each of the DmsD and TorD strictures are incomplete, missing electron density for regions of the protein such as the hinge region of the *S. typhimurium* DmsD, and several regions
across the *S. massilia* TorD structure. The limited sequence similarity between these proteins and the *E. coli* equivalents adds a degree of difficulty to the interpretation of the functional data in terms of protein structure.

In addition to the limitations to the structural data, the binding site that was mapped by mutagenesis was limited to a small selection of both conserved and randomly selected residues of DmsD (Chan et al., 2008). This provides a good, albeit incomplete, picture of the DmsA/DmsD interaction. The entire leader peptide is protected by REMP interaction (Genest et al., 2006); this suggests that the “hot pocket” identified by mutagenesis is insufficient for the full interaction that protects a 45 amino acid residue leader peptide from proteases.

Finally, REMPs are responsible for executing a number of functions for their substrate proteins, including leader peptide binding (Pommier et al., 1998; Oresnik et al., 2001), cofactor loading (Genest et al., 2008), membrane targeting (Papish et al., 2003), and proofreading (Jack et al., 2004). The structural basis for these roles is still not understood. There is no evidence for how DmsD binds to TatBC, the Molybdenum cofactor biosynthetic enzymes or other accessory chaperones such as trigger factor. Nor has there been any study of the mechanism for substrate release upon the satisfaction of the proofreading function. Because all of these functions are likely to require or cause conformational changes in response to a protein-protein interaction the identification of the flexible regions of the proteins as candidates for these interactions is a starting point for their understanding.
1.14. Objectives

The primary objectives of this study were:

1. To solve the structure of *E. coli* DmsD
2. To understand the dynamics of the DmsD molecule
3. To examine the molecular details of the DmsA leader peptide – DmsD binding interaction

Objective 1: The structure of *E. coli* DmsD was addressed through X-ray crystallography. The structure was solved in space group P3_121 using molecular replacement to generate the phases of the structure factors. The DmsD protein from *S. typhimurium* PDB: 1S9U (Qiu et al., 2008) was used as a search model, and 2 molecules were found in the asymmetric unit. An additional structure was solved in space group P2_12_12 with 8 molecules in the asymmetric unit. The satisfaction of this objective gave a structure of DmsD from the same organism that the biochemical are conducted in, and an accurate structure upon which to interpret existing biochemical data. These structures are the first with electron density present for the entire DmsD molecule.

Objective 2: The dynamics of the DmsD molecule were analyzed through a multidisciplinary approach that used data from the X-ray crystal structures as well as NMR backbone amide relaxation experiments and molecular dynamics simulation. Taken together, these methods demonstrate which regions of the protein are dynamic, and when combined with biochemical and phylogenetic data elucidate the functional importance of these regions. The completion of this objective identifies regions that are likely to be involved in the function of the protein, as the interaction with other proteins, or the release of the substrate will probably involve a structural change. These flexible regions may be a starting point for further study of DmsD as a model REMP.
Objective 3: The details of leader peptide binding were examined using NMR chemical shift perturbation analysis by comparing between the spectra produced by the DmsD molecule alone and the DmsA leader peptide / DmsD complex. This was augmented by docking and molecular dynamics simulation to construct a model for the chaperone / peptide complex. This new model builds on the previously identified “hot pocket” (Chan et al., 2008), and provides an expanded view of the interaction that is consistent with the understanding that the hydrophobic region of the leader peptide is important in the binding interaction (Shanmugham et al., 2012). This model was produced with chemical shift information from 160 non-proline residues, in addition to those that were identified previously.

The results that address each of these objectives are divided into four chapters according to the methods undertaken at each stage. Structural information about the DmsD molecule was obtained through X-ray crystallographic analysis (chapter 2). This yielded limited information about the dynamics of the protein, which were further explored thorough NMR backbone relaxation analysis (chapter 4) and Molecular Dynamics Simulation (chapter 5). The exploration of the DmsD/DmsA leader peptide interaction required that the peptide moiety be purified separately from any auxiliary fusion protein such as the Glutathione-S-transferase that has been used in previous studies (Chan et al., 2008). This method is described in chapter 3. Finally the DmsA/DmsD interaction is examined through NMR chemical shift perturbation analysis (chapter 4) and Molecular Dynamics Simulation (chapter 5).
2. **X-ray crystallographic analysis of *E. coli* DmsD**

2.1. **Publication & contributions**

Portions of this chapter have been published in the Journal of Molecular Biology with the following citation:


Cloning of the pTDms67 construct used in this chapter was performed by Tara Winstone, the remainder of the experimental work was my own, under the guidance of Dr. Paetzel with whom I collaborated in the preparation of the above mentioned manuscript.

2.2. **Introduction**

Redox enzyme maturation proteins (REMPs) play an essential role in the proofreading and membrane targeting of protein substrates to the (TAT). The REMP family of chaperones consists of several proteins that each bind exclusively to a single substrate. The chaperone has a proofreading function, which protects the leader peptide from degradation and prevents translocation until all folding and assembly is complete (Genest et al., 2009). REMPs may also have a role in targeting the substrate proteins to the TAT, as they interact with channel proteins (Papish et al., 2003), although they are not essential for the translocation of properly folded substrates (Ray et al., 2003). Members of the REMP family are able to bind to redox cofactors such as molybdopterin guanine dinucleotide (MGD), and to interact with the MGD biosynthesis enzyme MobA (Genest et al., 2008).
Functionally, the most thoroughly characterized REMP to date is *E. coli* DmsD. DmsD is the chaperone specific for the leader peptide of the DmsA pre-protein, which is the catalytic subunit of DMSO reductase, a membrane bound hetero-trimer that consists of DmsA, DmsB, and DmsC, and a molybdopterin cofactor. The DmsD protein has high affinity for the DmsA leader peptide (DmsA$_L$), requiring 4 M urea to disrupt the interaction (Winstone et al., 2006). Several residues have been determined, by mutagenesis, to be involved in the DmsA$_L$-DmsD interaction.

In this chapter we report the crystal structure of *E. coli* DmsD in two different space groups at 2.0 Å and 2.5 Å resolution. The DmsD structures are unique among the REMP structures solved to date, in that they provide clear electron density for the full 204 residues of the protein. The 2.0 Å structure was also the first structure of a TAT chaperone from the model organism *E. coli*, with DmsD being one of the most thoroughly biochemically characterized TAT chaperones. Additionally, comparison and analysis of all of the available X-ray crystallographic data for DmsD provides insight into the regions of DmsD that may be involved in the function of the protein.

### 2.3. Materials & methods

#### 2.3.1. Cloning

The *E. coli* *dmsD* gene (615 bp) was amplified from the plasmid pTDms67, provided by Dr. Raymond J. Turner at the University of Calgary, using the forward primer: 5’ - ACT GGA ATT CAT GAC CCA TTT TTC ACA GCA AGA TAA TTT TTC TG - 3’ and the reverse primer: 5’ - AGC TCT CGA GCT ATC GAA ACA GCG GTT TAA CCG CG – 3’ containing the Ndel and Xhol restriction enzyme sites in the forward and reverse primers respectively. The PCR product was cloned into pET-15b (Novagen). The construct was sequenced and compared to the sequence reported in the Uni-Prot database (accession number P69853) to verify the correct amplification and insertion of the gene.
2.3.2. Expression and purification of E. coli DmsD

The DmsD protein was expressed using the plasmid pTDms67 generated previously by Winstone and colleagues (Winstone et al., 2006) in the host strain C41(DE3) (Miron and Walker, 1996). Overnight cultures were diluted (1%) into LB broth containing ampicillin (50 µg/mL), grown at 37°C for 3 h and induced with a final concentration of 1 mM IPTG for a further 3 h. Cells were harvested by centrifugation and lysed with an Avestin Emulsiflex-3C cell homogenizer. The lysate was clarified by centrifugation (30,000 x g) for 30 minutes. The supernatant was applied to a Ni\(^{2+}\) NTA column (5 mL column volume, Qiagen) equilibrated with TBS (100 mM NaCl; 20 mM Tris-HCl pH 8.0). The column was then washed with ten column volumes of TBS, followed by two column volumes of TBS with 50 mM imidazole and elution was carried out with a stepwise gradient of imidazole to 500 mM at 100 mM increments. DmsD was eluted from the column between 100 mM and 400 mM imidazole, and fractions containing DmsD were identified by SDS-PAGE, pooled and concentrated using an Amicon ultra-centrifuge filter with a MWCO of 10 kDa (Millipore). The concentrated protein was then applied to a Sephacryl S-100 HiPrep 26/60 size-exclusion chromatography column on an AKTA Prime system (Pharmacia Biotech) at 1 mL/min using TBS as the buffer. Fractions containing DmsD were analyzed by SDS-PAGE, pooled and concentrated to 36 mg/mL. The hexa-histidine tag was removed by digestion with one unit of recombinant enterokinase (Novagen) per 500 µg of protein for 24 h at 25°C according to the manufacturer's instructions. The free hexa-histidine affinity tag and un-cleaved protein were removed by application of the protein mixture to Ni\(^{2+}\) NTA resin (5 mL column volume, Qiagen) equilibrated with TBS. The protein was further purified on a Sephacryl S-100 HiPrep 26/60 size-exclusion chromatography column a second time. Fractions containing purified DmsD were analyzed by SDS-PAGE (15%), pooled and concentrated to 14.5 mg/mL using an Amicon ultra-centrifuge filter (Millipore). The final 208 amino acid protein construct consisting of the full 204 residues of DmsD and a four residue N-terminal extension (RWGS), by-product of the proteolytic removal of the hexa-histidine tag. This sequence has a calculated molecular mass of 23,831 Da and a theoretical isoelectric point of 5.0.

The plasmid pH\(_{6}\)DmsD/pET-15b prepared as described in the previous paragraph with the following modifications. The E. coli expression strain BL21(DE3) was
used instead of C31(DE3), a 1:1000 ratio of thrombin was used to remove the hexahistidine tag. Fractions containing pure DmsD were concentrated as before to 15 mg/ml.

2.3.3. **Crystallization**

P3$_{1}$2$_{1}$2$_{1}$ Crystals of DmsD were produced through hanging drop vapour diffusion. Drops were prepared by combining 1 µL of protein solution (14.5 mg/mL) with 1 µL of reservoir solution. The optimized reservoir conditions were: 100 mM Bis-tris (Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane) pH 6.5; 12 % glycerol; 1.25 M (NH$_4$)$_2$SO$_4$. The crystals were grown at 18°C. This condition was derived from an initial hit in the Hampton Research sparse matrix crystal screen #2. Optimized crystals appeared after 72 h.

Crystals in the P2$_{1}$2$_{1}$2$_{1}$ space group were grown by the hanging drop method. One microliter of 15 mg/ml protein in TBS was mixed with 1 µL of reservoir solution, which was then equilibrated overnight, with 1 mL of reservoir solution at room temperature (21°C). The condition that yielded crystals was 1.25 M (NH$_4$)$_2$SO$_4$, 100 mM bis-tris pH 6.5, 11 % (v/v) glycerol. A cryo-protectant solution was created by replacing 30 % of the water in the reservoir solution with glycerol. Crystals were washed in cryo-protectant solution for 30 seconds prior to flash cooling and storage in liquid nitrogen.

2.3.4. **Data collection**

P3$_{1}$2$_{1}$2$_{1}$ Crystals were incubated for five minutes in a cryo-protectant solution consisting of the mother liquor in which 20 % of the water was replaced with glycerol. Diffraction data were collected at the Simon Fraser University Macromolecular X-ray Diffraction Data Collection Facility using a MicroMax-007 rotating-anode micro focus generator operating at 40 keV and 20 mA, VariMax Cu HF optics, an X-stream 2000 cryo-system and an R-AXIS IV++ imaging-plate area detector (MSC-Rigaku). The data were collected and processed using the CrystalClear software package (Pflugrath, 1999). Reflections were collected beyond 2.0 Å for 180° of rotation using 0.5° oscillations. See Table 3.1 for data-collection statistics.

Diffraction data for the crystals in the P2$_{1}$2$_{1}$2$_{1}$ space group were collected at the Canadian Macromolecular Crystallography Facility (CMCF) of the Canadian Light
Source (CLS) on beam-line 08B1-1 with a MarMosaicRayonix MX300HE CCD X-ray detector. The distance from the crystal to the detector was 260.0 cm and the X-ray wavelength was 1.1049 Å. Five hundred and forty five frames were collected with an oscillation of 0.35° and a 0.5 s exposure time. Data were processed with HKL2000 (Otwinowski and Minor, 1997) and the statistics for the data collection can be found in Table 2.1.
### Table 2.1. Crystallographic statistics

<table>
<thead>
<tr>
<th>Crystal Parameters</th>
<th>3efp</th>
<th>3u41</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P3;21 (152)</td>
<td>P2;2;2,1 (19)</td>
</tr>
<tr>
<td>a,b,c (Å)</td>
<td>128.0, 128.0, 78.7</td>
<td>93.59, 95.87, 209.59</td>
</tr>
</tbody>
</table>

**Data Collection Statistics**

<table>
<thead>
<tr>
<th></th>
<th>3efp</th>
<th>3u41</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.5418</td>
<td>1.1049</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>30.7-2.0 (2.08-2.01)</td>
<td>48.34 - 2.5 (2.59 - 2.50)*</td>
</tr>
<tr>
<td>Total reflections</td>
<td>278055</td>
<td>920851</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>48639</td>
<td>107698</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.089 (0.362)</td>
<td>0.109 (0.354)</td>
</tr>
<tr>
<td>Mean (I)/(I)</td>
<td>11.5 (4.5)</td>
<td>35.8 (9.1)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.7 (95.6)</td>
<td>98.4 (98.1)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.7 (5.6)</td>
<td>9.11 (8.7)</td>
</tr>
</tbody>
</table>

**Refinement Statistics**

<table>
<thead>
<tr>
<th></th>
<th>3efp</th>
<th>3u41</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein molecules (chains)</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Residues</td>
<td>412</td>
<td>2219</td>
</tr>
<tr>
<td>Water molecules</td>
<td>411</td>
<td>539</td>
</tr>
<tr>
<td>Ligands</td>
<td>11</td>
<td>14035</td>
</tr>
<tr>
<td>Total number of atoms</td>
<td>3860</td>
<td>14035</td>
</tr>
<tr>
<td>R / Rfree (%)</td>
<td>17.8/21.2</td>
<td>22.75 / 26.32</td>
</tr>
<tr>
<td>Average B-factor (Å²) (Protein)</td>
<td>25.8</td>
<td>27.49</td>
</tr>
<tr>
<td>Average B-factor (Å²) (Waters)</td>
<td>36.9</td>
<td>24.79</td>
</tr>
<tr>
<td>Average B-factor (Å²) (Ligands)</td>
<td>57.0</td>
<td>64.6</td>
</tr>
<tr>
<td>Rms deviation on angles (°)</td>
<td>1.1</td>
<td>0.897</td>
</tr>
<tr>
<td>Rms deviation on bonds (Å)</td>
<td>0.009</td>
<td>0.006</td>
</tr>
<tr>
<td>Residues in most favoured regions (%)*</td>
<td>188 (92)</td>
<td>1341 (93.8)</td>
</tr>
<tr>
<td>Residues in additional allowed regions (%)*</td>
<td>16 (8)</td>
<td>89 (6.2)</td>
</tr>
</tbody>
</table>

The data collection statistics in brackets are the values for the highest resolution shell.

\[
R = \sum_h \frac{|F_{\text{obs}}| - |F_{\text{calc}}|}{\sum_h |F_{\text{obs}}|} \quad R_{\text{merge}} = \frac{\sum_h \sum_i |I_i - \langle I \rangle|}{\sum_h \sum_i I_i}
\]

R_{\text{free}} is calculated as R, but uses a 5% subset of randomly selected reflections that are set aside from the beginning and not used in the refinement of the structural model.

* Ramachandran plot.
2.3.5. **Structure determination and refinement**

The structure of DmsD in P3₁2₁2₁ was solved using the molecular replacement program *Phaser* (McCoy et al., 2007). The search model was provided by the structure 1s9u, the DmsD homolog from *S. typhimurium* (Qiu et al., 2008). The model was adjusted manually using the program *coot* (Emsley and Cowtan, 2004), and refinement was carried out using *refmac5* (Murshudov et al., 2011). The final round of restrained refinement with TLS restraints used TLS models generated by the TLS motion determination server (Painter and Merritt, 2006). The final refined structure was evaluated by *procheck* (Laskowski et al., 1993).

The structure of DmsD in the P2₁2₁2₁ space group was solved by molecular replacement with PHENIX (Adams et al., 2010) using chain A of a previous DmsD (pdb:3efp) as a search model. Eight molecules were found in the asymmetric unit. Density is observed for residues 2-204 in all chains, with additional density corresponding to residue 1, observed in chains A-F. The structure was refined using PHENIX (Adams et al., 2010) and *refmac5* (Murshudov et al., 2011) with manual adjustments to the model with *coot* (Emsley and Cowtan, 2004). The final iteration of refinement incorporated TLS restraints generated by the TLSMD server (Painter and Merritt, 2006). Statistics for refinement are available in Table 2.1.

2.3.6. **Structural analysis**

Superpositions were carried out using SSM superposition in the program *coot* (Emsley and Cowtan, 2004). Volume and surface area calculations were performed with UCSF chimera (Pettersen et al., 2004). Intramolecular interaction and fold analysis was performed with PROMOTIF 3.0 (Hutchinson and Thornton, 1996) The surface electrostatics analysis was performed with the adaptive Boltzmann-Poisson solver plugin (Baker et al., 2001) and displayed using Pymol (DeLano, 2002). B-factor analysis was performed by the program Baverage within the CCP4 suite of programs (Collaborative Computational Project Number 4, 1994).

Superpositions for the structural comparisons of proteins were performed using the Superpose webservice (Maiti et al., 2004). The protein co-ordinates were analysed with *procheck* (Laskowski et al., 1993), while protein-protein and protein-ligand
interaction surfaces were examined using PISA (Krissinel and Henrick, 2007). B-factor plots were generated using the NIH-Helixweb program StrucTools. The calculation of RMSD per residue was performed with Mplot, a part of the LSQMAN package (Kleywegt, 1996). The Consurf web-server was used to generate conservation data (Ashkenazy et al., 2010), while pockets were probed using the CASTp webserver (Dundas et al., 2006) with a probe radius of 1.4 Å.

2.3.7.  **Protein data bank accession codes**

Atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank with accession code 3u41 for the structure in space group P2_12_1 and 3efp for the structure in space group P3_121.

2.3.8.  **Figure preparation**

Figures were prepared using PyMol (DeLano, 2002). The alignment figure was prepared using the programs ClustalW (Thompson et al., 2002) and ESPript (Gouet et al., 1999) based on a multiple sequence alignment generated by PSI-BLAST with 5 iterations (Altschul et al., 1997).

2.4.  **Results and discussion**

2.4.1.  **P3_121 crystal Form**

2.4.1.1.  **Purification & crystallization**

The DmsD protein was purified to homogeneity, as visualized by SDS PAGE, and eluted as a single peak from size exclusion chromatography. Crystals were observed after 72 hours. The crystals belong to the space group P3_121 with unit cell dimensions of 128.0 Å x 128.0 Å x 78.7 Å with two molecules in the asymmetric unit and a Matthews coefficient of 3.9 Å³Da⁻¹ (68.6 % solvent) (Figure 2.1).
Figure 2.1. Purification and crystallization of DmsD expressed in pTDms67

A. SDS PAGE of Ni²⁺-NTA Purification, The wash step was performed with TBS, and elution buffers one and two were TBS with 50 mM and 500 mM imidazole respectively. DmsD eluted as a single band in the “Elution 2” fraction and is visible at approximately 25 kDa. B. Size exclusion chromatogram of DmsD in TBS. DmsD eluted as a single peak at 165 mL. A crystal of DmsD formed after 72 hours at 18°C using the crystallization condition 11% Glycerol, 1.25 M Ammonium sulfate and 100 mM Bis-Tris (Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane) at pH 6.5.

2.4.1.2. Structure solution

The DmsD crystal structure was refined to 2.0 Å resolution. Clear electron density was observed for all 204 amino acid residues for both molecules in the asymmetric unit (Figure 2.2). The final refined model has an R_factor of 17.8% with an R_free of 21.2%. The average B-factor of the structure is 25.8 Å² (Table 2.1). The two molecules in the asymmetric unit superposed with an RMSD of 1.11 Å for all atoms, or 0.68 Å when the superposition is restricted to the polypeptide backbone.

2.4.1.3. Protein fold

The DmsD structure is mainly α-helical (Figure 2.3). There are 11 α-helices and one 3₁₀-helix arranged in a single globular domain with approximate dimensions of 49 Å x 37 Å x 34 Å. The DmsD monomer has a surface area of approximately 8,894 Å² and a volume of approximately 26,120 Å³. The fold can be classified within the TorD-like family of proteins according to the SCOP data base (Murzin et al., 1995) but there are no domain swapping interactions in this DmsD structure as were observed in the S. massilia TorD crystal structure (Tranier et al., 2003).
Figure 2.2. Sample $2F_o-F_c$ electron density map contoured at 2.0 $\sigma$ for E. coli DmsD (PDB:3efp,3u41) structures.

The sample given is centered on residues 126 and 127 of the DmsD protein from the 3efp structure (above, green) and the 3u41 structure (below, yellow), and uses a $2F_o-F_c$ map generated from the fully refined protein models.
Figure 2.3. The fold of the DmsD protein.

A. Cartoon rendering of the protein fold of DmsD coloured spectrally from the N-terminus (Blue) to the C-terminus (Red) with the amino and carboxy termini labeled. The 11 α-helices are also numerically labeled. B. A schematic diagram of the DmsD topology and cartoon representation of DmsD. Helices represented as cylinders are labeled, as are the amino and carboxy termini, shaded and white helices correspond to the domain swapped dimer conformation observed in *S. massilia* TorD and the conserved loops that make up the putative leader peptide binding site are coloured red, green and blue. C. A cross-eye stereo view of DmsD. Every 20th residue is labeled with a sphere.
Figure 2.4. Sequence alignment of *E. coli* DmsD with homologous proteins.

The secondary structure is shown above the clustalW alignment, formatted by ESPript. The sequences were acquired from the Swissprot/TrEMBL database. Absolutely conserved residues are colored white with red fill, similar residues within groups are colored red and similar residues across groups are surrounded by a blue box. Sequences for which three-dimensional coordinates are available are highlighted in green.
Figure 2.5. DmsD conservation mapped onto the E. coli DmsD surface

Conservation scores were generated by consurf using the alignment in Figure 2.4. Individual amino acid residues are colored according to the degree to which they are conserved: absolutely conserved residues are colored maroon while highly variable residues are colored green.

2.4.1.4. Structural comparison to other REMPss

Alignment of the DmsD amino acid sequence with that of sequences for DmsD molecules from other species shows a moderate level of sequence identity (15.7 %), yet the alignment also reveals two highly conserved regions that map to 3 loops (residues 77-88, 93-100 and 113-128) on the surface of the protein (Figure 2.4 A). When the conservation is mapped on to the molecular surface of DmsD, it is clear that the region of high conservation corresponds with many of the residues that were previously determined by mutagenesis to be important for TAT leader peptide binding (Chan et al., 2008) (Figure 2.4 B). This conserved region on the surface of DmsD also corresponds to the location of the most significant depression on the DmsD molecular surface.

The superposition of E. coli DmsD on S. typhimurium DmsD yields a Ca RMSD of 0.71 Å. The E. coli DmsD structure shows the path of a presumably flexible solvent exposed loop that was unresolved in S. typhimurium DmsD (residues 116-122) (Figure 2.6). This region was not resolved in the previously solved structure of S. typhimurium
DmsD (Qiu et al., 2008). These residues lie on one of the three conserved surface loops (Figure 2.3 B), part of which forms the putative leader peptide binding site (Chan et al., 2008). In the S. massilia TorD structure a homologous loop is involved in bridging the two domains that are swapped to form the dimer (Tranier et al., 2003). The E. coli DmsD structure presented here is the first structure of a DmsD to have experimental electron density for the complete protein and therefore the complete resulting refined model. Another notable difference between the two structures is the presence of an N-terminal $3_{10}$-helix in the E. coli DmsD structure; the corresponding residues in the S. typhimurium DmsD structure are part of the neighbouring $\alpha$-helix (helix 1).

![Image of electron density for the hinge region of E. coli DmsD.](image)

**Figure 2.6. Electron Density for the hinge region of E. coli DmsD.**

All of the loops that make up the putative leader peptide binding site of E. coli DmsD are visible in the electron density. A sample of the $2F_o - F_c$ electron density map contoured at 1.0 $\sigma$ is shown here for the loop between residues 116 and 121.

When comparing E. coli DmsD with the AF0173 protein from A. fulgidus, the two structures superpose with a Cα RMSD of 2.4 Å despite having only 17 % sequence identity and AF0173 being 45 amino acid residues shorter. Interestingly, the structure of AF0173 was solved such that the AF0173 protein was bound to the TEV protease
recognition sequence of a symmetry related molecule. This is despite limited homology between the TEV protease cleavage sequence: E-N-L-Y-F-Q-S and the twin arginine motif from the preAF0174 sequence: S-R-R-D-F-I-K (Kirillova et al., 2007) in the AF0173 structure, the TEV protease cleavage sequence binds into a region of AF0173 that is on the opposite pole of the structure from the putative binding site described by Chan and colleagues (Chan et al., 2008).

When *E. coli* DmsD is compared with the TorD structure from *S. massilia*, there are a number of minor differences that can be observed. The superposition of *E. coli* DmsD with a single domain swapped monomer of *S. massilia* TorD, containing residues 1-129 from one chain and residues 130-211 from the complementary chain, yields a RMSD of 2.6 Å. The major differences arise primarily in the bridge point between the two domains (Figure 2.2 B) and at the C-termini. Notably, both *S. massilia* TorD and *E. coli* DmsD have been reported to exist in monomeric and dimeric forms (Tranier et al., 2002; Sarfo et al., 2004).

### 2.4.1.5. The putative TAT leader peptide binding pocket on *E. coli* DmsD

Previous mutagenesis work by Chan and colleagues (Chan et al., 2008) identified a number of residues on the surface of DmsD that are important for DmsA leader peptide binding. Most of these residues map to a pocket on the surface of DmsD. This putative leader peptide binding site is comprised of sections of three conserved loops (Figure 2.4). The first loop is made up of residues 77-88 and is contained between helix α5 and α6. The next loop lies between helices α6 and α7 and encompasses residues 93-100. The third loop lies between helix α7 and α8, and is made up of residues 113-128 (Figure 2.6). These loops form a curved trench along the surface of the protein approximately 17.1 Å in length (Arg204 NH1 to Leu82 Cδ1 to Trp72 CH2) and approximately 8.5 Å in width at the narrowest point (Val77 O to Glu123 Cδ). The pocket is predominantly hydrophobic with small regions of positive charge (Arg204 and Lys120) as well as a region of negative charge centered at residue Glu29 (Figure 2.7). Interestingly, in our crystal structure of DmsD we find strong electron density for five small molecules (three glycerol molecules and two tris(hydroxymethyl)aminomethane molecules) within the proposed leader peptide binding pocket of DmsD (Figure 2.7).
Electron density for a PEG molecule was found in a similar region on the structure of *S. typhimurium* DmsD (Qiu et al., 2008).

Figure 2.7. *The putative leader peptide binding pocket on E. coli DmsD.*

A. Ribbon diagram of DmsD with stick representation of the glycerol and tris(hydroxymethyl)aminomethane molecules bound in the putative leader peptide binding site. The conserved loops that make up this pocket are shown in red, green and blue. B. Surface representation of the putative leader peptide binding site on a ribbon rendering of DmsD. The electrostatic potential is mapped onto the surface. The residues that make up the proposed pocket are shown within the semi-transparent surface and labeled. Those residues previously shown, by mutagenesis, to be important for leader peptide binding are labeled with a larger font and an asterisk.
2.4.2. \textit{P2}_12\textit{1}2\textit{1} crystal form

2.4.2.1. Purification & crystallization

The DmsD protein was purified to homogeneity assessed by SDS PAGE and size exclusion chromatography, and crystals were observed after overnight incubation (Figure 2.8). These crystals were in the space group \textit{P2}_12\textit{1}2\textit{1}, with unit cell dimensions: 93.59 Å x 95.87 Å x 209.59 Å. Eight molecules were observed in the asymmetric unit and the crystal had a Matthews coefficient of 2.59 Å\textsuperscript{3}Da\textsuperscript{-1}. The crystals in the \textit{P2}_12\textit{1}2\textit{1} form were grown in identical conditions as those in the \textit{P3}_21 space group, though the protein was expressed on a different plasmid (pET15b), and had a different series of residues at the N-terminus after proteolytic removal of the hexa-histidine tag. The crystals diffracted to 2.5 Å resolution, and after refinement contained clear electron density for the entire protein molecule (Figure 2.2).

\textbf{Figure 2.8. Purification and crystallization of DmsD}

\textbf{A.} SDS PAGE of Ni\textsuperscript{2+}-NTA Purification, The wash step was performed with TBS, and elution buffers 1-5 were TBS with 100, 200, 300, 400 & 500 mM imidazole respectively. DmsD eluted as a in each of the elution fractions and is visible at approximately 25 kDa. \textbf{B.} Size exclusion chromatogram of DmsD in TBS. The DmsD monomer eluted as a single peak at 165 mL, while the DmsD dimer eluted at 136 mL. \textbf{C.} A crystal of DmsD formed after 24 hours at room temperature using the crystallization condition 12% Glycerol, 1.25 M Ammonium Sulfate and 100 mM Bis-Tris at pH 6.5.
2.4.2.2. Overall protein fold & structural features

The P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1} crystal form of \textit{E. coli} DmsD has eight molecules in the asymmetric unit (Figure 2.9 A). The structure of each individual molecule is comprised of 11 helices arranged in a bundle such that the six N-terminal \(\alpha\)-helices are attached to the C-terminal five \(\alpha\)-helices by a long loop that connects the only two parallel \(\alpha\)-helices present in the structure (Figure 2.9 B). The core of the protein is made up of helices 2, 8, 9, 10 and 11, while the remaining helices are packed up against the core but connected by longer, more flexible loops. The loops between helices 5, 6 and 7 form a groove along the surface of the protein, with a deeper pocket in the centre of the groove. This pocket has a volume of 40.8 \(\text{Å}^3\) and a solvent accessible surface area of 57.2 \(\text{Å}^2\).

\textit{Figure 2.9. General structural features of the \textit{E. coli} DmsD structure in the P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1} space group}

A. The structure of DmsD in the P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1} space group has eight molecules arranged as two groups of four in the asymmetric unit. B. Each of the monomers consists of 11 \(\alpha\)-helices, arranged in a bundle such that the N-terminal helices 1-6 are connected to the C-terminal helices 7-11 by a long loop region.

2.4.2.3. Comparison with previously solved \textit{E. coli} DmsD structures

Each of the DmsD structures superpose quite well with an RMSD of 0.5 Å. The areas with the greatest variability are at the N-terminus of the protein, the N-terminal
region of helix 4, the loops between helices 5-6, 6-7, and 7-8, and finally at the C-terminus of the protein (Figure 2.10). When the thermal motion of each of the solved DmsD structures are plotted against residue number (Figure 2.10), the result is largely in agreement with the above plot, though some differences do exist, notably between helices 2 & 3, and helices 10 & 11. This comparison also demonstrates that there is a region of lower thermal motion within helices 1, 8, 9 and 11. It is interesting that the relative increase and decrease in thermal motions are similar across structures, though the magnitudes differ. These plots indicate the greatest stability exists within helices 1, 8, 9 & 11, and a greater range of motion is available for the loop regions and helices 3, 4 & 6 (Figure 2.10). As a point of comparison, the Consurf score plotted by residue indicates that clusters of conserved residues occur at helix 1 and helices 5-7, with a number of individual conserved residues throughout the protein.
Figure 2.10. Comparison of flexibility in each of the DmsD structures.

RMSD (root mean squared deviation) of each Cα when superposed using the Superpose (Maiti et al. 2004) multiple structure alignment. Next, the superposition of plots for the deviation of the B-factor of each residue from the average B-factor of the source structure. Finally, the Consurf (Ashkenazy et al. 2010) score from most (9) to least (1) conserved is plotted as a function of residue number.

The helices of the DmsD molecule are arranged such that there are two interacting bundles of alpha helices, connected by a hinge region between helices 7 and 8 (Figure 2.9 B). This fold has been observed in each of the TorD-like REMP structures solved to date: DmsD from *E. coli* (3efp, 3cw0) (Ramasamy and Clemons, 2009), *S. typhimurium* (1s9u) (Qiu et al., 2008), *A. fulgidus* AF0173 (2o9x) (Kirillova et al., 2007), AF0160 (2idg) and *S. massilia* TorD (1n1c) (Tranier et al., 2003). The DmsD and TorD
proteins are both found in monomeric and multimeric forms (Tranier et al., 2002; Sarfo et al., 2004), although structural data is only available for DmsD in the monomeric form, and for TorD as a domain swapped homodimer (Tranier et al., 2003).

2.4.2.4. Potential protein-protein interaction site

There was electron density in this DmsD structure that did not correspond to the DmsD molecule. This density was accounted for by modelling glycerol, Tris(hydroxymethyl)aminomethane, polyethylene glycol molecules and chloride ions. Comparable density has been observed in each of the DmsD molecules solved to date and has been accounted for, in some structures, by modeling ligand molecules into the density (Figure 2.11). The solvent accessible electrostatic potential was calculated for the DmsD molecule, and there are several charged regions on the surface that form pockets corresponding to ligand interactions (Figure 2.11). The remainder of the protein carries a neutral or weakly negative charge.
Figure 2.11. The presence of ligands observed on the surface of the crystalized E. coli DmsD molecule.

A. White indicates that ligands are not found, and dark blue indicates that the ligands are frequently observed making contact with these residues. Calculations were weighted according to the number of molecules in the asymmetric unit of each structure such that one crystal form was not over-represented. B. Ligands from all solved structures are superposed around a single chain of DmsD to illustrate the frequency with which they are observed in X-ray structural data. C. The surface of the DmsD molecule colored using the solvent accessible electrostatic potential with the APBS plugin for Pymol (Baker et al. 2001). This indicates that there is a region of negative charge at the putative binding pocket, as well as along one side of the protein. These regions are also found to be involved in ligand coordination in the X-ray crystal structures.
The putative leader peptide binding pocket is made up of the loops between helices 5-6-7-8, a region shown to be important by mutagenesis studies (Chan et al., 2008) and molecular dynamics simulation that will be discussed in detail in chapter 5. This corresponds to a negatively charged pocket on the surface of the DmsD protein with a CASTp calculated volume of 58.53 Å³. This is very close to the volume of a guanidinium molecule at 48.79 Å³, suggesting that this pocket may be involved in binding one of the arginine molecules present in the leader peptide.

The homologous protein, *E. coli* TorD, has been shown to have a GTPase activity that is present only in the homodimeric form with the region responsible for this activity residing in the loop region between helices 4 & 5 (Guymer et al., 2010). This region is not shown to be involved in ligand binding in the DmsD structures, nor is it highly conserved among DmsD molecules; however, it does contribute to a negatively charged pocket on the surface of the DmsD molecule. There has not been any evidence to date of nucleotide binding or hydrolysis activity for DmsD, though members of the REMP family have been shown to interact with the molybdopterin guanine dinucleotide (MGD) cofactor and the MGD synthesis enzyme MogA (Genest et al., 2008; Li et al., 2010) as well as directly with the TAT apparatus (Papish et al., 2003); however, the molecular details of these interactions are not yet understood.

When the B-factor, RMSD of superposition, and conservation of residues are considered together, there are several regions of the DmsD protein that are both flexible and conserved (Figure 2.10). The loop regions between alpha helices 5, 6 & 7 are such regions and contain many of the residues shown to be involved in DmsAL binding (Chan et al., 2009). Conversely, the region between helices 2-5 is very flexible and the least conserved region of the protein, suggesting that if any function can be associated with this region, it would be exclusive to *E. coli* DmsD.

The arrangement of the ligands bound to the surface of the DmsD molecule, in the various crystal structures that have been solved to date, give further insight into the protein interaction surfaces that may be present. In each of the solved structures, there are ligands or un-modeled electron density in the putative binding pocket (Figure 2.11), as well as a high density of ligands bound in the area along helix 5 and from the putative
binding pocket along the loop between helices 7 & 8 towards the N-terminus of the protein. These regions are also highly conserved among DmsD orthologs (Figure 2.10), supporting their importance in the function of DmsD.

The accumulated structural and functional data pertaining to DmsD and other REMP family proteins points to gaps in the analysis that still need to be addressed. For example: the function and mechanism of dimerization of the REMP molecule, and the effect of the GTPase activity on the role of the REMP in the cell. Additionally, the molecular details of the REMP-leader peptide interaction, and those of the other interaction surfaces, are unclear. There is insight into each of the REMP functions, but the mechanism and consequences of these functions remains unresolved. The REMP is involved in each step of substrate maturation, from the release of the nascent substrate protein from the ribosome, to delivery of the folded protein at the TAT channel. This is an interesting example of a complex interplay of proteins. In the following chapters more attention will be placed on the interaction between the DmsA leader peptide and DmsD.
3. Purification of the DmsA leader peptide by co-expression with the DmsD chaperone

3.1. Publication & contributions

Portions of this chapter have been published in the journal Protein Expression and Purification with the following citation:


All of the experimental work discussed in this chapter was my own, and performed under the guidance of Dr. Paetzel, with whom I collaborated in the preparation of the above mentioned manuscript.

3.2. Introduction

This chapter demonstrates a method for the purification of the 45 residue long leader peptide of *E. coli* dimethyl sulfoxide reductase subunit A (DmsA<sub>L</sub>), a substrate of the twin arginine translocase, by co-expressing the leader peptide with its specific chaperone protein, DmsD. The peptide can be isolated from the soluble DmsA<sub>L</sub>/DmsD complex or conveniently from the lysate pellet fraction. The recombinant leader peptide is functionally intact as the peptide-chaperone complex can be reconstituted from purified DmsA<sub>L</sub> and DmsD. A construct with DmsA<sub>L</sub> fused to the N-terminus of DmsD (DmsA<sub>L</sub>-DmsD fusion) was created to further explore the properties of the leader peptide-chaperone interactions. Analytical size-exclusion chromatography in-line with multi-angle light scattering reveals that the DmsA<sub>L</sub>-DmsD fusion construct forms a dimer wherein each protomer binds the neighbouring leader peptide. A model of this homodimeric interaction is presented.
Figure 3.1. Schematic of the DmsA leader peptide.
The twin arginine sequence is highlighted in red. In this construct, Ala45 was mutated to proline.

Many substrates of the TAT system have specific chaperone proteins that bind to the immature substrate protein, primarily via the leader peptide (Figure 3.1) (Winstone et al., 2006; Genest et al., 2008), and prevent translocation until it has been properly folded, cofactors have been inserted and multi-protein complexes have been assembled (Hatzixanthis et al., 2005; Genest et al., 2009).

These chaperone proteins are able to protect the leader peptide moiety of the pre-protein from proteolytic degradation. This protection has been demonstrated using the model chaperone protein TorD and the pre-TorA substrate protein (Pommier et al., 1998; Genest et al., 2006). Co-expression of the TorD protein has also been used to enhance secretion of a chimeric TorA leader peptide fused to the GFP protein (Li et al., 2006b).

E. coli dimethylsulfoxide reductase subunit A (DmsA) is one of the most studied TAT substrates. In chapter 4, NMR is used to probe the interactions between the 45 residue leader peptide from the DmsA pre-protein (DmsAL) and its specific chaperone DmsD. To do this, it is essential to have a method that can produce milligram quantities of the peptide. It is difficult and costly to chemically synthesize DmsAL due to its length and hydrophobicity. Recombinant production of the peptide is less expensive and provides the possibility of producing isotopically labeled versions of the peptide for NMR studies. Attempts to express DmsAL alone using recombinant technology have failed. This study demonstrates that the REMP DmsD (Oresnik et al., 2001), when co-expressed with DmsAL, aids in the production of recombinant DmsAL.
3.3. Materials and methods

3.3.1. Cloning DmsA_L

The DmsA leader peptide (DmsA_L), residues 1-45 of the dmsA gene (Uni-Prot ID: P18775), was PCR amplified from a plasmid that codes for the DmsA leader peptide sequence, pTDms35, a gift from Dr. R. J. Turner. The sense primer, 5'-AGCTCATATGATGAAAACGAAAATCCCTGATGCGGTATTGG-3', contained an NdeI restriction enzyme recognition site. The antisense primer, 5'-AGCTCTCGAGGCTGCCGCGGCACCAGGGGCGCAATCGACTAAAAGGTAATG-3', contained an XhoI restriction enzyme recognition site and a codon change that mutated the last residue of the leader peptide from an alanine to a proline (Ala45Pro). This mutation was introduced to limit the level of cleavage of the leader peptide by type I leader peptidase. The antisense primer also encoded a C-terminal thrombin protease recognition site before the C-terminal hexa-histidine tag that is present in the pET-24a vector. The PCR product was purified with a GeneJet PCR purification kit (Fermentas) and inserted into the kanamycin resistant expression plasmid pET-24a (Novagen). The sequence of this C-terminal extension was: LVPRGSLEH6 (Figure 3.1). This plasmid pDmsA_L-H6-24a was verified by DNA sequencing and transformed into competent BL21(DE3) host cells for expression.

3.3.2. Cloning DmsD

The E. coli dmsD gene was PCR amplified from the plasmid pTDms67, a gift from Dr. R. J. Turner, using the primers 5'-ACTGCATATGACCCATTTTTCACAGCAAGATAATTTTTCTG-3' and 5'-ACTGCATATGACCCATTTTTCACAGCAAGATAATTTTTCTG-3'. The resulting PCR product, containing 5' NdeI and 3' XhoI restriction sites, was inserted into the ampicillin resistant pET-15b expression vector (Novagen). The sequence for this plasmid, pH6-DmsD-15b, was verified by DNA sequencing. The construct encoded by pH6-DmsD-15b consists of an N-terminal extension (MGSH6SSGLVPRGSHM) containing a hexa-histidine tag and a thrombin protease recognition site, followed by the 204 residues of the dmsD gene product (UniProt ID: P69853). This construct was then modified by site directed mutagenesis using the QuickChange protocol (Stratagene) using the primer 5'
GAAGGAGATATACCAAGGCAGCAGCCATC 3’ and its complement. Successful mutagenesis was verified by DNA sequencing. The modification converted the translation initiation residue -19 M (ATG) residue to a K (AAG), causing the N-terminal extension not to be expressed. The resulting plasmid was named pDmsD-15b.

### 3.3.3. Cloning the DmsA<sub>L</sub>-DmsD fusion

The DmsA<sub>L</sub>-DmsD fusion construct was generated by cloning the DmsA<sub>L</sub> sequence (residue 45 mutated to proline as described above) using the primers: 5’- AGCTCATATGATGAAAACGAAAATCCCTGATGCGGTATTGG-3’, containing an NdeI restriction enzyme recognition site and 5’- AGCTGAATTCGGGCGCAATCCGACTAAAAGGTAATG-3’, containing an EcoRI restriction enzyme recognition site. The dmsD gene was amplified from the plasmid pTDms67, using the primers: 5’- ACTGGAATTCATGACCCATTTTTCACAGCAAGATAATTTTTCTG-3’ containing an EcoRI restriction enzyme recognition site and 5’- AGCTCTCGAGCTATCGAAACAGCGTTTAACCGCG-3’ containing an XhoI restriction enzyme recognition site. The amplified and digested products were ligated into the kanamycin resistant expression vector pET-28a (Novagen) to create a construct that produces a protein consisting of an N-terminal hexa-histidine tag and a thrombin protease recognition site (MGS<sub>6</sub>SSGLVPRGSHM), residues 1-44 of the DmsA preprotein, a proline residue, and finally, the DmsD protein. This plasmid was designated pH<sub>6</sub>-DmsA<sub>L</sub>-DmsD-28a.

### 3.3.4. Expression

To prepare for expression, pH<sub>6</sub>-DmsA<sub>L</sub>-DmsD-28a (DmsA<sub>L</sub>-DmsD fusion) was transformed into the host strain BL21(DE3). To prepare cells capable of co-expressing DmsA<sub>L</sub> and DmsD, the plasmids pDmsA<sub>L</sub>-H<sub>6</sub>-24a (leader peptide, kanamycin resistant) and either pH6DmsD-15b or pDmsD-15b (DmsD, ampicillin resistant) were co-transformed into the host strain BL21(DE3) and grown in LB media containing final concentrations of 0.05 mg/mL kanamycin and 0.1 mg/mL ampicillin.

To express each construct separately or to co-express the DmsA<sub>L</sub>/DmsD complex, 10 mL of an overnight culture was diluted into 1 L of LB media containing the
appropriate antibiotic, and incubated at 37°C with shaking at 250 rpm until OD$_{600}$ = 0.6. The culture was induced with IPTG to a final concentration of 0.1 mM, incubated for a further 3 h, and harvested by centrifugation at 5000 x g. The cell pellet was re-suspended in 30 mL of 20 mM Tris pH 8.0, 100 mM NaCl (TBS) and stored at -80°C. The cells were lysed by sonication for one minute at 30 % amplitude using 5 second pulses at 5 second intervals. This was followed by treating the lysate with three cycles through an Avestin Emulsiflex-3C cell homogenizer. The lysate was clarified by centrifugation at 45,000 x g for 60 minutes at 4°C, and the pellet and supernatant fractions were recovered.

3.3.5. **Purification of DmsD, the DmsA$_L$-DmsD fusion or the DmsA$_L$/DmsD complex**

To isolate DmsD, the DmsA$_L$-DmsD fusion protein or the soluble DmsA$_L$/DmsD complex, the lysate supernatant was applied to a 5 mL Ni-NTA (Qiagen) column equilibrated with TBS, washed with 10 column volumes (CV) of TBS, and eluted with 2 CV of 500 mM imidazole in TBS. The eluted protein was concentrated with a centrifugal filter concentration apparatus with a molecular mass cut-off of 10 kDa (cellulose membrane, Millipore) to 5 mL and further purified by size-exclusion chromatography (Sephacryl S-100 HiPrep 26/60, TBS equilibrated) using an AKTA Prime system (GE Healthcare) at a flow-rate of 1.0 mL/min. Fractions containing the purified protein of interest were concentrated as before to 15 mg/mL. Hexa-histidine tags were removed by incubation with a 1:1000 ratio of thrombin protease (Sigma) overnight at room temperature, followed by application of the digested protein sample to a 1 mL bed volume Ni-NTA (Qiagen) column, previously equilibrated with TBS, and washed with 4 mL of TBS. The flow-through and wash fractions were collected and further purified by size-exclusion chromatography (Sephacryl S 100 HiPrep 26/60, equilibrated with TBS, flow-rate 1.0 mL / min). Fractions containing the protein of interest were pooled and concentrated as above. The final protein concentration was measured using a NanoDrop spectrophotometer. The extension coefficient was calculated based on the protein sequence using the program ProtParam (Wilkins et al., 1999). The protein purity was confirmed by the presence of a single band on a 15 % SDS-PAGE gel stained with PageBlue (Fermentas).
3.3.6. **Purification of DmsA_L**

To isolate the DmsA_L peptide, the pellet fraction from the lysate clarification stage in the purification of the DmsA_L/DmsD complex was solubilized in 8 M urea in TBS and centrifuged at 45,000 x g for 60 min at 4°C. The supernatant was applied to a Ni-NTA (Qiagen) column (5 mL bed volume) equilibrated with TBS containing 8 M urea, and washed with 5 CV of TBS containing 8 M urea. The urea was removed by washing with 5 CV of TBS containing 0.1 % (v/v) Triton X-100 to stabilize the hydrophobic leader peptide. The peptide was eluted with 2 CV of TBS containing 0.1 % (v/v) Triton X-100 and 500 mM imidazole and the resulting purified peptide was visualized using SDS-PAGE stained with PageBlue (Fermentas).

To purify the DmsA_L peptide from the culture that co-expressed the DmsA_L-H_6 with DmsD without a His_6 Tag, the same protocols were followed for lysis and clarification, followed by application of the lysate supernatant to a Ni-NTA (Qiagen) column (5 mL bed volume) equilibrated with TBS. The column was then washed with 5 CV of TBS to remove nonspecific binding of unwanted protein, and the re-solubilized lysate pellet, prepared as above, was added to the column. The column was then washed with 20 CV of TBS containing 8 M urea, and the urea removed with by washing with 5 CV of TBS containing 0.1 % (v/v) Triton X-100 to stabilize the hydrophobic leader peptide. The peptide was eluted with 2 CV of TBS containing 0.1 % (v/v) Triton X-100 and 500 mM imidazole and the resulting purified peptide was visualized using SDS-PAGE stained with PageBlue (Fermentas).

3.3.7. **Amino-terminal sequencing of the DmsA_L peptide**

After purification, SDS-PAGE, and blotting of the peptide sample onto a PVDF membrane, samples of the DmsA_L peptide were analyzed by N-terminal sequencing at the Ohio State University Protein Facility.

3.3.8. **Formation of DmsA_L/DmsD complex from purified peptide and protein**

The DmsA_L/DmsD complex was reconstituted by first adding the DmsA_L peptide to the Ni-NTA (Qiagen) column with a 1 mL bed volume as described above, followed by
washing the column with 2 CV of TBS, and applying an excess of purified DmsD without a hexa-histidine tag. This was incubated at room temperature for 30 min before being washed with 5 CV of TBS to remove the unbound DmsD protein. The complex was then eluted with 2 CV of TBS containing 500 mM imidazole, and further purified by size exclusion chromatography (Sephacryl S 100 HiPrep 26/60) using an AKTA Prime system (GE Healthcare) equilibrated with TBS and run at 1.0 mL / min. Fractions containing the proteins of interest were pooled and concentrated, and the purity was verified by visualization using SDS-PAGE stained with PageBlue (Fermentas).

3.3.9. **Analytical size-exclusion chromatography in-line with multi-angle light scattering analysis (SEC-MALS)**

SEC-MALS was carried out using a Superdex 200 size-exclusion column (GE Healthcare) in-line with a multi-angle light scattering system (Wyatt Technologies Inc.). Purified DmsD, the DmsA<sub>L</sub>-DmsD fusion protein, and a mixture of equal volumes of DmsD and the DmsA<sub>L</sub>-DmsD fusion protein, each at a concentration of 5 mg/mL incubated at room temperature for 30 min, were assayed to probe the association of the dimeric DmsA<sub>L</sub>-DmsD fusion protein. A 100 µl sample of purified protein (5 mg/mL) was injected and resolved at a flow rate of 0.5 mL/min in a TBS buffer at pH 8.0. The molecular masses for each protein population in each sample were determined by a multi-angle light scattering DAWN EOS instrument with a 684 nm laser (Wyatt Technologies Inc.) coupled to a refractive index instrument (Optilab rEX, Wyatt Technologies Inc.). The molar mass of the protein was calculated from the observed light scattering intensity and differential refractive index using ASTRA v5.1 software (Wyatt Technologies Inc.), based on the Zimm fit method with a refractive index increment of dn/dc = 0.185 mL/g (Zimm, 1948).

### 3.4. Results & discussion

#### 3.4.1. **Expression and purification**

DmsD, the TAT chaperone (REMP) for DmsA, was successfully over-expressed from the plasmid pH₆-DmsD-15b with a yield of approximately 2.5 mg of pure protein per litre of culture. The purified DmsD was visualized as a single band by SDS-PAGE
stained with PageBlue. Attempts to express DmsA leader peptide (DmsA_L) from the plasmid pDmsAL-H6-24b failed to produce detectable quantities of product when expression was carried out at 37°C with induction at OD_{600} = 0.6 for 3 hours, or at 25°C, with induction at OD_{600} = 1.0 overnight. However, co-expression of the DmsD and DmsA_L, followed by purification of the lysate supernatant fraction, produced the DmsA_L/DmsD complex visible as two distinct bands on an SDS-PAGE gel (Figure 3.2) with a yield of approximately 1 mg of pure protein per litre of culture. Isolation of the lysate pellet fraction produced pure DmsA_L based on SDS-PAGE gel results, though a less intense band corresponding to the DmsD molecular mass was visible (Figure 3.2). The absence of tryptophan residues in the peptide makes direct spectrophotometric measurement of the concentration difficult, yet based on the concentration of DmsD recovered after assembly of the DmsA_L/DmsD complex (discussed below), and assuming a 1:1 molar ratio of DmsA_L (6.4 kDa) to DmsD (23.3 kDa) in the complex, approximately 0.02 µmol of peptide (129 µg) was recovered from the pellet fraction of each litre of culture, and a further 0.04 µmol (230 µg of peptide was recovered from the supernatant fraction from each litre of culture (Figure 3.3).

Amino-terminal sequencing of the first six residues of the purified DmsA_L recovered from the co-expressed protein complex yielded the sequence: MMKTKI, which corresponds exactly to the DmsA_L amino-terminal sequence (with an additional N-terminal methionine residue that was introduced by the pET-24a expression vector).
Figure 3.2. Co-expressed Dms\textsubscript{A\_L} and DmsD

Two populations of the DmsD and DmsA\textsubscript{L} peptide are present. 15 % SDS-PAGE stained with PageBlue shows the purification of the soluble DmsA\textsubscript{L}/DmsD complex from the lysate supernatant fraction and insoluble DmsA\textsubscript{L} from the pellet fraction.
Figure 3.3. Denatured purification of the complex
15 % SDS-PAGE gel stained with PageBlue of the results of the purification of the peptide after co-expression with untagged DmsD and chaotropic denaturation.

3.4.2. Reassembly of the DmsA\textsubscript{L}/DmsD complex

The purified DmsA\textsubscript{L} appears to be functional, in that it is able to bind to its chaperone DmsD. Adding DmsD to a Ni-NTA column containing bound DmsA\textsubscript{L} yielded a population of the DmsA\textsubscript{L}/DmsD complex that was distinct from the unbound DmsD when analyzed by size exclusion chromatography and could be visualized by SDS-PAGE (Figure 3.4).
Figure 3.4. Reconstitution of the DmsA<sub>L</sub>/DmsD complex from purified components.

Analytical size-exclusion chromatogram of DmsD before (Blue) and after (Red) a 30 min incubation with purified DmsD with the hexa-histidine tagged DmsA leader peptide attached to a Ni-NTA column. Each peak was visualized by 15 % SDS-PAGE gel stained with PageBlue.
3.4.3. **The DmsA\textsubscript{L}-DmsD fusion protein**

*Figure 3.5. Schematic diagram of the DmsA\textsubscript{L}-DmsD fusion protein construct.*

The DmsA leader peptide fused to N-terminus of DmsD (DmsA\textsubscript{L}-DmsD fusion protein) (Figure 3.5) was purified such that it appeared as a single band when visualized by SDS-PAGE (Figure 3.6) with an approximate yield of 0.75 mg/L of culture.

*Figure 3.6. Purification of the DmsA\textsubscript{L}-DmsD fusion protein.*

A single band is observed when pure DmsA\textsubscript{L}-DmsD fusion protein is analyzed by SDS-PAGE (15 %) stained with PageBlue.

Analysis of this fusion protein by SEC-MALS reveals that it purifies as a dimer. Incubation of the DmsA\textsubscript{L}-DmsD fusion protein with DmsD followed by SEC-MALS analysis reveals no higher order species, suggesting that there are no exposed leader peptide moieties for DmsD to bind (Figure 3.7). It is likely that each N-terminal leader peptide is bound to the neighbouring DmsD in the dimer, and that the dimer is likely held together via this trans-interaction. If there was a single DmsD molecule attached to the
DmsA_L-DmsD dimer molecule, the calculated molecular mass would be approximately 80.3 kDa, and it would be visible within the 10 kDa to 600 kDa separation range of the Superdex 200 column.

Figure 3.7. Leader peptide moieties on the DmsA_L-DmsD fusion protein dimer are not accessible to free DmsD.

The DmsA_L-DmsD fusion protein was analyzed by analytical size-exclusion chromatography in-line with multi-angle light scattering (SEC-MALS). The inaccessible peptide moiety is demonstrated by the absence of a peak corresponding to 80.3 kDa or higher, which would be present if a DmsD-DmsA_L fusion protein dimer contained a free DmsA leader capable of binding a DmsD monomer.

The molecular masses determined from SEC-MALS and SEC alone for each of the proteins is reported in Table 3.1. The observed molecular masses differ somewhat from those calculated but no major peaks are observed that correspond to higher order complexes. When DmsD alone was analyzed, a higher molecular mass peak was observed at the expected elution point of a homodimeric DmsD species that has been previously characterized (Sarfo et al., 2004).
Figure 3.8. Schematics of the possible interactions between the protomers in the fusion protein dimer.

The boxed model is consistent with the experimental results, and a model of the DmsA<sub>L</sub>-DmsD fusion protein is presented. In this model, each of the fusion protein monomers binds the leader peptide of the partner molecule. One molecule is shown as a white surface, with red indicating the residues shown to be important for DmsA<sub>L</sub> binding (Chan et al. 2008) and the neighboring molecule is shown as a black ribbon, with the DmsA<sub>L</sub> moiety shown as sticks and colored according to the region of the peptide, blue represents the N-terminal region, the RR motif is colored purple, the hydrophobic region is yellow, and the C-terminal region is red.

Table 3.1. Molecular mass of protein constructs: calculated and observed by SEC-MALS

<table>
<thead>
<tr>
<th>Protein</th>
<th>Calculated M.W. (kDa)</th>
<th>MALS M.W. (kDa)</th>
<th>SEC M.W. (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DmsD monomer</td>
<td>23.3</td>
<td>22.3</td>
<td>21.5 (Superdex 200)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>22.8 (Sephacryl S-100)</td>
</tr>
<tr>
<td>DmsAL / DmsD complex</td>
<td>28.5</td>
<td>N/A</td>
<td>32.7 (Sephacryl S-100)</td>
</tr>
<tr>
<td>DmsD dimer</td>
<td>46.6</td>
<td>52.8</td>
<td>39.1 (Superdex 200)</td>
</tr>
<tr>
<td>DmsAL-DmsD fusion protein dimer</td>
<td>57.0</td>
<td>43.6</td>
<td>52.0 (Superdex 200)</td>
</tr>
</tbody>
</table>
3.5. Concluding remarks

The methods described in this chapter demonstrate a way to produce the TAT leader peptide through co-expression and co-purification of the leader peptide with its chaperone protein. This method was used to generate the constructs used for NMR analysis described in chapter 4. In principle these techniques should be applicable to any of the REMP family proteins and other peptide binding molecular chaperones.
4. **NMR backbone amide assignment, backbone relaxation and chemical shift perturbation analysis of DmsD**

4.1. **Publication & contributions**

Portions of this chapter have been published in Biomolecular NMR Assignments with the following citation:


I was responsible for all of the experimental work to create the samples for NMR analysis. The NMR spectra were collected by Dr. Mark Okon. I assigned and analyzed the spectra under the guidance of Dr. Lawrence McIntosh. The above-mentioned manuscript was prepared by myself in collaboration with Drs. McIntosh and Paetzel.

4.2. **Introduction**

This chapter reports the main-chain $^1$H, $^{13}$C and $^{15}$N chemical shift assignments and amide $^{15}$N relaxation data for *E. coli* DmsD, a 23.3 kDa protein responsible for the correct folding and translocation of the DMSO reductase enzyme complex. In addition, the observed amide chemical shift perturbations resulting from complex formation with the reductase subunit DmsA leader peptide support a model in which the 44 residue peptide makes extensive contacts across the surface of the DmsD protein.

Although a wealth of structural information is available for many REMPs, including X-ray crystal structures of DmsD from *E. coli* and TorD from *S. massilia* (Tranier et al., 2003; Ramasamy and Clemons, 2009), the molecular details of how
these REMP's perform their functions remain largely unknown. There are many studies exploring the interaction of DmsD with the leader peptide from the DmsA subunit of dimethylsulfoxide reductase, including the identification and biochemical characterization of several residues that may play a role in binding and a molecular dynamics simulation with a *de novo* generated peptide discussed in chapter 5 (Winstone et al., 2006; Chan et al., 2008); However, a complete picture of the interactions between DmsD and its substrates has not yet been revealed.

In this chapter, the chemical shift assignments of the main-chain $^1$H, $^{15}$N and $^{13}$C nuclei are reported, along with backbone amide relaxation data for *E. coli* DmsD. This chapter also describes a comparison of spectra collected for the chaperone in the presence and absence of the DmsA leader peptide (DmsAL) that was purified using the methods discussed in chapter 3. The resulting data support the proposed peptide-binding interface of DmsD as a shallow groove extending around the surface of the protein and open the door for future NMR spectroscopic experiments to probe the function of DmsD and elucidate its role in the secretion of the DmsABC complex.

### 4.3. Methods and experiments

#### 4.3.1. Cloning, expression and purification

The DmsD gene from *E. coli* was cloned into the pET-15b expression vector and transformed into BL21(DE3) cells. A 100 mL culture of these cells was grown in Luria broth with ampicillin overnight at 30°C. The bacteria were harvested by centrifugation at 5000 x g, re-suspended in 1 L of unlabeled M9 minimal media, grown at 37°C for 7 hours to OD$_{600}$ = 0.9, harvested again by centrifugation, and finally re-suspended in 300 mL of 99 % D$_2$O M9 minimal media with 1 g $^{15}$NH$_4$Cl and 1 g $^2$H$_7$/13C$_6$-glucose. After 1 hour of incubation at 25°C, isopropyl-β-thiogalactoside (IPTG, 0.1 mM final) was added to the culture, followed by growth overnight at 25°C. The His$_6$-tagged DmsD protein was purified using Ni$^{2+}$-NTA affinity chromatography (Qiagen), as outlined previously in chapter 2. However, to ensure amide protonation, the bound protein was unfolded by treatment with 5 CV of TBS supplemented with 8 M urea and refolded by incubation with 10 CV of TBS for 30 min. prior to elution from the affinity column. Further purification to
homogeneity was achieved using size exclusion chromatography on a Sephacryl S-100 column, equilibrated with TBS and run with a flow rate of 1.0 ml per min using an ÄKTA prime system. The purified $^2$H/$^{15}$N/$^{13}$C-labeled protein, containing the full 204 residues of wild type DmsD (UniProt P69853) and a 19 residue N-terminal extension, comprised of a His$_6$ tag and thrombin protease recognition site (MGSH$_6$SSGLVPRGSH), was concentrated to 0.8 mM in 50 mM sodium phosphate buffer at pH 6.5 with 5% D$_2$O using a 10 kDa MWCO Amicon centrifugal filter (Millipore). A non-deuterated $^{15}$N/$^{13}$C-labeled sample was prepared as above using H$_2$O M9 media and $^{13}$C$_6$-glucose.

A complex of labeled DmsD with the unlabeled DmsA$_L$ peptide (residues 1-44 of DmsA followed by a proline, a thrombin protease cleavage site, and a His$_6$ tag) was obtained as described in chapter 3. Note that DmsA$_L$ must be co-expressed and co-purified with DmsD. Thus $^2$H/$^{15}$N/$^{13}$C-labeled DmsD was exchanged for unlabeled DmsD while the DmsA$_L$ was immobilized on a Ni$^{2+}$-NTA affinity column. After elution with imidazole, the desired sample was concentrated to 5 mL and applied to a Sephacryl S-100 column, equilibrated with 50 mM sodium phosphate buffer at pH 6.5, and run with a flow rate of 1.0 ml/min. Resolved fractions containing the complex of unlabeled DmsA$_L$ and $^2$H/$^{13}$C/$^{15}$N-labeled DmsD were concentrated to 400 µM in 50 mM sodium phosphate buffer at pH 6.5 using a 10 kDa MWCO Amicon centrifugal filter.

4.3.2. **NMR data collection**

NMR spectra were recorded at 25°C using a Bruker 850 MHz Avance III spectrometer equipped with a $^1$H/$^{13}$C/$^{15}$N TCI cryoprobe. The spectra were processed with NMRPipe (Delaglio et al., 1995) and analyzed with Sparky (Goddard and Kneller, 2008). External 2,2-dimethyl-2-siliapentane-sulfonic acid (DSS) was used as a direct reference for $^1$H and indirect for $^{13}$C and $^{15}$N. The signals from the main-chain nuclei of $^2$H/$^{13}$C/$^{15}$N-labeled DmsD at pH 6.5 were assigned using $^2$H-decoupled TROSY-based $^{15}$N-HSQC, HNCO, HN(CA)CO, HNCACB, HN(CO)CAB, and non-TROSY C(CO)-TOCSY-NH spectra (Sattler et al., 1999), combined with 150 ms mixing time 3D $^1$H- and $^{15}$N-resolved NOESY-HSQC spectra (Ikura et al., 1990). Amide $^{15}$N heteronuclear NOE relaxation data for a non-deuterated sample of $^{15}$N/$^{13}$C-labeled DmsD were recorded with a 600 MHz Varian Inova spectrometer (Farrow et al., 1994).
Amide $^{1}\text{H}N$ and $^{15}\text{N}$ assignments for labeled DmsD in complex with unlabeled DmsA$_l$ at pH 6.5 were assigned from $^2\text{H}$-decoupled TROSY-based HNCA CB and HNCA(CO)CB spectra, using the spectra for the free protein as a guide. The chemical shift perturbation for each residue with assigned signals in the spectra of both free and complexed DmsD was calculated with the following equation:

$$\Delta\delta = \left[ (\Delta\delta^{1}\text{H})^2 + \left( \frac{\Delta\delta^{15}\text{N}}{5} \right)^2 \right]^{1/2}$$
4.4. Results & discussion

4.4.1. Assignments

Figure 4.1. Assigned $^{15}$N-HSQC TROSY spectrum of DmsD (pH 6.5, 25°C).
Backbone $^{1}H^{N},^{15}N$ peaks are labeled by residue and the boxed region is expanded in the additional panel. 'B' denotes the weaker peak for amides yielding two signals.

Overall, DmsD yielded a $^{15}$N-HSQC spectrum with excellent dispersion and line shape, indicating that this 23.3 kDa protein is stably folded and well behaved in solution (Figure 4.1). Assignment of signals from main chain $^{1}H^{N},^{15}N,^{13}C,^{13}C'$ and/or $^{15}C'$ nuclei were obtained for 166 of 190 non-proline residues using standard triple resonance
approaches and have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) with accession number 18257. Signals from residues 4-9, 16, 74-78, 87-88, 97, 128-131, and 153 could not be assigned unambiguously, most likely due to spectral overlap. A second set of weaker $^1$H$^N$-$^{15}$N peaks was also identified for 63 residues broadly clustering near helix 5. The origin of this conformational heterogeneity is unclear as this region is lacking in prolines and, although containing an exposed cysteine, addition of excess DTT did not alter the spectra of DmsD. Parenthetically, although deuteration and the use of a 850 MHz NMR spectrometer with TROSY-based pulse sequences improved the quality of the DmsD spectra, a total of 161 assignments, including those from duplicate peaks, were obtained from data recorded for an initial $^{15}$N/$^{13}$C-labeled sample of the protein using a 600 MHz spectrometer.

4.4.2. Secondary structure

The secondary structural elements of DmsD derived from main chain $^{13}$C chemical shifts using the SSP (Marsh et al., 2006) and MICS (Shen and Bax, 2012) algorithms (Figure 4.2) are in good agreement with those observed in the X-ray crystallographic model of the helical protein (PDB: 3efp and 3cw0) (Ramasamy and Clemons, 2009); However, some small differences are observed, such as the indication from chemical shifts of $\beta$-strand character in the flexible loop region between helices 7 and 8. The SSP, but not MICS, algorithm also suggests that helices 10 and 11 extend longer than observed within the crystallized protein.

4.4.3. Dynamics

Insights into the local backbone dynamics of DmsD are provided by the steady-state heteronuclear $^{15}$N NOE values shown in Figure 4.2. In general, the backbone of the protein is well ordered, with enhanced fast timescale flexibility only observed in the loop regions between helices 5 and 6 and between helices 7 and 8. The former contribute to the proposed binding site for the DmsA leader peptide (Chan et al., 2008), and the latter comprises the hinge region between two subdomains of the protein that is flexible in the X-ray crystal structures of all of the TorD-like REMP proteins (Tranier et al., 2003; Ramasamy and Clemons, 2009).
Figure 4.2. Secondary structure prediction and NOE relaxation of the DmsD protein

Top: Confidence scores for the predicted secondary structure as helix (red), loop (green) and strand (blue) generated from main chain chemical shifts by the MICS program (Shen and Bax, 2012). Middle: Secondary structure propensity (SSP) scores based on backbone $^{13}\text{C}_a$, $^{13}\text{C}_b$, $^{13}\text{C}'$ chemical shifts (Marsh et al., 2006). Values approaching +1 and -1 are diagnostic of α-helical and β-strand conformations, respectively. Lower: Amide heteronuclear $^{15}\text{N}$-NOE values for DmsD. Blank points correspond to unassigned amides and prolines. Reduced NOE values, indicative of increased sub-nanosecond timescale motions, are observed at the termini and the loops between helices 5 and 6 and between helices 7 and 8. The cylinders denote α-helices observed in the X-ray crystallographic structure of DmsD (PDB: 3efp).
4.4.4. Chemical shift perturbation

To investigate the mechanism of leader peptide binding, a complex of unlabeled DmsA_L and $^2$H/$^{13}$C/$^{15}$N-labeled DmsD was prepared. As shown in Figure 4.3, the $^{15}$N-HSQC TROSY spectra of free and bound DmsD differ significantly. Unfortunately, titration experiments to aid spectral assignments were not possible as DmsA_L binds with 0.2 µM affinity (Winstone et al., 2006), and is not soluble in isolation. Although the spectra of the complex were of poorer quality than those of the free protein, we were able to confidently assign signals from 160 amides and thereby obtain chemical shift perturbations due to peptide binding. When mapped onto the structure of free DmsD, amides showing the greatest chemical shift perturbations cluster to one broad surface region (Figure 4.3 B C). This region overlaps, in part, with residues (18, 72, 75, 76, 86, 87, 123, 124, 126, 127, 147, 151, and 172) shown previously by mutagenesis to be important for leader peptide binding. Some discrepancies between the two approaches arise, possibly due to indirect effects of mutations and incomplete spectral assignments for the complex. For example, residues 72, 126, 127, 150, and 172 did not undergo large chemical shift perturbations, even though they were shown to be important by mutagenesis. Additionally, perturbations for Pro86, Pro124, and the unassigned Glu123 could not be assessed by these NMR methods despite being identified in the previous mutagenesis study. Nevertheless, these data are consistent with the hypothesis that DmsA_L binds DmsD along a hydrophobic surface groove that extends from the previously identified “hot pocket” between helices 5, 6 & 7 (Chan et al., 2008) outward across the “front” surface of the protein toward helices 3 & 4 (Figure 4.3 C left), as well as towards Arg161 on the “back” face of the protein (Figure 4.3 C right).
Figure 4.3. Chemical shift perturbation analysis of DmsD in the presence and absence of the DmsA<sub>L</sub> peptide.

A. Superimposed $^{15}$N-HSQC TROSY spectra of DmsD in the absence (red) and presence of unlabeled DmsA<sub>L</sub> (blue). B & C. Chemical shift perturbations $\Delta\delta$ for amides with assigned signals in both forms of DmsD are mapped onto a ribbon and surface diagram of free DmsD (PDB: 3efp) with the indicated color code. In panel C, blank points correspond to prolines and unassigned amides in the free protein, and an average shift was used for residues with two signals. In panels B and C, blue arrows and blue surface regions indicate residues that were assigned in the free DmsD and have a perturbed, albeit unassigned, chemical shift in the complex. Residues shown previously to be important for binding by mutagenesis are highlighted by an asterisk (Chan et al., 2008).

4.4.5. Structure of the bound leader peptide

It is also noteworthy that a $^{13}$C/$^{15}$N-filtered NOESY spectrum yielded only signals from the unlabeled peptide with poorly dispersed $^1$H$^N$ chemical shifts (not shown). This suggests that the DmsA<sub>L</sub> peptide does not adopt any predominant helical or stand conformation when bound to DmsD. This is similar to what was observed in our molecular dynamics simulation discussed in chapter 5, but is in contrast to predictions that the leader peptide adopts a helical conformation upon binding to the REMP (Buchanan et al., 2008).
5. Molecular dynamics simulation

5.1. Publication & contributions

Portions of this chapter have been published in the Journal of Molecular Biology and Protein Expression and Purification:


All of the molecular dynamics simulations were performed by myself using starting coordinates from the structures discussed in chapter 2. The above mentioned manuscripts were written in collaboration with Dr. Paetzel.

5.2. Introduction

The dynamics of the DmsD molecule were explored by monitoring the internal mobility of the protein over the course of a simulation of the X-ray structure of DmsD in solution, three regions of high mobility were found that correlate well with the regions of high crystallographic B factor and NMR relaxation measurements. The interaction of DmsD with the DmsA leader peptide was also probed through simulation. This used a generated DmsA leader peptide that was docked with the DmsD protein. The docked conformation placed the conserved RR motif in direct contact with the previously identified hot pocket residues. In a subsequent simulation the docked interaction was modified to take into account the NMR perturbation data discussed in chapter 4, and a further simulation was used to refine the model for peptide binding. In addition, the simulated binding interaction was used to model the DmsA leader peptide-DmsD fusion
protein dimer that was discussed in chapter 3, and predict the relative orientation of the molecules.

In solution, proteins are dynamic structures that have regions of both low and high internal flexibility. This flexibility is what allows many enzymes to bind to substrates and catalyze chemical reactions, or permits signal transduction proteins to change conformation and execute their function. Traditional forms of protein structural analysis give valuable information about the dynamic nature of proteins through NMR backbone amide relaxation studies and the B-factor of X-ray crystal structures. With current techniques these motions cannot be visualized directly, nor how they might change when presented with substrates, or disrupted with mutations. Molecular dynamics allows these properties of proteins to be calculated. The molecules are parameterized based on experimental data, and the behavior of the protein can be predicted in different environments, or with mutations or modifications. Simulations can be used to generate testable hypotheses about these effects, or model interactions for which there are limited data available.

The DmsD protein was simulated in the presence and absence of the docked DmsA leader peptide. These simulations were analyzed and compared, producing a series of residues that are likely involved in the binding interaction, and demonstrating that the flexible regions of DmsD are important for this interaction. The simulation data was also compared with the data produced from mutagenesis studies, NMR chemical shift perturbation, and relaxation data to propose a new model for leader peptide binding that is consistent with everything that is currently known about the DmsA_l/DmsD interaction. Molecular dynamics simulation was also employed to model the dimer interaction of the DmsA_l/DmsD fusion protein described in chapter 3, leading to a predicted orientation for the molecules.

5.3. Materials & methods

5.3.1. Molecular dynamics simulation of DmsD

The molecular dynamics simulation was carried out and analysed using the GROMACS 3.3 suite of programs (Lindahl et al., 2008). Chain A of the DmsD structure
was the subject of the simulation, wherein the protein was first energy minimized in \textit{vacuo}, using the steepest descent algorithm such that the maximum force on any atom did not exceed 250 kJ mol\(^{-1}\) nm\(^{-1}\). A cubic box was then constructed around the protein so that 9 Å of space was present between the edge of the box and the nearest protein atom. The system was solvated using the spc216 water model and the net charge of the system was made zero by replacing randomly selected water molecules with Na\(^+\) or Cl\(^-\) ions. The system was energy minimized by the steepest descents algorithm to a maximum force of 1000 kJ mol\(^{-1}\) nm\(^{-1}\) and simulated for 1 ns using a 0.002 ps time-step with all protein atoms constrained. This was followed by a simulation in which the constraints were replaced by LINCS bond length and angle restraints (Hess et al., 1997). The simulation was run on the Westgrid computing cluster “matrix” for a total simulation time of 53.8 ns, followed by steepest descent energy minimization to a maximum force of 250.0 kJ mol\(^{-1}\) nm\(^{-1}\). This simulation used the GROMOS96 43a1 force field and was run in an environment that holds the number of atoms, temperature and pressure of the system constant (NPT). Interactions were calculated using a twin range pair list with a long range cut-off of 10 Å and a short range cut-off at 0.8 Å. Temperature and pressure coupling used the Berendsen method (Berendsen et al., 1984) at 300 K, with a \(\tau_T\) value of 0.1 and a \(\tau_P\) value of 1.0. The simulation box was periodic in all dimensions.

### 5.3.2. Docking & molecular dynamics of the DmsA\(_L\)/DmsD complex

Docking was performed on the 3D-Garden webserver (Lesk and Sternberg, 2008). Chain A of the \textit{E. coli} DmsD (PDB: 3efp) asymmetric unit was used for the docking experiment. A \textit{de novo} generated polypeptide corresponding to the sequence of the pre-DmsA leader peptide (MKTKIPDAVLAAEVSRRGLVKTTAIGGLAMASSALTLPFSRIAHA) in an extended conformation were submitted to 3D-Garden using the default set of parameters. After docking the leader peptide was truncated at residue 29. The package GROMACS version 3.3.3 (Van Der Spoel et al., 2005) was used to perform the simulations. The docked DmsA\(_L\)/DmsD complex was processed using the GROMOS96 G43a2 force field and simulations were run in an environment that keeps the number of atoms, the pressure and temperature constant (NPT). The complex was energy minimized in \textit{vacuo} using steepest descents such that the maximum of force on any atom (\(F_{\text{max}}\)) was less than 250.0 kJ mol\(^{-1}\) nm\(^{-1}\). The complex was then embedded in a cubic box with a 9 Å
space between the edge of the protein and the edge of the box and solvated using the spc216 (simple point charge) water model, and the net charge of the system was made zero by replacing solvent molecules with sodium or chloride ions. The solvated system was energy minimized using the steepest descents algorithm to an $F_{\text{max}} < 1000.0 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ and equilibrated for 1 ns with a time step of 0.002 ps with position restraints placed on all atoms of the protein and peptide. Interactions were calculated using a twin range pair list with long and short range cut offs at 10 Å and 0.8 Å respectively. Berendsen coupling was applied for temperature and pressure coupling at 300 K using a $\tau_T$ value of 0.1 and a $\tau_P$ value of 1.0. The simulation cube was periodic in all dimensions. After equilibration the position restraints on the protein atoms were replaced with LINCS bond length constraints and bond angle restraints. The simulation was run on the Westgrid computing cluster “matrix” at variable intervals for a total of 63,500 ps followed by steepest descent energy minimization to an $F_{\text{max}}$ of $250.0 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ prior to analysis. Analyses of the simulations was carried out using Visual Molecular Dynamics (VMD) (Humphrey et al., 1996), and the GROMACS suite of programs (Van Der Spoel et al., 2005).

### 5.3.3. Molecular model for the DmsA\textsubscript{L}-DmsD fusion protein

A model for the DmsA\textsubscript{L}-DmsD fusion protein dimer was constructed based on the crystal structure of the DmsD monomer (PDB ID: 3efp). The molecules were positioned such that, in the dimer, the leader peptide moiety from each molecule was bound by the partner molecule in a manner based on our original molecular dynamics simulation of the binding event. Coordinates were manipulated using Coot (Emsley and Cowtan, 2004) and the final model was energy minimized, using the steepest descents algorithm in the GROMACS 3.3 suite of molecular dynamics programs, so that the maximum force on any atom did not exceed 250 kJ mol\textsuperscript{-1} nm\textsuperscript{-1} (Lindahl et al., 2008). A periodic simulation box, 9 Å larger than the protein complex along each dimension, was formed and the system was solvated using the spc216 water model. The net charge of the system was made zero by replacing randomly selected water molecules with Na\textsuperscript{+} or Cl\textsuperscript{-} ions. The system was energy minimized, as before, to a maximum force of 1000 kJ mol\textsuperscript{-1} nm\textsuperscript{-1} and the system was simulated for 1 ns using a 0.002 ps time-step with all protein atoms constrained. In subsequent simulation steps, the constraints were replaced by LINCS bond length and angle restraints. The simulation was run on the Westgrid
computing cluster “matrix” for a total simulation time of 6 ns, followed by steepest
descent energy minimization to a maximum force of 250.0 kJ mol$^{-1}$ nm$^{-1}$. This simulation
used the GROMOS96 43a1 force field and was run in an environment that holds the
number of atoms, temperature and pressure of the system constant. Interactions were
calculated using a twin range pair list with a long range cut-off of 10 Å and a short range
cut-off at 0.8 Å. Temperature and pressure coupling used the Berendsen method at 300
K (Berendsen et al., 1984), with a $\tau_T$ value of 0.1 and a $\tau_P$ value of 1.0. The figure was
generated by converting the output of the simulation from the GROMACS file format to
the PDB format with edition (Lindahl et al., 2008) and the image was rendered using
Pymol (DeLano, 2002).

5.3.4. **Generation of a new model for DmsD peptide binding**

Each of these simulations was repeated with the more recent GROMACS 4.5.4
(Lindahl et al., 2008). GROMACS 4.5.4 contains a number of improvements over version
3.3, and the procedure for those simulations was identical to the following method used
for the comprehensive model for DmsA_L/DmsD interaction.

The output of the simulation described in 5.3.2, was re-run using GROMACS
4.5.4 with the parameters described below, and was used as a starting point to manually
reposition the DmsA_L peptide such that it was in close proximity to the residues
highlighted by chemical shift perturbation analysis in 4.4.4. The repositioning was done
using UCSF chimera (Pettersen et al., 2004). The pdb file was converted to GROMACS
format and parameterized using pdb2gmx with the G53a6 force field. The system was
placed in an octahedral box 9 Å larger than the protein in each direction and solvated
using the spc216 water model. The net charge of the system was made 0 by replacing
randomly selected water molecules with Na$^+$ and Cl$^-$ ions to compensate for the charge
of the protein. The system was energy minimized to a maximum force of 1000 kJ mol$^{-1}$
nm$^{-1}$, using the steepest descents method, and used the Particle Mesh Ewald method
(PME) to calculate electrostatic interactions. The output of this minimization was energy
minimized again using the conjugate-gradient method to a maximum force of 50 kJ mol$^{-1}$
nm$^{-1}$. This was followed by position-restrained equilibration for 1 ns with a 2 fs time-step.
The electrostatic interaction energies were calculated using the PME method (Essmann
et al., 1995), a velocity rescaling thermostat and a Parrinello-Rahman barostat.
(Parrinello and Rahman, 1981; Nosé and Klein, 1983). P-LINCS restraints were used to restrain bond lengths (Hess, 2008). The production simulation ran for 50 ns using the same parameters as the restrained equilibration with the exception of the removal of position restraints on the protein molecule. Analyses were performed as discussed in the previous sections.

5.4. Results & discussion

5.4.1. Molecular dynamics of the DmsD protein

Simulation of the DmsD protein has demonstrated that regions of the DmsD protein that are involved in crystal contacts, or heteroatom binding in the X-ray structure exhibit some movement when those interactions are removed. The backbone of the protein is largely unchanged with the most fluctuation observed in the same regions that have been identified in our NMR relaxation analyses and with the B-factor of the X-ray crystal structures. These are the N-terminus of the protein, the N-terminal region of helix 4, the loops between helices 5-6, 6-7, and 7-8, and finally at the C-terminus of the protein. This was assessed in the simulation by measuring the root mean squared fluctuation (RMSF) of each alpha carbon over the course of the simulation (Figure 5.1).
**Figure 5.1. Root mean squared fluctuation (RMSF) plotted against residue number for the DmsD protein simulation.**

RMSF is the total fluctuation of each residue from the average structure; in this case, each frame was superposed onto the final frame of the simulation. Greater fluctuation is indicative of more flexible regions in the protein and is located primarily in loop regions. The largest fluctuation occurs between helices 2 & 3, 4 & 5, and at the termini.

**Figure 5.2. DmsD after 58 ns of simulation.**

A. The DmsD X-ray crystal structure (3efp:A) with residues important for binding colored red. B. The DmsD protein adopts a tight globular form, with fewer large cavities and very few of the residues important for binding are accessible to the solvent. C. A superposition of the structures shown in panels A and B – The crystal structure is colored in black, and the structure after simulation is shown in white.
The side chains of the residues that were involved in these crystal packing and heteroatom interactions, and those that were solvent exposed, were much more dynamic than the backbone or the buried residues. The result of simulating the fully solvated protein was a much more compact globular protein surface with fewer cavities and many of the residues that were shown to be important for DmsA\textsubscript{L} binding became less accessible (Figure 5.3).

![Figure 5.3. Docking and molecular dynamics simulation experiments to predict the DmsD/DmsA leader peptide interactions.](image)

The molecular surface of DmsD is shown in white with red highlights for those residues previously shown by mutagenesis to be important for DmsA leader peptide binding. A. Surface representation of the X-ray crystal structure of DmsD (pdb: 3efp) (chain A). B. Molecular dynamics simulation of DmsD with the region of DmsA leader peptide that harbors the twin arginine motif. Residues that are part of the twin arginine consensus motif are labeled. C. The sequence of the DmsA leader peptide. The twin arginine consensus motif is underlined.
5.4.2. Docking & molecular dynamics simulation of the DmsA<sub>L</sub>/DmsD complex

The mutagenesis studies by Chan and colleagues (2008) identified several residues that, if mutated, disrupted the interaction between DmsD and the DmsA leader peptide. We have mapped these residues onto the surface of the DmsD crystal structure (Figure 5.3). We used the docking server 3D-Garden (Lesk and Sternberg, 2008) to dock a DmsA leader peptide, modeled in an extended conformation, onto the crystal structure of DmsD. Previous work with TAT leader peptides has shown them to be unstructured in aqueous solution (San Miguel et al., 2003). The docking procedure repeatedly resulted in the leader peptide oriented across the DmsD surface, with the twin arginine motif placed between the three conserved surface loops. This is roughly the position taken by the glycerol and tris molecules that were modeled into the electron density in the putative leader peptide binding pocket of the DmsD crystal structure. The position of the docked peptide is consistent with several studies that have highlighted the importance of the twin arginine motif of the leader peptide in REMP-substrate interaction and those that demonstrate the role of the hydrophobic region of the leader peptide (Gross et al., 1999; Stanley et al., 2002; Li et al., 2006a; Shanmugham et al., 2012).

The docked peptide was used as a starting point for a molecular dynamics simulation (Figure 5.3 A). After 63.5 ns of simulation the DmsD/DmsA leader peptide complex became considerably more integrated (Figure 5.3 B), with a series of hydrogen bonds and van der Waals interactions formed between the DmsA leader peptide and DmsD (Figure 5.3 C). The dynamics simulation experiment resulted in small adjustments in the DmsD structure at helices 2, 3, 4, 7, the N-terminal coil, and various side chain rotamers on the DmsD molecular surface. Interestingly, the regions that moved the most in the simulation corresponded to the sites of intermolecular contact within the crystalline lattice. Some of the residues that were previously shown by mutagenesis to be important for leader peptide binding are not present on the solvent accessible molecular surface of the crystal structure (Figure 5.3 A). After the simulation, more of the proposed functionally important residues are exposed, interacting with the bound peptide (Figure 5.3 B). Table 5.1 lists the putative molecular interactions between DmsD and the leader peptide of DmsA.
When this simulation was repeated with the full length DmsA leader peptide, the DmsA leader peptide was extended further around the surface of the DmsD protein (Table 5.2) (Figure 5.4). The result of this simulation suggests that the hydrophobic region of the leader peptide has an important role in this interaction, a suggestion that is supported by the findings of Shanmugham et al., (2012).

**Table 5.1. Proposed interactions between DmsD and the truncated DmsA**

<table>
<thead>
<tr>
<th>DmsA leader peptide</th>
<th>DmsD</th>
<th>Type of Interaction*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met 1</td>
<td>Pro201 Arg 204</td>
<td>vdW</td>
</tr>
<tr>
<td>Lys 4</td>
<td>Leu 202</td>
<td>vdW</td>
</tr>
<tr>
<td>Pro 6</td>
<td>Phe203 Arg204</td>
<td>HB</td>
</tr>
<tr>
<td>Asp7</td>
<td>Asp93</td>
<td>HB</td>
</tr>
<tr>
<td>Val9</td>
<td>Pro83 Trp91</td>
<td>HB</td>
</tr>
<tr>
<td>Glu13</td>
<td>Arg94</td>
<td>vdW</td>
</tr>
<tr>
<td>Arg 16</td>
<td>Gln121 Asn122 Glu123 Glu125 Glu123 Glu125</td>
<td>HB / vdW</td>
</tr>
<tr>
<td>Arg17</td>
<td>Trp80 Asn122</td>
<td>HB</td>
</tr>
<tr>
<td>Gly18</td>
<td>Phe76</td>
<td>HB / vdW</td>
</tr>
<tr>
<td>Leu19</td>
<td>Phe76 His127 Tyr22</td>
<td>vdW</td>
</tr>
<tr>
<td>Val20</td>
<td>Trp72</td>
<td>vdW</td>
</tr>
<tr>
<td>Thr22</td>
<td>His68 Ala69</td>
<td>HB</td>
</tr>
<tr>
<td>Thr23</td>
<td>Glu65</td>
<td>HB</td>
</tr>
<tr>
<td>Ala24</td>
<td>Glu65</td>
<td>HB</td>
</tr>
<tr>
<td>Ile25</td>
<td>Glu65</td>
<td>vdW</td>
</tr>
<tr>
<td>Ala29</td>
<td>Gln73</td>
<td>vdW</td>
</tr>
</tbody>
</table>

* vdW: van der Waals interactions; HB: hydrogen bond
Figure 5.4. The DmsD protein and the full length DmsA leader peptide after 50 ns of simulation.

The peptide makes extensive contact around the surface of DmsD. There are notable polar interactions between the RR motif and the hot pocket and hydrophobic interactions with nonpolar regions of the DmsD molecule. The leader peptide is oriented such that the N-terminus is to the left and the C-terminus is to the right. DmsD is in standard orientation.

5.4.3. Molecular model for the DmsA<sub>L</sub>-DmsD fusion protein

The model for the DmsD-DmsA<sub>L</sub> dimer generated by docking and molecular dynamic simulation, and discussed in detail in chapter 3, provides a proposed relative orientation for the DmsD molecules within the dimer based on the mutual docking of the amino-terminal DmsA leader peptide of the neighbouring protomer within the dimer. The protein-protein interactions in the DmsA<sub>L</sub>-DmsD fusion protein dimer were modeled and
subjected to molecular dynamics simulation. The resulting model is consistent with the model wherein the leader peptide of each molecule is bound by the partner molecule, and predicts their relative orientation (Figure 5.5).

The interaction between the residues of DmsA_L and the DmsD surface is consistent with the those observed previously when the same analysis was performed for the DmsD alone and DmsA_L, and the modeled interactions are consistent with the DmsD contact sites proposed from previous site-directed mutagenesis studies (Figure 5.3) (Chan et al., 2008). The twin arginine motif is located within a crevice formed by three conserved surface loops that are observed to bind ligands in the X-ray crystal structure of *E. coli* DmsD (77-87, 93-100, 113-127) (PDB ID: 3efp).

*Figure 5.5. A proposed model for the DmsA_L-DmsD fusion protein dimer.*

The DmsA_L peptide is wrapped around the DmsD molecule as was observed in monomeric simulations. This model is consistent with SEC/MALS data described in chapter 3.
The N-terminal region of the DmsA_L peptide interacts with the region of DmsD near the unstructured C-terminus of the DmsD polypeptide. This includes the previously described hot pocket residues (Chan et al., 2008), the area between helix 5 and the loop between helices 1 & 2, which bind to the arginine of the SRRGLVK motif. Helix 5 has been shown by mutagenesis to be important for the homologous TorD / TorA_L interaction (Winstone et al., 2006). The hydrophobic region of the DmsA_L peptide also binds to helix 5, the loop between helices 4 and 5, and interacts with residues on helices 2 and 4. Each of the leucine residues in the hydrophobic region of the DmsA_L peptide interact with hydrophobic pockets on the surface of the DmsD molecule, consistent with observations that conserved leucines are involved in the interaction, and are similar across TAT leader peptides (Buchanan et al., 2008; Shanmugham et al., 2012). The C-terminal region of the DmsA_L peptide interacts in part with the C-terminal region of helix 2, though a large portion of the C-terminal region acts as a linker between the interacting portions and the partner protomer.
### Table 5.2. Contacts between the full length DmsA leader and DmsD

<table>
<thead>
<tr>
<th>DmsA&lt;sub&gt;L&lt;/sub&gt;</th>
<th>DmsD</th>
<th>DmsA&lt;sub&gt;L&lt;/sub&gt;</th>
<th>DmsD</th>
<th>DmsA&lt;sub&gt;L&lt;/sub&gt;</th>
<th>DmsD</th>
<th>DmsA&lt;sub&gt;L&lt;/sub&gt;</th>
<th>DmsD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met 1</td>
<td>Trp 91</td>
<td>Val 14</td>
<td>Trp 80</td>
<td>Ile 25</td>
<td>His 68</td>
<td>Leu 35</td>
<td>Gln 63</td>
</tr>
<tr>
<td></td>
<td>Arg 204</td>
<td></td>
<td>Ala 81</td>
<td>Ala 69</td>
<td>Gln 70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys 2</td>
<td>Asp 93</td>
<td>Asp 16</td>
<td>Gly 27</td>
<td>Leu 28</td>
<td>Ser 27</td>
<td>Pro 38</td>
<td>Thr 38</td>
</tr>
<tr>
<td></td>
<td>Arg 204</td>
<td>Val 77</td>
<td>Gly 26</td>
<td>Ala 28</td>
<td></td>
<td>Pro 38</td>
<td></td>
</tr>
<tr>
<td>Thr 3</td>
<td>Arg 204</td>
<td>Arg 16</td>
<td>Trp 7</td>
<td>Gly 27</td>
<td>Leu 37</td>
<td>Val 34</td>
<td></td>
</tr>
<tr>
<td>Lys 4</td>
<td>Pro 83</td>
<td>Val 165</td>
<td>Val 77</td>
<td>Ala 29</td>
<td>Ser 27</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asp 164</td>
<td>Gln 121</td>
<td>Gly 78</td>
<td>Ala 29</td>
<td>Ser 27</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Val 165</td>
<td>Asn 122</td>
<td>Pro 79</td>
<td>Pro 25</td>
<td>Glu 66</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glu 168</td>
<td>Pro 124</td>
<td>Trp 80</td>
<td>Ser 27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arg 204</td>
<td>Asp 126</td>
<td>Ala 69</td>
<td>Pro 25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Met 30</td>
<td>Gln 73</td>
<td>Ser 27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile 5</td>
<td>Ala 81</td>
<td>Gly 18</td>
<td>Gln 121</td>
<td>Ala 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu 82</td>
<td>Val 77</td>
<td>Gln 121</td>
<td>Asn 122</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pro 83</td>
<td>Gly 18</td>
<td>Gln 121</td>
<td>Ala 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro 6</td>
<td>Val 165</td>
<td>Gly 18</td>
<td>Gln 121</td>
<td>Met 30</td>
<td>Gln 63</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arg 204</td>
<td>Met 30</td>
<td>Gln 63</td>
<td>Arg 41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp 7</td>
<td>Asp 93</td>
<td>Ser 27</td>
<td>Ala 31</td>
<td>Ser 40</td>
<td>Thr 38</td>
<td>Leu 37</td>
<td></td>
</tr>
<tr>
<td>Val 9</td>
<td>Asp 93</td>
<td>Ala 69</td>
<td>Gln 31</td>
<td>Arg 41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu 10</td>
<td>Asp 93</td>
<td>Leu 19</td>
<td>Ala 69</td>
<td>Thr 38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu 92</td>
<td>Gln 70</td>
<td>Ala 31</td>
<td>Thr 38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ser 89</td>
<td>Trp 7</td>
<td>Ser 32</td>
<td>Gln 61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala 11</td>
<td>Val 90</td>
<td>Gln 73</td>
<td>Ser 32</td>
<td>Gln 61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trp 91</td>
<td>Gln 121</td>
<td>Ser 32</td>
<td>Gln 61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu 92</td>
<td>Val 20</td>
<td>Ser 33</td>
<td>Gln 61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asp 93</td>
<td>Tyr 23</td>
<td>Ser 33</td>
<td>Gln 61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala 12</td>
<td>Leu 82</td>
<td>Ala 69</td>
<td>Ser 33</td>
<td>Gln 61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pro 85</td>
<td>Gln 121</td>
<td>Ser 33</td>
<td>Gln 61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trp 87</td>
<td>Ala 24</td>
<td>Ser 33</td>
<td>Gln 61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lys 21</td>
<td>Val 34</td>
<td>Ser 33</td>
<td>Gln 61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ser 89</td>
<td>Ala 43</td>
<td>Ser 33</td>
<td>Gln 61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu 13</td>
<td>Val 90</td>
<td>Thr 23</td>
<td>Ala 43</td>
<td>Leu 48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thr 23</td>
<td>Ala 69</td>
<td>Ala 43</td>
<td>Leu 48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ala 24</td>
<td>Leu 35</td>
<td>Ala 43</td>
<td>Leu 48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val 14</td>
<td>Trp 87</td>
<td>Leu 35</td>
<td>Ala 43</td>
<td>Leu 48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pro 79</td>
<td>Ala 24</td>
<td>Ala 43</td>
<td>Leu 48</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

89
5.4.4. **A new model for DmsD/DmsA\textsubscript{L} peptide binding**

![Diagram of protein structures]

**Figure 5.6. Proposed model for DmsA\textsubscript{L}/DmsD interaction**

In the left panel, the DmsD molecule (3efp:A) coloured such that residues shown to be important by mutagenesis are red, residues that were perturbed between 0.01 and 0.02 ppm in the NMR experiments are yellow, and those perturbed beyond 0.02 ppm are green. In the center panel, the DmsA\textsubscript{L}/DmsD complex after 68 ns of simulation, the DmsA\textsubscript{L} peptide is black. The right panel shows the proposed orientation of the DmsA\textsubscript{L} peptide (black) manually docked with DmsD and simulated for 20 ns. The RR motif makes contact with the mutagenesis data hot pocket, while the hydrophobic region makes contacts with many of the residues that were identified through chemical shift perturbation analysis.

From the NMR and mutagenesis data a new model for the DmsA\textsubscript{L}/DmsD interaction is proposed. The mutagenesis data indicated several residues that are likely important for binding (Figure 5.6), which was complementary to the NMR chemical shift perturbation data that did not provide assignments for many of the hot pocket residues. Taken together these data provide a path along the surface of the DmsD molecule that was used to manually dock the DmsA\textsubscript{L} peptide as the starting point for a simulation.

The simulation indicated that the surface of the DmsD molecule, including the hot pocket, part of the hinge loop between helices 7 and 8, and the surface between helices 1, 2, and 8 make extensive contact with the N-terminal and hydrophobic regions of the DmsA\textsubscript{L} peptide. This includes a rearrangement of the solvent exposed side chains of the DmsD molecule in the proximity of DmsA\textsubscript{L}, and the flexible region...
between helices 3 and 4, but very little change in the secondary structure or backbone of DmsD (Figure 5.6).

It is tempting to speculate that the original simulation discussed in 5.4.2 would, given sufficient time, approach the peptide binding conformation that is observed in the new model presented here. The timescale of the simulations presented in this chapter are a limitation to their efficacy as a predictive tool, though having strong data with which to dock the ligand as closely as possible to the binding site greatly reduces the time required to simulate the interaction.

In this chapter, molecular dynamics was used to address the flexibility of the DmsD protein; to predict and model the interaction of the DmsD protein with the DmsA₆ peptide with both a de novo generated peptide and the peptide manually docked according to biochemical data. It was also used to model the interaction that was observed in the DmsA₆-DmsD fusion protein that was described in chapter 3. These simulations provide a theoretical framework for the design of future experiments, in DmsD as a model for the REMP family of chaperones.
6. Summary and conclusion

The crystal structure along with docking experiments, molecular dynamics simulation experiments, and the use of previous biochemical data (Chan et al., 2008) were used to generate a proposed model for how this chaperone is able to recognize and bind to the leader peptide of its substrate. The NMR and molecular dynamics simulation based description of the behaviour of the solvated DmsD protein, and a prediction of the structure of the DmsA_L/DmsD leader peptide complex is consistent with previous biochemical predictions of this interaction, and allows for the rational design of experiments that will add to the understanding of the formation and function of this complex.

6.1. The structure & dynamics of the DmsD protein

Prior to the work presented here, the biochemical data that was generated with E. coli DmsD was being interpreted based on the structure of a homologous DmsD protein from S. typhimurium. A structure that was missing electron density for a portion of the DmsD molecule. The solution of the complete E. coli DmsD structure provides a way to directly interpret the existing, and new biochemical data regarding its function.

The two crystal structures of DmsD that are presented in the preceding chapters, and a third structure solved by another team (Ramasamy and Clemons, 2009) contribute to the understanding of the dynamic properties of the protein in addition to providing structural information. Details regarding dynamics and flexibility can be gleaned from X-ray structures by studying the thermal motion of each of the structures and identifying regions with high variability when the individual chains of the structures are superposed. These regions include the residues encompassing helices 3 and 4, the loops between helices 5 & 6, and 6 & 7, and the hinge region between helices 7 & 8. This is discussed in greater detail in chapter 2.
To complement the crystallographic information, NMR analysis of the NOE relaxation revealed a very similar pattern of flexibility. The same regions that are flexible were also demonstrated to have a high B-factor, and the termini are also demonstrably less rigid than the core of the protein. More detail regarding this set of experiments is presented in chapter 4. Regions of functional importance may be predicted through mapping areas of high conservation onto the surface of the protein, this analysis is enhanced when paired with data regarding the flexibility of the structure. The conservation of flexible regions could be attributed to involvement in a function that is conserved between the aligned proteins, as the conserved sequence has no structural role.

Differences between the structures, or regions that have poor electron density are good indicators of flexibility, as are differences that can be attributed to differences in solvent exposure and crystal contacts across different structures. We can also use the interactions observed in the dimeric form of TorD (Tranier et al., 2003) (PDB: 1n1c) as a template to model the proposed domain swapped dimeric form of DmsD and predict the regions that would be required to make this conformational change.

Each of the X-ray crystal structures, NMR analysis and molecular dynamics simulation contributes to the understanding of the dynamic nature of the DmsD protein. The agreement between the data presented in this thesis reinforces the identification of the flexible and rigid regions of the DmsD molecule. By combining this data with the previously collected biochemical data, we gain a greater understanding of the structural basis of REMP function. This includes the regions of the protein that are less rigid and are involved in leader peptide binding comprised of the loops between helices 5, 6 & 7. The hinge region between helices 7 & 8 was also flexible, as expected given that it has been shown to connect two lobes of the domain swapped dimer in the homologous TorD structure (Tranier et al., 2003).

The final flexible region that was identified in this work encompasses helices 3 & 4, along with the surrounding loop regions. The sequence of these regions are not well conserved among REMPs, though this region was implicated in the guanosine binding and GTPase function in studies of the TorD protein (Hatzixanthis et al., 2005; Guymer et al., 2010), suggesting that a similar function may be observed in DmsD.
This structural and dynamics data can be used to generate new hypotheses about the molecular basis of the other REMP functions. For example DmsD has been shown to bind to the TatBC proteins (Papish et al., 2003), but there is no available information regarding the structural basis for this function. Additionally, despite a good understanding of the biophysical properties of the REMP-substrate interaction, the mechanism for substrate release is poorly understood. This may involve flexible regions of the REMP adjacent to the binding site and the regions identified in this work are good candidates for this function. The REMPs have also been shown to interact with other proteins as a part of their function, including trigger factor and cofactor biosynthetic enzymes (Jong et al., 2004; Genest et al., 2008). These interactions may require, or cause, a conformational change in the REMP protein further illustrating the importance of the flexible character of DmsD as a model REMP.

6.2. The DmsA_L/DmsD interaction

Before this thesis, the presence of the mutagenesis derived “hot pocket” was all that was known about the DmsA_L/DmsD (Chan et al., 2008), and that the leader peptide was protected from proteolytic degradation by a member of the REMP family (Genest et al., 2006). The molecular details of this interaction, and the remainder of the binding site for the leader peptide were unknown.

By studying the structures of DmsD and related proteins, we have gained insight into regions of the protein that form crystal contacts, both with adjacent protein molecules, and with molecules found in the crystallization solution. There are crystal contacts between loop 5 and the pocket between helices 2, 3, 4- as well as between symmetry related hinge loops- and finally a pocket comprised of helices 3, 4 and 11 near the N-terminus of the protein that binds to a symmetry related N-terminus. This demonstrates that these regions have a propensity for interactions, and taken together with conservation and flexibility, highlight them as potential protein binding sites that may be explored in the future. Interactions between the DmsD molecule and tris and glycerol molecules from the crystallization solution were observed in the putative leader peptide binding site. The presence of these coordinated molecules with functional groups that
are found in protein molecules such as amides, hydroxyl groups and ketones are indications that these regions may also have a role in protein-protein interaction.

The NMR chemical perturbation analysis of the DmsA<sub>L</sub>/DmsD interaction demonstrates a region of residues in DmsD whose chemical shifts have been perturbed by the presence of the DmsA<sub>L</sub> peptide. This patch of residues overlaps, in part, with those identified by mutagenesis, as well as with residues that are predicted to be involved by molecular dynamics simulation (chapter 5). Many of the residues in the putative binding site that were identified by mutagenesis (Chan et al., 2008) were not able to be assigned by NMR, while many of the residues that were identified by chemical shift perturbation were not selected as candidates for mutagenesis studies. In contrast, there were five residues that did not produce large chemical shifts that were identified by mutagenesis, these were residues 72, 126, 127, 150, and 172, which are either in the binding “hot pocket,” or adjacent to it. This may be due to the nature of the chaperone peptide interaction not changing the chemical environment of those backbone amides sufficiently to produce a chemical shift change, or perhaps due to effects of the mutagenesis that alter the binding pocket without making direct contact with the leader peptide.

This limitation could be addressed experimentally by continuing the mutagenesis experiments to include the NMR data, though a high-resolution structure of the DmsA<sub>L</sub>/DmsD complex through x-ray crystallography or an NMR solution structure would be the most informative means of obtaining the molecular details of this interaction. Additional methods for the continued analysis of this interaction are proposed in the future directions section below.

By combining all of the available data, a new model for DmsD/DmsA<sub>L</sub> peptide interaction was constructed (Figure 5.6). To generate this model the DmsA<sub>L</sub> peptide was positioned with the twin arginine motif in close proximity to the hot pocket, as the docking discussed in chapter 5 had placed it, with the hydrophobic region of the DmsA<sub>L</sub> peptide placed along the hinge loop on the hydrophobic patch on the surface of DmsD. This model was solvated and simulated for 20 ns, which allowed the surface of the DmsD molecule to accommodate the presence of the peptide. In the final model the peptide is positioned such that the N-terminus rests “behind” the protein, interacting with a small
patch of residues adjacent to the “hot pocket.” The RR motif of the leader peptide makes contact with the hot pocket as was described in chapter 5, while the hydrophobic region of the peptide makes contact with the groove defined between helices 1, 2 & 8. There are several contacts that are made between the residues of the hinge region, as was noted based on the NMR chemical shift perturbation data described in detail in chapter 4. The proposed path for the DmsA_L peptide accounts for the polar and non-polar contacts between the peptide and DmsD and makes contact with the regions that were identified by both the NMR and biochemically derived data.

6.3. Future directions

Given the wealth of data that is available regarding this interaction, the next logical step is to test the proposed new model for leader peptide binding in which the RR motif of the peptide makes contact with the “hot pocket” region, and the hydrophobic region lies along a groove formed by the hinge loop and helices 1 & 2. The most comprehensive methods that will provide that information are to solve the structure of the DmsA_L/DmsD complex through either X-ray crystallography or NMR.

There are other means of examining this interaction that have not yet been explored. One such method is hydrogen-deuterium exchange combined with mass spectrometry. By creating a deuterium labeled sample of the complex and permitting the exchange of amide deuterons with protons, a detectible shift in the mass of the protein will be observed when analyzed by mass spectrometry, as the buried amides will not be available for exchange, and remain deuterated (Zhang and Smith, 1993). This has the advantage of identifying a series of residues on both the leader peptide and the REMP that are involved in binding. There are some technical limitations to this technique, the protein must be digested by a protease at low pH to prevent the back-exchange of protons and deuterons and each mass spectrometry peak must be unambiguously assigned based on calculated expected masses for each of the digested peptides in variable states of deuteration. The complexity of the task may limit the information that can be collected, though it has the potential be very informative.
Hydrogen-deuterium exchange can also be used in concert with NMR, as the $^{15}$N-HSQC will contain weakened or broadened signals that correspond to any amide nitrogens that are deuterated depending on the rate of exchange, while those that are protonated will appear unaltered in the spectrum (Craven et al., 2000). Comparison of these spectra, combined with the assignments provided in chapter 4, will produce a list of residues that are not solvent accessible. When this list is compared with a similar experiment from un-complexed DmsD, it will allow the identification of residues involved in the interaction. This will only be useful for those residues that have assigned peaks on the 2 dimensional $^{15}$N-HSQC spectrum and should produce data similar to the chemical shift perturbation data that was presented in chapter 4, but will be more complete in that only the residues that are buried will be reported rather than all of those whose chemical environment had changed after forming the DmsA$_L$/DmsD complex. These methods are not as valuable as a high resolution structure of the complex, but when combined with the existing data, it will produce a more complete model of this interaction.

The biochemical role of REMP dimerization is another interesting aspect of this project as its biological role is still largely unknown. The DmsD dimer is capable of binding to the DmsA$_L$ peptide (Sarfo et al., 2004), however, during the purification of the DmsA$_L$/DmsD complex the dimeric species of DmsD was not observed. This is in contrast to the variable amount of dimeric DmsD that is observed when DmsD is expressed and purified alone. This suggests that the formation of the complex in some way prevents the formation of the dimer. A prospect that would be interesting to explore through SEC or MALS based binding assays in which the quantity of dimeric species is compared before and after the introduction of the leader peptide species.

There are a number of other interesting prospects for the study of the REMP family of proteins such as the mechanism for releasing the substrates once the proofreading has requirement is satisfied. This is a particularly interesting problem, given the strength of the REMP-substrate interaction at approximately 0.2 µM (Winstone et al., 2006). It is also possible that interaction of the REMP with the TatBC complex facilitates the release of the substrate (Papish et al., 2003); this could be tested with binding assays that are evaluated by size exclusion chromatography, or through the generation and characterization of mutants that are unable to release the substrates.
The importance of the TAT and related chaperone proteins as a fundamental part of protein translocation and their role in the biotechnological production of difficult to fold proteins, or as a target for antibiotics, continues to make them an attractive subject for study.
References


Rambaut, A. (2007). FigTree.


Appendices
# Appendix A.
## Cloning details

<table>
<thead>
<tr>
<th>Construct</th>
<th>Residues</th>
<th>Vector</th>
<th>Primers Forward (5’ to 3’)</th>
<th>Primers Reverse (5’ to 3’)</th>
<th>Restriction Enzymes</th>
<th>UniProt ID</th>
<th>pI**</th>
<th>Extinction Coefficient (M⁻¹·cm⁻¹)**</th>
</tr>
</thead>
<tbody>
<tr>
<td>H&lt;sub&gt;6&lt;/sub&gt;-T7-DmsD</td>
<td>1-204</td>
<td>pRSETa</td>
<td>Gift from Dr. R. J. Turner</td>
<td>BclI / EcoRI</td>
<td>P69853</td>
<td>4.98</td>
<td>65095</td>
<td></td>
</tr>
<tr>
<td>H&lt;sub&gt;6&lt;/sub&gt;-DmsD-28a</td>
<td>1-204</td>
<td>pET28a</td>
<td>ACTGCATA TGACCCAT TTTTCACA GCAAGATA ATTTTTCTG</td>
<td>AGCTCTCG AGCTATCG AACAGGG GTTTAACC GGG</td>
<td>Ndel / Xhol</td>
<td>P69853</td>
<td>5.69</td>
<td>64970</td>
</tr>
<tr>
<td>H&lt;sub&gt;6&lt;/sub&gt;-DmsD-15b</td>
<td>1-204</td>
<td>pET15b</td>
<td>ACTG CATATGAC CCATTTCAC ACAGCAGA ATAATTTTCTG</td>
<td>AGCTCTCG AGCTATCG AACAGGG GTTTAACC GGG</td>
<td>Ndel / Xhol</td>
<td>P69853</td>
<td>4.91</td>
<td>64970</td>
</tr>
<tr>
<td>DmsD-15b</td>
<td>1-204</td>
<td>pET15b</td>
<td>GAGGGAGA TACATGAC GGCAGCA GCATCTG</td>
<td>GATGGCTG CTGCCCTT GGTATATC TCCTTC</td>
<td>SDM*</td>
<td>P69853</td>
<td>4.62</td>
<td>58105</td>
</tr>
<tr>
<td>DmsA&lt;sub&gt;L&lt;/sub&gt;-GST</td>
<td>1-43</td>
<td>pGEX-5T</td>
<td>Gift from Dr. R. J. Turner</td>
<td>BamHI / EcoRI</td>
<td>P18775</td>
<td>7.73</td>
<td>29910</td>
<td></td>
</tr>
<tr>
<td>DmsA&lt;sub&gt;L&lt;/sub&gt;-H&lt;sub&gt;6&lt;/sub&gt;-24a</td>
<td>1-44</td>
<td>pET24a</td>
<td>ACTGCATA TGATGAAA AGAAAT CTCTGATG CGGTATTG G</td>
<td>AGCTCTCG AGCTGTGCC GCAGGCG ACCAGGGG CGCAATCC GACATAAA GTGAATTG</td>
<td>Ndel / Xhol</td>
<td>P18775</td>
<td>10.27</td>
<td>Not Visible by UV</td>
</tr>
<tr>
<td>H&lt;sub&gt;6&lt;/sub&gt;-DmsA&lt;sub&gt;L&lt;/sub&gt;-DmsD</td>
<td>1-44 (DmsA) - 1-204 (DmsD)</td>
<td>pET28a</td>
<td>ACTGGAAT CGCGGGCC CAAATCCGA CTAAMCGT TAAATG &amp; AGCTGAAT TCGGGGCG ACACAGGC TAAAAGGT AATG</td>
<td>AGCTGGAAT CGGGGGCC CAAATCCGA CTAAMCGT TAAATG &amp; AGCTCTCG AGCTATCG AACAGGGG GTTTAACC GGG</td>
<td>Ndel / EcoRI</td>
<td>P18775</td>
<td>6.05</td>
<td>65095</td>
</tr>
</tbody>
</table>

*Site directed mutagenesis product, no restriction enzymes required.

**pI and Extinction coefficient were calculated with ExPasy Tools ProtParam.
Appendix B.

Statistics in crystallography

B-Factor
A measurement of vibration or oscillation about a position specified in molecular coordinates, an increase in B-factor is indicative of increased disorder. The B-factor is a function of the magnitude of vibration according to the following equation:

\[ B_i = 8\pi^2 U_j \]

Where \( \{u_j^2\} \) is the mean-square displacement of the atom \( j \) from its rest position.

Completeness
The number of reflections measured, as a percentage of the total expected number of reflections at a given resolution.

Figure of Merit (FOM)
Describes the weight given to structure factor magnitudes based on the phase probability distribution. When the phase angle is perfect, FOM =1, and trends toward 0 as error increases. It is calculated as:

\[
\left( \frac{\sum P(\alpha) e^{i\alpha}}{\sum P(\alpha)} \right) \left( \frac{\sum P(\alpha) e^{i\alpha}}{\sum P(\alpha)} \right)
\]

such that \( \alpha \) is the phase angle, and \( P(\alpha) \) is the phase probability distribution.

Mean I/\( \sigma \)I
The average intensity of reflections over the standard deviation of the reflections. This describes the average signal to noise ratio of the data.

Multiplicity (Redundancy)
The average number of times each reflection is measured in a dataset. It is calculated as the number of measured reflections over the number of unique reflections.

\( R_{\text{merge}} \)
A data quality indicator that represents the agreement between multiple measurements of the same reflection (not symmetry related reflections). The measurements are taken from different frames or different datasets.

Calculated as:

\[
R_{\text{merge}} = \frac{\sum_i \sum_h |I_i - \langle I_i \rangle|}{\sum_h \sum_i I_i}
\]
Such that: $I_i$ denotes the $i$th intensity measurement of reflection ‘h’ and $<I>$ denotes the average intensity for the measurements of the reflection.

$R_{\text{symm}}$

A data quality indicator that represents the agreement between symmetry related reflections. These values should be identical.

Calculated as:

$$R_{\text{symm}} = \frac{\sum_h (I_h - \bar{I}_h)}{\sum_h I_h}$$

Such that $I$ and $\bar{I}$ represent intensities of symmetry-related reflections.

$R$-factor

A measure of fit between the model and the diffraction data. The expected diffraction data is calculated from the model and compared to the original experimental data.

Calculated as:

$$R = \frac{\sum_h |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum_h |F_{\text{obs}}|}$$

such that $F_{\text{obs}}$ is an observed structure factor and $F_{\text{calc}}$ is a calculated structure factor.

$R_{\text{free}}$

$R_{\text{free}}$ is calculated as $R$, but uses a 5% subset of randomly selected reflections that are set aside from the beginning and not used in the refinement of the structural model.

RMSD from ideal values

The ideal values for bond length and bond angle are compared with those from the model, and the root mean squared deviation from those values is calculated. The RMSD for the bond lengths should be below 0.02 Å for bond lengths, and below 4 degrees for bond angles.
Appendix C.
Size-exclusion chromatography standard curves

Sephacryl S-100

Standard curve of Sephacryl S-100

<table>
<thead>
<tr>
<th>Log(MW)</th>
<th>Kav</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

Column: HiPrep 26/60 Sephacryl S-100 High Resolution Column attached to the AKTAprime system
Buffer: 20 mM Tris-HCl (pH 8.0), 100 mM NaCl (TBS)
Temperature: 4°C
Flow Rate: 1 mL/min
Standards: From Amersham Biosciences’ LMW Calibration Kit (product ID: 17-0442-01)
Void Volume: 98 mL
Column Volume: 320 mL

Protein Standards for Sephacryl S-100.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Molecular Mass (kDa)</th>
<th>Elution Volume on S-100 (mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue Dextran 2000</td>
<td>2000</td>
<td>98*</td>
</tr>
<tr>
<td>Albumin</td>
<td>67.0</td>
<td>118</td>
</tr>
<tr>
<td>Ovalbumin</td>
<td>43.0</td>
<td>133</td>
</tr>
<tr>
<td>Chymotrypsinogen A</td>
<td>25.0</td>
<td>168</td>
</tr>
</tbody>
</table>
* The elution volume of Blue Dextran was used to determine the void volume ($V_0$ = the volume of the mobile phase in the column) as its size is much larger than what this column can retain.

**Concentration:** 5 mg/mL (400 µL final volume) were used except for Blue Dextran for which was at 1 mg/mL (1 mL final volume).

Three runs were performed. Blue Dextran and Chymotrypsinogen A were run independently. Chymotrypsinogen A was run separately to prevent cleavage of the other samples.

The standard curve for the Sephacryl S-100 column presented here was generated by Suraaj Aulakh.

### Superdex 200

| Ribonuclease A | 13.7 | 189 |

### Standard curve of Superdex 200

Column: Superdex 200 in line with Agilent multi-angle light scattering.  
Buffer: 20 mM Tris-HCl (pH 8.0), 100 mM NaCl (TBS)  
Temperature: 21°C  
Flow Rate: 0.5 mL/min  
Standards Used: BamE monomer, BamE Dimer, BamC and the BamCD complex from *E. coli.*
## Protein Standards for Sephacryl S-100.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Molecular Mass (kDa)</th>
<th>Elution Volume on S-100 (mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BamE Monomer</td>
<td>12</td>
<td>17.5</td>
</tr>
<tr>
<td>BamE Dimer</td>
<td>24</td>
<td>118</td>
</tr>
<tr>
<td>BamC</td>
<td>34</td>
<td>14.15</td>
</tr>
<tr>
<td>BamCD Dimer</td>
<td>53</td>
<td>12.38</td>
</tr>
</tbody>
</table>

**Concentration:** 100 µL final volume at 5 mg /mL.

**Column Volume:** 24 mL

**Void Volume:** 8 mL

### Molecular mass calculation based on the elution volume

Kav values were calculated as follows:

\[
K_{av} = \frac{V_e - V_o}{V_t - V_o} \quad \text{Equation #1}
\]

Where \( V_o \) is the void volume and \( V_t \) is the column volume and \( V_e \) is the elution volume.

The resulting Kav vs. Log(MW) plot produced the equation Log(MW) = M(Kav) + B, which can be rewritten as:

\[
MW = 10^{(M \times Kav + B)} \quad \text{Equation #2}
\]

For each protein sample examined by size exclusion chromatography, the Kav value was calculated based on the elution volume with Equation #1. The measured molecular mass was calculated with Equation #2.
Appendix D.
Alignment of protein sequences for *E. coli* REMPs.

**Figure D 1. Sequence Alignment for all of the TorD-like REMPs in *E. coli***

The sequences were aligned using clustal omega (Sievers et al., 2011). All of the proteins are found in *E. coli* and have the Uniprot accession numbers: P69853, P36662, P0AF26, P19317, P75915, P0A915, P0AAN1, P19931 and P13024 for DmsD, TorD, NarJ, NarW, YcdY respectively.
Figure D 2. Sequence alignment for all of the *E. coli* REMPs.

(Continued from previous page) The sequences were aligned using clustal omega (Sievers et al., 2011). All of the proteins are found in *E. coli* and have the Uniprot accession numbers: P69853, P36662, P0AF26, P19317 and P75915 for DmsD, TorD, NarJ, NarW, YcdY, NapD, HybE, HyaE and FdhE respectively.

![Sequence Alignment for all of the *E. coli* REMPs.](image)

Figure D 3. Sequence Alignment for all of the α/β REMPs in *E. coli*.

The sequences were aligned using clustal omega (Sievers et al., 2011). All of the proteins are found in *E. coli* and have the Uniprot accession numbers: P0A9I5, P0AAN1, P19931 and P13024 for NapD, HybE, HyaE and FdhE respectively.